Российская ассоциация искусственного интеллекта

Российская ассоциация нечетких систем и мягких вычислений

Администрация Коломенского городского округа

Главное управление научно-исследовательской деятельности и технологического сопровождения передовых технологий (инновационных исследований) Министерства обороны Российской Федерации

Департамент информационных систем Министерства обороны Российской Федерации

Московский государственный технический университет им. Н.Э. Баумана

Военный инновационный технополис «ЭРА»

Российский национальный исследовательский медицинский университет им. Н.И. Пирогова

Юбилейная X-я Международная научно-практическая конференция

«ИНТЕГРИРОВАННЫЕ МОДЕЛИ И МЯГКИЕ ВЫЧИСЛЕНИЯ В ИСКУССТВЕННОМ ИНТЕЛЛЕКТЕ»

(17-20 МАЯ 2021 г.)

СБОРНИК НАУЧНЫХ ТРУДОВ

Том 2

УНИВЕРСУМ

2021

УДК 004.8 : 681.3.01
ББК 32.813

Интегрированные модели и мягкие вычисления в искусственном интеллекте. Сборник научных трудов X-й Международной научно-технической конференции (ИММВ-2021, Коломна, 17-20 мая 2021 г.). В 2-х томах. Т2. – Смоленск: Универсум, 2021. – 404 с.

Научно-практическая конференция посвящена вопросам построения интегрированных систем и реализации мягких вычислений в искусственном интеллекте. Основные направления работы конференции:

- Гибридные интеллектуальные системы, синергетический искусственный интеллект (ИИ), объяснимый ИИ.
- Мягкие вычисления, измерения и оценки. Интеллектуальные вычисления. Вычисления, инспирированные природой.
- Неклассические логики и семантики в ИИ. Многозначные и модальные логики в ИИ.
- Нечеткие множества, нечёткие отношения, нечёткие графы и метаграфы.
- Нечёткие логики и приближённые рассуждения.
- Нечёткое управление, нечёткая оптимизации, нечёткие модели принятия решений.
- Когнитивные модели в искусственном интеллекте.
- Вероятностные модели в информационных технологиях и искусственном интеллекте. Байесовские сети.
- Нейросетевые технологии. Нейрологические модели.
- Бионические подходы, эволюционное моделирование, генетические алгоритмы.
- Автономные агенты и многоагентные системы, распределенный и децентрализованный ИИ.
- Интеллектуальные, коллаборативные, «понимающие» роботы.
- Искусственная жизнь и адаптивное поведение. Роевая робототехника.
- Технологии искусственного интеллекта военного, двойного и специального назначения.
- Медицинские интеллектуальные системы.
- Интеллектуальные производства, интернет вещей, интеллектуальные мехатронные и киберфизические системы.
- Инженерия знаний, полученных на основе измерений. Когнитивные и мягкие измерения. Сенсорные сети.

© Универсум, 2021
© РАИИ, 2021
© РАНСМВ, 2021

ISBN 978-5-91412-469-1

Отец Нечёткой Логики
Лотфи Заде

Программный комитет ИММВ-2021

Сопредседатели:

Кобринский Б.А., д.м.н., проф. (Россия, Москва, ФИЦ ИУ РАН)
Тарасов В.Б., к.т.н., доц. (Россия, Москва, МГТУ им. Н.Э. Баумана)

Заместители председателей:

Борисов В.В., д.т.н., проф. (Россия, Смоленск, филиал НИУ МЭИ)
Ковалёв С.М., д.т.н., проф. (Россия, Ростов-на-Дону, РГУПС)

Члены программного комитета:

Аверкин А.Н., к.ф.-м.н., доц. (Россия, Москва, ФИЦ ИУ РАН)
Алиев Р.А., д.т.н., проф. (Азербайджан, Баку, АзГНА)
Батыршин И.З., д.ф.-м.н., проф. (Мексика, Мехико, MPI)
Гладков Л.А., к.т.н., доц. (Россия, Таганрог, ЮФУ)
Голенков В.В., д.т.н., проф. (Беларусь, Минск, БГУИР)
Гончаров А.М., д.в.н., проф. (Россия, Москва, ГУНИД МО РФ)
Грибова В.В., д.т.н., проф. (Россия, Владивосток, ИАПУ ДВО РАН)
Еремеев А.П., д.т.н., проф. (Россия, Москва, НИУ МЭИ)
Зарубина Т.В., д.м.н., проф. (Россия, Москва, РНИМУ им. Н.И. Пирогова)
Карпов В.Э., к.т.н., доц. (Россия, Москва, НИЦ КИ)
Колесников А.В., д.т.н., проф. (Россия, Калининград, БФУ им. И. Канта)
Колоденкова А.Е., д.т.н., доц. (Россия, Самара, СамГТУ)
Кузнецов О.П., д.т.н., проф. (Россия, Москва, ИПУ РАН)
Кулинич А.А., к.т.н., доц. (Россия, Москва, ИПУ РАН)
Курейчик В.М., д.т.н., проф. (Россия, Таганрог, ЮФУ)
Михеенкова М.А., д.т.н., проф. (Россия, Москва, РГГУ)
Морозов А.В., д.т.н., проф. (Россия, Анапа, ВИТ «ЭРА»)
Осадчук А.В., к.т.н. (Россия, Москва, ДИС МО РФ)
Палюх Б.В., д.т.н., проф. (Россия, Тверь, ТвГТУ)
Петровский А.Б., д.т.н., проф. (Россия, Москва, ФИЦ ИУ РАН)
Плесневич Г.С., к.ф.-м.н., доц. (Россия, Москва, НИУ МЭИ)
Редько В.Г., д.ф.-м.н., проф. (Россия, Москва, ФНЦ НИИСИ РАН)
Ройзензон Г.В., к.т.н., доц. (Россия, Москва, ФИЦ ИУ РАН)
Рыбина Г.В., д.т.н., проф. (Россия, Москва, НИЯУ МИФИ)
Стефанюк В.Л., д.т.н., проф. (Россия, Москва, ИППИ РАН)
Тельнов Ю.Ф., д.э.н., проф. (Россия, Москва, РЭУ им. Г.В. Плеханова)
Тулупьев А.Л., д.ф.-м.н., проф. (Россия, Санкт-Петербург, СПбГУ)
Федунов Б.Е., д.т.н., проф. (Россия, Москва, ФГУП ГосНИИАС)
Финн В.К., д.т.н., проф. (Россия, Москва, ФИЦ ИУ РАН)
Фоминых И.Б., д.т.н., проф. (Россия, Москва, НИУ МЭИ)
Юсупбеков Н.Р., акад. АН РУз (Узбекистан, Ташкент, ТашГТУ)
Ющенко А.С., д.т.н., проф. (Россия, Москва, МГТУ им. Н.Э. Баумана)
Язенин А.В., д.ф.-м.н., проф. (Россия, Тверь, ТвГУ)
Ярушкина Н.Г., д.т.н., проф. (Россия, Ульяновск, УлГТУ)

Организационный комитет ИММВ-2021

Сопредседатели:

Лебедев Д.Ю.	(Россия, Коломна, Глава Коломенского городского округа)
Гаврюшин С.С., д.т.н., проф.	(Россия, Москва, МГТУ им. Н.Э. Баумана)

Заместители председателей:

Зафиров Э.Г., к.т.н.	(Россия, Коломна, ОАО «Коломенский завод РТИ»)
Тарасов В.Б., к.т.н., доц.	(Россия, Москва, МГТУ им. Н.Э. Баумана)

Члены организационного комитета:

Ахтямова М.Л.	(Россия, Москва, МГТУ им. Н.Э. Баумана)
Дмитриева И.В.	(Россия, Администрация Коломенского городского округа)
Королева М.Н., к.т.н., доц.	(Россия, Москва, МГТУ им. Н.Э. Баумана)
Мурзак Н.А., к.э.н.	(Россия, Коломна, Коломенский институт (филиал) Московского политехнического университета)
Федотова А.В., к.т.н., доц.	(Россия, Москва, МГТУ им. Н.Э. Баумана)
Колтунов С.С.	(Россия, Москва, МГТУ им. Н.Э. Баумана)
Луферов В.С.	(Россия, Смоленск, филиал НИУ МЭИ)
Луферова Е.А.	(Россия, Смоленск, филиал НИУ МЭИ)
Мажуга В.М.	(Россия, Коломна, ООО «Инлайт»)

СЕКЦИЯ 5 | **ИНТЕЛЛЕКТУАЛЬНЫЕ АГЕНТЫ, РОБОТЫ И КОЛЛЕКТИВЫ РОБОТОВ**

УДК 681.51

ОБУЧЕНИЕ С ПОДКРЕПЛЕНИЕМ В ЗАДАЧЕ УПРАВЛЕНИЯ ИНТЕЛЛЕКТУАЛЬНЫМ РОБОТОМ

***Добрынин Д.А.**, к.т.н.*
e-mail: rabota51@mail.ru
ЦИР РГГУ,
ФИЦ ИУ РАН

1. ВВЕДЕНИЕ

В настоящее время бурно развиваются исследования в области интеллектуальных роботов, что связано с большими перспективами применения таких систем. Одной из важных характеристик интеллектуального робота является адаптация к меняющимся условиям среды, в которой функционирует робот. Система управления робота должна сама настраиваться на эти изменения, что предполагает наличия у нее свойств обучения.

Обучение на заранее классифицированных примерах, известное как обучение с «учителем» имеет серьезные ограничения в тех случаях, когда заранее неизвестно – как учить робота. Желательным является способность робота обучаться самостоятельно. При этом роботу ставится цель, а последовательность действий, необходимых для достижения цели, робот вырабатывает сам. Учителем в данном случае выступает сама среда, в которой функционирует робот, а наказание и

поощрение задается с помощью целевых функций. Обучение с подкреплением или “reinforcement learning” относится к таким методам.

В настоящее время обучение с подкреплением активно развивается для нейронных сетей, особенно для глубинных нейронных сетей [1]. Особенностями такого обучения являются низкая скорость обучения, которую пытаются повысить различными методами, ограничения по сложности моделей мира, сложности в интерпретации результатов, зависимость качества обучения от условий обучения.

Для логических методов искусственного интеллекта обучение с подкреплением развито значительно слабее. Автором был предложен подход к обучению с учителем для динамического ДСМ метода в задачах обучения роботов [2]. Настоящая работа является попыткой развития этого направления в сторону обучения с подкреплением.

2. МОДЕЛЬНАЯ ЗАДАЧА

В качестве модельной задачи для обучения робота был выбрана задача движения робота по черной полосе (рис.1). Эта задача часто встречается на соревнованиях мобильных роботов. Несмотря на внешнюю простоту, она является хорошим примером динамической среды (так как сам робот перемещается) с ограниченным количеством действий. Робот (рис.1) представляет собой трехколесную тележку с дифференциальными колесами и одним опорным колесом. Для обнаружения полосы робот имеет четыре датчика. Для упрощения наблюдения за роботом первый датчик выделен желтым цветом. На рис.1 цифрами обозначены номера датчиков.

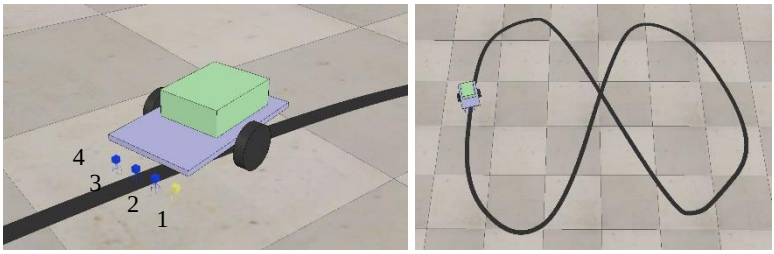


Рис.1. Модель робота и модель полигона

Текущее состояние робота определяется сенсорной информацией – состоянием датчиков полосы, и вектором управления – состоянием приводов робота. В терминах ДСМ-системы [2] текущее состояние представляет собой один пример, который имеет следующее представление:

$$P_i = (S_i, C_i)$$

где S_i – вектор сенсорной информации, который включает в себя всю текущую информацию о состоянии датчиков; C_i – вектор управления, который включает в себя всю текущую информацию о состоянии приводов.

Каждый датчик полосы кодируется двумя битами: {1 0} – не видит полосу, {0 1} – видит полосу. Надчеркивание над номером датчика означает инверсию состояния. Такое представление используется в операции пересечения над битовыми строками. На рис.2 приведен пример возможного представления одного примера.

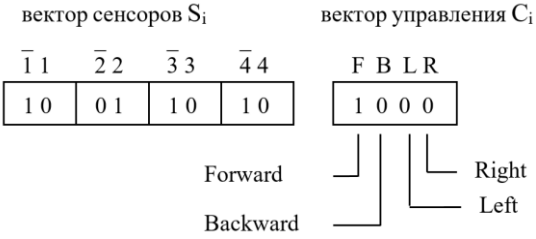


Рис.2. Структура примеров и гипотез

Вектор управления может содержать единицу только в одной позиции, соответствующей определенному движению робота. В данной модели используется четыре заранее определенных движения робота: вперед (Forward), назад (Backward), влево (Left) и вправо (Right). Этих движений достаточно для выполнения задачи – движения по полосе. Для удобства структура ДСМ гипотезы имеет тот же вид, что и пример.

3. ОБУЧЕНИЕ С УЧИТЕЛЕМ В ДИНАМИЧЕСКОМ ДСМ-МЕТОДЕ

Динамический ДСМ метод был предложен в работе [2]. В отличие от классического ДСМ метода [3], динамический ДСМ метод может быть использован в открытой среде с неполной информацией и с неизвестным заранее количеством примеров. В ходе работы динамического ДСМ метода база знаний является открытой и формируется в ходе обучения робота, т.е. по мере поступления новой информации. Это свойство динамического ДСМ дает возможность использовать данный метод для построения систем управления интеллектуальным роботом, способным обучаться.

Кратко рассмотрим режимы обучения с учителем и режим автономной работы с использованием динамического ДСМ–метода [6].

Это позволит нам в дальнейшем перейти к самообучению, которое оказывается похожим на обучение с учителем.

В режиме обучения (рис.3а) «учитель» получает на вход сенсорную информацию и вырабатывает управляющие сигналы, необходимые для адекватного поведения робота на полигоне. Одновременно сенсорная информация и состояние приводов подается в ДСМ-систему. Совокупность сигналов рецепторов и текущих управляющих воздействий определяет один обучающий пример P_i. ДСМ система проверяет его на уникальность и заносит в свою базу фактов. После занесения нового примера в список обучающих примеров производится поиск гипотез. Для данной задачи использовался простой ДСМ метод без запрета на контрпримеры [4], поскольку в данной задаче нет отрицательных обучающих примеров.

Обучение должно производиться до тех пор, пока база знаний не перестанет пополняться новыми гипотезами. Очевидно, что в этом случае обучающий алгоритм перебрал все возможные варианты входных воздействий, на которые он способен реагировать, и можно считать, что база фактов достаточно полна.

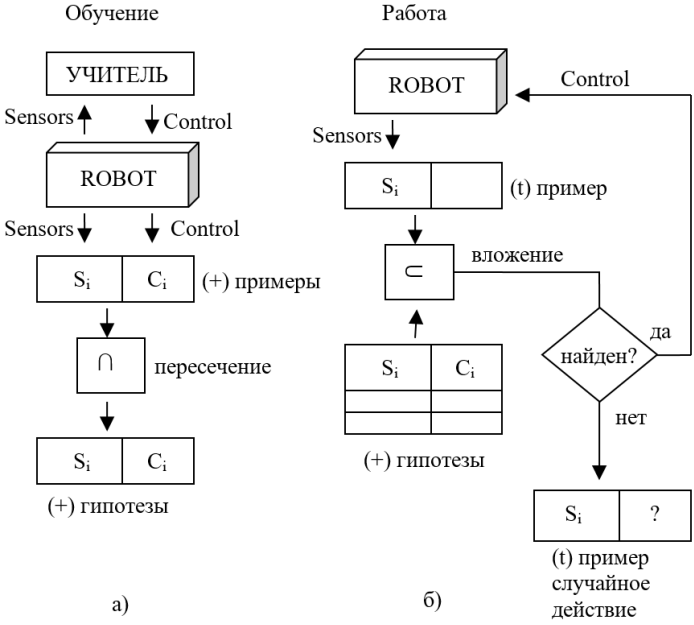


Рис.3. а) Режим обучения; б) рабочий режим

В рабочем режиме (рис.3б) ДСМ система получает на вход сенсорную информацию S_i, из которой формируется тестовый пример. Принятие решения происходит путем проверки вложения ранее полученных гипотез в этот вектор. Если в тестовый вектор сигналов рецепторов вкладывается гипотеза, то робот должен действовать в соответствии с ней. Если же ни одной гипотезы не найдено, то это неизвестное состояние, для которого нужно сформировать случайный вектор управления (рис.3б). Если база фактов полна, то характер поведения робота в рабочем режиме под управлением ДСМ системы ничем не должен отличаться от управления «учителем».

Для проверки проводились различные эксперименты, как на программном симуляторе, так и на минироботе «Амур», созданном в Творческой научно-технической лаборатории Политехнического музея [5]. Эксперименты показали, что таким образом можно обучать робота нужному поведению в реальном времени. Обученный робот внешне ничем не отличался по поведению от управления с помощью «учителя».

4. САМООБУЧЕНИЕ В ДИНАМИЧЕСКОМ ДСМ-МЕТОДЕ

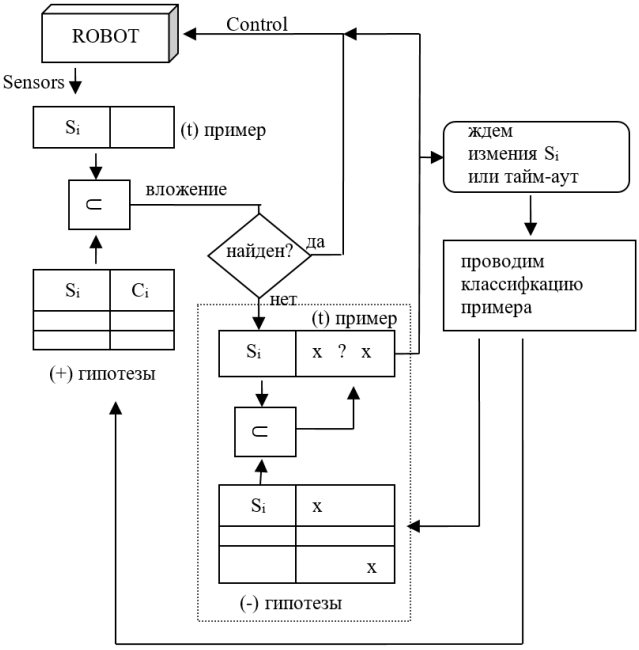


Рис.4. Самообучение ДСМ-системы

Рассмотренный выше способ обучения с учителем ДСМ системы можно назвать пакетным. При таком способе обучение и применение полученных знаний разделены во времени. Для перехода к самообучению необходимо в рабочем режиме каким-то образом классифицировать нераспознанные примеры. Дополнительно следует ввести правила поощрения и наказания за действия робота в среде.

Пусть в начальный момент времени база (+) гипотез робота пуста (рис.4). Тогда первый пример, сформированный из сенсорной информации, не будет распознан и система управления перейдет в режим классификации (t) примера. Для нового неизвестного примера необходимо случайным образом сформировать вектор управления. При этом сначала ищутся отрицательные гипотезы, которые вкладываются в текущий вектор сенсоров. Если такие гипотезы найдены, то соответствующие им действия в поле управления помечаются как недопустимые. Из оставшихся полей вектора управления случайным образом выбирается одно действие. Таким образом, для текущего состояния вектора сенсоров будет случайным образом сгенерировано действие, которое не приводило ранее к неправильным результатам.

Сформированный вектор управления управляет роботом и далее ДСМ-система ждет изменения вектора сенсоров S_i. Одновременно начинает отрабатывать тайм-аут. Когда вектор сенсоров изменится или сработает тайм-аут (кончится отведенное время) ДСМ система переходит к классификации примера.

5. КЛАССИФИКАЦИЯ НОВЫХ СОСТОЯНИЙ

Для определения качества действия необходимо ввести целевую функцию (Target), которая будет выдавать информацию о том, достиг робот поставленной цели или не достиг. Достижение цели определяется по вектору сенсоров. Целевая функция определяется как:

- $T(S_i) = 1$ в случае, если цель достигнута;
- $T(S_i) = 0$ в случае, если цель не достигнута.

Классификация примеров для вектора сенсоров S_i проводится следующим образом:

1) $S_i <> S_i'$, $T(S_i) = 0$, $T(S_i') = 0$, пример $\rightarrow P(-)$;
2) $S_i <> S_i'$, $T(S_i) = 0$, $T(S_i') = 1$, пример $\rightarrow P(+)$;
3) $S_i <> S_i'$, $T(S_i) = 1$, $T(S_i') = 0$, пример $\rightarrow P(+)$;
4) $S_i <> S_i'$, $T(S_i) = 1$, $T(S_i') = 1$, пример $\rightarrow P(+)$;
5) $S_i = S_i'$, timeout = true, пример $\rightarrow P(-)$, robot $\rightarrow$ home.

Рассмотрим более подробно правила классификации.

В правиле 1 робот двигался вне полосы ($T(S_i) = 0$) и полосу не нашел ($T(S_i') = 0$), потому такой пример будет отнесен к отрицательным

примерам (P(-)). В правиле 2 робот двигался вне полосы ($T(S_i)$=0) и полосу нашел ($T(S_i')$ = 1), потому такой пример будет отнесен к положительным примерам (P(+)).

В правиле 3 робот двигался по полосе ($T(S_i)$ = 1) и полосу потерял ($T(S_i')$ = 0), потому такой пример отнесен к положительным примерам (P(+)). Потеря полосы в данном случае не является плохим, поскольку могла произойти от не зависящим от робота обстоятельствам – например, робот двигался прямо по полосе, но встретил поворот.

В правиле 4 робот двигался по полосе ($T(S_i)$ = 1) и полосу не потерял ($T(S_i')$ = 1), потому такой пример отнесен к положительным примерам (P(+)).

Правило 5 работает в случаях, когда робот длительное время совершает действие, которое не приводит к изменению состояния. Тогда робота нужно возвращать в исходное состояние (home), а действие классифицировать как неправильное (P(-)) (приводящее к зависанию).

После классификации примера, т.е. отнесения его к положительным или отрицательным примерам, производится поиск гипотез. Таким образом, базы знаний (+) примеров и (-) примеров пополняются новой информацией. Критерием завершения процесса самообучения робота является отсутствие неизвестных (неклассифицированных) примеров при работе системы управления.

6. РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

Для проверки предлагаемого метода был использован симулятор роботов, в котором проводилось обучение (рис.1). Робот начинал движение вне полосы, попадал на полосу, а затем в процессе самообучения набирал гипотезы.

В качестве целевой функции использовались различные условия нахождения робота на полосе. Приведем несколько примеров.

Пример 1.

Целевая функция $T(S_i)$:

```
if  Photo2  then true
            else false
endif
```

Если фотодатчик 2 видит полосу, то функция выдает значение *true*, иначе выдает *false*.

Можно предположить, что для реализации такой целевой функции достаточно поворачивать налево, если фотодатчик 2 видит полосу. И поворачивать направо, если фотодатчик 2 не видит полосу.

Однако вопреки ожиданиям, в ходе экспериментов были сгенерированы только некоторые из этих гипотез.

По результатам самообучения в нескольких экспериментах формируется различные варианты положительных и отрицательных гипотез (таблица 1). В первом эксперименте формируются только движение направо и налево. Такой набор правил позволяет роботу устойчиво двигаться вдоль полосы. Как видно из таблицы 1, в этом варианте присутствуют действия, для которых целевая функция дает *false* – это гипотеза 1. При этом робот видит полосу датчиком 1, хотя целевая функция требует, чтобы робот видел полосу датчиком 2.

Таблица 1.

	Эксперимент 1		Эксперимент 2		Эксперимент 3	
	11223344	FBLR	11223344	FBLR	11223344	FBLR
(+) примеры	01001010 10010010 00010000	0001 0010 0001	10010010 00100010 00010000 10000101	0010 0001 1000 1000	00010110 00010000 00000010	0010 0001 0001
(-) примеры	10100000 10100001 10100000 10101000	0010 1000 0001 1000	10100001	0001	10100001 10100001 10100001 00101001 01100001	0001 0010 1000 1000 1000
	хорошо		хорошо		плохо	

Во втором эксперименте (таблица 1) набор управляющих правил содержит повороты налево, направо и движение прямо, что позволяет роботу устойчиво двигаться по полосе. Отметим, что даже в этом случае присутствуют гипотезы, для которых целевая функция дает *false*. В третьем эксперименте (таблица 1) набор правил не позволяет роботу устойчиво двигаться по полосе, хотя присутствуют повороты налево и направо. Набор отрицательных гипотез в данном эксперименте самый большой из представленных.

Финальный результат обучения и различия в составе гипотез сильно зависят от начальных условий – точки старта до входа на полосу, действия, которое робот выполнял при этом и состояния датчика случайных чисел, который использовался для выбора случайных вариантов.

Пример 2.

Целевая функция $T(S_i)$:

```
if  Photo1 OR Photo2 OR Photo3 OR Photo4
      then true
      else  false
endif
```

Если любой из фотодатчиков видит полосу, то функция выдает значение *true*, иначе выдает *false*.

Результаты некоторых экспериментов приведены в таблице 2. Как и для первого примера формируются различные наборы гипотез с разным финальным результатом.

Таблица 2.

	Эксперимент 4		Эксперимент 5		Эксперимент 6	
	$\overline{1}1\overline{2}2\overline{3}3\overline{4}4$	FBLR	$\overline{1}1\overline{2}2\overline{3}3\overline{4}4$	FBLR	$\overline{1}1\overline{2}2\overline{3}3\overline{4}4$	FBLR
(+) примеры	10000000	0010	00010000	1000	01000000	0001
	01000000	0001	00100000	0010	00000101	0001
					00100000	0001
					00010000	0010
	хорошо		плохо		хорошо	

В эксперименте 4 (таблица 2) получен практически идеальный результат для управления по одному фотодатчику. В эксперименте 5 (таблица 2) получились плохие правила, которые не позволяют роботу устойчиво двигаться по полосе. В эксперименте 6 (таблица 2) два последних правила являются идеальными правилами для движения робота по полосе, используя фотодатчик 2.

Интересным фактом является отсутствие отрицательных гипотез в этих экспериментах. Можно предположить, что это происходит из-за того, что целевая функция очень «мягкая», поскольку любой датчик может видеть полосу. Поэтому неправильного поведения в результате обучения не возникало, вследствие чего отрицательные примеры и гипотезы не генерировались.

Симуляция показала высокую скорость процесса обучения. Так для всех экспериментов достаточное количество гипотез собиралось при движении робота по одному кругу. Время, затраченное на обучение, составляло менее одной минуты.

7. ЗАКЛЮЧЕНИЕ

Эксперименты на симуляторе показали возможность обучения с подкреплением для динамического ДСМ-метода. Качество обучения сильно зависит от начальных условий обучения, что не всегда приводит к хорошим результатам обучения. Это свойство обучения с подкреплением наблюдается также и для нейронных сетей, что говорит о глубинных свойствах обучения, которые еще недостаточно изучены.

Особенностью обучения с подкрепления для ДСМ-метода является высокая скорость обучения. Практически робота можно учить в реальном времени во время работы на полигоне. Поскольку для получения ДСМ гипотезы достаточно двух похожих примеров, набор базы знаний происходит за несколько похожих случаев.

Предложенный подход к обучению с подкреплением для динамического ДСМ-метода может быть использован для построения эффективно обучаемых систем управления интеллектуальными роботами. Очевидно, что этот подход необходимо развивать, чтобы снизить зависимость качества обучения от начальных условий.

Литература

1. Richard S. Sutton, Andrew G. Barto Reinforcement learning: an introduction. Second edition//MIT Press, Cambridge, MA, 2018.
2. Добрынин Д.А. Динамический ДСМ-метод в задаче управления интеллектуальным роботом // Десятая национальная конференция по искусственному интеллекту КИИ-2006, 25-28 сентября 2006 г., Обнинск, Труды конференции, М: Физматлит 2006, т.2.
3. Финн В.К. Правдоподобные рассуждения в интеллектуальных системах типа ДСМ //Итоги науки и техники. Сер. «Информатика». Т. 15. – М.: ВИНИТИ, 1991.
4. Автоматическое порождение гипотез в интеллектуальных системах/под. общ. ред. В.К.Финна. – М.:Книжный дом «ЛИБРОКОМ», 2009. – 528 с.
5. Добрынин Д.А., Карпов В.Э. Моделирование некоторых простейших форм поведения: от условных рефлексов к индуктивной адаптации //Сб. научных трудов I Международной конференции «Системный анализ и информационные технологии САИТ-2005», М.: КомКнига, 2005, Т.1.
6. Д.А. Добрынин Использование обучаемой системы управления для реализации движения роботов // Завалишинские чтения: молодежная секция: сб. докл. СПб.: ГУАП, 2020, с. 128-134. ISBN 978-5-8088-1503-2

УДК 004.896

МОДЕЛИ КОМАНДНОГО ПОВЕДЕНИЯ РОБОТОВ

***Кулинич А.А.**, к.т.н.*
e-mail: alexkul@rambler.ru
ИПУ РАН

1. ВВЕДЕНИЕ

Тенденции развития групповой робототехники в настоящее время предполагают создание множества простых роботов с реактивной архитектурой. Поведение большого количества реактивных агентов организуется на основе их взаимодействия, построенного, например, на биологических или социальных принципах [1]. Считается, что у стаи или роя агентов за счет самоорганизации появятся эмерджентные свойства, позволяющие такой команде агентов решать сложные задачи. В качестве математической модели поведения стаи или роя роботов часто используют модель поведения стаи птиц, предложенную Рейнольдсом [2] или ее модификаций.

Задачи, которые могут решать стаи роботов – специфические – это движение строем, охрана территории, и др. предполагают действия роботов по алгоритмам, заложенным разработчиком.

Расширить возможности команд роботов возможно применением интеллектуальных агентов с когнитивными возможностями. Архитектура такого агента основана на исследованиях когнитивных функций человека: восприятия; убеждений; предпочтений и др.

Архитектура, позволяющая реализовать свойства интеллектуальных агентов – это ментальная *BDI* (*Belief-Desire-Intention*) архитектура, в которой выделены составляющие: убеждения (*Belief*), характеризующие знания о предметной области; желания (*Desire*), отражающие цели агентов и намерения (*Intention*) - это возможные действия агентов для достижения поставленных целей [3].

Несмотря на то, что в настоящее время создание основных элементов интеллектуальных роботов – это подсистем: представления знаний; языка коммуникаций и обмена знаниями и др. усилиями ассоциации FIPA доведены до уровня стандартов, трудоемкость создания и исследования команд интеллектуальных роботов на различных принципах командного поведения остается высокой.

В этой работе исследуются модели командного поведения реактивных роботов и интеллектуальных роботов с качественной BDI-архитектурой, основанные на моделях социального поведения.

2. СТАИ РЕАКТИВНЫХ РОБОТОВ

В работе [1] были рассмотрены алгоритмы стайного поведения роботов, которые предполагали обмен информации между роботами о собственных целях и ресурсах. В этой работе предложены принципы поведения стихийных команд реактивных роботов, не предполагающих обмен информацией. Поведение роботов сформулировано в форме принципов, основанных на анализе состояния среды функционирования и собственных целей и ресурсов и эвристиках поведения роботов в условиях неопределенности [5].

2.1. Модель среды функционирования реактивных роботов

Мы будем опираться на математическую постановку группового управления в динамической системе «группа роботов-среда» [4]. Рассмотрим множество роботов $A=\{R_i\}$, обладающих свойствами (параметрами) $F=\{f_i\}$. Для каждого свойства каждого робота определено упорядоченное множество их возможных значений, $Z=\{Z_i\}$, где $Z_i=\{z_{i1},\dots z_{iq}\}$, $z_{iq+1} \succ z_{iq}$, $q=0\dots n\text{-}1$. Кроме этого, определено множество объектов $B=\{b_j\}$, каждый из которых имеет свойства из множества Z свойств роботов.

Среда функционирования роботов определяется как прямое произведение множеств значений всех свойств роботов, $SF=\underset{i}{\times} Z_i$.

Вектор значений всех свойств роботов и свойств всех объектов $Y(t) = (Y_1(t),\dots, Y_n(t),\dots, Y_{b1}(t),\dots, Y_{bm}(t))$ определяет состояние среды функционирования, где $Y_1(t)$, $Y_n(t)$ – свойства *i-о* робота; $Y_{b1}(t), Y_{bm}(t)$ – свойства объектов; $Y_i(t)=(z_{1j},\dots, z_{nb}), z_{ij} \in Z_i, \forall i$. Изменение состояния среды функционирования происходит в случаях изменения роботами своих свойств и представляется как отображение:

$$W: Y(t) \rightarrow Y(t+1),$$

где W – система правил поведения роботов, заданных на множестве возможных состояний среды $W: \underset{i}{\times} Z_i \rightarrow \underset{i}{\times} Z_i$; $Y(t)$, $Y(t+1)$ – состояния среды в моменты времени t.

Каждый робот характеризуется следующим кортежем (1):

$$\langle g_q, r_q, \mu_q(Y_q, g_q), O(r_q)\rangle \qquad (1)$$

где:

- $g_q = (z_{1j}{}^g, \dots, z_{nb}{}^g)$ – вектор целевых значений робота q, где $g_q \in SF$;

- $r_q = (z_{1j}^r, \ldots, z_{nb}^r)$ – стратегия достижения цели робота q, где $r_q \in U_q$, $U_q = \underset{i}{\times} Z_i^r$, $Z_i^r \subseteq Z_i$– ресурсы робота q.

Считается, что робот q применяет стратегию r_q для достижения своей цели g_q, предполагая, что другие роботы бездействуют.

1) $\mu_q(Y_q(n), g_q)$ – возможность достижения роботом q целевого состояния g_q за счет собственных ресурсов. Возможность достижения цели определяется как близость прогнозной $Y_q(n)$ и целевой ситуации g_q робота:

$$\mu_q(Y_q(n), g_q) = \rho(Y_q(n), g_q)^{-1}.$$

где $\rho(a, b), a, b \in \underset{i}{\times} Z_i$ - метрика, определенная в пространстве состояний.

По сути, этот показатель определяет потенциальную «силу» каждого робота команды без поддержки других членов команды.

2) O(rq) – полезность целевой ситуации для робота q. Под полезностью целевой ситуации для робота здесь понимается параметр, характеризующий рациональность поведения робота в процессе достижения цели.

Задача заключается в изменении состояния среды функционирования из текущего $Y(t)$ в целевое $Y^*(n)$ путем изменения собственных свойств роботов и свойств объектов. Считается, что зная состояние среды функционирования $Y(t)$, робот, используя ресурсы r_q, пытается достичь своей цели g_q, осознавая при этом возможность $\mu_q(Y_q(n), g_q)$ и полезность $O(r_q)$ достижения цели самостоятельно.

Для задач групповой робототехники [1], в качестве принципов поведения команды роботов выбираются принципы социального поведения насекомых, животных или людей.

2.2. Принципы и эвристики функционирования стихийных команд реактивных роботов

Под стихийной командой реактивных роботов понимается множество роботов имеющих общую цель, возможность наблюдать за состоянием среды функционирования и не имеющих возможности общаться с другими роботами, координировать свои действия и строить общий план.

Под принципом формирования и функционирования команды роботов мы будем понимать поведение каждого робота, основанное на анализе среды функционирования, имеющейся у робота информации о собственном состоянии и некоторых эвристиках [5].

Основываясь на работах социальных психологов, исследовавших вопросы самоорганизации и сплоченности в малых социальных группах, будем рассматривать такие принципы поведения роботов в команде:

1. Принцип самостоятельного достижения цели;
2. Принцип взаимной полезности роботов;
3. Принцип «ленивости» робота;
4. Принцип «эгоистичности» робота.

Рассмотрим названные принципы.

1) Принцип самостоятельного достижения цели. Этот принцип предполагает, что роботы, не зная о существовании других роботов, их целей и возможностей достижения цели, пытаются достичь собственной цели самостоятельно. Все роботы применяют одновременно собственную стратегию r_i с целью достичь целевого состояния g_i. Поскольку цели у роботов могут быть разными то, агрегация их стратегий $\bigoplus_i r_i$ не гарантирует, что каждый робот достигнет цели. Уравнение динамики изменения состояния такой стихийной команды роботов имеет следующий вид:

$$Y^*(t+1)=W^\circ Y(t)\oplus(\bigoplus_i r_i)$$

Для роботов, имеющих близкие цели, т.е., $\forall R_i \in K, K \subseteq A, \rho(g_i, g_q) \le \varepsilon$, $\forall R_i, R_q \in K$, ε – критерий близости целей, достижение их цели возможно, если их агрегированная стратегия доминирует стратегии всех остальных роботов, т.е., если $\bigoplus_{R_i \in K} r_i > \bigoplus_{R_j \notin K} r_j$.

2) Принцип взаимной полезности. В работе [1] был сформулирован принцип полезности роботов: роботы R_i, R_q с близкими целями называются взаимно полезными, если объединение их ресурсов увеличивает возможность достижения общей цели g_Σ.

Реализация этого принципа в работе [1] предполагает обмен информацией между роботами о собственных целях, ресурсах и стратегиях достижения цели. Если обмена информацией нет, то для реализации этого принципа будем использовать эвристику 1.

Эвристика 1. Робот, имеющий большие возможности достижения цели, имеет большую полезность для командной работы. Будем считать, что, возможность достижения цели у робота тем выше, чем он ближе находится к целевому объекту, т.е. к объекту или роботу, параметры которого ($Y_q(t)$) нужно изменить до целевых значений (g_q).

Зададим окрестность близости робота к целевому объекту ε_k, в которой роботы считаются полезными, т.е. $\rho(Y_i(t), Y_q(t))<\varepsilon_k$. Тогда команду полезных роботов, образуют роботы R_i для которых верно

$\forall R_i \in K, K \subseteq A, \rho(Y_i(t), Y_q(t)) < \varepsilon_k$, где $Y_i(t)$ параметры робота, $Y_q(t)$ – параметры целевого объекта. Уравнение динамики стаи следующее:

$$Y_{iq}^*(n) = W^{\circ} Y(t) \bigoplus_{R_i \in K} r_i \bigoplus_{R_j \notin K} r_j.$$

Цель команды роботов K может быть достигнута, если ее агрегированные ресурсы доминируют ресурсы других роботов.

3) Принцип «ленивости» робота. При командной работе на основе принципа взаимной полезности (эвристика 1), в условиях, когда неизвестны ресурсы групп роботов, не входящих в команду K, объединение ресурсов для достижения цели роботов команды K может быть избыточными. Принцип «ленивости» робота позволяет объединять ресурсы роботов в количестве необходимом для достижения цели. Принцип «ленивости» робота основан на контроле им состояния среды в последовательные моменты времени. Если состояние среды $Y^*(t)$ изменяется в направлении цели робота R_i, то он бездействует, $r_i^* = 0$. Эвристика ленивого робота такая.

Эвристика 2. Робот не предпринимает никаких действий, если состояние среды функционирования изменяется в направлении к его цели.

Формально принцип «ленивости» робота выражается в виде проверки условия $\rho(g_i, Y_i(t)) > \rho(g_i, Y_i(t+1))$ каждым роботом, выполнение которого приводит к его бездействию, $r_i^* = 0$.

Процесс образования такой «ленивой» команды происходит на основе контроля состояния среды и не предполагает обмен информацией между роботами команды. В стихийных командах роботов на основе полезности и «ленивости» работу по достижению цели будут выполнять лучшие роботы, в то время как роботы, имеющие меньшие возможности достижения цели будут «простаивать». Этот недостаток можно устранить, если к этим принципам добавить еще и принцип «эгоистичности» роботов.

4) Принцип «эгоистичности» робота основан на том, что за достижение цели роботы получают вознаграждение $O(r_i)$. Все роботы, независимо от их возможностей по достижению цели стремятся достичь ее и получить вознаграждение. В командах, построенных на принципах полезности и «ленивости» вознаграждения будут получать только лучшие роботы. Т.е. $\rho(g, Y_i(t)) > \rho(g, Y_j(t)) \Rightarrow O(r_i) > O(r_j)$. Эвристика «эгоистичного» робота следующая.

Эвристика 3. Роботы, имеющие не лучшие возможности достижения цели, пытаются достичь ее в надежде получить вознаграждение.

Формально команду эгоистичных роботов, образуют роботы R_i, для которых верно $\forall R_i \in K$, $K \subseteq A$, $\rho(Y_i(t), Y_q(t)) > \varepsilon_k$, где $Y_i(t)$ параметры робота, $Y_q(t)$ – параметры целевого объекта.

Принцип «эгоистичности» роботов (эвристика 3) позволяет подключить к команде роботов, не обладающих большими возможностями достижения цели, балансируя тем самым загрузку всей команды роботов K.

Принципы функционирования стихийных команд роботов предполагают, что каждый робот обладает полной информацией о состоянии среды функционирования $Y(t)$. Для выполнения своей задачи достижения цели роботу достаточно знать текущие $Y(t)$ и целевые $Y^*(t)$ значения целевых объектов. Будем считать, что неопределенность возможна как в целевых значениях параметров робота, так в текущих состояниях целевых объектов. Правила поведения робота в условиях неопределенности строится на принципе обобщения неопределенного параметра, который заключается в том, что этот параметр может принимать любое значение из множества его возможных значений. Поведение робота на основе принципа обобщения заключается в поиске по неопределенному параметру объектов, удовлетворяющих шаблону поиска [5].

В работах [1, 5] рассмотрены принципы функционирования стихийных команд роботов, основанные на анализе роботами состояния среды функционирования и не предполагающие обмен информацией роботами о своих целях и ресурсах. Рассмотрены варианты поведения роботов в условиях неопределенности в параметрах целевого объекта. Проводятся анализ возможности образования стихийных команд роботов в условиях неопределенности.

3. РОБОТЫ С КАЧЕСТВЕННОЙ BDI АРХИТЕКТУРОЙ

Задачи, которые могут решать стаи реактивных роботов – это специфические задачи, не предполагающие действий роботов за пределами действий, заложенных в алгоритмы разработчиком.

Расширить возможности групп роботов возможно применением интеллектуальных роботов с когнитивными возможностями, приближенными к возможностям человека. Интеллектуальные роботы имеют ментальную BDI (*Belief-Desire-Intention*) архитектуру [3].

В этой работе предложена архитектура качественного когнитивного робота, позволяющего реализовать и исследовать разнообразные принципы командного поведения множества интеллектуальных роботов с BDI-архитектурой с меньшей трудоемкостью при сохранении их основных свойств.

Упрощение когнитивной архитектуры достигается за счет применения в качестве модели представления знаний агента о среде функционирования качественной онтологии, в виде так называемого концептуального каркаса среды функционирования.

3.1. Архитектура качественного когнитивного робота

Основные элементы архитектуры интеллектуальных агентов были определены в архитектуре InteRRap, предложенной в работе [6]. Это - иерархическая база знаний и связанная с ней компонента управления агентом. Далее при построении архитектуры качественного когнитивного робота мы будем ориентироваться на эту архитектуру. Рассмотрим организацию основных компонент качественного робота: модели представления знаний, планирования поведения, кооперативного взаимодействия агентов и управляющую систему для реализации индивидуального и кооперативного поведения робота.

3.2. Модель представления знаний качественного робота

В модели представления знаний качественного когнитивного робота будем определять количественную и качественную модели динамической системы «группа роботов-среда».

Ранее была уже рассмотрена динамическая система «группа роботов-среда», определенная в работе [4]. Напомним, что среда функционирования роботов определяется прямым произведением множеств значений их свойств, $SF = \underset{ij}{\times} Z_{ij}$, а состояние среды функционирования агентов в момент времени t – это вектор значений свойств всех роботов: $Z(t)=(z_{1e}, \ldots z_{1q}, \ldots, z_{ne}, \ldots z_{nq},)$.

Построение качественной среды функционирования интеллектуальных BDI-роботов основано на интерпретации пространства состояний SF как признакового семантического пространства, в котором состояния среды функционирования $Z(t)$ (векторы значений свойств роботов), определяются именами и векторами значений признаков, определяющие их содержание (смысл). То есть состояния системы «группа роботов-среда» представляются в символьном виде как знаки-символы. В определении немецкого логика Г. Фреге [7] знак - это тройка: имя, смысл и значение знака. Имя – это символ обозначающий объект реального мира, смысл определяет свойства этого объекта, а значение – это и есть сам объект – денотат. Применение знаковых моделей в управлении приведено в работе [8], а модели знаковых картин мира, которые могут использоваться для описания знаний агентов - роботов, рассмотрены в работе [9].

В этой работе в пространстве состояний динамической системы «группа роботов-среда», в семантическом пространстве, выделяются вложенные подпространства, определяющие классы возможных состояний системы «группа роботов-среда» [10]. Эти подпространства структурируются в виде концептуального каркаса, в котором все классы возможных состояний имеют имена d^H и образуют частично упорядоченное множество подпространств $SS(d^H)$ пространства состояний динамической системы FS, т.е. $SS(d^H) \subseteq FS$. Формально качественный концептуальный каркас определен как частично упорядоченное множество имен классов состояний:

$$KK^W=(\{d^H\}, \wedge, \vee),$$

где $\{d^H\}$ – множество имен классов состояний d^H, которые однозначно определяют подпространства среды функционирования - $d^H \Leftrightarrow SS(d^H)$, $SS(d^H) \subseteq FS$, и объем класса состояний $V(d^H)=R_i | R_i=(z_{1e}, \dots z_{1q}) \in SS(d^H)$.

3.3. BDI-архитектура качественного когнитивного робота

В работе [10] в терминах концептуального каркаса KK^W определены элементы архитектуры BDI-робота.

Знания (убеждения) BDI-робота определяются кортежем:

$$\langle BEL_i, W_i^{BEL} \rangle,$$

где BEL_i – убеждения *i-о* агента представляются в виде частично упорядоченного множества имен классов состояний среды функционирования FS, т.е.

$$BEL_i=(\{d^H\}, \leq)\}, \; d^H \Leftrightarrow SS_i(d^H) \subseteq KK^W, \; BEL_i \subseteq KK^W;$$

W_i^{BEL} – это знания агента о закономерностях среды функционирования представляются как отображение:

$$W_i^{BEL}: \underset{i}{\times} Z_{ji} \rightarrow \underset{i}{\times} Z_{ji},$$

где $\underset{i}{\times} Z_i$ - векторы значений признаков агентов, $\underset{i}{\times} Z_{ji} \in SS_{ji}(d^H)$, $d^H \in BEL_j$; W_i^{BEL}-множество правил продукций (Если, То), отражающих закономерности среды функционирования.

Цели агентов (*Desire*) – это вектор $G=(z_{1i}, \dots, z_{1h}; \dots; z_{nk}, \dots, z_{nr})$, $z_{ij} \in Z_i$, элементы которого определяют желаемые значения свойств каждого из роботов в среде функционирования FS и представляются в терминах имен классов среды функционирования:

$$DES_i=(d_1^{Gi}; d_2^{Gi}; \dots; d_n^{Gi}),$$

где $d_j^{Gi} \Leftrightarrow SS_j(d^H) \subseteq KK^W$ – имена классов целевых состояний, $G_j \in SS_j(d^H)$ – вектор цели *j-го* агента ,$\forall j, j=1, \dots, N$.

Действия агентов (*Intention*) в среде функционирования *FS* представлены в виде множества векторов: $U_i(t)=(u_{i1}, \ldots, u_{in}) \in \underset{j}{\times} Z_{ij}^{R}$, а в терминах концептуального каркаса представлены в виде множества имен классов состояний среды функционирования:

$$INT_i=\{d_j^{Ui}\},$$

где d_j^{Ui} – имя класса состояний действия, которое определяет его содержание $SS(d^{Uj}) | U_i \in \underset{j}{\times} Z_{ij}^{R} \in SS(d^{Uj})$.

3.4. Связь количественной и качественной моделей динамической системы «группа роботов-среда»

Концептуальный каркас KK^W можно представлять как гомоморфное отображение среды функционирования *FS*, т.е. $\Psi: FS \rightarrow KK^W$. Т.е. любая точка среды $Z(t) \in FS$ однозначно отображается в класс состояний с именем d^H, при этом обратное отображение класса состояний d^H представлено множеством точек подпространства $SS(d^H) \subseteq FS$. Такие же отображения действительны и для убеждений $\langle BEL_i, W_i^{BEL} \rangle$, целей DES_i и INT_i - действий BDI-архитектуры агентов.

Основная идея качественного представления интеллектуального робота заключается в том, что описание среды функционирования с помощью имен классов состояния позволяет предложить простые логические условия формирования и функционирования команды роботов. Таким образом, в архитектуре качественного интеллектуального робота в рамках качественной модели динамической системы решаются задачи командного взаимодействия роботов, а реализация этого взаимодействия осуществляется в терминах количественной модели системы «группа роботов-среда».

3.5. Подсистема планирования поведения робота

В модели среды функционирования «группа роботов-среда» в виде качественного концептуального каркаса решение задачи планирования поведения робота можно представить, как решение обратной задачи в уравнениях динамики этой системы.

Изменение состояния среды *i-м* агентом, с учетом его знаний W_i^{BEL} представляется системой логико-лингвистических уравнений:

$$W_i^{BEL}:(Z^*(t), INT^*_i)) \rightarrow Z^*(t+1), \forall i,$$

где W_i^{BEL} – знания *i*-го робота (система правил «*Если, То*»), $Z^*(t)=Z(t)$ – начальный вектор состояния среды в терминах имен классов состояний, INT^*_i – управление (действие агента), $Z^*(t+1)$ состояние среды после управления в терминах имен классов состояний [10].

Формально поиск действий для достижения заданного вектора цели каждым агентом сводится к решению им обратной задачи (2):

$$INT_i^* = DES_i \circ W_j^{BEL}. \quad (2)$$

где INT_i^* - действия роботов, позволяющие достичь цели DES_i, $\circ$ – процедура обратного вывода [10].

Действия INT_i^*- это множество имен классов состояния $\{d_j^{Ui}\}$, в которые должен перейти робот, для того, чтобы достичь цели DES_i.

3.6. Модель кооперативного взаимодействия качественного когнитивного робота

Возможны два подхода к планированию кооперативного поведения агентов для достижения общей цели. Первый предполагает построение общего плана до начала решения задачи достижения общей цели, второй – это ситуационная кооперация между роботами, когда решение о совместных действиях принимаются роботами на основе анализа состояния среды функционирования и наличия у роботов условий для кооперации. Выбор действия робота, основанный на решении обратной задачи (1), позволяет выбрать лучшее действие.

В качестве условия для кооперации роботов рассмотрен критерий взаимной полезности агентов, предложенный в работе [1]. По этому критерию в качестве партнера выбирается робот, взаимодействие с которым максимально полезно обоим роботам.

В работе [1, 10] были сформулированы следующие условия кооперации роботов:

- Робот *j* считается привлекательным для кооперации для робота *i*, если элемент его цели $d_i^H \notin BEL_i$ существует в системе убеждений (знаний) робота *j*, т.е. $d_i^H \in BEL_j$.
- Робот *j* считается привлекательным для кооперации для робота *i*, если в его системе убеждений существуют элементы $d_i^H \in BEL_i$, которые является также элементами системы убеждений робота *j*, т.е. $d_i^H \in BEL_j$.
- Робот *j* может изменить свойства агента *i*, если робот *i* включен в объем одного из возможных классов состояний робота *j*, т.е. $a_i \in V(d_j^H) | d_j^H \in BEL_j$.

Первое условие определяет, что у потенциального партнера должна быть возможность (ресурсы) для оказания помощи. Второе, у потенциальных партнеров должны быть пересечения их понятийных систем (убеждений-знаний). Тогда они смогут говорить на одном понятном им языке. Третье условие, в случае если выполняются первые два условия, то робот, попросивший о помощи должен сам создать условия того, чтобы совместные действия были возможны.

Командная работа роботов в данном случае заключается в обмене информацией о своих убеждениях и целях в терминах имен классов состояний, и проверки условий о взаимной полезности, и, следовательно, возможности кооперации.

4. ЗАКЛЮЧЕНИЕ

В работе рассмотрены модели командного поведения реактивных роботов и интеллектуальных роботов с качественной BDI-архитектурой. Для реактивных роботов определены принципы формирования стихийных команд на основе принципов социальной организации людей. Предложена архитектура интеллектуального робота с качественной BDI – архитектурой. Для интеллектуальных роботов предложены модели планирования их поведения и условия кооперации. Предложенные принципы формирования команд реактивных роботов и условия кооперации интеллектуальных роботов были исследованы при имитационном моделировании механизмов образования и функционирования команд роботов.

Литература

1. Карпов В.Э., Карпова И.П., Кулинич А.А. Социальные сообщества роботов. М.: URSS, ООО «ЛЕНАНД». – 2019. – 352 с.
2. Reynolds, C.W. Flocks, Herds, and Schools: A Distributed Behavioural Model // Computer Graphics. – 1987. – Vol. 21, No. 4, – P. 25-34.
3. Rao A.S., Georgeff M.P. BDI Agents: From Theory to Practice // Proc. First International Conference on Multi-Agent Systems (ed. V.Lesser). – AAAI Press/The MIT Press, – 1995. – P. 312-319.
4. Каляев И.А., Гайдук А.Р., Капустян С.Г. Модели и алгоритмы коллективного управления в группах роботов/М.: Физматлит, 2009. – 279 с.
5. Кулинич А. А. Когнитивная архитектура качественного интеллектуального агента. Труды конгресса по интеллектуальным системам и информационным технологиям «AIS-IT' 17». Научное издание в 2-х томах. Таганрог: Из-во ЮФУ. 2017. – Т.1. – С. 235-241.
6. Müller J.P., Pischel M. The Agent Architecture InteRRaP: Concept and Application. 1993. C.99. https://books.google.ru/books?id=gTNFOAAACAAJ.
7. Бирюков Б.В. Теория смысла Готлоба Фреге / В кн.: Применение логики в науке и технике. – М.: Изд-во АН СССР. – 1960. – С. 502-555.
8. Поспелов Д.А., Осипов Г.С. Прикладная семиотика // Новости искусственного интеллекта. – 1999. – № 1. – С. 9-35.
9. Осипов Г. С. Панов А. И. Отношения и операции в знаковой картине мира субъекта поведения// Искусственный интеллект и принятие решений. – 2017. – № 4. – С. 5-22.
10. Кулинич А.А. Модель командного поведения агентов в качественной семиотической среде. Часть 2. Модели и алгоритмы формирования и функционирования команд агентов. Искусственный Интеллект и Принятие Решений. – 2018. – № 1. – С. 29-40.

УДК 004.8;17

ПРИМЕНЕНИЕ МОРАЛЬНЫХ НОРМ ДЛЯ УПРАВЛЕНИЯ ПОВЕДЕНИЕМ ИСКУССТВЕННЫХ АГЕНТОВ

***Карпов В.Э.**, к.т.н.*
e-mail: karpov_ve@mail.ru
Сорокоумов П.С.
e-mail: petr.sorokoumov@gmail.com
НИЦ «Курчатовский институт»

1. ВВЕДЕНИЕ

Исследование социального поведения агентов является одним из важнейших направлений искусственного интеллекта, потому что, как известно из примеров естественных социумов, оно позволяет добиться синергетического эффекта – повышения возможностей группы по сравнению с суммой индивидуальных возможностей её участников. С технической точки зрения образование социальных сообществ является одним из эффективных методов адаптации групп агентов. Важное место в таких исследованиях занимают модели, в которых целенаправленное поведение всего социума определяется моделями индивидуального поведения и локального взаимодействия членов группы.

Применения таких моделей могут быть весьма разнообразны, но в основном разделяются по решаемой задаче на две группы: создание устойчивых социумов из индивидов и разрушение уже существующих социумов. К задачам первой группы относится, например, моделирование естественных эусоциальных сообществ с переносом полученных моделей на группы роботов [1, 2]; к задачам второй – моделирование сообществ бактерий (биоплёнок) с целью их разрушения [3]. Устойчивость социума в данном случае эквивалентна способности к поддержанию гомеостаза – сохранению характера функционирования даже при изменениях деталей поведения.

Таким образом, очевидно, что моделирование социума требует применения адекватных моделей индивидуального поведения особей, в которых используются методы поддержания гомеостаза. Одним из направлений исследований группового взаимодействия является т.н. парадигма моделей социального поведения в группах агентов или роботов. В рамках этой парадигмы решаются задачи создания моделей

индивидов, способных к социальному (межагентному) взаимодействию, строятся модели механизмов, определяющих условия формирования устойчивых социумов, исследуются вопросы целенаправленного управления характером поведения социумов искусственных агентов. При этом в рамках задачи управления поведения социума особый интерес представляет исследование регуляторного механизма, который присущ высокоразвитым социумам и называется моралью. В рамках концепции моделей социального поведения мораль рассматривается в ее непосредственном смысле, как некий регуляторный механизм, определяющий способы разрешения конфликтов внутри группы (социума).

В предлагаемой работе рассматривается вопрос применимости эмоционально-потребностной системы управления индивида-агента к решению задачи организации взаимодействия агентов и разрешения конфликтных ситуаций на основе механизмов, которые принято относить к моральным регулятивам.

2. ЭМОЦИОНАЛЬНО-ПОТРЕБНОСТНАЯ АРХИТЕКТУРА

Индивидуальное поведение агентов (и живых, и искусственных) определяется особенностями строения их базового, «физиологического» уровня, который оперирует такими сущностями, как стимулы, потребности и эмоции. Одной из наиболее конструктивных моделей является потребностно-информационноя теория эмоций П. В. Симонова [4]. Согласно ей, интегральная оценка ситуации определяется оценкой баланса между необходимыми и имеющимися средствами удовлетворения актуальных потребностей. Предлагаемое Симоновым соотношение можно записать как:

$$E=f(N, p(I_{need}, I_{has})) \quad (1)$$

где E – оценка эмоции (и её величины, и знака); N – сила текущей необходимости в удовлетворении потребности; $p(I_{need}, I_{has})$ – оценка возможности удовлетворить потребность на базе врожденного и полученного жизненного опыта; I_{need} – информация о способе удовлетворения потребности; I_{has} – информация об имеющихся у агента средствах, требуемых для удовлетворения актуальных потребностей. Очевидно, что подобный подход позволяет рассчитать эмоциональную реакцию агента по оценкам потребностей и средств их удовлетворения. Формализация этого качественного, оценочного соотношения для определения эмоций выглядит следующим образом.

Пусть поведение агента определяется тройкой:

$$Agent = <S, N, R> \quad (2)$$

где R – множество продукционных правил, N – множество потребностей, S – множество сенсоров. Представим правила R в MYCIN-подобной форме [5], т.е. как множество продукций с коэффициентами уверенности:

$$R_i: (Cond_1 \ \& \ \dots \ Cond_k) \rightarrow f \mid w_i$$

Здесь R_i – правило, $\mathrm{Cond_i}$ – коньюнкты условия, $Cond_i \in [0,1]$; f – номер выполняемого действия, $f = 1..M$; w_i – коэффициент уверенности (к.у.) правила, $w_i \in [0,1]$, M – количество действий (поведенческих процедур), которое выполняет агент. Каждая продукция оценивает уверенность в том, что агент совершит указанное в ней действие f. Коэффициент уверенности заключения вычисляется по обычным правилам нечёткой логики; в разработанной модели оказалось удобно выбирать эти правила так, чтобы все к.у. попадали в диапазон [0;1]. Кроме того, укажем для каждой продукции потребность $Need(R_i)$, на удовлетворение которой направлено действие.

