



# XXI



**НАЦИОНАЛЬНАЯ  
КОНФЕРЕНЦИЯ  
ПО ИСКУССТВЕННОМУ  
ИНТЕЛЛЕКТУ  
С МЕЖДУНАРОДНЫМ  
УЧАСТИЕМ**



## КИИ-2023



**ТРУДЫ КОНФЕРЕНЦИИ**

**Том 2**

16-20 октября 2023 г.  
Смоленск

# XXI

Российская ассоциация  
искусственного интеллекта

Федеральный  
исследовательский центр  
«Информатика  
и управление» РАН

Филиал Национального  
исследовательского  
университета «МЭИ»  
в г. Смоленске

## НАЦИОНАЛЬНАЯ КОНФЕРЕНЦИЯ ПО ИСКУССТВЕННОМУ ИНТЕЛЛЕКТУ С МЕЖДУНАРОДНЫМ УЧАСТИЕМ

### КИИ-2023

ТРУДЫ КОНФЕРЕНЦИИ

Том 2

16-20 октября 2023 г.  
Смоленск,  
филиал НИУ «МЭИ» в г. Смоленске

Смоленск  
Издательство «Принт-Экспресс»  
2023

УДК 004.8+004.89+004.82+004.032.26(045)+004.9  
ББК 32.813  
Д 22

Организаторы конференции:

*Российская ассоциация искусственного интеллекта  
Федеральный исследовательский центр «Информатика и управление» РАН  
Филиал Национального исследовательского университета «МЭИ»  
в г. Смоленске*

*При поддержке Администрации Смоленской области*

Д 22 Двадцать первая Национальная конференция по искусственному интеллекту с международным участием, КИИ-2023 (Смоленск, 16-20 октября 2023 г.). Труды конференции. В 2-х томах. Т.2. – Смоленск: Принт-Экспресс, 2023. – 381 с.

**ISBN 978-5-91812-231-0**

Двадцать первая Национальная конференция по искусственному интеллекту с международным участием КИИ-2023 продолжает традицию советских (российских) конференций, организуемых Российской ассоциацией искусственного интеллекта.

Во втором томе трудов публикуются доклады участников конференции, представленные на следующих секциях:

Секция 5 «Нечеткие модели, мягкие измерения и вычисления»,

Секция 6 «Инструментальные системы, фреймворки для создания интеллектуальных систем. Технологии проектирования и разработки интеллектуальных систем. Прикладные интеллектуальные системы»,

Секция 7 «Прикладные биомедицинские системы».

**ББК 32.813**

Рецензенты: академик РАН, ИПУ РАН *С.Н. Васильев*,  
д.т.н., доц., ФИЦ ИУ РАН *В.Н. Захаров*

ISBN 978-5-91812-231-0

© Авторы, 2023

© Российская ассоциация искусственного интеллекта, 2023

© Издательство «Принт-Экспресс», 2023

## ПРОГРАММНЫЙ КОМИТЕТ КОНФЕРЕНЦИИ

### *Сопредседатели Программного комитета*

Грибова В.В., член-корр. РАН, ИАПУ ДВО РАН, Владивосток  
Кобринский Б.А., д.м.н., проф., ФИЦ ИУ РАН, Москва

### *Заместители председателя Программного комитета*

Забейайло М.И., д.ф.-м.н., проф., ФИЦ ИУ РАН, Москва  
Тельнов Ю.Ф., д.э.н., проф., РЭУ им. Г.В. Плеханова, Москва

### *Ответственный секретарь Программного комитета*

Колоденкова А.Е., д.т.н., доц., СамГТУ, Самара

### *Члены программного комитета*

Аверкин А.Н., к.ф.-м.н., доц., ФИЦ ИУ РАН, Москва  
Афанасьева Т.В., д.т.н., РЭУ им. Г.В. Плеханова, Москва  
Болодурина И.П., д.т.н., проф, Оренбургский ГУ, Оренбург  
Боргест Н.М., к.т.н., доц., Самарский НИУ им. акад. С.П. Королева  
Борисов В.В., д.т.н., проф., филиал НИУ «МЭИ», Смоленск  
Васильев С.Н., акад. РАН, ИПУ РАН, Москва  
Визильтер Ю.В., д.ф.-м.н., проф., ГосНИИ АС, Москва  
Гаврилова Т.А., д.т.н., проф., СПбГУ, Санкт-Петербург  
Городецкий В.И., д.т.н., проф., АО «Эврика», Санкт-Петербург  
Дли М.И., д.т.н., проф. филиал НИУ «МЭИ», Смоленск  
Еремеев А.П., д.т.н., проф., НИУ «МЭИ», Москва  
Желтов С.Ю., акад. РАН, ГосНИИАС, Москва  
Загорулько Ю.А., к.т.н., ИСИ СО РАН, Новосибирск  
Каляев И.А., акад. РАН, ЮФУ, Ростов-на-Дону  
Ковалев С.М., д.т.н., проф., РГУПС, Ростов-на-Дону  
Колесников А.В., д.т.н., проф., БФУ им. И. Канта, Калининград  
Козенко И.В., д.т.н., проф., СПИИРАН, Санкт-Петербург  
Кузнецов С.О., д.ф.-м.н., проф., НИУ ВШЭ, Москва  
Курейчик В.В., д.т.н., проф., ЮФУ, Таганрог  
Лебедев О.Б., д.т.н., доц., ИКТИБ ЮФУ, Таганрог  
Лукашевич Н.В., д.т.н., проф., МГУ им. М.В. Ломоносова, Москва  
Мисник А.Е., к.т.н., доц., БРУ, Республика Беларусь, Могилев  
Михеенкова М.А., д.т.н., проф., ФИЦ ИУ РАН, Москва  
Мошкин В.С., к.т.н., доц., УлГТУ, Ульяновск  
Пальчунов Д.Е., д.ф.-м.н., доц., ИМ СО РАН, Новосибирск  
Палюх Б.В., д.т.н., проф., ТвГТУ, Тверь  
Панов А.И., к.ф.-м.н., доц., ФИЦ ИУ РАН, Москва  
Подвесовский А.Г., к.т.н., доц., БГТУ, Брянск  
Редько В.Г., д.ф.-м.н., проф., НИИСИ РАН, Москва  
Ройзензон Г.В., к.т.н., доц., ФИЦ ИУ РАН, Москва

Рыбина Г.В., д.т.н., проф., НИЯУ МИФИ, Москва  
Смирнов И.В., к.ф.-м.н., доц., ФИЦ ИУ РАН, Москва  
Соколов И.А., акад. РАН, ФИЦ ИУ РАН, Москва  
Стефанюк В.Л., д.т.н., проф., ИППИ РАН, Москва  
Сулейманов Д.Ш., акад. АН РТ, ИПС АН РТ, Казань  
Федунов Б.Е., д.т.н., проф., ГосНИИ АС, Москва  
Финн В.К., д.т.н., проф., ФИЦ ИУ РАН, Москва  
Хачумов В.М., д.т.н., проф., ФИЦ ИУ РАН, Москва  
Яковлев К.С., к.ф.-м.н., ФИЦ ИУ РАН, Москва  
Ярушкина Н.Г., д.т.н., проф., УЛГТУ, Ульяновск

### **ОРГАНИЗАЦИОННЫЙ КОМИТЕТ КОНФЕРЕНЦИИ**

*Председатель Организационного комитета*

Борисов В.В., д.т.н., проф., филиал НИУ «МЭИ», Смоленск

*Заместитель председателя Организационного комитета*

Федулов А.С., д.т.н., проф., филиал НИУ «МЭИ», Смоленск

*Члены организационного комитета*

Гладков Л.А., к.т.н., доц., ИКТИБ ЮФУ, Таганрог

Кириллова Е.А., к.э.н., доц., филиал НИУ «МЭИ», Смоленск

Королева М.Н., к.т.н., доц., МГТУ им. Н.Э. Баумана, Москва

Луферов В.С., к.т.н., ООО «Сбер бизнес софт», Смоленск

Рысина Е.И., к.т.н., филиал НИУ «МЭИ», Смоленск

Синяевский Ю.В., к.т.н., доц., филиал НИУ «МЭИ», Смоленск

Солопов Р.В., к.т.н., доц., филиал НИУ «МЭИ», Смоленск

Тютюнник А.А., к.э.н., доц., филиал НИУ «МЭИ», Смоленск

### **ИНФОРМАЦИОННАЯ ПОДДЕРЖКА КОНФЕРЕНЦИИ**

ООО «Лаборатория информационных технологий», Смоленск

## ПРЕДИСЛОВИЕ

Двадцать первая Национальная конференция по искусственному интеллекту с международным участием КИИ-2023 продолжает традицию советских (российских) конференций, организуемых Российской ассоциацией искусственного интеллекта (РАИИ).

Федеральный проект «Искусственный интеллект» направлен на решение междисциплинарных проблем, сформулированных в Национальной стратегии развития искусственного интеллекта, включая как исследования в области фундаментальной и прикладной науки, так и в образовательной сфере. Эти проблемы постоянно находятся в фокусе внимания членов РАИИ и нашли отражение в докладах конференции, организаторами которой являются Российская ассоциация искусственного интеллекта, Федеральный исследовательский центр «Информатика и управление» РАН, филиал Национального исследовательского университета «МЭИ» в г. Смоленске.

Тематика конференции охватывает следующие основные направления искусственного интеллекта: инженерия знаний; интеллектуальный анализ данных; интеллектуальные агенты, роботы, интеллектуальное управление; компьютерное зрение; машинное обучение, нейросетевые методы; нечеткие модели, мягкие измерения и вычисления; фреймворки для создания интеллектуальных систем; технологии проектирования и разработки интеллектуальных систем, прикладные интеллектуальные системы, включая биомедицинские системы.

Представлены пленарные доклады видных ученых и специалистов в области искусственного интеллекта. Из поданных на конференцию 90 секционных докладов Программным комитетом после рецензирования отобраны 64 доклада.

В конференции принимают участие более 160 ученых, исследователей и специалистов из академических институтов, университетов, отраслевых институтов и организаций Москвы, Санкт-Петербурга, Белгорода, Брянска, Владивостока, Владимира, Волгограда, Донецка, Иркутска, Казани, Калининграда, Красноярска, Новосибирска, Ростова-на-Дону, Самары, Смоленска, Таганрога, Твери, Ульяновска, Ханты-Мансийска, Челябинска, Могилёва (Республика Беларусь).

*Программный комитет КИИ-2023*

---

Секция 5

**НЕЧЕТКИЕ МОДЕЛИ, МЯГКИЕ  
ИЗМЕРЕНИЯ И ВЫЧИСЛЕНИЯ**

**УДК 004.415.2**

**ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ НЕЙРО-НЕЧЁТКИХ  
СЕТЕЙ ДЛЯ ЗАДАЧИ КЛАССИФИКАЦИИ  
ОФТАЛЬМОЛОГИЧЕСКИХ ИЗОБРАЖЕНИЙ**

А.Н. Аверкин (*averkin2003@inbox.ru*)  
Федеральный исследовательский центр  
«Информатика и управление» РАН, Москва

Е.Н. Волков (*envolkoff@gmail.com*)  
Государственный университет «Дубна», Дубна

С.А. Ярушев (*Yarushev.sa@rea.ru*)  
Российский экономический университет им. Г.В. Плеханова,  
Москва

Применение технологий искусственного интеллекта в задачах анализа изображений стремительно развивается с каждым годом благодаря чему возможности использования искусственных нейронных сетей (ИНС) в диагностике различных заболеваний вышли на совершенно новый уровень. Результаты тысяч исследований с применением ИНС в офтальмологии показывают результаты точности постановки диагноза сопоставимые, а иногда и превосходящие таковые при работе исполнителя-человека. Однако, прозрачность получаемых результатов и адаптивность работы таких систем остаются низкими. Использование свёрточных нейро-нечётких сетей (СННС) в задачах анализа изображений, в перспективе, может улучшить результаты. В работе рассмотрен имеющийся опыт при-

менения нейро-нечётких сетей для задач классификации медицинских изображений, продемонстрированы возможности практической реализации нейро-нечёткой сети для задач классификации офтальмологических изображений глаукомы.

**Ключевые слова:** искусственный интеллект, объяснительный искусственный интеллект, нечёткая логика, нейро-нечеткие сети, здравоохранение, медицина, персонализированная медицина, офтальмология, офтальмоскопия, глаукома, диагностика, ХАГ.

## 1. Глубокие нейросети и объяснительный интеллект

Один из основных принципов развития и использования технологий искусственного интеллекта (ИИ), приведенных в Национальной стратегии развития искусственного интеллекта на период до 2030 года является прозрачность: объяснимость работы искусственного интеллекта и процесса достижения им результатов, недискриминационный доступ пользователей продуктов, которые созданы с использованием технологий искусственного интеллекта, к информации о применяемых в этих продуктах алгоритмах работы искусственного интеллекта.

Каждое десятилетие в технологиях происходят революционные сдвиги, которые становятся новыми платформами, на которых строятся новые прикладные технологии. Так, искусственный интеллект перешел от экспертного обучения 1-го поколения и экспертных систем к системам искусственного интеллекта 2-го поколения на основе нейросетей глубокого обучению, требующих больших обучающих выборок. Теперь мы вступаем в 3-е поколение ИИ, в котором система искусственного интеллекта сможет интерпретировать и объяснить алгоритм принятия решений, даже если он имеет природу черного ящика. Объяснимый искусственный интеллект являются основной частью 3-го поколения ИИ. В 2030-х годах мы увидим ИИ 4-го поколения с интеллектуальными системами, которые будут сами обучаться учиться и динамически накапливать новые знания и навыки. К 2040-м годам ИИ 5-го поколения создаст системы искусственного интеллекта с воображением, которые больше не будут полагаться на людей в обучении.

Объяснимый ИИ устроен таким образом, что наблюдающий со стороны пользователь может понять, почему именно алгоритм принял то или иное решение. Такие методы приходят на смену принципу «чёрного ящика», в котором даже сами создатели ИИ не всегда в состоянии объяснить принципы его работы. Объяснительный интеллект - важная часть общего тренда, который Gartner называют «Trust in Algorithm», то есть «доверие алгоритму». Чем больше ИИ заменяет человеческие решения, тем больше он усиливает положительные и отрицательные последствия таких реше-



ний. Оставленные без контроля подходы на основе ИИ могут увековечить предвзятость, ведущую к проблемам, потере производительности и доходов. Двигаясь вперед, организации должны разрабатывать и эксплуатировать системы искусственного интеллекта на принципах справедливости и прозрачности и заботиться о безопасности, конфиденциальности и общест- ве в целом.

Большинство существующих моделей объяснимого ИИ связаны с начавшейся в 2018 году программой Управления перспективных исследовательских проектов Министерства обороны США (DARPA), которая в настоящее время инвестирует более 2 миллиардов долларов в создание искусственного интеллекта третьего поколения, которую DARPA назвало контекстуальной адаптацией, требующей не только прогресса в представлении знаний и рассуждений, машинном обучении, технологиях человеческого языка, зрения и робототехники, но также тесной интеграции каждого компонента для реализации надежных и прозрачных интеллектуальных систем, способных работать автономно или в команде с людьми и уметь объяснить людям свои решения. Программа DARPA по созданию систем объяснимого ИИ объединяет 11 групп университетов США и стремится создать такие системы ИИ, чьи модели обучения и решения могут быть понятны и должным образом проверены конечными пользователями. Команды разработчиков объяснимого ИИ решают эту проблему путем создания и развития технологий объяснимого машинного обучения, разрабатывая принципы, стратегии и методы взаимодействия человека и компьютера для получения эффективных объяснений.

В России также уделяется большое внимание направлению объяснимого ИИ. Так, Нижегородский государственный университет в 2020 г. стал победителем в конкурсе крупных научных проектов от Минобрнауки РФ с проектом «Надежный и логически прозрачный искусственный интеллект: технология, верификация и применение при социально-значимых и инфекционных заболеваниях». Главным результатом проекта стала разработка новых методов и технологий, позволяющих преодолеть два основных барьера систем машинного обучения и искусственного интеллекта: проблему ошибок и проблему явного объяснения решений. В 2022 году РЭУ им. Г.В. Плеханова выиграл проект РНФ «Гибридные модели поддержки принятия решений на основе методов дополненного искусственного интеллекта, когнитивного моделирования и нечеткой логики в задачах персонализированной медицины», посвященный созданию медицинских систем нового поколения, основанных на применении методов объяснимого ИИ для нейросетей глубокого обучения в задачах медицинской диагностики.

Следует отметить роль Л. Заде, в создании Z Advanced Computing, Inc. (ZAC). ZAC является пионером технологий когнитивного объяснимого ИИ (искусственного интеллекта) (Cognitive XAI), например, для детального распознавания сложных 3D-изображений / объектов под любым углом обзора.

ZAC продемонстрировала крупные достижения в области искусственного интеллекта и машинного обучения (ML) в недавних проектах для ВВС США (USAF) и для Bosch / BSH (крупнейшего производителя бытовой техники в Европе): ZAC добился детального сложного распознавания 3D-изображений, используя всего несколько обучающих образцов и используя только средний ноутбук с процессором с низким энергопотреблением, как для обучения, так и для распознавания. Это резко контрастирует с другими алгоритмами в отрасли (такими как Глубокие сверточные нейронные сети (CNN) или ResNets), которые требуют от тысячи до миллиардов обучающих выборок, обучаемых на больших серверах GPU.

В настоящее время XAI является выдающейся и плодотворной исследовательской областью, где многие результаты Л. Заде могут стать решающими, если они будут тщательно рассмотрены и разработаны. Стоит отметить, что около 30% публикаций в Scopus, связанных с XAI, датированных 2017 годом или ранее, принадлежат авторам, хорошо признанным в области нечеткой логики. В основном это связано с приверженностью нечеткого сообщества созданию интерпретируемых нечетких систем, поскольку интерпретируемость глубоко укоренена в основах нечеткой логики.

Главной составляющей второй и третьей волны ИИ являются технологии глубокого обучения [Gunning et al., 2019].

Множество методов глубокого обучения можно разделить на методы обучения с учителем и методы обучения без учителя. Методы неразрывно связаны с некоторым набором моделей глубоких нейросетей специального вида.

Первая группа методов требует для определения параметров сети наличие размеченных тренировочных наборов данных больших объемов. Как правило, размер наборов для обучения глубоких нейронных сетей на порядки превышает размеры тренировочных баз, используемых для обучения традиционных моделей машинного обучения. К моделям, которые обучаются с учителем, относятся многослойные полностью связанные нейронные сети, сверточные нейронные сети, рекуррентные нейронные сети. Эти модели различаются способами связи нейронов, находящихся на соседних слоях сети, и количеством нейронов предыдущего слоя, влияющих на состояние отдельных нейронов текущего слоя. При этом для начальной инициализации весов этих моделей – предварительного обучения – могут использоваться модели второй группы.

Вторая группа моделей не требуют наличия размеченных тренировочных наборов данных для настройки параметров сети. К данной группе моделей относятся различные виды автокодировщиков, ограниченные машины Больцмана, разверточные нейронные сети, генеративно-состязательные сети. Указанные модели в процессе обучения тем или иным образом пытаются восстановить исходный сигнал, пришедший на вход сети.

Глубокие нейронные сети (ГНС) добились значительного прогресса в приложениях, связанных с большими и сложными наборами данных, но при этом подвергаются критике как "черный ящик". Этот недостаток в последнее время стал побуждать исследовательское сообщество к реализации идей гибридных подходов,

В результате появились новые гибридные системы, классифицируемые как глубокие нейро-нечеткие системы (ГННС). Исследования, посвященные реализации ГННС, стремительно растут в таких областях, как вычислительная техника, здравоохранении, транспорте и финансах с высокой интерпретируемостью и приемлемой точностью.

Однако в литературе найдено сравнительно мало исследований, дающих полное представление об этой области. Поэтому целью данного исследования является проведение систематического обзора для оценки текущего прогресса, тенденций, возникающих проблем, пробелов в исследованиях, задач и перспектив, связанных с исследованиями в области ГННС.

## **2. Нечёткая логика и нейро-нечёткие сети**

Нейро-нечёткие сети – это системы из области искусственного интеллекта, которые были предложены Ж.С.Р. Чангом и комбинируют методы искусственных нейронных сетей и систем на основе нечёткой логики. Нейро-нечёткие сети являются результатом попытки создания гибридной интеллектуальной системы, которая бы давала синергетический эффект этих двух подходов путём комбинирования человекоподобного стиля рассуждений нечётких систем с обучением и коннекционистской структурой нейронных сетей. Основная сила нейро-нечётких систем состоит в том, что они являются универсальными аппроксиматорами со способностью порождать интерпретируемые правила ЕСЛИ-ТО.

Преимущество нейро-нечётких систем включает в себя две противоречивые необходимости нечёткого моделирования: интерпретируемость (объяснимость) и точность. На практике, одно из них всегда преобладает. Нас в настоящей работе больше интересуют вопросы объяснимости нейросетей.

Наиболее интересным направлением в области разработки объяснительных моделей ИИ является извлечение правил с использованием нейронечетких моделей. Системы, основанные на нечетких правилах (FRBS), разработанные с помощью нечеткой логики, стали полем активных исследований за последние несколько лет. Эти алгоритмы доказали свои сильные стороны в таких задачах, как управление сложными системами, создание нечетких элементов управления. Взаимоотношения между обоими подходами (ANN и FRBS) были тщательно изучены и показана их эквивалентность. Это позволяет сделать два важных вывода. Во-первых, можно применить то, что было обнаружено для одной из моделей, к другой. Во-вторых, мы можем перевести знания, встроенные в нейронную сеть, на более когнитивно-приемлемый язык – нечеткие правила. Другими словами, получаем семантическую интерпретацию нейронных сетей [Averkin et al., 2018].

Для того, чтобы получить семантическую интерпретацию черного ящика глубокого обучения, нейронечеткие сети могут быть использованы вместо последнего полносвязного слоя. Например, ANFIS (адаптивная нейронечеткая система) является многослойной сетью прямого распространения. Эта архитектура имеет пять слоев, таких как нечеткий слой, продукционный слой, слой нормализации, слой дефаззификации и выходной слой. ANFIS сочетает преимущества нейросети и нечеткой логики. Далее приведем классификацию наиболее известных нейронечетких подходов.

Рассматривая архитектуры нейронечетких моделей, можно выделить три методики объединения искусственных нейронных сетей (ИНС) и нечетких моделей [Jin, 2011]:

- неуго-FIS, в которых ИНС используется как инструмент в нечетких моделях;
- нечеткие ИНС, в которых классические модели ИНС фаззифицированы;
- нейронечеткие гибридные системы, в которых нечеткие системы и ИНС объединены в гибридные системы [Борисов и др., 2017].

Исходя из данных методик, нейронечеткие модели можно разделить на три класса [Rajab et al., 2018].

- Кооперативные нейронечеткие модели. В данном случае часть ИНС изначально используется для определения нечетких множеств и / или нечетких правил, где впоследствии выполняется только полученная нечеткая система. В процессе обучения определяются функции принадлежности, а также формируются нечеткие правила на основе обучающей выборки. Здесь основная задача нейронной сети заключается в подборе параметров нечеткой системы.

- Параллельные нейронечеткие модели. Нейронная сеть в данном типе модели работает параллельно с нечеткой системой, предоставляя входные данные в нечеткую систему или изменяя выходные данные нечеткой системы. Нейронная сеть может являться также и постпроцессором выходных данных из нечеткой системы.
- Гибридные нейронечеткие модели. Нечеткая система использует метод обучения, как это делает и ИНС, чтобы настроить свои параметры на основе обучающих данных. Среди представленных классов моделей наибольшей популярностью пользуются модели именно данного класса, доказательством тому служит их применение в широком спектре реальных задач [Shihabudheen et al., 2018].

Среди наиболее популярных гибридных моделей можно выделить следующие архитектуры:

- Сеть управления нечетким адаптивным обучением (FALCON) [Lin et al., 1991], которая имеет пятислойную архитектуру. На одну выходную переменную приходится по два лингвистических узла. Первый узел работает с обучающей выборкой (паттерном обучения), второй является входным для всей системы. Первый скрытый слой размечает входную выборку в соответствии с функциями принадлежности. Второй слой задает правила и их параметры. Обучение происходит на основе гибридного алгоритма без учителя для определения функции принадлежности, базы правил и использует алгоритм градиентного спуска для оптимизации и подбора итоговых параметров функции принадлежности.
- Адаптивная нейронечеткая система вывода ANFIS – это хорошо известная нейронечеткая модель, которая применялась во многих приложениях и исследовательских областях [Jang, 1993]. Более того, сравнение архитектур нейронечетких сетей показало, что ANFIS показывает минимальную ошибку в задаче прогнозирования. Основным недостатком модели ANFIS является то, что она предъявляет серьезные требования к вычислительной мощности.
- Система обобщенного приближенного интеллектуального управления на основе рассуждений (GARIC) [Berenji et al., 1992] представляет собой нейронечеткую систему, использующую два нейросетевых модуля, модуль выбора действия и модуль оценки состояния, который отвечает за оценку качества выбора действий предыдущим модулем. GARIC – пятислойная сеть прямого распространения.
- Нейронный нечеткий регулятор (NEFCON) [Nauck et al., 1999] был разработан для реализации системы нечеткого вывода типа Мамдани. Связи определяются с помощью нечетких правил. Входной слой является фаззификатором, а выходной решает задачу дефаз-

зификации. Обучается сеть на основе гибридного алгоритма обучения с подкреплением и алгоритма обратного распространения ошибки.

- Система нечеткого вывода и нейронной сети в программном обеспечении нечеткого вывода (FINEST) [Tano et al., 1996] представляет собой систему настройки параметров. Производится настройка нечетких предикатов, функции импликации и комбинаторной функции.
- Система для автоматического построения нейронной сети нечеткого вывода (SONFIN) [Juang et al., 1998] по своей сути аналогична NEFCON контроллеру, но вместо реализации нечеткого вывода типа Мамдани реализует вывод типа Такаги-Сугено. В данной сети входная выборка обрабатывается с помощью алгоритма выровненной кластеризации. При идентификации структуры части предварительного условия входное пространство разделяется гибким образом в соответствии с алгоритмом, основанным на выровненной кластеризации. Настройка параметров системы частично реализована на базе метода наименьших квадратов, предварительные условия настраиваются с помощью метода обратного распространения ошибки.
- Динамически развивающаяся нечеткая нейронная сети (dmEfuNN) и (EFuNN) [Kasabov et al., 1999]. В EFuNN все узлы создаются в процессе обучения. Первый слой передает обучающие данные на второй, который вычисляет степень соответствия с заранее определенной функцией принадлежности. Третий слой содержит в себе наборы нечетких правил, являющихся прототипами входных-выходных данных, которые можно представить в качестве гиперсфер нечеткого входного и выходного пространств. Четвертый слой рассчитывает степень, с которой выходная функция принадлежности разметила входные данные, а пятый слой производит дефаззификацию и подсчитывает числовые значения выходной переменной. DmEfuNN представляет собой модифицированную версию EFuNN. Основная идея состоит в том, что для всех входных векторов динамически подбирается набор правил, значения активации которых используются для расчета динамических параметров выходной функции. В то время как EFuNN реализует нечеткие правила типа Мамдани, dmEFuNN применяет тип Такаги-Сугено.

Извлечение правил является одним из способов помочь понять нейронные сети. Повышение прозрачности нейронных сетей путем извлечения из них правил имеет два основных преимущества. Это дает пользователю некоторое представление о том, как нейронная сеть использует

входные переменные, чтобы принять решение, и позволяет выявить скрытые функции в нейросетях, когда правила применяются для объяснения работы отдельных нейронов. Таким образом, в дальнейших исследованиях мы планируем заняться разработкой моделей извлечения правил из различных архитектур нейронных сетей путем создания гибридных нейро-нечетких глубоких моделей искусственных нейронных сетей.

Значительный вклад в изучение гибридных интеллектуальных систем в отечественной исследовательской практике, в том числе нейро-нечетких систем, внесли Борисов В.В. [Борисов, 2020], [Борисов и др., 2018] и Ярушкина Н.Г. [Батыршин и др., 2007].

### 3. Описание и архитектура СНС и СННС данного исследования

Представленная в исследовании свёрточная сеть-основа состоит из 5 свёрточных слоёв разных размерностей. Для создания СННС имеющаяся сеть была дополнена тремя слоями (Fuzzifier, Inference Layer, Defuzzifier). Согласно обзору разновидностей нейро-нечётких сетей [Talpur et al., 2023] созданная для исследования сеть относится к типу последовательных поскольку нечёткие слои присоединяются сразу за свёрточными. На рис. 1 представлено подробное схематичное описание архитектуры используемой СННС.

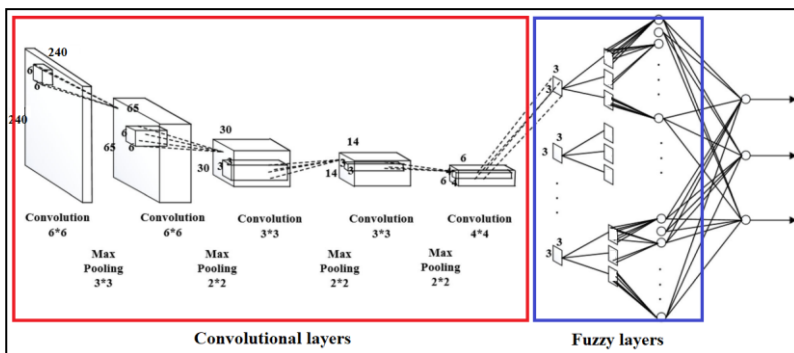


Рис. 1. Схематичное представление архитектуры СННС

В свою очередь, Борисов В.В. показывает особенности и общие принципы работы нейро-нечётких сетей в исследовании [Borisov et al., 2019].

## 4. Набор данных

Особое внимание при решении задачи классификации медицинских (офтальмологических) изображений стоит обратить на качество данных. Однако, в сфере интеллектуального анализа медицинских изображений существует значительные трудности с поиском общедоступных наборов данных, что связано как с редкостью отдельных патологий, так и трудностью сборки и аннотации данных из-за необходимости включения в процесс клиницистов. Так, согласно обзору наборов данных [Khan et al., 2021], проиндексировано всего 10 наборов данных, содержащих аннотированные изображения (снимки ретинальной камеры), свидетельствующие о поражении глазного нерва – глаукоме. Исходя из критериев доступности и общего объёма изображений был выбран набор данных [Song, 2020], содержащих два класса изображений (0 = «Glaucoma», 1 = «Normal»). Характеристики набора данных представлены в табл. 1.

Таблица 1

Набор данных	Соотношение классов		
	Glaucoma	Normal	Всего
[Wheyming Song, 2020]	899 / 62%	551 / 38%	1450

## 5. Метрики качества

Учитывая несбалансированность имеющихся наборов данных для оценки качества получаемого предсказания были использованы метрики: *F1-мера* и *loss*.

*F1-мера* – это гармоническое среднее точности (*Precision*) и полноты (*Recall*), вычисляемая по формуле:

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall}, \quad (5.1)$$

Метрика *loss* (потеря) используется для оценки ошибки модели машинного обучения. Она позволяет измерить, насколько хорошо модель соответствует тренировочным данным и насколько точно она предсказывает целевую переменную.

## 6. Результаты и обсуждение

### 6.1. Предподготовка данных

Для грамотного извлечения признаков с минимальными потерями и оптимизации процессов обучения нейросети необходимо выполнить предобработку входящего изображения. При работе с медицинскими снимками, имеющими высокое разрешение и огромное количество мелких элементов, этот шаг становится особенно актуальным.



Стоит отметить, что в рассматриваемом случае помимо изменения размера согласно входному слою и нормализации изображения необходимо использовать возможности цветowych каналов. Признано, что для работы с офтальмологическими изображениями наиболее предпочтительно использовать «зелёный канал» [Amin et al., 2016]. При поступлении входящего изображения, имеющего все три цветowych канала (RGB) необходимо выделить G-канал и дальше продолжать работать в преобразованном изображении.

Для шумоподавления в «зелёном канале» используется слияние исходного одноканального изображения ( $src1$ ) с изображением, прошедшим фильтр Гаусса ( $src2$ ) [Lestari et al., 2019]:

$$dst = \alpha \times src1 + \beta \times src2 + \gamma . \quad (6.1)$$

## 6.2. Обучение и валидация СНС и СННС

Обучение и валидация сетей производились на выборках подготовленных изображений, разделенных в пропорции 80 на 20 процентов соответственно. Для обучения и валидации сетей также были использованы одинаковые параметры процессов. При этом процесс обучения и валидации СННС занял несколько больше времени чем обычной СНС. Результаты обучения и валидации сетей представлены в табл. 2.

Таблица 2

СНС				СННС			
Train		Valid		Train		Valid	
F1	loss	F1	loss	F1	loss	F1	loss
0.98	0.12	0.9	0.17	0.96	0.1	<b>0.94</b>	<b>0.09</b>

Как можно увидеть из данных табл. 2, СННС имела незначительное преимущество перед обычной СНС только в процессе валидации результатов. При этом, данное преимущество не является статистически значимым.

## Заключение

Таким образом, в ходе исследования был изучен имеющийся опыт использования СННС в задачах классификации изображений. Создана модель СННС, обученная для задачи бинарной классификации изображений, содержащих глаукому. Полученные результаты говорят о незначительном превосходстве СННС в значениях F1 и loss при валидации над обычной СНС (СНС:  $F1-0.9$ ,  $loss-0.17$ ; СННС:  $F1-0.94$ ,  $loss-0.09$ ). Однако, стоит учесть, что базовая модель СНС, ставшая основой для СННС, может быть значительно улучшена и превзойти результаты нейро-нечеткой сети, что говорит о необходимости фундаментального подхода к разработке ИНС.

Проведённое исследование подтверждает целесообразность использования СННС в решении задач интеллектуального анализа медицинских изображений за счёт возможности повышения точности классификации по сравнению с обычными СНС. В дальнейшем, для повышения результатов работы СННС необходимо отойти от использования универсальных правил, генерируемых для работы с наборами данных состоящих из тысяч классов и перейти к формализации диагностических критериев искомых заболеваний в виде нечётких правил.

Дополнительно стоит упомянуть возможности использования СННС в составе гибридных систем поддержки принятия врачебных решений наряду с методами объяснительного искусственного интеллекта, что, например, позволило бы получать два типа объяснений.

### Список литературы

- [Gunning et al., 2019] Gunning D., Aha D. DARPA's Explainable Artificial Intelligence (XAI) Program // AI Magazine. – 2019. – Vol. 40 (2). – P. 44-58.
- [Averkin et al., 2018] Averkin A., Yarushev S. Hybrid Neural Networks and Time Series Forecasting // Artificial Intelligence. Communication in Computer and Information Sciences. – 2018. – Vol. 934. – P. 230-239.
- [Jin, 2011] Jin X.H. Model for efficient risk allocation in privately financed public infrastructure projects using neuro-fuzzy techniques // Journal of Construction Engineering and Management. – 2011. – Vol. 137, No. 11. – P. 1003-1014.
- [Борисов и др., 2017] Борисов В.В., Федулов А.С., Зернов М.М. Основы гибридизации нечетких моделей. Серия «Основы нечеткой математики». Книга 9. – М.: Горячая линия–Телеком, 2017. – 100 с.
- [Rajab et al., 2018] Rajab S., Sharma V. A review on the applications of neuro-fuzzy systems in business // Artificial Intelligence Review. – 2018. – Vol. 49. – P. 481-510.
- [Shihabudheen et al., 2018] Shihabudheen K.V., Pillai G.N. Recent Advances in Neuro-Fuzzy System: A Survey // Knowl.-Based Syst. – 2018. – Vol. 152. – P. 136-162.
- [Lin et al., 1991] Lin C.T., Lee C.S.G. Neural-network-based fuzzy logic control and decision system // IEEE Transactions on computers. – 1991. – Vol. 40, No. 12. – P. 1320-1336.
- [Jang, 1993] Jang J.S.R. ANFIS: adaptive-network-based fuzzy inference system // IEEE transactions on systems, man, and cybernetics. – 1993. – Vol. 23, No. 3. – P. 665-685.
- [Berenji et al., 1992] Berenji H.R., Khedkar P. Learning and tuning fuzzy logic controllers through reinforcements. – 1992. – No. NASA-TM-107875.
- [Nauck et al., 1999] Nauck D., Kruse R. Neuro-fuzzy systems for function approximation // Fuzzy sets and systems. – 1999. – Vol. 101, No. 2. – P. 261-271.
- [Tano et al., 1996] Deep combination of fuzzy inference and neural network in fuzzy inference software—FINEST // Fuzzy Sets and Systems. – 1996. – Vol. 82, No. 2. – P. 151-160.
- [Juang et al., 1998] Juang C.F., Lin C.T. An online self-constructing neural fuzzy inference network and its applications // IEEE transactions on Fuzzy Systems. – 1998. – Vol. 6, No. 1. – P. 12-32.

- [**Kasabov et al., 1999**] Kasabov N.K., Song Q. Dynamic Evolving Fuzzy Neural Networks with "m-out-of-n" Activation Nodes for On-line Adaptive Systems. – Department of Information Science, University of Otago, 1999.
- [**Борисов, 2020**] Борисов В.В. Систематизация нечетких и гибридных нечетких моделей // Мягкие измерения и вычисления. – 2020. – Т. 29, № 4. – С. 98-120.
- [**Борисов и др., 2018**] Борисов В.В., Круглов В.В., Федулов А.С. Нечеткие модели и сети. – М.: Горячая линия–Телеком, 2018. – 284 с.
- [**Батыршин и др., 2007**] Батыршин И.З., Недосекин А.О., Стецко А.А. и др. Нечеткие гибридные системы. Теория и практика / под ред. Н.Г. Ярушкиной. – М.: Физматлит, 2007. – 208 с. – ISBN 978-5-9221-0786-0.
- [**Talpur et al., 2023**] Deep Neuro-Fuzzy System application trends, challenges, and future perspectives: A systematic survey // Artificial intelligence review. – 2023. – Vol. 56, No. 2. – P. 865-913.
- [**Borisov et al., 2019**] Borisov V.V., Korshunova K.P. Multiclass Classification Based on the Convolutional Fuzzy Neural Networks // Proceedings of the 17th Russian Conference on Artificial Intelligence / In: Kuznetsov S., Panov A. (eds). Artificial Intelligence. RCAI 2019.
- [**Khan et al., 2021**] Khan S.M., Liu X., Nath S. et al. A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability // The Lancet Digital Health. – 2021. – Vol. 3, No. 1. – P. e51-e66. – DOI: 10.1016/S2589-7500(20)30240-5
- [**Song, 2020**] Song W. 1450 fundus images with 899 glaucoma data and 551 normal data // IEEE Dataport, – DOI: 10.21227/4bcp-2z21.
- [**Amin et al., 2016**] Amin J., Sharif M., Yasmin M. A review on recent developments for detection of diabetic retinopathy // Scientifica. – 2016. – Vol. 2016.
- [**Lestari et al., 2019**] Lestari T., Maison M. Retinal blood vessel segmentation using Gaussian filter // Journal of physics: conference series. – IOP Publishing, 2019. – Vol. 1376, No. 1. – P. 012023.

УДК 004.942

## МОДЕЛИРОВАНИЕ ПРОЦЕССОВ ПЛАНИРОВАНИЯ НА ОСНОВЕ ТЕМПОРАЛЬНЫХ НЕЙРО-НЕЧЕТКИХ СЕТЕЙ ПЕТРИ

А.В. Бобряков (*avbob@mail.ru*)

Национальный исследовательский университет «МЭИ», Москва

С.А. Прокопенко (*puss95@yandex.by*)

Национальный исследовательский университет «МЭИ», Москва  
Белорусско-Российский университет, Могилёв

Современные технологии производства позволяют выпускать разнообразные виды продукции без необходимости обладания огромными производственными мощностями и при ограниченных ресурсах. Такие производства находятся под влиянием сложного и быстро меняющегося комплекса внешних условий, где эффективное планирование и оптимизация ресурсов становятся ключевыми факторами успеха, при этом учитываются возможные траектории производства сложных технических объектов, план занятости оборудования и персонала, что способствует оптимизации процессов и снижению себестоимости продукции. Эффективное планирование такой сложности невозможно без соответствующего модельного, алгоритмического и программного обеспечения. В статье предложен подход к моделированию процессов планирования производства сложных технических объектов с использованием предложенной разновидности темпоральных нейро-нечетких сетей Петри, рассмотрено алгоритмическое и программное обеспечение для реализации и сопровождения процессов планирования.

**Ключевые слова:** процессы планирования, темпоральные нейро-нечеткие сети Петри.

### Введение

Можно выделить особый класс производств, которые выпускают широкий ассортимент продукции, не обладая большим запасом ресурсов. На таких производствах одно и то же оборудование используется для выпуска различных сложных технических объектов (СТО).

Изготовление СТО проходит в соответствии с одной из возможных траекторий производства. Траектория производства СТО – это последовательный путь, который определяет все этапы, фазы и процессы, необхо-

димые для изготовления и сборки СТО и его компонентов: конкретные производственные задачи, с указанием технологий, методов и средств их достижения, перечень производственных ресурсов и оборудования, а также временных интервалов, выделяемых на их решение. Траектория производства является комплексным и многомерным процессом, требующим точного планирования, координации и управления. Она может варьироваться в зависимости от специфики СТО, технологий, регулирования и других факторов. Производство разных единиц одного и того же класса СТО может проходить по разным траекториям. Ошибки или задержки на любом этапе производства могут повлечь за собой увеличение себестоимости и срыва поставок.

Процесс планирования производства – это процесс создания детального плана действий, включающего множество траекторий производства СТО, необходимых для выпуска сложных технических объектов за определенный промежуток времени [Bobryakov et al., 2021]. Эффективные процессы планирования позволяют производствам выпускать на рынок конкурентоспособную продукцию – это достигается за счет снижения внутренних расходов на производство, отсутствия лишних затрат на ресурсы и т.д. Эффективность процесса планирования может определяться следующими критериями:

- своевременный выпуск продукции. Выполненные своевременно заказы позволяют повысить рейтинг производства среди однотипных производств, избежать штрафных санкций от заказчиков;
- количество бракованной (не соответствующей законодательным, технологическим требованиям) продукции. Большое количество бракованной продукции снижает рейтинг производства, увеличивает себестоимость продукции, повышает риск возникновения непрогнозируемых затрат в будущем, которые могут привести к банкротству производства;
- простой оборудования. Работа оборудования на холостом ходу и количество циклов «запуск/остановка» должно быть минимальным. Рациональное использование оборудования позволяет увеличить время между техническим обслуживанием, снизить риск внезапной поломки оборудования;
- себестоимость выпущенной продукции. Итоговые затраты на продукт не должны превышать критических значений;
- сокращение времени производства продукции (от начала производства – до передачи заказчику) – позволяет снизить количество «замороженных» активов, снизить затраты на хранение готовой продукции.

С каждым годом ужесточаются требования к качеству и срокам получения производимых СТО, возрастает внутренняя и внешняя конкуренция, оказывают своё влияние эпидемиологические, политические, экономические, социальные, технологические и экологические факторы. Производствам необходимо постоянно модернизировать свои процессы планирования, чтобы производимые сложные технические объекты были качественными и конкурентоспособными [Vobryakov et al., 2019].

Для выпуска продукции на таких производствах в условиях высокой конкуренции, ограниченных ресурсов, постоянно меняющихся внутренних и внешних условиях необходимо располагать модельным, алгоритмическим и программным обеспечением, которое позволяет в кратчайшие сроки адаптировать процессы планирования производства.

## 1. Алгоритмическое обеспечение процесса планирования

Планирование, как подход к управлению производством, представляет собой процесс формирования согласованного множества траекторий производства СТО определяющих рациональную последовательность действий по производству и реализации продукции. Основные этапы процесса планирования представлены на рис. 1.



Рис. 1. Процесс составления плана производства

Использование механизма планирования при производстве СТО позволяет:

- определить оптимальную последовательность производственных операций и ресурсов, необходимых для их выполнения;
- определить, какие ресурсы будут необходимы для производства продукции, и как их можно оптимально использовать;
- сократить затраты на производство благодаря оптимизации процессов и использованию ресурсов с максимальной эффективностью;
- контролировать процесс производства на всех его этапах, что позволяет быстро реагировать на возможные проблемы и сбои.

Процесс планирования при производстве продукции должен использовать множество информационно-аналитических (ИАП) и производственно-технологических процессов (ПТП), а также учитывать особенности протекающих процессов:

- информационно-аналитические процессы (сбора, обработки и анализа информации) предоставляют разнородную слабоструктурированную информацию в больших объемах;
- производственно-технологические процессы могут быть реализованы с использованием различных траекторий производства;
- ИАП характеризуются высокой динамикой изменения системных и внешних факторов;
- ПТП должен осуществляться в соответствии с законодательными ограничениями;
- между процессами могут быть временные отношения. К примеру, процесс должен начинаться сразу после окончания предыдущего или спустя какой-то промежуток времени;
- возможность возникновения рисков и проблем (поломка оборудования, некачественные ресурсы, порча комплектующих и т.д.).

Существующие распространенные подходы (системы массового обслуживания, MRP (Material Requirements Planning, планирование потребностей в материалах), ERP (Enterprise Resource Planning, планирование ресурсов предприятия), ЛТ (Just-in-Time, точно в срок), Lean-подход) к планированию в производстве имеют недостатки и ограничения, которые могут привести к неэффективным результатам функционирования производства. Многие подходы к планированию:

- не учитывают изменения внешних условий и требований рынка;
- не учитывают все аспекты производства и не могут обеспечить полное представление о том, как изменения в одной части производства могут повлиять на производство в целом;

- используют ручной или полуавтоматический подход к планированию, что приводит к ошибкам и задержкам в производственных процессах;
- не обеспечивают достаточной гибкости для адаптации к управления изменениям внешних условий и требований рынка, что может привести к дополнительным затратам и задержкам в производстве;
- дороги в эксплуатации и могут потребовать дополнительных затрат для внедрения в существующие производства.

Для моделирования процессов планирования необходим инструментарий, который:

- обладает проработанным математическим аппаратом;
- способен учитывать особенности процессов;
- обладает наглядным графическим представлением;
- допускает использование расширений и модификаций;
- прост во внедрении и эксплуатации.

## **2. Темпоральная нейро-нечеткая сеть Петри**

Одним из популярных подходов для моделирования и управления процессами в сложных системах являются сети Петри [Wang et al., 2018]. Сети Петри являются мощным инструментом исследования моделируемых систем благодаря их возможности описания многих классов дискретных, асинхронных, параллельных, распределенных, недетерминированных систем, благодаря наглядности представления их работы, развитому математическому и программному аппарату анализа [Balbiani et al., 2019]. Сети Петри – достаточно выразительная модель параллелизма, обладающая в то же время значительным набором разрешимых свойств [Andrade et al., 2010].

Применение классических сетей Петри для моделирования процессов планирования затруднительно, однако одной из особенностей сетей Петри является возможность использования различных расширений (цветные, нечеткие и т.д.), которые позволяют их использовать для решения задач.

Нейро-нечеткая сеть Петри в качестве переходов может использовать нейро-нечеткие компоненты. Использование нейро-нечетких компонентов необходимо для работы с большим количеством разнородной слабоструктурированной информации [Kwan et al., 1994], а также для выявления ресурсной комбинации, которая может привести к браку, чрезвычайной ситуации.

В представленной разновидности сетей Петри используются нейро-нечеткие компоненты в переходах, основанные на нейроне Квана и Кэи, представлен на рис. 2, для выявления ситуаций, которые могут привести к получению брака. Данный нейро-нечеткий компонент способен обучаться, чтобы выбирать траекторию в необходимый момент времени при заданных условиях с наименьшей вероятностью брака.



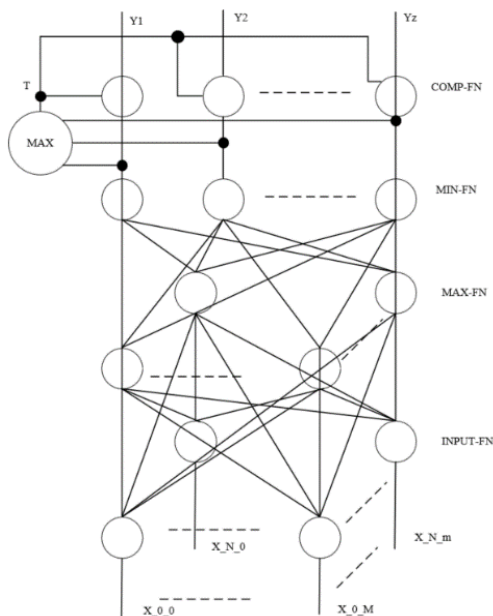


Рис. 2. Структура нейрона Квана и Кэи

Темпоральная логика Аллена основана на определении 13 базовых отношений [Allen, 1983] между временными интервалами, таких как «пересекаются», «начинаются», «заканчиваются» и другие. Каждое из этих отношений может быть выражено в терминах других отношений, что обеспечивает полноту и замкнутость системы. Использование темпоральной логики Аллена необходимо для задания отношений между процессами. Применение темпоральной логики Аллена позволяет снизить количество возможных траекторий производства.

Темпоральная нейро-нечеткая сеть Петри в общем виде может быть представлена следующим образом:

$$TNPN = (P, T, I, O, C, \alpha, f, B, R, Pl, m_0),$$

где  $P = \{P_1, P_2, \dots, P_{np}\}$  – конечное непустое множество вершин сети Петри. Вершина в сети Петри показывает текущее состояние процесса;  $T = \{T_1, T_2, \dots, T_{nt}\}$  – конечное непустое множество переходов сети Петри. Переход в сети Петри представляет собой действие, которое приводит к изменению состояния процесса;  $I = \{I_1, I_2, \dots, I_{nc}\}$  – конечное непустое множество входных функций сети Петри;  $O = \{O_1, O_2, \dots, O_{nc}\}$  – конечное непустое

тое множество выходных функций сети Петри. Выходная функция определяет, какие изменения произойдут в системе или процессе, когда переход будет выполнен. Например, выходная функция может уменьшать количество ресурсов, изменять значения меток на вершинах, увеличивать количество задач в очереди или запускать определенный процесс;  $C = \{C_1, C_2, \dots, C_{nc}\}$  – конечное непустое множество времен, которые могут быть присвоены дугам в сети Петри. Если время указано 0, то им можно пренебречь;  $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_{nc}\}$  – вектор значений порогов нечеткого срабатывания переходов. Значения порогов нечеткого срабатывания в сети Петри могут использоваться для определения степени соответствия меток на вершинах сети нечеткому множеству, а также для принятия решений о том, какой переход должен быть выполнен в текущий момент времени. На значение данного параметра влияет текущее состояние оборудования, квалификация оператора, степень готовности заготовки и т.д.;  $f = \{f_1, f_2, \dots, f_{nc}\}$  – вектор значений функции принадлежности нечеткого срабатывания переходов. Вектор значений функции принадлежности нечеткого срабатывания переходов в сети Петри используется для описания неопределенности или нечеткости в процессе моделирования системы;  $B = \{B_1, B_2, \dots, B_{nb}\}$  – множество базисных интервальных отношений логики Аллена;  $R = \{R_1, R_2, \dots, R_{nr}\}$  – конечное непустое множество ресурсов (комплектующие, вспомогательное оборудование и т.д.), необходимых для функционирования сети Петри;  $Pl$  – план выпуска, который включает в себя множество сложных технических объектов, которые необходимо произвести, дата начала действия плана, дата окончания действия плана;  $m_0$  – вектор начальной маркировки, каждая компонента которого определяется значением функции принадлежности нечеткого наличия маркера в соответствующей позиции.

Переход  $T$  является активным (разрешенным или возбужденным), если выполняется условие:

$$\begin{cases} \min\{m_i\} \geq \alpha_k; \\ R_f(t_j) = true; \\ B_f(\{t\}_j, Time_0) = true. \end{cases} \quad (2.1)$$

Переход  $T$  является потенциально активным, если выполняется условие:

$$\begin{cases} \min\{m_i\} \geq \alpha_k; \\ R_f(t_j) = true. \end{cases} \quad (2.2)$$

Если переход активен при некоторой маркировке  $m$ , то для него выполняется правило активности перехода. Срабатывание перехода приведет к новой маркировке  $m$  для каждой входной позиции:

$$m_i = 0. \quad (2.3)$$

Для каждой из выходной позиций:

$$m_j = \max\{m_j, \min\{m_i, f_k\}\}. \quad (2.4)$$

### 3. Пример планирования производства СТО на основе ТННСП

Рассмотрим использование ТННСП для моделирования процесса планирования на примере производства компонента «Д00.02.01.015 (М22х1,5-6Н) Корпус» устройства «Насос Дозатор Д125-01» в момент времени  $t$ .

#### 3.1. Определение ассортимента выпуска

Производству необходимо выпустить деталь «Д00.02.01.015 (М22х1,5-6Н) Корпус». Особенность данной детали в том, что она может быть произведена по одной из трех возможных траекторий производства. Технологическая схема производства представлена на рисунке 3. Пунктирной линией обозначены отношения между операциями в нотации темпоральной логики Аллена.

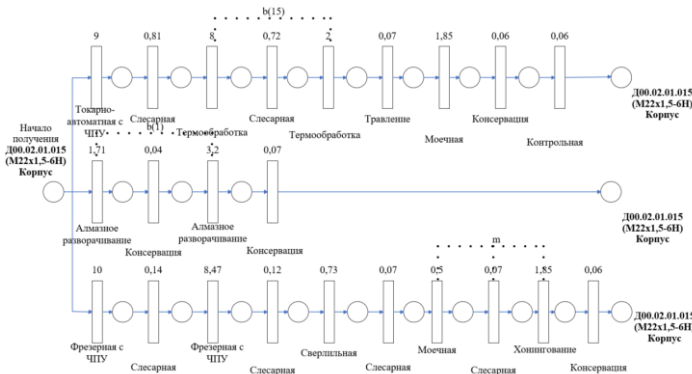


Рис. 3. Технологическая карта производства «Д00.02.01.015 (М22х1,5-6Н) Корпус» в нотации сети Петри

#### 3.2. Определение доступных стационарных ресурсов

Стационарные ресурсы – это ресурсы, которые не могут быть легко перемещены с одного места на другое. Они обычно связаны с определенным местоположением или фиксированной структурой (к примеру, слесарный станок закреплен за определенными сотрудниками). К стационарным ресурсам относятся станки, здания, персонал, рабочие места и т.д.

В момент времени  $t$  для проведения операции «Токарно-автоматная с ЧПУ» доступно 3 рабочих места, «Фрезерная с ЧПУ» доступно 2 рабочих места, «Алмазное разворачивание» нет доступных рабочих мест.

### 3.3. Формирования траекторий

Проецируя технологическую карту на имеющиеся возможные ресурсы, мы получаем схему на рис. 4. На схеме видно, что выполнение операции «Алмазное разворачивание» блокируется выполнением следующей операцией, так как она должна выполняться через минуту, а это невозможно из-за занятости оборудования.

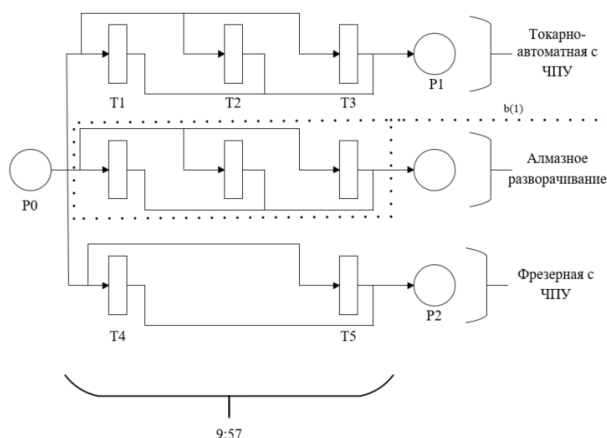


Рис. 4. Проекция части технологической карты производства «Д00.02.01.015 (М22х1,5-6Н) Корпус» в нотации сети Петри на ресурсы

Для определения какой из переходов  $T_1$ – $T_5$  может сработать – рассмотрим процесс функционирования темпоральной нейро-нечеткой сети Петри при следующих параметрах:  $m_0 = \{0.25, 0, 0\}$ ,  $f = \{0.23, 0.23, 0.22, 0.22, 0.22\}$ ,  $\alpha = \{0.7, 0.6, 0.49, 0.24, 0.3\}$ . Активным переходом согласно формулам 2.1 и 2.2 является переход  $T_4$ . Срабатывание перехода  $T_4$  согласно формулам 2.3 и 2.4 приведет к маркировке:

$$m_0 = \{0, \max\{0, \min(0.25, 0.22)\}, 0\} = \{0, 0.22, 0\}.$$

Поэтапно выполняя все переходы, будет получена траектория производства «Д00.02.01.015 (М22х1,5-6Н) Корпус» в момент времени  $t$ .

### 3.4. Определения необходимых ресурсов

Получив траекторию производства компонента «Д00.02.01.015 (М22х1,5-6Н) Корпус», можно определить какие ресурсы и в какой момент времени необходимы для реализации получения компонента.

### 4. Разработанные программно-инструментальные средства

Для реализации ТННСП были разработаны программно-инструментальные средства, позволяющие реализовать и сопровождать процесс планирования и производства сложных технических объектов. Экранная форма инструмента построения траектории производства СТО представлена на рис. 5.

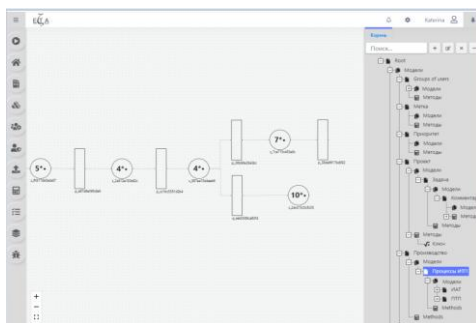


Рис. 5. Экранная форма разработанного инструмента

По окончании функционирования инструмента построения плана выдается отчет показывающий потребность в ресурсах, а также временные промежутки, когда они должны быть использованы для производства СТО.

### Заключение

Для получения результатов применения темпоральных нейро-нечетких сетей Петри при реализации механизма планирования и управления на производственно-сборочном предприятии было проанализирована деятельность предприятия за последние 3 года (2 года без применения ТННСП). За 3 года функционирования производства было проведено 6 523 621 технологических операций, выпущено 119 856 единиц продукции. В результате применения ТННСП удалось:

- снизить простой оборудования на 13%;
- сократить на 9% количество технологических операций, которые приводили к браку;
- снизить количество просроченных заказов на 49% (на 245 заказов за год).

## Список литературы

- [Allen 1983] Allen J.F. Maintaining knowledge about temporal intervals // Communications of the ACM. – 1983. – Vol. 26, No. 11. – P. 832-843.
- [Andrade et al. 2010] Andrade E., Maciel P., Callou G., Araujo C. An Approach Based in Petri Net for Requirement Analysis. – 2010. – DOI: 10.5772/7495.
- [Balbiani et al. 2019] Balbiani P., Boudou J., Diéguez M., Fernández-Duque D. Intuitionistic Linear Temporal Logics // ACM Transactions on Computational Logic. – 2019. – 21. – P. 1-32.
- [Bobryakov et al. 2019] Bobryakov A.V., Krutolevich S.K., Misnik A.E., Prakapenka S.A. Methodology for Development of Industrial Analytical Systems for Data Collection and Processing // Proceedings of the 14th International Conference on Interactive Systems. – 2019.
- [Bobryakov et al. 2021] Bobryakov A.V., Krutalevich S.K., Misnik A.E., Prokopenko S.A. Modeling of Industrial and Technological Processes in Complex Systems Based on NeuroFuzzy Petri Nets // Journal of Physics: Conference Series, Volume 2096, International Conference on Automatics and Energy (ICAE 2021) 7-8 October 2021, Vladivostok, Russia.
- [Kwan et al. 1994] Kwan H., Cai L. A fuzzy neural network and its application to pattern recognition // IEEE Transactions on Fuzzy Systems. – 1994. – Vol. 2, No. 3. – P. 185-193. – DOI: 10.1109/91.298447.
- [Wang et al. 2018] Wang S., You D. and Seatzu C. A Novel Approach for Constraint Transformation in Petri Nets With Uncontrollable Transitions // IEEE Trans. Systems, Man, and Cybernetics: Systems. – 2018. – Vol. 48, No. 8. – P. 1403-1410.

УДК 004.94

## СЦЕНАРНОЕ МОДЕЛИРОВАНИЕ НА ОСНОВЕ НЕЧЕТКИХ РЕЛЯЦИОННЫХ ТЕМПОРАЛЬНЫХ КОГНИТИВНЫХ МОДЕЛЕЙ\*

В.В. Борисов (*vbor67@mail.ru*)

А.П. Жарков (*antonzharkov@yandex.ru*)

Филиал Национального исследовательского университета «МЭИ»  
в г. Смоленске, Смоленск

В.С. Луферов (*lyferov@yandex.ru*)

ООО «Сбер бизнес софт», Смоленск

Описаны нечеткие реляционные темпоральные когнитивные модели (НРТКМ) и основанный на их использовании метод сценарного моделирования, включающий в себя: построение структуры НРТКМ и ее структурную настройку; выбор моделей сценарной динамики для всех концептов НРТКМ; параметрическую настройку НРТКМ; задание сценариев моделирования; моделирование на основе обученной НРТКМ в соответствии с заданным сценарием. Рассмотрен пример использования предложенного метода для сценарного моделирования распределения электрической нагрузки для потребителей региональной энергетической подсистемы.

**Ключевые слова:** нечеткая реляционная темпоральная когнитивная модель, сценарное моделирование, модель сценарной динамики.

### Введение

Сценарное моделирование направлено на решение целого спектра задач по управлению рисками и эффективностью для широкого класса систем, процессов, проблемных ситуаций. Известны различные подходы к сценарному моделированию [Guo et al., 2019], [Savushkin, 2020], [Савушкин и др., 2020], [Katz, 2021].

Нечеткие когнитивные модели хорошо зарекомендовали себя для анализа и моделирования (в том числе, для сценарного моделирования) систем, процессов, проблемных ситуаций в условиях неточности, неполноты

---

\* Исследование выполнено в рамках государственного задания, проект № FSWF-2023-0012.

и различного типа неопределенности данных и отношений между разнокачественными системными и внешними факторами, нелинейности взаимозависимостей между ними. Они также обеспечивают хорошую визуализацию и интерпретируемость процессов и результатов сценарного моделирования [Подгорская и др., 2019]. Однако эти модели не в полной мере учитывают взаимовлияние системных и внешних факторов с различными временными лагами друг относительно друга.

Для снятия указанных ограничений в работе [Borisov et al., 2020] предложены нечетко-нейросетевые когнитивные темпоральные модели, учитывающие взаимовлияние концептов с различными временными лагами друг относительно друга, однако, требующие ресурсозатратной структурно-параметрической настройки для всех концептов, в свою очередь реализуемых достаточно сложными рекуррентными нечетко-нейросетевыми моделями.

Необходимость в расширении возможностей нечетких когнитивных моделей для решения задач сценарного моделирования привела к созданию нечетких реляционных темпоральных когнитивных моделей (НРТКМ), обеспечивающих:

- гибкое изменение значений концептов (системных и внешних факторов) моделей как на этапе сценарного планирования, так и непосредственно в ходе моделирования;
- возможность идентификации, выбора и изменения зависимостей между концептами моделей при изменении условий сценарного моделирования и сценариев;
- типизацию настройки моделей за счет: «персонализации» сценарной динамики для взаимодействующих концептов; оперативной подстройки нечетких отношений влияния между концептами с использованием постоянно формируемых обучающих выборок; вычисления зависимостей между нечетко заданными значениями концептов в векторно-матричном виде [Борисов и др., 2022a].

В статье предложен метод сценарного моделирования на основе НРТКМ и рассмотрен пример моделирования сценариев распределения электрической нагрузки для потребителей региональной энергетической подсистемы.

## 1. Построение структуры и структурная настройка НРТКМ

Нечеткая реляционная темпоральная когнитивная модель представляется в виде [Борисов и др., 2022a]:

$$FRTCM = \langle C, R \rangle,$$
$$C = \{C_i \mid i = 1, \dots, I\}, \quad R = \{R_i \mid i = 1, \dots, I\},$$



$$R_i = \{ \tilde{r}_{ij}(t-l) \mid l=0, \dots, L_j^i, j=1, \dots, J^i \},$$

$$c_i : \tilde{c}_i(t) = \tilde{F}_i \left( \begin{array}{l} \{ \tilde{c}_i(t-k), \tilde{r}_{ii}(t-k) \mid k=1, \dots, L_i^i \}, \\ \{ \tilde{c}_j(t-l), \tilde{r}_{ij}(t-l) \mid j=1, \dots, J^i, l=1, \dots, L_j^i \} \end{array} \right), \quad i=1, \dots, I,$$

где  $C$  – множество концептов НРТКМ;  $I$  – количество концептов НРТКМ;  $R$  – множество нечетких отношений влияния концептов НРТКМ друг на друга;  $R_i$  – подмножество нечетких отношений влияния концептов, воздействующих на концепт  $C_i$ ;  $J^i$  – число концептов, воздействующих на концепт  $C_i$ ;  $\tilde{r}_{ii}(t-k)$  – нечеткое отношение влияния концепта  $C_i$  на себя в момент  $(t-k)$ ;  $L_i^i$  – максимально значимое значение временного лага при влиянии концепта  $C_i$  на себя;  $\tilde{r}_{ij}(t-l)$  – нечеткое отношение влияния концепта  $C_j$  на концепт  $C_i$  в момент  $(t-l)$ ;  $L_j^i$  – максимально значимое значение временного лага при влиянии концепта  $C_j$  на концепт  $C_i$ ;  $\tilde{c}_i(t)$ ,  $\tilde{c}_i(t-k)$ ,  $\tilde{c}_j(t-l)$  – нечеткие значения концептов  $C_i$  и  $C_j$  в соответствующие моменты времени.

Структурная настройка НРТКМ заключается в определении диапазонов временных лагов  $L_i^i$  и  $L_j^i$  при взаимовлиянии концептов, в которых воздействие нечетких отношений взаимовлияния  $\tilde{r}_{ii}(t-k)$  и  $\tilde{r}_{ij}(t-l)$  из соответствующих подмножеств  $R_i$  ( $i=1, \dots, I$ ) является значимым, а также для настройки этих отношений могут быть использованы различные статистические и экспертные методы. Так, при наличии обучающих выборок для отдельных концептов НРТКМ с целью определения нечетких отношений взаимовлияния применима нечеткая множественная линейная регрессия [Abdalla et al., 2007]:

$$\tilde{c}_i(t) = \bigoplus_{j=1}^{J^i} \bigoplus_{l=1}^{L_j^i} (\tilde{a}_{ij}(t-l) \tilde{c}_j(t-l) + \tilde{b}_{ij}(t-l)), \quad i=1, \dots, I,$$

где  $\bigoplus$  – операция нечеткого агрегирования (например, нечеткое сложение);  $\tilde{a}_{ij}(t-l)$  – нечеткие регрессионные коэффициенты;  $\tilde{b}_{ij}(t-l)$  – нечеткие свободные члены (как правило, равны 0).

Значения нечетких регрессионных коэффициентов  $\tilde{a}_{ij}(t-l)$  нормируются  $\tilde{a}_{ij}'(t-l)$  в диапазоне  $[0, 1]$ . На их основе задаются подмножества

$$R_i = \{ \tilde{r}_{ij}(t-l) = \tilde{a}_{ij}'(t-l) \mid l=0, \dots, L_j^i, j=1, \dots, J^i \} \quad \text{нечетких отношений}$$

взаимовлияния концептов НРТКМ. Далее исключаются те отношения, модальные значения которых меньше установленного порога (например, меньше 0,15).

## 2. Модели сценарной динамики для НРТКМ

При сценарном моделировании с использованием НРТКМ передачу влияния между концептами целесообразно представить с использованием нечетких композиционных правил ([Борисов и др., 2022b]), в соответствии с чем возможные модели сценарной динамики примут следующий вид:

$$\begin{aligned}\tilde{c}_i(t) &= \bigoplus_{j=1}^{J^i} \left( \bigoplus_{l=1}^{L_j^i} (\tilde{c}_j(t-l) \circ \tilde{r}_{ij}^l(t-l)) \right), \\ \tilde{c}_i(t) &= \left( \bigoplus_{m=1}^{L_i^i} (\tilde{c}_i(t-m) \circ \tilde{r}_{ii}^m(t-m)) \right) \oplus \left( \bigoplus_{j=1}^{J^i} \left( \bigoplus_{l=1}^{L_j^i} (\tilde{c}_j(t-l) \circ \tilde{r}_{ij}^l(t-l)) \right) \right), \\ \tilde{c}_i(t) &= \left( \bigoplus_{m=1}^{L_i^i} (\Delta\tilde{c}_i(t-m) \circ \tilde{r}_{ii}^m(t-m)) \right) \oplus \left( \bigoplus_{j=1}^{J^i} \left( \bigoplus_{l=1}^{L_j^i} (\Delta\tilde{c}_j(t-l) \circ \tilde{r}_{ij}^l(t-l)) \right) \right), \quad (1)\end{aligned}$$

где  $\Delta\tilde{c}_i(t-m)$  – нечеткое приращение значения концепта  $C_i$  в момент  $(t-m)$ ;  $\Delta\tilde{c}_j(t-l)$  – нечеткое приращение значения концепта  $C_j$  в момент  $(t-l)$ ;  $\circ$  – операция нечеткой композиции;  $\bigoplus_{m=1}^{L_i^i}$  – нечеткое агрегирование влияний концепта  $C_i$  на самого себя в диапазоне учитываемого временного лага ( $k=1, \dots, L_i^i$ );  $\bigoplus_{l=1}^{L_j^i}$  – нечеткое агрегирование влияний концепта  $C_j$  на концепт  $C_i$  в диапазоне учитываемого временного лага ( $l=1, \dots, L_j^i$ );  $\bigoplus_{j=1}^{J^i}$  – нечеткое агрегирование влияний концептов  $C_j$  ( $j=1, \dots, J^i$ ), непосредственно воздействующих на концепт  $C_i$ ;  $\oplus$  – нечеткое агрегирование совокупных влияний.

## 3. Параметрическая настройка НРТКМ

Допустим, множество концептов НРТКМ соответствует множеству компонентов многомерного временного ряда (МВР). При этом в НРТКМ изменение значений ее концептов может быть представлено в виде значений компонентных временных рядов, образующих МВР:

$$C = \{C_i \mid i = 1, \dots, I\},$$

$$C_i = \{\tilde{c}_i(t) \mid t = 1, \dots, T, \dots\}, i = 1, \dots, I,$$

$$\forall t \in \{1, \dots, T, \dots\}$$

$$\tilde{C}(t) = \left\{ \begin{array}{l} \tilde{c}_1(t) = F_1(\varphi_{1,1}(\tilde{c}_1(t-1), \dots, \tilde{c}_1(t-L_1^1)), \dots, \varphi_{1,l}(\tilde{c}_l(t), \dots, \tilde{c}_l(t-L_l^1))), \\ \dots \\ \tilde{c}_i(t) = F_i(\varphi_{i,1}(\tilde{c}_1(t-1), \dots, \tilde{c}_1(t-L_i^1)), \dots, \varphi_{i,l}(\tilde{c}_l(t), \dots, \tilde{c}_l(t-L_l^i))), \\ \dots \\ \tilde{c}_I(t) = F_I(\varphi_{I,1}(\tilde{c}_1(t-1), \dots, \tilde{c}_1(t-L_1^I)), \dots, \varphi_{I,l}(\tilde{c}_l(t), \dots, \tilde{c}_l(t-L_l^I))) \end{array} \right\},$$

где  $\tilde{C}(t) = \{\tilde{c}_1(t), \dots, \tilde{c}_I(t)\}$  – «временной срез» нечетких значений концептов НРТКМ (компонентов МВР) в  $t$ -й момент;  $\tilde{c}_i(t)$  – нечеткое значение концепта  $C_i$  в момент времени  $t$ ;  $\varphi_{i,j}$  и  $F_i$  – оператор для учета влияния  $\{\tilde{c}_j(t-1), \dots, \tilde{c}_j(t-L_j^i)\}$  на  $\tilde{c}_i(t)$ , а также преобразование для вычисления  $\tilde{c}_i(t)$ , соответственно, реализуемые с использованием выбранной модели сценарной динамики.

Тогда постановка задачи параметрической настройки НРТКМ сводится к минимизации ошибок прогнозирования одновременно для всех компонентов МВР и представляется следующим образом:

$$\delta C = (\delta C_1, \dots, \delta C_I),$$

$$\delta C \rightarrow \min,$$

$$\forall i \in 1, \dots, I, \delta C_i \rightarrow \min,$$

где  $\delta C$  – ошибка прогнозирования МВР в целом;  $\delta C_i$  – ошибка прогнозирования  $i$ -го компонента МВР.

Многокритериальный характер параметрической настройки НРТКМ обуславливает необходимость минимизации ошибок прогнозирования одновременно для всех компонентов МВР. Однако этого, как правило, невозможно достичь в условиях неопределенности, нелинейности взаимовлияния, частичной несогласованности и существенной взаимозависимости концептов НРТКМ.

Поэтому применим компромиссный подход к многокритериальному оцениванию точности (ошибки) прогнозирования  $\delta C$  МВР на основе обобщенного принципа Эджворта–Парето ([Noghin, 2018]), предполагающего невозможность максимального повышения точности прогнозирования какого-либо компонента МВР в ущерб точности прогнозирования другого(-их) компонента(-ов) МВР.

Параметрическая настройка НРТКМ заключается в настройке параметров моделей сценарной динамики всех концептов НРТКМ.

При этом процедуре параметрической настройки модели сценарной динамики каждого концепта  $C_i$  НРТКМ предшествует формирование обучающей выборки из ретроспективных данных соответствующего ей компонента МВР:

$$\left( \left\{ \left( \tilde{c}_i^{(k)}(t-1), \dots, \tilde{c}_i^{(k)}(t-L_i^j) \right) \mid j \in 1, \dots, I_i \right\}, \tilde{c}_i^{(k)}(t) \right), \quad k = 1, \dots, K,$$

где  $\left\{ \left( \tilde{c}_i^{(k)}(t-1), \dots, \tilde{c}_i^{(k)}(t-L_i^j) \right) \mid j \in 1, \dots, I_i \right\}, \tilde{c}_i^{(k)}(t)$  – значения (в общем случае, нечеткие) входных и выходной переменных модели сценарной динамики  $\tilde{c}_i(t)$  в  $k$ -м примере;  $K$  – число примеров обучающей выборки.

Для модели сценарной динамики каждого концепта  $C_i$  НРТКМ настраиваемыми параметрами являются нечеткие отношения влияния  $\tilde{r}_{ii}(t-m)$  ( $m = 1, \dots, L_i^i$ ) и  $\tilde{r}_{ij}(t-l)$  ( $j = 1, \dots, I_i, l = 1, \dots, L_j^i$ )

В общем виде процедура параметрической настройки модели сценарной динамики концепта  $C_i$  НРТКМ состоит из следующих шагов.

*Шаг 1.* По каждому примеру обучающей выборки для значений входных переменных  $\left\{ \left( \tilde{c}_i^{(k)}(t-1), \dots, \tilde{c}_i^{(k)}(t-L_i^j) \right) \mid j \in 1, \dots, I_i \right\}$  модель сценарной динамики концепта  $C_i$  НРТКМ рассчитывает значение выходной переменной  $\tilde{c}_i^{r(k)}(t)$ .

*Шаг 2.* По всем примерам обучающей выборки вычисляется ошибка прогнозирования компонента МВР, зависящая от настраиваемых параметров модели сценарной динамики концепта  $C_i$  НРТКМ:

$$\delta C_i = \sqrt{\frac{1}{K} \sum_{k=1}^K \left( \tilde{c}_i^{r(k)}(t) - \tilde{c}_i^{(k)}(t) \right)^2}.$$

*Шаг 3.* В соответствии с алгоритмом обучения (например, генетическим алгоритмом) выполняется корректировка настраиваемых параметров модели сценарной динамики концепта  $C_i$  НРТКМ.

Шаги 1–3 итеративно повторяются, и процедура параметрической настройки модели сценарной динамики концепта  $C_i$  НРТКМ считается успешно завершенной, если итоговая ошибка прогнозирования МВР не превышает заданного порогового значения.

#### 4. Сценарное моделирование с использованием НРТКМ

Предлагаемый метод сценарного моделирования с использованием НРТКМ включает в себя:

- во-первых, построение структуры НРТКМ и ее структурную настройку, заключающуюся в определении диапазонов временных лагов при взаимовлиянии концептов, в которых воздействие нечетких отношений взаимовлияния является значимым (см. 1-й раздел статьи);
- во-вторых, выбор модели(-ей) сценарной динамики для всех концептов НРТКМ (см. 2-й раздел статьи);
- в-третьих, параметрическую настройку НРТКМ, то есть настройку параметров моделей сценарной динамики всех концептов НРТКМ (см. 3-й раздел статьи);
- в-четвертых, задание сценария моделирования с использованием обученной НРТКМ, включающего в себя задание начальных значений концептов, а также изменение значений концептов и/или нечетких отношений влияния между концептами в соответствующие моменты модельного времени;
- в-пятых, непосредственно сценарное моделирование с использованием обученной НРТКМ в соответствии с заданным сценарием.

В качестве примера рассмотрим задачу сценарного моделирования распределения электрической нагрузки для 4-х потребителей, объединенных в региональную энергетическую подсистему.

Для определения множества нечетких отношений влияния  $R$  между этими потребителями (концепты  $C_1$ – $C_4$  НРТКМ) предварительно был выполнен анализ многомерного временного ряда их электрической нагрузки, включавшей в себя 720 (30 дней по 24 часа) измерений.

На рис. 1 проиллюстрирован суточный график распределения электрической нагрузки для 4-х потребителей.

В соответствии с анализом полученных нечетких регрессионных коэффициентов обоснован выбор максимального значения временного лага (равного 4). В табл. 1 показана сформированная матрица нечетких отношений влияния между концептами НРТКМ, значения которой представлены соответствующими нечеткими треугольными числами.

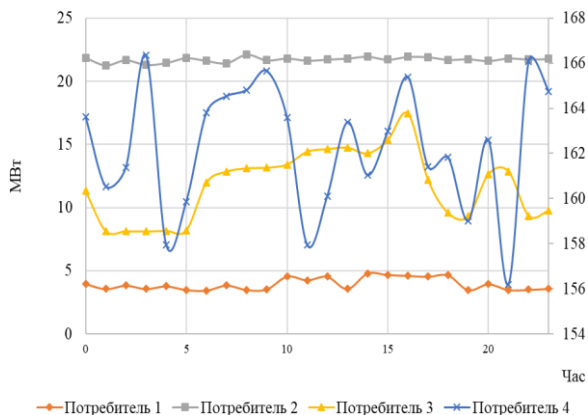


Рис. 1. Иллюстрация графика электрической нагрузки в марте 2023 года

Таблица 1

$R$	$l$	$C_2$	$C_2$	$C_3$	$C_4$
$C_1$	1	[0; 0; 0]	[0; 0; 0]	[0; 0; 0]	[0,74; 1,00; 1,29]
	2	[0; 0; 0]	[0; 0; 0]	[0; 0; 0]	[0,53; 0,67; 0,76]
	3	[0; 0; 0]	[0; 0; 0]	[0,59; 0,62; 0,69]	[0,54; 0,71; 0,84]
	4	[0,70; 0,75; 0,78]	[0; 0; 0]	[0; 0; 0]	[0; 0; 0]
$C_2$	1	[0,72; 0,86; 1,00]	[0; 0; 0]	[0,48; 0,51; 0,55]	[0; 0; 0]
	2	[0,64; 0,73; 0,85]	[0,54; 0,72; 0,81]	[0,58; 0,70; 0,81]	[0; 0; 0]
	3	[0; 0; 0]	[0; 0; 0]	[0; 0; 0]	[0; 0; 0]
	4	[0; 0; 0]	[0; 0; 0]	[0; 0; 0]	[0; 0; 0]
$C_3$	1	[0,47; 0,66; 0,70]	[0,56; 0,69; 0,89]	[0,52; 0,58; 0,71]	[0,58; 0,59; 0,66]
	2	[0,48; 0,60; 0,77]	[0; 0; 0]	[0,55; 0,66; 0,81]	[0; 0; 0]
	3	[0,46; 0,65; 0,77]	[0; 0; 0]	[0; 0; 0]	[0; 0; 0]
	4	[0; 0; 0]	[0; 0; 0]	[0; 0; 0]	[0,49; 0,62; 0,75]
$C_4$	1	[0,87; 1,00; 1,00]	[0,44; 0,61; 0,63]	[0,36; 0,50; 0,64]	[0,45; 0,59; 0,64]
	2	[0; 0; 0]	[0,84; 1,00; 1,00]	[0,70; 0,90; 0,92]	[0,39; 0,41; 0,51]
	3	[0; 0; 0]	[0,52; 0,61; 0,66]	[0,90; 1,00; 1,00]	[0,50; 0,65; 0,71]
	4	[0; 0; 0]	[0; 0; 0]	[0; 0; 0]	[0; 0; 0]

В соответствии с полученными значениями нечетких отношений влияния между концептами выполнена структурная настройка НРТКМ, структура которой представлена на рис. 2.

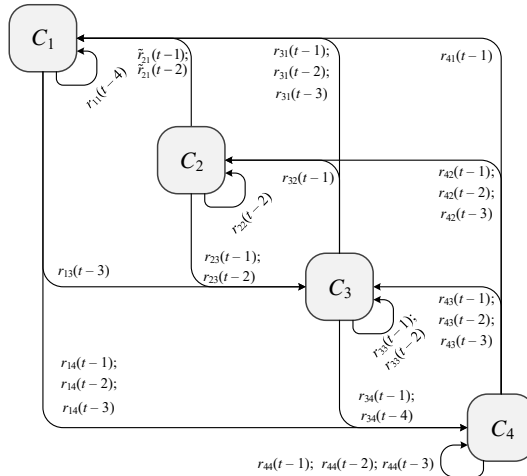


Рис. 2. Структура НРТКМ для сценарного моделирования электрической нагрузки для потребителей региональной энергетической подсистемы

Далее в соответствии в процедурой, рассмотренной в разделе 3, была выполнена параметрическая настройка НРТКМ.

При этом при прогнозировании МВР (с целью параметрической настройки НРТКМ) и для последующего сценарного моделирования распределения электрической нагрузки в качестве моделей сценарной динамики для всех концептов НРТКМ использовалось выражение (1), учитывающее нечеткие значения и приращения концептов и предъявляющее следующие требования к операциям над нечеткими значениями: операция агрегирования должна удовлетворять свойствам коммутативности и ассоциативности для независимости результата операции от порядка расположения нечетких переменных; операции агрегирования и нечеткой композиции должны иметь аддитивный характер накопления результатов [Федулов, 2006].

С учетом этого для нечеткого приращения использовалась операция нечеткого вычитания, для нечеткой композиции – векторно-матричные операции нечеткого умножения и сложения, а для нечеткого агрегирования – операция нечеткого сложения.

Сравнительная оценка результатов прогнозирования МВР электрической нагрузки, иллюстрирующая положительные итоги выполненной параметрической настройки НРТКМ, приведена в табл. 2. Эти результаты показывают, что использование НРТКМ позволяет повысить точность прогнозирования по сравнению с рассмотренными моделями, в том числе по сравнению с продемонстрировавшей наилучшие результаты моделью ANFIS-типа.

Таблица 2

Компоненты МВР (концепты НКРТМ)	Точность прогнозирования, <i>MAPE</i> , %						
	<i>LR</i>	<i>SR</i>	<i>ARIMA</i>	<i>GARCH</i>	ИНС	<i>ANFIS</i>	НКРТМ
Потребитель 1	85,2	85,6	90,7	85,0	89,8	88,9	92,6
Потребитель 2	90,9	93,1	95,4	92,1	93,6	93,1	97,4
Потребитель 3	77,1	78,2	86,8	82,8	86,9	88,3	93,3
Потребитель 4	98,6	98,6	98,6	98,6	98,4	98,6	99,7
МВР в целом	77,1	78,2	86,8	82,8	86,9	88,3	92,6

**Примечание:** *LR* – множественная линейная регрессия; *SR* – линейная регрессия со стохастическим градиентным спуском; *ARIMA* – авторегрессионное интегрированное скользящее среднее; *GARCH* – обобщенная авторегрессионная модель гетероскедастичности; ИНС – искусственные нейронные сети (многослойный персептрон); *ANFIS* – адаптивная сеть нечеткого вывода.

Экспертами были сформированы следующие типовые сценарии при распределении электрической нагрузки для потребителей:

- изменение потребления электроэнергии из-за изменения погоды;
- повреждение электрической сети, приводящее к дефициту мощностей;
- аварийный/плановый ремонт сетевого оборудования;
- регулирование напряжения в контрольных точках электрической сети у потребителей;
- обеспечение для каждого потребителя установленной мощности.

На рис. 3 показана временная «развертка» сценарного моделирования, соответствующая структуре НКРТМ из рис. 2.

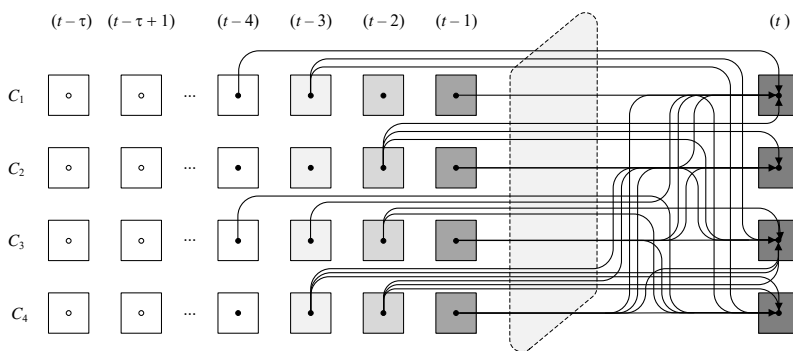


Рис. 3. Временная «развертка» сценарного моделирования, соответствующая структуре НКРТМ



В ходе предварительных исследований обоснована достоверность результатов сценарного моделирования с использованием обученной НРКТМ для всех типовых сценариев распределения электрической нагрузки по потребителям региональной энергетической подсистемы.

### **Заключение**

В статье описаны предложенные нечеткие реляционные темпоральные когнитивные модели, обеспечивающие: учет взаимовлияния концептов с различными временными лагами друг относительно друга; гибкое изменение значений концептов моделей как на этапе сценарного планирования, так и непосредственно в ходе моделирования; идентификацию, выбор и изменение зависимостей между концептами при изменении условий сценарного моделирования и сценариев; типизацию структурно-параметрической настройки.

Предложен метод сценарного моделирования на основе НРКТМ, включающий в себя:

- построение структуры НРКТМ и ее структурную настройку, заключающуюся в определении диапазонов временных лагов при взаимовлиянии концептов, в которых воздействие нечетких отношений взаимовлияния является значимым;
- выбор модели(-ей) сценарной динамики для всех концептов НРКТМ;
- параметрическую настройку НРКТМ, заключающуюся в настройке параметров моделей сценарной динамики всех концептов НРКТМ;
- задание сценариев моделирования с использованием обученной НРКТМ, заключающееся в задании начальных значений концептов, а также в изменении значений концептов и/или нечетких отношений влияния между концептами в соответствующие моменты модельного времени;
- непосредственно сценарное моделирование с использованием обученной НРКТМ в соответствии с заданным сценарием.

Рассмотрен пример использования предложенного метода на основе НРКТМ для сценарного моделирования распределения электрической нагрузки потребителей региональной энергетической подсистемы. В ходе предварительных исследований обоснована достоверность результатов сценарного моделирования с использованием обученной НРКТМ для типовых сценариев распределения электрической нагрузки по потребителям региональной энергетической подсистемы.

## Список литературы

- [Борисов и др., 2022a] Борисов В.В., Курилин С.П., Жарков А.П., Соколов А.М. Многомерное прогнозирование состояния неоднородных электромеханических систем для управления рисками нарушения их работоспособности на основе нечетких темпоральных онтологических и когнитивных моделей // Системы управления, связи и безопасности. – 2022. – № 4. – С. 83-102.
- [Борисов и др., 2022b] Борисов В.В., Курилин С.П., Луферов В.С. Нечеткие реляционные когнитивные темпоральные модели для анализа и прогнозирования состояния сложных технических систем // Прикладная информатика. – 2022. – Т. 17, № 1. – С. 27-38.
- [Подгорская и др., 2019] Подгорская С.В., Подвесовский А.Г., Исаев Р.А., Тарасов А.С., Бахматова Г.А. Моделирование сценарного развития сельских территорий на основе нечеткой когнитивной модели // Проблемы управления. – 2019. – № 5. – С. 49-59.
- [Савушкин и др., 2020] Савушкин С.А., Цыганов В.В. Сценарии развития транспортного комплекса макрорегиона // Информационные технологии в науке, образовании и управлении. – 2020. – № 1 (15). – С. 13-18.
- [Федулов, 2006] Федулов А.С. Вид взаимодействия нечетких чисел, ограничивающий возрастание неопределенности при выполнении операций нечеткой арифметики // Вестник Московского энергетического института. – 2006. – № 1. – С. 101-109.
- [Abdalla et al., 2007] Abdalla A., Buckley J.J. Monte Carlo Methods in Fuzzy Linear Regression // Soft Computing. – 2007. – Vol. 11. – P. 991-996.
- [Borisov et al., 2020] Borisov V., Lufarov V. Neuro-Fuzzy Cognitive Temporal Models for Predicting Multidimensional Time Series with Fuzzy Trends // Computación y Sistemas. – 2020. – Vol. 24, No. 3. – P. 1165-1177.
- [Guo et al., 2019] Guo Y., Kalidindi V.V., Arief M., Wang W., Zhu J., Peng H., Zhao D. Modeling Multi-Vehicle Interaction Scenarios Using Gaussian Random Field // Proc. of the IEEE Intelligent Transportation Systems Conference (ITSC). – 2019. – P. 3974-3980.
- [Katz, 2021] Katz G.. Towards Repairing Scenario-Based Models with Rich Events // In Proc. of the 9th International Conference on Model-Driven Engineering and Software Development – MODELSWARD. – 2021. – P. 362-372.
- [Noghin, 2018] Noghin V.D. Edgeworth–Pareto principle // Studies in Systems, Decision and Control. – 2018. – Vol. 126. – P. 1-22.
- [Savushkin, 2020] Savushkin S. Problems of Scenario Modeling of the Transport Complex // Proc. of the Conf. on Management of Large-Scale System Development. – Moscow: IEEE, 2020. – P. 1-5.

УДК 004.23, 004.72

## РАСШИРЕНИЕ ФУНКЦИОНАЛЬНОСТИ УЗЛОВ В БЕСПРОВОДНЫХ СЕНСОРНЫХ СЕТЯХ

Г.П. Виноградов (*wgp272ng@mail.ru*)

Тверской государственный технический университет, Тверь  
Научно-исследовательский институт программных систем, Тверь

Цель работы – разработка методов расширения функциональности в системах управления узлами беспроводной сенсорной сети на базе нечеткого логического вывода и теории паттернов. Приведена формальная модель паттерна, включающая ряд ментальных параметров. Показан способ ее реализации в форме алгоритмической модели и разработан алгоритм ее реализации. В статье используются методы теории нечетких множеств, теории построения нечетких моделей и сетей. Описанный подход реализован в форме программного продукта для TinyOS. Программный продукт написан на языке программирования nesC – диалекте C.

**Ключевые слова:** модель, нечеткая логика, паттерн, беспроводная сеть, датчик, искусственный интеллект.

### Введение

Беспроводные сенсорные сети (БСС) за счет таких характеристик, как низкое потребление энергии, низкая стоимость, самоорганизация, распределение и небольшой размер, возможность определять местоположение событий получили применение в различных областях [Лихтциндер и др., 2020], [Ефименко, 2018], [Kumar et al., 2016], [Zhao et al., 2016]. Полученная ими информация чрезвычайно важна для мониторинга среды средствами беспроводной сенсорной сети [Chen et al., 2019], [Huang et al., 2019]. Увеличение объемов информации в БСС породило ряд проблем, которые стали центром исследований в области беспроводных сенсорных сетей [Li et al., 2015], [Kumar et al., 2014]. Они направлены на поиск путей расширения функциональности узлов сети, особенно при их применении в средах, характеризующихся неопределенностью [Виноградов, 2022b]. Основные из них: узел сети должен быть способен на предварительную обработку информации, включающую кластеризацию, агрегацию и слияние разнородных данных [Аверкин и др., 2015], [Виноградов и др., 2021], изучению обстановки и принятию решений, как самостоятельно, так и в со-

ставе группы с учетом получаемых от других узлов сети данных. Кроме того, система управления узлом должна осуществлять распределенную обработку сложных запросов внутри сети и осуществлять маршрутизацию с оптимизацией расхода энергии. Реализация этих требований возможна путем совершенствования интеллектуальной составляющей системы управления как сети в целом, так и отдельным узлом в частности. Работа посвящена разработке методов, расширяющих функциональность беспроводных сенсорных сетей и превращающих их в *реагирующие беспроводные сенсорные сети* (РБСС) [Виноградов и др., 2021]. Наиболее удобным для построения интеллектуальных систем распознавания и поддержки принятия решений на основе сенсорных сетей является модель нечеткого сенсора [Виноградов и др., 2021], ядром которого является машина вывода [Аверкин и др., 2015]. В узлы по сети могут пересылаться *шаблоны правил* (номера переменных и термов, входящие в левую и правую часть базы знаний, параметры новых термов, параметры треугольных норм). При решении задачи, например, слежения за целью, такие пересылки могут осуществляться узлами самостоятельно, на основании имеющихся правил, в зависимости от положения цели и содержания запроса. Однако в более сложных случаях используется еще один уровень вершинкоординаторов, которые управляют пересылкой знаний по сети или самообучением узлов и сети. Для этого они, в общем случае, обладают данными о распределении знаний по сети, у них имеются внутренние модели среды и наблюдаемых объектов.

## **1. Проблемы обработки данных в БСС, требующие интеллектуализации сенсоров**

Основными факторами, влияющими на функционирование БСС являются [Виноградов и др., 2021]: ограниченные вычислительные возможности узлов БСС; радиус коммуникации между узлами в большинстве приложений ограничен (в военных приложениях ограничен ближайшими соседями из-за требований скрытности); узлы и/или отдельные каналы связи могут временно прекращать функционировать, что приводит к изменению топологии БСС и необходимости иметь множество дублирующих маршрутов; неравномерный расход энергии в узлах; рост энергетических и временных затрат на передачу данных с увеличением размеров сети.

Две последние проблемы представляют для пользователя особую важность. Малая пропускная способность сети предполагает уменьшение потока данных через сеть за счет добавления распределенной иерархической обработки данных внутри сети и передачи пользователю лишь релевантных ответов на запрос. Известные алгоритмы предполагают только пересылку по сети больших объемов данных. Для решения данной проблемы

предлагается дополнить пересылку данных пересылкой знаний, требуемых для необходимой настройки локальных узлов. При таком подходе сенсор становится интеллектуальным агентом. Он обладает знаниями о себе и о среде и способен к самостоятельному принятию решений. У сенсора возникает возможность управления имеющимися знаниями и их передачи другому сенсору.

Реализация подхода предполагает разработку относительно простых алгоритмов интеллектуализации системы управления узлом РБСС в условиях жестких требований к времени отклика, вычислительным ресурсам, радиусу радиообмена и энергопотреблению. Предлагаемая технология основана на использовании методов нечеткой логики и эффективного опыта в форме паттернов и является объединением таких компонентов как пользователь (компьютер) и беспроводная сенсорная сеть (БСС). Пользователь задает характеристики объекта (виртуальные атрибуты), значения которых надо вычислить, и правила, при помощи которых эти атрибуты будут подсчитаны. Если узел не имеет достаточно данных для вычисления атрибутов, он может запросить их у другого узла.

## 2. Формальное определение паттерна

*Паттерн* – это модель поведения, получаемая как результат деятельности естественного или искусственного объекта, связанной с мониторингом, принятием решения, его реализацией и т.д. [Виноградов, 2022b]. Паттерн рассматривается как шаблон (образец) для повторных действий в аналогичных типовых ситуациях (ТС). Поведение в ТС складывается из следующих компонент: субъект, осуществляющий выбор (агент)  $k \in K$ ; окружение выбора  $S$  – множество элементов среды и их существенных свойств, изменение в любом из которых может стать причиной или продуцировать изменение состояния выбора; доступные способы действий  $c_j^k \in C^k, j = \overline{1, n}$   $k$ -го агента, которые ему известны и могут быть использованы для достижения  $i$ -го результата; возможные в окружении  $S$  результаты существенные для агента –  $o_i^k \in O^k, i = \overline{1, m}$ ; способы оценки свойств получаемых результатов, отражающих их ценность для агента; модель ограничений, как требования на значения выходных переменных и выбор способов действия; модель предметной области, представляющая собой множество соотношений, описывающих зависимость результатов от выбора способа действия и возмущений (база знаний).

В соответствии с гипотезой о рациональном поведении агент формирует решение в соответствии с моделью [Vinogradov et al., 2022]:

$$\begin{aligned}
P_i^k (s \in S) &= \text{Arg max} ( \sum_{c_j^k} E\varphi_i(o_i^k(c_j^k)) - EE_i^k(o_i^k(c_j^k)) ) \\
c_j^k &\in C^k(I_t^i), I_t^i \subseteq M, o_i^k \in O^k \\
\sum_i \chi_{i1}^k(E\varphi_i^k) &\geq \chi_1^0, \quad \sum_i \chi_{i2}^k(EE_i^k) \geq \chi_2^0 \\
\sigma^k(Es^k(X)) &\geq \sigma_0^k
\end{aligned} \tag{1}$$

Такая модель индивидуального поведения агента предполагает формирование базы знаний путем обучения на основе, как экспериментального опыта, так и опыта эксперта, что позволяет реализовать эволюцию «кооперативного интеллекта», аналогичного тому, что имеет место у естественных существ. Как показано в [Vinogradov et al., 2022] описание паттерна имеет форму алгоритмической модели. Набор моделей или поведенческих паттернов формирует опыт или базу паттернов. В случае использования знаний эксперта связь между конкретным классом ситуации и конкретным паттерном устанавливается при работе с оператором - реальным носителем знаний в процессе формирования базы знаний. Соответствующая методика построения алгоритмической модели паттерна описана в [Виноградов, 2022b]. Все компоненты такой модели, за исключением логических связей, могут быть языковыми конструкциями на естественном или естественно-профессиональном языке.

Реализация предложенного подхода связана с: 1) правильным соотношением вектора координат ситуации с тем или иным паттерном, как с точки зрения их количества, так и с точки зрения формы представления каждой координаты; 2) полнотой описания ситуации.

На узле установлены сенсоры, воспринимающие различные сигналы физической природы. Поскольку оператор описывает ситуацию средствами естественного языка, то координаты ситуационного вектора являются лингвистическими переменными. Для этого количественные шкалы в датчиках преобразуются в лингвистические. Результат затем может использоваться, например, для определения типа цели, ее степени угрозы и т.п. Вычисления значений в таком виртуальном сенсоре целесообразно выполнять средствами нечеткого вывода. Для двух типов данных в узле  $k$  должны храниться списки имен, таблицы значений атрибутов. Для вычисляемых атрибутов должны храниться нечеткие правила для вычисления атрибутов, параметры функций принадлежности термов.

### 3. База знаний на основе паттернов

Пусть состояние ситуации описывается ситуационным вектором  $x_i, i = \overline{1, n}$ , каждая координата которого лингвистическая переменная  $x_i$  с множеством термов  $A_i = \{a_i^k, k = \overline{1, K^i}\}$ . Пусть определен набор классов реализаций ситуационного вектора, для которых есть паттерны поведения успешного разрешения любой ситуации класса. При этом значения удельной ценности по результату и удельной ценности по эффективности превышают некоторый пороговый уровень при реализации паттерна.

Пусть накоплено множество  $d_j, j = \overline{1, p}$  паттернов, каждый из которых связан с множеством ситуаций кластера, для разрешения которых он был выбран. Тогда можно составить матрицу соответствия между классами ситуаций и располагаемым множеством паттернов поведения. Если сгруппировать строки матрицы по паттернам, где каждая строка матрицы представляет собой конкретный ситуационный кластер, при котором успешно реализован соответствующий паттерн, то в результате можно получить кластер типовой ситуации. Для него паттерн поведения по структуре будет один и тот же, отличие будет в значениях параметров. Пронумеруем строки блока паттерна двумя индексами: первый индекс – номер паттерна (он же номер блока); второй индекс – это номер ситуации, входящий в  $j$ -й кластер. Такая матрица определяет систему логических высказываний вида Если ... То ..., Иначе ... (их примеры приведены ниже). Полученную упорядоченную таким образом систему логических высказываний будем называть нечеткой матрицей знаний или базой паттернов.

Первоначально каждый узел  $k$  имеет некоторый список физических атрибутов, который соответствуют его набору датчиков, и пустую оболочку машины вывода и пустой список лингвистических атрибутов. Пользователь может добавить в список лингвистических атрибутов узла  $k$  лингвистические внутренние (выходные) атрибуты и послать правило его вычисления. После установления базы правил лингвистических атрибутов в машину вывода в узел  $k$  начинает измерять значения внутренних физических атрибутов, получать от сети и пользователя значения внешних физических и лингвистических атрибутов [Виноградов, 2022а] и производить вычисления внутренних лингвистических атрибутов. Полученные значения внутренних физических и лингвистических атрибутов сохраняются в соответствующих списках и по требованию предоставляются пользователю или другим узлам сети. Пользователь или другие узлы сети могут добавлять в узел  $k$  или исключать лингвистические атрибуты базы правил.

#### 4. Алгоритм вычисления функции принадлежности паттерна

Бортовая интеллектуальная система с помощью своих измерительных средств формирует вектор  $x_i^*, i = \overline{1, n}$  с количественными характеристиками координат. В этой точке, полученной на момент  $t$ , требуется определить значение функции принадлежности паттерна  $d_j(x_i, i = \overline{1, n}), j \in \overline{1, p}$ . В этом случае функция принадлежности рассчитывается для каждого блока.

Наиболее предпочтительный паттерн для разрешения проблемной ситуации будет паттерн  $d_j^*$ , вычисленный по правилу [Виноградов и др., 2021]

$$d^*(x_i^*, i = \overline{1, n}) = \max_{j \in \overline{1, p}} d_j(x_i, i = \overline{1, n}).$$

Блок логических высказываний, относящийся к паттерну  $d_j$ , представляет собой объединение соответствующих нечетких множеств, образованных строками выделенного блока. Тогда функция принадлежности этого объединения будет определяться

$$\mu_{d_j}(x_i, i = \overline{1, n}) = \bigvee_{j=1}^p \bigwedge_{i=1}^n \mu_i^{jK_j}(x_i),$$

где « $\wedge$ » и « $\vee$ » означают операции *min* и *max* соответственно.

Формально представленный алгоритм определения функции принадлежности паттерна  $d_j$  можно записать в виде:

а) фиксируем произвольную точку  $x_i^*, i = \overline{1, n} \in \prod_{i=1}^n U_i$ .

б) для каждого блока матрицы знаний, соответствующего  $d_j$ , определим  $\mu_{d_j}(x_i, i = \overline{1, n})$  в этой точке согласно правилу, приведенному в [Виноградов и др., 2021].

Проводя такие вычисления для каждой точки универсального множества, получим требуемые значения функции принадлежности. Это позволяет алгоритм определения функции принадлежности паттерна  $d_j$  записать в виде:

а) фиксируем произвольную точку  $x_i^*, i = \overline{1, n} \in \prod_{i=1}^n U_i$ ; б) для каждого блока матрицы знаний, соответствующего  $d_j$ , определим  $\mu_{d_j}(x_i, i = \overline{1, n})$  в этой точке.



Операция  $\min_i (a_i^{js})^*$  производится над числами, стоящими в строках «i»  $\overline{1, n}$  и в столбец «min» заносится минимальное число в соответствующей строке. Операция  $\max_{j_s} \min_i (a_i^{js})^*$  определяет в полученном массиве строчных минимумов  $j_s \in \overline{1, K_j}$  наибольшее значение. Проводя такие вычисления для каждой точки универсального множества, получим требуемые значения функции принадлежности.

Код каждого порождающего правила является самодостаточным, т.е. весь необходимый контекст активизации правила содержится только в его предпосылках. Нет способа вызова из одного правила другое как если бы правила были процедурами. Единственный способ сделать это – использовать метаправила. Они направляют ход измерений и вычислений атрибутов, но не принимают непосредственное участие в процессе их вычисления. Как показано в [Аверкин и др., 2015] для управления передачей баз правил по сети могут использоваться нечеткие продукционные правила, что позволяет считать их метаправилами и ввести в сеть два типа нечетких сенсоров: сенсоры-координаторы и сенсоры, соответствующие уровням представления данных. Следовательно, при управлении РБСС каждый пакет данных может сопровождаться капсулами с правилами для их обработки узлом. Таким образом, в сети существуют два вида уровня трафика и трафик знаний, причем трафик знаний ставит целью минимизацию трафика данных. Это требует использования специализированного нечеткий языка представления знаний.

## 5. Использование нечеткого сенсора для агрегации данных

Использование в информационных моделях обстановки лингвистических переменных предполагает: 1) представление первичной информации от датчиков в вербальной форме, например, *объект проникновения находится близко, над* и т.п.; 2) связывание первичных данных и лингвистических данных для определения идентичности или распознавание объектов, например, *объект проникновения танк, животное, солдат и т.п.* [Khalegi et al., 2013]. В последнем случае, используется числовая информация и информация, представленная в нечисловой форме. В этих двух случаях целесообразно использовать нечеткие сенсоры, которые вычисляют и передают лингвистические значения результатов агрегации и слияния численных измерений, причем могут быть использованы следующие два метода для агрегации базовых измерений. Первый из них осуществляет комбинацию существенных параметров с помощью выраженных связей между ними в виде правил. Во втором методе агрегация и слияние данных реализовано с помощью нечеткого разбиения числового многомерного пространства измерений [Lei et al., 2010]. Наилучший результат дает гибридизация этих двух подходов.

Слияние данных внутри сети может быть осуществлено иерархией узлов, между которыми динамически распределяются знания, необходимые для обработки данных в конкретном узле и в конкретной ситуации. Следующий пример иллюстрирует механизм вычисления значения лингвистического атрибута степень опасности от объекта проникновения на основе измерения параметров вибрации, акустики и магнитного поля путем агрегирования данных.

*R(1): ЕСЛИ (мощность сигнала сейсмодатчика – СРЕДНЯЯ И мощность акустического сигнала – ВЫСОКАЯ И сила сигнала магнитного датчика – ВЫСОКАЯ), ТО (опасность – ОЧЕНЬ ВЫСОКАЯ)*

*R(2): ЕСЛИ (мощность сигнала сейсмодатчика – НИЗКАЯ И мощность акустического сигнала – НИЗКАЯ И сила сигнала магнитного датчика – НИЗКАЯ), ТО (опасность – ОЧЕНЬ НИЗКАЯ)*

*R(3): ЕСЛИ (мощность сигнала сейсмодатчика – СРЕДНЯЯ И мощность акустического сигнала – СРЕДНЯЯ И сила сигнала магнитного датчика – НИЗКАЯ), ТО (опасность – НИЗКАЯ)*

На рис. 1 представлена диаграмма передачи данных и знаний по сети при обработке запроса «Сообщите ID ближайшего узла, вблизи которого находится объект проникновения, и соответствующую степень опасности».

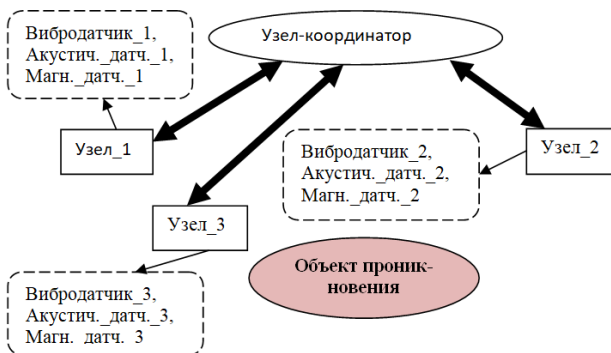


Рис. 1. Пример обработки запроса

Узел-координатор получил сообщение о фиксации тремя узлами факта нахождения объекта проникновения в их зоне ответственности. Очевидно, что выполнение запроса о его нахождении равносильно задаче поиска максимума в массиве (параметр степень опасности имеет числовой тип) для случая, когда его элементы распределены по узлам. Для решения задачи поиска узел-координатор запрашивает у ближайшего к нему узла ID и значение параметра степень опасности, запоминает его в переменной St0, формирует правило

**ЕСЛИ** ( $St0 < \text{степень опасности}$ ), **ТО** ( $St0 = \text{степень опасности}$ ,  $Id0=Id$ ) и посылает это правило в базу знаний второму ближайшему узлу и активизирует его. Полученные данные запоминаются, формируется правило для следующего узла и т.д. В результате формируется массив значений и будет установлено, что ближайший узел к объекту проникновения – это узел\_3. Очевидно, что узел-координатор должен иметь таблицу подчиненных ему узлов с соответствующими данными.

## **6. Обнаружение и локализация вторжения**

Одной из важнейших функциональностей РБСС является реализация двух задач: обнаружение объекта и локализация его в пространстве. Если, по крайней мере, три интеллектуальных узла (ИСУ) фиксируют это событие, то для локализации объекта вторжения можно использовать эффективный метод триангуляции для локализации объекта. Три сенсора с достаточной степенью точности локализуют объект. При наличии сигнатур возможных объектов вторжения и при известном законе изменения мощности трех указанных выше сигналов от узла от расстояния можно математически описать область локализации объекта. В этом случае вычисление положения объекта можно производить в одном ИСУ [Виноградов и др., 2022a].

Если сенсоры ИСУ рассчитаны на диапазон сигналов, то путем установки различных порогов можно обнаруживать и идентифицировать объекты с сигнатурами, характеризующимися отличающимися частотными диапазонами и мощностями спектров излучения.

При использовании в ИСУ TinyOS сенсор имеет возможность хранить массив, например, измерений акустического поля: частота излучения, мощность излучения, формируемые с заданным интервалом дискретизации (определяется точностью измерения).

Поскольку предполагается, что ИСУ будут активизироваться в сети по событиям, поток данных по сети будет направляться по нескольким ветвям сети от различных ИСУ из зоны вторжения. В связи с чем возникает задача (например, в пункте управления, или в головном узле кластера, или в обоих местах одновременно) обработки таких данных. Например, вторжение группы танков или бронемашин, или вторжение танков с мотопехотой и т.п. В этом случае от различных ИСУ, фиксирующих факт проникновения, будут поступать данные, предполагающие как слияние данных, так и слияние решений. При этом данные могут образовывать комбинацию зависимых (коррелированных) или независимых (некоррелированных) компонент. Корреляция будет свидетельствовать об обнаружении одного и того же объекта различными ИСУ. Тогда обработка данных будет включать алгоритмы триангуляции с последующим агрегированием

данных для уточнения сигнатуры объекта вторжения. В противном случае либо уточняются сигнатуры, либо фиксируется новый объект и выполняется его локализация.

Полученные и обработанные данные в пункте управления БСС используются для отображения ситуации вторжения и активизации задачи слежения за объектами вторжения. Алгоритм слежения включает в себя: идентификацию таким ИСУ, который близок к объекту вторжения (сбор данных об изменении положения объекта, локализацию зоны вторжения, оценку основных параметров объекта вторжения, формирование сигнатур объектов вторжения, моделирование сигнатур, сопоставление записей с модельными, прогнозирование и оценка возможной траектории, создание и обновление записей о траектории объекта).

На рис. 2 иллюстрирована процедура активизации узлов ближайших к узлу, обнаружившему факт проникновения объекта противника.

*R(1): ЕСЛИ (степень опасности > d), ТО (активизировать узлы согласно таблице связей с соседними узлами).*

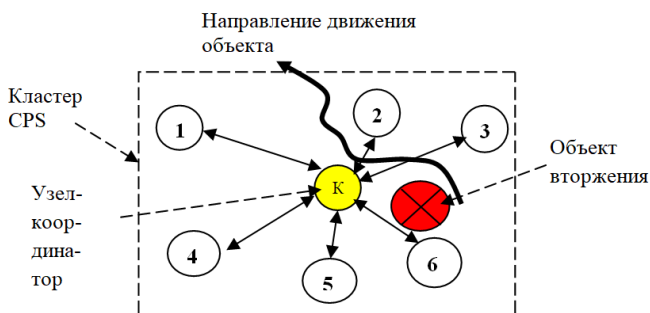


Рис. 2. Активизация ближайших узлов к узлу, зафиксировавшему факт проникновения в зону защиты

Пусть несколько ИСУ засекли цель и ее локализовали [Виноградов и др., 2022b]. Возникает проблема, какой из этих ИСУ будет ее сопровождать. Пусть выполняется в сети БСС отслеживание одной цели. Каждый объект вторжения в поле сенсора ИСУ генерирует изменяющееся во времени поле пространственной сигнатуры, которое воспринимается с использованием нескольких модальностей. Тогда движущийся объект – это пространственный пик в поле сигнатуры, который перемещается во времени. Следовательно, отслеживание цели – это отслеживание местоположения пика в течение времени. Для этого с помощью правила в примере 2 в поле БСС динамически определяется пространственная ячейка. Внутри ячейки узел ИСУ с максимальной степенью опасности координирует за-

дачи слежения. Можно сформулировать пять основных шагов алгоритма слежения: 1) ИСУ вблизи центра кластера потенциальных целевых траекторий должны быть активированы. ИСУ внутри кластеров должны взаимодействовать для определения факта присутствия цели в кластере; 2) Ближайший ИСУ должен быть активирован, начать передавать информацию об объекте (тип и т.п.), инициировать отслеживание данным ИСУ; 3) ИСУ передает в центр информацию: оценку местонахождения цели, направление движения, скорость. Вычисляется прогноз будущей позиции цели; 4) Эта информация передается в соседние ИСУ и кластеры (для облегчения слежения); 5) Если цель обнаружена в зоне ответственности предупрежденного ИСУ, этот ИСУ активируется и работа алгоритма повторяется с п. 1. Шаг траектории цели отображается в пункте управления.

### **Заключение**

Предложен достаточно простой способ представления знаний в виде нечетких продукций. Так как задание функций принадлежности нечетких лингвистических переменных, извлечения правил из данных представляет достаточно длительный процесс, то использование обучающей выборки, отражающей мнение эксперта о процессе измерения и поведения, может значительно его ускорить.

Ближайшей целью разработчиков является использование, помимо нечетких сенсоров, нейро-нечетких сетей (вложенных в один узел или распределенных по БСС), алгоритмов для управления адаптацией и обучением нечетких сенсоров. Это существенно повысит эффективность РБСС в решении для таких задач, как слежение за стационарными и мобильными объектами, мониторинг внешней среды. Описанный подход реализован в форме программного продукта для TinyOS. Он написан на языке программирования nesC – диалекте C, как набор взаимодействующих задач и процессов. Его дополнительные инструменты представлены в форме Java-приложений и интерфейсов для shell-скриптов. TinyOS содержит встроенные интерфейсы, модули и специфические настройки, которые позволяют создавать программу как набор модулей, реализующие заданные функциональности. Модули TinyOS обеспечивают интерфейс к стандартным аппаратным средствам ввода, вывода и датчикам.

### **Список литературы**

- [Аверкин и др., 2015] Аверкин А.Н., Лавров Г.К. Создание универсального интерфейса высокого уровня для умного узла беспроводной сенсорной сети // Системный анализ в науке и образовании: сетевое научное издание. – Дубна, 2015. – № 2. – [Электронный ресурс]. – URL: <http://sanse.ru/download/239>.

- [Виноградов и др., 2021]** Виноградов Г.П., Емцев А.С., Федотов И.С. Беспроводные сенсорные сети в защищаемых зонах // Известия ЮФУ. Технические науки. – 2021. – № 1..
- [Vinogradov et al., 2022]** Gennady P. Vinogradov, Igor A. Konyukhov. Patterns in Smart Wireless Sensor Network Nodes. International Scientific Conference “Information Technologies for Industry” (ITI’22). – Lecture Notes in Networks and Systems. Springer Nature Switzerland AG. – 2022.
- [Виноградов, 2022а]** Виноградов Г.П. Алгоритмы управления процессами в реагирующих сенсорных сетях для задач защиты объектов // Программные продукты и системы. – 2022. – № 2.
- [Виноградов, 2022б]** Виноградов Г.П. Паттерны в интеллектуальных системах управления автономными сущностями // Прикладная математика и вопросы управления. – 2022. – № 3.
- [Виноградов и др., 2022а]** Виноградов Г.П., Емцев А.С. Методы локализации и отслеживания целей в беспроводных сенсорных сетях: Сборник материалов XVII Всероссийской научно-практической конференции «Перспективные системы и задачи управления». – Таганрог: ИП Марук М.Р., 2022. – С. 52-65.
- [Виноградов и др., 2022б]** Виноградов Г.П., Конохов И.А.. Система управления FUZZY SMART MOT WSN // Труды 6-й Всероссийской Поспеловской конференции с международным участием «Гибридные и синергетические интеллектуальные системы» ГИСИС’2022. – Калининград, 2022.
- [Мальцева и др., 2018]** Мальцева Н.С., Зубова А.Д., Марышева И.Н. Анализ способов построения беспроводных сенсорных сетей // Инженерностроительный вестник Прикаспия: научно-технический журнал / Астраханский государственный архитектурностроительный университет. – Астрахань: ГАОУ АО ВО «АГАСУ», 2018. – № 2 (24). – С. 31-36.
- [Ефименко и др., 2018]** Ефименко М.С., Клымыв С.И Саткенов Р.Б. Беспроводные сенсорные сети. – Текст: непосредственный // Молодой ученый. – 2018. – № 51 (237). – С. 40-42. – URL: <https://moluch.ru/archive/237/55115/> (дата обращения: 07.08.2023).
- [Лихтциндер и др., 2020]** Лихтциндер Б.Я., Киричек Р.В., Федотов Е.Д. и др. Беспроводные сенсорные сети: учебное пособие для вузов. – М.: Горячая линия–Телеком, 2020. – 236 с.
- [Kumar et al., 2016]** Kumar S., Lobiyal D.K. An advanced DV-Hop localization algorithm for wireless sensor networks. Wirel // Pers. Commun. – 2013. – 71 (2). – P. 1365-1385.
- [Zhao et al., 2016]** Zhao Q.S., Hu Y.L. An improved DV-Hop localisation algorithm // Int. J. Wirel. Mob. Comput. – 2016. – 10 (1). – P. 20-25.
- [Huang et al., 2019]** Huang Z., Tang J., Shan G., Ni J., Chen Y. and Wang C. An efficient passenger-hunting recommendation framework with multi-task deep learning // IEEE Internet of Things Journal. – 2019. – DOI: <https://doi.org/10.1109/IIOT.2019.2901759>.

- [**Chen et al., 2019**] Chen Z., Zhang Y., Wu C., Ran B. Understanding individualization driving states via latent Dirichlet allocation model // IEEE Intell. Transp. Syst. Mag. – 2019. – 11 (2). – P. 41-53.
- [**Li et al., 2015**] Li X., Yan L., Pan W., et al. Optimization of DV-HOP localization algorithm in hybrid optical wireless sensor networks [J] // J. Heuristics. – 2015. – 21 (2). – P. 177-195.
- [**Kumar et al., 2014**] S. Kumar, D.K. Lobiyal, Power efficient range-free localization algorithm for wireless sensor networks // Int. J. Adv. Res. Comput. Sci. – 2014. – 20 (4). – P. 681-694.
- [**Khalegi et al., 2013**] Khalegi B., Khamis A., Karey F.O., Razavi S.N. Combining multi-sensory data: a review of the current state // Information. Unification. – 2013. – 14. – P. 28-44. – DOI: 10.1016/j.inffus.2011.08.001.
- [**Lei et al., 2010**] Lei A., Schober R. Merging multi-symbol solutions for mobile wireless sensor networks. Translated by IEEE. Wirel. Common. – 2010. – 9. – P. 778-790. – DOI: 10.1109/TWC.2010.02.090365.

## РАЗРАБОТКА И ИССЛЕДОВАНИЕ МЕТОДА РЕШЕНИЯ ЗАДАЧИ РАЗМЕЩЕНИЯ ЭЛЕМЕНТОВ ЭВА НА ОСНОВЕ ГИБРИДНОГО ПОДХОДА

Л.А. Гладков (*leo\_gladkov@mail.ru*)

М.Д. Ясир (*yasir\_82@mail.ru*)

Н.В. Гладкова (*nadyusha.gladkova77@mail.ru*)

Южный федеральный университет, Таганрог

В статье рассматривается задача размещения элементов цифровой электронной техники. Приведена постановка задачи, а также основные ограничения и критерий оптимизации. Предложен метод решения задачи размещения на основе гибридного подхода модель гибридного алгоритма решения задачи размещения. Разработана схема гибридного нечеткого генетического алгоритма. Предложено использовать островную модель параллельного многопопуляционного генетического алгоритма для организации вычислений. Разработана процедура, позволяющая реализовать принцип многопоточности на локальном уровне. Описаны принципы работы и структура блока нечеткого управления. Определены управляющие параметры генетического алгоритма. Выполнена программная реализация предложенного гибридного алгоритма. Проведены серии вычислительных экспериментов для определения эффективности разработанного алгоритма и выбора оптимальных значений управляющих параметров. Проведен анализ полученных данных и выработаны рекомендации по дальнейшему продолжению исследований.

**Ключевые слова:** автоматизированное проектирование, задача размещения, нечеткие генетические алгоритмы, нечеткое управление, гибридные модели, параллельные вычисления.

### Введение

Проектирование новых видов оборудования, приборов и их блоков других изделий представляет сложный процесс, включающий в себя разработку технического задания, анализ и синтез возможных вариантов решения поставленной задачи, разработку комплекта технической документации, необходимого для изготовления опытных образцов и последующего производства и эксплуатации объектов проектирования. При проекти-



ровании новой электронной аппаратуры важным моментом является повышение эффективности существующих методов автоматизированного проектирования, разработка новых гибридных подходов и алгоритмов, которые позволят создавать высоконадежные сверхбольшие интегральные схемы (СБИС) в короткие сроки и при сравнительно низких затратах. Стоимость кристаллов СБИС складывается из стоимости их проектирования и стоимости производства. Существующая тенденция к росту степени интеграции СБИС приводит к существенному увеличению трудоемкости процесса проектирования. Производство современных СБИС можно условно разделить на три этапа: проектирование, изготовление, тестирование. В виду значительной сложности ни один из этих этапов не может быть выполнен без средств автоматизированного проектирования. Одним из важнейших этапов проектирования СБИС является этап конструкторского (физического) проектирования, который включает в себя задачи: компоновки, размещения, трассировки и др. [Shervani, 1995].

Разработка топологии СБИС включает: определение взаимного расположения компонентов с минимальным числом пересечений; размещение элементов схемы в системе координат на рабочем поле с учетом схемотехнических, технологических, нормативных ограничений и требований по энергосбережению; трассировку (проведение внутрисхемных соединений). При проектировании топологии необходимо учитывать, что взаимное расположение компонентов оказывает большое влияние на динамические характеристики схемы. К задачам синтеза топологии относятся разработка библиотеки элементов и стандартных IP-блоков; декомпозиция электрической схемы; размещение элементов на кристалле; проведение соединений между эквипотенциальными выводами элементов [Charles et al., 2009].

В настоящее время эволюционные и биоинспирированные алгоритмы широко используются для решения различных задач проектирования и оптимизации [Гладков и др., 2009]. Благодаря своему разнообразию подходов к нахождению решений, эволюционные алгоритмы представляют собой весьма эффективное средство решения задач оптимизации и проектирования [Cohon et al., 2003]. Новым этапом развития теории эволюционного моделирования стали нечеткие генетические алгоритмы. Они основаны на интеграции методов эволюционного поиска, методов нечеткого управления и, в последнее время, нейронных сетей.

Одним из основных трендов современной науки является гибридизация [Ярушкина и др., 2007], т.е. создание новых систем, моделей и методов, подразумевающих интеграцию научных подходов различной физической природы. Особенно эффективно использование таких систем для решения сложных слабо формализованных задач, в том числе задач проектирования и оптимизации. Таким образом, эта тенденция естественным образом вписывается в общемировую.

В данной статье предлагается гибридный подход к решению задачи размещения элементов ЭВА на основе интеграции различных подходов, таких как, эволюционные алгоритмы поиска, нейронные сети, методы нечеткого управления параметрами алгоритма [Гладков, 2000].

## 1. Постановка задачи

В общем виде задача размещения заключается в определении оптимального значения заданного критерия, отражающего положение элементов и взаимосвязи между ними. При этом должны соблюдаться заданные конструктивно-технологические ограничения. Задачу размещения можно условно разбить на две части: размещение конструктивных элементов и оценка возможной длины связей между ними [Charles et al., 2009].

Возможно два варианта постановки задачи размещения стандартных ячеек: с постоянным и переменным расстоянием между рядами элементов. Актуальность задачи размещения с постоянным расстоянием между рядами обусловлена использованием иерархического стиля проектирования. В этом случае решается задача размещения отдельных фрагментов (блоков), а не всей схемы. В сравнении с первой постановкой задачи размещения вторая наиболее соответствует текущим техническим тенденциям. При такой постановке задачи размещение производится в условиях, когда расстояние между рядами, а также число и расположение допустимых посадочных мест являются фиксированными величинами.

При решении задачи размещения с фиксированным расстоянием между рядами процесс может оказаться очень сложным, если область, выделенная для размещения элементов ограничена. Поэтому необходим механизм обратной связи, позволяющий регулировать процесс размещения. Оценки времени прохождения сигнала и загрузки области размещения очень важны на ранних стадиях размещения и способны существенно улучшить производительность схемы.

С другой стороны, не менее важной является задача минимизации максимальной задержки прохождения сигнала в цепи. Алгоритмы размещения должны учитывать оценки качества последующей трассировки, иначе полученное решение ухудшит трассируемость и производительность схемы.

При таком подходе задача размещения сводится к нахождению оптимального положения элементов и внешних контактов. Исходными данными для задачи размещения являются схема соединения элементов, метрические параметры и топологические свойства рабочего поля.

Наиболее часто используются следующие критерии: суммарная длина всех соединений; длина самой длинной связи; оценка числа пересечений; равномерность распределения элементов и связей; тепловая и электрическая совместимость элементов и т.д.

Самым распространённым критерием является критерий суммарной длины соединений, поскольку его оптимизация положительно влияет на другие критерии: суммарная задержка сигналов, производительность схемы. В общем виде задача размещения может быть сформулирована следующим образом.

Пусть задано множество элементов  $E = (e_1, \dots, e_n)$ , отнесённых к некоторому ряду и множество возможных позиций  $P = (p_1, \dots, p_m)$  размещения элемента  $e_i$ . Задача решается в два этапа: задается начальное размещение, а затем алгоритм пытается его улучшить.

Первым в ряду размещается элемент  $e_i$ , имеющий максимальное число связей с другими элементами из множества  $E$ . Тогда на каждом последующем шаге алгоритма у нас будет также подмножество размещённых  $E_h$  и неразмещённых элементов  $E_k$  (рис. 1).

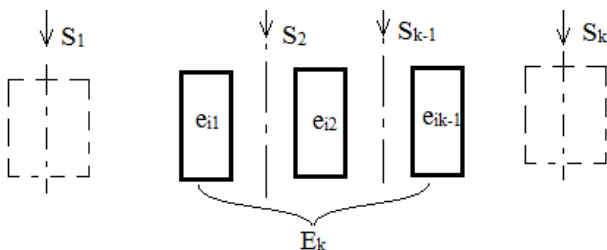


Рис. 1. Позиции для размещения очередной ячейки

Обозначим через  $V_{ik}$  множество связей, соединяющих элемент  $e_i$  и элементы из подмножества  $E_h$ . На  $k$ -м шаге вычисляется суммарная длина связей для множества  $V_{ik}$  при условии размещения элемента  $e_i$  в позицию  $p_j$  ( $j = 1, \dots, m$ ):

$$L_{ij} = \sum_{v_i \in V_{ik}} (x_{i \max} - x_{i \min}),$$

где  $x_{i \max}$  и  $x_{i \min}$  – координаты крайнего правого и крайнего левого выводов связи  $v_i$  при линейном расположении элементов.

Очередной шаг алгоритма состоит в выборе элемента  $e_i^*$  и позиции  $p_j^*$ , для которых справедливо:

$$L_{i^*j^*} = \max_{e_i \in E_k} \min_{p_j \in P_k} L_{ij}$$

После размещения всех элементов из множества  $E$  фиксируются их координаты в текущем ряду.

Данный алгоритм используется в комбинации с алгоритмами разбиения ячеек на ряды или блоки, при этом процессы формирования рядов (блоков) и размещения в них элементов выполняются параллельно.

В процессе улучшения начального размещения будет использоваться оценка качества возможной трассировки соединений, включающую в себя оценки возможного числа не выполненных связей и возможной суммарной длины соединений.

Тогда критерий оптимизации для данной задачи можно представить как нормированный аддитивный критерий:

$$F = k_1q_1 + k_2q_2,$$

где  $k_1, k_2$  – весовые коэффициенты, отражающие важность и влияние данного критерия на качество итогового решения;  $q_1$  – оценка возможного числа не построенных связей;  $q_2$  – оценка суммарной длины соединений между элементами схемы.

## 2. Описание гибридного алгоритма

Для решения поставленной задачи предложен гибридный алгоритм размещения на основе интеграции принципов генетического поиска и нечеткого управления [Michael et al., 1993]. Поиск и генерация новых вариантов решений осуществляет генетический алгоритм. Для повышения эффективности поиска были предложены новые модификации генетических операторов кроссинговера, мутации и др., а также разработана оригинальная методика кодирования альтернативных решений популяции. Одним из недостатков классического генетического алгоритма как метода оптимизации является его быстрая сходимость. Для преодоления данного недостатка в структуру генетического алгоритма вводится нечеткий логический контроллер (НЛК), который изменяет параметры генетического алгоритма в ходе эволюции, что позволяет избежать преждевременной сходимости [Herrera et al., 2003].

Наличие нечеткого логического контроллера является важной особенностью разработанного гибридного алгоритма. В процессе работы алгоритма, на вход НЛК поступают данные о текущем состоянии популяции, после чего происходит операция фазификации, т.е. преобразование входных данных к нечеткому виду, обработка их нечетким контроллером, оценка качества текущих решений. После чего НЛК на основе хранящихся в системе знаний и правил определяет управляющее воздействие и возвращает скорректированные значения контрольных параметров [Herrera et al., 1996].

В качестве управляющих параметров генетического алгоритма могут использоваться оценки разнообразия генотипа и фенотипа текущей популяции, динамика изменения лучшего и (или) среднего решения, а также динамика изменения значения функции пригодности популяции в целом.

В предложенном алгоритме для оценки текущего состояния были выбраны 4 параметра [Gladkov et al., 2019]:

$$e_1(t) = \frac{f_{ave}(t) - f_{best}(t)}{f_{ave}(t)}; \quad e_2(t) = \frac{f_{ave}(t) - f_{best}(t)}{f_{worst}(t) - f_{best}(t)};$$

$$e_3(t) = \frac{f_{best}(t) - f_{best}(t-1)}{f_{best}(t)}; \quad e_4(t) = \frac{f_{ave}(t) - f_{ave}(t-1)}{f_{ave}(t)},$$

где  $t$  – номер текущей популяции,  $f_{best}(t)$ ,  $f_{worst}(t)$ ,  $f_{ave}(t)$  – лучшее, худшее и среднее значение целевой функции на итерации  $t$  соответственно,  $f_{best}(t-1)$ ,  $f_{ave}(t-1)$  – лучшее и среднее значение ЦФ на итерации  $(t-1)$ .

Для построения оценки возможной суммарной длины соединений и числа не реализованных соединений используется волновой алгоритм. Работа алгоритма включает в себя три этапа: инициализацию, распространение волны и восстановление пути [Норенков, 2002]. Во время инициализации строится образ множества ячеек на рабочем поле, каждой ячейке приписываются атрибуты проходимости/непроходимости, запоминаются начальная и конечная ячейки.

Далее, от стартовой ячейки выполняется шаг к соседней ячейке, при этом проверяется, проходима ли она, и не была ли она ранее помечена. Для классификации соседних ячеек используется окрестность фон Неймана. Соседними ячейками считаются 4 ячейки по вертикали и горизонтали. При выполнении условий проходимости и непринадлежности её к ранее помеченным в процессе построения пути ячейкам, ячейке ставится в соответствие число, равное количеству шагов от начальной ячейки. Каждая ячейка, имеющая оценку числа шагов сама становится начальной на новой итерации и из неё порождаются очередные шаги в соседние ячейки. При такой последовательности шагов в конце либо будет построен путь от начальной до конечной ячейки, либо выяснится невозможность дальнейшего распространения текущей связи.

Восстановление кратчайшего пути происходит в обратном направлении: при выборе ячейки от конечной ячейки к начальной на каждом шаге выбирается ячейка, имеющая атрибут расстояния от стартовой на единицу меньше текущей ячейки. В качестве окончательного значения суммарной длины пути выбирается минимальное из полученных значений. В случае, если получено несколько одинаковых значений, выбор осуществляется произвольным образом. Структурная схема волнового алгоритма представлена на рис. 2.

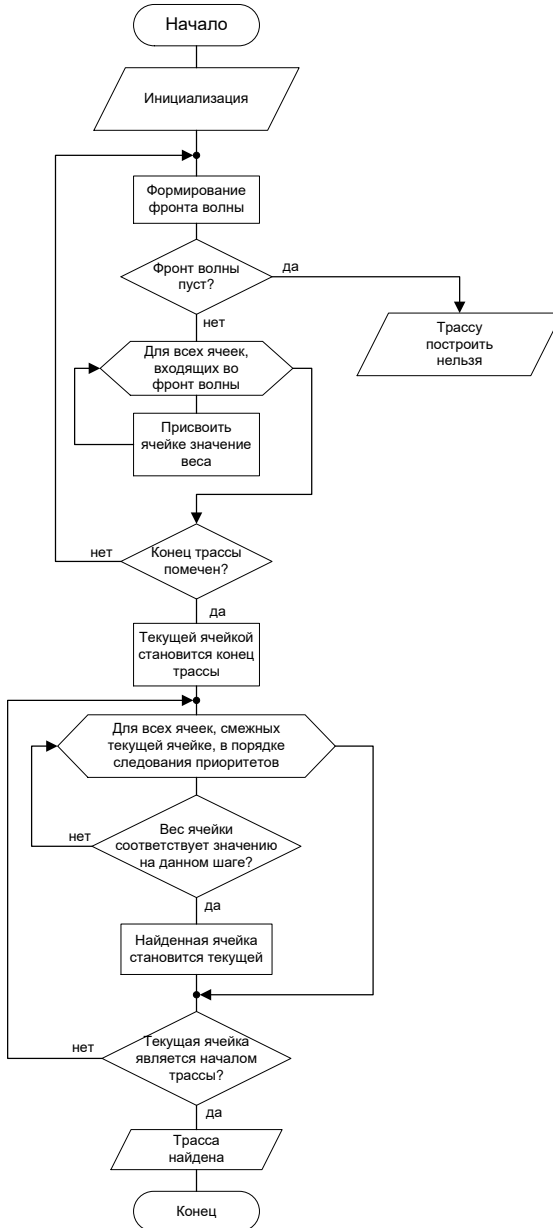


Рис. 2. Структурная схема волнового алгоритма

Одним из способов повышения производительности эволюционных алгоритмов является модель параллельных вычислений [Alba et al., 2002]. Для решения задачи размещения предложена модифицированная архитектура организации вычислительного процесса на основе использования островной модели параллельного многопопуляционного генетического алгоритма (рис. 3).

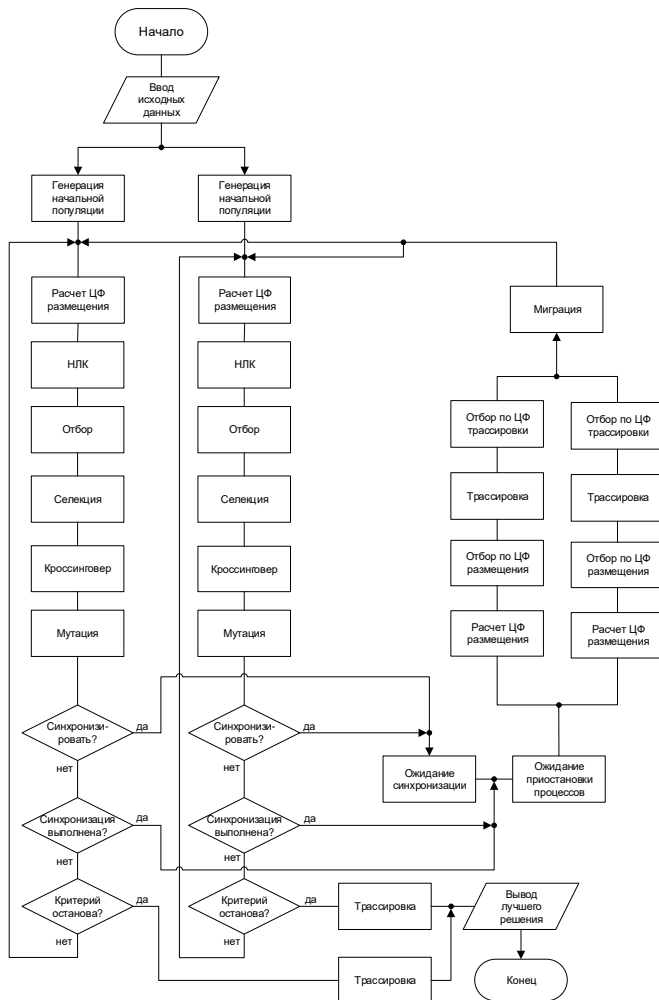


Рис. 3. Структурная схема островной модели параллельного генетического алгоритма размещения

В данной модели для синхронизации различных популяций в процессе эволюции используется оператор миграции. С помощью оператора миграции осуществляется обмен решения между популяциями [Zhongyang et al., 2006]. Для участия в обмене выбираются индивидуумы с наилучшим функцией приспособленности. После проведения обмена происходит редукция популяции для сохранения постоянного размера.

### 3. Программная реализация

При разработке программы, реализующей предложенный алгоритм, был использован фреймворк Qt 5.6. Он является полностью объектно-ориентированным, легко расширяемым и поддерживает возможности компонентного программирования.

Для хранения данных о топологии печатной платы используется LEF/DEF спецификация. LEF (Library Exchange Format) – это спецификация для представления физической структуры интегральной схемы в формате ASCII. Она включает правила оформления и абстрактную информации об элементах. LEF используется в сочетании с DEF (Design Exchange Format) спецификацией, которая используется для представления полного размещения элементов интегральной схемы.

В программе предусмотрена возможность гибкой настройки основных параметров гибридного алгоритма: первоначальные значения вероятностей генетических операторов, интенсивность миграции, число параллельных популяций, возможность включения (отключения) нечеткого управления для каждого параметра генетического алгоритма.

Для визуализации и анализа работы гибридного алгоритма размещения можно вывести на экран графики изменения входных и выходных параметров контроллера (рис. 4) [Gladkov et al., 2020].

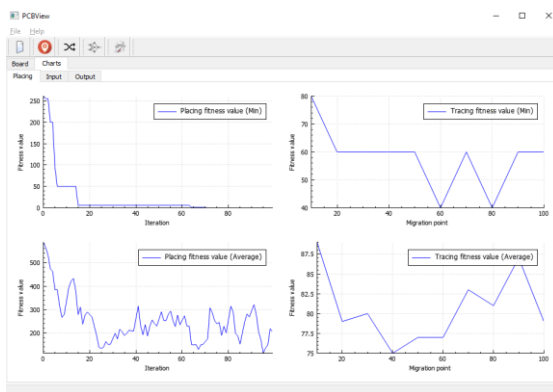


Рис. 4. Пример отображения графиков с использованием расширения *QCustomPlot*



Проведенные вычислительные эксперименты показали, что разработанный алгоритм имеет в среднем временную сложность  $VCA \approx O(n^2)$ .

Сравнение результатов работы алгоритма с использованием нечеткого нечеткого контроллера и без него показало, что, при одинаковом числе итераций, использование НЛК позволяет улучшить качество получаемых решений задачи в среднем на 25% (табл. 1).

Таблица 1

### Сравнение качества решений с использованием НЛК и него

	Без НЛК (N = 50)	С НЛК (N = 50)	Без НЛК (N = 100)	С НЛК (N=100)	Без НЛК (N=150)	С НЛК (N=150)
1	4585	3147	29658	21296	67953	48509
2	3870	3330	31145	23582	64311	51737
3	4245	2724	28192	23145	68989	50901
4	4056	3425	31632	23481	65576	50798
5	3774	2885	29761	21844	65184	48973
6	4896	2984	28487	23148	67925	49752
7	4129	2873	31845	22946	65427	52164
8	4812	3776	29145	21941	64964	48862
9	3981	3145	29411	22157	65817	50314
10	3876	3168	30491	22981	68482	50957
Среднее значение	4222,4	3145,7	29976,7	22652,1	66862,8	50296,7
Улучшение решения	25,6%		24,44%		24,78%	

При этом использование нечеткого контроллера практически не оказывает влияния на время выполнения алгоритма (увеличение времени выполнения составляет не более 1%).

### Заключение

В процессе работы была предложена структура гибридного алгоритма решения задачи размещения, предложена функция оценки и методика кодирования альтернативных решений, разработаны модифицированные генетические операторы, а также компоненты гибридного алгоритма. Предложена архитектура параллельного генетического алгоритма для повышения эффективности и быстродействия. Выполнена программная реализация и проводятся серии вычислительных экспериментов для определения оптимальных параметров алгоритма и изучения эффективности предложенного подхода. Перспективным направлением развития исследований можно считать введение в структуру гибридного алгоритма нейро-нечеткого блока для обучения нечеткого контроллера. Проведенные эксперименты показали, что эффективность контроллера повышается после введения блока обучения.

## Список литературы

- [**Shervani, 1995**] Shervani N. Algorithms for VLSI physical design automation. – USA, Kluwer Academy Publisher, 1995.
- [**Charles et al., 2009**] Charles J. Alpert, Dinesh P. Mehta, Sachin S. Sapatnekar. Handbook of algorithms for physical design automation. – CRC Press, New York, USA, 2009.
- [**Гладков и др., 2009**] Гладков Л.А., Курейчик В.М., Курейчик В.В., Сороколетов П.В. Биоинспирированные методы в оптимизации. – М.: Физматлит, 2009.
- [**Cohoon et al., 2003**] Cohoon J.P., Karro J., Lienig J. Evolutionary Algorithms for the Physical Design of VLSI Circuits // Advances in Evolutionary Computing: Theory and Applications. Ghosh A., Tsutsui S. (eds.). – Springer Verlag, London, 2003. – P. 683-712.
- [**Ярушкина и др., 2007**] Нечеткие гибридные системы. Теория и практика / под ред. Н.Г. Ярушкиной. – М.: Физматлит, 2007.
- [**Гладков, 2000**] Гладков Л.А. О некоторых подходах к построению гибридных интеллектуальных систем для решения графовых задач // Новости искусственного интеллекта. – 2000. – № 3. – С. 71-90.
- [**Michael et al., 1993**] Michael A., Takagi H. Dynamic control of genetic algorithms using fuzzy logic techniques // Proc. of the 5th International Conference on Genetic Algorithms. – Morgan Kaufmann, 1993. – P. 76-83.
- [**Herrera et al., 2003**] Herrera F., Lozano M. Fuzzy Adaptive Genetic Algorithms: design, taxonomy, and future directions // Soft Computing. – Springer-Verlag, 2003. – No. 7. – P. 545-562.
- [**Herrera et al., 1996**] Herrera F., Lozano M. Adaptation of genetic algorithm parameters based on fuzzy logic controllers / In: F. Herrera, J. L. Verdegay (eds.) // Genetic Algorithms and Soft Computing. – Physica-Verlag, Heidelberg, 1996. – P. 95-124.
- [**Gladkov et al., 2019**] Gladkov L.A., Gladkova N.V., Leiba S.N., Strakhov N.E. Development and research of the hybrid approach to the solution of optimization design problems. Advances in Intelligent Systems and Computing. Vol. 875 // International Conference on Intelligent Information Technologies for Industry IITI'18. – Vol. 2. – Springer Nature Switzerland AG, 2019. – P. 246-257.
- [**Норенков, 2002**] Норенков И.П. Основы автоматизированного проектирования: учебник для ВУЗов. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2002.
- [**Alba et al., 2002**] Alba E., Tomassini M. Parallelism and evolutionary algorithms // IEEE T. Evolut. Comput. – 2002. – Vol. 6. – P. 443-461.
- [**Zhongyang et al., 2006**] Zhongyang X., Zhang Y., Zhang L., Niu S. A parallel classification algorithm based on hybrid genetic algorithm // Proc. of the 6th World Congress on Intelligent Control and Automation, Dalian, China. – 2006. – P. 3237-3240.
- [**Gladkov et al., 2020**] Gladkov L.A., Gladkova N.V., Gusev N.Y., Semushina N.S. Integrated approach to the solution of computer-aided design problems // Proceedings of the 4th International Scientific Conference “Intelligent Information Technologies for Industry” (IITI'19). Advances in Intelligent Systems and Computing. – Vol. 1156. – Springer, Cham, 2020. – P. 465-476.

УДК 519.816

## НЕЧЕТКО-ДИНАМИЧЕСКИЙ МЕТОД АНАЛИЗА ГЛАВНЫХ КОМПОНЕНТ В СИСТЕМАХ ИНТЕЛЛЕКТУАЛЬНОГО МОНИТОРИНГА\*

А.И. Долгий  
АО «НИИАС», Москва

С.М. Ковалев (*ksm@rfniias.ru*)  
Ростовский филиал АО «НИИАС», Ростов-на-Дону

А.Е. Колоденкова (*anna82\_42@mail.ru*)  
Самарский государственный технический университет, Самара

В статье разработан новый подход к интеллектуальному мониторингу состояний сложных технологических объектов, к которым относятся слабо формализованные технологические процессы, сложные технические комплексы и устройства. В основу предлагаемого подхода положена цепочка преобразований между пространствами первичных параметров контролируемого объекта, обобщенных мета-параметров и лингвистических описаний технологических ситуаций и состояний контролируемого объекта. Центральным элементом в цепочке преобразований является предложенный авторами новый метод нечетко-динамического анализа главных компонент, основанный на объединении нелинейного анализа главных компонент и нечетких эволюционирующих систем. В силу естественной интерпретируемости переменных семантического пространства удается не только обнаруживать нештатные состояния и факты нарушений в поведении контролируемого объекта, но и идентифицировать типы, масштабы и направленность нарушений. Приводится модельный пример использования предлагаемой технологии интеллектуального мониторинга для оценки состояний и прогнозирования поведения технологических процессов, протекающих на железнодорожных станциях.

**Ключевые слова:** диагностирование сложных объектов, анализ главных компонент, множество параметров.

---

\* Работа выполнена при финансовой поддержке РФФ (проект № 23-29-00415).

## Введение

На современном этапе развития информационно-диагностических и мониторинговых технологий возникает ряд нерешенных проблем, среди которых важное место занимают задачи интеллектуального мониторинга, направленные на своевременное, а часто и заблаговременное получение достоверной информации о состоянии контролируемых объектов – технических систем и оборудования, производственных и технологических процессов (ТСТП). Раннее обнаружение нештатных состояний ТСТП, в то время, когда они еще обратимы, позволит снизить риски, избежать материальных потерь и повысить эффективность принятия управляющих и диагностических решений.

Большинство ТСТП относятся к классу слабо формализованных объектов, нелинейных по своей природе и не имеющих адекватных математических описаний, особенно в преддверии нештатных технологических ситуаций и состояний ТСТП. Вследствие этого значительная часть работ, посвященных вопросам мониторинга состояния ТСТП, базируется не на классических методах математического моделирования [Каштаева, 2020], а на статистических методах контроля и технологиях интеллектуального анализа данных [Роджерс, 2020]. К ним относятся методы многофакторного анализа, включая *PCA* и нелинейный *PCA*-анализ [Yu et al., 2022], нейросетевые и нечетко-логические технологии [Zhao et al., 2020], дискриминантный анализ Фишера [Chiang et al., 2000], [Md Nor N et al., 2020], методы опорных векторов, вероятностные подходы, включая байесовские сети [Hamza, 2022] и модели байесовского вывода [Yu et al., 2009], и многие другие.

За последние десятилетия интеллектуальные методы мониторинга и диагностики ТСТП, основанные на интеллектуальном анализе данных, привлекли значительное внимание как в исследовательской, так и в прикладных областях. При этом особый интерес вызвали технологии на основе анализа главных компонент (*PCA*-анализ).

В настоящей статье предлагается новый подход к диагностированию и мониторингу сложных объектов ТСТП на основе объединения динамического *PCA*-анализа и технологий эволюционирующих нечетких систем.

### 1. Лине́йный *PCA*-анализ

Известный *PCA*-анализ основан на ортогональном разложении ковариационной матрицы переменных процесса по направлениям, имеющих максимальные вариации данных. Основная цель использования *PCA* состоит в том, чтобы найти множество факторов, имеющих гораздо меньшую размерность, чем исходный набор данных, и способных правильно описать основные тенденции в исходном наборе данных.

Пусть  $p$  – количество измеряемых переменных;  $X$  –  $(n \times p)$  – матрица, представляющая нормированные измерения;  $\Sigma$  – ковариационная  $(p \times p)$  – матрица для  $X$ .

Строками  $X$  являются  $p$ -мерные векторы, соответствующие наблюдениям, а столбцами –  $n$ -мерные векторы, соответствующие переменным. Известно, что матрица  $\Sigma$  может быть преобразована в диагональную матрицу  $L$  с помощью ортонормированной  $(p \times p)$  – матрицы  $U$ , а именно,  $\Sigma = ULU^T$ . Столбцы  $u_1, u_2, \dots, u_p$  матрицы  $U$  известны как вектора загрузки главных компонент. Диагональные элементы  $\lambda_1, \lambda_2, \dots, \lambda_p$  матрицы  $L$  являются упорядоченными собственными значениями матрицы  $\Sigma$ . Они представляют дисперсии переменных  $x_j$  в новом ортонормированном базисе, определяемом собственными векторами матрицы  $\Sigma$ .

Использование *PCA*-анализа в методах мониторинга и диагностики реализуется в несколько этапов. Вначале на основе множества данных  $X$ , соответствующих нормальному состоянию ТСТП, с использованием *PCA*-анализа формируется минимальный набор новых оценочных переменных  $\theta_1, \theta_2, \dots, \theta_p$ , описывающих основные тенденции в  $X$ . Затем формируется статистика вариаций переменных в пространстве  $X_{pc}$  или  $X_{rez}$  для целей диагностики.

## 2. Динамический *PCA*-анализ

Основным инструментом методов мониторинга и диагностики является динамический *PCA*-анализ [Park et al., 2020], ориентированный на обработку потоковых данных, представленных в виде многомерных временных рядов  $X = \{x_i\}$ . Элементами временного ряда  $X$  являются вектора числовых значений  $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$ , описывающих параметры контролируемого объекта в  $i$ -й момент времени.

В основу разработки динамического *PCA*-анализа положен принцип «скользящего окна» – трансформация исходного потока данных  $\{x_i\}$ , представленного в виде многомерного временного ряда, в последовательность коротких наблюдений – временных рядов  $\{x_i, x_{i+1}, \dots, x_{i+s-1}\}$ . Последняя является окном анализа размерностью  $s$ , зависящей специфики задачи. Поступающий на вход динамического *PCA*-анализа поток данных преобразуется в матрицу наблюдений за потоком данных  $x(t)$  с окном анализа  $s$ .

$$X_s = \begin{bmatrix} x_1^T & x_2^T & \dots & x_s^T \\ x_2^T & x_3^T & \dots & x_{s+1}^T \\ \dots & \dots & \dots & \dots \\ x_N^T & x_{N+1}^T & \dots & x_{N+s-1}^T \end{bmatrix} \in R^{N \times sm},$$

где  $x_i \in R^m$ ,  $(i = 1, \dots, N+s-1)$  обозначает нормализованный вектор наблюдений из  $m$  параметров ТСТП.

В основу динамического *РСА*-анализа положены выражения на основании которых осуществляется преобразование входного потока данных  $X$  в новое редуцированное пространство переменных  $Z_{pc}$ . Благодаря тому, что размерность нового пространства переменных намного меньше размерности исходного потока данных, то существенно упрощается исследование процессов средствами дескриптивного анализа. Кроме того, вновь образованные на основе динамического *РСА*-анализа пространства переменных  $Z_{pc}$ ,  $Z_{rez}$  дают возможность на их основе формирование нескольких важных статистик для переменных  $z \in Z_{pc}$ , и  $z \in Z_{rez}$  в целях интеллектуального мониторинга.

### 3. Статистики динамического *РСА*-анализа

Для обнаружения и идентификации нарушений на основе технологий динамического *РСА*-анализа используются диагностические статистики – вероятностные распределения целевых параметров контролируемого объекта. На практике наиболее часто используются две статистики:  $T^2$  – статистика, контролирующая отклонения параметров  $x_i$  в пространстве главных компонент  $Z_{pc}$  и  $SPE$  – статистика, контролирующая отклонения параметров  $x_i$  в пространстве невязок  $Z_{rez}$ . Статистики  $T^2$  и  $SPE$  формируются на данных, полученных в нормальных состояниях контролируемого ТСТП. При превышении значениями  $x_i$  пороговых оценок хотя бы для одной из статистик принимается гипотеза о возникновении нештатной ситуации или нештатного состояния ТСТП.

В нормальных условиях функционирования ТСТП матрица главных компонент  $P_{pc}$  в силу свойств *РСА*-разложения несет основную информацию о процессе, а, следовательно, при стабильных параметрах ТСТП, находящегося в нормальном состоянии, значения новых параметров  $z_{pc}$ , извлеченных путем *РСА*-редукции из исходных параметров  $x$ , подчиняются нормальному закону распределения. Для переменных  $z_{rez}$  в остаточном пространстве невязок  $Z_{rez}$  также имеет место нормальный закон распределения.

Для оперативно поступающих на вход мониторинговой системы данных  $x \in R^m$  критерии  $T^2$  и  $SPE$  рассчитываются в режиме *on-line*.

### 4. Общий подход к нечетко-динамическому *РСА*-анализу

В основу нечетко-динамического *РСА*-анализа положен двух этапный процесс реализации прямой и обратной цепочки преобразований между первичными и обобщенными параметрами ТСТП.

На первом этапе на основе динамического *РСА*-анализа осуществляется цепочка преобразований первичных параметров ТСТП в новое семантическое пространство

$$J^k = \{(\mu_{r_1}, \mu_{r_1}, \dots, \mu_{r_k}) \mid r_i \in KB, KB \in eTS\}$$

активаций нечетких правил специальной модели, принадлежащей классу эволюционирующих нечетких системы (*eTS*). Нечеткие правила *eTS* представляют собой лингвистические описания причинных связей между первичными параметрами ТСТП, с одной стороны, и состояниями ТСТП либо целевыми переменными, сопровождающими поведение контролируемого объекта. На втором этапе на основе динамического *РСА*-анализа производится редукция пространства активаций нечетких правил  $J^k$  в ограниченный набор обобщенных мета-параметров  $\omega \in \Omega$ . Далее в пространстве  $\Omega$  формируются статистики изменений мета-параметров  $\omega$  для нормальных состояний ТСТП и определяются статистические критерии для диагностики нарушений. На основе сформированных критериев выявляются отклонения обобщенных параметров  $\omega$  от их пороговых значений, обозначающих границы нормальных режимов функционирования объекта, и выявляются факты нарушений или появления нештатных состояний.

На втором этапе осуществляется обратное преобразование обобщенных мета-параметров  $\omega$  в семантическое пространство  $J^k$  активаций нечетких правил  $r_i \in KB_i$ , после чего методом взятия первых производных от функций-критериев по переменным  $r_i$  определяются оценки влияния нечетких правил на значения статистических критериев. Учитывая, что нечеткие правила (переменные пространства  $J^k$ ) являются интерпретируемыми объектами, можно понятным образом идентифицировать и описать конкретные типы нарушений или нештатных состояний путем выявления наиболее «влиятельных» нечетких правил.

Рассмотрим реализацию данных этапов.

#### 4.1 Общее представление eTS

Данный класс моделей базируется на нечетко-продукционных правилах вида:

$$\begin{aligned} r_i : IF \ x_1 \ IS \ \alpha_{i1} \ AND \ x_2 \ IS \ \alpha_{i2} \ AND \dots \ AND \ x_n \ IS \ \alpha_{in} \\ THEN \ C_i = a_{i0} + a_{i1}x_1 + \dots + a_{in}x_n \end{aligned} \quad , \quad (1)$$

где  $x_i = (x_1, x_2, \dots, x_n)$  – вектор входных переменных;  $\alpha_{ij}$  – лингвистические значения переменных  $x_i$ ,  $\alpha_{ij}$  – параметры консеквента.

Нечеткие термы, входящие в антецедент нечеткого правила (1), описываются Гауссовыми функциями принадлежности.

В рассматриваемом классе моделей нечеткое правило (1) выступает в качестве модели эмпирического знания о причинной связи между технологической ситуацией, описанной в антецеденте нечеткого правила (1), и значением целевой переменной  $C_i$ , указанной в консеквенте нечеткого правила. В качестве целевой переменной могут выступать прогнозируемые параметры ТСТП и/или классы состояний ТСТП вместе с их вероятностными оценками.

## 4.2. Обучение eTS-модели

*eTS*-модель относится к классу адаптивных эволюционирующих систем, основанных на гибридизации нечетких *TS*-моделей с методами машинного *on-line* обучения. Механизм *on-line* обучения придает *eTS*-модели способность оперативно в пошаговом (инкрементном) режиме подстраиваться под изменяющийся входной поток данных. Подстройка *eTS*-модели осуществляется на основе инкрементного алгоритма обучения под управлением потоковых данных [Kovalev et al., 2020], реализованного с использованием принципов субтрактивной кластеризации [Yager et al., 1994].

На практике для оценки потерь часто используется безразмерный индекс ошибок (*NDEI*) [Kasabov et al., 2002], определяемый как отношение среднеквадратичной ошибки к стандартному отклонению целевой переменной:

$$NDEI = \frac{RMSE}{\sigma_y}.$$

## 4.3. Функционирование eTS и формализация состояний ТСТП

Информация для обучения *eTS* накапливается в режиме *on-line* по мере поступления новых данных. Каждый образец  $z_i = (x_i, y_i)$  характеризует некоторую новую информацию о текущем состоянии ТСТП, представленную точкой  $z_i$  в признаковом пространстве числовых параметров  $Z = X^{n+1}$ .

По мере эволюционирования *eTS*-модель приобретает, так называемое, свойство структурной чувствительности – способность реагировать на дрейфы в потоке данных синхронными изменениями активаций нечетких правил. Эффективность диагностирования на основе нечетко-динамического *РСА*-анализа во многом зависит от структурной чувствительности *eTS*-модели, а именно, с повышением структурной чувствительности увеличивается коррелированность нечетких правил  $r_i \in KB_i$  и, как следствие, эффективность сжатия данных.

## 5. Обнаружение и идентификация нарушений

Возникновение нарушений или нештатных состояний в поведении ТСТП обычно приводит к изменениям статистик: либо *SPE*, либо в  $T^2$ , либо обоих вместе. В этом случае статистики *SPE* и/или  $T^2$  превышают пороговые оценки (контрольные пределы). Идентификация и реконструкция нарушений является более сложной задачей, которая должна быть сформулирована в терминах обоих показателей. В данной статье используется подход на основе взятия первой производной функционала статистического критерия по обобщенным параметрам, полученным в результате редукции семантического пространства нечетких правил.



Интерпретируемость нечетких правил позволяет организовать прозрачный механизм реконструкции нарушений и идентификации типов нестандартных состояний.

### 5.1. Обнаружение нарушений на основе диагностических статистик

Пусть  $X$  – матрица исходных данных (наблюдений) размера  $n \times m$ , в которой строками являются отдельные выборки данных, а столбцами – переменные. Предполагается, что значения в  $X$  центрированы к нулевому среднему значению.  $PCA$  разлагает  $X$  на две компоненты:

$$X = TP^T + E = TP^T + \tilde{T}\tilde{P}^T,$$

где  $T \in R^{n \times I}$  и  $P \in R^{m \times I}$  – матрицы оценок и нагрузок соответственно, а  $I$  – количество главных компонент, сохраненных в модели.

Матрица остатков  $E$ , в свою очередь, может быть представлена в виде произведения остаточных матриц оценок  $\tilde{T}$  и нагрузок  $\tilde{P}$ . Область значений  $P$  является подпространством главных компонент  $P_{pc}$  размерностью  $I$ , а область значений  $\tilde{P}$  образует остаточное подпространство  $P_{res}$  размерностью  $m - I$ .  $P$  и  $\tilde{P}$  также можно интерпретировать как собственные вектора ковариационной матрицы.

### 5.2. Идентификация нарушений

Идентификация нарушений требует дополнительного анализа соответствующей информации о компонентах разложения  $PCA$ . Рассмотрим подход к идентификации нарушений на основе обратной цепочки преобразований обобщенных мета-параметров  $\omega \in \Omega$  в пространство активаций нечетких правил  $J^k$  с использованием обратного  $PCA$ -преобразования и последующей интерпретации влияния нечетких правил.

$T^2$ -статистика для пространства главных компонент  $\Omega$ , полученного на основе нечетко-динамического  $PCA$ -анализа из семантического пространства активаций нечетких правил  $J^k$ , может быть преобразована следующим образом:

$$T^2 = \omega^T P \Delta P^T \omega = \omega^T M \omega, \quad (2)$$

где  $\omega$  – обобщенные мета-параметры;  $M = P \Delta P^T$  – матрица преобразований.

Из (2) можно выразить отдельно каждую переменную  $\omega_i \in \omega$ :

$$\omega_i = \sum_1^n \omega_j m_{ji} \omega_j = \sum_1^n \omega_j^2 m_{ji}.$$

Путем взятия первой производной  $T^2$  по переменной  $x_i$  получим:

$$\frac{dT^2}{d\omega_i} = \frac{d}{d\omega_i} \sum_1^n \omega_j^2 m_{ji} = 2 \sum_1^n \omega_j m_{ji}. \quad (3)$$

Величина  $\gamma_i = \omega_j m_{ji}$  является индексом влияния, характеризующим вклад каждой переменной  $\omega_i \in \omega$  в критерий  $T^2$ .

Обратное преобразование индексов влияния обобщенных параметров  $\omega$  в пространство активаций нечетких правил  $J^k$  осуществляется на основании выражения:

$$[\mu_{r_1}(z_t), \mu_{r_2}(z_t), \dots, \mu_{r_k}(z_t)] = \gamma^T P^T, \quad (4)$$

где  $\gamma$  – вектор индексов влияния обобщенных параметров  $\omega \in \Omega$ ,  $P$  – матрица PCA-преобразования (собственных векторов) пространства  $J^k$  в пространство  $\Omega$ .

Путем реализации цепочки преобразований (2)-(4) можно получить оценки влияния нечетких правил (являющихся переменными в семантическом пространстве  $J^k$ ) на величину критерия  $T^2$ .

Таким образом, на основе цепочки обратных преобразований индексов влияния мета-параметров  $\omega \in \Omega$  в семантическое пространство  $J^k$  активаций нечетких правил  $r_i$  удастся выявить нечеткое правило  $r_i^*$ , оказывающее наибольшее влияние на оценку диагностического  $T^2$ -критерия.

## 6. Экспериментальное исследование

Для оценки возможности и эффективности применения разработанной технологии нечетко-динамического PCA-анализа к интеллектуальному мониторингу была проведена серия экспериментов с реальными и синтетическими данными, описывающими технологический процесс (ТП) обработки вагонопотоков на сортировочной станции.

Начальные входные данные для эксперимента генерировались в *on-line* режиме на основе имитационной модели, разработанной в среде *MATLAB*, генерирующей потоки данных о поведении ТП в разных его состояниях ТП. Поток данных представляется временным рядом параметров ТП:

$$\{x_t^T\} x_t^T = \left( \tau, N, \frac{dN}{d\tau} \right), \quad (5)$$

Одновременно с имитацией ТП осуществлялся нечетко-динамический PCA-анализ потоковых данных, реализация основных этапов которого рассматривается ниже.

1. Первая серия данных, описывающих поведение ТП в разных состояниях, генерировалась на основе имитационной модели. Одновременно с имитацией потока данных формировалась *eTS*-модель с использованием

алгоритма инкрементной кластеризации. В результате эволюционирования *eTS* была сформирована достаточно устойчивая база знаний, содержащая следующие три нечетких правила:

$r1$ : If ( $\tau = Med$ ) and ( $N = Sm$ ) and ( $\frac{dN}{d\tau} = -Big$ ) THEN  $St = 1$  or  $St = 2$ ,

$r2$ : If ( $\tau = Sm$ ) and ( $N = Med$ ) and ( $\frac{dN}{d\tau} = 0$  or  $\frac{dN}{d\tau} = -Sm$ ) THEN  $St = 3$ ,

$r3$ : If ( $\tau = Sm$ ) and ( $N = Big$ ) and ( $\frac{dN}{d\tau} = Big$ ) THEN  $St = 4$ .

Здесь *Sm*, *Med* и *Big* обозначают нечеткие термы соответственно: Малое, Среднее, Большое.

Одновременно с генерацией первой серии данных осуществлялся нечетко-динамический *PCA*-анализ потока данных, в результате чего были получены следующие промежуточные результаты:

1. было сформировано семантическое пространство активаций нечетких правил  $J^3 = \mu_{r1}(x_t) \times \mu_{r2}(x_t) \times \mu_{rk}(x_t)$ ;

2. на основе нечетко-динамического *PCA*-анализа были сформированы матрица собственных векторов-столбцов *P*, вектор собственных значений  $[\lambda_1, \lambda_2, \lambda_3]$  и матрица оценок *T*.

На основе первого столбца оценочной матрицы *T*, (соответствующего максимальному собственному значению) был сформирован обобщенный мета-параметр  $\omega$ , а на основе остальных столбцов (второй) – параметр невязок  $\nu$  (переменная в пространстве невязок  $P_{rez}$ ).

II. Вторая серия экспериментов проводилась на реальных данных. Мониторинг ТП осуществлялся путем нечетко-динамического *PCA*-анализа входного потока в режиме *on-line* и отслеживания значений обобщенного мета-параметра  $\omega$  вместе с параметром невязок  $\nu$  с целью обнаружения и идентификации нештатных состояний.

Решались три разные задачи. Первая заключалась в выявлении предикторных состояний ТП на основе идентификации перехода  $St_1 \rightarrow St_2$  с использованием первой нулевой точки параметра невязок  $\nu$  на шкале  $\omega$ . Вторая задача заключалась в обнаружении нештатного состояния ТП в начале его развития путем фиксации второй нулевой точки параметра невязок  $\nu$  на шкале  $\omega$ . Третья задача заключалась в прогнозировании нештатного состояния не менее чем за 7 временных тактов с использованием нулевой точки обобщенного параметра  $\omega$  на шкале  $\omega$ . Серии включали по 300 испытаний. В экспериментах участвовали три мониторинговых модели: стандартная модель на основе линейного *PCA*-анализа; модель динамического *PCA*-анализа; разработанная нечетко-динамическая *PCA*-

модель. Одновременно оценивалась погрешность прогнозирования целевой переменной  $\tau$  на основе точностного критерия в виде безразмерного индекса ошибок *NDEI*. Сводная таблица результатов приведена ниже.

Таблица 1

Тип модели	Ошибки обнаружения предикторов. Нештатных состояний (%)	Ошибки обнаружения нештатных состояний (%)	Ошибки прогнозирования нештатных состояний (%)	NDEI $\tau$
Линейная <i>РСА</i> -модель	12.5	11	11.5	0.8
Динамическая <i>РСА</i> -модель	9.3	9.3	8	0.75
Нечетко-динамическая <i>РСА</i> -модель	7	4.5	3.75	0.35

Приведенные результаты экспериментов показывают убедительное преимущество разработанной нечетко-динамической *РСА*-технологии в решении мониторинговых задач перед существующими моделями на основе *РСА*-анализа.

### Заключение

В статье разработан новый подход к интеллектуальному мониторингу состояний сложных технологических объектов и процессов. В основу предлагаемого подхода положена цепочка преобразований между пространствами первичных параметров контролируемого объекта, обобщенных мета-параметров и лингвистических описаний технологических ситуаций и состояний контролируемого объекта.

Достоинствами предлагаемого подхода являются:

1) Интерпретируемость нечетко-динамического *РСА* –преобразования позволяет не только осуществлять редукцию переменных и диагностировать в «сжатом» пространстве факты нарушений, но и путем обратного преобразования идентифицировать и понятным образом описывать типы нарушений, масштабы и их направленность.

2) Возможность обработки потоковых данных на основе объединения динамического *РСА*-анализа и эволюционирующих нечетких систем, что позволяет адаптировать технологию мониторинга под медленные дрейфы данных.

3) За счет интерпретируемости эволюционирующей модели, являющихся одним из ключевых элементов предлагаемой технологии, удастся не только осуществлять редукцию исходных многомерных данных, но и извлекать знания о состояниях контролируемого объекта и причинах их изменений.

### Список литературы

- [Каштаева, 2020] Каштаева С.В. Математическое моделирование. – 2020.
- [Роджерс, 2020] Роджерс С. Использование облачных технологий для ускорения интеллектуального анализа данных // Control Engineering Россия. – 2020. – № 1.
- [Chiang et al., 2000] Chiang L.H., Russell E.L., Braatz R.D. Fault diagnosis in chemical processes using Fisher Discriminant Analysis, discriminant Partial Least Squares, and Principal Component Analysis // Chemometrics and Intelligent Laboratory Systems. – 2000. – Vol. 50, No. 2.
- [Hamza, 2022] Hamza Z. A Review of Bayesian Networks Applications for Electrical Systems // Recent Advances in Electrical & Electronic Engineering (Formerly Recent Patents on Electrical & Electronic Engineering). – 2022. – Vol. 15.
- [Kasabov et al., 2002] Kasabov N.K., Song Q. DENFIS: dynamic evolving neural-fuzzy inference system and its application for time-series prediction // IEEE transactions on Fuzzy Systems. – 2002. – Vol. 10.
- [Kovalev et al., 2020] Kovalev S., Kolodenkova A., Sukhanov A. Incremental Structure-Evolving Intelligent Systems with Advanced Interpretational Properties // Artificial Intelligence. RCAI 2020. Lecture Notes in Computer Science. – 2020. – Vol. 12412.
- [Md Nor N. et al., 2020] Md Nor N., Che Hassan C.R., Hussain M.A. A review of data-driven fault detection and diagnosis methods: Applications in chemical process systems // Reviews in Chemical Engineering. – 2020. – Vol. 36, No. 4.
- [Park et al., 2020] Park Y.J., Fan S.K.S., Hsu C.Y. A review on fault detection and process diagnostics in industrial processes // Processes. – 2020. – Vol. 8, No. 9.
- [Yager et al., 1994] Yager R.R., Filev D.P. Approximate clustering via the mountain method // IEEE Transactions on systems, man, and Cybernetics. – 1994. – Vol. 24, No. 8.
- [Yu et al., 2022] Yu E. et al. A multigroup fault detection and diagnosis framework for large-scale industrial systems using nonlinear multivariate analysis // Expert Systems with Applications. – 2022. – Vol. 206.
- [Yu et al., 2009] Yu J., Qin S.J. Multiway gaussian mixture model based multiphase batch process monitoring // Industrial & Engineering Chemistry Research. – 2009. – Vol. 48, No. 18.
- [Zhao et al., 2020] Zhao Y. et al. A review of data mining technologies in building energy systems: Load prediction, pattern identification, fault detection and diagnosis // Energy and Built Environment. – 2020. – Vol. 1, No. 2.

УДК 004.896

## ОПТИМИЗАЦИЯ МЕТОДОМ РОЯ ТРАНСФОРМИРУЮЩИХСЯ ХРОМОСОМ

Б.К. Лебедев (*lebedev.b.k@gmail.com*)

О.Б. Лебедев (*lebedev.ob@mail.ru*)

Южный федеральный университет, Таганрог

М.А. Ганжур (*mganzhur@yandex.ru*)

Донской государственный технический университет,  
Ростов-на-Дону

Предложена новая стратегия поиска эффективных решений на основе модификации и гибридизации канонических представлений алгоритма роя частиц. Описывается парадигма роя трансформирующихся хромосом, обеспечивающая возможность представления решений в аффинном пространстве хромосом с целочисленными значениями параметров. Рассмотрены механизмы трансформации хромосом в аффинном пространстве для увеличения веса аффинных связей. Обновленные после трансформации хромосомы являются, в свою очередь, базовыми точками в последующих трансформациях (непрерывная цепочка постоянных перерождений в соответствии с принципом сансара). Общая оценка временной сложности лежит в пределах  $O(n^2)$ - $O(n^3)$ .

**Ключевые слова:** рой трансформирующихся хромосом, статическая популяция, стайное поведение, аффинное пространство, структуры хромосом, вес аффинных связей, трансформация, оператор направленной мутации, оптимизация.

### Введение

Разработка метода и средств оптимизации принятия решений осуществлялась на основе общего концептуального подхода, базирующегося на применении методов искусственного интеллекта, инспирированных природными системами, сочетании принципов адаптации, самообучения, самоорганизации, использовании моделей биологических адаптивных систем. В состав общей концепции включена концепция построения архитектуры популяционных алгоритмов на основе метаэвристик природных вычислений. Авторы предлагают новые стратегии поиска эффективных решений на основе модификаций канонических представлений.

Метод роя частиц (*Particle Swarm Optimization, PSO*) является методом стохастической оптимизации схожий с эволюционными алгоритмами. Этот метод моделирует роевое и стайное поведение животных [Clerc, 2006]. В отличие от канонических популяционных методов *PSO* работает с одной статической популяцией, члены которой постепенно с появлением информации о пространстве поиска улучшаются. Данный метод представляет собой вид направленной мутации (*directed mutation*). Решения в *PSO* мутируют в направлении наилучших найденных решений. Частицы никогда не умирают (так как нет селекции).

В канонической парадигме *PSO* [Kennedy et al., 1995] многомерное, вещественное, метрическое пространство поиска  $X$  населяется перемещающимися в том пространстве роем частиц  $P = \{p_i / i = 1, 2, \dots, n\}$ . Каждая частица  $p_i$  размещается в текущей позиции  $x_i$ , связана и может взаимодействовать со всеми частицами роя, она тяготеет к лучшему решению роя. По аналогии с эволюционными стратегиями, рой частиц можно трактовать как популяцию, а частицу как индивида (хромосому).

Каноническая парадигма роя частиц предусматривает использование вещественных значений параметров в многомерных, вещественных, метрических пространствах. Однако в большинстве генетических алгоритмов гены в хромосомах имеют целочисленные значения. В свою очередь хромосомы являются некоторыми интерпретациями решений, которые трансформируются в решения путем декодирования хромосом.

Анализ существующих методов и алгоритмов решения комбинаторных задач показал, что в качестве структуры данных, несущих информацию о решении, чаще всего используются списки, фактически являющимися интерпретациями решений. Данное представление удобно для его использования в различных метаэвристиках (генетические алгоритмы, муравьиные алгоритмы) [Карпенко, 2016].

На списки, используемые в качестве интерпретаций решений, в зависимости от специфики задачи, накладываются ограничения. Существенную роль в общем процессе нахождения решения играет декодер, осуществляющий трансформацию от списка к решению. Важной характеристикой декодера является его способность получить оптимальное решение по заранее известному оптимальному коду (приоритетному списку), т.е. способность правильно декодировать.

В отличие от канонической парадигмы роя частиц гибридные алгоритмы в качестве моделей для представления решений используют широкий диапазон графовых структур (маршрут, дерево, двудольный граф, паросочетание, внутренне-устойчивое множество и т.д.). Такой подход является эффективным способом поиска рациональных решений для задач оптимизации, допускающих интерпретацию решений в виде различ-

ного рода графовых структур. Это не позволяет напрямую использовать каноническую парадигму роя частиц (например, задача направленной мутации одного дерева в направлении другого, с формальной точки зрения, является весьма нетривиальной).

В работе предлагается подход к построению модифицированной парадигмы роя трансформирующихся хромосом  $H = \{H_i | i = 1, 2, \dots, n\}$ ,  $H_i \in H$ ,  $H_i = \{h_{il} | l = 1, 2, \dots, n_l\}$  с целочисленными значениями генов  $h_{il}$  в алгоритмах генетического поиска по аналогии с парадигмой роя частиц. Хромосома  $H_i$  является генотипом некоторого решения (фенотипа  $\Phi_i$ ). Оценка целевой функции фенотипа  $\Phi_i$  является оценкой целевой функции генотипа  $H_i$ . Хромосома  $H_i$  полностью определяет состояние, в котором находится фенотип  $\Phi_i$ .

В работе представлена архитектура многоагентной системы на основе природных вычислений, выполняющая поиск решения, используя в качестве модели роевого интеллекта рой трансформирующихся хромосом.

## 1. Разработка метода оптимизации роем трансформирующихся хромосом

Для организации поисковой процедуры необходимо представление объекта оптимизации, наличие операторов, преобразующих одно решение в другое (т.е. осуществляющих переход в пространстве решений от одного решения к другому), а также наличие стратегии поиска в пространстве решений. Какие изменения и на сколько – определяет стратегия поиска. В работе в качестве объекта оптимизации рассматривается хромосома [Карпенко, 2016].

Пусть имеется линейное векторное пространство (ЛВП), элементами которого являются  $n$ -мерные точки. Каждым любым двум точкам  $p$  и  $q$  этого пространства однозначным образом сопоставим единственную упорядоченную пару этих точек, которую в дальнейшем будем называть геометрическим вектором (вектором).  $p, q \in V(p, q)$  – геометрический вектор (упорядоченная пара).

Совокупность всех точек ЛВП, пополненная геометрическими векторами, называют точечно-векторным или аффинным пространством. Аффинное пространство является  $n$ -мерным, если соответствующее ЛВП так же является  $n$ -мерным.

В качестве аффинно-релаксационной модели (АРМ) роя хромосом используется граф  $G_a$ , вершины которого соответствуют хромосомам роя, а ребра соответствуют аффинным связям между хромосомами в аффинном пространстве.

В каноническом методе роя частиц [Clerc, 2006], [Kennedy et al., 1995] используются геометрические меры близости решений. Решение, соответствующее частице, полностью определяется параметрами позиции, в ко-



торой размещена частица. Изменение решения реализуется путем перехода частицы в новую позицию. Средство изменения решения представлено в виде универсального аналитического выражения синтеза геометрических параметров новой позиции (координат в пространстве поиска) в предположении (с надеждой), что решение в данной позиции будет лучше (что не гарантировано).

В работе агентами роя являются трансформирующиеся хромосомы, являющиеся генотипами решения [Лебедев и др., 2018]. В качестве меры близости (сходства) двух хромосом используется аффинность – мера эквивалентности, родственности (сходства, близости) двух хромосом. Вес ребра в графе  $G_a$ , связывающего две вершины графа  $G_a$ , соответствующие двум хромосомам  $H_i$  и  $H_j$ , равен аффинности, количественно описывающей степень родственности между  $H_i$  и  $H_j$ . Эквивалентом оценки, степень родственности (аффинность) двух хромосом  $H_i$  и  $H_j$ , служит параметр степень различия двух хромосом  $H_i$  и  $H_j$ . В работе в качестве оценки аффинной связи используется параметр – степень различия двух хромосом  $H_i$  и  $H_j$ . Максимизация аффинности двух хромосом  $H_i$  и  $H_j$  соответствует минимизации степени различия.

В качестве средства изменения решения (трансформации) служит оператор трансформации (направленной мутации), суть которого заключается в изменении генотипа путем изменения целочисленных значений генов  $h_{il}$  в хромосоме  $H_i(t) = \{h_{il}(t) | l=1, 2, \dots, n\}$ . Целью перехода хромосомы в новое состояние является увеличение веса аффинной связи между хромосомами, что равносильно сокращению степени различия [Лебедев и др., 2019].

Введём на множестве трансформирующихся хромосом  $H_i(t)$  отношение предпочтения  $\succ$ . Пусть имеются две хромосомы  $H_i(t)$  и  $H_z(t)$ . Хромосома  $H_i(t)$  считается предпочтительней (лучшей), т.е.  $H_i(t) \succ H_z(t)$ , если выполняется приведённое ниже выражение:  $f_i(t) > f_z(t)$ , где  $f_i(t)$  и  $f_z(t)$  – значение целевой функции  $H_i(t)$  и  $H_z(t)$ .

Введем обозначения:

$H(t) = \{H_i(t) | i=1, 2, \dots, n\}$  – рой хромосом,  $H_i \in H$ ,  $H_i = \{h_{il} | l=1, 2, \dots, n_l\}$ ;

$f_i(t)$  – значение целевой функции хромосомы  $H_i(t)$ ;

$T$  – число выполненных итераций;

$H_i^*(t)$  – трансформирующая хромосома  $H_i$  в лучшем состоянии, среди состояний, в которых она находилась с начала первой итерации до  $t$ -й итерации:  $H_i^*(t) = \{h_{il}^*(t) | l=1, 2, \dots, n_l\}$ ;  $(H_i^* \succ H_i(t) \rightarrow f_i^*(t) > f_i(t))$ ;

$f_i^*(t)$  – значение целевой функции хромосомы в позиции  $H_i^*(t)$ :

$$f_i^*(t) = \min f_i(t), \quad t \in [0:T]. \quad (1)$$

$H^*(t)$  – хромосома роя в лучшем состоянии среди всех состояний всех хромосом роя, в которых они находилась с начала первой итерации до  $t$ -й итерации;  $H^*(t) = \{h^*_i(t) | i=1, 2, \dots, n\}$ ;

$f^*(t)$  – значение целевой функции хромосомы  $H^*(t)$ .

$$f^*(t) = \min f_i(t), \quad i \in [1:N] \quad t \in [0:T]. \quad (2)$$

Парадигма оптимизации роем трансформирующихся хромосом имеет вид:

1. Формирование исходного роя трансформирующихся хромосом  $H = \{H_i | i=1, 2, \dots, n_o\}$ .  $t=1$ .

2. Расчет целевой функции  $f_i(t)$  для всех  $H_i$ .

3. Выбор лучших хромосом  $H^*_i(t)$  и  $H^*(t)$ .

4. Если  $t < T$ , то  $t=t+1$  и переход к пункту 5, иначе переход к пункту 7.

5. Трансформация всех хромосом роя в новые состояния.

6. Переход к пункту 2.

7. Конец работы алгоритма.

$t$  – номер итерации,

$T$  – число итераций.

После ввода исходных данных формируется структура хромосомы и производится формирование (разработка) оператора трансформации. Далее формируется рой хромосом. Для всех хромосом рассчитывается значение целевой функции  $f_i(t)$ .

Если вес аффинной связи между хромосомами рассматривать как длину, то в процессе поисковой процедуры происходит релаксационное сжатие аффинного графа.

Процесс поиска решений итерационный. В процессе поиска осуществляется трансформация (переход) хромосом в состояния с лучшим значением целевой функции решения (градиентная стратегия).

На каждой итерации  $t$  «лучшие», с точки зрения целевой функции хромосомы, объявляются «центром притяжения». Векторы трансформации всех хромосом в аффинном пространстве устремляются к этим центрам. Каждая хромосома  $H_i(t)$  на шаге  $t$  переходит в аффинном пространстве с помощью релаксационной процедуры трансформации в новое состояние  $H_i(t+1)$ , при котором вес аффинной связи (степень различия между хромосомой  $H_i(t)$  и «центром притяжения») уменьшается.

Переход хромосомы в новые состояния возможен с учетом степени различия к одной хромосоме центра притяжения либо к группе соседних с учетом вероятности перехода в новое состояние.

Переход в новые состояния выполняется путем изменения значений генов и структуры хромосом. Целью трансформации хромосомы  $H_i(t)$ , тяготеющей к лучшей хромосоме  $H^*(t)$ , в новое состояние  $H_i(t+1)$  является минимизация степени различия  $R_i^*$  между  $H_i(t)$  и  $H^*(t)$ , что соответствует увеличению веса аффинной связи.

Эволюция роя базируются на коллективном обучающем процессе внутри популяции хромосом, каждая из которых представляет собой поисковую точку в пространстве допустимых решений данной задачи. Рой хромосом случайно инициализируется, и затем охватывает лучшие регионы поискового пространства посредством случайных процессов трансформации. Окружающая среда представляет качественную информацию (степень пригодности) о поисковых точках (хромосомах), а процесс трансформации переводит хромосомы в новые состояния, у которых значение пригодности выше. Обновленные после трансформации хромосомы являются, в свою очередь, базовыми точками в последующих трансформациях (непрерывная цепочка постоянных перерождений в соответствии с принципом сансара) [Курейчик и др., 2006].

Механизм трансформации перемешивает генетическую информацию роя и способствует обновлению генетической информации индивидуумов. Методологическая основа роя хромосом зиждется на гипотезе селекции, которая в самом общем виде может быть сформулирована так: чем выше приспособленность особи, тем выше вероятность того, что в результате трансформации (переходе в новое состояние) признаки, определяющие приспособленность, будут выражены еще сильнее. Трансформация, т.е. переход хромосомы  $H_i(t)$  в новое состояние  $H_i(t+1)$  осуществляется с учетом «степени различия» с лучшей хромосомой  $H^*(t)$ , либо с интегральной оценкой «степени различия» от группы соседних хромосом и с учетом вероятности перехода в новое состояние.

В работе рассматриваются два подхода к трансформации хромосомы – численный и структурный. В обоих подходах на базе  $H_i(t)$  формируется такая хромосома  $H_i(t+1)$ , у которой степень различия с лучшей хромосомой  $H^*(t)$  меньше.

При численном подходе изменяются количественные параметры хромосомы в заданном диапазоне. В качестве степени близости служит степень различия параметров, описывающих хромосомы.

При структурном подходе производится трансформация хромосомы на основе изоморфных преобразований структуры и параметров хромосомы. В качестве степени близости рассматривается степень различия изоморфных представлений и параметров, описывающих сравниваемые хромосомы. Процедура трансформации включает ряд операторов, структура которых связана со структурой хромосомы и генов и ориентирована на изменении близости (сходства) двух агентов.

## 2. Процедуры трансформации хромосом

В качестве средства трансформации – перехода хромосомы  $H_i(t)$  в хромосому  $H_i(t+1)$  служит оператор направленной мутации (ОНМ), суть которого заключается в изменении целочисленных значений генов в хромосоме  $H_i(t)$ . ОНМ применяется одновременно ко всем генам хромосомы. Решением задачи является хромосома, построенная в процессе трансформации роя хромосом в аффинном пространстве, с лучшим значением целевой функции.

При использовании первого подхода в качестве степени различия между двумя хромосомами  $H_i(t)$  и  $H^*(t)$  используется величина  $R_i^*$ :

$$R_i^* = \sum_l |h_{il} - h_{il}^*(t)| = \sum_l \delta_l \quad (3)$$

где  $h_{il} \in H_i$ ;  $h_{il}^* \in H^*(t)$ ;  $h_{il}$ ,  $h_{il}^*$  – значения генов в локусе  $l$ ;  $\delta_l^* = |h_{il} - h_{il}^*|$ .

В работе используется градиентный подход к организации поисковой процедуры. Градиентный метод поиска экстремума заключается в последовательном улучшении решения за счет движения по направлению к оптимуму целевой функции.

При градиентном подходе трансформация гомологичной хромосомы выполняется как:

$$H_i(t+1) = H_i(t) + \omega_i(t+1), \quad (4)$$

где  $\omega_i(t+1)$  шаг трансформации хромосомы  $H_i(t)$  в хромосому  $H_i(t+1)$ . Начальное (стартовое) состояние определяется, как  $H_i(0)$ ,  $\omega_i(0)$ .

Приведенная формула представлена для хромосомы в целом. Для отдельного гена  $h_{il}(t) \in H_i(t)$  формула примет вид:

$$h_{il}(t+1) = h_{il}(t) + \omega_{il}(t+1). \quad (5)$$

В работе предложены два подхода к расчету параметра  $\omega_{il}(t+1)$ .

Расчет по формуле (5) полностью *детерминированный*, расчет по формуле (6) включает *элементы случайности*.

$$\omega_{il}(t+1) = k_1 \cdot (\text{sgn}(h_{il}^*(t) - h_{il}(t)) + k_2 \cdot (\text{sgn}(h_{il}^*(t) - h_{il}(t))). \quad (6)$$

$$\omega_{il}(t+1) = L(\text{sgn}(h_{il}^*(t) - h_{il}(t)), P, 0) + L(\text{sgn}(h_{il}^*(t) - h_{il}(t)), P, 0), \quad (7)$$

где  $k_1$  и  $k_2$  коэффициенты пропорциональности;  $\text{sgn}[y]$  – функция знака числа: если  $y > 0$ , то  $\text{sgn}(y) = 1$ ; если  $y < 0$ , то  $\text{sgn}(y) = -1$ ; если  $y = 0$ , то  $\text{sgn}(y) = 0$ ;  $L(y_1, p, y_2)$  вероятностное событие, имеющее два возможных исхода  $y_1$  и  $y_2$ , вероятности наступления которых соответственно  $p$  и  $(1-p)$ .

На каждой итерации  $t$  вычисляются пригодности  $f_i(t)$  позиций  $H_i(t)$  (значение целевой функции  $f_i(t)$ ) и при необходимости обновляется информация о наилучших найденных хромосомах  $H^*(t)$  и  $H_i^*(t)$  роя. Каждая позиция роя  $H_i(t)$ , кроме лучших –  $H^*(t)$  и  $H_i^*(t)$ , путем изменения значений  $h_{il}$  генов  $h_{il}$  трансформируется в новую позицию  $H_i(t+1)$ .

Цель трансформации гомологичной  $H_l(t)$  заключается в уменьшении разности между значениями каждой пары генов  $(h_{il}, h_{il}^*)$  у двух выбранных хромосом  $H_l(t)$  и  $H^*(t)$ ,  $l=1, \dots, n$ .

Последовательно просматриваются (начиная с первого) локусы хромосом  $H_l(t)$  и  $H^*(t)$ , и сравниваются соответствующие им гены  $h_{il}(t)$  и  $h_{il}^*(t)$ . Если в ходе последовательного просмотра локусов в текущем локусе  $l$  выпадает с вероятностью  $P$  событие «мутация», то ген  $h_{il}(t) \in H_l(t)$  мутирует.

Вероятность мутации  $P$  зависит от степени различия  $R_i^*(t)$  между хромосомами  $H^*(t)$  и  $H_l(t)$ , и определяется следующим образом:

$$P = \alpha \cdot R_i^*(t) / (n). \quad (8)$$

Чем больше  $P$ , тем больше вероятность, что значение  $h_{il}(t)$  будет изменено.

**Пример.**

Пусть  $H^*(t) = \langle 5, 8, 17, 20 \rangle$ ,  $H_l(t) = \langle 4, 10, 15, 22 \rangle$ . В соответствии с выражением (6)

$$S_i^*(t) = |5-4| + |8-10| + |17-15| + |20-22| = 7.$$

Примем  $P=1$ , тогда в соответствии с правилами направленной мутации:

$$H_l(t+1) = \langle 5, 9, 16, 21 \rangle, \quad R_i^*(t+1) = |5-5| + |8-9| + |17-16| + |20-21| = 3.$$

Различие между хромосомами  $H_l(t)$  и  $H^*(t)$  сократилось с 7 до 3.

Для учета одновременного тяготения хромосомы  $H_l(t)$  к хромосомам роя  $H^*(t)$  и  $H_{i'}^*(t)$  на первом этапе формируется виртуальный центр притяжения роя  $H^v(t+1)$ . Формирование виртуального центра  $H^v(t+1)$  осуществляется путем применения процедуры виртуальной трансформации к хромосоме  $H_l(t)$ , тяготеющей к хромосоме  $H_{i'}^*(t)$ , в виртуальную хромосому  $H^v(t+1)$ .

После определения центра притяжения  $H^v(t+1)$  хромосома  $H_l(t)$  с помощью процедуры трансформации в направлении виртуальной позиции  $H^v(t+1)$  трансформируется в хромосому  $H_l(t+1)$ .

После трансформации хромосомы  $H_l(t)$  в новую  $H_l(t+1)$  виртуальная хромосома  $H^v(t+1)$  исключается.

Локальная цель трансформации хромосомы достижение ею позиции с наилучшим значением целевой функции. Глобальная цель роя позиций – формирование оптимального решения задачи.

В ряде алгоритмов качество структуры данных, несущей информацию о решении, используется последовательность (приоритетный список) код, представляющую собой очередность формирования решения. Переход от приоритетного списка к решению производится с помощью декодера. Хромосома имеет следующую структуру:  $H_i(t) = \langle h_{il}(t) | l=1, 2, \dots, n \rangle$ .

Авторами разработана процедура трансформации, которая заключается в изменении взаимного расположения элементов в хромосоме. Хромосома  $H_l(t)$  переходит в новое состояние  $H_l(t+1)$  с новым взаимным расположением элементов.

Подсчитывается число  $Q_i(t)$  непересекающихся пар соседних генов в хромосомах  $H_i(t)$  и  $H^*(t)$ , у которых взаимное расположение элементов пары в хромосоме  $H_i(t)$  и  $H^*(t)$  не совпадает.

Вычисляется степень различия  $R_i^*(t_i) = Q_i^*(t)$ .

### **Пример.**

Процедура трансформации реализуется за два такта.

Пусть  $H_i(t)$  и  $H^*(t)$  имеют вид:  $H_i(t) = \{1, 3, 2, 10, 8\}$ ,  $H_j^*(t) = \{1, 10, 2, 3, 8\}$ .

На первом такте в  $H_i(t)$  формируется множество непересекающихся пар генов  $D_1 = (1, 3), (2, 10)$ . Взаимное расположение элементов пары  $(1, 3)$  в  $H_i(t)$  и  $H^*(t)$  совпадают, пары  $(2, 10)$  – нет. В  $H_i(t+1)$  переставляются местами элементы пары  $(2, 10)$ .  $H_i(t+1) = \{1, 3, 10, 2, 8\}$ .

На втором такте в  $H_i(t+1)$  формируется множество пар  $D_2 = (3, 10), (2, 8)$ . Взаимное расположение элементов пары  $(2, 8)$  в  $H_i(t+1)$  и  $H_j^*(t)$  совпадают, пары  $(3, 10)$  – нет. В  $H_i(t+1)$  переставляются местами элементы пары  $(3, 10)$ .  $H_i(t+2) = \{1, 10, 3, 2, 8\}$ .

Хромосома  $H_i(t)$  переходит в новое состояние  $H_i(t+1)$  с новым взаимным расположением элементов в хромосоме и новым значением степени близости.

Если вес аффинной связи между хромосомами рассматривать как длину, то в процессе поисковой процедуры происходит релаксационное сжатие аффинного графа.

## **3. Экспериментальные исследования**

Метод роя трансформирующихся хромосом был использован в разработанном алгоритме размещения ТХАР.

Для экспериментальных исследований разработанной программы размещения были применены схемы набора РЕКУ с известным оптимумом  $F_{opt}$ : Ex.1 на 30 блоков, Ex.2 – 60, Ex.3 – 90, Ex.4 – 120, Ex.5 – 150. Для сравнения были выбраны современные алгоритмы размещения: Dragon v2.20, Capo v.8., mPL v.2.0, mPG v1.0, QPlace v.5.1. [Cong J. et al., 2004], [CADENCE DESIGN SYSTEMS, INC].

Для определения оптимальности достигнутых значений, рассчитывался параметр: наилучшая длина соединений  $F_{opt}$  к полученной длине соединения  $F$  (для РЕКО) или (для G-РЕКУ и РЕКУ). Это отношение называется «степенью качества».

В табл. 1. приведены полученные значения показателя степень качества  $F_{opt}/F$  ряда известных алгоритмов и алгоритма ТХАР. Степень качества у разработанной программы ТХАР на 10% выше, чем у программ Dragon, Caro, mPL, mPG и Qplace.

В результате экспериментов установлено, что при объеме популяции  $M=100$  алгоритм сходится в среднем на 115 итерации. При этом отклонения в сторону увеличения этой оценки составлял до 10%, а в сторону уменьшения до 35%. Оценкой качества служит величина  $F_{\text{опт}}/F$ , где  $F$  – оценка полученного решения. ВСА  $O(n^2)$ .

Таблица 1

### Значения показателя «степень качества»

Тест	Dragon	Сапо	mPL	mPG	Qplace	ТХАР
Ex.1	0.71	0.73	0.81	0.71	0.73	0.92
Ex.2	0.72	0.72	0.84	0.72	0.78	0.9
Ex.3	0.81	0.8	0.83	0.81	0.82	0.9
Ex.4	0.83	0.84	0.85	0.83	0.81	0.92
Ex.5	0.82	0.81	0.86	0.82	0.84	0.92

### Заключение

Суть поисковой процедуры заключается в последовательной смене посредством трансформации состояний объекта оптимизации (хромосомы) и поиске оптимального состояния. В поисковом популяционном методе оптимизации роem хромосом агентами популяция являются хромосомы. Хромосома является генотипом объекта оптимизации.

Предложена аффинно-релаксационная модель (АРМ) роя хромосом – граф вершины которого соответствуют хромосомам, а дуги соответствуют аффинным связям между ними. В качестве веса аффинной связи между хромосомами на аффинно-релаксационной модели роя хромосом используются оценки степени «различия» между ними.

Ключевая проблема, которая была решена в данной работе связана с разработкой методики организации процесса перехода хромосомы в новое состояние. Переход хромосомы в новое состояние осуществляется с помощью релаксационной процедуры трансформации. Разработаны аналитические выражения и процедуры для расчета оценки степени близости (аффинности), зависящей от вида структуры хромосомы: вектор, матрица, дерево и их совокупности, являющейся интерпретацией решений

В работе в качестве средства изменения решения служит процедура трансформации, суть которой заключается в изменения структуры и целочисленных значений генов в хромосоме. Целью перехода является сокращении веса аффинной связи между хромосомами. Описаны механизмы трансформации. В качестве основных структур данных, несущих информацию о решении, используются списки, фактически являющимися интерпретациями решений.

Улучшить качество работы разработанного алгоритма можно при помощи настройки значений управляющих параметров. Временная сложность алгоритма при фиксированных значениях размера популяции и количества генераций составляет  $O(n)$ . В общем зависимость времени работы гибридного алгоритма составляет  $O(n^2) - O(n^3)$ .

### Список литературы

- [Clerc, 2006] Clerc M. Particle Swarm Optimization. – ISTE, London, UK, 2006. – 97 p.
- [Kennedy et al., 1995] Kennedy J., Eberhart R.C. Particle swarm optimization // In Proceedings of IEEE International Conference on Neural Networks. – 1995. – P. 1942-1948.
- [Карпенко, 2016] Карпенко А.П. Современные алгоритмы поисковой оптимизации. Алгоритмы, вдохновленные природой: учебное пособие. Второе издание. – М: Изд-во МГТУ им. Н.Э. Баумана, 2016. – 448 с.
- [Лебедев и др., 2018] Лебедев Б.К., Лебедев О.Б., Лебедев В.Б. Гибридный метод стохастической оптимизации на основе интеграции моделей эволюции и роевого (стайного) поведения животных в аффинных пространствах поиска // Сборник трудов Шестнадцатой национальной конференции по искусственному интеллекту с международным участием. – М.: ФГП ИТАР-ТАСС филиал РКП, 2018. – С. 148-156.
- [Лебедев и др., 2019] Лебедев Б.К., Лебедев О.Б., Лебедева Е.О Модифицированные механизмы перемещения роя частиц (агентов) в аффинном пространстве с целочисленными параметрами // Известия ЮФУ. Технические науки. – 2019. – № 4. – С. 121-136.
- [Курейчик и др., 2006] Курейчик В.М., Лебедев Б.К., Лебедев О.Б. Поисковая адаптация: Теория и практика. – М.: Физматлит, 2006. – 272 с.
- [Cong et al., 2004] Cong J., Romesis M., Xie M. Optimality, Scalability and Stability Study of Partitioning and Placement Algorithms // Proc. of the International Symposium on Physical Design. – Monterey, CA, 2004. – P. 88-94.
- [CADENCE DESIGN SYSTEMS, INC] QPlace version 5.1.55, compiled on 10/25/1999. Envisia ultra-placer reference.



## ПОВЫШЕНИЕ КАЧЕСТВА ПРОЦЕССОВ КОНТРОЛЯ ПРОДУКЦИИ НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ НЕЙРОСЕТЕВЫХ И НЕЙРО-НЕЧЕТКИХ МОДЕЛЕЙ И СРЕДСТВ

А.Е. Мисник (*anton@misnik.by*)

М.А. Шалухова (*shaluhova.m@yandex.ru*)

Белорусско-Российский университет, Могилёв

В статье описывается способ повышения качества процессов контроля выпускаемой продукции на пищевом производстве с помощью использования нейросетевых и нейронечетких методов, моделей и средств. Предложено использовать выделение признаков с помощью свёрточных сетей с дальнейшей постобработкой в системе нечёткого вывода. В ходе работы предложенной системы был получен высокий процент верных распознаваний (91,9%), за время эксплуатации возвраты покупателем продукции по причине брака на снизились на 63% к аналогичному периоду прошлого года. Полученные результаты показывают, что идентификация дефектов с использованием адаптивной нейро-нечеткой системы вывода является подходящим инструментом при решении задач анализа дефектов.

**Ключевые слова:** нейросетевой подход, нечеткая логика, распознавание дефектов, нейронная сеть, ANFIS.

### Введение

При производстве продукции, требующей сложного технологического процесса, контроль качества в целом и своевременное обнаружение дефектов на каждой стадии играет решающую роль в конечном качестве продукции и удовлетворенности заказчика. Процесс контроля качества является сложным процессом, заключающим в выявлении явных дефектов и несоответствий заданным требованиям [Кушнаренко и др., 2011]. Как правило, контролем качества на производстве занимается отдел контроля качества, состоящий из специалистов-экспертов. При классическом подходе к проверке качества, проверки зачастую носят периодический либо плановый характер, так как сплошной контроль или разбраковка всех изделий невозможны в силу экономической нецелесообразности проверки больших партий готовой продукции или неизбежного разруше-

ния изделий при контроле [Юдин, 2022]. Часть функций контроля способны взять на себя экспертные системы, использующие нейросетевой подход. Заинтересованность крупных производств в построении эффективных систем и процессов контроля качества ежегодно растет, ведь с повышением качества отгружаемой продукции увеличивается вероятность повторных закупок, дополнительных продаж и, в целом, улучшает имидж организации [Дунченко и др., 2012].

Искусственные нейронные сети позволяют решать задачи, на решение которых другими способами уходит существенное количество времени и средств [Николенко и др., 2018]. Сочетание искусственных нейронных сетей с использованием нечеткой логики позволяет описывать сложные трудно формализуемые процессы, автоматизировать сложные производственные задачи, такие как обнаружение дефектов и определении брака [Гифт, 2019]. Применение возможностей нейросетевого подхода для процессов контроля качества способно снять часть задач с отдела контроля, а также уменьшить количество брака, фиксируемого на стороне заказчика.

В данной рассматривается способ повышения качества процессов контроля выпускаемой продукции на пищевом производстве, за счёт использования нейросетевых и нейронечетких методов, моделей и средств. Целью разработки способа является уменьшение количества возвратов покупателем продукции по причине брака.

## **1. Разработка способа и средств контроля качества продукции**

Автоматизация части процессов по поиску и классификации дефектов за счёт нейросетевых технологий позволят существенно сократить затраты предприятия на выявление дефектов в произведенной продукции. Применение подобных систем позволит так же осуществлять переход к сплошному контролю над качеством выпускаемой продукции и своевременно выявлять дефекты на стадии незавершенного производства.

Одним из способов выявления дефектов производства является визуальный осмотр и сверка полученных размеров с эталоном (контроль по контрольному образцу, измерительный контроль, визуальный контроль) [Понамарёв и др., 2012]. Для поиска и классификации дефектов, обнаруживаемых при классическом подходе в ходе визуального контроля, предлагается использовать свёрточную нейронную сеть: для обработки дефектоскопических изображений свёрточные нейронные сети могут иметь в своем составе несколько слоев: входной, свёрточный, субдискретизации, полносвязный [Замятин, 2016].

Свёрточные сети, использующие признаковый метод анализа изображений в автоматизированной системе, хорошо справляются с поиском и определением дефектов, требующими выделения узкоспецифичных при-

знаков. Это позволяет их применять при визуальном и измерительном контроле, однако для построения полноценной экспертной системы полученные результаты следует подвергнуть дальнейшему анализу [Мисник и др., 2021].

Для анализа результатов, полученных свёрточной сетью, как правило, необходимо применение дополнительных средств интеллектуального анализа. При отборе образцов брака специалистами отдела контроля качества было отмечено, что различные специалисты оценивают степень выраженности дефекта и его исправимости неодинаково. В результате оценки разными экспертами один и тот же образец продукции мог быть признан неисправимым браком и годной продукцией одновременно. Расхождение наблюдалось в случае, если на продукции присутствовали признаки нескольких дефектов, что добавляет данным неопределенности. Экспертное мышление основано на неточных, приблизительных данных, полученном ранее опыте в сфере, анализ которых позволяет формировать четкие решения. Благодаря своим особенностям, адаптивная нейро-нечеткая система вывода хорошо справляется с задачами интерпретации полученного результата классификации [Misnik et al., 2019]. Для построения нейро-нечетких систем могут использоваться как предварительные знания, так и данные процесса. Предлагается использование для анализа полученных данных адаптивную нейро-нечеткая система вывода (ANFIS) [Шалухова, 2021].

Для построения нейро-нечетких систем могут использоваться как предварительные знания, так и данные процесса. Предварительное знание может иметь довольно приблизительный характер, можно выделить два основных подхода к интеграции знаний и данных: формулирование знаний эксперта как совокупность лингвистических правил “ЕСЛИ, ТО” и построение с нуля нечетких правил, включая связанные параметры. В первом случае таким образом создается исходная модель, а функции принадлежности и последующие параметры затем настраиваются с использованием данных процесса, во втором случае преимуществом использования нейро-нечеткой модели является возможность интерпретировать полученный результат. Эксперт может справиться с информацией, хранящейся в базе правил, своими собственными знаниями, изменять правила или предоставлять дополнительные, чтобы продлить срок действия модели [Борисов и др., 2021]. Вышеупомянутые методы, в зависимости от проблемы, могут быть объединены.

ANFIS является нейронной сетью с несколькими входами, представляющими собой нечеткие лингвистические переменные, и единственным выходом. Термы входных лингвистических переменных описываются стандартными функциями принадлежности, термы выходной переменной

представляются линейной или постоянной функцией принадлежности. Также ANFIS представляет собой систему нечеткого вывода, в которой каждое из правил нечетких продукций имеет постоянный вес, равный 1 [Горбаченко, 2022].

В адаптивных системах веса изменяются в процессе обучения для достижения наилучшей приспособленности модели к заданным значениям, а не уточняются в процессе анализа данных для определения уровня активации условия в правиле вывода. Такая сеть может быть успешно использована для настройки функций принадлежности и настройки базы правил в нечеткой экспертной системе. Предлагаемая схема работы системы представлена на рис. 1.

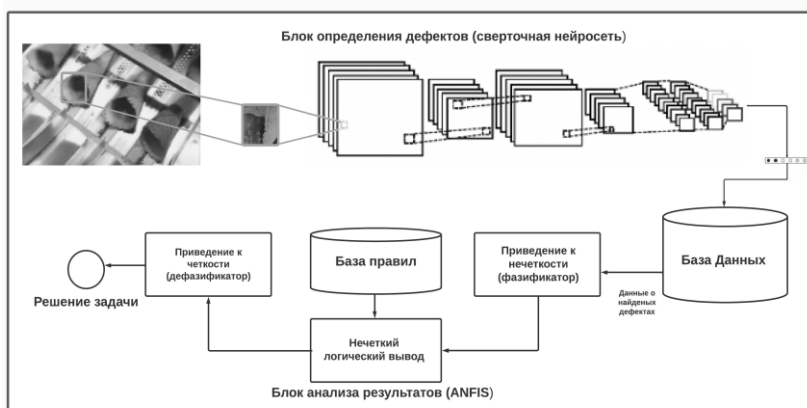


Рис. 1. Схема работы экспертной системы распознавания дефектов

Над производственной линией, с дополнительным источником света, размещена камера. Изображение с камер поступает на вход к сверточной нейронной сети, где происходит обнаружение дефекта (при наличии такового), его классификация, определяется его расположение. Сверточные нейронные сети обеспечивают частичную устойчивость к изменениям масштаба, смещениям, поворотам, смене ракурса и прочим искажениям. Выходные данные, полученные от сверточной нейронной сети, передаются в блок анализа, анализ полученных данных проводится с помощью ANFIS.