Действия агента могут быть сложными процедурами, реализующими целые комплексы действий (в биологии – фиксированные комплексы действий, ФКД) или даже поведенческими последовательностями. Иначе говоря, предполагается, что речь идет о высокоуровневой архитектуре, оперирующей именно поведением агента.

После вычисления всех к.у. заключений агент будет способен выбрать, какую из процедур запускать. Таким образом, множество продукций R порождает вектор к.у. выполняемых действий A. Если два разных правила требуют совершать одно и то же действие, их к.у. A_1 и A_2 сочетаются по правилу:

$$A_{sum} = A_1 + A_2 - A_1 A_2$$

После выбора действия можно по формуле (1) оценить эмоции, вызываемые в агенте после такого выбора. При этом I_{need} определяется как вектор коэффициентов уверенностей заключений действий ($I_{need} = A$), а I_{has} определяется как вектор актуально исполняемых в данный момент времени действий той же размерности, что и A, в котором для реально исполняемых действий указано значение 1, для остальных – 0. Функцию f в формуле (1), сочетающую силу потребности и возможности по её удовлетворению, будем для простоты считать произведением. Таким образом, вычисляя к.у. A_i для всех действий в текущий момент, мы можем определить их эмоциональные оценки E_i, составляющие вектор эмоций E:

$$E = k\,(I_{has} - I_{need}),\ i = 1 \dots M \qquad (3)$$

где k – постоянный коэффициент. Видно, что если какое-то действие выполнить не получается (т.к. в каждый момент времени выполняется

только одной действие), то у агента возникает отрицательная эмоция. Сумма эмоций, связанных со всеми действиями, позволяет оценить общее эмоциональное состояние агента.

Допустим далее, что эмоциональное состояние агента влияет на его восприятия, но не непосредственно на сенсоры, а через промежуточные элементы – шлюзы *G*. В результате агент воспринимает от каждого сенсора не непосредственно показания *S*, а некоторое эмоционально окрашенное восприятие: $G_i = G_i(S_i, E)$. В этой модели все функции шлюзов G_i определялись как линейные комбинации S_i и компонентов *E*.

Видно, что в конечном счёте эмоции образуют обратную связь от выбранных действий на вход механизма выбора – к условиям правил *Cond*, использующим не непосредственные (*S*), а эмоционально окрашенные показания сенсоров (*G*). В итоге функционирование эмоционально-потребностной схемы поведения агента выглядит как бесконечный циклический процесс:

1. Формирование вектора эмоционально окрашенных восприятий *G*.
2. Вычисление вектора коэффициентов уверенности заключения правил *A*.
3. Выбор действия для выполнения.
4. Формирование вектора эмоций *E*.

Схематически данная схема показана на рис.1.

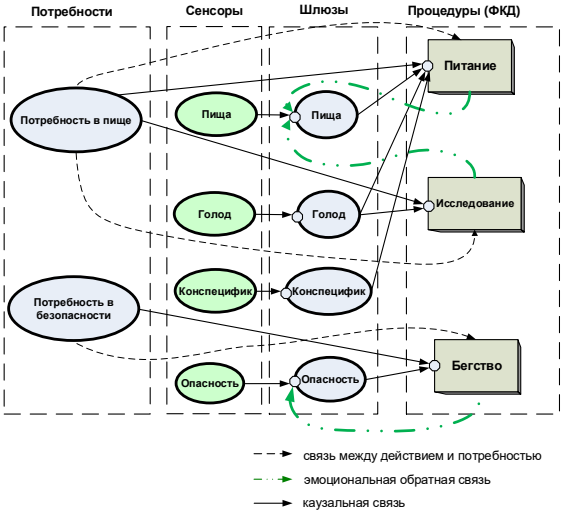


Рис.1. Эмоционально-потребностная архитектура агента

3. МОРАЛЬНЫЕ НОРМЫ КАК СИСТЕМА ПРАВИЛ

Мораль, если рассматривать её как свод правил поведения верхнего уровня, является одним из механизмов адаптации. Мораль – это наиболее гибкая и вариативная надстройка системы управления. Нормы морали могут меняться многократно в течение жизненного цикла индивида. Управление потребностями индивида или характером его поведения – это достаточно грубые способы вмешательства, чреватые нарушениями работоспособности. Гораздо удобнее оперировать надстройкой, отвечающей за характер внутрисоциальных отношений, т.е. тем, что мы называем в человеческом обществе моралью.

Важнейшей особенностью морали как адаптационного и регуляционного механизма является ее гибкость, вариативность. Эта надстройка над базовыми моделями поведения чрезвычайно «легковесна» и может варьироваться в широких пределах на протяжении жизненного цикла индивида.

Считается, что в основе морального поведения лежит эмпатия – способность к отзывчивости на эмоциональное состояние конспецифика (другого, представителя того же вида). Частным случаем или проявлением эмпатии является т.н. симпатия, т.е. сочувствие, взаимовлечение и т.п. положительные реакции. В основе симпатии-эмпатии лежат три базовых механизма: (1) способность отождествлять конспецифика с собой (компонент «Я» знаковой картины мира, если агент обладает таковой), (2) наличие эмоционально-потребностной системы управления и (3) наличие сигнальной коммуникации.

Механизмы (2) и (3) служат для того чтобы агент мог генерировать некие воспринимаемые другими сигналы, соответствующими его некоторому эмоциональному состоянию. В рамках эмоционально-потребностной архитектуры это означает генерацию коммуникационных сигналов, связанных с максимально активными в данный момент времени шлюзами (рис.1). Именно такой механизм работает в животном мире – сигнальная коммуникация отображает эмоциональное состояние животного.

Механизм сопоставления основан на том, что система управления агента оперирует знаковыми структурами, и в них работает механизм внутренней активизации компонентов знака. В данном случае, наблюдение за конспецификом приводит, если агент воспринимает того как своего (знакомого, себя самого) определяет возбуждение перцепта знака Я, что влечет ассоциативное возбуждение вершин «значение» и «смысл» знака Я. В свою очередь, это приводит к активизации соответствующих процедур, иных элементов знаковой

сети и т.п. В рассмотренной выше архитектуре речь идет о возбуждении шлюзовых элементов, т.е. возникновению ситуации, в которой отсутствие явно наблюдаемого стимула (сенсор *S*) все равно формирует отклик (возбуждение шлюза G и соответствующей ему процедуры).

Итак, введем некий метапараметр, который назовем склонностью к симпатии *S*, $S \in [-1, 1]$. Этот параметр определяет восприятие других агентов с точки зрения «свой-чужой». Агент, у которого $S>0$, отождествляет наблюдаемого конспецифика с собой с коэффициентом уверенности S; механизм подобного отождествления описан выше. В результате такого отождествления агент способен делиться пищей с этим конспецификом. Агент с отрицательной симпатией не отождествляет себя с другими агентами, рассматривая их как источник ресурсов (пищу) с коэффициентом уверенности $|S|$. При нулевой симпатии агент безразличен (толерантен) к другим агентам. При этом если агент видит чужого, ресурс которого больше, чем у него самого, тогда этот чужой воспринимается как опасность.

Интерпретацию влияния "моральных норм" на описанную ранее эмоционально-потребностную схему выполним через шлюзы. Пусть они не только окрашивают сенсорные восприятия эмоционально, но и модифицируют их в соответствие с некоторым набором нечётких правил морали, например:

ЕСЛИ наблюдаю_конспецифика И S<0, ТО возбудить шлюз «Опасность»

ЕСЛИ наблюдаю_конспецифика И S>0, ТО возбудить шлюз «Друг»

4. МОДЕЛЬНЫЕ ЭКСПЕРИМЕНТЫ

Для оценки работоспособности предложенной эмоционально-потребностной архитектуры и механизмов реализации эмпатии к реализации феноменов морального поведения была рассмотрена задача по моделированию пищевого поведения сообщества агентов – аниматов. Мир, в котором моделируется социум, представляет собой клеточное поле, в котором находятся агенты и частицы пищи. В каждой клетке могут находиться частица пищи и произвольное количество агентов. Для исключения краевых эффектов поле имеет тороидальную топологию, то есть верхняя и нижняя, а также левая и правая стороны попарно соединены между собой.

Моделирование проводится пошагово. На каждом шаге агент выполняет одну из допустимых элементарных операций: передвижение на одну из восьми соседних клеток, поедание пищи либо кормление другого агента. Поедание пополняет количество ресурсов у агента, выражающееся величиной $R \in [0, 1]$. В качестве еды агент может воспринимать как частицы

пищи, так и других агентов; при этом допустимо только поедание более слабого агента более сильным (т.е. имеющим большее значение ресурса *R*).

Сенсоры анимата определяют восприятие внешнего мира, а также внутреннего состояния и задаются числами в диапазоне [0, 1]: наличие пищи, оценка чувства голода (1-*R*), наличие дружественного агента и наличие враждебного агента.

У аниматов имеются две потребности: в пище и в безопасности – постоянные величины в диапазоне [0, 1], характеризующие значимость соответствующих факторов. Агенту доступны три действия: питание, бегство и исследование. За каждое перемещение агент тратит фиксированное количество ресурса *R*; при нулевом значении R агент не может совершать действия. Альтруистичные агенты способны кормить других агентов, при этом их запасы пищи делятся пополам. Погибнуть агенты не способны.

Описанная среда была реализована в виде приложения на языке *Python*. Пример вида среды с агентами и пищевыми частицами в его интерфейсе показан на рис.2. В начале работы агенты распределяются по миру случайно. Запасы пищевых частиц на карте пополняются после каждого шага моделирования: в каждой клетке, соседней по отношению к съедобной частице, может появиться новая частица с некоторой заданной вероятностью. Эта вероятность при фиксированном начальном количестве еды на карте определяет т.н. ресурсное богатство среды обитания.

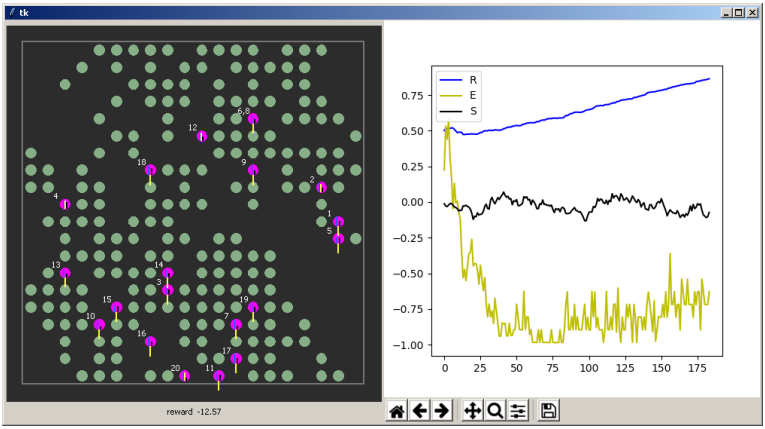


Рис.2. Внешний вид модели в разработанной среде моделирования

Зелёными кругами обозначены частицы пищи, пронумерованными розовыми кругами – агенты. Параметры *R* (количество ресурса), *E* (эмоциональный уровень), *S* (уровень

симпатии) показаны на каждом агенте столбиками. В правой части приведены графики зависимости *R, E, S* от шага моделирования.

Для оценки влияния эмпатии *S* на эффективность социума проведена серия экспериментов с разработанной моделью. Группы агентов при этом имели одинаковую численность (50 участников), но разную долю альтруистичных агентов (от 0 до 100%). При этом у всех эгоистичных агентов склонность к симпатии *S*=-0,75, а у всех альтруистов *S*=0,75. Средние значения *R* и *E* на последнем шаге моделирования изображены на рис.3.

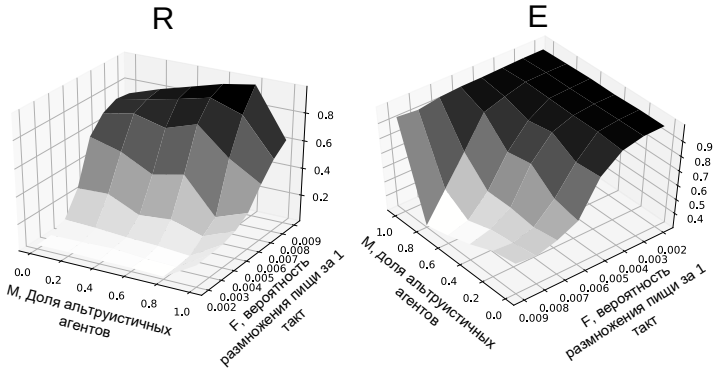


Рис.3. Графики зависимостей успешности группы от условий существования *F* и доли альтруистов *M*. Направления изменения *F* на графиках противоположны для более явной демонстрации характера зависимости

При возрастании продуктивности среды и постоянной симпатии *S* средние значения *R*, как и следовало ожидать, монотонно растут. Однако зависимость *R* от доли альтруистичных агентов оказалась немонотонна; в частности, для самых продуктивных сред (вероятность размножения пищи 0.009) наибольшее среднее значение *R* достигалось при 80% альтруистичных агентов. Полностью альтруистичный коллектив оказался в целом менее продуктивен в богатых средах, чем коллективы с агрессивными агентами, но в бедных средах его продуктивность выше. Наиболее вероятной причиной такой зависимости является постоянная раздача ресурсов агентами друг другу вместо сбора новых ресурсов при полном отсутствии хищников; это понижает как потери ресурса при хищничестве, так и интенсивность их сбора.

Анализ значений *E* в этих же результатах показывает, что в целом при отсутствии ресурсов эмоциональный уровень значительно выше, чем при их наличии. Объясняется это наличием в такой ситуации только одного варианта действий – поиска пищи; если же агенту приходится выбирать из множества вариантов, часть из них остаётся неиспользованной, и это понижает *E*.

Изменение моральных норм со временем можно промоделировать путём модификации *S* по какому-либо закону. Например, можно исследовать эффективность социума при изменении моральных норм на более или менее альтруистичные (рис.4).

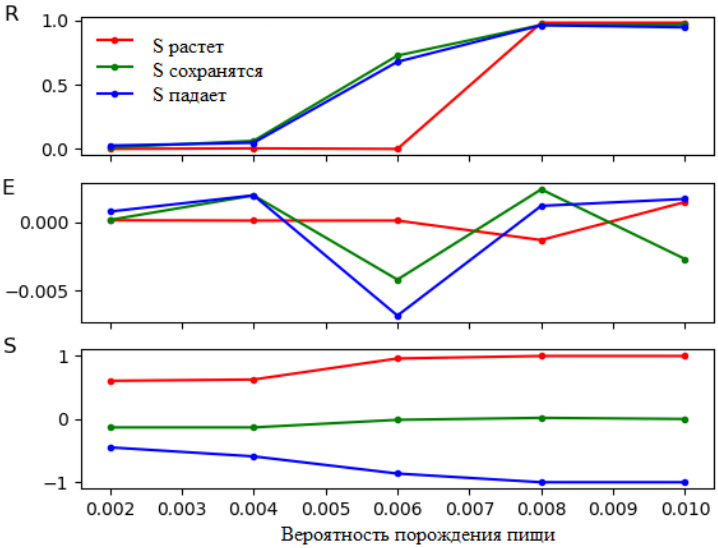


Рис.4. Зависимость параметров эффективности социума от направления изменения моральных норм

Видно, что рост альтруизма в среде средней продуктивности приводит к падению *R* (среднего количества ресурса по группе на последнем шаге моделирования), при этом подобный рост не сопровождается значительными изменениями эмоционального уровня. Из этого можно сделать вывод, что модель воспроизводит наблюдаемое в природе падение жизнеспособности социума при резком снижении уровня опасности для индивидов (и, как следствие, повышении альтруистичности).

5. ЗАКЛЮЧЕНИЕ

Отметим, то в работе не ставились задачи исследования влияния количества эгоистов/альтруистов на эффективность социума, того, как зависит это соотношение от условий среды и т.п. Таких работ крайне много и социологии, и в психологии. Цель была лишь в демонстрации работоспособности некоторых моделей индивидуальной организации и механизмов межагентного взаимодействия при решении задач управления характером поведения социума на самом верхнем уровне, на котором аспекты взаимодействия обычно рассматриваются и оцениваются с точки рения морали.

Разработанная эмоционально-потребностная архитектура позволила воспроизвести в модели некоторые характерные особенности поведения реальных социумов, показав потенциал включения формализованных моральных норм в архитектуры агентов. Перспективной выглядит экспериментальная проверка данной архитектуры на реальных группах роботов, выполняющих практически важные задачи.

Литература

1. Бургов Е.В., Малышев А.А. Качественные и количественные характеристики биониспирированных моделей групповой робототехники // V Всероссийский научно-практический семинар Беспилотные транспортные средства с элементами искусственного интеллекта. – 2019. – P. 139-148.
2. Воробьев В.В., Ровбо М.А. Применение передачи обучения в семиотических моделях к проблеме фуражирования с реальными роботами // Международный журнал “Программные продукты и системы.” – 2020. Vol. 21. P. 413-419.
3. Koshy-Chenthittayil S. et al. Agent Based Models of Polymicrobial Biofilms and the Microbiome-A Review. // Microorganisms. – 2021. Vol. 9. – № 2.
4. Симонов П.В. Потребностно-информационная теория эмоций // Вопросы психологии. – 1982. Vol. 6. – P. 44-56.
5. Джексон П. Введение в экспертные системы. М.: Вильямс, 2001. 624 p.

УДК 004.5

ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА УПРАВЛЕНИЯ ДЛЯ ТАКТИЛЬНОГО МУЛЬТИМОДАЛЬНОГО ПОЛЬЗОВАТЕЛЬСКОГО ИНТЕРФЕЙСА*

Голубурдо Е.И.
e-mail: vixryawa@yandex.ru
Леушина В.В.
e-mail: wandbpand@gmail.com
Сорокоумов П.С.
e-mail: petr.sorokoumov@gmail.com
НИЦ «Курчатовский институт»

1. ВВЕДЕНИЕ

Для передачи информации пользователю сложной системы чаще всего используются видео- и аудиоинтерфейсы, однако сегодня все более явно прослеживается тенденция к применению и других каналов связи и способов воздействия на пользователя. Можно, например, посылать пользователю тепловые, вибрационные сигналы, надавливать на участки кожи и даже пропускать через тело электрический ток – как для управления мышцами, так и для возбуждения болевого ощущения. Эти типы интерфейсов способны снизить нагрузку на зрение и слух пользователя и сделать работу более естественной, «погружая» оператора в контур управления. В целом очевидно, что осязательные воздействия могут облегчить пользователю работу со сложной системой, повысить его производительность, ускорить обучение.

Однако внедрение тактильных интерфейсов в практику осложняется тем, что в каждом конкретном случае их использование при некорректной настройке может ухудшить условия работы пользователя: сигналы могут восприниматься как слишком внезапные, чужеродные, отвлекающие, субъективно лишние и неинформативные. При этом мировой опыт использования этих систем, как будет показано ниже, не позволяет сформулировать конструктивные рекомендации по практическому использованию, не говоря о сочетаемости друг с другом и традиционными интерфейсами. Поэтому может оказаться полезным

* Работа выполнена при финансовой поддержке НИЦ «Курчатовский институт» (приказ № 1055 от 02.07.2020)

использовать гибкую систему для их настройки, которая позволяет задавать требуемые реакции каждого из интерфейсов на события в управляемой системе и предсказуемо изменять их.

В данной работе описана система управления набором тактильных интерфейсов (вибрационным, температурным, компрессионным и миоэлектрическим), основанная на использовании системы нечётких правил. Показано, как такой подход позволяет сочетать входные данные разной природы и формировать управляющие сигналы для контроллеров всех типов устройств по единому протоколу. Описана также экспериментальная проверка работоспособности предложенного подхода на примере задачи управления моделью мобильного робота.

2. РАСПРОСТРАНЁННЫЕ ПОДХОДЫ К ПРИМЕНЕНИЮ ТАКТИЛЬНЫХ ИНТЕРФЕЙСОВ

Наряду с модернизацией технологий, позволяющих считывать информацию об окружающей среде, в последнее время развиваются и новые типы оповещающих интерфейсов. Например, с появлением виртуальной реальности в качестве визуального способа взаимодействия с оператором стали применять VR-шлем. Однако для некоторых задач это либо слишком ресурсозатратно, либо просто ограничено с точки зрения широты репертуара воздействия.

Тактильные ощущения, вызываемые исследуемыми в данной работе интерфейсами, достаточно разнообразны: вибрация, давление, боль, температура и т. д. Это достаточно сильный инструмент, при помощи которого можно передавать сигналы как от устройств к людям, так и в обратном направлении. Многие современные системы имитируют только один определенный вид тактильных ощущений. В повседневной жизни наиболее часто встречается вибротактильная обратная связь: она присутствует во всех современных устройствах (смартфонах, умных часах, джойстиках и т. д.), которыми мы постоянно пользуемся.

В 2008 году Линетт Джонс и Надин Сартер описали основные принципы реализации вибротактильных интерфейсов [1]. Воздействие вибрацией на человека достигается путем закрепления устройства с виброактуаторами на участках его тела и использовании широтно-импульсной модуляцией (ШИМ), позволяющей свободно регулировать силу ее воздействия. На основе других работ они изучили четыре главных параметра (частота, интенсивность, продолжительность и месторасположение) и оценили их влияние на характер работы вибротактильной стимуляции, что позволило им сделать некоторые

немаловажные выводы, которые необходимо учитывать при проектировании таких систем. Во-первых, человек довольно плохо распознает незначительные частотные изменения. Во-вторых, вибросигналы одинаковой частоты, воздействующие на разные участки человеческого тела, могут ощущаться по-разному.

По сравнению с тактильной обратной связью электростимуляция не получила пока широкого распространения. В основном ее активно применяют в медицинской области: она используется в качестве тренировочного и терапевтического приспособления. В сфере медицинской науки существует огромное количество статей, описывающих эффект различных токов на те или иные участки тела; целью таких воздействий может быть либо восстановление утраченных возможностей организма (например, при атрофии мышц), либо тренировка мышечных волокон (применяется спортсменами в дополнение к обычным упражнениям). В качестве интерфейса электрическое воздействие применяют весьма ограниченно; например, в 2015 году было создано устройство «Affordance++» [2], способное заставить человека при определенных обстоятельствах выполнять конкретные действия: например, если пользователь взял в руку баллончик, то этой рукой необходимо потрясти; или, если пользователь тянется к горячей кружке, то необходимо проследить, что пользователь возьмет ее только за ручку.

Термальные (использующие нагревание) и компрессионные (использующие сдавливание) интерфейсы исследованы намного меньше, хотя и существуют, например, проекты перчаток, использующих нагреваемый воздух для сдавливания и нагрева рук для имитации восприятия поверхностей [3]. Люди относительно плохо определяют конкретное место теплового раздражителя, что объясняется малой плотностью терморецепторов на коже: плотность холодовых рецепторов 1-5 рец/см^2, и они расположены в эпидермисе; плотность тепловых терморецепторов около 0,4 рец/см^2, и расположены они глубоко в коже, в дерме. Наиболее чувствительные к воздействию теплом части тела – это лицо, шея, уши, запястье и область ладони, прилегающая к большому пальцу. Кожные сенсоры давления располагаются ещё реже.

Разработка термальных и компрессионных интерфейсов осложняется недостатком научных знаний о восприятии таких воздействий людьми [4], а также высокими требованиями к мощности нагревателя либо компрессора, чтобы реализовывать перемены температуры и давления быстро и реалистично.

Таким образом, общепринятые стандартные подходы к применению подобных интерфейсов ещё не выработаны. Из-за этого необходимо при экспериментальных исследованиях тактильных систем применять максимально гибкие и интуитивно понятные непрофессионалу средства настройки. Крайне желательно ещё и автоматически учитывать эргономические ограничения, накладываемые нормативными документами [5]. Всем этим требованиям удовлетворяет система нечётких правил, описанная ниже.

3. ФОРМИРОВАНИЕ СИГНАЛОВ ПО НЕЧЁТКИМ ПРАВИЛАМ

В качестве тестовой задачи для системы тактильных интерфейсов было выбрано управление моделью мобильного робота, оснащённого дальномерами. Пользователь должен, пользуясь данными дальномера, направить модель к обозначенной цели, избегая препятствий. При этом предполагалось, что дополняющие визуальное восприятие модели тактильные воздействия могут помочь или помешать ему действовать эффективно.

Исходя из требований к гибкости настройки системы было решено формировать сигналы для всех видов интерфейсов на основе единого набора нечётких правил. Каждое правило позволяет оценить истинность выходных лингвистических переменных, зная истинность входных лингвистических переменных. Множество входных лингвистических переменных формируется по описанию управляемой оператором системы; в рассматриваемом случае в это описание входили следующие элементы:

- дальность препятствия для каждого дальномера;
- дальность цели;
- направление на цель;
- предсказываемое с применением фильтра Калмана столкновение модели с препятствием через заданный промежуток времени (логическое значение; истинно, если модель предсказывает столкновение с известными препятствиями).

По каждому из перечисленных параметров оценивается достоверность одной входной лингвистической переменной. На рис.1 проиллюстрировано такое вычисление для лингвистической переменной *k* «Препятствие близко к роботу по данным дальномера». Набор входных переменных может быть дополнен при настройке системы благодаря использованному интерфейсному решению: набор параметров вычисляется отдельными программными компонентами и может быть переслан ими общему сервису учёта состояния системы.

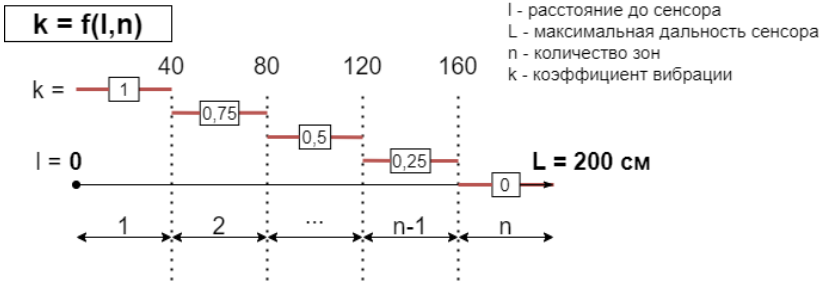


Рис.1. Пример оценки достоверности лингвистической переменной «Препятствие близко» по показаниям дальномерного сенсора

После получения численных значений достоверности каждой входной переменной по обычным правилам нечёткой логики оценивается достоверность выходных переменных. Набор выходных переменных определяется имеющимися тактильными сенсорами и используемыми в них режимами работы; интерпретация выходных переменных в разных случаях различается; скажем, режим работы температурного элемента задаётся одним параметром «Интенсивность», однозначно определяющим его целевую температуру. Например, правило, по которому вибросигнал подаётся тем сильнее, чем ближе препятствие, выглядит как:

ЕСЛИ препятствие_по_дальномеру_X_близко

ТО вибрация_актуатора_Y_включена.

Дополнительно к системе могут быть добавлены правила, исключающие недопустимые по эргономическим соображениям [5] воздействия на пользователя.

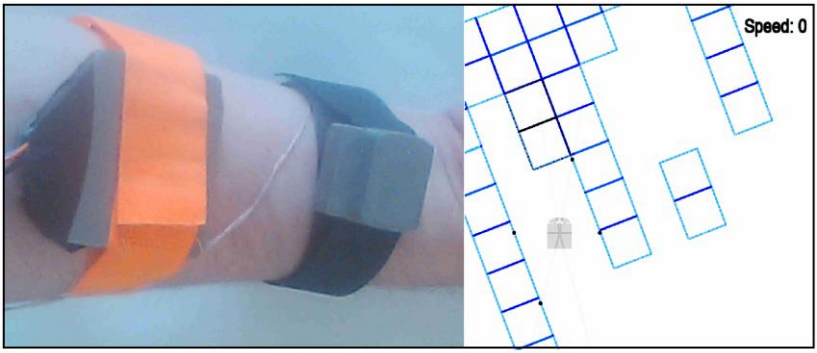
После дефаззификации, проводимой обычным образом, формируются управляющие сигналы для каждого из компонентов интерфейса.

Достоинством нечёткой системы, определившим её выбор, является гибкость настройки. Изменение параметров фаззификации позволяет предсказуемо менять силу каждого из воздействий, адаптируя систему к задаче и добиваясь наилучших результатов при сохранении хорошего самочувствия пользователя.

4. ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ МОДАЛЬНОСТЕЙ

Интерфейсы всех четырёх указанных типов (вибрационный, температурный, компрессионный, электрический) были аппаратно реализованы на основе контроллеров Atmega238. На рис.2 показан

внешний вид температурного и вибрационного актуаторов на запясть испытуемого. Напечатанные на 3D принтере корпуса удерживают активные элементы и датчики на запястье.



а) б)

Рис.2. а) Внешний вид температурного (слева) и вибрационного (справа) актуаторов на запястье испытуемого; б) внешний вид интерфейса тестовой программы (цвета для удобства воспроизведения инвертированы)

Реализация термо-интерфейса выполнена при помощи элементов Пельтье TEC1-12706. Модули позволяют получать температуры от -55 до 83°C, однако исследования показывают, что комфортные для восприятия человеком температуры находятся в диапазоне 15-45°C [5], поэтому и рабочие температуры устройства не должны превышать 45°C. Компрессионный интерфейс выполнен в виде пневматической системы, под действием управляющих сигналов подающей воздух в надуваемую манжету. В качестве виброактуаторов использованы моторы QX-6A-3V.

Программная часть интерфейса реализована следующим образом. Программа, связанная с роботом, фиксирует некоторые триггерные события на данных навигации и сенсоров. Триггерными событиями могут являться, например, обнаружение преграды или опасности рядом с роботом, приближение к цели и т.д. Далее эти события передаются сообщениями на драйвер интерфейса и там обрабатываются, чтобы привести данные к нужному виду согласно разработанному протоколу, а затем передать информацию на контроллер. Контроллер принимает текущие состояния устройств от датчиков и инструкции от драйвера.

На контроллере информация с датчиков обрабатывается и по принципу петли гистерезиса принимается решение: менять режим

работы или нет. Гистерезис требуется для предотвращения постоянного переключения сигналов. Например, для температурного интерфейса алгоритм следующий: если требуемая температура меньше текущей с учетом некоего диапазона (рис.3), контроллер отдает команду остановить нагрев; если требуемая температура больше текущей с учетом диапазона, то отдается команда начать нагрев. Информация о текущих и требуемых температурах для удобства может быть выведена на вспомогательный LED дисплей.

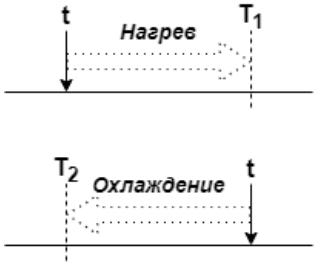


Рис.3. Визуализация алгоритма работы температурного интерфейса; T_1, T_2 – требуемая температура, $T_1 < T_2$; t – текущаяя температура

Для работы системы вибротактильной обратной связи использовалось описание подаваемого сигнала в виде упорядоченной четверки (рис.4):

$$S=<T1, P, N, T0>$$

Здесь:

- *T1* – длительность сигнала высокого уровня t_1,
- *P* – интервал между ними t_0,
- *N* – количество сигналов в группе (далее пачка) *N*,
- *T0* – интервал между пачками T_0.

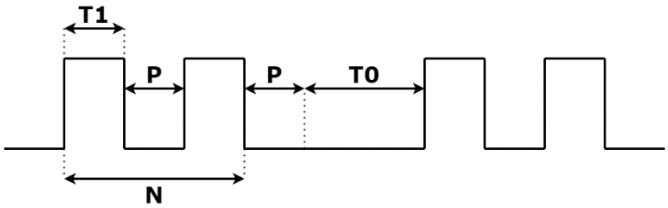


Рис.4. Параметры сигнала актуатора системы вибротактильной обратной связи

Формирование сигнала ШИМ выполняется на контроллере под управлением команд драйвера.

Для подключения тактильных систем к экспериментальной программе, запуск которой происходит на персональном компьютере, был разработан и унифицирован протокол взаимодействия между этими двумя частями экспериментальной системы.

Разработанный протокол описывает минимально допустимый набор сообщений, необходимых для корректного взаимодействия систем, а также порядок обработки ответных сообщений. Пакет данных для вибротактильной системы, содержит в себе четыре параметра, описание которых было представлено выше; для миостимулятора, термо- и компрессионного воздействия – один параметр, обозначающий наличие сигнала, температуру и давление для каждого вида интерфейса соотвественно.

Экспериментальная программа каждые 30 секунд должна отправлять пакет структурированных данных, в которых описывается режим работы тактильной системы. После того, как микроконтроллер прочитает полученный пакет, он должен послать ответ, содержащий в себе информацию о результате считывания команды.

Полученный программно-аппаратный комплекс был использован для экспериментального исследования качества управления. Испытуемые должны были несколько раз решить тестовую задачу (провести модель робота через препятствия к цели). Как и ожидалось, использование тактильных сигналов не всегда давало положительный эффект: например, при исследовании эффективности виброактуаторов среднее время выполнения задания с ними увеличилось по сравнению с их отсутствием (130с ± 20 против 110с ± 20, доверительная вероятность 0,95). При этом из семи испытуемых двое жаловались на неприятные ощущения от вибрации, а остальные – на непривычность задачи; после настройки режимов работы моторов с помощью модификации процедуры фаззификации субъективно воспринимаемые неприятные ощущения значительно снизились. Это лишний раз показывает правильность выбора максимально гибкой и легко настраиваемой системы формирования сигналов для исследований данного типа новых, малоосвоенных интерфейсов.

5. ЗАКЛЮЧЕНИЕ

Разработанный аппаратно-программный комплекс для исследований тактильных интерфейсов оказался пригоден для выполнения поставленных задач. В его состав входит

экспериментальная программа и группа исследуемых аппаратных устройств с нечёткой системой управления, состоящая из системы вибротактильной обратной связи, миостимуляции, термо- и компрессионного интерфейса. Функционирование этих элементов осуществляется при помощи разработанного унифицированного протокола взаимодействия. Основным преимуществом применяемого подхода к управлению оказались гибкость и простота настройки параметров комплекса. Запланировано проведение широкомасштабных исследований применимости тактильных интерфейсов к различным системам с использованием полученного решения.

Литература

1. Jones N., Sarter N. Tactile Displays: Guidance for Their Design and Application. In: Human Factors The Journal of the Human Factors and Ergonomics Society, March 2008, 50(1):90-111. DOI: 10.1518/001872008X250638.
2. Lopes P., Jonell P., Baudisch P. Affordance++: Allowing Objects to Communicate Dynamic Use. In: CHI '15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. – April 2015. – P. 2515-2524. DOI 10.1145/2702123.2702128
3. Shaoyu Cai, Pingchuan Ke, Takuji Narumi et al. ThermAirGlove: A Pneumatic Glove for Thermal Perception and Material Identification in Virtual Reality. In: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), March 2020. DOI:10.1109/VR46266.2020.00044
4. L. A. Jones and H. N. Ho, "Warm or cool large or small? The challenge of thermal displays", IEEE Trans. Haptics. – 2008. vol. 1. – no. 1. – P. 53-70.
5. ГОСТ Р ИСО 9241-910-2015. Эргономика взаимодействия человек-система. Часть 910. Основы тактильных и осязательных взаимодействий

УДК 004.81

ПОДХОД К МОДЕЛИРОВАНИЮ ПОВЕДЕНИЯ ИСКУССТВЕННЫХ СУЩНОСТЕЙ

Виноградов Г.П., *д.т.н.*
e-mail: wgp272ng@mail.ru
ТвГТУ

1. ВВЕДЕНИЕ

Интерес к проблеме построения «интеллектуальных» машин связан с задачами управления их поведением в организационно-технических системах [1, 2]. Поведение искусственных сущностей на основе формальных моделей исследуется в теории агентов, теории принятия решений, поведенческой робототехники и ряда других направлений [3, 5]. В классической BDI архитектуре агента рассуждения выполняются с использованием механизмов логического вывода на основе ментальных понятий, представленных некоторыми структурами знаний [3]. Последовательная критика такого подхода дана в [2]. Поэтому в процессе проектирования интеллектуальных систем разработчики стали апеллировать к поведению человека в конкретной ситуации. За основу моделирования берется прототип – человек (естественная сущность). Развитие этого подхода на базе теории нечетких множеств было связано с обоснованием понятия субъективно рационального выбора, зависящего от субъективных представлений о ситуации выбора лица, принимающего решение [1, 6]. Там же было показано, что в условиях ограниченных ресурсов мозг использует энергосберегающие и когнитивные технологии в форме доступных паттернов. В связи с этим актуальной становится проблема построения моделей паттернов поведения, которые отражают реальные аналоги, имеющие наибольшую эффективность.

2. ЗАДАЧИ ПРОЕКТИРОВАНИЯ ПОВЕДЕНИЯ СУЩНОСТЯМИ НА ОСНОВЕ ТЕОРИИ ПАТТЕРНОВ

Разработка цифровых продуктов в виде паттернов поведения в системах управления искусственными сущностями предполагает решение, по крайней мере, двух задач. Во-первых, это задача переноса наблюдаемого паттерна поведения субъекта-лидера другим субъектом (выступающего в роли сенсора). Во-вторых, – это задача подбора

понятий для формирования цифрового описания модели паттерна разработчиками, менеджерами проекта, конструкторами, пользователями, так как они используют разные онтологии для описания идей и понятий. Основные критерии здесь следующие: 1) поведение искусственной сущности должно быть понятно и прогнозируемо; 2) поведение должно быть эффективно; 3) поведение должно быть корректируемым, то есть допускать коррекцию опыта.

3. СОСТАВНЫЕ ЭЛЕМЕНТЫ ПАТТЕРНА ПОВЕДЕНИЯ

Формальная модель паттерна поведения приведена в работе [8]. Она предполагает разработку следующих общих ментальных моделей, предназначенных для структурирования объема информации, полученной по результатам наблюдения: 1) восприятия состояний в предметной области; 2) мотивации: зачем что-то делать; 3) способов действия: чем и как будем воздействовать на среду и на себя; 4) процессов в предметной области: модели объектов управления – как это устроено; 5) данных: база типовых ситуаций; 6) ценности текущих и возможных состояний: почему это нужно; 7) целевой и текущих тактических ситуаций; 8) ценности ситуаций целеустремленного состояния (тактических и стратегических) по результатам и эффективности; 9) удовлетворенности ситуацией целеустремленного состояния по ценности и эффективности [1]; 10) принятия решений с изменяющейся структурой предпочтений.

Эти элементы образуют то, что можно назвать «Интеллектуальной цифровой машиной», предназначенной для автоматизации выполнения миссии в интересах субъекта-хозяина. Перечисленные выше модели всегда присутствуют, возможно неосознанно, при выполнении миссии человеком, в том числе и у всех участников проекта при разработке искусственной сущности.

Согласование представлений об архитектуре искусственной сущности путем обмена информацией при обсуждении позволяет определить наиболее эффективную модель паттерна поведения и всех ее составляющих для реализации ее в искусственной сущности.

Полученная таким образом цифровая машина позволяет проводить с ней эксперименты, используя пространственную и визуальную логику для интерпретации результатов, как в методе TOTE (набор входных данных $\rightarrow$ воздействие $\rightarrow$ результат) в смысле Миллера, Галантера, Прибрама [9]. Логика интерпретации базируется на следующих предположениях.

- Выбор субъектом осуществляется на основе представлений о ситуации целеустремленного состояния, которая формально определена в [1].
- Компоненты представления отражают различные аспекты понимания субъектом ситуации целеустремленного состояния и образуют информационную структуру представлений. Множество возможных вариантов представлений обозначим через X.
- Для множества состояний окружения S множество наблюдаемых состояний окружения удовлетворяют условию $S \cap X \neq \varnothing$, то есть представления субъекта могут содержать как объективную составляющую, так и фантомную.
- Альтернативы субъект выбирает в зависимости от оценок удовлетворенности значениями свойств ситуации целеустремленного состояния.
- Формирование представлений осуществляется на основе процедур восприятия, осознания и анализа согласно с когнитивными возможностями субъекта.

В соответствии с введенными предположениями субъект при принятии решений использует три множества альтернатив: управляющие C (способы действия), структурные G (интересы, предпочтения) и идентификации X. Следовательно, можно предположить существование трех виртуальных сторон, осуществляющих выбор соответствующих альтернатив. Правила выбора таких альтернатив в зависимости от понимания субъектом обстановки и структуры своих интересов будем называть *стратегиями*.

Модель восприятия состояний в предметной области. Представления о ситуации выбора основываются на входной информации, которую субъект извлекает из окружения и опыта. Основное предположение состоит в том, что представления не являются точной копией реальности. Динамика процессов в окружении субъекта недоступна прямому восприятию, поэтому представления о ней формируется путем применения процедур идентификации, суть которых сводится к выбору варианта представлений на используемом *языке описания* в зависимости от наблюдаемого состояния и структуры интересов $g \in G$. При этом существуют и известны ограничения $X_s \subseteq X$ на допустимость представлений в качестве альтернатив идентификации в зависимости от наблюдаемых состояний $s \in S$.

Однозначное монотонное отображение $\xi: S \to X$ такое, что $\xi(s) \in X_s$, $s \in S$, называется функцией идентификации; упорядоченный набор $(\xi_1, ..., \xi_n) \equiv \xi_1^n$ – стратегией идентификации на горизонте длины $n < \infty$;

последовательность $\{\xi_1^n,\ n = 1,2,...\}$ – стратегией идентификации на ограниченном горизонте. Поскольку субъект стремится к формированию полезных представлений, то существует $lim\{\xi_1^n\} = \xi^\infty$ при $n \to \infty$.

Так как множества S и X удовлетворяют условию $|S| > |X|$, то однозначное отображение $\xi: S \to X$ порождает разбиение множества S на подмножества $\xi^{-1}(x) = \bigcup\{s \in S :\ \xi(s) = x\} \subset S,\ x \in X$.

Подмножества $\xi^{-1}(x) \subset S,\ x \in X$, являются связными множествами, то есть любой элемент $s \in \xi^{-1}(x)$ однозначно определяет соответствующее представление $x \in X$. Следовательно, можно говорить, что подмножества $\xi^{-1}(x) \subset S,\ x \in X$, образуют классы (паттерны) эквивалентных представлений.

Представления – это форма описания субъектом окружения и себя в нем в форме паттернов. Они играют (если использовать терминологию теории выбора) роль моделей, с помощью которых человек предсказывает и оценивает возможные результаты от выбираемых способов действия. Первоочередная задача стратегий управления представлениями состоит в их расширении, дополнении, что создает большие возможности для выбора.

Пусть агент способен выделять факторы – характеристики окружения $X^k = \left\{ x_i^k,\ i = \overline{1, N} \right\}$. Влияние каждого фактора агент оценивает с помощью лингвистической переменой степень влияния фактора $\mu_x^k(x_i^k) : x_i^k \to [0,1]$. Введем параметр, с помощью которого агент оценивает свою ситуационную осведомленность в ситуации целеустремленного состояния

$$Es^k = \frac{\sum_{i=1}^{N} \mu_x^k(x_i^k) x_i^k}{\sum_{i=1}^{N} \mu_x^k(x_i^k)}.$$

Можно определить следующее ограничение: $\sigma^k(Es^k) \geq \sigma_0^k$, где σ_0^k – некоторый пороговый уровень осведомленности агента от использования собственных источников информации.

Представления о ситуации выбора основываются на входной информации, которую субъект извлекает из окружения и опыта. Модель объекта управлени**я** для «цифровой машины» определяет структуру паттернов**,** структуру и взаимосвязь первичных элементов паттерна.

Будем предполагать, что для описания влияния выделенных факторов на результаты o_i^k, $i=\overline{1,m}$ агент использует аппроксимацию в виде продукционных правил, которые имеют вид (1):

Если x_1 есть A_{r1}^k и если x_2 есть A_{r2}^k и … и если x_N есть A_{rN}^k, то $o_i^k = f_{ir}^k(x1, x2, \ldots, xN)$, $r=\overline{1,R}$, $i=\overline{1,m}$, (1)

где R – количество продукционных правил, r – номер текущего продукционного правила, $o_i^k = f_{ir}^k(x1, x2, \ldots, xN)$ – четкая функция, отражающая представление агента о причинно-следственной связи входных факторов с возможными результатами для *r*-го правила; A_{ri}^k – нечеткие переменные, определенные на $X^k = \left\{ x_i^k,\ i=\overline{1,N} \right\}$.

Поскольку c_j^k является функцией параметров состояния внешней среды, свойств системы, принимаемых во внимание, то набор предположений об их возможных значениях образует сценарий возможного состояния внешней среды, функциональных возможностей системы. Реализация сценариев, с помощью правил (1) позволяет сформировать представление о возможных результатах o_i^k.

Модель пользователя. Модель пользователя описывает потребителя, использующего искусственную сущность (например, дрон). Она моделирует мотивацию, цели, ценности и оценки. Для проекта создания искусственной сущности построение такой модели должно быть отражено в документе «*Техническое задание*». Модели пользователя в практике проектирования интерактивных цифровых продуктов, сред, систем и сервисов рассмотрены в [10].

Модель ценности. Ценность искусственной сущности как продукта определяется набором ее характеристик (параметров). Их значения позволяют пользователю вынести суждение: 1) что это за продукт, то есть что искусственная сущность может делать – набор функциональности; 2) за счет чего она позволит автоматизировать выполнение миссии; 3) почему я должен использовать этот продукт, чем он полезен для меня. Ответ на первый вопрос, который выступает как фильтр, позволяет отнести искусственную сущность к определенной категории. И если

соответствующая категория интересует пользователя, то он будет искать ответы на следующие вопросы модели ценности.

Формально модель ценности можно представить в виде оценки удельной ценности по результату [1]. Так как организация создает условия для роста удельной ценности по эффективности у пользователя, то ценность результата для его первична, а ценность для организации носит производный (зависимый) характер. То есть организация получает ценность как вознаграждение за создание ценности у пользователя.

Проблема – найти формулировку, которая определяет ценность для пользователя, а не для организации, создающей пользовательский опыт. Решение этой проблемы предполагает: 1) выполнение мысленной имитации с помощью бизнес-модели, в которой искусственная сущность лишь одна составляющая общего ценностного предложения организации для пользователя; 2) для создания убедительной модели ценности нужно выстроить ее общими усилиями, применяя человеко-ориентированные методы согласования в процессе разработки, привлекая людей, похожих на целевого пользователя.

Модель взаимодействия. Формализация модели взаимодействия строится на результатах работы [11]. Процесс проектирования можно представить, как процесс реализации последовательности ситуаций. На каждом этапе состояние «цифровой машины» меняется и одновременно меняется состояние среды, которое является генератором условий, определяющих последовательность выполнения этапов.

Пусть принятие решения выполняется в несколько циклических этапов, и способы действия выбираются на каждом этапе $n = 1, 2, ...$ из множества C в зависимости от представления о состоянии окружения $x \in X$. Это связано с тем, что совместный над сознательный (интуитивный) и сознательный (формальный) анализы состояния окружения позволяют в многократных итерациях принять сначала смутно осознаваемое, а затем все более четко формулируемое и обоснованное решение. При этом существуют ограничения $C_x \subseteq C$ на допустимость выбора альтернатив в зависимости от представлений о состоянии окружения $x \in X$.

Исходя из этих предположений, следуя [7], введем определения стратегий. Однозначное отображение $\lambda: X \to C$ такое, что $\lambda(x) \in C_x$, $x \in X$, называется функцией выбора или управления; упорядоченный набор $(\lambda_1, ..., \lambda_n) \equiv \lambda_1^n$ – стратегией выбора на горизонте длины $n < \infty$; $\lim\{\lambda_1^n\} = \lambda_1^{n\,\infty}$ при $n \to \infty$ будем называть стратегией, направленной на достижение локального идеала, определяющего смысл существования субъекта.

Выбранная в момент n структурная альтернатива $\gamma_n \in G$ является *структурным выбором* на n-м шаге принятия решений; упорядоченный набор $(\gamma n, ..., \gamma_1) \equiv \gamma_1^n$ – стратегией *структурного выбора* на горизонте принятия решений длины $n<\infty$; последовательность $\{\gamma_1^n, n = 1, 2, ...\}$ – стратегией структурного выбора на ограниченном горизонте. Поскольку субъект стремится к соответствию своей структуры интересов требованиям принятой им этической системы, то существует $\lim\{\gamma_1^n\} = \gamma^\infty$ при $n\to\infty$.

Модель принятия решений с изменяющейся структурой предпочтений. Согласно [1] критерий выбора стратегии управления имеет смысл ожидаемой удельной ценности целеустремленного состояния по результату, формализация которой имеет вид функции полезности $E\varphi^g(C\times S\times X)$, зависящей от структурной альтернативы $g\in G$ как от параметра. Поскольку процесс управления начинается с некоторой ситуации $x\in X$, то критерий $E\varphi_n(\lambda_1^n \mid \gamma_1^n)$ будет также зависеть и от ситуации $x\in X$ как от начального условия. Так как при этом множество ситуаций X конечно, то критерий $E\varphi n(\lambda_1^n \mid \gamma_1^n)$ будет окончательно представляться вектором в пространстве R^X размерности $|X|$. Его компоненты будем записывать в виде $E\varphi_n(\lambda_1^n \mid \gamma_1^n)(x)$, $x \in X$. По результату выбора субъект испытывает эмоциональное переживание, поэтому качество стратегии структурного выбора γ_1^n следует описывать в виде критерия, имеющего смысл «*удовлетворенности результатами выбора*». Следовательно, качество стратегии γ_1^n естественно описывать сверткой вектора ожидаемой полезности $E\varphi_n(\lambda n \mid \gamma_1^n) \in R^X$ в некоторый функционал $\mu : R^X \to R^1$. Тогда критерий качества стратегии γ_1^n можно записать в виде $\mu_n(\lambda_1^n \mid \gamma_1^n) = \mu(E\varphi_n(\lambda_1^n \mid \gamma_1^n)) \in R^1$.

Качество своих представлений субъект связывает с оценками возможности достижения желаемых состояний от управления $c\in C$, а также с возможностью расширения множества $C\uparrow$ путем включения в него эффективных альтернатив. В работе [6] в качестве критерия оценки представлений использовать термы лингвистической переменной

«*полезность*», которые строятся на значениях $E\varphi_n(\lambda_1^n \mid \gamma_1^n)$. При этом оценки полезности будут зависеть от стратегий управления λ_1^n, структурного выбора γ_1^n как от заданных условий. Обозначим критерий «полезность» следующим образом $\psi n(\xi_1^n \mid \lambda_1^n, \gamma_1^n)$. Поскольку процесс идентификации начинается с некоторого состояния $s \in S$, то этот критерий будет зависеть от состояния $s \in S$, задаваемого в качестве начального условия. Так как при этом множество состояний S конечно, то критерий идентификации будет представляться вектором $\psi n(\xi_1^n \mid \lambda_1^n, \gamma_1^n)$ в пространстве R^S размерности $|S|$.

В ситуации целеустремленного состояния качество стратегий управления и структурного выбора описывается соответственно критериями $E\varphi_n(\lambda_1^n \mid \gamma_1^n) \in R^X$ и $\mu_n(\gamma_1^n \mid \lambda_1^n) \in R^1$, имеющими смысл удельной ценности по результату и удовлетворенности результатами выбора, а качество стратегии идентификации– критерием $\psi_n(\xi_1^n \mid \lambda_1^n, \gamma_1^n) \in R^S$, имеющим смысл полезности представлений для достижения желаемых состояний. Использование введенных критериев предполагает определение соответствующих информационных структур или моделей, позволяющих выполнить соответствующий выбор.

Будем предполагать существование информационной структуры представлений *I*, которая отражает знания и опыт субъекта о: способах действия (управления), своих интересах и предпочтениях, динамике перехода окружения в различные состояния. Тогда можно предположить, что существует структурное преобразование этой структуры в информационную структуру, обеспечивающую возможность построения критерия удельной ценности $E\varphi_n(\lambda_1^n \mid \gamma_1^n)$ и модели предметной области. Такое преобразование будем называть «*преобразованием удельной ценности*», а индуцируемую им информационную структуру будем называть «информационной структурой удельной ценности ситуации целеустремленного состояния по результату» и обозначать $U = U(I)$.

Аналогично, если существует структурное преобразование структуры *I* в информационную структуру, обеспечивающей возможность построения критерия идентификации $\psi n(\xi_1^n \mid \lambda_1^n, \gamma_1^n)$ и модели

процедур идентификации, то такое преобразование будем называть «преобразованием идентификации» и обозначать *R*, а индуцируемую им информационную структуру будем называть «*информационной структурой идентификации*» и обозначать *R*= *R*(*I*).

Представления субъекта о ситуации целеустремленного состояния является субъективными и качественными, построенными на основе наблюдений и анализа процесса перехода окружения под действием управления $c \in C$ в различные состояния $s \in S$. Обозначим правило такого перехода через $q\mathcal{G}(S \mid S \times C)$ из $S \times C$ в S. Фактически субъект для оценки ценности возможных результатов использует построенную по результатам стратегии идентификации ξ_1^n модель $Q\mathcal{G}(X \mid X \times C)$ из $X \times C$ в X. При ее построении учитываются стратегии управления λ_1^n, структурного выбора γ_1^n, либо такими стратегиями он задается. Это означает, что преобразование фактической функции $q\mathcal{G}(S \mid S \times C)$ в функцию понимания субъектом процессов в его окружении $Q\mathcal{G}(X \mid X \times Y)$ возможно лишь в апостериорном режиме в зависимости от используемых стратегий $(\lambda_1^n, \gamma_1^n, \xi_1^n)$. Такое преобразование и построение критерия ожидаемой удельной ценности $E\varphi n(\lambda \mid \gamma_1^n)$ возможно при последовательном формировании информационных структур «полезности» в зависимости от используемых стратегий. Это условие будем записывать в виде $Un = U(\lambda_1^n, \gamma_1^n, \xi_1^n)(I),\ n = 1,2,...$ Поскольку это условие является необходимым для формирования критерия ожидаемой полезности и модели предметной области, то оно должно указываться всякий раз при его использовании. Отметим, что критерий $E\varphi n(\lambda_1^n \mid \gamma_1^n)$ неявно зависит от стратегии идентификации ξ_1^n за счет введения в модель выбора индуцированной структуры U_n. Как было отмечено выше, критерий $\mu_n(\gamma_1^n \mid \lambda_1^n) \in R^1$ качества структурного выбора определяется сверткой критерия $E\varphi n(\lambda_1^n \mid \tau_1^n) \in R^X$.

Для построения критерия идентификации требуется использование некоторой функции, которая имела бы смысл «полезности». Для этого надо построить вербальные оценки на значениях функции полезности

$E\varphi^{g}(S \times X \times Y)$. Требуемое преобразование существует и может выполняться в априорном режиме (т. е. до выбора решений).

Такое преобразование определяется субъектом относительно нечеткой меры, которая может быть построена, если задана функция $q^{g}(S \mid S \times C)$ из $S \times C$ в *S*. Поскольку ее аналог в сознании субъекта имеет вид $Q^{g}(X \mid X \times C)$ и он может ее однозначно задать в информационной структуре *I*, то, следовательно, не требуется дополнительных преобразований. Построение функции «полезности представлений» исчерпывает необходимое структурное преобразование, которое будем называть структурным преобразованием «идентификации» и обозначать *R*, а индуцируемую им информационную структуру называть информационной структурой «полезности представлений» и обозначать *R* = *R*(*I*).

Из введенных определений и построений следует, что критерии качества этих видов стратегий различны и взаимозависимы. Тогда проблема выбора имеет игровое содержание и сводится к отысканию устойчивого компромисса между стремлением к максимизации ожидаемой удельной ценности целеустремленного состояния по результату и минимизации возможных потерь от неправильных представлений. Такой компромисс называется *равновесием*.

Заметим, что поскольку информационная структура «удельной ценности» $Un = U(\lambda_1^n, \gamma_1^n, \xi_1^n)(I)$, в условиях которой строится критерий $\mu_n(E\varphi_n(\gamma_1^n \mid \lambda_1^n))$, должна формироваться последовательно в зависимости от используемых стратегий, то искомые равновесия будут взаимозависимы не только на каждом этапе *n* = 1, 2, ... формирования решений, но они будут зависеть также и от решений, выбираемых на предшествующих, шагах. Таким образом, содержание проблемы моделирования выбора состоит в отыскании компромисса между стремлением к достижению максимальной ожидаемой удельной ценности по результату и минимальных потерь от неверных представлений с учетом их взаимной зависимости. При достижении такого компромисса можно утверждать, что интересы субъекта реализуются с «наилучшим результатом». С учетом этого динамические равновесия естественным образом определяют *внутреннюю цель* при принятии решений.

4. ЗАКЛЮЧЕНИЕ

Формальная модель паттерна поведения субъекта-лидера позволяет сформировать архитектуру искусственной сущности, как некоторой «цифровой машины», представленной как совокупность десяти моделей. Показано, что эти модели в той или иной форме, в том числе и неосознанно, присутствуют, как у разработчиков, так и потребителей этой машины. Достичь желаемого резкого повышения эффективности подобных систем и комплексов возможно, главным образом, путем организации по определенным правилам коммуникации разработчиков отдельных элементов модели паттерна. Это позволяет выявить неосознанные предположения и допущения на ранней стадии и оценить степень согласованности решений, предложенных участниками проекта. Использование модели заказчика образует так называемый «корпоративный интеллект», что позволяет обеспечить проектирование в соответствии с функциональным назначением искусственной сущности в неопределенной и плохо формализуемой среде.

Литература

1. Виноградов Г.П. Моделирование принятия решений интеллектуальным агентом. Программные продукты и системы. – 2010. – № 3. – С. 45-51.
2. Городецкий В.И., Самойлов В.В., Троцкий Д.В. Базовая онтология коллективного поведения автономных агентов и ее расширения / Известия РАН, Теория и системы управления. – 2015. – №5. – С. 102-121.
3. Вулдридж, М. (2000). Рассуждения о рациональных агентах. MIT Press. ACL – Agent Communication Language. – URL: http://fipa.org/specs/fipa00061/SC00061G.pdf (дата обращения: 19.01.2021).
4. Bernon C. et al. ADELFE: A Methodology for Adaptive Multi-Agent Systems Engineering // Proc. of the 3th International Workshop on Engineering Societies in the Agents World. – 2002. – P. 156-169.
6. G.P. Vinogradov. A Subjectve Rational Choice. IOP Conf. Series: Journal of Physics: Conference Series 803 (2017) 012176 DOI: 10.1088/1742–6506/803/1/012176.
7. Баранов В.В. Динамические равновесия в задачах стохастического управления и принятия решений при неопределенностях // Изв. РАН. Теория и системы управления. 2002. №3. С. 77-93.
8. Vinogradov G.P. Pattern control system/Software Journal: Theory and Applications. № 2, –2020. – №5-6. – P. 10-19.
9. Miller G.A., Gflfnter G., Pribram K.H. Plans and Strukture of Behaviour. Holt Druclen, New York. 1960.
10. Johnson J., Henderson A. Conceptual Models: Core to Good Design, Penn. State University. 2012. DOI: 10.2200/S00391ED1V01Y201111HCI012
11. Cooper A., Reimann R., Cronin D. About Face 3: The Essentials of Interface Design. Wiley Publisheng, Inc. 2007. 651 p.

УДК: 004.896

ПАТТЕРНЫ В КОГНИТИВНЫХ МОДЕЛЯХ ПОВЕДЕНИЯ АВТОНОМНЫХ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ

Шапель Д.А.
e-mail: shapel@cps.tver.ru
НИИ «Центрпрограммсистем»

1. ВВЕДЕНИЕ

Автономная интеллектуальная система обычно реализуется как совокупность подвижной механической базы, бортовых систем, систем управления, пульта управления с информационно-управляющим полем, и при этом предполагает самодостаточность поведения, гарантирующего выполнение определенной миссии. Задачи, возникающие перед автономной системой, как правило, достаточно сложны для формализации традиционными методами, их многочисленность делает невыполнимой формирование базы знаний на их основе. В то же время они могут быть описаны на естественном языке, а также средствами нечеткой логики. Автоматизация управления автономной системой или робототехническим комплексом (РТК) с использованием такого подхода не устраняет всех проблем традиционной алгоритмизации, интерпретации, защиты и анализа данных, влияния ошибок обработки данных и алгоритмов. Она позволяет строить эффективные логические структуры и реализовывать их в аппаратном обеспечении, приспособленном для массового серийного выпуска и практически достижимых на сегодняшний день вычислительных возможностей. Кроме того, такой способ описания типовых задач – это путь к совершенствованию человеко-машинных интерфейсов.

2. КОГНИТИВНЫЕ МОДЕЛИ ПОВЕДЕНИЯ

Характерной особенностью когнитивных систем является тот факт, что поведенческие процессы протекают в подобных системах в некотором оптимуме, заданном системой управления [1]. В теории систем управление рассматривается как некое воздействие на вход системы, которое перевело бы ее в желаемое состояние. Для моделирования такого оптимального управления применяются

показатели качества управления и критерии (правила), в соответствии с которым выбираются наилучшие управляющие воздействия. При разработке модели функционирования нечетких динамических моделей для сложных систем и процессов различной природы необходимо учитывать такие составляющие, как управление органами восприятия, формирование управляющих воздействий информационной системой управления и другие. Для компенсации допущенных ошибок в реализации управляющих воздействий или при наличии внешних возмущений осуществляется адаптация информационной системы.

3. НЕЧЕТКИЕ ИНТЕРВАЛЫ ОЦЕНКИ

В пространстве оценок системы X существует некое подпространство $X_1 \in X$, в которое необходимо перевести систему в результате процесса управления за конечное время управления, т.к. сам процесс может настолько затянуться, что цель управления потеряет актуальность. Чаще всего моделируется ситуация, в которой решение первоначальной задачи достигается только после решения всех подзадач [2]. Однако в условиях неопределенности получить конечный результат затруднительно, и фактически, приходится упрощать ситуацию применения.

Использование интервального оценивания для каждой проблемы позволяет привести разбиение первоначально высокого уровня неопределенности информации, необходимой для решения задачи, к множеству более мелких неопределенностей (определить альтернативные управляющие решения с помощью нечетких интервалов).

Такую иерархию проблем принято называть иерархией слоев принятия решений [3], а всю систему − многослойной системой принятия решений, изображенной на рис.1, где S_i является решающей подсистемой и/или ситуацией.

Следует отметить, что подавляющее большинство, например, всех морских происшествий на 80 % связано с человеческим фактором (ошибка экипажа на борту судна или береговой службы). Автоматизация судов обычно рассматривается как способ повышения эффективности и безопасности системы за счет сокращения риска ошибки человека [4]. Аналогичные задачи решаются операторами авиационных, железнодорожных перевозок и другими специалистами [10, 14].

Получить требуемые результаты безопасности и эффективности в практическом применении автономных интеллектуальных устройств в настоящее время возможно при использовании так называемого

кооперативного интеллекта. Человек-оператор должен получить все преимущества роботизированной системы, такие как быстродействие, точность и др. Их взаимодействие позволит передавать опыт не только от человека к машине, но и в обратном направлении, обеспечивая тем самым приспособительное поведение всей системы.

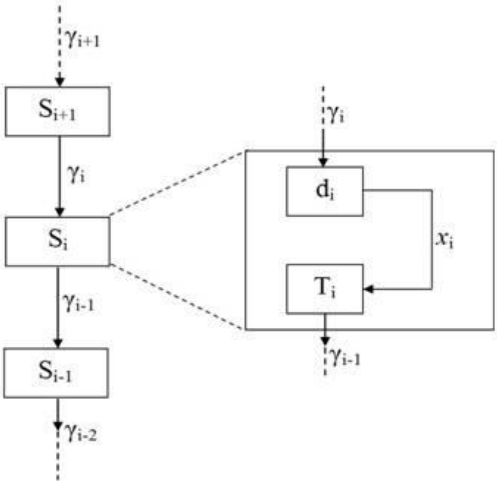


Рис.1. Многоуровневое представление решающих подсистем

Можно использовать сенсорно-моторные рефлексы и когнитивные способности человека-оператора (включая интуицию, инстинкт и эмоции) для принятия решения в сложных ситуациях, т.к. они являются дополнительным информационным ресурсом, позволяющим учитывать влияние человеческого фактора в контуре управления и тем самым повысить эффективность разработки и применения ИСУ. Таким образом, управление реализуется по принципу «человек в контуре управления» [5].

Экспериментальные исследования коры головного мозга и поведенческих реакций человека-оператора [6] подтвердили гипотезу о взаимосвязи электрического возбуждения отдельных участков коры головного мозга (нейронов или группы нейронов) с формированием и прогнозом поведенческих реакций человека-оператора. Выполняемые оператором задания, принадлежащие разным типам когнитивной деятельности, соотносятся с принципиально различными физиологическими паттернами. Паттерны, соответствующие конкретным видам когнитивных заданий, принадлежащим одному типу когнитивной деятельности, различаются между собой в деталях.

Различия между такими паттернами существенно меньше, чем между паттернами, принадлежащими различным типам когнитивной деятельности.

Процессы, происходящие в сознании человека при выполнении определенных действий, приводят к расширению его структуры способностей (знания, опыт). Модель поведения автономной интеллектуальной системы так же должна учитывать этот феномен взаимовлияния. Существует большой потенциал применения ещё относительно слабо изученных перспективных видов человеко-машинных интерфейсов (прежде всего, голосовых и нейрокомпьютерных интерфейсов), возможностей применения заранее зафиксированных реакций на типовые ситуации (ТС), а также практической подготовки операторов в комбинированных с машиной системах.

Поведение в типовых ситуациях требует особого внимания в рассматриваемом подходе и связано с выбором, который происходит в ситуации ***целеустремленного состояния*** [2]. Если рассматривать модель поведения в виде нечеткого описания модели ситуации выбора, возможный вариант такой конструкции можно строить путем так называемой парадигмальной прививки идей, транслируемых из других наук [8]. Так, например, целеустремленное состояние складывается из следующих компонентов:

- Субъект, осуществляющий выбор (агент), $k \in K$.
- Окружение выбора (S), под которым понимается множество элементов и их существенных свойств, изменение в любом из которых может стать причиной или продуцировать изменение состояния целеустремленного выбора. Часть этих элементов может не являться элементами системы и образует внешнюю среду для нее. Воздействие внешней среды описывается с помощью некоторого набора переменных.
- Доступные способы действий k-го агента, которые ему известны и могут быть использованы для достижения i-го результата (их еще называют альтернативами). Каждый способ этого множества характеризуется набором параметров, называющихся управляющими воздействиями:

$$c_j^k \in C^k, \quad j = \overline{1,n}$$

- Возможные при окружении S результаты, существенные для агента:

$$o_i^k \in O^k, \; i = \overline{1,m}$$

Модель субъективного представления (восприятия) о ситуации целеустремленного состояния можно описать с точки зрения среды следующим образом:

Характеристики окружения:

$$X^k = \left\{ x_i^k,\ i = \overline{1, N} \right\}$$

Степень влияния каждого фактора:

$$\mu_x^k(x_i^k) : x_i^k \to [0,1]$$

Оценка ситуационной осведомленности в ситуации целеустремленного состояния:

Ограничение на пороговый уровень осведомленности агента при

$$Es^k = \frac{\sum_{i=1}^{N} \mu_x^k(x_i^k) x_i^k}{\sum_{i=1}^{N} \mu_x^k(x_i^k)}$$

использовании собственных источников информации:

Оценка результатов производится с помощью определенного

$$\sigma^k(Es^k) \geq \sigma_0^k,$$

набора характеристик, которые называются выходными параметрами ситуации целеустремленного состояния.

- Метод оценки свойств получаемых результатов вследствие выбора способа действия.
- Ограничения, отражающие требования, накладываемые ситуацией выбора на выходные переменные и управляющие воздействия.
- Модель предметной области, которая представляет собой множество соотношений, описывающих зависимость управляющих воздействий, параметров и возмущений с выходными переменными.
- Модель ограничений агента (подробное описание дается в работе [6]). Независимо от используемого вида описания ограничений можно предполагать наличие у агента определенной степени уверенности о возможности изменения части ограничений в сторону расширения множества возможных вариантов выбора.

Моделирование паттернов выполняется с ограниченным подмножеством естественного языка, как и моделирование рассуждений на паттернах.

Модель паттерна следует рассматривать как единицу человеческого опыта, для которой в ситуации, схожей с типовой, у человека сформирована определенная степень уверенности в получении желаемых состояний. Человек, приобретая опыт, нацелен на агрегирование его путем создания моделей паттернов.

Для оценки целеустремленного состояния агент способен выделять характеристики окружения [9]. В качестве функции, отражающей представление агента о причинно-следственной связи входных факторов, могут использоваться, например, математические модели, словесное описание, графики, таблицы, алгоритмы т.д. Влияние каждого фактора агент может оценивает с помощью лингвистической переменой.

4. ОБРАБОТКА ЕСТЕСТВЕННО-ЯЗЫКОВОЙ ИНФОРМАЦИИ

В настоящее время многие исследования и разработки посвящены вопросам категоризации текстов. В классификаторах, основанных на традиционных алгоритмах, предполагается, что каждый класс текстов однозначно отделяется от других классов, т.к. классификация базируется на предположении, что разные классы являются взаимоисключающими и каждый объект не может принадлежать более чем к одному классу [9].

Сложность систематизации естественно-языковой информации состоит в нечеткости смысловых значений словесных выражений. Описание объектов исследования на основе лингвистических правил и традиционного представления в виде функциональных зависимостей сочетает в себе модель нечеткого вывода Сугено. Такой подход значительно упрощает интерпретацию полученных результатов, делает их понятными и логически обоснованными [16]. Модель Сугено, более простая с точки зрения математической обработки текстов, может применяться как альтернатива классическим методам машинного обучения (Байеса, К-ближайших соседей, деревьев решений, опорных векторов) и позволяет добиться точности классификации выше 80 %. Прежде всего, это относится к коротким техническим текстам и описанию технических состояний.

Особый интерес с точки зрения обработки входящей информации представляет получение знаний из семантически неоднородных и распределенных источников данных. Данные, сгенерированные из электронных систем автономных устройств и робототехнических комплексов, часто имеют определенные особенности:

- разнородность вследствие разнообразия устройств сбора данных;

- разномасштабность с сохранением не только полученных в настоящее время, но и архивных данных в течение заданного периода времени;
- эффект значительной корреляции между временем и пространством, поскольку каждый набор данных относится к определенному пространственному положению и каждый фрагмент данных имеет соответствующую отметку времени.

Кроме того, существует проблема эффективности и ценности данных. Из большого количества собранных данных лишь незначительная часть имеет ценность для принятия решений или дальнейшей обработки. Зачастую, чем выше объем данных, тем большее количество шумов (помех) или ненужных данных может быть изъято во время сбора и передачи данных.

Из-за разнообразия типов данных некоторые базы данных могут содержать сложные объекты данных, включая временные данные, пространственные данные, данные транзакций, графические и другие медиаданные [11]. Обработка нечисловых и числовых переменных отличается. Нечисловые значения можно обрабатывать определенными инструментами, например, обработчиком деревьев решений. Другой инструмент, например, нейросеть, может обрабатывать только числовое представление значения. Следовательно, для разных видов данных существуют разные системы их интеллектуального анализа. Критика аналитики больших данных заключается в том, что существование массивных наборов данных не устраняет традиционные статистические ошибки и смещения выборки. Даже если корреляция может оказаться надежной в течение определенного периода времени, аналитика больших данных сама по себе не может дать представление о том, что приведет к нарушению корреляции, или о том, какая модель может появиться на ее месте.

Моделирование паттернов, в том числе и моделирование рассуждений на паттернах, выполняется с ограниченным подмножеством естественного языка [2], что образует специфическую часть человеческого опыта – метаопыт.

В предлагаемой модели все составляющие, кроме логических связок, могут представлять собой языковые конструкции на естественном или естественно-профессиональном языке. Поэтому такая модель определяется как типовая естественно-языковая модель паттерна. Соответственным образом выстраиваются модели памяти интеллектуальной системы, изображенные на рис.2.

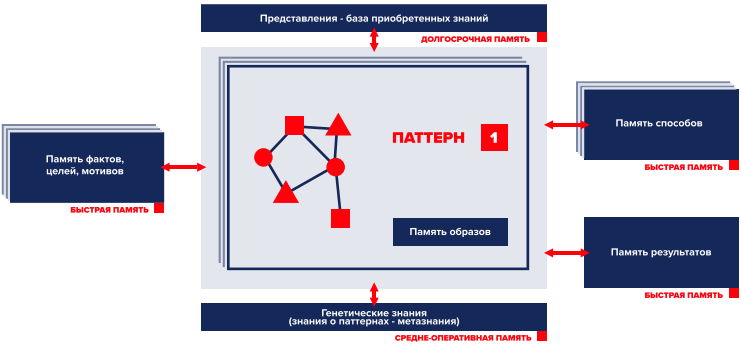


Рис.2. Модели памяти и мышления интеллектуальной системы

5. ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА

Автономная система, проявляющая поведение, направленное на выполнение миссии, реализуется, как правило, как интеллектуальный агент (ИА), взаимодействует с другими агентами, со средой при выполнении задач и обязательств с помощью доступных способов действия и имеет в своём составе:

- бортовые измерительные устройства (или комплекс бортовых измерительных устройств), позволяющих получить информацию о состоянии внешней среды и собственном состоянии;
- бортовые исполнительные устройства, с помощью которых система воздействует на внешнюю среду и на саму себя, выполняющие роль эффекторов;
- средства коммуникации с другими системами;
- бортовой интеллект, составляющими которого могут быть бортовые вычислители, их программное обеспечение, а также операторы пункта управления, являющиеся носителем набора алгоритмов для решения задач предметной области, полученный за счет обучения, тренировок и накопления опыта.

Общая схема ИА представлена на рис.3.

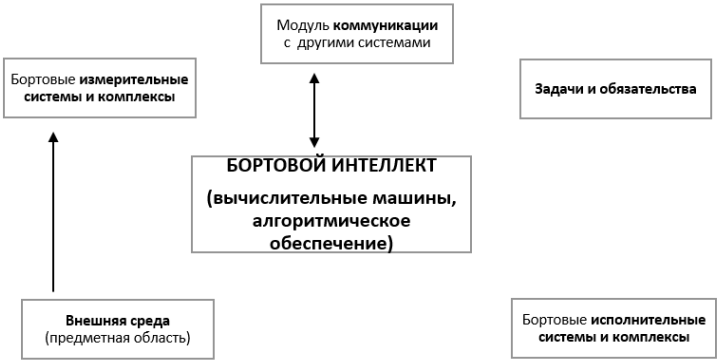


Рис.3. Интеллектуальная автономная система

Поставленные задачи интеллектуальный агент выполняет исходя из понимания своего состояния и субъективных представлений о состоянии среды и развитии боевой ситуации, а также информации, полученной через модуль коммуникации. Агент способен прогнозировать изменение среды от своих действий и оценивать их полезность.

Типовая ситуация как единое целое и как функционально замкнутая с четко обозначенной значимой целью часть работы встречается в различных (реальных) сеансах и вместе со ***способом действия*** (реакции на типовую ситуацию) образует **паттерн поведения**. Учитывая ранее описанные особенности агрегирования опыта оператора-человека, модель паттерна следует также рассматривать как единицу человеческого опыта, для которой в ситуации, схожей с типовой, у человека сформирована определенная степень уверенности в получении желаемых состояний.

Моделью выбора агента в типовой ситуации можно называть множество структурных и функциональных свойств, которыми, по его убеждению, обладает ситуация выбора и которые влияют на его удовлетворенность или неудовлетворенность ситуацией. Есть группа факторов [18], определяющих реализацию результата – воля, склонность к рискам, самооценка, мотивированность (рис.4). Эти факторы позволяют говорить о показателе уверенности в получении результата в ситуации выбора при использовании одного из возможных способов действия. Агент рассматривает паттерн как способ описания задачи, принцип и алгоритм ее решения [5]. Такая задача возникает часто, причем ее решение можно использовать много раз, ничего не изобретая заново. Для достижения желаемого результата агент располагает несколькими альтернативными способами с разной эффективностью, поэтому его уверенность в получении желаемого результата значительна.

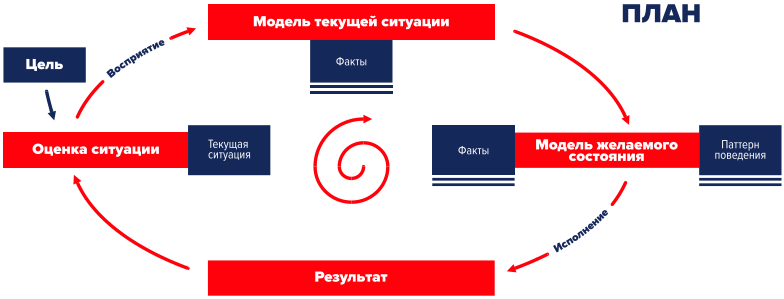


Рис.4. Схема мышления интеллектуального агента

Такая модель индивидуального поведения автономного агента предполагает формирование базы знаний путем обучения на основе экспериментального опыта, что позволяет реализовать эволюцию кооперативного интеллекта благодаря использованию искусственного когнитивного процесса, аналогичного тому, что имеет место у естественных сущностей [18]. При этом указанная возможность отсутствует в системах на основе знаний, так как в ней нет компьютерной модели адаптивного поведения.

Таким образом, общие принципы мышления агента являются вполне традиционными и включают три основные фазы:

- восприятие – получение данных и построение модели сцены в загруженном мире;
- познание – анализ и формирование сценария действий субъекта для достижения поставленных целей;
- исполнение намеченного сценария с постоянным сопоставлением ожидаемых и наблюдаемых результатов.

В отличие от других в рассматриваемой системе реализация этих фаз осуществляется через базовые механизмы абстрагирования и конкретизации, тесно связанные между собой. Таким образом, паттерн можно описать представленным на рис.4 типовым образом:

Имя паттерна:
так как [мотивы M]
поскольку [цели G]
если [предусловия U']
то способ действия
из-за чего [постусловие U"]
...
есть альтернатива [ ]

Схема 1. Логическая схема описания паттерна

В теории паттернов используются паттерны четырех уровней. Первый уровень – это паттерны, называемые образующими. Они моделируют относительно простые объекты реального мира, обладают связями, которые могут соединяться со связями конструирования регулярных конфигураций, которые составляют объекты второго уровня. Объектами третьего уровня являются изображения, получаемые путем объединения сходных между собой регулярных конфигураций в множество, на котором определяются классы эквивалентности, что и называется изображением.

Четвертый уровень объектов составляют образы – множество изображений, инвариантных относительно преобразований подобия. Если на некотором множестве регулярных конфигураций существует только одно изображение, то оно представляет собой единственный образ этого множества. Следовательно, должны существовать методы, позволяющие моделировать связи, соединения и преобразования подобия логических объектов реального мира. Задачи рассуждений можно решать путем обработки образов на нейросетевых структурах.

Синтез всех полученных представлений дает возможность получить модель, характеризуемую максимальными значениями показателей удельной ценности по результату и эффективности. Этот подход предполагает использование имплицитной и эксплицитной информации.

Агент может знать или понимать суть некоторой деятельности, однако может быть не способен осуществлять ее (осознанная некомпетентность). И наоборот, агент может оказаться способен хорошо выполнять некоторые действия, но не понимать, как это делается (неосознанная компетентность). Владение навыком в совершенстве подразумевает не только способность «делать то, что знаешь», но и способность «знать, что делаешь». Тем не менее, многие поведенческие и психологические элементы, обеспечивающие успешность действий агентов, остаются по большей части неосознанными и лишь интуитивно понятными для них самих. В результате они неспособны напрямую описать механизмы, лежащие в основе каких-либо умений. Поэтому одной из целей моделирования является выявление и идентификация неосознанной компетентности, а также доведение ее до сознания с целью лучшего понимания, совершенствования и передачи навыка.

Когнитивная и поведенческая компетентности могут быть смоделированы либо имплицитно, либо эксплицитно. ***Имплицитное моделирование*** предполагает занятие по отношению к субъекту моделирования такой позиции, которая позволит добиться

интуитивного понимания субъективных переживаний данного человека. ***Эксплицитное моделирование*** состоит в описании явной структуры переживаний моделируемого агента так, чтобы ее можно было передать другим. Имплицитное моделирование является в первую очередь индуктивным процессом, с помощью которого люди принимают и воспринимают структуры окружающего мира. Эксплицитное моделирование, по существу, является дедуктивным процессом, с помощью которого человек описывает и осуществляет это восприятие. Оба процесса необходимы для успешного моделирования (рис.5). Без имплицитной стадии не может быть эффективной интуитивной базы, на основе которой можно построить эксплицитную модель. С другой стороны, без эксплицитной фазы смоделированная информация не сможет воплотиться в приемах или средствах и быть переданной другим. Имплицитное моделирование само по себе помогает человеку развивать личное, неосознанное умение в связи с желаемым поведением (так обычно учатся маленькие дети). Однако создание техники, механизма или навыка, которому можно научить или передать его другим, в каком-то смысле требует применения эксплицитного моделирования.

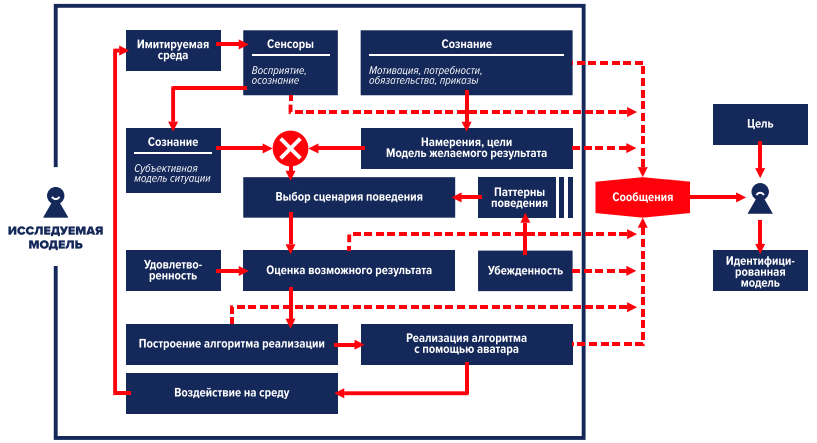


Рис 5. Схема рефлексивного подхода при идентификации модели паттерна поведения

Результатом становится модель, в которой синтезированы интуитивное понимание способностей агента, непосредственные наблюдения за работой агента, эксплицитные познания исследователя в предметной области агента.

6. ЗАКЛЮЧЕНИЕ

Интеллектуальные технологии, использующие теорию паттернов, имеют значительные перспективы, так как позволяют решить практические проблемы программирования реальных устройств, устранить проблемы вычислительной сложности. Формальная модель паттерна поведения для системы управления интеллектуальным автономным объектом, представленная в статье, описывает механизмы формирования субъективных представлений и оценок компонент ситуации выбора, а также модель выбора, учитывающую мотивы и обязательства. Построение модели предполагает выявление носителя наиболее успешной модели поведения. В результате идентификация паттерна поведения оператора может осуществляется на основе выявления паттернов команд управления. Это открывает новые возможности в развитии перспективных человеко-машинных интерфейсов.

Обработка результатов предполагает синтез интуитивного понимания способностей агента, непосредственных наблюдений за работой агента и эксплицитных познаний исследователя в предметной области агента.

Литература

1. Кузнецов О.П. Когнитивная семантика и искусственный интеллект // Искусственный интеллект и принятие решений. –2012. – № 4. – С. 32-42.
2. Виноградов Г.П., Прохоров А.А., Шепелев Г.А. Паттерны в системах управления автономными системами // Информационные и математические технологии в науке и управлении. – 2020. – № 1 (17). – С. 40-54. DOI: 10.38028/ESI.2020.17.1.003
3. Карпович С.Н. Многозначная классификация текстовых документов с использованием вероятностного тематического моделирования ml-PLSI // Тр. СПИИРАН. – 2016. – № 47 (4). – С. 92-104.
4. Юрин И.В. Лебедев Г.В., Лившиц И.И. Перспективы использования безэкипажных транспортных судов в морях арктического бассейна России // Научно-технический вестник информационных технологий, механики и оптики. – 2021. – Том 2 №1. – С. 73.
5. Виноградов Г.П. Моделирование принятия решений интеллектуальным агентом. Программные продукты и системы. – 2010. – № 3. – С. 45-51.
6. Волобуев А.Н., Романов Д.В., Романчук П.И. Природа и мозг человека: парадигмы обмена информацией // Бюллетень науки и практики. – 2021. – Т. 7. №1. – С. 59-76. https://doi .org/10.33619/2414-2948/62/06
7. Федунов Б.Е. Конструктивная семантика для разработки алгоритмов бортового интеллекта антропоцентрических объектов. Известия РАН. Теория и системы управления. 1998. №5

8. Mikolov T., Sutskever I., Chen K. Distributed representations of words and phrases and their composition-ality // Advances in neural information processing systems. 2013. P. 3111-3119
9. Боровский А.В., Раковская Е.Е., Бисикало А.Л. Классификация коротких технических текстов с применением системы нечеткого вывода Сугено // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. – 2021. – №1. – С. 16-27. DOI: 10.24143/2072-9502-2021-1-16-27.
10. Аллакулиев Ю.Б. Идентификация системы сбора, передачи и отображения информации узла связи Берегового центра управления автономными необитаемыми подводными аппаратами // Инженерный вестник Дона. – 2019. – №4.
11. Национальный стандарт Российской Федерации 433-2020 Информационные технологии. Интернет вещей. Требования к платформе обмена данными для различных служб интернета вещей (дата введения 01.01.2021)
12. Shin J., Kwon J., Choi J., Im C.-H. Performance enhancement of a brain-computer interface using high-density multi-distance NIRS. Sci Rep 2017; 7(1): 16545, https://doi. org/10.1038/s41598-017-16639-0
13. Багутдинов Р.А. Разработка мультисенсорной системы для задач мониторинга и интерпретации разнородных данных // Системный администратор. – 2019. – №3 (196). – С. 82-85
14. Непряев А.А. Применение принципа целенаправленного поведения в когнитивной системе управления радиолокационной станцией // Программные продукты и системы. – 2021. – №1. DOI:10.15827/0236-235X.133.047-052
15. Fazli, S., Dähne, S., Samek, W., Bießmann, F. & Müller, K.-R. Learning from more than one data source: data fusion techniques for sensorimotor rhythm-based brain–computer interfaces. Proc. IEEE 103, 891–906 (2015)
16. Заде Л. Понятие лингвистической переменной и ее применение к принятию приближенных решений. М.: Радио и связь. 1996
17. Бархатова И.А., Соколов И.А., Шмыков Г.Ю., Ульянов С.В. Интеллектуальные самоорганизующиеся когнитивные регуляторы// УДК 658.618.53, «Системный анализ в науке и образовании». – 2015. – №1.
18. Городецкий. В.И. Самоорганизация и многоагентные системы. II. Приложения и технология разработки // Известия РАН. Теория и системы управления. – 2012. – № 3. – С. 102-123.

СЕКЦИЯ 6 | **НЕЙРОСЕТЕВЫЕ ТЕХНОЛОГИИ И ИХ ПРИЛОЖЕНИЯ**

УДК 004.08

МОНИТОРИНГ СИТУАЦИИ НА ОСНОВЕ ДВУМОДАЛЬНОЙ МОДЕЛИ МИРА

Харламов А.А. *д.т.н.*
e-mail: kharlamov@analyst.ru
ИВНД и НФ РАН, МГЛУ, ВШЭ, МФТИ
Жаркой Р.М.
e-mail: roman.jarkoi@iss.ru
ISS (International group of companies)

1. ВВЕДЕНИЕ

Человек манипулирует внешним для него и интероцептивным миром посредством его моделирования. Модель мира в сознании человека представляется двумя частями: языковой моделью мира и экстралингвистической (многомодальной) моделью мира, которые действуют как единое целое.

Обычно экстралингвистическая часть модели мира используется на нижних (сигнальных) уровнях обработки экстралингвистической (например – зрительной) информации, в то время как языковая модель мира, в силу невысокой вариативности языковых конструкций, – на верхних (символьных) уровнях. Текстовое представление ситуации удобно для ее (ситуации) описания при взаимодействии с пользователем. Такая форма представления информации также удобна и в случае, когда при анализе ситуации используется текстовая информация для представления ограничений ситуации (например, в виде инструкций).

Кортикоморфная реализация экстралингвистической модели мира требует использования очень сложных механизмов структурной обработки информации. Поэтому в качестве такой модели используется так называемое 2.5D представление ситуации. Оно легко интерпретируется пользователем, поэтому может служить основой для эффективного понимания ситуации.

2.5D модель мира обладает еще одним свойством, делающим ее незаменимой в случае ее использования при интерпретации экстралингвистической реальности: эта модель является физически непрерывной, в отличие от сенсорного представления реальности, которая наблюдается сенсорами лишь в некоторых временных и пространственных точках. Это позволяет использовать модель для эффективного трекирования объектов реальности: на ее основе реализуется интерполяция информации, полученной дискретно в разных точках реального пространства и времени.

2. МОДЕЛЬ МИРА ЧЕЛОВЕКА

Поставленная в работе задача предполагает создание модели мира человека, которая совмещает детальность представления мира с простотой ее использования. С этой целью предпринята попытка воспроизведения модели мира, как она формируется в сознании человека. Такое воспроизведение, с одной стороны, позволяет реализовать эффективные механизмы моделирования и манипулирования моделью, а с другой, использовать модель, как удобный интерфейс взаимодействия ее с социумом (и пользователем, как его представителем).

Модель мира человека включает в свой состав языковую часть, описывающую мир в текстах языка, и экстралингвистическую модель, которая представляет мир в терминах других сенсорных модальностей, с помощью которых воспринимает мир человек, в первую очередь – зрительной модальности.

Моделирование мира человека является сложной задачей [1]. И если технология создания языковых моделей в настоящий момент позволяет формировать языковые модели более или менее адекватные задаче моделирования мира [2], то технология создания экстралингвистических моделей позволяет сформировать модели, значительно уступающие языковым по качеству из-за сложности описания мира в неязыковых категориях [3].

Модель мира делится на две части, не одинаковые с точки зрения их участия в формировании представлений о мире у пользователя. Экстралингвистическая часть модели мира формирует, как правило,

нижние уровни представления ситуации, но представленные детально. Языковая часть модели мира формирует верхние (обобщенные) уровни представления.

Таким образом, экстралингвистическая модель содержит всю необходимую информацию о ситуации, языковая же обходится допустимым минимумом, опуская известные социуму (пользователю) детали – так называемые базовые знания.

2.5D модель мира как экстралингвистическая часть модели мира. Интересные возможности моделирования мира в этом случае предоставляют так называемые 2.5 модели [4]. Они, при сравнительной простоте реализации, включают физические свойства реального мира, чем восполняют недостаток учета этих свойств в языковых моделях (вследствие отсутствия в текстах так называемых базовых знаний).

И хоть такие модели своими механизмами отличаются от представления мира в сознании человека, они легко вписываются как часть в модель мира, формируя единые представления с языковой моделью.

Представление мира на основе текстов – языковая часть модели мира. Языковая часть модели мира формируется с использованием хорошо известных механизмов формирования семантических сетей [3], которые позволяют не только наглядно представить ситуацию, но и осуществлять манипулирование ситуациями, в том числе сравнивать их между собой, и потому, кластеризовать и классифицировать.

Такое представление ситуаций позволяет описывать поведение моделей с целью его интерпретации пользователю, но также и вносить в модели ограничения, представленные, например, в нормативных актах, также являющихся текстами.

3. МОДЕЛИРОВАНИЕ В ANYLOGIC

2.5D модель может быть получена разными способами. Например, с использованием инструментария фирмы AnyLogic [5]. В этом случае модель учитывает геометрию пространства и некоторые физические законы, а написанное в ее рамках специальное ПО позволяет естественно представлять поведение агента.

Поведение агента на модели синхронизируется с его поведением в реальном мире посредством сенсоров – видеокамер, находящихся в разных точках реального пространства.

С помощью программного обеспечения AnyLogic строится 2.5D модель агента, который движется в заданном направлении с заданной скоростью с учетом геометрии плана помещения и движением других

агентов. Модель агента позволяет предсказать его поведение в любой момент времени экстраполяцией его параметров во времени в рамках заданной геометрии ограничений пространства.

Сенсорика поведения реального агента дискретна во времени: на входе модели появляются параметры реального агента только в некоторые моменты времени в некоторых точках пространства. В эти моменты времени происходит верификация реального агента его синхронизацией с его модельным эквивалентом. Таким образом осуществляется эффективное трекирование агента.

Наряду с физическими параметрами агента оцениваются некоторые характеристики состояний объекта, например, поза агента.

Поза агента оказывается не только эффективным средством для описания ситуации, но и удобным инструментом для последующего отнесения ситуации к классу. Так две ситуации, неотличимые с помощью простой классификации (например, с применением сверточных искусственных нейронных сетей), легко относятся к двум разным классам ситуаций при использовании категорий позы.

4. ОПИСАНИЕ СИТУАЦИИ ПОСЛЕДОВАТЕЛЬНОСТЬЮ ПРЕДЛОЖЕНИЙ-ШАБЛОНОВ ИЗ ОТДЕЛА «ТО» ПРОДУКЦИОННЫХ ПРАВИЛ

Поведение агента в любой момент времени соотносится с поведением модели не только с учетом представленных выше параметров, но также и с учетом позы агента в этот момент времени. Использование соответствующего продукционного правила из базы знаний позволяет выдать соответствующее предложение естественного языка, описывающее ситуацию, при условии появления на видеокадре определенной позы агента.

Например, если агент курит, и была распознана поза «агент курит», на выход системы выдается предложение «Агент курит». Если агент говорит по телефону – выдается предложение «Агент говорит по телефону».

5. ТЕКСТОВОЕ ОПИСАНИЕ ПОСЛЕДОВАТЕЛЬНОСТИ ПРОСТЫХ СИТУАЦИЙ КАК СЦЕНАРИЙ ПОВЕДЕНИЯ

Ситуация, как правило, это не только объединение нескольких объектов в пространстве, но и некоторое развитие этой конфигурации во времени. Например, представлена последовательность кадров, на которой рука человека находится в течение некоторого времени в одном положении. В другой последовательности кадров чередуются

положения руки у рта и не у рта. Можно рассматривать и другие варианты сценариев ситуаций.

Ситуации, появляющиеся на последовательности видеокадров, делят его на отдельные сцены, которые идентифицируются шаблонами ситуаций. Имена этих шаблонов в динамике развертывания видеоряда фактически являются сценарием поведения агента. Пример такого сценария для видеоряда выглядит следующим образом: «Курит-норма-курит-норма». Время на соответствующих видеокадрах соответствует реальному времени развертывания сценария.

6. ПРЕДСТАВЛЕНИЕ СИТУАЦИИ СЕМАНТИЧЕСКОЙ СЕТЬЮ ТЕКСТА, ОПИСЫВАЮЩЕГО СЦЕНАРИЙ ПОВЕДЕНИЯ АГЕНТА

Сценарий как текст является основой для построения семантической сети, которая используется для распознавания ситуации.

Ситуации, представленные однородными семантическими сетями, легко классифицируются их сравнением с семантическими сетями классов [4].

7. ОПИСАНИЕ ДИНАМИКИ СИТУАЦИИ ПОСЛЕДОВАТЕЛЬНОСТЬЮ СЕМАНТИЧЕСКИХ СЕТЕЙ

В случае необходимости описания динамически развертывающейся ситуации она представляется не одной семантической сетью, а некоторым их кортежем. Одинаково поименованные вершины семантических сетей кортежа связываются друг с другом и характеризуют динамику объектов в ситуации: их появление, существование во времени, исчезновение. Это привносит изменение в описание ситуации как ее динамику во времени.

8. ЗАКЛЮЧЕНИЕ

Двумодальное представление ситуации – в виде ее текстового описания и в виде ее представления на 2.5D модели – является удобным инструментом для моделирования ситуации. Поэтому модель мира, состоящая из двух частей – языковой и экстралингвистической – позволяет использовать возможности видеоаналитики (в том числе на основе искусственных нейронных сетей) и семантического анализа текстовой информации (также с использованием искусственных нейронных сетей). Совместное использование этих двух частей модели мира позволяет также реализовать эффективное взаимодействие такого представления с пользователем при сохранении детальности представлений о мире.

Литература

1. Neuroinformatics and Semantic Representations. Theory and Applications. Alexander Kharlamov & Maria Pilgun eds. – Tyne upon Evon: Cambridge Scholars Publishing, 2020.
2. Харламов А.А. Ассоциативная память — среда для формирования пространства знаний. От биологии к приложениям. – Дюссельдорф: Palmarium Academic Publishing, 2017.
3. Панов Александр Игоревич Одновременное обучение и планирование в когнитивной робототехнике http://www.raai.org/news/pii/ppt/2021/pii2101_Panov.pdf.
4. Харламов А.А. Нейросетевой подход к распознаванию ситуации по тексту. Речевые технологии, № 1 2019 г. С. 22-29.
5. Anastasiia Zhiliaeva Network port: Learning to use the Material Handling Library (Part 5) https://www.anylogic.com/blog/network-port-learning-to-use-the-material-handling-library-part-5/.

УДК 004.056.55, 004.032.26

НЕЙРОСЕТЕВАЯ КРИПТОГРАФИЧЕСКАЯ ОБФУСКАЦИЯ ДЛЯ ДОВЕРЕННЫХ ОБЛАЧНЫХ ВЫЧИСЛЕНИЙ

***Елисеев В.Л.**, к.т.н.*
e-mail: vlad-eliseev@mail.ru
АО «ИнфоТеКС»,
НИУ «МЭИ»
Милюкова Е.А.
e-mail: k.milyush@bk.ru
НИУ «МЭИ»
Колпинский С.В.
e-mail: skolpinsky@bk.ru
АО «ИнфоТеКС»,
НИУ «МЭИ»

1. ВВЕДЕНИЕ

Развитие современных технологий вычислительной техники и телекоммуникаций привело к существенному изменению парадигмы обработки данных. Широкое распространение получили облачные вычисления, при которых обработка данных осуществляется удаленно в крупных вычислительных центрах, а источниками данных являются компьютеры и мобильные устройства пользователей. Центры обработки данных (ЦОД), являющиеся основой инфраструктуры облачных вычислений, функционируют путем совместного использования и виртуализации ресурсов – сетевых адресов, трафика, процессоров, памяти, программного обеспечения и т.п. Краеугольным аспектом является доверие к поставщику облачных услуг, поскольку алгоритмы и обрабатываемые данные размещаются в полностью контролируемой им инфраструктуре.

Алгоритмы обработки данных в зачастую представляют собой объекты интеллектуальной собственности, а иногда и коммерческой тайны, поэтому не должны быть доступны для изучения посторонними. Примерами являются алгоритмы банковского и страхового скоринга, проверки условий лицензирования программного обеспечения, а также

некоторые прикладные ноу-хау, например, в анализе медицинских и маркетинговых данных.

Сами данные, обрабатываемые в облачных ЦОД, в подавляющем большинстве случаев также представляют собой ценность, требующую той или иной степени защиты. Это могут быть персональные данные людей, включая биометрическую и медицинскую информацию, а также коммерчески чувствительная информация. В настоящее время криптографическая защита данных, дающая гарантии конфиденциальности, целостности и аутентичности, обеспечивается только при передаче данных, но не при их обработке.

В связи со всем вышесказанным, представляется актуальной разработка такого подхода, который бы обеспечивал защиту как алгоритмов, так и обрабатываемых им данных, а также был достаточно эффективным в условиях реальной эксплуатации в недоверенной среде облачного агента.

2. ОБЗОР И АНАЛИЗ ТЕХНОЛОГИЙ ЗАЩИТЫ

2.1 Обфускация

Для сохранения секретности алгоритмов могут применяться методы обфускации кода программ. Обфускация (сокрытие) представляет собой процесс преобразования алгоритма к форме, сохраняющей его функциональность, но затрудняющей понимание его работы. Для обфускации разработаны многочисленные эвристические методы, автоматизирующие эквивалентное преобразование программы с целью запутывания оригинальной структуры алгоритма путем изменения представления используемых данных и констант. Такое преобразование дает визуально впечатляющие результаты, однако для каждого из эвристических обфусцирующих алгоритмов может быть создан обратный алгоритм, восстанавливающий исходную структуру. Поэтому эвристическая обфускация не может считаться надежным подходом для защиты алгоритмов, содержащих конфиденциальную информацию.

Задача сокрытия логики работы программы впервые была сформулирована в 1970-е годы, однако до появления работы [1] все методы обфускации представляли собой эвристики, которые было трудно оценивать и сравнивать друг с другом. В работе [1] было впервые введено строгое понятие стойкой обфускации, однако было продемонстрировано, что полное сокрытие логики программы невозможно. В [2] было доказано, что наилучшим из возможных обфускаторов является обфускатор неразличимости, то есть такой,

который сообщает об исходной программе не больше, чем любая другая программа с функциональностью, эквивалентной исходной.

В 2013 году был представлен первый метод обфускации, доказуемо реализующий свойство неразличимости, называемый градуированным кодированием (graded encoding) [3]. Реализация этого подхода для прикладных применений существенно неэффективна с вычислительной точки зрения. В 2019 году в качестве обфускатора неразличимости были предложены искусственные нейронные сети [4], применение которых подробно рассмотрено в п. 4.

2.2 Гомоморфная криптография

В настоящее время единственным подходом, обеспечивающим выполнение доверенных вычислений в недоверенной среде, является гомоморфная криптография, которая подразумевает выполнение операций над зашифрованными данными, при этом сами операции реализуются способом, не зависящим от ключа шифрования. Разработанный в 2009 году алгоритм гомоморфного шифрования был крайне неэффективным: по сравнению с обычными вычислениями производительность снижалась в 10^{12} раз [5]. Продолжение исследований позволило на один-два порядка улучшить производительность [6], но всё равно до практически полезных применений всё ещё далеко.

В работе [7] был предложен и опробован подход по обработке зашифрованных медицинских данных с помощью нейросетей. С помощью данного подхода была продемонстрирована 99% точность классификации рукописных символов над зашифрованными данными MNIST [8]. Отмечалось, что дополнительной проблемой является замедление обучения при использовании зашифрованных данных.

Перспективы гомоморфной криптографии продолжают оцениваться достаточно высоко, что подтверждается как финансированием исследований в крупнейших корпорациях (IBM, Microsoft, Intel, Google), так и появлением стартапов (Duality, Enveil, CryptoNext Security) [9].

3. МОДЕЛЬ НАРУШИТЕЛЯ

При рассмотрении вопроса доверенной обработки данных в публичном облаке необходимо определить местоположение потенциального нарушителя и его возможности для атаки на передаваемые и обрабатываемые данные. Базовая модель нарушителя, рассматриваемая повсеместно, подразумевает его расположение за пределами физической инфраструктуры облачного центра обработки

данных. Облачный провайдер как юридическое лицо, принадлежащая ему инфраструктура и обслуживающий персонал в этом случае рассматриваются как вполне доверенное окружение для программ и данных пользователя. Внешний злоумышленник может атаковать данные, передаваемые в облако и из него (рис.1). Для обеспечения конфиденциальности, целостности и аутентичности данных в этом случае уместно применение криптографических протоколов (IPsec, TLS и др.). Однако фактически публичное облако является местом присутствия многих пользователей, среди которых может быть и нарушитель.

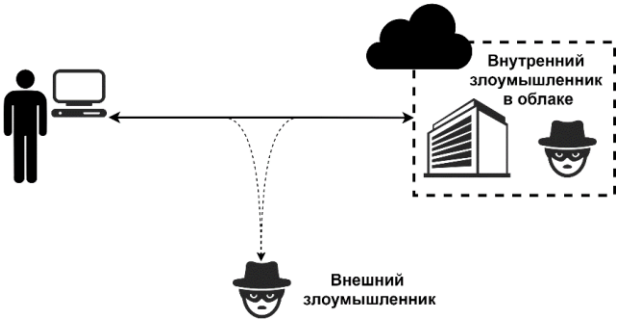


Рис.1. Модель нарушителя при обработке данных в облаке

Для повышения доверия к облачным вычислениям в базовую модель нарушителя добавляют возможность его присутствия в инфраструктуре ЦОД. Традиционные меры защиты в этом случае включают в себя управление идентификацией и доступом (Identification and Access Management, IAM), что позволяет изолировать пользователей внутри принадлежащих только им доменов облачных ресурсов. Это не гарантирует абсолютной безопасности. В частности, широко применяемые в облачных системах токены аутентификации OAuth в случае их хищения открывают нарушителю доступ к облачным ресурсам пользователя. Данная атака известна под именем Man-in-the-Cloud, однако рассматриваемая модель нарушителя могла считаться вполне удовлетворительной до обнаружения уязвимостей современных процессоров, атаки на которые получили имена Spectre и Meltdown. Эти атаки в некоторых условиях позволяют читать память других процессов и даже ядра операционной системы, обходя ограничения IAM. Единственным действенным способом избежать атак Spectre и Meltdown является запуск программ пользователя на выделенном только для него сервере, что фактически разрушает саму основу экономической эффективности облачной модели обработки данных.

Модель нарушителя, допускающая наличие у него прав администратора частей инфраструктуры ЦОД, требует глубоко продуманной системы IAM для персонала ЦОД и значительных ресурсов на её поддержку. При этом сохраняется риск халатности, либо сговора сотрудников ЦОД с целью координированных действий по атаке на облачные ресурсы пользователя.

Одной из проблем облачной обработки данных в современном мире являются трансграничные риски, обусловленные тем, что пользователь и физическая инфраструктура облака зачастую находятся в различных юрисдикциях. Совершенство облачных технологий позволяет не замечать, где сейчас происходит обработка данных: в США, Европе или на территории России, однако с точки зрения законодательства большинства развитых стран мира есть категории данных, перемещение которых за границу даже в зашифрованном виде запрещено. Очевидно, что этот запрет обусловлен тем, что перемещение информации через границу для обработки в облаке делает её доступной иностранным государствам вопреки национальным интересам.

Таким образом, можно констатировать, что современные облачные технологии несут в себе множество рисков для данных пользователя, который хотел бы для их обработки воспользоваться публичными облаками. Кардинальным способом разрешения рисков является эффективная гомоморфная криптографическая система, однако, как было рассмотрено выше, степень зрелости данного подхода пока недостаточна для практически полезных применений.

4. НЕЙРОСЕТЕВАЯ КРИПТОГРАФИЧЕСКАЯ ОБФУСКАЦИЯ

Рассмотренные выше технологии в настоящее время не могут в полной мере решить проблему защищенной обработки данных в недоверенной облачной среде. Вместе с тем, потенциальным решением данной задачи может стать применение искусственных нейронных сетей одновременно как механизма обфускации алгоритмов и как механизма обработки зашифрованных данных. В комплексе это обеспечит свойства, напоминающие гомоморфное шифрование, но на основе искусственных нейронных сетей в качестве инструмента обфускации.

4.1 Искусственные нейронные сети

Хехт-Нильсен в 1987 году на основе работ А.Н. Колмогорова доказал представимость любой непрерывной функции многих

переменных с помощью двухслойного многослойного персептрона (multilayer perceptron, MLP) с любой наперед заданной точностью. Для синтеза искусственной нейронной сети (ИНС) под конкретную прикладную задачу обычно применяется оптимизационный процесс, называемый обучением.

Общеизвестны трудности с извлечением знаний из обученной нейросети, в связи с чем в [10] ставится вопрос, можно ли считать их черным ящиком? Отсутствие четкого понимания внутренних механизмов ИНС даже ставило под сомнение возможность их использования в ответственных прикладных областях [11]. Эту особенность можно использовать во благо, скрывая в ИНС информацию о выполняемых ею действиях, а также для работы с зашифрованными данными.

Идея обработки зашифрованных данных с помощью нейросетей по-разному была сформулирована в [10] и в [11], однако при этом предлагались промежуточные формальные конструкции в виде полиномов и гомоморфных криптографических алгоритмов. Это свидетельствует о недооценке авторами этих работ возможностей нейросетей по обработке зашифрованных данных.

4.2 Нейросетевая обфускация

Как уже отмечалось выше, нейронная сеть обеспечивает обфускацию неразличимости согласно [1]. Таким образом, нейросети могут использоваться для сокрытия алгоритмов внутри структуры весовых коэффициентов. Практическая возможность нейросетевой обфускации была продемонстрирована на примере упрощенного алгоритма банковского скоринга [14]. Разработанная методика основывается на декомпозиции алгоритма на простые функциональные блоки и замену вычислений в этих блоках на вычисления с помощью нейронных сетей. Алгоритм реализован в целых числах, при этом нейронные сети обрабатывали битовое представление обрабатываемой информации.

В [15] предложенная методика была уточнена и развита с целью применения обфускации к произвольным алгоритмам. Были введены общие требования, предъявляемые к функциям, для которых могут строиться нейросетевые аналоги. Одним из ключевых требований является ограничение диапазона аргументов функции, так как количество возможных аргументов функции определяет размер обучающей выборки, которая должна быть конечной. Объем обучающей выборки влияет на длительность обучения и структурную сложность результирующей ИНС. Для векторной булевой функции

объем обучающей выборки, а значит, длительность обучения, будет иметь экспоненциальную сложность от размерности вектора.

Синтез архитектуры искусственных нейронных сетей для конкретных прикладных задач является малоисследованной областью, в которой часто применяются эвристические подходы и большое значение имеет опыт разработчика. Возможность синтеза нейронных сетей для вычисления булевых функций различных классов была продемонстрирована в [16]. Также было проведено исследование зависимости между сложностью булевой функции и сложностью эквивалентной ей нейросети [17]. Под сложностью нейронной сети понималось количество весовых коэффициентов. В качестве метрики сложности булевой функции была выбрана степень нелинейности булевой функции. Выяснилось, что для линейных и бент-функций сложность эксвивалентной ИНС примерно одинакова. Также в [17] были предложены метрика сложности вычислений и итеративная методика синтеза нейросетей для реализации заданной булевой функции.

4.3 Обработка зашифрованных данных

Полученные результаты по нейросетевой обфускации, а также возможности ИНС по аппроксимации произвольных функций позволяют поставить задачу обработки зашифрованных данных. Например, можно использовать таблицы случайных подстановок для шифрования входных и выходных данных нейронной сети, а сам алгоритм представить по методике, изложенной в [14]. В работе [18] было показано, что при размере входного вектора $n \geq 6$ переборная стойкость шифра случайной подстановки, являющегося общим случаем зашифрованного входа нейросетевого криптографического обфускатора, превышает переборную стойкость современных шифров с ключом длины 256 бит.

На основе полученных в [18] теоретических результатов в работе [19] было продемонстрировано практическое применение ИНС по обработке зашифрованных данных. Для трех функций разных классов были проведены эксперименты с созданием эквивалентных нейронных сетей, обрабатывающих открытые и зашифрованные данные. Для шифрования применялись таблицы случайных подстановок. Проведенные исследования показали, что нейронным сетям, обрабатывающим зашифрованные данные, требуется больше нейронов в скрытом слое, то есть в общем случае метрика сложности вычислений для зашифрованных данных выше, чем для открытых. В общем случае обучение нейронных сетей обработке зашифрованных

данных является более трудоемким процессом, чем для открытых данных.

Накладные расходы на обработку зашифрованных данных нейронными сетями можно разделить на два этапа: обучение ИНС и непосредственно обработка зашифрованных данных. Как уже отмечалось, обучение является достаточно длительным и ресурсоемким процессом и должно проводиться в доверенной среде, поскольку строится с использованием таблиц подстановок, являющихся ключом шифрования. Сами вычисления, осуществляемые уже обученной нейросетью над зашифрованными данными, уже не являются экспоненциально сложными и пропорциональны числу весовых коэффициентов ИНС.

На основе предложенных подходов по применению нейронных сетей для криптографической обфускации возможно разработать систему доверенных облачных вычислений. Проект такой системы был предложен в работе [20]. Система состоит из трех подсистем (рис.2): подготовки нейросетевых алгоритмов, обмена данными с облаком и облачных вычислительных агентов. Система включает полный цикл действий по подготовке и работе некоторого вычислительного алгоритма в недоверенной вычислительной среде. В системе производится обработка исходного вычислительного алгоритма с целью его дальнейшей нейросетевой обфускации, создание обфусцированного нейросетевого аналога алгоритма, шифрование исходных данных, передача данных и обфусцированного алгоритма в инфраструктуру облачного агента, проведение вычислений в облаке, передача зашифрованных результатов на компьютер пользователя, расшифрование и получение результата вычислений на компьютере пользователя.

В недоверенной облачной среде функционируют специальный вычислительной агент, в который загружается синтезированная ИНС и с её помощью осуществляются вычисления на основе зашифрованных данных, переданных пользователем агенту по открытым каналам. Полученный результат вычислений также зашифрован и может быть передан пользователю обратно также по открытым каналам. Вычислительный агент не содержит в себе секретов и представляет собой виртуальную машину, выполняющую обмен данными и вычисляющий выход нейросети по входу. Передаваемая в облако нейросеть не содержит в себе ключа шифрования. Более того, сам алгоритм обработки обфусцирован, что обеспечивает его защиту. Таким образом, спроектированная система позволяет проводить доверенные

вычисления в инфраструктуре облачного агента без необходимости доверять ему.

Рис.2. Состав системы доверенных облачных вычислений

Предложенный подход неделимо совмещает шифрование и обфускацию, защищая конфиденциальность как для данных, так и для алгоритма, который их обрабатывает. В то же время, предложенный подход удовлетворяет известному принципу Керкгоффса, поскольку знание злоумышленником алгоритма шифрования, например, факт использования таблицы случайных подстановок, не делает предложенный подход нестойким. Аналогично, знание злоумышленником принципов обучения нейросетей и полный доступ к нейросетевой реализации алгоритма не делает шифрование и обфускацию менее стойкими.

5. ОБСУЖДЕНИЕ

Анализируя предложенный подход нейросетевой обработки зашифрованных данных, следует отметить следующие достоинства:

- высокое быстродействие при выполнении обфусцированных алгоритмов по сравнению с методами гомоморфной криптографии;
- в случае применения таблиц случайных подстановок стойкость имеет факториальную оценку переборной сложности;
- кроме защиты данных обеспечивается защита алгоритма вычислений.

Вместе с тем, этот подход требует дальнейших исследований, поскольку некоторые его недостатки препятствуют широкому внедрению:

- высокая вычислительная сложность обучения НС;
- нет методики обоснованного синтеза архитектуры НС;
- применение для шифрования данных шифрования подстановкой подвержен атаки с помощью частотного анализа;
- вопрос потери стойкости и обфускации в случае перепараметризации остается открытым.

6. ЗАКЛЮЧЕНИЕ

В работе представлен анализ проблем обеспечения доверенных облачных вычислений. Рассмотрены различные подходы к решению данной задачи, а также предложен и обоснован подход на основе искусственных нейронных сетей.

Литература

1. Barak B., Goldreich O., Impagliazzo R., Rudich S., Sahai A., Vadhan S., Ke Yang. On the (im)possibility of obfuscating programs. J. Kilian (ed.) // Advances in Cryptology – Crypto'01. LNCS, v.2139, Springer-Verlag. – 2001. – P. 1-18.
2. Garg S., Gentry C., Halevi S., Raykova M., Sahai A., Waters B. Candidate indistinguishability obfuscation and functional encryption for all circuits // Proc. of the 54th IEEE Annual Symposium on Foundations of Computer Science, 2013, pp. 40-49. 10.1109/FOCS.2013.13.
3. M. R. Albrecht, C. Cocis, F. Laguillaumie, A. Langlois Implementing Candidate Graded Encoding Schemes from Ideal Lattices // International Conference on the Theory and Application of Cryptology and Information Security, ASIACRYPT. – 2015. – P. 752-775.
4. Елисеев В.Л. Искусственные нейронные сети как механизм обфускации вычислений // Труды всероссийской конференции «XVIII Сибирская научная школа-семинар с международным участием «Компьютерная безопасность и криптография» – SIBECRYPT'19», Томск, 9–14 сентября 2019 г., Прикладная дискретная математика. Приложение. – 2019. – №12. – С. 165-169, doi: 10.17223/2226308X/12/46.
5. Microsoft researchers smash homomorphic encryption speed barrier [Электронный ресурс]. URL: https://www.theregister.com/2016/02/09/researchers_break_homomorphic_encryption/ (дата обращения 02.04.2021).
6. S. Halevi, V. Shoup. Faster Homomorphic Linear Transformations in HElib // Cryptology ePrint Archive, Report 2018/244. – 2018.
7. Graepel T., Lauter K., Naehrig M. ML Confidential: Machine Learning on Encrypted Data // Information Security and Cryptology – ICISC 2012. Lecture Notes in Computer Science, vol 7839. Springer, Berlin, Heidelberg, 2012, pp.1-21. doi: 10.1007/978-3-642-37682-5_1.
8. Gilad-Bachrach, R., Dowlin, N., Laine, K., Lauter, K., Naehrig, M. & Wernsing, J. CryptoNets: Applying Neural Networks to Encrypted Data with High Throughput and

Accuracy // Proceedings of The 33rd International Conference on Machine Learning, in Proceedings of Machine Learning Research 48:201-210. – 2016.
9. Duality, a security startup co-founded by the creator of homomorphic encryption, raises $16M [Электронный ресурс]. URL: https://techcrunch.com/2019/10/30/duality-cybersecurity-16-million/ (дата обращения 02.04.2021).
10. J. M. Benitez, J. L. Castro, I. Requena Are artificial neural networks black boxes? // Trans. Neur. Netw. 8, 5 (September 1997). – P. 1156-1164.
11. J.T. Yao Knowledge Extracted from Trained Neural Networks What's next? // Department of computer Sciences University of Ragina Canada. – 2006.
12. Xie, P., Bilenko, M., Finley, T., Gilad-Bachrach, R., Lauter, K., Naehrig, M. Crypto-nets: Neural networks over encrypted data // arXiv preprint arXiv:1412.6181, 2014.
13. R. Dathathri, O. Saarikivi, H. Chen, K. Laine, K. Lauter, S. Maleki, M. Musuvathi, T. Mytkowicz CHET: an optimizing compiler for fully-homomorphic neural-network inferencing // In Proceedings of the 40th ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI 2019). ACM, New York, NY, USA, 142-156. doi: 10.1145/3314221.3314628.
14. Колпинский С.В., Милюкова Е.А., Елисеев В.Л. Нейросетевая обфускация алгоритма банковского скоринга // Сборник материалов XXV Международной научно-технической конференции «Информационные системы и технологии – 2019», Нижний Новгород, 19 апреля 2019 г. – С. 725-732.
15. Милюкова Е.А., Елисеев В.Л. Разработка способа представления вычислительного алгоритма для нейросетевой обфускации // Тезисы докладов XXVII Международной научно-технической конференции студентов и аспирантов «Радиоэлектроника, электротехника и энергетика», Москва. – 11-12 марта 2021. – С. 274.
16. Милюкова Е.А., Елисеев В.Л. Исследование сложности нейросетевой обфускации векторных булевых функций // Сборник трудов XXVIII Международной научно-технической конференции "Современные технологии в задачах управления, автоматики и обработки информации", Алушта, 14-20 сентября 2019 г. – С. 75-76.
17. Милюкова Е.А., Елисеев В.Л. Нейросетевая реализация некоторых классов булевых функций // Сборник материалов XXVI Международной научно-технической конференции «Информационные системы и технологии – 2020», Нижний Новгород, 24,27,28 апреля 2020 г. С. 786-797.
18. Елисеев В.Л. Нейросетевая обфускация вычислений над зашифрованными данными // Прикладная дискретная математика. Приложение, 2020, №13, с. 85-93. doi: 10.17223/2226308X/13/25.
19. Милюкова Е.А., Елисеев В.Л. Исследование нейросетевой обфускации для обработки конфиденциальных данных // в публикации.
20. Милюкова Е.А., Елисеев В.Л. Проектирование системы доверенных облачных вычислений с использованием нейросетевой обфускации // Сборник трудов XXIX Международной научно-технической конференции «Современные технологии в задачах управления, автоматики и обработки информации», Алушта, 14-20 сентября 2020 г. – С. 166-167.