## 2. Определение лингвистических переменных

Лингвистические переменные предназначены в основном для анализа сложных или плохо определенных явлений, их определение требует анализа как поставленной задачи, так и текущей ситуации на рассматриваем-

мом производстве. При подготовке к разработке одной из сложностей является определение необходимости и достаточности данных, так как экспертное мнение часто является плохо формализуемым, эксперт часто не может четко обозначить приоритетность признаков, отсюда возникает проблема избыточного набора данных. Для уменьшения времени обучения и повышения качества работы системы требуется определить, какие данные необходимы для решения конкретной задачи. Рассматриваемое предприятие (СООО “Айспак”) производит несколько основных видов продукции: вафельный стаканчик и конус, лист вафельный, вафельная крошка, сухари панировочные и вафельные сахарные рожки. Соответствие качества продукции определяется согласно государственному стандарту Республики Беларусь (СТБ 2400-2015) “Вафли”, и межгосударственному стандарту ГОСТ 14031-2014 в котором содержатся общие технические условия, правила приемки, методы контроля, упаковка, маркировка, транспортирование, хранение.

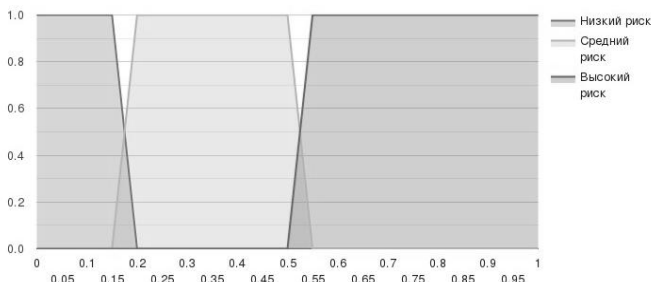
Целевой продукцией средств визуального контроля разработанной системы являются сахарные рожки (в том числе и окрашенные), вафельные стаканчики и вафельные листы. Визуальный контроль способен выявить следующие разновидности дефектов у произведенных вафель: разрывы либо разломы вафельного листа, неровный обрез края вафельного листа, пятна, просвечивание корпуса с доньшка изделий (для вафельных стаканчиков). Тогда, для свёрточной нейронной сети задача включает в себя обнаружение одного из перечисленных дефектов, а также варианта, при котором искомые дефекты отсутствуют. Распространённость различных дефектов продукции в общей доле всех выявляемых дефектов на производстве приведена в табл. 1.

Таблица 1

Виды дефекта			
Разрывы, разломы	Неровный обрез	Пятна	Просвечивание корпуса
57,63%	17,27%	11,40%	13,70%

При производстве различных видов продукции, процент частоты выявления характерных дефектов в каждом подвиде неодинаков. Различие в содержании сахара (массовая доля от 0 до 45%, согласно технологическим картам производства вафельной продукции, применяемым на рассматриваемом производстве) и допустимом стандартом проценте влаги обуславливает отличия в прочности изделий. В результате анализа полученных данных, а также интервью с технологами производства и специалистами контроля качества, были выявлены следующие факторы риска возникновения брака: сложность формы, содержание сахара, сложность линии обрезки (неровный обрез),

наличие разрывов/разломов и пятен, просвечиваний корпуса [Мисник и др., 2022]. Факторы риска являются входными переменными для нечеткой системы. Каждая переменная принимает три нечетких значения, которые описываются трапецевидными функциями принадлежности. Выходной переменной является нечеткая переменная «риск появления брака», которая принимает нечеткие значения: низкий, средний и высокий (рис. 2).



*Рис. 2. Переменная риска: низкий риск [0–20%], средний риск [20–55%] и высокий риск [55–100%]*

Эти нечеткие значения определяются также трапецевидной функцией принадлежности, соответствующей четкому значению процента риска. Диапазоны каждой переменной риска определяются как низкий риск [0–20%], средний риск [20–55%] и высокий риск [55–100%].

Ниже приведено несколько примеров правил, используемых в работе. Нечеткие правила и факторы риска сформированы методом изучения нормативной документации, а также по результатам интервью со специалистами контроля качества.

Примеры полученных нечетких правил:

R1: ЕСЛИ (форма – простая) И (содержание сахара – низкий) ТО (риск появления брака – низкий).

R2: ЕСЛИ (форма – простая) И (содержание сахара – высокий) И (сложность линии обрезки – средний) ТО (риск появления брака – низкий).

R3: ЕСЛИ (форма – сложная) И (содержание сахара – высокий) И (сложность линии обрезки – средний) И (просвечиваний корпуса – низкий) И (Наличие пятен – много) ТО (риск появления брака – высокий).

R4: ЕСЛИ (форма – сложная) И (содержание сахара – низкий) И (сложность линии обрезки – высокий) И (наличие разрывов – средний) ТО (риск появления брака – высокий), и т.д. правил в базе знаний должно быть достаточным, для описания лингвистических правил, однако нужно избегать избыточности.

Для определения типа дефектов и дальнейшей классификации использовалась структура с многими слоями, не имеющими обратных связей, использующих нечёткие сигналы. Операции суммирования производятся на основе фиксированной t-нормы и s-конормы или другой операции. Нечёткость в сетях этого типа проявляется при использовании ретроспективной информации, при формировании параметров функций принадлежности. Сами же параметры находятся на основе методов обучения нейронных сетей.

На основе полученной информации, были подготовлены три матрицы данных: обучающая, тестовых данных и матрица проверочных значений. В обучающей выборке содержится 7200 изображений вафельной продукции (4600 позитивных примеров, с низким риском выявления брака, 1800 примеров продукции, где риск возникновения брака средний (наличие неявных дефектов), и 800 единиц с высоким риском признания браком (явные дефекты либо сочетание нескольких видов дефектов), проверочная выборка составила 900 изображений (500, 280 и 120 соответственно).

Адаптивная нейро-нечеткая система вывода на достаточном уровне справляется с задачами высокоточного распознавания, которые требуют выделения узкоспецифичных признаков, что позволит их применять при визуальном и измерительном контроле. При выявлении несоответствия, решается задача классификации, согласно месту обнаружения, типа, и процента выявленного несоответствия.

### **3. Оценка эффективности способа и средств контроля качества выпускаемой продукции**

При использовании в тестовом режиме на пищевом производстве ("ООО "Айспак"), разработанная по предложенной схеме система, в период с сентября 2022 по январь 2023, позволила сократить затраты, связанные с возвратом продукции покупателем по причине брака на 63%, по сравнению с аналогичным периодом прошлого года.

Полученная точность распознавания обучающей и проверяющей выборки рассматривается в табл. 2.

Таблица 2

Тип выборки	Точность распознавания выборки			
	Неровный обрез	Пятна	Просвечивание корпуса	Разрывы, разломы
Обучающей	92,73%	90,96%	89,04%	93,83%
Проверяющей	91,70%	89,3%	88,77%	91,90%

Была достигнута цель в обеспечении непрерывного контроля за производственными линиями аппаратными средствами, что позволило разгрузить отдел контроля качества. В дальнейшей работе планируется улучшить точность распознавания, введя новые уточнения в множества лингвистических правил.

## Заключение

В рамках данной работы была рассмотрена возможность применения системы с использованием ANFIS, ориентированной на распознавания дефектов вафельной продукции. При применении на производстве система, показала высокий процент верных распознаваний (91,9%), за время эксплуатации возвраты покупателем продукции по причине брака на снизились на 63% к аналогичному периоду прошлого года. Возможность автоматической отбраковки дефектной продукции способно существенно сократить дальнейшие издержки, связанные с внутренним и внешним браком, а также расширяет возможности автоматизации линий производства.

Полученные результаты показывают, что идентификация дефектов с использованием адаптивной нейро-нечеткой системы вывода достаточно эффективна, что достигается благодаря ее устойчивости к искажениям входных данных, а также является подходящим выбором при решении задач обнаружения и классификации дефектов.

## Список литературы

- [Misnik et al., 2019] Misnik A., Krutalevich S., Prakapenka S., Borovykh P. and Vasiliev M. Impact Analysis of Additional Input Parameters on Neural Network Cryptocurrency Price Prediction // 2019 XXI International Conference Complex Systems: Control and Modeling Problems (CSCMP), Samara, 2019.
- [Борисов и др., 2021] Борисов В.В., Бобряков А.В., Мисник А.Е. Экспертные системы. – Смоленск: Универсум, 2021.
- [Гифт, 2019] Гифт Н. Прагматичный ИИ. Машинное обучение и облачные технологии // Серия «Библиотека программиста». – СПб.: Питер, 2019.
- [Горбаченко, 2022] Горбаченко В.И. Интеллектуальные системы: нечеткие системы и сети: учебное пособие для вузов. – 2-е изд., испр. и доп. – М.: Изд-во «Юрайт», 2022.
- [Дунченко и др., 2012] Дунченко Н.И., Магомедов М.Д., Рыбин А.В., Управление качеством в отраслях пищевой промышленности: учебное пособие. – 4-е изд. – М.: ИТК «Дашков и К°», 2012.
- [Замятин, 2016] Замятин А.В. Введение в интеллектуальный анализ данных: учебное пособие. – Томск: Издательский Дом Томского государственного университета, 2016.
- [Кушнаренко и др., 2011] Кушнаренко В.М., Репях В.С., Чирков Е.Ю. Кушнаренко Е.В. Дефекты и повреждения деталей и конструкций: монография. – Оренбург: ОГУ, 2011.



- [Мисник и др., 2021] Мисник А.Е., Шалухова М.А., Использование нейронных сетей для идентификации дефектов вафельного производства // Энергетика, информатика, инновации – 2021 (электроэнергетика, электротехника и теплоэнергетика, математическое моделирование и информационные технологии в производстве, микроэлектроника и оптотехника): Сборник трудов XI Международ. науч.-технич. конф., филиал ФГБОУ ВО «НИУ «МЭИ» в г. Смоленске. – Смоленск, 2021.
- [Мисник и др., 2022] Мисник А.Е., Шалухова М.А., Применение ANFIS при моделировании кибер-физических систем выявляющих дефекты незавершенного производства // XX Всероссийская научная конференция «Нейрокомпьютеры и их применение», Московский государственный психолого-педагогический университет: – М., 2022.
- [Николенко и др., 2018] Николенко С., Кадури А., Архангельская Е. Глубокое обучение // Серия «Библиотека программиста». – СПб.: Питер, 2018.
- [Пономарёв и др., 2012] Пономарев С.В., Мищенко С.В., Мищенко Е.С., Гребенникова Н.М., Балабанов П.В., Евлахин, Р.Н., Злобин Э.В., Копышева Н.А., Мозгова Г.В., Чуриков А.А., Шишкина Г.В. Управление качеством процессов и продукции. В 3-х кн. Кн. 1: Введение в системы менеджмента качества процессов в производственной, коммерческой и образовательной сферах: учебное пособие / под ред. д-ра техн. наук, проф. С.В. Пономарева. – Тамбов: Изд-во ФГБОУ ВПО «ТГТУ», 2012.
- [Шалухова, 2021] Шалухова М.А. Использование нейронных сетей для идентификации дефектов незавершенного производства // Международная научно-техническая конференция молодых ученых БГТУ им. В.Г. Шухова - 2021 г. (24-25 октября 2021 г.). – Белгород: Федеральное государственное бюджетное образовательное учреждение высшего образования «Белгородский государственный технологический университет им. В.Г. Шухова», 2021.
- [Юдин, 2022] Юдин С.В. Управление качеством: теоретические основы и практические рекомендации: учебное пособие для вузов. – СПб.: Лань, 2022.

## ПОДХОД К ГЕНЕРАЦИИ ФОРМАЛЬНЫХ НЕЧЕТКИХ КОГНИТИВНЫХ КАРТ ДЛЯ ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ В КОГНИТИВНОМ МОДЕЛИРОВАНИИ\*

А.Г. Подвесовский (*apodv@tu-bryansk.ru*)

Р.А. Исаев (*ruslan-isaev-32@yandex.ru*)

И.А. Копелиович (*igor.copeliovitich@yandex.ru*)

Брянский государственный технический университет, Брянск

В работе рассматриваются задачи экспериментального исследования новых методов когнитивного моделирования путем их апробации на множестве искусственно созданных когнитивных моделей. Исследованы особенности процесса получения таких моделей, введены понятия формальной когнитивной карты и формальной когнитивной модели. Предложен и частично формализован подход к генерации формальной когнитивной карты с заданными характеристиками, основанный на использовании моделей случайных графов. Выполнена модификация конфигурационной модели, и приведены примеры генерации формальных нечетких когнитивных карт на ее основе.

**Ключевые слова:** когнитивное моделирование, нечеткая когнитивная карта, модели случайных графов, конфигурационная модель.

### Введение

В настоящее время наблюдается устойчивый интерес к моделированию слабо структурированных проблем и сложных систем на основе когнитивного подхода, о чем свидетельствуют многочисленные публикации как методологического, так и прикладного характера. При этом в ходе соответствующих исследований часто возникают новые задачи, связанные с развитием математического и алгоритмического обеспечения когнитивных моделей, с целью совершенствования методов когнитивного моделирования и придания им новых возможностей. Имеющийся у авторов опыт работы с когнитивными моделями, основанными на нечетких когнитивных картах (НКК) В.Б. Силова [Силов, 1995], включая создание информа-

---

\* Работа выполнена при финансовой поддержке РФФ (проект № 23-19-00342).

ционных технологий и программных средств поддержки когнитивного моделирования, а также решения прикладных задач на их основе, позволил сформировать ряд актуальных направлений развития математического аппарата когнитивного моделирования применительно к данному типу НКК. За последние годы был предложен ряд решений в части структурной и параметрической идентификации НКК Силова, верификации, сценарного анализа, а также визуализации и визуальной аналитики когнитивных моделей [Isaev et al., 2017], [Подвесовский и др., 2019], [Исаев и др., 2020] и др.

Опыт исследований в перечисленных направлениях показал, что важным этапом разработки или модификации любого метода или алгоритма работы с когнитивной моделью является экспериментальная проверка его эффективности и корректности. В большинстве случаев подобная проверка проводится на специально создаваемом тестовом наборе когнитивных моделей, структура и параметры которых определяют особенностями экспериментальной задачи. Ручное формирование такого набора представляет собой достаточно трудоемкую задачу, особенно если требуется, чтобы он содержал большое число тестовых примеров.

В настоящей работе предлагается подход к автоматизированной генерации когнитивных моделей для использования в указанных экспериментальных исследованиях, и рассматривается его ключевой этап – генерация структуры НКК на основе моделей случайных графов.

## **1. Особенности экспериментального исследования методов нечеткого когнитивного моделирования**

Процесс экспериментального исследования методов когнитивного моделирования в общем случае характеризуется следующими чертами:

- Структура и, при необходимости, параметры когнитивной карты подбираются с учетом сути и специфики разработанного метода, этапа моделирования, на котором он применяется, а также тех или иных аспектов и особенностей его работы, тестированию и исследованию которых требуется уделить внимание. Так, в работе [Isaev et al., 2017], затрагивающей этап сценарного анализа когнитивной модели, ключевым требованием к структуре когнитивного графа являлось наличие в нем цикла. Тестирование методов идентификации НКК, рассмотренных в [Подвесовский и др., 2019], подхода к верификации НКК, предложенного в [Исаев и др., 2020], как и других методов и алгоритмов, также накладывало ряд требований к структуре НКК.
- Наличие у когнитивных моделей смыслового содержания, как правило, не является обязательным условием в рамках задачи тестирования формальных методов. Так, искусственно созданные «обезли-

ченные» когнитивные модели (отметим, что именно такие использовались в упомянутых авторских статьях) оказываются в данном случае пригодными в той же мере, что и «реальные» когнитивные модели конкретных систем и проблемных ситуаций.

- Полнота экспериментальной проверки и исследования разработанного метода (и, следовательно, степень доверия к результатам его работы при дальнейшем использовании в реальных задачах) естественным образом зависит от числа и от разнообразия проведенных экспериментов – то есть, в общем случае, от количества различных когнитивных моделей, на множестве которых происходило тестирование метода.

Сочетание перечисленных факторов обуславливает актуальность автоматизации процесса генерации нечетких когнитивных моделей с заданными характеристиками для обеспечения проводимых экспериментальных исследований при разработке методов поддержки когнитивного моделирования.

С учетом структуры НКК Силова, очевидно, что задача генерации когнитивной модели, основанной на данном типе НКК, в основной своей части сводится к задаче генерации связного ориентированного взвешенного графа, с учетом ряда возможных ограничений, требований и параметров [Касьянов и др., 2003].

Дальнейшее изложение материала построено следующим образом. Вводится и описывается понятие формальной нечеткой когнитивной карты (везде далее, без дополнительных уточнений, подразумеваются НКК Силова) и основанной на ней формальной нечеткой когнитивной модели. Предлагается математическая модель генерации формальных нечетких когнитивных моделей. Основным компонентом модели является алгоритм, генерирующий когнитивный граф в соответствии с определенной моделью случайных графов. Проводится обзор ряда моделей случайных графов и реализующих их алгоритмов, а также предлагается модификация алгоритма, признанного наиболее подходящим для применения к решаемой задаче, и приводятся примеры его применения в базовой и модифицированной версиях.

## **2. Формализованное представление задачи генерации формальных нечетких когнитивных моделей**

Под *формальной нечеткой когнитивной картой* будем понимать такую НКК, структура и параметры которой удовлетворяют формальным требованиям, предъявляемым к данному классу моделей, и которая при этом не является моделью какой-либо реальной системы или ситуации.

Последнее условие подразумевает, в первую очередь, *формальность концептов* – то есть отсутствие связи между концептами такой НКК и семантическими категориями «внешнего» по отношению к модели мира. С технической точки зрения это может выражаться в отсутствии у концептов имен или, если наличие у концепта имени требуется по тем или иным причинам (например, для его идентификации), в типовом «обезличенном» формате их имен.

Таким образом, формальная НКК может быть получена следующими способами:

- сформирована на основе любой предметной (т.е. построенной для решения прикладной задачи) НКК, путем отбрасывания имен концептов или их замены на типовые имена («Концепт 1», «Концепт 2», ...);
- построена исследователем самостоятельно, с учетом целей ее дальнейшего использования и вытекающих из этих целей требований к структуре и параметрам НКК;
- создана путем автоматизированной генерации когнитивного графа, удовлетворяющего как общим формальным требованиям к НКК, так и, опционально, специфическим требованиям, которые отдельно формулируются исследователем с учетом особенностей решаемой задачи.

Практический интерес в рамках данной работы представляет, очевидно, последний способ, и далее рассмотрим математическую модель соответствующего процесса.

Следуя приведенному в [Подвесовский и др., 2019] определению НКК, обозначим формальную НКК следующим образом:

$$G_F = \langle E_F, W \rangle,$$

где  $E_F$  – множество формальных концептов,  $E_F = \{e_1, e_2, \dots, e_n\}$ , а  $W$  – нечеткое бинарное отношение на множестве  $E_F$ , задающее набор влияний между его элементами.

Тогда под *формальной нечеткой когнитивной моделью* будем понимать когнитивную модель, основанную на формальной НКК:

$$M_F = \langle G_F, E_C, E_T, E_O, V_1, V_T, I_E \rangle,$$

где

$E_C \subseteq E_F$  – множество управляемых концептов;

$E_T \subseteq E_F$  – множество целевых концептов;

$E_O \subseteq E_F$  – множество концептов, подверженных воздействиям со стороны внешней среды;

$V_1 = \{v_1^1, v_2^1, \dots, v_n^1\}$  – множество начальных состояний всех концептов,  $v_i^1 \in [0, 1]$  для всех  $i \in [1, n]$ ;

$V_T = \{v_1^T, v_2^T, \dots, v_k^T\}$  – множество целевых состояний целевых концептов,  $v_i^T \in [0, 1]$  для всех  $i \in [1, k]$ ;

$I_E = \{i_1, i_2, \dots, i_q\}$  – множество воздействий внешней среды на концепты, подверженные этим воздействиям, при этом  $i_j = \langle t, e, \Delta v \rangle$ ,

где

$t$  – такт моделирования, на котором осуществляется воздействие ( $t = 1, 2, \dots$ );

$e$  – концепт, его испытывающий ( $e \in E_O$ );

$\Delta v$  – изменение состояния концепта в результате воздействия.

Следует отметить, что допускается пересечение множеств  $E_C$ ,  $E_T$  и  $E_O$ .

Тогда *модель генерации* формальной нечеткой когнитивной модели можно представить следующим образом:

$$P = \langle R, A, p_A; M_F \rangle,$$

где

$R$  – набор формализованных требований к получаемой модели;

$A$  – алгоритм генерации случайных графов, используемый для получения когнитивного графа (выбирается с учетом набора требований  $R$ );

$p_A$  – параметры и настройки алгоритма  $A$  (их наличие и состав определяются спецификой выбранного алгоритма, а значения определяются с учетом набора требований  $R$ ).

При этом набор требований  $R$  может включать в себя множество различных видов информации о структуре и параметрах когнитивной модели, которую требуется сгенерировать. Перечислим наиболее важные из этих видов информации.

- Требуемое число концептов  $n$  в составе  $G_F$ . Допускается как указание конкретного значения  $n$ , так и задание диапазона  $[n_{\min}, n_{\max}]$  с указанием типа и параметров распределения значения  $n$  как случайной величины на данном диапазоне.
- Требуемое число отношений влияния в составе  $G_F$  (т.е. число ненулевых элементов  $W$ ). Как правило, явное указание этого значения не является интуитивно понятным вариантом, и целесообразнее использовать косвенные способы его задания. Самым простым из таких способов является указание средней степени  $\deg(e_i)$  вершин когнитивного графа (с каким количеством других концептов связан, в среднем, концепт в данной НКК). Более сложный и гибкий способ состоит в описании типа и параметров вероятностного рас-

пределения степеней вершин – это, в частности, позволит формировать модели с более или менее выраженными концептами-«ядрами». Также допустимым вариантом может являться прямое задание степени для каждой из  $n$  вершин графа.

- Информация о допустимости или предпочтительности наличия в структуре когнитивного графа следующих ситуаций:
  - петли (ситуация влияния концепта на этот же концепт);
  - кратные дуги (ситуация взаимного прямого влияния двух концептов друг на друга);
  - ориентированные циклы (соответствуют контурам обратной связи в НКК, интерес также может представлять задание их количества);
  - узлы без исходящих дуг (отражают концепты способные только испытывать влияния со стороны других концептов);
  - узлы без входящих дуг (отражают концепты, способные только оказывать влияния на другие концепты)

Отметим, что все указанные виды информации могут влиять как на выбор используемого алгоритма генерации графа  $A$ , так и на задание его параметров. При этом добавление в модель  $M_F$  остальных требуемых элементов (интенсивностей влияний, значений концептов и т.д.) представляет собой следующий после генерации графа этап, который выходит за рамки рассмотрения данной работы. Далее будут рассмотрены некоторые существующие модели случайных графов, с выбором наиболее подходящей для решения задачи генерации формальной НКК.

### **3. Исследование моделей случайных графов в контексте задачи генерации нечетких когнитивных карт**

#### **3.1. Обзор моделей случайных графов применительно к задаче генерации нечетких когнитивных карт**

Отметим, что все рассматриваемые далее модели являются моделями неориентированных графов. В то же время, преобразование такого графа в ориентированный (путем указания направлений ребер), как правило, не вызывает затруднений.

Одна из наиболее изученных и простых в реализации моделей случайных графов – модель Эрдеша-Реньи [Райгородский, 2011]. Алгоритм генерации случайного графа в соответствии с данной моделью состоит в следующем. задается желаемое число вершин графа  $n$ . Далее составляется множество  $K_n$ , включающее все возможные ребра в графе с числом вершин  $n$ . Каждое ребро из множества  $K_n$  включается в граф с заданной вероятностью  $p \in [0, 1]$ , независимо от остальных ребер. В результате выполнения данного алгоритма формируется случайный граф, средняя степень вершины в котором равна  $np$ .

Другая заслуживающая рассмотрения модель случайных графов – модель Барабаши-Альберт [Albert et al., 2002]. Графы, построенные согласно данной модели, близки по своим свойствам к безмасштабным сетям [Barabási et al., 1999]. Алгоритм генерации случайного графа в данном случае подразумевает последовательное добавление вершин в граф и формирование новых ребер в соответствии с принципом предпочтительного присоединения. Согласно данному принципу, чем большую степень имеет каждая уже добавленная в граф вершина, тем более вероятно присоединение новой вершины к ней.

В своем исходном виде обе рассмотренных модели слабо применимы к задаче генерации НКК по следующим причинам.

1. Лежащие в их основе алгоритмы не гарантируют получения связанного графа.
2. Если количество вершин в генерируемом графе невелико (так, для НКК характерно наличие не более чем нескольких десятков концептов), то фактически получаемая средняя степень вершины может значительно отличаться от заданного желаемого значения.
3. Отсутствует возможность задавать тип распределения степеней вершин: так, в модели Эрдеша-Реньи получаемый граф всегда имеет биномиальное распределение, а для модели Барабаши-Альберт характерно получение степенного распределения.
4. Отсутствует возможность задавать значения других важных параметров графа (например, количество циклов).

Наиболее гибкой моделью случайных графов в контексте решаемой задачи представляется конфигурационная модель [Hofstad, 2017], поскольку она позволяет генерировать графы с любым желаемым распределением степеней вершин. При этом допускается задание информации о степенях вершин путем явного перечисления требуемых степеней всех вершин либо путем задания типа и параметров вероятностного распределения степеней вершин.

Алгоритм генерации случайного графа в рамках конфигурационной модели включает в себя следующие этапы.

1. Создается список  $L$ , в котором содержится по  $d_i$  копий каждой  $i$ -й вершины (где  $d_i$  – степень данной вершины).
2. Пока список  $L$  не пуст, из него случайным образом извлекается пара копий, и соответствующая пара вершин соединяется ребром.

В итоге формируется случайный граф, степени вершин которого строго соответствуют заданным.

Кроме того, модель допускает генерацию графа с заданной средней степенью вершины, однако для этого также необходимо сформировать набор степеней вершин, и данная операция потребует информации о типе



и параметрах распределения этих степеней. Количество циклов, как и в предыдущих рассмотренных моделях, задать не представляется возможным. Также не гарантируется связность графа. В частности, чем большее число вершин имеют малую степень, тем выше риск того, что граф будет иметь больше одной компоненты связности. Таким образом, использование конфигурационной модели требует ряда доработок, главной из которых является обеспечение связности графа.

### **3.2. Модификация алгоритма конфигурационной модели для обеспечения связности получаемого графа**

Чтобы снизить вероятность образования нескольких компонент связности, можно выбирать пары вершин не случайным образом, как в базовой версии алгоритма, а с учетом степеней обеих вершин в составе каждой такой пары.

Модифицированный алгоритм включает в себя следующие шаги.

1. Создается список  $L$ , в котором содержится по  $d_i$  копий каждой  $i$ -й вершины (где  $d_i$  – степень данной вершины).
2. Пока список  $L$  не пуст, из него извлекается копия той вершины, которая на данный момент имеет наименьшее число копий в этом списке (если таких вершин несколько, случайным образом выбирается одна из них).
3. Аналогичным образом из списка  $L$  извлекается копия той вершины, которая на данный момент имеет наибольшее число копий в списке.
4. Вершины, копии которых выбраны на шаге 2 и шаге 3, соединяются ребром. Далее происходит возврат к шагу 2.

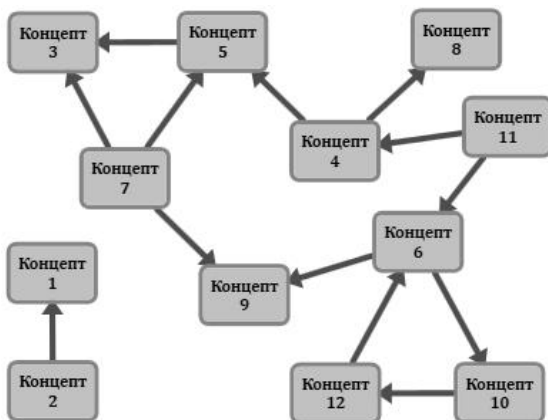
Таким образом, модификация алгоритма гарантирует соединение между собой вершин с малыми степенями и вершин с относительно большими степенями. Тем самым, уменьшается вероятность того, что вершины с малыми степенями будут образовывать ребра друг с другом, обособляясь в отдельную компоненту связности.

### **3.3. Примеры генерации нечеткой когнитивной карты на основе базовой и модифицированной конфигурационных моделей**

Рассмотрим результаты работы двух вариантов конфигурационной модели на примере генерации НКК, состоящей из 12 концептов ( $n = 12$ ).

Информацию о количестве отношений влияния между концептами зададим путем явного указания последовательности степеней вершин: 1, 1, 2, 3, 3, 4, 3, 1, 2, 2, 2, 2 (в данной последовательности  $i$ -й элемент задает степень  $i$ -й вершины когнитивного графа, то есть число связей  $i$ -го концепта с другими концептами). Другие типы информации о структуре и параметрах НКК не задавались.

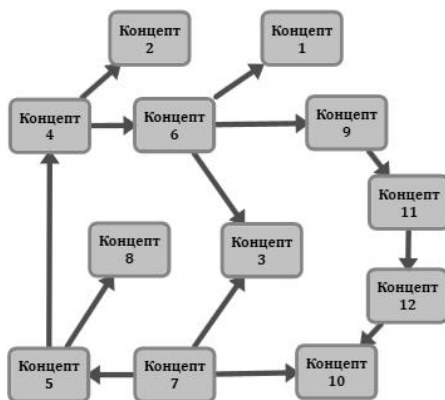
На рис. 1 приведена формальная НКК, сгенерированная с применением базового алгоритма конфигурационной модели. Полученный граф имеет две компоненты связности. При выборе пар из списка  $L$  (в соответствии с описанным выше алгоритмом) были случайно выбраны вершины 1 и 2. Поскольку обе вершины имеют степень 1, то их соединение между собой привело к появлению дополнительной компоненты связности.



*Рис. 1. Формальная нечеткая когнитивная карта, полученная на основе базового алгоритма конфигурационной модели*

При этом при генерации модифицированным алгоритмом (рис. 2) получаемый граф имеет одну компоненту связности, т.е. выполняется ключевое для формирования корректной НКК условие связности ее графа. В частности, это было достигнуто за счет того, что вершина 1, согласно алгоритму, была соединена с вершиной 6, имеющей наибольшую степень среди всех вершин графа.

Следует отметить, что при использовании базовой конфигурационной модели генерируемый граф каждый раз получается случайным, даже в условиях одинаковых параметров. В случае модифицированной модели генерируемые с одинаковыми параметрами графы часто получают идентичными по структуре. Различия же могут возникать за счет ситуаций, когда в процессе выполнения алгоритма какой-либо выбор делается между двумя или более вершинами с одинаковыми степенями.



*Рис. 2. Формальная нечеткая когнитивная карта, полученная на основе модифицированного алгоритма конфигурационной модели*

## Заключение

В работе была рассмотрена задача обеспечения экспериментальных исследований методов когнитивного моделирования, и показано, что важной составляющей ее решения является автоматизация построения искусственных когнитивных моделей с заданными структурой и параметрами. С этой целью был предложен и частично формализован подход, основанный на применении моделей случайных графов. Среди ряда рассмотренных моделей была выбрана конфигурационная модель как наиболее перспективная для использования в рамках указанного подхода, и предложена ее модификация с учетом специфики решаемой задачи.

Ключевым направлением дальнейшей доработки конфигурационной модели является обеспечение возможности задавать количество ориентированных циклов в генерируемой НКК, поскольку рассмотренные варианты данной модели не позволяют задать даже примерное количество циклов в получаемом графе. Тем не менее, данный параметр представляет значительный интерес при генерации когнитивных моделей, особенно при исследовании методов сценарного анализа. Одним из решений здесь может стать вспомогательный алгоритм, выполняемый перед реализацией основного алгоритма, результатом которого будет являться базовый граф с требуемым количеством ориентированных циклов, и далее к нему могут добавляться остальные необходимые вершины и дуги.

Основным прикладным направлением дальнейших исследований является создание специальной исследовательской подсистемы для разрабатываемой авторским коллективом системы поддержки принятия решений

на основе нечетких когнитивных моделей «ИГЛА»<sup>1</sup>. Эта подсистема будет использована для автоматизации проведения экспериментальных исследований методов и алгоритмов когнитивного моделирования, и ее важнейшим элементом станет модуль генерации формальных нечетких когнитивных моделей с заданными параметрами, основанный на доработанной конфигурационной модели (в том числе с учетом ее перспективных доработок). Такая подсистема должна иметь развитый пользовательский интерфейс, позволяющий в удобной форме задавать интересующие пользователя параметры создаваемой когнитивной модели.

### Список литературы

- [Albert et al., 2002] Albert R., Barabási A.-L. Statistical mechanics of complex networks // *Reviews of Modern Physics*. – 2002. – Vol. 74, Issue 1.
- [Barabási et al., 1999] Barabási A.-L., Albert R. Emergence of scaling in random networks // *Science*. – 1999. – Vol. 286.
- [Hofstad, 2017] Hofstad R. *Random graphs and complex networks*. Vol. 1. – New York, Cambridge University Press, 2017.
- [Isaev et. al., 2017] Isaev R.A., Podvesovskii A.G. Generalized Model of Pulse Process for Dynamic Analysis of Sylov’s Fuzzy Cognitive Maps // *CEUR Workshop Proceedings of the Mathematical Modeling Session at the International Conference Information Technology and Nanotechnology (MM-ITNT 2017)*. – Vol. 1904.
- [Исаев и др., 2020] Исаев Р.А., Подвесовский А.Г. Верификация причинно-следственных связей в когнитивных моделях на основе применения метафор визуализации нечетких когнитивных карт // *Научная визуализация*. – 2020. – Т. 12, № 4.
- [Касьянов и др., 2003] Касьянов В.Н., Евстигнеев В.А. *Графы в программировании: обработка, визуализация и применение*. – СПб.: БХВ-Петербург, 2003.
- [Подвесовский и др., 2019] Подвесовский А.Г., Исаев Р.А. Идентификация структуры и параметров нечетких когнитивных моделей: экспертные и статистические методы // *International Journal of Open Information Technologies*. – 2019. – Т. 7, № 6.
- [Райгородский, 2011] Райгородский А.М. *Модели случайных графов*. – М.: МЦНМО, 2011.
- [Силов, 1995] Силов В.Б. *Принятие стратегических решений в нечеткой обстановке*. – М.: ИНПРО-РЕС, 1995.

---

<sup>1</sup> <http://iipo.tu-bryansk.ru/quill/>.

УДК 004.822

## МЕТОД ОЦЕНИВАНИЯ СЛОЖНЫХ СИСТЕМ И ПРОЦЕССОВ НА ОСНОВЕ ОНТОЛОГИЧЕСКОГО И НЕЙРО-НЕЧЕТКОГО МОДЕЛИРОВАНИЯ\*

И.А. Прокопенков (*prokopenkoff.ivan@yandex.ru*)

И.А. Фролов (*igor-frolov-81@mail.ru*)

Военная академия войсковой противовоздушной обороны ВС РФ,  
Смоленск

В.И. Меденников (*dommed@mail.ru*)

Федеральный исследовательский центр  
«Информатика и Управление» РАН, Москва

Для различных отраслей экономики (в промышленности, в топливно-энергетическом комплексе, в агропромышленном комплексе, на транспорте, в сфере обслуживания) обеспечение взаимодействия информационных, управленческих, технологических процессов остро ставит проблему качества и оперативности их оценивания. В работе представлен метод оценивания сложных систем и процессов, использующий представление знаний проблемной области в виде онтологической модели, а также нейро-нечеткий классификатор для оценивания имеющихся прецедентов.

**Ключевые слова:** онтологическая модель, прецедент, нейро-нечеткий классификатор.

### Введение

Практически для всех отраслях экономики (в промышленности, в топливно-энергетическом комплексе, в агропромышленном комплексе, на транспорте, в сфере обслуживания) тенденция к тесному взаимодействию информационных, информационно-аналитических, управленческих, основных технологических и вспомогательных процессов остро ставит проблему качества и оперативности оценивания этих процессов, обусловленную:

---

\* Работа поддержана грантом Министерства науки и высшего образования РФ (внутренний номер 00600/2020/51896, договор № 075-15-2022-319).

- необходимостью обработки больших объемов разнородной и разноформатной информации;
- противоречием между постоянно накапливаемыми объемами данных, поступающими из различных источников, и отсутствием типовых подходов к оцениванию, агрегированию и обобщению данных;
- противоречием между, с одной стороны, сложностью и многоаспектностью моделей, описывающих отдельные аспекты сложной системы, недостаточной точностью моделирования, большим количеством параметров и показателей, используемых в этих моделях, и, с другой стороны, отсутствием эффективных методов обобщения, анализа и извлечения знаний из больших информационных массивов разнокачественных данных, позволяющих построить комплексную информационную модель для оперативного оценивания показателей эффективности системы;
- сложностью интеграции информационных ресурсов, приложений и инструментов в единую информационно-управленческую среду [Зацаринный и др., 2023].

В работе представлен метод, направленный на повышение качества и оперативности оценивания сложных систем и процессов за счет представления знаний проблемной области в виде онтологической модели, формирования относительно ее классов баз нечетких прецедентов, построения и обучения модели оценивания в виде каскадного нейро-нечеткого классификатора.

## **1. Построение онтологической модели проблемной области**

Анализ существующих подходов по представлению знаний позволяет обосновать целесообразность применения методов и средств онтологического инжиниринга для формализованного представления различных проблемных областей. Его целью является повышение эффективности информационного поиска, обеспечение интеграции разнородной информации, выявления закономерностей на основе единого семантического описания пространства знаний для принятия эффективных управленческих решений.

Онтологический подход обеспечивает целостный взгляд на проблемную область, позволяет сделать знания доступными, обеспечивает их неоднократное использование, тиражируемость, адаптивное наследование.

Она представляется в виде иерархической структуры, концепты которой представляют сущности проблемной области, а связи между ними характеризуют отношения различной семантики [Гаврилова и др., 2016], [Ярушкина и др., 2018], [Грибова и др., 2020], [Грибова и др., 2022].

В качестве иллюстрации предлагаемого метода рассмотрим задачу взаимодействия поставщиков, логистических компаний и потребителей при выполнении заказов.

*Этап 1.* Построение онтологической модели проблемной области.

Онтологию можно представить в виде [Загоруйко и др., 2020]:

$$O = \langle C, E, R, T, D, A, F, Ax \rangle,$$

где  $C = \{C_1, \dots, C_i\}$  – конечное множество классов и подклассов;

$E = \{E_1, \dots, E_j\}$  – конечное множество экземпляров классов и подклассов описывающих конкретные понятия проблемной области;

$R = \{R_1, \dots, R_g\}$  – конечное множество отношений между классами и подклассами проблемной области;

$T = \{T_1, \dots, T_k\}$  – множество стандартных типов данных;

$D = \{C_i\{A_1, \dots, A_p\}, \dots, C_i\{A_1, \dots, A_h\}\}$  – множество доменов, которые позволяют объединять классы (подклассы) и их экземпляры по совокупности характеризующих их атрибутов;

$Ac = \{Ac_1, \dots, Ac_n\}$  – конечное множество атрибутов, описывающих свойства понятий  $C$  и отношений  $R$ . Среди множества атрибутов выделяют ключевые, позволяющие однозначно идентифицировать класс, подкласс и их экземпляры;

$fc = \{fc_1, \dots, fc_m\}$  – множество ограничений значения атрибутов;

$Ax = \{Ax_1, \dots, Ax_q\}$  – множество аксиом, определяющих дополнительную семантику классов, подклассов и отношений.

Формализация и представление проблемной области подразумевает работу инженера по знаниям с применением коммуникативных и текстологических методов для извлечения знаний, его последующего анализа и формализации [Гаврилова и др., 2001].

Для оперативного оценивания сложных систем и процессов, необходимо выделять знания, характеризующие задачи и подзадачи в отдельные классы и подклассы. Пример формализации задач проблемной области представлен на рис. 1.

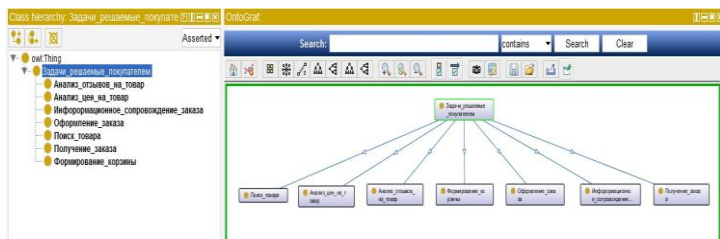


Рис. 1. Формализованное представление задач взаимодействия поставщиков, логистических компаний и потребителей с использованием возможностей Protege 5.0

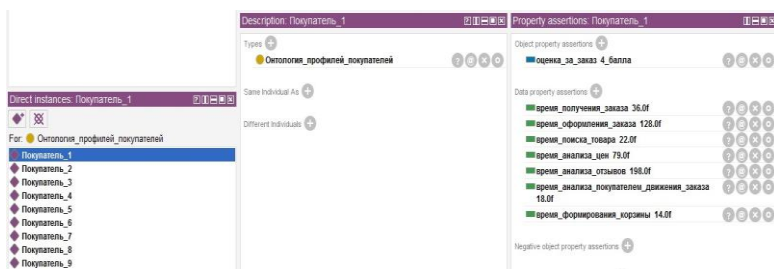
Помимо классов и подклассов задач необходимо сформировать класс профилей потребителей.

*Этап 2.* Формирование базы прецедентов.

Для хранения информации о выполняемых заказах, в классе профилей потребителей необходимо сформировать базу прецедентов. Прецедентом в данном случае будет являться набор входных атрибутов, характеризующих время, затраченное на задачи, решаемые в ходе выполнения заказа, и отзыв потребителя на выполненный заказ. Пример «наполнения» класса онтологии прецедентами выполнения заказа поставщиком представлен на рис. 2.

*Этап 3.* Формирование базы нечетких прецедентов.

Эффективность реализации прецедентного подхода зависит от размеров базы прецедентов [Прокопенков и др., 2020].



*Рис. 2. Пример наполнения класса «Онтология профилей потребителей» прецедентами выполненных заказов*

Нечеткая грануляция информации позволяет существенно сократить количество прецедентов в базе за счет нечеткого представления значений атрибутов в виде функций принадлежности термов лингвистических переменных. Для их построения могут быть использованы как прямые, так и косвенные методы.

Прямые методы основаны на непосредственном задании функций принадлежности нечетких множеств на основе коммуникативных действий при извлечении экспертных знаний. Их целесообразно использовать для измеримых атрибутов (переменных, показателей).

Косвенные методы определения значений функций принадлежности используются в случаях, когда нет измеримых показателей. Эти методы реализуются на основе: попарных сравнений, статистических данных, типовых *L-R* функций [Борисов и др., 2014].

В качестве примера рассмотрим использование прямого подхода к построению функций принадлежности термов лингвистической переменной для времени, затрачиваемого на выполнение заказа, на основе прямого подхода.



Формализация такого описания может быть проведена с помощью лингвистической переменной  $\langle \beta, Q, X, P, M \rangle$ , со следующими параметрами:  $\beta$  – наименование лингвистической переменной «Временные затраты потребителей на оформление заказа»;  $Q$  – терм-множество лингвистической переменной {«Малые временные затраты», «Средние временные затраты», «Большие временные затраты»};  $fc_1$  = «Малые временные затраты»,  $fc_2$  = «Средние временные затраты»,  $fc_3$  = «Большие временные затраты» – нечеткие множества, соответствующие термам лингвистической переменной  $Q$  (определяются на основе значений атрибутов онтологической модели);  $X$  универсальное множество, являющееся областью определения нечетких множеств каждой переменной из терм-множества  $Q$ ;  $P$  – синтаксическая процедура образования новых термов с помощью связок «И», «ИЛИ» и модификаторов типа «ОЧЕНЬ», «НЕ», «ПОЧТИ», например: «Малые ИЛИ Средние», «НЕ Большие»;  $M$  – процедура задания на  $X$  нечетких множеств  $fc_1$  = «Малые временные затраты»,  $fc_2$  = «Средние временные затраты»,  $fc_3$  = «Большие временные затраты», а также нечетких множеств для термов из  $P(Q)$  в соответствии с правилами трансляции нечетких связок и модификаторов «И», «ИЛИ», «НЕ», «ОЧЕНЬ», «ПОЧТИ» над нечеткими множествами, например, вида:  $fc_1 \cap fc_2, fc_1 \cup fc_2$ .

Перспективным представляется декомпозиция прецедентов относительно взаимосвязанных частных задач оценки, что в последующем позволяет выполнить декомпозицию соответствующих модели оценивания.

Вариант представления базы нечетких прецедентов, ориентированной на последующую декомпозицию модели оценивания, представлен на рис. 3.

Исходные данные (атрибуты) $A_{c_i}$ для решения задачи «Поиск товара 1»	Прецедент 1 (значения атрибутов) ( $f_c$ )	...	Прецедент N (значения атрибутов) ( $f_c$ )
Артикул товара	1	...	100
Уровень активности покупателя	активный	...	неактивный
Временные затраты на поиск товара	малые	...	средние
Выходные данные (атрибуты)	значения атрибутов	...	значения атрибутов
Товар найден	да	...	нет
Покупатель приступил к анализу отзывов и цен	да	...	нет
Количество кликабельных действий при поиске товара	10	...	37

Исходные данные (атрибуты) $A_{c_i}$ для решения задачи «Анализ отзывов и цен на товар 2»	Прецедент 1 (значения атрибутов) ( $f_c$ )	...	Прецедент N (значения атрибутов) ( $f_c$ )
Артикул товара	1	...	100
Покупатель приступил к анализу отзывов и цен	да	...	нет
Количество кликабельных действий при поиске товара	10	...	37
Временные затраты на анализ отзывов и цен на товар	средние	...	нет
Выходные данные (атрибуты)	значения атрибутов	...	значения атрибутов
Количество кликабельных действий при анализе отзывов и цен на товар	45	...	0
Покупатель добавил товар в корзину	да	...	нет

Рис. 3. Пример представления базы нечетких прецедентов для задач «Поиск заказа» и «Анализ отзывов и цен на заказ», ориентированной на декомпозицию модели оценивания

## 2. Построение нейро-нечеткого классификатора

Указанные ранее особенности обуславливают целесообразность использования каскадного нейро-нечеткого классификатора для оценивания прецедентов.

Традиционный подход к классификации прецедентов основан на предварительной кластеризации обучающих примеров и отнесении их к заданным классам. Ограничения при осуществлении этого предварительного этапа в большой степени обусловлены сложностью определения границ между кластерами. Эта проблема становится еще более трудно решаемой, когда число используемых параметров существенно возрастает. В отличие от традиционного подхода, нечеткая классификация допускает непрерывность границы между двумя соседними классами с наложением областей, в каждой из которых классифицируемый прецедент характеризуется своей степенью принадлежности. Данный подход не только подходит для сравнения прецедентов, характеризующихся нечеткими границами между классами, но также обеспечивает достаточно простое представление потенциально сложного разделения пространства признаков [Борисов и др., 2016].

Например, для  $H$  прецедентов в виде  $b$ -мерных четких векторов, где  $h=1, \dots, H$ , которые относятся к двум классам, типичные нечеткие правила классификации для  $b=2$  имеют следующий вид:

*Пр<sub>1</sub>* :ЕСЛИ  $Ac_1$  есть *малые врем. затраты* И  $Ac_2$  есть *средние врем. затраты*, ТО  $PE_{11}$  принадлежит к классу оценок  $W_4$ ;

*Пр<sub>2</sub>* :ЕСЛИ  $Ac_1$  есть *малые врем. затраты* И  $Ac_2$  есть *большие врем. затраты*, ТО  $PE_{12}$  принадлежит к классу оценок  $W_3$ ,

где  $Ac = \{Ac_1, \dots, Ac_n\}$  – конечное множество атрибутов, описывающих временные затраты на отдельные задачи выполняемые пользователем при совершении покупки;

$PE = \{PE_{11}, \dots, PE_{sj}\}$  – конечное множество прецедентов выполненных пользователем  $E_j$ ;

$W = \{W_1, \dots, W_3\}$  – классы оценок за выполненный заказ (1 балл, ..., 5 баллов, соответственно).

Аналогичным образом осуществляется генерирование правил для нечеткого разбиения пространства входных атрибутов.

Нейро-нечеткий классификатор имеет многослойную структуру с прямым распространением сигнала.

*Слой 1.* На выходе элементов этого слоя формируется степени принадлежности входных переменных к нечетким множествам  $fc_c$ :

$$\mu_{f_{c_{ie}}}(Ac'_e) = \exp\left(-\frac{1}{2}\left(\frac{Ac'_e - a_e}{b_{ie}}\right)^2\right),$$

где  $a_{ie}, b_{ie}$  – параметры функции принадлежности колоколообразного типа.

Начальные значения этих параметров определяются экспертным методом, чтобы функции принадлежности удовлетворяли свойствам полноты, нормальности и выпуклости. Значения параметров могут корректироваться в процессе обучения.

Колоколообразный тип функции принадлежности позволяет оперативно осуществлять ее настройку, так как она по сравнению с аналогами имеет всего два настраиваемых параметра  $a_{ie}, b_{ie}$ .

*Слой 2.* Каждый элемент данного слоя является нечетким нейроном, реализующим операцию «И» над нечеткими множествами. Структура нечеткого нейрона «И» представлена на рис. 4.

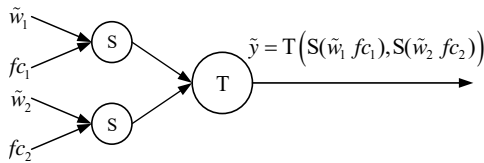


Рис. 4. Структура нечеткого нейрона «И»

В данном нечетком нейроне нечеткая выходная переменная формируется за счет комбинирования значений нечетких входных переменных  $fc_i$  и нечетких весов  $w_i$  с помощью операции s-нормы, и агрегирования полученных промежуточных результатов с помощью операции t-нормы:

$$\tilde{y} = T(S(\tilde{w}_1, fc_1), S(\tilde{w}_2, fc_2)).$$

Если  $T - \min$ ,  $S - \max$ , то нечеткий нейрон «И» реализует операцию min-max-композиции:

$$\tilde{y} = \min(\max(\tilde{w}_1, fc_1), \max(\tilde{w}_2, fc_2)).$$

Выходной сигнал второго слоя представляет «силу» срабатывания нечеткого правила относительно классифицируемого прецедента.

*Слой 3.* Элементы этих слоев предназначены для взвешенного суммирования значений выходов элементов предыдущего слоя.

*Слой 4.* Значения на выходах элементов слоя 4 формируются с использованием активационных функций сигмоидального типа. Функция сигмоидального типа представляет собой гладкую монотонно возрастающую нелинейную функцию:

$$\sigma(fc_i) = \frac{1}{1 + e^{-fc_i}}.$$

Выбор данного типа функции обусловлен необходимостью сглаживания выходных значений.

Выходы четвертого слоя трактуются как степени принадлежности предъявленной ситуации к соответствующему прецеденту [Борисов и др., 2016].

Так как для решаемых задач «Поиск заказа» и «Анализ отзывов и цен на заказ» сформированы связанные между собой базы нечетких прецедентов, то оценивания используется двухкаскадный нейро-нечеткий классификатор.

В 1-м каскаде классификатора осуществляется оценивание прецедентов по задаче «Поиск заказа». Выходы 1-го каскада, совместно с дополнительными значениями атрибутов, подаются на входы 2-го каскада. На выходе двухкаскадного нейро-нечеткого классификатора формируются степени принадлежности влияния атрибутов прецедента на отнесение его к соответствующему классу оценок.

Обобщенная структура двухкаскадного нейро-нечеткого классификатора представлена на рис. 5.

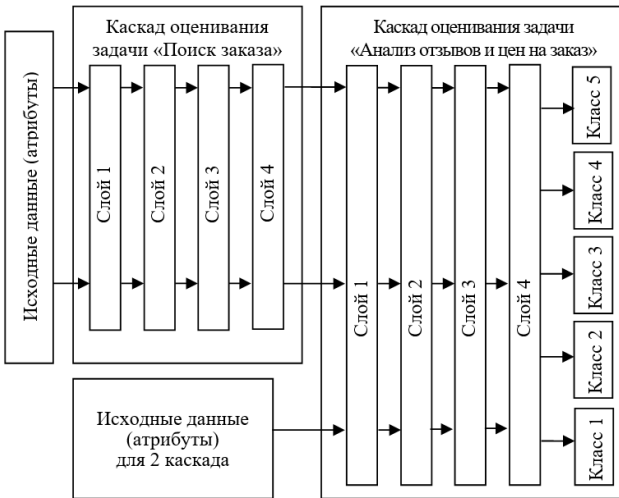


Рис. 5. Структура двухкаскадного нейро-нечеткого классификатора для оценивания зависимости времени выполнения заказа

Особенностью реализуемого подхода является, то, что при появлении новых атрибутов, изменения в онтологической модели, автоматически будут учтены в базе прецедентов.

Так как нейро-нечеткий классификатор представляется в виде многослойной структуры с настраиваемыми параметрами элементов слоев, то для его структурно-параметрической настройки можно использовать широкий спектр алгоритмов обучения искусственных нейронных сетей [Борисов и др., 2016].

### 3. Реализация предлагаемого метода

Выполнено прототипирование системы, реализующей предлагаемый метод оценивания сложных систем и процессов на основе онтологического и нейро-нечеткого моделирования. Структура системы представлена на рис. 6.

Метод может быть использован для оценивания взаимодействующих информационных, информационно-аналитических, управленческих, технологических процессов для различных отраслей экономики (в промышленности, в топливно-энергетическом комплексе, в агропромышленном комплексе, на транспорте, в сфере обслуживания).

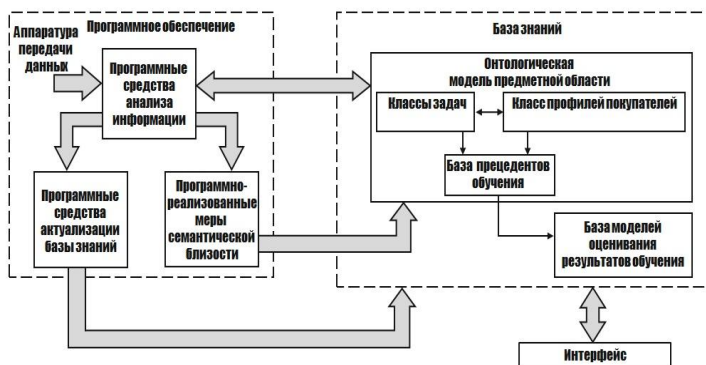


Рис. 6. Структура системы оценивания сложных систем и процессов на основе онтологического и нейро-нечеткого моделирования

Проверка работоспособности предлагаемого метода и оценивания оперативности выполнения логистических заказов с его использованием была осуществлена на примере выполнения заказов из онлайн-магазинов.

Результаты позволили сделать вывод, что использование предлагаемого метода позволяет повысить оперативность оценивания сложных систем и процессов в среднем в два раза.

## Заключение

В работе представлен предлагаемый метод оценивания сложных систем и процессов, использующий представление знаний проблемной области в виде онтологической модели, а также нейро-нечеткий классификатор для оценивания имеющихся прецедентов.

Для различных отраслей экономики (в промышленности, в топливно-энергетическом комплексе, в агропромышленном комплексе, на транспорте, в сфере обслуживания) обеспечение взаимодействия информационных, управленческих, технологических процессов остро ставит проблему качества и оперативности их оценивания.

## Список литературы

- [Борисов и др., 2014] Борисов В.В., Федулов А.С., Зернов М.М. Основы нечеткой математики. Книга 4. Основы нечеткого логического вывода. – М.: Горячая линия – Телеком, 2014. – 102 с.
- [Борисов и др., 2016] Борисов В.В., Федулов А.С., Зернов М.М. Основы нечеткой математики. Книга 5. Основы гибридизации нечетких моделей. – М.: Горячая линия – Телеком, 2016. – 105 с.
- [Гаврилова и др., 2001] Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. – СПб.: Питер, 2001. – 384 с.
- [Гаврилова и др., 2016] Гаврилова Т.А., Кудрявцев Д.В., Муромцев Д.И. Инженерия знаний. Модели и методы. 2016. – СПб: Лань.
- [Грибова и др., 2020] Грибова В.В., Тимченко В.А. Концепция поддержки лазерного аддитивного производства на основе онтологического подхода // Онтология проектирования. – 2020. – Т. 10, № 2 (36). – С. 176-189.
- [Грибова и др., 2022] Грибова В.В., Москаленко Ф.М., Тимченко В.А., Шалфеева Е.А. Платформа IASaaS для разработки систем на основе онтологий: десятилетие использования // Искусственный интеллект и принятие решений. – 2022. – № 4. – С. 55-65.
- [Загорюлько и др., 2020] Загорюлько Ю.А., Загорюлько Г.Б. Искусственный интеллект. Инженерия знаний. – М.: Юрайт, 2020. – 93 с.
- [Зацаринный и др., 2023] Зацаринный А.А., Меденников В.И., Райков А.Н. Интеграция приложений искусственного интеллекта в единую цифровую платформу АПК // Информационное общество. – 2023. – № 1. – С. 127-138.
- [Прокопенков и др., 2020] Прокопенков И.А., Котов Д.В., Молявко А.А. Способ формирования обобщенных прецедентных решений на основе композиционного онтологического подхода // Сборник научных трудов VIII Международной научно-практической конференции «Нечеткие системы, мягкие вычисления и интеллектуальные технологии». – Смоленск: Универсум, 2020. – Ч. II. – С. 128-135.
- [Ярушкина, 2018] Ярушкина Н.Г. Анализ слабоструктурированных текстов на основе нечетких онтологий // Сб. трудов IV Всероссийской Поспеловской конференции с международным участием «Гибридные и синергетические интеллектуальные системы» ГИСИС-2018. – Калининград: Изд-во БФУ им. И. Канта, 2018. – С. 133-139.

## МЕТОД ВЫВОДА НЕЧЕТКИХ СИСТЕМ ЛОГИЧЕСКОГО ТИПА ПРИ НЕСИНГЛТОННОЙ ФАЗЗИФИКАЦИИ\*

В.Г. Синюк (*vgsinuk@mail.ru*)

С.А. Каратач (*karatach1998@yandex.ru*)

Белгородский государственный технологический университет  
им. В.Г. Шухова, Белгород

В статье получен результат вывода для широко используемых нечетких систем логического типа при несинглетонной фаззификации. Он достигнут на основании подхода с применением нечеткого значения истинности, что дало возможность свести задачу к полиномиальной вычислительной сложности. В статье также описана возможность использования технологий OpenGL и CUDA при реализации алгоритма представленного нечеткого вывода.

**Ключевые слова:** несинглетонная фаззификация, нечеткое значение истинности, вычислительная сложность, эффективные вычисления.

### Введение

Изначально описанные в [Mamdani, 1974] нечеткие системы являлись композицией фаззификатора, базы знаний, модуля нечеткого вывода и дефаззификатора. Дальнейшее развитие теории нечетких систем заключалось в модификации каждого из этих компонентов.

Теоретические основы *нечеткого вывода* были введены Заде [Zadeh, 1973], были предложены и другие решения этой задачи. Самыми популярными в приложениях являются подходы, представленные Мамдани, Ларсеном [Larsen, 1980], Такаги-Сугено [Takagi et al., 1985]. В перечисленных подходах при применении нечетких систем использовалась singleton-фаззификации. Это решение, как правило, используются из-за простоты и эффективности их реализации. Вместе с тем, эти методы не в полной мере соответствуют теории Заде из-за значительного упрощения.

---

\* Работа выполнена при финансовой поддержке РФФИ (проект № 20-07-00030).

Альтернативным подходом является фаззификация типа non-singleton (NS). NS-фаззификация используется в нечетких системах на основе правил, когда измерения, которые их активируют, несовершенны или неопределенны (из-за шума измерений, дефектов или ухудшения качества датчиков и т.д.), или когда их входными данными являются понятия естественного языка. Она моделирует такие измерения или понятия нечеткими множествами или более общими лингвистическими переменными. Данные объекты теории нечетких множеств и нечетких систем в большей степени отражают исходную природу содержащих неопределенность или несовершенство измерения числовых данных или понятий естественного языка.

Основной проблемой, возникающей при реализации вывода в нечетких системах MISO-структуры, использующих NS-фаззификацию, является экспоненциальный рост вычислительной сложности [Mendel, 2021]. В данной статье приведен метод вывода для нечетких систем логического типа который снижает вычислительную сложность до полиномиальной. Предложенный метод позволяет реализовать работу алгоритма как последовательность параллельных вычислений и сверток. Использование параллельных технологий вычисления становится особенно оправданным, когда при реализации данного алгоритма в обобщенном виде выполняется дискретизация функций принадлежности.

В первом разделе статьи формулируется задача вывода, решаемая нечеткой системой. Во втором разделе описан метод нечеткого вывода с использованием нечеткого значения истинности и NS-фаззификации. В третьем разделе описаны особенности эффективной реализации данного подхода с использованием технологий CUDA и GPGPU. Затем, в четвертом разделе для описанной реализации проведен вычислительный эксперимент, по результатам которого были сделаны выводы об оправданности предложенных техник эффективной реализации.

## 1. Постановка задачи

Лингвистическая модель представляет собой базу нечетких правил  $R_k$ ,  $k = \overline{1, N}$  вида:

$$R_k: \text{Если } x_1 \text{ есть } A_{k1} \text{ и } x_2 \text{ есть } A_{k2} \text{ и } \dots \text{ и } x_n \text{ есть } A_{kn}, \text{ то } y \text{ есть } B_k, \quad (1.1)$$

где  $N$  – количество нечетких правил,  $A_{ki} \subseteq X_i$ ,  $i = \overline{1, n}$ ,  $B_k \subseteq Y$  – нечеткие множества, которые характеризуются функциями принадлежности  $\mu_{A_{ki}}(x_i)$  и  $\mu_{B_k}(y)$  соответственно;  $x_1, x_2, \dots, x_n$  – входные переменные лингвистической модели, причем

$$[x_1, x_2, \dots, x_n]^T = \mathbf{x} \in X_1 \times X_2 \times \dots \times X_n.$$



Символами  $X_i$ ,  $i = \overline{1, n}$  и  $Y$  обозначаются соответственно пространства входных и выходной переменных. Если ввести обозначения  $\mathbf{X} = X_1 \times X_2 \times \dots \times X_n$  и  $\mathbf{A}_k = A_{k1} \times A_{k2} \times \dots \times A_{kn}$ , причем

$$\mu_{A_k}(\mathbf{x}) = T_{i=\overline{1, n}} \mu_{A_{ki}}(x_i),$$

где  $T_1$  – произвольная t-норма, то правило (1.1) представляется в виде нечеткой импликации

$$R_k : \mathbf{A}_k \rightarrow B_k, \quad k = \overline{1, N}.$$

Правило  $R_k$  можно формализовать как нечеткое отношение, определенное на множестве  $\mathbf{X} \times Y$ , т.е.  $R_k \subseteq \mathbf{X} \times Y$  – нечеткое множество с функцией принадлежности

$$\mu_{R_k}(\mathbf{x}, y) = \mu_{\mathbf{A}_k \rightarrow B_k}(\mathbf{x}, y).$$

Модель логического типа определяет задание функции  $\mu_{\mathbf{A}_k \rightarrow B_k}(\mathbf{x}, y)$  на основе известных функций принадлежности  $\mu_{A_k}(\mathbf{x})$  и  $\mu_{B_k}(y)$  с помощью одной из предложенных в [Рутковский и др., 2010] функций импликации:

$$\mu_{\mathbf{A}_k \rightarrow B_k}(\mathbf{x}, y) = I(\mu_{A_k}(\mathbf{x}), \mu_{B_k}(y)),$$

где  $I$  – некоторая импликация.

Ставится задача определить нечеткий вывод  $B'_k \subseteq Y$  для системы, представленной в виде (1.1), если на входах – нечеткие множества  $A'_i = A'_1 \times A'_2 \times \dots \times A'_n \subseteq \mathbf{X}$  или  $x_1$  есть  $A'_1$  и  $x_2$  есть  $A'_2$  и ... и  $x_n$  есть  $A'_n$  с соответствующей функцией принадлежности  $\mu_{A'_i}(\mathbf{x})$ , которая определяется как

$$\mu_{A'_i}(\mathbf{x}) = T_{i=\overline{1, n}} \mu_{A'_i}(x_i). \quad (1.2)$$

Несингтонный фаззификатор отображает измеренное  $x_i = x'_i$ ,  $i = \overline{1, n}$  в нечеткое число, для которого  $\mu_{A'_i}(x'_i) = 1$  и  $\mu_{A'_i}(x_i)$  уменьшается от единицы по мере удаления от  $x'_i$ .

Концептуально [Mouzouris et al., 1997] NS фаззификация подразумевает, что данное входное значение  $x'_i$  есть значение, которое будет наиболее возможным из всех значений, находящихся рядом; однако, так как вход является неопределенным, соседние значения также могут быть возможными, но в меньшей степени.

В соответствии с обобщенным нечетким правилом *modus ponens* [Рутковский и др., 2010], нечеткое множество  $B'_k$  определяется композицией нечеткого множества  $A'$  и отношения  $R_k$ , т.е.

$$B'_k = A' \circ (A_k \rightarrow B_k),$$

или, на уровне функций принадлежности

$$\mu_{B'_k}(y | \mathbf{x}') = \sup_{x \in X} \left\{ \mu_{A'}(\mathbf{x}') * I(\mu_{A_k}(\mathbf{x}), \mu_{B_k}(y)) \right\}. \quad (1.3)$$

В (1.3) применена условная нотация, так как ввод в нечеткую систему происходит при определенном значении  $\mathbf{x}$ , а именно  $\mathbf{x}'$ . Обозначение  $\mu_{B'_k}(y | \mathbf{x}')$  показывает, что  $\mu_{B'_k}$  изменяется с каждым значением  $\mathbf{x}'$ . Вычислительная сложность выражения (1.3) составляет  $O(|X_1| \cdot |X_2| \cdot \dots \cdot |X_n| \cdot |Y|)$  т.е. экспоненциальная.

## 2. Метод вывода с использованием нечеткого значения истинности

Применяя правило истинностной модификации [Борисов и др., 1990]

$$\mu_{A'}(\mathbf{x}) = \tau_{A_k|A'}(\mu_{A_k}(\mathbf{x})),$$

где  $\tau_{A_k|A'}(\cdot)$  — нечеткое значение истинности (НЗИ) нечеткого множества  $A_k$  относительно  $A'$ , представляющее собой функцию принадлежности совместимости  $CP(A_k, A')$   $A_k$  по отношению к  $A'$ , причем  $A'$  рассматривается как достоверное [Дюбуа и др., 1990]:

$$\tau_{A_k|A'}(v) = \mu_{CP(A_k, A')}(v) = \sup_{\substack{\mu_{A_k}(x)=v \\ x \in X}} \{ \mu_{A'}(x) \}.$$

Таким образом *НЗИ* отражает совместимость факта с посылкой в нечеткой форме. Упрощенные подходы отображают совместимость в одно значение из диапазона впервые представленное в [Mamdani, 1974].

Перейдем от переменной  $\mathbf{x}$  к переменной  $v$ , обозначив  $\mu_{A_k}(\mathbf{x}) = v$ .

Получим:

$$\mu_{A'}(\mathbf{x}) = \tau_{A_k|A'}(\mu_{A_k}(\mathbf{x})) = \tau_{A_k|A'}(v), \quad (2.1)$$

тогда (1.3) примет вид:

$$\mu_{B'_k}(y | \mathbf{x}') = \sup_{v \in [0,1]} \left\{ \tau_{A_k|A'}(v) * I(v, \mu_{B_k}(y_r)) \right\}. \quad (2.2)$$

При вербализации импликации в (2.2) она представится в виде:

$$\text{Если } n\text{зи есть истинно, то } u \text{ есть } B'_k. \quad (2.3).$$

Таким образом, (2.3) представляет собой еще одну структуру правил в отличие от канонических структур Заде [Zadeh, 1973] и Такаги-Сугено [Takagi et al., 1985]. Применение данного правила не зависит от количества входов в нечетких системах.

Выражение (2.2) характеризуется сложностью порядка  $O(|v| \cdot |Y|)$ . Как следует из [Синюк и др., 2018], [Sinuk et al., 2016]:

$$\begin{aligned} \mu_{CP(A_k, A')} (v) &= \tilde{T}_1 \mu_{CP(A_k, A'_i)} (v_i) = \\ &= \left( \dots \left( \left( \mu_{CP(A_{k1}, A'_1)} (v_1) \tilde{T}_1 \mu_{CP(A_{k2}, A'_2)} (v_2) \right) \tilde{T}_1 \mu_{CP(A_{k3}, A'_3)} (v_3) \right) \tilde{T}_1 \dots \right) \tilde{T}_1 \mu_{CP(A_{kn}, A'_n)} (v_n), \end{aligned}$$

где  $\tilde{T}_1$  – расширенная по принципу обобщения  $n$ -местная  $t$ -норма и

$$\mu_{CP(A_{ki}, A'_i)} (v_i) = \sup_{\substack{\mu_{A_{ki}}(x_i)=v_i \\ x_i \in X_i}} \left\{ \mu_{A'_i}(x_i) \right\}.$$

Например, двухместная расширенная  $t$ -норма имеет вид:

$$\mu_{CP(A_k, A')} (v) = \tilde{T}_1 \mu_{CP(A_{ki}, A'_i)} (v_i) = \sup_{\substack{v_1 \tilde{T}_1 v_2 = v \\ (v_1, v_2) \in [0,1]^2}} \left\{ \mu_{CP(A_{k1}, A'_1)} (v_1) \tilde{T}_3 \mu_{CP(A_{k2}, A'_2)} (v_2) \right\}.$$

Последнее соотношение характеризуется сложностью порядка  $O(|v|^2)$ .

Таким образом данный подход позволяет переместить процесс вывода в единое пространство *НЗИ*, где функции истинности, в отличии от различных пространств в подходе Заде, могут быть объединены в более эффективный вычислительный процесс.

Выходное значение нечеткой системы, описываемой (1.1), при дефазификации по методу центра тяжести с учетом (2.2) определяется [Рутковский и др., 2010]:

$$\bar{y}(\mathbf{x}') = \frac{\sum_{r=1, N} \bar{y}_r T_4 \left\{ \sup_{v \in [0,1]} \left[ \tau_{A_k / A'}(v) \overset{T_2}{*} I(v, \mu_{B_k}(\bar{y}_r)) \right] \right\}}{\sum_{r=1, N} T_4 \left\{ \sup_{v \in [0,1]} \left[ \tau_{A_k / A'}(v) \overset{T_2}{*} I(v, \mu_{B_k}(\bar{y}_r)) \right] \right\}}. \quad (2.4)$$

Поскольку импликация в формуле (2.4) не зависит от входных данных (1.2), то предварительно, т.е. до использования композиционного правила (2.2), вычисляется  $\tau_{k,r}(v) = I(v, \mu_{B_k}(\bar{y}_r))$  при  $k = \bar{1}, \bar{N}$ ,  $r = \bar{1}, \bar{N}$ ,  $v \in [0,1]$ .

Используя наиболее часто применяемые в нечетких системах логического типа нечеткие импликации [Рутковский и др., 2010] определим  $\tau_{k,r}(v)$  следующим образом:

1. бинарная импликация Клини-Динеса;

$$\tau_{k,r}(v) = \max(1 - v, b_{k,r}), \text{ где } b_{k,r} = \mu_{B_k}(\bar{y}_r);$$

2. импликация Лукасевича:

$$\tau_{k,r}(v) = \min(1, 1 - v + b_{k,r});$$

3. импликация Райхенбаха:

$$\tau_{k,r}(v) = 1 - v + vb_{k,r};$$

4. импликация Гогуена:

$$\tau_{k,r}(v) = \begin{cases} \min\left(1, \frac{b_{k,r}}{v}\right), & \text{если } v > 0; \\ 1, & \text{если } v = 0 \end{cases};$$

5. импликация Геделя:

$$\tau_{k,r}(v) = \min(1, b_{k,r} + \rho(v, b_{k,r})),$$

где  $\rho(a, b) = \begin{cases} 1, & \text{если } a \leq b; \\ 0, & \text{если } a > b; \end{cases}$

6. импликация Фодора:

$$\tau_{k,r}(v) = \min(1, \max(1 - v, b_{k,r})) + \rho(v, b_{k,r});$$

7. импликация Заде:

$$\tau_{k,r}(v) = \max(1 - v, \min(v, b_{k,r}));$$

8. импликация Вильмогта:

$$\tau_{k,r}(v) = \min(\max(1 - v, b_{k,r}), \max(v, 1 - b_{k,r}, \min(1 - v, b_{k,r}))).$$

С учетом данных определений (2.4) принимает вид:

$$\bar{y}(\mathbf{x}') = \frac{\sum_{r=1, N} \bar{y}_r T_4 \left\{ \sup_{v \in [0,1]} \left\{ \tau_{A_k / A'}^{T_2}(v) * \tau_{k,r}(v) \right\} \right\}}{\sum_{r=1, N} T_4 \left\{ \sup_{v \in [0,1]} \left\{ \tau_{A_k / A'}^{T_2}(v) * \tau_{k,r}(v) \right\} \right\}} \quad (2.5)$$

Одна из возможностей упрощения состоит в том, что функции принадлежности консеквентов правил (1.1) достаточно удалены друг от друга т.е. выполняется следующее соотношение  $b_{k,r} = 0, k \neq r$  что проиллюстрировано на рис. 1:

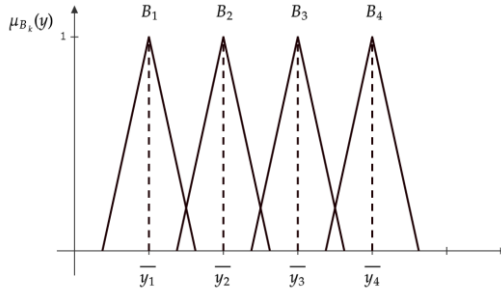


Рис. 1. Пример нечетких множеств, удовлетворяющих условию  $\mu_{B_k}(y_r) = 0$  для  $k \neq r$

При этом условии, если используется S-импликация Клини-Динеса, Лукасевича, Рейхенбаха, Фодора, то:

$$\tau_{k,r}(v) = \begin{cases} 1 - v, & \text{если } k \neq r \\ 1, & \text{если } k = r \end{cases}$$

и (2.5) имеет вид:

$$\bar{y}(\mathbf{x}') = \frac{\sum_{r=1,N} \bar{y}_r T_4 \left\{ \sup_{v \in [0,1]} \left\{ \tau_{A_k/A'}^{T_2}(v) * (1-v) \right\} \right\}}{\sum_{r=1,N} T_4 \left\{ \sup_{v \in [0,1]} \left\{ \tau_{A_k/A'}^{T_2}(v) * (1-v) \right\} \right\}}.$$

Если используется R-импликация Гогуена, Геделя, то:

$$\tau_{k,r}(v) = \begin{cases} \delta(v), & \text{если } k \neq r \\ 1, & \text{если } k = r \end{cases}, \text{ где } \delta(a) = \begin{cases} 1, & \text{если } a = 0 \\ 0, & \text{если } a > 0 \end{cases}$$

и (2.5) имеет вид:

$$\bar{y}(\mathbf{x}') = \frac{\sum_{r=1,N} \bar{y}_r T_4 \left\{ \tau_{A_k/A'}(0) \right\}}{\sum_{r=1,N} T_4 \left\{ \tau_{A_k/A'}(0) \right\}}.$$

Если используется Q-импликация Заде или Вильмотта, то:

$$\tau_{k,r}(v) = \begin{cases} 1 - v, & \text{если } k \neq r \\ \max(v, 1 - v), & \text{если } k = r \end{cases}$$

и (2.5) примет вид:

$$\bar{y}(\mathbf{x}') = \frac{\sum_{r=1,N} \bar{y}_r T_4 \left\{ \sup_{v \in [0,1]} \left\{ \tau_{A_k/A'}(v) \overset{T_2}{*} \max(v, 1-v) \right\} T_4 \left\{ \sup_{v \in [0,1]} \left\{ \tau_{A_k/A'}(v) \overset{T_2}{*} (1-v) \right\} \right\} \right\}_{k \neq r}}{\sum_{r=1,N} T_4 \left\{ \sup_{v \in [0,1]} \left\{ \tau_{A_k/A'}(v) \overset{T_2}{*} \max(v, 1-v) \right\} T_4 \left\{ \sup_{v \in [0,1]} \left\{ \tau_{A_k/A'}(v) \overset{T_2}{*} (1-v) \right\} \right\} \right\}_{k \neq r}}.$$

Выражение (2.5) в фигурных скобках, которое представляет композицию двух нечетких множеств, есть определение *возможности* [Борисов и др., 1990] то есть соответствие того, что  $\tau_{A_k/A'}(v)$  есть  $\tau_{k,r}(v)$  или наоборот. Обозначим:

$$\sup_{v \in [0,1]} \left\{ \tau_{A_k/A'}(v) \overset{T_2}{*} \tau_{k,r}(v) \right\} = \Pi_{k,r}$$

тогда (2.5) примет вид:

$$\bar{y}(\mathbf{x}') = \frac{\sum_{r=1,N} \bar{y}_r T_4 \left\{ \Pi_{k,r} \right\}_{k=1,N}}{\sum_{r=1,N} T_4 \left\{ \Pi_{k,r} \right\}_{k=1,N}}. \quad (2.6)$$

Сетевая структура соотношения (2.5) с учетом (2.6) представлена на рис. 2.

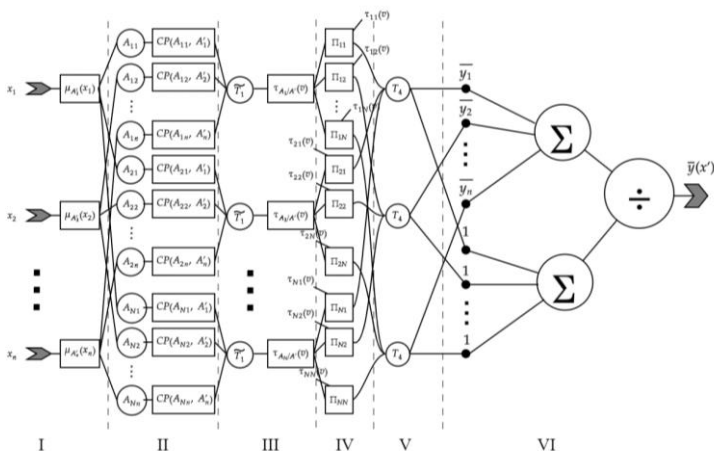


Рис. 2. Сетевая структура выражения (2.5) с учетом (2.6)

### 3. Особенности реализации метода вывода с использованием нечеткого значения истинности

Отдельные этапы вычислений соответствуют различным слоям представленной на рис. 2 сетевой структуры, на каждом из которых решаются различные алгоритмические задачи, поэтому при реализации каждого этапа вычислений могут использоваться различные подходы по организации эффективных вычислений.

#### 3.1. Использование OpenGL для вычисления нечеткого значения истинности

При программной реализации предложенного метода выполнялась дискретизация функций принадлежности (ф. п.) нечетких множеств. Тогда, сложность вычисления НЗИ по данному входу для данных посылки и факта –  $O(D_{ftv} \times D_{fset})$ , где  $D_{ftv}$  – размер сетки дискретизации ф.п. нечеткого значения истинности, а  $D_{fset}$  – размер сетки дискретизации ф.п. нечеткого множества входной переменной.

Вычисление в разных точках расчетной сетки НЗИ можно распределить между нитями блока CUDA-ядра, тогда сложность вычислений в одной нити составит  $O(D_{fset})$ .

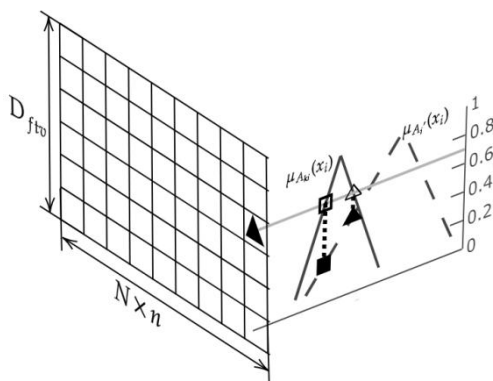


Рис. 3. Использование OpenGL для вычисления НЗИ

Другой, представленный на рис. 3, подход к реализации вычисления НЗИ состоит в том, чтобы вместо программирования потокового мультипроцессора внутри CUDA-ядра использовать аппаратную реализацию внутренних механизмов шейдеров OpenGL. В данном случае процесс вычисления НЗИ для всех входов и всех правил представляет собой нахождение интенсивности свечения одноканальных пикселей изображения

размера  $D_{ftv} \times (N \times n)$ , где  $N$  – количество правил в базе правил,  $n$  – количество входов. Вычисления НЗИ организуются на основе спецификации конвейера отрисовки трехмерной графики в OpenGL, в соответствии с которой вычисление значения этой формулы разбивается между вершинным и фрагментным шейдером.

### 3.2. Особенность реализации свертки НЗИ, когда $T_1 = \min$ , а $T_3$ является функцией неубывающей по всем аргументам

Свертку НЗИ по входам можно эффективно вычислить попарно находя  $T_3$ -норму до получения результата, в результате чего образуется дерево промежуточных  $T_3$ -норм НЗИ. Вычислительная сложность такого подхода  $O(\log n \times D_{ftv}^2)$ . В ситуации, когда в качестве  $T_1$  выбрана функция  $\min$ , то опираясь на свойство не возрастания  $T_1$  и  $T_3$ , предложен алгоритм, позволяющий реализовать вычисление свертки со сложностью  $O(D_{ftv} \times \log n)$ . Данный алгоритм представлен на рис. 4.

---

**Алгоритм 1** Свертка НЗИ, когда  $T_1 = \min$  и  $T_3(a, b) \geq T_3(c, d)$  при  $a > c$  или  $b > d$

---

**Вход:**  $ftv_i, i = \overline{1, n}$   
 $max\_ftv[i] = 0$ ;  
**for**  $v = 1 \dots 0$  **do**  
     $s \leftarrow \{ftv_i[v] \mid ftv_i[v] \geq max\_ftv[i]\}$ ;  
     $max\_ftv[i] \leftarrow \max(max\_ftv[i], ftv_i[v])$ ;  
     $v\_max \leftarrow \max_i \{ftv_i[v]\}, v\_max\_index \leftarrow_i \{ftv_i[v]\}$ ;  
    **if**  $s = \emptyset$  &  $i = v\_max\_index$  **then**  
         $r[i] \leftarrow v\_max$ ;  
    **else**  
         $r[i] \leftarrow max\_ftv[i]$ ;  
    **end if**  
     $ftv[v] \leftarrow T_3 \{r[i]\}$ ;  
**end for**  
**return**  $ftv$

---

Рис. 4. Процедура свертки НЗИ, когда  $T_1 = \min$

### 3.3. Реализация эффективной свертки внутри CUDA-ядра

После нахождения свертки НЗИ для получения окончательного результата вывода необходимо найти значение формулы (2.4), выполнив свертку по всем правилам для всех центров на основании значений предыдущей свертки и предварительно вычисленных значений  $\tau_{k,r}$ .



При реализации свертки внутри CUDA-ядра важно учитывать особенности вычислений на потоковых мультипроцессорах: необходимо исключать бездействия нитей внутри одного пакета из 32 нитей, а доступ к данным должен избегать возникновения конфликта банков.

#### 4. Сравнительный анализ различных способов реализации

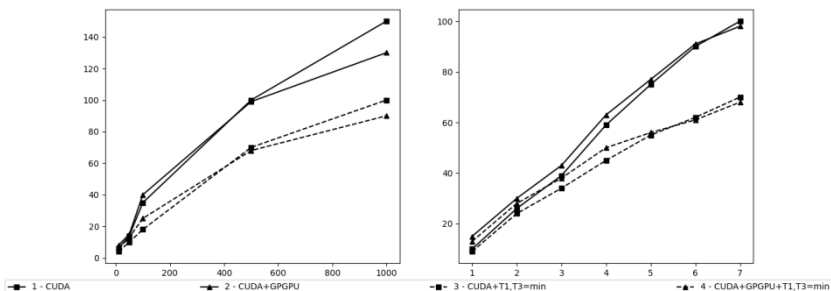


Рис. 5. Графики времени выполнения времени работы различных модификаций при различных значениях степени дискретизации НЗИ (слева) и при различном количестве входов (справа)

В данном разделе проводится сравнительный анализ результатов нескольких вычислительных экспериментов, проведенных для различных модификаций параллельной реализации нечеткого классификатора на основе предложенного метода. Нечеткая классификация выполнялась на наборе данных KDD99, содержащем 494021 записей, каждая из которых фиксирует различные атрибуты TCP пакетов.

Сравнению подвергались различные модификации алгоритма вывода. В 1-й и 3-й модификациях все вычисления выполняются с помощью технологии CUDA, во 2-й и 4-й модификациях для нахождения НЗИ была задействована технология OpenGL (EGL), а для остальных вычислений также использовалась технология CUDA. В 1-й и 2-й модификациях ограничений на выбора  $T_1$  и  $T_3$  не накладывается, в 3-й и 4-й модификациях в качестве  $T_1$  и  $T_3$  используется функция  $\min$ . В обеих модификациях использовалась импликация Лукасевича.

В первой серии вычислительных экспериментов оценивается время работы алгоритма при различных размерах расчетной сетки НЗИ при количестве входов равному 7. Во второй серии экспериментов сравнивается время работы алгоритма при различном количестве входов и размере расчетной сетки НЗИ равному 500.

По полученным результатам экспериментов можно заметить, что использование GPGPU становится оправданным при размере расчетной сетки больше 500 и количестве входов больше 6, а использование оптимального алгоритма свертки НЗИ дает преимущество во времени вычислений практически на всем диапазоне параметров экспериментов.

### Заключение

В статье приведен метод вывода для нечетких систем логического типа MISO-структуры использующий NS-фаззификацию. Метод основан на применении нечеткого значения истинности, что дает возможность снизить сложность с экспоненциальной до полиномиальной за счет перевода процесса вывода в единое пространство *НЗИ* в отличие от различных пространств в подходе Заде.

В результате была представлена другая форма нечетких правил, которое не зависит от количества входов, что повлияло на классическую структуру нечетких систем.

Применение НЗИ отражает совместимость факта с посылкой в нечеткой форме. Упрощенные подходы отображают совместимость в одно значение в диапазоне  $[0, 1]$  и теряет нечеткость.

В статье описана возможность использования технологии OpenGL при реализации вычисления НЗИ, а также ключевые особенности реализации эффективной свертки с помощью технологии CUDA. В случае, когда на, используемые при свертке НЗИ, *t*-нормы наложены определенные ограничения, предложен эффективный алгоритм свертки вычисленных по всем входам НЗИ. На основе результатов вычислительных экспериментов, проведенных для предложенных модификаций, были сделаны выводы о целесообразности использованием GPGPU для вычисления НЗИ и эффективности процедуры свертки, достигаемой за счет предложенной оптимизации.

### Список литературы

- [Борисов и др., 1990] Борисов А.Н., Крунберг О.А., Федоров И.П. Принятие решений на основе нечетких моделей. – Рига: Зинатне, 1990.
- [Рутковский и др., 2010] Рутковский Л. Методы и технологии искусственного интеллекта. – М.: Горячая линия – Телеком, 2010.
- [Синюк и др., 2018] Синюк В.Г., Михелев В.В. Методы вывода для систем логического типа на основе нечеткой степени истинности // Известия РАН. Теория и системы управления. 2018, № 3.
- [Larsen, 1980] Larsen P.M. Industrial Applications of Fuzzy Logic Control // Intern. J. Man-Machine Studies. – 1980. Vol. 12, No. 1.

- [**Mamdani, 1974**] Mamdani E.H. Applications of Fuzzy Algorithm for Control a Simple Dynamic Plant // Proc. IEEE. – 1974. –Vol. 121, No. 12.
- [**Mendel, 2021**] Mendel J.M. Non-Singleton Fuzzification Made Simpler // Information Sciences. – 2021. – Vol. 559.
- [**Mouzouris et al., 1997**] Mouzouris G.C., Mendel J.M. Non-Singleton Fuzzy Logic Systems: Theory and Application // IEEE Transactions on Fuzzy Systems. – 1997. – Vol. 5, No. 1.
- [**Sinuk et al., 2016**] Sinuk V.G., Polyakov V.M., Kutsenko D.A. New Fuzzy Truth Value Based Inference Methods for Non-singleton MISO Rule-Based Systems // Proceedings of the First International Scientific Conference “Intelligent Information Technologies for Industry” (IITI '16). – 2016.
- [**Takagi et al., 1985**] Takagi T., Sugeno M. Fuzzy Identification of Systems and Its Application to Modeling and Control // IEEE Transactions on Systems, Man and Cybernetics. – 1985. – Vol. 15, No. 1.
- [**Zadeh, 1973**] Zadeh L.A. Outline of a New Approach to the Analysis of Complex Systems and Decision Processes // IEEE Transactions on Systems, Man and Cybernetics. – 1973. – Vol. 3, No. 1.

УДК 004.89

## НЕЧЕТКИЙ СИТУАЦИОННО-ПРЕЦЕДЕНТНЫЙ АНАЛИЗ И МОДЕЛИРОВАНИЕ ПРОЦЕССОВ В СЛОЖНЫХ ТЕХНИЧЕСКИХ СИСТЕМАХ\*

А.М. Соколов (*andreisokol98@gmail.com*)

Национальный исследовательский университет «МЭИ», Москва

В.И. Будзко (*vbudzko@ipiran.ru*)

А.А. Зацаринный (*azatsarinny@frcscs.ru*)

Федеральный исследовательский центр  
«Информатика и управление» РАН, Москва

Современные производственно-технологические системы, объекты агропромышленного комплекса характеризуются сложными зависимостями между системными и внешними факторами, неполнотой и неопределенностью сведений об их состоянии и функционировании. В качестве основы для анализа и моделирования таких систем, с целью поддержки эффективных управленческих решений, в статье предлагается развитие нечеткого ситуационного подхода, ориентированного на гибкое построение и адаптивную структурно-параметрическую настройку нечетких ситуационных моделей (сетей), учитывающую обобщение и актуализацию сведений о прецедентных ситуациях, имеющих место в процессе функционирования и развития таких объектов и систем.

**Ключевые слова:** нечеткая ситуационная сеть, прецедентный подход, сложная система, агропромышленный комплекс.

### Введение

В настоящее время с развитием высокотехнологических производств все более широкое распространение получают сложные технические системы (СТС), которые характеризуются наличием большого числа входящих в них компонентов, взаимодействующих между собой и образующих сложную структуру связей, а также комплексным характером протекаю-

---

\* Работа поддержана грантом РФФ (проект № 22-61-00096), а также грантом Министерства науки и высшего образования РФ (внутренний номер 00600/2020/51896, договор № 075-15-2022-319).

щих в них процессов [Бобряков и др., 2022]. К таким системам могут быть отнесены, например, химико-электротехнологические системы термической переработки рудного сырья [Пучков и др., 2022], производственные объекты агропромышленного комплекса [Будзко и др., 2023], автоматизированные и роботизированные компоненты современных «умных» производств [Зацаринный и др., 2023], [Bosdelekidis и др., 2022], комплексы агрегатов бесперебойного электропитания в составе энергетической инфраструктуры [Рожков и др., 2023] производственно-технологических систем и агропромышленного комплекса.

Важной особенностью указанных систем является возможность их функционирования во множестве различных технологических состояний, которые могут характеризоваться неполнотой информации, описываться с помощью разнокачественных данных и быть связаны посредством сложной системы управляющих переходов. Каждое состояние может характеризоваться набором явных и неявных внутренних и внешних параметров, сочетание которых оказывает влияние на характеристики функционирования системы в целом. При этом часто не представляется возможным получение аналитического и статистического описания зависимостей ее параметров. В этой связи для решения задачи комплексного управления подобной системой с учетом особенностей процессов работы всех ее компонентов должны применяться специальные методы моделирования. На практике одним из перспективных подходов к моделированию таких систем является применение нечетких ситуационных сетей (НСС) [Борисов и др., 2021]. Наиболее трудоемким при реализации данного подхода является этап построения ситуационной сети, который, как правило, требует привлечения эксперта. Однако из-за сложности современных СТС указанный подход к построению НСС не позволяет в полной мере учесть особенности функционирования таких систем.

Целью данной работы является развитие подходов к моделированию функционирования СТС на основе комплексного использования нечетких ситуационных сетей и прецедентного подхода.

Построение указанных моделей сложных производственных технических систем позволит, с одной стороны, повысить качество работы систем поддержки принятия управляющих решений на их основе. С другой стороны, будет способствовать развитию и более широкому практическому применению современных инструментов цифровизации производств в соответствии с основными направлениями концепции «Индустрия 4.0». К таким направлениям может быть отнесено, например, применение промышленного Интернета вещей (IoT), предполагающего объединение входящих в систему устройств в единую сеть для сбора, обработки и передачи данных [Гумеров и др., 2022].

## 1. Существующие подходы к построению нечетких ситуационных сетей

НСС представляет собой граф, узлами которого являются элементы множества нечетких ситуаций  $S = \{S_i | i = 1, \dots, I\}$ , соединенные нечеткими управляющими переходами  $W = \{w_{j,k} | S_j, S_k \in S\}$ . Каждая нечеткая ситуация из множества  $S$  определяется набором ситуационных признаков  $P = \{p_j | j = 1, \dots, J\}$  и отражает совокупность внутренних параметров и внешних условий, в которых потенциально может оказаться система в процессе своего функционирования [Борисов и др., 2021]. В целом ситуация определяется совокупностью обстоятельств, приводящих систему в определенное состояние. Если указанные признаки могут быть представлены только в виде нечетких описаний, то рассматриваются так называемые нечеткие ситуации.

Каждому управляющему переходу из множества  $W$  ставится в соответствие влияющее на определенный ситуационный признак управляющее решение из множества  $U = \{u_h^{(p_j)} | h = 1 \dots H; j = 1, \dots, J\}$  и соответствующие веса, которые определяются экспертом на основе выбранной системы параметров и могут использоваться для выбора стратегии управления, оптимальной с точки зрения одного или комбинации критериев [Evstifeev et al., 2021]. В качестве таких критериев могут рассматриваться, например, оперативность управления или минимизация затрачиваемой на выполнение операций энергии.

Для каждого переход из одной ситуации в другую осуществляется посредством выбора и применения соответствующего управляющего решения  $u_h^{(p_j)}$ , к исходной нечеткой ситуации. Воздействие управляющего решения на заданный нечеткий ситуационный признак  $p_j$  реализуется с помощью операции нечеткой композиции:

$$p'_j = p_j \circ r_{k_j}^{(p_j)}.$$

Предполагается, что НСС строится заранее на основе экспертной оценки состояний и процессов функционирования реальной СТС, при этом множество эталонных ситуаций детерминировано и определяется всеми возможными комбинациями ситуационных признаков из рассматриваемого набора. Как отмечалось ранее, при моделировании более сложных технических и кибер-физических систем, обладающих широким набором учитываемых внутренних параметров и внешних факторов, построение НСС в соответствии с описанным подходом сталкивается с рядом трудностей. В первую очередь данные трудности

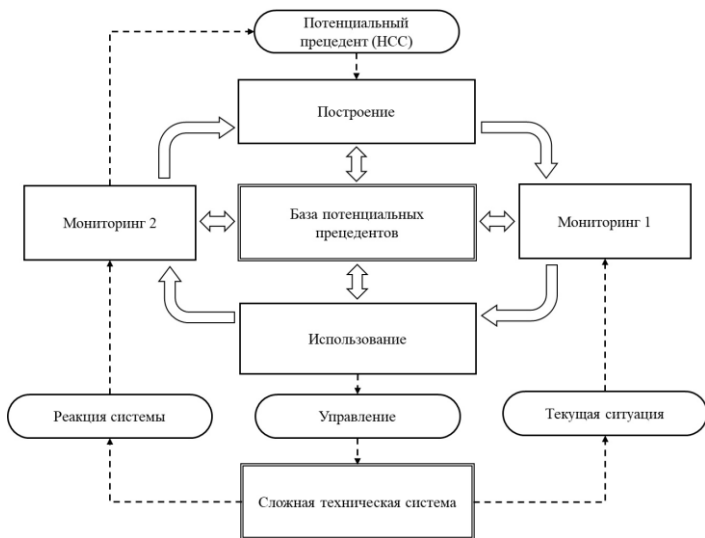
связаны со значительным ростом необходимого для подготовки системы поддержки принятия решений объема экспертной информации, что, с одной стороны, может привести к снижению точности прогнозирования, а с другой – к увеличению затрат на создание модели. Также стоит отметить, что на практике при управлении СТС могут возникать ситуации, при которых одно и те же управляющие решения могут с некоторой вероятностью приводить к различным ситуациям в виду наличия неучтенных внешних и системных нечетких факторов. В этой связи аккумулирование и использование исторических данных об эксплуатации СТС для выявления и учета указанных случаев также является актуальной задачей [Dobrynin et al., 2020].

В соответствии с вышесказанным представляется целесообразным применение более гибкого подхода к построению и адаптации НСС. Для моделирования работы комплексных систем, для которых приведение полного описания невозможно или затруднено в виду их структурной и параметрической сложности, а также наличия влияния на их функционирование стохастических и нечетких внешних воздействий, на практике часто применяется прецедентный подход, который предполагает [Черновалова и др., 2023].

## **2. Прецедентный подход для построения нечеткой ситуационной модели**

В работе предлагается применение прецедентного подхода для построения НСС, предполагающий накопление ретроспективной информации о ситуациях, в которых оказывается система в процессе своего функционирования, о принимаемых оператором системы управляющих решениях, а также затрат и эффекта от реализации управления с целью последующего использования полученного опыта при выработке управляющих решений для текущей ситуации. К достоинствам применения прецедентного подхода можно отнести возможность построения системы управления в условиях неполноты информации о текущих параметрах СТС и характеристиках окружающей среды ее функционирования, структурной сложности взаимодействия отдельных компонентов СТС, а также отсутствия возможности получения полной экспертной оценки всех возможных эксплуатационных состояний [Еремеев и др., 2021].

На рис. 1 приведена предложенная обобщенная схема цикла принятия решений по управлению СТС на основе комплексного применения нечётких ситуационных сетей и прецедентного подхода.



*Рис. 1. Обобщенная схема управления СТС на основе комбинирования нечётких ситуационных сетей и прецедентного подхода*

Данная схема отражает полный цикл принятия решений в соответствии с прецедентным подходом, включая этапы построения базы потенциальных прецедентов и ее применения для формирования сигналов управления СТС.

В данной работе под прецедентом понимается ситуация, информация о которой была получена ранее в ходе эксплуатации СТС, и которая является наиболее близкой к текущей ситуации. Каждая нечеткая ситуация описывается набором нечетких признаков, характеризующих как состояние самой СТС, так и условий внешней среды ее функционирования.

Как отмечалось ранее, ключевым и наиболее трудоемким является этап построения НСС. В данной работе предлагается процедура формирования сети на основе поиска в базе последовательности прецедентов, позволяющих осуществить перехода из текущей нечеткой ситуации в целевую.

Пусть в процессе управления СТС было установлено, что имеет место следующий кортеж:  $(S_n^1, S_k^1, w^1, q^1)$ , где  $S_n^1$  – начальная нечеткая ситуация;  $S_k^1$  – конечная нечеткая ситуация;  $w^1$  – управляющий переход из  $S_n^1$  в  $S_k^1$ , которому поставлено в соответствие некоторое управляющее решение и затраты на его осуществление;  $q^1$  – возможность перехода из ситуации  $S_n^1$  в  $S_k^1$ . Данный кортеж может быть визуализирован как фрагмент НСС, представленный на рис. 2,а.



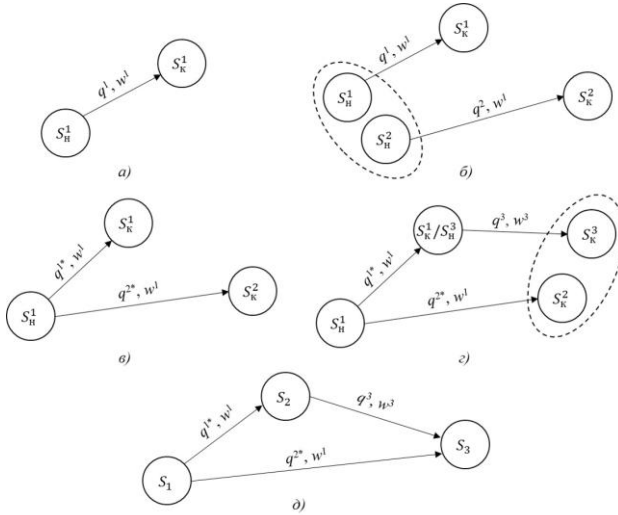


Рис. 2. Этапы построения нечеткой ситуационно-прецедентной сети в соответствии с прецедентным подходом

Пусть также был получен кортеж  $(S_H^2, S_K^2, w^1, q^2)$ , для которого начальные и конечные ситуации не совпадают с кортежем  $(S_H^1, S_K^1, w^1, q^1)$ , но при этом применяемое управляющее решение является идентичным. Визуализация данного кортежа в виде участка НСС приведена на рис. 2, б.

Далее определяется попарная схожесть нечетких ситуаций с помощью вычисления нечеткой степени близости, на основе, например, псевдометрического расстояния Евклида для двух нечетких множеств:

$$d_E(S_A, S_B) = \sqrt{\sum_{j=1}^J (\mu_A(p_j) - \mu_B(p_j))^2}, p_j \in P.$$

Если полученное значение меры близости не превышает значение заданного порога, то выполняется слияние нечетких ситуаций в единую ситуацию с агрегированными значениями ситуационных признаков.

Если было произведено слияние двух нечетких ситуаций, выходным переходам которых соответствуют одинаковые управляющие решения, то осуществляется пересчет возможностей срабатывания соответствующих переходов по формуле:

$$q^{i*} = q^i / \sum_{j=1}^n q^j,$$

где  $q_i$  – возможность срабатывания текущего перехода;  $q_j$  – возможность срабатывания перехода, которому соответствует совпадающее с текущим управляющее решение,  $n$  – количество выходящих из ситуации переходов.

Полученный фрагмент НСС приведен на рис. 2,в.

Пусть далее был получен кортеж  $(S_n^3, S_k^3, w^3, q^3)$ , у которого начальная ситуация  $S_n^3$  совпадает с конечной в первом кортеже  $S_k^1$ . В данном случае расчет меры близости проводить не целесообразно и ситуации могут сразу быть объединены. В случае, если полное совпадение отсутствует, то определение близости ситуаций осуществляется в соответствии с предыдущим примером, как показано на рис. 2,г.

Фрагмент НСС для случая совпадения ситуаций  $S_k^2$  и  $S_k^3$  приведен на рис. 2,д.

В общем виде процедура построения НСС на основе применения прецедентного подхода может быть формализована в виде последовательности шагов, приведенной ниже.

1. Формирование кортежа  $(S_n^i, S_k^i, w^i, q^i)$  некоторого  $i$ -ого этапа управления реальной СТС.
2. Поиск прецедентов для входных и выходных ситуаций кортежа в существующем фрагменте НСС.
3. Если прецедент не был найден, то осуществляется дополнение сети фрагментом текущего кортежа.
4. Если был найден прецедент, то производится слияние соответствующих нечетких ситуаций.
5. Если найден прецедент, выходной переход которого имеет управляющее решение  $w^i$ , то осуществляется пересчет значений возможности срабатывания соответствующих переходов.

На этапе применения для текущей ситуации ищется соответствующий ей прецедент, а далее с помощью НСС определяется оптимальный с точки зрения заданного критерия маршрут до целевой ситуации.