УДК 004.912

ПРИМЕНЕНИЕ ПОДХОДА НА ОСНОВЕ ГИБРИДНЫХ ИНТЕЛЛЕКТУАЛЬНЫХ ИНФОРМАЦИОННЫХ СИСТЕМ ДЛЯ ГЕНЕРАЦИИ ВОПРОСОВ К РУССКОЯЗЫЧНЫМ ТЕКСТАМ

Белянова М.А.
e-mail: belyanovama@student.bmstu.ru
МГТУ им. Н.Э. Баумана

1. ВВЕДЕНИЕ

Генерация вопросов – перспективная задача обработки естественного языка, которая относится к группе задач синтеза текста. С помощью вопросов обогащается речь, способность их задавать – одна из отличительных особенностей сознания, позволяющих ему получать дополнительную информацию.

В данной статье рассматривается классификация способов генерации вопросов, архитектура гибридной интеллектуальной системы генерации вопросов по текстам, а также приведены результаты экспериментов для различных конфигураций модуля формирования вопросов на основе машинного обучения в применении к русскоязычным текстам.

2. КЛАССИФИКАЦИЯ СПОСОБОВ ГЕНЕРАЦИИ ВОПРОСОВ

Классификация способов генерации вопросов, представленная на рис.1, разделяет эти способы по двум признакам: по методам и по структуре исходных данных.

По методам способы генерации вопросов разбиваются на логические, то есть методы на основе правил и условий, методы на основе машинного обучения: статистические, нейронные сети, нечёткая логика, и гибридные методы, которые определённым образом совмещают в себе и логические методы, и методы на основе машинного обучения.

По структуре исходных данных способы генерации вопросов можно разделить на те, что используют в качестве входных данных сырые текстовые данные из различных источников (это может быть онлайн-энциклопедия Wikipedia, тексты книг, история диалогов и т.д.),

графы знаний (в формате RDF-триплетов и т.д.), а также комбинации текстовых данных и графов знаний, которые в классификации обозначены как гибридные данные.

Рис.1. Классификация способов генерации вопросов [2]

Таким образом, способы генерации вопросов можно распределить по 9 различным группам, которые представляют собой комбинации двух категорий из трёх пунктов.

3. АРХИТЕКТУРА ГИБРИДНОЙ ИНТЕЛЛЕКТУАЛЬНОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ

В данной статье предложен метод, позволяющий генерировать вопросы по русскоязычным документам, который является частью системы, в основе которой лежит архитектура Гибридной Интеллектуальной Информационной Системы (ГИИС) [1], представленная на рис.2. Система использует гибридные методы генерации вопросов, на вход которым, в свою очередь, подаются гибридные текстовые и графовые данные.

Базовыми компонентами ГИИС являются «подсознание» системы, «сознание» системы, а также граничная модель сознания и подсознания.

Подсознание системы включает модуль извлечения текстов, модуль извлечения графов знаний, модуль извлечения концептов и связей, модуль векторизации текстов, модуль векторизации графов знаний.

Сознание системы включает модуль логического формирования вопросов, модуль формирования вопросов на основе машинного обучения, модуль логической коррекции вопросов, модуль гибридного формирования вопросов, модуль оценки качества.

Предлагаемый подход на основе ГИИС позволяет проводить эксперименты с различными вариантами архитектур интеллектуальной системы формирования вопросов [2].

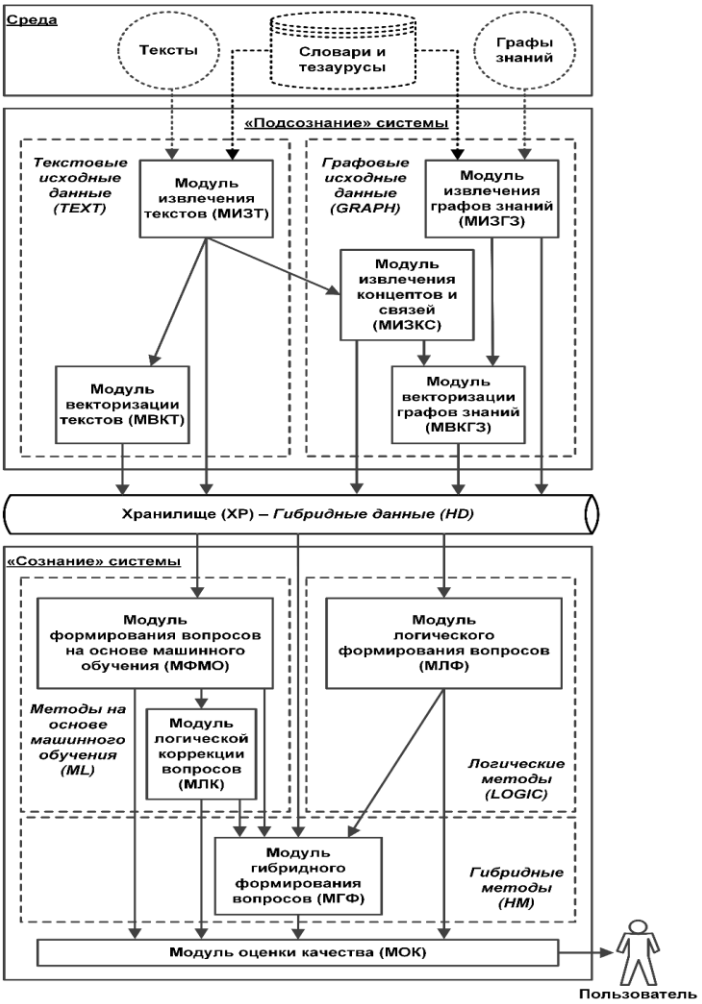


Рис.2. Архитектура гибридной интеллектуальной информационной системы генерации вопросов по тексту

4. РЕАЛИЗАЦИЯ

В данной статье предлагается рассмотреть модификацию системы из [2], позволяющую генерировать вопросы по русскоязычным текстам.

В качестве обучающей выборки используется набор данных DaNetQA, состоящий из текстов и вопросов к этим текстам, основывающихся на русскоязычном содержимом онлайн-энциклопедии Wikipedia [3].

Метрикой качества для моделей генерации последовательностей традиционно является BLEU-4 [4], метрика, предложенная для оценки качества онлайн перевода, которая учитывает пересечения n-грамм размера 4 между сгенерированным текстом и написанным вручную эталонным текстом.

В рамках исследования был изучен процесс обучения моделей нейронных сетей для формирования вопросов на основе машинного обучения.

В работе [5] была предложена архитектура, которая позволяла генерировать вопросы по графовой структуре текста, тогда как в этой работе был сделан акцент на генерацию по текстовым данным. Реализация модуля логического формирования вопросов в данной работе затронута не была.

Для обучения использовано два набора данных: DaNetQA – набор данных, содержащий в себе текст, взятый с русскоязычной онлайн-энциклопедии Wikipedia, вопрос к тексту и ответ на него, а также RuBQ [6], который изначально представляет из себя набор данных с вопросами и ответами в виде запросов к графовой базе знаний, но также имеет дополнительную информацию в виде текстов, которые иллюстрируют вопросы и содержат в себе ответы на них.

В наборе данных DaNetQA содержится 1065 текстов и вопросов к ним. Особенностью этого набора данных является то, что вопросы в нём таковы, что предполагают ответом либо да, либо нет. В наборе данных RuBQ – 2910 общих вопросов, при этом определить тексты, содержащие ответы, удалось лишь для 791 вопроса.

Примеры данных представлены ниже в таблице 1.

Архитектура кодировщик-декодировщик работает следующим образом. Кодировщик получает на вход последовательность и переводит её в некоторое векторное представление, а декодировщик после этого считывает входную последовательность в этом векторном представлении и по ней строит выходную последовательность.

Таблица 1. Примеры данных из наборов

Набор данных	Текст	Вопрос
DaNetQA	«Вы́ставочный центр» – станция Московского монорельса. Расположена между станциями «Улица Академика Королёва» и «Улица Сергея	Вднх – это выставочный центр?

	Эйзенштейна». Находится на территории Останкинского района Северо-Восточного административного округа города Москвы. Переход на станцию ВДНХ Калужско-Рижской линии. Названа в честь Всероссийского выставочного центра – названия ВДНХ с 1992 по 2014 год. 20 ноября 2004 года линия монорельса начала работать в «экскурсионном режиме» и перевезла первых пассажиров.	
RuBQ	«Моссáд» (ивр. המוסד למודיעין ולתפקידים מיוחדים, ха-Мосад ле-модиин у-ль-тафкидим меюхадим – «Ведомство разведки и специальных задач») – политическая разведка Израиля, по своему назначению и функциям сравнимая с ЦРУ.	Как звучит название политической разведки Израиля, по своему назначению и функциям сравнимой с американским ЦРУ?

В качестве модели для обучения использовалась нейронная сеть, генерирующая последовательность по заданной последовательности (Seq2Seq), архитектуры кодировщик-декодировщик [7] с механизмом внимания.

У данного подхода есть одно ограничение, обойти которое можно с помощью механизма внимания. Ограничение связано с тем, что кодировщик может работать с данными одной, фиксированной длины, тогда как какая-то часть данных может быть более значимой для алгоритма, чем другая, при этом данные считаются однозначными.

Механизм внимания в нейронных сетях, генерирующих данные, необходим для того, чтобы лучше и тоньше определять значимые для алгоритма места в исходных данных, которые могут существенно влиять на качество генерации.

В данной статье использовался алгоритм для реализации механизма внимания из статьи [8], который позволяет «обращать внимание» на отдельные части входных данных, а не на весь вектор. При каждом шаге декодировщик выбирает отдельную часть предложения.

В реализации слой нейронной сети, с помощью которого применяется внимание – это маска, которая задаёт коэффициенты значимости для каждого из нейронов во внутреннем представлении сети.

Для предобработки текстов и представления их в векторном виде были использованы методы токенизации и фильтрации текстов, были

отфильтрованы знаки препинания, мусорные слова, токенизация была проведена с отсечением самых популярных слов, чтобы они не засоряли выборку.

Было решено использовать методы, учитывающие русскоязычную специфику, то есть изменяемость по числу, падежу у существительных, по времени, числу и роду у глаголов, поэтому нормализация не проводилась.

Обучение моделей происходило на графической карте NVIDIA GeForce RTX 3060Ti с использованием CUDA 10.2, Python 3.8 и фреймворка Tensorflow версии 2.1.

Подробные параметры нейронной сети показаны на рис.3.

```
Model: "encoder_15"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_31 (Embedding)     multiple                  4696832
_________________________________________________________________
gru_31 (GRU)                 multiple                  99072
=================================================================
Total params: 4,795,904
Trainable params: 4,795,904
Non-trainable params: 0
_________________________________________________________________
Model: "decoder_16"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_32 (Embedding)     multiple                  332672
_________________________________________________________________
gru_32 (GRU)                 multiple                  148224
_________________________________________________________________
dense_121 (Dense)            multiple                  335271
_________________________________________________________________
bahdanau_attention_35 (Bahda multiple                  33153
=================================================================
Total params: 849,320
Trainable params: 849,320
Non-trainable params: 0
```

Рис.3. Параметры кодировщика и декодировщика нейронной сети

5. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

Сравнение результатов, полученных для наборов данных DaNetQA и RuBQ, а также для набора, содержащего в себе одновременно и первый, и второй наборы данных, представлено в таблице 2.

Для каждого набора данных модель обучалась в течение 60 эпох, ранняя остановка была предусмотрена в случае, если функция потерь (в данном случае, категориальная кроссэнтропия) не уменьшалась на протяжении 5 эпох.

Таблица 2. Результаты экспериментов на наборах данных

Текст	Вопрос
Пожар в театре Ирокез катастрофа произошедшая декабря года в Чикаго штата Иллинойс В результате пожара погибло более человек этот театральный пожар стал самым смертоносным за всю историю США Театр Ирокез в Чикаго был построен строительной компанией и открыт ноября года после многочисленных проблем и задержек в строительстве Несмотря на это многие американские критики назвали этот театр самым изящным и архитектурно совершенным в Чикаго Театр имел вместимость в места с тремя уровнями На первом этаже было около мест на том же уровне что и фойе и зал	**DaNetQA: разрешены ли в театрах пожары?** **RuBQ:** какой тембр голоса у Пласидо Доминго? **DaNetQA+RuBQ:** был ли на самом деле?
Абсолютное большинство населения Маршалловых Островов составляют маршалльцы	**DaNetQA:** был ли в России испанский грипп? **RuBQ: какой город является столицей республики?** **DaNetQA+RuBQ:** какой город является столицей Исландии?
Список гор Узбекистана – это перечень вершин всех горных систем, расположенных в пределах Узбекистана	**DaNetQA:** была ли что курение вредит здоровью? **RuBQ:** в каком году образована кубанская область? **DaNetQA+RuBQ:** в каком году был образован СССР?

Можно отметить, что из-за маленького количества текстов, модель показывает низкое качество, это видно по тому, что тексты вопросов слабо связаны с текстами ответов, но при этом они верно согласованы, кроме случая с последним примером, косвенно связаны с текстами, к которым предполагается их задавать.

Из плюсов можно отметить то, что нейронная сеть верно определила и выучила структуру вопросов, учитывая специфику наборов данных.

В дальнейшем планируется продолжение исследования, расширение модели для применения логических методов, а также методов на основе продукционных правил, а также применение

гибридных методов генерации вопросов для улучшения качества смыслового наполнения генерируемого алгоритмом вопроса по заданному тексту.

6. ВЫВОДЫ

Таким образом, в данной статье предложена модель на основе ГИИС для генерации вопросов по русскоязычным текстам, проведены эксперименты на основе двух наборов данных DaNetQA и RuBQ, показано, что при использовании комбинированного подхода и обучении на совокупности этих двух наборов данных, качество генерируемых вопросов выше, чем при обучении на каждом из наборов данных отдельно.

Результаты на каждом из отдельных наборов данных показывают, что наборы такого размера не дают удовлетворительного качества для генеративных моделей, и необходимо использовать агрегированные наборы данных, чтобы получить приемлемое качество.

Литература

1. Черненький В. М., Терехов В. И., Гапанюк Ю. Е. Структура гибридной интеллектуальной информационной системы на основе метаграфов // Нейрокомпьютеры: разработка, применение. – 2016. – №. 9. – С. 3-13.
2. Belyanova M. и др. Metagraph Based Approach for Neural Text Question Generation //Analysis of Images, Social Networks and Texts: 9th International Conference, AIST 2020, Skolkovo, Moscow, Russia, October 15-16, 2020, Revised Selected Papers. – Springer Nature. – С. 82.
3. Белянова М. А. и др. Автоматическая генерация вопросов на основе текстов и графов знаний //Динамика сложных систем-XXI век. – 2020. – Т. 14. – №. 4. – С. 55-64.
4. Ignatov D. I. DaNetQA: A Yes/No Question Answering Dataset for the Russian Language //Analysis of Images, Social Networks and Texts: 9th International Conference, AIST 2020, Skolkovo, Moscow, Russia, October 15-16, 2020, Revised Selected Papers. – Springer Nature. – С. 57.
5. Papineni K. и др. Bleu: a method for automatic evaluation of machine translation //Proceedings of the 40th annual meeting of the Association for Computational Linguistics. – 2002. – С. 311-318.
6. Rybin I. и др. RuBQ 2.0: An Innovated Russian Question Answering Dataset. – 2020.
7. Cho K. и др. On the properties of neural machine translation: Encoder-decoder approaches //arXiv preprint arXiv:1409.1259. – 2014.
8. Bahdanau D. и др. End-to-end attention-based large vocabulary speech recognition //2016 IEEE international conference on acoustics, speech and signal processing (ICASSP). – IEEE, 2016. – С. 4945-4949.

УДК 004.9

РАЗРАБОТКА МНОГОЯЗЫЧНОЙ НЕЙРОСЕТЕВОЙ СИСТЕМЫ МАШИННОГО ПЕРЕВОДА ДЛЯ 7 ТЮРКСКО-РУССКИХ ПАР ЯЗЫКОВ*

***Хусаинов А.Ф.**, к.т.н.*
e-mail: khusainov.aidar@gmail.com
***Сулейманов Д.Ш.**, д.т.н.*
e-mail: dvdt.slt@gmail.com
Институт прикладной семиотики
Академии наук Республики Татарстан

1. ВВЕДЕНИЕ

Область машинного перевода значительно изменилась в последние годы с развитием нейросетевых технологий. Разработка новых методов и алгоритмов обучения моделей, создание больших коллекций параллельных и моноязычных данных, а также увеличение вычислительных возможностей привели одновременно к нескольким результатам: значительно возросло качество машинного перевода для крупных мировых языков (к примеру, для английско-китайской языковой пары для предметной области новостных сообщений было заявлено о достижении уровня человеческого качества перевода [1]); были разработаны машинные переводчики для множества языков с меньшим количеством доступных лингвистических данных.

При выборе нейросетевого подхода к решению задачи машинного перевода следует иметь в виду, что до опреленного объема обучающих данных данный подход показывает меньшее качество работы, чем другие статистические алгоритмы [2]. Поэтому разработки нейросетевых систем машинного перевода для малоресурсных пар языков по-прежнему сопряжены с проблемой поиска данных для создания параллельных корпусов необходимого размера.

В данной работе мы представляем результаты работы, направленной на использование комплексного подхода к построению комплекса машинных переводчиков для 7 русско-тюркских пар языков: русско-татарской, русско-казахской, русско-чувашской,

* Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-07-00823

русско-башкирской, русско-крымскотатарской, русско-киргизской и русско-узбекской. В разделе 2 данной статьи приведены результаты работы по построению параллельных корпусов для заявленных пар языков, а также планы по использованию подхода, основанного на правилах, для унификации собранных многоязычных параллельных корпусов (на основе структурно-функциональной модели тюркских морфем [3]). В разделе 3 описываются использованные технологии обучения нейросетевых моделей переводчика. Раздел «Эксперименты» содержит результаты оценки качества работы созданных моделей для каждой из языковых пар.

2. ПАРАЛЛЕЛЬНЫЕ КОРПУСА ДЛЯ РУССКО-ТЮРКСКИХ ПАР ЯЗЫКОВ

Ключевый этапом создания нейросетевых систем машинного перевода для малоресурсных пар языков является этап сбора данных. Для решения этой задачи мы использовали несколько подходов. Прежде всего, были выделены несколько основных источников параллельной информации: новости и веб-сайты государственных организаций, переведенные книги, уже существующие корпуса. Для всех заявленных пар языков был сформирован перечень источников двуязычных текстов. Следует заметить, что для разных языковых пар разные типы источников содержат большее количество данных. Например, для крымско-татарского языка основным источником параллельных текстов на данный момент являются переведенные книги, для казахского и чувашского – существующие корпуса, для всех остальных языков проекта (киргизский, башкирский, татарский и узбекский) – двуязычные Интернет-сайты.

Для осуществления процесса сбора данных с веб-сайтов нами был разработан и внедрен полуавтоматический процесс обработки данных. Он включал в себя первоначальный ручной поиск потенциальных Интернет-ресурсов, анализ структуры сайтов, проверку наличия файлов карты сайта (sitemap.xml), поиск способов автоматического установления пар переведенных страниц. Следующим шагом было создание списка необходимых для загрузки URL-адресов. Процедура загрузки проводилась с помощью инструмента Trafilatura [4], который показал эффективность извлечения основной текстовой информации с веб-страницы для всех проанализированных веб-сайтов за исключением официального веб-сайта Министерства юстиции Киргизской Республики (для этого сайта инструмент выделения основной информации был разработан отдельно). Загруженные текстовые материалы обрабатывались с помощью

инструмента razdel [5]; в результате обработки каждая строка текстового файла содержала отдельное предложение. Последним шагом на этом этапе была фильтрация и унификация вспомогательных и непечатных элементов («°», «■», тире, дефисы).

Следующим этапом создания параллельных корпусов был этап выравнивания документов и предложений. Все загруженные и обработанные текстовые документы были преобразованы в файл формата WARC, который является стандартным форматом для веб-архивов. В зависимости от наличия систем машинного перевода для конкретного языка нами был использован один из двух подходов для выравнивания документов и предложений:

- В случае наличия системы машинного перевода для конкретной пары языков все документы веб-сайта переводились с одного языка на другой, при этом в качестве исходного языка для перевода выбирался язык с меньшим объёмом данных на этом сайте. Для языков, кроме крымско-татарского, татарского и башкирского, в качестве систем перевода были использованы системы Яндекс.Переводчик и Google Переводчик. Для татарского языка был использован переводчик Tatsoft [6]. Для башкирского – татарский переводчик Tatsoft с предварительной обработкой башкирских текстов (обработка включала преобразование башкирских букв, отсутствующих в татарском языке, к наиболее близким татарским аналогам). Такая процедура последовательного преобразования и перевода «башкирский-татарский-русский» показала достаточное качество работы для задачи выравнивания документов и предложений.
- Для крымско-татарского языка не существует доступных систем машинного перевода, поэтому мы подготовили и использовали двуязычный лексикон для поиска пар документов и сегментов. На этом этапе использовались некоторые части системы Bitextor [7] и инструмента Bleualign [8].

Последний шаг процесса создания корпусов включал создание и применение алгоритмов удаления дубликатов предложений.

Разработанные алгоритмы позволили осуществить процесс сбора данных для всех 7 языковых пар. Текущие результаты показывают, что объем собранных данных может позволить построить базовые системы машинного перевода для большинства выбранных языковых пар.

Суммарно по всем языкам подготовлен корпус объемом более 7,9 миллионов пар предложений. Краткая статистика собранных данных по языкам представлена в таблице 1.

Таблица 1. Объем собранных параллельных данных

Язык	Источники данных	Собрано пар	Русская часть	Тюркская часть
киргизский язык	8 сайтов: sti.gov.kg, www.kenesh.kg, minjust.gov.kg, novosti.kg, edu.gov.kg, mineconom.gov.kg, med.kg, ru.sputnik.kg; корпус: JW300	418 тысяч пар	6,4 млн слов	6,1 млн слов
башкирский язык	7 сайтов: bash.news, ufacity.info, glavarb.ru, bashinform.ru, bashdram.ru, house.bashkortostan.ru, pravitelstvorb.ru; печатная энциклопедия; корпус: JW300	352 тысячи пар	5,2 млн слов	4,9 млн слов
татарский язык	3 сайта: tatar-inform.tatar, tatarstan.ru, kiziltan.rbsmi.ru; корпус: JW300; существующие корпуса	2 млн пар	32,8 млн слов	31,1 млн слов
узбекский язык	5 сайтов: kun.uz, www.uzdaily.uz, www.gazeta.uz, uza.uz, xabar.uz	404 тысячи пар	7,7 млн слов	7,6 млн слов
крымско-татарский язык	1 сайт crimeantatars.club; 2 корпуса: OPUS-GNOME и OPUS-Ubuntu; печатная литература (9 книг)	26 тысяч пар	0,17 млн слов	0,16 млн слов
чувашский язык	1 существующий корпус	241 тысяча пар	2,9 млн слов	2,8 млн слов
казахский язык	1 существующий корпус WMT	4.5 млн пар	80,2 млн слов	83,9 млн слов

Перспективное направление, которое в будущем позволит увеличить объем доступных для обучения переводчика данных, заключается в разработке средств унификации параллельных корпусов для различных тюркских языков на основе структурно-функциональной модели тюркской морфемы. Этот подход должен позволить за счет использования особенностей тюркских языков значительно увеличить объем обучающих данных. В состав модели тюркской морфемы входят несколько модулей, работающих с дву- и многоязычными словарями,

система описания правил морфотактики для разных языков, а также модули морфологического анализа и синтеза слов. Нами предлагается программный модуль для перевода между тюркскими языками, который будет работать по следующему принципу: на начальном шаге происходит разбор синтаксических конструкций и отдельных слов в исходном предложении, за этим следует этап перевода отдельных элементов на целевой язык с последующим синтезом слов целевого языка на основе отдельных морфем. Ключевым аспектом, на котором было сосредоточено внимание, стал учёт многозначности, возникающий в процессе перевода. Так, для слова из исходного предложения может быть доступно несколько вариантов морфологического разбора, что в случае отсутствия системы снятия многозначности для данного языка приведет к многозначному переводу на целевой язык. Наличие нескольких многозначных слов в предложении значительно увеличит количество итоговых вариантов перевода. На текущий момент идёт разработка скрипта перевода между тюркскими языками, который сможет генерировать варианты перевода исходного предложения следующим образом:

- генерируются все возможные варианты перевода;
- для каждого случая многозначности выбирается первый вариант разбора;
- для каждого случая многозначности выбирается случайный вариант разбора;
- генерируются все возможные варианты перевода, которые затем ранжируются обученной на моноязычном корпусе статистической языковой моделью.

3. ТЕХНОЛОГИИ ОБУЧЕНИЯ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ МАШИННОГО ПЕРЕВОДА

Ключевая задача построения моделей машинного переводчика решалась на основе нейросетевого подхода. Мы использовали архитектуру нейронной сети Transformer, ключевой особенностью которой является использование механизма внимания (multihead attention).

Было проведено три серии экспериментов:

- обучение базовых моделей машинного перевода для каждой из языковых пар;
- эксперимент по использованию материалов всех родственных языков для обучения базовой «общей» модели с последующим дообучением до модели целевого языка;

- обучение единой многоязычной модели в направлении перевода с русского на тюркские языки с добавлением тега языка в исходное предложение.

Для всех экспериментов собранные параллельные корпусы были случайным образом поделены на обучающую, тестовую и валидационную выборки; минимальный размер тестовой и валидационной выборок составляет 1000 пар предложений (для крымско-татарского, киргизского, узбекского, башкирского языков), 2000 пар – для казахского, 2500 пар – для татарского и 2900 пар – для чувашского языков.

В рамках первой серии экспериментов для направлений перевода с русского на тюркский и с тюркского на русский языки были обучены ансамбли нейросетевых моделей (14 ансамблей для 7 пар языков). Каждый ансамбль состоял из 8 независимых нейросетевых моделей, 4 из которых были обучены в направлении текста справа налево. Для контроля процесса обучения использовались 3 критерия: ce-mean-words, perplexity, BLEU. В качестве критериев выхода было задано максимальное значение количества эпох обучения, равное 300 (3000 - для крымско-татарского), и максимальное количество итераций, в ходе которых не было достигнуто улучшения целевого критерия ce-mean-words, равное 5.

Вторая серия экспериментов заключалась в проверке гипотезы, согласно которой малоресурсные пары языков могут выиграть в качестве благодаря использованию предобученной многоязычной нейросети. Эта «общая» нейросеть обучается на всём массиве собранных данных для всех 7 языков. Таким образом, на начальном этапе был сформирован единый русско-тюркский корпус (случайный выбор предложений для валидационных и тестовых выборок повторно не проводился, выборки были получены объединением уже сформированных выборок по каждому из языков). На основе этого корпуса было проведено обучение одной нейросетевой модели в направлении слева направо и одной – справа налево. Затем данная модель использовалась в качестве базовой для дообучения на основе параллельного корпуса для каждой конкретной пары языков. При этом дообучение проводилось по тем же правилам, что и в рамках первого эксперимента за исключением того, что максимальные значения для количества эпох обучения были увеличены и текущие лучшие значения критериев обнулены.

Для третьего эксперимента ко всем исходным русским предложениям был добавлен тег целевого тюркского языка перевода в формате «<language_code> Исходное предложение». Ещё одно отличие

этого эксперимента заключалось в использовании подхода SentencePiece [9] к разбиению слов на части (в первых двух экспримениах деление осуществлялось с помощью алгоритма BPE [10]).

4. ЭКСПЕРИМЕНТЫ

Для оценки качества построенных систем перевода мы использовали показатель BLEU.

По результатам первого эксперимента можно отметить зависимость между качеством перевода и объемом обучающих данных. В то же время на довольно высокие значения метрики BLEU повлияло небольшое количество источников, послуживших созданию параллельных корпусов. В связи с чем автоматически составленный тестовый корпус имел высокое сходство с обучающим подкорпусом.

В таблице 2 представлены результаты значений метрики качества перевода BLEU для базовых моделей из первого эксперимента. Приведены значения BLEU, рассчитанные для ансамбля из 8 нейросетей, а также для для 4 отдельных моделей «слева направо».

Таблица 2. Значения качества перевода BLEU для базовых моделей по результатам первого эксперимента

Направление перевода	Объем обучающих данных	BLEU для ансамбля из 8 нейросетей	BLEU для 4 отдельных моделей
Русско-казахский	4 513 тыс.	48,2	45,6; 47,8; 43,9; 46,3
Казахско-русский		64,3	62,6; 62,3; 62,2; 62,2
Русско-татарский	1 994 тыс.	34,6	32,6; 32,6; 32,4; 32,6
Татарско-русский		37,5	35,3; 34,8; 35,4; 34,8
Русско-киргизский	416 тыс.	19,7	16,7; 17,6; 17,7; 18,4
Киргизско-русский		21,6	18,6; 18,0; 17,8; 18,5
Русско-узбекский	404 тыс.	32,8	30,4; 30,2; 30,1; 30,2
Узбекско-русский		34,2	31,7; 31,1; 31,2; 31,9
Русско-башкирский	351 тыс.	45,7	42,7; 43,8; 43,1; 42,6
Башкирско-русский		45,4	40,8; 40,2; 40,3; 40,0

Русско-чувашский	236 тыс.	21,9	18,1; 18,3; 18,1; 18,2
Чувашско-русский		24,8	20,6; 20,9; 20,6; 20,6
Русско-крымско-татарский	26 тыс.	13,5	12,3; 12,6; 12,5; 12,4
Крымско-татарско-русский		15,7	13,9; 13,5; 13,9; 14,3

В таблице 3 представлены результаты значений метрики качества перевода BLEU для моделей из второго эксперимента, полученных путем дообучения общетюркской модели, а также для многоязычной модели третьего эксперимента, полученной при обучении с добавлением тегов языков (отмечена в таблице как многоязычная).

Таблица 3. Результаты второго эксперимента по построению общих русско-тюркской и тюркско-русской моделей переводчика и их дообучению для конкретных языковых пар

Направление перевода	BLEU	Изменение значения BLEU относительно базовых моделей, %
Русско-казахский	47,8	+0%
Казахско-русский	61,9	-1,1%
Русско-казахский (многоязычная)	48,4	+1,3%
Русско-татарский	33,6	+3,1%
Татарско-русский	36,4	+3,1%
Русско-татарский (многоязычная)	33,2	+1,8%
Русско-киргизский	22,2	+20,7%
Киргизско-русский	25,0	+34,4%
Русско-киргизский (многоязычная)	22,5	+22,3
Русско-узбекский	33,4	+9,9%
Узбекско-русский	35,5	+11,3%
Русско-узбекский (многоязычная)	31,1	+2,3%
Русско-башкирский	45,9	+4,8%
Башкирско-русский	47,3	+15,9%
Русско-башкирский (многоязычная)	47,3	+8,0%
Русско-чувашский	28,0	+53%
Чувашско-русский	30,4	+45,4%

Русско-чувашский (многоязычная)	25,8	+41%
Русско-крымско-татарский	22,7	+80,2%
Крымско-татарско-русский	24,4	+70,6%
Русско-крымско-татарский (многоязычная)	15,0	+19%

Результаты второго эксперимента демонстрируют, что для языковых пар с наименьшим объемом обучающих данных использование предобученной «общетюркской» модели позволяет значительно улучшить качество перевода. Так, для перевода с русского на чувашский показатель BLEU вырос с 18,3 до 28,0, а с русского на крымско-татарский – с 12,6 до 22,7. Отдельно стоит отметить качество перевода единой многоязычной модели с русского на 7 тюркских языков. Для всех 7 языков качество работы единой модели превзошло качество работы базовых моделей, а для 3 языков – также и дообученных версий. При этом единая многоязычная модель имеет преимущества на этапе реализации программных сервисов для машинного перевода.

5. ЗАКЛЮЧЕНИЕ

В данной работе представлены результаты создания лингвистических материалов для 7 тюркских языков и разработки программных средств обучения нейросетевых моделей машинного перевода. Был сформирован перечень источников параллельных данных, разработаны вспомогательные программные средства, которые позволили сформировать параллельный корпус общим объемом более 7 миллионов пар предложений. Для некоторых языков, например, крымско-татарского, параллельный корпус был построен впервые; для других языков – объёмы существовавших корпусов удалось значительно увеличить. Впервые на основе методов машинного обучения были разработаны необходимые программные средства и проведены комплексные эксперименты по построению современных нейросетевых моделей перевода для группы тюркских языков.

Литература

1. Hassan H. et al. Achieving Human Parity on Automatic Chinese to English News Translation. arxiv:1803.05567. – 2018.
2. Koehn P., Knowles R. Six Challenges for Neural Machine Translation. arXiv:1706.03872. – 2017.
3. Gatiatullin A., Suleymanov D., Prokopyev N., Khakimov B. About Turkic Morpheme Portal // Proc. of 2nd Workshop CMLS'20, Kazan, November 12-13, 2020, CEUR-WS.org, online.
4. trafilatura: Web scraping tool for text discovery and retrieval. URL: https://trafilatura.readthedocs.io/.
5. razdel: rule-based system for Russian sentence and word tokenization. URL: https://github.com/natasha/razdel.
6. Русско-татарский машинный переводчик Tatsoft. URL: translate.tatar.
7. Bitextor: tool to automatically harvest bitexts from multilingual websites. URL: https://github.com/bitextor/bitextor.
8. Bleualign: an MT-based sentence alignment tool. URL: https://github.com/rsennrich/Bleualign.
9. Kudo T., Richardson J SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing. Arxiv: 1808.06226, 2018.
10. Sennrich, R., Haddow, B., Birch, A Neural Machine Translation of Rare Words with Subword Units // Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. – Vol. 1. – 2016. – P. 1715-1725.

УДК 004.934.2

САМООБУЧЕНИЕ НА НЕРАЗМЕЧЕННЫХ ДАННЫХ ДЛЯ СИСТЕМЫ РАСПОЗНАВАНИЯ ТАТАРСКОЙ РЕЧИ

***Хусаинов А.Ф.**, к.т.н.*
e-mail: khusainov.aidar@gmail.com
***Сулейманов Д.Ш.**, д.т.н*
e-mail: dvdt.slt@gmail.com
Мухаметзянов И.Р.
e-mail: iilnur.mukhametzyanov@gmail.com
ИПС Академии наук Республики Татарстан

1. ВВЕДЕНИЕ

Исследования в области автоматического распознавания речи активно развиваются по нескольким основным направлениям: системы становятся устойчивее к различным фоновым шумам, диалектам, особенностям произношения, а также разрабатываются подходы для работы с малоресурсными языками. И если несколько лет назад абсолютное большинство систем распознавания было основано на «классическом» подходе разделения на акустические модели, модель произношения и языковую модель, то в последнее время на первый план вышли end-to-end системы (E2E).

Применение E2E-систем распознавания позволяет получать более качественный результат, однако требуют большого объёма обучающих данных, который недоступен для малоресурсных языков. Одним из способов преодоления нехватки обучающих данных является предобучение системы на данных для родственных языков или использование модели, преодобученной для другого языка с большим количеством размеченных данных, например, для английского.

Кроме того, вслед за областями анализа изображений и естественно-языковых текстов в области речевых технологий активно развивается подход, при котором модели предобучаются на большом количестве неразмеченных данных. В области анализа речи такой подход был реализован в рамках модели wav2vec2, которая позволила получить для английского языка качественные результаты с минимальным количеством (от 10 минут записей) размеченных данных [1]. Суть технологии заключается в использовании большого

объёма неразмеченных данных для построения акустического представления исходных значений отсчетов речевого сигнала.

В данной работе мы описываем результаты экспериментов по построению системы распознавания татарской речи на основе технологии предварительного самообучения на неразмеченных данных. Были созданы неразмеченный аудиокорпус для татарского языка общим объемом 340 часов, сформирован аннотированный речевой корпус объемом 128 часов. Проведён эксперимент, в рамках которого многоязычная модель XLSR [2], обученная на 56 тысячах часов речи на 53 языках, была последовательно предобучена (pretrain) с использованием неразмеченных татарских данных и дообучена (fine-tune) на аннотированном татарском корпусе. Проведенное тестирование предложенной системы распознавания речи подтвердило высокое качество работы для разных типов речи (читаемой и спонтанной) и шумовых условий. Этап самообучения на неразмеченных данных позволил снизить значение ошибок распознавания WER (word error rate) для корпуса CommonVoice с 26,76 [3] до 6,48; предыдущий лучший опубликованный результат для корпуса читаемой татарской речи TatarCorpus, равный 12,89 WER [4], был снижен до 6,46.

2. ПОДХОД К ОБУЧЕНИЮ СИСТЕМЫ РАСПОЗНАВАНИЯ РЕЧИ

В данной статье используется подход с постепенным дообучением базовой многоязычной модели XLSR [2]. На первом этапе применяется подход самообучения (self-supervised learning), который заключается в решении псевдо-задачи, не требующей ручного аннотирования корпуса. В данном случае используется критерий CPC (Contrastive Predictive Coding), а модель решает задачу классификации, определяя фрагмент, следующий непосредственно за текущим, от других фрагментов [5-7]. В работе [8] показано, что характеристики аудио, выявляемые моделью в процессе решения поставленной задачи, демонстрируют устойчивость к изменению предметной области и даже языка речи. Иллюстрация работы модели из оригинальной статьи wav2vec2 [1] предствлена на рис.1.

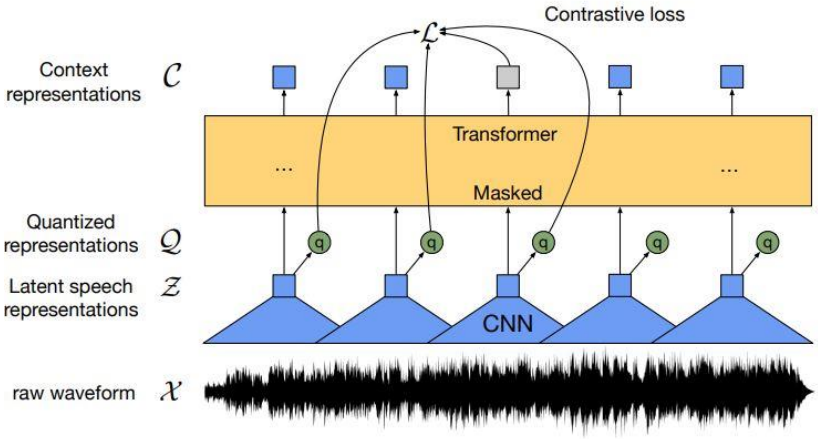


Рис.1. Иллюстрация работы модели wav2vec2, которая обучается контекстному представлению аудио-фрагментов на основе неразмеченных данных

На втором этапе используется аннотированный речевой корпус для дообучения полученной на первом этапе модели. Дообучение происходит на основе алгоритма CTC (Connectionist Temporal Classification) [5, 9]. К модели первого этапа добавляется случайно инициализированный слой размерностью, равной количеству элементов в словаре. Для случая татарского языка словарь состоит из 39 элементов: 38 букв и дополнительного символа-разделителя для слов '|'.

3. КОРПУС ЗВУЧАЩЕЙ ТАТАРСКОЙ РЕЧИ

В состав созданного корпуса неразмеченной татарской речи вошло 4 подкорпуса:

1. Подкорпус аудиокниг (читаемая речь, записанная в студийных условиях);
2. Подкорпус телевизионной трансляции (спонтанная речь, многообразие внешних шумов, фоновой музыки);
3. Подкорпус записей эфира двух радиостанций (читаемая и спонтанная речь, фоновая музыка);
4. Подкорпус научных видеолекций с платформы YouTube (слитная речь, хорошее качество записи).

На этапе предобработки из видео файлов была выделена аудиодорожка, все аудио файлы затем были приведены к единому формату: 16 бит на отсчет, 16 кГц моно PCM.

С учетом специфики исходных данных (длинные аудиокниги, 12-часовые фрагенты ТВ-эфира, 40 минутные ролики YouTube) стояла задача разделения аудио файлов на короткие фрагменты, содержащие речь. При этом желательно, чтобы каждый фрагмент содержал речь только одного диктора. Для решения этой задачи был использован инструмент Silero-VAD [10]. Выборочный анализ фрагментов показал, что модель справилась с фильтрацией музыкального контента, который присутствовал в радио и ТВ-эфире, сохранив при этом речевые сегменты с фоновой музыкой.

На основе рекомендаций разработчиков модели wav2vec2 [11] в качестве дополнительной меры были отфильтрованы короткие (меньше 4,5 секунд) и длинные (длиннее 30 секунд) аудио файлы. Итоговая статистика по количеству файлов и их продолжительности по каждому из подкорпусов представлена в таблице 1.

Таблица 1. Статистика неразмеченного корпуса татарской речи

Корпус	Исходные данные	Выделенные речевые фрагменты	Итоговый корпус после фильтрации по длительности
Аудиокниги	114:28:10 (520 файлов)	105 часов (36 712 файлов)	58 часов (17 563 файла)
ТВ-трансляции	732:59:29 (62 файла)	472 часа (263 466 файлов)	202 часа (67 065 файлов)
Записи двух радиостанций	215:36:04 (398 файлов)	146 часов (29 778 файлов)	29 часов (8 941 файл)
Научные лекции с YouTube	87:18:42 (100 файлов)	81 час (31 437 файл)	39 часов (12 764 файла)
Итоговый корпус	1 150:22:25	804 часа	328 часов

Аннотированный корпус звучащей татарской речи, который был использован для дообучения модели, состоит из 3 частей:

1. Корпус читаемой татарской речи Tatar Corpus [12];
2. Корпус размеченных фрагментов ТВ эфира [13];
3. Татарская часть корпуса CommonVoice [14].

Статистика по продолжительности каждого из подкорпусов представлена в таблице 2.

Таблица 2. Статистика размеченного корпуса татарской речи

Корпус	Длительность
Tatar Corpus	99:09:59
ТВ-трансляции	1:33:28
Common Voice	28:47:26
Итоговый корпус	129:30:53

Из итогового корпуса был выделен тестовый подкорпус общей продолжительностью 7 часов. Из корпуса читаемой речи Tatar Corpus случайным образом были выделены записи 10 дикторов (5 мужских, 5 женских) продолжительностью 1 час 37 минут. Для корпуса Common Voice нами было принято разделение на обучающую и тестовую выборки, предложенное составителями корпуса. Для подкорпуса ТВ-трансляций не было подготовлено аннотации по дикторам, поэтому выбор тестовых фрагментов осуществлялся случайным образом по всей коллекции.

4. ЯЗЫКОВАЯ МОДЕЛЬ

В качестве языковой модели для системы распознавания речи была построена 4-граммная статистическая модель на основе инструмента KenLM [15]. В качестве текстового корпуса для построения языковой модели был собран Интернет-корпус татарских текстов. Были скачаны и обработаны архивы ведущих новостных агенств, газет, журналов, сайты государтсвенных учреждений и ведомств, форумы. В состав обучающей базы были также включены части национального корпуса татарского языка «Туган тел» [16]. Суммарный объём обучающих данных составил 8 760 330 предложений, содержащих 116 миллионов словоформ.

5. ЭКСПЕРИМЕНТЫ

Эксперименты проводились на платформе fairseq [17]. В качестве архитектуры нейросети для предобучения была выбрана архитектура модели XLSR: блок кодировщика состоял из 24 слоев размерностью 1024, количество блоков механизма внимания – 16, dropout не использовался. Предобучение осуществлялось на 8 видеокартах V100 32 ГБ в течение 4 дней. На рис.2 представлены графики изменений параметров train_loss и valid_loss.

Рис.2. График изменения значений параметров train_loss и valid_loss в процессе предобучения модели на неразмеченном корпусе

Дообучение модели на размеченном корпусе осуществлялось со значением параметра learning rate – 3e-5; в качестве оптимизационного алгоритма использовался Adam [18]. Как и в оригинальной работе wav2vec2.0, в течение первых 10 тысяч итераций обучался только последний слой нейросети, в дальнейшем обучении участвовали также параметры слоев трансформера. Блок выделения признаков оставался без изменений в течение всего времени обучения.

Дообучение также проводилось на 8 видеокартах V100 32 ГБ, общее время расчетов составило 26 часов, время расчетов до получения оптимальной модели – 9 часов. На рис.3 представлены графики изменений параметров train_loss и valid_loss в процессе обучения.

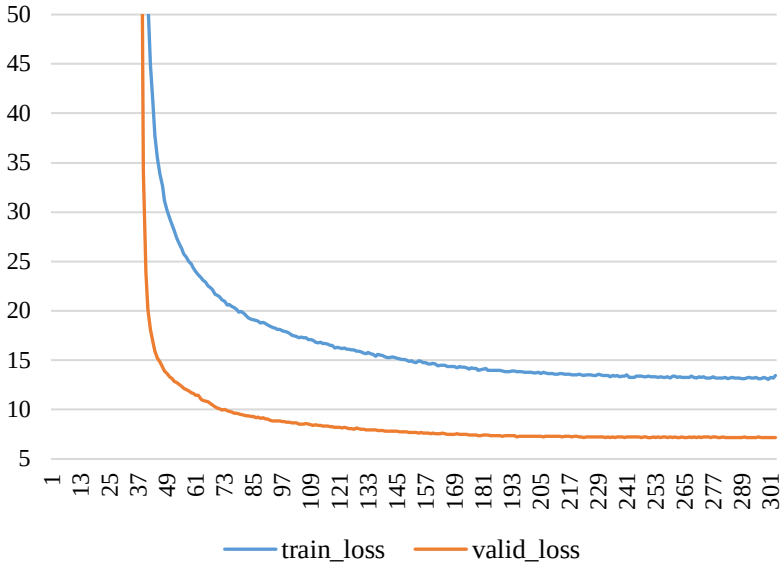


Рис. 3. График изменения значений параметров train_loss и valid_loss в процессе дообучения модели на аннотированном корпусе

Качество работы итоговой системы распознавания татарской речи было проверено на тестовом корпусе и составило 6,89 WER. Отдельно были рассчитаны значения качества распознавания для тестовой части Common Voice и Tatar Corpus с целью сравнения с другими опубликованными ранее моделями.

Лучшее значение, показанное «классической» системой распознавания слитной речи, построенной на отдельных акустических моделях, модели произношения и языковой модели, на тестовом корпусе Tatar Corpus равно 12,89 WER [4]. Предложенная в этой работе модель на том же тестовом подкорпусе показала значение 6,46 WER.

В качестве базовых значений для сравнения качества на тестовом корпусе Common Voice были взяты значения из системы [19]. Лучшее представленное там значение равняется 26,76 WER, в то время как предложенная нами система продемонстрировала значение 6.48 WER. Сводные значения результатов тестирования представлены в таблице 3.

Таблица 3. Результаты тестирования качества системы распознавания татарской речи

Тестовый корпус	Модель	Ошибки распознавания, WER
Common Voice test (читаемая речь)	**Предложенная в данной работе модель**	**6,48**
	Лучшая модель из опубликованных ранее [3]	26,76
ТВ-трансляции (спонтанная речь)	**Предложенная в данной работе модель**	**28,73**
Tatar Corpus test (читаемая речь)	**Предложенная в данной работе модель**	**6,46**
	Лучшая модель из опубликованных ранее [4]	12,89
Объединенный тестовый корпус	**Предложенная в данной работе модель**	**6,89**

6. ЗАКЛЮЧЕНИЕ

В данной работе представлены результаты экспериментов по построению системы распознавания речи с использованием подхода самообучения на неразмеченных аудио данных. Были сформированы необходимые размеченные и неразмеченные речевые корпуса. Результаты тестирования обученных моделей распознавания показали улучшение качества распознавания тестовых баз данных слитной татарской речи.

Литература

1. A. Baevski, H. Zhou, A. Mohamed, and M. Auli. Wav2vec 2.0: A framework for self-supervised learning of speech representations // In Proc. NeurIPS. – 2020.
2. Alexis Conneau, Alexei Baevski, Ronan Collobert, Abdelrahman Mohamed, and Michael Auli. Unsupervised cross-lingual representation learning for speech recognition. arXiv, 2020.
3. Wav2Vec2-Large-XLSR-53-Tatar. URL: https://huggingface.co/anton-l/wav2vec2-large-xlsr-53-tatar.
4. A. Khusainov. Recent Results in Speech Recognition for the Tatar Language // Text, Speech, and Dialogue, Lecture Notes in Artificial Intelligence. – 2017. Vol. 10415. – P. 183-191.
5. A. Baevski, M. Auli, and A. Mohamed. Effectiveness of selfsupervised pre-training for speech recognition. – 2019.
6. S. Schneider, A. Baevski, R. Collobert, and M. Auli. Wav2vec: Unsupervised Pre-Training for Speech Recognition // In Proc. Interspeech 2019. – 2019. – P. 3465-3469.

7. J. Kahn, M. Riviere, W. Zheng, E. Kharitonov, Q. Xu, P.-E.Mazare, J. Karadayi, V. Liptchinsky, R. Collobert, C. Fuegen, T. Likhomanenko, G. Synnaeve, A. Joulin, A. Mohamed, and E. Dupoux. Libri-light: A benchmark for asr with limited or no supervision. – 2019.
8. K. Kawakami, L. Wang, C. Dyer, P. Blunsom, and A. van den Oord. Learning robust and multilingual speech representations. – 2020.
9. A. Graves, S. Fernández, and F. Gomez. Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks // In Proc. of ICML. – 2006.
10. Silero VAD: pre-trained enterprise-grade Voice Activity Detector (VAD), Number Detector and Language Classifier. URL: https://github.com/snakers4/silero-vad.
11. FAIR-seq, Wav2vec 2.0 PyTorch example. URL: https://github.com/pytorch/fairseq/tree/master/examples/wav2vec.
12. Khusainov, A. Design and creation of speech corpora for the Tatar speech recognition and synthesis tasks / A. Khusainov // Proceedings of the 3rd International Conference on Turkic Languages Processing (TURKLANG-2015. (Kazan, September 17-19, 2015). – Kazan: Academy of Sciences of the Republic of Tatarstan Press, 2015. – P. 475-484.
13. Инструмент для распределенного создания аннотированных речевых корпусов / Хусаинов Айдар // Proceedings of the 8th International Conference on Turkic Languages Processing (TURKLANG-2020). (Ufa, October 18-20, 2020). – Ufa, 2020.
14. CommonVoice. URL: https://commonvoice.mozilla.org/.
15. Heafield, Kenneth. KenLM: Faster and Smaller Language Model Queries // Proceedings of the Sixth Workshop on Statistical Machine Translation. 2011. – Vol. 7. – P. 187-197.
16. Сулейманов Д.Ш., Хакимов Б.Э., Гильмуллин Р.А. Корпус татарского языка: концептуальные и лингвистические аспекты // Вестник ТГГПУ. – 2011. – № 4 (26). – С. 211-216.
17. M. Ott, S. Edunov, A. Baevski, A. Fan, S. Gross, N. Ng, D. Grangier, and M. Auli. Aairseq: A fast, extensible toolkit for sequence modeling // In Proc. of NAACL System Demonstrations. – 2019.
18. Diederik P. Kingma, Jimmy Ba. Adam: A Method for Stochastic Optimization. // ICLR-2015. – 2015.
19. CommonVoice Tatar Benchmark. URL: https://paperswithcode.com/sota/speech-recognition-on-common-voice-tatar.

УДК 004.8

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ГИБКИХ АЛГОРИТМОВ ПОИСКА РЕШЕНИЯ ДЛЯ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ РЕАЛЬНОГО ВРЕМЕНИ*

***Еремеев А.П.**, д.т.н*
e-mail: eremeev@appmat.ru
Полюшкин И.А.
e-mail: poliushkinia@yandex.ru
Сергеев М.Д.
e-mail: sergeevmd@mpei.ru
НИУ «МЭИ»

1. ВВЕДЕНИЕ

В работе дается сравнительный анализ гибких алгоритмов поиска решения, построенных в результате модификации (конвертации) алгоритмов поиска решения в пространстве состояний. Проведены компьютерные испытания данных алгоритмов с целью их применения в составе интеллектуальных систем реального времени (ИС РВ) типа интеллектуальных систем поддержки принятия решений реального времени (ИСППР РВ).

ИСППР РВ предназначены для помощи в принятии решений в условиях жестких временных ограничений [1]. В таких условиях найти оптимальное решение в установленные временные ограничения, как правило, не удаётся, а решение, найденное несвоевременно, может не иметь практической пользы. В частности, при управлении энергоблоком АЭС время, отводимое оперативно-диспетчерскому персоналу (лицам, принимающим решения, ЛПР) для нормализации возникшей нештатной ситуации, может ограничиваться одной минутой, после чего включается автоматическая система защиты [2]. Применение так называемых «гибких» алгоритмов поиска решения позволяет найти приемлемое решение за меньшее время по сравнению с временем, затрачиваемым «строгим» алгоритмом для поиска оптимального решения.

* Работа выполнена при финансовой поддержке РФФИ, проект №20-07-00498 А, 18-29-03088 МК

Гибкими (anytime) называются алгоритмы, качество найденных результатов которых улучшается с увеличением времени работы алгоритма [3]. Гибкие алгоритмы можно выполнять двумя способами: либо с заранее предусмотренным лимитом времени выполнения («контрактные вычисления»), либо с возможностью останова в любой момент времени («гибкие вычисления»). Оба этих вида алгоритмов всегда можно продолжить после того, как они были остановлены (либо по истечении лимита времени, либо из-за прерывания), с целью поиска лучшего (вплоть до оптимального) решения. Гибкий алгоритм поиска решения находит субоптимальные решения, качество которых улучшается с течением времени. Так как гибкие алгоритмы относятся к классу монотонных алгоритмов, то при предоставлении им необходимых вычислительных ресурсов нахождение оптимального решения гарантировано.

2. КОНВЕРТИРОВАНИЕ ЭВРИСТИЧЕСКИХ АЛГОРИТМОВ ПОИСКА В ПРОСТРАНСТВЕ СОСТОЯНИЙ В ГИБКИЕ АЛГОРИТМЫ

Для решения задачи поиска в пространстве состояний разработаны такие алгоритмы поиска на основе эвристической функции, как A*, IDA*, RBFS и другие [4]. Данные алгоритмы не относятся к гибким, однако, применение к ним подхода на основе методов перезапуска, продолжения или восстановления позволяет их конвертировать в гибкие.

Согласно [4] эвристическая функция называется *допустимой*, если она не переоценивает стоимость пути от любой вершины до цели. Соответственно, эвристическая функция, не обладающая данным свойством, называется *недопустимой*. При этом недопустимая функция должна быть таковой, чтобы её применение обеспечивало поиск решения за более короткое время, чем применение допустимой функции.

Метод перезапуска состоит в том, чтобы повторить алгоритм поиска после нахождения первого решения с использованием недопустимой эвристической функции. Нахождение решения стоимостью не ниже, чем уже найденное, обеспечивается с помощью отсечения тех вершин соответствующего дерева (дерева решений, ДР), для которых допустимая оценка проходящего через них решения будет не ниже, чем стоимость найденного на предыдущем шаге решения.

Метод продолжения отличается от метода перезапуска тем, что после нахождения очередного решения списки открытых и закрытых вершин ДР не очищаются.

Метод восстановления похож на метод продолжения, но с двумя важными отличиями. Первое отличие заключается в том, что, при нахождении более короткого пути к вершине ДР, которая уже содержится в открытом или закрытом списках, эта вершина не перераскрывается сразу, а откладывается в специальный список, вершины из которого переносятся в открытый список после нахождения очередного решения. Второе отличие заключается в том, что параметры алгоритма могут меняться при нахождении нового решения. Например, может изменяться вес эвристической функции.

3. ГИБКИЙ ВЗВЕШЕННЫЙ АЛГОРИТМ A* С ИТЕРАТИВНЫМ УГЛУБЛЕНИЕМ

Алгоритм A с итеративным углублением* (*Iterative-Deepening A*, IDA**) отличается от стандартного алгоритма итеративного углубления тем, что применяемым условием останова развертывания в алгоритме IDA* служит минимальная f-стоимость любого узла (вершины ДР), превышающего остановочное значение, достигнутое в предыдущей итерации, а не глубина. На рис.1 приведён псевдокод алгоритма IDA*.

В качестве функции f выбрана традиционная для оценки эвристических алгоритмов функция

$$f(n) = g(n) + w \times h(n),$$

где g и h функции оценки пройденного и оставшегося до цели пути на ДР, n – число оцениваемых (раскрываемых) вершин в ДР, w – вес эвристической функции g.

Проведенное компьютерное моделирование алгоритмов эвристического поиска с конвертацией. показало, что применение метода перезапуска к алгоритму IDA* позволяет получить *гибкий взвешенный алгоритм A* с итеративным углублением* (*Anytime Weighted Iterative Deepening A*, AWIDA**). Данный алгоритм может быть использован для поиска решения в ИСППР РВ.

```
path          current search path (acts like a stack)
node          current node (last node in current path)
g             the cost to reach current node
f             estimated cost of the cheapest path (root..node..goal)
h(node)       estimated cost of the cheapest path (node..goal)
cost(node, succ)  step cost function
is_goal(node)     goal test
successors(node)  node expanding function, expand nodes ordered by g + h(node)
ida_star(root)    return either NOT_FOUND or a pair with the best path and its cost

procedure ida_star(root)
  bound := h(root)
  path := [root]
  loop
    t := search(path, 0, bound)
    if t = FOUND then return (path, bound)
    if t = ∞ then return NOT_FOUND
    bound := t
  end loop
end procedure

function search(path, g, bound)
  node := path.last
  f := g + h(node)
  if f > bound then return f
  if is_goal(node) then return FOUND
  min := ∞
  for succ in successors(node) do
    if succ not in path then
      path.push(succ)
      t := search(path, g + cost(node, succ), bound)
      if t = FOUND then return FOUND
      if t < min then min := t
      path.pop()
    end if
  end for
  return min
end function
```

Рис.1. Псевдокод алгоритма IDA*

4. ГИБКИЙ ВЗВЕШЕННЫЙ РЕКУРСИВНЫЙ АЛГОРИТМ ПОИСКА ПО ПЕРВОМУ НАИЛУЧШЕМУ СОВПАДЕНИЮ

Одним из алгоритмов эвристического поиска в пространстве состояний с ограничением на доступную память является рекурсивный поиск по первому наилучшему совпадению [6]. Он отличается от стандартного алгоритма поиска в глубину тем, что оценивает длину пути, проходящего через раскрываемую в данный момент вершину ДР. Если данная оценка становится больше, чем оценки длин, проходящих через

альтернативные пути, то алгоритм переключается на альтернативный путь. На рис.2 представлен псевдокод данного алгоритма.

```
Input: A node n, F'(n), and a threshold B'
begin
    if n is a goal node then output solution path and exit algorithm
    if Successors(n) = ∅ then return ∞
    foreach n_i ∈ Successors(n), i = 1, 2, ···, |Successors(n)| do
        g(n_i) ← g(n) + c(n, n_i),  f'(n_i) ← g(n_i) + w × h(n_i)
        if f'(n) < F'(n) then F'(n_i) ← max{F'(n), f'(n_i)}
        else F'(n_i) ← f'(n_i)
    sort n_i in increasing order of F'(n_i)
    if |Successors(n)| = 1 then F'(n_2) ← ∞
    while F'(n_1) < ∞ and F'(n_1) ≤ B' do
        F'(n_1) ←RBFS(n_1, F'(n_1), min{B', F'(n_2)})
        insert n_1 in sorted order of F'(n_i)
    return F'(n_1)
end
```

Рис.2. Псевдокод взвешенного рекурсивного алгоритма поиска по первому наилучшему совпадению

Псевдокод предлагаемого гибкого взвешенного рекурсивного алгоритма поиска по первому наилучшему совпадению (*Anytime Weighted Recursive Best-First Search, Anytime-WRBFS*) представлен на рис.3.

```
Input: A node n, F(n), and a threshold B'
begin
    if n is a goal node then return f(n)
    if Successors(n) = ∅ then return ∞
    foreach n_i ∈ Successors(n), i = 1, 2, ···, |Successors(n)| do
        g(n_i) ← g(n) + c(n, n_i),  f(n_i) ← g(n_i) + h(n_i)
        if n_i is a goal node and f(n_i) < f(incumbent) then
            incumbent ← n_i
            save (or output) incumbent solution path
        if f(n_i) ≥ f(incumbent) then F(n_i) ← ∞
        else if f(n) < F(n) then F(n_i) ← max{F(n), f(n_i)}
        else F(n_i) ← f(n_i)
    sort n_i in increasing order of F(n_i)
    if |Successors(n)| = 1 then F(n_2) ← g(n_2) ← ∞
    while F(n_1) < f(incumbent) and g(n_1) + w × (F(n_1) − g(n_1)) ≤ B' do
        F(n_1) ←Anytime-WRBFS(n_1, F(n_1), min{B', g(n_2) + w × (F(n_2) − g(n_2))})
        insert n_1 in sorted order of F(n_i)
    return F(n_1)
end
```

Рис.3. Псевдокод гибкого взвешенного рекурсивного алгоритма поиска по первому наилучшему совпадению

5. КОМПЬЮТЕРНЫЕ ИСПЫТАНИЯ АЛГОРИТМОВ

Качество найденного решения определяется как отношение стоимости найденного решения к стоимости оптимального решения.

Зависимость качества решения от времени поиска для алгоритмов *WRBFS* и *AWIDA** при $w = 2$ представлена на рис.4.

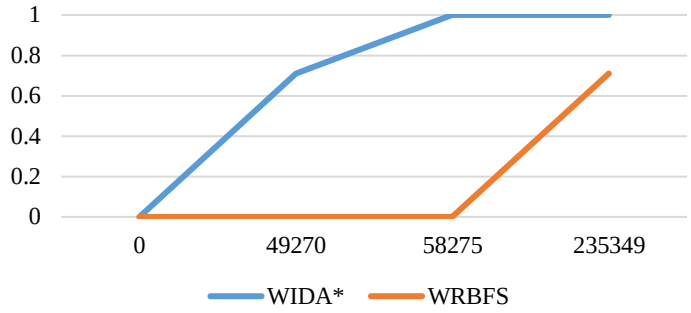


Рис.4. Сравнение алгоритмов при $w = 2$ (по оси x отложено время выполнения алгоритма)

Зависимость качества решения от времени поиска для алгоритмов *WRBFS* и *AWIDA** при $w = 3$ представлена на рис.5.

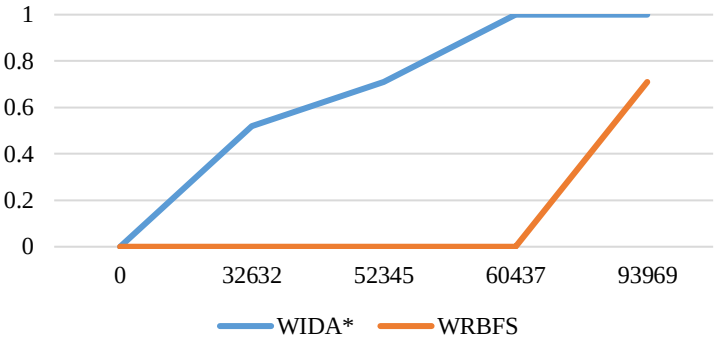


Рис.5. Сравнение алгоритмов при $w = 3$ (по оси x отложено время выполнения алгоритма)

Зависимость качества от числа раскрытых вершин для алгоритмов *WRBFS* и *AWIDA** при $w = 2$ представлена на рис.6.

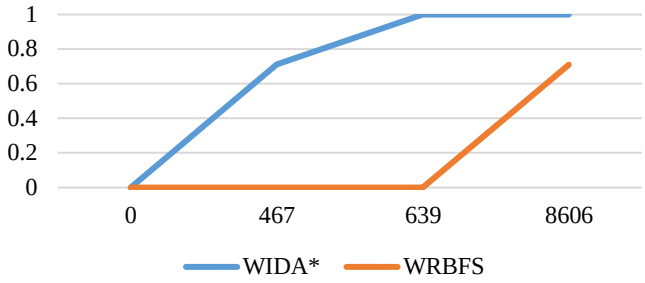


Рис.6. Сравнение алгоритмов при *w = 2* (по оси *x* отложено число раскрываемых при поиске решения вершин ДР)

Зависимость качества от числа раскрытых вершин для алгоритмов *WRBFS* и *AWIDA** при *w = 3* представлена на рис.7.

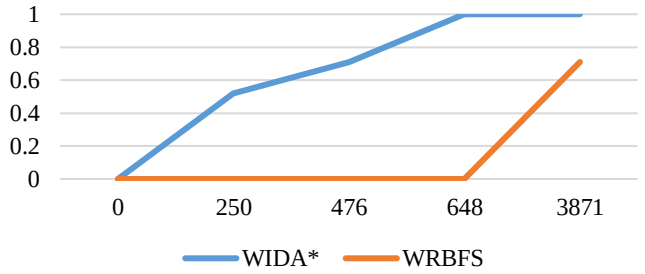


Рис.7. Сравнение алгоритмов при *w = 3* (по оси *x* отложено число раскрываемых при поиске решения вершин ДР)

В результате испытаний установлено, что предложенный гибкий алгоритм *AWIDA** находит оптимальное решение за время, меньшее, чем время работы алгоритма WRBFS. При этом гибкий алгоритм WRBFS на модельной задаче проигрывает не только в скорости поиска решения, но и в качестве.

Гибкий алгоритм *AWIDA* также показал лучшее качество результата*, чем *WRBFS*, при меньшем раскрытии вершин для *w = 2* и *w = 3*.

Оба этих алгоритма также могут быть включены в состав алгоритмов поиска решения ИСППР РВ.

6. ЗАКЛЮЧЕНИЕ

В плане разработки интегрированной инструментальной среды для реализации ИС РВ типа ИСППР РВ был разработан, реализован и испытан

(протестирован) ряд гибких алгоритмов поиска решения, являющихся конвертацией (модификацией) известных и активно используемых эвристических алгоритмов поиска решения.

В работах [7-9] обоснована с применением компьютерного моделирования перспективность применения в ИСППР РВ и ИС РВ в целом гибких алгоритмов поиска решения в сочетании с алгоритмами машинного обучения на примере алгоритмов обучения с подкреплением на основе темпоральных различий. Гибкие алгоритмы могут быть применены для прогнозирования и оценки получаемых с помощью обучения с подкреплением решений для выявления и предоставления ЛПР наиболее предпочтительного согласно заданным критериям (например, требуемому для нормализации ситуации времени).

Литература

1. Вагин В.Н., Еремеев А.П. Некоторые базовые принципы построения интеллектуальных систем поддержки принятия решений реального времени // Известия РАН. Теория и система управления. – 2001. – № 6. – С. 114-123.
2. Башлыков А.А. Основы конструирования интеллектуальных систем поддержки принятия решений в атомной энергетике: учебник /А.А. Башлыков, А.П. Еремеев. - М.: ИНФРА-М. – 2017.
3. Zilberstein S. Using anytime algorithms in intelligent systems. // In Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, 1996. – Pp. 528-544.
4. Рассел С., Норвиг П. Искусственный интеллект: современный подход, 2-е изд.: Пер. с англ. – М.: Издательский дом «Вильямс». – 2006.
5. Hart P. E., Nilsson, N. J., Raphael, B. A Formal Basis for the Heuristic Determination of Minimum Cost Paths // IEEE Transactions on Systems Science and Cybernetics SSC4. – 1968. – № 2. – P. 100-107.
6. Hansen E. A., Zhou R. Anytime Heuristic Search // Journal of Artificial Intelligence Research 28. – 2007. – P. 267-297.
7. Еремеев А.П., Кожухов А.А. Реализация методов обучения с подкреплением на основе темпоральных различий и мультиагентного подхода для интеллектуальных систем реального времени // Программные продукты и системы. – 2017. – № 1. – С. 28-33.
8. Eremeev A.P., Kozhukhov A.A., and Gerasimova A.E. Implementation of the Real-Time Intelligent System Based on the Integration Approach //Advances in Intelligent Systems and Computing 1156 / Sergey Kovalev, Valery Tarassov, Vaclav Snasel, Andrey Sukhanov Editors. Proceedings of the Fours International Scientific Conference “Intelligent Information Technologies for Industry” (IITI’19). Springer Nature Switzerland AG. – 2020.– P. 99-108.
9. Еремеев А.П., Герасимова А.Е. Применение нейросетевого подхода к методам обучения с подкреплением // VIII Международная научно-практическая конференция «Нечеткие системы, мягкие вычисления и интеллектуальные технологии» НСМВИТ-2020 (29 июня – 1 июля 2020 г., г. Смоленск, Россия). Труды конференции. В 2-х томах. Т 1. – Смоленск: Универсум. – 2020. – С. 135-143.

УДК 004.8, УДК 004.896

ПРИМЕНЕНИЕ ТЕХНОЛОГИЙ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ОБНАРУЖЕНИЯ АТАК НА ОТКАЗ В ОБСЛУЖИВАНИИ

***Соловьев Д.В.**, к.т.н.*
e-mail: 9218964588@mail.ru
СПбГУТ им. проф. М.А. Бонч-Бруевича
***Гатчин Ю.А.**, д.т.н.*
e-mail: gatchin1952@mail.ru
***Поляков В.И.**, к.т.н.*
e-mail: v_i_polyakov@mail.ru
Университет ИТМО

1. ВВЕДЕНИЕ

Как показывает практика, распределенные атаки на отказ в обслуживании – серьезный инструмент в руках злоумышленников, успешным результатом которого для организации-жертвы является потеря денег, репутации, а возможно даже и кража данных в случаях, когда атака является фоновой для другого серьезного киберпреступления [1]. Никого не удивит тот факт, что в сети интернет уже давно существуют ресурсы, позволяющие заказать распределенную атаку, где любой желающий может создать себе личный кабинет, с помощью которого можно управлять силой атаки, распределением по времени и т.д. Перед непосредственно самой атакой злоумышленник сканирует сеть с целью нахождения потенциальных будущих участников для атаки, так называемых компьютеров-зомби, они же – участники сети ботнета [2]. Найденные незащищенные узлы сети включаются в ботнет посредством эксплуатации уязвимостей в программном обеспечении, например, уязвимостей самой операционной системы, прикладных программ, используемых протоколов передачи данных, интернет-браузеров и т.п. Нередко обычные пользователи сети, сами ничего не подозревая, становятся участниками ботнета, играя, например, в интернет-игры. Получив так называемый «рутовый» доступ к удаленной машине, злоумышленник устанавливает троянские программы, которые запускаются фоном и, как правило, совершенно не заметны для рядовых пользователей компьютеров. В свою очередь это программное обеспечение ждет

команды и цели для начала проведения DDoS-атаки. Кроме этого, в последнее время с развитием устройств интернета-вещей (IoT), умные устройства так же могут являться «хостами» или источниками распределенной атаки на отказ в обслуживании. Как показывает практика, хакерам даже не нужно прикладывать много сил для получения полного удаленного доступа к таким устройствам, так как эти устройства либо работают с несмененным паролем администратора по умолчанию, либо имеют крайне слабые пароли, подбираемые методом полного перебора максимум за несколько часов [3-5].

2. МЕХАНИЗМЫ ПРОВЕДЕНИЯ АТАК НА ОТКАЗ В ОБСЛУЖИВАНИИ

В настоящее время существует несколько механизмов реализации атак на отказ в обслуживании, приведем некоторые из них:

1. HTTP-флуд. Механизм основан на переполнении полосы пропускания сервера с помощью ping-запросов. Является самым примитивным механизмом и результативен только в случае, если канал связи сервера намного уже канала атакующего.

2. ICMP-флуд. Данный механизм является одним их самых опасных и приводит к негативным последствиям практически в 100% случаев. Сущность механизма кроется в использовании усиливающей сети, по которой широковещательно рассылается пинг-запрос, где в адресе отправителя такого запроса указывается адрес жертвы. В результате такой атаки все узлы сети пришлют ответ на поступивший запрос, перегрузив тем самым ресурсы машины-жертвы.

3. UDP-флуд. Механизм проведения такой атаки аналогичен ICMP-флуду, однако в качестве транспортного протокола используется соответственно протокол UDP. Результатом атаки является практически полное насыщение полосы пропускания и, следовательно, отказ в обслуживании легитимных пользователей.

4. SYN-флуд. Механизм такой атаки основывается на особенностях в установлении TCP-соединений и может быть назван «механизмом трех рукопожатий». Смысл метода в следующем: на компьютер жертвы посылаются пакеты SYN с подделанным и несуществующим адресом отправителя пакета. Понятно, что ответа на такой поддельный запрос никогда не поступит, что приводит к постановке такого соединения в очередь. Выполнив тысячи или десятки тысяч таких запросов, хакер добьется переполнения сетевого буфера на удаленной машине, что с большой вероятностью приведет к отказу удаленной системы.

5. Атака на отказ в обслуживании посредством тяжелых запросов. Как следует из названия, цель такой атаки –перегрузить процессор (процессоры) удаленной системы тяжелыми вычислениями. Как правило, в качестве таких нагрузок используются вычисления сложной математики в циклах по сотни тысяч итераций.

3. ПРОВЕДЕНИЕ ЭКСПЕРИМЕНТА

Экспериментально было установлено, что совершенно бесполезно анализировать весь поступающий трафик для определения факта наличия или отсутствия атаки, а достаточно выделения так называемых сигнатур атак. Например, для атак на отказ в обслуживании, для которых транспортной средой является протокол TCP, достаточно лишь анализировать некоторые поля из заголовков IP-пакетов. Опытным путем было установлено, что наиболее оптимальными с точки зрения результативности решений для анализа являются поля в пакете IP: «Identification», «Flags», «Source IP», «Source port number» и «Destination port number». Как правило, все IP-пакеты, сгенерированные злоумышленником, сильно фрагментированы, имеют набор одинаковых IP-адресов источников с одинаковыми портами назначения и портами источника, что и обусловило наш выбор данных полей для анализа и формирования обучающих выборок. Однако в некоторых изученных нами работах [6-8] предлагалось анализировать 55 байт пакета IP, что представляет собой 5 байт из заголовка (IP-адрес с номером порта) плюс 50 байт полезной нагрузки, что, на наш взгляд, является тяжелой вычислительной задачей, требующей больших вычислительных ресурсов и достаточно длительного времени для получения качественного решения. На наш взгляд, анализ первых 50 байт полезной нагрузки (первых 50 байт тела пакета) никак не укажет на факт наличия или отсутствия атаки, а только потребует дополнительных затрат вычислительных мощностей.

Далее мы приступили к выбору структуры искусственной нейронной сети (ИНС) и алгоритма обучения, наиболее подходящих под нашу задачу. В ходе анализа работ [9-12] было решено использовать структуру многослойного персептрона с одним скрытым слоем, активационную функцию нейронов скрытого слоя – сигмоиду и алгоритм обучения обратного распространения ошибки. Структура ИНС представлена на рис.1.

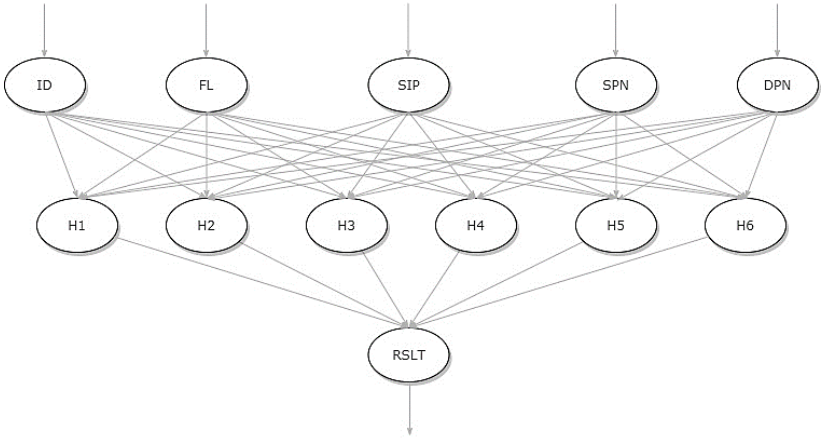


Рис. 1. Схема ИНС

Структура ИНС состоит из 5 нейронов во входном слое (нейроны: ID, FL, SIP, SPN, DPN), 6 нейронов в скрытом слое (H1-H6) и одного нейрона в выходном слое (RSLT). Из-за использования сигмоиды в качестве функции активации для нейронов скрытого слоя, входные данные ID, FL, SIP, SPN, DPN нормировались в интервал (0, +1). В свою очередь сигнал выходного нейрона определял наличие (+1) или отсутствие атаки (0). Количество нейронов в скрытом слое подбиралось экспериментально, и было установлено, что 6 нейронов оптимально для получения качественного решения за приемлемое время.

Для проведения эксперимента был собран виртуальный стенд, представленный на рис.2. Сбор всех экспериментальных данных производился на одном компьютере благодаря использованию средств виртуализации VMware Workstation. В качестве машины-жертвы использовалась виртуальная машина под управлением ОС Ubuntu Server с веб-сервером Nginx, в качестве атакующей машины использовалась виртуальная машина с установленной ОС Kali Linux с инструментом для проведения стресс-тестирования веб-серверов SlowHTTPTest. Так же была создана виртуальная машина-арбитр под управлением ОС Ubuntu Desktop для наблюдения за результативностью проведения атаки и записи сетевого трафика.

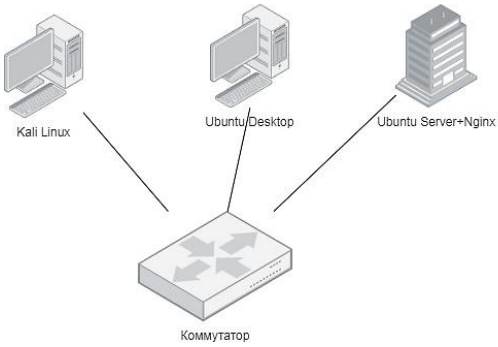


Рис. 2. Схема экспериментального стенда

В начале с виртуальной машины Kali Linux запускалась атака Slowloris для генерирования DoS-трафика, а затем запускался скрипт для генерирования легитимного трафика. Трафик записывался и анализировался с помощью сетевого снифера Wireshark. Далее формировалась обучающая выборка. Выборка представляла собой массив данных размерностью 1000 строк на 6 столбцов. Выборка была сформирована из записанного трафика путем выделения нужных полей пакета IP и перевода данных в десятичную систему счисления с нормированием в интервал (0, +1). Каждая нечетная строка выборки содержала данные, выделенные из зараженного трафика, а каждая четная – данные из легитимного трафика. Было решено 85% выборки использовать для обучения ИНС и 15% – для тестирования. После обучения ИНС на выборке из 850 строк при тестировании ИНС обнаруживала зараженные трафик с вероятностью 97%, что является достаточно качественным результатом.

4. СТРУКТУРА АВТОМАТИЗИРОВАННОЙ НЕЙРОСЕТЕВОЙ СИСТЕМЫ ОБНАРУЖЕНИЯ АТАК

Нами была разработана структура автоматизированной нейросетевой системы обнаружения атак на отказ в обслуживании (АНСОА). Структура показана на рис.3. Модуль обнаружения атак – основной нейросетевой вычислительный блок в данной структуре, он в реальном времени идентифицирует тип атаки на отказ в обслуживании и вырабатывает решение по противодействию, формируя рекомендации для системного администратора информационной системы с занесением данных об атаке в базу знаний системы. Модуль слежения взаимодействует программно как с сетевым интерфейсом напрямую,

так и с журналом событий сервера. Задача модуля реагирования состоит в выработке такого управляющего воздействия на информационную систему, которое приводит к нивелированию последствий атаки. Модуль управления данными отвечает за сохранение данных, полученных в результате функционирования АНСОА. Управляющий модуль реализует управление и согласование всех остальных модулей АНСОА.

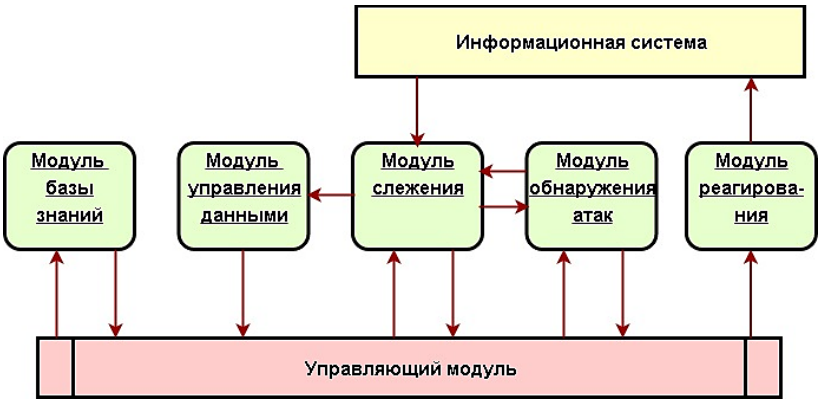


Рис.3. Структура АНСОА

5. ЗАКЛЮЧЕНИЕ

Как показано выше, распределенные атаки на отказ в обслуживании – крайне опасный вид атак на информационную систему. Актуальной задачей является их своевременное обнаружение для возможности оперативного противодействия. Успешное проведение таких атак опасно для организаций финансовыми и репутационными потерями. Технологии искусственного интеллекта в общем и искусственные нейронные сети в частности, позволяющие решать задачи не поддающиеся или слабо поддающиеся формализации – эффективный инструмент для решения задач по обнаружения аномалий в сетевом трафике, а в нашем случае наличия «зараженного» трафика с атакой. Результаты нашей работы показывают возможность разработки и успешного применения автоматизированных нейросетевых систем обнаружения атак для оперативного выявления факта наличия атаки и применения соответствующих механизмов противодействия.

Литература

1. Отчет «Лаборатории Касперского»: «DDoS-атаки в IV квартале 2020 года» [Электронный ресурс]. – Режим доступа: https://securelist.ru/ddos-attacks-in-q4-2020/100469/ (дата обращения: 11.04.2021).
2. Статья «Свободной энциклопедии Википедии»: «Ботнет» [Электронный ресурс]. – Режим доступа: https://en.wikipedia.org/wiki/Botnet (дата обращения: 11.04.2021).
3. Статья «Хабра»: «Взлом камер: векторы атак, инструменты поиска уязвимостей и защита от слежки» [Электронный ресурс]. – Режим доступа: https://habr.com/ru/company/ivideon/blog/443462/ (дата обращения: 11.04.2021).
4. Статья «Cnet»: «Why it was so easy to hack the cameras that took down the web» [Электронный ресурс]. – Режим доступа: https://www.cnet.com/how-to/ddos-iot-connected-devices-easily-hacked-internet-outage-webcam-dvr/ (дата обращения: 11.04.2021).
5. Статья «Известий»: «Хакеры получили доступ к 15 тысячам камер видеонаблюдения в Москве» [Электронный ресурс]. – Режим доступа: https://iz.ru/1079917/2020-10-29/khakery-poluchili-dostup-k-15-tysiacham-kamer-nabliudeniia-v-moskve (дата обращения: 11.04.2021).
6. Тарасов Я.В. Метод обнаружения низкоинтенсивных DDoS-атак на основе гибридной нейронной сети // Известия ЮФУ. Технические науки. – 2014. С. 47-57.
7. Тарасов Я.В., Абрамов Е.С., Тумоян Е.П. Нейросетевой метод обнаружения низкоинтенсивных атак типа «отказ в обслуживании» // Известия ЮФУ. Технические науки. – 2016. – С. 58-71.
8. Тарасов Я. В. Исследование применения нейронных сетей для обнаружения низкоинтенсивных DDoS-атак прикладного уровня // Вопросы кибербезопасности. – 2017. – № 5 (24). – С. 23-28.
9. Булдакова Т.И. Нейросетевая защита ресурсов автоматизированных систем от несанкционированного доступа // Наука и образование: научное издвание МГТУ им. Н.Э. Баумана. – 2013. – №5. – С. 269-278.
10. Saied A., Overill R.E., Raszik T, Detection of known and unknown DDoS attacks using Artifical Neural Networks // Neurocomputing. – 2016. – 172. – P. 385-393.
11. Igor H. [et all.]. Application of Neural Networks in Computer Security // Procedia Engineering. – 2014. 69. – P.1209-1215.
12. Alfantookh A.A. DoS Attacks Intelligent Detection using Neural Networks // Journal of King Saud Universite – Computer and Information Sciences. – 2006. – Vol. 18. – P. 31-51.

УДК 004.8

РАЗРАБОТКА И ТЕСТИРОВАНИЕ СИСТЕМЫ МОНИТОРИНГА ДВИЖЕНИЯ ПЕШЕХОДОВ НА ПЛАТФОРМЕ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ

***Меженин А.В.**, к.т.н.*
e-mail: mejenin@mail.ru
***Поляков В.И.**, к.т.н.*
e-mail: v_i_polyakov@mail.ru
Университет ИТМО
***Извозчикова В.В.**, к.т.н.*
e-mail: viza-8.11@mail.ru
Оренбургский государственный университет
***Зыков А.Г.**, к.т.н.*
e-mail: zykov_a_g@mail.ru
Университет ИТМО

1. ВВЕДЕНИЕ

Распространение систем автоматического слежения, в том числе систем мониторинга или подсчета пешеходов, стало возможно благодаря развитию технологий распознавания образов на изображении. Такие системы находят применение в различных областях - подсчет посетителей торговых площадок, сбор статистики посещеня массовых мероприятий и культурных объектов.

Как правило, такие системы используют веб-камеру и встроенное программное обеспечение. Однако предлагаются системы, обрабатывающие входное видео уже после съемки, на отдельно расположенных персональных компьютерах. В таких случаях важным является достижение разумного компромисса между точностью обнаружения и подсчета и быстродействием работы программы. В данной работе рассматриваются существующие решения для мониторинга движения пешеходов, их особенности, и предлагается подход на основе использования сверточных нейронных сетей, позволяющий повысить точность работы, с сохранением приемлемой скорости работы.

С ростом числа предлагаемых систем мониторинга, важным становится вопрос оценки их качества. Существующие наборы тестов

(data sets) не всегда удовлетворяют необходимым требованиям. Создание тестовых видео требует значительных ресурсов: обеспечение необходимого количества людей, выбор места съемки, различные конфигурации расположения камеры. В данной работе предлагается подход к тестированию систем мониторинга движения пешеходов и классификации объектов на основе использования синтетических наборов данных (виртуально построенных наборов видео данных), позволяющий сократить затраты на создание тестовых видео.

2. СОВРЕМЕННЫЕ СИСТЕМЫ МОНИТОРИНГА ДВИЖЕНИЯ ПЕШЕХОДОВ

В настоящий момент существует большое количество приложений, предназначенных для мониторинга и подсчета пешеходов на видео.

Комерческие решения, выпускаемые производителями специализированных камер. Macroscop – система, которая позволяет определять количество вошедших и вышедших посетителей при помощи IP-камер видеонаблюдения. Ключевые возможности данного решения – определение количества вошедших и вышедших посетителей при помощи IP-камер как через один, так и через несколько входов и подсчет людей в движущихся группах [1]. Domination – приложение, которое осуществляет статистический подсчет посетителей, анализ пиковой нагрузки. Определяет количество вошедших и вышедших людей, показывает отчет в виде графиков [2].

Решения, запускаемые при помощи различного вспомогательного оборудования. Система CleverCamera работает с IP-камерами фирм Axis, SpezVizion, Beward и других производителей, по протоколам ONVIF, RTSP. Также система включает в себя сервера вычисления и обработки, на которые устанавливается ПО [3]. Компания SimbirSoft предланает решение на плате UpBoard [4]. Реализованный алгоритм, включает в себя: выделение движущихся объектов из общего видеопотока; сегментация объектов и поиск людей среди них; разделение групп на отдельных людей; трекинг людей; подсчет прохождения посетителей через определенные ворота.

Решения с открытым исходным кодом. Footfall – система, реализованная на C++, использующая RaspberryPi, PiCam и openFrameworks [5]. Overhead-camera-people-counter – решение на языке C++, с использованием библиотеки OpenCV [6]. Решения с открытым исходным кодом – «People-detecting» [7] и «People_counter» [8]. Приложение «People-detecting», использует алгоритмы: *accumulateWeighted*. *Pedestrian, FollowMovement*. Первый, наиболее

простой из трех, обладает низкой точностью. Метод *Pedestrian*, использует предварительно подготовленную модель гистограммы ориентированных градиентов (HOG) и метод опорных векторов (SVM) [9]. Показывает более точные результаты, но потребляет больше ресурсов и работает медленно на обычных CPU. Может быть использован, если обработка происходит на GPU. Метод FollowMovement. В данном решении реализован метод из статьи «PeopleCounterwithOpenCVPython» [10]. Алгоритм включает выделение фона, фильтрацию, поиск контуров, наиболее важный шаг - слежение за движениями. Результат является достаточно точным, метод обладает приемлемой производительностью. Приложение «People_counter» для обнаружения людей используется классификатор на основе гистограмм ориентированных градиентов.

Апробация существующих решений. Для анализа были выбраны Особенности этих решений: People-detecting включает в себя операции выделение фона, фильтрацию, поиск контуров и слежение за движениями, основанную на методах OpenCV. В алгоритме

Было произведено тестирование перечисленных решений на видео, снятых с разными условиями съемки. Результаты апробации рассмотренных методов представлены на рис.1 и рис.2.

Тестирование трех выбранных решений позволяет сделать следующие выводы: в большинстве случаев метод FollowMovement обнаруживает и подсчитывает большое количество лишних фрагментов, не являющихся людьми, метод accumulateWeighted обнаруживает меньшее количество людей, чем имеется в кадре, даже в случае с небольшим количеством людей (1-2). Наиболее точные результаты показал метод Pedestrian, использующий предварительно подготовленную модель гистограммы ориентированных градиентов (HOG) и метод опорных векторов (SVM). Однако скорость работы этого метода является самой низкой и представляется невозможным использование данного метода без специального мощного оборудования.

Таким образом, все рассмотренные решения обладают определенными недостатками – большое количество ошибок при распознавании или крайне медленная скорость работы при отсутствии специализированного оборудования.

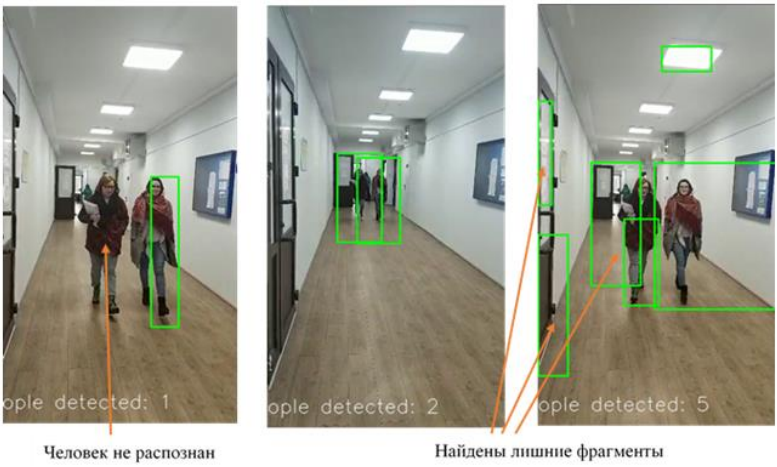


Рис.1. Результаты мониторинга с малым количеством людей

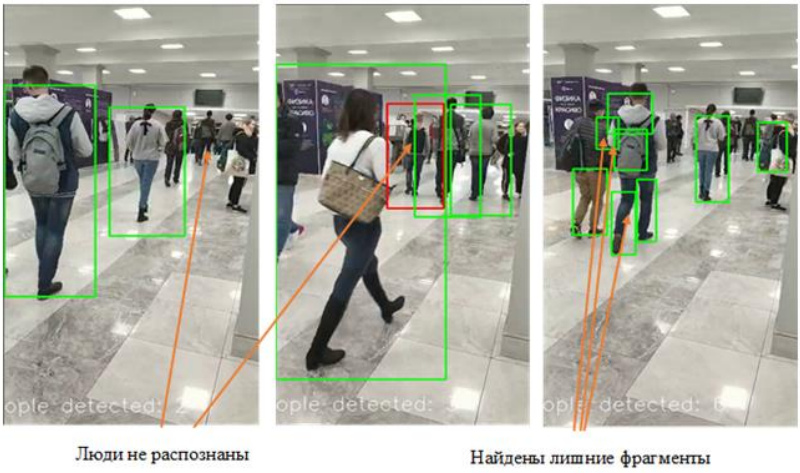


Рис.2. Результаты мониторинга с большим количеством людей

3. СВЕРТОЧНЫЕ НЕЙРОННОЙ СЕТИ В СИСТЕМАХ МОНИТОРИНГА

Существует множество средств, выполняющих распознавание образов на изображении. К ним относятся, в первую очередь, различные методы машинного обучения, включающие в себя классификацию, кластеризацию, нейронные сети и комбинации нескольких

методов [11]. В последнее время использование нейронных сетей становится все более успешным. Для анализа изображений в первую очередь предназначены сверточные нейронные сети. Основной особенностью сверточных нейронных сетей является чередование сверточных слоев и слоев подвыборки (poolinglayers, или слой пулинга), по аналогии с чередованием, так называемых простых и сложных клеток зрительной коры [12]. Суть операции свертки заключается в пошаговом перемножении фрагментов изображения на некоторую матрицу – ядро свертки.

Свертка – это первый слой, направленный на выделение признаков из входного изображения. Свертка сохраняет связь между пикселями, выделяя признаки на маленьких квадратах входного изображения. На вход операции свертки подаются фрагменты изображения и ядро свертки, или фильтр [13]. В общем виде операцию свертки можно представить в виде следующей формулы:

$$(f * g)[m,n] = \sum_{k,l} f[m-k, n-l] * g[k,l] \quad (1)$$

где f – исходная матрица изображения, g – ядро свертки. Применяя операцию свертки с различными фильтрами, можно выполнять такие операции, как обнаружение границ объектов, размытие и повышение резкости. Слой пулинга выполняет уплотнение фрагментов изображения, при котором каждая непересекающаяся группа пикселей уплотняется до одного пикселя. Существует несколько алгоритмов выбора – MaxPooling (функция максимума), AveragePooling (функция среднего), SumPooling (функция суммы). Смысл пулинга заключается в сокращении избыточной информации и сохранении только важных признаков для дальнейшей обработки, так же это позволяет избежать переобучения. Чередование сверточных слоев и слоев пулинга повторяется до тех пор, пока не будет достигнут необходимый результат. Далее результат последовательной свертки и пулинга подается на вход полносвязной нейронной сети, которая тоже может состоять из нескольких слоев – FullyConnectedlayers. Наконец, для итоговой классификации используется функция активации нейрона – некоторая нелинейная функция, которая определяет выходной сигнал по набору входных сигналов. Очень часто в качестве функции активации используется сигмоидальная функция:

$$f(s) = \frac{1}{1+e^{-S}} \quad (2)$$

YOLO (You Only Look Once) – система обнаружения объектов в реальном времени. Представляет собой сверточную нейронную сеть, относящуюся к классу SingleShot детекторов. При помощи YOLO можно выполнять распознавание 80 классов объектов, включая людей,

машины, предметы мебели и т.д. Благодаря высокой скорости работы данная архитектура подходит для обработки и обнаружения объектов на видео [14]. YOLOv3 – это усовершенствованная версия архитектуры YOLO. Она состоит из ста шестисвёрточных слоев и лучше детектирует небольшие объекты по сравнению с её предыдущей версией – YOLOv2. Основная особенность YOLOv3 состоит в том, что на выходе есть три слоя, каждый из которых рассчитан на обнаружения объектов разного размера – использована идея архитектуры FeaturePyramidNetworks. Также количество опорных прямоугольников увеличено до 9, а также используется более точная функция классификации [15]. В разрабатывемом приложении была использована библиотека компютерного зрения OpenCV, в последних версиях которой добавлена поддержка YOLOv3, что обеспечило удобство их интеграции [16].

4. РЕАЛИЗАЦИЯ АЛГОРИТМА МОНИТОРИНГА

Опишем основные этапы реализации предлагаемого алгорима мониторинга пешеходов на основе выбранных технологий – YOLOv3 и OpenCV. На первом этапе выполняются предварительные действия по инициализации библиотек и входных данных. Загрузка конфигурации и весов YOLO выполняются при помощи функции OpenCVdnn.readNetFromDarknet из модуля dnn – DeepNeuralNetwork. Результатом выполнения данной функции является объект, представляющий нейронную сеть, считанную из входных конфигурационных файлов; инициализация входного и выходного видеофайлов, функцией OpenCVVideoCapture. Результат выполнения данного метода – объект входного видеопотока. Считываются высота и ширина кадров входного видеопри помощи метода get() входного видеопотока. Происходит определение используемых далее переменных. Задается расположение линии подсчета, цвет для обозначения ограничивающих прямоугольников в кадре. Помимо этого, инициализируются массивы для сохранения координат найденных объектов в текущем и предыдущем кадрах и счетчик, необходимый для суммарного подсчета людей. Далее в цикле выполняется покадровая обработка. Кадр считывается при помощи функции read(), вызываемой для объекта входного видеопотока. Если очередной кадр считать не удалось, обработка завершается, происходит выход из цикла. Из считанного исходного кадра формируется двоичный объект при помощи метода dnn.blobFromImage. Он подается на вход нейронной сети при помощи методов setInput и forward для обработки и получения ограничивающих прямоугольников и соответствующих им вероятностей. Далее для всех найденных объектов выполняется

фильтрация по классу – остаются только объекты класса «Person», а также отбрасываются объекты, вероятность попадания в класс «Person» для которых меньше 0,5. По вычисленным координатам ограничивающие прямоугольники наносятся на выходной кадр.

В рассмотренных ранее решениях по подсчету людей основные ошибки возникали на этапе обнаружения – были найдены лишние объекты, которые являются неподвижными или не являются людьми. Части человека подсчитывались как несколько отдельных людей или напротив, движущийся человек не был обнаружен, и, следовательно, не подсчитан. Наконец, в решениях с достаточно большой точностью ошибки наиболее вероятны, если в кадре находится несколько людей, идущих близко друг к другу. В разрабатываемом решении предлагается повысить точность итогового подсчета людей в первую очередь за счет применения средств, позволяющих осуществить более точное распознавание людей в каждом отдельном кадре.

5. ТЕСТИРОВАНИЕ РАЗРАБОТАННОЙ СИСТЕМЫ

Тестирование разработанного приложения проводилось с использованием «живых» тестовых видео и синтетических наборов данных, полученных в компьютерных программах 3D моделирования и персонажной анимации. Сравнение осуществлялось с двумя рассмотренными ранее решениями с открытым исходным кодом – «People-detecting» и «People_counter».

Тестирование на фактических данных. Была произведена съемка набора тестовых видео людей, двигающихся по одной траектории, в которых использовались различные параметры: высота расположения камеры и интервал между идущими людьми (рис.3). На кадре обработанного видео размещена вертикальная линия подсчета. Рамкой желтого цвета обведены обнаруженные люди, для каждого объекта сверху от ограничивающего прямоугольника выводится вероятность попадания в класс Person, в кадре в правом нижнем углу расположена надпись с текущим значением счетчика людей, пересекших линию.

Рис.3. Фрагменты тестового видео и результаты тестирования

Предлагаемые рещения позволили получить большую наглядность, а также помогли конкретизировать моменты, в которых допускались ошибки.

Тестирование с использованием синтетических данных. Далее рассматривается предлагаемый авторами подход - использование синтетических наборов данных (виртуально построенных наборов данных) для тестирования программного обеспечения обнаружения, мониторинга движения и классификации объектов. Создавать тестовые видео предлагается средствами специального программного обеспечения персонажного компютерного 3D моделирования и анимации, так называемых «аниматиков». В проводимых исследованиях были использованы тестовые видео, созданные в программе Antics3D AnticsTechnologies [17]. Antics3D была разработана для создания трехмерной персонажной анимации и визуализации событий в кинематографии, криминалистической анимации, образовании. В настоящее время поддержка программы прекращена, однако она находится в свободном доступе. Программное обеспечение применяет парадигмы перетаскивания, наведения и щелчка для анимации, позволяя пользователям создавать виртуальные сцены, направлять персонажей для выполнения действий и использовать виртуальные камеры для записи сцен, которые можно редактировать на временной шкале [18]. Результат обработки тестовых видео представлен на рис.4.

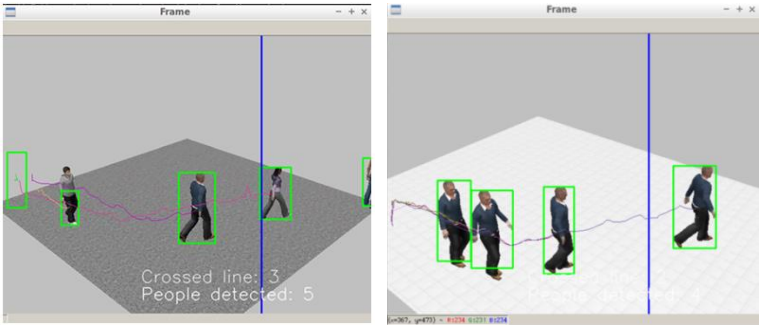


Рис.4. Результаты обработки тестовых видео

Для тестирования были созданы видео с различными интервалами между пешеходами, различными расположениями камер и углами обзора. Результаты одного из тестов (съемка под углом больше 90 градусов) представлены на диаграмме рис.6. На диаграмме отображены полученные значения количества людей в сравнении с реальным значением. Данные сгруппированы по высоте расположения камеры над землей.

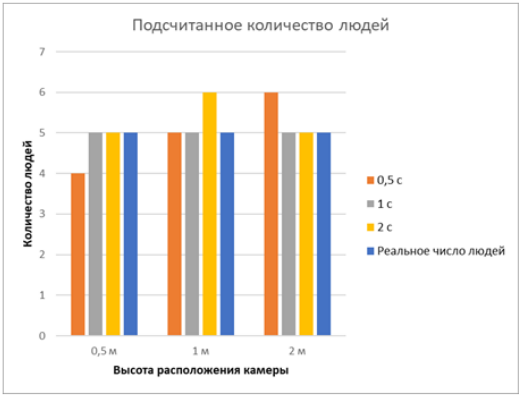


Рис.5. Результаты подсчета количества людей

На основе полученных данных можно сделать следующие выводы: для разработанного алгоритма характерно небольшое количество ошибок при подсчете, не зависящее от расположения камеры и рассмотренных интервалов между людьми. Уменьшение количества ошибок, которые происходчт при пересечении людей в кадре, получено благодаря использованию архитектуры YOLOv3.

Сравнения результатов работы разработанного алгоритма с существующими решениями. Сравнение точности подсчета в рассматриваемых решениях производилось на основе величины относительной ошибки. Для каждого приложения было вычислено суммарное количество ошибок, допущенное во всех тестах, а затем вычислена относительная ошибка, итоговая формула имеет вид:

$$\delta = \frac{\Delta n}{n_\partial} \tag{3}$$

где n – количество подсчитанных людей. Для всех случаев n_∂ – действительное значение величины, равно 45, как общее количество людей, проходившее в тестовых видео. Результаты вычисления относительной ошибки определения количества пешеходов, для реальных и синтетических данных, представлены в виде столбчатой диаграммы (рис.6). Столбцы сгруппированы по тестируемому приложению.

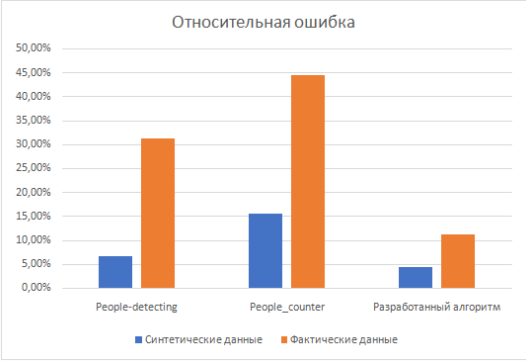


Рис.6. Сравнение относительной ошибки для синтетических и фактических данных

На основании полученной диаграммы можно сделать следующие выводы. При тестировании на фактических данных возникло большее количество ошибок при распознавании и подсчете по сравнение с тестированием на виртуальных сценах. Это связано в первую очередь с особенностями фона, одежды и походки различных людей при съемке в реальных условиях. Разработанный алгоритм показал лучшие результаты при сравнении с рассмотренными решениями с открытым исходным кодом, при тестировании на реальных и синтетических данных. Наблюдается взаимосвязь между количеством ошибок, допущенных при тестировании двумя разными способами. Чем хуже

результат, демонстрируемый приложением при тестировании на синтетических данных, тем хуже результат и при тестировании на видео с реальными людьми. Таким образом, несмотря на то, что при тестировании двумя способами количество ошибок не совпадает, наличие качественной связи между результатами позволяет использовать предлагаемй метод тестирования как менее затратный, позволяющий моделировать различные ситуации и условия съемки.

6. ЗАКЛЮЧЕНИЕ

В ходе выполнения работы был разработан и протестирован алгоритм мониторинга движения пешеходов. Удалось повысить точность распознавания и дальнейшего подсчета пешеходов, обеспечив приемлемую скорость работы. Разработанный алгоритм показал сопоставимые по точности результаты в случаях, когда пересечения пешеходов в кадре не происходит. В случае с небольшим интервалом между людьми и их пересечением в кадре разработанный алгоритм демонстрирует значительно меньшее количество ошибок по сравнению с двумя протестированными приложениями – для выполненных тестов количество ошибок уменьшилось в несколько раз.

Сравнение двух методов тестирования – на синтетических и реальных данных показало, что несмотря на наличие количественных расхождений в точности подсчета пешеходов, наблюдается связь между качеством работы приложения, демонстрируемом при тестировании разными способами. Так результаты тестирования на синтетических данных корелируют с результатами тестирования на реальных данных. Наличие такой взаимосвязи позволяет сделать вывод о том, что использование синтетических данных может применяться для систем автоматического мониторинга движения пешеходов, так как благодаря данному подходу можно сократить затраты на тестирование, при этом создавая тестовые данные с условиями съемки, которые было бы трудно воспроизвести в реальных условиях. В дальнейшем возможно более подробное исследование влияния условий съемки на совпадение результатов при тестировании двумя различными способами.

Литература

1. «MacroscopUltra 3.0. Подсчет посетителей» – URL: https://macroscop.com/assets/documentation/ultra-3.0/analytics/counting/counting.htm.
2. «Модуль аналитики «Подсчет посетителей» Domination» – URL: https://domination.one/products/analitycs/modul-analitiki-podschet-posetiteley-domination-n0000131235/.

3. «CleverCamera. Инструкция по использованию» URL: http://www.clevcam.ru/assets/files/how_to_use.pdf.
4. «Детектирование и подсчет посетителей в режиме реального времени на одноплатном компьютере «Up-board»» – URL: https://habr.com/company/simbirsoft/blog/359304/.
5. «Footfall. Application that allows you to monitor the traffic in and out of your building, using the RPi Camera and openFrameworks» – URL: https://github.com/WatershedArts/Footfall.
6. «Overhead-camera-people-counter. People counting algorithm using an overhead video camera» – URL: https://github.com/koba/overhead-camera-people-counter
7. «People-detecting. Detect people from video/camera using OpenCV. FollowMovement» – URL: https://github.com/ndaidong/people-detecting/tree/FollowMovement.
8. «Counting people in a video stream» – URL: https://github.com/shashgupta/people_counter.
9. «Pedestrian Detection OpenCV by Adrian Rosebrock» – URL: https://www.pyimagesearch.com/2015/11/09/pedestrian-detection-opencv/
10. «OpenCVPeopleCounter» – [2018] – URL: https://www.pyimagesearch.com/2018/08/13/opencv-people-counter/.
11. Howard, W.R. Pattern Recognition and Machine Learning. Kybernetes, Vol. 36 No. 2 – [2007].
12. «Understanding of Convolutional Neural Network (CNN) – URL: https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148.
13. «Сверточная нейронная сеть, часть 1: структура, топология, функции активации и обучающее множество» – URL: https://habr.com/ru/post/348000/
14. «YOLO: Real-Time Object Detection» // [Электронныйресурс] / https://pjreddie.com/ – [2018] – URL: https://pjreddie.com/darknet/yolo/.
15. «What's new in YOLO v3? » – URL: https://towardsdatascience.com/yolo-v3-object-detection-53fb7d3bfe6b.
16. «Darknet Yolo v2 is added to the OpenCV» – URL: https://github.com/pjreddie/darknet/issues/242.
17. «Antics3D» – URL: https://en.wikipedia.org/wiki/Antics3D.
18. «Antics' new premium content and free 3D animation software» – URL: http://www.studiodaily.com/2008/03/antics-technologies-opens-3d-animation-to-all/.

УДК 656.61.052:629.5.072:004.032.26

РАСЧЁТ ПРОДОЛЬНОЙ СКОРОСТИ СУДНА НА ОСНОВЕ НЕЙРОННОЙ СЕТИ

***Дерябин В.В.**, д.т.н.*
e-mail: gmavitder@mail.ru
ГУМРФ имени адмирала С.О. Макарова

1. ВВЕДЕНИЕ

Решение задачи определения продольной скорости судна относительно воды может возникнуть в случае выхода из строя датчика скорости – лага. В открытом море знание продольной скорости необходимо для реализации режима счисления пути, который используется обычно для контроля спутниковой радионавигационной системы (GPS NAVSTAR, ГЛОНАСС), а также применяется автономно, когда использование спутниковой системы и других неавтономных методов невозможно.

Известные методы определения продольной относительной скорости базируются, как правило, на составлении регрессионных зависимостей скорости хода судна от частоты вращения гребного винта (или оборотов главного двигателя). Такие зависимости, как правило, позволяют учитывать только режим работы движителя. Более того, они предполагают статичный, установившийся характер движения судна. Более совершенными в этом смысле являются расчётные методы, основанные на использовании аппарата обыкновенных дифференциальных уравнений [1, 2].

К недостаткам методов, основанных на использовании дифференциальных уравнений, можно отнести трудности с определением вида их правых частей, т.е. алгоритмов расчёта отдельных сил, действующих на корпус судна. Как правило, области применимости данных алгоритмов ограничены конкретными условиями плавания, и указанные методы не всегда являются универсальными, адекватными для всех условий. Более того, гидродинамические силы часто раскладываются в ряд Тейлора, в результате чего возникают дополнительные ошибки, т.е. ошибки метода.

Можно показать [3], что задача определения относительной скорости судна есть задача аппроксимации (приближения) функций нескольких переменных.

Из теории нейронных сетей известно [4-6], что сеть прямого распространения с одним скрытым слоем и полными связями может равномерно приблизить любую непрерывную функцию с любой наперёд заданной точностью. Нейроны скрытого слоя такой сети не должны являться полиномами, но быть при этом локально интегрируемыми по Риману [6]. Нейроны выходного слоя имеют тождественную функцию активации.

2. ПОСТАНОВКА ПРОБЛЕМЫ

Продольное ускорение судна относительно воды может быть представлено следующим образом [7]:

$$\frac{dV_x}{dt} = f\left(V_x, V_y, \omega, n, \delta, V_r, \alpha_r, h, \gamma, \lambda\right), \tag{1}$$

где V_x – продольная составляющая скорости судна относительно воды, V_y – поперечная составляющая скорости относительно воды, ω – угловая скорость поворота судна вокруг вертикальной оси, n – частота вращения винта, δ – угол перекладки руля, V_r – модуль вектора скорости относительного ветра, α_r – направление вектора скорости относительного ветра (курсовой угол), h – высота волн, γ – направление волн относительно корпуса судна (курсовой угол), λ – длина волн.

Уравнение (1) определяет список факторов, которые должны учитываться при расчёте продольной скорости судна. Однако, можно учитывать не все из них, а только наиболее существенные, что может упростить процесс обучения нейронной сети. Более того, для сокращения размерности задачи целесообразно некоторые из переменных объединить. На данном этапе исследования будем ограничиваться учётом лишь ветра в качестве внешнего фактора. Волнение учитываться не будет. С учётом указанных соображений, переходя к дискретному виду, получаем следующее:

$$y := V_x(t_1) \approx f_y\left(X\right), \tag{2}$$

где вектор X определяется следующим образом:

$$X = \begin{pmatrix} V_x(t_0), \omega(t_0), \omega(t_1), n(t_0), n(t_1), \delta(t_0), \delta(t_1), \\ V_r(t_0)\cos\alpha_r(t_0), V_r(t_1)\cos\alpha_r(t_1) \end{pmatrix},$$

t_0, t_1 – два последовательный момента времени, разделенных шагом $\Delta t = t_1 - t_0$.

Таким образом, возникает задача разработки нейронной сети, реализующей преобразование $f_y(X)$ в соотношении (2). Данная задача включает в себя ряд последовательных этапов:

- определение архитектуры нейронной сети;
- формирование набора учебных данных;
- обучение (настройка) нейронной сети;
- тестирование нейронной сети.

3. АРХИТЕКТУРА НЕЙРОННОЙ СЕТИ

В качестве архитектуры нейронной сети выбирается модель полносвязного многослойного персептрона с одним скрытым слоем. Первый слой содержит N нейронов с функциями активации в виде гиперболического тангенса. Второй слой представлен единственным нейроном с тождественной функцией активацией. В среде разработки MATLAB 2020b нейронная сеть задаётся функцией feedforwardnet(). Общий вид архитектуры такой сети представлен на рис.1.

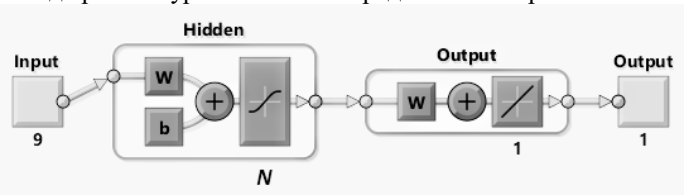


Рис.1. Общий вид архитектуры нейронной сети для расчёта продольной скорости судна

Важным параметром, определяющим скорость сходимости алгоритма настройки нейронной сети и, следовательно, конечную точность системы, в целом, является число нейронов в скрытом слое N. Каких-либо точных рекомендаций, позволяющих определить, сколько нейронов необходимо для решения данной задачи аппроксимации с заданной точностью, не существует. В связи с этим, остаётся в распоряжении лишь эмпирический путь. В рамках настоящего исследования были рассмотрены нейронные сети с количеством нейронов N 30, 40, 50, 60 и 70.

4. ПРОЦЕДУРА ФОРМИРОВАНИЯ ОБРАЗЦОВ

Для настройки нейронной сети необходим набор данных, содержащий достаточное количество учебных примеров (образцов). В идеальном случае, набор учебных данных должен быть получен в ходе проведения экспериментов на реальном судне. К сожалению, такая

возможность не всегда имеется, и приходится прибегать к имитационному моделированию.

В качестве реализации имитационной модели движения выбирается Marine System Simulator (MSS) (T. I. Fossen and T. Perez (2004). Marine Systems Simulator (MSS) URL: https://github.com/cybergalactic/MSS). MSS представляет собой набор MATLAB-файлов, позволяющих моделировать движение судна в условиях воздействия управляющих и внешних факторов. В качестве модели судна выбирается контейнеровоз серии SR 108 [8], т.е. используется функция container.m. Данная функция учитывает гидродинамические силы, воздействующие на корпус судна, а также силы со стороны винто-рулевого комплекса. Для учёта воздействий ветра к силам и моменту вокруг вертикальной оси были добавлены слагаемые, включающие силы, получаемые на выходе функции blendermann94m. Данная функция принимает на вход следующие параметры:

- площадь боковой парусности (ALw): 3000 м2;
- площадь фронтальной парусности (AFw): 700 м2;
- абсцисса центра боковой парусности (sL): –3,7 м;
- аппликата центра боковой парусности (sH): 9 м.

На выходе определяются продольная и поперечная силы, а также момент вокруг вертикальной оси, вызванные воздействием ветра на судно [9].

Таким образом, используемая имитационная модель учитывает ветровые воздействия, но не учитывает волнение, поэтому её применение ограничено лишь умеренными ветровыми воздействиями, когда волнение не достигает интенсивности, достаточной для того, чтобы заметным образом влиять на скорость судна. Для существенного же волнения необходим и его учёт в модели. Также благоприятными условиями применения разрабатываемого алгоритма будут плавание вдоль берега или кромки льда при направлениях ветра от них.

Использование имитационного моделирования позволяет не только предварительно апробировать нейросетевой алгоритм перед планированием натурных наблюдений, но также нейронная сеть может быть обучена на имитируемых данных заранее, а затем её состояние будет корректироваться по мере поступления образцов в ходе реальной навигации судна.

Следует отметить также, что при моделировании движения судна используется ПД-регулятор, стабилизующий судно на заданном курсе. Пропорциональный коэффициент принимается равным 1, дифференциальный – 10.

Сама технология получения образцов состоит в следующем. Рассматриваются всё возможные комбинации «заданный курс+частота вращения винта *n*». Заданный курс выбирается из следующей

последовательности: 0, 15, 30, …, 345°. Частота n из следующей: 30, 40, …, 100 об. /мин. Каждая комбинация соответствует плаванию судна в течение 1800 с, т.е. модельной ситуации. В каждой модельной ситуации начальный курс отличается от заданного на 30°. Причём, если в одной модельной ситуации курс меньше заданного на 30°, то в следующей он непременно на 30° больше, и наоборот. Начальная скорость в каждой модельной ситуации устанавливается равной пропорционально оборотам из следующей последовательности 6, 7, 9, 11, 13, 15, 17 и 19 уз. (1 узел (уз.) = 1 морской миле в час). Ветер во всех модельных ситуациях выбирается северного направления со скоростью 15 м/с.

Таким образом, образцы формируются в результате работы имитационной модели на каждой итерации вложенных циклов:

```
for i = 1:24%заданный курс k_ref
    for j = 1:8%частота вращения винта rpm
        k_ref(i), rpm(j); % комбинация для индексов i, j
        % работа имитационной модели (1800 с)
    end
end
```

В результате работы данного алгоритма были получено 345600 образцов. Для непосредственной настройки (коррекции весовых коэффициентов и пороговых значений сети) случайным образом из них были выбраны 241919 образцов (70%). Так была сформирована обучающая выборка. Остальные использовались для проверки обобщающих свойств и формировали контрольную выборку.