### 3. Результаты моделирования

В работе с использованием предложенного подхода было проведено моделирование функционирования сложной химико-технологической системы производства окатышей из мелкодисперсных отходов добычи апатит-нефелиновых руд [Пучков и др., 2022]. Данная СТС состоит из тарельчатого гранулятора, определяющего габариты и свойства сырого окомкованного сырья, а также обжиговой машины конвейерного типа, состоящей из отдельных вакуум-камер для определенной стадии производства конечного продукта, и имеет собственные характеристики. Важной особенностью данной системы является комплексный характер взаимодействия компонентов, а также наличия множества внутренних и внешних параметров определяющих возможную вариативность её технологических состояний.

В качестве ситуационных признаков  $p_j$  данной системы могут рассматриваться скорость и температура газа теплоносителя в каждой камере, а также степень увлажнения окатышей на каждой стадии технологического процесса. Под нечеткими ситуациями  $S_i$  понимаются ситуации, определяющие качество готового продукта (механические свойства, уровень остаточной влажности, количество брака). Управляющие переходы  $W_{j,k}$  включают управляющие решения (например, «малое снижение температур газа в вакуум камере 1») и веса, учитывающие совокупные затраты тепловой и электрической энергии на производство готового продукта, а также режимы работы оборудования с точки зрения их влияния на степень износа. В качестве основной критерии при управлении подобными СТС, как правило, рассматривается минимизация удельных затрат тепловой и электрической энергии на единицу массы готовой продукции.

База потенциальных прецедентов была сформирована на основе имеющихся исторических данных и результатов применения разработанных ранее математических моделей процессов функционирования рассматриваемой СТС [Bobkov et al., 2022].

Для сравнения полученных результатов был также использован известный подход к построению НСС. Результаты эксперимента показали, что применение предлагаемого варианта построения НСС позволяет обеспечить снижение энергетических затрат на переход системы в требуемого состояния в среднем на 7%. Также было зафиксировано уменьшение быстродействия при поиске решения в среднем на 3% по сравнению с применением классического подхода, что вызвано общим усложнением структуры итоговой НСС. Стоит отметить, что для рассматриваемой СТС выявленное снижение оперативности управления не является существенным, поскольку протекание её процессов носит инерционный характер.

## Заключение

В статье предлагается развитие нечеткого ситуационного подхода к анализу и моделированию процессов в сложных системах. Предлагаемый подход ориентирован на гибкое построение и адаптивную структурно-параметрическую настройку нечетких ситуационных моделей (сетей), учитывающую обобщение и актуализацию сведений о прецедентных ситуациях, имеющих место в процессе функционирования и развития современных производственно-технологических систем, объектов агропромышленного комплекса.

Отличием данного подхода от известных является возможность динамической настройки и корректировки нечеткой ситуационной сети на основе актуальных данных, а также ретроспективной информации об эксплуатации системы в различных условиях. Предложенный подход позво-

ляет учесть сложные зависимости между системными и внешними факторами, неполноту и неопределенность сведений о состоянии и функционировании систем. В работе предложена обобщенная схема управления СТС на основе комплексного применения нечетких ситуационных сетей и прецедентного подхода. Описана процедура построения нечеткой ситуационной сети, отличающаяся от известных учетом ретроспективной информации о функционировании системы и позволяющая динамически осуществлять настройку и корректировку нечеткой ситуационной сети на основе прецедентов, а также минимизировать необходимость использования экспертной информации.

### Список литературы

- [Бобряков и др., 2022] Бобряков А.В., Борисов В.В., Мисник А.Е., Прокопенко С.А. Моделирование и проектирование информационно-аналитических производственных процессов на основе нейронечетких темпоральных сетей Петри // Прикладная информатика. – 2022. – Т. 17, № 2 (98).
- [Борисов и др., 2021] Борисов В.В., Авраменко Д.Ю. Нечеткое ситуационное управление сложными системами на основе их композиционного гибридного моделирования // Системы управления, связи и безопасности. – 2021. – № 3.
- [Будзко и др., 2023] Будзко В.И., Меденников В.И. Условия результативного применения технологий искусственного интеллекта в агропромышленном комплексе ЕАЭС // Труды Института системного анализа Российской академии наук. – 2023. – Т. 73, № 1. – С. 148-158.
- [Гумеров и др., 2022] Гумеров Э.А., Алексеева Т.В. Конкурентоспособность системы промышленного интернета вещей // Современная конкуренция. – 2022. – Т. 16, № 1 (85).
- [Еремеев и др., 2021] Еремеев А.П., Варшавский П.Р., Поляков С.А. Программная реализация модуля анализа данных на основе прецедентов для распределенных интеллектуальных систем // Программные продукты и системы. – 2021. – № 3.
- [Зацаринный и др., 2023] Зацаринный А.А., Меденников В.И., Райков А.Н. Интеграция приложений искусственного интеллекта в единую цифровую платформу АПК // Информационное общество. – 2023. – № 1. – С. 127-138.
- [Пучков и др., 2022] Пучков А.Ю., Дли М.И., Прокимов Н.Н., Шутова Д.Ю. Многоуровневые алгоритмы оценки и принятия решений по оптимальному управлению комплексной системой переработки мелкодисперсного рудного сырья // Прикладная информатика. – 2022. – Т. 17, № 6 (102).
- [Рожков и др., 2023] Рожков В.В., Крутиков К.К., Федотов В.В., Бутримов С.Г. Динамическое имитационное моделирование системы возбуждения синхронных генераторов стационарных дизель-генераторных установок аварийного электроснабжения атомной станции // Прикладная информатика. – 2023. – Т. 18, № 1 (103).
- [Черновалова и др., 2023] Черновалова М.В., Борисов В.В., Власова Е.А. Интеллектуальная поддержка управления процессами обработки рудного сырья на основе прецедентного подхода и онтологических моделей // Прикладная информатика. – 2023. – Т. 18, № 2 (104).

- [Bobkov et al., 2022] Bobkov V.I., Chernovalova M.V., Sokolov A.M. Mathematical models for an energy intensive multistage roasting process of pelletized industrial waste from apatite-nepheline ores // Materials Science Forum. – 2022. – Vol. 1052.
- [Dobrynin et al., 2020] Dobrynin A.S., Gudkov M.Y., Koynov R.S. A precedent approach to incident management in automated process control systems // Software systems and computational methods. – 2020. – No. 2.
- [Evstifeev et al., 2021] Evstifeev A. A., Zaeva M. A. Method of Applying Fuzzy Situational Network to Assess the Risk of the Industrial Equipment Failure // Procedia Computer Science. – 2021. – Vol. 190.
- [Bosdelekidis et al., 2022] Bosdelekidis V., Torleiv H.B., Sokolova N., Johansen T.A. Navigation algorithm-agnostic integrity monitoring based on solution separation with constrained computation time and sensor noise overbounding // Journal of Intelligent & Robotic Systems. – 2022. – Vol. 106, No. 7.

## НЕЧЕТКАЯ МОДЕЛЬ ДИАГНОСТИКИ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ

К.Б. Фам (*phambang79520897405@gmail.com*)

П.М. Мурашев (*myptver@gmail.com*)

В.Н. Богатиков (*vnbgtk@mail.ru*)

Тверской государственной технической университет, Тверь

В этой работе рассмотрено формирование критерия оценки состояний на основе нечёткого индекса безопасности на примере промышленных технологий сушки зелёного чая и шаровой мельницы измельчения. Рассмотрена формализованная система обеспечения безопасности работы технологических процессов. Для представления упорядоченного множества неисправностей использован алгоритм построения диаграммы Хассе. Задача диагностики рассмотрена на примере промышленных процессов сушки зелёного чая и измельчения апатитонефелиновой руды.

**Ключевые слова:** диагностика технологических процессов, диаграмма Хассе, нечеткое равенство, нечеткое включение, поиск в графе.

### Введение

В связи с тенденциями развития, как самих технологий, так и систем управления ими, приходится решать ряд параллельных задач – модернизации самого технологического процесса, оборудования и систем управления процессами. В связи с этим, актуализируются задачи устранения неисправностей. Становится все более важной для промышленных систем управления задача роботизации методов управления и обнаружения неисправностей. Для решения этой задачи были разработаны различные методы, которые включают подходы, основанные на моделях знаний, качественном моделировании, нейронных сетях, и так же классические многомерные статистические методы.

Решая слабоструктурированные, неформализованные проблемы, человек просто вынужден оперировать качественными суждениями. Однако качественное мышление также нуждается в опоре на вспомогательные средства, которые облегчают сложный концептуальный анализ, позволяют выявить границы возможных действий и, главное, помогают найти перспективное направление поиска решений.

Онлайн-реализация методов диагностики неисправностей становится важной темой исследований в связи с растущим спросом на создание промышленных технологий с более высокой производительностью, эффективностью, надежностью как систем управления, так и оборудования технологических процессов. Классические методы обнаружения отказов основаны на проверке предельных значений некоторых важных измеряемых переменных. В настоящее время наблюдается тенденция слияния решения задач диагностики контролируемых параметров самих технологических процессов, систем управления ими, состояния основного оборудования процессов. Такое слияние опирается методологию технологий основанных на знаниях – экспертные системы, интеллектуальная обработка сигналов.

### **1. Построение математического описания нечёткой оценки состояний промышленной технологии на основе индекса безопасности**

В работе предлагается сравнительно новый подход в оценке состояний промышленной системы. В первоначальных вариантах предлагаемый критерий использовался в основном в экспертных системах [Мелихов и др., 1990], [Богатиков и др., 2022] В данной работе такой критерий предложено использовать как задание регулятору. В процессе функционирования технологического процесса за счет изменения значений его параметров происходит непрерывное изменение состояния систем. Каждое из этих состояний системы характеризуется набором значения технологических параметров –  $T = \{T_1, T_2, \dots, T_p\}$ . Каждый параметр из этого набора имеет наиболее благоприятную область работы. Во время его рабочего (штатного) функционирования, множество  $T_{ii} = \{T_{i1}, T_{i2}, \dots, T_{ip}\}$  является центром этой области [Тоичкин и др., 2005]. Обычно эта точка определяется исходя из свойств технологического процесса. В реальных технологических условиях существуют различные методы оценки центра безопасности. Например: экспертный опрос [Борисов и др., 1990], метод разделения состояний [Богатиков и др., 2022]. Выходы параметров из области безопасности может вызывать критические ситуации, поэтому для оценки степени безопасности системы при текущим состоянии необходимо определить удаленность текущей рабочей точки процесса  $s^*$  от точки центра безопасности  $s_0$  [Богатиков и др., 2022]. При этом каждый параметр  $T_i$  описывается соответствующей лингвистической переменной  $\langle \beta_i, E_i, D_i \rangle$  [Тоичкин и др., 2005], [Борисов и др., 1990] где:

$\beta_i$  – название переменной;

$E_i = \{E_i^1, E_i^2, \dots, E_i^M\}$  – терм-множество переменной  $\beta_i$ ;

$D_i$  – базовое множество лингвистической переменной  $\beta_i$ ;

На рис. 1 представлен пример функций принадлежности одного параметров.

Нечеткие ситуации S описываются таким образом:

$$\tilde{S} = \{ \langle \mu_s(T_i)/T_i \rangle, T_i \in T \},$$

где  $\mu_s(T_i) = \{ \langle \mu_{\mu_s(T_i)}(E^j_i)/E^j_i \rangle, j = 1 \dots M; i = 1 \dots p \}$ .

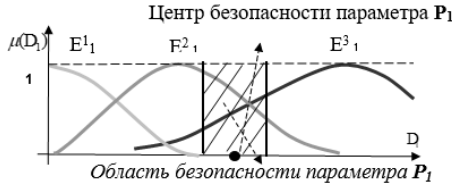


Рис. 1. Представление функции принадлежности параметров диагностики

Уровень безопасности системы в данный момент времени можно определить через степень нечеткого равенства [Мелихов и др., 1990] между текущим состоянием и центром безопасности и в работе названа индексом безопасности. Данная степень рассматривает два критерия: степени нечеткого включения текущего состояния в центр безопасности и центра безопасности в текущее состояние. Процедура определения степени равенства между текущим состоянием и центром безопасности осуществляется следующим образом:

$$\mu(\tilde{S}_{\text{теку}} \leftrightarrow \tilde{S}_{\text{Ц}}) = \vartheta(\tilde{S}_{\text{теку}}, \tilde{S}_{\text{Ц}}) \& \vartheta(\tilde{S}_{\text{Ц}}, \tilde{S}_{\text{теку}}), \quad (1.1)$$

где  $\vartheta(\tilde{S}_{\text{теку}}, \tilde{S}_{\text{Ц}})$  — степень включения текущего состояния  $\tilde{S}_{\text{теку}}$  в центр безопасности  $\tilde{S}_{\text{Ц}}$ ;

$\vartheta(\tilde{S}_{\text{Ц}}, \tilde{S}_{\text{теку}})$  — степень включения центра безопасности  $\tilde{S}_{\text{Ц}}$  в текущее состояние  $\tilde{S}_{\text{теку}}$ .

$$\vartheta(\tilde{S}_{\text{теку}}, \tilde{S}_{\text{Ц}}) = \& \left( \mu_{S_{\text{теку}}}(T_i) \rightarrow \mu_{S_{\text{Ц}}}(T_i) \right), \quad i = 1 \dots n \quad (1.2)$$

$$\mu_{S_{\text{теку}}}(T_i) \rightarrow \mu_{S_{\text{Ц}}}(T_i) = \max \left\{ 1 - \mu_{S_{\text{теку}}}(T_i), \mu_{S_{\text{Ц}}}(T_i) \right\} \quad (1.3)$$

$$\vartheta(\tilde{S}_{\text{Ц}}, \tilde{S}_{\text{теку}}) = \& \left( \mu_{S_{\text{Ц}}}(T_i) \rightarrow \mu_{S_{\text{теку}}}(T_i) \right), \quad i = 1 \dots n \quad (1.4)$$

$$\mu_{S_{\text{Ц}}}(T_i) \rightarrow \mu_{S_{\text{теку}}}(T_i) = \max \left\{ 1 - \mu_{S_{\text{Ц}}}(T_i), \mu_{S_{\text{теку}}}(T_i) \right\} \quad (1.5)$$

Полученное нечеткое равенство между текущим состоянием и центром безопасности сравнивается с установленным значением порога (примерно 0,6). В случае, когда данное равенство больше порога, можно делать вы-



вод, что текущее состояние системы находится в рамках безопасности, а когда уровень безопасности ниже порога, то работа оборудования нарушается и необходимо устранить неисправности.

## 2. Обобщённая методика применение нечеткого математического описания к обнаружению нарушения работы различных промышленных процессов на примере процессов сушки зелёного чая и измельчения апатитонепелиновой руды

### 2.1. Для процесса сушки зелёного чая

Рассмотрим пример применения диагностики к сушильному оборудованию по данному подходу [Черноусов, 2020]. На рис. 2 представлена технологическая схема данного промышленного оборудования.

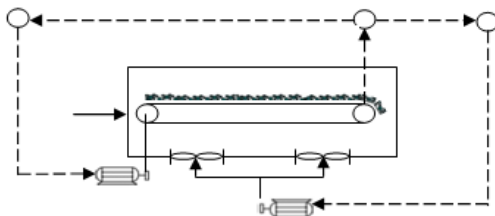


Рис. 2. Схема конвейерной сушилки зеленого чая

На основании знаний экспертов и отклонения технических параметров в работе агрегата при возникновении критической ситуации были выбраны параметры, которые оценивают степень безопасности. Ими являются сильное повышение давления в камере сушилки, и также отклонения температуры, скоростей ленты и воздуха от задания. Нечеткая ситуация представляется в виде:

{«0/ “малый –  $E_1^1$ ”, 80/ “средний –  $E_1^2$ ”, 150/ “большой –  $E_1^3$ »}/ температура – признак 1;  
 «0/ “малый –  $E_2^1$ ”, 0,3/ “средний –  $E_2^2$ ”, 0,6/ “большой –  $E_2^3$ »}/ давление пары – признак 2;  
 «0/ “малый –  $E_3^1$ ”, 1/ “средний –  $E_3^2$ ”, 2/ “большой –  $E_3^3$ »}/ скорость воздуха – признак 3;  
 «0/ “малый –  $E_4^1$ ”, 0,5/ “средний –  $E_4^2$ ”, 1/ “большой –  $E_4^3$ »}/ скорость ленты – признак 4}

После получения текущего значения и заданного значения (центра безопасности) необходимых параметров, данные преобразуются в нечетком виде для процесса вычисления степени нечеткого равенства между текущей точкой системы и точкой центра безопасности.

Рассмотрим следующую конкретную ситуацию параметров процесса сушки. На рис. 3 представлены примеры процесс преобразования полученных данных в нечетком виде.

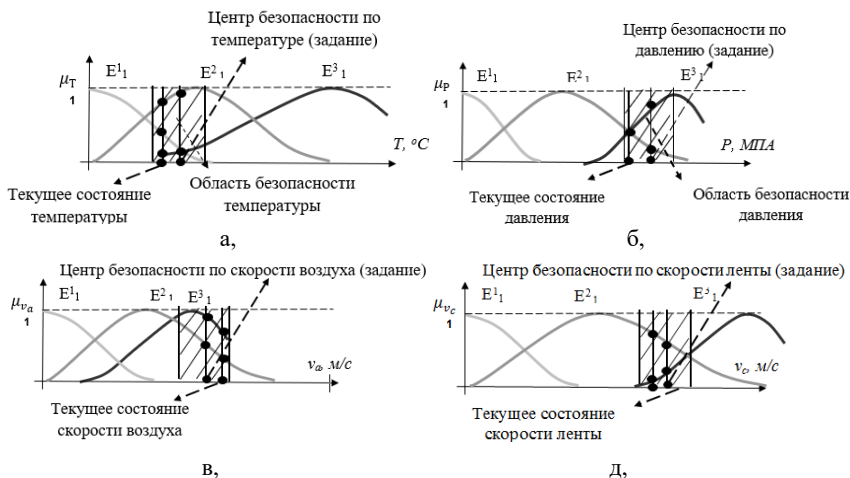


Рис. 3. Функции принадлежности параметров диагностики процесса сушки: а – температуры, б – давления, в – скорости воздуха и д – скорости ленты

В данном примере нечеткие параметры диагностики имеет следующей значения (табл. 1):

Таблица 1

### Нечеткие значения по параметрам

Парам.	T <sub>1</sub> – Отклонение температуры			T <sub>2</sub> – Отклонение давления			T <sub>3</sub> – Отклонение скорости воздуха			T <sub>4</sub> – Отклонение скорости ленты		
	E <sub>1</sub> <sup>1</sup>	E <sub>1</sub> <sup>2</sup>	E <sub>1</sub> <sup>3</sup>	E <sub>2</sub> <sup>1</sup>	E <sub>2</sub> <sup>2</sup>	E <sub>2</sub> <sup>3</sup>	E <sub>3</sub> <sup>1</sup>	E <sub>3</sub> <sup>2</sup>	E <sub>3</sub> <sup>3</sup>	E <sub>4</sub> <sup>1</sup>	E <sub>4</sub> <sup>2</sup>	E <sub>4</sub> <sup>3</sup>
Терм	0,15	0,9	0,15	0	0,2	0,8	0	0,5	0,9	0	0,6	0,2
Тек.	0,4	0,8	0,1	0	0,4	0,6	0	0,3	0,65	0	0,65	0,1

Степень включения текущего состояния  $\tilde{S}_{\text{теку}}$  в центр безопасности  $\tilde{S}_{\text{ц}}$ :

$$\vartheta(\tilde{S}_{\text{теку}}, \tilde{S}_{\text{ц}}) = \& \left( \mu_{s_{\text{теку}}}(T_i) \rightarrow \mu_{s_{\text{ц}}}(T_i) \right) = 0,6.$$

Степень включения центра безопасности  $\tilde{S}_{Ц}$  в текущее состояние  $\tilde{S}_{теку}$ :

$$\vartheta(\tilde{S}_{Ц}, \tilde{S}_{теку}) = \& \left( \mu_{S_{Ц}}(T_i) \rightarrow \mu_{S_{теку}}(T_i) \right) = 0,6.$$

Степень равенства между текущим состоянием и центром безопасности:

$$\mu(\tilde{S}_{теку} \leftrightarrow \tilde{S}_{Ц}) = \vartheta(\tilde{S}_{теку}, \tilde{S}_{Ц}) \& \vartheta(\tilde{S}_{Ц}, \tilde{S}_{теку}) = 0,6 \geq 0,6 \text{ (порог)}.$$

Видно, что полученные степени равенства между текущим состоянием и центром безопасности равно порогу (0,6) поэтому данная ситуация находится в зоне безопасности.

После обнаружения нарушения безопасности ТП необходимо найти место неисправности в оборудовании. Рассмотрим дефекты, которые возможны в сушильном аппарате (табл. 2).

Таблица 2

### Дефекты, возникающие в сушилке

Код	Описание
O <sub>0</sub>	Нарушение работы камеры сушилки
O <sub>1</sub>	Нарушение работы двигателя для ленты в камере сушилки
O <sub>2</sub>	Нарушение работы нагревателя
O <sub>3</sub>	Нарушение работы вентилятора
O <sub>4</sub>	Неисправность системы регулирования двигателя ленты
O <sub>5</sub>	Неисправность системы регулирования вентилятора
O <sub>6</sub>	Неисправность регулятора температуры воздуха
O <sub>7</sub>	Неисправность датчика измерения скорости ленты
O <sub>8</sub>	Неисправность датчика измерения скорости воздуха
O <sub>9</sub>	Неисправность датчика измерения влажности
O <sub>10</sub>	Неисправность датчика измерения давления
O <sub>11</sub>	Неисправность датчика измерения температуры
O <sub>12</sub>	Неисправность клапана воздуха
O <sub>13</sub>	Неисправность клапана зеленого чая
O <sub>14</sub>	Разбито смотровое окно
O <sub>15</sub>	Повышение значения давления
O <sub>16</sub>	Изменение скорости движения ленты
O <sub>17</sub>	Изменение значения температуры воздуха
O <sub>18</sub>	Остановка ленты
O <sub>19</sub>	Остановка вентилятора
O <sub>20</sub>	Остановка нагревателя

Для изучения взаимодействия между дефектами построена матрица нечеткого отношения, которая характеризуется степенями порождения одного дефекта к другому дефекту (табл. 3). Нечеткий граф развития опасностей изображен на рис. 4.

Нечеткое отношение  $\mu_{R_1}(o_i, o_j)$ 

	O <sub>0</sub>	O <sub>1</sub>	O <sub>2</sub>	O <sub>3</sub>	O <sub>4</sub>	O <sub>5</sub>	O <sub>6</sub>	O <sub>7</sub>	...	O <sub>14</sub>
O <sub>0</sub>	1	0	0	0	0	0	0	0	...	0
O <sub>1</sub>	0,9	1	0	0	0	0	0	0	...	0
O <sub>2</sub>	0,9	0	1	0	0	0	0	0	...	0
O <sub>3</sub>	0,9	0	0	1	0	0	0	0	...	0
O <sub>4</sub>	0,6	0,8	0	0	1	0	0	0	...	0
O <sub>5</sub>	0,6	0	0	0,9	0	1	0	0	...	0
O <sub>6</sub>	0,5	0	0,9	0	0	0	1	0	...	0
O <sub>7</sub>	0,5	0,8	0	0	0	0,8	0	1	...	0
O <sub>8</sub>	0,5	0	0	0,8	0,8	0	0	0	...	0
O <sub>9</sub>	0,5	0	0	0	0	0	0	0	...	0
O <sub>10</sub>	0,7	0	0	0,5	0	0	0	0	...	0
O <sub>11</sub>	0,6	0	0,8	0	0	0	0,8	0	...	0
O <sub>12</sub>	0,5	0	0,7	0	0	0	0	0	...	0
O <sub>13</sub>	0,5	0,5	0	0	0	0	0	0	...	0
O <sub>14</sub>	0,7	0	0	0	0	0	0	0	...	1

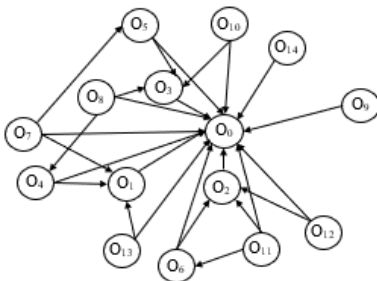


Рис. 4. Нечеткий граф развития неисправностей сушильного оборудования

Далее проводится анализ и преобразование нечеткого графа развития дефектов, целью которого является выделение в нем иерархической структуры, позволяющей проводить процедуру диагностирования внештатных ситуаций на множестве опасностей. Данная иерархическая структура является диаграммой Хассе [Попов и др., 2015], [Порешин и др., 2014] Алгоритм построения этой диаграммы заключается в следующем: на первом уровне диаграммы находятся только вершины, из которых не выходит ни одной дуги, на втором – вершины, из которых выхо-

дят дуги, инцидентные только вершинам первого уровня и на следующих уровнях есть вершины, из которых выходят дуги, инцидентные вершинам, расположенным на уровнях с первого по данный уровень (рис. 5).

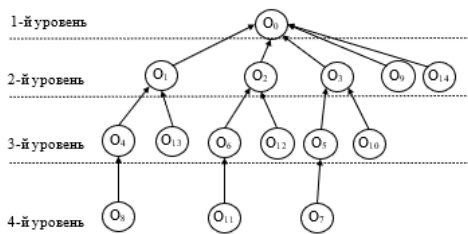


Рис. 5. Диаграмма Хассе [Черноусов, 2020] развития неисправностей сушильного оборудования

## 2.2. Для процесса измельчения

Как для процесса сушки, так и для процесса измельчения, используя знания экспертов были выбраны параметры, которые оценивают степень безопасности. Это давление в клиновом зазоре, и также температура и скорость возвращения барабана. Порядок проверки наличия неисправностей такой же как для процесса сушки. Далее рассмотрим построение диаграмма Хассе для поиска дефекта для данного процесса. На рис. 6. представлена схема цепи аппаратов технологического процесса измельчения (ГЦ – гидроциклон).

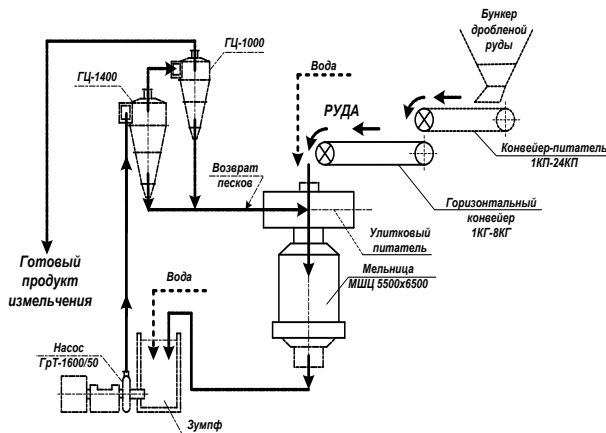


Рис. 6. Схема цепи аппаратов технологического процесса измельчения (ГЦ – гидроциклон)

В табл. 4 перечислены дефекты, которые могут возникнуть в ходе работы мельницы.

Таблица 4

**Дефекты, возникающие в барабанной мельнице**

Обозначение дефекта	Дефект
O <sub>0</sub>	Повышение температуры
O <sub>1</sub>	Снижение давления в клиновом зазоре
O <sub>2</sub>	Отсутствие смазки
O <sub>3</sub>	Забит трубопровод
O <sub>4</sub>	Износ насоса
O <sub>5</sub>	Забит фильтр (раз в неделю)
O <sub>6</sub>	Наличие большого количества воды в масле (постоянно)
O <sub>7</sub>	Забиты трубы
O <sub>8</sub>	Износ маслонасоса
O <sub>9</sub>	Низкое качество шабровки
O <sub>10</sub>	Нарушение баббитовой заливки
O <sub>11</sub>	Наличие инородных тел в баббитовой заливке
O <sub>12</sub>	Неравномерное прилегание цапфы на баббитовую поверхность

Матрица нечеткого отношения, которая характеризуется степенями порождения одного дефекта к другому дефекту (табл. 5). Диаграмма Хасе развития дефектов процесса изменения изображена на рис. 7.

Таблица 5

**Нечеткое отношение  $\mu_{R_1}(\sigma_i, \sigma_j)$**

	O <sub>0</sub>	O <sub>1</sub>	O <sub>2</sub>	O <sub>3</sub>	O <sub>4</sub>	O <sub>5</sub>	O <sub>6</sub>	O <sub>7</sub>	O <sub>8</sub>	O <sub>9</sub>	O <sub>10</sub>	O <sub>11</sub>
O <sub>0</sub>	1	0	0	0	0	0	0	0	0	0	0	0
O <sub>1</sub>	0,8	1	0	0	0	0	0	0	0	0	0	0
O <sub>2</sub>	0,7	0	1	0	0	0	0	0	0	0	0	0
O <sub>3</sub>	0,5	0,7	0	1	0	0	0	0	0	0	0	0
O <sub>4</sub>	0,6	0,8	0	0	1	0	0	0	0	0	0	0
O <sub>5</sub>	0,5	0,7	0,8	0	0	1	0	0	0	0	0	0
O <sub>6</sub>	0,3	0,7	0	0	0	0	1	0	0	0	0	0
O <sub>7</sub>	0,5	0,8	0,8	0	0	0	0	1	0	0	0	0
O <sub>8</sub>	0	0	0,9	0	0	0	0	0	1	0	0	0
O <sub>9</sub>	0,1	0	0	0	0	0	0	0	0	1	0	0,7
O <sub>10</sub>	0	0	0	0	0	0	0	0	0	0	1	0,8
O <sub>11</sub>	0,9	0	0	0	0	0	0	0	0	0	0	1

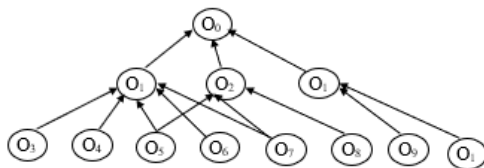


Рис. 7. Диаграмма Хассе развития неисправностей оборудования измельчения

### 3. Порядок поиска неисправности при выходе системы за ограничения области безопасности по графу Хассе

Для нахождения неисправности по полученную графу существуют достаточно много методов, например поиск в ширину (BFS), поиск в глубину (DFS), алгоритм Дейкстры, поиск по первому наилучшему совпадению, алгоритм Беллмана, и т.д. [Листопад и др., 2016]. В данной работе предположено использования метода поиска в ширину для построения последовательности проверки узлов графа при возникновении признаки выхода за ограничения области безопасности системы. Поиск в ширину проиллюстрирован на рис. 8.

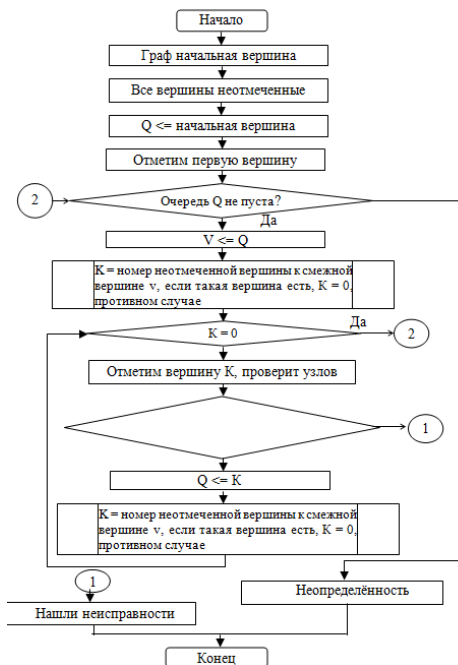


Рис. 8. Алгоритм поиска неисправности в графу

Алгоритм данного метода осуществляется следующим образом [Алексеев, 2017]:

1. Начинаем поиск с произвольной вершины  $g$ . Формируем множество текущих вершин  $A$ , включив в него вершину  $g$ .

2. Идем в ранее не пройденные вершины по всем дугам, имеющим начальную вершину, принадлежащую множеству  $A$ . Запоминаем эти дуги. Формируем множество  $A$ , включив в него конечные вершины пройденных дуг.

3. Процесс поиска (пункт 2) заканчивается, когда множество  $A$  станет пустым.

### Заключение

В данной работе была рассмотрена методика создания системы безопасности технологических процессов с использованием нечеткой логики. Поиск первопричин возникновения внештатных ситуаций осуществляется на основе диаграммы Хассе, которая отражает структуру опасностей. В заключении доклада приводится алгоритм поиска неисправностей с использованием диаграммы Хассе. Применение алгоритма показано на примере двух технологических процессов – сушки зелёного чая и процесса измельчения апатитонелефиновой руды.

### Список литературы

- [Алексеев, 2017] Алексеев В.Е. Дискретная математика: учебное пособие. – Нижний Новгород: Нижегородский госуниверситет, 2017.
- [Богатиков и др., 2022] Богатиков В.Н., Виноградов Г.П., Волков В.Ю. и др. Разработка интеллектуального управления в многоуровневых промышленных системах в условиях неполной информации на основе нечёткой формализации представлений о параметрах технологических процессов. – Новомосковск: Новомосковский институт (филиал) РХТУ имени Д.И. Менделеева, 2022.
- [Борисов и др., 1990] Борисов А.Н., Крумберг О.А., Федоров И.П. Принятие решений на основе нечетких моделей: Примеры использования. – Рига: Зинатне, 1990.
- [Листопад и др., 2016] Листопад Н.И., Карук И.А., Хайдер А.А. Алгоритмы поиска кратчайшего пути и их модификация // Информатизация образования. – 2016. – № 1 (77).
- [Мелихов и др., 1990] Мелихов А.Н., Берштейн Л.С., Коровин С.Я. Ситуационные советующие системы с нечеткой логикой. – М.: Наука, 1990.
- [Попов и др., 2015] Попов Д.И., Виноградов Д.В. Анализ алгоритмов поиска в пространстве состояний // Вестник Московского государственного университета печати. – 2015. – № 6.
- [Порешин и др., 2014] Порешин П.П., Попов Б.Н. Дискретная математика: множества, отношения, логика, автоматы: учебное пособие. – М.: МАИ-ПРИНТ, 2014.
- [Тоичкин и др., 2005] Тоичкин Н.А., Богатиков В.Н. Алгоритм определения центра безопасности для оценки состояния технологического процесса // Информационные технологии в региональном развитии. – 2005. – Вып. 5.
- [Черноусов, 2020] Черноусов, Я.О. Изучение отдельных компонентов химического состава чайного листа. Качественный анализ алкалоидов // Чай в историческом, культурном, медицинском аспекте: Материалы I научно-теоретической онлайн-конференции с международным участием. – 2020.



УДК 004.272

## ОРГАНИЗАЦИЯ ПАРАЛЛЕЛЬНЫХ ВЫЧИСЛЕНИЙ НА ГИБРИДНЫХ ВЫЧИСЛИТЕЛЬНЫХ КЛАСТЕРАХ НА ОСНОВЕ НЕЧЕТКОГО АНАЛИЗА

Я.А. Федулов (*fedulov\_yar@mail.ru*)

А.С. Федулова (*voitsitskay@mail.ru*)

Филиал Национального исследовательского университета «МЭИ»  
в г. Смоленске, Смоленск

В работе предложена программно-аппаратная модель организации параллельных вычислений на гибридных вычислительных кластерах, ориентированная на создание средств преобразования параллельных программ в гибридную форму. Разработан метод организации выполнения параллельных программ на гибридных вычислительных кластерах с применением совместимых нечетких когнитивных карт, ориентированный на расширенные возможности предлагаемой модели программно-аппаратной среды с учетом различной аппаратных и программных показателей, позволяющий снизить долю обменных операций, выполняемых через медленные сетевые интерфейсы.

**Ключевые слова:** нечеткая логика, параллельные вычисления, технологии параллельного программирования, нечеткий анализ данных, преобразование программ.

### Введение

В настоящее время проводится активная работа по разработке параллельных высокопроизводительных вычислительных систем различных типов (вычислительные кластеры, мультимпьютеры, грид системы, гетерогенные параллельные вычислительные системы, облачные платформы) и по процессу их внедрения в промышленность [Торожков et al., 2021].

В качестве целевой исследуемой параллельной вычислительной системы в данной работе был выбран получивший наибольшее распространение класс гибридных вычислительных кластеров (ГВК). Высокопроизводительные гибридные вычислительные кластеры интенсивно развиваются и используются для решения прикладных и научных задач в различных сферах деятельности человека, где требуется расчет корректных и вывод быстрых результатов, полученных с помощью высокоточного и

обладающего высокими техническими характеристиками оборудования, работающего с большим количеством входных данных. ГВК дают возможность эффективного выполнения алгоритмов научно-технических задач, при этом используются различные методы работы с вычислительными и обменными операциями. Данный класс вычислительных систем характеризуется распределенным доступом к памяти при взаимодействии параллельной программы между узлами кластера, а также, зачастую, неоднородным доступом к памяти внутри узлов с общей памятью.

Постоянная потребность улучшения структуры и программного обеспечения ГВК, связанная с нерешенной проблемой эффективного распределения задач на имеющихся вычислительных узлах, приводит к необходимости разработки новых методов и моделей работы с программами и узлами кластера, а также способов распараллеливания кода, учитывающих архитектурные особенности устройства, позволяющих работать с большим количеством входных данных.

В таких условиях важное значение приобретает разработка универсальной модели и метода для функционирования среды распараллеливания программного кода.

Дополнительные требования сокращения временных затрат на решение прикладных задач, упрощения процедур сопровождения и внедрения программно-аппаратных сред распараллеливания на ГВК с учетом рассмотренных особенностей, обосновывают актуальность применения методов нечеткой логики и нечетких множеств [Борисов и др., 2012].

## **1. Программно-аппаратная модель организации параллельных вычислений на гибридных вычислительных кластерах**

Программно-аппаратная модель организации параллельных вычислений в общем случае должна включать в себя следующие составляющие:

- архитектурный компонент – представляется, как правило, в форме графа (схемы), узлы которого соответствуют модулям кластерной архитектуры с различной функциональностью (узлам кластера, процессорам, сетевым устройствам и компонентам более низкого уровня), а дуги – соединениям между модулями, отображающим передачу данных;
- программный компонент, моделирующий программы / алгоритмы, выполняемые на кластере;
- компонент отображения, задающий соответствие между выполняемыми частями параллельной программы и архитектурными модулями, обеспечивающими корректное выполнение алгоритма / программы для определенных входных данных;

- компонент затрат, позволяющий оценить затраты на выполнение параллельного алгоритма / программы.

В работе предложена программно-аппаратная модель организации параллельных вычислений на гибридных вычислительных кластерах. Предлагаемая модель основывается на интегральных показателях, каждый из которых отражает значение отдельной характеристики рассматриваемой программно-аппаратной среды. Среди показателей рассматривались факторы, влияющие на эффективность выполнения программ в гибридных вычислительных кластерах, которые условно можно разделить на два класса: аппаратные и программные [Ежова и др., 2019].

К аппаратным факторам следует отнести следующие: производительность процессоров и их количество; пропускная способность коммуникационной среды; задержка (латентность) при отправке сообщений в системах с распределенной памятью; скорость обмена данными между ядрами процессоров через разделяемую память; объем и уровни кэш-памяти, тактовая частота и число ядер процессоров на узлах кластера. В отдельную группу аппаратных факторов можно выделить характеристики графических процессорных устройств, к которым относятся: тактовая частота видеочипа и видеопамати; пропускная способность видеопамати и ее объем; пропускная способность интерфейса взаимодействия центрального процессора с графическим процессорным устройством (наиболее распространенный вариант интерфейса – PCI-Express); размерность сетки блоков видеокарты, размер отдельного блока, максимальное число нитей в блоке, выделяемых для вычислений [Juan et al., 2019].

Отличительной чертой факторов аппаратного класса, влияющих на эффективность выполнения параллельных программ на ГВК, является сложность их оперативного изменения.

Совокупность аппаратных факторов совместно с графическим представлением определяют архитектурную составляющую системы [Sokolinsky, 2021].

Пример типовой архитектуры ГВК [Борисов и др., 2016а] представлен на рис. 1. Каждый узел состоит из двух многоядерных процессоров. Внутри одного узла центральные процессоры имеют неоднородный доступ к общей памяти по технологии Non Uniform Memory Access (NUMA). Оба центральных процессора имеют быстрый доступ напрямую к ассоциированной с ними оперативной памяти, в то время как доступ к оперативной памяти другого процессора внутри узла осуществляется с дополнительными накладными расходами через двухканальный интерфейс Quick Path Interconnect (QPI). Графический процессор имеет в доступе собственную память и обменивается данными с процессором по интерфейсу PCI Express.

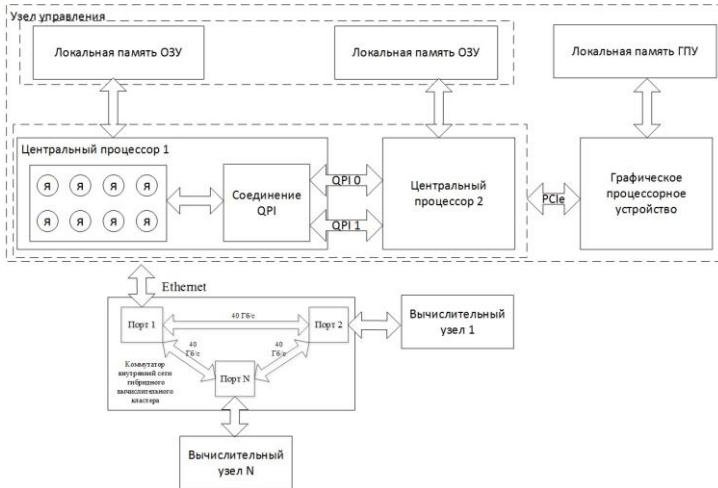


Рис. 1. Типовая архитектура ГВК

В качестве программного компонента в работе предложено использовать новую модель гибридной параллельной программы. Она основана на совокупности программных показателей (факторов) и имеет иерархическую (двухуровневую) структуру. К программным факторам следует отнести: особенности технологий параллельного программирования; возможности операционной системы; особенности компилятора; характеристики и особенности параллельных алгоритмов и программ.

Двухуровневая схема построения модели гибридной параллельной вычислительной программы представлена на рис. 2.

Предлагаемая структура параллельной программы состоит из двух уровней. На первом (внешнем) уровне применяется технология передачи сообщений в параллельных вычислительных системах Message Passing Interface (MPI). В программе порождается число MPI процессов равное доступному числу узлов кластера  $N$ , на которых планируется запуск программы, этому уровню на рис. 2 соответствуют закрашенные полужирные стрелки, закрашенные круги отображают процессы MPI на узлах кластера. Исходная программа разбивается на  $N$  равных участков, которые затем рассылаются на каждый из задействованных узлов кластера. Основной способ передачи данных между порождаемыми MPI процессами – широковещательные рассылки, использующие функцию `MPI_Broadcast`. Результаты вычислений агрегируются на узле управления (процесс с нулевым порядковым номером) с помощью директив редукции по определенной операции агрегирования `MPI_Reduce`.

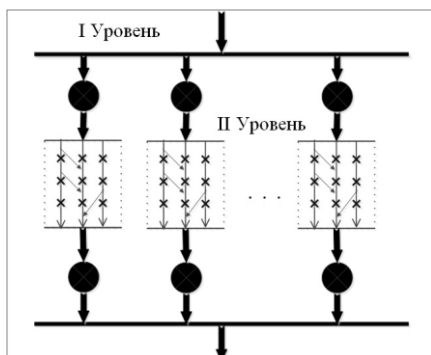


Рис. 2. Модель гибридной параллельной программы

На втором, внутреннем уровне модели, происходит дополнительное разделение уже распределенного по узлам кластера кода на число параллельных нитей (потоков), доступное в выбранной технологии распараллеливания внутри одного кластерного узла (OpenMP, CUDA, OpenCL). Второму уровню соответствуют тонкие линии, вертикальные стрелки отражают параллельно выполняемые подзадачи, крестики отображают вычислительные операции, косые стрелки отображают передачу промежуточных результатов между потоками внутри одного кластера. Результаты вычислений, полученные на втором внутреннем уровне программы агрегируются средствами выбранной технологии в управляющий процесс на узле кластера, данные с которого затем собираются на нулевом процессе узла управления. Узел управления также может содержать достаточные вычислительные мощности и сам участвовать в вычислениях.

Компонент отображения гибридной параллельной программы на архитектуру кластера в программно-аппаратной модели представлен на рис. 3.

Необходимо отметить, что предлагаемое преобразование параллельной программы к гибриднему виду имеет смысл производить, лишь в случае достаточно весомой доли обменных операций в исходном коде программы. Разработанная модель также содержит ряд специальных показателей, вычисляемых с учетом обязательных характеристик эффективного распараллеливания, предоставляя следующие возможности:

- определение преобладающего типа операций в решаемой задаче (вычислительные операции с операндами действительного, строкового или целого типа);

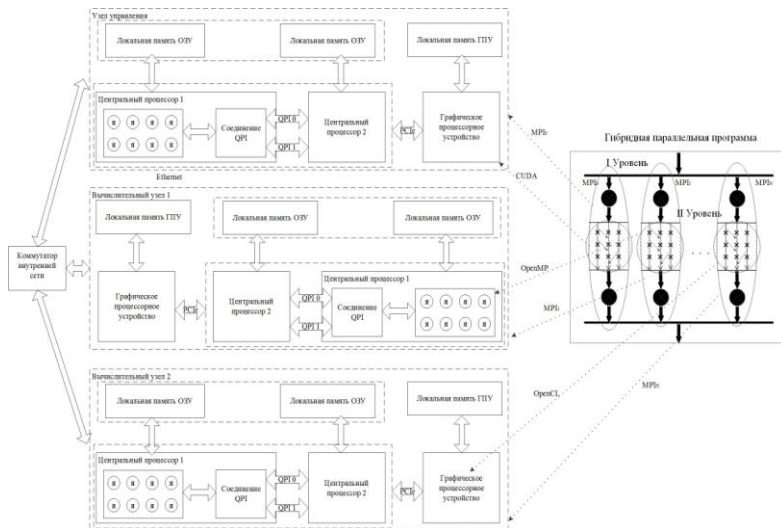


Рис. 3. Отображение гибридной программы на архитектуру кластера

- возможность разделения алгоритма вычислительных и обменных операций (декомпозиция кода) на части (подзадачи), которые выполняются независимо друг от друга (параллельно);
- определение типов взаимодействия для полученного набора подзадач, которые осуществляются в ходе решения поставленной задачи;
- определение параллельной вычислительной системы, её характеристик и технологий, на которых будет производиться распараллеливание алгоритма между процессорами системы.

## 2. Метод организации параллельных вычислений на гибридных вычислительных кластерах

По результатам проведенных исследований был предложен метод организации выполнения параллельных программ, применяемый на гибридных вычислительных кластерах. Метод заключается в преобразовании исходной параллельной программы, построенной с использованием технологии передачи сообщений MPI в гибридную форму с целью сокращения числа медленных обменов данными, как между узлами, так и между процессами внутри одного узла кластера. В результате учитывается главный недостаток интерфейсов передачи сообщений, заключающийся в использовании для обменных операций сокетов и сетевых устройств с заведомо более низкой пропускной способностью, чем шины данных, соединяющие оперативную память с процессорами в структурах с общей памятью.

В качестве альтернативы передачам сообщений на основе сетевых компонентов предлагается преобразовывать параллельную программу таким образом, чтобы обменные операции, по возможности, производились на основе общих переменных либо с использованием оперативной памяти компьютера (разделяемые (shared) переменные в технологии распараллеливания OpenMP), либо на основе разделяемой памяти графического процессорного устройства (обмены между потоками внутри одной связки (block) с синхронизацией записи в общую переменную по функции synctreads() в программно-аппаратной платформе параллельных вычислений на графических процессорах CUDA или обращения к глобальной памяти (`_global_`) из локальной (`_local_`) внутри ядровых функций (`_kernel`) на платформе параллельных вычислений OpenCL).

В случае гибридизации программы с использованием графических процессорных устройств добавляется ряд ограничений, определяемых спецификой работы с GPU.

Сложность операций по обработке данных в исходной параллельной программе должна соответствовать цене перемещения данных на графическое устройство и обратно. Перенос данных для кратковременного использования небольшим числом потоков, не позволит получить выигрыш производительности, или этот выигрыш будет небольшим. В идеальном случае большое число потоков должно произвести большой объем обработки с достаточной весомой долей обменных операций [Федулов и др., 2021].

Для повышения эффективности выполнения программы данные следует сохранять на графическом устройстве как можно дольше. В силу указанного ограничения программы, использующие на графическом устройстве несколько вычислительных ядер, должны быть написаны с предпочтением сохранения данных на устройстве между вызовами ядер повторному копированию промежуточных результатов на хост и обратно.

Метод использует имеющуюся структуру гибридного вычислительного кластера, с установленным набором программ и технологий программирования, а также известными характеристиками вычислительных узлов, узла управления и коммуникационной сети, а также параллельную программу, разработанную с использованием технологии передачи сообщений MPI.

Предлагаемый метод организации выполнения параллельных программ на гибридных вычислительных кластерах состоит из следующих основных этапов.

*Этап 1.* Построение программно-аппаратной модели.

*Этап 2.* Определение значений аппаратных характеристик на основе построенной модели.

*Этап 3.* Вычисление значений показателей исходной параллельной программы на основе построенной модели.

*Этап 4.* Построение модуля интеллектуального анализа данных для определения необходимости проведения преобразования.

*Этап 5.* Преобразование исходной параллельной программы в гибридную параллельную форму.

*Этап 6.* Получение рекомендаций и выполнение гибридной формы параллельной программы на ГВК.

Результатом работы метода является набор рекомендаций по модификации исходного кода, с разметкой в виде комментариев в файле исходного кода также предлагаются наиболее подходящие параметры запуска полученной версии параллельной программы. В зависимости от настроек, заданных пользователем метода и имеющихся в распоряжении программно-аппаратных возможностей, получается гибридная программа, составленная с учетом совместного использования технологий параллельного программирования средствами центрального процессора (CPU) MPI и OpenMP, либо с использованием графических процессорных устройств (GPU) MPI и CUDA, MPI и OpenCL.

В случае достаточной сложности исходного алгоритма программы и доступности соответствующих технологий на узлах кластера, может быть предложен комбинированный вариант гибридной программы, где каждая отдельная часть общего алгоритма выполняется на своем узле с применением выбранной технологии (OpenMP/CUDA/OpenCL), затем полученные результаты агрегируются на управляющем узле кластера с использованием интерфейса передачи сообщений MPI.

Общая схема метода организации параллельных вычислений на гибридных вычислительных кластерах представлена на рис. 4.

В качестве реализации модуля интеллектуального анализа для последующего преобразования параллельной программы в гибридную форму и получения рекомендаций по модификации программы используется предложенные в работе [Борисов и др., 2016b] нечеткие «совместимые» когнитивные карты, позволяющие при анализе и моделировании проблемных ситуаций учитывать различную степень совместимости концептов при выборе операций для оценки непосредственного и опосредованного влияния концептов друг на друга.



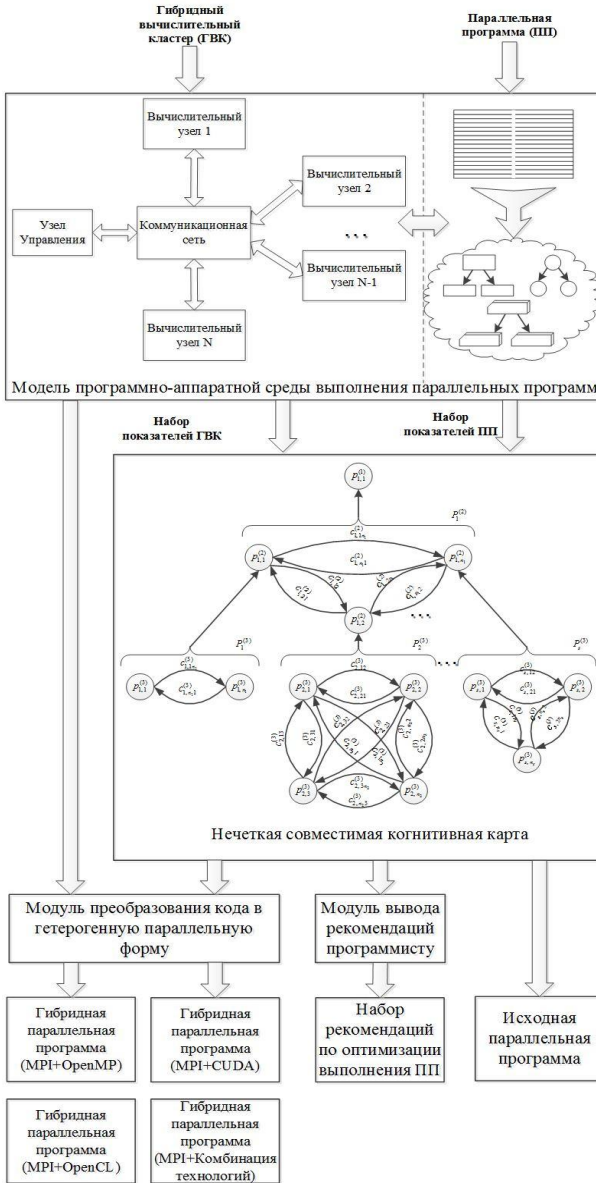


Рис. 4. Схема метода организации параллельных вычислений на ГВК

## Заключение

В работе предложена программно-аппаратная модель организации параллельных вычислений на гибридных вычислительных кластерах, ориентированная на создание программных средств преобразования параллельных программ в гибридную форму. Предлагаемая модель основывается на интегральных показателях, каждый из которых отражает значение отдельной характеристики рассматриваемой программно-аппаратной среды. Предложен метод организации выполнения параллельных программ на гибридных вычислительных кластерах с применением совместимых нечетких когнитивных карт, ориентированный на расширенные возможности предлагаемой модели программно-аппаратной среды с учетом различной аппаратных и программных показателей, позволяющий снизить долю обменных операций, выполняемых через медленные сетевые интерфейсы.

В дальнейшей работе планируется уточнение четвертого и пятого этапов предлагаемого метода. Планируется разработка способа преобразования исходной параллельной программы в гибридную параллельную форму на основе показателей модуля интеллектуального анализа и нечетких «совместимых» когнитивных карт.

## Список литературы

- [Борисов и др., 2012] Борисов В.В., Круглов В.В., Федулов А.С. Нечеткие модели и сети. – 2-е изд. стереотип. – М.: Горячая линия – Телеком, 2012.
- [Борисов и др., 2016a] Борисов В.В., Федулов А.С. «Совместимые» нечеткие когнитивные модели: прямые и обратные задачи // Нечеткие системы и мягкие вычисления. – 2016. – Т. 11, № 2. – С. 103-114.
- [Борисов и др., 2016b] Борисов В.В., Зернов М.М., Федулов А.С., Якушевский К.А. Исследование характеристик гибридного вычислительного кластера // Системы управления, связи и безопасности. – М., 2016. – № 4. – С. 129-146.
- [Ежова и др., 2019] Ежова Н.А., Соколинский Л. Б. Обзор моделей параллельных вычислений // Вестник ЮУрГУ. Серия «Вычислительная математика и информатика». – 2019. – Т. 8, № 3. – С. 58-91.
- [Федулов и др., 2021] Федулов А.С., Федулов Я.А., Федулова А.С. Совместное использование технологий OpenMP и MPI на узлах вычислительного кластера // Прикладная информатика. – 2021. – Т. 16, № 2. – С. 60-71..
- [Juan et al., 2019] Juan A., Juan C., Ravi R., Lastovetsky A. A Survey of Communication Performance Models for High-Performance Computing // ACM Comput. Surv. – 2019. – Vol. 51. – P. 36.
- [Sokolinsky, 2021] Sokolinsky L. BSF-skeleton: A Template for Parallelization of Iterative Numerical Algorithms on Cluster Computing Systems // Journal of Parallel and Distributed Computing. – 2021. – Vol. 149. – P. 193-206.
- [Toporkov et al., 2021] Toporkov V., Toporkova A., Yemelyanov D. Coordinated global and private job-flow scheduling in grid virtual organizations // Simulation Modelling Practice and Theory. Vol. 107, Netherlands, 2021.

---

Секция 6

**ИНСТРУМЕНТАЛЬНЫЕ  
СИСТЕМЫ, ФРЕЙМВОРКИ  
ДЛЯ СОЗДАНИЯ  
ИНТЕЛЛЕКТУАЛЬНЫХ  
СИСТЕМ. ТЕХНОЛОГИИ  
ПРОЕКТИРОВАНИЯ  
И РАЗРАБОТКИ  
ИНТЕЛЛЕКТУАЛЬНЫХ  
СИСТЕМ. ПРИКЛАДНЫЕ  
ИНТЕЛЛЕКТУАЛЬНЫЕ  
СИСТЕМЫ**

УДК 004.822

**ИНТЕЛЛЕКТУАЛЬНАЯ АППАРТНО-ПРОГРАММНАЯ  
ПОДДЕРЖКА И ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ  
СВАРОЧНЫХ ПРОЦЕССОВ**

С.В. Болотов (*s.v.bolotov@mail.ru*)

К.В. Захарченков (*zaharchenkovkv@mail.ru*)

С.К. Крутолевич (*s\_krutolevich@tut.by*)

Белорусско-Российский университет, Могилёв

Проблема повышения эффективности сварочных процессов на современных машиностроительных предприятиях неразрешима без применения аппаратно-программных комплексов, позволяющих отслеживать в реальном масштабе времени параметры работы каждого сварщика, однако, помимо отслеживания и журналирования та-

ких работ, крайне важным является рациональный подбор параметров сварочного процесса – фактически, необходимым компонентом таких комплексов становятся интеллектуальные помощники технологов, основанные, в том числе, и на нейросетевых технологиях. В данной статье рассматривается комплексный подход к аппаратно-программному сопровождению сварочных процессов, описывается разработанное аппаратное и программное обеспечение, а так же нейросетевой подход к подбору рациональных параметров сварочного процесса.

**Ключевые слова:** сварочные процессы, аппаратно-программные комплексы, нейронные сети, LSTM.

## Введение

В настоящее время на крупных машиностроительных предприятиях актуальной является проблема повышения эффективности управления сварочными процессами с использованием интеллектуальных технологий. Актуальность проблемы может быть количественно оценена низким значением коэффициента эффективности использования сварочного оборудования, что обусловлено:

- 1) необходимостью частого переключения режимов работы;
- 2) перемещениями сварщиков между швами для выполнения заданий;
- 3) отсутствием технологических инструкций, однозначно определяющих оптимальные параметры сварочных процессов для каждого прохода каждого сварного шва каждого изделия, узла, детали;
- 4) нарушением сварщиками режимов сварки швов, приводящим к появлению дефектов сварных соединений и в результате к появлению брака.

Нарушение сварщиками режимов с целью ускорения выполнения работ и выполнения большего объема работ в течение рабочего времени на предприятиях машиностроения встречается достаточно часто и является серьезной проблемой [Burlaka et al., 2019]. Отклонение температуры предварительного подогрева, силы тока и напряжения на дуге от нормативных значений приводит к образованию дефектов сварных соединений. Выход из строя узлов и деталей машин в процессе эксплуатации из-за разрушения сварных швов приводит к необходимости ремонта в «полевых» условиях, который не всегда возможен [Khmelev et al., 2018]. Результатом является выход машины из строя на длительный период времени, срывы планов и сроков работ, уплата штрафов и неустоек, потеря деловой репутации потребителей и завода-производителя. Кроме того, найти ответственного за сварку шва, вызвавшего выход из строя узла или детали, по истечении некоторого периода времени при отсутствии автоматизации также достаточно сложно. Решением данной проблемы являет-

ся создание эффективной системы регистрации сварочных процессов, обеспечивающей идентификацию каждого сварщика с использованием удостоверений с RFID-метками, автоматический интеллектуальный контроль выхода параметров сварочных процессов за пределы допустимых значений как со стороны сварщика в реальном времени, так и со стороны мастера и контролера путем интеллектуальной оценки результатов выполнения сварочных работ на основании паспортов сварных швов.

Отсутствие экспериментально подтвержденных исследований и методик создания технологических инструкций, определяющих оптимальные диапазоны параметров сварочных процессов, приводит к тому, что технологические инструкции создают технологи, которые не всегда являются экспертами в области решения подобных задач. В результате диапазоны допустимых значений параметров сварочных процессов устанавливаются достаточно широкими, и попадание значений параметров сварочных процессов в диапазоны, установленные технологической инструкцией, не гарантирует качество сварного соединения [Lin et al., 1997]. Некорректный выбор диапазонов допустимых значений параметров сварочных процессов приводит к систематическому появлению дефектов сварных соединений и брака. Одним из путей решения данной проблемы является создание методики интеллектуального анализа дефектов сварных соединений для выбора оптимальных диапазонов допустимых значений параметров сварочных процессов для каждого сварного шва каждого изделия, узла и детали.

Комплексная оценка эффективности работы оборудования осуществляется на основе коэффициента ОЕЕ (Overall Equipment Efficiency). Данный коэффициент учитывает доступность и производительность оборудования, качество продукции [Lindgren, 2011]. Комплексная оценка эффективности работы оборудования позволяет определить пути повышения количества времени выполнения сварочных работ за счет сокращения простоев, переналадок, устранения брака.

## 1. Аппаратная часть



Рис. 1. Регистратор сварочных процессов RSP-BPV-01

Регистратор сварочных процессов РСР-БРУ-01 включает три основных блока:

- основной блок обеспечивает идентификацию сварщиков с использованием RFID-меток, автоматическое получение с сервера, на котором развернута система, сменных заданий, закрепленных мастером за каждым сварщиком, выбор сварщиком заданий для исполнения, регистрацию и усреднение значений параметров сварки, полученных с блока датчиков, с последующей передачей через GSM связь или WI-FI на сервер для последующей обработки;
- блок датчиков обеспечивает регистрацию сварочного тока и напряжения на дуге с точностью  $\pm 1,5\%$  и передачу сигнала в основной блок регистратора. Экспериментально установлено, что для исследования процессов дуговой сварки плавящимся электродом оптимальной частотой опроса при измерениях регистратором сварочного тока и напряжения на дуге является 10-20 кГц. При этом сохранение данных о параметрах режима сварки (усреднение данных) следует проводить с частотой 10 Гц;
- пульт сварщика предназначен для дистанционного переключения сварочных швов (слоёв), звукового оповещения сварщика о выходе параметров режима за установленные пределы, контроля температуры предварительного подогрева.

## 2. Программная часть

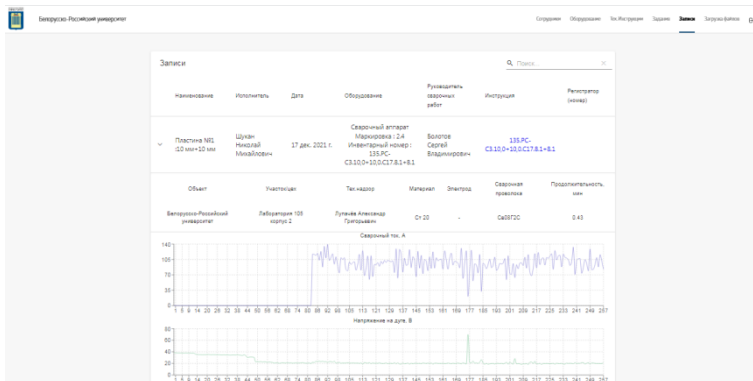


Рис. 2. Программный комплекс сопровождения сварочных процессов

Программная часть автоматизированной системы включает следующие основные модули:

- программный модуль автоматизации работы мастеров. Мастер распределяет задания сварщикам на выполнение каждого шва каждого изделия, узла, детали;
- программный модуль автоматизации работы технологов. Технолог определяет структуру производства (состав изделий, узлов, деталей) и задает допустимые диапазоны значений параметров сварочных процессов для каждого прохода (корневого, заполняющего, облицовочного) каждого изделия, узла и детали. Для технолога актуальными задачами являются: выбор рациональных значений диапазонов параметров сварочных процессов, обеспечивающих минимальную вероятность появления дефектов сварных соединений; рациональный выбор оборудования и материалов, обеспечивающих приемлемое качество соединения при минимальном расходе ресурсов (электроэнергии, газа, сварочной проволоки). Выбор данных значений может осуществляться как в ручном режиме, так и с помощью нейросетевого модуля;
- программный модуль автоматизации работы сварщика. Данный модуль обеспечивает идентификацию сварщика на регистраторе с помощью удостоверения с RFID-меткой, выбор сварщиками заданий на сварку, автоматический контроль параметров сварочного процесса;
- программный модуль автоматизации работы контролера. Контролеры вводят данные о количестве забракованных изделий по результатам визуального, рентгенографического и ультразвукового контроля. При выявлении брака и/или дефектов сварных швов контролеры фиксируют определяют причину брака или выявленных дефектов сварного соединения;
- программный модуль руководителей сварочных работ. На основании анализа паспортов сварных швов и отчёта об эффективности использования оборудования руководителем сварочных работ определяются причины снижения эффективности использования оборудования и разрабатываются способы устранения выявленных причин.

### **3. Методика повышения эффективности управления сварочными процессами**

Методика повышения эффективности управления сварочными процессами с использованием дистанционного контроля состоит из следующих этапов.

1. Ввод с помощью клиентского Web-приложения сменных заданий руководителем сварочных работ (мастером). Распределение заданий между сварщиками определяет количество рабочего времени,

которое каждый сварщик тратит на перемещения между изделиями, узлами, деталями, сварными швами, подлежащими сварке в соответствии со сменными заданиями. При изменении режимов сварки дополнительное рабочее время тратится на переключение режимов сварки. Эффективность распределения заданий между сварщиками определяется на основании отчета о работе оборудования, в котором указывается в минутах и процентах время нахождения сварочного оборудования в каждом из состояний: выключено, включено, сварка, вынужденный простой. Оптимизация распределения заданий между сварщиками обеспечивает сокращение времени простоев оборудования.

2. В системе осуществляется автоматический контроль сроков прохождения аттестации сварщиками. Для выполнения работ в соответствии с технологической инструкцией сварщик должен иметь соответствующую область распространения квалификации, при отсутствии которой сварщик не может быть допущен к выполнению работ. За месяц до истечения срока аттестации система автоматически выделяет цветовой индикацией сварщика, у которого заканчивается срок действия удостоверения об аттестации. Автоматический контроль сроков аттестации сварщиков и соответствия выполняемых работ квалификации обеспечивает повышение качества сварных швов и сокращает количество брака при выполнении сварочных работ.
3. Ввод данных о технологических процессах на каждое изделие в программном модуле технологов. Для каждого технологического процесса вводятся технологические инструкции, содержащие диапазоны допустимых значений контролируемых параметров (температуры предварительного подогрева, силы тока и напряжения на дуге) для каждого прохода (корневого, заполняющего, облицовочного) каждого сварного шва каждого изделия, узла и детали. На данном этапе задача технологов состоит в определении оптимальных и допустимых диапазонов параметров сварочного процесса, обеспечивающих минимизацию количества дефектов сварных соединений.
4. Рациональный выбор сварочного оборудования и материалов для выполнения заданий сварщиками. Выбранное оборудование должно содержать технические характеристики, обеспечивающие допустимые значения параметров сварочных процессов. Выбранные сварочные материалы должны обеспечивать минимизацию дефектов сварных соединений при минимальной стоимости и расходе. Для каждого используемого материала указывается номер партии и акт входного контроля.



5. Сварщиком на регистраторе подтверждает запуск процесса сварки. С использованием термопары, установленной на свариваемом изделии, регистрируется температура предварительного нагрева области сварки. Полученное значение сравнивается с указанным в технологической инструкции для соответствующего прохода сварного шва, соответствующего выбранному ранее заданию. В случае выхода температуры за пределы допустимых значений сварочный процесс не запускается.
6. В процессе сварки с блока датчиков на регистратор и пульт сварщика передаются параметры сварочных процессов (сварочный ток, напряжение на дуге). Переданные с блока датчиков параметры сравниваются с допустимыми значениями, указанными в технологической инструкции для соответствующего прохода.
7. Автоматический контроль простоев оборудования. Регистратором осуществляется постоянный мониторинг состояния сварочного оборудования.
8. Передача данных с регистратора на сервер с использованием GSM или WI-FI канала. В случае отсутствия соединения с сервером параметры сварочных процессов сохраняются в энергонезависимой памяти регистратора с последующей передачей на сервер после восстановления соединения.
9. Выставление оценки сварщику за выполнение каждого прохода (корневого, заполняющего, облицовочного) сварного шва по результатам анализа изменения значений сварочного тока и напряжения на дуге.
10. Формирование паспорта сварного шва для документального подтверждения качества каждого сварного соединения.
11. Визуальный и выборочный неразрушающий контроль (рентген, УЗК) сварных швов.
12. Формирование руководителями отчёта об эффективности использования оборудования.

#### **4. Использование нейросетевого подхода для выбора рациональных значений диапазонов параметров сварочных процессов**

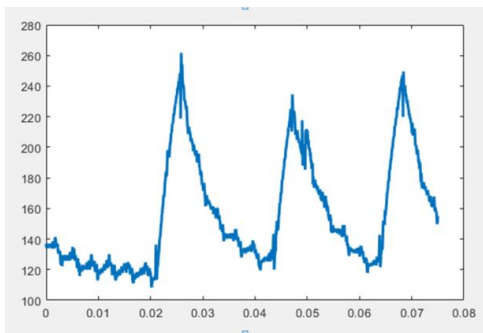
На каждом режиме сварочный процесс имеет длительность 150-200 циклов. При использовании классического подхода к проектированию нейронно-сетевой модели, количество входных параметров и, соответственно, размер входного слоя нейронной сети будет достаточно большим, что приведет к снижению производительности нейронной сети [Хайкин, 2006], [Tang et al., 2015].



*Рис. 3. Пример входного сигнала*

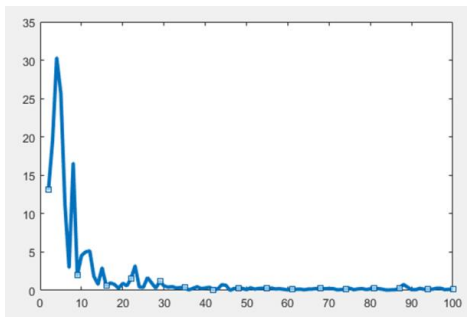
Таким образом, желательно уменьшить пространство входных признаков без потери информационной составляющей. Частотный анализ является одним из доступных инструментов, позволяющих выполнить поставленную задачу [Галушкин, 2000], [Мисник и др., 2019].

Для обработки сигнал был разделен на 70 отрезков по 3-4 цикла в каждом.



*Рис. 4. Пример отрезка цикла для преобразования*

Числовой ряд был подвергнут одномерному прямому преобразованию Фурье [Misnik et al., 2018]. График спектральной плотности сигнала представлен на рис. 5.



*Рис. 5. График спектральной плотности сигнала*

Из графика очевидно, что для обучения нейронной сети можно ограничиться диапазоном от 1 до 30 Гц [Misnik et al., 2019a].

Пример значений действительной и мнимой части для первых девяти гармоник представлен в табл. 1.

Таблица 1

Частота	Действительная часть	Мнимая часть
1	-4916,0	6748,8i
2	2205,2	12264,2i
3	18392,6	5836,4i
4	14399,6	7770,9i
5	4867,2	5221,6i
6	1738,1	803,82i
7	8823,6	5782,4i
8	1028,1	723,8i
9	2826,0	570,7i

Поскольку нейросетевому анализу подвергается процесс происходящий в течение определенного времени, каждый следующий шаг которого зависит от текущего шага, были выбраны сети долгой краткосрочной памяти (LSTM) [Misnik et al., 2019b].

Для обучения нейронной сети были созданы обучающие выборки с экспертной оценкой более чем 4 000 сварных швов.

В качестве входных параметров нейронной сети были использованы: тип материала, толщина материала, наличие или отсутствие предварительного прогрева, тип газовой среды и действительная и мнимая часть первых 30 гармоник, соответствующих частотам спектральной плотности сигнала от 1 до 30 Гц.

На выходе нейронной сети формировались рациональные значения напряжения и силы тока для сварочного процесса.

Метрикой качества была выбрана F1-мера. При обучении нейронной сети было достигнуто значение F1-меры в 0,89, что является достаточным для данной задачи. Улучшение данной метрики возможно путём увеличения обучающей выборки.

## Заключение

Решением проблемы повышения эффективности управления сварочными процессами является создание эффективной системы регистрации сварочных процессов, обеспечивающей идентификацию каждого сварщика, контроль параметров сварочных процессов и оценку результатов выполнения сварочных работ.

Нейросетевой подход на основе частотного анализа способен обеспечить подбор рациональных параметров сварочного процесса, обеспечивающих приемлемое качество соединения при минимальном расходе ресурсов.

Внедрение интеллектуальных технологий в управлении сварочными процессами на крупных машиностроительных предприятиях может повысить эффективность производства, уменьшить количество брака и снизить затраты на ремонт и обслуживание машин. Кроме того, это позволит улучшить деловую репутацию предприятия и повысить удовлетворенность потребителей продукции.

### Список литературы

- [Burlaka et al., 2019] Burlaka V., Lavrova E. Improving energy characteristics of the welding power sources for TIG-AC welding // Eastern-European Journal of Enterprise Technologies. – 2019. – Vol. 5, No. 5-101. – P. 38-43. – DOI: 10.15587/1729-4061.2019.180925. – EDN EFUOHE.
- [Lin et al., 1997] Lin Y.C., Lee K.H. Effect of welding parameters on the residual stress by the parallel heat welding // International Journal of Pressure Vessels and Piping. – 1997. – Vol. 71, No. 2. – P. 197-202. – EDN AITIVN.
- [Lindgren, 2011] Lindgren L.E. Understanding welding stress and distortion using computational welding mechanics // Minimization of Welding Distortion and Buckling: Modeling and Implementation, 2011. – P. 22-78. – DOI: 10.1533/9780857092908.1.22. – EDN PAMKIR.
- [Khmelev et al., 2018] Khmelev V.N., Slivin A.N., Abramov A.D. [et al.]. The ultrasonic device and the positioning system of the welding tool for welding of automobile bumpers // International Conference of Young Specialists on Micro/Nanotechnologies and Electron Devices, EDM: EDM 2018 - Proceedings, Erlagol, Altai, 29 июня – 03 2018 года. – Erlagol, Altai: IEEE Computer Society, 2018. – P. 318-321. – DOI: 10.1109/EDM.2018.8435068. – EDN YBXSbv.
- [Misnik et al., 2018] Misnik A., Krutalevich S., Prakapenka S. [et al.]. Neural network approximation precision change analysis on cryptocurrency price prediction // Fuzzy Technologies in the Industry - FTI 2018: Proceedings of the II International Scientific and Practical Conference, Ulyanovsk, 23–25 октября 2018 года. – Ulyanovsk: Ульяновский государственный технический университет, 2018. – P. 96-101. – EDN REYSPU.
- [Misnik et al., 2019a] Misnik A., Krutalevich S., Prakapenka S. [et al.]. Impact Analysis of Additional Input Parameters on Neural Network Cryptocurrency Price Prediction // Proceedings - 2019 21st International Conference "Complex Systems: Control and Modeling Problems", CSCMP 2019, Samara, 03–06 сентября 2019 года. – Samara: Institute of Electrical and Electronics Engineers Inc., 2019. – P. 163-167. – DOI: 10.1109/CSCMP45713.2019.8976486. – EDN VIKJCV.
- [Misnik et al., 2019b] Misnik A., Krutalevich S., Prakapenka S. [et al.]. Comparison of the predictions of convolutional neural networks with image arguments and long short-term memory neural networks with time-series arguments for cryptocurrency

- markets // CEUR Workshop Proceedings : IS 2019 - Proceedings of the 14th International Conference on Interactive Systems: Problems of Human-Computer Interaction, Ulyanovsk, 24–27 сентября 2019 года. Vol. 2475. – Ulyanovsk: Без издательства, 2019. – P. 214-222. – EDN RNCLVY.
- [Галушкин, 2000]** Галушкин А.И. Теория нейронных сетей. Кн. 1. – М.: ИПРЖР. 2000. – 416 с.
- [Мисник и др., 2019]** Мисник А.Е., Боровых П.А., Крутолевич С.К. [и др.]. Анализ изменения точности прогнозов нейронной сети, в зависимости от используемых входных параметров // Проблемы управления и моделирования в сложных системах: Труды XXI Международной конференции. В 2-х т. Т. II. Самара, 03–06 сентября 2019 года / под редакцией С.А. Никитова, Д.Е. Быкова, С.Ю. Боровика, Ю.Э. Плешивцевой. – Самара: Общество с ограниченной ответственностью "Офорт", 2019. – С. 293-296. – EDN YIGKIP.
- [Хайкин, 2006]** Хайкин С. Нейронные сети: полный курс. – 2-е изд., испр. – М.: Ж ООО «И.Д. Вильямс», 2006. – 1104 с.
- [Tang et al., 2015]** Tang, Jiliang; Chang, Shiyu; Aggarwal, Charu; Liu, Huan (2015). "Negative Link Prediction in Social Media" (PDF). In Proceedings of ACM International Conference on Web Search and Data Mining.

УДК 004 891.3

## ДЕТЕКТИРОВАНИЕ БИФУРКАЦИЙ МНОГОМЕРНЫХ СЛУЧАЙНЫХ ПРОЦЕССОВ В ДИНАМИЧЕСКИХ ЭКСПЕРТНЫХ СИСТЕМАХ

А.Н. Ветров (*vetrov\_48@mail.ru*)

А.А. Мальков (*kja@list.ru*)

Б.В. Палюх (*pboris@tstu.tver.ru*)

Тверской государственный технический университет, Тверь

В работе описан метод и процедура определения точки бифуркации непрерывного технологического процесса, в которой наиболее вероятен переход системы в критическое (паредаварийное) состояние. Кратко описан метод двойного скользящего окна, позволяющий выявлять аномалии в непрерывном многомерном потоке информации, формируемой сенсорами технологического оборудования. Приводится описание программно-технической платформы для анализа потоковых данных.

**Ключевые слова:** прикладные системы искусственного интеллекта, динамические экспертные системы, диагностика, точка бифуркации.

### Введение

Мы рассматриваем непрерывные многостадийные процессы. К ним относятся процессы химической промышленности, нефтепереработки и нефтехимии, фармакологическое производство, атомные электростанции, металлургия и т.д. Одной из важных задач управления такими процессами является обеспечение их стабильности. Для этого производится непрерывный мониторинг состояния технологического процесса путем периодического измерения диагностических переменных. При нормальной работе оборудования технологический процесс функционирует в стационарном режиме. Имеются слабые колебания значений диагностических показателей, которые укладываются в нормативные границы. При возникновении неисправности в работе оборудования, технологический процесс переходит в нестационарный режим, и он становится неуправляемым. Время обнаружения аномального поведения технологического процесса существенно влияет на способность системы к восстановлению.

Определение точек бифуркации в эволюционном управлении непрерывными многостадийными технологическими процессами, подверженными случайным флуктуациям, является основой для обеспечения их стабильности и безопасности.

Проблема обнаружения аномалий в сложных динамических системах является развивающейся областью научных исследований и активно обсуждается в литературе [Malhotra et al., 2016], [Yangdong et al., 2019], [Dohyung et al., 2018], [Zhao et al., 2020]. В проведенных ранее исследованиях в области обнаружения и диагностики аномальных событий подчеркивается, что вопросы раннего предупреждения об опасной ситуации являются очень актуальными. Традиционно обнаружение аномалий относится к проблеме выявления паттернов в данных, полученных на основе мониторинга деятельности системы, которые не соответствуют ожидаемому поведению. Эти несоответствующие паттерны часто называют аномальными. Системы обнаружения аномалий во временных рядах используются в самых разных приложениях [Dan et al., 2019], [Dohyung et al., 2018], [Lin et al., 2020], [Malhotra, 2016], [Su et al., 2019], [Yangdong et al., 2019], [Zekai et al., 2021], [Zhao et al., 2020]. Проблема поиска аномалий интерпретируется как проблема классификации. Для ее решения используются методы классификации, такие как метод  $k$ -ближайшего соседа, кластерный анализ, классификаторы, построенные с использованием искусственных нейронных сетей, метод случайного леса и тому подобное. Хотя существующие системы поиска аномалий во временных рядах доказали свою эффективность, существует ряд проблем, связанных с их практическим применением. Эти проблемы включают высокий уровень ложных срабатываний, ограниченные типы аномалий, которые система может обнаружить. Кроме того, такие системы не могут выполнять обнаружение в реальном времени. Наиболее существенным недостатком, с точки зрения определения бифуркаций в непрерывных многостадийных технологических процессах, является необходимость предварительного обучения рассмотренных систем поиска аномалий. Поскольку аварийные ситуации относятся к классу редких событий, формирование базы данных для обучения проблематично.

Для решения задачи управления непрерывными многостадийными технологическими процессами, авторами был разработан прототип динамической экспертной системы (ЭС) ExpSystem, совмещающей методы экспертного и количественного анализа состояния технологического процесса в условиях неопределенности. ЭС ExpSystem позволяет не только обнаруживать и диагностировать переход технологического процесса в нестационарное состояние, но и определять причину возникновения аномалии, находить неисправное оборудование, являющееся источником

критической ситуации. Более подробно она описана в работах [Shabanov et al., 2019], [Palyukh et al., 2021]. Важную роль в системе играет процедура определения момента перехода технологического процесса из штатного в критический (предварийный) режим. В данной статье описан метод и процедура определения точки бифуркации в технологическом процессе, при которой наиболее вероятен переход системы в критическое состояние. Достоинствами предложенного метода является возможность обнаружения изменения состояния технологического процесса в режиме реального времени, а также отсутствие необходимости предварительного обучения системы поиска аномалий.

## 2. Проблема определения точки бифуркации

Будем рассматривать непрерывный многостадийный технологический процесс как сложную динамическую систему, состояние которой определяется вектором  $Y(t) \in R^n$ . Совокупность всех возможных состояний системы  $Y(t)$  образует фазовое пространство. Эволюция системы соответствует движению изображающей точки в фазовом пространстве вдоль траектории, определяемой семейством эволюционных операторов. Динамическая система может быть в двух состояниях: стационарном и не стационарном. Стационарность динамической системы предполагает неизменность ее структуры и параметров. Под влиянием возмущающих воздействий, изменяющих её состояние, структуру или параметры, система может перейти в нестационарное состояние. Переходный процесс определяет новое установившееся стационарное состояние системы, которое не зависит от начального. Бифуркация – это такой вариант развития ситуации, когда система переходит от устойчивости к хаосу. Таким образом задача поиска аномалий непрерывного многостадийного технологического процесса сводится к обнаружению бифуркации процесса и формирования флуктуаций, определяющих новую траекторию (аттрактор). Для решения задачи обнаружения точки бифуркации в непрерывном многостадийном производственном процессе используются данные объективного контроля, получаемые от сенсоров технологического оборудования. Контроль состояния технологического процесса осуществляется на основе наблюдаемых значений диагностических переменных. Наблюдения осуществляются через равные промежутки времени  $\Delta t$  (или с частотой  $f = 1/\Delta t$ ). Множество наблюдаемых за период времени  $T$  значений образуют многомерный ( $D$ -мерный) временной ряд. Этот временной ряд является реализацией многомерного случайного процесса. Измеренные значения технологических параметров отражают динамику его состояния. Система измерения включает  $D$  сенсоров технологического оборудования. Для любого senso-



ра  $d = 1, 2, \dots, D$  временной ряд  $y_1^{(d)}, y_2^{(d)}, \dots, y_t^{(d)}$  представляет собой множество значений технологического параметра  $y_t^{(d)}$  измеренных в момент времени  $t$ . На значения этих параметров накладываются технологические ограничения в виде верхних  $y_u^{(d)}$  и нижних  $y_l^{(d)}$  границ. Эти нормативные границы параметров определяют область работоспособного состояния технологической системы. Выход за пределы нормативных границ означает переход системы в критическое состояние. Если технологический процесс протекает в штатном режиме, значения диагностической переменной колеблются в заданных пределах. Можно считать, что случайный процесс, который описывается набором значений диагностической переменной, в этом случае является стационарным. Выход диагностической переменной за пределы нормативных границ указывает на возможную неисправность технологического оборудования. Однако даже при нормально работающем оборудовании могут иногда возникать кратковременные случаи, когда значения диагностической переменной выходят за нормативные границы. Переход технологического процесса из штатного режима в критическое состояние возможен при наличии определенной устойчивой тенденции движения значений диагностических переменных за пределы нормативных границ. Следовательно, точка бифуркации определяется моментом перехода случайного процесса из стационарного в нестационарное состояние. Слабую стационарность случайного процесса можно проверить на основе того факта, что среднее значение

$$\mu_y(t_1) = \lim \frac{1}{T} \sum_{k=1}^T y_k(t_1)$$

и автокорреляционная функция

$$R_y(t_1, t_1 + \tau) = \lim \frac{1}{T} \sum_{k=1}^T y_k(t_1) y_k(t_1 + \tau)$$

не зависят от времени  $t_1$ . Эти условия стационарности лежат в основе предлагаемого метода двойного скользящего окна для поиска точки бифуркации в непрерывных многостадийных технологических процессах.

### 3. Метод двойного скользящего окна

Работа экспертной системы ЭС ExpSystem основана на анализе многомерного потока данных, поступающих от сенсоров технологического оборудования. Для определения точки бифуркации предлагается метод двойного скользящего окна, суть которого заключается в проверке условий стационарности на основе выборочных данных для коротких временных

рядов. При этом мы исходим из того, что измеряемые значения диагностических переменных образуют концептуально бесконечный непрерывный поток данных  $S$ . Под потоком данных  $S$  здесь понимается [Tangwongsan et al., 2015], [Traub et al., 2019] последовательность записей, поступающих в систему непрерывной обработки запросов. Записи из  $S$  помечаются временем прибытия с использованием системных часов.

Рассмотрим поток  $S$  с последовательностью записей  $s_1, s_2, \dots$  и соответствующими временными метками  $\tau_1, \tau_2, \dots$ . Для анализа непрерывных потоков данных в режиме реального времени применяется потоковая обработка, включающая агрегацию информации с использованием скользящих окон. Для любых двух моментов времени  $T_L > T_R$ , обозначим  $S[T_L, T_R]$  скользящее окно на основе временных меток в  $S$ . Поскольку информация от сенсоров поступает в дискретные моменты времени, целесообразно использовать скользящее окно на основе кортежей над  $S - S_H[H_L, H_R]$ , где  $H_L > H_R$  любые два неотрицательных целых числа. Аналогично скользящему окну на основе временных меток в  $S$ ,  $S_H[H_L, H_R]$  содержит набор кортежей  $\{y_i: \{(H - H_L + 1) \leq i \leq (H + H_R)\}$  когда текущая длина  $S$  равна  $H$ . В этом случае кортежи нумеруются целыми числами.

Для обработки запросов потоковых приложений используется операция агрегации кортежей в скользящем окне. В каждый момент времени текущая длина  $S_H$  определяется количеством поступивших кортежей. Данные в бесконечном потоке фиксируются скользящим окном и агрегируются по мере того, как элементы данных вставляются и удаляются из скользящего окна.

Обозначим  $O_1, \dots, O_n$  набор операторов агрегации. Для входного потока  $S$   $O_i = f_A(S[T_{L_i}, T_{R_i}])$ , где  $f$  – функция агрегации, а  $A$  – атрибут агрегации. Каждое значение атрибута агрегации  $a_i$  представляет собой элемент потока. В процессе обработки потоковых данных соответствующие значения элементов потока записываются в циклический буфер и обрабатываются операторами агрегации  $O_i$ , а результат агрегации сохраняется в памяти.

Рассматриваемый метод двойного скользящего окна предполагает использование системы агрегации, состоящей из двух скользящих окон. Первое скользящее окно  $W_1$  используется как низкочастотный фильтр, подавляющий кратковременные случайные выбросы значений диагностических переменных. Второе скользящее окно  $W_2$  аккумулирует информацию для анализа состояния технологического процесса. Скользящее окно  $W_1$  делится на  $N$  равных интервалов  $I_k$  ( $k = 1, \dots, N$ ), так что  $\bigcup I_k = S$ ,  $\bigcap I_k = \emptyset$ . Каждый из них имеет длину равную  $h$ . Длина интервала  $h = H/N$  кратна длине окна  $W_1$ . Для любых двух целых чисел  $l$  и  $r$  ( $l \leq r$ ), обозначим  $I[l, r]$  интервал, который содержит набор кортежей в позициях  $l, l+1, \dots, r$  в  $S$ . Элементы  $a_l, \dots, a_r$  принадлежат интервалу  $I \subset S$ . Для ин-

тервала  $I f_A(I)$  обозначает агрегацию по элементам из  $S$ , принадлежащим  $I$ . Скользящее окно  $W_2$  содержит значения  $N$  агрегатов, вычисленных в окне  $W_1$ . Обозначим его  $S^{(2)}[T_L^{(2)}, T_R^{(2)}]$ . Длина его равна  $N$ . Текущее состояние скользящего окна  $W_1$  определяется набором элементов потока, хранящихся в циклическом буфере окна в данный момент. Аналогично текущее состояние скользящего окна  $W_2$  определяется набором агрегированных значений, вычисленных в окне  $W_1$ .

Алгоритм обработки потоковых данных содержит следующие шаги.

1. Считываем текущее состояние скользящего окна  $W_1$  из циклического буфера.
2. Вычисляем  $f_A(I_k)$  для всех элементов потока, соответствующих текущему состоянию скользящего окна  $W_1$ .
3. Добавляем все вычисленные значения  $f_A(I_k)$  в окно  $W_2$ . Обозначим эти значения  $X_1, X_2, \dots, X_N$ . Они определяют текущее состояние скользящего окна  $W_2$ .
4. Считываем текущее состояние скользящего окна  $W_2$  из циклического буфера окна.
5. Вычисляем агрегатное значение для элементов потока, соответствующих текущему состоянию скользящего окна  $W_2$ .
6. Если полученное значение соответствует условию поиска, логической переменной присваивается значение 1, в противном случае 0.
7. Удаляем значения “самого старого” интервала из окна  $W_1$ .
8. Добавляем значения следующего интервала в окно  $W_1$ .
9. Вычисляем значения  $f_A(I_k)$  для последнего добавленного интервала.
10. Добавляем последнее вычисленное значения  $f_A(I_k)$  в окно  $W_2$ .
11. Удаляем значения “самого старого” значения агрегата из окна  $W_2$ .
12. Возвращаемся на шаг 5.

Более подробно процедура агрегации потоковых данных описана в статье [Palyukh, 2021].

#### **4. Программно-техническая платформа для анализа потоковых данных**

В прототип динамической экспертной системы включены алгоритмы обнаружения аномалий многомерного случайного процесса и определения точки бифуркации в соответствии с методом двойного скользящего окна. Система может работать как с сенсорами технологического оборудования, так и в режиме имитации. Для работы в режиме имитации был разработан испытательный стенд на основе программируемых внешних интерфейсных плат (ВИП) для чтения и передачи показания сенсоров с низкоуровневым интерфейсом GPIO для считывания и передачи данных сенсоров.

Он позволяет считывать потоки данных параллельно для каждого сенсора для передачи их на сервер с помощью механизмов потоковой параллельной и распределенной обработки [Dagit et al., 2016], [Hirzel et al., 2016]. Сами потоковые данные моделируются генератором данных на основе марковского случайного процесса, который строится на основе модели авторегрессии порядка  $p$ :

$$\Phi(B)y_t = \varepsilon_t.$$

где  $\Phi(B)$  оператор авторегрессии,

$$\Phi(B) = (1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_p B^p).$$

а  $B$  – оператор сдвига назад.

Если все корни характеристического уравнения  $\Phi(B) = 0$  по модулю больше единицы и различны, то процесс является стационарным. В противном случае процесс становится нестационарным. Это свойство используется для имитации различных режимов работы технологического оборудования. Задавая различные значения коэффициентов авторегрессии можно менять режим случайного процесса от стационарного до нестационарного. Испытательный стенд является частью динамической экспертной системы и служит для проверки, отладки и настройки блока обнаружения аномалий многомерного случайного процесса и определения точки бифуркации.

Для управления испытательным стендом в состав экспертной системы включено приложение для моделирования работы ВИП.

Особенность указанного приложения заключается в параллельной работе генераторов данных на одном устройстве ВИП с последующей их передачей на главный сервер. Кроме того, предусмотрена возможность управления процессом генерации/передачи данных на платах со стороны клиента ЭС.

Поскольку на реальном производстве количество сенсоров достаточно большое, а их отключение/подключение возможно в режиме реального времени, то предусмотрена возможность моделировать эти условия на некотором количестве интерфейсных плат, подключаемых/отключаемых к главному серверу в любой момент времени. Сервер реализован в виде службы операционной системы, работает постоянно и может управлять работой программно-технической платформой в режиме реального времени, поддерживать все указанные соединения и информировать клиента экспертной системы обо всех таких подключениях/отключениях.

В состав сервера включен программный модуль, отвечающий за запуск/остановку генераторов потоковых данных на интерфейсных платах, а также за управление режимом моделирования online путем настройки различных параметров. Этот модуль получает команды управления режимом работы конкретной ВИП от клиента ЭС.

Клиент ЭС представляет из себя клиентское приложение, организованное в виде простой консоли, куда выводятся модельные данные (рис. 1).

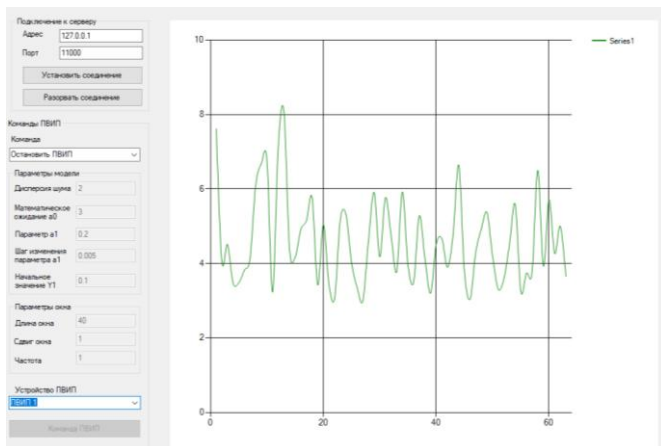


Рис. 1. Окно управления испытательным стендом

На графике рис. 1 по оси абсцисс откладывается время, а по оси ординат – значение технологической переменной. Значения этих параметров задаются при настройке приложения.

Возможности этого приложения следующие:

- подключение/отключение к серверу;
- получение от сервера информации о количестве подключенных интерфейсных плат с их идентификацией, что дает возможность управления каждой интерфейсной платой в отдельности;
- формирование набора команд для управления интерфейсными платами посредством сервера:
- параметры моделирования случайного процесса ;
- параметры временного окна (длина, шаг сдвига окна);
- частота генерации данных;
- получение данных от интерфейсных плат через сервер;
- визуализация результатов генерации данных на интерфейсных платах в виде графиков.

Эти возможности отображены на рис. 1. Приложение позволяет не только моделировать случайный процесс, но и настраивать характеристики (параметры) метода двойного скользящего окна, добиваясь максимальной эффективности его работы. Настройка может производиться как в тестовом режиме, так и на работающем технологическом оборудовании.

## Заключение

В настоящее время, в рамках фундаментальной проблемы обеспечения устойчивости системы управления эволюцией непрерывных многостадийных процессов к возникающим критическим ситуациям, недостаточно проработаны методы раннего обнаружения момента перехода технологического процесса из штатного режима в предаварийное состояние. В статье представлен новый подход для решения этой актуальной задачи, основанный на использовании метода двойного скользящего окна для выявления начала переходного процесса в системе управления. Достоинство этого подхода заключается в том, что в отличие от существующих разработок предлагаемая система детектирования бифуркаций не требует предварительного обучения. Обнаружение аномалий в технологическом процессе происходит online на основе анализа бесконечного многомерного потока технологических данных, получаемых от сенсоров оборудования. Кроме того, в системе предусмотрена возможность тестирования и настройки параметров метода двойного скользящего окна на конкретный объект. Для этой цели разработан испытательный стенд, имитирующий работу сенсоров технологического оборудования и генерирующий многомерные временные ряды, в которых переключение с одного состояния на другое является вероятностным процессом. Проведенные испытания подтвердили эффективность предложенной системы детектирования бифуркаций.

Полученные результаты могут служить методологической основой для разработки динамических экспертных систем, работающих в режиме реального времени в условиях флуктуации технологического процесса.

## Список литературы

- [Dagit et al., 2016] Dagit D., Evans B. Testing the Mechanisms of Streaming International Parallel and Distributed Processing Symposium Workshops (IPDPSW) // IEEE. – 2016. – P. 1789-1792. – DOI: 10.1109 / IPDPSW. 2016. 138.
- [Dan et al., 2019] Dan Li, Dacheng Chen, Baihong Jin, Lei Shi, Jonathan Goh, and See-Kiong Ng. MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks // In Artificial Neural Networks and Machine Learning: Text and Time Series. – 2019. – P. 703-716.
- [Dohyung et al., 2018] Dohyung Kim, Hyochang Yang, Minki Chung, Sungzoon Cho, Huijung Kim, Minhee Kim, Kyungwon Kim, and Eunseok Kim. Squeezed convolutional variational autoencoder for unsupervised anomaly detection in edge device industrial internet of things // In International Conference on Information and Computer Technologies. – 2018. – P. 67-71.
- [Hirzel et al., 2016] Hirzel Martin, Rabbah Rodric, Suter Philippe, Tardieu Olivier, and Vaziri Mandana. Spreadsheets for Stream Processing with Unbounded Windows and Partitions // In Conference on Distributed Event-Based Systems (DEBS). – 2016. – P. 49-60.

- [**Lin et al., 2020**] Lin Shuyu, Clark Ronald, Birke Robert, Schonborn Sandro, Trigoni Niki, Roberts Stephen. Anomaly detection for time series using vae-lstm hybrid model // ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – DOI: 10.1109/ICASSP40776.2020.9053558.
- [**Malhotra, 2016**] Malhotra P., Ramakrishnan A. LSTM-based encoder-decoder for multi-sensor anomaly detection. Available: arXiv:1607.00148 [cs.AI]. – 2016.
- [**Palyukh et al., 2021**] Palyukh B., Merkurjev S., Vetrov A., Shabanov B. and Sotnikov A. Methods for Forecasting Critical States of the Technological Process in the Evolutionary Management of Continuous Multi-Stage Production // 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus). – 2021. – DOI: 10.1109/ElConRus51938.2021.9396111.
- [**Shabanov et al., 2019**] Shabanov B., Sotnikov A., Palyukh B., Vetrov A., Alexandrova D. Expert System for Managing Policy of Technological Security in Uncertainty Conditions: Architectural, Algorithmic, and Computing Aspects // Proceedings of the 2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus) January 28-30, 2019.
- [**Su et al., 2019**] Su Y., Zhao Y., Niu C., Liu R., Sun W., and Pei D. Robust anomaly detection for multivariate time series through stochastic recurrent neural network // In ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. – 2019. – P. 2828-2837.
- [**Tangwongsan et al., 2015**] Tangwongsan Kanat, Hirzel Martin, Schneider Scott, and Wu Kun-Lung // General Incremental Sliding-Window Aggregation. In Conference on Very Large Data Bases (VLDB). – 2015. – P. 702-713.
- [**Traub et al., 2019**] Traub Jonas, Grulich Philipp, Cuéllar Alejandro, Breß Sebastian, Katsifodimos Asterios, Rabl Tilmann, Markl Volker. Efficient Window Aggregation with General Stream Slicing // Published in Proceedings of the 22nd International Conference on Extending Database Technology (EDBT), March 26-29, 2019.
- [**Yangdong et al., 2019**] Yangdong He and Jiabao Zhao. Temporal convolutional networks for anomaly detection in time series // Journal of Physics: Conference Series. – 2019. – 1213 (4).
- [**Zekai et al., 2021**] Zekai Chen, Dingshuo Chen, Zixuan Yuan, Xiuzhen Cheng, and Xiao Zhang. Learning graph structures with transformer for multivariate time series anomaly detection in IoT // ArXiv, abs/2104.03466. – 2021.
- [**Zhao et al., 2020**] Zhao H., Wang Y., Duan J., Huang C., Cao D., Tong Y., Xu B., Bai J., Tong J., and Zhang Q. Multivariate time-series anomaly detection via graph attention network // In 20th IEEE International Conference on Data Mining. – 2020. – P. 841-850.

## СОЗДАНИЕ ВИРТУАЛЬНЫХ ПОМОЩНИКОВ НА ОСНОВЕ ТРАНСФОРМАЦИИ МОДЕЛЕЙ\*

Н.О. Дородных (*tualatin32@mail.ru*)

О.А. Николайчук (*nikoly@icc.ru*)

А.Б. Столбов (*stolboff@icc.ru*)

А.Ю. Юрин (*iskander@icc.ru*)

Институт динамики систем и теории управления  
им. В.М. Матросова СО РАН, Иркутск

В работе описывается оригинальный подход, использующий принципы модельно-ориентированной разработки и трансформаций, для создания прототипов виртуальных помощников. Представлены модели всех уровней абстракции (конкретизации), согласно методологии MDA (Model-Driven Architecture). Показаны фрагменты трансформации элементов моделей. Приведено описание основных этапов предложенного подхода в контексте решения задачи диагностики системы электроснабжения воздушного судна.

**Ключевые слова:** онтология задачи, онтология диалога, модельно-ориентированная разработка, трансформация моделей, виртуальный помощник.

### Введение

Целью работы является разработка элементов технологии создания виртуальных помощников для решения производственных проблемно-ориентированных задач на основе трансформации концептуальных моделей.

Под производственными задачами авторы понимают задачи, решаемые при создании и эксплуатации технических объектов. Особенностью данных задач является то, что данные и знания, необходимые для их решения, во-первых, содержатся в ограниченном пространстве технической документации, большей частью слабоструктурированной; во-вторых, ошибки в решениях таких задач не желательны или не допустимы; в-третьих, применение экспертных знаний может значительно повысить эффективность их решения. Последний фактор обуславливает необходимость непосредственного участия экспертов в процессе создания таких систем, что также связано с различными проблемами извлечения знаний.

---

\* Работа выполнена при финансовой поддержке РФФ (проект № 22-21-00099).



Подход к созданию программных систем, основанный на трансформации моделей, позволяет снизить требования к уровню подготовки разработчиков в области навыков программирования, ускорить процесс разработки и уменьшить количество ошибок, связанных как с кодированием, так и системным анализом задачи.

Для создания виртуальных помощников предлагается развить авторскую специализацию модельно-ориентированного подхода (Model-Driven Architecture, MDA) [Da Silva, 2015], [Pérez-Soler et al., 2020], ориентированную на создание производственных экспертных систем [Yurin et al., 2018], [Дородных и др., 2019].

Данная специализация определяет процесс создания как последовательность этапов, включающих построение определенных моделей и их трансформации (рис. 1):

- Этап создания вычислительно-независимых моделей (CIM), обеспечивающих формулировку ключевых абстракций решаемой задачи, включая: общую (типовую) модель интеллектуального помощника, онтологию задач, онтологию предметной области и онтологию диалога.
- Этап создания проблемно-ориентированных платформенезависимых моделей (PIM), учитывающих специфику решаемой задачи, включая базу знаний интеллектуального помощника и модель проблемно-ориентированного диалога.
- Этап создания платформозависимых моделей (PSM), ориентированных на определенную программную платформу функционирования интеллектуального помощника.
- Этап генерации кодов и спецификаций интеллектуального помощника для определенной платформы.

Далее опишем модели и их трансформации, а также примеры применения предложенного подхода.