5. ПРОЦЕДУРА НАСТРОЙКИ НЕЙРОСЕТИ

Для настройки нейронной сети использовался метод Левенберга-Марквардта [10], который успел хорошо зарекомендовать себя для решения подобного класса задач. Он обладает быстрой сходимостью по сравнению с обычным «дельта-правилом», т.к. является алгоритмом второго порядка. Однако, применение данного метода имеет свои особенности. Дело в том, что в качестве функционала, подлежащего минимизации, как известно, выступает средний квадрат ошибки (точнее, его половина). Выбор данного критерия для алгоритма обратного распространения ошибки продиктован, в первую очередь, удобством существования и вычисления частных производных на пространстве оптимизируемых весовых коэффициентов нейронной сети. Тем не менее, с точки зрения судовождения более надёжным представляется всё-таки критерий максимума модуля ошибки определяемой скорости $e_{\max}$ (естественно, на контрольной выборке). Его и будем использовать.

С учётом вышеизложенного, процесс обучения построим следующим образом. Зададим общее количество итераций обучения (10000). После каждой итерации обучения, т.е. однократной коррекции весов нейронной сети методом Левенберга-Марквардта, рассчитывается характеристика точности $e_{\max}$ на образцах из контрольной выборки. Если значение данной характеристики уменьшается, то состояние сети сохраняется, и выполняется переход к следующей итерации обучения. Если нет, то переход к следующей итерации обучения выполняется без сохранения состояния сети.

Помимо $e_{\max}$ в ходе обучения контролируется также e_{mean} – среднее значение модуля ошибки скорости, а также минимизируемая функция – половина среднего квадрата ошибки p.

Результаты обучения представлены в таблице 1. В ней значения e_{mean} и p указаны для той итерации, для которой были получены лучшее значение $e_{\max}$ (колонка «№ итерации»), т.е. не обязательно минимально достигнутые значения за 10000 итераций.

Таблица 1. Результаты обучения нейронных сетей

Кол-во нейронов N	Контрольная выборка			
	№ итерации	$e_{\max}$, м/с	e_{mean}, м/с	p, $(\text{м/с})^2$
30	9220	$6,2\cdot10^{-4}$	$2,0\cdot10^{-5}$	$1,1\cdot10^{-9}$
40	8717	$5,5\cdot10^{-4}$	$1,9\cdot10^{-5}$	$8,5\cdot10^{-10}$
50	10000	$5,5\cdot10^{-4}$	$1,8\cdot10^{-5}$	$7,3\cdot10^{-10}$
60	10000	$5,1\cdot10^{-4}$	$1,4\cdot10^{-5}$	$4,9\cdot10^{-10}$
70	10000	$5,5\cdot10^{-4}$	$1,3\cdot10^{-5}$	$4,1\cdot10^{-10}$

Как видно из таблицы, лучшие результаты были достигнуты для нейронной сети, содержащей 60 скрытых нейронов. Наименее точной оказалась сеть, содержащая 30 нейронов в скрытом слое. Для большинства сетей наилучшая точность была достигнута для последней итерации, что позволяет надеяться на возможное увеличение точности по мере дальнейшего обучения. Для двух сетей было обнаружено переобучение, причём эти сети, как ни странно, имеют наименьшие количества нейронов в скрытом слое (30 и 40).

Следует отметить, что перед обучением каждой из сетей выполнялось своё случайное разбиение набора учебных данных на обучающую и контрольную выборки. Более объективные результаты

могут быть получены, если для каждого N провести процедуру обучения несколько раз, каждый раз случайно перемешивая образцы. Оставлять же одно случайное разбиение набора данных для каждого количества N будет, скорее всего, неправильным, т.к. одно и то же разбиение может хорошо подходить для одной сети, но быть менее удачным для другой.

6. ТЕСТИРОВАНИЕ НЕЙРОННОЙ СЕТИ

После обучения нейронная сеть должна пройти процедуру тестирования в таких ситуациях, которые не наблюдались в процессе её настройки. В большой степени, обеспечить такой вариант тестирования можно, например, если задавать заданный курс из промежуточных значений, по сравнению с теми, которые использовались для формирования набора учебных данных. Т.е. задавать последовательность курса следующую 0; 7,5°; 22,5°; …, 352,5°. При этом начальный курс совпадает с заданным в каждой модельной ситуации, что более соответствует установившемуся режиму удержания судна на курсе. Длительность каждой ситуации задаётся равной 4 часа. Частота вращения винта выбирается 90 об. /мин. Скорость и направление истинного ветра прежнее (северный, 15 м/с).

В процессе тестирования фиксируются $e_{\max}$ и e_{mean}, а также СКП, которые определяются по всем определяемым значениям во всех модельных ситуациях.

Работа нейронных сетей сравнивалась с метрическим алгоритмом (методом ближайшего соседа), в котором в качестве функции расстояния использовалось расстояние Евклида. Естественно, что данный алгоритм использует весь набор учебных данных, т.е. все образцы из обучающей и контрольной выборок.

Общее число модельных ситуаций составляет 24. Результаты тестирования представлены в таблице 2.

Таблица 2. Результаты тестирования нейронных сетей

Кол-во нейронов N	Характеристика точности (24 четырёхчасовые выборки)		
	$e_{\max}$, уз.	e_{mean}, уз.	СКП, уз.
30	0,0079	$6,2 \cdot 10^{-4}$	0,0010
40	0,0103	0,0010	0,0015
50	0,0061	$9,5 \cdot 10^{-4}$	0,0013

60	0,0165	$9,7 \cdot 10^{-4}$	0,0014
70	0,0144	$6,1 \cdot 10^{-4}$	$8,7 \cdot 10^{-4}$
Метод ближайшего соседа	1,7546	1,0334	0,5293

Как видно из таблицы 2, все нейронные сети демонстрируют высокую точность прогноза продольной скорости судна, в отличие от метода ближайшего соседа. Следует отметить, что в соответствии с Резолюцией Комитета по безопасности на море при Международной морской организации MSC.96(72), погрешность определения скорости при помощи лага должна быть не более 0,2 уз. или 2% от измеряемой скорости (смотря по тому, что больше). Точность нейронных сетей получилась на порядок выше. Метрический же алгоритм оказался не способным удовлетворять указанным требованиям точности.

Необходимо отметить также, что лучшие и худшие результаты тестирования показали не те сети, которые обнаружили такие результаты на контрольной выборке в процессе обучения. Например, сеть с 60 нейронами должна была быть самой точной, но ею оказалась сеть с 50 нейронами. То же относится и к сети 30 нейронами, которая на этапе настройки выглядела как наименее точная.

7. ЗАКЛЮЧЕНИЕ

Разработана нейронная сеть, рассчитывающая продольную скорость судна с учётом управляющих воздействий со стороны движительно-рулевого комплекса, а также с учётом ветровых воздействий. Сеть построена на основе дискретной модели, предполагающей расчёт продольной скорости на текущий момент времени на основе факторов, измеряемых в текущий и предшествующий моменты времени. Набор учебных данных был сформирован на основе имитационной модели для судна-контейнеровоза (MSS toolbox). Она же использовалась и для тестирования. Результаты тестирования говорят о высокой точности определения продольной скорости в рассмотренных модельных ситуациях. Разработанную систему можно условно назвать «нейросетевой лаг», который может использоваться для определения продольной скорости судна в случае выхода из строя лага.

В качестве перспектив исследования можно выделить переход от имитационного моделирования к использованию натурных данных, что потребует, возможно, доработки процедуры получения образцов. Также целесообразным будет исследование возможностей использования

нейронных сетей с более чем одним скрытым слоем (глубоких сетей) для решения задачи оценки продольной скорости судна.

Литература

1. Fossen T.I. Handbook of Marine Craft Hydrodynamics and Motion Control. – Chichester: John Wiley & Sons Ltd, – 2011. – XVIII, – P. 575.
2. Guide for vessel maneuverability (Updated Feb. 2017). – Houston: American Bureau of Shipping, – 2017.
3. Дерябин В.В. Модель движения судна на основе нейронных сетей // Навигация и гидрография. – 2020. – №60. – С. 32-42.
4. Bianchini M., Maggini M., Jain L.C. Handbook on neural information processing. – NY: Springer. – 2013. – XX. – P. 538.
5. Haykin S. Neural Networks and Learning Machines: Third Edition. – NY: Prentice Hall. – 2009. – XXX. – P. 906.
6. Hornik K. Some new results on neural network approximation// Neural Networks. – 1993. – Vol.6. – №8. – P. 1069-1072.
7. Deryabin V.V. Neural Networks Based Prediction Model for Vessel Track Control// Automatic Control and Computer Sciences. – 2019. – Vol.53. – Is.6. – P. 502-510.
8. Son K-H., Nomoto S. On the Coupled Motion of Steering and Rolling of a High-Speed Container Ship. // Naval Architecture and Ocean Engineering. – 1982. – Vol.20. – P. 73-83.
9. Blendermann W. Parameter identification of wind loads on ships// Journal of Wind Engineering and Industrial Aerodynamics. – 1994. – Vol.51. – Is.3. – P. 339-351.
10. Yu H. Advanced Learning Algorithms of Neural Networks. PhD dissertation, Auburn, USA, 2011.

СЕКЦИЯ 7 | **ЭВОЛЮЦИОННОЕ МОДЕЛИРОВАНИЕ И БИОНИЧЕСКИЕ АЛГОРИТМЫ**

УДК 573.2

МОДЕЛЬ ЭВОЛЮЦИИ И ОБУЧЕНИЯ NK-АВТОМАТОВ КАУФМАНА[*]

Редько В.Г.
e-mail: vgredko@gmail.com
НИИ системных исследований РАН

1. ВВЕДЕНИЕ

В настоящей работе построена компьютерная модель обучения и эволюции автономных агентов, генотипы и фенотипы которых определяются NK-автоматами Стюарта Кауфмана [1-5].

NK-автомат есть сеть из N булевых логических элементов. Каждый логический элемент имеет K входов и один выход. Сигналы на входах и выходах элементов бинарные, т.е. принимают значения 0 либо 1. Выходы одних элементов поступают на входы других, эти связи случайны, но число входов K каждого элемента фиксировано. Сами логические элементы также выбираются случайно. Автоматы автономны, т.е. внешние входы отсутствуют. Число логических элементов, входящих в автомат, предполагается большим, $N >> 1$. Автомат функционирует в дискретном времени: $t = 1, 2,\ldots$ Состояние автомата в каждый момент времени t определяется вектором $\mathbf{X}(t)$ – совокупностью выходных сигналов всех логических элементов. В процессе функционирования последовательность состояний сходится к

[*] Работа выполнена при финансовой поддержке РФФИ, проект №19-01-00331

аттрактору – предельному циклу. Число аттракторов M и типичная длина аттрактора L – важные характеристики NK-автоматов. При больших K ($K = N$) «жизнь» автоматов стохастична. Длина аттракторов L очень велика: $L \sim 2^{N/2}$. Число аттракторов M порядка N. При $K \sim 2$ поведение автоматов принципиально меняется: стохастичность значительно уменьшается. Длина аттракторов L и число аттракторов M уменьшаются до величин порядка $N^{1/2}$ Такое поведение характеризуют как жизнь на границе хаоса и порядка (“at the edge of chaos”). NK-автоматы могут рассматриваться как модель генетической регуляторной системы живых клеток.

С.А. Кауфман аргументирует, что именно случай $K \sim 2$ есть адекватная модель молекулярно-генетических систем управления биологических клеточных организмов. В частности, если сравнить число различных аттракторов NK-автомата M при $K = 2$ с числом различных типов клеток биологических организмов разного эволюционного уровня, то мы получим близкие цифры. Например, для человека мы имеем $N \sim 10^5$. При этом число аттракторов для NK-автомата M приближенно равно 370, число различных типов клеток равно 254 [1-5].

В ряде работ исследовались компьютерные модели взаимодействия между обучением и эволюцией [6-11]. В работах [6, 7] были продемонстрированы основные эффекты взаимодействия между обучением и эволюцией: 1) генетическая ассимиляция приобретаемых в результате индивидуального обучения навыков в течение ряда поколений дарвиновского эволюционного процесса, 2) эффект экранирования, который приводит к тому, что сильное обучение тормозит эволюционный поиск оптимального генотипа, 3) эффект нагрузки на обучение, который приводит к ускорению эволюционного процесса нахождения оптимума. Однако в этих работах были использованы довольно сложные схемы генетического алгоритма со скрещиванием, поэтому механизмы эффектов, характеризующих взаимодействие между обучением и эволюцией, не были четко представлены. В наших работах [8-11] для изучения механизмов взаимодействия между обучением и эволюцией была использована одна из наиболее четких моделей эволюции – модель квазивидов Манфреда Эйгена [12, 13]. При этом в [8-11] задача оптимизации была сведена к одномерной задаче: к задаче уменьшения расстояния до единственного оптимума [8, 9] или к задаче поиска минимумов энергии спинового стекла [10, 11]. То есть в процессах эволюции и обучения использовалась только одна существенная переменная (расстояние до оптимума или энергия спинового стекла), что позволило четко

проанализировать эффекты взаимодействия между обучением и эволюцией.

В настоящей работе используется опыт исследований [8-11]. Рассматриваемая компьютерная модель основана на модели квазивидов [12, 13]. В процессах эволюции и обучения используется одна существенная переменная – число аттракторов M анализируемого NK-автомата.

2. КОМПЬЮТЕРНАЯ МОДЕЛЬ ЭВОЛЮЦИИ И ОБУЧЕНИЯ NK-АВТОМАТОВ

Изложим модель эволюции и обучения NK-автоматов. Имеется популяция, состоящая из n агентов. Генотипы и фенотипы агентов задаются NK-автоматами. Сначала рассмотрим эволюцию агентов без обучения, введя только генотипы агентов. Затем рассмотрим фенотипы агентов, механизм обучения агентов и совместные процессы эволюции и обучения.

Генотип каждого агента задается определенным автоматом. Число генов автомата равно N. Каждому гену G_i соответствует свой логический элемент, i – номер гена, $i = 1, 2, \ldots, N$. Число входов элемента равно K. Для определённости считаем, что $K = 2$, т.е. каждый логический элемент имеет два входа от определенных генов. Такие автоматы могут эффективно моделировать молекулярно-генетические системы биологических клеток [1-5]. Гены принимают бинарные значения: $G_i = 0$ либо 1. В начальной популяции значения генов для всех агентов случайны. Логическая функция каждого логического элемента изначально (в начальной популяции) также считается случайной, т.е. она случайно выбирается одной из 16 возможных булевых функций двух переменных. В начальной популяции входы определенного логического элемента от других генов также выбираются случайно. Автомат каждого агента функционирует в дискретном времени: $t = 1, 2, \ldots$ При переходе к следующему такту времени каждый логический элемент срабатывает, выходы элементов определяют новые значения генов G_i. Естественно, что новые значения генов равны 0 либо 1: $G_i = 0$ либо 1.

Итак, в начальной популяции генотип каждого агента определяется случайным NK-автоматом. Логический элемент для каждого из генов выбирается случайно. Связи между генами и входами логических элементов также случайны.

При функционировании автоматов значения генов G_i меняются во времени. Обозначим вектор генов $\mathbf{G} = \{G_i\}$. Для каждого агента изменения вектора $\mathbf{G}$ приводит к определенному аттрактору. В

зависимости от начального значения вектора $\mathbf{G} = \{G_i\}$ и структуры NK-автомата аттракторы могут быть различными. В компьютерной модели определялось число аттракторов M для каждого из агентов. И чем больше число таких аттракторов у агента, тем больше преимущество данного агента. При биологической интерпретации NK-автоматов число различных аттракторов соответствует числу различных клеток организма (см. выше). То есть число аттракторов характеризует число различных функций, выполняемых рассматриваемым автоматом. Поэтому число различных аттракторов качественно характеризует эффективность рассматриваемого агента.

Процесс *эволюции* популяции агентов моделируется следующим образом. Рассматриваются поколения популяции. В начальной популяции, как было изложено выше, все агенты имеют случайные генотипы. При переходе к новому поколению осуществляются отбор и мутации генотипов агентов. Численность популяции n сохраняется постоянной.

При отборе оценивается число аттракторов для каждого из агентов M_k, k – номер агента, $k = 1, 2,\ldots, n$. Далее определяются приспособленности агентов:

$$f_k = \exp(\beta M_k), \qquad (1)$$

где β – параметр интенсивности отбора. Отбор особей в новое поколение производится хорошо известным в теории эволюционного моделирования рулеточным методом, в котором отбор осуществляется вероятностно. При этом особи отбираются в новое поколение с вероятностями, пропорциональными их приспособленностям f_k. Выбор новой особи в следующее поколение производится ровно n раз. Каждый раз вероятность выбора k-й особи равна

$$q_k = f_k \left[\sum_{l=1}^{n} f_l \right]^{-1}.$$

Вероятности q_k пропорциональны приспособленностям особей f_k. В результате такого отбора численность популяции во всех поколениях постоянна и равна n.

Отобранные в новое поколение агенты мутируют. Мутации осуществляются следующим образом. Мутация для каждого из генов осуществляется с вероятностью P_M. Если данный ген выбран для мутации, то для него случайно меняются номера генов, значения которых подаются на вход логического элемента, соответствующего данному гену. Также случайно выбирается новый логический элемент данного выбранного для мутации гена. Таким образом, при мутациях

меняются (с определенными вероятностями, как правило, с малыми вероятностями) как связи между генами, так и логические элементы генов.

Теперь перейдем к описанию фенотипов агентов и к изложению совместного процесса обучения и эволюции агентов. Считаем, что в течение поколения фенотипы агентов могут изменяться путем обучения. Но генотипы агентов при этом в течение поколения не меняются и передаются потомкам агентов, рождающимся в следующем поколении. Имеются этапы поколения $T = 1, 2, \ldots, T_{max}$, T – номер этапа, T_{max} – длительность поколения, длительность жизни каждого агента. В начале поколения фенотипы агентов **Ph** определяются их генотипами **G**. В нашей модели мы считаем, что начальный фенотип каждого агента равен его генотипу: $\mathbf{Ph}_k(T = 1) = \mathbf{G}_k$, k – номер агента, $k = 1, 2,\ldots, n$. При этом мы вводим понятие фен, аналогичное понятию ген. Фен, как и ген, в нашей модели принимает значение 0 либо 1. Фенотип, так же, как и генотип, определяет структуру NK-автомата, с логическими элементами и связями между ними. Структура фенотипа точно такая же, как и структура генотипа. Основное отличие генотипа и фенотипа состоит в том, что генотип в течение жизни агента не меняется, а фенотип модифицируется в процессе обучения агента. Отбор в следующее поколение осуществляется в соответствии с конечными фенотипами агентов, получившимися в процессе обучения, а в новое поколение передаются генотипы агентов (с малыми мутациями), которые агенты получили от своих родителей. Конечные фенотипы определяют своё число аттракторов M_k, по которым определяются приспособленности агентов в соответствии с выражением (1).

Обучение агентов происходит методом проб и ошибок. На каждом этапе T обучение происходит с определенной вероятностью P_L для каждого из фенов Ph_i. i – номер фена, $i = 1, 2,\ldots, N$. Сначала осуществляется проба для выбранного фена. Проба делается так же, как и при мутациях, только не для генотипов, а для фенотипов. А именно, происходит случайная замена номеров фенов, значения которых подаются на вход логического элемента, соответствующего выбранному фену. Также происходит случайная замена логического элемента для выбранного фена. Затем происходит проверка сделанной пробы, а именно, оценивается число аттракторов для нового NK-автомата, кодируемого фенотипом. Если в результате пробы число аттракторов увеличивается, то новые связи к логическому элементу выбранного фена и сам этот новый логический элемент сохраняются. Если в результате пробы число аттракторов уменьшается, то считается,

что проба была неудачной и восстанавливаются старые связи к логическому элементу и старый логический элемент.

В результате в процессе обучения у агентов увеличивается число аттракторов, определяемых фенотипами. Приспособленности агентов, как и в случае эволюции без обучения, определяются выражением (1), в котором M_k есть число аттракторов NK-автомата, определяемого конечным фенотипом при $T = T_{max}$.

Итак, в случае совместного функционирования обучения и эволюции отбор агентов в новое поколение производится по конечным фенотипам агентов, полученным в течение жизни агента в результате обучения. Но при размножении агентов в следующее поколение передаются с мутациями генотипы агентов, которые они получили от своих родителей. Так как потомкам передаются именно генотипы, а не фенотипы, то эволюционный процесс носит дарвиновский характер.

3. НАЧАЛЬНЫЕ РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

Было проведено начальное моделирование процессов эволюции в рамках рассматриваемой модели. Обучение пока не моделировалось. Были использованы следующие параметры: число генов в одном автомате: N =100, число входов каждого логического элемента: K = 2, численность популяции: n =100, параметр эффективности отбора: β = 10. Интенсивность мутаций варьировалась, было P_M = 0 либо P_M = 0,01.

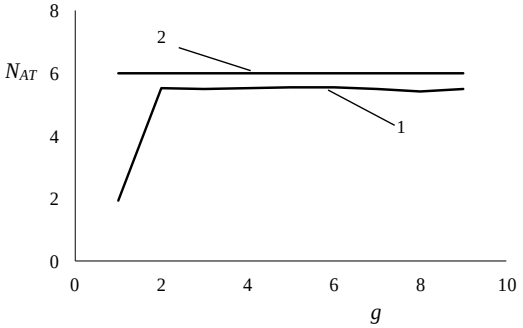


Рис.1. Зависимость числа аттракторов N_{AT} для NK-автоматов от номера поколения g. 1 – среднее по популяции значение N_{AT}, 2 –максимальное в популяции значение N_{AT}. P_M = 0

Типичные результаты моделирования приведены на рис.1 и рис.2. Рис.1 показывает зависимость от номера поколения средней по популяции и максимальной в популяции длины аттрактора при

отсутствии мутаций: $P_M = 0$. Рис.2 показывает такие же зависимости при наличии мутаций: P_M = 0,01. Видно, что число аттракторов для NK-автоматов в процессе эволюции растет. Сопоставление рисунков показывает, что мутации обеспечивают определенный поиск новых удачных автоматов.

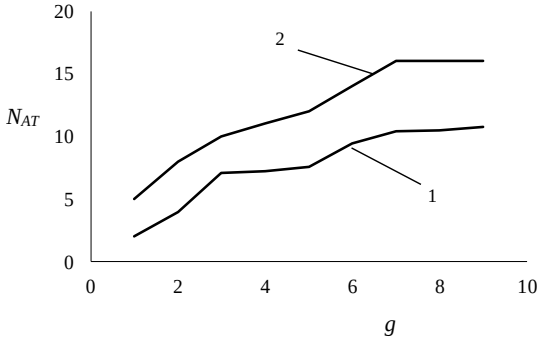


Рис.2. Зависимость числа аттракторов N_{AT} для NK-автоматов от номера поколения g. 1 – среднее по популяции значение N_{AT}, 2 – максимальное в популяции значение N_{AT}. P_M = 0,01

4. ЗАКЛЮЧЕНИЕ

Итак, построена модель эволюции и обучения NK-автоматов Кауфмана. Эволюция состоит из ряда поколений. В каждом поколении происходит отбор и мутации агентов, генотип и фенотип которых кодируются NK-автоматами. Отбор агентов происходит с вероятностями, которые пропорциональны приспособленностям агентов. Приспособленность агентов определяется числом аттракторов в пространстве цепочек генотипов или фенотипов. Чем больше число аттракторов, тем выше приспособленность соответствующего агента. Мутации представляют собой случайные замены связей и логических элементов в генотипе агента.

В течение каждого поколения агенты обучаются. При обучении изменяются фенотипы. В начале поколения фенотипы равны генотипам. Обучение происходит путём модификации фенотипов методом проб и ошибок. При пробе случайно меняются связи и логические элементы в NK-автомате, кодируемом фенотипом. Если при пробе увеличивается число аттракторов для фенотипа, то соответствующие пробе изменения автомата фиксируются в фенотипе. Если при пробе число аттракторов уменьшается, то фенотип агента не меняется.

Отбор в следующее поколение осуществляется в соответствии с числом аттракторов, определяемых конечными фенотипами, полученными в результате обучения. При этом потомкам агентов передаются генотипы (модифицируемые малыми мутациями), которые эти агенты получили от своих родителей.

Произведены начальные расчеты эволюции популяции агентов, колируемых NK-автоматами. Эти расчеты показывают важную роль мутаций при оптимизации структуры агентов.

Литература

1. Kauffman S.A., Smith R.G. Adaptive Automata Based on Darwinian Selection // Physica D. – 1986. – Vol.22. – №1-3. – P.68-82.
2. Кауфман С. Антихаос и приспособление // В мире науки. – 1991. – №10. – С.58-65.
3. Kauffman S.A. Origins of Order: Self-organization and Selection in Evolution. – New York: Oxford Univ. Press, 1993.
4. Редько В.Г. NK-автоматы С. Кауфмана (Интернет-лекция, 1999): https://www.keldysh.ru/pages/BioCyber/Lectures/Lecture9/Lecture9.html
5. Редько В.Г. Эволюция, нейронные сети, интеллект. Модели и концепции эволюционной кибернетики. – М.: КомКнига/URSS, 2005. (Раздел 3.5).
6. Hinton G.E., Nowlan S.J. How Learning Can Guide Evolution // Complex Systems. – 1987. – Vol.1. – №3. – P. 495-502.
7. Mayley G. Guiding or Hiding: Explorations into the Effects of Learning on the Rate of Evolution // Proceedings of the Fourth European Conference on Artificial Life (ECAL 97). Husbands P., Harvey I. (eds.). – 1997. – Cambridge, Massachusetts: MIT Press. – P. 135-144.
8. Редько В.Г. Модель взаимодействия между обучением и эволюционной оптимизацией // Математическая биология и биоинформатика. – 2012. – Т.7. – №2. – С. 676-691. См. также: http://www.matbio.org/2012/Redko_7_676.pdf
9. Red'ko V.G. Mechanisms of Interaction between Learning and Evolution // Biologically Inspired Cognitive Architectures. – 2017. – Vol.22. – P. 95-103.
10. Редько В.Г. Моделирование эффектов взаимодействия между обучением и эволюцией // Журнал общей биологии. – 2020. – Т.81. – №6. – С. 469-480.
11. Red'ko V.G. Spin Glass Energy Minimization through Learning and Evolution // Optical Memory and Neural Networks. – 2020. – Vol.29. – №3. – P. 187-197.
12. Эйген М. Самоорганизация материи и эволюция биологических макромолекул. – М.: Мир, 1973.
13. Эйген М., Шустер П. Гиперцикл. Принципы самоорганизации макромолекул. – М.: Мир, 1982.

УДК 519.712.2

ГИБРИДНЫЙ АЛГОРИТМ РЕШЕНИЯ ЗАДАЧ ПРОЕКТИРОВАНИЯ ЭЛЕКТРОННО-ВЫЧИСЛИТЕЛЬНОЙ ТЕХНИКИ*

***Гладков Л.А.**, к.т.н.*
e-mail: leo_gladkov@mail.ru
Ясир М.Д.
e-mail: yasir_82@mail.ru
Гладкова Н.В.
e-mail: nadyusha.gladkova77@mail.ru
Южный федеральный университет

1. ВВЕДЕНИЕ

Одной из сложнейших задач, стоящих перед специалистами в различных областях техники является обеспечение эффективного процесса проектирования современных технических систем высокой степени интеграции в различных отраслях техники и экономики. Эффективная организация процесса проектирования невозможна без применения вычислительной техники. Это обусловлено наличием большого числа иерархических уровней проектируемых объектов и необходимостью учитывать взаимосвязи между ними, а также большим числом различных элементов, входящих в состав объекта проектирования. Например, если речь идет о проектировании отдельных устройств и блоков в микроэлектронике или вычислительной технике, то размерность каждого отдельного блока может составлять десятки тысяч элементов. В то же время, текущий уровень развития вычислительных ресурсов и технологии пока не позволяет нам решать возникающие задачи за счет наращивания вычислительных мощностей, решать сложные задачи проектирования методом полного перебора всех возможных вариантов. Также полностью автоматизировать процесс проектирования не позволяет отсутствие эффективных детерминированных алгоритмов решения отдельных задач и даже этапов проектирования [1-3].

* Работа выполнена при финансовой поддержке РФФИ, проекты №19-01-00715 и №19-01-00559

Совершенствование технологий изготовления привели к пересмотру некоторых базовых предположений, на которых было основан процесс физического проектирования. В результате прошлые оценки, построенные на модели задержки межсоединения, пропорциональной длине соединений стали неактуальными, изменилась концепция процесса физического проектирования вычислительных устройств.

Одним наиболее известных подходов к решению задачи снижения общих затрат вычислительных ресурсов и времени в процессе проектирования сверхбольших интегральных схем (СБИС), систем на кристалле и печатных плат и т.д. является блочно-иерархический подход. Данный подход позволяет разбить процесс проектирования на этапы и произвести декомпозицию объекта проектирования, что в итоге дает возможность снизить итоговую сложность и размерность решаемых задач [1-3].

Одной из основных тенденций развития современной научной мысли является создание и совершенствование гибридных систем, моделей и методов, подразумевающих объединение в одной системе различных научных подходов, а также моделей и методов различной физической природы. Особенно эффективно использование такого рода систем для решения сложных слабо формализованных задач, что характерно для процесса проектирования [4, 5].

2. ПОСТАНОВКА ЗАДАЧИ

Цель этапа физического проектирования СБИС – заключается в построении некоего абстрактного описания проектируемой схемы, например, в виде списка соединений, создание детальной геометрической схемы расположения элементов на кристалле.

Цикл проектирования элементов цифровой вычислительной техники включает следующие этапы: спецификация системы, функциональное проектирование, логическое проектирование, схемное проектирование, конструкторское проектирование, изготовление, сборка, тестирование и контроль. В свою очередь конструкторский этап проектирования включает задачи: компоновки, размещения, трассировки и т.д.

Размещение – одна из важнейших задач этапа физического проектирования в процессе решения которой выполняется построение макета проектируемой схемы на основе заданного списка соединений, а также оценка времени прохождения сигналов и возможных задержек межсоединений. Важность данной задачи определяется тем, что на данном этапе строится пространственная модель расположения

элементов, которая является основой выполнения всех последующих задач проектирования. В современных электронных устройствах задержки межсоединений становятся определяющим фактором, а поскольку взаимное расположение элементов схемы определяется на этапе размещения, это оказывает существенное влияние на качество проектирования [1-3].

Задача размещения состоит в задании такого расположения элементов схемы на рабочем поле, которое приводит к минимизации общей длины соединений схемы. Это связано с тем, что длина соединений может быть легко смоделирована и служит хорошим приближением первого порядка к реальным целевым функциям, таким как синхронизация, мощность и управляемость проекта. Длина соединений имеет решающее значение при проектировании современных изделий микроэлектроники, поскольку она напрямую влияет на задержки электрических сигналов. Длина соединений также влияет на качество решения задачи трассировки. Трассировка выполняется сразу после размещения, и нет смысла создавать решение для размещения без учета условий для последующей трассировки.

Для определения местоположения элементов схемы на кристалле необходимо сначала определить область размещения. Как правило, это прямоугольную область, границы которой задаются координатами угловых точек (x_{low}, y_{low}, x_{high}, y_{high}). Данное условие не является обязательным, и в последнее время наблюдается более широкое разнообразие областей размещения, например, использование областей размещения *L*-образной или *T*-образной формы. Исходными данными для начала решения задачи размещения является список соединений схемы, который может быть задан в виде графа $G = (X, U)$, где множество X – это элементы проектируемой схемы, а множество U – множество соединений элементов схемы [1-3]. В свою очередь множество вершин X состоит из двух непересекающихся подмножеств: $X = X_1 \cup X_2$, где X_1 – множество элементов, которым еще не назначены позиции, а X_2 – множество элементов, позиции которых уже определены. Местоположение каждого элемента x_i схемы должно находиться в границах заданной области размещения (рабочего поля). Для представления элементов множества соединений u_{ij} могут использоваться различные графовые модели, например, звездный граф или гиперграф.

Итак, исходными данными задачи являются: a, b – размеры (длина и ширина) рабочего поля; $\{z, z_2, ..., z_n\}$ – множество элементов схемы; U – список связности, отражающий связи элементов.

Необходимо найти такой вариант размещения элементов на рабочем поле, чтобы не происходило взаимного перекрытия элементов и суммарная длина соединений была минимальной

$$Z = \{(x_1, y_1), ..., (x_i, y_i), ..., (x_n, y_n)\},$$

где (x_i, y_i) – координаты центра тяжести посадочных мест, предназначенных для размещения элементов.

С точки зрения теории оптимизации задача размещения может быть интерпретирована как задача оптимизации аддитивной целевой функции, включающей нормированное значение штрафа за перекрытие площадей размещаемых элементов и оценку суммарной длины соединений:

$$F = \min_{z_j \in Z}(k \cdot O(L(z_j)) + P(S_{общ}(z_j))),$$

где z_j – текущий вариант размещения; k – весовой коэффициент; $S_{общ}$ – суммарная площадь перекрытия элементов; $O(L(z_j))$ – оценка длины соединений; $P(S_{общ}(z_j))$ –штрафа за перекрытие площадей.

3. СТРУКТУРА АЛГОРИТМА

Стратегической тенденцией при разработке новых методов и подходов к проектированию является их интеллектуализация. Под интеллектуализацией мы будем понимать наличие возможностей выполнять ряд функций и свойств, обычно выполняемых человеком, например, решение задач анализа и принятия решений при наличии неполных или противоречивых входных данных, оценка и интерпретация выявленных фактов. Одним из наиболее эффективных средств решения *NP*-полных комбинаторных задач проектирования являются различные метаэвристические методы и модели, в том числе давно зарекомендовавшие себя и доказавшие свою эффективность генетические, эволюционные, бионические, адаптивные и другие методы поиска [6, 7].

Термин «биоинспирированные алгоритмы» (или алгоритмы вдохновленные природой), включает в себя генетические, эволюционные, роевые и некоторые другие типы алгоритмов. Также к биоинспирированным относятся алгоритмы воспроизводящие процессы, происходящие в живой и неживой природе [8, 9].

На практике часто возникают ситуации, когда гибридные системы показывают достаточно хорошие результаты при решении различных практических задач. Гибридные модели и системы базируются на научных методах, относящихся к вычислительному интеллекту: нечеткие модели и методы, эволюционные и генетические алгоритмы, многоагентные системы и методы. Они позволяют эффективно работать с нечеткой, плохо

формализованной информацией и, в то же время, имеют объективный математический фундамент [4, 5, 10, 11].

Среди научных подходов, связанных с разработкой нечетких гибридных моделей и методов можно выделить две основные группы методов:

1. Разработка методологии обучения и настройки системы нечетких правил, а генетические алгоритмы играют роль инструмента поиска оптимальных решений в условиях, когда информации об объекте оптимизации либо недостаточно, либо она размыта или недоопределена.

2. Оптимизация поискового процесса и динамическое изменение управляющих (свободных) параметров алгоритма поиска за счет использования методов, основанных на нечеткой логике и нечетком выводе.

Также популярным научным трендом является разработка новых принципов организации параллельных вычислений и использование многопроцессорных вычислительных машин [12, 13].

Хороший эффект при создании алгоритмов проектирования и оптимизации дает гибридизация популяционного алгоритма с двумя и более популяционными и/или не популяционными алгоритмами [11]. Структура гибридного алгоритма позволяет сохранить преимущества популяционных алгоритмов, использовав их на начальном этапе для эффективного сужения пространства поиска, а затем применить один из «классических» методов оптимизации для нахождения глобального экстремума.

Одной из гибридных моделей является нечеткий генетический алгоритм. Он сочетает поисковые возможности генетических алгоритмов и возможности систем нечеткого вывода для оценки и изменения управляющих параметров. Для этого используется специальный блок – нечеткий логический контроллер (НЛК), который использует имеющиеся нечеткие правила, оценивает ход процесса эволюции и разнообразие текущей популяции решений и, при необходимости, изменяет значения свободных переменных, таких, например, как вероятности выполнения генетических операторов.

Схема НЛК (рис.1) построена основе управления с обратной связью. НЛК преобразует текущую информацию о ходе эволюции и состоянии популяции решений к нечеткому виду, затем на основе заданных нечетких правил оценивает ее, определяет управляющее воздействие и возвращает скорректированные значения контрольных параметров.

Рис.1. Структура нечеткого генетического алгоритма

Нечеткое управление выполняется на основе использования следующего правила [14, 15]:

$$u(t) = f(e(t), e(t-1), \ldots, e(t-r), u(t-1), \ldots, u(t-r)),$$

где e – разница между ожидаемым и фактическим значениями выходного параметра объекта управления y^*; f – нелинейная функция, отражающая отношение между входом и выходом НЛК; t – время.

Функционирование НЛК определяется набором лингвистических правил, которые записываются в следующей форме: Если (набор условий), то (заключение). НЛК оперирует понятиями нечетких множеств. Поэтому входные значения в процессе фаззификации, преобразуются в лингвистические переменные. Они передаются в блок нечеткого вывода, где происходит генерация множества нечетких наборов с необходимыми значениями функции принадлежности. После выполнения дефаззификации, полученный набор нечетких множеств преобразуется в управляющее воздействие.

В генетических алгоритмах (ГА) управляющими параметрами, как правило, являются значения вероятности выполнения генетических операторов кроссинговера (P_c) и мутации (P_m), а также размер популяции. Взаимодействие блоков нечеткого генетического алгоритма показано на рис.2.

4. РАЗРАБОТКА АЛГОРИТМА

Обобщенная схема организации поиска в популяционных алгоритмах может быть представлена следующим образом:

Рис.2. Схема взаимодействия блоков нечеткого генетического алгоритма

Этап 1. Создание начального множества решений. При этом следует отметить, что качество начальной популяции во многих случаях имеет большое значение для последующей работы алгоритма.

Этап 2. Получение новых решений путем применения к выбранным из начальной популяции индивидам неких эвристических процедур (стандартных либо модифицированных операторов) и оценка изменения значения целевой функции (как отдельных индивидов и средней по популяции).

Этап 3. Проверка условия завершения поиска. Если заданные условия выполнены, вычисления завершаются, и лучшее из решений текущей популяции сохраняется в качестве «квазиоптимального» решение задачи. В противном случае процесс повторяется.

Общая схема работы предложенного гибридного алгоритма представлена на рис.3.

Гибридный алгоритм начинает работу с задания исходных данных, определения критерия оптимизации и ограничений задачи. После этого происходит создание стартовой популяции решений (начальное размещение) и оценка качества начальной популяции (как отдельных индивидов, так и среднего значения по популяции) с точки зрения выбранного критерия оптимизации.

После этого выполняется генетический алгоритм, происходит создание новой популяции решений и возможная корректировка управляющих параметров алгоритма с помощью нечеткого логического контроллера.

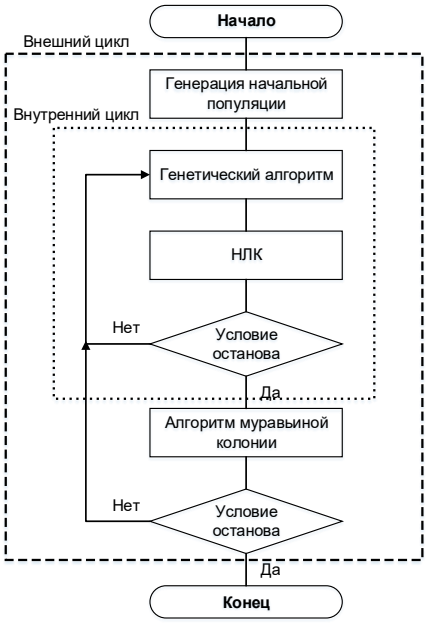


Рис.3. Общая схема гибридного алгоритма

Следующий блок представляет собой модифицированный муравьиный алгоритм. Его задача попытаться улучшить полученное размещение.

В конце происходит проверка выполнения критерия останова алгоритма, условия не выполнены, алгоритм переходит к следующей итерации, в противном случае выполнение алгоритма завершается.

Нечеткий логический контроллер (НЛК) реализует описанные выше требования с помощью гибкой системы правил, с помощью которых удовлетворяются описанные выше условия, что предотвращает ранний локальный оптимум. Правила, по которым работает НЛК, используют несколько характеристик [16]:

- степень разнообразия популяции. В качестве целевой функции для оценки качества решения используется фенотип:

$$VF = \frac{F_{best} - F_{aver}}{F_{best}},$$

где F_{best} – решение с лучшим значением целевой функции, F_{aver} – среднее значение целевой функции для решений текущей популяции.

- скорость роста среднего значения ЦФ популяции, позволяет оценить, насколько динамику улучшения качества решений в популяции:

$$S_{aver} = \frac{F_{aver} - F_{avprev}}{F_{aver}},$$

где $F_{averprev}$ – среднее значение ЦФ решений популяции на предыдущей итерации.

- граничное значение *GAP* – некий лимит, с которым будут сравнивать предыдущие две характеристики, определяющий скорость роста ЦФ. Значение *GAP* должно быть низким, не более 5-10%.

5. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

В ходе проведения вычислительных экспериментов проводилось исследование зависимостей основных параметров разработанного алгоритма от времени, числа итераций и т.д. Исследования проводились при следующих фиксированных значениях параметров: размер популяции – 100; число элементов схемы – 1200. Результаты отображены на рис.4-6.

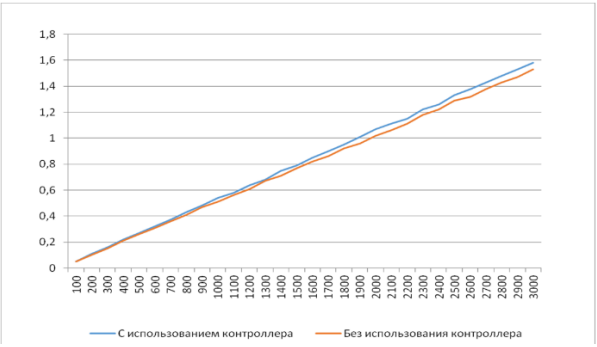


Рис.4. Зависимость времени выполнения алгоритма от количества итераций с использованием НЛК и без него

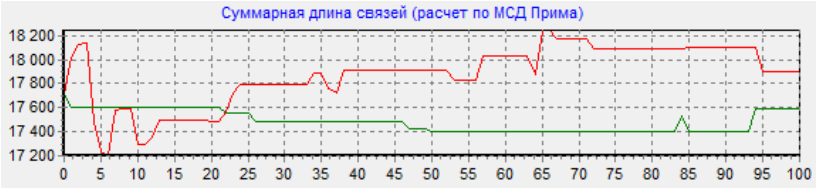


Рис.5. Эффективность работы различных моделей

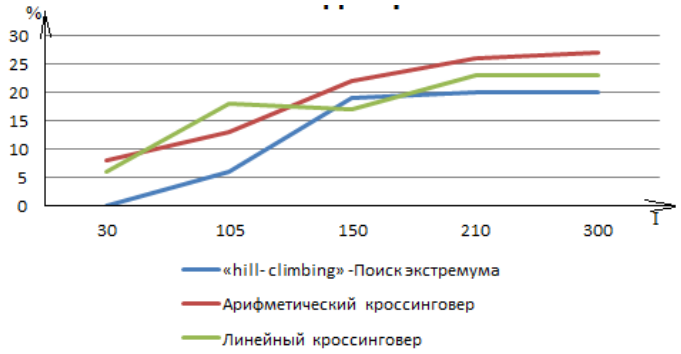


Рис.6. Эффективность различных генетических операторов

Проводились исследования эффективности алгоритма для задачи размещения 300 случайно сгенерированных элементов. Алгоритм выполнялся на одном и том же наборе исходных данных при различном количестве итераций. Популяции генетического алгоритма состояли из 20 хромосом. Результаты экспериментов представлены на рис.7.

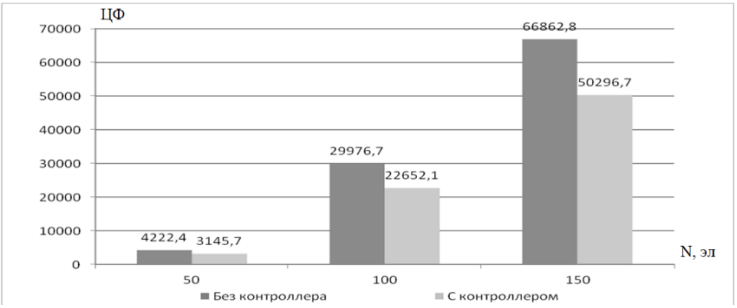


Рис.7. Сравнение качества решений с контроллером и без его использования

6. ЗАКЛЮЧЕНИЕ

Предложена структура гибридного алгоритма решения задачи размещения разногабаритных элементов, разработаны модифицированные генетические операторы, а также компоненты гибридного алгоритма. Гибридный алгоритм реализован в виде прикладной программы, проведены серии вычислительных экспериментов для определения эффективности разработанного алгоритма и выбора оптимальных значений управляющих параметров.

Необходимо дальнейшее продолжение исследований разрабоатнного алгоритма для выявления имеющихся закономерностей, совершенствования механизма реализации разработанных моделей и подходов.

Литература

1. Sherwani Naveed. Algorithms for VLSI Physical Design Automation, Kluwer Academic Publishers, Boston/Dordrecht/London. – 1995.
2. 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.
3. Charles J. Alpert, Dinesh P. Mehta, Sachin S. Sapatnekar. Handbook of algorithms for physical design automation. CRC Press, New York, USA. – 2009.
4. Ярушкина Н.Г. Основы теории нечетких и гибридных систем. М.: Финансы и статистика. – 2004.
5. Борисов В.В., Круглов В.В., Федулов А.С. Нечеткие модели и сети. – М.: Горячая линия – Телеком, 2007. – 284 с.
6. Гладков Л.А., Курейчик В.В., Курейчик В.М. Генетические алгоритмы. М.: Физматлит. – 2010.
7. Емельянов В.В., Курейчик В.В., Курейчик В.М. Теория и практика эволюционного моделирования. М.: Физматлит. – 2003.
8. Гладков Л.А., Курейчик В.М., Курейчик В.В., Сороколетов П.В. Биоинспирированные методы в оптимизации. М.: Физматлит. – 2009.
9. Карпенко А. П. Современные алгоритмы поисковой оптимизации. Алгоритмы, вдохновленные природой. – М.: изд-во МГТУ им. Баумана. –2016.
10. Пегат А. Нечеткое моделирование и управление. М.: БИНОМ. Лаборатория знаний. – 2009.
11. Herrera F., Lozano M. Fuzzy Adaptive Genetic Algorithms: design, taxonomy, and future directions. // Soft Computing 7(2003), Springer-Verlag. – 2003. – P. 545-562.
12. Alba E., Tomassini M. Parallelism and evolutionary algorithms // IEEE T. Evolut. Comput. – 2002. – Vol. 6. – P. 443-461.
13. Кныш Д.С., Курейчик В.М. Параллельные генетические алгоритмы: Проблемы, обзор и состояние // Известия РАН. Теория и системы управления. – 2010. – № 4. – С. 72-82.
14. Гладков Л.А. Интегрированный алгоритм решения задач размещения и трассировки на основе нечетких генетических методов. // Известия ЮФУ. Технические науки, № 7 (120), 2011. – Таганрог: Изд-во ТТИ ЮФУ, 2011. – С. 22-30.
15. Michael A., Takagi H. Dynamic control of genetic algorithms using fuzzy logic techniques. Proc. of the 5th International Conference on Genetic Algorithms. Morgan Kaufmann. P. 76-83 (1993).
16. Гладков Л.А., Гладкова Н.В. Решение задач поиска и оптимизации решений на основе нечетких генетических алгоритмов и многоагентных подходов. // Известия ТРТУ. Тематический выпуск «Интеллектуальные САПР». – Таганрог: Изд-во ТРТУ, 2006. №8(63). – С. 83-88.

УДК 658.512.2.011.56

ИНТЕГРАЦИОННЫЙ ПОДХОД К РЕШЕНИЮ ЗАДАЧ ПРОЕКТИРОВАНИЯ (РАЗМЕЩЕНИЕ)*

Щеглов С.Н., *к.т.н.*
e-mail: srg_sch@mail.ru
ИТА ИКТИБ ЮФУ

1. ВВЕДЕНИЕ

Проектирование БИС и СБИС в настоящее время зависит от возрастания информационной сложности объектов автоматизированного проектирования и практически исключает возможность использования традиционных методов. Распространение в САПР начинает получать новая технология – гибридное моделирование, основная на методах, инспирированных природными системами. Это особенно актуально при переходе на современные нормы проектирования. При конструкторском проектировании основные задачи – это компоновка, размещение, трассировка, верификация и экстракция. В настоящее время широкое распространение получили новые идеи и методы реализующие задачи компоновки, размещения, а также построения стандартных и IP-блоков [1]. Принципиально новая методология проектирования систем на кристалле основана на многократном повторном использовании на всех этапах ранее созданных IP-блоков и их совокупности [2]. Это аналогично эволюции в живой и неживой природе и обосновывает разработку и применение инспирированных природой алгоритмов в САПР.

Одним из важнейших этапов автоматизации проектирования является размещение элементов на непрерывном монтажном пространстве [1-3]. Каждый год появляются новые технологические процессы со сложнейшим перечнем технологических ограничений, которые делают процесс проектирования более трудоемким. Уменьшение размеров, переход к проектированию в наносекундном диапазоне приводит к тому, что на первый план выходят проблемы учета задержек распространения сигналов и энергопотребления блоков. Это приводит к необходимости разработки новых параллельно-последовательных алгоритмов размещения с доработкой технологии.

* Работа выполнена при финансовой поддержке РФФИ, проект № 19-01-00715

2. ПОСТАНОВКА ЗАДАЧИ

Пусть дано множество конструктивных элементов (блоков), связанных между собой в соответствии с принципиальной схемой узла. Требуется разместить элементы на некотором плоском коммутационном поле (КП) так, чтобы выполнялись все ограничения на взаимное расположение элементов, а некоторый функционал достигал экстремального значения. Структура задачи задается посредством ограничений пространства размещения (габариты кристалла, конкретного корпуса или монтажной платы), характеристик элементов размещения, а также множеством цепей, соединяющих данные элементы. Главные метрические критерии задачи размещения: суммарная длина межсоединений и площадь области размещенных элементов и межсоединений, длина задержки сигналов, энергосбережение, а также их производные [2-5]. Основной целью решения задачи размещения следует считать создание наилучших условий для последующей трассировки.

Представление устройства в виде совокупности блоков разного уровня определяет формальную структурную модель, в которой каждый блок содержит блоки нижних уровней. Такой иерархический подход позволяет снизить временную сложность задач автоматизированного проектирования, разбивая её на подзадачи меньшей размерности. Таким образом, иерархическое деление позволяет организовать описание и хранение данных, при котором конструкторско-технологические ограничения не нарушаются [6-9].

При построении целевой функции (ЦФ) размещения будем использовать единый функционал на основе аддитивных и мультипликативных критериев.

В общем виде задачи размещения неформально можно представить следующим образом. Дано множество элементов, находящихся в отношении связности, в соответствии с принципиально-электрической схемой создаваемого объекта. Требуется разместить элементы внутри коммутационного пространства, таким образом, чтобы заданная целевая функция достигала локального или оптимального значения [1, 2, 7].

На прямоугольную конструкцию накладывается декартова система координат с осями s и t, определяющая граф G_r, представляющий собой координатную решетку.

Задача размещения сводится G_r к отображению заданного графа-модели $G = (X, U)$ схемы в решетку таким образом, чтобы множество вершин $X = \{x_i\}$, $i = \overline{1, n}$, графа G размещалось в узлах решетки, число

которых конечно, а также соблюдался интегрированный критерий *Compl(G)*, представляющий собой интегрированную ЦФ [1, 2, 7, 9]:

$$\begin{cases} Compl(G) = Compl_Kr(G), \text{ если } N_Kr > 0 \text{ и } N_GA = 0; \\ Compl(G) = L_GA(G), \text{ если } N_Kr = 0 \text{ и } N_GA > 0; \\ Compl(G) = \alpha L_GA(G) + \beta Compl_Kr(G), \text{ если } N_Kr > 0 \text{ и } N_GA > 0, \end{cases}$$

где

- *L_Kr(G)* – длина критических связей;
- *L_GA(G)* – величина суммарной длины соединений;
- *L_Dr(G)* – суммарная длина задержек на КС;
- *L_Er(G)* – суммарное электропотребление на КС (оно связано с суммарной площадью областей, занятых цепями КС);
- *N_Kr>0 и N_GA=0* – условие оптимизации ЦФ, основанное на соблюдении критерия длин критических связей;
- *N_Kr=0 и N_GA>0* – условие оптимизации ЦФ, основанное на выполнении критерия суммарной длины соединений;
- *N_Kr>0 и N_GA>0* – условие оптимизации ЦФ, основанное на одновременном соблюдении критериев длин критических связей и суммарной длины соединений;
- *α, β* – коэффициенты полезности вышеуказанных критериев, которые определяются на основании конструкторского опыта или экспертных оценок;
- *Compl_Kr(G)* – комплексный критерий оценки качества размещения определяется формулой:

$$Compl_Kr(G) = W_1 L_Kr(G) + W_2 L_GA(G) + W_3 L_Dr(G) + W_4 L_Er(G).$$

Коэффициенты $W_1 + W_2 + W_3 + W_4 = 1$ и определяются на основе нечеткого множества {0,1}.

Требование оптимизации: Compl_Kr(G)→min, т.е. необходимо, чтобы весовая функция была наименьшей для всевозможных способов отождествления вершин графа и узлов решетки. Приведенные критерии косвенно связаны с критерием суммарной длины соединений. Они получили распространение в существующих эвристиках размещения. При использовании этих критериев происходит учет тепловой и электромагнитной совместимости элементов, повышается эффективность упаковки, трассировки и сжатия области коммутационного пространства.

Использование комплексного критерия при разработке алгоритмов размещения позволяет повысить качество полученной конструкции за счет учета реальных знаний эксперта–конструктора. Это связано с тем, что основной задачей размещения элементов

является получение конструкции не только с минимальной длиной соединений, но и с созданием условий для последующих этапов трассировки соединений, сжатия и верификации.

3. МОДИФИЦИРОВАННАЯ ТЕХНОЛОГИЯ ИПА

В современных системах конструкторского проектирования эффективны различные стратегии, концепции, методы, механизмы. В задачах размещения любое альтернативное решение представляется набором параметров x_i, представляющих один элемент некоторого множества альтернативных решений. Поиск решений в случайно выбранном направлении часто не приводит к квазиоптимальному реальному решению. В этой связи наиболее интересной является модифицированная технология ИПА (инспирированные природой алгоритмы), приведенная на рис.1.

Рис.1. Технология ИПА

Изначально на основе конструкторско-технологических ограничений при взаимосвязи с внешней средой сокращается область поиска допустимых решений. Далее, эта область анализируется, случайным или направленным образом генерируется начальная популяция или стартовое множество альтернативных решений, которое в дальнейшем развивается на основе выбранной эволюционной стратегии.

4. ИНТЕГРАЦИОННЫЙ МЕТОД

Рассмотренная ранее технология ИПА широко позволяет применять биоинспирированный подход к решению задач проектирования и управления. В инспирированных природой алгоритмах каждое альтернативное решение рассматривается как особь, развивающаяся в популяции – сообществе себе подобных. Для каждой хромосомы в популяции определяется значение целевой функции в соответствии с выбранным или построенным критерием оптимизации задачи размещения.

Предлагается следующая технология поиска при размещении элементов. Сначала на основе конструкторско-технологических ограничений сокращается область поиска допустимых решений. Далее анализируя эту область, случайным, направленным или комбинированным образом создаётся начальная популяция или стартовое множество альтернативных решений, которое в дальнейшем развивается на основе выбранной эволюционной теории [9, 10]. Поиск решений в случайно выбранном направлении часто не приводит к эффективному реальному решению. В этой связи здесь вводится специальные этапы направления поиска. На основе постановки задачи размещения определяется ЦФ. Далее на основе ЦФ производится анализ популяции альтернативных решений и селекция (отбор) хромосом для дальнейшего поиска оптимальных решений задачи размещения. В рассматриваемой схеме бионического поиска в зависимости от экспертной системы (шкалы) выбирается генетический или эволюционный алгоритм. В генетическом алгоритме новые решения в популяции формируются путём реализации различных генетических операторов (кроссинговер, мутация, инверсия, сегрегация, транслокация, транспозиция, удаления и вставки) [11]. В эволюционном алгоритме новые решения в популяции формируются путём реализации модификаций одного оператора мутации. После репродукции размер популяции остается постоянным. Для этого производится его уменьшения до прежних размеров с помощью механизма естественного отбора или принципа «выживания сильнейших». В данной работе используется и другой способ, когда популяция изменяется динамически после реализации каждого оператора. Далее вся процедура повторяется на новом шаге эволюции.

На рис.2 приведена укрупненная схема бионического алгоритма размещения. Здесь БА, ГА и ЭА – соответственно бионический, генетический и эволюционный алгоритмы размещения, ОР – оператор репродукции, изменяющий размер популяции альтернативных решений, ЭС – экспертная система, определяющая дальнейший ход поиска. Следовательно, БА состоит из пяти основных блоков [8].

БА = (ГА,ЭА,ОР,ЭС, правила останова).

Тогда ЭА и ГА можно представить в следующем виде:

ЭА = (P, N, МК, С, ЦФ, ОГР, ГУ, ОМ,ОР, G, p_i, L),

ГА = (P, N, МК, С, ЦФ, ОГР, ГУ, ГО (ОК, ОМ, ОИ, ОТ, ОС, ОТР, ОУ, ОВ), ОР, *G*, p_i, *L*). Здесь МК- метод кодирования хромосомы (альтернативного решения), С – селекция, ОГР и ГУ – ограничения и граничные условия на задачу размещения, ГО – генетические операторы, G – число генераций или поколений алгоритма, $p_i \in$ P – хромосома, а *L* – длина хромосомы.

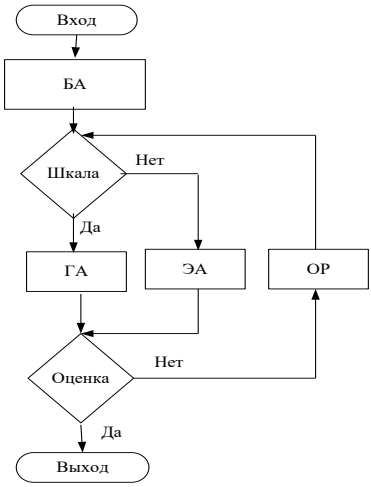


Рис.2. Укрупненная схема ИПА алгоритма размещения

Процесс в том случае, когда достигнута заданная точность, т.е. за несколько генераций сумма изменений значений лучших решений меньше или равна заданному значению. Процесс может также остановиться после заданного количества генераций алгоритма или после получения оптимального решения.

Принципиальных отличием БА является то, что возможно выполнение двух видов поиска быстрого эволюционного (ВСА одной генерации $\approx$ O(n)) и более медленного, но точного генетического

(ВСА одной генерации ≈ O(n) – O(n^3)). Кроме того, поиск решений осуществляется не из единственной точки, а из множества решений, и анализируются одновременно различные области пространства решений. БА более приспособлены к нахождению областей работоспособности за счёт объединения решений из различных подобластей.

5. ВЫЧИСЛИТЕЛЬНЫЕ ЭКСПЕРИМЕНТЫ

Для реализации бионического алгоритма размещения была разработана программа на языке программирования Borland C++ Builder 6.0. Рабочие окна программы показаны на рис.3.

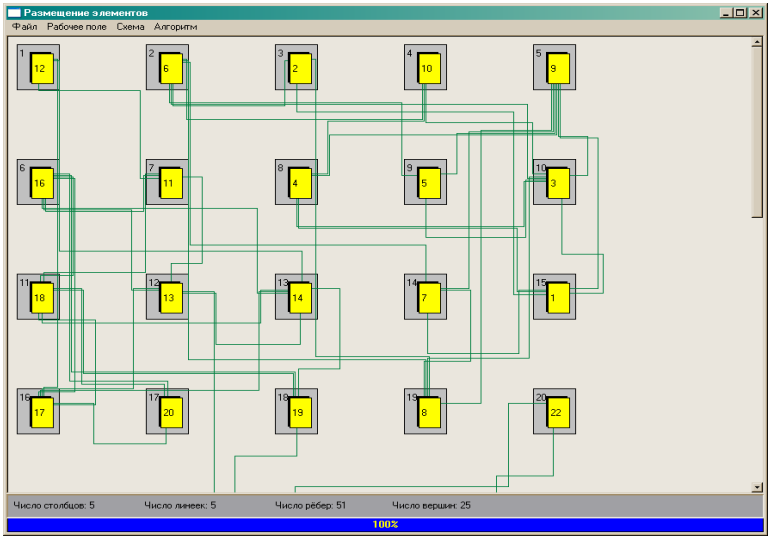


Рис.3. Результат работы алгоритма

В качестве изменяемых параметров выступают вероятности операторов, численность общей популяции, число итераций поиска. После нескольких запусков алгоритма получается окончательное размещение, выводится график работы бионического алгоритма. Помимо результата работы бионического алгоритма дается статистика: число элементов схемы, общая длина цепей (начальная и конечная), время работы алгоритма.

Эксперименты показали, что при размещении элементов СБИС на рабочем пространстве использование модифицированных ГО, нестандартных методов поиска и комбинированных моделей эволюции позволяет получать набор оптимальных решений. При этом с большой

вероятностью среди этих решений может быть найден глобальный экстремум. Из проведенных экспериментов следует, что в общем случае время решения линейно зависит от количества генераций и ВСА приближенно $O((n\log n) - (n^2))$. Анализ экспериментов позволяет отметить, что инспирированные природой алгоритмы требуют больших затрат времени, но позволяют получать набор локально-оптимальных решений, в частном случае – оптимальных решений.

На рис.4 построены графики зависимости времени получения лучшего решения от количества элементов для различных алгоритмов размещения.

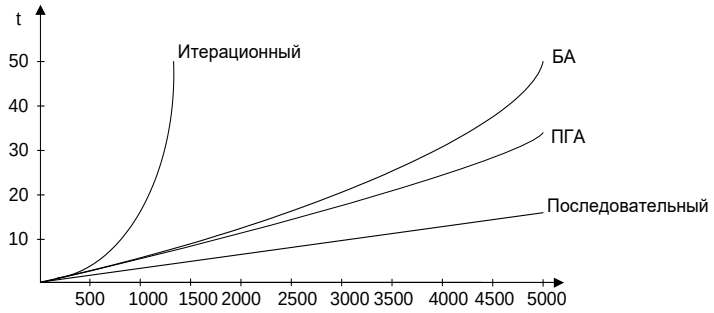


Рис.4. График зависимости времени решения от числа вершин для различных алгоритмов

6. ЗАКЛЮЧЕНИЕ

Таким образом, следует заметить, что с увеличением числа итераций время решения конечно повышается, но это повышение незначительное и компенсируется получением множества локально-оптимальных решений. Причем разработанные алгоритмы имеют большее быстродействие, чем метод ветвей и границ, метод моделирования отжига и другие аналогичные методы.

Литература

1. Курейчик В.В., Курейчик В.М., Гладков Л.А., Сороколетов П.В. Бионспирированные методы в оптимизации. – М.: Физмалит. – 2019.
2. Норенков, И.П. Основы автоматизированного проектирования [Текст]: учеб. пособие для вузов/ И.П. Норенков; ред. коллегия: И.Б. Федоров, В.В. Девятков, И.П. Иванов. и др. – 3-е изд., перераб. и доп. – М.: Изд. МГТУ им. Баумана, 2006. – 448 с.: ил. ; 24 см. – Библиогр.: с. 442. – 2000 экз. – ISBN 5-7038-2892-9.
3. Sherwani Naveed. Algorithms for VLSI Physical Design Automation, Kluwer Academic Publishers, Boston/Dordrecht/London. – 1995.

4. Gladkov L.A., Sheglov S.N., Gladkova N.V. The application of bioinspired methods for solving vehicle routing problems. // Procedia Computer Science, 120 (2017). 9th International Conference on Theory and Application of Soft Computing, Computing with Words and Perception, ICSCCW. – 2017. – P. 39-46.
5. Cohoon J.P. and Paris. W.D., Genetic Algorithms in Engineering Systems, The Institute of Electrical Engineers, London. – 1997.
6. P. Mazumder and E.M. Rudnick. Genetic Algorithms for VLSI Design, Layout and Test Automation. Prentice Hall, Toronto, Canada. – 1999.
7. Кулиев Э.В., Шеглов С.Н., Пантелюк Е.А., Логинов О.А. Адаптивный алгоритм стаи серых волков для решения задач проектирования. // Известия ЮФУ. Технические науки. № 7 (192). – 2017. С. 28-38.
8. Гладков Л.А., Курейчик В.В., Курейчик В.М., Сороколетов П.В. Биоинспирированные методы в оптимизации. Монография. – М.: Физматлит. –2010.
9. Гладков Л.А., Курейчик В.В., Курейчик В.М. и др. Анализ и исследование эволюционных методов решения задач разбиения СБИС. Монография. – Таганрог, ТТИ ЮФУ. – 2010.
10. Гладков Л.А., Курейчик В.В., Курейчик В.М., Родзин С.И. Основы теории эволюционных вычислений. Монография. – Ростов-на-Дону, изд-во ЮФУ. – 2010.

УДК 004.8; 519.6

СРАВНИТЕЛЬНАЯ ОЦЕНКА АЛГОРИТМОВ АВТОМАТИЧЕСКОГО ПОСТРОЕНИЯ НЕЛИНЕЙНЫХ МОДЕЛЕЙ НА ОСНОВЕ МЕТАЭВРИСТИЧЕСКОГО ПРОГРАММИРОВАНИЯ С ЭКСПРЕССИЕЙ ГЕНОВ*

***Монахов О.Г.**, к.т.н.*
e-mail: monakhov@rav.sscc.ru
***Монахова Э.А.**, к.т.н.*
e-mail: emilia@rav.sscc.ru
ИВМиМГ СО РАН

1. ВВЕДЕНИЕ

Рассмотрено решение задачи автоматического построения нелинейных моделей в виде математических выражений (функций, формул, алгоритмов и программ) на основе заданных наборов переменных, базовых функций, операций и экспериментальных данных. Необходимо найти математическое выражение f для наилучшего описания модели нелинейных вычислений, определяемой множеством входных *(X)* и выходных *(Y)* экспериментальных данных, т.е. необходимо найти функцию $Y = f(X)$, чтобы представить зависимость Y от X с минимальной ошибкой. Эта проблема также называется символьной регрессией или идентификацией системы.

Поиск выражения f осуществляется по заданному терминальному множеству переменных и констант $T_1 = \{x_i, c_i\}$ и заданному множеству базовых функций и операций $F_1 = \{f_i \mid f_i: R \times \ldots \times R \rightarrow R\}$. Эти множества используются для автоматического создания аналитических выражений (формул), представляющих модель. Мы предполагаем, что целевая функция (функция пригодности) FF вычисляет сумму квадратов отклонений выходных данных выражения f от заданных эталонных значений Y_i для определенных подмножеств входных данных выражения X_i , $1 \le i \le N$: $FF = \sum_{i=1}^{N} (f(X_i) - Y_i)^2$, где N – количество экспериментальных данных. Целью алгоритма поиска

* Исследование выполнено в рамках бюджетного проекта ИВМиМГ СО РАН № 0251-2021-0005

является определение минимума рассогласования (невязки): $min_{f\in D\,(F_1,T_1)}FF\,(f)$, где $D\,(F_1,T_1)$ – множество моделей, определяемых множествами переменных и констант, а также множеством базовых функций.

Алгоритмы генетического программирования (Genetic Programming – GP) [5], [8], [13] – один из известных подходов к этой проблеме. Они ориентированы в первую очередь на решение задач автоматического синтеза программ на основе обучающих данных с использованием генетического алгоритма в качестве механизма поиска моделей, минимизирующих ошибку представления функции. Хромосомы автоматически генерируются с использованием генетических операторов и имеют древовидные структуры, которые при интерпретации представляют собой некоторые выражения. Программирование экспрессии генов (Gene Expression Programming – GEP) [2], [3] является членом семейства генетического программирования, которое широко используется для поиска функций и открытия закономерностей. GEP имеет линейную хромосому, содержащую терминальные и функциональные символы в строке фиксированной длины, называемой *K*-выражением. *K*-выражение может быть переведено в индивидуальный фенотип в виде дерева экспрессии (ET) с использованием синтаксического анализа по ширине. Алгоритм поиска оптимума для GEP – это генетический алгоритм, который работает с хромосомами GEP для создания новых индивидуумов посредством операций кроссовера, мутации и транспозиции.

В этой работе мы предлагаем новый унифицированный подход метаэвристического программирования (МП) с экспрессией генов для автоматического синтеза нелинейных моделей в виде аналитического выражения, который использует: (1) генотип нелинейных моделей, представленный простым вектором действительных чисел (мы не используем сложные структуры, такие как деревья, сети или программы, традиционные для генетического программирования), (2) новые простые алгоритмы для преобразования этих векторов (генотип) в фенотип (в *K*-выражения) для представления нелинейной модели, (3) почти любой[1] метаэвристический, биоинспирированный алгоритм непрерывной глобальной оптимизации [4], [11], [14] с простыми операторами поиска (а не специализированными операторами для более сложных структур) для поиска оптимальной модели, (4) метод многовариантного кодирования нескольких решений

[1] Все алгоритмы (порядка 100), просмотренные авторами из [4], [11], [14]

в одном генотипе. Таким образом, одной из отличительных особенностей предложенного подхода является его универсальность при использовании различных метаэвристических алгоритмов непрерывной глобальной оптимизации для автоматизации синтеза нелинейных моделей.

Отметим, что метод МП является обобщением и дальнейшим развитием метода многовариантного эволюционного синтеза (МВЭС), предложенного авторами [6], [7]. Метод МП: (1) более унифицирован и универсален для использования со многими метаэвристическими (био- и природно-инспирированными) алгоритмами, а не только использует один дискретный генетический алгоритм (GA) или только одну дифференциальную эволюцию (DE) в качестве метода поиска как в МВЭС, (2) использует в качестве хромосомы для представления фенотипа с помощью экспрессии генов вектор действительных чисел, а не в виде последовательности инструкций, как в методе МВЭС, и, соответственно, (3) использует другие операции для декодирования хромосом. От метода, предложенного авторами в [12], новый метод отличается представлением фенотипа с помощью экспрессии генов и *K*-выражения, а не в виде программы (последовательности) трехадресных инструкций.

В следующем разделе мы описываем унифицированный МП с алгоритмом экспрессии генов, основанным на четырнадцати различных природно-инспирированных алгоритмах для синтеза нелинейных моделей. В последнем разделе мы сравниваем наш алгоритм со стандартным программированием экспрессии генов и с применением различных алгоритмов, вдохновленных природой, и показываем, что алгоритм МП имеет более высокую эффективность эволюционного поиска для многих случаев использования алгоритмов, вдохновленных природой.

2. МЕТАЭВРИСТИЧЕСКОЕ ПРОГРАММИРОВАНИЕ С ЭКСПРЕССИЕЙ ГЕНОВ

Метаэвристическое программирование с экспрессией генов для синтеза нелинейных моделей основано на моделировании природных процессов в популяции особей [4], [11], [14]. Отдельные особи в популяции представляют собой хромосомы, а именно векторы действительных чисел, которые кодируют математические выражения, и каждая хромосома в этом алгоритме определяет набор выражений (формул), возникающих из нее после декодирования. Алгоритм начинается с создания начальной популяции случайным оператором, затем отбираются лучшие особи путем декодирования генотипа (хромосомы) в фенотип (выражение) и

вычисления функции пригодности (приспособленности). Для создания следующего поколения новые особи создаются с использованием операторов миграции (в случае стандартного генетического алгоритма операторами миграции являются отбор, мутация и кроссовер). Цель операторов миграции – продвижение к экстремуму целевой функции.

Представленный в статье алгоритм метаэвристического программирования основан на следующих основных этапах.

- начальная популяция генерируется случайным образом из векторов (хромосом) с дейсвительными компонентами в интервале [0, 1]. Хромосома (генотип) кодирует решение (фенотип).
- пригодность каждой особи популяции вычисляется после процедуры декодирования генотипа в фенотип.
- для создания популяции следующего поколения используются операторы миграции биоинспирированного алгоритма, чтобы переместить особей в этой популяции к экстремуму целевой функции.
- повторение шагов 2 и 3 до тех пор, пока не будет найдено решение, удовлетворяющее указанным критериям, или пока не будет достигнуто максимальное количество поколений.

В классических GEP [2], [3] хромосомы (гены) состоят из головы (head) и хвоста (tail). Голова содержит символы, которые представляют как функции (элементы из базового множества функций *F*), так и терминалы (элементы из множества терминалов *T*), тогда как хвост содержит только терминалы. Следовательно, два разных алфавита встречаются в разных частях гена. Для каждой задачи выбирается длина головы *h*, тогда как длина хвоста *t* является функцией *h* и числа аргументов базовой функции с наибольшим количеством аргументов *n* и оценивается уравнением $t = h\,(n-1) + 1$.

В предлагаемом методе метаэвристического программирования с экспрессией генов предлагается единый подход к декодированию основных структур данных, хромосом, для различных природно-инспирированных алгоритмов. Последовательность действительных чисел (генотип) делится на три группы элементов (для головы, хвоста и констант):

$$\left(\{m_{1j}\},\{m_{2j}\},\{m_{3j}\}\right),\ 0 < m_{ij} < 1;\ \left|\{m_{1j}\}\right| = h;\ \left|\{m_{2j}\}\right| = t;\ \left|\{m_{3j}\}\right| = t.$$

Каждая такая группа интерпретируется разными алфавитами, которые рассчитываются по следующим формулам:

$$h_j = \left\lfloor m_{1j} * (|F| + |T|)\right\rfloor;\ t_j = \left\lfloor \left(m_{2j} * |T|\right)\right\rfloor;\ c_j = \left\lfloor \left(m_{3j} * |C|\right)\right\rfloor,$$

где | *F* | – количество элементов множества базовых функций, | *T* | – количество элементов терминального множества, | *C* | – число констант.

Таким образом, декодирование хромосомы приводит к тому, что функция представляется как интерпретируемое выражение. Но мы использовали многовариантный алгоритм для декодирования выражения в несколько подвыражений. В этом многовариантном алгоритме мы начинаем чтение с первого символа головной части хромосомы, чтобы построить первое выражение, которое полностью аналогично классическому GEP. Затем мы продолжаем читать второй символ головы, чтобы получить второе выражение. И повторяем чтение символов головы начиная с каждого символа до конца головы. Генетическим решением в этом случае является набор функций экспрессии. Это позволяет, в отличие от стандартного GEP, одновременно оценивать множество выражений в виде последовательности операторов и может сократить время на поиск оптимального решения. Здесь оценка определенной хромосомы из полученного множества выбрана как оценка выражения, имеющего минимальное значение функции приспособленности. Обратим внимание, что для данной хромосомы K существует множество выражений для $\{E_i; 1 \leq i \leq h\}$; значение целевой функции вычисляется для каждого выражения, и пригодность хромосомы определяется как соответствие наилучшему выражению, кодируемому этой хромосомой. Таким образом, в отличие от алгоритма GEP, который кодирует одно решение в хромосоме, новый алгоритм кодирует несколько решений в хромосоме, что также выполняется в программировании экспрессии генов на основе множественных выражений (MGEP) [1] и в [6], [7], но без единого подхода для применения различных природно-инспирированных алгоритмов в их эволюции.

Предлагаемое метаэвристическое программирование с экспрессией генов для синтеза нелинейных моделей объединяет: (1) линейное представление хромосомы, постоянный размер которой предотвращает эффект необоснованного роста выражений, (2) простые операции при расшифровке генотипа в фенотип для интерпретации хромосомы как выражения и (3) многовариантный метод для представления множества моделей (выражений) через одну хромосому. Простой алгоритм декодирования хромосомы (вектора) с действительными значениями в множество выражений позволяет использовать единый подход к поиску оптимальных моделей с использованием различных популяционных алгоритмов, вдохновленных природой [4], [11], [14].

3. ЭКСПЕРИМЕНТАЛЬНЫЕ РЕЗУЛЬТАТЫ

Мы исследовали эффективность метаэвристического программирования с экспрессией генов (mpge), используя

четырнадцать различных алгоритмов, вдохновленных природой, описанных в [4], [11], [14]: (1) Artificial Bee Colony (ABC), (2) Continuous Ant Colony Optimization (ACOR), (3) Bees Algorithm (BA), (4) Biogeography-based Optimization (BBO) [9], (5) Covariance Matrix Adaptation Evolution Strategy (CMA-ES), (6) differential evolution (DE) [10], (7) Firefly Algorithm (FA), (8) genetic algorithm (GA), (9) Harmony Search (HS), (10) Imperialist Competitive Algorithm (ICA), (11) Invasive Weed Optimization (IWO), (12) Particle Swarm Optimization (PSO),
(13) Simulated Annealing (SA), (14) Teaching-Learning based Optimization (TLB), для задачи поиска аналитических функций и параметров моделей на основе заданных экспериментальных данных, наборов переменных, базовых функций и операций. Мы использовали следующие десять тестовых функций (таблица1):

Таблица 1.

№	Определение функции	Число переменных	Диапазон
f1	$f1(x) = x^4 + x^3 + x^2 + x$	1	[1,10]
f2	$f2(x) = \sin(x^2 + x^4)$	1	[1,10]
f3	$f3(x) = \sin(\exp(\sin(\exp(\sin(x)))))$	1	[1,10]
f4	$f4(x) = \sin(x^3) + e^x$	1	[1,10]
f5	$f5(x) = x^5 - 2x^3 + x$	1	[1,10]
f6	$f6(x) = \sin(x) + \sin(x^2 + x)$	1	[1,10]
f7	$f7(x) = \sin(x^2)\exp(x) - 1$	1	[1,10]
f8	$f8(x, y) = 2\sin(x)\exp(y)$	2	[1,10]
f9	$f9(x, y) = \sin(x) + \sin(y^2)$	2	[1,10]
f10	$f10(x, y) = x^3 + y * x^2 + y$	2	[1,10]

В экспериментах использовались значения каждой функции в 10 случайных точках из диапазона [1,10]. Набор основных функций для тестовых функций: F = {+, -, ∗, sin, exp}, терминальные символы для *f1-f7 – T1 = {x}*, для *f8 – f10 – T2 = {x, y}*.

Применяются следующие параметры сравниваемых алгоритмов: вероятность кроссовера 0,80, вероятность мутации 0,15, размер популяции 200 и максимальное количество поколений 200. Длина хромосом 40. Для GEP мы использовали два варианта параметров: размер популяции 200, максимальное количество поколений 200 (для *classGEP1*), размер популяции 1200 и максимальное количество поколений 600 (для *classGEP2*). Для каждого алгоритма и каждой тестовой функции эксперименты проводились 50 раз, и результаты усреднялись. В экспериментах мы использовали значения параметров алгоритмов, рекомендованные их авторами в

цитируемых книгах и статьях для корректного сравнения эффективности. Параметры алгоритма отдельно не оптимизировались.

Метаэвристическое программирование с использованием четырнадцати различных естественных алгоритмов было реализовано с использованием ресурсов Информационно-вычислительного центра (ИВЦ) Новосибирского национального исследовательского государственного университета (НГУ) на процессорах Intel Xeon X5670, 2,93 ГГц (Westmere). Мы также сравнивали МП, основанное на вдохновленных природой алгоритмах, реализованных в MATLAB, с классической реализацией алгоритма программирования экспрессии генов (GEP) в MATLAB.

Один из основных показателей для измерения эффективности алгоритмов синтеза – вероятность (частота) успеха, то есть вероятность того, что алгоритм найдет выражение, точно совпадающее с тестовой функцией. Это – отношение количества успешных экспериментов, когда алгоритм нашел правильное выражение, к общему количеству экспериментов с заданными параметрами. В таблице 2 представлены вероятности (частоты) успеха для различных алгоритмов, где $\bar{p}$ – средний показатель успеха для каждого алгоритма по набору тестовых функций. Сравнение всех четырнадцати различных природно-инспирированных алгоритмов по средней вероятности успеха в порядке убывания показано на рис.1. Видно, что вероятность успеха в МР выше, чем у *classGEP1* в 11 случаях (в следующем порядке: *mpgeABC* > *mpgeTLB* > *mpgeFA*> *mpgeDE* > *mpgeBBO* > *mpgeHS* > *mpgeACO* > *classGEP2* > *mpgePSO* > *mpgeGA* > *mpgeCMA* > *classGEP1*), а вероятность успеха в МП выше, чем у *classGEP2* в 6 случаях. Этот результат объясняется способностью МП представлять несколько выражений на одной хромосоме, что приводит к более высокому шансу найти решение при одинаковом числе итераций.

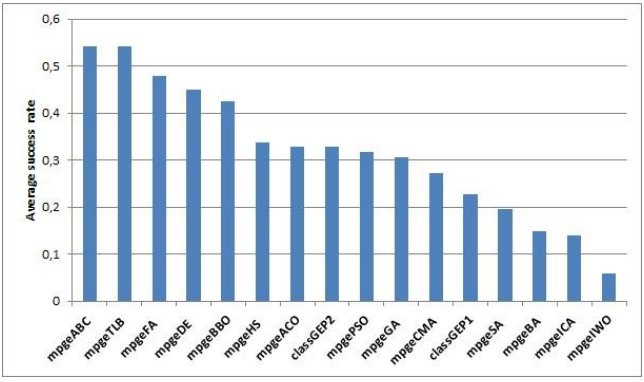


Рис.1. Средняя вероятность успеха для каждого алгоритма

Таблица 2. Вероятность успеха для каждого алгоритма на наборе тестовых функций

	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	$\bar{p}$
mpgeABC	0,63	0,08	0,96	0,33	0,08	0,63	0,38	0,79	0,92	1	0,542
mpgeTLB	0,67	0,08	1	0,08	0	0,75	0,54	0,96	0,75	1	0,541
mpgeFA	0,17	0,04	1	0,38	0,13	0,46	0,13	0,63	1	1	0,479
mpgeDE	0,67	0	1	0,04	0	0,42	0,42	0,75	0,5	1	0,45
mpgeBBO	0,92	0,04	0,88	0,08	0	0,2	0,38	0,79	0,33	1	0,425
mpgeHS	0,42	0,08	0,96	0,04	0	0,2	0	0,33	0,33	1	0,337
mpgeACO	0,5	0	0,5	0,04	0	0,33	0,17	0,65	0,46	0,96	0,328
classGEP2	0.13	0	1	0,25	0	0,21	0,04	0,33	0,67	0,46	0,329
mpgePSO	0,33	0,04	0,97	0	0	0,21	0,13	0,38	0,38	0,79	0,317
mpgeGA	0,63	0	0,45	0,04	0,29	0,17	0,25	0,75	0,13	0,79	0,306
mpgeCMA	0,42	0	0,96	0,13	0	0	0,08	0,54	0,33	0,33	0,273
classGEP1	0	0	0,96	0,17	0	0,08	0,04	0,29	0,33	0,41	0,228
mpgeSA	0,38	0	0,42	0	0	0,04	0,17	0,29	0	0,46	0,195
mpgeBA	0	0	0,79	0,04	0	0,08	0	0,04	0,29	0,04	0,15
mpgeICA	0,08	0	0,17	0	0	0,04	0	0,25	0,08	0,88	0,139
mpgeIWO	0	0	0,21	0	0	0	0	0,08	0,04	0,25	0,058

В таблице 3 показано время выполнения (в секундах) для различных алгоритмов, т. е. среднее время выполнения алгоритма до того, как алгоритм найдет (синтезирует) заданное выражение или выполнит необходимое количество поколений, где $\bar{t}$ – среднее время выполнения каждого алгоритма для набора тестовых функций. Сравнение всех четырнадцати различных биоинспирированных алгоритмов по среднему времени выполнения $\bar{t}$ в порядке возрастания показано на рис.2. Этот результат показывает, что у МП меньше времени на поиск решения, чем у *classGEP2* во всех случаях, и меньше времени на поиск решения, чем у *classGEP1* в трех случаях (в следующем порядке: *mpgeHS* < *mpgeACO* < *mpgePSO* < *classGEP1*), но алгоритм *classGEP1* имеет низкую среднюю вероятность успеха. Например, *mpgeTLB* имеет средний уровень успеха на 60% выше, чем *classGEP2*, и на 40% меньше среднего времени выполнения, чем *classGEP2*. В случае с *classGEP1*, *mpgeTLB* имеет средний показатель успешности на 134% выше и только на 11% выше среднего времени выполнения, чем *classGEP1*. Отметим, что два алгоритма (*mpgeHS и mpgeACO*) лучше, чем *classGEP* по обоим критериям одновременно. Этот результат можно объяснить простотой структур данных и операций в МП.

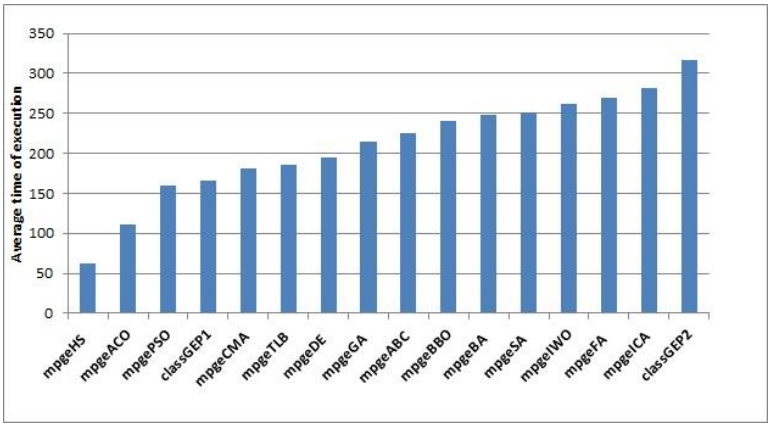


Рис.2. Среднее время выполнения каждого алгоритма

Таблица 3. Среднее время выполнения каждого алгоритма для набора тестовых функций

	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	$\bar{t}$
mpgeDE	204	295	24	265	289	233	214	145	167	66.7	195,3
mpgeGA	164	298	195	232	263	258	231	81	214	70,6	214
mpgePSO	201	203	36	174	232	195	209	150	131	58.9	160
mpgeTLB	229	278	7,8	287	348	125	236	65	99	55	185
mpgeSA	202	303	248	250	310	306	268	211	301	164	251,3
mpgeIWO	280	310	256	287	275	301	297	207	209	192	261,4
mpgeICA	88	381	334	307	379	330	372	267	264	80	281,6
mpgeHS	77	73	26	71	84	62	90	57	54	25	62
mpgeFA	314	323	199	293	332	293	334	220	156	186	270
mpgeBBO	79	424	82	369	400	329	308	100	156	19	240,7
mpgeBA	286	273	139	259	280	266	295	201	184	198	247,8
mpgeACO	114	158	9,19	149	165	126	140	97	105	34	111
mpgeABC	264	326	79	285	340	223	314	157	112	41	225,4
mpgeCMA	216	195	13,7	236	332	164	267	92	117	115	202
classGEP1	202	178	46,7	161	211	164	284	172	115	129	166.3
classGEP2	387	345	47,7	309	425	313	564	318	196	259	316,4

4. ЗАКЛЮЧЕНИЕ

Мы рассмотрели новый подход метаэвристического программирования с экспрессией генов для решения задачи автоматического построения нелинейных моделей (математических выражений и программ) на основе заданных базовых функций и операций, экспериментальных данных и наборов переменных. Для

синтеза таких моделей было разработано метаэвристическое программирование как единый подход для использования почти любых алгоритмов, вдохновленных природой. Этот подход использует линейное (в виде вещественных векторов) представление хромосомы как генотипа, простые операции для декодирования генотипа в фенотип для интерпретации хромосомы как математического выражения и многовариантный метод для представления множества моделей (выражений) с использованием единственной хромосомы. Предложенный подход МП с экспрессией генов был реализован с использованием четырнадцати различных алгоритмов, вдохновленных природой, и эти алгоритмы сравнивались друг с другом и со стандартным алгоритмом программирования экспрессии генов. Эксперименты показывают, что мы всегда можем выбрать наиболее подходящий алгоритм, лучше, чем алгоритм GEP, на основе предпочтительного критерия. Показано, что предложенный подход МП с экспрессией генов при использовании различных биоинспирированных алгоритмов превосходит алгоритм GEP либо по времени нахождения решения, либо по вероятности нахождения функции (модели), либо по обоим критериям одновременно.

Литература

1. Deng W., He P., Huang Z. Multi-Expression Based Gene Expression Programming// Proceedings of 2013 Chinese Intelligent Automation Conference. Lecture Notes in Electrical Engineering. – 2013. – Vol 256. – P. 439-448.
2. Ferreira C. Gene expression programming: A new adaptive algorithm for solving problems// Complex Systems. – 2001. – Vol.13. – P. 87-129.
3. Ferreira C. Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence. 2nd Edition. – Heidelberg: Springer, 2006.
4. Ke-Lin Du, Swamy M. N. S. Search and Optimization by Metaheuristics. Techniques and Algorithms Inspired by Nature. – Basel: Birkhauser, 2016.
5. Koza J. Genetic Programming II: Automatic Discovery of Reuseable Programs. – Cambridge, MA: MIT Press, 1996.
6. Monakhov O.G.: Differential Evolution for Multi-Variant Evolutionary Synthesisof Nonlinear Models// Proceedings of 2018 XIV International Scientific-Technical Conference on Actual Problems of Electronic Instrument Engineering (APEIE2018). – 2018. – P. 487-491.
7. Monakhov O., Monakhova E. An Algorithm of Multi-Variant Evolutionary Synthesis of Nonlinear Models with real-valued chromosomes// Decision Science in Action: Theory and Applications of Modern Decision Analytic Optimisation. – Singapure: Springer, 2019 – P. 41-49.
8. Poli R., Langdon W. B., McPhee N. F. A Field Guide to Genetic Programming. – San Francisco: Lulu.com, 2008.
9. Simon D. Evolutionary Optimization Algorithms. – Hoboken: Wiley, 2013.

10. Zaheer H., Pant M., Kumar S., Monakhov O., Monakhova E., Deep K. A newguiding force strategy for differential evolution// International Journal of System Assurance Engineering and Management. – 2017. – Vol.8. – Suppl.4. – P. 2170-2183.
11. Yang X.S. Nature-Inspired Optimization Algorithms. – London: Elsevier, 2014.
12. Monakhov O.G., Monakhova E.A. Development of a Metaheuristic Programming Method for Synthesis of Nonlinear Models// Numerical Analysis and Applications. – 2020. – Vol.13. – №4. – P. 349-359.
13. Емельянов В.В., Курейчик В. В., Курейчик В. М. Теория и практика эволюционного моделирования. – М.: ФИЗМАТЛИТ, 2003.
14. Карпенко А.П. Современные алгоритмы поисковой оптимизации. Алгоритмы, вдохновленные природой. – М.: Издательство МГТУ им. Н.Э. Баумана, 2014.

УДК 519.673

МУЛЬТИМЕМНАЯ МОДИФИКАЦИЯ КВАНТОВОГО АЛГОРИТМА ПРЕСНОВОДНЫХ ГИДР ДЛЯ РЕШЕНИЯ ЗАДАЧИ ОПТИМИЗАЦИИ

Майков Д.В.
e-mail: maykovd@yandex.ru
Ижевский торгово-экономический техникум

1. ВВЕДЕНИЕ

Решение ряда практически важных задач сопряжено с нахождением точки экстремума некоторой функции нескольких переменных (целевой функции). Такими задачами являются, например, оптимизация параметров технических систем [1], обучение нейронных сетей [2], идентификация параметров математической модели по экспериментальным данным [3], нахождение оптимального управления для динамических систем [4] и др.

В общем случае задача условной многомерной оптимизации (на примере задачи минимизации) имеет вид:

$$\mathbf{x}^{opt} = \arg \min_{\mathbf{x} \in D} f(\mathbf{x}).$$

Необходимо найти точку $\mathbf{x}$, в которой функция $f(\mathbf{x})$ достигает глобального экстремума в замкнутой n-мерной области D. Функция $f(\mathbf{x})$ может иметь несколько точек локальных экстремумов и не выражаться аналитически.

Решение данной задачи может быть получено с помощью биоинспирированных популяционных алгоритмов, таких, как алгоритм роя частиц, алгоритм колонии искусственных пчел и др. [5]. Эти алгоритмы обрабатывают множество особей – допустимых векторов из области D. Особи образуют популяцию.

Для повышения скорости сходимости глобального популяционного алгоритма часть особей подвергается улучшению с помощью алгоритмов локального поиска, которые называются мемами [5]. Эти алгоритмы могут быть как классическими (метод сопряженных градиентов и др.), так и популяционными (алгоритм роя частиц и т.д.).

Обычно применяются гиперэвристические адаптивные мультимемные алгоритмы, в которых мемы выбираются с помощью специальных правил из роя мемов. Подобные модификации известны для большинства популяционных алгоритмов, например, для алгоритмов дифференциальной эволюции [6], эволюции разума [7], муравьиного алгоритма непрерывной оптимизации [8] и др. В данной работе этот подход применен к квантовому алгоритму пресноводных гидр [9].

2. КВАНТОВЫЙ АЛГОРИТМ ПРЕСНОВОДНЫХ ГИДР

Данный алгоритм разработан в двух вариантах [9]. В первом варианте (который рассматривается в настоящей работе) выбор направления движения особи осуществляется с помощью метода анализа иерархий (QH-AHP-алгоритм), тогда как во втором варианте для этого применяется байесовский подход (QH-B-алгоритм).

На первом этапе работы алгоритма случайным образом формируется популяция из S особей $\mathbf{x}_j$, находящихся в области D.

На каждой итерации алгоритма выявляется лидер $\mathbf{x}^{**}$ – особь с наилучшим значением целевой функции и вектор $\mathbf{C}^k$ средней лучшей позиции всех особей популяции:

$$C_i^k = \frac{1}{S}\sum_{j=1}^{S} x_{ij}^* , \qquad (1)$$

где x_{ij}^* – координаты j-й особи с наилучшим значением целевой функции за все время поиска.

Предполагается, что особь гидры может двигаться в следующих направлениях: к средней лучшей позиции всех особей, к лидеру и случайном. В QH-AHP-алгоритме направление выбирается методом анализа иерархий. При этом предпочтения каждого направления выражаются элементами $a_{rp} \in \left\{\frac{1}{10}, \frac{1}{9}, ..., 1, 2, ..., 10\right\}$ матрицы парных сравнений A. Каждый из них показывает, насколько r-е направление предпочтительнее p-го $\left(r, p = \overline{1, m}\right)$, причем $a_{rp} = \frac{1}{a_{pr}}$ и $a_{rr} = 1$. Каждый элемент матрицы A делится на сумму элементов столбца, в котором он находится. При этом формируется матрица B. Приоритет каждого направления выражается его весовым коэффициентом

$$w_r = \frac{1}{m}\sum_{p=1}^{m} b_{rp} \,. \qquad (2)$$

Исходные значения элементов матрицы A показаны в таблице 1, они получены в результате решения задачи метаоптимизации [9].

Таблица 1. Начальные значения элементов матриц парных сравнений

Направление	К средней лучшей позиции	К лидеру	Случайное
К средней лучшей позиции	1	$\frac{1}{3}$	$\frac{1}{6}$
К лидеру	3	1	$\frac{1}{4}$
Случайное	6	4	1

С помощью матрицы парных сравнений согласно выражению (2) определяются приоритеты направлений. Перемещение особи осуществляется в направлении с наибольшим приоритетом.

Необходимо сгенерировать случайные величины α_{ij}, β_{ij}, γ_{ij} принадлежащие отрезку [0, 1], а также ξ_{ij} и η_{ij} со стандартным нормальным распределением.

Координаты аттрактора каждой особи имеют вид:

$$p_{ij}^k = \gamma_{ij}^k x_{ij}^* + \left(1 - \gamma_{ij}^k\right) x_i^{**},$$

где x_i^{**} – координаты лучшего решения (лидера) за все время поиска.

Координаты особей находятся следующим образом при их движении:

1) к вектору $\mathbf{C}$ лучших положений особей:

$$x_{ij}^{k+1} = \begin{cases} p_{ij}^k + \lambda^k \cdot \left| C_i^k - x_{ij}^k \right| \cdot \ln \dfrac{1}{\alpha_{ij}^k}, \beta_{ij}^k \geq 0{,}5, \\ p_{ij}^k - \lambda^k \cdot \left| C_i^k - x_{ij}^k \right| \cdot \ln \dfrac{1}{\alpha_{ij}^k}, \beta_{ij}^k < 0{,}5, \end{cases} \qquad (3)$$

2) в направлении к лидеру:

$$x_{ij}^{k+1} = x_{ij}^{k} + \lambda^{k} \cdot \xi_{ij}^{k} \cdot \left(x_{i}^{**} - x_{ij}^{k}\right) \cdot \alpha_{ij}^{k}. \quad (4)$$

3) в случайном направлении:

$$x_{ij}^{k+1} = \begin{cases} x_{ij}^{*} + \lambda^{k} \cdot \eta_{ij}^{k}, \ \beta_{ij}^{k} \geq 0{,}5, \\ x_{ij}^{*} - \lambda^{k} \cdot \eta_{ij}^{k}, \ \beta_{ij}^{k} < 0{,}5. \end{cases} \quad (5)$$

Параметр λ изменяется по формуле:

$$\lambda^{k} = \lambda_{1} + \frac{\lambda_{0} - \lambda_{1}}{k}, \quad (6)$$

где λ_{0} – его начальное значение (при $k = 1$), а к значению λ_{1} он стремится при $k \to \infty$.

Оператор переноса применяется, когда число итераций τ_{j}, в течение которых целевая функция j-й особи не улучшалась, становится равным пороговому значению $\tau_{\max}$. При этом текущая особь заменяется новой с координатами:

$$x_{ij}^{k+1} = \begin{cases} x_{ij}^{k} + \alpha_{ij} \cdot \left(x_{i}^{\max} - x_{ij}^{k}\right) \cdot g(k), \ \beta_{ij} \geq 0{,}5 \\ x_{ij}^{k} - \alpha_{ij} \cdot \left(x_{ij}^{k} - x_{i}^{\min}\right) \cdot g(k), \ \beta_{ij} < 0{,}5. \end{cases} \quad (7)$$

В данной работе использована функция вида:

$$g(q) = \frac{1}{2} \cdot \left(1 - \frac{q}{\mu + |q|}\right), \ q = k - \frac{K}{2},$$

с параметрами $\mu = 89$ и $K = 1000$.

В качестве критерия окончания поиска было выбрано условие:

$$\chi = \max_{q} \left| \frac{f^{k} - f^{k-q}}{f^{k}} \right| < 10^{-4}, \quad (8)$$

где $q = 1, ..., \min(k, \omega)$, $\omega = 100$.

Общая схема алгоритма имеет вид:

I. Создать начальную популяцию. Для каждой особи задать матрицу парных сравнений A_{j} (см. таблицу 1), определить значение целевой функции и положить количество итераций τ_{j}, в течение которых не происходило улучшение значения целевой функции, равным нулю. Найти лидера. Положить номер итерации $k = 1$.

II. По формуле (1) определить средний вектор **C** лучших положений особей популяции и значение параметра λ (6).

III. Для каждой j-й особи выполнить следующее.

1. Найти вектор приоритетов $\mathbf{w}$ (2).

2. Выполнить шаг в направлении с наибольшим приоритетом $p_{\max} = \arg\max_p w_p$ (формулы (3), (4), (5)). Если он одинаков у нескольких направлений, то выбрать любое из них.

3. Вычислить значение целевой функции. Если оно улучшилось, то в строке $i = p_{\max}$ матрицы A увеличить элементы a_{ij} (например, на 1, если $a_{ij} > 1$, и на 0,1, если $0 < a_{ij} < 1$) и положить τ_j равным нулю. В случае ухудшения уменьшить соответствующие элементы матрицы A и увеличить τ_j на единицу. Если при этом значение a_{ij} оказывается вне диапазона [0,1, 10], то положить его равным ближайшему граничному значению. Обновить лучшее положение особи $\mathbf{x}_j^*$, если это возможно.

4. Пересчитать элементы матрицы A, симметричные тем, которые изменились по формуле $a_{ji} = \frac{1}{a_{ij}}$.

5. Если $\tau_j = \tau_{\max}$, то применить оператор переноса (7), рассчитать значение целевой функции в новой точке и вернуть элементам матрицы A_j исходные значения (таблица 1).

IV. Найти лидера и проверить выполнение критерия останова (8). В случае выполнения завершить поиск, иначе положить $k = k + 1$ и перейти к шагу II (следующей итерации алгоритма).

3. МУЛЬТИМЕМНЫЙ КВАНТОВЫЙ АЛГОРИТМ ПРЕСНОВОДНЫХ ГИДР

В концепции С.Р. Докинза [5] мемы представляют собой единицу культурной информации (навыки, приемы и т.д.), влияют на поведение всей популяции и служат аналогами генов. Мемы могут быть переданы как в вертикальном направлении (от предков к потомкам), так и в горизонтальном (от одной особи к другой). Множество M мемов образует рой мемов.

Мемы (как алгоритмы локального поиска) применяются не ко всем особям популяции, а к некоторой ее части, например, к заданному количеству E особей с лучшим значением целевой функции (элите). Если в результате этого значение целевой функции особи улучшается, то ее координаты заменяются новыми.

Правило выбора мема из роя принято называть гиперэвристикой [5]. В данной работе был выбран подход на основе концепции самоадаптирующихся мультимемных алгоритмов. При этом для каждой j-й особи мерой эффективности мема являлось удельное приращение значений целевой функции до ($f_0^k(\mathbf{x}_j)$) и после ($f_1^k(\mathbf{x}_j)$) применения m-го мема:

$$\varphi_{jm}^k = \left| \frac{f_0^k(\mathbf{x}_j) - f_1^k(\mathbf{x}_j)}{f_0^k(\mathbf{x}_j)} \right|. \qquad (9)$$

В данной работе предлагается следующая модификация самоадаптирующегося мультимемного алгоритма:

1. Выполнить пункт I QH-AHP- или QH-B-алгоритма. Создать новую популяцию с заданным числом E особей с лучшим значением целевой функции (элиту) и разбить ее на M групп. К первой группе особей применить первый мем, ко второй – второй мем и т.д. Указать номер применяемого мема N_j для каждой особи и оценку его эффективности (9).

2. С заданной вероятностью случайным образом выбирать пары элитных особей и осуществлять "дискуссию" внутри каждой пары, т.е. сравнивать эффективность (9) применяемых мемов. Оба участника получают один номер N_j мема, у которого эффективность максимальна. Таким способом осуществляется горизонтальная передача мемов.

3. С определенной вероятностью осуществить мутацию мемов элитных особей, когда основной мем N_j случайно заменяется другим.

4. К каждой j-й элитной особи применить N_j-й мем. При этом для каждого алгоритма локального поиска (мема) сформировать свою субпопуляцию из элиты. По результатам обновить значение эффективности (9) соответствующего мема и значение N_j принять равным номеру лучшего мема. При этом происходит вертикальная передача мемов.

5. Выполнить итерацию алгоритма глобального поиска для основной популяции (пункты II и III QH-AHP- или QH-B-алгоритма).

6. Обновить состав элиты, выбрав E лучших особей среди прежней элиты и особей основной популяции.

7. Найти лидера и проверить выполнение критерия останова (8). В случае выполнения завершить поиск, иначе положить $k = k + 1$ и перейти к шагу 2 (следующей итерации алгоритма).

Пункты 2, 3, 4 можно выполнять через заданное (в общем случае динамическое) число итераций основного алгоритма.

4. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ

Для сравнения скорости сходимости различных алгоритмов применялись следующие тестовые функции [5]:

1) $f_1(\mathbf{x}) = \sum_{i=1}^{n-1} \left(x_{i+1}^2 + x_i^2\right)^{0,25} \times \left(\sin^2\left(50\left(x_{i+1}^2 + x_i^2\right)^{0,1}\right) + 1\right)$ – функция Дэвиса. Оптимальное решение $x_i^{opt} = 0$, $i = \overline{1, n}$, $f_2\left(\mathbf{x}^{opt}\right) = 0$.

2) $f_2(\mathbf{x}) = \sum_{i=1}^{n-1}\left(100 \cdot \left(x_{i+1} - x_i^2\right)^2 + \left(1 - x_i\right)^2\right)$ – функция Розенброка. Оптимальное решение $x_i^{opt} = 1$, $i = \overline{1, n}$, $f_1\left(\mathbf{x}^{opt}\right) = 0$.

3) $f_3(\mathbf{x}) = -20\exp\left(-0,2 \cdot \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}\right) -$

$- \exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos\left(2\pi x_i\right)\right) + \exp(1) + 20$ – функция Экли. Оптимальное решение $x_i^{opt} = 0$, $i = \overline{1, n}$, $f_3\left(\mathbf{x}^{opt}\right) = 0$.

4) $f_4(\mathbf{x}) = \sum_{i=1}^{n} 10\left(1 - \cos\left(2\pi x_i\right)\right) + x_i^2$ – функция Растригина. Оптимальное решение $x_i^{opt} = 0$, $i = \overline{1, n}$, $f_4\left(\mathbf{x}^{opt}\right) = 0$.

Рассматривались следующие алгоритмы оптимизации: квантовые алгоритмы пресноводных гидр на основе метода анализа иерархий (QH-AHP) и байесовского подхода (QH-B), а также меметические модификации соответствующих алгоритмов (MQH-AHP и MQH-B).

В качестве мемов применялись следующие алгоритмы: алгоритм роя частиц (PSO), алгоритмы пресноводных гидр на основе метода анализа иерархий (H-AHP) и байесовского подхода (H-B) [11]. Два последних алгоритма использовались в версии для локального поиска, т.е. без оператора переноса. Каждый мем выполнял 10 итераций в течение одной итерации основного алгоритма.

Графики скорости сходимости и количество особей в популяции каждого мема представлены на рис.1 и рис.2.

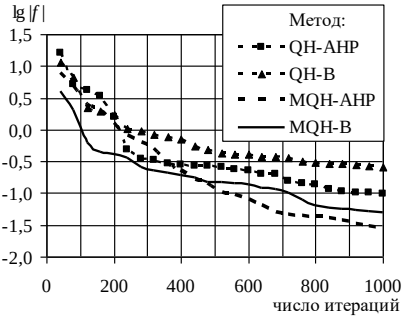


Рис.1. График скорости сходимости алгоритмов на примере тестовой функции Дэвиса

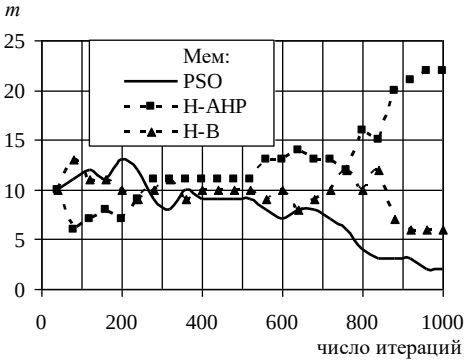


Рис.2. Количество особей в субпопуляциях, соответствующих каждому мему

Для всех тестовых функций меметические алгоритмы показали лучшие результаты по сравнению с исходными. В большинстве случаев скорость сходимости MQH-AHP-алгоритма была выше, чем у остальных. В процессе самоадаптации чаще других применялся мем H-AHP.

5. ЗАКЛЮЧЕНИЕ

В работе предложена меметическая модификация квантового алгоритма пресноводных гидр для решения задачи оптимизации. При этом применялась концепция самоадаптирующихся мультимемных

алгоритмов. В качестве мемов использованы алгоритмы роя частиц и пресноводных гидр. Выполнено тестирование скорости сходимости рассмотренных алгоритмов на примере различных многоэкстремальных функций. Лучшие результаты показала меметическая модификация квантового алгоритма пресноводных гидр на основе метода анализа иерархий. При этом в роли мема чаще всего выбирался алгоритм пресноводных гидр на основе метода анализа иерархий.

Литература

1. Дарьина А.Н., Прокопьев И.В. Метод оптимизации параметров контроллера беспилотного летательного средства на основе оптимизации методом роя частиц // Надежность и качество сложных систем. – 2020. – № 3 (31). – С. 80-87.
2. Загинайло М.В. Обучение искусственных нейронных сетей с помощью математического аппарата генетических алгоритмов // Инновации. Наука. Образование. – 2020. – № 12. – С. 384-389.
3. Королев С.А., Майков Д.В. Идентификация математической модели метаногенеза и исследование различных режимов метаногенеза в мезофильной среде // Компьютерные исследования и моделирование. – 2012. – Т. 4, № 1. – С. 131–141. DOI: 10.20537/2076-7633-2012-4-1-131-141
4. Королев С.А., Майков Д.В. Решение задачи оптимального управления процессом метаногенеза на основе принципа максимума Понтрягина // Компьютерные исследования и моделирование. – 2020. – Т. 12, № 2. – С. 357–367. DOI: 10.20537/2076-7633-2020-12-2-357-367.
5. Карпенко А.П. Современные алгоритмы поисковой оптимизации. Алгоритмы, вдохновленные природой. – М.: Изд-во МГТУ имени М.Э. Баумана, 2017.
6. Kumar S., Sharma V.K., Kumari R. Memetic search in differential evolution algorithm // International Journal of Computer Applications. – 2014. – Vol. 90, no. 6. – P. 40-47.
7. Сахаров М.К., Пономаренко А.В. Исследование эффективности мультимеметического алгоритма эволюции разума // Математика и математическое моделирование. – 2017. – № 06. – С. 70-82. DOI: 10.24108/mathm.0617.0000090
8. Карпенко А.П., Чернобривченко К.А. Мультимемеевая модификация гибридного муравьиного алгоритма непрерывной оптимизации HCIAC // Наука и образование: научное издание МГТУ имени Н. Э. Баумана. – 2012. – №9. – С. 261-296. DOI: 10.7463/0912.0470529
9. Королев С.А., Майков Д.В. Квантовая модификация алгоритма пресноводных гидр для решения задачи оптимизации // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. – 2020. – № 2. – С. 37-48. https://doi.org/10.17308/sait.2020.2/2914
10. Майков Д.В. Метаоптимизация алгоритма пресноводных гидр XXIV Международ. конф. «Информационно-вычислительные технологии и их приложения: сб. тр. Международн. конф. (Пенза, 27-28 августа 2020 г.). – Пенза: «РИО ПГАУ», 2020. – С. 77-84.

СЕКЦИЯ 8 | **ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ВОЕННОГО, ДВОЙНОГО И СПЕЦИАЛЬНОГО НАЗНАЧЕНИЯ**

УДК 004.891.3

ТЕОРЕТИЧЕСКИЕ И ПРИКЛАДНЫЕ ОСНОВЫ ПОСТРОЕНИЯ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ УПРАВЛЕНИЯ ЗАЩИТОЙ КОСМИЧЕСКИХ АППАРАТОВ ОТ УДАРОВ АНТРОПОГЕННЫХ (ТЕХНОГЕННЫХ) ЧАСТИЦ*

***Палюх Б.В.**, д.т.н.*
e-mail: pboris@tstu.tver.ru
***Кемайкин В.К.**, к.т.н.*
e-mail: vk-kem@mail.ru
ТвГТУ
***Цурков М.Л.**, к.т.н.*
e-mail: mlts69@yandex.ru
НИЦ (г.Тверь) ЦНИИ ВКС Минобороны России

1. ВВЕДЕНИЕ

Основные недостатки существующей системы управления КА, в том числе и защиты от ударов АЧ, определяются сложившимися жесткими

* Работа выполнена при финансовой поддержке Минобрнауки РФ (проект № 2.1777.2017/4.6)

принципами построения таких структур. Так автономное управление техническим состоянием КА реализуется на основе проводимого на борту оперативного контроля. Параметры контроля, обработанные по логическим уравнениям, позволяют определить причину неисправности и запустить на выполнение одну из аварийных программ. Перечень типовых неисправностей известен, и аварийные программы, обеспечивают восстановление работоспособности КА путем резервирования систем, снижения уровня энергопотребления, изменения положения внешних конструкций и т.д.

Такой подход не позволяет успешно противостоять случайным внешним воздействиям (в том числе ударам АЧ), определение угрозы от которых основано на оценке риска, и не может быть проведено средствами параметрического контроля.

Следовательно, для управления защитой КА от ударов АЧ необходимо использовать исключительно упреждающую стратегию, основу которой должна составлять способность адаптации к изменениям условий функционирования КА, обусловленным возможными ударами антропогенных (техногенных) космических частиц.

Анализ результатов работ в области искусственного интеллекта (ИИ) в сочетании с основными положениями теории автоматического управления позволяет ожидать позитивных результатов в обоснованной интеллектуализации систем автоматического управления защитой на основе применения современных методов и технологий обработки знаний [7].

В рамках создания данной теории необходимо комплексное решение ряда фундаментальных научных проблем, направленных на исследование и разработку новых теоретических моделей и методов для создания на их основе нового поколения интеллектуальных систем управления защитой КА:

- разработка моделей внешних механических угроз, здесь нужно выделить классы управляемых или неуправляемых внешних угроз соударения с АЧ, движение которых осуществляется по баллистическим и небаллистическим траекториям относительно разных притягивающих центров;
- создание математической теории прогнозирования, идентификации результатов возможных ударов АЧ, оценки и выбора эффективных защитных барьеров КА;
- создание методологии построения структуры интеллектуальных систем управления защитой КА от ударов АЧ;
- создание методологии построения систем поддержки принятия решений по управлению защитой КА от ударов АЧ.

Решение обозначенных фундаментальных научных проблем позволит создавать перспективные интеллектуальные системы управления защитой КА, основным преимуществом которых будет способность обнаруживать, предотвращать и компенсировать угрозы возможных ударов АЧ в условиях исходной внешней неопределённости.

Учитывая тематику Конференции, представляемый материал посвящен новым результатам решения задачи – обоснованию построения интеллектуальной системы управления защитой КА.

2. МЕТОДЫ И АЛГОРИТМЫ ПОСТРОЕНИЯ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ УПРАВЛЕНИЯ ЗАЩИТОЙ КА ОТ УДАРОВ АЧ

Прогнозирование результатов воздействия АЧ реализовано в виде иерархической структуры дерева нечеткого логического вывода. Логико-лингвистическая аппроксимация позволяет наглядно представлять входные параметры и их взаимосвязи на основе логически непротиворечивых и интерпретируемых высказываний о влиянии факторов на результат прогнозирования. Гибридная технология, дополненная формализмом нейронных сетей, позволяет не только построить модель, но и обеспечивает ее настройку методами машинного обучения [4].

Алгоритм прогнозирования и идентификации результатов возможных ударов АЧ реализован в виде нейро-нечеткой сети (ННС) - сеть прямого распространения сигнала, которая построена на основе многослойной архитектуры с использованием И, ИЛИ нейронов. Структура ННС представлена на рис.1.

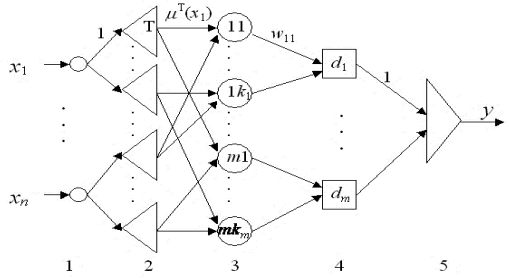


Рис.1 Структура нейро-нечеткой сети

Число узлов в ННС:

- слой 1 – по количеству входных параметров прогнозирования;
- слой 2 – по количеству нечетких термов параметров в базе знаний;
- слой 3 – по количеству правил (строк-конъюнкций) в базе знаний;

- слой 4 – по количеству критериев повреждения, выбранных для прогнозирования;
- слой 5 – уровень, формирующий ошибку прогноза.

В качестве показателя прогноза удара АЧ выбран коэффициент истинности критерия повреждения d_j (K_d), вычисляемый в слое 4:

$$K_d = \max_{j=1,m}\left\{\mu^{d_j}(x1, x2, \ldots xn)\right\} = \max_{j=1,m}\left\{\max_{p=1,k_j}\left\{w^{jp} \min_{i=1,n}\left(\mu_i^{a^{jp}}(xi)\right)\right\}\right\},$$

$$i=\overline{1,n},\ j=\overline{1,m},$$

где x_i, w_{jp}, μ^{d_j}, $\mu_i^{a^{jp}}$ – значения входных параметров, весов правил и параметров функций принадлежности для входного значения параметра и критерия повреждения

Показатель рассчитывается на основе композиционного правила Заде: при прогнозе результата воздействия используется идея определения лингвистического терма критерия повреждения по максимуму функции принадлежности, и обобщение этой идеи на всю базу знаний.

Для функции принадлежности параметров x выбрана двухпараметрическая форма, удобная для дальнейшей настройки:

$$\mu^j(x) = \frac{1}{1+\left(\frac{x-b_j}{c_j}\right)^2},$$

где b_j и c_j – параметры функции принадлежности: координата максимума и коэффициент концентрации.

Обучение ННС проводится методом обратного распространения ошибки с использованием двойственной сети [6], которая представлена на рис.2.

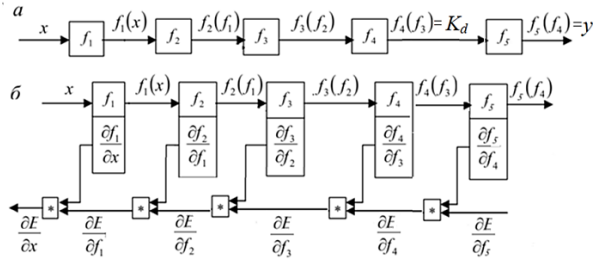


Рис.2. Структура и режимы функционирования двойственной сети: а – прямое функционирование сети, б – прямое нагруженное и обратное функционирование сети

Для настройки использована функциональная зависимость на выходе ННС, которая участвует в формировании значения ошибки $E_t = \frac{1}{2}\left(\hat{y}_t - y_t\right)^2$, где: $\hat{y}_t$ и y_t – теоретический и экспериментальный выходы сети на *t*-м шаге обучения.

Отличительной особенностью алгоритма является то, что метод обратного распространения ошибки при обучении ННС последовательно применяется в два этапа:

- $\sum_{p=1}^{M}\left[\mu^y\left(X_p, W, B, C\right) - \mu^y{}_p\right]^2 = \min_{W,B,C}$ – обучение синоптической карты ННС;
- $\sum_{p=1}^{M}\left[\mu^y\left(X, W_p, B_p, C_p\right) - \mu^y{}_p\right]^2 = \min_{X}$ – обучение параметров на входе ННС.

Применение алгоритма обучения позволило автоматически проводить настройку БЗ на изменения в предметной области защиты от АЧ и повысить качество прогнозирования результатов ударов АЧ (таблица 1).

Таблица 1. Качество прогнозирования результатов воздействия антропогенных частиц, полученные в среде разработки БЗ

Критерии повреждения	Вероятность безошибочного прогнозирования до обучения БЗ	Вероятность безошибочного прогнозирования после обучения БЗ
Аберрация, царапина, вмятина (d1)	0,60	0,90
Глухой канал (кратер) (d2)	0,71	0,92
Сквозной канал, (пробой) (d3)	0,58	0,83
Разрушение элемента, (d4)	0,51	0,88
Осредненная оценка	0,60	0,88

Оценка и выбор защитного барьера является многокритериальной задачей, структура которой содержит набор созданных на КА защитных барьеров $S=\{s_v\}$, множество критериев их оценки $C=\{c_l\}$ и коэффициенты важности для каждого критерия $F=\{f_l\}$. Оценка и выбор S^* проводится при обнаружении каждого опасного удара АЧ в автоматическом режиме.