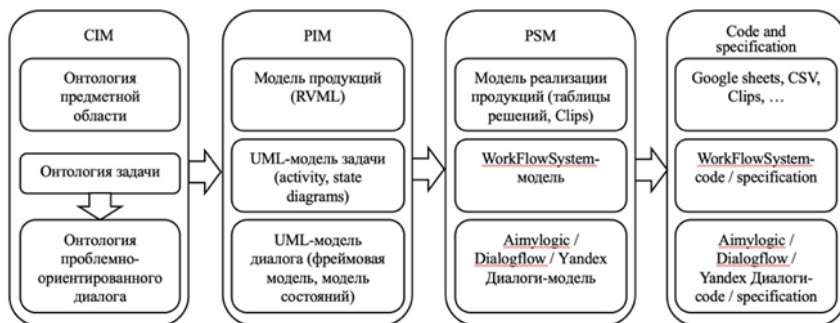


Рис. 1. Этапы подхода MDA для создания интеллектуальных помощников

## 1. Онтологические модели

Виртуальный помощник (ВП) – это набор контекстно-ориентированных сервисов, которые реагируют на команды на естественном языке. Наиболее распространенными задачами, выполняемыми ВП, являются: интерпретация команд, поддержка интерактивного диалога и выполнение задач, в том числе от имени пользователя. Типичный алгоритм принятия решения виртуальным помощником включает следующие шаги: прием и распознавание голосового ввода; получение текстового представления голосового ввода; выполнение обработки текста на естественном языке; определение намерений пользователя; определение решаемой задачи, ее параметров и этапов решения; выполнение этапов решения задачи; генерирование диалогового ответа; представление результата пользователю.

Для реализации приведенного выше общего алгоритма современные ВП должны включать: компоненты обработки естественного языка, обеспечивающие распознавание и генерацию речи (голоса и текста); коммуникационные компоненты для взаимодействия с пользовательскими устройствами и внешними сервисами; компоненты, относящиеся к системе управления диалогами; компоненты для решения задач и принятия решений.

Первые два компонента предлагается реализовать сторонними системами и использовать эти компоненты при создании платформозависимых моделей, последние компоненты опишем более детально в виде моделей предметной области, задач и диалога.

Модель предметной области имеет классический вид онтологии, описывающей понятия и отношения между ними [Berman et al., 2018].

Существует несколько общих моделей системы управления диалогами. Следует упомянуть международный стандарт ISO 24617-2 [ISO, 2012] для сематических аннотаций диалоговых актов и современных OWL-онтологий: VOnDA [Kiefer et al., 2021] – фреймворк для онтологического диалога и TODA для задачно-ориентированных диалоговых систем [Aceta et al., 2021]. В описываемом подходе предлагается использовать основные элементы диалога перечисленных моделей (рис. 2).

Для апробирования подхода использовано программное средство Aimylogic [Aimylogic, 2023], для него создана соответствующая модель (рис. 3).

Модель задачи позволяет описать структуру типа «задача – подзадача» с указанием используемых понятий онтологии предметной области, методов решения и программного обеспечения, реализующего эти методы (рис. 4) [Berman et al., 2018].

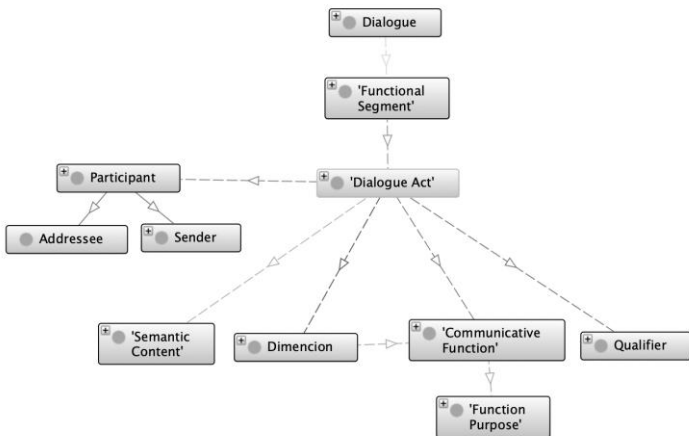


Рис. 2. Фрагмент модели диалога (скриншот программной системы Protege) [Berman et al., 2018]

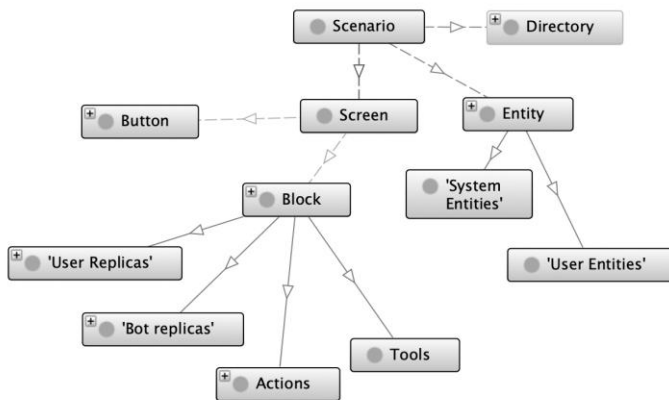


Рис. 3. Фрагмент онтологии Aimylogic tool [Dorodnykh et al., 2022].

На основе моделей диалога и задачи создается проблемно-ориентированная модель диалога для решения данной задачи. Причем, представленные модели являются метамоделями или шаблонами для создания моделей следующего уровня детализации.



Рис. 4. Фрагмент модели задачи (скриншот программной системы Protege)

Например, рассмотрим одну из задач технического диагностирования – поиск и устранение неисправностей технического объекта. Модель данной задачи опишем в различных аспектах: свойства задачи (рис. 5), структура подзадач (отношение «часть-целое») (рис. 6) и последовательность их выполнения (отношение «последовательность») между стадиями алгоритма (рис. 7).



Рис. 5. Фрагмент модели свойств задачи поиска и устранения неисправности (скриншот программной системы Protege)

Основываясь на данных моделях, получаем модель проблемно-ориентированного диалога, где алгоритм решения каждой задачи дополняется стадиями получения и представления данных и знаний, которые реализуются в виде диалога. На рисунке представлен фрагмент модели для одной из проблемно-ориентированных стадий алгоритма (рис. 8).



Рис. 6. Фрагмент модели структуры задачи поиска и устранения неисправности (скриншот программной системы Protege)

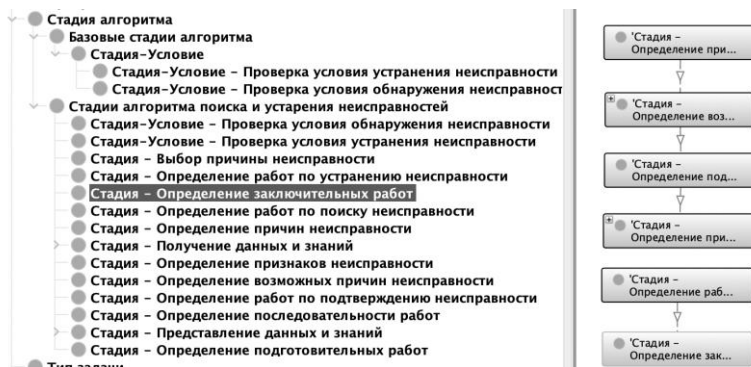


Рис. 7. Фрагмент модели стадий алгоритма решения задачи поиска и устранения неисправности (скриншот программной системы Protege)

## 2. Трансформации

Несколько слов о трансформации моделей. Последовательность данных трансформаций представим в виде модели:

$$M_T \times M_{Domain} \times M_D \rightarrow M_{D_T} \rightarrow M_{Aimylogic}$$

где  $M_T$  – модель задачи,  $M_{Domain}$  – модель предметной области,  $M_D$  – модель диалога,  $M_{D_T}$  – модель проблемно-ориентированного диалога,  $M_{Aimylogic}$  – модель диалога для программного средства, реализующего диалог, в частности, Aimylogic.



Рис. 8. Фрагмент модели проблемно-ориентированного диалога (скриншот программной системы Protege)

В таблице представлены фрагменты трансформации элементов представленных моделей (табл. 1).

Таблица 1

### Примеры трансформаций элементов моделей

Модель задачи ( $M_T$ )	Модель диалога ( $M_D$ )	Модель диалога помощника ( $M_{D_T}$ )	Модель инструментального средства ( $M_{Aimylogic}$ )
	Functional segment	Приветствие помощника	Screen – Приветствие помощника (Блок типа Bot Replicas)
	Functional segment	Представление помощника	Screen – Представление помощника (Блок типа Bot Replicas)

Подзадача – Определение последовательности работ	Functional segment		
Получение входных данных и знаний задачи: признаки неисправности, работа и ее свойства.	Functional segment (итерация сегмента в зависимости от количества входных данных и знаний)	Dialogue act (Sender – Assistant, Addressee - User, Dimension - Auto-Feedback, Communicative Function – AutoPositive, Purpose - Inform)	Screen – Информирование о необходимости выбора/ввода входных данных и знаний задачи (Блок типа Bot replicas)
		Dialogue act (Sender – Assistant, Addressee - User, Dimension - Time Management, Communicative Function - Pausing)	Screen – Ответ пользователя (Блок типа User Replicas)
		Dialogue act (Sender – User, Addressee – Assistant, ..., Purpose - Inform)	
		Dialogue act (Sender – User, Addressee - Assistant, Dimension - Auto-Feedback, Communicative Function – AutoPositive, Purpose - Request)	
Определение последовательности работ	Functional segment		
	Functional segment	Dialogue act (Sender – Assistant, Addressee – User, Dimension – Auto-Feedback, Communicative Function – AutoPositive, Purpose - Request)	Screen – (Блок типа Интент – Запрос на обработку входных данных и знаний, взаимодействие с базой знаний)
Предоставление выходных данных и знаний задачи: работа и ее свойства.	Functional segment		
	Functional segment	Dialogue act (Sender – Assistant, Addressee - User, ..., Purpose - Inform)	Screen – Информировании о содержании работы, которую необходимо выполнить (Блок типа Bot replicas)

Подзадача - Определение признаков неисправности	Functional segment		
Получение входных данных и знаний задачи: объект поиска и устранения неисправности, признаки неисправности.	Functional segment (итерация сегмента в зависимости от количества входных данных и знаний)	(см. выше, описание аналогичное)	(см. выше, описание аналогичное)
Подзадача – Определение подготовительных работ	Functional segment		
	Functional segment	Dialogue act (Sender – Assistant, Addressee - User, ... , Purpose - Inform)	Screen – Информирование о начале выполнения рассматриваемого типа работ (Блок типа Bot replicas)
	...		
	Functional segment	Assistant не понял ответ User	Screen – Assistant не понял ответ User (Блок типа Bot Replicas; Блок типа User Replicas)
	Functional segment	User не понял ответ Assistant	Screen – User не понял ответ Assistant (Блок типа User Replicas; Блок типа Bot Replicas)
	Functional segment	Завершение работы помощника	Screen – Завершение работы помощника (Блок типа Bot Replicas)

### 3. Предлагаемый метод и его реализация

На основе трансформации моделей предлагается следующий метод создания виртуального помощника:

1. Создание пользователем-экспертом модели предметной области на основе предложенного проблемно-ориентированного онтологического шаблона.
2. Автоматизированное формирование модели проблемно-ориентированной задачи на основе онтологического шаблона.
3. Автоматизированное формирование модели проблемно-ориентированного диалога на основе модели задачи.



4. Генерация спецификации проблемно-ориентированного диалога на основе моделей проблемно-ориентированного диалога и модели программного средства создания виртуального помощника.
5. Создание виртуального помощника.

Для реализации этапов метода предложено использовать набор программных решений: Protégé, Aimylogic [Aimylogic, 2023], а также авторский инструментарий [KMS, 2023].

Основные компоненты инструментария:

- редактор диаграмм переходов состояний, предназначенный для визуального моделирования диалога ВП в соответствии с принципами конечного автомата;
- подсистема хранения информации о проектах и моделях (СУБД Postgres);
- подсистема сопряжения с концептуальными моделями (включая OWL онтологии) и электронными таблицами в части импорта их содержимого, предназначенная для наполнения предварительно разработанных и встроенных в ВП моделей;
- подсистема настройки ВП, обеспечивающая специализацию существующей модели ВП в части подключаемых внешних модулей (например, для распознавания текста и голоса) и выбора платформы;
- подсистема генерации кодов и спецификаций.

В настоящий момент проработаны первые три компонента, разработка инструментария продолжается. При этом с использованием инструментария и предложенной технологии реализован прототип ВП в форме чат-бота (рис. 9).

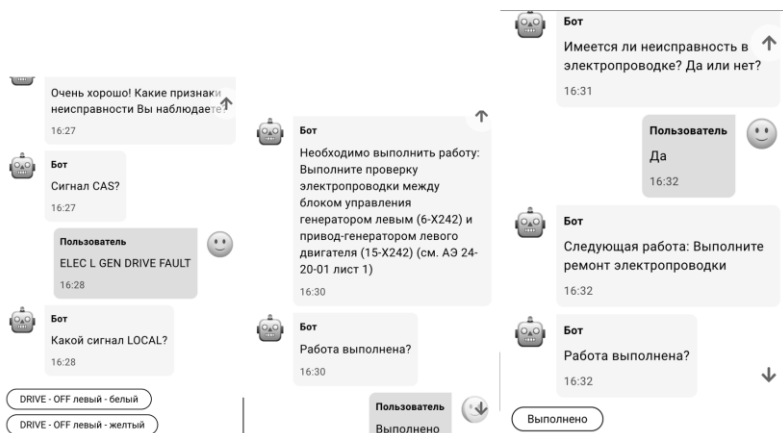


Рис. 9. Фрагменты диалога чат-бота

## Заключение

В данной работе предлагаются технология создания проблемно-ориентированного виртуального помощника (в форме чат-бота) на основе трансформации онтологических моделей. Преимуществом данного подхода является обеспечение возможности непрограммирующему пользователю на основе понятий предметной области и онтологических шаблонов получить готовое программное приложение для решаемой проблемно-ориентированной задачи. В качестве примера для демонстрации моделей использована проблемно-ориентированная задача диагностирования, а объектом задачи рассмотрена система электроснабжения воздушного судна. Созданная модель задачи является онтологическим шаблоном и может быть повторно использована для различных технических объектов.

## Список литературы

- [Da Silva, 2015] Da Silva A.R. Model-driven engineering: A survey supported by the unified conceptual model // Computer Languages, Systems & Structures. – 2015. – Vol. 43.
- [Pérez-Soler et al., 2020] Pérez-Soler S., Guerra E. and De Lara J. Model-Driven Chatbot Development // Lecture Notes in Computer Science. – 2020. – Vol. 12400.
- [Yurin et al., 2018] Yurin A.Yu., Dorodnykh N.O., Nikolaychuk O.A., Grishenko M.A. Designing rule-based expert systems with the aid of the model-driven development approach // Expert Systems. – 2018. – No. 5 (35).
- [Дородных и др., 2019] Дородных Н.О., Юрин А.Ю. Технология создания производственных экспертных систем на основе модельных трансформаций / под ред. О.А. Николайчук. – Новосибирск: СО РАН, 2019.
- [ISO, 2012] ISO 24617-2: A semantically-based standard for dialogue annotation [Электронный ресурс]. – URL: [http://www.lrec-conf.org/proceedings/lrec2012/pdf/530\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2012/pdf/530_Paper.pdf) (Дата обращения: 29.09.2022).
- [Kiefer et al., 2021] Kiefer B., Welker A. and Biber C. VOnDA: A Framework for Ontology-Based Dialogue Management // Lecture Notes in Electrical Engineering. – 2021. – Vol. 714.
- [Aceta et al., 2021] Aceta C., Fernandez I. and Soroa A. TODO: A Core Ontology for Task-Oriented Dialogue Systems in Industry 4.0 // Studies on the Semantic Web. – 2021. – Vol.53.
- [Berman et al., 2018] Berman A.F., Nikolaychuk O.A., Pavlov A.I. The Ontology Model for Automating the Solution of Multidisciplinary Research Tasks // Proc. the V Intern. Workshop Critical Infrastructures: Contingency Management, Intelligent, Agent-Based, Cloud Computing and Cyber Security (IWCI 2018). – 2018. – Vol. 158.
- [Dorodnykh et al., 2022] Dorodnykh N.O., Nikolaychuk O.A., Yurin A.Yu., Stolbov A.B. Towards knowledge-based virtual assistant development with the aid of ontology transformations // Proceedings of the 2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON). – 2022.
- [Aimylogic, 2023] Visual builder for AI-powered bots. Aimylogic [Электронный ресурс]. URL: <https://aimylogic.com/> (дата обращения: 30.04.2023).
- [KMS, 2023] Knowledge Modeling System [Электронный ресурс]. – URL: <http://kms.knowledge-core.ru/> (дата обращения: 30.04.2023).

## ПРЕДСТАВЛЕНИЕ И ОБРАБОТКА ТЕМПОРАЛЬНЫХ ПРЕЦЕДЕНТОВ В ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ РЕАЛЬНОГО ВРЕМЕНИ

А.П. Еремеев (*eremeev@appmat.ru*)

Н.А. Панявин (*PaniavinNA@mpei.ru*)

М.В. Фомина (*m\_fomina2000@mail.ru*)

Национальный исследовательский университет «МЭИ», Москва

Рассматривается проблема повышения эффективности принятия решений на основе прецедентов (Case-Based Reasoning, CBR) в интеллектуальных системах реального времени. Предлагается метод предварительной обработки базы прецедентов (БП), включающий этапы обобщения свойств динамических параметров, формирование классов схожих ситуаций, связанных с прецедентами БП, построение деревьев решений для эффективного поиска решений в пределах каждого класса. Приводятся примеры обработки прецедентов, результаты машинного эксперимента. Рассматривается возможность реализации предложенного подхода на основе темпоральной БД, создаваемой в среде нереляционной графовой СУБД Neo4j.

**Ключевые слова:** интеллектуальная система, реальное время, вывод на основе прецедентов, темпоральные данные, нереляционная база данных.

### Введение

В настоящее время активно развивается направление, связанное с разработкой и внедрением интеллектуальных систем (ИС), реализующих возможности интеллектуального анализа данных (ИАД), включая ИС, ориентированные на анализ ситуаций в открытых и динамических предметных областях [Варшавский и др., 2009]. Поиск решения в ИС связан со структурированием, упорядочением данных и знаний, поступивших от различных источников. Такие сложно организованные знания удобно представлять в виде базы прецедентов (БП). При работе с БП возникает проблема поиска «наиболее подходящего» на основе выбранного критерия прецедента с дальнейшей его адаптацией к сложившейся ситуации.

Эффективность методов поиска решений на основе прецедентов (CBR-методов), связана с наличием удобных средств поиска аналогов среди имеющихся примеров. В работе предлагается исследовать методы улучшения поиска прецедентов за счёт предварительной структуризации БП, которая содержит темпоральные данные. Предлагается организации темпоральной базы данных на основе нереляционной (NoSQL) графовой СУБД для поддержки темпоральных прецедентов.

## 1. Представление темпоральной информации

Поиск решения на основе прецедентов связан с поиском ситуации, наиболее близкой к той, которая наблюдается в текущий момент. Система ИАД может включать методы классификации прецедентов с последующим уточнением решения, связанного с определённым классом ситуаций (например, штатные, нештатные, аварийные). Описание некоторой ситуации (объекта БП)  $s$  традиционно задается в виде набора признаков:  $s = \langle x_1, x_2, \dots, x_n \rangle$ . Однако, при реализации современных ИС, следует учитывать возможность работы со сложными динамическими объектами и процессами, поэтому нужны средства для представления временных (темпоральных) зависимостей, отражающих динамику изменения ситуации (поведения объекта) во времени. Таким образом, возникает задача интеллектуального анализа темпоральных данных, что требует разработки новых методов анализа таких данных [Lin et al., 2002].

В общем виде прецедент можно задать как

$$CASE = (Situation, Solution, Result), \quad (1.1)$$

где *Situation* – ситуация  $s$ , для которой создан прецедент, *Solution* – решение (например, диагноз, причина возникновения аномальной ситуации, рекомендуемые действия и т.п.), *Result* – результат (прогноз) применения предлагаемого решения (рекомендации). Для динамических предметных областей в описаниях ситуаций будет присутствовать фактор времени.

Пусть имеется  $n$  параметров, значения которых были измерены в некоторые дискретные моменты времени:  $t = 0, 1, 2, 3, \dots$ . Тогда ситуацию  $s(t)$  в момент времени  $t$  можно представить в виде:

$$s(t) = \langle x_1(t), x_2(t), \dots, x_n(t) \rangle. \quad (1.2)$$

Очевидно, чтобы проследить динамику развития ситуаций и тенденции, необходимо проанализировать упорядоченное множество таких векторов, полученных на интересующем временном интервале  $(t_0, t_r)$ ,  $r > 1$ .

Введем следующие определения. Параметр  $x_j(t)$ , значения которого зависят от  $t = 0, 1, 2, 3, \dots$ , назовем *динамическим параметром*. Прецедент *CASE*, представленный в виде (1.1), назовем *темпоральным прецедентом*, если описание  $s$  в (1.2) содержит хотя бы один динамический параметр. Тогда можно говорить о *динамическом описании ситуации*.

Каждый динамический параметр  $x_j(t)$  может быть рассмотрен как *временной ряд* (последовательность наблюдений, упорядоченных по времени), показывающий изменение значений определённого параметра на интервале  $(t_0, t_r)$ . Предлагается использовать предварительную обработку (препроцессинг) и свёртку массивов темпоральных данных с выделением наиболее существенных характеристик наблюдаемой ситуации.

## 2. Предварительная обработка (свёртка) динамической информации

Модель динамического описания ситуации связана с непрерывным поступлением больших массивов информации, что затрудняет хранение прецедентов, а также выбор управляющих воздействий, так как на развитие ситуации могут влиять сразу несколько динамических параметров. Введем следующие ограничения.

Темпоральный прецедент, заданный (1.1), включает описание ситуации *Situation*, где помимо статических параметров имеется по крайней мере один динамический параметр, контроль за которым наиболее важен:

$$s(t) = \langle x_1, x_2, \dots, x_{n-1}, y(t) \rangle \quad (2.1)$$

Таким образом в описание ситуации входит временной ряд  $y(t)$ .

Рассмотрим проблему качественного оценивания временного ряда  $y(t)$ . Предлагается использовать описанный в [Антипов и др., 2012] алгоритм TS-Adeep, включающий процедуры снижения размерности временного ряда (подавление помех), нормализацию числовых значений временного ряда, переход от числового представления временного ряда к символьному. В основу последнего преобразования положен алгоритм SAX [Keogh et al., 2003]. Цель перехода к символьному представлению – получение возможности качественно оценить процесс изменения динамического параметра.

В табл. 1 представлены основные этапы предобработки темпоральных данных. Первая строка таблицы – это исходный ряд, где представлены абсолютные значения *Хабс* вторая строка содержит нормализованные данные, третья строка представляет перевод числовых значений в символьные. Обработке предшествует этап предварительного сжатия информации с целью подавления помех и выбросов, но позволяя сохранить основную тенденцию развития процесса во времени. Тенденция может быть охарактеризована качественно: «резкий спад параметра», «ступенька», «пила» и т.п. [Saito, 1994]. На этапе нормализации выполняется переход от абсолютных значений ряда к относительным: временной ряд приводится к такому виду, что среднее для него было бы равно нулю, а среднеквадратичное отклонение – единице [Keogh et al., 2003]. Преобразование динамического параметра в нормализованный вид, а затем в набор символов

дает возможность использовать методы и алгоритмы обобщения с целью получения критериев разделения параметров на классы [Антипов и др., 2012]. При этом каждому классу экспертом приписывается качественное значение, характеризующее тенденцию развития динамического процесса.

Таблица 1

Время $t$	1	2	3	4	5	6	7	8	9
.Абсолютные значения $Хабс$	512	1448	88	1448	1448	1448	1448	1024	512
Нормализация $Хабс$	-1.04	0.75	-1.85	0.75	0.75	0.75	0.75	-0.06	-0.04
Символьные значения	<i>C</i>	<i>G</i>	<i>A</i>	<i>G</i>	<i>G</i>	<i>G</i>	<i>G</i>	<i>B</i>	<i>C</i>

Рассмотрим обработку динамического параметра на примере набора данных User Knowledge Modeling (UKM) из репозитория Калифорнийского университета UCI Repository of Machine Learning Datasets [UCI, 1998]. Набор данных содержит информацию об уровне знаний обучающихся (студентов) по дисциплине «Электрические машины постоянного тока». Набор данных включает 258 примеров, принадлежащих одному из 4 решений: 1 – уровень знаний очень низкий (Very low), 2 – низкий (Low), 3 – средний (Middle) и 4 – высокий (High).

На результат влияют следующие 5 параметров, заданных в числовой форме: STG – доля учебного времени, затраченного для изучения материалов по дисциплине; SCG – доля повторов, допустимых при изучении материалов; STR – доля учебного времени, затраченного для изучения материалов по смежным дисциплинам; LPR – результаты экзаменов по смежным дисциплинам; PEG – результаты экзаменов по дисциплине. Все параметры принимают значения в интервале [0, 1].

Каждый пример набора данных рассматриваем как прецедент, хранящийся в БП. Введенные выше пять параметров трактуем как описание ситуации, каждой из которых сопоставлено Solution из множества {High, Middle, Low, Very low}.

Из всех параметров, влияющих на результат, для предварительного анализа был выбран параметр STG, рассматриваемый как динамический параметр  $y(t) = STG(t)$ , представляющий распределение времени, затраченного студентом на изучение предмета в течение 10 учебных недель. Пример временного ряда представлен в табл. 2.

Здесь  $t$  – учебная неделя, строка  $STG(t)$  – доля учебного времени, затраченная в каждую неделю на освоение материала. В строке  $STG'(t)$  каждое последующее значение суммируется с предыдущими значениями. Соответственно значение  $STG'(10)$  – это доля времени, затраченная сту-

дентом на изучение дисциплины в течение всего семестра. Временные ряды были построены для каждого объекта набора данных, следовательно, мы получили динамические объекты.

Для объектов обучающей выборки было выполнено обобщение с помощью алгоритма TS-Adeep.

Таблица 2

$t$	1	2	3	4	5	6	7	8	9	10
$STG(t)$	0	0.1	0.01	0.05	0	0.08	0.01	0.05	0	0.1
$STG^*(t)$	0	0.1	0.11	0.16	0.16	0.24	0.25	0.3	0.3	0.4

В результате были выделены три типа временных рядов:

- TS1 – студенты, которые занимались равномерно, результат высокий;
- TS2 – студенты, которые занимались неравномерно; результат высокий;
- TS3 – студенты, которые занимались неравномерно, результат низкий.

На рис. 1 показаны в сравнении временные ряды, относящиеся к типам TS1, TS2 и TS3.

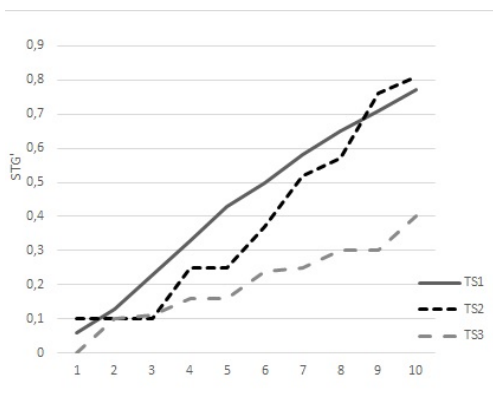


Рис. 1. Примеры временных рядов TS1, TS2 и TS3

Следует отметить, что не были выделены в отдельный класс студенты с низкими результатами, которые занимались равномерно, из-за нетипичности этой ситуации, так, из 258 примеров, включенных в набор, только 3 могли быть оценены таким образом.

График TS3 – студенты с низким результатом, которые занимались неравномерно – соответствует данным из табл. 2. Получение качественных характеристик темпоральных зависимостей дало возможность провести следующий этап структуризации базы прецедентов.

### 3. Структуризация базы прецедентов

Эффективность поиска прецедента в БП напрямую зависит от организации хранения прецедентов, на множестве которых устанавливается отношение сходства (близости). На первом этапе предварительной обработки динамического параметра были сформированы классы объектов, обладающих сходными свойствами. Таким образом выполнен переход от табличного задания темпорального (динамического) атрибута к качественному описанию такого атрибута. Динамический параметр  $y(t)$  заменяется на несколько параметров  $y_1, y_2, \dots, y_m$ , не представляющих время явно, но отражающих наиболее важные закономерности, связанные с  $y(t)$ . Назовём такие параметры *обобщенными* параметрами. Теперь описание *Situation* (2.1) примет вид

$$s' = \langle x_1, x_2, \dots, x_{n-1}, y_1, y_2, \dots, y_m \rangle. \quad (3.1)$$

Для рассмотренного примера в результате применения алгоритма TS-Adeer для каждого объекта (ситуации) в БП динамический параметр  $STG(t)$  представляется набором из трёх значений:

$$\langle \text{Тип временного ряда, } STG'(10), Interval \rangle. \quad (3.2)$$

Тип временного ряда обозначается как TS1, TS2 или TS3 и качественно характеризует ситуацию, числовой параметр  $STG'(10)$  показывает общее время (долю времени), затраченное на освоение дисциплины, параметр *Interval* – длительность наблюдений ( $Interval = 10$ ).

Далее для поиска групп близких (схожих) ситуаций используется итерационный алгоритм MAXMIN, который детально описан в [Astahova et al., 2018]. Для разделения объектов БП на классы алгоритм MAXMIN использует атрибуты  $x_1, x_2, \dots, x_{n-1}, y_1, y_2, \dots, y_m$  относящиеся к описанию *Situation*. Отнесение каждого объекта к одному из классов выполняется на основе критерия минимального расстояния от точек-прототипов этих классов. В качестве таких точек первоначально выбираются прецеденты, для которых были приняты различные решения *Solutions*. Затем в каждом классе выбирается объект, наиболее удаленный от своего прототипа, при этом могут быть созданы новые классы и корректируется значение порога  $T$ . Окончательным считается разбиение, для которого в каждом классе расстояние от точки-прототипа до всех объектов этого класса не превышает значения  $T$ .



Результатом работы алгоритма MAXMIN будет разбиение БП на множество классов  $M_1, M_2, \dots, M_K$ . Каждый  $M_i$  ( $i = 1, 2, \dots, K$ ) включает прецеденты, обладающие наибольшим сходством с «одноклассниками» по параметрам, которые использовались для описания Situation. Однако, в соответствии с (1.1) выбор прецедента (CASE) в конкретном случае может потребовать дополнительного анализа с использованием сведений о *Solution* и *Result*.

При поступлении нового прецедента  $C$  (CASE), не совпадающего точно ни с одним из прецедентов БП, его можно отнести к одному из полученных классов  $M_1, M_2, \dots, M_K$  на основе анализа описания ситуации *Situation*. Тогда целью дальнейших действий будет поиск подходящего решения *Solution* и оценка последствий *Result*. Возможна следующая ситуация.

А. Для нового прецедента  $C$  ситуация наиболее близка к ситуациям класса  $M_i$ . Если все прецеденты, включенные на этапе структуризации в класс  $M_i$ , были отнесены к сходным ситуациям *Situation*, и для них были приняты одинаковые решения *Solution*, решение для прецедента  $C$  найдено.

В. Новый прецедент  $C$  приписан к классу  $M_i$ , где прецеденты, вошедшие в класс  $M_i$  были отнесены к сходным ситуациям *Situation*, но для них были приняты различные решения *Solution* (возможная причина – неполнота или нечеткость описания *Situation* в БП). Требуется дополнительный анализ.

Одним из подходов к поиску решения в последнем случае может стать построение дерева решений (ДР). Было предложено использовать алгоритм С4.5 [Quinlan, 1996], который, строит классификатор на основе обучающей выборки.

Процесс построения ДР для множеств  $M_i$ , которые относятся к случаю В, заключается в следующем. Все  $s' \in M_i$  образуют обучающую выборку. Параметры ситуации  $s' = \langle x_1, x_2, \dots, x_{n-1}, y_1, y_2, \dots, y_m \rangle$  трактуются как атрибуты объектов обучающей выборки. Класс объекта (решающий атрибут) определяется тем решением (*Result*), которое было принято для конкретного  $s'$ . ДР  $Tree_i$ , построенное алгоритмом С4.5, представляет собой классификатор, который применяется для прецедентов конкретного класса  $M_i$ . Важно, что при построении ДР используются как статические параметры  $x_1, x_2, \dots, x_{n-1}$ , так и обобщенные параметры  $y_1, y_2, \dots, y_m$ , полученные при обработке динамических параметров. Новая ситуация первоначально относится к конкретному классу  $M_i$  на основании критерия максимальной близости к его центру. Когда класс  $M_i$  определен, дальнейшее решение уточняется с помощью ДР  $Tree_i$ , связанного с классом  $M_i$ .

#### 4. Эксперименты и оценка результатов

В ходе исследования было выполнено программное моделирование изложенных выше методов и алгоритмов. На рис. 2 представлены результаты разбиения объектов на классы алгоритмом MAXMIN, полученные при работе с набором данных УКМ после свертки динамического параметра и перехода к качественным значениям для STG.

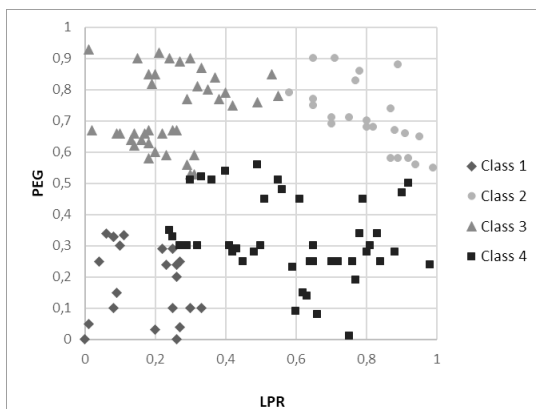


Рис. 2. Результаты разбиения обучающего множества на классы

При разбиении обучающей выборки на классы алгоритмом MAXMIN с учётом всех атрибутов, использованных для описания объектов (примеров) этого набора данных, первоначально было получено излишне большое число «мелких» классов (свыше 40). Поэтому были выделены два наиболее важных параметра из пяти – PEG и LPR, наиболее близко расположенных к корню ДР, построенного алгоритмом С 4.5. На их основе обучающая выборка была разбита на 4 класса, представленных на рис. 2.

Теперь сопоставим классам ситуаций решения, приписанные объектам обучающей выборки; это оценка результативности обучения High, Middle, Low, Very Low. Результаты представлены на рис. 3.

Из рис. 3 видно, что все объекты, отнесенные на рис. 2 к классу 2 (круг), представляют ситуации, для которых было принято единственное общее решение (Solution = High). Для элементов классов 1 и 3, представленных на рис. 2 (ромб и треугольник), принятие решения потребовало проверки одного дополнительного условия.

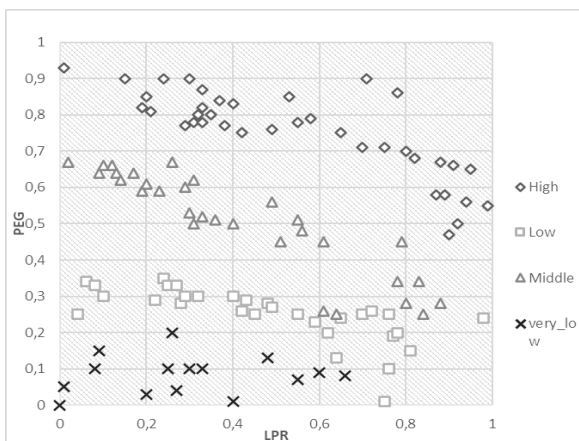


Рис. 3. Решения, приписанные элементам обучающей выборки

Для элементов класса 4 было построено ДР Tree<sub>4</sub> (рис. 4). Объекты этого класса представлены на рис. 2 символами ■.

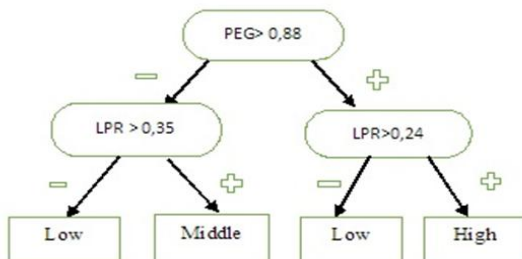


Рис. 4. Дерево решений для объектов класса 4

Отметим, что ДР, построенное на полной обучающей выборке без предварительного разбиения на классы, существенно сложнее: имеется 30 конечных вершин и в большинстве случаев достижение конечной вершины требует проверки 5 условий (глубина поиска – 5) [Astahova et al., 2018]. Аналогичные исследования были выполнены для ряда других наборов из репозитория [UCI, 1998]. Алгоритмы сравнивались по критерию точности классификации экзаменационных примеров. Результаты эксперимента на различных обучающих выборках представлены ниже в табл. 3.

Таблица 3

Наборы данных	Использование алгоритма C 4.5	Использование алгоритмов C 4.5 и MAXMIN
UKM	64,69%	65,06%
Iris	96,67%	95%
Glass	65,89%	64,95%
Transfusion	74%	76%

Табл. 3 показывает точность классификации тестовых примеров в обучающих выборках для двух вариантов: построения классификатора: без использования алгоритма MAXMIN и с предварительным использованием этого алгоритма. В результате эксперимента установлено, что дополнительное использование алгоритма MAXMIN при построении классификатора существенно сокращает количество проверок, необходимых для принятия решения (достижения листа на ДР), что обеспечит ускоренный поиск нужного прецедента в БП. Также можно сделать вывод, что такой подход позволяет не только сократить количество проверок условий при поиске решений, но и повысить точность классификации для наборов UKM, Transfusion. Для наборов Iris и Glass наблюдалось незначительное снижение точности классификации, что в определенной степени объясняется спецификой используемых в этих наборах данных.

## 5. Использование нереляционной СУБД для представления данных БП

Рассмотрим проблему организации хранения прецедентов в БП при наличии данных, зависящих от времени (темпоральных данных). Для представления сложно организованных данных, которые используются для описания прецедентов, проводилось сравнение возможностей реляционной и нереляционной моделей СУБД. Было обосновано использование нереляционной (NoSQL) СУБД Neo4j [Azzini et al., 2019]. Главное отличие двух моделей заключается в том, что нереляционная графовая модель для узлов-сущностей и ребер-отношений использует явно определенные правила. Поскольку все отношения между сущностями в графовой модели определяются явно, отсутствует необходимость моделирования внешних ключей, а также не нужно прописывать всю логику их хранения. Достаточно определить схему сущностей и отношений между ними. Наибольшая эффективность такого подхода хранения возникает в задачах моделирования сильно связанных данных.

Вместо явного указания и хранения дополнительных ссылок, как в реляционных, таблицах, в качестве атрибутов, графовая СУБД позволяет хранить указатели на ячейку памяти, содержащую ближайшую связанную сущность. Большинство существующих графовых СУБД хранят прямые

ссылки на данные, которые связаны (в отличие от организации хранения ссылок на аналогичные объекты для представления этих связей в реляционных табличных моделях).

Преимущество предложенного подхода в том, что графовая модель представления данных, в том числе и темпоральных, позволит хранить как уже обработанные данные, так и данные в состоянии предобработки. Такая возможность минимизирует необходимость строго фиксировать формат хранения данных на разных этапах [Еремеев и др., 2022]. В то же время, внутренняя гибкая структура представления данных (что особо важно для динамических данных) внутри узла/ребра позволяет хранить данные произвольной сложности в json формате, что также избавляет от необходимости объявления полей, где значение не определено, и хранения null значений, что требуется в реляционных моделях данных.

### Заключение

В работе рассмотрены проблемы обработки темпоральных данных, связанных с организацией хранения прецедентов, в том числе и темпоральных прецедентов в БП. Предложен метод ускорения поиска прецедентов в БП, за счёт предварительной структуризации примеров и построения системы дополнительных классификаторов в виде множества деревьев решений. Приведены результаты машинных экспериментов на наборах данных из коллекции UCI Repository of Machine Learning Datasets [UCI, 1998]. Для поддержки предложенного подхода и создания темпоральной БД предлагается применить нереляционную (NoSQL) СУБД. Выбор данной СУБД основывается на отсутствии необходимости строго фиксировать структуру данных на начальном этапе, а также на гибкости самой структуры представления динамических данных, которая может меняться по мере поступления новой информации.

Данные исследования и разработки направлены на повышение эффективности создаваемых методов и инструментальных средств конструирования интеллектуальных систем поддержки принятия решений, работающих в режиме реального времени.

### Список литературы

- [Варшавский и др., 2009] Варшавский П.Р., Еремеев А.П. Моделирование рассуждений на основе прецедентов в интеллектуальных системах поддержки принятия решений // Искусственный интеллект и принятие решений. – 2009. – № 2.
- [Lin et al., 2002] Weiqiang Lin, Mehmet A. Orgun, Graham J. Williams. An Overview of Temporal Data Mining // Proceedings of the 1st Australasian Data Mining Workshop. – 2002.

- [**Антипов и др., 2012**] Антипов С.Г., Фомина М.В. Проблема обнаружения аномалий в наборах временных рядов // Программные продукты и системы. – 2012. – № 2 (2).
- [**Keogh et al., 2003**] Keogh E., Kasetty Sh. On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration/ Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. – 2003. – Vol. 7 (4).
- [**Еремеев и др., 2021**] Еремеев А.П., Варшавский П.Р., Поляков С.А. Программная реализация модуля анализа данных на основе прецедентов для распределенных интеллектуальных систем // Программные продукты и системы. – 2021. – № 34 (3).
- [**UCI, 1998**] UCI Repository of Machine Learning Datasets. Available: <http://archive.ics.uci.edu/ml/>.
- [**Saito, 1994**] Saito N. Local feature extraction and its application using a library of bases: Ph.D. thesis / Yale University. 1994.
- [**Quinlan, 1996**] Quinlan J.R. Improved Use of Continuous Attributes in C 4.5. // Journal of Artificial Intelligence Research. – 1996. – No. 4.
- [**Astahova et al., 2018**] Astahova I.I., Fomina M.V. and Shcherbakova V.N. Development of Precedents Searching Methods Based on Decision Trees/ Proceedings of the II International Scientific and Practical Conference “FTI 2018”. CEUR.Workshop Proceedings. – 2018. – Vol. 2258 (42).
- [**Azzini et al., 2019**] Azzini A., Ceravolo P., Colella M. Performances of OLAP Operations in Graph and Relational Databases //International Conference on Knowledge Management in Organizations. – Springer Cham, 2019.
- [**Еремеев и др., 2022**] Еремеев А.П., Панявин Н.А. Унификация модели представления данных и преобразование форматов на основе нереляционной СУБД Neo4j // Программные продукты и системы. – 2022. – Т. 35 (4).

УДК 004.421.6

## ФРЕЙМВОРК ДЛЯ РАЗРАБОТКИ ИНТЕЛЛЕКТУАЛЬНЫХ ОБУЧАЮЩИХ СИСТЕМ НА ОСНОВЕ МОДЕЛЕЙ ПРЕДМЕТНЫХ ОБЛАСТЕЙ В ВИДЕ ДЕРЕВЬЕВ РЕШЕНИЙ\*

А.И. Крыгин (*andrey\_krygin@outlook.com*)

М.Р. Гумеров (*max\_person@mail.ru*)

Н.А. Москаленко (*nekit\_vlg@mail.ru*)

О.А. Сычев (*oasychev@gmail.com*)

Волгоградский государственный технический университет,  
Волгоград

Формальные модели предметных областей позволяют строить интеллектуальные обучающие системы основанные на символическом искусственном интеллекте с надежным определением ошибок обучаемых, генерацией объясняющей обратной связи, наводящих вопросов и другими свойствами. Но разработка таких моделей для современных машин вывода очень трудоемка. В работе описывается метод представления решения и проверки задач предметных областей в виде деревьев решений и прототип фреймворка для разработки интеллектуальных обучающих систем на его основе. При опытной разработке четырех ядер интеллектуальных обучающих систем фреймворк показал умеренную производительность и значительное сокращение сроков разработки.

**Ключевые слова:** деревья решений, интеллектуальные обучающие системы, модели предметных областей, фреймворк.

### Введение

Интеллектуальная обучающая система (ИОС, Intelligent Tutoring System) – это программное обеспечение, созданное для имитации навыков, поведения и указаний преподавателя. ИОС разрабатываются с использованием технологий ИИ, пытаясь приблизиться к техникам обучения, которые, при их выполнении человеком, можно было бы назвать “ин-

---

\* Исследование выполнено при финансовой поддержке ВолгГТУ в рамках научно-го проекта № 60/438-22.

теллектуальными” [Nwana, 1990]. Преимуществами ИОС по сравнению с преподавателем-человеком являются возможность охвата большого количества учащихся, а также ведение проработки темы до овладения ею (training until mastery) без ограничений по времени.

Частыми функциями ИОС являются способность анализировать сложные ответы учащегося и адаптироваться к ним в процессе работы, возможность определить, где и почему учащийся допустил ошибку, и дать необходимые подсказки, чтобы помочь ему в понимании материала.

В традиционной форме ИОС описываются как оперирующие информацией трех ключевых видов: информацией о предметной области, учащемся и способах или стратегиях обучения – именно эти данные обеспечивают “интеллектуальность” системы. В классических архитектурах ИОС выделяют 3 соответствующих информационных модели (модель предметной области, модель учащегося и педагогическая модель), а также модуль, ответственный за пользовательский интерфейс и его особенности [Nwana, 1990], [Alkhatlan et al., 2018].

В силу такого разделения, большинство независимых ИОС включают в себя специфическую реализацию трех описанных моделей. Большинство современных ИОС преуспевают только в узком кругу задач и неразрывно связаны с представляемой ими предметной областью [Soofi et al., 2019], [Mousavinasab et al., 2018].

Для создания универсальных ИОС, основанных на моделях предметных областей, необходимо формальное описание модели предметной области. Оно используется как входные данные для обучающей системы, делая саму систему не зависящей от специфики изучаемых проблем. Небольшое количество ИОС реализует данный подход [Freedman et al., 2000], [Singh et al., 2022]. При этом по-прежнему может стоять необходимость разработки отдельного пользовательского интерфейса для каждой предметной области, однако наибольшие трудозатраты всегда связаны с моделированием предметных областей. Их минимизация наиболее актуальна для широкого распространения ИОС.

В работе [Turner et al., 2005] реализован инструмент ASSISTment Builder для создания быстрых “псевдо-тренажеров”, использующих когнитивную модель, основанную на графе состояний, настроенном для конкретной задачи ее автором. Однако следует отметить, что ИОС, разработанные с использованием этого инструмента, имеют ограниченную область применения, поскольку каждая предназначена для решения одной учебной задачи. Процесс создания ИОС приходится повторять для каждой задачи.

В работе [O'Rourke et al., 2015] предлагается формализация процессов решения класса проблем в виде кода на независимом от предметной области языке, называемом языком мыслительных процессов (Thought



Process Language, TPL). По мере выполнения TPL кода система регистрирует факты выполнения определенных операторов языка, которые затем могут быть отображены с помощью “защепок” интерфейса, специфичных для предметной области. Такое представление мыслительного процесса в виде абстрактного кода, хотя и является перспективным подходом, ограничено представлением TPL – представлением мыслительного процесса обучаемого в ИОС, базирующееся на прямом соответствии программному языку, заставляет его мыслить в терминах этого языка (присвоениях, развилках, циклах), что не является естественным для человеческого мышления (основанного на задавании вопросов, отбору по критериям и принятии решений). Кроме того, TPL не описывает поиск семантических ошибок и генерацию их объяснений. Он позволяет только подсказывать правильные решения.

В работе [Singh et al., 2022] описывают высокоуровневую модель предметной области, состоящая из тем, зависящих друг от друга (если для понимания одной темы необходимо первоначально ознакомиться с другой) и объединенных в курсы, также зависящие от тем и друг от друга. Данная схема используется в совокупности с классификацией запросов учащихся на соответствие темам, а использование экспертов предметной области для реализации новых тем обеспечивает системе расширяемость. Эта высокоуровневая модель не моделирует мышление учащихся при освоении конкретных тем и может использоваться только при планировании занятий.

Рыбина [Рыбина, 2023] описывает модели и технологию построения обучающих интегрированных экспертных систем на основе онтологического подхода, реализованную в комплексе АТ-ТЕХНОЛОГИЯ. Но в этой технологии базовая модель онтологии курса сводится к моделированию структуры изучаемых тем курса и не затрагивает уровень моделирования собственно вырабатываемых умений и навыков.

Одной из основных проблем использования ИОС является трудоемкость их разработки, особенно при использовании стандартных машин вывода. Без использования заменяемой модели предметной области, ИОС должна обладать всеми необходимыми экспертными знаниями; примерное время разработки составляет от 200 до 300 часов разработки на 1 час обучающего контента [Aleven et al., 2006]. Хотя использование формальной модели предметной области дает возможность повторного использования и быстрого изменения необходимой модели, ее построение остается трудозатратным процессом, что обусловлено тем, что подобные модели должны быть описаны в более общих терминах, нежели интегрированные, специфичные модели [Martin et al., 2003], [Mitrovic et al., 2007]. Поэтому важно облегчить труд разработчиков обучающих моделей предметных областей и повысить их производительности [O'Rourke et al., 2015], [Aleven et al., 2006], [Mitrovic et al., 2007].

В данной работе мы рассматриваем новый подход к представлению формальных моделей предметных областей, а также фреймворк, использующий данный подход для облегчения построения ИОС, независимых от предметной области.

## **1. Деревья решений и задачи в интеллектуальных обучающих системах, решаемые с их помощью**

При определении формата описания предметной области для ИОС необходимо учитывать стоящие перед системой задачи, так как ее описание должно включать в себя всю необходимую информацию для их решения.

В данной работе мы основываемся на моделировании предметных областей в виде деревьев решений. В общем случае, такие деревья моделируют процессы рассуждений предметной области, что открывает возможности решений различных педагогических задач. Деревья могут отвечать на такие вопросы как: “Какой следующий шаг является правильным для решения задачи?” или “Корректен ли данный шаг в решении задачи и если нет, то почему?”.

Деревья решений описываются узлами (шагами рассуждения) и переходами между ними:

- Узлы вопросов – принять решение в зависимости от некоторого условия (ответа на некоторый вопрос).
- Узлы ветвления – принять решение в зависимости от результатов нескольких подзадач (рассуждений более низкого уровня).
- Узлы результатов означают получение ответа на (под)задачу.
- Узлы действий – определить ключевые объекты задачи для использования в дальнейших рассуждениях.
- Циклические узлы – выполнить некоторые рассуждения для набора ключевых объектов.

Для связи с формальным условием задачи и ответом обучаемого, эти деревья сопровождаются словарями понятий предметной области:

- Классы определяют категории или типы объектов в предметной области.
- Они поддерживают наследование позволяя определять общие характеристики для группы понятий.
- Свойства описывают атрибуты или характеристики объектов некоторого класса.
- Отношения задают связи и взаимоотношения между объектами указанных классов.

Процесс рассуждения в узлах дерева решений представляет собой вычисление выражений логики первого порядка над базой знаний о конкретной задаче. Структура связей между узлами задает общий процесс рассуждений.

Деревья решений могут моделировать две категории процессов:

- Процесс решения задач – рассуждения, которые необходимы непосредственно для нахождения (создания или определения) правильного ответа на конкретную задачу. Деревья этой категории назовем деревьями вычислений.
- Процесс проверки ответов на задачи – рассуждения, необходимые для проверки частичного ответа обучаемого на корректность. Такие деревья назовем деревьями проверок.

Эти категории деревьев предоставляют различные возможности разрабатываемой ИОС. Деревья вычислений дают возможность:

- “Прямого” решения задач путем автоматического выполнения рассуждений, предписанных деревом вычислений. ИОС может найти корректный ответ на любую из своих задач и сравнить ответ учащегося с корректным. Это значительно облегчает как ввод задач учителем, так и их автоматическую генерацию.
- Генерации интеллектуальных подсказок, объясняющих один правильный шаг в решении задачи. Задействованные узлы дерева решений позволяют генерировать объяснения, почему данный шаг в данных обстоятельствах правильный.

Деревья проверок позволяют:

- Определять семантические ошибки обучаемого – узлы результаты в дереве проверок прямо соответствуют ограничениям (constraints) предметной области, которые были нарушены в данном ответе, что позволяет организовать обучающую обратную связь.
- Генерировать наводящие вопросы для определения конкретных ошибок в мышлении пользователя. С их помощью можно установить ошибки в знании фактов, правил или порядка их применения.

Таким образом, моделирование предметных областей с помощью деревьев решений обеспечивает хорошую основу для построения ИОС, участвующих процессы рассуждений учащихся.

## **2. Разработанный фреймворк**

Разработанный фреймворк состоит из трех частей: графического конструктора деревьев решений предметной области, машины вывода и генератора наводящих вопросов.

Графический конструктор был реализован на базе редактора диаграмм Draw.io [diagrams.net, 2021]. Данный редактор имеет открытый исходный код и поддерживает плагины. Мы разработали плагин дающий возможность создания и редактирования узлов деревьев решений, и элементов, содержащих словари используемых классов, свойств и отношений. Узлы содержат дополнительную информацию в зависимости от их вида. Связи между узлами представлены стрелками, которые хранят значения возможного результата выполнения узла.

Формальным описанием действий с базой знаний, необходимых для выполнения узлов, служат выражения рассуждений. Для записи этих выражений был разработан специальный язык логики первого порядка для удобной обработки графов знаний (например, представленных в формате RDF). Он включает в себя логические операции и кванторы и позволяет выполнять операции проверки, получения и изменения информации в графовой базе знаний, а также оперировать переменными как в рамках одного выражения, так и в рамках всего дерева решений. Значениями переменных являются объекты базы знаний задачи. Свойства объектов могут хранить данные различных типов: целые и дробные числа (ограниченные в некоторых пределах или неограниченные), булевы значения, строки и заданные автором перечисления.

На рис. 1 приведены примеры выражений на разработанном языке.

<pre> find C {   \$C is operand and   \$C-&gt;isInside(var:X) and   \$C.state == state::evaluated } </pre>	<pre> forall X_t [   var:X-&gt;has(\$X_t) ] {   \$X_t-&gt;leftOf(var:Y2) and   \$X_t-&gt;rightOf(var:Y1) } </pre>
(а)	(б)

Рис. 1. Пример выражения-действия (а) и выражения-утверждения (б)

Выражение, изображенное на рис. 1,а, осуществляет поиск объекта С, который имеет класс operand, находится внутри объекта, сохраненного в переменную X (связан отношением isInside с этим объектом) и имеет свойство state со значением unused. Выражение, изображенное на рис. 1,б, проверяет верность утверждения, что для всех X\_t, связанных с переменной X отношением has верно, что X\_t находится справа от объекта Y1 (связан отношением leftOf) и слева (отношение rightOf) от Y2.

Выражения для узлов могут задаваться как с помощью текстовой записи, так и визуально, посредством блоков из библиотеки Blockly [Pasternak et al., 2017]. Блоковое представление позволяет уменьшить время, требуемое для изучения синтаксиса текстовой записи разработанного языка, и исключить синтаксические ошибки.

Созданное в данном конструкторе дерево решений является основой для решения всех описанных в предыдущем разделе задач. Для этого необходимо выполнять как отдельные узлы, так и ветви дерева в целом; этим занимается машина вывода. Простейшим вариантом ее реализации является интерпретатор дерева решений, который был реализован авторами для оценки перспектив фреймворка.

Более эффективной реализацией является преобразование дерева решений или отдельных его узлов в продукционные правила, используемые машинами логического вывода, в частности Apache Jena Generic Rule Reasoner [Jena, 2023], показывающий лучшие результаты по скорости работы на больших графах знаний среди сравнимых машин вывода [Sychev et al., 2021]. Но это непростая задача. В Apache Jena ограничены способы конструирования правил. Основными проблемами являются отсутствие дизъюнкции, отсутствие отрицания над группой утверждений, отсутствие возможности задания приоритета правилам, а также сложности при организации итеративных действий. Например, при преобразовании выражения, изображенного на рис. 1,а, в продукционные правила Apache Jena, получается правило, изображенное на рис. 2.

```
[
  (?C rdf:type ?var2...)
  isReachable(?var2...,rdfs:subClassOf,poas:element,"true"^^xsd:boolean)
  noValue(?C,poas:state,poas:unevaluated)
  (?var3... poas:var... "X"^^xsd:string)
  (?C poas:isInside ?var3...)
  (?C poas:state ?var5...)
  equal(?var5...,poas:evaluated)
  makeSkolem(?var1...)
  ->
  (?var1... poas:predicate1... ?C)
]
```

*Рис. 2. Пример правила Jena для вычисления выражения, изображенного на рис. 1,а*

Преобразование выражения, изображенного на рис. 1,б, потребует несколько наборов правил, объем которых в 4.3 раза превышает объем правила на рис. 2. Составление этих правил требует от разработчика четкого понимания механизма работы продукционных правил Jena и умения организовывать работу независимо выполняющихся правил. Это делают генерацию правил Jena из деревьев решений предпочтительной.

Формальное представление процесса решения задач в виде дерева решений также открывает возможности для решения более сложных задач. Например, с их помощью можно генерировать наводящие вопросы. [Sychev et al., 2023] Структура дерева проверок для конкретной задачи служит входными данными для создания конечного автомата вопросов, позволяющего в каждой конкретной ситуации ошибки учащегося эффективно определить, в чем именно была ошибка рассуждений учащегося, и объяснить ему ее причину. Возможность ведения диалога с помощью наводящих вопросов приближает интеллектуальные тренажеры к взаимодействию с преподавателем-человеком.

### 3. Апробация

С использованием разработанного фреймворка были разработаны четыре ядра ИОС для обучения следующим интеллектуальным навыкам: определение порядка вычисления операций в выражении на языках программирования, определения порядка прилагательных в словосочетании на английском языке, составление выражений доступа к данным в языках программирования, разрешение имен в программных выражениях. Последние три из них были разработаны авторами, не входящими в коллектив разработчиков фреймворка.

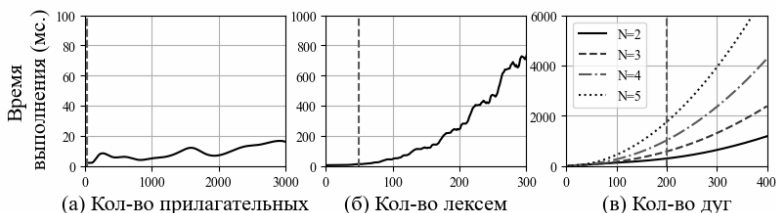


Рис. 3. Графики зависимости времени выполнения рассуждений, описанных деревом решений, от показателя сложности задачи

На первых трех предметных областях были проведены замеры скорости проверки ответа студента; ее результаты для трех предметных областей в зависимости от сложности задач представлены на рис. 3.

Вертикальная черта на каждом из графиков обозначает максимальную сложность задач, которую эксперты в областях сочли подходящей для обучения. Для навыка “определение порядка прилагательных” реалистичный размер словосочетания составляет не более 20 слов. Для определения порядка вычисления выражения объем задачи измерялся в количестве лексем в выражении (до 50 лексем). Для навыка “составление выражений доступа к данным” сложность задачи зависит от количества дуг в графе

возможных путей к данным, т.к. производительность обучающей системы в этой задаче зависит от структуры графа путей, были проведены замеры доступа к элементам массивов увеличивающейся размерности из N элементов по каждому измерению (см. рис. 3,в).

Из графика можно видеть, что на понятных человеку объемах данных в первых двух областях (см. рис. 3,а,б) интерпретатор тратит на рассуждения приемлемое количество времени, однако в области выражений доступа (см. рис. 3,в), на решение реальной по объему задачи, может уходить до 2-х секунд. Соответственно, интерпретатор, в целом, позволяет разрабатывать эффективные тренажеры с пошаговой оценкой данных, но в некоторых предметных областях он не является масштабируемыми для более сложных задач. Поэтому возможность автоматического преобразования логических выражений в продукционные правила современных машин вывода является перспективным направлением дальнейших исследований.

Для определения порядка операторов в выражении было разработано дерево проверок, состоящее из 62 узла и 41 логического выражения. Результирующая ИОС может использовать 74 наводящих вопроса, каждый из которых распознает одну уникальную ошибку в мышлении учащегося. Разработка интеллектуального ядра тренажера с помощью дерева решений и фреймворка заняла около 20 человеко-часов. Ручное написание продукционных правил Apache Jena Generic Rule Reasoner для этой задачи включало составление 230 правил и потребовало около 150 человеко-часов. Это показывает, что фреймворк позволяет значительно снизить трудозатраты на разработку моделей предметных областей.

Оценки времени, затраченного на разработку модели предметной области тремя различными разработчиками, не являвшимися авторами фреймворка, представлены в табл. 1.

Таблица 1

Предметная область	Кол-во узлов в дереве решений	Кол-во логических выражений	Примерное время разработки, ч.
Порядок прилагательных	47	21	10
Выражения доступа	25	11	20
Разрешение имен	136	72	50

Разработанное ядро предметной области порядка прилагательных было интегрировано в ИОС “Как это работает” (HowItWorks) [Sychev et al., 2022]. Для областей выражений доступа и разрешения имен были созданы

онлайн-тренажеры, что потребовало разработки интерфейса пользователя для ввода и решения задач. В ходе дальнейших исследований также планируется интеграция этих тренажеров в ИОС “Как это работает”.

## Заключение

Мы предложили новый подход к представлению предметных областей ИОС на основе деревьев решений, позволяющий решить серию задач, необходимых при разработке интеллектуальных обучающих систем, при существенном снижении трудоемкости разработки. Данный подход применим к разработке ИОС в которых процесс проверки ответа обучаемого можно представить в формальном виде, главным образом к хорошо определенным предметным областям (формальные языки, математические и технические науки, некоторые технические языки). Возможность применения этого подхода к задачам с менее четкими критериями является одним из направлений дальнейших исследований.

С использованием данного подхода был разработан фреймворк для создания ИОС на базе деревьев решений. Данный фреймворк решает задачи выполнения необходимых рассуждений на базе знаний, а также генерации наводящих вопросов в разработанной ИОС. Однако используемый на данный момент в качестве машины вывода интерпретатор не масштабируем для работы с большими базами знаний. Преобразование разработанных логических выражений в продукционные правила Apache Jena или других машин вывода, таких как SWI-Prolog [Wielemaker et al., 2011], Answer Set Programming (ASP) [Gebser et al., 2018], Pellet [Sirin et al., 2007] и SPARQL [SPARQL, 2008], является приоритетным направлением дальнейших исследований.

Другим направлением дальнейших исследований является расширение функциональных возможностей фреймворка – например, генерация интеллектуальных подсказок, преобразование деревьев проверок в деревья вычислений и наоборот, а также определение списка задач, которые необходимо проверить для оценки усвоения предметной области.

## Список литературы

- [Зверев, 2007] Зверев Г.Н. К обобщенной теории обработки наблюдений // Нефтепромысловая геофизика. – М.: ИГ и РГИ, 2007.
- [Зверев и др., 2008] Зверев Г.Н., Петров И.И. Оценка надежности и оптимизация качественной интерпретации // Техника и технология геофизических исследований скважин. – Уфа: БНИПИ, 2008.
- [Kowalsky, 1975] Kowalsky R. A proof procedure using connection graphs // Journal of the ACM. – 1975. – No. 22 (4).
- [Elliot, 1984] Elliot H. and Goodwin G.C. Adaptive implementation of the internal model principle // Proc. 23rd IEEE Conference on Decision and Control, USA. – 1984.



- [FIPA, 1997] FIPA'97, Specification Part 2. – <http://drogo.cselst.stet.it/uvf/fipa712.zip>.
- [Завидов, 2020] Завидов Б.Д. Менеджмент в РФ [Электронный ресурс] // Электрон. жур. – 2020. – № 12. – URL: <http://www.meviz.ru/articles/2020/1432.htm> (дата обращения: 01.04.2023).
- [Nwana, 1990] Nwana H. Intelligent tutoring systems: an overview // *Artificial Intelligence Review*. – 1990. – No. 4 (4).
- [Alkhatlan et al., 2018] Alkhatlan A. and Kalita J.K. Intelligent Tutoring Systems: A Comprehensive Historical Survey with Recent Developments // *International Journal of Computer Applications*. – 2018. – No. 181 (43).
- [Freedman et al., 2000] Freedman R., Ali S.S. and McRoy S. Links: what is an intelligent tutoring system? // *Intelligence*. – 2000. – No. 11 (3).
- [Singh et al., 2022] Singh S. and Singh V. An Architecture of Domain Independent and Extensible Intelligent Tutoring System based on Concept Dependencies and Subject Paths // *International Journal of Advanced Computer Science and Applications*. – 2022. – No. 13 (5).
- [Рыбина, 2023] Рыбина Г.В. Интеллектуальная технология построения обучающих интегрированных экспертных систем: новые возможности // *Открытое образование*. – 2017. – No. 21 (4).
- [Turner et al., 2005] Turner T.E., Macasek M.A., Nuzzo-Jones G., Heffernan N.T. and Koedinger K. The Assistent Builder: A Rapid Development Tool for ITS // *Proceedings of the 12th Artificial Intelligence In Education*. – 2005.
- [O'Rourke et al., 2015] O'Rourke E., Andersen E., Gulwani S. and Popović Z.A. Framework for Automatically Generating Interactive Instructional Scaffolding // *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 2015.
- [Aleven et al., 2006] Aleven V., McLaren B.M., Sewall J. and Koedinger K.R. The Cognitive Tutor Authoring Tools (CTAT): Preliminary Evaluation of Efficiency Gains // *In the Proceedings of the 8th International Conference on Intelligent Tutoring Systems*. – 2006.
- [Martin et al., 2003] Martin B. and Mitrovic A. ITS Domain Modeling: Art or Science? // *In the Proceedings of the 11th International Conference Artificial Intelligence in Education*. – 2003.
- [Mitrovic et al., 2007] Mitrovic A., Brent M. and Pramuditha S. Intelligent Tutors for All: The Constraint-Based Approach // *Intelligent Systems, IEEE*. – 2007. – No. 22 (4).
- [Sychev et al., 2021] Sychev O.A., Anikin A. and Denisov M. Inference Engines Performance in Reasoning Tasks for Intelligent Tutoring Systems // *International Conference on Computational Science and Its Applications*. – 2021.
- [Soofi et al., 2019] Soofi A.A. and Ahmed M.U. A Systematic Review of Domains, Techniques, Delivery Modes and Validation Methods for Intelligent Tutoring Systems // *International Journal of Advanced Computer Science and Applications*. – 2019. – No. 10 (3).
- [Mousavinasab et al., 2018] Mousavinasab E., Zarifasanaiey N., Niakan Kalhori S.R., Rakhshan M., Keikha L. and Ghazi Saedi M. Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods // *Interactive Learning Environments*. – 2018. – No. 29 (1).

- [**diagrams.net, 2021**] diagrams.net, previously draw.io, is an online diagramming web site that delivers the source in this project. [Электронный ресурс]. – URL: <https://app.diagrams.net/> (дата обращения: 27.05.2023).
- [**Pasternak et al., 2017**] Pasternak E., Fenichel R. and Marshall A.N. Tips for creating a block language with blockly // IEEE Blocks and Beyond Workshop. – 2017.
- [**Sychev et al., 2023**] Sychev O. and Gumerov M. Generating Pedagogical Questions to Help Students Learn // International Conference on Intelligent Tutoring Systems. – 2023.
- [**Sychev et al., 2022**] Sychev, O., Penskoj, N., Terekhov, G. Using Software Reasoning to Determine Domain-law Violations and Provide Explanatory Feedback: Expressions Tutor Example // Proceedings of the International Conference on Computer Supported Education. – 2022.
- [**Wielemaker et al., 2011**] Wielemaker J., Schrijvers T., Triska M. and Lager T. SWI-Prolog // Theory and Practice of Logic Programming. – 2011. – No. 12 (1-2).
- [**Gebser et al., 2018**] Gebser M., Kaminski R., Kaufmann B. and Schaub T. Multi-shot ASP solving with clingo // Theory and Practice of Logic Programming. – 2018. – No. 19 (1).
- [**Sirin et al., 2007**] Sirin E., Parsia B., Grau B.C., Kalyanpur A. and Katz Y. Pellet: A practical OWL-DL reasoner // Web Semantics: Science, Services and Agents on the World Wide Web. 2007. № 5(2).
- [**Jena, 2023**] Jena – a free and open source Java framework for building Semantic Web and Linked Data applications. [Электронный ресурс]. – URL: <https://jena.apache.org> (дата обращения: 29.05.2023).
- [**SPARQL, 2008**] SPARQL Update. A language for updating RDF graphs. W3C Member Submission 15 July 2008. [Электронный ресурс]. – URL: <https://www.w3.org/Submission/SPARQL-Update/> (дата обращения: 29.05.2023).

## БАЗА ЗНАНИЙ ЭКСПЕРТНОЙ СИСТЕМЫ ДЛЯ АНАЛИЗА SQL-КОДА ПРОМЫШЛЕННЫХ БАЗ ДАННЫХ\*

В.С. Мошкин (*v.moshkin@ulstu.ru*)

М.М. Бажутин (*myodo@mail.ru*)

Н.Г. Ярушкина (*jng@ulstu.ru*)

Ульяновский государственный технический университет,  
Ульяновск

В статье описан подход к повышению эффективности запросов на языке SQL с целью оптимизации времени выборки и обработки данных из высоконагруженных СУБД. Предлагаемый подход предполагает формирование базы знаний экспертной системы в виде наборов правил, позволяющих оптимизировать время выборки данных из реляционных баз данных. Набор правил, повышающих эффективность обработки данных из БД, можно разделить на наборы, решающие следующие проблемы: отсутствие профилирования, проблемы с производительностью, вызванные совмещением SQL и алгоритмической части задачи, структура выполнения SQL. Также в работе представлены результаты экспериментов по обработке данных информационной системы Филиала ПАО «Ил» – Авиастар.

**Ключевые слова:** база знаний, экспертная система, SQL, базы данных, СТЕ, оптимизация.

### Введение

Проблема эффективного хранения, вывода и обработки больших объемов данных производственных предприятий в настоящее время актуальна в связи с необходимостью ускорения производственного процесса за счет концепции управления данными. Таким образом, возникает задача

---

\* Исследование выполнено при поддержке Российского научного фонда (проект № 23-71-01101 «Разработка моделей и методов повышения производительности хранилищ данных посредством предиктивного анализа темпоральной диагностической информации»).

оптимизации времени отклика базы данных путем настройки методологии, программирования или комплектующих с учетом имеющихся ресурсов [Ioannidis, 1996], [Nosova et al., 2019], [Jarke et al., 1984].

Рассмотрим типичные примеры неэффективных методов и проверенных способов поиска оптимального решения типовых задач:

1. Нет профилирования. Разработчики редко подробно анализируют реализованный алгоритм с точки зрения его потребления ресурсов. Соответственно, они не понимают, какие операции, почему и как потребляют ресурсы и не думают о том, как этого избежать [Mishra et al., 2015], [Mehenni, 2015], [Мошкин и др., 2014].

Для решения данной задачи необходимо:

- Оказывать разработчикам систематическую методическую помощь в эффективной организации профилирования.
- Выступать за использование передовых методов профилирования, демонстрируя, как алгоритмы могут быть более эффективными и упростить решение общих проблем.
- Регулировать использование инструментов и методов профилирования для поиска оптимального решения задач на основе профилирования [Misaki et al., 2017], [Darabant et al., 2017], [Yang et al., 2017].

2. Проблема технологичности совмещения SQL и алгоритмической части задачи [Nasser et al., 2017], [Ярушкина и др., 2013], [Al-Sanhani, et al., 2017].

Это проблема планирования на начальном этапе решения задачи. Недостаток заключается в отсутствии планирования работы алгоритма в экстремальных режимах и на больших наборах данных.

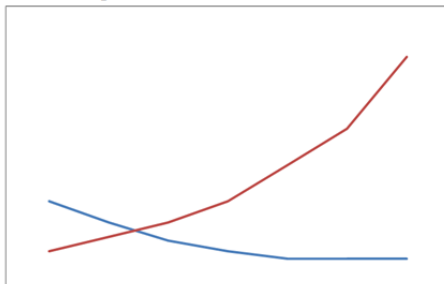
Для решения задачи необходимо уменьшить количество запросов, желательно до одного для решения проблемы. Это сразу снизит нагрузку на сервер и сеть в десятки тысяч раз, повысит производительность всей системы и ускорит получение результатов пользователем.

В настоящее время существует ряд подходов к повышению производительности промышленных информационных систем (ИС) по критериям оптимизации среднего времени отклика на запросы к базе данных [Misaki et al., 2016], [Sun et al., 2016], [Liu et al., 2016]. Это:

- Повышение производительности оборудования.

Обычно это наиболее технологически простой способ. Однако существенное изменение, как правило, можно получить только на начальных этапах жизненного цикла системы. Основная причина заключается в уменьшении относительного времени отклика при росте производительности оборудования [Lakshmi et al., 2016], [Abbas, 2016], [Pourhaji et al., 2015].

Особенно выраженный эффект такого рода можно наблюдать на растущих базах данных, что характерно для промышленных систем. Соответствие аппаратных затрат и среднего времени отклика информационной системы показано на рис. 1.



*Рис. 1. Соответствие аппаратных затрат и среднего времени отклика информационной системы*

В этих случаях наблюдается увеличение среднего времени отклика, несмотря на рост затрат на оборудование [Kanungo et al., 2015], [Yarushkina et al., 2018], [Ban et al., 2015].

- Реинжиниринг информационной системы.

Обычно на начальных этапах жизненного цикла ИС многие варианты использования и функции не планируются. Поэтому система неоднократно модифицируется. Эти улучшения редко происходят по единому плану, что приводит к дисбалансу, росту стоимости владения и стоимости модификаций. Следовательно, промышленные ИС требуют реинжиниринга для устранения подобных недостатков. Однако реинжиниринг связан с большими затратами, которые сопоставимы со стоимостью создания новой ИС, поэтому такой подход используется редко [Kumar et al., 2010], [Li et al., 2010], [Mafla et al., 2018].

- Совершенствование методологии создания программного обеспечения.

Одна и та же проблема может быть решена многими способами. По критерию оптимального времени отклика простой или традиционный способ решения этой задачи обычно не эффективен. При этом использование эффективных решений зачастую не требует чрезмерных трудозатрат. Требуется только совершенствование методологии. Таким образом, этот метод наиболее технологически прост в реализации, однако требует исследований, изобретений и подготовки кадров [Kamatkar et al., 2018], [Waluyo et al., 2005], [Khan et al., 2013].

## 1. Техника повышения производительности запросов

В рамках данного исследования была разработана экспертная система, которая позволяет решить представленные выше проблемы путем анализа SQL-кода промышленных баз данных и вывода рекомендаций по его корректировке.

Аналитическое представление зависимости времени выполнения от количества данных для традиционного подхода линейное:

$$f_1(n) = n \cdot x,$$

где

$n$  – количество обработанных строк данных,

$x$  – среднее время отклика.

Следовательно, чем больше у нас набор данных, тем дольше время выполнения. Время обработки одной строки данных не изменилось.

Аналогично, аналитическое представление зависимости времени выполнения от количества данных для нового подхода представляет собой показательное уравнение:

$$f_2(n) = \frac{1}{e^n} x + x$$

Время выполнения также увеличивается с увеличением количества обрабатываемых строк. Однако относительное время обработки одной строки экспоненциально уменьшается с увеличением количества обрабатываемых строк данных. Следовательно, чем больше строк данных обрабатывается, тем быстрее обрабатывается каждая строка. Чем больше набор данных, тем больше выигрыш.

Граф зависимостей использования ресурсов при традиционном подходе к решению задачи показан на рис. 2.



Рис. 2. График зависимости использования ресурсов при традиционном подходе к решению задачи

База знаний экспертной системы, предложенной в рамках данной работы, включает в себя набор правил, обеспечивающих значительное повышение эффективности реализации алгоритма обработки данных, извлеченных из реляционных баз данных.

Набор правил, повышающих эффективность обработки данных из базы данных, может быть представлен в соответствии со следующей информационной моделью, которая предполагает разделение правил на наборы в зависимости от решаемых ими задач:

$$R = \{R_{prof}, R_{alg}, R_{sql}\},$$

где

- $R_{prof}$  – набор правил, решающих проблему отсутствия профилирования.

Эта проблема может быть решена путем детального анализа алгоритма извлечения и обработки данных из базы данных по критериям потребления ресурсов в разрезе операций.

- $R_{alg}$  – набор правил, решающих проблему производительности, вызванную совмещением SQL и алгоритмической части задачи.

Проблема заключается в отсутствии планирования работы алгоритма в экстремальных режимах, как правило, на больших наборах данных. Эта проблема решается модификацией алгоритма обработки данных за счет уменьшения количества запросов (удаления набора запросов в цикле) и реинжиниринга SQL-запроса с оператором select. Эту манипуляцию можно осуществить с помощью механизма CTE.

Почти все современные системы управления базами данных (СУБД) предлагают механизм структурирования сложного SQL-выражения выбора, который называется Common Table Expressions (CTE) [Oracle DBMS] (рис. 3).

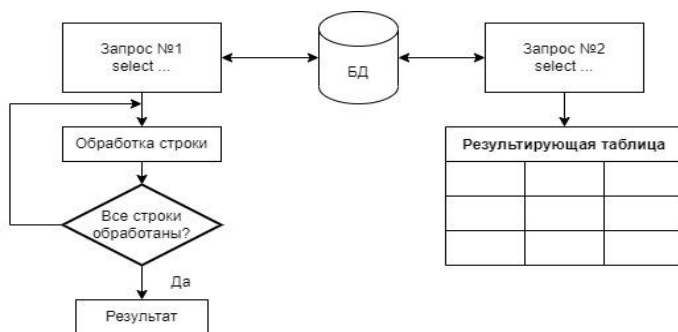


Рис. 3. Разница между традиционным (слева) и CTE (справа) подходами

- $R_{sql}$  – набор правил, решающих проблему структуры выполнения SQL. В свою очередь, этот набор правил включает в себя два подмножества, обеспечивающие решение задач соответствующих типов:

$$R_{sql} = \{R_q, R_{plan}\},$$

где

- $R_q$  – набор правил, решающих проблему варианта построения запроса.

Традиционно SQL-запросы пишутся без учета избыточных и неэффективных операций. В высоконагруженной системе большинство запросов выполняются много раз в день, стабильно потребляя ресурсы и замедляя работу информационной системы в целом. Поэтому эффективность потребления ресурсов необходимо планировать на начальном этапе создания каждого запроса [Zarubin et al., 2022].

Типичные задачи такого рода и методы их решения представлены в табл. 1.

Таблица 1

### Типичные проблемы и их решения

№	Проблема	Описание	Возможные решения, правила
1	Поиск соответствия по большому набору данных	Чем больше таблица, тем больше ресурсов нужно для выборки из нее, тем дольше будет работать поиск.	Выполнение выборки не по полному набору, а по небольшой его части, содержащей все нужные нам строки, ограниченной по индексируемому полю и/или разделу (partition).
2	Последовательность фильтрации и объединения (join)	Операция объединения для большого набора данных потребляет много ресурсов.	Объединение необходимо выполнять после фильтрации.
3	Оператор like	Оператор like требует больших ресурсов.	<ul style="list-style-type: none"> <li>– Применять поиск по точному совпадению;</li> <li>– Использовать регулярные выражения (regex);</li> <li>– Использовать оператор like при поиске по индексированному полю с начала строки (например, like '12345%')</li> </ul>
4	Избыточные данные	Извлечение данных ненужных полей.	Исключить извлечение данных ненужных полей. Особенно актуально для полей типа LOB.



5	Взятие максимума и минимума	Операции получения минимума и максимума требуют больших ресурсов.	Заменить операции минимума и максимума сортировкой с извлечением первого элемента. <i>with t1 as (select 123 a from dual order by a desc) select a from t1 where rownum=1</i>
6	Подсчет строк по группам (функция count)	Операция получения количества строк требуют больших ресурсов в задачах подсчета по группам.	Заменить операцию получения количества строк (count) операцией row_number и сортировкой. <i>with t1 as (select name, row_number() over (partition by name order by null) ord from table1), t2 as (select name, ord, row_number() over (partition by name order by ord desc) ord2 from t1) select name, ord from t2 where ord2=1</i>
7	Сокращение числа операций	При работе с битовой информацией использование ресурсоемких преобразований данных (например, hex-bin-string).	Снизить число преобразований (в том числе и к формату строки), использовать битовую арифметику для битовых масок.
8	Join вместо subselect	Subselect предполагает последовательность проведения операций.	Заменить subselect на join, наиболее широко используемой возможности аппаратной и программной параллелизации.
9	Обоснованное использование индекса	Объединение таблиц не по индексу.	При использовании операции join и фильтрации нужно максимально использовать существующие индексы с явным указанием объединения по ним в условиях для join.
10	Необоснованное использование индекса	Сканирование таблиц по индексу неэффективно при извлечении более 50% данных таблицы.	Когда необходимо извлечь более данных 50% таблицы, эффективным является операция сканирования полным доступом.

- *Rplan* – набор правил, решающих проблемы с режимом работы планировщика запросов.

SQL – это не алгоритм. Алгоритм строится планировщиком запросов на основе того, что необходимо получить. Сам планировщик запросов представляет собой алгоритм со своими особенностями, ограничениями и ошибками [ISO/IEC 9075:1999]. Типичные задачи такого рода и правила их решения представлены в табл. 2.

Таблица 2

### Типичные проблемы и их решения

№	Проблема	Описание	Возможные решения, правила
1	Предел сложности запроса	При достижении некоторого уровня сложности планировщик запросов теряет способность эффективно планировать исполнение запроса.	Хранение в памяти промежуточных результатов запросов (использование <i>global temporary table</i> или <i>hint materialize</i> ) [Yang, 2017].
2	Кэширование на уровне запроса	Кэширование эффективно далеко не всегда, потому обычно в больших базах кэширование всего не применяется, ибо не дает выигрыша в быстродействии, а иногда и негативно на нем отражается	Применение или неприменение принудительного кэширования в зависимости от характера задачи. Например: <i>select /*+ RESULT_CACHE */ * from dual</i>
3	Режим работы планировщика	Неоптимальный режим работы планировщика.	Методом, аналогичным принудительному включению кэширования на уровне запроса, можно менять подход планировщика запросов к построению алгоритма исполнения. Если в сложном запросе нужно заставить планировщик использовать индексы, <i>rule</i> – простейший вариант это сделать. Подход <i>choose</i> использует подход, основанный на статистике, если статистика есть, и <i>rule</i> , если ее нет.
4	Точная настройка	Отсутствие дополнительной настройки для оптимизации времени запроса.	С помощью подсказок ( <i>hints</i> ) можно явным образом обязать планировщик в определенном элементе <i>select</i> использовать определенный индекс, метод доступа к таблице и др.

## 2. Эксперименты

В рамках данного исследования была разработана экспертная система, которая позволяет решить задачи повышения эффективности работы СУБД путем анализа SQL-кода промышленных баз данных и вывода рекомендаций по его корректировке.

В качестве объекта применения предложенных методов использовалась база данных информационной системы Филиала ПАО «Ил» – «Авиастар». В этой базе данных хранятся данные для построения комплекта технологической документации и отображения трехмерной геометрии агрегата и оборудования самолета Ил-76-МД-90А. Эта база данных построена на СУБД Oracle [Oracle DBMS].

Интерпретатор SQL в СУБД Oracle называется «планировщик запросов». Планировщик запросов Oracle преобразует код SQL, написанный разработчиком, в определенный алгоритм, как и любой интерпретатор с функциями оптимизации.

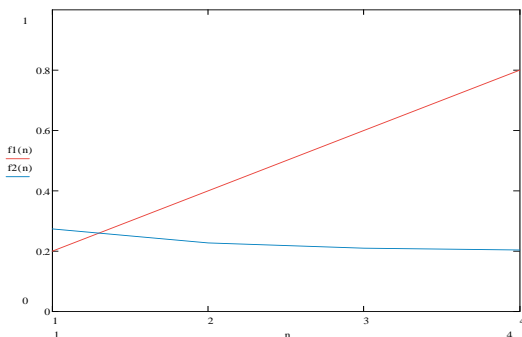
Полученный алгоритм выполняется машиной с процессором x86 и соответствует программно-аппаратному распараллеливанию выполнения.

Преобразование кода SQL в алгоритм также представляет собой алгоритм, аналогичный алгоритму компиляции. Этот алгоритм выполняет сложную задачу преобразования неалгоритмического кода, описывающего результат, в оптимальный алгоритм. Поэтому алгоритм планировщика не в состоянии построить оптимальный план выполнения некоторых запросов, особенно при увеличении их сложности. Этот факт подтверждается экспериментально.

По условиям задачи необходимо контролировать 10 000 строк данных по 24 критериям в каждой. При традиционном подходе для этого требуется 240001 запросов. Если каждый из этих запросов будет выполняться за 0,2 секунды, то весь алгоритм будет выполняться более 13 часов, обеспечивая при этом постоянную высокую нагрузку на сервер и сеть, что неизбежно замедлит работу всех остальных.

Применение предложенного подхода позволило сократить количество запросов до одного. Кроме того, значительно сократился объем кода и, соответственно, время разработки и отладки сократилось в 5-10 раз.

Сравнение аналитического представления зависимости времени выполнения от количества данных по традиционной схеме реализации задачи и предложенному подходу показано на рис. 4.



*Рис. 4. Сравнение аналитического представления зависимости времени выполнения запроса от количества обрабатываемых строк по традиционной схеме и предложенному подходу*

Как видно из графика, нагрузка на сервер и сеть радикально снижена.

### **Заключение**

Таким образом, в данном исследовании обсуждались вопросы повышения эффективности извлечения и обработки данных из высоконагруженных промышленных баз данных. Эта цель может быть достигнута за счет оптимизации времени отклика базы данных путем модификации методологической, программной или аппаратной составляющей с учетом задействованных ресурсов. Реинжиниринг ИС, как и увеличение аппаратной мощности, являются универсальными методами, которые можно применять к информационным системам из разных областей и с любым объемом данных. Корректировка методики построения запросов и управления работой планировщика запросов помогает значительно снизить затраты ресурсов при сохранении того же аппаратного и программного контекста.

В рамках исследования была разработана база знаний экспертной системы, включающая стек правил использования и модификации SQL-запросов и работы планировщика, благодаря чему удалось снизить нагрузку на сервер и время на извлечение и обработку данных из СУБД большого промышленного предприятия в сотни тысяч раз.

Следующими шагами в этом проекте является расширение набора правил с возможностью маневрирования параметрами загрузки сервера баз данных и их адаптация к реляционным СУБД, отличным от Oracle.

## Список литературы

- [Ioannidis, 1996] Ioannidis Y.E. Query optimization // ACM Comput. Surv. – 1996. – 28, 1. – P. 121-123. – <https://doi.org/10.1145/234313.234367>.
- [Nosova et al., 2019] Nosova T., Kalugina O., Barankova I. Increase the effectiveness of search databases queries using the algorithm of bitmap scales // Journal of Physics: Conference Series. 1210. 012109. – 2019. – 10.1088/1742-6596/1210/1/012109.
- [Jarke et al., 1984] Jarke M., Koch J. Query Optimization in Database Systems // ACM Comput. Surv. – 1984. – 16, 2. – P. 111-152. – <https://doi.org/10.1145/356924.356928>.
- [Mishra et al., 2015] Mishra Richa & Singh Sanjeev & Singh P. & Singh R. Dynamic load balancing using buffer management in distributed database environment. – 2015. – P. 787-790. – 10.1109/ICACEA.2015.7164810.
- [Mehenni, 2015] Mehenni Tahar. Integration of useful links in distributed databases using decision tree classification. – 2015. – P. 5-9. – 10.1109/ISEI.2015.7358717.
- [Мошкин и др., 2014] Мошкин В.С., Ярушкина Н.Г. Система онтологического анализа временных рядов // Автоматизация процессов управления. – 2014. – № 2 (36). – С. 78-85.
- [Misaki et al., 2017] Misaki Makoto & Tsuda Tomio & Inoue Shinji & Sato Shintaro & Kayahara Akihiro & Imai Shin-ichi. Distributed Database and Application Architecture for Big Data Solutions // IEEE Transactions on Semiconductor Manufacturing. – 2017. – P. 1-1. – 10.1109/TSM.2017.2750183.
- [Darabant et al., 2017] Darabant Adrian & Varga Viorica & Tambulea Leon. A linear approach to distributed database optimization using data reallocation. – 2017. – P. 1-6. – 10.23919/SOFTCOM.2017.8115503.
- [Yang et al., 2017] Yang Chaozhi & Cai Tingting & Li Zhihui. Research on a tunable consistency strategy of the distributed database. – 2017. – P. 533-538. – 10.1109/LCCSS.2017.8091473.
- [Nasseri et al., 2017] Nasseri, Mahnaz & Jameii, Seyed Mahdi. Concurrency control methods in distributed database: A review and comparison. – 2017. – P. 200-205. – 10.1109/COMPTELIX.2017.8003964.
- [Ярушкина и др., 2013] Ярушкина Н.Г., Мошкин В.С. Онтологический подход к анализу временных рядов // Интегрированные модели и мягкие вычисления в искусственном интеллекте. Сб. научных трудов VII-й Международной научно-практической конференции. – Коломна, 2013. – № 2. – С. 529-537.
- [Al-Sanhani et al., 2017] Al-Sanhani Asma & Hamdan Amira & Al-Thaher Ali & Al Dahoud Ali. A comparative analysis of data fragmentation in distributed database. – 2017. – P. 724-729. – 10.1109/ICITECH.2017.8079934.
- [Misaki et al., 2016] Misaki Makoto & Tsuda Tomio & Inoue Shinji & Sato Shintaro & Kayahara Akihiro & Imai Shin-ichi. Distributed database and application architecture for big data solutions. – 2016. – P. 1-4. – 10.1109/ISSM.2016.7934509.
- [Sun et al., 2016] Sun Qiao & Fu Lan-mei & Deng Bu-qiao & Sun Jiasong. An efficient transaction processing method on the distributed database. – 2016. – P. 1916-1920. – 10.1109/CISP-BMEI.2016.7853031.
- [Liu et al., 2016] Liu, ShaoHua & Xu, Xing. Distributed Database Query Based on Improved Genetic Algorithm. – 2016. – P. 348-351. – 10.1109/ICISCE.2016.84.
- [Lakshmi et al., 2016] Lakshmi Sanapala & Vatsavayi Valli Kumari. Query optimization using clustering and Genetic Algorithm for Distributed Databases. – 2016. – P. 1-8. – 10.1109/ICCCI.2016.7479934.

- [**Abbas, 2016**] Abbas Qasim & Shafiq Hammad & Ahmad Imran & Tharanidharan Sridevi. Concurrency control in distributed database system. – 2016. – P. 1-4. – 10.1109/ICCCI.2016.7479987.
- [**Pourhaji et al., 2015**] Pourhaji Soheil & Moattar Mohammad. Developing a pieces of data allocation method in distributed databases using Bayesian networks. – 2015. – P. 117-122. – 10.1109/ICTCK.2015.7582656.
- [**Kanungo et al., 2015**] Kanungo Sonal & Morena Rustom. Analysis and Comparison of Concurrency Control Techniques. IJARCCCE. – 2015. – P. 245-251. – 10.17148/IJARCCCE.2015.4360.
- [**Yarushkina et al., 2018**] Nadezhda Yarushkina, Vadim Moshkin, Aleksey Filippov Development of a knowledge base based on context analysis of external information resources // DS-ITNT 2018 // Proceedings of the International conference Information Technology and Nanotechnology. Session Data Science, Samara, Russia, 24-27 April, 2018. – P. 328-337.
- [**Ban et al., 2015**] Ban W., Lin J., Tong J. and Li S. Query Optimization of Distributed Database Based on Parallel Genetic Algorithm and Max-Min Ant System // 2015 8th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China, 2015. – P. 581-585. – DOI: 10.1109/ISCID.2015.199.
- [**Kumar et al., 2010**] Kumar T.V.V., Singh V. and Verma A.K. Generating Distributed Query Processing Plans Using Genetic Algorithm // 2010 International Conference on Data Storage and Data Engineering, Bangalore, India, 2010. – P. 173-177. – DOI: 10.1109/DSDE.2010.56.
- [**Li et al., 2010**] Li D., Han L., & Ding Y. SQL query optimization methods of relational database system // In 2010 Second International Conference on Computer Engineering and Applications. – 2010, March. – Vol. 1. – P. 557-560. IEEE.
- [**Mafla et al., 2018**] Mafla Enrique, Toapanta Moisés, Gallegos Luis, Mendoza Felix, Orizaga Trejo, Jose Antonio. Analysis of Efficient Processes for Optimization in a Distributed Database. – 2018. – P. 1-5. – 10.1109/CITS.2018.8440153.
- [**Kamatkar et al., 2018**] Kamatkar S.J., Kamble A., Viloría A., Hernández-Fernandez L., Cali E.G. Database Performance Tuning and Query Optimization / In: Tan, Y., Shi, Y., Tang, Q. (eds) // Data Mining and Big Data. DMBD 2018. Lecture Notes in Computer Science. Vol 10943. – Springer, Cham, 2018. – [https://doi.org/10.1007/978-3-319-93803-5\\_1](https://doi.org/10.1007/978-3-319-93803-5_1).
- [**Waluyo et al., 2005**] Waluyo, Agustinus Borgy, Srinivasan, Bala, and Taniar David. Research in Mobile Database Query Optimization and Processing. 1 Jan. 2005. – P. 225-252.
- [**Khan, 2013**] Majid Khan, Khan M.N.A. Exploring Query Optimization Techniques in Relational Databases // International Journal of Database Theory and Application. – 2013. – 6.3. – P. 11-20.
- [**Oracle DBMS**] Oracle DBMS technical and operational documentation. – <https://docs.oracle.com>.
- [**Zarubin et al., 2022**] Zarubin A., Moshkin V., Koval A. Features of Software Development for Data Mining of Storage System State. In: Radionov A.A., Gasiyarov V.R. (eds) Advances in Automation III. RusAutoCon 2021. Lecture Notes in Electrical Engineering. Vol 857. – Springer, Cham., 2022. – P. 489-499. – [https://doi.org/10.1007/978-3-030-94202-1\\_45](https://doi.org/10.1007/978-3-030-94202-1_45).
- [**ISO/IEC 9075:1999**] ISO/IEC 9075:1999 – fourth revision of the international standard for the SQL language.

## ОТКРЫТЫЕ ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ: ПОНЯТИЕ И ПОДХОДЫ К РАЗРАБОТКЕ\*

С.Б. Румовская (*sophiyabr@gmail.com*)

Калининградский филиал Федерального исследовательского центра «Информатика и управление» РАН, Калининград

Сложные практические проблемы отличаются субъективным и динамическим характером. Подобные проблемы никогда в точности не повторяются, поэтому метод решения необходимо конструировать каждый раз заново. В связи с этим актуальна разработка интеллектуальных систем – целенаправленных, обладающих способностью приобретать, воспроизводить и использовать знания, функции которых оцениваются с позиции обеспечения устойчивости, открытости и работы в реальном времени. В работе рассмотрено понимание интеллектуальных систем и их «открытости», а также подходы к их разработке.

**Ключевые слова:** интеллектуальная система, открытая система, подходы к разработке.

### Введение

Слабая формализация, комплексное строение, сетевой характер условий и целей, существенная непрозрачность (неопределенность), субъективность и динамичность определяют сложность решения практических проблем. Проблема никогда в точности не повторяется, поэтому методы, выработанные однажды для ее решения, со временем становятся нерелевантными. В результате возникает необходимость выработки новых методов или изменения старых с учетом новой информации, знаний и опыта. Всё это требует новых решений, разработки целенаправленных систем, обладающих способностью приобретать, воспроизводить и использовать знания – интеллектуальных систем (ИС). Такие ИС должны иметь развитые возможности и средства адаптации к изменениям среды, в том числе путем модификации своей структуры и параметров [Тарасов, 1998], т.е. быть «открытыми». Это предполагает следующее: взаимодействие системы с окружающей средой; наличие мотивации; использование знаний для синтеза цели и оценки;

---

\* Работа выполнена при финансовой поддержке РФФ (проект № 23-21-00218).

принятие решения и выработка управления; контроль реальных результатов управления и сопоставление их с прогнозированными результатами действия [Пупков, 2008]. В работе рассмотрены: понятие ИС, интерпретация её «открытости», а также подходы к разработке.

## 1. Понятие интеллектуальной системы и её «открытости»

Общепринятое определение интеллектуальной системы (от лат. *intellectus* – ум, рассудок) отсутствует. Однако превалирует следующее понимание ИС [Самодумкин и др., 2014; с. 13] – это «...объединенная информационным процессом совокупность технических средств и программного обеспечения (ПО), работающая во взаимодействии с человеком (коллективом людей) или автономно, способная на основании сведений об окружающей среде и собственном состоянии при наличии знаний и мотивации синтезировать цель, принимать решение о действии и находить рациональные способы достижения цели». В технологическом аспекте вычислительные части ИС должны гибко обрабатывать информацию о реальном мире подобно человеку, поскольку многие задачи этого мира плохо определены [Пупков, 2008]. Функции ИС необходимо оценивать с точки зрения обеспечения устойчивости, открытости и работы в реальном времени. Источник возмущающего воздействия для ИС – окружающая среда, а функционирование системы должно компенсировать это воздействие при достижении цели. Так в [Тихонов и др., 2010] ИС определили, как адаптивную систему, позволяющую строить программы целесообразной деятельности по решению поставленных задач на основании конкретной ситуации. Адаптивность здесь – это сохранение работоспособности при непредвиденных изменениях свойств объекта управления (ОУ), целей управления или окружающей среды путем смены алгоритма функционирования, поведения или поиска оптимальных (или эффективных) решений и состояний. Различают самоадаптирующиеся, самообучающиеся и самоорганизующиеся системы.

Для практических исследований наиболее технологичным является определение ИС В.К. Финна [Финн, 2023], как средства имитации и усиления умственной активности человека. К интеллектуальным способностям человека относятся: способности к рассуждению и порождению гипотез, к выдвиганию цели и рефлексии, к поиску объяснений и коррекции знаний в связи с изменением имеющихся ситуаций и другие (всего их тринадцать). Архитектура таких ИС (допускает вариации) следующая: ИС = Решатель задач + Информационная среда + Интеллектуальный интерфейс (ИНИ). Решатель задач = Рассуждатель (Р) + Вычислитель (В) + Синтезатор (С). Информационная среда = База фактов + База знаний (БЗ). ИНИ включает представление результатов, диалог на естественном



языке и научение работе с ИС. (P) реализует правдоподобные и достоверные рассуждения и синтез познавательных процедур посредством когнитивных рассуждений. (B) и (C) – вычислительные процедуры и комбинирование рассуждений и вычислений, а также выбор стратегий решения.

Таким образом, открытость ИС [Ruzic, 2013], [Остроух, 2015], [Тарасов, 1998] проявляется: *в развитых коммуникативных способностях* для связей между системой и средой; *в самоорганизации* – способность на основе сформированной или заданной цели определять методы, пути и средства достижения конечного результата; *в способности к развитию и самообучению* – аппроксимация, обобщение, концептуализация, извлечение знаний из накопленного опыта, а также получение и использование информации, которая выводится из имеющихся в базе данных (увеличивает мобильность и гибкость системы); *в самонастройке* параметров и структуры, исходя из изменений в модели предметной области или параметров среды (возможно наличие модели самой себя – позволяет определять границы своей компетентности); *в информационной открытости* – способность расширять модели мира и других систем и представлять события, которые могут повлиять на систему.

При построении ИС для решения практических проблем обычно требуется объединять в составе системы и согласовывать работу моделей, созданных различными командами разработчиков. Потому ИС должны быть открытыми и как прикладное окружение [Бойченко и др., 2004], основанное на интерфейсных стандартах, обеспечивающих: переносимость ПО, информации и пользователей с одной платформы на другую; масштабируемость; стандартизуемость; способность к интеграции; высокую отказоустойчивость и интероперабельность.

## **2. Подходы к разработке открытых интеллектуальных систем**

Подходы к разработке открытых интеллектуальных систем (ОИС) сочетают методы различных направлений искусственного интеллекта и универсальные подходы [Загоруйко, 2020] создания ПО и ИС.

### **2.1. Универсальные подходы**

*По Power D.J.* выделяется три основных подхода: обеспечивающий полный жизненный цикл системы, быстрое прототипирование и подход конечного пользователя, который сам выбирает средства для решения задач. *Гибкая разработка ПО* – итеративная разработка, динамическое формирование требований, тесное взаимодействие разработчиков различного профиля, использование гибких, легких сервисов, позиционируемых как отдельные приложения и интегрируемых между собой. *Каркасная разработка* ИС, обеспечивающая её базовую функциональность и под-

держивающая вариативность поведения пользователя через точки расширения. *Сервис-ориентированный подход* – используются распределённые, слабо связанные, заменяемые компоненты (сервисы), которые могут принадлежать доменам различных владельцев.

В рамках *онтологического подхода* была предложена *технология OSTIS компонентного и платформенно-независимого проектирования семантически совместимых ИС* [Голенков и др., 2021], имеющих БЗ любого уровня сложности. В основе лежат формальные онтологии логико-семантических моделей ИС, средств и методов их проектирования, а также проектных действий. Используется модель асинхронного управления знаниями – все процессы их обработки неким набором агентов инициируются соответствующими ситуациями или событиями. Последние возникают в семантической памяти в ходе обработки БЗ. OSTIS обеспечивает высокий уровень информационной и функциональной гибкости и позволяет разрабатывать совместимые, платформенно независимые, адаптивные, обучаемые ИС.

В [Харин и др., 2014] предложена *концепция и принципы функционирования универсальной открытой многоагентной среды (МС)*. В рамках МС смогут существовать и взаимодействовать различные многоагентные системы (МАС), построенные на основе разных многоагентных платформ. Это реализуется с помощью коммуникаций программных объектов подобного общению в реальном времени между людьми с помощью XMPP-протокола и конструирования организаций с использованием HTTP-протокола на основе одной или более организационных единиц, которые строятся с помощью соответствующих XMPP-сущностей. Для решения текущих задач агенты и сервисы объединяются, взаимодействуя между собой. В рамках этой архитектуры разработаны базовые компоненты платформы eSocialGRIDs, в основе которой лежат кластеры Jabber-серверов, использующие XMPP протокол, и могут быть организованы в «облако». В этих кластерах коммуникация между абстрактными агентами и другими объектами может, кроме XMPP-сообщений, производиться и посредством распределенной базы данных.

## 2.2. Интеллектуальные системы управления

Открытые интеллектуальные системы управления (ИСУ) предлагается разрабатывать [Вагин и др., 2001] в соответствии с пятью принципами: 1) непрерывное взаимодействие с реальным внешним миром с использованием информационных каналов связи – ИСУ получают из него информацию в виде извлеченных знаний и оказывают на него целенаправленное воздействие; 2) самонастройка, самоорганизация и самообучение; 3) механизмы прогноза изменения внешнего мира и собственного поведения сис-

темы в динамически меняющемся внешнем мире; 4) многоуровневая иерархическая структура по IPDI (Increasing Precision with Decreasing Intelligence) – неточность знаний об объекте или его поведении компенсируют увеличением числа уровней интеллектуальности или использованием механизмов принятия решений в условиях неопределенности; 5) постоянство функционирования при разрыве связей или потере управляющих воздействий.

На практике эти принципы выполняются частично. Например, в [Чертовской, 2017] предлагается *адаптивная оптимальная трехуровневая модель*, отличающаяся системным характером описания процесса планирования в автоматизированной ИСУ производством. Здесь согласованы и объединены метод генерации данных в технологической цепочке и верхнем уровне системы, а также методы согласования экономических интересов структурных элементов (возможно переключение между согласованным и несогласованным режимами по экономическим интересам). У [Бейнарович, 2010] описаны принципы построения и методы расчета *самонастраивающихся систем (СнС) автоматического управления с эталонной моделью* для объектов с изменяющимся коэффициентом передачи и изменяющейся постоянной времени. [Зайцев и др., 2022] предлагают *системы переменной структуры с нелинейным бинарным алгоритмом «квадрантно-избирательного» управления* с адаптацией в особых фазовых состояниях в условиях неопределенностей управленческих моделей. За счет нелинейного сочетания различных линейных структур организован скользящий режим перехода от движения, соответствующего одной структуре, к движению, соответствующему другой, с помощью логического переключения связей относительно фазового состояния системы. В [Фролов и др., 2005] представлена *СнС управления*, включающая ОУ, датчик, элемент сравнения, блок зоны нечувствительности (БЗН), нейронную сеть (НС) типа многослойного персептрона, блок корректировки весовых коэффициентов (БКВК). Выход с БЗН подается на вход БКВК, в котором происходит корректировка весовых коэффициентов НС (на вход НС в дискретный момент времени подается сигнал ошибки регулирования, который преобразуется сетью в управляющий сигнал). В [Константинов и др., 2001] предлагается структура *самообучающейся ИСУ без предварительного обучения*. При помощи блока-анализатора система отслеживает действия лица, принимающего решения (ЛПР), и в ходе диалога выясняет, какие переменные побудили к данному решению, как соотносятся текущие с соответствующими лингвистическими переменными и нечеткими множествами. Накапливая информацию, она формирует функции принадлежности нечетких множеств блоков ввода и вывода знаний, а также таблицу нечетких отношений, отражающую решения ЛПР. При любом несоответствии решений системы и ЛПР, система иницирует диалог для уточнения БЗ.

### 2.3. Холистический и многоагентный подходы

В [Pan et al., 2017] предлагается подход к разработке самоорганизующихся логистических систем (СЛС) как открытых ИС с децентрализованным управлением – как холоническую многоагентную систему. Здесь контроль осуществляется на индивидуальном уровне холонов. Холоны – это самоподобные сущности с конструкцией «целое-часть». Они могут рассматриваться либо как системные компоненты более высокого уровня, либо как целое, состоящее из других холонов в качестве подструктур. Каждый холон состоит из физической части, отражаемой агентом в модели. Каждый агент – самоконтролируемый.

[Скобелев, 2003] предлагает открытые МАС оперативной обработки информации для поддержки процессов принятия решений, базирующиеся на модели сети потребностей и возможностей (ПВ) предприятия и холистическом подходе. В рамках последнего предприятие декомпозируется до уровня сети отдельных автономных «физических сущностей» (станки, детали и т.д.), каждая из которых получает своих агентов ПВ. Агенты (как автономные компании) взаимодействуют между собой, идентифицируя нужные ПВ и устанавливая временные связи, обеспечивающие бронирование ресурсов под заказы. Текущая конфигурация ПВ-сети, задаваемая набором связей между агентами, отражает временный баланс интересов участников, поэтому предлагаются методы взаимодействия агентов, позволяющие динамически создавать сети и реконфигурировать их под действием изменений в среде.

В [Городецкий и др., 2007] авторами предложена интеграция МАС и систем распределенных вычислений на основе парных взаимодействий (P2P computing). Её экземпляры установлены в узлах сети поверх стандартного P2P сервиса и образуют распределенную БЗ. Прикладные агенты устанавливаются в узлах сети поверх экземпляров агентской платформы (АП). На АП есть сервис белых и желтых страниц: желтые содержат информацию о сервисах агентов экземпляра платформы в стандартной форме, в терминах общей онтологии сети агентов и об именах агентов, которые эти сервисы предоставляют; белые – информацию об адресах последних. В основе – функциональная архитектура, разработанная рабочей группой FIPA. АП может быть снабжена компонентой, отслеживающей потребности установленных на ней агентов. Она предоставляет возможность динамического формирования содержания страниц, что важно для открытой сети агентов, где агенты и узлы могут появляться и уходить в любой момент времени.

В [Машонин и др., 2016] рассмотрены принципы построения распределенной МАС для управления группировкой автономных необитаемых подводных аппаратов – использовался сетевый подход в созда-

нии мультиагентной «системы систем», согласованно взаимодействующих на базе р2р через общую шину. Командный центр не командует процессами, а использует, катализируя, прямое взаимодействие систем для согласования решений, что позволяет моделировать появление феномена нового типа – «командного интеллекта роя» у группировки.

В [Городецкий и др., 2017] предлагается концепция группового управления в системе космического наблюдения на базе принципа самоорганизации группового поведения автономного кластера спутников (в основе концепция В2В-сети [Городецкий, 2014]). Система в реальном времени распределяет задачи, планирует и составляет расписание, выполняет оперативное управление распределенным исполнением построенного расписания, а также коррекцию распределения задач и расписания их выполнения при возникновении нештатных ситуаций.

[Швецов и др., 2018] предложили многоуровневую архитектуру интеллектуального агентно-ориентированного учебного комплекса (ИАОУК) для подготовки специалистов технического профиля в рамках многоагентного и сервис-ориентированного подходов. Агенты опираются на уровень сервисов, обеспечивающий все необходимые базовые функции учебного комплекса. Агенты гетерогенные с иерархической моделью взаимодействия. Выделен метаагент (агент-координатор), координирующий распределенное решение задач агентами и межагентное взаимодействие. Гетерогенные агенты также используются и в [Колесников и др., 2014] – предложены гибридные интеллектуальные МАС, которые объединяют преимущества МАС и гибридных интеллектуальных систем, имитируя эффекты самоорганизации (явно управляется одним из агентов), конфликтов и сплоченности в коллективах специалистов, а также интегрируя разнородные технологии моделирования интеллектуальной деятельности разных специалистов.

#### **2.4. Интеллектуальные системы с активными знаниями**

В [Нестерова и др., 2012], [Гулюк, 2015] рассмотрены принципы самообучающихся экспертных систем, имеющих веб-интерфейс и доступных любому пользователю Интернета (например, развлекательные сервисы «Акинатор», «20q» и «Энди» – все ответы учитываются в последующих играх, т.к. БЗ постоянно уточняется и расширяется в процессе взаимодействия с пользователями).

В [Золотарев и др., 2010] предлагается самообучающаяся система (СС), выдающая рекомендации по технической защите информации. Опросные листы, статистика и нормативные документы поступают в блок обучения, реализованный НС. Механизм нечеткой логики используется для оценки уровня риска. Рекомендации формулируются на основе баз экспертных знаний, ранжированных угроз и обработанных нормативных актов.

В [Волобуев, 2009] предлагается СС идентификации известных данных с использованием НС с обратным распространением ошибки и неизвестных с использованием методов, основанных на теории самозарождающихся и саморазрушающихся структур, которые способны изменять свою структуру адекватно воздействию входных данных, создавая и/или уничтожая компоненты.

В [Аверченков и др., 2014] предложена концептуальная архитектура самообучающихся МАС на основе знаний из сети Интернет. Главный агент координации формирует задания для агентов и их групп, распределяет нагрузки между узлами, разрешает конфликты, реализует балансировку между стадиями обработки документов. Основной узел – агент обучения онтологий: формирует поисковое задание и запросы на анализ и оценку полученных документов и на пополнение онтологии. Концепция позволит создавать ИС с активно наращиваемыми знаниями.

В [Насыпный и др., 2008] предложен способ синтеза самообучающейся аналитической вопросно-ответной системы с извлечением знаний из текстов, в том числе содержащихся в неявном виде. Поиск основан на эквивалентных преобразованиях предложений в совокупность индексированных предикатов и словосочетаний, сформированных стохастически. В процессе самообучения формируется система классификации, БЗ предметной области текста и метазнаний для повышения эффективности семантического анализа как основы поиска.

В [Баранюк и др., 2016] представлена ИС информационной поддержки процессов создания и развития перспективных бионических технологий с модулями добавления информационных ресурсов и автоматизированного пополнения БЗ (из открытых источников). Методы информационного поиска и машинного обучения используются при решении задач автоматической классификации информационных ресурсов.

ДСМ-метод [Финн, 2023] позволяет разрабатывать ИС, в которых БФ содержит представление элементарных высказываний, используемых в качестве посылок правдоподобных рассуждений, включающих индукцию, реализующих машинное обучение. БФ является исходным материалом для машинного обучения, посредством которого извлеченные из неё знания пополняют БЗ. Возможное пополнение БЗ – порожденные гипотезы, выражающие эмпирические закономерности.

В таблице представлен анализ подходов к построению ОИС. ИС [Скобелев, 2003] должны быть построены как открытые системы, способные самостоятельно приобретать новые знания и изменять свою структуру и функции, развиваясь и адаптируясь к изменению условий во внешней среде. Взаимодействуя между собой, такие системы стремятся достичь повышения эффективности своего функционирования и для этого изме-

няются сами – в свою очередь, изменяя среду окружения для всех участников взаимодействия. Анализ работ (табл.) показал – пока что подобное понимание «открытости» выходит далеко за рамки современных систем, «открытых» с позиции какого-то одного аспекта – открытых баз знаний, открытых сетей агентов, самонастройки и т.д.

Таблица

### Анализ подходов к построению ОИС

Методы и подходы		«Открытость»	Как прикладное окружение	Самонастройка		Самообучение	Самоорганизация	
				П	С		Э-К	Э
УНВ	[Загоруйко, 2020]		√					
	[Голенков и др., 2021], [Харин и др., 2014]		√			√	√	√
ИСУ	[Чертовской, 2017]			√				
	[Бейнарович, 2010]			√				
	[Зайцев и др., 2022]				√			
	[Фролов и др., 2005]			√				
	[Константинов и др., 2001]							
Холистический и многоагентный	[Pan et al., 2017]							√
	[Скобелев, 2003]							√
	[Городецкий и др., 2007]							√
	[Машонин и др., 2016]							√
	[Городецкий и др., 2017]							√
	[Швецов и др., 2018]						√	
[Колесников и др., 2014]						√		
Системы с активными знаниями	[Нестерова и др., 2012]					√		
	[Гулюк, 2015]					√		
	[Трофимов и др., 2010]					√		
	[Золотарев и др., 2010]					√		
	[Волобуев, 2009]				√	√		
	[Аверченков и др., 2014]					√	√	
	[Насыпный и др., 2008]					√		
	[Баранок и др., 2016]					√		
[Финн, 2023]					√			

Обозначения: УНВ – универсальные, ИСУ – интеллектуальные системы управления, П – параметры, С – структура, Э-К – элемент-координатор, Э – на уровне каждого элемента

### Заключение

В работе рассмотрено понятие открытых ИС и аспекты их «открытости». Анализ работ показал, что все ИС – «открыты» частично в каждом конкретном случае, и исследования в данном направлении актуальны, в

том числе и систем с открытым составом моделей, так как при построении ИС для решения практических проблем обычно требуется объединять работу моделей от различных команд разработчиков. Дальнейшее исследование направлено на разработку нового класса ИС – рефлексивно-активных систем гетерогенных интеллектуальных агентов, агенты которых взаимно моделируют рефлексивные позиции друг друга, динамически вырабатывают тактики и стратегии своего поведения, релевантные моделям своих контрагентов, по мере необходимости в процессе решения проблем привлекают/исключают агентов от различных разработчиков. Предлагается также использовать технологию OSTIS. Разработка подобных систем позволит повысить релевантность интеллектуальных автоматизированных систем коллективам специалистов, решающих проблемы «за круглым столом», а также получить новые и верифицировать существующие знания о процессах самоорганизации малых коллективов специалистов. Исследование данных методов на примере искусственных интеллектуальных рефлексивных агентов позволит в дальнейшем адаптировать их для включения в состав системы специалиста-человека. Это, в свою очередь, позволит агентам системы самостоятельно строить модели поведения реальных специалистов, непрерывно обучаясь в процессе решения проблем.

### Список литературы

- [Аверченков и др., 2014] Аверченков А.В., Леонов Е.А. Архитектура и принципы самообучения конструкторско-технологических экспертных систем, основанных на знаниях с использованием мониторинга сети Интернет // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2014. – №. 4 (32). – С. 53-61.
- [Баранюк и др., 2016] Баранюк В.В., Смирнова О.С., Володина А.М., Блинов М.Ю. Подсистема пополнения базы знаний интеллектуальной системы информационной поддержки создания и развития перспективных бионических технологий // International Journal of Open Information Technologies. – 2016. – Т. 4, №. 11.
- [Бейнарович, 2010] Бейнарович В.А. Самонастраивающиеся системы с эталонной моделью // Доклады Томского государственного университета систем управления и радиоэлектроники. – 2010. – № 1-1 (21).
- [Вагин и др., 2001] Вагин В.Н., Еремеев А.П. Некоторые базовые принципы построения интеллектуальных систем поддержки принятия решений реального времени // Известия РАН. Теория и системы управления. – 2001. – № 6.
- [Волобуев, 2009] Волобуев Е.С. Самообучающаяся система идентификации данных, поступающих в АСУ физической защитой критического объекта: дисс. ... к.т.н.: 05.13.19. – Обнинск, 2009.
- [Голенков и др., 2021] Голенков В.В., Гулякина Н.А., Шункевич Д.В. Открытая технология онтологического проектирования, производства и эксплуатации семантически совместимых гибридных интеллектуальных компьютерных систем. – Минск: Бестпринт, 2021.



- [Городецкий, 2014] Городецкий В.И. Многоагентная самоорганизация в В2В-сетях // Тр. Всероссийского совещания по проблемам управления ИПУ. – М., 2014.
- [Городецкий и др., 2007] Городецкий В.И., Карсаев О.В., Самойлов В.В., Серебряков С. В. Открытые сети агентов // Труды СПИИРАН. Вып. 4. – СПб.: Наука, 2007.
- [Городецкий и др., 2017] Городецкий В.И., Карсаев О.В. Самоорганизация группового поведения кластера малых спутников распределенной системы наблюдения // Известия ЮФУ. Технические науки. – 2017. – № 2 (187).
- [Гулюк, 2015] Гулюк Н.В. Применение байесовского подхода в интеллектуальных мультиагентных системах // Бизнес-образование в экономике знаний. – 2015. – № 1 (1).
- [Загорулько, 2020] Загорулько Г.Б. Модель, методы и средства комплексной поддержки разработки СППР в слабоформализованных предметных областях: дисс. ... канд. тех. наук. – Новосибирск, 2020.
- [Зайцев и др., 2022] Зайцев А.В., Канушкин С.В. Системы переменной структуры при управлении дронами в условиях информационной неопределенности // Правовая информатика. – 2022. – № 2.
- [Золотарев и др., 2010] Золотарев А.В., Жукова М.Н. Разработка самообучающейся системы на прецедентах для решения задачи комплексной оценки в сфере информационной безопасности // Решетневские чтения. – 2010. – Т. 2, № 14.
- [Колесников и др., 2014] Колесников А.В., Кириков И.А., Листопад С.В. Гибридные интеллектуальные системы с самоорганизацией: координация, согласованность, спор. – М.: ИПИ РАН, 2014.
- [Константинов и др., 2001] Константинов И.С., Филатов А.Г., Касьянов Ю.В. Принципы построения интеллектуальных автоматизированных систем управления с нечетким регулированием на базе логико-лингвистических моделей представления знаний // Сборник трудов Седьмых Академических чтений РААСН «Современные проблемы строительного материаловедения». – Белгород, 2001.
- [Нестерова и др., 2012] Нестерова Е.В., Зайцева Т.В., Смородина Н.Н., Маматов Р.А., Слободок А.А. О некоторых подходах к построению самообучающихся экспертных систем // Экономика. Информатика, – 2012. – Т. 21, № 1-1 (120).
- [Машонин и др., 2016] Машошин А.И., Скобелев П.О. Применение мультиагентных технологий для управления группой автономных необитаемых подводных аппаратов // Известия ЮФУ. Технические науки. – 2016. – № 1 (174).
- [Насынный и др., 2008] Насынный В.В., Насыпная Г.А. Способ синтеза самообучающейся аналитической вопросно-ответной системы с извлечением знаний из текстов, заявка на патент №2007120344/09 от 06.08.2007.
- [Остроух, 2015] Остроух А.В. Интеллектуальные системы. – Красноярск: Научно-инновационный центр, 2015.
- [Пупков, 2008] Пупков К.А. О некоторых этапах развития теории и техники интеллектуальных систем // Вестник Российского университета дружбы народов. Серия: Инженерные исследования. – 2008. – № 4.
- [Самодумкин и др., 2014] Самодумкин С.А., Степанова М.Д., Колб Д.Г. Прикладные интеллектуальные системы: учебно-метод. пособие. – Минск: БГУИР, 2014.

- [Скобелев, 2003] Скобелев П.О. Открытые мультиагентные системы для оперативной обработки информации в процессах принятия решений: автореф. ... д.т.н.: 05.13.01. – Самара, 2003.
- [Тарасов, 1998] Тарасов В.Б. Агенты, многоагентные системы, виртуальные сообщества: стратегическое направление в информатике и искусственном интеллекте // Новости искусственного интеллекта. – 1998. – № 2.
- [Тихонов и др., 2010] Тихонов А.Н., Иванников А. Д., Соловьёв И.В., Цветков В.Я. Основы управления сложной организационно-технической системой. Информационный аспект. – М.: МаксПресс, 2010.
- [Финн, 2023] Финн В.К. Искусственный интеллект: Методология, применения, философия. № 36. – 2-е изд. испр. и доп. – М.: URSS, 2023.
- [Фролов и др., 2005] Фролов С.В., Елизаров И.А., Солуданов М.Н. Самонастраивающаяся нейросетевая система управления // Вестник Тамбовского государственного технического университета. – 2005. – Т. 11, № 4.
- [Харин и др., 2014] Харин И.Г., Яновский А.В. Открытая многоагентная среда: концепция и платформа // ИТиВС. – 2014. – № 4.
- [Чертовской, 2017] Чертовской В.Д. Математическое описание и компьютерная реализация модели адаптивной автоматизированной системы управления производством // Информационно-управляющие системы. – 2017. – № 1 (86).
- [Швецов и др., 2018] Швецов А.Н., Ржеуцкая С.Ю., Сергушичева А.П., Суконщиков А.А. Архитектура интеллектуального агентно-ориентированного учебного комплекса для подготовки специалистов технического профиля // Открытое образование. – 2018. – 22 (3). – С. 14-24.
- [Pan et al., 2017] Pan S., Trentesaux D., Sallez Y. Specifying Self-organising Logistics System: Openness, Intelligence, and Decentralised Control / In T. Borangiu, Trentesaux D., et al. (eds) // Service Orientation in Holonic and Multi-Agent Manufacturing. SOHOMA 2016. Studies in Computational Intelligence. – Springer, Cham, 2017. – No. 694.
- [Ruzic, 2013] Ruzic F. Openness of Information-Communications Systems: The Rescue Tool for Preserving Information Age Heritage // Informing Science: the International Journal of an Emerging Transdiscipline. – 2013. – No. 16.

## ИНТЕЛЛЕКТУАЛЬНАЯ ТЕХНОЛОГИЯ ПОСТРОЕНИЯ ДИНАМИЧЕСКИХ ИНТЕГРИРОВАННЫХ ЭКСПЕРТНЫХ СИСТЕМ: ОСОБЕННОСТИ ПОСТРОЕНИЯ ИМИТАЦИОННЫХ МОДЕЛЕЙ ВНЕШНЕЙ СРЕДЫ

Г.В. Рыбина (*gvrybina@yandex.ru*)

А.А. Слиньков (*sanch971@mail.ru*)

Д.Д. Белов (*dancher.00@gmail.com*)

Национальный исследовательский ядерный университет «МИФИ»,  
Москва

В работе рассматриваются вопросы, связанные с дальнейшим развитием интеллектуальной технологии построения динамических интегрированных экспертных систем (ИЭС) с акцентом на методы и средства построения имитационных моделей внешней среды как неотъемлемого этапа разработки динамических ИЭС на основе задачно-ориентированной методологии и инструментального комплекса АТ-ТЕХНОЛОГИЯ. В качестве экспериментальной проблемной области для прототипирования динамической ИЭС используется бортовой программный комплекс космического аппарата.

**Ключевые слова:** динамические интегрированные экспертные системы, интеграция, задачно-ориентированная методология, интеллектуальный планировщик, имитационное моделирование, имитационная модель, темпоральный решатель, космический аппарат.

### Введение

Инструментальный подход в программной инженерии (этому термину в нашей стране соответствовал термин «технология программирования» [Лаврищева, 2016]) основан на автоматизации процессов создания программного обеспечения (ПО) путем использования различных CASE-инструментов, фреймворков, Workbench-систем и др. типов средств поддержки процессов проектирования, разработки и сопровождения прикладных программных систем на всех этапах жизненного цикла (ЖЦ).

Поэтому исследования и разработки в области создания моделей, методов и инструментальных средств автоматизации процессов построения интеллектуальных систем различной архитектурной типологии, в частности, интегрированных экспертных систем (ИЭС) [Рыбина, 2008], всегда были направлены на решение комплекса научных и технологических проблем, связанных со снижением трудоемкости и интеллектуальной нагрузки на разработчиков, уменьшением рисков ошибочных действий, срывов сроков сдачи проектов и др. аспектами влияния человеческого фактора на разработку ПО [Рыбина, 2015].

В лаборатории «Интеллектуальные системы и технологии» кафедры «Кибернетика» НИЯУ МИФИ создана, динамически развивается и активно используется на протяжении ряда лет оригинальная задачно-ориентированная методология автоматизированного построения ПО прикладных ИЭС на всех этапах ЖЦ [Рыбина, 2008]. Принципиальная особенность этой методологии заключается в возможности концептуального и программного моделирования архитектуры разрабатываемых прикладных ИЭС (статических, динамических, обучающих) на всех уровнях конкретизации процессов интеграции в ИЭС с ориентацией на наиболее распространенные типы неформализованных (НФ) задач (диагностика, управление, проектирование, планирование, обучение), релевантных технологии систем, основанных на знаниях. Накоплен значительный опыт разработки прикладных ИЭС для этих задач и различных проблемных областей (Про) [Рыбина, 2015], [Рыбина, 2023].

Соответственно, постоянно усовершенствуется и экспериментально исследуется на практике созданный для автоматизированной поддержки задачно-ориентированной методологии уникальный инструментарий типа Workbench (комплекс (АТ-ТЕХНОЛОГИЯ), на базовой функциональной платформе которого создано несколько поколений инструментальных средств, обеспечивающих автоматизацию процессов проектирования ПО прикладных ИЭС с расширяемой функциональностью и масштабируемой архитектурой на всех этапах ЖЦ построения и сопровождения ИЭС. Описание задачно-ориентированной методологии и комплекса АТ-ТЕХНОЛОГИЯ приводится в [Рыбина, 2008], [Рыбина, 2015].

Несмотря на то, что была показана достаточно высокая эффективность автоматизированной технологии прототипирования прикладных ИЭС с использованием базовых средств комплекса АТ-ТЕХНОЛОГИЯ, акцент в исследованиях последних лет был сделан на создание методов и средств интеллектуальной поддержки процессов прототипирования ИЭС за счет разработки новых методов интеллектуального планирования [Рыбина и др., 2018] и онтологического подхода, [Rybina et al., 2021], что в целом отражает постепенный переход от автоматизации к интеллектуализации базовых процессов проектирования ПО прикладных ИЭС.

В качестве концептуальной основы интеллектуальной технологии использовалось введенное в работе [Рыбина, 2008] понятие модели интеллектуальной программной среды, основными компонентами которой являются интеллектуальный планировщик и технологическая база знаний (БЗ), которая включает: совокупность типовых проектных процедур (ТПП) и готовых решений на различных этапах прототипирования ИЭС, включая их повторное использование в виде отдельных программных и информационных компонентов (ПИК) комплекса АТ-ТЕХНОЛОГИЯ, реализующих базовую функциональность ИЭС различной архитектурной типологии (статические, динамические, обучающие и др.); прикладную онтологию типовых архитектур ИЭС, концептуально расширяющую уровни взаимодействия интеллектуального планировщика с семантически разнородными информационно-программными ресурсами технологической БЗ в виде ТПП и ПИК [Rybina et al., 2021].

Опыт использования интеллектуальной технологии для проектирования ПО динамических ИЭС показал необходимость решения проблемы многоуровневой интеграции моделей, методов и программных средств, совместно определяющих логику принятия решений в реальном времени (РВ), поскольку архитектуры динамических ИЭС, (т.е. ИЭС, использующих динамическое представление предметной области и решающих динамические задачи в РВ [Рыбина, 2015]), представляют собой существенно более сложный тип гетерогенной архитектуры, особенность которой заключается в моделировании в РВ внешней среды любого типа и включения в состав архитектуры соответствующих подсистем, адекватно отражающих (имитирующих) все процессы и законы функционирования среды. Поэтому важное место отводится вопросам интеграции методов и средств представления и обработки темпоральной информации с методами и средствами имитационного моделирования внешней среды в РВ [Рыбина, 2015].

В настоящее время проводится комплекс исследований по методам реализации ТПП «Построение динамических ИЭС», включая моделирование совместного функционирования подсистемы имитационного моделирования с темпоральным ешателем и другими компонентами под управлением интеллектуального планировщика и онтологии типовых архитектур ИЭС. Актуальной задачей в рамках конкретизации процессов данной ТПП является развитие базовых методов и средств построения имитационных моделей (ИМ) внешней среды и проведение имитационных экспериментов на этих моделях с помощью средств подсистемы имитационного моделирования.

В данной работе в качестве экспериментального «полигона» для исследования технологии построения ИМ внешней среды с помощью имеющихся средств, выбрана достаточно новая (с точки зрения примене-

ния методов и средств разработки динамических ИЭС) область космонавтики, а бортовой программный комплекс космического аппарата (КА) в качестве конкретной ПрО.

Этот выбор не случаен, т.к. в настоящее время благодаря бурному развитию микроэлектроники современные КА стали способны принимать, обрабатывать и отправлять значительные объемы телеметрических данных о состоянии КА и внешней среды. Поэтому в космической индустрии начали активно рассматриваться возможности применения интеллектуальных систем для контроля, диагностики и управления техническим состоянием КА [Демиденко и др., 2019], что связано с повышением автономности КА при решении этих задач. Здесь важно, что контроль технического состояния производится на борту КА, а диагностика и управление осуществляется на наземных комплексах центра управления полетами, что вызывает длительные задержки в выработке управляющих воздействий и их передаче на борт КА.

Анализ исследований и разработок в ракетно-космической области, отраженный в трудах предприятий госкорпорации «Роскосмос» [Балухто и др., 2019], [Иванов, 2020], [Данилкин и др., 2021], [Лоскутов и др., 2021], материалах профильных конференций и научных журналах [Vincent et al., 2021], [Oiu et al., 2022], [Гарбук, 2022], внушает определенный оптимизм, что применение методов и средств искусственного интеллекта (ИИ) для управления техническим состоянием КА позволит: значительно сократить время и стоимость эксплуатации КА; повысить устойчивость к отказам бортовых систем; уменьшить нагрузку на наземный комплекс управления и др. возможности.

Поскольку описанные действия систем КА должны осуществляться непрерывно в РВ, то наиболее целесообразным подходом здесь является использование методов и средств динамических интеллектуальных систем различной архитектурной типологии [Рыбина, 2015], в частности, динамических ИЭС, которые могут функционировать в РВ, используя измерения параметров систем КА и обеспечивая совместную обработку изменяющихся в РВ данных и знаний из различных источников. Для решения задач, связанных с уже выведенными на орбиту КА, применяются бортовые системы, в состав которых могут быть включены динамические ИЭС [Балухто и др., 2019], [Vincent et al., 2021].

Данная работа посвящена экспериментальному исследованию интеллектуальной технологии построения динамических ИЭС на примере разработки и конкретизации ТПП «Построение динамических ИЭС» с акцентом на имитационное моделирование внешней среды как неотъемлемого этапа построения динамических ИЭС.

## **1. Базовые положения общего подхода к прототипированию прикладных ИЭС с использованием интеллектуальной программной среды комплекса АТ-ТЕХНОЛОГИЯ**

Концептуально-функциональные и технологические вопросы, связанные с планированием и управлением процессами проектирования ПО прикладных ИЭС с использованием интеллектуального планировщика и технологической БЗ детально обсуждались в ряде публикаций, [Рыбина, 2008],[Рыбина и др., 2018а], поэтому здесь позиционированы только принципиально важные положения разработанного подхода.

1. Исходные данные для генерации планов разработки прототипов ИЭС: модель архитектуры прототипа ИЭС в виде иерархии расширенных диаграмм потоков данных (РДПД), и технологическая БЗ ( множество ТПП, ПИК , прикладная онтология типовых архитектур ИЭС).

2. Модель прототипирования ИЭС включает функцию планирования для генерации текущего прототипа ИЭС для конкретной ПрО, поэтому постановка задачи интеллектуального планирования процессов разработки ИЭС рассматривается в контексте этой модели.

3. Для реализации эффективного метода планирования действий инженеров по знаниям использован достаточно известный подход, связанный с планированием в пространстве состояний [Nau, 2007].

4. Основная задача интеллектуального планировщика заключается в автоматической генерации планов (глобального и детальных) на основе модели архитектуры ИЭС и набора ТПП из технологической БЗ, что значительно снижает риски ошибочных действий инженеров по знаниям.

5. Выполнение задач плана осуществляется с помощью совокупности операционных ПИК (каждая плановая задача связана с функцией конкретного ПИК). Оперируя ТПП как основным алгоритмическим элементом, интеллектуальный планировщик производит построение плана разработки ИЭС в зависимости от текущего состояния проекта (типа НФ-задачи в модели архитектуры) и особенностей ПрО (в настоящее время используется около 100 операционных и информационных ПИК).

6. Основная суть разработанного метода генерации плана заключается в выполнении последовательности следующих преобразований модели архитектуры текущего прототипа ИЭС: получение обобщенной РДПД в виде графа; генерация точного покрытия (т.е. множества экземпляров ТПП с взаимно непересекающимися фрагментами, содержащими все вершины графа) с помощью эвристического поиска; генерация плана на основе полученного покрытия; генерация представления плана.

7. Все этапы, указанные в п.6, выполняются интеллектуальным планировщиком, полностью реализующим функциональность, связанную с планированием процессов прототипирования ИЭС, в результате чего

строится точное покрытие, преобразуемое в дальнейшем в глобальный план разработки, и осуществляется детализация всех элементов покрытия, формируя тем самым предварительный детальный план, где каждая задача адресована вызову конкретного ПИК (на основе анализа доступности ПИК с помощью прикладной онтологии типовых архитектур ИЭС), а затем производится формирование необходимого представления конечного плана для его использования другими компонентами.

## 2. Разработка и конкретизация ТПП «Построение динамических ИЭС»

Для генерации планов разработки прототипа динамической ИЭС используется модель архитектуры динамической ИЭС и ТПП «Построение динамической ИЭС», модель которой в соответствии с [Рыбина, 2008] представляется в виде  $TPP = \langle C, L, T \rangle$ , где  $C$  – множество условий, при выполнении которых возможна реализация данной ТПП;  $L$  – сценарий выполнения ТПП, представляемый на внутреннем языке описания действий ТПП;  $T$  – множество параметров, инициализируемых интеллектуальным планировщиком при включении ТПП в план разработки прототипов ИЭС.

При разработке данной ТПП и ее конкретизации на основе базовой модели учитывались особенности всех этапов ЖЦ и распределенная функциональность архитектуры комплекса АТ-ТЕХНОЛОГИЯ, «разнесенная» на зарегистрированные компоненты (ПИК), использующиеся для построения прототипов динамических ИЭС.

**Набор условий  $C$ .** Для включения в план разработки должны быть выполнены все следующие условия: текущий этап ЖЦ – анализ системных требований; в составе модели архитектуры динамической ИЭС (иерархия РДПД) присутствует элемент, указывающий на наличие ИМ внешней среды; на иерархии РДПД присутствует хотя бы один элемент, связанный с решением НФ-задачи; на иерархии РДПД указан источник темпоральных знаний (эксперты, ЕЯ-тексты, БД).

**Сценарий  $L$ .** Основные этапы сценария ТПП:

**1. Стадия анализа системных требований.** Выполняются действия: автоматизированное приобретение знаний на основе темпоральной версии комбинированного метода приобретения знаний [Рыбина, 2015] и построение фрагментов темпоральной БЗ; верификация темпоральной БЗ; построение ИМ внешней среды.

**2. Стадия проектирования.** Выполняются действия: формирование средств вывода (поддерживается связка универсального АТ-РЕШАТЕЛЯ и темпорального решателя); разработка модели рабочей памяти и организация совместного функционирования подсистемы имитационного моде-



лирования, темпорального решателя и АТ-РЕШАТЕЛЯ; разработка объяснений (опционально); конфигурирование компонентов ядра динамической ИЭС в зависимости от специфики ПрО.

**3. Стадия реализации.** Выполняются действия: проведение имитационного эксперимента для построения ИМ; разработка визуального представления ИМ (опционально); синхронизация работы темпорального решателя и АТ-РЕШАТЕЛЯ; разработка средств общения с использованием языка описания сценариев диалога [Рыбина, 2015]; интеграция при необходимости с внешними системами (БД, ППП и др.).

**Набор параметров Т.** Включение ТПП в план разработки ИЭС инициализируется как параметром планировщика, в котором хранится информация о выполняемой в данный момент ТПП, так и параметром, в котором хранится текущий шаг сценария ТПП. В данном случае выполнение ТПП обеспечивается с помощью следующих операционных ПИК, реализующих объявленную функциональность: три универсальных ПИК («Темпоральный решатель», «Рабочая память», «АТ-РЕШАТЕЛЬ»); два специализированных ПИК («Расчет ИМ», «Разработка ИМ»).

Основные этапы сценария данной ТПП и соответствующие ПИК связаны с методами и средствами построения ИМ внешней среды, поэтому ниже рассматриваются особенности используемого подхода.

### **3. Особенности применения имитационного моделирования при построении динамических ИЭС на основе задачно-ориентированной методологии**

В рамках задачно-ориентированной методологии для программного моделирования динамики внешней среды используется концепция имитационного моделирования на основе развития оригинального отечественного подхода «Ресурсы-Действия-Операции» (РДО) [Емельянов и др., 2009], являющегося наиболее приемлемым как с точки зрения концептуальной совместимости с формализмом представления темпоральных знаний в динамических ИЭС, так и за счет разработки подсистемы имитационного моделирования, реализованной путем развития базовых принципов РДО-подхода и удовлетворяющей требованиям со стороны темпорального решателя [Рыбина, 2015].

Этот подход оказался достаточно эффективным, исходя из принципов глубинной интеграции всех компонентов ядра динамической ИЭС и возможностей передачи данных от ИМ в рабочую память при совместном функционировании подсистемы имитационного моделирования с темпоральным решателем, АТ-РЕШАТЕЛЕМ и другими компонентами комплекса АТ-ТЕХНОЛОГИЯ. Общая функциональность подсистемы имитационного моделирования разделена между двумя глобальными модулями

– «Разработка ИМ», обеспечивающим разработку, отладку и визуальный интерфейс ИМ, и «Расчет ИМ», предназначенным для расчета состояний ИМ в каждый такт времени функционирования динамической ИЭС.

Объединяющей концептуальной основой для этих модулей является разработка объектно-ориентированного полнофункционального языка высокого уровня РДО<sup>AT</sup> [Рыбина и др., 2014] для описания ИМ на основе развития элементов РДО-метода и с учетом требований к реализации темпорального вывода в динамических ИЭС и создания соответствующего транслятора в составе модуля «Расчет ИМ».

#### **4. Разработка и исследование технология построения имитационных моделей средствами подсистемы имитационного моделирования**

Рассмотрим технологию построения ИМ, которая разрабатывалась и экспериментально исследовалась как в автономном режиме построения ИМ, так и в режиме совместного функционирования подсистемы имитационного моделирования и темпорального решателя, в том числе с использованием некоторых примеров построения ИМ внешней среды при прототипировании динамической ИЭС для бортового комплекса КА.

**Описание элементов ИМ.** Производится выбор объектов моделирования (ресурсов) и параметров ресурсов, после чего определяются правила взаимодействий между ресурсами, т.е. события, и выполняется разделение событий на регулярные и нерегулярные, причем с указанием релевантных событию ресурсов.

**Построение описания ИМ на языке РДО<sup>AT</sup>.** Строится описание объектов языка РДО<sup>AT</sup> (типы ресурсов, ресурсы, образцы операций, операции, функции) и проводится контроль корректности описания объектов РДО<sup>AT</sup> (незаполненный параметр объекта, нарушение логики объекта и т.д.). Каждый описанный объект добавляется в соответствующую коллекцию, а затем компонентом синтеза производится преобразование коллекций объекта в описание ИМ в формате РДО<sup>AT</sup> (осуществляется сохранение в соответствующую XML-структуру).

**Трансляция ИМ на языке РДО<sup>AT</sup> во внутреннее представление на языке C#.** С помощью транслятора РДО<sup>AT</sup> описание ИМ на языке РДО<sup>AT</sup> проходит этапы лексического анализа (формируются списки объектов), синтаксического анализа (полученные списки объектов преобразуются в дерево синтаксического анализа), семантического анализа (выполняется проверка правильности описания ИМ в соответствии с логикой языка РДО<sup>AT</sup>). Затем осуществляется преобразование исходного описания ИМ в код на языке C# и полученное описание компилируется в .dll.

**Конфигурирование параметров и расчет состояний ИМ.** Полученная ИМ в виде .dll поступает на вход компонента поддержки расчета состояний ИМ для конфигурирования параметров ИМ (длина и количество тактов), производится запуск ИМ и управление ИМ путем остановки/приостановки. Расчет состояний выполняется на основе состояния ИМ на предыдущем такте и с учетом операций, выполняемых ИМ. При достижении последнего такта процесс завершается и формируется файл значений параметров ресурсов на всех тактах. Если необходимость визуализации состояний ИМ отсутствует и результат работы ИМ устраивает специалистов, то первая итерация считается пройденной, т.е. получена первая версия ИМ.

**Визуализация ИМ.** Для графического представления ИМ используется компонент визуализации ИМ, включающий редактор кадров анимации и правил отображения и визуализатор. Редактор получает на вход описание ИМ на языке РДО<sup>AT</sup> и производится добавление кадров анимации (путь к изображению, ширина, высота и т.д.) и правил отображений, определяющих поведение кадров анимации, после чего соответствующие XML-структуры передаются на вход визуализатору вместе со значениями параметров ресурсов на всех тактах.

Технологический цикл автономного построения ИМ на этом завершается, а для организации дальнейшей работы полученная версия ИМ в виде .dll передается в рабочую память с помощью средств поддержки совместного функционирования (СПСФ), обеспечивающих синхронизацию процессов взаимодействия подсистемы имитационного моделирования и средств вывода (темпоральный решатель).

Исследования показали, что технология построения ИМ определяет по сути содержание и последовательность этапов имитационного эксперимента, выполняемого в рамках итеративного процесса построения ИМ, когда при возникновении ошибок или неудовлетворительных результатов работы (прогона) ИМ можно осуществить перестройку ИМ.

## **5. Пример построения имитационной модели внешней среды для демонстрационного прототипа динамической ИЭС**

В качестве примера построения ИМ на основе представленной выше технологии рассмотрим кратко особенности выбранной модельной ПрО, связанной с диагностикой бортового программного комплекса (БПК) КА, включающего в себя различные модули и компоненты, в том числе средства для управления энергопотреблением, системой связи, системой навигации, контроля и управления ориентацией, сбора данных и др.

Основное внимание в процессе разработки отдельных компонентов прототипа динамической ИЭС, предназначенного для функционирования в составе БПК, было уделено проведению модельного эксперимента, имитирующего работу КА в космическом пространстве. Целью моделирования являлось создание условий, максимально приближенных к реальным летным испытаниям КА, включающим трудно-прогнозируемые космические явления, такие как: солнечные вспышки, гамма-всплески, геомагнитные бури, осколки космического мусора и т.д.

Для этих целей, используя технологию, описанную в разделе 4., была построена ИМ внешней среды, названной «Орбитальные условия», где в качестве ресурсов (объектов) ИМ исследовались:

- поток заряженных частиц (электронов и протонов) в околоземном пространстве, измеряемый в единицах флюенса, в  $\text{MeV} / (\text{cm}^2 * \text{сек})$ ;
- магнитное поле Земли и его параметры, такие как интенсивность, направление и сила, измеряемые в единицах магнитной индукции, в нанотеслах (нТл);
- уровень радиации на орбите, который может быть выражен в единицах дозы, в раддах (rad) или грей (Gr);
- концентрация космического мусора в определенной области космоса, измеряемая в единицах количества мусора на кубический километр;
- солнечная активность, измеряемая в единицах интенсивности солнечного излучения, в ваттах на квадратный метр и др.

Далее в табл. 1 приводится пример описания объекта типа ресурса на языке РДО<sup>АТ</sup>.

Таблица 1

Описание объекта на языке РДО <sup>АТ</sup>	Комментарий
<pre> Тип ресурса: \$Resource_type Солнечная_радиация: permanent \$Parameters Интенсивность_солнечной_радиации: real Излучение_в_ультрафиолетовом_диапазоне: real Излучение_в_инфракрасном_диапазоне: real Космические_лучи: real \$Modifiers Constructor Begin Интенсивность_солнечной_радиации set 0 Излучение_в_ультрафиолетовом_диапазоне set 0 Излучение_в_инфракрасном_диапазоне set 0 Космические_лучи set 0 End \$End </pre>	<p>Тип ресурса "Солнечная радиация" представляет собой постоянный ресурс, который характеризуется параметрами, связанными с радиацией, которой подвергается космический аппарат в орбитальных условиях. Модификаторы в данном контексте выполняют роль конструктора, устанавливая начальные значения для параметров типа ресурса "Солнечная радиация", которые равны 0. В ходе имитационной моделирования эти параметры могут изменяться, отражая изменение условий внешней среды на орбите.</p>

С целью упрощения эксперимента, в качестве сущности КА рассматривался малый КА стандарта CubeSat-3U. Состояния объекта ИМ характеризовались параметрами, близкими к низкой околоземной орбите (НОО). Для ИМ «Орбитальные условия» были созданы 5 типов ресурсов, 15 ресурсов, 21 образец операций и 21 операция. В табл. 2 представлены экспериментальные данные, полученные при разработке элементов ИМ «Орбитальные условия» с использованием средств подсистемы имитационного моделирования.

Таблица 2

Элементы ИМ орбитальных условий	Кол-во элементов	Кол-во ошибок, обнаруженных транслятором языка РДО <sup>АТ</sup> при трансляции (мин.)	Время для устранения ошибок (мин.)	Общее время разработки элементов ИМ (мин.)	Общее время разраб. ИМ (мин.)
Типы ресурсов	5	-		32	393
Ресурсы	15	2	4	34	
Образцы операций	21	6	30	290	
Операции	21	-		37	

Общий сценарий модельного эксперимента, включал выполнение следующих этапов.

**Построение ИМ «Орбитальные условия».** В рамках проведения эксперимента была построена ИМ «Орбитальные условия», реализующая несколько сценариев полета КА. Данная ИМ способна имитировать различные действия, включающие изменения параметров орбиты, солнечной активности, магнитного поля Земли и т.д.

**Проведение интеграционного тестирования.** Исследовалось взаимодействие построенной ИМ с темпоральным решателем с помощью средств, обеспечивающих синхронизацию работы путем отправки текстовых сообщений на ограниченном подмножестве языка XML (компонент СПСФ). Моделировалась генерация и отправка управляющих воздействий (запуск, остановка, приостановка) к ИМ и темпоральному решателю, а также отсчет тактов дискретного модельного времени.

Проведенные эксперименты показали универсальные возможности подсистемы имитационного моделирования как для построения ИМ внешней среды, так и совместного функционирования с «темпорально-ориентированными» компонентами динамических ИЭС, что позволяет использовать базовые модули подсистемы как операционные ПИК при их инициализации интеллектуальным планировщиком на соответствующем этапе сценария ТПП «Построение динамических ИЭС».

## Заключение

Данная работа продолжает цикл исследований, связанных с многоаспектным анализом проблемы интеграции в гетерогенных архитектурах интеллектуальных систем (на примере динамических ИЭС). Рассматривался конкретный оригинальный подход к интеграции методов имитационного моделирования с методами построения динамических ИЭС, который был изначально реализован средствами базовой версии инструментального комплекса АТ-ТЕХНОЛОГИЯ, а теперь развивается и экспериментально исследуется как логически-связанная основа ТПП «Построение динамических ИЭС» уже в рамках интеллектуальной программной среды комплекса АТ-ТЕХНОЛОГИЯ.

## Список литературы

- [Балухто и др., 2019] Балухто А.Н., Романов А.А. Искусственный интеллект в космической технике: состояние, перспективы развития // Ракетно-космическое приборостроение и информационные системы. – 2019. – Т. 6, № 1.
- [Гарбук, 2022] Гарбук С.В. Задачи нормативно-технического регулирования интеллектуальных систем обработки данных дистанционного зондирования Земли // Современные проблемы дистанционного зондирования Земли из космоса. – 2022. – Т. 19, № 1.
- [Данилкин и др., 2021] Данилкин А.П. и др. Космическая система оперативного мониторинга земной поверхности на базе малых космических аппаратов // Космическая техника и технологии. – 2021. – № 1 (32).
- [Демиденко и др., 2019] Демиденко Ю.А., Ключников В.Ю., Смагин Ю.Н. Состояние и перспективы интеллектуализации космических систем // Космонавтика и ракетостроение. – 2019. – № 1.
- [Емельянов и др., 1998] Емельянов В.В., Ясиновский С.И. Введение в интеллектуальное имитационное моделирование сложных дискретных систем и процессов. Язык РДО. – М.: АНВИК, 1998.
- [Иванов, 2020] Иванов И. Г. Поддержка принятия решений в процессе испытаний перспективных космических средств на основе прецедентов // Вопросы электротехники. Труды ВНИИЭМ. – 2020. – Т. 175, № 2.
- [Лаврищева, 2016] Лаврищева Е.М. Программная инженерия. Парадигмы, технологии и CASE-средства: учебник для вузов. – М.: Изд-во Юрайт, 2016.
- [Лоскутов и др., 2021] Лоскутов А. И. и др. Нейросетевой подход к контролю технического состояния бортовой радиоэлектронной аппаратуры изделий ракетно-космической техники в условиях многозначности функции выходов // Вестник НПО им.С.А. Лавочкина. – 2021. – № 1.
- [Рыбина, 2008] Рыбина Г.В. Теория и технология построения интегрированных экспертных систем. Монография. – М.: Научтехлитиздат, 2008.
- [Рыбина, 2015] Рыбина Г.В. Интеллектуальные системы: от А до Я. Серия монографий в трех книгах. Кн. 2. Интеллектуальные диалоговые системы. Динамические интеллектуальные системы. – М.: Научтехлитиздат. 2015.

- [Рыбина, 2023]** Рыбина Г.В. Интеллектуальная технология построения интегрированных экспертных систем различной архитектурной типологии: особенности разработки прототипа для управления медицинскими силами и средствами при крупных дорожно-транспортных происшествиях // Информационно-измерительные и управляющие системы. – 2023. – Т. 21, № 1.
- [Рыбина и др., 2014]** Рыбина Г.В., Рыбин В.М., Паронджанов С.С., Со Ти Ха Аунг. Имитационное моделирование внешнего мира при построении динамических интегрированных экспертных систем // Информационно-измерительные и управляющие системы. – 2014. – Т. 12, № 12.
- [Рыбина и др., 2018]** Рыбина Г.В., Блохин Ю.М. Методы и программные средства интеллектуального планирования для построения интегрированных экспертных систем // Искусственный интеллект и принятие решений. – 2018. – № 1.
- [Nau, 2007]** Nau D.S. Current trends in automated planning // AI Magazine. – 2007. – Vol. 28, № 4.
- [Oiu et al., 2022]** Qiu Y. et al. Exploration and Application of Information Technology in Quality Control over the Development of Spacecraft // 2022 International Conference on Computer Science, Information Engineering and Digital Economy (CSIEDE 2022). – Atlantis Press, 2022.
- [Rybina et al., 2021]** Rybina G.V., Slinkov A.A. The Implementation of the Ontological Approach to Control of the Processes of Designing Integrated Expert Systems Based on the Problem-Oriented Methodology // Proceedings 19th Russian Conference Artificial Intelligence (RCAI 2021). – Springer Nature Switzerland AG, 2021.
- [Vincent et al., 2021]** Vincent N.C. et al. Impact of artificial intelligence in the aviation and space sector // Artificial Intelligence. – CRC Press, 2021.

## РЕАЛИЗАЦИЯ МЕХАНИЗМА ПРИНЯТИЯ РЕШЕНИЙ В ИНТЕЛЛЕКТУАЛЬНОЙ АВТОМАТИЗИРОВАННОЙ ОБУЧАЮЩЕЙ СИСТЕМЕ НА БАЗЕ МОДУЛЯ ЭКСПЕРТНЫХ СИСТЕМ

В.А. Углев ([vauglev@sfu-kras.ru](mailto:vauglev@sfu-kras.ru))

Сибирский федеральный университет, Железногорск

В статье рассматривается общий принцип организации интеллектуального планировщика (решателя) экспериментальной интеллектуальной автоматизированной обучающей системы, опирающейся на механизмы экспертных систем и когнитивной визуализации. В качестве основы для реализации планировщика использован модуль программы FLM\_Builder. Описана специфика решателя для интерпретатора на языке Python и приведены этапы формирования моделей экспертных систем. В качестве иллюстрации работы механизма вывода рассмотрена задача индивидуализации состава учебного курса с последующей визуализацией в нотации когнитивных карт диагностики знаний, соответствующих сквозному подходу к анализу учебной ситуации в обучающей системе.

**Ключевые слова:** электронное обучение, интеллектуальные автоматизированные обучающие системы, принятие решений, экспертные системы, когнитивная визуализация.

### Введение

Дистанционная и электронная формы обучения сделали востребованными такие программные комплексы, как интеллектуальные автоматизированные обучающие системы. В большинстве случаев их делают на базе развитых средств управления обучением (LMS Moodle или iSpring). Но у большинства реализаций таких систем подход к принятию решений достаточно прост и сводится к технологиям программированного обучения [Skinner, 1986], [Беспалько, 1970]. Повысить результативность функционирования таких систем можно за счёт методов искусственного интеллекта [Brusilovsky et al., 1999], [Gavrilova, 2019], [Рыбина, 2023], что привело к формированию отдельного класса обучающих систем – *интеллектуальных автоматизированных обучающих систем* (ИАОС или ITS).



Спецификой работы ИАОС является возможность использования явных знаний и опыта при выработке управляющего (педагогического) воздействия на учащегося. При этом решения со стороны ИАОС должны быть не только проверяемы, но и автоматически аргументируемы (объяснимы, в соответствии с подходом ХАИ [Arrieta, 2020]). По этой причине *планировщик* (решатель интеллектуального ядра ИАОС) реализует логику с опорой на базу знаний [Гаврилова и др., 2001], представленной в виде сложных структур (коллекций правил, графов, семантических сетей, сценариев и онтологий). Подобные механизмы выработки решений принято называть экспертными (рекомендательными) системами.

При комплексной обработке обстоятельств сложившейся учебной ситуации важно выявить не только текущие показатели (оценки, частотно-временные показатели и характер реакций), но и свести воедино текущие (оперативные), тактические и стратегические задачи всего процесса обучения, формируемые «на протяжении всего образовательного цикла по конкретным направлениям / специальностям подготовки выпускников» [Рыбина и др., 2022]. Такие возможности пока не реализованы в полной мере в используемых в учебных заведениях ИАОС, а соответствующие методы и средства требуют дополнительного изучения и методической проработки.

Снижение сложности процесса комплексной диагностики учебной ситуации можно достичь за счёт дополнительных механизмов параметрического картирования и когнитивной визуализации. Мы объединяем данные методы в рамках единого подхода к принятию решений, названного сквозным [Uglev et al., 2022].

Рассмотрим в данной статье вариант организации механизма принятия решений в ИАОС, опирающийся как на механизм экспертных систем, так и на средства когнитивной визуализации.

## **1. Принятие решений в сквозном подходе**

Преобразование текущей информации об учебной ситуации к решению предполагает со стороны планировщика осуществить концентрацию информации, извлеченной из образовательного цифрового следа, и её обобщение. Этапы обобщения, как было показано в [Углев и др., 2022], во многом зависят от того, в каких масштабах и аспектах необходимо проанализировать учебную ситуацию. Локализуем задачи, которые предполагают привлечение интеллектуального решателя для обеспечения основных этапов (табл. 1): инициализация моделей (а), синтез шаблонов визуализации (б), статистическая концентрация (в), метрическая концентрация (г), семантическая концентрация (д), логическая концентрация (е), синтез обратной связи (реакция ИАОС, ж) и синтез диалога (з).

Таблица 1

Механизмы / Этапы	(а)	(б)	(в)	(г)	(д)	(е)	(ж)	(з)
Фашификация			Ф1	Ф2	Ф3	Ф4		Ф5
Экспертный вывод				Л1	Л2	Л3	Л4	Л5
Параметрическое картирование		П1		П2	П3	П4	П5	П6
Когнитивная визуализация		К1						К2

Из таблицы видно, что преодоление нечёткой неопределенности (фашификация по Л. Заде [Zadeh et al., 2018]) и картирование (параметрическое и визуальное) тесно связаны друг с другом и должны координироваться общей логикой работы планировщика. Как следствие, этапы (а)-(г) реализуются планировщиком ИАОС при каждой загрузке системы и обновляются после каждого значимого события (фиксации изменения в модели учащегося): на выходе имеется набор параметрических частных карт (атлас) и их обобщение (сводная карта). Непосредственное принятие решения (экспертный вывод) реализуется на этапах (д) и (е), предполагающее типовую последовательность:

- 1) выбирается задача принятия решений относительно текущего события, целей обучения и состояния модели учащегося;
- 2) для каждой модели ученика, а затем учителя, реализуется подбор значимых параметров по данным из карт атласа или упрощенной сводной карты для аргументации частного решения и установления семантических связей между аспектами анализа (семантическая концентрация);
- 3) вектор значимых параметров подаётся в экспертную систему и для моделей ученика и учителя строятся частные решения со своей системой аргументации;
- 4) частные решения подаются на вход базе знаний тьютора, который реализует поиск компромисса между ними;
- 5) полученное решение подвергается верификации со стороны моделей ученика и учителя и, если нет принципиальных противоречий, то она принимается в работу на этапы (ж) и (з), иначе идет итерация, оттачиваясь от текущей оценки ситуации (возвращение к шагу 3).

Предложенная схема реализует вариант экспертной системы с доской объявлений (blackboard) [Jackson, 1999], где в основе баз знаний *Q*-моделей лежат коллекции правил (продукционная модель вывода знаний). Важной спецификой такого подхода является то, что за счёт метри-

ческой концентрации (картирования) появляется возможность быстрой проверки гипотез, сопоставления различных “точек зрения”  $Q$ -моделей, и смены аспектов аргументации (сквозной подход).

Каждый фрагмент базы знаний можно представить в виде чёрного ящика, для которого зафиксировано пространство входов и выходов. Процесс фашификации входных количественных данных (шкалы, число качественных состояний) является настраиваемыми на уровне базы знаний. Также и переход от входов к выходам может реализовать любую логику, отвечающую правилу вывода *modus ponens*, без ограничений на число промежуточных логических гипотез. Это позволяет не фиксировать жестко методики принятия решений и настраивать их под специфику места локализации ИАОС.

Подробнее об этапах (б), (в) и (г) можно посмотреть в [Углев и др., 2022]. Описание этапов (ж) и (з) приведено в [Uglev, 2023]. Поэтому обратимся к описанию технической реализации этапов (д) и (е) в экспериментальной ИАОС AESU.

## 2. Инструментарий

Реализация механизмов работы решателя, обеспечивающего логику работы экспертной системы для планировщика ИАОС, должно исходить из следующих условий:

- 1) отсутствует качественно проработанная онтология, позволяющая формализовать все особенности учебного процесса (затруднительно реализовать интегрированную экспертную систему по [Рыбина, 2023]);
- 2) анализ учебной ситуации должен комплексно охватывать учебный процесс ступени образования, т.е. единое описание метаданных охватывает лишь методическую сторону организации взаимодействия учащегося с ИАОС (даны параметры о электронных курсах и учебном плане);
- 3) интеграция показателей факторного пространства предполагает настройку характеристических функций, обеспечивающих механизм нечёткой логики при фашификации значений количественных переменных (минимизируется объём правил в базе знаний и учитывается субъективная неопределенность);
- 4) базы знаний должны иметь возможность редактироваться в дружелюбном интерфейсе и, по возможности, иметь редактор визуального проектирования;
- 5) просчёт условий учебной ситуации должен быть ориентирован на быстрый вызов (в 2-5 строк программного кода);

- б) при реализации ядра экспертной системы желательно применять широко распространенный язык программирования с поддержкой работы на backend сетевых реализаций ИАОС.

Несмотря на то, что конструкторы экспертных систем широко распространены (например, пакеты AT-ТЕХНОЛОГИЯ, СОВЕТ, FuziCalc, iThink, Knowledge Craft, ILOG RULES, движки Jena, Drools и даже MATLAB), мы использовали версию программы FLM\_Builder для языка Python. Версия 4.1b реализует решатель в виде модуля на языке Python (FLM\_modul.py), поддерживающий продукционные базы знаний с элементами нечёткой логики; а также WYSIWYG конструктор, развернутый на платформе нашей экспериментальной ИАОС AESU (<https://aesfu.ru/method/flm-builder/>) в виде web-сервиса.

Проект экспертной системы реализуется в конструкторе, последовательно:

- 1) формируя граф решений (рис. 1, справа), обеспечивающий выполнение правила *modus ponens* от входов к выходам (коэффициенты уверенности  $C_f$  пересчитываются в соответствии с композиционным правилом вывода);
- 2) настраивая характеристические функции, реализующие фазификацию количественных данных к набору качественных состояний (рис. 1, слева);
- 3) заполняя коллекции продукционных правил.

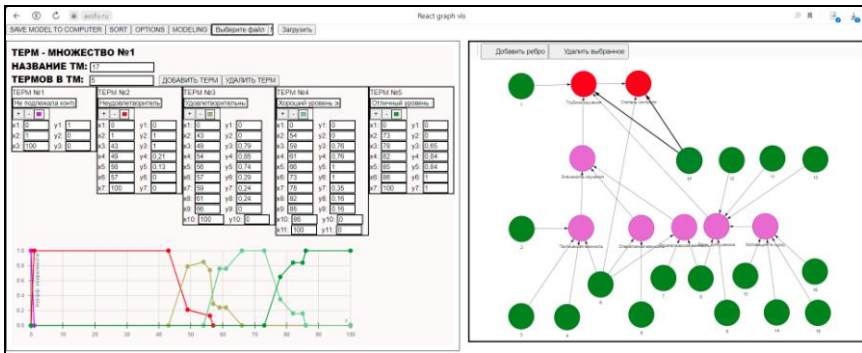


Рис. 1. Пример экранной формы конструктора моделей экспертных систем в браузере

Экспортируемый проект экспертной системы с базой знаний сохраняется в flm-формат (на базе XML). Для интеграции экспертной системы в ноутбук Jupiter языка Python или вызвать его интерпретатор на сервере, необходимо осуществить (на примере проекта в Python):

- 1) присоединение модуля FLM\_Builder к использующей системе (например, командой `import "FLM_modul as f"`);
- 2) загрузку целевой модели экспертной системы с базой знаний через команду `arr = f.startModelling(S)`, где строка, содержащая путь к файлу в формате flm (например, "ESModel.flm");
- 3) просчет целевой модели экспертной системы с базой знаний через команду `p = f.mod({P}, F, R)`, где

$P$  – вектор входных параметров, соответствующих числу входов в формате “номер узла в графе: 'значение'” (например, для экспертной системы, имеющей три входа, запишем “{1:'2',4:'1',5:'-27.5'}”);

$F$  – флаг, отвечающий за формат вывода результата (True – решение в естественно-языковой форме, False – только параметры решения);

$R$  – режим разрешения конфликтов при их возникновении равных  $C_f$  в процессе фазификации (1 – выбор терма слева, 2 – выбор терма справа).

На выходе формируется решение в виде массива, выводящего номер решения для каждого выхода и значений его  $C_f$  (например, для каждого выхода после команды “print(p)” будет выведено “[1', '0.7]” ).

Применение решателя FLM\_Builder, реализованного сначала в настольной, а затем и в сетевой версии, с 2005 года применялось в учебном процессе и научных проектах (подробнее см. в [Болсуновский, 2022]). Сформировано несколько десятков блоков баз знаний для экспериментальной ИАОС AESU, реализующих элементы этапов (д) и (е) сквозного подхода. Рассмотрим в качестве примера один из них.

### 3. Пример реализации проекта экспертной системы для ИАОС

Пусть *имеется* модель электронного курса «Имитационное моделирование» (читается магистрантам направления «Информатика и вычислительная техника» Сибирского федерального университета) и его связи с иными объектами учебного плана (другими курсами, компетенциями), а также данные анкетирования учащегося и результаты входного контроля. На основании этих данных *необходимо* рассмотреть базовую модель электронного курса на предмет индивидуализации. Курс состоит из 19 дидактических единиц  $u_i$  (где  $i \in [1..19]$ ), сгруппированных в 4 раздела. Для каждой из них имеются семантические связи внутри курса и с внешними сущностями (включая указанные в анкете цели учащегося), а также ряд априорных характеристик (базовая глубина изучения, тип, глубина контроля, сложность, важность, включенность в ядро курса и пр.). Из образовательного цифрового следа учащегося дополнительно извлекаются данные о результатах входного контроля (относительно каждой  $u_i$ ), личный приоритет, отношение к профилированию обучения и отношение к указанным личностным целям. Итого, имеется характеристика оценок/значений

из 17 показателей, рассчитываемых/получаемых на момент, предшествующий формированию индивидуализированного состава учебного курса и индивидуальной учебной траектории.

Сформируем модель экспертной системы с базой знаний для модели учителя. Упуская процесс извлечения/расчёта оценок и структурирование поля знаний, покажем результирующий граф решений (рис. 2), являющийся фрагментом окна конструктора с рис. 1. Зелёным автоматически отрисовываются входные данные из  $\{P\}$ , розовым – промежуточные гипотезы (об оперативной важности  $u_i$  в рамках темы, о тактической значимости в рамках курса, о стратегической важности в рамках учебного плана, об уровне мотивации учащегося к изучению курса и субъективной значимости дидактической единицы), а красным – итоговое решение о статусе дидактической единицы. Решение по статусу  $u_i$  определяется парой глубина изучения/степень контроля. Первый выход может принимать одно из четырех качественных состояний: “Исключена из курса (справочный режим)” (ответ с номером 1), “Включить в курс в кратком изложении” (2), “Включена в курс в стандартном изложении” (3), “Включена в курс в расширенном изложении” (4). Второй выход может принимать одно значение из “Исключена из контроля как справочная” (1), “Минимальный уровень” (2), “На оценку” (3), “Не контролировать как освоенную” (4).

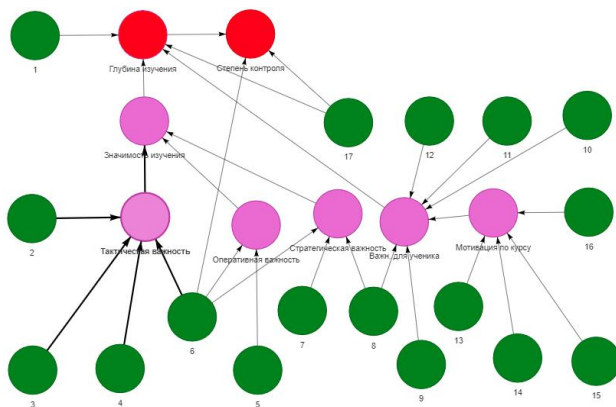


Рис. 2. Модель экспертной системы для индивидуализации состава учебного материала (модель учителя)

Рассмотрим пример результата просчёта для первой дидактической единицы “Имитационное моделирование, его объект и предмет” ( $u_1$ ), вектор  $\{P\}$  которого задан следующими значениями семнадцати входов:

1. “Базовая глубина изучения”  $u_1$  с качественным значением 2, т.е. «Включить в курс в кратком изложении»;
2. “Сила семантических связей  $[u_1]$  с материалом курса” за рамками текущей темы с количественным значением 21, т.е. после фазификации вход получил состояние 2 “Низкая” с коэффициентом уверенности  $C_f=0,86$ ;
3. “Сила семантических связей  $[u_1]$  с целями реализации курса” с количественным значением 14, т.е. после фазификации вход получил состояние 2 “Низкая” с коэффициентом уверенности  $C_f=0,92$ ;
4. “Сила семантических связей  $[u_1]$  с умениями компетенциями курса” с количественным значением 21.1, т.е. после фазификации вход получил состояние 2 “Низкая” с коэффициентом уверенности  $C_f=0,84$ ;
5. “Сила семантических связей  $[u_1]$  с материалом курса” в рамках текущей темы с количественным значением 100, т.е. после фазификации вход получил состояние 4 “Высокая” с коэффициентом уверенности  $C_f=1$ ;
6. “Тип”  $u_1$  с качественным значением 3, т.е. «Не включено в ядро курса»;
7. “Сила семантических связей  $[u_1]$  с иными учебными элементами курсов учебного плана” с количественным значением 0, т.е. после фазификации вход получил состояние 1 “Нет” с коэффициентом уверенности  $C_f=1$ ;
8. “Профильность дисциплины” с качественным значением 1, т.е. “Да”;
9. “Наличие связей  $[u_1]$  с личностными предпочтениями по компетенциям” с качественным значением 2, т.е. “Опосредованные”;
10. “Априорная сложность изучения  $[u_1]$ ” с количественным значением 12, т.е. после фазификации вход получил состояние 1 “Низкая” с коэффициентом уверенности  $C_f=1$ ;
11. “Связь с индивидуальными целями ученика” дидактической единицы с качественным значением 2 “Опосредованная”;
12. “Выбранный учеником приоритет  $[u_1]$  в анкете” с качественным значением 3 “Не указан”;
13. “Выбранный учеником приоритет дисциплины в анкете” с качественным значением 2 “Приоритетная”;
14. “Ведущая мотивация изучения дисциплины” с качественным значением 3 “Использование для выпускной работы и применение в будущей профессии”;
15. “Предпочитаемая глубина изучения курса учеником по анкете” с качественным значением 2 “Профильная”;

16. “Активность учащегося при записи на курс” с качественным значением 1, т.е. “Низкая активность”;
17. “Результаты входного контроля” [ $u_i$ ] дидактической единицы с количественным значением 0, т.е. после фазификации вход получил состояние 1 “Не подлежала контролю” (на этапе входного контроля) с коэффициентом уверенности  $C_f=1$  (см. характеристические функции на левой части окна с рис. 1).

При вызове просчёта экспертной системы через модуль FLM\_modul.py с параметрами

$$p = f.mod(\{1:'2', 2:'21', 3:'14', 4:'21.1', 5:'100', 6:'3', 7:'0', 8:'1', 9:'2', 10:'1', 11:'2', 12:'3', 13:'2', 14:'3', 15:'2', 16:'1', 17:'0'\}, False, 2)$$

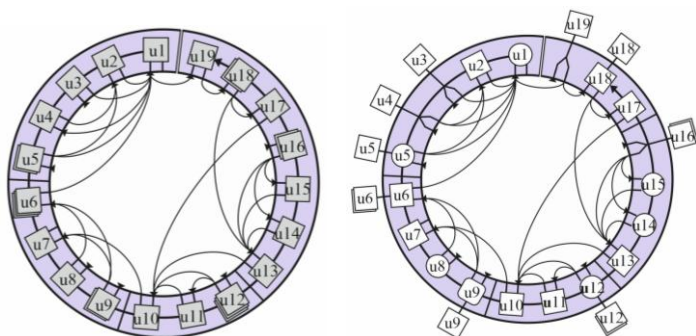
получим ответ  $[[3, 0.62][2, 1]]$ , т.е. для  $u_1$  принято решение “включить в стандартном изложении” с  $C_f=0,62$  (степень контроля для данной дидактической единицы будет принят на минимальном уровне, т.е. в режиме зачета  $C_f=1$ ). По аналогии планировщик определяет уровень включенность для всех 19 дидактических единиц (см. данные из табл. 2): обновленный статус дидактических единиц курса «с точки зрения» модели учителя можно сравнить с базовым значением из модели курса (первая позиция вектора  $P$ ).

Таблица 2

$i$	$P$							Результующий статус	$C_f$
	1	2	3	4	...	16	17		
1	2	21	14	21.1	...	1	0	3	0,62
2	3	0	84.1	44	...	1	1	4	0,84
3	2	0	12.2	15.5	...	1	1	1	0,71
...	...	...	...	...	...	...	...	...	...
18	3	14	69	71.2	...	1	1	3	0,94
19	2	12	14.8	14.1	...	1	0	1	0,76

Полученный результат можно визуализировать с помощью графической нотации когнитивных карт диагностики знаний (ККДЗ [Углев и др., 2022]). На рис. 3 (слева) показана визуализация базовой параметрической карты (по метаданным курса), предполагающей изучение всех 19 дидактических единиц. На правом рисунке показан результат отработки планировщика с базой знаний: из траектории изучения исключены дидактические единицы  $u_3$ ,  $u_4$ ,  $u_{16}$  и  $u_{19}$  (переведены в режим “справочных”), а ряд элементов курса изменил свою глубину контроля (на картах закодировано формой блоков, т.е. чем более округлая форма, тем менее требовательна ИАОС к уровню овладения материалом). Подробнее об интерпретации ККДЗ и их характеристиках см. в [Uglev et al., 2022].





*Рис. 3. Базовая (слева) и индивидуализированная (справа) когнитивная карта диагностики знаний курса “Имитационное моделирование”*

Приведенный пример показывает то, как на этапе семантической концентрации (д) был сгруппирован и предобработан блок параметров учебной ситуации и подан на этап логической концентрации (е), т.е. выработки решения с помощью базы знаний. Очевидно, что для разъяснения решений (например, при последующем диалоге согласования состава курса с учащимся), достаточно воспользоваться обратным проходом по протоколу работы решателя и автоматически синтезировать объяснение (см. методику в [Ughev, 2023]). Это позволяет воспользоваться преимуществами сквозного подхода как на этапе выработки решений, так и при организации результативного диалога с учащимся, подкрепленного когнитивной визуализацией.

### **Заключение**

Организация работы ИАОС, реализующих автоматически объяснимые решения на базе механизма экспертных систем, во многом зависит от гибкости интеграции и простоты использования планировщика (решателя). Модульное исполнение планировщика ИАОС позволяет использовать потенциал экспертных (рекомендательных) систем на любом этапе обработки образовательного цифрового следа или события. Для этого не требуется формировать классическую онтологию, хотя и требуется определенный уровень формализации метаданных.

Наш опыт реализации элементов ИАОС, включая разработку интеллектуального планировщика, показывает, что комплексное применение экспертных систем позволяет не только повысить адекватность решений обучающей системы, но и делает эти решения понятнее для учащегося.

В настоящее время нами осуществляется работа над методической проработкой баз знаний, касающихся вопросов функции рефлексивной оценки учебной ситуации и повышения эффективности взаимодействия моделей учащегося, учителя и тьютора при выработке обоснованных компромиссных решений. Также исследуется вопрос более гибкой визуализации различных аспектов учебной ситуации при автоматической аргументации принятых ИАОС решений.

**Благодарности.** Автор выражает благодарность Болсуновскому Николаю и Пронину Артёму за программную реализацию отдельных элементов экспериментальной платформы AESU и её сопровождение.

### Список литературы

- [Arrieta et al., 2020] Arrieta A., Díaz-Rodríguez N., Del Ser et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI // Information fusion. – 2020. – Vol. 58.
- [Brusilovsky, 1999] Brusilovsky P. Adaptive and intelligent technologies for web-based education. Ki, 1999. – 13 (4). – P. 19-25.
- [Gavrilova et al., 2019] Gavrilova T., Kokoulina L. Using Ontology Engineering to Design an Artificial Intelligence Course // Smart Education and e-Learning 2019. – Springer, 2019. – DOI: 10.1007/978-981-13-8260-4\_19.
- [Jackson, 1999] Jackson P. Introduction to Expert Systems, Addison-Wesley Pub. Co., Reading, 1999.
- [Skinner, 1986] Skinner B.F. Programmed instruction revisited // Phi Delta Kappan. – 1986. – Vol. 68 (2).
- [Ughev et al., 2022] Ughev V., Gavrilova T. Cross-Cutting Visual Support of Decision Making for Forming Personalized Learning Space // Novel & Intelligent Digital Systems: Proceedings of the 2nd International Conference (NiDS 2022). LNNS. 2022. – Vol. 556. – Springer. – DOI: 10.1007/978-3-031-17601-2\_1.
- [Ughev, 2023] Ughev V.A. Explanatory Didactic Dialogue in the Intelligent Tutoring Systems Based on the Cross-Cutting Approach // Proc. Augmented Intelligence and Intelligent Tutoring Systems. ITS 2023. LNCS. – Vol. 13891. – Springer, Cham. – DOI: 10.1007/978-3-031-32883-1\_34.
- [Zadeh et al., 2018] Zadeh L.A., Aliev R.A. Fuzzy logic theory and applications: Part I and Part II. – World Scientific Publishing, 2018.
- [Беспалько, 1970] Беспалько В.П. Программированное обучение. Дидактические основы. – М., 1970.
- [Болсуновский и др., 2022] Болсуновский Н.А., Пронин А.Д., Углев В.А. Конструктор производственных экспертных систем с элементами нечёткой логики FLM\_Builder и интеграция его моделей в пользовательские проекты // Нейроинформатика, ее приложения и анализ данных: XXX Всероссийский семинар. – Красноярск: ИВМ СО РАН, 2022. – С. 24-33.

- [Гаврилова и др., 2001]** Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. – СПб.: Питер, 2001. – 384 с.
- [Рыбина и др., 2022]** Рыбина Г.В., Григорьев А.А. Современные архитектуры интеллектуальных обучающих систем на основе интегрированных экспертных систем: особенности подхода к автоматизированному формированию онтологического пространства знаний и умений обучаемых // XX национальная конференция по искусственному интеллекту с международным участием (КИИ-2022). В 2 т. Т. 2. – М.: Изд-во МЭИ, 2022. – С. 218-231.
- [Рыбина, 2023]** Рыбина Г.В. Интеллектуальные обучающие системы на основе интегрированных экспертных систем: учебное пособие. – М.: Директ-Медиа, 2023. – 132 с.
- [Углев и др., 2022]** Углев В.А., Гаврилова Т.А. Подход к реализации сквозной визуальной поддержки процессов принятия решений для интеллектуальных автоматизированных обучающих систем // XX национальная конференция по искусственному интеллекту с международным участием (КИИ-2022). В 2 т. Т. 2. – М.: Изд-во МЭИ, 2022. – С. 413-426.

## ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА СЛУЖБЫ ПОДДЕРЖКИ ПОЛЬЗОВАТЕЛЕЙ ИНФОРМАЦИОННОГО ПОРТАЛА

Н.Ю. Щукин (*shchukinny@gmail.com*)

А.В. Бобряков (*avbob@mail.ru*)

А.С. Мохов (*asmokhov@mail.ru*)

Д.А. Щербаков (*shcherbakovdan@mpei.ru*)

Национальный исследовательский университет «МЭИ», Москва

Описываются структура интеллектуальной системы службы поддержки пользователей и исследования по решению задачи поиска релевантных ответов на запросы пользователей с использованием нейронной сети архитектуры «Трансформер», обеспечивающей представление запросов в векторном виде и дальнейший поиск наиболее близких векторов в базе данных часто задаваемых вопросов для формирования ответа.

**Ключевые слова:** служба поддержки, анализ текстовых данных, база данных часто задаваемых вопросов, релевантные ответы, нейросеть, трансформер.

### Введение

Важнейшей элементом современного информационного портала, выполняющего, как правило, функции единой точки доступа к различным корпоративным информационным системам, является служба поддержки пользователей (СПП). Основной задачей СПП является быстрое формирование релевантных ответов на поступающие запросы пользователей. Для решения названной задачи в настоящее время все чаще применяются различные интеллектуальные системы (ИС), обеспечивающие автоматическую или автоматизированную обработку запросов и формирование ответов пользователю [Al-Hawari et al., 2021], [Salman et al., 2023].

В работе описывается структура ИС службы методической, технической и консультационной поддержки пользователей (СМТКП) портала информационного взаимодействия с организациями и учреждениями, участниками бюджетного процесса Минобрнауки России (ПИВ МОН) [Bobryakov et al., 2019], [Kurorenkov et al., 2022], а также результаты исследований по решению задачи поиска наиболее релевантных ответов на поступающие

запросы пользователей. Важнейшей особенностью названной интеллектуальной системы является наличие базы данных часто задаваемых вопросов (ЧЗВ), содержащей вопросы методического и технического характера по эксплуатации различных информационных систем, входящих в состав портала, и экспертно выверенные релевантные ответы на эти вопросы.

ИС должна автоматически обрабатывать запрос, поступающий от пользователя, и пытаться найти релевантный ответ в базе данных ЧЗВ. Найденный ответ может быть как сразу автоматически направлен пользователю в случае высокой уверенности системы в релевантности ответа, так и направлен вместе с запросом специалисту СМТКП в качестве шаблона ответа, что ускорит время получения консультации пользователем. На практике, несмотря на то, что база данных ЧЗВ структурирована по разделам и задачам, зачастую пользователи не могут быстро найти ответ на свой вопрос самостоятельно и пишут запрос в СМТКП. Операторы СМТКП, в свою очередь, сталкиваются с тем, что приходится многократно отвечать на однотипные вопросы.

## **1. Структура интеллектуальной системы СМТКП ПИВ МОН**

Структура ИС СМТКП ПИВ МОН приведена на рис. 1 и содержит:

- базу данных часто задаваемых вопросов (БД ЧЗВ), содержащую вопросы методического и технического характера по эксплуатации различных информационных систем, входящих в состав портала, и экспертно выверенные релевантные ответы на эти вопросы;
- АРМ специалиста СМТКП, обеспечивающее возможность просмотра БД ЧЗВ при «ручном» поиске ответа на запрос пользователя (взаимодействие 1) или просмотр шаблона ответа подготовленного в автоматизированном режиме (взаимодействие 2), его коррекцию (в случае необходимости) и отправку пользователю через подсистему фиксации запросов (взаимодействие 3);
- подсистему фиксации запросов (ПФЗ), обеспечивающую фиксацию текста поступивших запросов и их атрибутов (время, отправитель и т.д.), а также ответов на запросы, сформированных специалистами СМТКП. Поступивший запрос пользователя передается на подсистему поиска релевантных ответов (взаимодействие 4);
- подсистему поиска релевантных ответов (ППРО), построенную на нейросети архитектуры «Трансформер», которая преобразует текст запроса в векторный вид и выполняет поиск по БД ЧЗВ для выявления похожего вопроса и ответа на него (взаимодействие 5). Найденный ответ может быть как сразу автоматически направлен пользователю в случае высокой уверенности системы в релевантности ответа (взаимодействие 6), так и направлен вместе с запросом специалисту СМТКП в качестве шаблона ответа (взаимодействие 2).

- подсистему ведения БД ЧЗВ, обеспечивающую периодическую актуализацию информации в БД ЧЗВ путем просмотра поступившей информации в ПФЗ (взаимодействие 7), нахождения схожих запросов, формирования из них типовых часто задаваемых вопросов и экспертно выверенных ответов на них и загрузки этой информации в БД ЧЗВ (взаимодействие 8).



Рис. 1. Структура ИС СМТКП ПИВ МОН

Основной подсистемой в рассматриваемой ИС является подсистема поиска релевантных ответов, для построения которой были проведены следующие исследования, описанные в работе:

- подготовка и обработка различных выборок из БД ЧЗВ;
- оценка применимости нейронной сети архитектуры «Трансформер» как модели представления текстов, обеспечивающей нахождение в базе данных вопроса, наиболее близкого к вопросу пользователя, и выдача релевантного ответа на вопрос пользователя.

## 2. Подготовка и обработка выборок из базы данных ЧЗВ

Исходная имеющаяся выборка из базы данных ЧЗВ содержит 447 вопросов и ответов на них, при этом длина вопроса варьируется от 5 до 119 слов, средняя длина – 27 слов. Также, особенностью задачи является достаточно специфичная предметная область, связанная с обеспечением мероприятий по финансированию подведомственных учреждений.

В ходе обработки выборки были проведены следующие работы:

- Перевод выборки ЧЗВ в векторную форму с помощью нейросети архитектуры «Трансформер».

- Формирование выборки для настройки системы. Для формирования данной выборки использовался сервис <https://neuralwriter.com/ru/>, позволяющий на основе заданного текста создавать его перефразировку с разными уровнями уникальности: «низкий», «средний» и «высокий». Таким образом была создана выборка, содержащая 201 перефразированный оригинальный вопрос из БД ЧЗВ. Перефразировки сохраняют структуру вопроса, для ее формирования чаще всего используются синонимы.
- Оценка качества различных подходов к поиску в БД ЧЗВ, наиболее близких к поступившему запросу пользователя. При этом на вход системы подавался перефразированный вопрос из выборки для настройки системы. Качество работы оценивалось как степень соответствия подобранных системой наиболее близких вопросов к исходному вопросу, из которого был перефразирован поступающий на вход вопрос.
- Формирование тестовой выборки для оценки качества работы системы на реальных запросах пользователей. Для формирования данной выборки из базы данных подсистемы фиксации запросов пользователей экспертно были отобраны 114 запросов, реально поступивших от пользователей портала, которым были экспертно сопоставлены соответствующие вопросы и ответы на них из БД ЧЗВ.

### 3. Модели представления текстовых запросов

Для работы с текстовыми данными их необходимо преобразовать – привести к математическому виду. Помимо стандартной модели «мешок слов» (bag of words) [Маннинг и др., 2014] в последнее время получили распространение векторные модели (word embedding), в которых каждое слово представляется в виде вектора, причем близкие термины с точки зрения контекста и смысла должны иметь близкие вектора. Для получения таких векторов могут применяться различные подходы на основе нейронных сетей: Word2Vec, GLOVE, FastText, [Mikolov et al., 2013], [Pennington et al., 2014], [Joulin et al., 2016], но в последнее время наилучшее качество показывают подходы на основе нейросетей с архитектурой «Трансформер», основанные на механизме внимания (attention model) [Vaswani et al., 2017].

Для реализации модуля поиска интеллектуальной системы службы поддержки пользователей информационного портала была выбрана нейронная сеть архитектуры «Трансформер» SBERT (sbert\_large\_mt\_nlu\_ru) для русского языка от Sberbank AI [SBERT]. Ее особенностью является то, что векторное представление данного трансформера является контекстным – это значит, что в зависимости от контекста документа, одно и то же слово (токен) может иметь разный вектор [Devlin et al., 2018], [Shao, 2020]. Каждый текст представлял собой среднее значение вектора слов, входящих в его состав, его размерность равна 1024.

#### 4. Исследование подходов к поиску схожих вопросов из БД ЧЗВ

На первом этапе были исследованы различные подходы к поиску наиболее близких вопросов в базе данных ЧЗВ. При этом на вход системы подавался вопрос из контрольной выборки, сформированной из перефразированных запросов. Далее с помощью трансформера SBERT получалось его векторное представление, и для полученного вектора искался наиболее близкий вектор из БД ЧЗВ.

Ответ системы считался правильным, если найденный наиболее близкий вектор из БД ЧЗВ соответствовал оригиналу перефразированного вопроса из выборки.

Исследовалось качество работы модуля в зависимости от применяемых мер близости (косинусная мера или евклидово расстояние [Маннинг и др., 2014]), а также от фрагмента БД ЧЗВ, используемого для поиска – поскольку БД ЧЗВ состоит из вопросов и ответов, то мы можем для поступившего вопроса искать наиболее близкий текст среди вопросов, среди ответов или среди объединенного текста из БД ЧЗВ, содержащего вопрос и ответ на него.

Таким образом рассматривалось четыре подхода:

1. Использование только вопроса из БД ЧЗВ и сравнение векторов по евклидову расстоянию.
2. Использование только вопроса из БД ЧЗВ и сравнение векторов по косинусной мере.
3. Использование только ответа из БД ЧЗВ и сравнение векторов по косинусной мере – на вход системы подавался вопрос пользователя, переводился в векторный вид, и наиболее близкий вектор к нему искался среди векторов ответов ЧЗВ.
4. Использование вопроса и ответа из БД ЧЗВ и сравнение векторов по косинусной мере – на вход системы подавался вопрос пользователя, переводился в векторный вид, и наиболее близкий вектор к нему искался среди векторов объединенного текста, состоящего из вопроса и ответа из ЧЗВ. Данный подход обусловлен тем, что в ответах на вопрос также содержится важная информация, которая может быть использована для улучшения качества формирования ответа.

Полученные результаты приведены в табл. 1.

Таблица 1

**Результаты для разных подходов к поиску в БД ЧЗВ**

Подход	Результат
Только вопрос, евклидово расстояние	0 из 201 (0%)
Только вопрос, косинусная мера	191 из 201 (95,0%)
Только ответ, косинусная мера	34 из 201 (16,9%)
Вопрос и ответ, косинусная мера	159 из 201 (79,1%)



Из таблицы видно, что применение евклидова расстояния не принесло абсолютно никаких результатов, поэтому дальнейшие исследования с использованием этой метрики не являются целесообразными. Скорее всего это связано с высокой размерностью задачи. Также не подтвердилась гипотеза о том, что для поступающего вопроса нужно искать похожий текст среди ответов из БД ЧЗВ. Такой подход выдал лишь 16,9% точности. Наилучший результат показал подход, когда мы сравниваем поступивший вопрос с имеющимся в ЧЗВ вопросом. В этом случае только для 10 перефразировок из 201 система не нашла оригинальный вопрос как самый близкий по косинусной мере. Из них, для 6 перефразировок оригинальный вопрос был найден в числе первых трех ближайших текстов.

Таким образом, для дальнейшей разработки системы выбран подход, заключающийся в сравнении векторов поступившего вопроса и вопроса из БД ЧЗВ по косинусной мере близости.

Также, отмечено, что чем короче длина сообщения, тем сильнее шум (добавление слов, перефразировки) влияет на близость текстов по косинусной мере.

Для коротких текстов даже небольшой шум сильно влияет на значение косинусной меры. При этом, для длинных текстов даже существенные перефразировки не так сильно сказываются на определении степени близости (табл. 2 и 3).

Таблица 2

**Значение косинусной меры для коротких текстов**

Оригинальный вопрос	Перефразировка	Значение меры
Пропала вкладка «Загрузить документы в формате ZIP или RAR», как загрузить подтверждающие документы?	Здравствуйте. Пропала вкладка «Загрузить документы в формате ZIP или RAR» где необходимо загрузить подтверждающие документы.	0.856
Необходимо утвердить приказами о стоимости до 2024/2025 учебного года включительно?	Нужно ли нам утверждать приказы о расходах до 2024/2025 года?	0.815

**Значение косинусной меры для длинных текстов**

Оригинальный вопрос	Перефразировка	Значение меры
Мониторинг налогообложения, разделы 3 и 4. В организации есть аспиранты, которые получают только стипендию, а есть аспиранты, которые трудоустроены на 4 часа и начисления проходят именно по их должностям. Вопрос – каких именно аспирантов нужно учитывать в данном отчете?	Налоговый мониторинг, разделы 3 и 4. В организации есть аспиранты, которые получают только зарплату, и аспиранты, которые работают по четыре часа, а начисления производятся по должности. Вопрос в том, каких аспирантов следует включить в этот отчет.	0.940
Скажите, пожалуйста, скан какого плана ФХД на 2021 год надо прикреплять на сайте sbias.ru: сформированного и утвержденного в ИС "Электронный бюджет" или сформированного на сайте sbias.ru и подписанного руководителем организации?	Подскажите какой план на 2021 год должен быть прикреплен на сайте sbias.ru, тот, который разработан и утвержден в ИС "электронный бюджет" или тот, который разработан на сайте sbias.ru и подписан ответственным лицом организации?	0.954

Из таблиц видно, что, для коротких текстов (менее 15 слов) даже небольшие перефразировки будут существенно влиять на значение косинусной меры. Чем длиннее текст, тем он более устойчив к перефразировкам и тем выше значение косинусной меры близости.

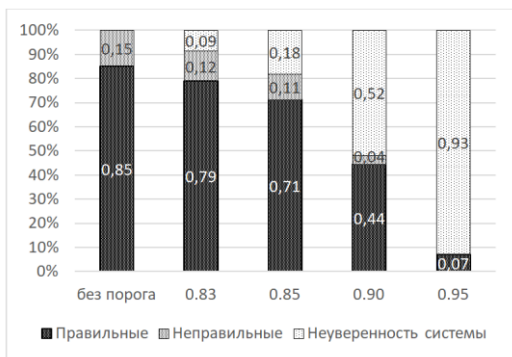
### **5. Апробация ИС СМТКП на реальных запросах пользователей**

Второй этап заключался в проверке работы ИС на тестовой выборке, содержащей реально поступившие запросы пользователей и экспертно подобранные для них вопросы из БД ЧЗВ. Ответ системы считался правильным, если найденный системой вопрос из БД ЧЗВ, как наиболее близкий к запросу пользователя, был сопоставлен и соответствовал вопросу, подобранному экспертами.

Также было исследовано влияние порога для значения косинусной меры, характеризующей уверенность модели в принятом решении, на распределение правильных/неправильных и неуверенных сопоставлений.

Результаты исследования для разных порогов приведены на рис. 2. Темным цветом показана доля верных сопоставлений, серым – доля неверных, светлым – доля сопоставлений с уверенностью системы ниже заданного порога.

Из результатов видно, что в целом, без учета степени уверенности, на 114 запросов тестовой выборки система нашла верные сопоставления для 97.



*Рис. 2. Распределение правильных/неправильных и неуверенных сопоставлений при различных порогах*

Установка значения порога равного 0.9, не имеет смысла, так как в этом случае более чем для половины вопросов системой не будет найдено правильного уверенного сопоставления, а 52% сопоставлений лежат в зоне неуверенного решения и их придется обрабатывать оператору (вместе с тем, при этом доля неверных сопоставлений составляет всего 4%).

Очевидно, рациональным значением порога можно считать 0.85, при этом доля правильных уверенных сопоставлений составляет более 70%, что позволит формировать ответ пользователю в полностью автоматическом режиме на 70% поступивших запросов.

## Заключение

В работе описаны результаты исследований по проектированию интеллектуальной системы службы поддержки пользователей информационного портала, базирующегося на нейронной сети архитектуры «Трансформер» SBERT.

Исследования показали, что для поиска наиболее релевантного ответа на поступивший запрос пользователя следует использовать косинусную меру близости.

При выборе значения порога уверенности системы нужно руководствоваться длиной сообщений, с которыми предстоит работать – для коротких сообщений порог нужно выставлять меньшим, поскольку в этом случае небольшие шумы существенно снижают значение косинусной меры. При этом, нужно учитывать критичность неправильного ответа. В случае, когда выдача правильного ответа является критичной, значение порога следует выбирать как можно большим.

Предложенные подходы показали высокую эффективность в ходе апробации при обработке запросов пользователей, поступающих в службу методической, технической и консультационной поддержки пользователей портала информационного взаимодействия с организациями и учреждениями, участниками бюджетного процесса Минобрнауки России (cbias.ru) и позволили значительно снизить время обработки входящего запроса пользователя.

В качестве дальнейших исследований планируется дообучение (fine tuning) нейросети на текстах предметной области, что обычно позволяет улучшить качество модели.

### Список литературы

- [Al-Hawari et al., 2021] Al-Hawari F., Barham H. A machine learning based help desk system for IT service management // J. King Saud Univ.-Comput. Inf. Sci. – 2021. – 33. – P. 702-718.
- [Salman et al., 2023] Salman Ahmed, Muskaan Singh, Brendan Doherty, Effirul Ramlan, Kathryn Harkin, Magda Bucholc and Damien Coyle. An Empirical Analysis of State-of-Art Classification Models in an IT Incident Severity Prediction Framework. // Applied Sciences. – 2023. – 13 (6):3843. – <https://doi.org/10.3390/app13063843>.
- [Bobryakov et al., 2019] Bobryakov A., Kuryliov V., Mokhov A., Stefantsov A. Approaches to automation processing of user requests in a multi-level support service using featured models // Proceedings of the 30th DAAAM International Symposium, pp. 0936-0944, B. Katalinic (Ed.). – Published by DAAAM International, 2019. – ISBN 978-3-902734-22-8. – ISSN 1726-9679, Vienna, Austria. – DOI: 10.2507/30th.daaam.proceedings.130.
- [Kuporenkov et al., 2022] Kuporenkov V., Mokhov A., Bobryakov A., Gavrilov A. and Naryadkina N. Approach to Develop an Intelligent User Support System of the Educational Platform // In 2022 VI International Conference on Information Technologies in Engineering Education (Inforino). – 2022. – P. 1-4. – DOI: 10.1109/Inforino53888.2022.9782934.
- [Маннинг и др., 2014] Маннинг К.Д., Рагхаван П., Шьюте Х. Введение в информационный поиск. – М.: Вильямс, 2014.
- [Mikolov et al., 2013] Mikolov T., Chen K., Corrado G., Dean J. Efficient Estimation of Word Representations in Vector Space // In Proceedings of Workshop at ICLR. – 2013.

- [Pennington et al., 2014]** Jeffrey Pennington, Richard Socher, and Christopher D. Manning. GloVe: Global Vectors for Word Representation // Conference: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). – 2014. – DOI: 10.3115/v1/D14-1162.
- [Joulin et al., 2016]** Joulin Armand, Grave Edouard, Bojanowski Piotr, Mikolov Tomas. Bag of Tricks for Efficient Text Classification // Conference: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. – 2017. – Vol. 2. – DOI: 10.18653/v1/E17-2068.
- [Vaswani et al., 2017]** Vaswani Ashish, Shazeer Noam, Parmar Niki, Uszkoreit Jakob, Jones Llion, Gomez Aidan N., Kaiser Lukasz, Polosukhin Illia. Attention Is All You Need: arXiv preprint arXiv:1706.03762, 2017.
- [SBERT]** AI-модель SBERT: <https://developers.sber.ru/portal/products/sbert>.
- [Devlin et al., 2018]** Devlin Jacob, Chang Ming-Wei, Lee Kenton, Toutanova Kristina. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding: arXiv:1810.04805v1, 2018.
- [Shao, 2020]** Shao H. Pretraining Deep Learning Models for Natural Language Understanding // Oberlin Honors. – Papers 709. – 2020. – <https://digitalcommons.oberlin.edu/honors/709>.

---

Секция 7 | ПРИКЛАДНЫЕ  
БИМЕДИЦИНСКИЕ СИСТЕМЫ

УДК 004.891.3

АВТОМАТИЗАЦИЯ ABCDE-АНАЛИЗА  
НОВООБРАЗОВАНИЯ С ИНТЕРПРЕТАЦИЕЙ  
РЕЗУЛЬТАТОВ\*

Е.И. Акимова

Г.Ю. Гуськов (*guskovgleb@gmail.com*)

Н.Г. Ярушкина (*jng@ulstu.ru*)

Ульяновский государственный технический университет,  
Ульяновск

В статье представлены работы по анализу макроизображений новообразований на основе анализа ABCDE. Основной целью исследования была попытка сформировать инструмент, позволяющий выявлять изменения новообразований по метрике ABCD, а метрику «Эволюция» отображать в виде прогноза тенденций развития новообразования. Также в статье представлена информационная система, позволяющая оценить изображение новообразования в режиме нечерного ящика и интерпретировать результат. Для реализации метрик необходимо было провести комплексную предобработку изображения и специальную обработку для каждой метрики. Ре-

---

\* Работа выполнена в рамках государственного задания № 075-03-2023-143 по проекту «Исследование интеллектуальной предиктивной аналитики на базе интеграции методов конструирования признаков гетерогенных динамических данных для машинного обучения и методов предиктивного мультимодального анализа данных».

зультаты работы исследования оценивались профильными экспертами. Пороговые значения были найдены при помощи материалов массива данных из открытых источников.

**Ключевые слова:** экспертная система, кластеризация, ABCDE-анализ, обработка изображений, новообразование.

## Введение

### ABCDE-анализ новообразований

Меланома кожи – злокачественная опухоль нейроэктодермального происхождения, исходящая из меланоцитов (пигментных клеток) кожи. Это одна из самых агрессивных и смертельно опасных видов опухолей, именно поэтому своевременно диагностировать её крайне важно.

Для выявления меланомы используют два вида методов: инвазивный и неинвазивный. Пациенты зачастую не соглашаются на проведение биопсии кожного новообразования без веских причин. Перед биопсией всегда применяется метод наружного осмотра, в ходе которого врач оценивает новообразование по нескольким критериям, распространенным в медицинской практике.

Систем критериев существует множество, однако одна из самых популярных – ABCDE, где:

- А (Asymmetry). Обычно проявление пигментации круглые и симметричные. Формы ранней меланомы зачастую бывают асимметричными: ось симметрии проводимая через центр очага пигментации, разделит очаг на две неодинаковые половины;
- В (Border). У обычных – границы четкие и ровные, у ранней меланомы же зачастую они нечеткие и изрезанные;
- С (Цвет) Как правило, родинки и пигментные пятна имеют равномерную коричневую окраску одного тона, злокачественные новообразования отличаются неравномерной окраской разных тонов черного или коричневого;
- D (диаметр). Обычные родинки или пигментные пятна обычно имеют диаметр до 5–6 мм, тогда как меланома на ранней стадии обычно больше 6 мм;
- E (Evolution) Изменения в родинке или пигментном пятне. Пигментные пятна, которые с течением времени меняют свою форму, размер или окраску, могут оказаться злокачественными.

В настоящее время ведутся работы по автоматизации анализа кожных новообразований. Однако чаще всего в исследованиях используется технология нейронных сетей, характеризующаяся сложностью интерпретации

результатов и их обоснованности. Подобные ограничения для области медицины крайне нежелательно: врач должен понимать, как именно система приняла то или иное решение.

В связи с этим возникает задача разработки метода автоматизации проведения и интерпретации ABCDE-анализа новообразования на изображении. Такой подход может быть использован как пациентом для первичного анализа, так и специалистом на приеме, что позволит сократить время, затрачиваемое специалистом на рутинные действия.

### **Статистика заболеваемости меланомой**

В Российской Федерации с меланомой ежегодно впервые сталкивается 12 тысяч человек (около 5 тысяч мужчин и 7 тысяч женщин), что составляет 5-8 человек на 100 000 населения для обоих полов. Этот показатель так же, как и в остальном мире, ежегодно увеличивается. С 2011 по 2021 годы заболеваемость меланомой в Российской Федерации выросла на 30-45% [Muntyanu et al., 2020]. Среднегодовой темп прироста заболеваемости за 10 лет составил 3,74% у мужчин и 3,04% у женщин [Muntyanu et al., 2020].

Стоит учесть, что чем больше возраст человека, тем выше вероятность возникновения меланомы. Кроме того, методы диагностики становятся всё более совершенными и эффективными, растёт осведомленность пациентов о заболевании [Каприн и др., 2022]. Доля меланомы в общем числе злокачественных новообразований в Российской Федерации сопоставима с мировыми показателями и составляет 1,8%. Средний возраст больных составил 61,6 года (оба пола), 61,3 года (мужчины), 61,8 года (женщины). Кумулятивный риск развития меланомы кожи (период 2007–2017 гг., возраст 0–74 года) составил 0,55% [Каприн и др., 2018].

Хотя на меланому приходится лишь 4% всех случаев рака кожи, она является причиной 80% случаев смерти от рака кожи. В абсолютном выражении меланома ежегодно вызывает смерть 57 тысяч человек в мире [Sung et al., 2021].

Представленные показатели показывают, что эффективность работы медицинских работников «первого контакта» по выявлению злокачественных новообразований довольно мала. Диагностика меланомы на ранней стадии крайне важна: чем раньше она будет выявлена, тем эффективнее будет результат лечения.

### **Обзор существующих программных решений**

Тенденция последних лет, обычно не связанная строго с системами диагностики новообразований кожи, заключается в том, что исследователи разрабатывают глубокие сети с большим количеством скрытых слоев (либо сверточных, либо полностью связанных слоев) для получения лучших результатов. Обычно, когда это происходило в начале, временная



сложность обучения, классификации, обнаружения или сегментации как-то игнорировалась, вся работа была больше ориентирована на лучшие статистические показатели (требуемые диагностами). В результате большинство работ, связанных с обнаружением, сегментацией и/или классификацией меланомы, основано на таких архитектурах нейронных сетей, как ResNet [El-Khatib et al., 2020], Inception/GoogLeNet, U-Net [Almaraz-Damian et al., 2020], DenseNet [Bajwa et al., 2020], AlexNet, VGG [Oliveira et al., 2016] и Mask R\_CNN.

## 1. Показатели ABCDE- анализа

### 1.1. Предобработка изображения для ABCDE-анализа

Каждый из существующих методов автоматизированной диагностики имеет этап предварительной обработки. Он заключается в применении следующих основных операций: удаление шума, выделение значимой части данных, изменение размера, преобразование яркости в оттенки серого или коррекция яркости, усиление интенсивности и контраста [Almeida et al., 2020]. Поскольку изображения новообразований кожи имеют высокую вариабельность содержания, этап сегментации является широко обсуждаемой темой и сложной задачей. Этот шаг является частью алгоритма, который позволяет разделить изображение на несколько наборов пикселей с выделением областей интереса с использованием автоматического или полуавтоматического процесса в качестве конечного результата. К числу наиболее часто используемых методов обнаружения и сегментации новообразований относятся методы, основанные на искусственных нейронных сетях.

### 1.2. Пример предварительной обработки изображения

Пиксели изображения группируются по цвету в зависимости от их близости друг к другу (рис. 1).

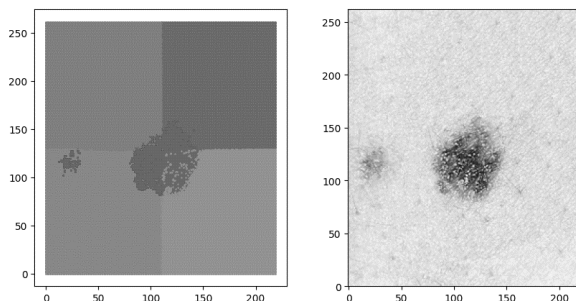


Рис. 1. Результат кластеризации изображений

На основании данных каждого кластера можно построить гистограмму рис. 2. Кластер, наиболее отличающийся от остальных по данным гистограммы, является кластером новообразования.

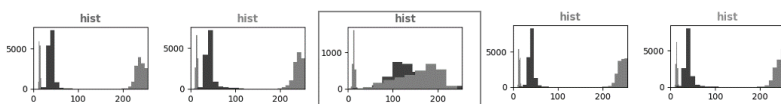


Рис. 2. Выделение кластера с новообразованием при анализе гистограммы

После этого на изображении выделяется новообразование кожи и убирается шум встроенной в OpenCV функцией для цветных изображений `fastNlMeansDenoisingColored`. Применяются морфологические трансформации: `ERODE`, удаляющая лишние пиксели вне контура новообразования, и `DILATE`, восстанавливающая все пиксели на контуре и внутри новообразования.

В этом случае на фотографии больше одного новообразования. Для обработки и оценки используется больший из них, поэтому на изображении находится и выделяется самый большой контур (рис. 3).

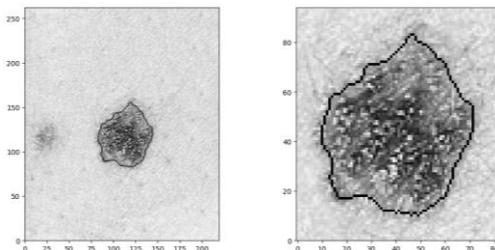


Рис. 3. Выделение основного очага новообразования

И только после проведения операций предобработки становится возможным рассчитать метрики. Далее по полученным результатам рассчитываются показатели ABCDE-анализа, после чего рассчитывается общий результат в формате вероятности перерождения новообразования в меланому в процентах.

### 1.3. Пример расчета ABCDE-метрик

**Асимметрия.** Выбирается точка из массива точек контура новообразования, показанного на рисунке синим цветом. Относительно него берутся пиксели справа и слева, показанные на рисунке оранжевым цветом, и от них рассчитывается расстояние до геометрического центра контура новообразования, показанного на рисунке красным цветом (рис. 4).

Дан набор, содержащий  $n$  точек  $x_1, x_2 \dots x_n$ , где  $x_i \in R^n$ , геометрический центр образования определяется:

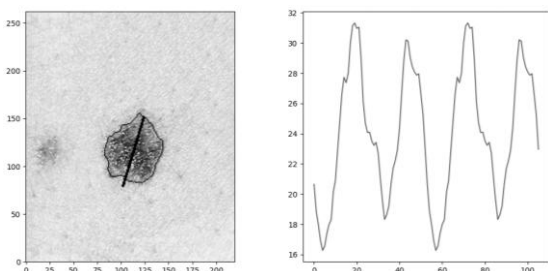
$$c = \arg \min_{y \in R^n} \sum_{i=1}^n \|x_i - y\|_2, \quad (1)$$

где  $\arg \min$  означает значение аргумента  $y$ , при котором достигается минимальная сумма. Это точка  $y$ , для которой сумма всех евклидовых расстояний до  $x_i$  минимальна.

Затем вычисляется модуль разности этих расстояний и прибавляется к конечному значению асимметрии. Чем выше это значение, тем более асимметрично новообразование.

Расчет происходит с шагом в один пиксель, значения метрик нормированы относительно длины контура новообразования.

В процессе расчета асимметрии также находится главная ось симметрии, где значение метрики асимметрии наименьшее, и выводится график рассчитанной асимметрии относительно каждой точки (рис. 4).



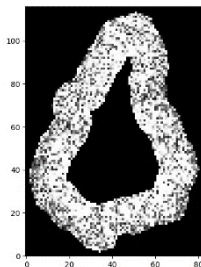
*Рис. 4. Ось симметрии, на которой значение метрики асимметрии наименьшее. И график рассчитанной асимметрии относительно каждой точки*

$$A = \arg \min_{j \in n} \sum_{i=1}^{n/2} |s_{(j-i)\%n} - s_{(j+i)\%n}|, \quad (2)$$

где  $\arg \min$  означает значение аргумента  $j$ , при котором достигается минимальное значение метрики,  $j$  – выбранная точка,  $i$  – индекс смещения,  $n$  – количество точек,  $s_{(j-i)\%n}$  – расстояние от левой точки до центра,  $s_{(j+i)\%n}$  – расстояние от правой точки до центра. На рис. 6 представлено изображение с выделенной осью сечения относительно которой новообразование представляется наименее симметричным, что позволит специалистам интерпретировать результаты автоматизированного анализа.

**Граница.** Для оценки ровности границ создается контурная маска и по этой маске вычисляется градиент, нормализованный по количеству пикселей. Значение метрики представляет собой сумму заданного градиента. Чем больше число, тем более неровной будет граница, поскольку градиент показывает разницу между соседними пикселями.

Граница на рис. 5 включает 15 краевых пикселей.

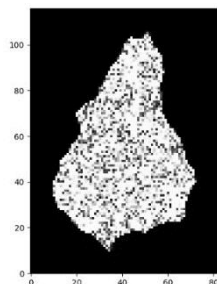


*Рис. 5. Граница, состоящая из 15 пикселей*

$$B = \sum_{i=1}^n \text{grad } x_i, \quad (3)$$

где  $n$  – количество пикселей на границе,  $x_i$  – конкретный пиксель.

**Цвет.** Для расчета цветовой метрики используется тот же метод, что и для границ, только расчет ведется по маске всего новообразования (рис. 6).



*Рис. 6. Маска новообразования для расчета метрики цвета*

$$C = \sum_{i=1}^n \text{grad } x_i, \quad (4)$$

где  $n$  – количество пикселей новообразования,  $x_i$  – конкретный пиксель.

**Размер.** Вычисляется радиус наименьшей описанной окружности в пикселях, затем умножается на размер пикселя, полученное значение удваивается (рис. 7).

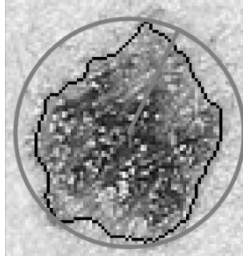


Рис. 7. Метрика диаметра: удвоенный радиус наименьшей описанной окружности

$$D = 2 * \min_{\vec{c} \in Z} (\max_{\vec{x} \in X} \|\vec{c} - \vec{x}\|), \quad (5)$$

где  $Z$  – множество двумерных векторов,  $X$  – множество точек, принадлежащих новообразованию. На рис. 7 представлено изображение, позволяющее специалистам интерпретировать результаты автоматизированного анализа. Так как на изображении поверх новообразования отрисована окружность в которую вписано новообразование.

## 2. Структура системы и процесс анализа изображений новообразований

Система представляет собой WEB-приложение, разработанное на Python с использованием OpenCV. Входная информация поступает от пользователей системы в виде изображений в форматах .jpg и .png.

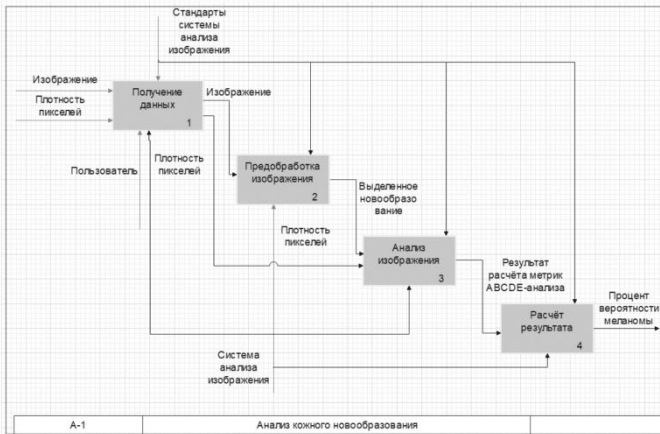


Рис. 8. Диаграмма IDEF0 level 1

На схеме (рис. 8) подробно описан процесс анализа кожных новообразований. Этот процесс состоит из четырех этапов, на каждом из которых элементом управления являются эталоны системы анализа изображений, а необходимыми механизмами – система анализа изображений:

- Входные данные. Ввод и вывод изображения и метаданных изображения. В этом процессе есть еще один необходимый механизм: пользовательские настройки.
- Предварительная обработка изображений. На входе – изображение, полученное от пользователя, на выходе – изображение изолированного новообразования кожи.
- Анализ изображения новообразования. Вход – выбранное новообразование и плотность пикселей, выход – результат расчета метрик ABCDE-анализа.
- Формирование прогноза эволюции новообразования. На входе показатели анализа ABCDE, на выходе – процент вероятности принадлежности новообразования к меланоме.

### 3. Эксперименты

Для определения эффективности предложенной модели и выбора порогового значения были проведены вычислительные эксперименты на открытом наборе данных. Под пороговым значением понимается граница, после превышения которой новообразование считается злокачественным, и система предлагает специалисту изучить новообразование ещё раз с учетом визуализации. Значения показателей суммируются с учетом коэффициентов подобранных при помощи модели линейной регрессии на основе изображений из размеченного датасета. Результаты опытов представлены на графике на рис. 9.

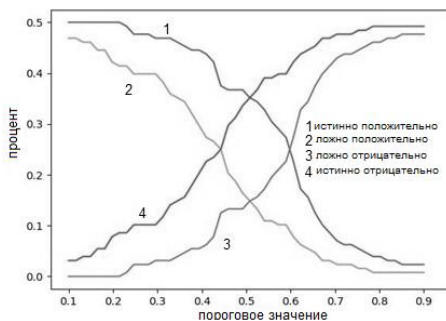


Рис. 9. График пороговой оценки

## Заключение

В ходе исследования был разработан программный продукт, позволяющий оценивать новообразование по изображению на основе критериев ABCDE-анализа. Подобное решение позволяет не просто рассчитать нормализованные показатели, привязанные к критериям ABCDE-анализа, но и позволить интерпретировать результаты в процессе работы специалистов в том числе за счёт визуализации дополнительных построений на изображении. Метрика «Эволюция» представлена в информационной системе последовательностью изображений, по каждому из которых проведён анализ. В данный момент существует набор правил, фиксирующий изменения новообразования, а обучить модель для поиска изменений невозможно по причине отсутствия подобного набора данных отражающего развитие новообразования в динамике. В статье представлена архитектура проектного решения, основные сценарии применения и общий вид приложения.

## Список литературы

- [Muntyanu et al., 2020] Muntyanu A., Savin E., Ghazawi F., Alakel A., Zubarev A., Litvinov I. Geographic Variations in Cutaneous Melanoma Distribution in the Russian Federation // *Dermatology*. – 2020. – No. 236 (6).
- [Каприн et al., 2022] Каприн А.Д., Старинский В.В., Петрова Г.В. Состояние онкологической помощи населению России в 2021 году. – М.: МНИОИ им. П.А. Герцена, 2022.
- [Каприн и др., 2018] Каприн А.Д., Старинский В.В., Петрова Г.В. Злокачественные новообразования в России в 2017 году (заболеваемость и смертность). – М.: МНИОИ им. П.А. Герцена, 2018.
- [Sung et al., 2021] Sung H., Ferlay J., Siegel R.L., Laversanne M., Soerjomataram I., Jemal A., Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries // *CA: a cancer journal for clinicians*. – 2021. – No. 71 (3).
- [El-Khatib et al., 2020] El-Khatib H., Popescu D., Ichim L. Deep learning-based methods for automatic diagnosis of skin lesions // *Sensors*. – 2020. – No. 20 (6).
- [Almaraz-Damian et al., 2020] Almaraz-Damian J.A., Ponomaryov V., Sadovnychiy S., Castillejos-Fernandez H. Melanoma and nevus skin lesion classification using handcraft and deep learning feature fusion via mutual information measures // *Entropy*. – 2020. – No. 22 (4).
- [Bajwa et al., 2020] Bajwa M.N., Muta K., Malik M.I., Siddiqui S.A., Braun S.A., Homey B., Dengel A., Ahmed S. Computer-aided diagnosis of skin diseases using deep neural networks // *Applied Sciences*. – 2020. – No. 10 (7).
- [Oliveira et al., 2016] Oliveira R.B., Filho E.M., Ma Z., Papa J.P., Pereira A.S., Tavares R.S. Computational methods for the image segmentation of pigmented skin lesions // *Computer Methods and Programs in Biomedicine*. – 2016. – No. 131.
- [Almeida et al., 2020] Almeida M.A.M., Santos I.A.X. Classification models for skin tumor detection using texture analysis in medical images // *Journal of Imaging*. – 2020. – No. 6 (6).

УДК 159.91+613.6

## МЕТОДЫ КОСМИЧЕСКОЙ ПСИХОЛОГИИ И ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА. ПЕРСПЕКТИВЫ ИНТЕГРАЦИИ\*

Ю.А. Бубеев (*aviamed@inbox.ru*)

А.В. Иванов (*avivanov@imbp.ru*)

Институт медико-биологических проблем РАН, Москва

В работе описываются психологические особенности космических полетов, некоторые экспериментальные модели для изучения факторов космического полета, описаны задачи биомедицинского мониторинга и психофизиологические методы оценки состояния с возможностью использования элементов искусственного интеллекта.

**Ключевые слова:** психофизиологическое состояние, дистанционный мониторинг, изоляция, астенизация, антиортостатическая гипокинезия, сухая иммерсия, коммуникативная активность, анализ мимики, анализ речи, актиграфия, системы ИИ в межпланетном полете.

### Введение

Космическая психология сфокусирована на изучении и контроле психического состояния человека в различных условиях деятельности, связанных с подготовкой и работой в космосе. Космическая психология и медицина уделяют особое внимание методам мониторинга и коррекции функционального состояния космонавтов и участников наземных экспериментов, с регистрацией различных физиологических данных, поскольку психофизиологические аспекты функционального состояния являются важной составляющей безопасной и эффективной деятельности. Мониторинг сердечного ритма, суточной физической активности, когнитивных показателей, речевых и мимических особенностей в процессе коммуникации осуществляется в постоянном или периодическом режиме. Обработка получаемых данных, как правило, требует определенного времени, что может влиять на оперативность получения оценок и основанных на них

---

\* Исследование выполнено в рамках научной программы Национального центра физики и математики (направление №9 «Искусственный интеллект и большие данные в технических, промышленных, природных и социальных системах»).



решений. Опыт использования моделей искусственного интеллекта (ИИ) в отдельных методиках (анализ психоэмоционального состояния по мимике) свидетельствует о перспективности внедрения ИИ в методический арсенал космической психологии для снижения рисков и изучения влияния факторов полета, дистанционного мониторинга и психологической поддержки членов экипажей.

### **Актуальность**

В структуре деятельности космонавта ведущее место занимает операторская деятельность, для эффективного выполнения которой требуется поддержание оптимального психического состояния, высокой и устойчивой работоспособности. Решение этой задачи требует проведения полетного мониторинга психического состояния членов экипажей.

В настоящее время такой мониторинг осуществляется экспертами группы медицинского обеспечения центра управления полетом (ЦУП). Результаты представляются в виде еженедельных заключений. Эти заключения подлежат не только оперативному, но и ретроспективному анализу, выполняемому после завершения полета. В ходе этого анализа необходимо:

- оценить полет в целом с точки зрения значимости проблем с психическим состоянием членов экипажа;
- определить наиболее важные причины отмеченных нарушений психического состояния.

Конечная цель ретроспективного анализа состоит в том, чтобы определить пути дальнейшего совершенствования системы психологического обеспечения космических полетов, что является одной из актуальных задач космической медицины. Сохраняет актуальность задача создания информационно-технических систем полетной количественной оценки когнитивных функций космонавтов с использованием специально отобранных и информационно значимых психологических когнитивных тестов, дополняющих данные дистанционного медицинского мониторинга.

### **Риски при психологическом обеспечении полета**

Реальная опасность развития психической астенизации космонавтов проявляющейся в виде нарушений сна, эмоциональных расстройств в виде раздражительности и депрессии, которые могут приводить к напряженности, ухудшению психологического климата в экипаже и развитию конфликтных ситуаций. В итоге это может приводить к снижению работоспособности и нарастанию числа ошибок при выполнении полетных операций [Мясников и др., 2000].

В межпланетном полете существенное отличие от орбитального приобретают мероприятия по диагностике психического состояния космонавтов, их психологической поддержке и сохранению необходимых профессиональных навыков непосредственно связанных с психикой. В этих условиях психически неуравновешенный человек должен рассматриваться не только как профессионально несостоятельное лицо, но и как потенциальный источник конфликтной напряженности в экипаже, что создает серьезную угрозу безопасности и выполнению программы полета. Для межпланетных экспедиций необходимо предусмотреть наличие в составе экипажа либо профессионального врача – психоневролога, либо сертифицированного в области клинической психологии специалиста, обеспеченного необходимыми средствами, включая технологии ИИ, предназначенными для оценки и коррекции состояния человека.

### **Психологические особенности межпланетного полета**

Несмотря на наличие общих с орбитальными психологических факторов, для межпланетных полетов характерны свои уникальные особенности.

- Замкнутое пространство обитания. При этом надо учитывать, что ограничения в размерах личного пространства и вынужденная длительная коммуникативная близость способна вызывать дискомфорт, ощущение тревожности и неврозы.
- Чувство отрыва от Земли, присутствие так называемого феномена "break-off" (оторванности, отрешенности) выражающегося в растерянности и мрачных предчувствиях ввиду отсутствия на трассе полета психологического «якоря», непосредственного визуального контакта с Землей.
- Постоянная угроза для жизни (радиация, метеориты, возможность опасного контакта с новыми формами жизни и др.), а также состояние тревоги и настороженности при этом, способны приводить к развитию разнообразных фобий/страхов.
- Абсолютная автономность функционирования и невозможность осуществления связи с наземными службами управления полетом в реальном масштабе времени.
- Монотонность выполняемой работы, а в определенные периоды полета и полное ее отсутствие.
- Ограниченный круг формального и неформального общения.
- Первостепенное значение совместимости членов экипажа.
- Невозможность экстренной эвакуации экипажа при аварии корабля.

- Отсутствие доказательных данных об устойчивости психики человека к осмыслению момента после ухода с земной орбиты и осознанию невозможности экстренной посадки (преодоление точки «не возврата»).
- Ностальгия при длительном пребывании и работе в составе экипажей межпланетных экспедиций и инопланетных баз.

### **Экспериментальные модели для изучения факторов космического полета**

Экспериментальные модели с изоляцией экипажа в ограниченных по объему герметично замкнутых объектах, также как модели «сухой» иммерсии и антиортостатической гипокинезии используются, в том числе для изучения социально-психологических и психофизиологических факторов, воздействующих на космический экипаж.

Сухая иммерсия (СИ) позволяет воспроизводить такие факторы космического полета как опорная и весовая аксиальная разгрузка, перераспределение жидких сред организма, гиподинамия. Также СИ даёт возможность решать космическим психологам такие задачи как: психологическая подготовка к условиям в изменённом пространстве; развитие стрессоустойчивости, саморегуляции и работоспособности в условиях невесомости, а также поло-ролевые особенности изменений в когнитивной, эмоциональной и поведенческой сферах личности в условиях невесомости. Модель позволяет проводить исследования эффективности различных средств и методов профилактики негативных влияний гравитационной разгрузки. Все более широкое применение метод находит в спортивной и клинической медицине [Григорьев и др., 2017].

Антиортостатическая гипокинезия (АНОГ) является одной из моделей, воспроизводящей физиологические эффекты невесомости у человека. Она приводит к перераспределению жидких сред организма в краниальном направлении, что вызывает взаимообусловленные адаптационные механизмы, в частности, сердечно-сосудистой, эндокринной, центральной и периферической нервных систем [Михайлов, 2001].

Преимущества экспериментальных моделей заключаются в возможности моделирования полного рабочего дня членов экипажа на фоне воздействия неблагоприятных факторов. В изоляционных экспериментах это факторы изоляции, скученности, сенсорной депривации, монотонии деятельности и однообразия обстановки, публичности деятельности, ограничения и навязанного характера социальных контактов, изменений гравитации. В этих экспериментах организация коммуникаций с внешним миром осуществляется по типу космической станции – с Центром управления и группой специалистов, контакты с родными и близкими лимитиру-

ются. Модельное исследование позволяет обеспечивать управление средой обитания, дает возможность контролировать воздействующие на организм и психику переменные (условия внешней среды, питание, величина физических нагрузок и пр.)

### **Системы биомедицинского психофизиологического мониторинга**

С помощью мониторинга медицинских и психофизиологических показателей в условиях космических и наземных экспериментов решаются следующие задачи:

1. Повышение профессиональной эффективности путем предоставления индивидуальной информации о физиологическом статусе каждого специалиста для оптимизации распределения задач и нагрузки, повышения ситуационной осведомленности руководства.

2. Оперативное прогнозирование приближающегося срыва систем адаптации от стрессовой нагрузки (физической, психологической и экологической).

3. Раннее обнаружение воздействия потенциально опасных агентов (химических, биологических, радиационных, электромагнитных, температурных, барометрических).

4. Детекция повреждений и ранений с подачей сигнала о необходимости оказания помощи, прогноза состояния, очередности оказания и объема необходимой медицинской помощи.

5. Обнаружение жертв и пострадавших во внешней среде, запрос и передача данных об их состоянии.

6. Обратная связь для оптимизаций физической нагрузки при тренировках и обучении.

7. Долгосрочный мониторинг и дозиметрия воздействия различных видов излучения с прогнозом рисков для здоровья.

8. Мониторинг психофизиологического состояния и сигнализация о пределах теплового воздействия, усталостных пределов опорно-двигательного аппарата, нейропсихологического статуса для выполнения конкретной задачи.

9. Периодический тренинг решения профессиональных задач, безопасности и выживаемости с целью сохранения и повышения эффективности важных навыков.

10. Информационная поддержка при оказании первой помощи при ранениях и поражениях.

Одним из существенных звеньев системы является психологический и психофизиологический контроль и прогнозирование состояния членов экипажа. В настоящее время с этой целью используются инструменталь-

ные методики и средства, позволяющие регистрировать объективные психофизиологические показатели в режиме периодического или постоянного мониторинга [Иванов и др., 2022]. Среди них отдельного рассмотрения заслуживают средства, которые могут использоваться в процессе операторской деятельности: с регистрацией показателей голоса и речи, выражений лица, жестов, двигательной активности (актиграфия), variability сердечного ритма. Для регистрации могут использоваться варианты «умных» тканей, активных электродов, бесконтактных датчиков, носимых мониторов физической активности. Для коррекции психологических проблем, связанных с изоляцией в экстремальных условиях, все более широко применяются технологии виртуальной реальности [Lockard et al., 2019], [Розанов и др., 2022], которые, в сочетании с регистрацией физиологических показателей, могут использоваться для периодического психофизиологического обследования, в ходе которого анализируются изменения состояния оператора при восприятии модельных виртуальных сценариев.

Любая система мониторинга должна обеспечивать возможность получения физиологических показателей без отвлечения специалистов от профессиональной деятельности; учитывать индивидуальные значения биометрических показателей конкретного специалиста; проводить анализ опасности выявленных изменений; выдавать рекомендации по необходимым мероприятиям оптимизации или коррекции; автоматически выдавать предупредительный сигнал о существенном нарушении работоспособности оператора, при возникновении состояний, представляющих реальную угрозу; давать возможность дополнительного углубленного анализа зарегистрированных биомедицинских показателей медицинским персоналом. Из применяемых в настоящее время систем мониторинга, элементы ИИ в виде нейронных сетей используются при оценке эмоциональных характеристик состояния по мимическим проявлениям при коммуникации членов экипажа с центром управления (ЦУП) по видеоканалу.

Принимая во внимание дистанционный режим бесконтактного измерения физиологических показателей для диагностики и мониторинга функционального состояния оператора в условиях космического полета, необходимо учитывать организационные сложности. Как показывает опыт [Вассерман и др., 2019], узкими местами телемониторинга остаются робастность получаемых оценок при измерениях в реальных условиях и точность вторичных показателей. Одним из показателей, обеспечивающих наибольшее разнообразие данных, является регистрация речи. Наименьшую организационную сложность измерений, как правило, обеспечивают методы, использующие видеокамеры.

Перечисленные особенности применения, а также дополнительные факторы (нестабильность и задержка связи с ЦУП по мере удаления от Земли и др.) определяют важность технологий на основе ИИ для оперативной оценки параметров функционального состояния экипажа, сохранения эффективности и безопасности деятельности в процессе выполнения миссии. В этих условиях применение ИИ как бортовым психологом, так и членами экипажа, перспективно для повышения эффективности психологической поддержки, психологической готовности экипажа к нештатной или аварийной, несущей угрозу жизни и здоровью ситуации. В настоящее время активно разрабатываются нейросетевые «виртуальные собеседники» (программы-чатботы) на основе ИИ для психологической помощи и психотерапии [URL: <https://smartreading.ru/bots/>], [URL: <https://www.wysa.com>].

Чатботы еще не используются в условиях реальных космических полетов, но рассматриваются на уровне планирования наземных экспериментов в качестве одного из средств психологической поддержки.

Развитие и внедрение в практику психологического обеспечения космических миссий подобных систем, способных вести диалог с пользователем, обученных классифицировать различные аспекты состояния, с учетом данных дополетного компьютерного моделирования [Поляков и др., 2003] может способствовать повышению уровня надежности и безопасности экипажа, поддержанию готовности космонавтов к экстремному реагированию, включающему:

- способность сохранять самообладание (стрессоустойчивость) и не допускать импульсивных поступков;
- слаженность взаимодействия при реализации принятого решения на устранение нештатной ситуации;
- понимание назначения и показаний на применение средств защиты и/или спасение;
- умение оказывать неотложную помощь по показаниям;
- четкое ведение коммуникации [Поляков и др., 2003].

## **Перспективы применения систем искусственного интеллекта в мониторинге состояния космонавтов**

Упомянутые технические и организационные особенности условий дистантного мониторинга определяют основные направления экспериментального развития ИИ в рассматриваемой области. Среди них: экспериментальные решения в области распознавания речи и жестов, экспериментальные модели рекомендательных систем, системы распознавания эмоциональной окраски речи, модели правдивости/ уверенности/ интонационной окраски, affective computing (эмоциональный ИИ).

Распознавание психоэмоционального состояния по голосу предполагает регистрацию громкости, тембра, спектральных характеристик. Имеют значение также темп речи и её ритмическое членение с помощью пауз и логиче-

ских ударений. Информативным источником психологической информации о человеке, о наличии у него проявлений нежелательных изменений психической активности могут быть результаты содержательного анализа речи. Существенными для выявления особенностей психического состояния здоровых людей при анализе переговоров могут быть такие содержательные параметры, как «эмоциональная окраска» высказывания, наличие шуток, способность выражать поддержку собеседникам, повторы информационных сообщений. По речевым характеристикам возможно выявление предвестников или коррелятов эмоциональной (стрессовой) напряженности, которая негативно влияет на работоспособность или вызывает нежелательные изменения в психической активности человека. Таким образом, проведение анализа ежедневной диалоговой речи в процессе переговоров позволяет определять психоэмоциональные реакции коммуникантов, отслеживать и оценивать их психологическое состояние [Гущин и др., 2020]. Одним из активно развиваемых направлений разработок являются интеллектуальные технологии по дистанционному распознаванию эмоциональных состояний и психической надежности по видеоизображению лица. Этому способствует, в частности, появление в открытом доступе ресурсов автоматической разметки на лице краудиометрических и кефалометрических точек, позволяющих обучать нейронные сети определять мимическую выразительность и, таким образом, создавать измерительные инструменты оценки психических состояний по видеоизображению. Применение этих инструментов показывает возможность оценки психоэмоционального состояния и психической надежности оператора [Жовнерчук и др., 2022], [Гущин и др. 2016].

## Заключение

Опыт модельных экспериментов свидетельствует об актуальности внедрения моделей ИИ, основанных на систематизированных и аннотированных данных по оценке выражений лица, двигательной активности, variability сердечного ритма, и других показателей, получаемых в моделируемых и реальных условиях деятельности. Такие модели, способные классифицировать психофизиологические особенности функциональных состояний, прогнозировать степень эффективности и безопасности деятельности членов космических экипажей, представителей других операторских профессий, могут стать значимым фактором повышения эффективности системы психологического обеспечения особых условий жизни и профессиональной деятельности.

## Список литературы

- [Вассерман и др., 2019] Вассерман Е.Л., Денисова Д.М., Рудницкий С.Б. Методы дистанционного бесконтактного измерения физиологических показателей человека для диагностики и мониторинга его функционального состояния //Авиакосмическая и экологическая медицина. – 2019. – Т. 53, №. 3. – С. 20-32.

- [Григорьев и др., 2017] Григорьев А.И., Потапов А.Н. Космическая физиология (рус.) // Вестник российского фонда фундаментальных исследований. – 2017. – № S1. – С. 21-38.
- [Гущин и др., 2020] Гущин В.И., Савинкина А.О., Швед Д.М. Анализ эмоционального состояния участников 4-х месячного изоляционного эксперимента с помощью методики оценки выражения лица FaceReader // Тезисы докладов IV Международной научно-практической конференции «Актуальные вопросы судебной психологической экспертизы и комплексной экспертизы с участием психолога. Современные компьютерные технологии в экспертной практике». – Калуга, 18-19 декабря 2020. – С. 9.
- [Гущин и др., 2016] Гущин В.И., Юсупова А.К., Швед Д.М., Чекалина А.Г., Шуева Л.В., Кузнецова П.Г., Виноходова А.Г., Карпова О.И., Чернова Е.Г. Методологические подходы к контент - анализу общения космического экипажа // Авиакосмическая и экологическая медицина. – 2016. – Т. 50, № 3. – С. 1-15.
- [Жовнерчук и др., 2022] Жовнерчук Е.В., Чичкалюк В.А., Жовнерчук И.Ю., Михайловский Н.Э., Мошкин В.В., Юрашку И.В. Методика определения психической надежности сотрудников транспортной безопасности с использованием машинного обучения // Психическое здоровье. – 2022. – Т. 17, № 12. – С. 3-10.
- [Иванов и др., 2022] Иванов А.В., Квасовец С.В., Бубеев Ю.А. Комплексное нейрокогнитивное тестирование в телемониторинге функционального состояния // Авиакосмическая и экологическая медицина. – 2022. – Т. 56, № 3. – С. 71-80.
- [Михайлов, 2001] Михайлов В.М. Гипокинезия как фактор риска в экстремальных условиях // Авиакосмическая и экологическая медицина. – 2001. – Т. 35, № 2.
- [Мясников и др., 2000] Мясников В.И., Степанова С.И., Сальницкий В.П., Козеренко О.П., Нечаев А.П. Проблема психической астенизации в длительном космическом полете. – М.: Фирма "Слово", 2000. – 224 с.
- [Поляков и др., 2003] Поляков А.В., Усов В.М., Крючков Б.И., Косачев В.Е., Михайлок М.В., Мотиенко А.И. Компьютерное моделирование жизнеугрожающих ситуаций и проведения аварийно-спасательных, медицинских и эвакуационных мероприятий на лунной базе // Авиакосмическая и экологическая медицина – 2003. – Т. 53, №. 3. – С. 13-19.
- [Розанов и др., 2022] Розанов И.А., Иванов А.В., Рюмин О.О., Бубеев Ю.А. Опыт применения виртуальной реальности для психологической коррекции в экспериментах с моделированием стрессоров космического полета // Методология современной психологии. Вып. 16 / под ред. Козлова В.В., Карпова А.В., Мазилова В.А., Петренко В.Ф. – М-Ярославль: ЯрГУ, ЛКИИСИ РАН, МАПН, 2022. – С. 333-344.
- [Lockard et al., 2019] Lockard E., Kaufman A. Bringing Nature into Space: The Restorative Potential of Virtual Environments for Long Term Space Travel // 49th International Conference on Environmental Systems. – 2019.
- [URL:<https://research.aimultiple.com/therapist-chatbot/>] <https://research.aimultiple.com/therapist-chatbot/> [Электронный ресурс] (дата обращения: 14.05.2023).
- [URL:<https://smartreading.ru/bots>] <https://smartreading.ru/bots> [Электронный ресурс] (дата обращения: 14.05.2023).
- [URL:<https://www.wysa.com>] <https://www.wysa.com> [Электронный ресурс] (дата обращения: 14.05.2023).



УДК 004.852

**МЕТОДЫ ПРОГНОСТИЧЕСКОЙ АНАЛИТИКИ  
В ОЦЕНКЕ ВЕРОЯТНОСТИ ВНУТРИГОСПИТАЛЬНОЙ  
ЛЕТАЛЬНОСТИ У БОЛЬНЫХ ИНФАРКТОМ  
МИОКАРДА ПОСЛЕ ЧРЕСКОЖНЫХ КОРОНАРНЫХ  
ВМЕШАТЕЛЬСТВ\***

Б.И. Гельцер (*boris.geltser@vvsu.ru*)

Дальневосточный федеральный университет, Владивосток

И.Г. Домжалов (*igor@domzhalov.ru*)

Владивостокский государственный университет, Владивосток

Н.С. Куксин (*Nikita.Kuksin@vvsu.ru*)

К.И. Шахгельдян (*carinash@vvsu.ru*)

Владивостокский государственный университет, Владивосток

Дальневосточный федеральный университет, Владивосток

Целью исследования является разработка объяснимой модели машинного обучения для прогнозирования внутригоспитальной летальности (ВГЛ) у больных инфарктом миокарда с подъёмом сегмента ST на электрокардиограмме (ИМпST) после операции чрескожного коронарного вмешательства (ЧКВ). В ходе работы проанализированы данные 4681 электронных историй болезни пациентов с ИМпST и выявлено 12 факторов риска ВГЛ. Были разработаны прогностические модели на основе методов многофакторной логистической регрессии, случайного леса и стохастического градиентного бустинга. Поиск пороговых значений на сетке при максимизации площади под ROC-кривой и их валидация методом аддитивных объяснений Шепли позволили верифицировать факторы риска ВГЛ. Модель, параметрами которой являлись факторы риска (AUC=0.917 на кросс-валидации и 0.903 при итоговом тестировании), превосходила по точности лучшую модель с непрерывными предикторами на основе стохастическо-

---

\* Работа выполнена при финансовой поддержке проекта FZNS-2023-0010 Госзадания Дальневосточного федерального университета (ДВФУ).

го градиентного бустинга. Использование в качестве предикторов факторов риска ВГЛ позволяет объяснить полученный прогноз и снизить риск неблагоприятных событий после ЧКВ.

**Ключевые слова:** прогностические модели машинного обучения, факторы риска, аддитивное объяснение Шепли, стохастический градиентный бустинг, поиск пороговых значений предикторов.

## Введение

Сердечно-сосудистые заболевания продолжают оставаться одной из наиболее актуальных проблем здравоохранения в большинстве стран мира [The World Health Organization, 2020]. Ишемическая болезнь сердца (ИБС) занимает одну из лидирующих позиций в структуре смертности населения от сердечно-сосудистых заболеваний и составляет 20% от всех летальных исходов [Ибанез и др., 2018]. К наиболее опасным клиническим формам ИБС относят инфаркт миокарда (ИМ) с подъемом сегмента ST на электрокардиограмме (ИМпST). Одним из наиболее эффективных методов лечения ИМпST является реваскуляризация миокарда посредством чрескожного коронарного вмешательства (ЧКВ), проведение которого необходимо в кратчайшие сроки от начала заболевания. Несмотря на совершенствование технологий ЧКВ внутригоспитальная летальность (ВГЛ) после ее выполнения остается высокой и варьирует от 4 до 7%.

В клинической практике применяются различные шкалы, позволяющие стратифицировать риск пациентов в отношении прогноза ВГЛ на основе анализа их клинических данных, результатов инструментальных и лабораторных исследований. К наиболее известным инструментам риск-метрии относятся шкалы TIMI, GRACE, PAMI, CADILLAC и другие [Granger et al., 2003], [Wilson et al., 2022], [Kao et al., 2020], [Bai et al., 2018], [Wang et al., 2021]. Чаще остальных на практике используется шкала GRACE (Global Registry of Acute Coronary Events), разработанная с помощью регрессии Кокса, совершенствование которой является целью ряда исследований последних лет [Wang et al., 2019], [Chen et al., 2020]. Для создания прогностических моделей на основе предикторов этой шкалы их структуру дополняют новыми переменными. При этом для моделирования ВГЛ всё чаще применяют более сложные алгоритмы машинного обучения (МО), в том числе случайный лес (СЛ) и стохастический градиентный бустинг (СГБ), которые позволяют достичь более высоких параметров качества моделей [Li et al., 2023]. Вместе с тем включение новых предикторов часто не поддерживается объяснением правил и оценкой степени их влияния на конечную точку, что ограничивает эффективность их применения в клинической практике.

Целью исследования является разработка объяснимой модели МО для прогнозирования ВГЛ у больных ИМпST после операции ЧКВ.

## 2. Материалы и методы

### 2.1. Данные

Проведено ретроспективное исследование, в рамках которого анализировали данные из 4681 историй болезни пациентов, проходивших лечение в Региональном сосудистом центре ГБУЗ «Приморская краевая клиническая больница № 1» г. Владивостока в период с 2015 по 2021 гг. Всем пациентам в первые сутки стационарного лечения была выполнена инвазивная коронарография с последующей транслюминальной баллонной ангиопластикой со стентированием инфаркт-связанных артерий. Из медицинской информационной системы DOCA+ с помощью парсинга html-файлов были извлечены все необходимые данные и сформирован датасет. Среди больных (3207 мужчин и 1474 женщины с медианой возраст 63 года) было выделено 2 группы лиц, первую из которых составили 318 (6.8%) человек, умерших в период госпитализации в стационаре, вторую – 4363 (93.2%) – с благоприятным исходом ЧКВ.

Клинико-функциональный статус больных оценивали в первые сутки стационарного лечения с помощью 136 факторов, основные из которых представлены в табл. 1. Помимо демографических (возраст, гендерный фактор), антропометрических (индекс массы тела – ИМТ) и анамнестических данных (фибрилляция предсердий – ФП, сахарный диабет второго типа – СД 2, хроническая болезнь почек – ХБП) анализировали систолическое (САД) и диастолическое артериальное давление (ДАД), частоту сердечных сокращений (ЧСС), а также показатели клинического анализа крови, концентрацию креатинина в сыворотке крови (Cr), уровень нейтрофилов и эозинофилов в крови, международное нормализованное отношение (МНО), протромбиновый индекс (ПТИ), тромбиновое время (ТВ), тромбокрит, уровень фибриногена (Фб), эритроцитов (RBC), гемоглобина (Hb), тромбоцитов (PLT), глюкозы (Glu), мочевины (Urea), лейкоцитов (WBC). Эхокардиографическое исследование проводилось с определением поперечного (LA1) и продольного (LA2) размеров левого и правого (RA1 и RA2) предсердий, конечного систолического (КСР) и диастолического (КДР) размеров левого желудочка (ЛЖ), фракции выброса (ФВ) ЛЖ методом Гейхольца, систолического давления в легочной артерии (СДЛА). Конечная точка исследования была представлена показателем ВГЛ больных после ЧКВ от всех причин в форме категориального бинарного признака («отсутствие» или «развитие»).

В связи с тем, что пациенты с ИМпСТ поступали в стационар экстренно, большая часть лабораторных исследований была выполнена сразу после ЧКВ. В датасете встречались пропуски, которые не заполнялись синтетическими значениями.

## 2.2. Дизайн исследования

Исследование состояло из 5 этапов. На первом из них, с помощью теста межгруппового сравнения на двух группах – умерших и выживших пациентов с ИМПСТ после ЧКВ формировался список потенциальных предикторов (табл. 1). Критерием включения являлось условие  $p$ -value  $< 0.05$ . Для непрерывных переменных использовали тест Манна-Уитни, а для категориальных –  $\chi^2$ . Отношение шансов (ОШ) и их 95% доверительные интервалы (ДИ) рассчитывали точным тестом Фишера.

На втором этапе определяли степень влияния изолированных потенциальных предикторов на конечную точку (табл. 2). Для этого оценивали весовые коэффициенты однофакторных моделей логистической регрессии (ЛР) по нормализованным данным.

На третьем этапе валидировали предикторы, используя базовую модель (baseline) из 5 предикторов шкалы GRACE и новые многофакторные модели. Последние разрабатывали путем пошагового включения в структуру базовой модели факторов, имеющих наибольший весовой коэффициент однофакторной ЛР, с последующей оценкой качества модели. Прогностическая значимость предиктора подтверждалась, если после его включения площадь под ROC-кривой (AUC) возрастала. Для разработки моделей датасет был разделен на 2 выборки: для обучения и кросс-валидации (80%) и для итогового тестирования (20%). Процедура обучения и кросс-валидации выполнялась методом Монте-Карло на 50 случайных выборках. Усредненные метрики качества AUC, чувствительность (Sen) и специфичность (Sp) использовали для выбора лучшей модели, подбирая предикторы и гиперпараметры. Порог отсечения для вычисления чувствительности и специфичности определяли путем поиска баланса между ними. Модели строили с помощью методов многофакторной ЛР (МЛ), СЛ и СГБ.

На четвертом этапе переменные, вошедшие в лучшую прогностическую модель, были преобразованы в факторы риска ВГЛ с помощью 4 методов: максимальное значение ОШ, минимальное значение  $p$ -value, определяемое с помощью теста  $\chi^2$ , наибольшее значение AUC в модели однофакторной ЛР и определение центроидов [Valente et al., 2021]. Факторы риска сравнивались с диаграммами аддитивного объяснения Шепли (SHAP), полученными на основе анализа изменения предсказания при изменении значения показателя и позволяющими оценивать степени влияния предикторов на конечную точку [Lundberg et al, 2017].

На завершающем этапе исследования была разработана прогностическая модель МЛР с использованием верифицированных факторов риска, прогностические свойства которой сравнивались с выше полученными моделями.

Анализ проводился на языке Python с открытым исходным кодом, версии 3.9.16.

### 3. Результаты

#### 3.1. Характеристики когорты пациентов и отбор предикторов

Первый и второй этапы анализа показателей включали медианные и частотные оценки, а также тесты на межгрупповые сравнения (табл. 1).

Таблица 1

#### Клинико-функциональная характеристика больных

Предиктор	Группа 1 (n=318)	Группа 2 (n=4363)	ОШ (95%) ДИ	p-value
Пол: Ж, абс. (%)	142 (44.65%)	1332 (30.5 %)	1.8 [1.5; 2.3]	<0.000001
Возраст, лет	71 [70; 72]	62[62;62]	-	<0.000001
Рост, см	168 [167;169]	170[169.8;170.3]	-	0.000001
Вес, кг	78 [76; 80]	80[79.5 80.5]	-	0.000005
ИМТ, кг/м <sup>2</sup>	26.8 [26.2;27.4]	27.7[27.5; 27.8]	-	0.039489
САД, мм рт.ст.	110 [106; 114]	130 [129;131]	-	<0.000001
ДАД, мм рт.ст.	72 [69.9;74.1]	80 [79.6; 80.4]	-	<0.000001
ЧСС, уд/мин	86 [83.3;88.7]	72 [71.6; 72.4]	-	<0.000001
Сг, мкмоль/л	130 [115;144]	97 [95.7; 98.2]	-	<0.000001
Класс ОШ по Т.Killip				
I	71 (22.33)	2745 (62.5%)	0.2[0.13;0.22]	<0.000001
II	58 (18.2%)	870 (20%)	0.9[0.67;1.20]	0.508052
III	66 (20.75%)	479 (11%)	2.1[1.59;2.8]	<0.000001
IV	123 (38.7%)	269(6.17%)	9.6[7.4;12.4]	<0.000001
GRACE, сумма баллов	203 [197; 211]	147 [147; 149]	-	<0.000001
ФВ ЛЖ, %	46.5[44.7; 48.3]	56 [55.7;56.3]	-	<0.000001
КДР ЛЖ, см	5[4.9;5.1]	5[4.98;5.02]	-	0.356921
КСР ЛЖ, см	3.7[3.6;3.8]	3.4[3.38;3.42]	-	<0.000001
СДЛА, мм рт.ст.	35[33.1;36.9]	28[27.8; 28.2]	-	<0.000001
LA1, см	4.10[4.02;4.18]	3.90[3.89;3.91]	-	<0.000001
LA2, см	5.20[5.09;5.31]	4.90[4.88;4.92]	-	<0.000001
RA1, см	3.80[3.71;3.89]	3.60[3.59;3.61]	-	<0.000001
RA2, см	4.80[4.70;4.90]	4.70[4.68;4.72]	-	0.00004
RBC, 10 <sup>12</sup> /л	4.25[3.96;4.53]	4.48[4.46;4.50]	-	<0.000001
Hb, г/л	132[129; 135]	141[140;142]	-	<0.000001
PLT, 10 <sup>9</sup> /л	228[215;241]	221[219;223]	-	0.020104
Нейтрофилы, %	81.3[79.9;82.7]	66.7[66.4; 67.1]	-	<0.000001
Эозинофилы, %	0.1[-0.02;0.22]	0.90[0.85;0.95]	-	<0.000001
Glucose, ммоль/л	7.95[6.89;9.01]	5.79[5.71;5.87]	-	<0.000001

Urea, мкмоль/л	12.12[7.8; 16.5]	6.7 [6.38;7.02]	-	<0.000001
Тромбокрит, %	0.22[0.21;0.23]	0.20[0.20;0.20]	-	0.0012
WBC, 10 <sup>9</sup> /л	14[13.2; 14.8]	10.5[10.3;10.6]	-	<0.000001
ПТИ, %	75.5[72.2; 78.8]	89.3[88.8;89.9]	-	<0.000001
МНО, ед.	1.26[1.10;1.42]	1.06[1.05;1.07]	-	<0.000001
Передний ИМ, абс. (%)	177(55.66%)	2023(46.37%)	1.5[1.15;1.83]	0.001647
ФП, абс. (%)	129(40.57%)	772(17.69%)	3.2[2.51;4.02]	<0.000001
СД 2, абс. (%)	99(31.13%)	831(19.05%)	1.9[1.50;2.46]	<0.000001
ХБП, абс. (%)	83 (26.1%)	677 (15.5%)	1.97 [1.5;2.6]	<0.000001

### 3.2. Обучение и валидация моделей

Для показателей, имеющих статистически значимые межгрупповые различия, были определены весовые коэффициенты однофакторной ЛР (табл. 2). Взаимосвязи с ВГЛ фиксировались при подъеме уровня показателей нейтрофилов, Ср, ЧСС, WBC, СДЛА, возраста больных, Glu, класса ОСН и др. Для предикторов САД, ДАД, ФВ ЛЖ, ПТИ, эозинофилов, гемоглобина и некоторых других факторов наблюдалась обратная зависимость: с увеличением значений этих переменных повышалась вероятность благоприятного исхода ЧКВ.

Таблица 2

#### Весовые коэффициенты в моделях однофакторной ЛР

Предиктор	Коэффициент	p-value
Нейтрофилы (отн. значение)	6.485 [6.481;6.489]	<0.000001
САД	-6.280 [-6.283;-6.277]	<0.000001
Ср	6.102 [6.100;6.103]	<0.000001
ЧСС	5.573 [5.570;5.575]	<0.000001
ФВ ЛЖ	-5.225 [-5.229;-5.221]	<0.000001
СДЛА	4.639 [4.637;4.642]	<0.000001
Эозинофилы (отн. значение)	-4.206 [-4.208;-4.204]	<0.000001
Возраст	4.171 [4.167;4.176]	<0.000001
Класс ОСН по Т.Killip	3.583 [3.575;3.590]	<0.000001
Нь	-3.088 [-3.091;-3.086]	<0.000001
Тромбокрит	2.063 [2.060;2.065]	<0.000001

Помимо базовой МЛР на основе 5 предикторов GRACE нами были разработаны модели МЛР, СЛ и СГБ с дополнительными факторами (табл. 3). Наилучший результат был достигнут при расширении спектра

предикторов за счет показателей ФВ ЛЖ, СДЛА, Нб, тромбокрит, относительного уровня нейтрофилов и эозинофилов крови и роста больных. Анализ данных демонстрировал устойчивое повышение качества прогноза для моделей МЛР, СЛ и СГБ как при кросс-валидации, так и при итоговом тестировании.

Таблица 3

### Оценка точности прогностических моделей

Метрики	Возраст, класс ОСН по Т.Killip, Cr, САД, ЧСС			Возраст, рост, класс ОСН по Т.Killip, ЧСС, Cr, САД, ФВ ЛЖ, СДЛА, Нб, тромбокрит, нейтрофилов и эозинофилов в крови		
	МЛР	СЛ	СГБ	МЛР	СЛ	СГБ
Усредненные метрики при кросс-валидации на валидирующих выборках						
AUC	0.855	0.853	0.859	0.915	0.919	0.915
Se	0.774	0.710	0.742	0.840	0.720	0.760
Sp	0.778	0.844	0.819	0.828	0.910	0.895
Метрики качества на итоговой тестовой выборке						
AUC	0.823	0.836	0.839	0.899	0.901	0.903
Se	0.711	0.711	0.689	0.771	0.686	0.714
Sp	0.772	0.836	0.831	0.846	0.899	0.896

### 3.3. Факторы риска

Определение факторов риска ВГЛ выполнялось методом поиска оптимального решения на сетке. Для таких показателей, как: возраст, класс ОСН по Т.Killip, ЧСС, Cr, САД, ФВ ЛЖ, СДЛА, Нб, тромбокрит, относительное значение нейтрофилов и эозинофилов в крови были определены пороговые значения с наибольшим предиктивным потенциалом, которые относили к факторам риска (табл. 4). Фактор риска принимает значение “1”, если значение предиктора превышает порог с постфиксом “+”, или не достигает его – с постфиксом “-”. В тех случаях, когда показатель имеет значение “0”, он не относится к факторами риска (табл. 4).

Таблица 4

**Определение факторов риска ВГЛ с помощью различных методов**

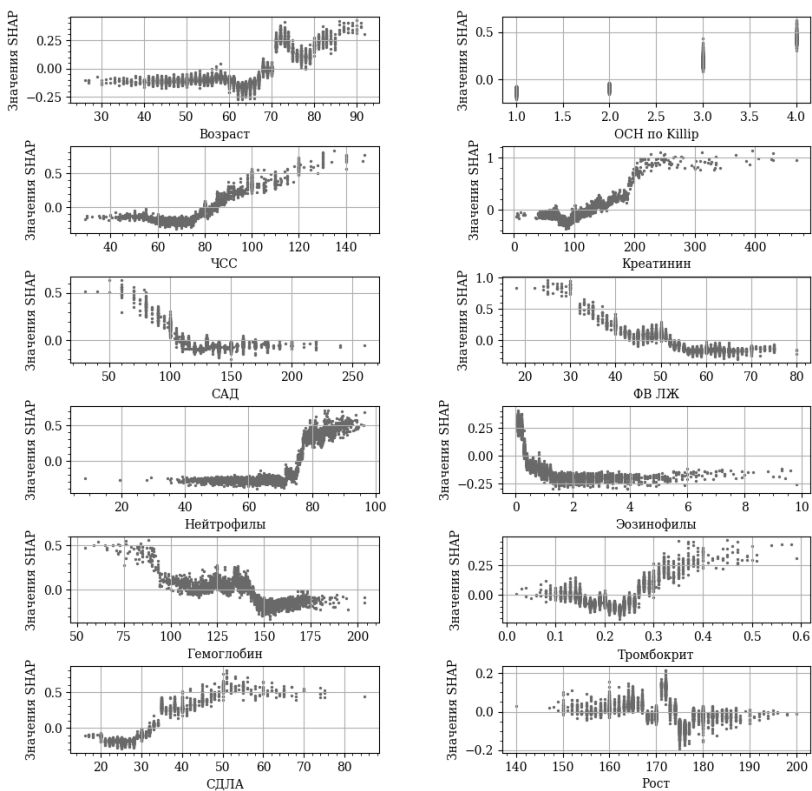
Предиктор	Метод	Порог	p-value	ОШ ДИ(95%)	AUC
Нейтрофилы	max(ОШ)	94.2+	<0.00001	23.01 [4.6;114.8]	0.513
	min(p-value)	78.8+	<0.00001	9.12 [6.63;12.53]	0.730
	max(AUC)	75.4+	<0.00001	11.28[7.83;16.2]	0.774
	Центроид	74.0+	<0.00001	9.92[6.86;14.36]	0.751
Эозинофилы	max(ОШ)	1.3-	<0.00001	9.79[5.3;18.08]	0.666
	min(p-value)	0.3-	<0.00001	7.87[5.62;11.03]	0.741
	max(AUC)	0.3-	<0.00001	7.87[5.62;11.1]	0.741
	Центроид	0.5-	<0.00001	7.630[5.239;11.112]	0.722
ФВ ЛЖ	max(ОШ)	31.0-	<0.00001	19.693[12.08;32.11]	0.514
	min(p-value)	31.0-	<0.00001	19.693[12.08;32.11]	0.594
	max(AUC)	51.0-	<0.00001	4.850[3.504;6.712]	0.690
	Центроид	51.25-	<0.00001	4.885[3.514;6.792]	0.683
Гемоглобин	max(ОШ)	94.0-	<0.00001	5.304[3.278;8.582]	0.540
	min(p-value)	94.0-	<0.00001	5.304[3.278;8.582]	0.541
	max(AUC)	144.0-	<0.00001	2.285[1.703;3.068]	0.614
	Центроид	136.5-	<0.00001	2.250[1.728;2.928]	0.585
Тромбокрит	max(ОШ)	0.36+	<0.00001	4.490[2.743;7.350]	0.500
	min(p-value)	0.36+	<0.00001	4.490[2.743;7.350]	0.524
	max(AUC)	0.22+	0.00002	1.816[1.379;2.391]	0.598
	Центроид	0.21+	0.0009	1.629[1.238;2.144]	0.576
СДЛА	max(ОШ)	34.0+	<0.00001	6.401[4.680;8.753]	0.500
	min(p-value)	34.0+	<0.00001	6.401[4.680;8.753]	0.728
	max( AUC)	34.0+	<0.00001	6.401[4.680;8.753]	0.728
	Центроид	31.5+	<0.00001	5.902[4.298;8.103]	0.701
Рост	max(ОШ)	152.0-	0.00003	4.174[2.118;8.228]	0.513
	min(p-value)	173.0-	<0.00001	1.890[1.451;2.464]	0.594



	max(AUC)	173.0-	<0.00001	1.890[1.451;2.464]	0.594
	Центроид	169.0-	0.00004	1.667[1.307;2.126]	0.559
САД	max(ОШ)	60.0-	<0.00001	31.538[10.89;91.35]	0.523
	min(p-value)	92.0-	<0.00001	11.014[8.28;14.65]	0.613
	max(AUC)	112.0-	<0.00001	5.361[4.237;6.783]	0.685
	Центроид	120.0-	<0.00001	4.777[3.781;6.035]	0.676
Креатинин	max(ОШ)	427.0+	<0.00001	30.523[10.16;91.74]	0.518
	min(p-value)	188.6+	<0.00001	13.054[9.44;18.04]	0.625
	max(AUC)	122.9+	<0.00001	5.774[4.477;7.448]	0.701
	Центроид	113.3+	<0.00001	4.605[3.559;5.960]	0.693
ЧСС	max(ОШ)	150.0+	<0.00001	41.635[4.32;401.42]	0.500
	min(p-value)	94.0+	<0.00001	6.225[4.860;7.972]	0.663
	max(AUC)	79.0+	<0.00001	4.019[3.170;5.096]	0.672
	Центроид	79.0+	<0.00001	4.019[3.170;5.096]	0.672
Возраст	max(ОШ)	85.0+	<0.00001	6.194[3.407;11.264]	0.513
	min(p-value)	70.0+	<0.00001	4.109[3.259;5.181]	0.644
	max(AUC)	65.0+	<0.00001	3.735[2.923;4.772]	0.667
	Центроид	66.5+	<0.00001	3.646[2.872;4.628]	0.660
Класс ОШ по Т.Killip	max(ОШ)	3.0+	<0.00001	9.600[7.423;12.416]	0.665
	min(p-value)	3.0+	<0.00001	9.600[7.423;12.416]	0.665
	max(AUC)	2.0+	<0.00001	7.081[5.585;8.977]	0.714
	Центроид	2.0+	<0.00001	7.081[5.585;8.977]	0.714

Факторы риска ВГЛ, полученные с помощью оптимизации целевой функции, формировались по оценке степени влияния изолированного предиктора на конечную точку. Вместе с тем при влиянии на результирующую переменную комплекса факторов их пороговые значения могут смещаться. Для определения пороговых значений в многофакторной модели использовали оценку важности предикторов с помощью метода аддитивного объяснения Шепли. Были созданы диаграммы значений SHAP для каждого из предикторов лучшей многофакторной модели СГБ (рис. 1).

обходимо также отметить, что пороговые значения факторов риска, полученные с помощью максимизации AUC однофакторной ЛР, соответствовали результатам оценки по методу Шепли. Так, значения SHAP больше 0 указывает на возрастающую вероятность ВГЛ. Например, максимизация AUC обеспечивала пороговое значение нейтрофилов как фактора риска на уровне 75%, что ассоциировалось с увеличением риска ВГЛ в 11 раз (табл. 4). При значении содержания в крови нейтрофилов более 75% SHAP резко возрастает и становится устойчиво больше 0 (рис. 1). Аналогичная взаимосвязь наблюдалась и для других предикторов.



*Рис. 1. Взаимосвязь предикторов ВГЛ со значениями SHAP*

Используя полученные с помощью максимизации AUC факторы риска, авторы разработали многофакторную модель прогноза ВГЛ с метриками качества при кросс-валидации: AUC=0.917, Sen=0.821, Sp=0.854, а

при итоговом тестировании:  $AUC = 0.903$ ,  $Sen=0.775$ ,  $Sp=0.851$ , что соответствует лучшей модели СГБ. В МЛР получены весовые коэффициенты при соответствующих факторах риска:  $w_{возраст}=0.902$ ,  $w_{рост}=0.49$ ,  $w_{Killip}=0.779$ ,  $w_{ЧСС}=1.19$ ,  $w_{САД}=0.877$ ,  $w_{Cr}=0.966$ ,  $w_{ФВ}=0.584$ ,  $w_{СДЛА}=0.819$ ,  $w_{Hb}=0.297$ ,  $w_{тромбокрит}=0.357$ ,  $w_{нейтрофилы}=0.566$ ,  $w_{эозинофилы}=1.16$  и свободный член  $e=-6.35$ . Преимуществом этой модели является ее объяснимость для клинических врачей.

К выделенным факторам риска ВГЛ у пациентов с ИМПСТ после ЧКВ относятся: уровни содержания в крови нейтрофилов более 75.4% и эозинофилов более 0.3%; ФВ ЛЖ менее 51%, уровня Hb в крови менее 144 г/л; уровня тромбокрита в крови более 0.22 %, СДЛА более 34 мм рт.ст.; рост менее 173 см, уровня САД менее 112 мм рт.ст.; креатинина в крови более 123 мкмоль/л, ЧСС более 79 уд/мин; возраст старше 65 лет; III-IV класс ОСН по T.Killip.

#### 4. Обсуждение

В работе был проведен интеллектуальный анализ данных на датасете пациентов с ИМПСТ после ЧКВ с целью поиска и валидации новых предикторов ВГЛ, верификации среди них факторов риска фатального исхода, разработки прогностической модели, превышающей по точности известную модель GRACE, и обладающей элементами объяснения. Выделенный пул потенциальных предикторов включал 5 факторов шкалы GRACE, а также ряд новых показателей, которые ранее не рассматривались в качестве предикторов. С помощью 3-х методов МО, процедур кросс-валидации и тестирования подтверждена их прогностическая значимость в отношении ВГЛ. Наилучшая точность прогноза была достигнута с помощью модели СГБ. При этом все три метода (МЛР, СЛ и СГБ) показывали устойчивое повышение качества прогноза как при кросс-валидации, так и при итоговом тестировании.

Метод аддитивного объяснения Шепли демонстрировал, что высокую степень влияния на ВГЛ помимо параметров GRACE (креатинина, ЧСС, САД и класс ОСН по Killip) оказывали также вновь выделенные предикторы: ФВ ЛЖ, СДЛА, уровни нейтрофилов, эозинофилов и тромбокрита. Сила влияния на конечную точку гемоглобина была соизмерима с возрастом больных, а влияние роста пациентов было минимальным. Повышенные значения параметров Cr, ЧСС, нейтрофилов, СДЛА, тромбокрита и возраста у больных с ИМПСТ увеличивают, а низкие у САД, ФВ ЛЖ и эозинофилов уменьшают вероятность смерти после ЧКВ (рис. 2).

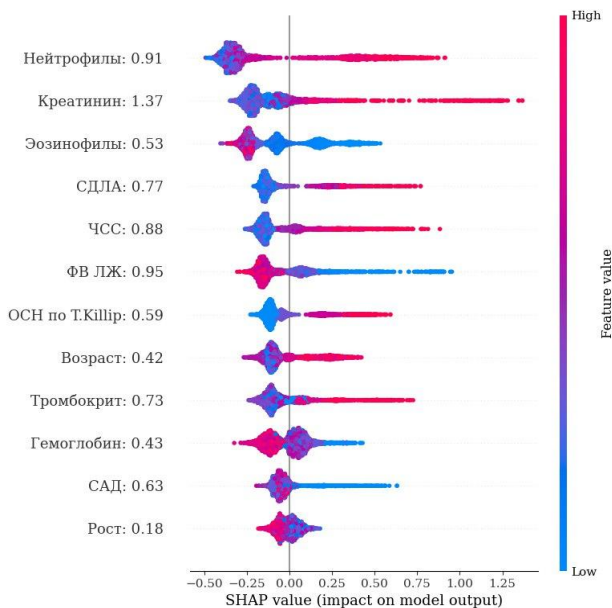


Рис. 2. Диаграмма значений SHAP по модели СГБ

С помощью максимизации AUC однофакторной ЛР, подтвержденного результатами аддитивного объяснения Шепли, были выделены пороговые значения предикторов, отклонение от которых относили к факторами риска ВГЛ. Данный подход позволяет врачам в реальном времени оценивать факторы риска ВГЛ и своевременно осуществлять их коррекцию.

Точность прогноза разработанной нами модели с факторами риска ВГЛ превосходит результаты, полученные другими авторами (авторами (AUC=0.891 и 0.913 при кросс-валидации) [Li et al., 2023], [Wang et al., 2022]).

### Заключение

В настоящем исследовании были выделены и верифицированы факторы риска ВГЛ у пациентов с ИМПСТ, оперированных методом ЧКВ. Разработанная на их основе модель имела лучшие прогностические свойства по сравнению с классической моделью GRACE, а также с моделями, представленными другими исследователями. Использование в предикторов качестве факторов риска ВГЛ обеспечивает прозрачность вычислений и позволяет практикующим врачам своевременно реализовывать мероприятия по снижению риска фатальных исходов.

## Список литературы

- [Ибанез и др., 2018] Ибанэз Б., Джеймс С., Аджевал С., Энтунэс М.Д., Баккиарэли-Дукки Ч., Бюэно Е., Кафорно А.Э.П., Крэа Ф., Гоудэвэнос Д. Э., Халворсэн С., Хиндрикс Д., А. Кастрати, Лензэн М.Д., Прэскотт И., Роффи М., Валгимигли М., Варэнхорст К., Вранккс П., Видимски П. Рекомендации ЕОК по ведению пациентов с острым инфарктом миокарда с подъемом сегмента ST // Российский кардиологический журнал. – 2018. – № 23 (5).
- [Bai et al., 2018] Bai X.F., Zhang Y.P., Zhou J., Wu Y., Li R.F., Sun L.Z., Ma Q.Q., Lou B.W., Zhai B.W., Liu M.P., Cheng L.L., Tong X.N., Yuan Z.Y. Combination of the CYP2C19 metabolizer and the GRACE risk score better predicts the long-term major adverse cardiac events in acute coronary syndrome undergoing percutaneous coronary intervention // *Thromb Res.* – 2018. – No. 170.
- [Chen et al., 2020] Chen X., Shao M., Zhang T., Zhang W., Meng Y., Zhang H., Hai H., Li G. Prognostic value of the combination of GRACE risk score and mean platelet volume to lymphocyte count ratio in patients with ST-segment elevation myocardial infarction after percutaneous coronary intervention // *Exp Ther Med.* – 2020. – No. 19 (6).
- [Fox et al., 2010] Fox K.A.A., Eagle K.A., Gore J.M., Steg Ph.G., Anderson F.A. The Global Registry of Acute Coronary Events, 1999 to 2009 – GRACE. *Heart* 2010. – 96 (14). – P. 1095-101.
- [Granger et al., 2003] Granger C.B., Goldberg R.J., Dabbous O., Pieper K.S., Eagle K.A., Cannon C.P. et al. Predictors of hospital mortality in the global registry of acute coronary events // *Arch Intern Med.* – 2003. – No. 163 (19).
- [Kao et al., 2020] Kao Y.T., Hsieh Y.C., Hsu C.Y., Huang C.Y., Hsieh M.H., Lin Y.K., Yeh J.S. Comparison of the TIMI, GRACE, PAMI and CADILLAC risk scores for prediction of long-term cardiovascular outcomes in Taiwanese diabetic patients with ST-segment elevation myocardial infarction: From the registry of the Taiwan Society of Cardiology // *PLoS One.* – 2020. – No. 15 (2).
- [Li et al., 2023] Li R., Shen L., Ma W., Yan B., Chen W., Zhu J., Li L., Yuan J., Pan C. Use of machine learning models to predict in-hospital mortality in patients with acute coronary syndrome // *Clin Cardiol.* – 2023. – No. 46 (2).
- [Lundberg et al., 2017] Lundberg S.M., Su-In Lee. A unified approach to interpreting model predictions // *Proceedings of the 31st International Conference on Neural Information Processing Systems.* – 2017. – P. 4768-4777.
- [The World Health Organization, 2020] The World Health Organization The top 10 causes of death [Электронный ресурс]. – URL: <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death> (дата обращения: 02.05.2023).
- [Valente et al., 2021] Valente F., Henriques J., Paredes S., Rocha T., de Carvalho P., Morais J. A new approach for interpretability and reliability in clinical risk prediction: Acute coronary syndrome scenario // *Artif Intell Med.* – 2021. – No. 117.
- [Wang et al., 2019] Wang J.J., Fan Y., Zhu Y., Zhang J.D., Zhang S.M., Wan Z.F., Su H.L., Jiang N. Biomarkers enhance the long-term predictive ability of the KAMIR risk score in Chinese patients with ST-elevation myocardial infarction // *Chin Med J (Engl).* – 2019. – No. 132 (1).

- [Wang et al., 2021] Wang L., Zhang Z., Hu T. Effectiveness of LODS, OASIS, and SAPS II to predict in-hospital mortality for intensive care patients with ST elevation myocardial infarction // *Sci Rep.* – 2021. – No. 11 (1).
- [Wang et al., 2022] Wang Y., Wang W., Jia S., Gao M., Zheng S., Wang J., Dang Y., Li Y., Qi X. Development of a nomogram for the prediction of in-hospital mortality in patients with acute ST-elevation myocardial infarction after primary percutaneous coronary intervention: a multicentre, retrospective, observational study in Hebei province, China // *BMJ Open.* – 2022. – No. 12 (2).
- [Wilson et al., 2022] Wilson R.S., Malamas P., Dembo B., Lall S.K., Zaman N., Peterson B.R. The CADILLAC risk score accurately identifies patients at low risk for in-hospital mortality and adverse cardiovascular events following ST elevation myocardial infarction // *BMC Cardiovasc Disord.* – 2021. – No. 21 (1).

## СОЦИАЛЬНО-ЭМОЦИОНАЛЬНЫЕ РАЗГОВОРНЫЕ АГЕНТЫ НА ОСНОВЕ КОГНИТИВНЫХ АРХИТЕКТУР И МАШИННОГО ОБУЧЕНИЯ\*

А.А. Долгих (*tolick.dolgi2013@yandex.ru*)

А.В. Самсонович (*alexei.samsonovich@gmail.com*)

Д.В. Тихомирова (*dvsulim@mail.ru*)

Национальный исследовательский ядерный университет «МИФИ»,  
Москва

Большие языковые модели (Large Language Models, LLM) могут распознавать тематику произвольных высказываний, а также их эмоциональную окраску, но не понимают логику эмоций несмотря на то, что могут генерировать часто адекватные ответы в заданном контексте. С другой стороны, такие когнитивные архитектуры как eBICA способны моделировать динамику эмоциональных состояний в общем случае, но требуют помощи для понимания смысла высказываний и генерации ответов на них. В данной работе представлен способ интеграции LLM и eBICA, позволяющий им дополнить друг друга. Исследованы парадигмы «Виртуальный рецепционист» и «Виртуальный психодиагност», результаты обнадеживают.

**Ключевые слова:** LLM, когнитивные архитектуры, эмоциональный интеллект, искусственный интеллект.

### Введение

Интеллектуальные разговорные агенты, способные поддерживать разговор с пользователем и обладающие элементами эмоциональности, играют все большую роль в жизни общества, и спрос на них стремительно растет. Среди примеров их применений – персональный агент-собеседник [Ali, 2020], виртуальный тьютор [Hartholt, 2020], реабилитационный разговорный агент [Diez Díaz et al., 2022], робот-рассказчик, стимулирующий развитие творческих способностей у детей [Elgarf et al., 2021], виртуальный пациент для обучения врачей [Blache, 2020] и другие. В то же время,

---

\* Работа выполнена при финансовой поддержке РФФ (грант № 22-11-00213, <https://rscf.ru/project/22-11-00213/>).

существующим диалоговым агентам (включая Siri, Алису, ChatGPT и т.д.) недостает социально-эмоционального интеллекта, который позволил бы им быть принятыми человеком на социальном уровне.

Задача, в решение которой дает вклад данная работа, заключается в создании искусственного социального агента, который, кроме выполнения функций в своей предметной области, способен (а) устанавливать и поддерживать контакт с пользователем на социальном уровне не ниже уровня человека, и (б) внутренне оценивать пользователя социально, с тем чтобы более эффективно с ним работать. Для определенных областей применения данные способности критичны.

Сегодня социальные разговорные агенты реализуются либо на базе статистических методов машинного обучения, прежде всего больших языковых моделей (Large Language Models, LLM) на основе глубоких нейросетей, либо на базе когнитивных архитектур, таких как ACT-RE, расширенный Soar, EMA [Marsella et al., 2009], eBICA [Samsonovich, 2013], [Samsonovich, 2020] и других. В том и другом случае трудности вызывает выбор адекватного поведения в непредвиденном социальном контексте. Когнитивные архитектуры приходится создавать вручную, и при этом нельзя предвидеть все возможные ситуации при их применении. Для обучения же статистических моделей нужны большие объемы данных, которые в сфере социального принятия решений обычно труднодоступны.

В настоящей работе развит комбинированный подход: а именно, способ интеграции методов машинного обучения и когнитивного моделирования, позволяющий им дополнить друг друга. Результат демонстрируется на парадигмах «Виртуальный рецепционист отеля» и «Виртуальный психодиагност».

## **1. Принципы когнитивной архитектуры eBICA**

Описанная неоднократно (например, [Samsonovich, 2013], [Samsonovich, 2020]) когнитивная архитектура eBICA не представляет собой большую воплощенную программную среду со своим языком программирования, подобно Soar, Act-R, Icarus. Мы преднамеренно пошли по другому пути, развивая eBICA как теоретический фреймворк, имеющий на сегодня десятки специализированных воплощений, но постоянно меняющийся и расширяющийся в том числе концептуально, адаптируясь к новым задачам и предметным областям. Такая гибкость была бы невозможна при реализации его как единой большой системы подобно Soar. В то же время в данном случае именно теоретическое понимание принципов решения представляет основную ценность для практики, позволяя создать технологическую линию разработки социально-эмоциональных агентов для разнообразных приложений.



Ключевыми элементами eVICA, помимо типичного для когнитивных архитектур набора систем памяти, являются семантические карты, мысленные перспективы, и моральные схемы [Samsonovich, 2020]. Принцип работы моральной схемы состоит в определении некоторого «нормального» положения дел в данном социальном окружении и в использовании доступных агенту средств для поддержания либо достижения такого «нормального состояния». Минимальное упрощенное объяснение базовых принципов построения теории eVICA дано в следующих параграфах.

Первый шаг состоит во введении оценок для всех значимых событий и акторов, действующих в окружении. Оценки вводятся как вектора, принимающие значения в семантическом пространстве, определенном набором актуальных для данной парадигмы семантических шкал (например, в модели VAD это Valence, Arousal, Dominance: т.е., «положительность», «возбужденность» и «доминантность»). Пусть  $X$  и  $Y$  – вектора оценок двух акторов, также обозначенных буквами  $X$  и  $Y$ . Предположим, актер  $X$  совершает дискретное действие  $a$  по отношению к актору  $Y$ . Пусть  $a$  и  $a^+$  – оценки действия  $a$ , определенные как матожидания эффекта действия  $a$  на  $X$  и  $Y$  соответственно. Постулируем следующее правило обновления оценок  $X$  и  $Y$  при совершении действия  $a$ :

$$X := (1-rw)X + rwa, \quad (1)$$

$$Y := (1-rw)Y + rwa^+. \quad (2)$$

Здесь  $r$  – константа, параметр модели, причем  $0 < r \ll 1$ , а  $w$  – это положительная величина, характеризующая значимость действия  $a$ . Таким образом, каждое действие характеризуется двумя векторами  $a$ ,  $a^+$  и скаляром  $w$ , причем эти величины могут зависеть от контекста.

Конфигурацию моральной схемы в eVICA определяют чувства к актерам – участникам схемы. Понятие чувства  $F_X$  вводится как значения оценки актора  $X$  в нормальном состоянии [Samsonovich, 2020]. Тогда можно показать, что для поддержания нормального состояния актору достаточно выбирать действия, оценки которых по возможности близки к значениям соответствующих чувств и симметрично распределены относительно них. В состоянии, близком к нормальному, чувства фиксированы. Они могут меняться в конфликтной ситуации, когда расхождение оценок и чувств существенно и не может быть устранено выбором допустимых действий при фиксированных чувствах. Условия активации и деактивации моральных схем описаны в [Samsonovich, 2020]. В то время как оценки и чувства имеют отношение к системе ценностей, существенную роль в выборе поведения играют также физиологические драйвы, имеющие отношение к соматической области.

Оценки и чувства могут быть разными в мысленных перспективах разных акторов. Тем не менее, проверенные экспериментально [Samsonovich, 2013], [Tikhomirova, 2020] уравнения (1), (2) вместе с законами моральных схем [Samsonovich, 2020] позволяют рассчитать динамику межличностных эмоциональных отношений и таким образом предсказать психические состояния агентов на основе их наблюдаемого поведения.

В данной работе eVICA была использована для генерации наборов данных для обучения нейросетевых компонентов, как описано ниже. Это позволило агенту адекватно реагировать на поведение пользователя.

## 2. Экспериментальные парадигмы

В рамках парадигмы виртуального регистратора отеля определяется следующая модельная ситуация: посетитель приходит заселяться в отель и на стойке ресепшена между ним и регистратором происходит диалог. Предметом рассмотрения здесь является динамика эмоционального взаимодействия пользователя и виртуального регистратора, причем в роли последнего может выступать как виртуальный актер (бот), так и скрытый живой человек (конфедерат).

Исследовательские вопросы таковы: можно ли считать виртуального актора социально приемлемым для пользователя? Заслуживающим доверие? Привлекательным? Способным вызывать эмоции? Имеющим индивидуальность? А также, можно ли определить характеристики личности человека по его поведению в данной парадигме? В итоге, можно ли сказать, что виртуальный актер в данной парадигме устанавливает социальное-эмоциональное взаимодействие на уровне человека?

Мотивация выбора парадигмы «Виртуальный психодиагност» может быть объяснена следующим образом. В настоящее время существует большое количество опросников, благодаря которым пользователь может узнать свои личностные характеристики (например, [Truity, 2023]), но для этого ему необходимо ответить на большое количество скучных вопросов. Парадигма «Виртуальный психодиагност» – это игра, в которой пользователю предлагается принять участие в диалоге, проходящем в воображаемой жизненной ситуации (например, заказ обеда в ресторане, выбор мебели в магазине и т.п.) для того, чтобы узнать свой тип личности. Главные достоинства с точки зрения пользователя – доступность, этичность, увлекательность и легкость прохождения теста при условии анонимности. Исследовательские вопросы те же.

### 3. Разработка и реализация систем

#### 3.1. Виртуальный регистратор

Структурно виртуальный регистратор можно разделить на две составляющие: графический интерфейс, реализованный средствами виртуальной реальности, и серверная компонента, управляющая диалогом с пользователем. Такая структура помогает отделить визуализацию регистратора от построения диалога с пользователем, что обеспечивает возможность интеграции сервера не только с виртуальным окружением, но и с другими средствами взаимодействия с пользователями.

Серверная часть (рис. 1) – основная часть виртуального регистратора. Она включает компонент классификации реплик пользователя, компонент генерации ответов регистратора и компонент обработки реплик (последний выполняет функцию автоматической правки текста).

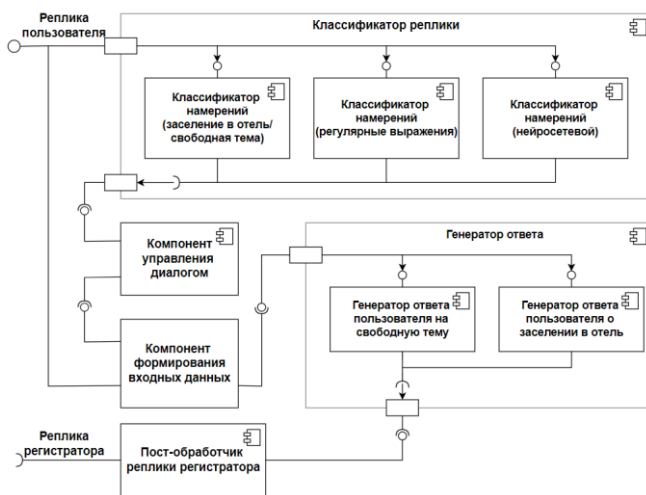


Рис. 1. Структура серверной части виртуального регистратора

Компонент классификации последовательно вызывает несколько классификаторов. В первую очередь, необходимо определить, к какому типу диалога относится данная реплика. В рамках парадигмы «Виртуальный регистратор» выделяется два типа диалога: диалог о заселении в отель и диалог на свободную тему. Так как в рамках диалога о заселении в отель значительно сужается список возможных фраз, данная классификация выполняется путем поиска ключевых слов в реплике пользователя. Далее происходит классификация некоторых специфических реплик, таких как

фамилия, имя и отчество пользователя. Для решения этой проблемы был использован классификатор на основе правил: проверялось наличие имени в заранее подготовленном списке.

Наконец, выполнялась классификация по намерению пользователя. Для этого использовался нейросетевой классификатор. Далее для выявления намерения по фразе пользователя была дообучена модель *rubert-base-cased-sentence*, которая входит в библиотеку *deeppavlov*. Данная модель построена на архитектуре BERT. Для обучения классификатора был создан набор данных из 231 реплики пользователя, разделенных на 5 категорий – приветствие, завершение диалога, бронирование номера, получение номера, общие вопросы. В качестве функции потерь использовалась кросс-энтропия. Рассчитанные метрики после дообучения нейронной сети в течение пяти эпох таковы:  $F1(\text{macro})=0.8902$ ,  $F1(\text{micro})=0.8972$ ,  $F1(\text{weighted})=0.8922$ .

Следующим важным компонентом является генератор ответов регистратора. Он был реализован с использованием двух нейронных сетей. Одна для генерации диалога на свободную тему, а другая – для генерации диалога на тему заселения в отель.

Нейронная сеть для генерации ответа регистратора на тему заселения в отель основана на архитектуре Трансформер. Для ее реализации сети были выбраны следующие параметры:

- размер эмбединга (обычного и для позиции слова): 256;
- число голов в механизме внимания: 8;
- число блоков кодировщиков: 6;
- число блоков декодировщиков: 6;
- размер одной порции данных, поступающих на вход нейронной сети (batch): 128.

Для получения результирующей последовательности токенов использовался алгоритм жадной генерации (*greedy decoding*). Особенность алгоритма, как можно судить из его названия, заключается в том, что на каждом шаге генерации выбирается тот токен, вероятность которого наиболее высока на данном шаге. Данный алгоритм хорошо работает для генерации небольших последовательностей слов, где длина предложений невелика.

Общее число обучаемых параметров составило 5686725. Причем для обучения нейронной сети было сгенерировано 10,000 диалогов с помощью алгоритма, построенного на базе когнитивной архитектуры eVICA (см. раздел 1). Данная когнитивная модель работала с заранее написанным сценарием диалога, допускающим небольшие вариации. Вариации состояли в выборе вариантов высказываний как пользователя, так и регистратора. Каждый вариант каждого высказывания был заранее оценен испытуемыми по шкалам «Валентность», «Доминантность» и «Возбужден-

ность» в отдельном эмпирическом исследовании. В итоге eVICA могла определять выбор одного из нескольких вариантов каждого высказывания для каждого виртуального собеседника, руководствуясь моделью развивающихся социальных отношений между собеседниками. Таким образом был получен набор синтетических диалогов. Каждый диалог был разбит на пары <фраза пользователя, фраза регистратора>. После этого было проведено обучение нейронной сети на наборе данных, состоящем из таких пар, в течение 20 эпох.

Для генерации ответов на вопросы на свободные темы была использована предобученная нейронная сеть, натренированная на наборах данных, составленных из русских диалогов. Ее, в силу большого количества параметров, не дообучали.

Функции компонента визуализации виртуального окружения включают преобразование реплики, полученной от пользователя – в текст, а реплики, полученной от серверной части – в голосовое сообщение, и также обмен данными с серверной частью.

### **3.2. Виртуальный психодиагност**

Система, реализующая Виртуального психодиагноста, была описана нами ранее в позиционной статье [Anisimova et al., 2022]. Здесь мы представляем предварительные результаты ее реализации, пока находящейся на промежуточной стадии. В частности, eVICA пока не использовалась для управления поведением психодиагноста, хотя возможность и целесообразность ее использования здесь вполне понятны.

Виртуальный психодиагност ведет беседу с пользователем по нескольким заданным сценариям (см. раздел 2), адресуя пользователю заранее написанные вопросы. Использование робота Ф-2, разработанного в лаборатории А. Котова в НИЦ «Курчатовский институт», позволяет озвучивать текст и выражать заданные эмоции. Робот Ф-2 [Котов, 2023] имеет возможность выражать большое число эмоций и коммуникативных намерений. Программный интерфейс на основе BML (Behavior Markup Language) позволяет работать с базой данных паттернов эмоциональной динамики. Для целей проекта нами с помощью испытуемых были оценены по эмоциональным шкалам жесты робота (ограниченный объем статьи не позволяет нам привести результаты данных оценок). Примеры жестов показаны на рис. 2. Робот также доступен в виде виртуального аватара.



*Рис. 2. Примеры жестов, выполняемых роботом Ф-2*

Основная сложность в данном подпроекте заключалась в анализе высказываний пользователя с целью определения типа личности человека. Задача решалась применением методов машинного обучения. Сегодня алгоритмы машинного обучения используются для создания моделей, которые классифицируют типы личности по модели Big5 на основе данных ответов на опросник, а также на основе произвольного текста большого объема (20 тысяч слов), написанного человеком [Martin, 2019], однако доступ к данной проприетарной системе ограничен.

Наша задача здесь состоит в обучении системы распознаванию типов личности на основе диалогов в рамках выбранных сценариев. Из-за отсутствия больших объемов данных нужного типа обучение производилось на основе массива данных открытого доступа, включающего в себя результаты теста Big5 отдельных испытуемых и текст, написанный ими [Kosinski, 2016]. Для реализации алгоритмов машинного обучения использовалась библиотека scikit-learn для Python. Для предобработки данных использовалась база данных “NRC Word-Emotion Association Lexicon” открытого доступа [Mohammad, 2013], [Mohammad, 2022], состоящая из 14,155 слов с оценками по категориям валентности (отрицательный окрас, положительный окрас) и эмоций, таких как гнев, предвкушение, отвращение, страх, радость, печаль, удивление, доверие, а также модель эмбединга GloVe.

## **4. Эксперимент: Материалы и методы**

В данном исследовании приняли участие 16 студентов НИЯУ МИФИ в возрасте от 20 до 22 лет обоих полов в равной пропорции. Тестирование реализованных систем подтвердило их соответствие функциональным требованиям и целям исследования.

### **4.1. Виртуальный регистратор**

Процедура эксперимента в парадигме «Виртуальный регистратор отелей» включала в себя три этапа.

*На первом этапе* участник эксперимента на протяжении двух минут оценивал внешний вид и окружение агента-регистратора в виртуальном пространстве, после чего проходил опросник Artificial Social Agent Questionnaire (ASA) [Fitriane et al., 2022], отвечая на вопросы, посвященные первому впечатлению об агенте.

Опросник ASA – это инструмент для оценки взаимодействия человека с виртуальным агентом, разработанный международной рабочей группой исследователей в области искусственного интеллекта. Нами использовался краткий вариант опросника, позволяющий составить профиль искусственного социального агента по пяти шкалам: привлекательность агента, его социальная приемлемость, доверие к агенту, наличие у агента индивидуальности, и способность вызвать эмоции у пользователя.

Регистратором в данном исследовании мог управлять как виртуальный актер, так и скрытый в другой комнате человек-конфедерат, который вводил текст своих вопросов на клавиатуре. В роли конфедерата были поочередно использованы два участника. Остальные участники эксперимента не знали, что в роли регистратора мог быть человек.

*На втором этапе* участник проходил процедуру регистрации в виртуальной реальности. Участнику озвучивались вопросы регистратора путем синтеза речи, речь участника распознавалась – происходил диалог, как описано в разделе 3. Эксперимент длился до завершения оформления номера регистратором. Если оформление номера не происходило в течение 10 минут, эксперимент завершался. В конце второго этапа участник снова проходил краткий вариант опросника ASA полностью, оценивая собеседника после взаимодействия с ним.

*На третьем этапе* участник снова проходил ту же самую процедуру регистрации в виртуальной реальности. Третий этап отличался от второго лишь тем, что если на втором этапе регистратором управлял конфедерат, то на третьем этапе им управлял виртуальный актер. И наоборот, если на втором этапе регистратором управлял виртуальный актер, то на третьем этапе им управлял конфедерат. По завершении третьего этапа участник эксперимента также проходил опросник.

#### **4.2. Виртуальный психодиагност**

Эксперимент в парадигме «Виртуальный психодиагност» проходил в форме беседы испытуемого с роботом по заранее написанному сценарию, как описано выше. После нескольких общих вопросов, задаваемых роботом, испытуемый и робот играли в игру в форме диалога (раздел 2). При этом фразы, озвучиваемые роботом, сопровождалась жестами, мимикой, взглядом и языком тела. К каждой фразе был подобран жест, соответствующий намерению и эмоциональной окраске фразы, согласно ранее установленным эмоциональным окраскам жестов (см. раздел 3.2).

## 5. Результаты и анализ

### 5.1. Виртуальный регистратор

В результате проведенного исследования были получены и сопоставлены психологические профили виртуального регистратора до и после общения с ним, а также профили регистратора, управляемого виртуальным агентом, и регистратора, управляемого человеком. Найдено, что впечатление о виртуальном регистраторе формируется постепенно, и до общения с ним впечатление определяют преимущественно характеристики внешности. После общения с регистратором на оценку влияет его поведение.

На рис. 3 представлены результаты сравнения профилей регистратора, управляемого виртуальным агентом до и после общения испытуемого с ним, а также профилей после общения с регистратором, управляемым виртуальным агентом конфедератом.

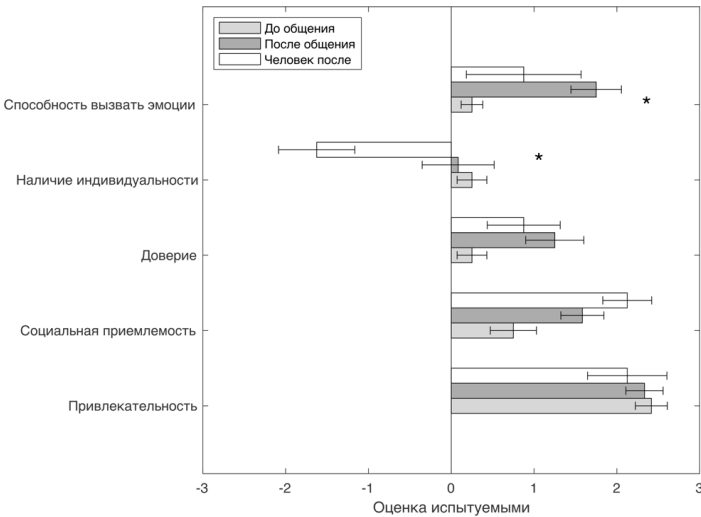


Рис. 3. Сравнение психологических профилей регистраторов, управляемых виртуальным агентом и человеком, до и после процедуры регистрации

Затемненные полосы на рис. 3 отображают характеристики виртуального агента. Значимые различия отмечены звездочками.

Расчеты проводились с использованием U-критерия Манна-Уитни при учете поправки Бонферрони. Найдено следующее. Способность виртуального актора вызывать эмоции возросла после общения с ним ( $P < 0.001$ ). Оценка наличия индивидуальности для виртуального актора оказалась выше, чем для человека-конфедерата ( $P < 0.0095$ ).



## 5.2. Виртуальный психодиагност

В случае Виртуального психодиагноста обученная система была применена для анализа реальных диалогов, проведенных Виртуальным психодиагностом в обличии робота Ф-2 с испытуемыми. По каждой из пяти шкал была получена бинарная оценка на основе анализа диалога. Результаты сравнивались с результатами стандартного теста Big5 в его адаптации на русском языке, также разбитыми на две категории по каждой шкале. Результаты сравнения представлены на рис. 4.

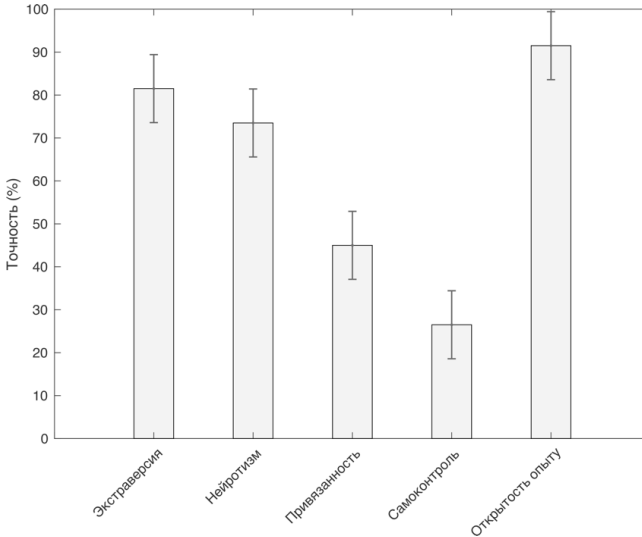


Рис. 4. Результаты распознавания типа личности на основе диалога с роботом. «Усы» показывают оценку стандартной ошибки

Максимальная точность классификации психотипа пользователя была достигнута для характеристики «открытость опыту» и оказалась выше 90%, а для половины остальных характеристик превысила 70% (рис. 4). Мы ожидаем, что подключение eVICA или ее нейросетевого эквивалента к управлению роботом позволит нам повысить точность.

## Заключение

В данной работе воплощен и исследован в экспериментах с испытуемыми способ интеграции методов машинного обучения (в частности, LLM) и когнитивного моделирования (на основе eVICA), позволяющий им дополнить друг друга. Обнадеживающие результаты получены в парадигмах «Вирту-

альный рецепционист отеля» и «Виртуальный психодиагност». Они указывают на то, что социально-эмоциональные разговорные агенты на основе когнитивных архитектур и машинного обучения при правильном подходе к их реализации могут быть социально совместимыми с человеком. В наших экспериментах артефакт воспринимался человеком как имеющий индивидуальность на уровне не ниже уровня человека. С другой стороны, при взаимодействии с артефактом человек проявлял свою индивидуальность на уровне, достаточном для воспроизведения результатов психологических тестов. По целому ряду шкал значимых различий между человеком и артефактом не было обнаружено, что говорит о человекоподобии артефакта. В целом, результаты указывают на реализуемость виртуального агента или робота, способного устанавливать социальный контакт на уровне человека и при этом также оценивать психологию пользователя.

**Благодарности.** Авторы считают своим приятным долгом поблагодарить тех, кто внес вклад в работу по данному проекту: студентов НИЯУ МИФИ Алену Анисимову, Алексея Михнева, и Владимира Царькова.

### Список литературы

- [Ali, 2020] Ali M.R. A Virtual Conversational Agent for Teens with Autism Spectrum Disorder: Experimental Results and Design Lessons // In Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents (IVA '20). Association for Computing Machinery, New York, USA, 2020.
- [Anisimova et al., 2022] Anisimova A.S., Potemkina A., Chervakov P., Komza V., Maksimov D., Panin I., Vaselyuk A., and Samsonovich A.V. Artificial Psychologist: An intelligent virtual/robotic assistant based on a cognitive modeling framework // Procedia Computer Science. – 2022. – Vol. 213. – P. 793-800.
- [Blache, 2020] Blache P. An integrated model for predicting backchannel feedbacks / Philippe Blache, Massina Abderrahmane, Stéphane Rauzy, and Roxane Bertrand. 2020 // In Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents (IVA '20). Association for Computing Machinery, New York, USA, 2020.
- [Díez Díaz et al., 2022] Díez Díaz F. et al. Empathic Smart Conversational Agent for Enhanced Recovery from Abdominal Surgery at Home // HCI International 2022–Late Breaking Posters: 24th International Conference on Human-Computer Interaction, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings, Part I. – Cham: Springer Nature Switzerland, 2022.
- [Elgarf et al., 2021] Elgarf M., Skantze G., Peters C. Once upon a story: Can a creative storyteller robot stimulate creativity in children? // Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents. – 2021.
- [Fitrianie et al., 2022] Siska Fitrianie, Merijn Bruijnes, Fengxiang Li, Amal Abdulrahman, and Willem-Paul Brinkman. The Artificial-Social-Agent Questionnaire: Establishing the long and short questionnaire versions // In ACM International Conference on Intelligent Virtual Agents (IVA '22), September, 2022, Faro, Portugal. ACM, New York, USA, 2022.

- [**Hartholt, 2020**] Hartholt A. Introducing Canvas: Combining Nonverbal Behavior Generation with User-Generated Content to Rapidly Create Educational Videos // In Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents (IVA '20). Association for Computing Machinery, New York, USA, 2020.
- [**Kosinski, 2016**] Kosinski M. MyPersonality Project [Электронный ресурс]. Last modified: 10.05.2016. – URL: [https://web.archive.org/web/20160518020419/http://mypersonality.org/wiki/doku.php?id=download\\_databases](https://web.archive.org/web/20160518020419/http://mypersonality.org/wiki/doku.php?id=download_databases) (дата обращения: 26.05.2023).
- [**Marsella et al., 2009**] Marsella S.C., Gratch J. EMA: A process model of appraisal dynamics // Cognitive Systems Research, 2009. Vol. 10, P. 70-90.
- [**Martin, 2019**] Martin J.A. Watson Personality Insights introduction and how to access Watson without SDK // Analytics Vidhya, 12.11.2019. – URL: <https://medium.com/analytics-vidhya/watson-personality-insights-introduction-and-how-to-access-watson-without-sdk-89eb8992fff2> (дата обращения: 26.05.2023).
- [**Mohammad et al., 2013**] Mohammad S., Turney P. Crowdsourcing a Word-Emotion Association Lexicon // Computational Intelligence. – 2013. – Vol. 29 (3). – P. 436-465.
- [**Mohammad, 2022**] Mohammad S.M. NRC Word-Emotion Association Lexicon [Электронный ресурс]. Last update: August 2022. – URL: <http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm> (дата обращения: 26.05.2023).
- [**Samsonovich, 2013**] Samsonovich A.V. Emotional biologically inspired cognitive architecture // Biologically Inspired Cognitive Architectures. – 2013. – No. 6. – P. 109-125.
- [**Samsonovich, 2020**] Samsonovich A.V. Socially emotional brain-inspired cognitive architecture framework for artificial intelligence // Cognitive Systems Research. – 2020. – Vol. 60. – P. 57-76.
- [**Tikhomirova et al., 2020**] Tikhomirova D.V., Chubarov A.A., Samsonovich A.V. Empirical and modeling study of emotional state dynamics in social videogame paradigms // Cognitive Systems Research. – 2020. – Vol. 60, – P. 44-56.
- [**Truity, 2023**] Truity. The Big Five Personality Test [Электронный ресурс]. – URL: <https://www.truity.com/test/big-five-personality-test> (дата обращения: 26.05.2023).
- [**Котов, 2023**] Котов А. [Электронный ресурс]. – URL: <http://www.f2robot.com> (дата обращения: 26.05.2023).

УДК 004.82

## МЕТОД РАЗРАБОТКИ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ КОМПЛЕКСНОГО НАЗНАЧЕНИЯ ПЕРСОНИФИЦИРОВАННОГО ЛЕЧЕНИЯ\*

Р.И. Ковалев (*koval-995@mail.ru*)

В.В. Грибова (*gribova@iacp.dvo.ru*)

Институт автоматики и процессов управления ДВО РАН,  
Владивосток

Дальневосточный федеральный университет, Владивосток

Д.Б. Окунь (*okdm@iacp.dvo.ru*)

Институт автоматики и процессов управления ДВО РАН,  
Владивосток

В работе описывается метод разработки систем поддержки принятия решений при назначении комплексного лечения. Он основывается на создании задаче-ориентированной оболочки, ключевым компонентом которой является обобщённая онтология, позволяющая формировать знания о различных видах лечения (медикаментозном, восстановительном, хирургическом) вне зависимости от раздела медицины. Это обеспечивает возможность создания единого решателя, что обеспечивает существенное снижение затрат на разработку системы. В работе описывается адаптация онтологии к различным видам лечения и терминологии разработчика базы знаний – эксперта предметной области. Онтология реализована на облачной платформе IACPaaS и в настоящее время уже активно используется для создания баз знаний в различных областях медицины.

**Ключевые слова:** система поддержки принятия решений, онтология, база знаний, онтолого-ориентированный решатель, персонализированное лечение.

---

\* Данная работа выполнена в рамках госзаданий FZNS-2023-0010 (разработка обобщённой онтологии), а также 0202-2022-0002 (разработка методов адаптации обобщённой онтологии).

## Введение

Назначение лечения является одним из этапов практической деятельности врача. Согласно различным публикациям, например, [Лудупова, 2016], [Кузьмина и др., 2016], на этом этапе совершается большое количество врачебных ошибок, поэтому системы поддержки принятия (СППР) решений должны помочь врачу в выборе оптимального, ориентированного на конкретного пациента плана лечения. Сложность этой задачи заключается в том, что врачу необходимо принимать во внимание специфику различных видов терапий, таких как медикаментозное, восстановительное лечение или же хирургическое вмешательство; учитывать большое количество различной информации: жалобы больного, результаты его объективного, лабораторного и инструментального исследований, сопутствующие заболевания, аллергии, совместимость лекарственных препаратов, противопоказания и многое другое [Федосеев, 2018].

К настоящему времени разработано большое количество различных СППР [Ефименко и др., 2017], однако реализуются они только для одного заболевания, реже для их группы, как правило, небольшой. При этом каждый вид лечения также реализован отдельной системой, а структура знаний, лежащая в их основе, является существенно упрощённой по сравнению с реальными знаниями и данными в этой области [Haendel, 2017]. Использование врачом комплекса различных систем на практике не представляется возможным. Их создание и сопровождение для разработчиков является очень дорогим и трудоемким процессом.

Следует отметить и различные подходы к реализации данного класса систем, среди которых основными являются системы на основе знаний и методов машинного обучения. Учитывая, что методы лечения постоянно уточняются и совершенствуются, появляются новые лекарственные препараты, врачам рекомендовано придерживаться клинических рекомендаций Минздрава, использование систем на основе знаний является наиболее адекватным методом реализации данного класса систем. В работе [Ковалев и др., 2020] авторами была предложена специализированная оболочка для создания систем поддержки принятия решений по назначению медикаментозного лечения на основе онтологических баз знаний. Однако она ограничивалась только одним видом лечения, в то время как для множества заболеваний необходимо сочетание нескольких видов одновременно или последовательно.

Целью работы является описание метода построения интеллектуальной СППР по назначению лечения вне зависимости от раздела медицины и его вида, а также способа представления знаний, в ней использующихся.

## 1. Постановка задачи

Создание программных оболочек для класса задач обеспечивает снижение трудоемкости разработки, сопровождения и масштабирования прикладных систем, созданных на ее основе. В данном случае рассматривается класс задач планирования лечения, которое объединяет его различные виды (медикаментозное, восстановительное и хирургическое) и не зависит от заболевания или раздела медицины в целом [Prasad, 2022].

Использование двухуровневого подхода к разработке знаний позволяет явно отделить базы знаний от их онтологий, что дает возможность формирования на основе одной онтологии множества баз знаний (БЗ). Ранее авторами были разработаны и успешно используются онтологии медикаментозного и восстановительного лечения [Грибова, 2018]. Каждая онтология обеспечивает возможность формализации БЗ по соответствующему виду лечения силами экспертов предметной области в привычной, понятной им терминологии и структуре.

Полученный опыт использования онтологий показал:

- для всех возможных видов лечения можно выделить общие понятия и структуру;
- врачебные назначения имеют сходный алгоритм применения.

На основе вышесказанного, требуется разработать обобщенную онтологию по комплексному воздействию на пациента, а также механизмы ее адаптации к конкретным видам лечения. Онтология должна иметь возможность задавать характеристики объекта или процесса, их свойства и ограничения, такие как периодичность, дозировка, способ применения, длительность, условия применения. По данной онтологии необходимо создать единый решатель, управляемый этой онтологией, благодаря чему модификация базы знаний не повлечет его изменения, а создание любой новой системы сводится к формированию ее базы знаний. Таким образом, обеспечивается создание задаче-ориентированных программных оболочек, которые не зависят от заболевания или раздела медицины.

## 2. Схема предлагаемого подхода

Технология разработки интеллектуальных систем с онтологическими базами знаний на платформе IASPaas [Клещев и др., 2018] поддерживает создание специализированных оболочек за счет двухуровневого подхода к разработке баз знаний. Решатель в таком подходе разрабатывается на основе онтологий. Проводя аналогию с программированием, можно считать, что онтология для решателя является формальным параметром, а конкретная база знаний – фактическим параметром. Таким образом, обеспечивается независимость решателя от конкретной базы знаний.

Суть предлагаемого метода заключается в следующем. Создается обобщённая онтология представления знаний о лечении, онтология таблицы соответствий. На их основе формируется решатель (онтологии являются формальными параметрами решателя). С использованием соответствующей онтологии формируется таблица соответствий, которая уточняет (адаптирует) обобщенную онтологию к любой специализированной онтологии – онтологии восстановительного, медикаментозного или хирургического лечения. Суть адаптации заключается в уточнении терминологии, расширении (добавлении) ее структурных элементов, определенных в обобщенной онтологии. Единый решатель, «зная» онтологии, а также соответствие между терминами, обрабатывает любую специализированную онтологию, поскольку все блоки и разделы обобщенной и специализированных онтологий имеют похожую структуру и семантику, и как следствие, идентичный алгоритм обработки. Любая модификация базы знаний не влечет изменения решателя. Эксперты предметной области формируют базы знаний в терминах специализированных онтологий, которые могут быть гибко (через таблицу соответствий) адаптированы к их требованиям. Принципиальная схема такого подхода приведена на рис. 1.

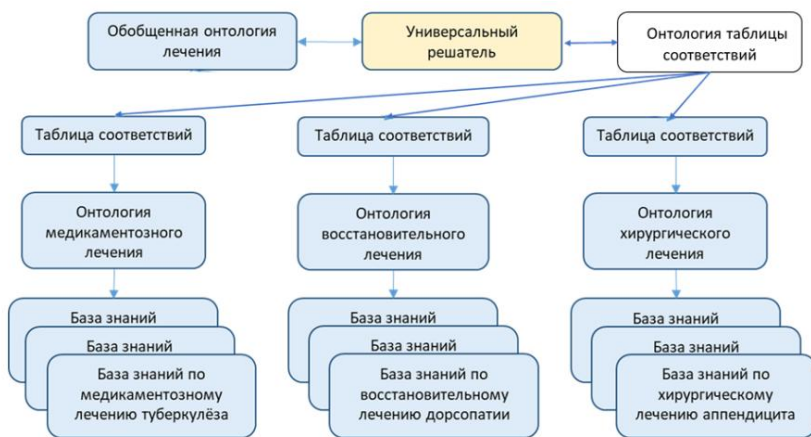


Рис. 1. Схема предлагаемого подхода

## 2.1. Описание обобщенной онтологии

Обобщенная онтология включает описание различных видов лечения в рамках конкретной патологии с учетом персональных данных пациента. Структурное представление онтологии лечения представлено на рис. 2.

Для каждого заболевания формируется своя *Модель лечения* (или несколько моделей), которая может включать либо восстановительное, либо медикаментозное, либо хирургическое. Возможна комбинация различных видов лечения.

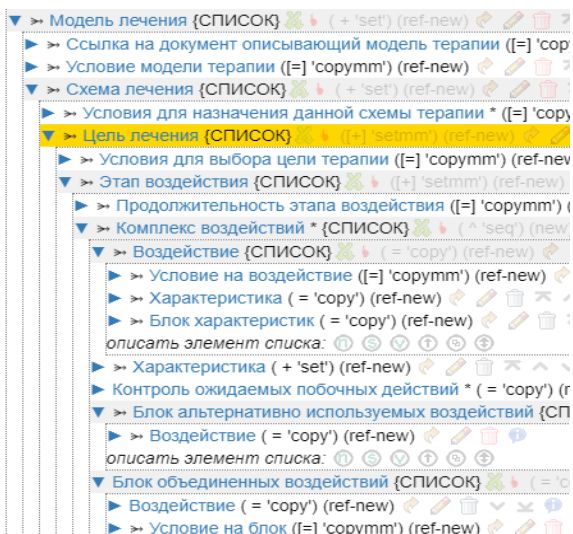


Рис. 2. Модель общей онтологии лечения

Раздел *Модель лечения* содержит описание совокупности медицинских мероприятий с доказанным клиническим эффектом. Вершина группирует конкретный подход или план лечения заболевания.

Основным структурным элементом онтологии является *Схема лечения*, которая объединяет планы лечения, сгруппированные согласно варианту протекания заболевания, который определяется набором целей терапии. Каждая *Цель лечения* содержит комплекс клинических мероприятий, направленный на конкретный аспект проводимого лечения (рис. 3).

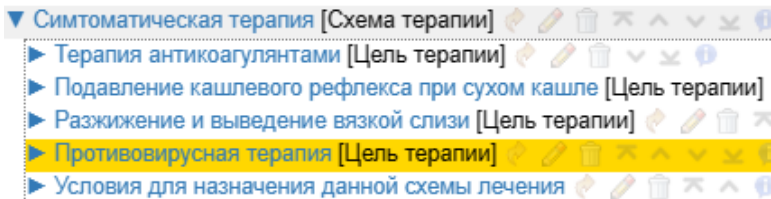


Рис. 3. Пример описания в базе знаний раздела *Схема терапии*



Лечение может быть разделено на этапы, каждый *Этап лечения* представляет собой набор клинических мероприятий, организованный либо по временным отрезкам, либо по достижению целевых показателей, соответствующий цели лечения. Каждый этап лечения содержит *Комплекс воздействий*, описывающий знания о любых манипуляциях с пациентом (рис. 4), и *Продолжительность этапа воздействия*.

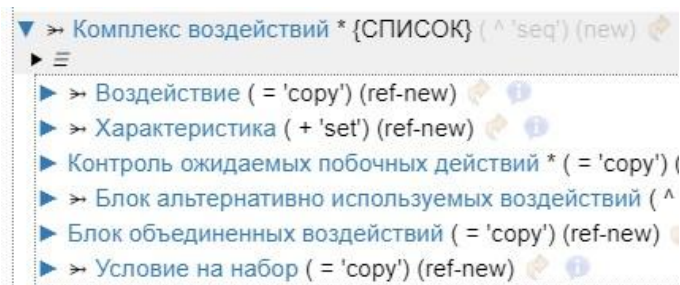


Рис. 4. Раздел Комплекс воздействий

Для решения вопросов оптимизации и персонификации лечения предусмотрен целый многоуровневый комплекс условий, контролирующих возможность проведения медицинских мероприятий (рис. 5). Каждое условие является формально представленным клиническим комплексом критериев с четко определенными значениями. Под критерием можно понимать любой элемент записи в истории болезни, относящийся к персональным данным пациента или отражающий особенность клинической картины заболевания.

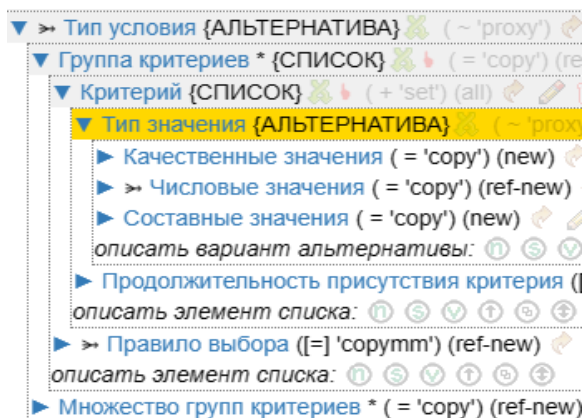


Рис. 5. Структура условия

*Воздействие* – основополагающий раздел онтологии, характеризующий тип манипуляций с пациентом (рис. 6). Вершина позволяет описать любой фактор, имеющий доказанное лечебное действие, например, действующее вещество, физиопроцедура, массаж, хирургические манипуляции (перевязка, наложение шва, обработка ран и т.д.).

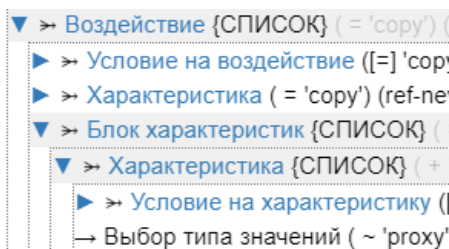


Рис. 6. Структура раздела *Воздействие*

Для персонификации применения воздействия в онтологии содержатся блок *Условие на воздействие*, описывающий перечень клинических показателей, при которых данное воздействие может быть рекомендовано к использованию, а также *Блок альтернативно используемых воздействий*, позволяющий описать некоторое множество альтернативных воздействий, когда необходимо выбрать одно из целого ряда, и *Блок объединенных воздействий*, предназначенный для описания некоторого множества воздействий, использующихся в связке, например в растворе.

Раздел *Характеристика* позволяет описывать параметры, свойства и ограничения какого-либо объекта или процесса (рис. 7), например, того же воздействия. Он так же содержит собственный блок условий, необходимый для выбора конкретной характеристики, соответствующей картине заболевания. Данный раздел имеет универсальную структуру и позволяет задавать любые типы значений и параметров.

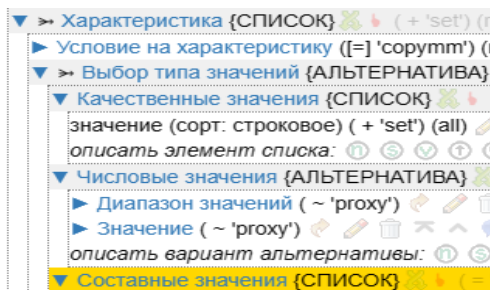


Рис. 7. Структура раздела *Характеристика*

## 2.2. Механизмы адаптации обобщенной онтологии к специализированным онтологиям

Представленная онтологическая модель имеет универсальную структуру представления знаний по назначению персонифицированного лечения. Для ее адаптации к специализированной онтологии по конкретному виду лечения используется таблица соответствий. Она позволяет гибко настроить как структуру, так и терминологию обобщенной онтологии под специфику решаемой задачи. Таблица соответствий сопоставляет абстрактные понятия обобщенной онтологии и конкретные понятия в целевой. Помимо задания узкой, ориентированной на тип лечения терминологии, в целевых онтологиях возможно различное количество вершин на одном структурном уровне, т.е. одному абстрактному понятию могут соответствовать несколько конкретных. Таким образом, формируются специализированные онтологии, адаптированные к конкретным видам лечения и требованиям пользователей. На основе данных онтологий формируются уже соответствующие базы знаний по лечению заболеваний из различных разделов медицины.

Основная спецификация онтологии медикаментозного лечения начинается с уровня *Комплекс воздействий*, где термин *Воздействие* заменен на *Действующее вещество*. Уточнение разделов, формирующих различные параметры, свойства и ограничения происходит следующим образом. Вершина *Блок характеристик* заменяется на *Вариант назначения*, а термину *Характеристика* сопоставляется сразу несколько разделов: *Разовая дозировка*, *Суточная дозировка*, *Вариант применения*, *Кратность приема*, *Продолжительность приема*.

Таким образом, изменяя количество вершин и терминологию, обобщенная онтология адаптируется к конкретной онтологии, формируя специализированную онтологию по назначению медикаментозного лечения (рис. 8).

При адаптации к онтологии по восстановительной терапии термин *Воздействие* заменяется на *Методику восстановительного лечения*, а *Блок характеристик* сопоставляется с *Режим проведения* и *Описание ожидаемой реакции при проведении процедуры*. Термин *Характеристика* заменяется на *Характеристика методики*, *Область применения*, *Количество процедур*, *Дни применения*.

В результате, формируется также специализированная онтология, соответствующая восстановительному лечению (рис. 9).

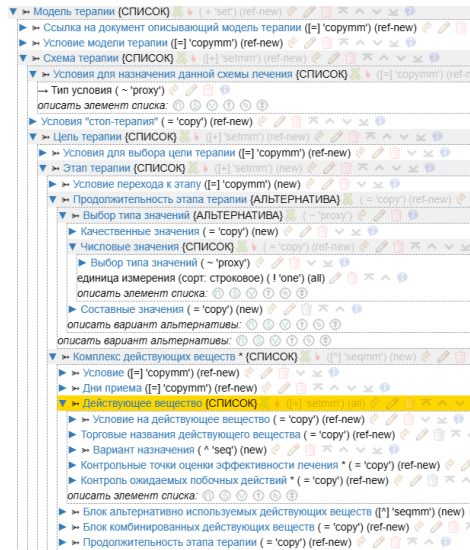


Рис. 8. Онтологии назначения медикаментозного лечения

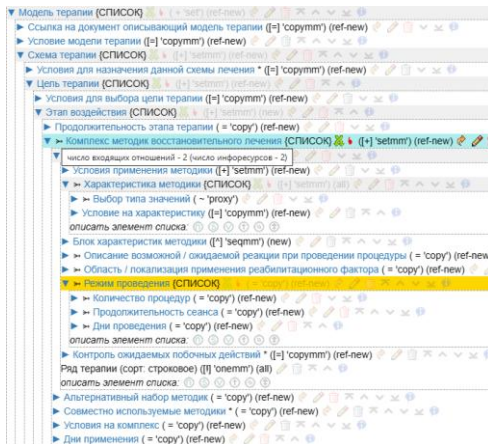


Рис. 9. Онтология назначения восстановительного лечения

Таким методом также возможна адаптация обобщённой онтологии к лечебным заведениям или специалистам, которые хотят использовать собственную терминологию. В целом возможна адаптация к любой предметной области со схожей структурой знаний, при этом сохраняется возможность использования единого решателя.

## Заключение

В работе представлен метод разработки систем поддержки при назначении комплексного (комбинированного) лечения пациента, включающее медикаментозное, восстановительное и хирургическое лечение. Его суть заключается в разработке обобщенной онтологии представления знаний, которая позволяет формировать знания о различных видах лечения по различным заболеваниям. Для адаптации онтологии к конкретным условиям необходимо лишь сформировать таблицу соответствий. Это открыло возможность создания единого решателя, работающего по различным специализированным онтологиям, и соответственно базам знаний, а также объединению различных видов лечения в одной системе.

Такой подход позволяет существенно снизить затраты на разработку и сопровождение интеллектуальных систем по назначению персонализированного лечения.

## Список литературы

- [Лудупова, 2016] Лудупова Е.Ю. Врачебные ошибки. Литературный обзор // Вестник Росздравнадзора. – 2016. – № 2. – С. 6-15.
- [Кузьмина и др., 2016] Кузьмина А.В. и др. Медицинские ошибки при применении лекарственных препаратов в практическом здравоохранении // Практическая пульмонология. – 2016. – № 3. – С. 76-83.
- [Федосеев, 2018] Федосеев Г.Б. Врачебные ошибки: характер, причины, последствия, пути предупреждения // Терапия. – 2018. – № 5 (23). – С. 109-115
- [Ефименко и др., 2017] Ефименко И.В. Интеллектуальные системы поддержки принятия решений в медицине: ретроспективный обзор состояния исследований и разработок и перспективы // Открытые семантические технологии проектирования интеллектуальных систем (OSTIS-2017). – Минск: БГУИР, 2017. – С. 251-260.
- [Haendel et al., 2018] Haendel M.A., Chute C.G., Robinson P.N. Classification, ontology, and precision medicine // New England Journal of Medicine. – 2018. – Vol. 379, No. 15. – P. 1452-1462.
- [Ковалев и др., 2020] Ковалев Р.И., Грибова В.В., Окунь Д.Б. Специализированная оболочка для построения интеллектуальных систем назначения медикаментозного лечения // Искусственный интеллект и принятие решений. – 2020. – № 4. – С. 66-79. – DOI 10.14357/20718594200407.
- [Prasad, 2022] Prasad G.N.R. A review on open source and free expert system shells // Multidisciplinary Research Journal. – 2022. – Vol. 1, Issue. V. July PC: PSR-220701.
- [Грибова и др., 2018] Грибова В.В., Окунь Д.Б. Онтологии для формирования баз знаний и реализации лечебных мероприятий в медицинских интеллектуальных системах // Информатика и системы управления. – 2018. – № 3 (57). – С. 71-80.
- [Клещев и др., 2018] Грибова В.В., Клещев А.С., Москаленко Ф.М., Тимченко В.А., Федорищев Л.А., Шалфеева Е.А. Облачная платформа IASaaS для разработки оболочек интеллектуальных сервисов: состояние и перспективы развития // Программные продукты и системы. – 2018. – Т. 31, № 3. – С. 527-536.

## ПРИНЦИПЫ ПОСТРОЕНИЯ СИСТЕМЫ КОНТРОЛЯ ПСИХОЭМОЦИОНАЛЬНОГО СОСТОЯНИЯ И МОНИТОРИНГА ЗДОРОВЬЯ ОПЕРАТОРОВ КРИТИЧЕСКОЙ ИНФРАСТРУКТУРЫ\*

Б.А. Кобринский (*kba\_05@mail.ru*)  
Н.А. Благосклонов (*nblagosklonov@frccsc.ru*)  
О.Г. Григорьев (*olegpolikvart@yandex.ru*)  
Федеральный исследовательский центр  
«Информатика и управление» РАН, Москва

В статье рассматриваются принципы построения интеллектуальной рекомендательной системы мониторинга операторов критической инфраструктуры, ориентированной на оценку их психоэмоционального состояния в режиме реального времени и периодическую оценку факторов риска хронических болезней. Описаны информационные потоки разнотипной информации и модули для преобразования и анализа разнородных данных. Представлена архитектура системы, в которой предусмотрена выдача информации об отклонениях в состоянии операторов для лиц принимающих решения на объекте и для медицинских работников.

**Ключевые слова:** интеллектуальная рекомендательная система, мониторинг состояния операторов, диагностика психоэмоционального состояния, оценка риска заболеваний, разнотипная информация, человеческий фактор, критическая инфраструктура.

### Введение

В настоящее время большое значение придается мониторингу объектов критической инфраструктуры [Puuska et al., 2018]. Однако, несмотря на прогресс систем безопасности, именно человеческий фактор, продолжает оставаться релевантным элементом, требующим особого внимания на объектах критической инфраструктуры. Для снижения рисков ошибки

---

\* Исследование выполнено в рамках научной программы Национального центра физики и математики, направление № 9 «Искусственный интеллект и большие данные в технических, промышленных, природных и социальных системах».

операторов на таких объектах актуальным является оперативный контроль их психоэмоционального состояния во время работы и мониторинг состояния их здоровья.

Необходимость быстрого реагирования на негативные изменения в состоянии операторов выдвигает требования постоянной актуализации данных о текущем состоянии как психоэмоционального состояния, так и здоровья. Для решения этой задачи требуется интеграция многообразной разнотипной информации, которая должна поступать из различных источников: медицинской документации, специализированных опросников, с носимых медицинских устройств съема физиологических данных, с видеобиометрии и др.

Такой многоаспектный контроль может быть реализован с использованием интеллектуальной системы, которая, в режиме реального времени, получает, обрабатывает и анализирует различную персональную информацию медицинского, психофизиологического и психоэмоционального характера. Выдаваемые на основе этого решения должны носить рекомендательный характер и обеспечивать точность, достоверность и конфиденциальность [Sahoo et al., 2019], [Кобринский, 2022].

## **1. Принципы сбора и анализа информации для контроля состояния операторов**

Источниками данных для интеллектуальной системы в настоящее время являются:

- Электронные медицинские карты (ЭМК), содержащие результаты прохождения периодических профилактических осмотров и данные обращений за медицинской помощью, а также анкеты-опросники для выявления донозологических изменений, угрожающих развитием хронических болезней [Бобров и др., 2023], [Кобринский и др., 2019].
- Результаты прохождения предсменных осмотров, содержащие текущие данные о состоянии оператора.
- Персональные носимые и стационарные устройства съема физиологических параметров, которые могут включать пульс, артериальное давление, оксигенацию крови, уровень глюкозы в крови, температуру тела, уровень стресса и др.
- Профили операторов в социальных сетях и мессенджерах, результат анализа которых может быть интерпретирован в виде личностных черт, склонности к депрессивности, агрессии и др. [Смирнов, 2023].
- Потоки аудио и видеoinформации с рабочего места оператора, позволяющие выделять характеристики лицевой мимики, движений глаз, речи, а также отдельные составляющие двигательной активности и положения тела в пространстве, что при комплексном анализе харак-

теризует определенные паттерны поведения [Александров и др., 2019], [Анищенко и др., 2018], [Бобе и др., 2016], [Кухарев и др. 2013], [Calder et al., 2001], [Korovin et al., 2020], которые могут указывать на отклонения, несущие потенциальный риск для контролируемого объекта (АЭС, центр мониторинга чрезвычайных ситуаций и др.).

- Данные бесконтактного анализа психофизиологических параметров организма оператора [Анищенко и др., 2018].

Для сбора и анализа разнотипной информации из разнородных источников должны быть разработаны соответствующие модули. Результаты первичной обработки и свертки данных должны поступать в интеграционный модуль, где информация, полученная из одного источника, будет корректироваться с учетом данных от других источников. Однако следует учитывать, что на объектах критической инфраструктуры нередко существуют ограничения на использование определенных электронных устройств и передачу информации даже внутри закрытого контура, что в реальности может ограничивать использование определенных устройств. Кроме того, данные ЭМК и опросные листы могут содержать неполную информацию. Также не все операторы представлены в социальных сетях.

Поэтому для объективной оценки психоэмоционального состояния оператора, необходимо использовать и совместно анализировать информацию из всех доступных на текущий момент источников. В случае выявления неадекватного поведения оператора будет поступать экстренное уведомление соответствующему лицу, принимающему решение, с рекомендацией обратить внимание и принять необходимые меры в отношении конкретного человека (временное отстранение от работы, непродолжительный отдых для восстановления работоспособности, оказание медицинской помощи). Кроме того, на основании имеющихся данных система будет формировать персональные рекомендации по профилактике хронических заболеваний при выявлении риска их возникновения, аналогично интеллектуальной системе здоровьесбережения [Grigoriev et al., 2018].

## **2. Архитектура системы мониторинга состояния операторов**

Гибридная система мониторинга психоэмоционального статуса и здоровья операторов, включающая анализ разнородной и разнотипной информации, включая аудио-видео данные, предполагает наличие следующих программных модулей и компонентов:

- API, обеспечивающее возможность получения информации об операторе из различных источников, преобразования первичной информации и обмена данными между модулями.



- Модуль сопоставления и интеграции обезличенной информации индивидуумов, поступающей со своими идентификаторами из различных источников.
- Модуль обработки физиологических параметров, осуществляющий непрерывное получение информации с датчиков, ее анализ и расчет необходимых интегрированных показателей за определенный период времени.
- Модуль обработки и анализа психоэмоциональных реакций, осуществляющий непрерывный анализ аудио- и видеопотоков информации с рабочего места оператора и позволяющий идентифицировать отклонения в поведении, которые представляют потенциальную опасность.
- Модуль обработки информации из социальных сетей и опросников для определения личностных черт и других аспектов психологического профиля оператора.
- Хранилище (база) данных, обеспечивающее загрузку и накопление разнотипной информации, получаемой из различных источников.
- База знаний, обеспечивающая классификацию состояний оператора и логический вывод на основе анализа многочисленных актуальных данных.
- Web-интерфейс, позволяющий просматривать и вводить информацию лицам с разными уровнями доступа, используя различные сценарии для пользователей-операторов, медицинских работников и ЛПР о продолжении работы оператора в текущей ситуации.

Рассмотрим организацию получения и анализа данных, поступающих из различных источников (рис. 1):

1. Получение разнородной информации осуществляется посредством API, с возможностью донастройки при появлении новых источников данных.
2. Информация собирается из разных мест в обезличенном виде с использованием различных персональных идентификаторов оператора. Это обусловлено тем, что один и тот же человек имеет различные уникальные идентификаторы в разных информационных системах. В специальной защищенной базе данных будет осуществляться их сопоставление.
3. Часть информации в хранилище системы (Цифровой Медицинский Профиль – ЦМП) передается непосредственно после сбора соответствующим модулем, а часть (физиологические параметры, биометрические данные и профили из социальных сетей) предварительно обрабатывается в проблемно-ориентированных аналитических модулях.

4. Многоаспектный анализ и диагностика состояния оператора осуществляется на основе правил базы знаний.
5. Результирующая текущая оценка состояния оператора с рекомендациями (при наличии отклонений в его поведении или состоянии) передается и сохраняется в Цифровом Профиле Здоровья (ЦПЗ) индивидуума, а медицинская часть информации (факторы риска и рекомендации) при необходимости корректируется врачом и также передается в ЦПЗ. Одновременно эта информация передается на постоянное хранение в ЦМП.

Рассмотрим специализированные модули анализа данных.

*Модуль сбора и обработки физиологических параметров.* Данный модуль в режиме реального времени получает данные, поступающие с персональных носимых или стационарных медицинских устройств и непосредственно осуществляет автоматический расчет минимального, максимального и среднего значений каждого из физиологических показателей. Рассчитанные величины поступают в хранилище ЦМП индивидуума.

*Модуль сбора и обработки данных ЭМК/опросников.* Данный модуль получает информацию из ЭМК, анкет-опросников и результатов предсменных осмотров.

*Модуль анализа адаптационных возможностей организма.* Данный модуль на основе информации из ЭМК, данных предсменных осмотров и обработанных показателей персонального мониторинга физиологических параметров рассчитывает текущий уровень адаптации организма к внешним стрессовым воздействиям.

*Модуль анализа психоэмоциональных реакций.* Этот аппаратно-программный комплекс анализирует видео и аудио с рабочего места оператора для выявления определенных мимических и речевых реакций, и на основе их комплексной оценки выдает характеристику психоэмоционального состояния оператора.

*Модуль анализа психологического профиля по социальным сетям.* Данный модуль анализирует профиль человека, полученный из социальных сетей, на основе частных и публичных сообщений, пересылаемых публичных сообщений других пользователей или групп, публикуемые фотографии и прочее. На основе этих данных определяется психологический профиль индивидуума. Это уменьшает необходимость заполнять большое количество опросных листов. Анализ профиля пользователя в социальной сети осуществляется после получения от него разрешения на использование его учетной записи и данных в обезличенном виде.



*Рис. 1. Структурная схема информационных потоков системы мониторинга здоровья и психоэмоционального состояния операторов*

Архитектура системы представлена на рис. 2.

База знаний обеспечивает формирование гипотез (заключений) в отношении:

- неадекватности в поведении;
- текущего уровня адаптации организма;
- психологического профиля человека;
- донозологических нарушений хронических неинфекционных заболеваний (ХНИЗ), факторов риска, уровней риска развития отдельных ХНИЗ;
- рекомендаций в отношении операторов при отклонениях в их поведении или опасных изменениях в их состоянии;
- рекомендаций по снижению негативного воздействия модифицируемых факторов риска.

Как представлено на рис. 2, информация из разных источников постоянно поступает в хранилище системы:

- данные из ЭМК о профилактических осмотрах должны поступать по мере их прохождения путем запроса из баз данных медицинских организаций;
- сведения о вновь возникших хронических заболеваниях будут поступать от врачей в процессе корректировки выдаваемых системой рекомендаций;
- данные предменструальных осмотров ежедневно автоматически;
- психофизиологические и психоэмоциональные параметры с фиксированной периодичностью после преобработки в специальных модулях, представленных выше.

Актуальные на текущий момент времени данные передаются в рабочую память и затем подвергаются анализу с использованием правил базы знаний для определения в оперативном режиме отклонений от нормы в состоянии и/или поведении оператора и в дискретном режиме для оценки рисков ХНИЗ.

С использованием механизма аргументационного рассуждения формируются и выдаются заключения, характеризующие уровни риска в текущей ситуации:

- о потенциально опасных отклонениях в состоянии/поведении оператора – с экстренным уведомлением для ЛПР,
- о психоэмоциональном состоянии, уровне адаптационных возможностей оператора, выявленных факторах и уровнях риска ХНИЗ – для лечащего врача/менеджера здоровья.

Наряду с заключениями выдаются рекомендации по психологической коррекции, по балансировке адаптационных механизмов организма, по снижению риска возникновения ХНИЗ.

Информация для врача поступает в его личный кабинет, где он ее просматривает и анализирует. Врач или принимает решение о подтверждении рекомендаций интеллектуальной системы или, при необходимости, может дополнить/скорректировать рекомендации, внести дополнения о состоянии здоровья пациента и запустить процедуру пересчета рисков. После окончательного одобрения рекомендации, подтвержденной электронной подписью, она поступает в ЦПЗ и передается на персональную страницу пользователя (оператора), а также передается на постоянное хранение в ЦМП.

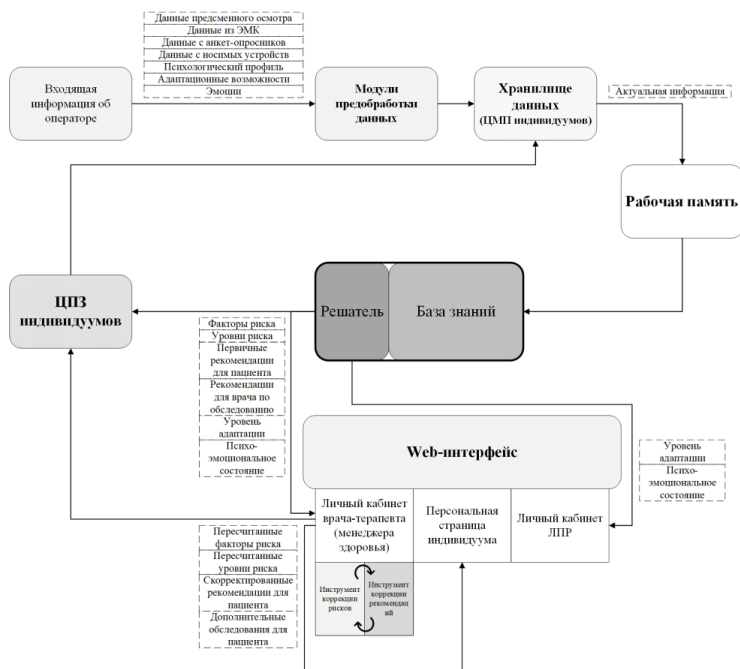


Рис. 2. Архитектура системы мониторинга психоэмоционального состояния и состояния здоровья операторов критической инфраструктуры

## Заключение

Интеллектуальная система для мониторинга состояния операторов критической инфраструктуры должна способствовать повышению безопасности объектов путем снижения негативных аспектов влияния человеческого фактора, обусловленных колебаниями психоэмоционального и психофизиологического состояния людей, контролирующих функциониру-

вание критически важных систем. Особенность предлагаемого подхода состоит в интегральном оперативном анализе разнотипной информации о ситуативном психоэмоциональном статусе операторов (с учетом состояния их здоровья) и выдаче рекомендаций для ЛППР при выявлении их неадекватного поведения или угрожающего психофизиологического состояния.

### Список литературы

- [Александров и др., 2019] Александров А.А., Кирпичников А.П., Ляшева С.А., Шлеймович М.П. Анализ эмоционального состояния человека на изображении // Вестник Технологического университета. – 2019. – Т. 22, № 8.
- [Анищенко и др., 2018] Анищенко Л.Н., Турецкая А.В. Программно-аппаратный комплекс для бесконтактной оценки психофизиологического состояния человека // Биомедицинская радиоэлектроника. – 2018. – № 10. – С. 61-67.
- [Бобе и др., 2016] Бобе А.С., Конышев Д.В., Воротников С.А. Система распознавания базовых эмоций на основе анализа двигательных единиц лица // Инженерный журнал: наука и инновации. – 2016. – № 9.
- [Бобров и др., 2023] Бобров А.Ф., Новикова Т.М., Седин В.И., Фортунатова Л.И. Системные критерии дифференциальной экспресс-диагностики донозологических нарушений профессионального здоровья работников объектов использования атомной энергии // Медицинская радиология и радиационная безопасность. – 2023. – Т. 68, № 2.
- [Кобринский и др., 2019] Кобринский Б.А., Кадыков А.С., Полтавская М.Г., Благосклонов Н.А., Ковелькова М.Н. Принципы функционирования интеллектуальной системы динамического контроля факторов риска и формирования рекомендаций по здоровьесбережению // Профилактическая медицина. – 2019. – Т. 22, № 5.
- [Кобринский, 2022] Кобринский Б.А. Интеллектуальные рекомендательные системы для медицины: особенности и ограничения // Искусственный интеллект и принятие решений. – 2022. – № 3.
- [Кухарев и др., 2013] Кухарев Г.А., Каменская Е.И., Матвеев Ю.Н., Щеголева Н.Л. Методы обработки и распознавания изображений лиц в задачах биометрии. – СПб: Политехника, 2013.
- [Смирнов, 2023] Смирнов И.В. Программные средства психоэмоционального анализа текстов // Информационные технологии и вычислительные системы. – 2023. – № 1.
- [Calder et al., 2001] Calder A.J., Burton A.M., Miller P., Young A.W., Akamatsu S. A principal component analysis of facial expressions // Vision Research. – 2001. – Vol. 41.
- [Grigoriev et al., 2018] Grigoriev O.G., Kobrinskii B.A., Osipov G.S., Molodchenkov A.I., Smirnov I.V. Health Management System Knowledge Base for Formation and Support of a Preventive Measures Plan // Procedia Computer Science. – 2018. – Vol. 145.
- [Korovin et al., 2020] Korovin I., Ivanov D. Human Pose Estimation Applying ANN While RGB-D Cameras Video Handling // Advances in Intelligent Systems and Computing. – 2020. – Vol. 1225.
- [Puuska et al., 2018] Puuska S., Rummukainen L., Timonen J., Lääperi L., Klemetti M., Oksama L., Vankka J. Nationwide critical infrastructure monitoring using a common operating picture framework // International Journal of Critical Infrastructure Protection. – 2018. – Vol. 20.
- [Sahoo et al., 2019] Sahoo A.K., Pradhan C., Barik R.K., Dubey H. Deepreco: Deep learning based health recommender system using collaborative filtering // Computation. – 2019. – Vol. 7.

## ТЕХНОЛОГИИ ДИСТАНЦИОННОГО МОНИТОРИНГА ПСИХОЭМОЦИОНАЛЬНОГО СОСТОЯНИЯ ОПЕРАТОРОВ КРИТИЧЕСКОЙ ИНФРАСТРУКТУРЫ\*

М.В. Хисамутдинов (*L-V-P@yandex.ru*)

Я.С. Коровин (*korovin\_yakov@mail.ru*)

Научно-исследовательский институт многопроцессорных  
вычислительных систем им. А.В. Каляева ЮФУ, Таганрог

Д.Я. Иванов (*donat.ivanov@gmail.com*)

Научно-исследовательский институт многопроцессорных  
вычислительных и управляющих систем, Таганрог

В работе описывается предлагаемый способ решения задачи дистанционного неинвазивного мониторинга признаков сохранения или утраты оператором профессионального здоровья и профессиональной надежности. Предложена концепция и архитектура автоматизированной системы дистанционного анализа признаков утраты работоспособности оператором критической инфраструктуры. Приведены краткие результаты экспериментальных исследований и направления дальнейших исследований.

**Ключевые слова:** мониторинг функционального здоровья сотрудника, дистанционный мониторинг, анализ видеопоследовательности.

### Введение

Для предприятий критической инфраструктуры важной задачей является контроль сохранения работниками профессиональной надежности и профессионального здоровья. Под профессиональным здоровьем понимают наличие и развитие регуляторных свойств организма работника, обеспечивающих высокую надежность профессиональной деятельности, определяемой, в свою очередь, вероятностью эффективной и безаварийной работы в течение определенного периода времени при сохранении жизни и здоровья самого работника и его окружения. При оценке профес-

---

\* Исследование выполнено в рамках научной программы Национального центра физики и математики, направление № 9 « Искусственный интеллект и большие данные в технических, промышленных, природных и социальных системах».

сионального здоровья учитывают физическое и психологическое состояние сотрудника, его функциональный резерв, адекватность уровня функционирования организма рабочим нагрузкам.

Так, в области атомной энергетики обязанность по организации медицинских осмотров и психофизиологических обследований персонала в целях снижения вероятности аварий, связанных с отклонением в состоянии здоровья работников, возлагается на работодателя. Предпринимается целый комплекс мероприятий, таких как предварительные медицинские осмотры при поступлении на работу; периодические медицинские осмотры в ходе трудовой деятельности с интервалами в 6-12 месяцев; психофизиологические обследования в лабораториях психофизиологического обеспечения при поступлении на работу и в ходе трудовой деятельности; проведение предсменных психофизиологических обследований.

При этом при проведении предсменных осмотров действуют жесткие лимиты времени в связи с необходимостью обеспечения высокой пропускной способности. В связи с этим в существующей системе психофизиологических обследований используются только простейшие контактные методы измерения, которые позволяют выявлять только самые грубые нарушения.

Применяемая в настоящее время методология проведения психофизиологических обследований исключает современные бесконтактные методы, а также не включает внутрисменные и/или послесменные осмотры. Также к недостаткам существующих подходов, применяемых в рамках современных психофизиологических обследований работников, следует отнести узкий спектр оцениваемых психофизиологических параметров индивидов и отсутствие сбора статистических данных для построения доверительных интервалов границ допустимых показателей отдельных психофизиологических параметров для групп специалистов и индивидуально для каждого работника (понятия «групповой нормы» и «индивидуальной нормы»); отсутствие комплексных методов оценки функционального состояния.

Таким образом, наблюдается относительно невысокая предиктивная способность существующих методов проведения психофизиологических обследований работников в отношении краткосрочного и долгосрочного риска критического снижения профессиональной надежности специалиста. В связи с этим актуальной является разработка концепции, методов, подходов и систем, которые бы позволяли осуществлять комплексный анализ совокупности психофизиологических параметров работника не только перед рабочей сменой, но и на протяжении всего рабочего процесса. Для этого необходима разработка автоматизированных систем дистанционного мониторинга психоэмоционального состояния операторов критической инфраструктуры.



## 1. Концепция АСДА

В рамках проводимого исследования предложена концепция создания автоматизированной системы дистанционного анализа (АСДА) психоэмоционального состояния (ПЭС) человека на основе дистанционно и не инвазивно собираемых данных.

Потенциально опасные состояния определяются на основе автоматического интеллектуального анализа видеоизображений, акустических и электромагнитных сигналов.

Принятие окончательного решения по поводу степени опасности состояния объекта наблюдения осуществляет оператор АСДА.

Функциями автоматизированной системы являются следующие:

- получение видеоизображений с камер наблюдения;
- получение акустических сигналов с микрофонов;
- получение электромагнитных сигналов от бесконтактного СВЧ/КВЧ-радара;
- обработка и хранение информации;
- оценка набора признаков, характеризующих состояние анализируемого объекта наблюдения;
- сигнал оповещения оператору в случае, если комплексная количественная оценка потенциальной опасности состояния индивида превышает заданное пороговое значение.

Основные группы конструктивных элементов АСДА:

- видеокамеры,
- микрофоны,
- СВЧ/КВЧ-радары,
- вычислительная среда
- АРМ-операторов.

Коммуникация и информационный обмен в АСДА осуществляется посредством проводных и/или беспроводных интерфейсов связи между устройствами. Обобщенная структура АСДА показана на рис. 1.

При этом в общем случае вычислительная среда может состоять из нескольких компьютеров. К каждому из таких компьютеров имеется опция подключения некоторого количества источников данных: камер, микрофонов и СВЧ/КВЧ-радаров. Целесообразно подключать источники данных, направленные на одно рабочее место к одному и тому же вычислительному устройству. В общем случае одно вычислительное устройство может анализировать параметры состояния одного или нескольких работников. Будем считать один компьютер с подключенными к нему источниками данных и интерфейсами работы АРМ-операторов одной единицей АСДА, как это показано на рис. 2.

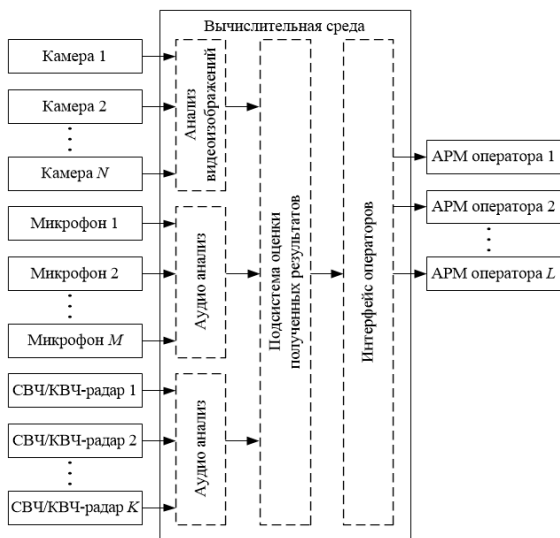


Рис. 1. Обобщенная структура АСДА

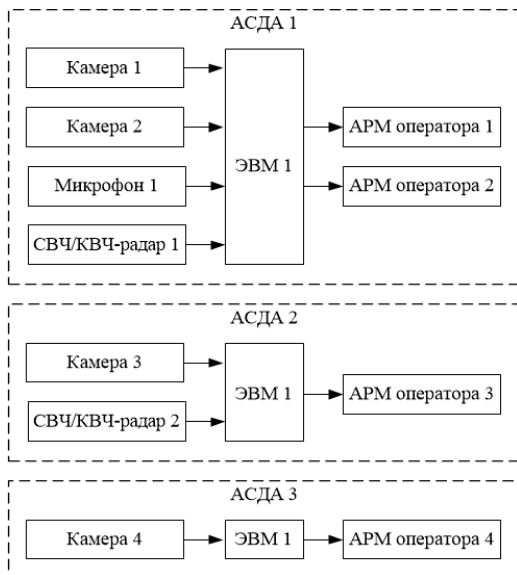


Рис. 2. Пример составов АСДА, применяемых на предприятии

Некоторые экземпляры АСДА могут обладать полным набором источников данных: видеокамеры, микрофоны, СВЧ/КВЧ-радары, в то время, как другие экземпляры АСДА имеют в своем составе только ограниченные наборы источников данных. Например, только видеокамеры, или видеокамеры и СВЧ/КВЧ-радары без микрофона. Такой подход целесообразен в связи с тем, что различные экземпляры АСДА могут применяться в различных сценариях.

## 2. Архитектура АСДА

В состав АСДА входят следующие подсистемы:

- регистрации видеоизображения;
- анализа видеоизображения;
- аудио анализа;
- дистанционной регистрации физиологических показателей на основе бесконтактного СВЧ/КВЧ-радара;
- оценки полученных результатов.

Схематично полный состав АСДА изображен на рис. 3.



Рис. 3. Схематическое изображение подсистем, входящих в состав АСДА

Подсистема регистрации видеоизображения предназначена для получения видеоданных об индивиде, состояние которого контролируется АСДА. В состав подсистемы регистрации видеоизображения включено следующее оборудование:

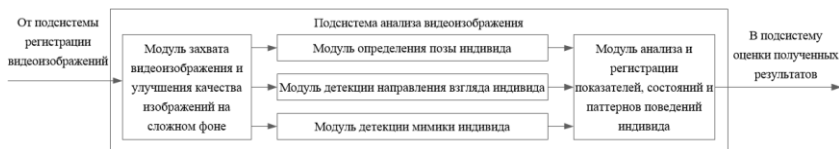
- RGB цифровые видеокамеры, в том числе высокоскоростные видеокамеры для получения данных о цвете;
- RGBD-камеры (камеры глубины) – для получения данных о расстоянии от точки до объектива (опционально);
- инфракрасные камеры (опционально);
- стереопары камер для формирования стереоизображения (опционально).

На вход подсистема анализа видеоизображений получает видеоданные от подсистемы регистрации видеоизображений. Выходными данными данной подсистемы являются регистрируемые показатели, состояния и паттерны поведения, представленные ниже.

Данная подсистема включает в себя следующие модули:

- модуль захвата видеоизображения и улучшения качества изображений на сложном фоне;
- модуль определения позы индивида;
- модуль детекции направления взгляда индивида;
- модуль детекции мимики индивида;
- модуль анализа и регистрации показателей, состояний и паттернов поведения индивида.

На рис. 4 схематично показаны направления информационных потоков между модулями в подсистеме анализа видеоизображений.



*Рис. 4. Модули подсистемы анализа видеоизображений*

Подсистема аудио анализа, включающая:

- микрофоны для записи голоса индивида;
- программные компоненты анализа аудиоинформации.

Технические требования к микрофонам:

- полоса частот от 30 Гц до 10 кГц;
- стереомикрофон с разнесенными каналами.

Расположение микрофонов должно обеспечивать возможность записи голоса индивида, состояние которого оценивается.

Принципы работы подсистемы дистанционной регистрации физиологических показателей на основе бесконтактного СВЧ/КВЧ-радар основываются на методе биорадиолокации [Anishchenko et al., 2018], [Evtveeva et al., 2019], [Абрамов и др., 2010], [Анищенко и др., 2018a], [Анищенко и др., 2018b].

### 3. Методы и алгоритмы

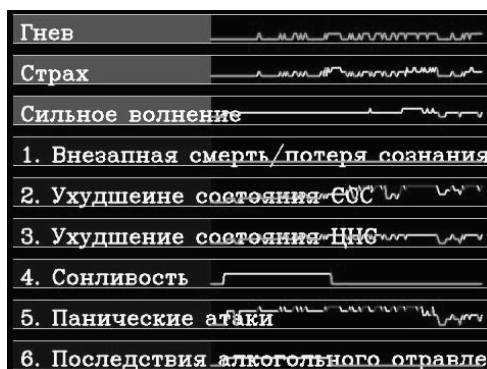
При разработке программного компонента подсистемы регистрации видеоизображений были разработаны следующие алгоритмы:

- алгоритм анализа конфигурации скелета человека;
- алгоритм детектирования области лица человека;
- алгоритм анализа мимики лица человека;
- алгоритм детектирования центров зрачков глаз;
- алгоритм определения направления взгляда человека.

### 4. Результаты экспериментальных исследований

На основе предложенных методов и алгоритмов было разработано программное обеспечение, позволяющее отслеживать признаки особых психоэмоциональных состояний человека.

На рис. 5 приведены графики во время онлайн-мониторинга состояния оператора. Полученные результаты передаются в блок анализа параметров.



*Рис. 5. Фрагмент экрана программного обеспечения во время экспериментальных исследований системы*

Для проведения экспериментальных исследований было использовано несколько датасетов (VAM [Grimm et al., 2008] и датасет собственной разработки с использованием добровольцев) и проведено сравнение с известным аналогом – утилиты на базе AffectNet [Mollahosseini et al., 2017].

Видеофрагменты датасета VAM слишком короткие, а разрешение видео слишком низкое для точного детектирования мимических признаков. Тем не менее, было проведено сравнение работы разработанного ПО и утилиты на базе AffectNet на данных датасета VAM.

Также был разработан новый датасет. Для этого было привлечено 10 добровольцев обоих полов. Проведена серия видеозаписей моделирования задач профессиональной деятельности с каждым из добровольцев. По итогам был сформирован датасет фрагментов видеоизображений, каждое из которых начинается и заканчивается нейтральным состоянием, а также содержит моделирование первичных мимических признаков.

Результаты сравнения, разработанного программного обеспечения с зарубежным открытым аналогом AffectNet определения эмоционального состояния по изображениям лица человека на двух описанных выше датасетах показали:

- одноименные психоэмоциональные состояния показали высокий уровень совпадения результатов детекции для обеих систем (более 70%);
- система AffectNet детектирует эмоциональные состояния только на изображениях с ярко выраженным мимическим проявлением эмоционального состояния индивида, а также не учитывает позу индивида. В остальных случаях результатом обработки является «нейтральное выражение лица».
- система AffectNet работает только с одиночными изображениями лица человека, а не последовательностью видеокадров, в отличие от программного обеспечения, описываемого в данной статье.

### Заключение

Предложенные в данной работе концепция, архитектура, методы и алгоритмы были использованы при разработке программного обеспечения онлайн-мониторинга состояния операторов оборудования областей ответственного применения. Результаты экспериментальных исследований показали его работоспособность и невысокую вычислительную сложность.

Дальнейшая работа направлена на совершенствование алгоритмов анализа физиологического и психоэмоционального состояния оператора с целью своевременного предотвращения возможных сбоев в работе оборудования. Такие системы предназначены для повышения безопасности использования оборудования и с целью своевременного оказания медицинской помощи операторам в случае болезни. Разрабатываемые системы найдут применение в таких областях, как энергетика, транспорт и др.

### Список литературы

- [Anishchenko, 2018] Anishchenko L. Challenges and potential solutions of psychophysiological state monitoring with bioradar technology // *Diagnostics*. – 2018. – Vol. 8, No. 4. – P. 73.
- [Evtseva et al., 2019] Evtseva K., Turetskaya A., Anishchenko L. Sensing of human breathing and heart beating at different orientation angles by CW Doppler radar // 2019 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBREIT). – 2019. – P. 36-39.

- [Абрамов и др., 2010] Абрамов А.В. и др. Биорадиолокация / под ред. А.С. Бугаева. – М.: МГТУ им. Баумана, 2010. – 396 с.
- [Анищенко и др., 2018a] Анищенко Л.Н., Васильев И.А., Чиж М.А. Радиолокатор ближнего действия с непрерывным излучением для исследования биологических объектов // Электромагнитные волны и электронные системы. – 2018. – Т. 23, № 6. – С. 20-24.
- [Анищенко и др., 2018b] Анищенко Л.Н., Турецкая А.В. Программно-аппаратный комплекс для бесконтактной оценки психофизиологического состояния человека // Биомедицинская радиоэлектроника. – 2018. – № 10. – С. 61-67.
- [Grimm et al., 2008] Grimm M., Kroschel K., Narayanan S. The Vera am Mittag German audio-visual emotional speech database // 2008 IEEE international conference on multimedia and expo. – 2008. – P. 865-868.
- [Mollahosseini et al., 2017] Mollahosseini A., Hasani B., Mahoor M.H. Affectnet: A database for facial expression, valence, and arousal computing in the wild // IEEE Trans. Affect. Comput. – 2017. – Vol. 10, Issue 1. – P. 18-31.

---

## ПРИЛОЖЕНИЕ

УДК 004.8

### ЛЕТНЯЯ ШКОЛА РОССИЙСКОЙ АССОЦИАЦИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ЛШ РАИИ-2023

Н.Г. Ярушкина (*jng@ulstu.ru*)

В.С. Мошкин (*v.moshkin@ulstu.ru*)

Ульяновский государственный технический университет, Ульяновск

Г.Е. Веселов (*info@ictis.sfedu.ru*)

Л.А. Gladkov (*lagladkov@sfedu.ru*)

Южный федеральный университет, Таганрог

Представлены итоги работы Летней школы Российской ассоциации искусственного интеллекта ЛШ РАИИ-2023, организованной и проведенной совместно с Институтом компьютерных технологий и информационной безопасности ЮФУ и Ульяновским государственным техническим университетом в г. Таганроге в июле 2023 года.

**Ключевые слова:** Летняя школа, искусственный интеллект, Российская ассоциация искусственного интеллекта, Ульяновский государственный технический университет, Институт компьютерных технологий и информационной безопасности ЮФУ.

#### Введение

С 17 по 28 июля 2023 г. в г. Таганроге прошла Летняя школа Российской ассоциации искусственного интеллекта ЛШ РАИИ-2023. Работа ЛШ РАИИ-2023 была организована Российской ассоциацией искусственного интеллекта совместно с Институтом компьютерных технологий и информационной безопасности (ИКТИБ) ЮФУ и Ульяновским государственным техническим университетом (УлГТУ).



В этом году ЛШ РАИИ-2023 была посвящена памяти выдающегося исследователя в сфере искусственного интеллекта, основателя научной школы в области нечетких систем – *Мелихова Аскольда Николаевича*.

Председатель Программного комитета ЛШ РАИИ-2023 – Ярушкина Надежда Глебовна, д.т.н., профессор, ректор УлГТУ, Ульяновск; заместитель председателя Программного комитета – Мошкин Вадим Сергеевич, к.т.н., доц., УлГТУ, Ульяновск.

Председатель Организационного комитета – Веселов Геннадий Евгеньевич, д.т.н., доцент, директор ИКТИБ ЮФУ; заместитель председателя Организационного комитета – Гладков Леонид Анатольевич, к.т.н., доцент, ИКТИБ ЮФУ.

Целью ЛШ РАИИ-2023 являлось получение участниками новых знаний, умений и навыков в области искусственного интеллекта, применение их на практике, знакомство с сообществом исследователей и друг другом.

Отбор участников осуществлялся через сайт ЛШ РАИИ-2023: <https://raai-school.ru>.

В результате отбора в работе Летней школе приняло участие 85 студентов, аспирантов и докторантов из 23 учебных заведений России:

- Белгородский государственный НИУ, Белгород;
- Военная академия ВПВО ВС РФ им. А.М. Василевского, Смоленск;
- Поволжский государственный технологический университет, Йошкар-Ола;
- Донской государственный технический университет, Донецк;
- Донецкий национальный университет, Донецк;
- Институтом компьютерных технологий и информационной безопасности ЮФУ, Таганрог;
- Иркутский государственный университет, Иркутск;
- МГТУ им. Н.Э. Баумана, Москва;
- МГУ им. М.В. Ломоносова, Москва;
- НИУ «МЭИ», Москва;
- НИТУ «МИСИС», Москва;
- НИЯУ «МИФИ», Москва;
- Приазовский государственный технический университет, Мариуполь;
- Пятигорский государственный университет, Пятигорск;
- Ростовский государственный университет путей сообщения, Ростов-на-Дону;
- РУДН, Москва;
- Таганрогский институт имени А.П. Чехова, Таганрог;
- Удмуртский государственный аграрный университет, Ижевск;
- Ульяновский государственный технический университет, Ульяновск;

- Филиал НИУ «МЭИ» в г. Смоленске, Смоленск;
- Финансовая академия при Правительстве РФ, Москва;
- Чувашский государственный университет им. И.Н. Ульянова, Чебоксары;
- Южный федеральный университет, Ростов-на-Дону.

## 1. Лекции и практические занятия

За 12 дней ведущими отечественными исследователями в области искусственного интеллекта были прочитаны 18 лекций, проведены 2 круглых стола, организовано по 8 практических занятий для каждого потока участников ЛШ РАИИ-2023 по следующим 4-м трекам:

- Нечеткие системы и Эволюционные вычисления;
- Инженерия знаний и Экспертные системы;
- NLP и Предиктивная аналитика;
- Робототехника и Computer Vision.

Темы лекций ЛШ РАИИ-2023 представлены в табл. 1.

Таблица 1

№ п/п	ФИО лектора	Краткие сведения	Тема лекции
1	<b>Борисов Вадим Владимирович</b>	д.т.н., профессор, профессор кафедры вычислительной техники, руководитель лаборатории интеллектуальных технологий и высокопроизводительных вычислений филиала НИУ «МЭИ» в г. Смоленске, президент РАИИ	Мягкие модели и методы в задачах искусственного интеллекта
2	<b>Кобринский Борис Аркадьевич</b>	д.м.н., профессор, зав. отделом систем поддержки принятия клинических решений Института проблем искусственного интеллекта ФИЦ ИУ РАН, профессор факультета ВМК МГУ им. М.В. Ломоносова, профессор кафедры медицинской кибернетики и информатики РНИМУ им. Н.И. Пирогова, председатель Научного совета РАИИ	Этапы и «подводные камни» при построении экспертных систем
3	<b>Ярушкина Надежда Глебовна</b>	д.т.н., профессор, ректор Ульяновского государственного технического университета, член Научного совета РАИИ, председатель Программного комитета ЛШ РАИИ-2023	Гибридные и интегрированные интеллектуальные системы
4	<b>Ефремов Денис Александрович</b>	директор Научно-образовательного центра «Точка кипения УлГТУ»	Теория и практика молодых стартапов
5	<b>Панов Александр Игоревич</b>	к.ф.-м.н., доцент, заведующий отделом интеллектуальных динамических систем и когнитивных исследований ФИЦ ИУ РАН, с.н.с. МФТИ	Обучение с подкреплением и большие языковые модели в приложении к робототехнике

6	<b>Гладков Леонид Анатольевич</b>	к.т.н., доцент, профессор кафедры систем автоматизированного проектирования ИКТИБ ЮФУ, член Научного совета РАИИ, зам. председателя Организационного комитета ЛШ РАИИ-2023	Генетические алгоритмы: эволюция от простых алгоритмов до нечетких гибридов
7	<b>Гуськов Глеб Юрьевич</b>	к.т.н., доцент кафедры «Информационные системы» УлГТУ	Применение моделей искусственного интеллекта и машинного обучения в задачах медицины и молекулярной биологии
8	<b>Андриянов Никита Андреевич</b>	к.т.н., доцент Департамента анализа данных и машинного обучения Финансового университета при Правительстве РФ	Развитие генеративных моделей изображений
9	<b>Еремеев Александр Павлович</b>	д.т.н., профессор, профессор НИУ «МЭИ», член Научного совета РАИИ	Методы и программные средства конструирования современных интеллектуальных (на основе экспертных знаний) СППР реального времени
10	<b>Забезайло Михаил Иванович</b>	д.ф.-м.н., профессор МФТИ, зав. отделом ФИЦ ИУ РАН, зам. председателя Научного совета РАИИ	Эволюция методов и инструментария ИИ-исследований и разработок: три «волны» развития ИИ
11	<b>Карпов Валерий Эдуардович</b>	д.т.н., доцент, начальник лаборатории робототехники НИЦ «Журчатовский институт», вице-президент РАИИ	Модели социального поведения в групповой робототехнике
12	<b>Курейчик Владимир Викторович</b>	д.т.н., профессор, заведующий профессор кафедрой систем автоматизированного проектирования ИКТИБ ЮФУ	Эволюционное моделирование и биоинспирированная оптимизация: теория и практика
13	<b>Лукашевич Наталья Валентиновна</b>	д.т.н., профессор, профессор кафедры теоретической и прикладной лингвистики МГУ им. М. В. Ломоносова	Языковые модели: этапы развития до ChatGPT. Часть 1
14	<b>Лукашевич Наталья Валентиновна</b>	д.т.н., профессор, профессор кафедры теоретической и прикладной лингвистики МГУ им. М. В. Ломоносова	Языковые модели: этапы развития до ChatGPT. Часть 2
15	<b>Подвесовский Александр Георгиевич</b>	к.т.н., доцент, профессор кафедры «Информатика и программное обеспечение» Брянского государственного технического университета	Подходы, методы и программные средства когнитивного моделирования слабоструктурированных систем
16	<b>Романов Антон Алексеевич</b>	к.т.н., доцент, заведующий кафедрой «Информационные системы» УлГТУ	Нечеткие модели временных рядов в предиктивной аналитике

17	<b>Святлов Кирилл Валерьевич</b>	к.т.н., доцент, декан факультета информационных систем и технологий УлГТУ, тренер студенческих команд-победителей Всероссийских соревнований по робототехнике и беспилотному транспорту	Основы программирования роботов в ROS
18	<b>Филиппов Алексей Александрович</b>	к.т.н., доцент, доцент кафедры «Информационные системы» УлГТУ	Современные программные средства для построения и использования баз знаний интеллектуальных систем

Практические занятия в рамках перечисленных треков проводились под руководством исследователей, практикующих разработчиков интеллектуальных информационных систем и технологий (табл. 2).

Таблица 2

№ п/п	ФИО руководители практических занятий	Краткие сведения
1	<b>Гуськов Глеб Юрьевич</b>	к.т.н., доцент кафедры «Информационные системы» УлГТУ
2	<b>Григорьев Андрей Александрович</b>	Ассистент кафедры «Кибернетика» НИЯУ «МИФИ», лаборатория «Интеллектуальные системы и технологии»
3	<b>Дырочкин Александр Александрович</b>	Ведущий программист Департамента цифровой трансформации УлГТУ
4	<b>Ефремов Денис Александрович</b>	Директор Научно-образовательного центра «Точка кипения УлГТУ»
5	<b>Желепов Алексей Сергеевич</b>	“Engineering Team Lead”
6	<b>Курило Дмитрий Сергеевич</b>	Ведущий программист Департамента цифровой трансформации УлГТУ
7	<b>Моисеев Владислав Валерьевич</b>	Ведущий инженер-программист компании «АИС Город»
8	<b>Михайлов Владислав Святославович</b>	Ведущий программист Центра новых информационных технологий УлГТУ
9	<b>Лукьянова Елена Николаевна</b>	Начальник отдела развития молодёжного технологического предпринимательства Научно-образовательного центра «Точка кипения УлГТУ»

## 2. Проектная работа

В рамках проектной работы все участники ЛШ РАИИ-2023 были разбиты на 9 команд, и в течение всех двух недель под руководством научных наставников разрабатывали следующие проекты с использованием методов, моделей и технологий искусственного интеллекта.

- «AI-support. Автоматизация обработки обращений в службу поддержки»;
- «Разработка чат-бота для рекомендации фильма по сюжету»;
- «Разработка системы составления писем»;
- «Система интеллектуального управления адаптивной подготовкой специалистов организаций»;
- «Фильтрация новостного потока на основе эмоциональной окраски и тематики на платформе Telegram»;
- «PlayVision AI. Веб-сайт для просмотра матчей по футболу»;
- «Гид Том. Система составления оптимального маршрута»;
- «Система экспертных оценок городских вывесок»;
- «The Phantome. Интеллектуальный поисковый комплекс для обнаружения людей».

На итоговом заседании ЛШ РАИИ-2023 команды представили проекты, уделив внимание не только инновационности разработок, но перспективам их дальнейшей коммерциализации.

Проекты оценивались по следующим критериальным показателям:

- технологический стек;
- целевая аудитория;
- качество плана разработки;
- MVP;
- экономические показатели;
- презентация результатов.

### **3. Культурная программа**

Помимо учебы для участников ЛШ РАИИ-2023 была организована насыщенная культурная программа: экскурсии по Таганрогу и в Народный военно-исторический музей Великой Отечественной войны «Самбекские высоты», посещение спектакля в Таганрогском театре им. А.П. Чехова.

### **Заключение**

По результатам защиты проектов экспертной комиссией были отобрали 3 лучших проекта, проектным командам которых организаторы вручили призы и памятные сувениры.

По итогам работы ЛШ РАИИ-2023 всем ее участникам были вручены сертификаты, а также удостоверения об успешном освоении программы дополнительного профессионального образования «Искусственный интеллект и анализ данных» (48 часов).

В результате проведения ЛШ РАИИ-2023 достигнута поставленная цель – получение участниками новых знаний, умений и навыков в области искусственного интеллекта, применение их на практике, а также знакомство, налаживание тесных контактов с сообществом исследователей и друг другом.

Работа по организации и проведению летних школ РАИИ должна иметь регулярный характер в тесном взаимодействии с учебными заведениями, к ней активно должны привлекаться как ведущие ученые, так и молодые исследователи, занимающиеся теоретическими и прикладными исследованиями и разработками в области искусственного интеллекта.

## *Abstracts of Sectional Reports*

### *Section 5*

#### *Fuzzy Models, Soft Measurements and Computing*

##### **POSSIBILITIES OF APPLICATION OF NEURO-FUZZY NETWORKS FOR OPHTHALMOLOGIC IMAGE CLASSIFICATION**

A.N. Averkin (*averkin2003@inbox.ru*)  
FRC "Informatics and Management" RAS, Moscow

E.N. Volkov (*envolkoff@gmail.com*)  
Dubna State University, Dubna

S.A. Yarushev (*Yarushev.sa@rea.ru*)  
Plekhanov Russian University of Economics, Moscow

The application of artificial intelligence technologies in image analysis tasks is rapidly developing every year, thanks to which the possibilities of using artificial neural networks (ANN) in the diagnosis of various diseases have reached a whole new level. The results of thousands of studies with the use of ANN in ophthalmology show the results of accuracy of diagnosis comparable, and sometimes even superior to those of the human performer. However, the transparency of the results obtained and the adaptability of such systems remain low. The use of convolutional neuro-fuzzy networks (CFNN) in image analysis tasks may improve the results in the future. The paper reviews the existing experience of applying neuro-fuzzy networks for medical image classification tasks, demonstrates the possibilities of practical realization of neuro-fuzzy network for classification tasks of ophthalmic glaucoma images.

**Keywords:** Artificial intelligence, explanatory artificial intelligence, fuzzy logic, neuro-fuzzy networks, healthcare, medicine, personalized medicine, ophthalmology, ophthalmoscopy, glaucoma, diagnosis, XAI.

##### **MODELING OF PLANNING PROCESSES BASED ON TEMPORAL NEURO- FUZZY PETRI NETS**

A.V. Bobryakov (*avbob@mail.ru*)  
National Research University "MPEI", Moscow

S.A. Prokopenko (*puss95@yandex.by*)  
National Research University "MPEI" Moscow  
Inter-state educational institution of higher education  
"Belarusian-Russian university" Mogilev

Modern production technologies make it possible to produce a variety of products without the need for huge production capacities and with limited resources. Such productions are influenced by a complex and rapidly changing set of external conditions, where ef-

fective planning and optimization of resources become key success factors, while considering the possible trajectories of the production of complex technical objects, the plan for the employment of equipment and personnel, which helps to optimize processes and reduce product costs. Effective planning of such complexity is impossible without appropriate modeling, algorithmic and software. The article proposes an approach to modeling the processes of planning the production of complex technical objects using the proposed variety of temporal neuro-fuzzy Petri nets, considers algorithmic and software for the implementation and maintenance of planning processes.

**Keywords:** planning processes, temporal neuro-fuzzy Petri nets.

### **SCENARIO MODELING BASED ON FUZZY RELATIONAL TEMPORAL COGNITIVE MODELS**

V.V. Borisov (*vbor67@mail.ru*)

A.P. Zharkov (*antonzharkov@yandex.ru*)

The Branch of National Research University

“Moscow Power Engineering Institute” in Smolensk, Smolensk

V.S. Lufarov (*lyferov@yandex.ru*)

“T1 Digital” LLC, Smolensk

Fuzzy Relational Temporal Cognitive Models (FRTCM) are described and a scenario modeling method based on these models is proposed. The method includes: firstly, the construction of the FRTCM structure and its structural adjustment; secondly, the choice of scenario dynamics models for all FRTCM concepts; thirdly, parametric adjustment of the FRTCM; fourthly, the designing of modeling scenarios; fifth, the modeling based on trained FRTCM in accordance with a scenario. An example of using the proposed method for scenario modeling of electric load distribution of consumers of a regional energy subsystem is considered.

**Keywords:** fuzzy relational temporal cognitive model, scenario modeling, scenario dynamics model.

### **EXPANDING THE FUNCTIONALITY OF NODES IN WIRELESS SENSOR NETWORKS**

G.P. Vinogradov (*wgp272ng@mail.ru*)

Tver State Technical University, Tver

Program Systems Research Institute, Tver

The purpose of the work is to develop methods for extending functionality in control systems of wireless sensor network nodes based on unclear logical inference and pattern theory. A formal model of the pattern is presented, which includes a number of mental parameters. The method of its implementation in the form of an algorithmic model is shown, and an algorithm for its implementation is developed. The article uses the meth-



ods of fuzzy set theory, the theory of constructing fuzzy models and networks. The described approach is implemented in the form of a software product for TinyOS. It is written in the nesC programming language, a dialect of C.

**Keywords:** model, fuzzy logic, pattern, wireless network, sensor, artificial intelligence.

### HYBRID ALGORITHM FOR SOLVING THE PROBLEM OF PLACEMENT OF ELEMENTS OF COMPUTING EQUIPMENT

L.A. Gladkov (*leo\_gladkov@mail.ru*)

M.D. Yasir (*yasir\_82@mail.ru*)

N.V. Gladkova (*nadyusha.gladkova77@mail.ru*)

Southern Federal University, Taganrog

The article deals with the problem of placing elements of digital electronic equipment. The statement of the problem, as well as the main limitations and optimization criterion is given. A method for solving the placement problem based on a hybrid approach is proposed, a model of a hybrid algorithm for solving the placement problem. A scheme of a hybrid fuzzy genetic algorithm has been developed. It is proposed to use the island model of a parallel multipopulation genetic algorithm for organizing calculations. A procedure has been developed that makes it possible to implement the principle of multi-threading at the local level. The principles of operation and the structure of the fuzzy control block are described. The control parameters of the genetic algorithm are determined. The software implementation of the proposed hybrid algorithm has been completed. A series of computational experiments were carried out to determine the efficiency of the developed algorithm and to select the optimal values of the control parameters. An analysis of the obtained data was carried out and recommendations for further continuation of the research were developed.

**Keywords:** computer-aided design, placement problem, fuzzy genetic algorithms, fuzzy control, hybrid models, parallel computing.

### FUZZY-DYNAMIC METHOD OF THE PRINCIPAL COMPONENT ANALYSIS IN INTELLIGENT MONITORING SYSTEMS

A.I. Dolgiy

JSC NIIAS, Rostov-on-Don

S.M. Kovalev (*ksm@rfniias.ru*)

Rostov branch JSC NIIAS, Rostov-on-Don

A.E. Kolodenkova (*anna82\_42@mail.ru*)

Samara State Technical University, Samara

The article develops a new approach to intelligent monitoring of the states of complex technological objects, which include poorly formalized technological processes, complex technical complexes and devices. The proposed approach is based on a chain of transformations between the spaces of primary parameters of the controlled object, generalized meta-parameters and linguistic descriptions of technological situations and states

of the controlled object. The central element in the chain of transformations is the new method of fuzzy dynamic analysis of principal component proposed by the authors, based on the combination of nonlinear analysis of principal component and fuzzy evolving systems. Due to the natural interpretability of semantic space variables, it is possible not only to detect abnormal states and facts of violations in the behavior of the controlled object, but also to identify the types, scale and direction of violations. A model example of using the proposed intelligent monitoring technology to assess the conditions and predict the behavior of technological processes occurring at railway stations is given.

**Keywords:** diagnostics of complex objects, principal component analysis, multiple parameters.

### OPTIMIZATION BY THE SWARM OF TRANSFORMING CHROMOSOMES METHOD

B.K. Lebedev (*lebedev.b.k@gmail.com*)

O.B. Lebedev (*lebedev.ob@mail.ru*)

Southern Federal University, Taganrog

M.A. Ganzhur (*mganzhur@yandex.ru*)

Don State Technical University, Rostov-on-Don

A new strategy for finding efficient solutions based on the modification and hybridization of canonical representations of the particle swarm algorithm is proposed. The paradigm of a swarm of transforming chromosomes is described, which provides the possibility of representing solutions in the affine space of chromosomes with integer values of the parameters. The mechanisms of transformation of chromosomes in the affine space to increase the weight of affine bonds are considered. The chromosomes updated after the transformation are, in turn, the base points in subsequent transformations (an uninterrupted chain of constant rebirths in accordance with the principle of samsara). The overall time complexity estimate lies within  $O(n^2)$ - $O(n^3)$ .

**Keywords:** swarm of transforming chromosomes, static population, flocking behavior, affine space, chromosome structures, weight of affine bonds, transformation, directed mutation operator, optimization.

### IMPROVING THE QUALITY OF PRODUCTION MANAGEMENT PROCESSES BASED ON NEURAL NETWORK AND NEURO-FUZZY MODELS AND TOOLS

A.E. Misnik (*anton@misnik.by*)

M.A. Shalukhova (*shalukhova.m@yandex.ru*)

Inter-state Educational Institution of Higher Education

"Belarusian-Russian University", Mogilev

The article describes a way to improve the quality of product control processes in food production by using neural network and neuro-fuzzy methods, models and tools. It is proposed to use feature extraction using convolutional networks with further post-processing in a fuzzy inference system. During the operation of the proposed system, a

high percentage of correct recognitions was obtained, as well as during the operation the customer returns due to rejects decreased by 63% over the same period last year. The results show that defect identification using adaptive neuro-fuzzy inference system is a suitable tool for solving defect analysis problems.

**Keywords:** neural network approach, fuzzy logic, defect recognition, neural network, ANFIS.

### **AN APPROACH TO GENERATING FORMAL FUZZY COGNITIVE MAPS FOR EXPERIMENTAL STUDIES IN COGNITIVE MODELING**

A.G. Podvesovskii (*apodv@tu-bryansk.ru*)

R.A. Isaev (*ruslan-isaev-32@yandex.ru*)

I.A. Kopeliovich (*igor.copeliovitch@yandex.ru*)

Bryansk State Technical University, Bryansk

The paper considers the tasks of experimental research of new methods of cognitive modeling by testing them on a set of artificially created cognitive models. The features of the process of obtaining such models are studied, and the concepts of formal cognitive map and formal cognitive model are introduced. An approach to generating a formal cognitive map with given characteristics, based on the use of random graph models, has been proposed and partially formalized. A modification of the configuration model is provided, and examples of generation of formal fuzzy cognitive maps based on it are given.

**Keywords:** cognitive modeling, fuzzy cognitive map, random graph models, configuration model.

### **THE METHOD OF EVALUATION FOR COMPLEX SYSTEMS AND PROCESSES BASED ON ONTOLOGICAL AND NEURO-FUZZY MODELING**

I.A. Prokopenkov (*prokopenkoff.ivan@yandex.ru*)

I.A. Frolov (*igor-frolov-81@mail.ru*)

Military Academy of Military Air Defense of the Armed Forces of the Russian Federation, Smolensk

V.I. Medennikov (*dommed@mail.ru*)

Federal Research Center for Informatics and Control  
of the Russian Academy of Sciences, Moscow

For various sectors of the economy (in industry, in the agro-industrial complex, in transport, in the service sector) there is a problem of quality and efficiency of evaluation of interacting information, management, technological processes. The paper presents the evaluating method for complex systems and processes using the representation of the problem domain in the form of an ontological model, as well as a neuro-fuzzy classifier for evaluating existing precedents.

**Keywords:** ontological model, precedent, neuro-fuzzy classifier.

## METHOD OF INFERENCE OF FUZZY LOGICAL TYPE SYSTEMS WITH NON-SINGLETON FUZZIFICATION

V.G. Sinuk (*vgsinuk@mail.ru*)  
S.A. Karatach (*karatach1998@yandex.ru*)  
BSTU after V.G. Shukhov, Belgorod

In the article, the output result is obtained for widely used fuzzy logic-type systems with non-singleton fuzzification. It was achieved on the basis of an approach using a fuzzy truth value, which made it possible to reduce the problem to polynomial computational complexity. The article also describes the possibility of using OpenGL and CUDA technologies when implementing the algorithm of the presented fuzzy inference.

**Keywords:** non-singleton fuzzification, fuzzy truth value, computational complexity, efficient calculations.

## FUZZY SITUATIONAL PRECEDENT ANALYSIS AND MODELING OF PROCESSES IN COMPLEX TECHNICAL SYSTEMS

A.M. Sokolov (*andreisokol98@gmail.com*)  
National Research University «MPEI», Moscow

V.I. Budzko (*vbudzko@ipiran.ru*)  
A.A. Zatsarinny (*azatsarinny@frccsc.ru*)  
Federal Research Center for Informatics and Control  
of the Russian Academy of Sciences, Moscow

Modern production and technological systems, objects of the agro-industrial complex are characterized by complex dependencies between systemic and external factors, incompleteness and uncertainty of information about their condition and functioning. Fuzzy situational models are one of the promising ways to model such systems and can be used to support control decision-making. However, existing approaches to these models constructing do not allow to completely consider the relationship between the parameters and conditions of the system transition from one state to another in the absence of the necessary amount of retrospective information. The paper shows that in order to solve this problem, it is advisable to develop these approaches based on taking into account precedents. The author suggests an approach to building a control model of a complex production and technological system based on the combination of the fuzzy situational networks and a precedent approach. The proposed approach involves the automated creation of fuzzy situational networks, the individual components of which describe the precedents that took place earlier.

**Keywords:** fuzzy situational network, precedent approach, complex technical system, agro-industrial complex.

## FUZZY MODEL OF DIAGNOSTICS OF TECHNOLOGICAL PROCESSES

K.B. Fam (*phambang79520897405@gmail.com*)

P.M. Murashev (*myptver@gmail.com*)

V.N. Bogatkov (*vnbgtk@mail.ru*)

Tver State Technical University, Tver

In this paper, the formation of a condition assessment criterion based on a fuzzy safety index is considered on the example of industrial technologies for drying green tea and ball grinding mill. The formalized system of ensuring the safety of technological processes is considered. An algorithm for constructing a Hasse diagram is used to represent an ordered set of faults. The diagnostic task is considered on the example of industrial processes of drying green tea and grinding of apatitinefelin ore.

**Keywords:** diagnostics of technological processes, Hasse diagram, fuzzy equality, fuzzy inclusion, graph search

## ORGANIZATION OF PARALLEL COMPUTING ON HYBRID COMPUTING CLUSTERS USING FUZZY INTELLECTUAL ANALYSIS

Y.A. Fedulov (*fedulov\_yar@mail.ru*)

A.S. Fedulova (*voitsitskay@mail.ru*)

Branch of the National Research University Moscow Power Engineering  
Institute in Smolensk, Smolensk

The paper proposes a hardware-software model for organizing parallel computing on hybrid computing clusters, focused on creating tools for converting parallel programs into a hybrid form. A method has been developed for organizing the execution of parallel programs on hybrid computing clusters using compatible fuzzy cognitive maps, focused on the extended capabilities of the proposed model of the hardware and software environment, taking into account various hardware and software indicators, which makes it possible to reduce the share of exchange operations performed through slow network interfaces.

**Keywords:** fuzzy logic, parallel computing, parallel computing technologies, program conversion, fuzzy data analysis.

## Section 6

### ***Frameworks and Technologies for the Design of Intelligent Systems. Applied Intelligent Systems***

#### **INTELLIGENT HARDWARE AND SOFTWARE SUPPORT AND INCREASING THE EFFICIENCY OF WELDING PROCESSES**

S.V. Bolotov (*s.v.bolotov@mail.ru*)

K.V. Zakharchenkov (*zakharchenkovkv@mail.ru*)

S.K. Krutolevich (*s\_krutolevich@tut.by*)

Belarusian-Russian University, Belarus, Mogilev

The problem of improving the efficiency of welding processes at modern machine-building enterprises is unsolvable without the use of hardware and software systems that allow real-time monitoring of the parameters of each welder, however, in addition to tracking and logging such work, it is extremely important to rationally select the parameters of the welding process – in fact, a necessary component such complexes become intelligent assistants to technologists, based, among other things, on neural network technologies. This article discusses an integrated approach to the hardware and software support of welding processes, describes the developed hardware and software, as well as a neural network approach to the selection of rational parameters of the welding process.

**Keywords:** welding processes, hardware and software systems, neural networks, LSTM.

#### **BIFURCATION DETECTIONS OF MULTIDIMENSIONAL RANDOM PROCESSES IN DYNAMIC EXPERT SYSTEMS**

A.N. Vetrov (*vetrov\_48@mail.ru*)

A.A. Malkov (*kja@list.ru*)

B.V. Paliukh (*pboris@tstu.tver.ru*)

Tver State Technical University, Tver

The paper describes the method and procedure for determining the bifurcation point in the technological process, at which the transition of the system to a critical (emergency) state is most likely. The method of a double sliding window is briefly described, which makes it possible to detect anomalies in a continuous multidimensional flow of information generated by sensors of technological equipment. The description of the software and technical platform for the analysis of streaming data is given.

**Keywords:** applied artificial intelligence systems, dynamic expert systems, diagnostics, bifurcation point.

## CREATING VIRTUAL ASSISTANTS BASED ON MODEL TRANSFORMATIONS

N.O. Dorodnykh (*tualatin32@mail.ru*)

O.A. Nikolaychuk (*nikoly@icc.ru*)

A.B. Stolbov (*stolboff@icc.ru*)

A.Yu. Yurin (*iskander@icc.ru*)

Matrosov Institute for System Dynamics and Control Theory Siberian Branch  
of the Russian Academy of Sciences, Irkutsk

The paper describes an original approach that uses the principles of Model-Based Development and transformations to create prototypes of virtual assistants. Models of all levels of abstraction are presented according to the methodology of MDA (Model-Driven Architecture). Fragments of the transformation of model elements are shown. A description of the main stages of the proposed approach is given. Detailed models were created to solve the problem of diagnosing the aircraft power supply system.

**Keywords:** task ontology, dialog ontology, model-based development, model transformation, virtual assistant.

## REPRESENTATION AND PROCESSING OF TEMPORAL CASES IN REAL-TIME INTELLIGENT SYSTEMS

A.P. Ereemeev (*eremeev@appmat.ru*)

M.V. Fomina (*m\_fomina2000@mail.ru*)

N.A. Paniavin (*PaniavinNA@mpei.ru*)

National Research University "MPEI", Moscow

The problem of increasing the efficiency of decision-making based on precedents (Case-Based Reasoning, CBR) in intelligent real-time systems is considered. A precedent base (BP) preprocessing method is proposed, which includes the stages of generalizing the properties of dynamic parameters, the formation of classes of similar situations associated with BP precedents, and the construction of decision trees for efficient search for solutions within each class. Examples of processing precedents, results of a machine experiment are given. The possibility of implementing the proposed approach based on a temporal database created in the environment of a non-relational graph DBMS Neo4j is considered.

**Keywords:** Intelligence System, Real Time, Case-Based Reasoning, Temporal Date, Nonrelational Data Base.

## **A FRAMEWORK FOR DEVELOPING INTELLIGENT TUTORING SYSTEMS BASED ON DOMAIN MODELS IN THE FORM OF DECISION TREES**

A.I. Krygin (*andrey\_krygin@outlook.com*)

M.R. Gumerov (*max\_person@mail.ru*)

N.A. Moskalenko (*nekit\_vlg@mail.ru*)

O.A. Sychev (*oasychev@gmail.com*)

Volgograd State Technical University, Volgograd

Formal domain models allow the development of intelligent tutoring systems based on symbolic AI capable of reliable identification of learners' errors, generation of explanatory feedback, pedagogical questions, and other predictable features. However, the development of such models for modern inference machines is time-consuming and labor-intensive. The paper describes a method for representing processes of solving and verifying problems of subject domains as decision trees and a prototype of the framework for developing intelligent tutoring systems based on it. According to the results of the development of four intelligent tutoring systems' cores, the framework showed moderate performance and a significant decrease of development time.

**Keywords:** decision trees, intelligent tutoring systems, domain models, framework.

## **KNOWLEDGE BASE OF EXPERT SYSTEM FOR ANALYSIS OF SQL-CODE IN INDUSTRIAL DATABASES**

V.S. Moshkin (*v.moshkin@ulstu.ru*)

M.M. Bazhutin (*myodo@mail.ru*)

N.G. Yarushkina (*jng@ulstu.ru*)

Ulyanovsk state technical university, Ulyanovsk

The article describes an approach to improving the efficiency of queries in the SQL language in order to optimize the time of sampling and processing data from highly loaded DBMS. The proposed approach involves the formation of an expert system knowledge base in the form of sets of rules that allow optimizing the time of data sampling from relational databases. The set of rules that increase the efficiency of processing data from the database can be divided into sets that solve the following problems: lack of profiling, performance problems caused by the combination of SQL and the algorithmic part of the task, SQL execution structure. The paper also presents the results of experiments on data processing of the information system of the Branch of PJSC "И" – Aviastar.

**Keywords:** knowledge base, expert system, SQL, databases, CTE, optimization.



## OPEN INTELLIGENT SYSTEMS: THE CONCEPT AND APPROACHES TO DEVELOPMENT

S.B. Rumovskaya (*sophiyabr@gmail.com*)  
Kaliningrad branch of FRC CSC RAS, Kaliningrad

Complex practical problems are subjective and dynamic. Such problems never repeat exactly, so the solution method must be designed anew each time. In this regard, it is relevant to develop intelligent systems that are purposeful, with the ability to acquire, reproduce and use knowledge. Its functions must be evaluated from the standpoint of ensuring sustainability, openness and real-time operation. The paper considers the understanding of intelligent systems and their "openness" as well as approaches to their development.

**Keywords:** intelligent system, open system, development approaches.

## INTELLIGENT TECHNOLOGY FOR DYNAMIC INTEGRATED EXPERT SYSTEMS CONSTRUCTION: FEATURES OF DESIGNING SIMULATION MODELS OF THE EXTERNAL ENVIRONMENT

G.V. Rybina (*gvrybina@yandex.ru*)  
A.A. Slinkov (*sanch971@mail.ru*)  
D.D. Belov (*dancher.00@gmail.com*)  
National Research Nuclear University "MEPhI", Moscow

This article focuses on the further development of intellectual technology for dynamic integrated expert systems (IES) construction with an emphasis on methods and tools for designing simulation models of the external environment. This is an essential step in the development of dynamic IES based on a problem-oriented methodology and the AT-TECHNOLOGY workbench. The onboard software complex of a spacecraft is used as an experimental problem area for prototyping a dynamic IES.

**Keywords:** dynamic integrated expert systems, integration, problem-oriented methodology, intelligent planner, simulation modeling, simulation model, temporal solver, spacecraft.

## IMPLEMENTATION OF DECISION-MAKING MECHANISM IN THE INTELLIGENT TUTORING SYSTEM BASED ON THE EXPERT SYSTEMS MODULE

V.A. Uglev (*vauglev@sfu-kras.ru*)  
Siberian Federal University, Zheleznogorsk

The paper deals with the general principle of organizing an intelligent software reasoner (solver) of an experimental Intelligent Tutoring System (ITS) based on the mechanisms of expert systems and cognitive visualization. The FLM\_Builder module is used as the basis for the software reasoner implementation. The specifics of the solver for the Python interpreter are described and the stages of forming the models of expert systems are given. The task of individualizing the composition of the learning course with subse-

quent visualization in the notation of Cognitive Maps of Knowledge Diagnosis, corresponding to the cross-cutting approach to the analysis of the learning situation in the tutoring system, is given as an illustration of the display mechanism operation.

**Keywords:** e-learning, Intelligent Tutoring Systems, decision making, expert systems, cognitive visualization.

## **INTELLIGENT USER SUPPORT SERVICE SYSTEM OF THE INFORMATION PORTAL**

N.Y. Shchukin (*shchukinny@gmail.com*)

A.V. Bobryakov (*avbob@mail.ru*)

A.S. Mokhov (*asmokhov@mail.ru*)

D.A. Shcherbakov (*shcherbakovdan@mpei.ru*)

NRU MPEI, Moscow

The article describes the structure of the intelligent user support service system and researches on solving the problem of finding relevant answers to user requests using the neural network of the Transformer architecture, which provides the representation of the requests in vector form and further search for the closest vectors in the database of frequently asked questions to form an answer.

**Keywords:** support service, text data analysis, frequently asked questions database, relevant answers, neural network, transformer.

## *Section 7*

### *Applied Biomedical Systems*

#### **AUTOMATION OF ABCDE-ANALYSIS OF NEOPLASM WITH INTERPRETATION OF RESULTS**

E.I. Akimova

G.Y. Guskov (*guskovgleb@gmail.com*)

N.G. Yarushkina (*jng@ulstu.ru*)

Ulyanovsk State Technical University, Ulyanovsk

The paper presents studies of neoplasms macroimages based on ABCD-analysis. The purpose of the study is to design tools for detecting changes in neoplasms according to the ABCD metric. The paper also presents an information system for evaluating and interpreting the result.

**Keywords:** expert system, clustering, ABCD analysis, image processing, neoplasm.

#### **METHODS OF SPACE PSYCHOLOGY AND ARTIFICIAL INTELLIGENCE TECHNOLOGIES. PROSPECTS FOR INTEGRATION**

Y.A. Bubeev (*aviamed@inbox.ru*)

SSC RF-IMBP RAS, Moscow

A.V. Ivanov (*avivanov@imbp.ru*)

SSC RF-IMBP RAS, Moscow

The paper describes the psychological features of space flights, some experimental models for studying the factors of space flight, describes the tasks of biomedical monitoring and psychophysiological methods of assessing the condition with the possibility of using elements of artificial intelligence.

**Keywords:** psychophysiological state, remote monitoring, isolation, asthenization, anti-orthostatic hypokinesia, dry immersion, communicative activity, facial expression analysis, speech analysis, actigraphy, AI systems in interplanetary flight.

## METHODS OF PROGNOSTIC ANALYSIS FOR THE PREDICTION OF IN-HOSPITAL MORTALITY IN PATIENTS WITH ACUTE ST-ELEVATION MYOCARDIAL INFARCTION AFTER PERCUTANEOUS CORONARY INTERVENTIONS

B.I. Geltser (*boris.geltser@vvsu.ru*)  
Valdivostok State University, Vladivostok

I.G. Domzhalov (*igor@domzhalov.ru*)  
Far Eastern Federal University, Vladivostok

N.S. Kuksin (*Nikita.Kuksin@vvsu.ru*)  
K.J. Shakhgeldyan (*carinash@vvsu.ru*)  
Far Eastern Federal University, Vladivostok  
Valdivostok State University, Vladivostok

The aim of the study is to develop an explainable machine learning model for predicting in-hospital mortality (IHF) in patients with ST-elevation myocardial infarction (STEMI) after percutaneous coronary intervention (PCI). We analyzed data from 4681 electronic medical records of patients with STEMI and identified 12 risk factors for IHF. The predictive models were developed based on multivariate logistic regression, random forest and stochastic gradient boosting methods. The search for threshold values on the grid while maximizing the area under the ROC-curve and their validation by Shapley's additive explanation method made it possible to verify the risk factors for IHF. The model, whose parameters were risk factors (AUC= 0.917 and 0.903 at cross-validation and final testing, respectively), was superior in accuracy to the best model with continuous predictors based on stochastic gradient boosting. The use of IHF risk factors as predictors makes it possible to explain the obtained prognosis and reduce the risk of adverse events after PCI.

**Keywords:** machine learning predictive models, risk factors, Shapley's additive explanation, stochastic gradient boosting, predictor thresholding.

## SOCIAL-EMOTIONAL CONVERSATIONAL AGENTS BASED ON COGNITIVE ARCHITECTURES AND MACHINE LEARNING

A.A. Dolgikh (*tolick.dolgih2013@yandex.ru*)  
A.V. Samsonovich (*alexei.samsonovich@gmail.com*)  
D.V. Tikhomirova (*dvsulim@mail.ru*)  
National Research Nuclear University MPhI, Moscow

Large Language Models (LLM) enable recognition of the topic of arbitrary statements, as well as their emotional coloring, but do not “understand” the logic of emotions, despite the fact that they can often generate adequate responses in a given context. On the other hand, cognitive architectures such as eBICA are able to model the dynamics of emotional states in the general case but require assistance in understanding the meaning

of statements and generating responses to them. This work introduces a new way to integrate LLM and eBICA, allowing them to complement each other. An experimental study based on the paradigms "virtual receptionist" and "virtual psychologist" is presented, showing encouraging results.

**Keywords:** LLM, cognitive architectures, emotional intelligence, artificial intelligence.

### **METHOD OF DEVELOPING AN INTELLIGENT SYSTEM FOR THE COMPLEX PURPOSE OF PERSONALIZED TREATMENT**

R.I. Kovalev (*koyal-995@mail.ru*)

V.V. Gribova (*gribova@iacp.dvo.ru*)

Institute of Automation and Control Processes Far Eastern Branch  
of the Russian Academy of Sciences, Vladivostok  
Far Eastern Federal University, Vladivostok

D.B. Okun (*okdm@iacp.dvo.ru*)

Institute of Automation and Control Processes Far Eastern Branch  
of the Russian Academy of Sciences, Vladivostok

The paper describes a method for developing decision support systems for the appointment of complex treatment. The approach is based on the creation of a task-oriented shell, the key component of which is a main ontology that allows you to form knowledge about various types of treatment (medication, restorative, surgical), regardless of the branch of medicine. This ensures the creation of knowledge bases in decision support systems for the complex purpose of patient treatment, thereby opening up the possibility of creating a single solver, which provides a significant reduction in system development costs. The paper describes the adaptation of ontology to various types of treatment and terminology of a knowledge base developer – an expert in the subject area. The ontology is implemented on the IACPaaS cloud platform and is currently already actively used to create knowledge bases in various fields of medicine.

**Keywords:** CDDS, ontology, knowledge base, problem domain ontology.

### **PRINCIPLES OF CONSTRUCTION SYSTEM FOR MONITORING THE PSYCHO-EMOTIONAL STATE AND HEALTH OF CRITICAL INFRASTRUCTURE OPERATORS**

B.A. Kobrinskii (*kba\_05@mail.ru*)

N.A. Blagosklonov (*nblagosklonov@frccsc.ru*)

O.G. Grigoriev (*oleggpolikvart@yandex.ru*)

Federal Research Center "Computer Science and Control",  
Russian Academy of Sciences, Moscow

The article discusses the principles of building an intelligent recommender system for monitoring critical infrastructure operators, focused on assessing their psycho-emotional state in real time and periodically assessing risk factors for chronic diseases. Information

flows of heterogeneous information and modules for preprocessing and analysis of heterogeneous data are described. The architecture of the system is presented, which provides for the issuance of information about deviations in the state of operators for decision makers at the facility and for medical workers.

**Keywords:** intelligent recommender system, operator state monitoring, psychoemotional state diagnosis, assessment of the risk of diseases, heterogeneous information, human factor, critical infrastructure.

## **TECHNOLOGIES FOR REMOTE MONITORING OF THE PSYCHO-EMOTIONAL STATE OF CRITICAL INFRASTRUCTURE OPERATORS**

M.V. Khisamutdinov (*L-V-P@yandex.ru*)

I.S. Korovin (*korovin\_yakov@mail.ru*)

Research Institute of Multiprocessor Computing Systems  
of the Southern Federal University, Taganrog

D.Ya. Ivanov (*donat.ivanov@gmail.com*)

Limited Liability Company "Research Institute of Multiprocessor Computing  
and Control Systems", Taganrog

The paper describes the proposed method for solving the problem of remote non-invasive monitoring of signs of preservation or loss of professional health and professional reliability by the operator. The concept and architecture of an automated system for remote analysis of signs of loss of operability by a critical infrastructure operator is proposed. Brief results of experimental studies and directions for further research are given.

**Keywords:** employee functional health monitoring, remote monitoring, video sequence analysis.

## *Abstract of Application*

### **SUMMER SCHOOL OF THE RUSSIAN ASSOCIATION FOR ARTIFICIAL INTELLIGENCE-2023**

N.G. Yarushkina (*jng@ulstu.ru*)

V.S. Moshkin (*v.moshkin@ulstu.ru*)

Ulyanovsk state technical university, Ulyanovsk

L.A. Gladkov (*lagladkov@sfedu.ru*)

G.E. Veselov (*info@ictis.sfedu.ru*)

South Federal University, Taganrog

The results of the Summer School of the Russian Association for Artificial Intelligence, organized and conducted jointly with the Institute of Computer Technology and Information Security of the Southern Federal University and Ulyanovsk State Technical University in Taganrog in July 2023, are presented.

**Keywords:** Summer School, artificial intelligence, Russian Association for Artificial Intelligence, Ulyanovsk State Technical University, Institute of Computer Technologies and Information Security of SFedU.

## АВТОРСКИЙ УКАЗАТЕЛЬ

Аверкин А.Н.	6	Зацаринный А.А.	131
Акимова Е.И.	277	Иванов А.В.	287
Бажутин М.М.	218	Иванов Д.Я.	342
Белов Д.Д.	242	Исаев Р.А.	97
Благосклонов Н.А.	333	Каратач С.А.	118
Бобряков А.В.	19, 267	Кобринский Б.А.	333
Богатиков В.Н.	141	Ковалев Р.И.	323
Болотов С.В.	162	Ковалев С.М.	66
Борисов В.В.	30	Колоденкова А.Е.	66
Бубеев Ю.А.	287	Копелиович И.А.	97
Будзко В.И.	131	Коровин Я.С.	342
Веселов Г.Е.	351	Крутолевич С.К.	162
Ветров А.Н.	173	Крыгин А.И.	206
Виноградов Г.П.	42	Куксин Н.С.	296
Волков Е.Н.	6	Лебедев Б.К.	77
Ганжур М.А.	77	Лебедев О.Б.	77
Гельцер Б.И.	296	Луферов В.С.	30
Гладков Л.А.	59, 351	Мальков А.А.	173
Гладкова Н.В.	55	Меденников В.И.	108
Грибова В.В.	323	Мисник А.Е.	88
Григорьев О.Г.	333	Москаленко Н.А.	206
Гумеров М.Р.	206	Мохов А.С.	267
Гуськов Г.Ю.	277	Мошкин В.С.	218, 351
Долгий А.И.	66	Мурашев П.М.	141
Долгих А.А.	310	Николайчук О.А.	183
Домжалов И.Г.	296	Окунь Д.Б.	323
Дородных Н.О.	183	Палюх Б.В.	173
Еремеев А.П.	194	Панявин Н.А.	194
Жарков А.П.	30	Подвесовский А.Г.	97
Захарченков К.В.	162	Прокопенко С.А.	19
			375



Прокопенков И.А.	108	Федулова А.С.	152
Румовская С.Б.	230	Фомина М.В.	194
Рыбина Г.В.	242	Фролов И.А.	108
Самсонович А.В.	310	Хисамутдинов М.В.	342
Синюк В.Г.	118	Шалухова М.А.	88
Слиньков А.А.	242	Шахгельдян К.И.	296
Соколов А.М.	131	Щербаков Д.А.	267
Столбов А.Б.	183	Щукин Н.Ю.	267
Сычев О.А.	206	Юрин А.Ю.	183
Тихомирова Д.В.	310	Ярушев С.А.	6
Углев В.А.	255	Ярушкина Н.Г.	218, 277, 351
Фам К.Б.	141	Ясир М.Д.	55
Федулов Я.А.	152		

## СОДЕРЖАНИЕ

### Секция 5. НЕЧЕТКИЕ МОДЕЛИ, МЯГКИЕ ИЗМЕРЕНИЯ И ВЫЧИСЛЕНИЯ

<b>А.Н. Аверкин, Е.Н. Волков, С.А. Ярушев</b> ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ НЕЙРО-НЕЧЁТКИХ СЕТЕЙ ДЛЯ ЗАДАЧИ КЛАССИФИКАЦИИ ОФТАЛЬМОЛОГИЧЕСКИХ ИЗОБРАЖЕНИЙ .....	6
<b>А.В. Бобряков, С.А. Прокопенко</b> МОДЕЛИРОВАНИЕ ПРОЦЕССОВ ПЛАНИРОВАНИЯ НА ОСНОВЕ ТЕМПОРАЛЬНЫХ НЕЙРО-НЕЧЕТКИХ СЕТЕЙ ПЕТРИ .....	19
<b>В.В. Борисов, А.П. Жарков, В.С. Луферов</b> СЦЕНАРНОЕ МОДЕЛИРОВАНИЕ НА ОСНОВЕ НЕЧЕТКИХ РЕЛЯЦИОННЫХ ТЕМПОРАЛЬНЫХ КОГНИТИВНЫХ МОДЕЛЕЙ .....	30
<b>Г.П. Виноградов</b> РАСШИРЕНИЕ ФУНКЦИОНАЛЬНОСТИ УЗЛОВ В БЕСПРОВОДНЫХ СЕНСОРНЫХ СЕТЯХ.....	42
<b>Л.А. Гладков, М.Д. Ясир, Н.В. Гладкова</b> РАЗРАБОТКА И ИССЛЕДОВАНИЕ МЕТОДА РЕШЕНИЯ ЗАДАЧИ РАЗМЕЩЕНИЯ ЭЛЕМЕНТОВ ЭВА НА ОСНОВЕ ГИБРИДНОГО ПОДХОДА .....	55
<b>А.И. Долгий, С.М. Ковалев, А.Е. Колоденкова</b> НЕЧЕТКО-ДИНАМИЧЕСКИЙ МЕТОД АНАЛИЗА ГЛАВНЫХ КОМПОНЕНТ В СИСТЕМАХ ИНТЕЛЛЕКТУАЛЬНОГО МОНИТОРИНГА .....	66
<b>Б.К. Лебедев, О.Б. Лебедев, М.А. Ганжур</b> ОПТИМИЗАЦИЯ МЕТОДОМ РОЯ ТРАНСФОРМИРУЮЩИХСЯ ХРОМОСОМ .....	77
<b>А.Е. Мисник, М.А. Шалухова</b> ПОВЫШЕНИЕ КАЧЕСТВА ПРОЦЕССОВ КОНТРОЛЯ ПРОДУКЦИИ НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ НЕЙРОСЕТЕВЫХ И НЕЙРО-НЕЧЕТКИХ МОДЕЛЕЙ И СРЕДСТВ .....	88
<b>А.Г. Подвесовский, Р.А. Исаев, И.А. Копелиович</b> ПОДХОД К ГЕНЕРАЦИИ ФОРМАЛЬНЫХ НЕЧЕТКИХ КОГНИТИВНЫХ КАРТ ДЛЯ ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ В КОГНИТИВНОМ МОДЕЛИРОВАНИИ.....	97

<b>И.А. Прокопенков, И.А. Фролов, В.И. Меденников</b> МЕТОД ОЦЕНИВАНИЯ СЛОЖНЫХ СИСТЕМ И ПРОЦЕССОВ НА ОСНОВЕ ОНТОЛОГИЧЕСКОГО И НЕЙРО-НЕЧЕТКОГО МОДЕЛИРОВАНИЯ.....	108
<b>В.Г. Синюк, С.А. Каратач</b> МЕТОД ВЫВОДА НЕЧЕТКИХ СИСТЕМ ЛОГИЧЕСКОГО ТИПА ПРИ НЕСИНГЛТОННОЙ ФАЗЗИФИКАЦИИ.....	118
<b>А.М. Соколов, В.И. Будзко, А.А. Зацаринный</b> НЕЧЕТКИЙ СИТУАЦИОННО-ПРЕЦЕДЕНТНЫЙ АНАЛИЗ И МОДЕЛИРОВАНИЕ ПРОЦЕССОВ В СЛОЖНЫХ ТЕХНИЧЕСКИХ СИСТЕМАХ .....	131
<b>К.Б. Фам, П.М. Мурашев, В.Н. Богатиков</b> НЕЧЕТКАЯ МОДЕЛЬ ДИАГНОСТИКИ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ .....	141
<b>Я.А. Федулов, А.С. Федулова</b> ОРГАНИЗАЦИЯ ПАРАЛЛЕЛЬНЫХ ВЫЧИСЛЕНИЙ НА ГИБРИДНЫХ ВЫЧИСЛИТЕЛЬНЫХ КЛАСТЕРАХ НА ОСНОВЕ НЕЧЕТКОГО АНАЛИЗА.....	152
 <b>Секция 6. ИНСТРУМЕНТАЛЬНЫЕ СИСТЕМЫ, ФРЕЙМВОРКИ ДЛЯ СОЗДАНИЯ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ. ТЕХНОЛОГИИ ПРОЕКТИРОВАНИЯ И РАЗРАБОТКИ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ. ПРИКЛАДНЫЕ ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ</b> 	
<b>С.В. Болотов, К.В. Захарченков, С.К. Крутолевич</b> ИНТЕЛЛЕКТУАЛЬНАЯ АППАРТНО-ПРОГРАММНАЯ ПОДДЕРЖКА И ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ СВАРОЧНЫХ ПРОЦЕССОВ.....	162
<b>А.Н. Ветров, А.А. Мальков, Б.В. Палюх</b> ДЕТЕКТИРОВАНИЕ БИФУРКАЦИЙ МНОГОМЕРНЫХ СЛУЧАЙНЫХ ПРОЦЕССОВ В ДИНАМИЧЕСКИХ ЭКСПЕРТНЫХ СИСТЕМАХ .....	173
<b>Н.О. Дородных, О.А. Николайчук, А.Б. Столбов, А.Ю. Юрин</b> СОЗДАНИЕ ВИРТУАЛЬНЫХ ПОМОЩНИКОВ НА ОСНОВЕ ТРАНСФОРМАЦИИ МОДЕЛЕЙ .....	183

<b>А.П. Еремеев, Н.А. Панявин, М.В. Фомина</b> ПРЕДСТАВЛЕНИЕ И ОБРАБОТКА ТЕМПОРАЛЬНЫХ ПРЕЦЕДЕНТОВ В ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ РЕАЛЬНОГО ВРЕМЕНИ .....	194
<b>А.И. Крыгин, М.Р. Гумеров, Н.А. Москаленко, О.А. Сычев</b> ФРЕЙМВОРК ДЛЯ РАЗРАБОТКИ ИНТЕЛЛЕКТУАЛЬНЫХ ОБУЧАЮЩИХ СИСТЕМ НА ОСНОВЕ МОДЕЛЕЙ ПРЕДМЕТНЫХ ОБЛАСТЕЙ В ВИДЕ ДЕРЕВЬЕВ РЕШЕНИЙ .....	206
<b>В.С. Мошкин, М.М. Бажутин, Н.Г. Ярушкина</b> БАЗА ЗНАНИЙ ЭКСПЕРТНОЙ СИСТЕМЫ ДЛЯ АНАЛИЗА SQL-КОДА ПРОМЫШЛЕННЫХ БАЗ ДАННЫХ .....	218
<b>С.Б. Румовская</b> ОТКРЫТЫЕ ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ: ПОНЯТИЕ И ПОДХОДЫ К РАЗРАБОТКЕ .....	230
<b>Г.В. Рыбина, А.А. Слиньков, Д.Д. Белов</b> ИНТЕЛЛЕКТУАЛЬНАЯ ТЕХНОЛОГИЯ ПОСТРОЕНИЯ ДИНАМИЧЕСКИХ ИНТЕГРИРОВАННЫХ ЭКСПЕРТНЫХ СИСТЕМ: ОСОБЕННОСТИ ПОСТРОЕНИЯ ИМИТАЦИОННЫХ МОДЕЛЕЙ ВНЕШНЕЙ СРЕДЫ .....	242
<b>В.А. Углев</b> РЕАЛИЗАЦИЯ МЕХАНИЗМА ПРИНЯТИЯ РЕШЕНИЙ В ИНТЕЛЛЕКТУАЛЬНОЙ АВТОМАТИЗИРОВАННОЙ ОБУЧАЮЩЕЙ СИСТЕМЕ НА БАЗЕ МОДУЛЯ ЭКСПЕРТНЫХ СИСТЕМ .....	255
<b>Н.Ю. Шукин, А.В. Бобряков, А.С. Мохов, Д.А. Щербаков</b> ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА СЛУЖБЫ ПОДДЕРЖКИ ПОЛЬЗОВАТЕЛЕЙ ИНФОРМАЦИОННОГО ПОРТАЛА .....	267

## **Секция 7. ПРИКЛАДНЫЕ БИМЕДИЦИНСКИЕ СИСТЕМЫ**

<b>Е.И. Акимова, Г.Ю. Гуськов, Н.Г. Ярушкина</b> АВТОМАТИЗАЦИЯ АВСДЕ-АНАЛИЗА НОВООБРАЗОВАНИЯ С ИНТЕРПРЕТАЦИЕЙ РЕЗУЛЬТАТОВ .....	277
<b>Ю.А. Бубеев, А.В. Иванов</b> МЕТОДЫ КОСМИЧЕСКОЙ ПСИХОЛОГИИ И ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА. ПЕРСПЕКТИВЫ ИНТЕГРАЦИИ .....	287

<b>Б.И. Гельцер, И.Г. Домжалов, Н.С. Куксин, К.И. Шахгельдян</b> МЕТОДЫ ПРОГНОСТИЧЕСКОЙ АНАЛИТИКИ В ОЦЕНКЕ ВЕРОЯТНОСТИ ВНУТРИГОСПИТАЛЬНОЙ ЛЕТАЛЬНОСТИ У БОЛЬНЫХ ИНФАРКТОМ МИОКАРДА ПОСЛЕ ЧРЕСКОЖНЫХ КОРОНАРНЫХ ВМЕШАТЕЛЬСТВ .....	296
<b>А.А. Долгих, А.В. Самсонович, Д.В. Тихомирова</b> СОЦИАЛЬНО-ЭМОЦИОНАЛЬНЫЕ РАЗГОВОРНЫЕ АГЕНТЫ НА ОСНОВЕ КОГНИТИВНЫХ АРХИТЕКТУР И МАШИННОГО ОБУЧЕНИЯ .....	310
<b>Р.И. Ковалев, В.В. Грибова, Д.Б. Окунь</b> МЕТОД РАЗРАБОТКИ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ КОМПЛЕКСНОГО НАЗНАЧЕНИЯ ПЕРСОНИФИЦИРОВАННОГО ЛЕЧЕНИЯ .....	323
<b>Б.А. Кобринский, Н.А. Благодослов, О.Г. Григорьев</b> ПРИНЦИПЫ ПОСТРОЕНИЯ СИСТЕМЫ КОНТРОЛЯ ПСИХОЭМОЦИОНАЛЬНОГО СОСТОЯНИЯ И МОНИТОРИНГА ЗДОРОВЬЯ ОПЕРАТОРОВ КРИТИЧЕСКОЙ ИНФРАСТРУКТУРЫ .....	333
<b>М.В. Хисамутдинов, Я.С. Коровин, Д.Я. Иванов</b> ТЕХНОЛОГИИ ДИСТАНЦИОННОГО МОНИТОРИНГА ПСИХОЭМОЦИОНАЛЬНОГО СОСТОЯНИЯ ОПЕРАТОРОВ КРИТИЧЕСКОЙ ИНФРАСТРУКТУРЫ .....	342

## ПРИЛОЖЕНИЕ

<b>Н.Г. Ярушкина, В.С. Мошкин, Г.Е. Веселов, Л.А. Гладков</b> ЛЕТНЯЯ ШКОЛА РОССИЙСКОЙ АССОЦИАЦИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ЛШ РАИИ-2023 .....	351
Abstracts of Sectional Reports .....	358
Abstract of Application .....	374
АВТОРСКИЙ УКАЗАТЕЛЬ .....	375

*Научное издание*

Двадцать первая Национальная конференция  
по искусственному интеллекту с международным участием  
КИИ-2023  
Труды конференции в 2-х томах  
Том 2

---

Подписано в печать 03.10.2023 г.  
Формат 60x84<sup>1</sup>/<sub>16</sub>. Тираж 300 экз. Усл. печ. л. 22,1.

Издательство «Принт-Экспресс»



ISBN 978-5-91812-231-0