Алгоритм оценки и выбор барьера безопасности реализован в виде семантической нечеткой сети (СНС).

Семантическая нечеткая сеть является носителем утверждения: «Эффективный барьер должен быть результативным, оперативным и экономичным». Оценка проводится на основе логического вывода с использованием критериев (значений параметров барьера) и коэффициентов важности, представленных в виде нечетких множеств.

Структура семантической нечеткой сети (СНС) оценки и выбора барьера безопасности представлена на рис.3:

- d_j – прогнозируемые критерии повреждения КА от воздействия АЧ;
- s_v – барьеры, реализованные на КА для каждого dj;
- c_1-c_3 – критерии (параметры) оценки эффективности защитного барьера, представленные в виде лингвистических переменных: «Результативность», «Оперативность», «Экономичность»;
- f_1-f_3 – коэффициенты важности параметров оценки барьера, определяемая текущими условиями, представленные в виде нечетких множеств (НМ);
- S^* – выбранный защитный барьер.

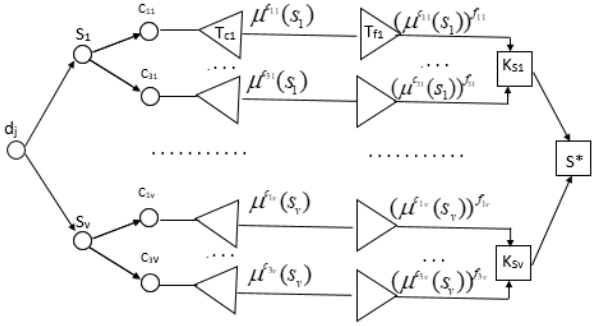


Рис.3. Структура семантической нечеткой сети (СНС) оценки и выбора защитного барьера

Базируясь на принципе Беллмана-Заде, лучший барьер имеет наибольшую гарантированную оценку по выбранным показателям эффективности: *c1*-оперативность, *c2*-экономичность, *c3*-результативность (надежность) барьера. Оценка эффективности определяется в виде их пересечения. Учитывая то, что в теории нечетких множеств операция пересечения соответствует min, расчетное соотношение для коэффициента защищенности при заблаговременной оценке, т.е. равновесных критериях имеет вид:

$$K_s = \left\{ \frac{\min\limits_{l=1..3} \mu^l(s_1)}{s_1}, \frac{\min\limits_{l=1..3} \mu^l(s_2)}{s_2}, ..., \frac{\min\limits_{l=1..3} \mu^l(s_v)}{s_v} \right\}.$$

Для учета условий обстановки, определяемых временными, вещественными и вероятностными характеристиками возможного удара вводятся коэффициенты важности параметров оценки. Расчет коэффициентов $F=\{f_l\}$ проводится в системе поддержки принятия решений (СППР) наземного командного управления (НКУ) с участием человека-оператора и без участия оператора – непосредственно на борту КА, в случае автономного участка функционирования. Для этого разработана база знаний важности параметров, с ее помощью, уточняются значения критериев и эффективность барьера [8].

Полученные значения f_l являются показателями степени для значений оценок параметров барьера и выступают коэффициентами концентрации нечетких множеств. С учетом важности параметров уточняется эффективность защиты КА и коэффициент защищенности примет вид:

$$K_s = \left\{ \frac{\min\limits_{l=1..3} \mu^l(s_1)^{f_l}}{s_1}, \frac{\min\limits_{l=1..3} \mu^l(s_2)^{f_l}}{s_2}, \frac{\min\limits_{l=1..3} \mu^l(s_v)^{f_l}}{s_v} \right\}.$$

Выбирается защитный барьер с наибольшим значением K_s.

3. ПРИНЦИПЫ ПОСТРОЕНИЯ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ УПРАВЛЕНИЯ ЗАЩИТОЙ КА ОТ УДАРОВ АЧ

Теоретические основы построения и функционирования интеллектуальной системы управления защитой КА учитывают особенности построения базы знаний интеллектуальной системы [2] и представляются следующими положениями:

1. Принцип наличия встроенного механизма прогнозирования:

- условий обстановки, связанных с возможными ударами АЧ;
- результатов таких ударов и их идентификация с заданными критериями повреждения(поражения) КА;
- оценки эффективности имеющихся защитных барьеров для различных условий обстановки.

Указанный принцип позволяет избежать ситуации, связанной с возможными ударами, для которых не предусмотрены барьеры в силу временных и иных ограничений на работу механизмов, формирующих управляющие воздействия и (или) подключающих защитный барьер. Особенно важен указанный принцип для участков автономного функционирования КА на орбите.

2. Возможность прогнозирования условий и результатов ударов АЧ позволяет сформулировать еще один принцип, внедрение которого

позволит предотвратить серьезные повреждения – принцип наихудшего результата. Принцип предполагает моделирование прогнозируемых ситуаций, с проверкой их на высокоточных моделях, причем в соответствии с этим принципом необходимо ориентироваться на самые неблагоприятные условия.

3. Принцип вариантного подхода к решению задач управления защитой КА, основанный на положениях теории ситуационного управления

Отличительной особенностью интеллектуальной системы управления от построенной по «традиционной» схеме является наличие механизмов хранения и обработки знаний по управления защитой в неопределенных условиях возможных ударов АЧ.

При обнаружении угрозы удара накопленная совокупность знаний (БЗ) должна обеспечить идентификацию возможного результата удара, по его принадлежности к выделенным классам критериев повреждения КА и провести оценку эффективности защитных барьеров с учетом складывающейся обстановки.

4. Принцип многоуровневого иерархического построения систем управления

Иерархия уровней управления защитой КА определяются исходя из совокупности (дерева) целей, которые должны быть достигнуты. Эти цели определяют:

- планирование целесообразного поведения КА с учетом возможного удара и дальнейшего выполнения задач по предназначению
- планирование целесообразных действия для эффективной защиты от возможного удара в текущих условиях
- планирование реализации выбранных действий (подключение защитных барьеров) компенсирующих или уменьшающих угрозу поражения от удара частиц.

5. Принцип непрерывности функционирования (возможно, с некоторой потерей качества или эффективности) при разрыве связей и невозможности получения управляющих воздействий от высших уровней иерархии системы.

Принцип устанавливает частичное снижение уровня интеллектуальности, но не прекращение функционирования системы в целом. Управление защитой КА осуществляется на основе оперативного прогнозирования угрозы с использованием актуальных (обученных) баз знаний, их настройки, работы с базами событий, знаний и адаптивных мер по защите, выбранных с использованием интеллектуальных механизмов, реализованных как на КА, так и с помощью НКУ. При таком подходе неточность знаний о ситуации

можно скомпенсировать за счет наличия интеллектуальности системы и соответствующих алгоритмов управления. Наличие механизма оценки и базы знаний по прогнозированию повреждений и оценки эффективности защитных барьеров на КА обеспечивает режим продолжительного функционирования и максимальную живучесть КА в целом.

6. Принцип адекватности уровня интеллектуальности системы управления факторам неопределенности, воздействующим на неё

В соответствие с этим принципом построенная система должна соответствовать правилу IPDI (Increasing Precision with Decreasing Intelligence) теории интеллектуальных машин и иметь степень интеллектуальности, как минимум на двух слоях [5].

Структурная схема системы управления представлена на рис.4.

Модуль обнаружения и распознавания, который входит в состав системы, решает задачу пространственной селекции, обнаружения, распознавания АЧ и оценку их параметров, включая: массо-габаритные характеристики, параметры движения и расстояние от КА, предполагаемое время до столкновения [3].

Рис.4. Иерархическая структура интеллектуальной системы управления защитой КА от ударов АЧ

На первом слое решаются задачи прогноза событий, построения и поддержания в актуальном состоянии базы знаний, автоматизированной выработки предложений по повышению живучести и эксплуатационной надежности КА. На этом уровне управление КА осуществляется по командной радиолинии при наличии

располагаемого времени на анализ, подготовку и выдачу команд и программ управления. Для этого с использованием модуля оценки результатов экспериментальных исследований и моделирования взаимодействия проводится расчетно-экспериментальный и компьютерный анализ напряженно-деформированных состояний материалов и конструкций КА с учетом механических и термических нагрузок, анализ критериев и условий появления повреждений КА, а также оценка долговечности материалов для стадий образования и развития кратеров и трещин

Указанный анализ с использованием разработанных расчетно-экспериментальных и компьютерных моделей позволяет детализировать прогноз событий и выработать обоснованные предложения по применению защитных мер [1].

Кроме этого, на верхнем слое реализована среда разработки и обучения базы знаний разработанной системы. Здесь проводится разработка логико-лингвистических моделей представления знаний и использование технологии нейронных сетей для их настройки. Это позволяет обеспечить надежную идентификацию предъявляемых ситуаций. При этом важнейшей особенностью нейросетевых структур является высокое быстродействие, достигаемое за счет параллельности обработки информации при их аппаратной реализации [4].

Отличительной чертой обучения БЗ является наличие второго этапа. На нем происходит решение обратной задачи распознавания путем подстройки значений входных параметров через градиент функции ошибки по входу сети для каждого из критериев повреждения. Для этого, генерируемые наборы значений входных параметров подаются на вход обученной сети. Имея заданный критерий повреждения и результат, выданный сетью, вычисляется градиент функции ошибки по входным параметрам сети и по технологии метода нагруженных двойственных сетей. В соответствии со значениями элементов градиента изменяются значения входных параметров в направлении снижения ошибки, что позволяет итерационно получать наборы входных векторов, порождающих требуемый результат. При этом синаптическая карта сети, настроенная на предыдущем этапе, остается неизменной.

Второй слой предназначен для прогнозирования событий, а также уточнения эффективности и подключение барьера безопасности в автоматическом режиме. Для этого разработан модуль интеллектуального прогнозирования результатов воздействия оценки и выбора защитного барьера. Ключевым элементом слоя является БЗ, которая поддерживается в актуальном состоянии путем учета всех

изменений в предметной области, привлекаемых при ее обучении. С использованием заложенных в базе знаний реализуется механизм нечеткого логического вывода. Максимум функции принадлежности определяет наиболее вероятный прогнозируемый критерий повреждения.

На втором уровне для выбранного критерия уточняются оценки эффективности всех барьеров, которые связаны с выделенным повреждением. Уточнение проводится автоматически с учетом располагаемых и требуемых ресурсов и надежности защиты.

Третий слой системы включает множество подготовленных к подключению защитных барьеров. Защитные барьеры, реализованные на КА в виде функций, изделий, материалов, программного обеспечения и т. д., создаются с целью предотвращения, остановки или замедления развития опасной ситуации. Все они выполнятся по жесткой программе и необходимо наличие команды на их подключение [1].

Отличительной особенностью представленной системы управления защитой является обеспечение автономного функционирования КА, при котором в цикле работы бортовой ЭВМ автоматически решаются задачи обнаружения, распознавания, оценки опасности и прогнозирование результатов удара АЧ, а также оценки эффективности защитных барьеров и реализация выбранного барьера для защиты КА, адаптивного к складывающейся обстановке.

4. ЗАКЛЮЧЕНИЕ

Разработана методология построения интеллектуальной системы управления защитой КА от ударов АЧ. Представлена структура и механизмы, реализующие интеллектуальную оценку, прогнозирование результатов гиперскоростного удара и выбора эффективного защитного барьера при воздействии АЧ.

Сформулированы принципы построения интеллектуальной системы, которые позволяют реализовать быстрый и медленный контуры управления защитой КА от ударов АЧ.

Интеллектуальная составляющая реализована в обоих контурах управления, при этом быстрый контур обеспечивает оперативную (в реальном масштабе времени) оценку уровня угрозы и позволяет сформировать эффективный механизм защиты КА без участия оператора и НКУ.

При наличии связи с КА и выполнении баланса времени на подготовку решения, медленный контур реализует человеко-машинные (включая СППР) процедуры управления защитой КА с использованием наземных высокоточных экспериментально-расчетных и

компьютерных моделей, на основе углубленного анализа, прогнозирования и обоснования мер по повышению живучести и эксплуатационной надежности КА.

Литература

1. Meshkov V.V., Fedosov Yu.A, Filatova N.N. An integrated system for modeling hypervelocity interaction of solids. Acta Astronautica. – 2021. – V. 178. – P. 606-615.
2. Paliukh B.V., Kemaykin V.K., Kozlova Y.G. Development principles of knowledge database of intelligent system for estimation of dynamical interaction of orbital systems with space debris // IT in industry. – 2017 Vol. 2. – No. 6. – P. 1-6.
3. Paliukh B.V., Zykov I. I. Метод распознавания антропогенных частиц в сложных фоновых условиях// Proceedings of the Third International Scientifig Conference “Intelligent Information Technologies for Industry” (IITI’19) Ostava, Чехия, Dezember 1-6, 2019. P. Advances in Intelligent Systems and Computing
4. Ярушкина Н. Г. Основы теории нечетких и гибридных систем / Н.Г. Ярушкина. – М. : Финансы и статистика. – 2004.
5. Saridis G.N. Analytical formulation of the principle of increasing precision with decreasing intelligence for intelligent machines // Automatics, Vol. 3, No. 25, 1989.
6. Горбань А.Н. Обучение нейронных сетей. М.: СП «ПараГраф». – 1990.
7. Интеллектуальные системы автоматического управления / Под ред. И.М. Макарова, В.М. Лохина. – М.: Физматлит, 2001.
8. Кемайкин В.К., Кожухин И.В. Формирование барьера безопасности на космическом аппарате при угрозе воздействия космического мусора методами нечеткой логики // Программные продукты и системы. – 2019. – № 32.

УДК 629.786, 531.58

ИНТЕЛЛЕКТУАЛЬНАЯ ОЦЕНКА УДАРНОГО ВОЗДЕЙСТВИЯ ГИПЕРСКОРОСТНЫХ АНТРОПОГЕННЫХ ЧАСТИЦ НА КОСМИЧЕСКИЙ АППАРАТ

***Мешков В.В.**, д.т.н.*
e-mail: msf-tgtu@yandex.ru
ТвГТУ
***Лопин Г.А.**, д.т.н.*
e-mail: georgi-l@mail.ru
НИЦ (г.Тверь) ЦНИИ ВКС Минобороны России
***Филатова Н.Н.**, д.т.н.*
e-mail: nfilatova99@mail.ru
ТвГТУ

1. ВВЕДЕНИЕ

Одним из актуальных направлений современного этапа освоения космического пространства является обеспечение эффективности и долговечности орбитальных космических аппаратов (КА) в условиях постоянного накопления в околоземном пространстве частиц космического «мусора». «Мусор» состоит из *антропогенных* (*техногенных) частиц*, возникших в результате активной космической деятельности человечества: операционных элементов ракет и КА и их обломков.

По прогнозам аналитиков, пассивное продолжение запусков ракет может в течение ближайших 50 лет привести к полной экранировке Земли космическим мусором на высотах до 2000 км и значительно ухудшить возможности эффективного и долговременного использования КА.

В ведущих странах мира, осуществляющих активную космическую деятельность в России (организациях Роскосмос, учреждениях Минобороны, Академии науки, Минобрнауки), а также в США, Китае предпринимаются активные попытки снижения негативного влияния частиц, а также повышения живучести КА.

Под живучестью КА понимают его свойство как системы сохранять способность функционирования с требуемой

эффективностью вовремя или после воздействия неблагоприятных внешних факторов космического пространства. Живучесть является сложным свойством, включающим *стойкость* к воздействию неблагоприятных факторов, *защищённость* КА, а также его *неуязвимость*, характеризующую способность КА избегать воздействия неблагоприятных факторов на уязвимые элементы за счёт маневрирования и активного противодействия неблагоприятным факторам [1].

Поэтому проблема повышения живучести КА длительной эксплуатации широкого класса и назначения в условиях воздействия антропогенных частиц за счёт оптимальной интеграции в конструкцию КА интеллектуальной системы управления маневрированием и активного противодействия ударам частиц является актуальной, но в тоже время чрезвычайно сложной.

2. ПРОБЛЕМА СОЗДАНИЯ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ УПРАВЛЕНИЯ МАНЕВРИРОВАНИЕМ И АКТИВНОГО ПРОТИВОДЕЙСТВИЯ УДАРАМ ЧАСТИЦ

Данная работа является продолжением и развитием исследований, выполненных в 2017-2019 годах Тверским государственным техническим университетом и Научно-исследовательским центром (г. Тверь) ЦНИИ ВКС Минобороны России по заданию Минобрнауки РФ на тему «Теория и адаптивные алгоритмы обнаружения антропогенных частиц и объектов, и оценка динамического взаимодействия с космическими аппаратами на основе интеллектуального анализа данных» [2].

Проведенный комплекс работ по теоретическому анализу данных о космической обстановке, анализу известных характеристик антропогенных частиц, результатам имитационного моделирования и экспериментального исследования, позволили сделать ряд принципиальных выводов.

Во-первых, наибольшую угрозу для КА могут представлять металлические антропогенные частицы размером 0,5 – 5 см, движущиеся на орбите с гиперзвуковой космической скоростью на встречных с КА курсах, или достаточно быстро приближающиеся к нему вдогон, или по касательной от витка к витку на близких орбитах. Размеры данных частиц, вероятность столкновения которых с КА велика, не позволяют достоверно их контролировать наземными радиолокационными на больших дальностях из-за малого уровня ЭПР частиц (менее 0,0001 м^2), а также наземными оптическими средствами.

Во-вторых, исходя из анализа характеристик опасных частиц, наиболее предпочтительным методом их обнаружения, селекции и

распознавания на фоне других частиц и помеховых излучений космоса, является использование бортовых оптико-электронных средств (ОЭС) видимого диапазона длин волн, качество обнаружения которых не зависит от искажений сигналов, вызванных неоднородностью среды передачи. Однако, для их реализации необходимы новые методы и алгоритмы, обеспечивающие решение информационных задач с требуемым качеством для большого числа (более 1000) одновременно наблюдаемых малоразмерных целей при ограничениях по времени и количеству ОЭС на борту КА. Теоретические основы разработки таких методов рассмотрены в работе [3].

В-третьих, входной информацией для предполагаемой бортовой интеллектуальной системы управления КА являются не только траекторные характеристики частиц, но и выделение из их множества наиболее опасных для живучести КА частиц, способных повредить покрытия или конструкцию, вывести из строя жизненно важные агрегаты. Для этого, необходимы данные об уязвимости конструкций, агрегатов и покрытий КА различных типов как объектов поражения от ударов гиперзвуковых частиц в условиях вакуума.

Получение таких данных в натурных условиях космоса практически не реализуемо. Возможности физического моделирования характеристик уязвимости материалов и конструкций КА в них ограничены по массе и скорости ускоряемых элементов. Предлагаемый в работе подход основан на применении комбинированных расчётно-экспериментальных методов с использованием математических моделей высокой степени адекватности на суперЭВМ, откалиброванных по результатам экспериментов на более низких скоростях.

Для достижения поставленной цели необходимо решение проблемы разработки теоретических и практических основ, обеспечивающих обоснование построения интеллектуальной системы для повышения живучести космических аппаратов длительного функционирования в условиях воздействия антропогенных частиц.

Решение поставленной проблемы и создание комплекса новых и усовершенствованных теоретических и экспериментальных методов, принципов построения рассматриваемых систем является развитием достижений в данной предметной области известных научных школ России, таких как ЦНИИМаш, Военно-космическая академия, Тульский государственный университет и других.

Декомпозиция проблемы создания интеллектуальных бортовых комплексов для обеспечения живучести КА в условиях воздействия космического мусора предполагает решение следующих основных задач:

1.Разработка теоретических и прикладных основ построения интеллектуальной системы управления защитой КА от ударов антропогенных частиц с обоснованием возможных вариантов структуры и состава системы, оценкой её эффективности на имитационных моделях.

2. Разработка методов и алгоритмов обнаружения и селекции бортовыми ОЭС из большого множества (более 1000) наиболее опасных гиперзвуковых антропогенных частиц малого размера (0,5-5 см), не обнаруживаемых наземными национальными средствами контроля космического пространства в сложных фоновых условиях внешней подсветки частиц Солнцем и Луной, механических шумов приборов.

3.Разработка экспериментально – теоретических методов, обеспечивающих верификацию компьютерных моделей для изучения антропогенных частиц и уязвимости материалов и конструкций КА как объектов гиперзвукового поражения.

Цель данной статьи – интеллектуальная оценка ударного воздействия гиперзвуковых антропогенных частиц на типовой космический аппарат в интересах обоснования построения его бортовой интеллектуальной системы управления защитой от антропогенных частиц.

3. ОСНОВЫ СОЗДАНИЯ СИСТЕМЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА РЕЗУЛЬТАТОВ УДАРА ГИПЕРСКОРОСТНЫХ АНТРОПОГЕННЫХ ЧАСТИЦ В ЭЛЕМЕНТЫ КА

Важным аспектом проблемы управления защитой КА при ударе гиперскоростных антропогенных частиц является разработка основ построения информационной системы оценки и прогнозирования результатов удара частиц по элементам КА.

Основными задачами при создании системы интеллектуального анализа результатов удара гиперскоростных антропогенных частиц в КА являются:

- исследование удара гиперскоростных частиц в элементы КА методами физического и математического моделирования;
- экспериментальное исследование удара гиперскоростных частиц в элементы КА в широком диапазоне параметров;
- интегральный анализ результатов моделирования и экспериментальных данных с использованием цифровой обработки изображений (оптических, рентгеновских и др.) удара гиперскоростных частиц в элементы КА.

Для исследования гиперскоростного взаимодействия частиц с КА используют экспериментальные методы и математическое моделирование [3-10].

Основные результаты в области экспериментального исследования получены путем натурного эксперимента с помощью ствольного метания твердых тел по мишеням. Недостатком экспериментальных методов является их высокая стоимость, сложность регистрации параметров и трудности или невозможность ускорения тел при скоростях порядка 7-14 км/с.

Для математического моделирования используют: а) модели течения и уравнения баллистического предела для определения параметров повреждения и пробития стенок КА и б) моделирование в современных программных средах, в том числе методом SPH [11]. Основной проблемой для математических моделей и их программных реализаций является точность задания параметров, которые необходимо корректировать на основе анализа экспериментов и верификации моделей.

Основное назначение предлагаемой интегрированной информационной интеллектуальной системы моделирования (ИИИС) заключается в объединении и анализе результатов экспериментов и разных видов моделирования, а также генерации структур знаний, используемых при формировании формализованных гипотез о взаимосвязи входных и выходных переменных. В состав системы ИИИС входят три программных комплекса (ПК) [9, 10, 13, 14], рис.1.: ПК-1, ориентированный на задачи расчетно-экспериментального моделирования и натурных экспериментов; ПК-2, решающий задачи математического моделирования и постановки вычислительных экспериментов с учетом результатов эксперимента, и ПК-3, осуществляющий обработку и анализ результатов моделирования.

Расчетно-экспериментальное моделирование гиперскоростного удара твердых тел

Комплекс ПК-1 предназначен для расчетно-экспериментального моделирования гиперскоростного удара частиц и для информационной поддержки экспериментов, реализуемых на специальном экспериментальном комплексе. Экспериментальное изучение удара гиперскоростных частиц производилось на аттестованном стрельбовом моделирующем комплексе СМК-1, обеспечивающем ускорение частиц диаметром до 2 см до скорости 5 км/с и оснащен высокоскоростной рентгеновской камерой, рис.2.

Комплекс ПК-1 предназначен для расчетно-экспериментального моделирования гиперскоростного удара частиц и для информационной поддержки экспериментов, реализуемых на специальном экспериментальном комплексе.

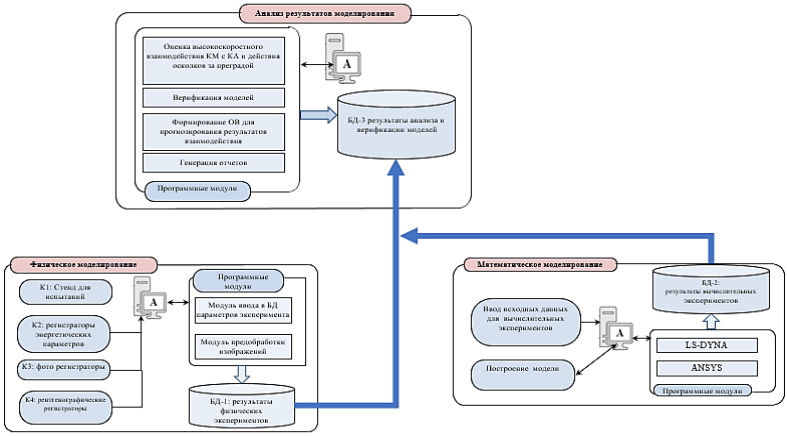


Рис.1. Принципиальная структура интегрированной информационной интеллектуальной системы оценки удара гиперзвуковых частиц с КА

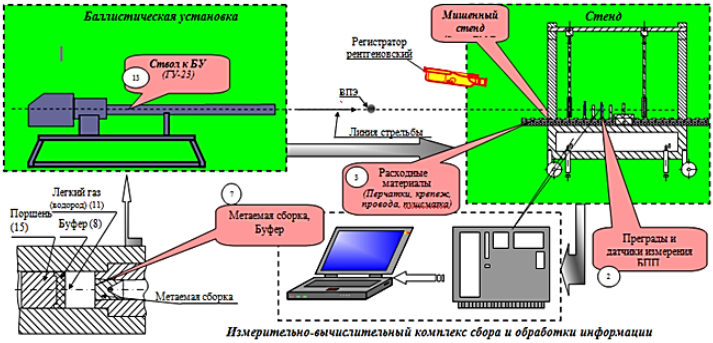


Рис.2. Схема стрельбового моделирующего комплекса СМК-1

Экспериментальное изучение гиперскоростного удара частиц производилось на аттестованном стрельбовом моделирующем комплексе СМК-1, НИЦ (г. Тверь) ЦНИИ ВКС Минобороны России, структурная схема которого представлена на рис.2. СМК-1 обеспечивает ускорение частиц диаметром до 2 см до скорости 5 км/с и оснащен высокоскоростной рентгеновской камерой.

Разработанная блок-схема объекта моделирования позволяет установить количественные соотношения, характеризующие взаимосвязь между входными переменными исследуемого процесса и основными характеристиками результатов воздействия. Результаты

расчетно-экспериментального моделирования могут быть представлены в виде набора параметров, характеризующих пробитие/повреждение пакета преград на СМК-1, рис.3.

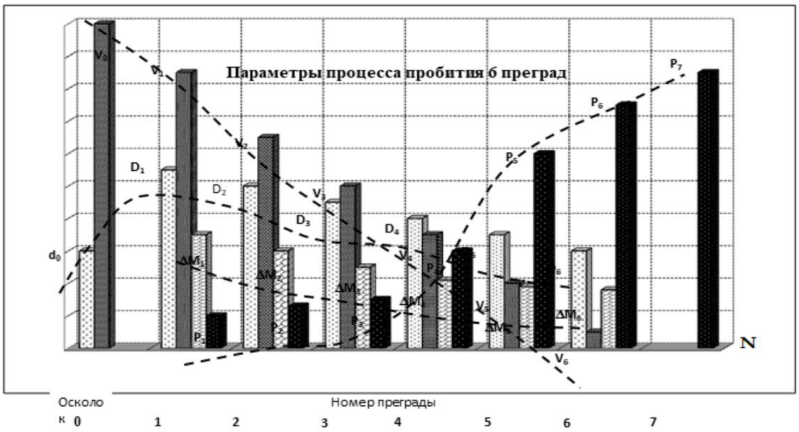


Рис.3. Набор параметров пробития пакета преград на СМК-1

Исследование явлений, возникающих при гиперскоростном ударе, только с помощью физических экспериментов зачастую не дает ожидаемого результата, несмотря на большие материальные и технические затраты. Решение этой задачи возможно путем использования компьютерного моделирования.

Компьютерное моделирование удара гиперскоростных частиц по КА

Для реализации этой задачи в состав ИИИС включен комплекс ПК-2, который предназначен для численной реализации компьютерных моделей процесса удара гиперскоростной частицы по элементу КА.

Применение методов конечных элементов и SPH в средах ANSYS/LS-DYNA/AUTODYN и FLOWVISION позволяет исследовать кинетику изменения напряженно-деформированного состояния и температурного поля техногенной частицы в гибридный элемент реальной оболочки КА, рис.4. По результатам численного моделирования для любого момента времени можно оценить:

- степень и параметры разрушения, повреждения и трещин в элементе, а также влияние мелких осколков (менее 0,5 мм) на повреждение элемента КА и внутреннего оборудования;
- кинетику изменения напряженно-деформированного состояния и температурного поля элемента КА и частицы;
- кинетическую энергию элемента КА и частицы, а также временные 3D-зависимости средних скоростей осколков в облаке.

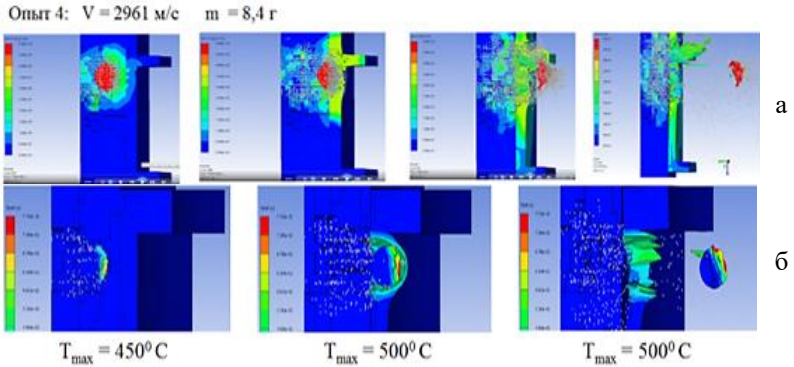


Рис.4. Кинетика изменения напряженно-деформированного состояния (а) и температурного поля (б) при ударе гиперскоростной частицы в гибридный элемент реальной оболочки КА

Изучение особенностей сложных физических процессов, протекающих при ударе гиперскоростной частицы в элемент КА, позволяет обосновать состав и конструкцию реальных элементов КА и в определенных пределах прогнозировать повреждение КА в зависимости от изменения материала, размеров и скорости антропогенных частиц и сформировать банк данных, который может уточняться и дополняться новыми данными.

Компьютерное моделирование позволяет также расширить ограниченный набор результатов расчетно-экспериментального моделирования и натурных экспериментов. В свою очередь, это позволяет разработать основы использования банка данных при разработке защиты и систем управления защитой перспективных КА.

Обработка и анализ результатов расчетно-экспериментального и компьютерного моделирования

Задачи обработки результатов экспериментов, полученных с применением разных видов моделирования, решаются в комплексе ПК-3, в состав которого входят: а) база данных с векторными моделями, построенными по результатам расчетно-экспериментального и компьютерного моделирования, б) расчетные модули, позволяющие сделать анализ результатов натурного и вычислительного экспериментов, в том числе высокоскоростной рентгенографии и верификации моделей, с помощью алгоритмов нечеткой логики [10, 11].

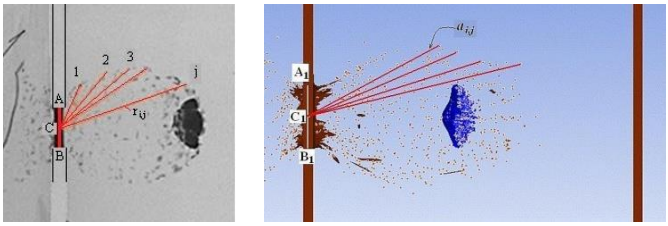
Использование двух видов моделирования (расчетно-экспериментального и компьютерного) при исследовании одного и того же объекта позволяет решать задачи:

- верификации математических моделей путем сравнения векторных моделей, построенных по результатам натурного и вычислительного экспериментов при условии, что эксперименты выполнены при одинаковых условиях;
- формирования обучающих выборок для подсистемы прогнозирования результатов удара частиц в элементы КА;
- формирования обобщающих высказываний с оценками результатов моделирования, которые будут использоваться при построении базы знаний в интегральной информационной интеллектуальной системы для оценки гиперзвукового удара частиц в элемент КА.

Анализ изображений результатов компьютерного моделирования и экспериментов методом нечетких множеств

Достоверность результатов, полученных с помощью конечно-элементной модели, реализуемых в среде ANSYS, зависит от результатов ее верификации. Решение этой задачи осуществляется путем сравнения изображений на оптических и импульсных рентгеновских снимках, полученных в натурном эксперименте, с результатами компьютерного моделирования. Для сравнения параметров изображений используются методы нечетких множеств [12-14], рис.5.

Окончательный результат оценки различий в оптических и рентгеновских изображениях отверстий, образованных при ударе в процессе натурного эксперимента и компьютерного моделирования, формируется путем логического вывода с помощью алгоритма Мамдани [15]. Анализ результатов сравнения параметров экспериментальных (оптических и рентгеновских) и компьютерных изображений позволяет сделать вывод о возможности использования компьютерной модели в программе ANSYS для описания повреждения элемента КА при ударе гиперскоростной частицей, а также для последующей верификации модели в комплексе ПК-3.



а б

Рис.5. Изображение облака осколков за первой пробитой преградой, полученное на рентгенограмме (а) и при моделировании в среде ANSYS (б)

4. ЗАКЛЮЧЕНИЕ

Разработана структура и описаны функции интегрированной интеллектуальной информационной системы, реализующей оценку и прогнозирование результатов ударного воздействия гиперскоростных антропогенных частиц по элементам КА.

Предложена новая методология оценки параметров ударного воздействия гиперскоростных частиц по элементам КА, объединяющая результаты взаимодополняющих методов расчетно-экспериментального и компьютерного моделирования. Использован аппарат нечеткой логики, а также оригинальная процедура перехода к индексной шкале при формировании общей оценки различий для множества точек соответствующих оптических и рентгеновских изображений. Применение аппарата нечеткой логики позволяет формировать обобщающие высказывания с оценками результатов моделирования, могут быть использованы при построении базы знаний в интегрированной информационной интеллектуальной системе оценки и прогнозирования последствий динамического удара частиц в КА в диапазоне скоростей и масс частиц, недоступных в экспериментах.

Предложенная интеллектуальная оценка ударного воздействия гиперскоростных антропогенных частиц на КА может быть использована при разработке теоретических и прикладных основ построения интеллектуальной системы управления защитой КА от ударов частиц, а также для оптимизации состава и конструкции элементов перспективных КА.

Литература

1. Абдурахимов А.А., Скворцов Д.В. Методологические основы обеспечения живучести космических аппаратов//Труды Военно-космической академии имени А.Ф. Можайского. / Под ред. М.М. Пенькова. – СПб.: ВКА имени А.Ф. Можайского. – 2013. -Выпуск 640. – С.7-20.
2. Теория и адаптивные алгоритмы обнаружения антропогенных частиц и объектов, и оценка их динамического взаимодействия с космическими аппаратами на основе интеллектуального анализа данных // Отчёт о НИР 2.1777.2017/4.6. – Тверь: ТвГТУ. – 2019.
3. Катулев А.Н., Храмичев А.А., Ягольников С.В. Цифровая обработка 2D слабоконтрастных изображений, формируемых оптико-электронным прибором в сложных фоновых условиях. Обнаружение, распознавание, сопровождение динамических объектов. – М.: Радиотехника, 2018.
4. Герасимов А.В., Пашков С.В., Христенко Ю.Ф. Защита космических аппаратов от техногенных и естественных осколков. Эксперимент и численное моделирование //Вестник Томского гос. техн. ун-та. Математика и механика. – 2011. – № 4(6). – С. 70-78.

5. Gong Z. Hypervelocity Impacts and Protection Research in CAST. – Beijing Orbital Debris Mitigation Workshop. – 2010.
6. Steinhauser M.O. Novel Computer Simulations Addressing the Impact Risks in Space from Orbiting Debris//Material Science & Engineering Journal.- 2017. – Vol. 1. – Is. 1. – Article 1005. – P. 1-4.
7. Schonberg W.P. Studies of hypervelocity impact phenomena as applied to the protection of spacecraft operating in the MMOD environment//Procedia Engineering. – 2017. – Vol. 204. – P 4-42.
8. Bergh M., Garcia V. A computational model for assessing high-velocity debris impact in space applications//Shock Waves. – 2017. – Vol.27 (4). – P. 675-684. https://doi. org/10.1007/s00193-017-0709-9.
9. Silnikov M.V., Guk I.V., Nechunaev A.F., Smirnov N.N. Numerical simulation of hypervelocity impact problem for spacecraft shielding elements//ActaAstronautica. – 2018. – Vol. 150. – P. 56–62. – https://doi.org/10.1016/j. actaastro. 2017.08.030.
10. Watson E, et al. Simulating hypervelocity impact phenomena with discrete elements // Procedia Engineering. – 2017. – Vol.204. – P. 75-82.
11. Зеленцов В.В., Маханьков А.В. Численное моделирование методом SPH взаимодействия фрагментов космического мусора с элементами конструкций космических аппаратов//Наука и Образование. V.: МГТУ им. Н.Э. Баумана. – Электрон. журн. – 2017. – № 05. – С. 42-56. DOI: 10.7463/0517.0001164.
12. Paliukh B.V., Kemaykin V.K., KozlovaYu.G., Kozhukhin I. V. Forecasting of results of dynamic interaction between space debris and spacecrafts on the basis of soft computing methods//Proceedings of the Third International scientific conference "Intelligent information technologies for industry" (IITI'18). – Russia, Sochi. – 2019. September 17-21. – P. 293-302.
13. Paliukh B.V., Meshkov V.V., Kemaykin V. K., Kozlova Yu. G. Fundamentals of intelligent system for estimation of dynamical interaction of space debris with spacecrafts//Advances in intelligent systems and computing. – 2018. – Vol. 658. – P.248-257.
14. Meshkov V.V., Fedosov Yu.A., Filatova N.N. An integrated system for modeling hypervelocity interaction of solids//ActaAstronautica. – 2021. Vol. – 178. – P. 606-615.
15. Zimmermann H-J. Fuzzy set theory and its applications. – Kluwer. Dordrecht. – 1991. – 315 P.

УДК 004:007

МОДЕЛИ ПРОТИВОДЕЙСТВИЯ УГРОЗАМ НАРУШЕНИЯ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ ПО ВИБРОАКУСТИЧЕСКОМУ КАНАЛУ В ПОМЕЩЕНИЯХ ОФИСНОГО ТИПА

Алексеев В.В., *д.т.н.*
e-mail: vvalex1961@mail.ru
Моисеева М.В.
e-mail: mariyamoiseeva@mail.ru
Яковлев А.В.*, к.т.н.*
e-mail:yava73@bk.ru
ТГТУ

1. ВВЕДЕНИЕ

В акустических каналах утечки информации техническим демаскирующим (разведывательным) признаком объектов защиты являются акустические волны. Такие каналы утечки характерны для акустической речевой разведки (для перехвата речевой информации из мест коммуникативной деятельности человека) и акустической сигнальной разведки (для получения разведданных об акустических «портретах» различных акустических устройств, работе которых сопутствуют акустические поля).

Речь является источником акустической информации, носителем которой являются акустические сигналы. В воздушной среде сигнал распространяется в виде продольной упругой волны, которая представляет собой колебания частиц воздуха вдоль направления распространения волны. Акустический сигнал может быть перехвачен злоумышленником при помощи использования микрофона, который преобразует сигнал в электрический и позволит передать сигнал за границы контролируемого помещения.

Основными средствами защиты речевой информации от утечки по данным каналам являются звукоизоляция помещений, поиск закладных устройств и активная акустическая маскировка. Основу средств маскировки составляют генераторы помех. На практике наиболее широкое применение нашли генераторы шумовых колебаний. Применение данного метода позволяет снизить отношение сигнал-шум на входе технического средства разведки за счет увеличения уровня помехи.

Однако, акустические помехи, создаваемые техническими средствами защиты информации, не должны приносить существенный дискомфорт присутствующим в помещении участникам диалога. Возникает проблема приведения мощности акустической помехи к оптимальному уровню, удовлетворяющему и требованиям защищенности помещения, и комфортному ведению разговора.

Процесс восприятия речи в шуме сопровождается потерями составных элементов речевого сообщения. Понятность речевого сообщения характеризуется количеством правильно принятых слов, отражающих качественную область понятности, которая выражена в категориях подробности справки о перехваченном разговоре, составляемой злоумышленником.

В качестве решения было предложено использовать инструментально-расчетный метод оценки разборчивости речи, позволяющий подобрать уровень шумовых помех средства защиты таким образом, чтобы излучение генератора было минимальным, но обеспечивало защиту речевой информации от утечки. Мощность, удовлетворяющую этим критериям, предложено считать оптимальной. Таким образом, необходимо провести измерение заданных методом параметров, произвести расчет словесной разборчивости и настроить средство маскировки на оптимальный уровень излучения.

Способ формирования коррелированной по уровню, спектру и времени излучения со скрываемым сигналом речеподобной помехи является наиболее эффективным способом активной защиты речевой информации.

Использование речеподобной помехи в системе защиты информации от утечки по акустическому каналу не только является эффективным методом предотвращения утечки речевой информации, но и является более «мягким», по отношению к находящимся в помещении людям способом. Более низкий уровень шума, необходимый для достижения поставленного уровня защиты (относительно использования генераторов другого типа), спектр сигнала, близкий к речи человека – эти факторы делают работу в защищаемом помещении более комфортной и спокойной.

В системах активной защиты речевой информации в помещениях для переговоров в качестве маскирующих сигналов возможно использование речеподобных сигналов и речевых последовательностей, сформированных с учетом лингвистических особенностей языка и статистических характеристик встречаемости фонем в данном языке, а также длины слов и предложений. Общее представление системы активной защиты от несанкционированного прослушивания представлено на рис.1.

Формирование речеподобных сигналов можно выполнить компиляционным методом по базе структурных единиц речи. В результате сформированные таким образом речеподобные сигналы сохраняют все оттенки речи определенного диктора, и их весьма сложно отличить от информационных сигналов этого же диктора.

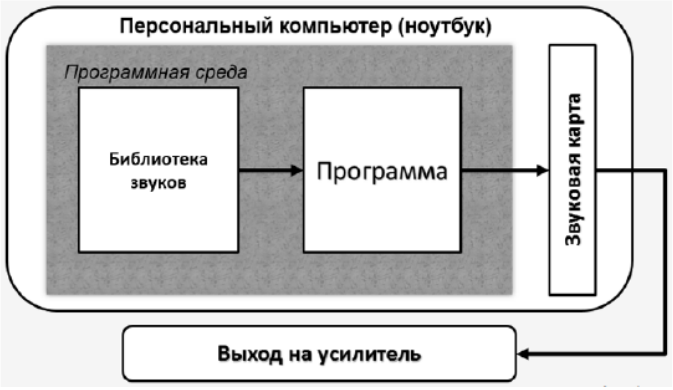


Рис.1. Общее представление системы активной защиты от несанкционированного прослушивания

Повысить качество синтезируемой речи при этом можно, используя экспоненциальные сплайн функции на границах перехода от одной фонемной структуры к другой и наложении при этом окончания одной фонемной структуры на начало второй фонемной структуры. При этом будет происходить некоторое более быстрое затухание амплитуд колебаний окончания одной фонемной структуры и увеличение амплитуды начала второй фонемной структуры. Такой механизм компиляционного синтеза речи позволит устранить скачки сигнала на границах фонемных структур. Для выполнения такого компиляционного синтеза речи нужна база фонемных структурных единиц речи с несколько увеличенными сегментами во временной области, так как при синтезе происходит наложение окончания одной фонемной структуры на начало второй фонемной структуры.

Анализ методов сегментации речи показал, что для формирования базы фонемных структурных единиц речи для синтеза речи компиляционным методом наиболее удобным является метод сегментации, использующий динамическое программирование. Для этого необходимо иметь фонетическую запись слитной речи, размеченной вручную на фонемные структурные элементы и содержащей все фонемные структурные единицы, необходимые для базы. Обычно это 300-400 аллофонов для русской, казахской, белорусской речи и около 1200 фонемных структурных единиц.

Так как фонетические базы структурных элементов речи в начале и в конце содержат переходные области, то при синтезе речи следует использовать сплайны для переходных областей. Рекомендуется использовать так называемое «сшивание» аллофонов при компиляционном синтезе речи. Переходной участок окончания предшествующего аллофона, умноженный на убывающую функцию, изменяющуюся от 1 до 0, накладывается на переходной участок последующего аллофона, умноженный на возрастающую функцию от 0 до 1. Если длины накладываемых друг на друга переходных участков не равны, то длина формируемой переходной области выбирается равной длине более длительного переходного участка (таблица 1).

Таблица 1. Значения отношений сигнал-шум, при которых обеспечивается требуемая эффективность защиты акустической (речевой) информации

Вид помехи	Словесная разборчивость W, %	Отношение сигнал/шум q_i в октавных полосах					Отношение с/ш в полосе частот 180…5600 Гц
		250	500	1000	2000	4000	
«Белый» шум	20	+0,8	-2,2	-10,7	-18,2	-24,7	-10,0
	30	+3,1	+0,1	-8,4	-15,9	-22,4	-7,7
	40	+5,1	+ 2,1	-6,4	-13,9	-20,4	-5,7
«Розовый» шум	20	-5,9	-5,9	-11,4	-15,9	-19,4	-8,8
	30	-3,7	-3,7	-9,2	-13,7	-17,2	-6,7
	40	-1,9	-1,9	-7,4	-11,9	-15,4	-4,9
Шум со спадом спектральной плотности 6 дБ на октаву	20	-14,1	-11,1	-3,6	-15,1	-15,6	-13,0
	30	-12,0	-9,0	-11,5	-13,0	-13,5	-10,8
	40	-10,0	-7,2	-9,7	-11,2	-11,7	-9,0
Шумовая речеподобная помеха	20	-3,9	-7,9	-12,9	-15,9	-16,9	-9,0
	30	-1,7	-5,7	-10,7	-13,7	-14,7	-6,8
	40	+0,1	-3,9	-8,9	-11,9	-12,9	-5,0

В системах акустической и виброакустической маскировки используются помехи как «белого» и «розового» шумов, так и

речеподобные помехи. В комплексах защиты применяют для маскировки речи помехи похожие по своей структуре на маскируемую речь. Это могут быть помехи от внешнего источника или помехи, создаваемые синтезатором речеподобных помех фонемным клонером. Помехи, создаваемые подобным синтезатором, являются не просто речеподобными, фонемный клонер обеспечивает формирование таких помех, которые в максимальной степени соответствуют звукам речи конкретного лица или группы лиц, чьи переговоры защищаются от подслушивания.

2. ОПИСАНИЕ МОДЕЛИ

Предложена модель генератора речеподобной помехи, полученная с применением методов спектрального анализа сигналов и реализована средствами математического пакета *Matlab*. Модель собрана на базе стандартных объектов среды визуального проектирования *Simulink*. В представленной модели предлагается выполнять умножение переходных участков на линейные функции, изменяющиеся от 1 до 0 и от 0 до 1. Однако в связи с тем, что чувствительность слуха имеет нелинейный характер, более эффективно применение сплайн функций более высокого порядка (до третьей степени).

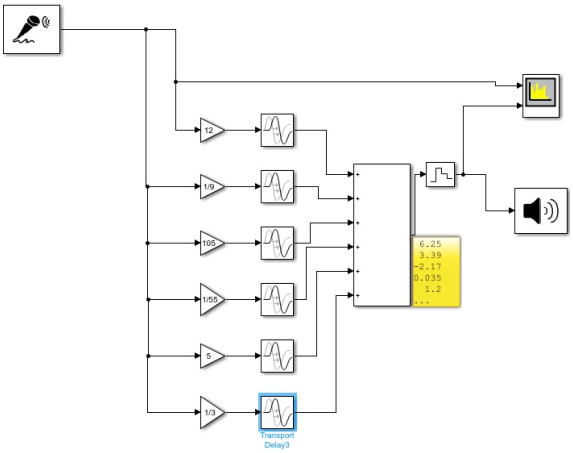


Рис.2. Модель генератора речеподобной помехи с шестью каскадами формирователя

Исследования модели показали, что увеличение количества аллофонных каскадов приводит к изменению как спектра, генерируемой речеподобной помехи, так и ее амплитуды. Эти

изменения хорошо заметны при переходе от трехкаскадной схемы к четырехкаскадной (рис.3).

На рис.4 представлены спектры речевого сообщения собеседников и синтезированный генератором речеподобная помеха. Результат моделирования показывает достаточную схожесть данных спектров.

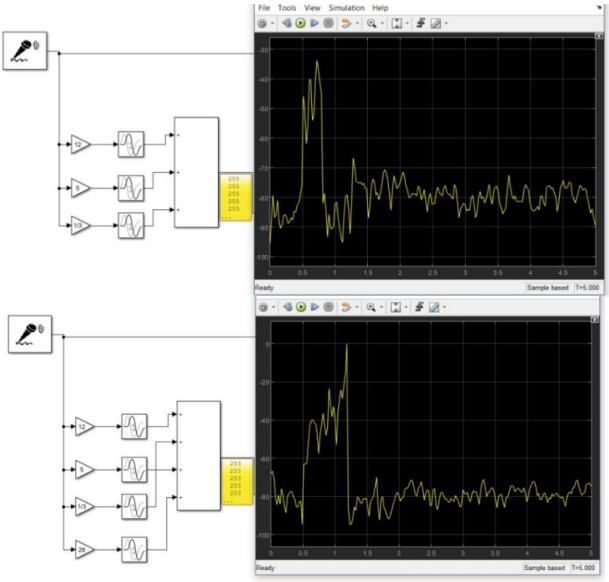


Рис.3. Результаты моделирования генератора речеподобной помехи с различным количеством каскадов формирователя

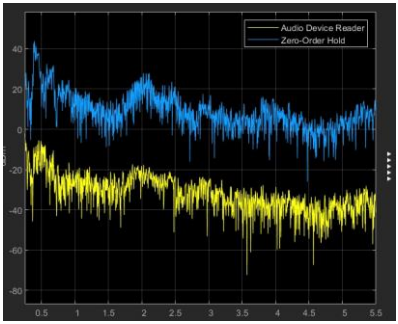


Рис.4. Спектры речевого сообщения собеседников и синтезированная генератором речеподобная помеха

3. ЗАКЛЮЧЕНИЕ

Таким образом, синтезирована модель генератора речеподобной помехи на основе методов спектрального анализа. Программная реализация модели осуществлена средствами математического пакета *Matlab*. Модель собрана на базе стандартных объектов среды визуального проектирования *Simulink*. Результаты моделирования показывают достаточную схожесть спектров речевого сообщения собеседников и синтезированная генератором речеподобная помеха. Дальнейшие исследования направлены на оценку эффективности применения такой помехи в различных ситуациях и условиях и разработку методики применения средств противодействия в помещениях офисного типа.

Литература

1. Каторин, Ю.Ф. Защита информации техническими средствами: Учебное пособие [Текст]/ Ю.Ф. Каторин, А.В. Разумовский, А.И. Спивак. – СПб: НИУ ИТМО, 2012. – 416 с.
2. Царегородцев, А.В. Методы и средства защиты информации в государственном управлении [Текст]/ А.В. Царегородцев, М.М. Тараскин. – Москва: Проспект, 2017. – 205 с.
3. Лыньков, Л.М. Основы защиты информации и управления интеллектуальной собственностью [Текст]/ Л.М. Лыньков, В.Ф. Голиков, Т.В. Борботько. – Минск: БГУИР, 2013. – 243 с.
4. Ворона, В.А. Способы и средства защиты информации от утечки по техническим каналам [Текст]/ В.А. Ворона, В.О. Костенко // Computational nanotechnology. 2016. №3. – С. 208-223.
5. Хорев, А. А. Техническая защита информации [Текст]/ А.А. Хорев. – М.: НПЦ «Аналитика», 2008. – 436 с.

УДК 004.8

ИНТЕЛЛЕКТУАЛЬНОЕ УПРАВЛЕНИЕ РОБОТАМИ-МАНИПУЛЯТОРАМИ ДВОЙНОГО НАЗНАЧЕНИЯ НА ОСНОВЕ КОМБИНИРОВАНИЯ НЕЙРОСЕТЕВЫХ И НЕЧЕТКИХ МЕТОДОВ*

***Борисов В.В.**, д.т.н.*
e-mail: vbor67@mail.ru
Филиал НИУ «МЭИ» в г. Смоленске
Соколов А.М.
e-mail: andreisokol98@gmail.com
Луферов В.С.
e-mail: lyferov@yandex.ru
НИУ «МЭИ»

1. ВВЕДЕНИЕ

Роботизация различных сфер человеческой деятельности является одним из ключевых факторов, обеспечивающих формирование и реализацию производственно-экономического и оборонного потенциалов государства. В первом случае использование современных роботизированных комплексов в производственных процессах позволяет значительно повысить производительность и эффективность функционирования узлов и агрегатов, входящих в состав промышленного оборудования. Во втором случае комплексная роботизация машин специального назначения – одна из приоритетных задач современной отечественной и зарубежной военной доктрины, основной целью которой является обеспечение безопасности жизни и здоровья личного состава, как важнейшей ценности, при осуществлении специальных задач, в том числе проведении контртеррористических операций. Автоматизированные системы способны функционировать в опасных и непригодных для пребывания человека условиях, характеризующихся высоким уровнем

* Работа выполнена в рамках проекта «Методы и технологии интеллектуального управления многозвенными роботами-манипуляторами на основе нейро-нечетких моделей» (ПНИ 20/22-0000028/47) при поддержке гранта НИУ «МЭИ» на реализацию программ научных исследований «Энергетика», «Электроника, радиотехника и IT» и «Технологии индустрии 4.0 для промышленности и робототехника» в 2020-2022 гг.

радиационного, биологического и химического заражения, а также снизить физическую нагрузку на личный состав при выполнении специальных заданий, благодаря чему находят все большее применение в составе различных формирований.

При создании роботизированных систем гражданского, двойного и специального назначения наибольшее распространение получили последовательные многозвенные роботы-манипуляторы (МРМ), которые представляют собой единую кинематическую цепь, образованную рядом последовательно соединенных звеньев. Среди МРМ особое место занимают роботы с угловой системой координат. Данная конструкция предполагает наличие звена-основания, осуществляющего поворот робота в горизонтальной плоскости относительно вертикальной оси, и прикрепленного к нему ряда последовательно соединенных звеньев с вращательными сочленениями. Манипуляторы с угловой системой координат при прочих равных условиях обладают наибольшей зоной обслуживания и при этом имеют сравнительно небольшие собственные габариты, что обеспечивает универсальность их применения при решении различных классов задач.

В этой связи разработка новых алгоритмов управления МРМ, в том числе на основе применения интеллектуальных технологий анализа информации, могут рассматриваться в качестве актуальной научной задачи.

2. ПОСТАНОВКА ЗАДАЧИ

На практике значительная часть задач управления робототехническими системами решается с учетом выбранных критериев, среди которых можно выделить:

- максимизацию скорости перемещения рабочего органа МРМ в заданную точку пространства с последующим его переводом в конечную позицию;
- обеспечение максимальной точности перемещения рабочего органа МРМ из одной заданной точки пространства, входящего в его рабочую зону, в другую.

Выбранные критерии определяют способ решения обратной задачи кинематики (ОЗК), которая заключается в определении конфигурации МРМ (вектора обобщенных координат $\vec{q}$, который в случае ПРМ с угловой системой координат представляет собой вектор присоединенных углов поворота звеньев) для обеспечения заданного пространственного положения рабочего органа $\vec{k}$.

Так с точки зрения первого критерия целесообразно максимально быстро определить решение ОЗК для перемещения рабочего органа в заданную точку пространства. Второй критерий в формализованном виде может быть записан в следующем виде:

$$r^* \to \min_{r \in R} \left\{ \sum_{i=0}^{N} \Delta q_i^r \right\},$$

$$\left\| \Phi(\vec{q}_{\text{н}} + \Delta \vec{q}^r) - \vec{k}_{\text{к}} \right\| \le \varepsilon \to r.$$

где R – множество векторов обобщенных координат, являющихся решениями ОЗК; $r = \Delta \vec{q}^r = \{\Delta q_0^r, \Delta q_1^r, \dots, \Delta q_N^r\}$ – конкретное решение ОЗК; Δq_i^r – изменение положения i-ого звена, необходимое для получения решения r, $\vec{q}_{\text{н}}$ – вектор углов, характеризующий начальное положение ПРМ, $\vec{k}_{\text{к}}$ – вектор координат целевой точки перемещения, N – количество звеньев МРМ, не считая звена-основания; ε – точность нахождения решения.

Сутью критерия является выбор из множества возможных решений ОЗК R единственный вектор r^*, который обеспечить минимальное перемещение подвижных частей МРМ относительно исходного положения.

При разработке алгоритмов управления МРМ в режиме реального времени целесообразно применять комбинацию двух указанных ранее критериев [1, 2]. Данный режим предъявляет особые требования к показателям быстродействия ПРМ, которое в свою очередь зависит от скорости обработки информации и показателей динамики составных частей робота. Это обусловило разработку так называемых комбинированных интеллектуальных алгоритмов, предполагающих выработку решений по выводу рабочего органа МРМ в область пространства (например, с помощью нейросетевой модели), максимально близкую к заданной координате, с последующей более точной подстройкой вектора углов сопряжения звеньев при помощи численных методов [3].

Анализ работ, связанных с указанным подходом [3, 4], показывает, что некоторые аспекты требуют дополнительного обоснования. К указанным аспектам относится выбор архитектуры нейронной сети, а также типа численного метода при функционировании МРМ в условиях реального времени. Данная статья посвящена решению указанных задач.

3. ПРЕДЛАГАЕМЫЙ МЕТОД

В соответствии с выбранной постановкой задачи был проведен анализ структур нейронных сетей для определения вектора

обобщенных координат , задающего конфигурацию МРМ для позиционирования его рабочего органа в заданной точке пространства $\vec{k} = (x, y, z)^T$ в абсолютной системе координат.

Для мобильных роботов специального и двойного назначения одним из ключевых факторов является требуемый объем памяти вычислителя. В этой связи предложено в качестве архитектуры нейронной сети, применяемой на первом этапе комбинированного метода, использовать полносвязную сеть прямого распространения типа многослойный персептрон.

Основными преимуществами данной сети над рассматриваемой в статьях [3-5] нейронной сетью нечеткого вывода ANFIS является относительная простота архитектуры. Также стоит отметить, что многослойный персептрон имеет достаточно высокое быстродействие и не требует сложной предварительной настройки.

На втором этапе для уточнения полученного решения предлагается использовать численный метод Гаусса-Ньютона, поскольку он в отличие от применяемого в статье [3] метода Ньютона-Рафсона является менее ресурсоемким, поскольку не требует вычисления матрицы вторых производных. При этом данный метод может применяться для часто используемых на практике избыточных ПРМ, для которых характерно превышение размерности вектора обобщенных координат $\vec{q}$ над размерностью вектора $\vec{k}$, содержащего характеризующие положение схвата значения.

Для реализации предложенного комбинированного метода была разработана компьютерная программа, блок-схема алгоритма которой приведена на рис.1.

В соответствии с блок-схемой на первом этапе пользователем задаются параметры, описывающие физическую модель МРМ рассматриваемого типа. Далее следует этап обучения нейросетевой модели, после чего программа ожидает получения координат точки пространства, в которую необходимо переместить рабочий орган МРМ, а также значение требуемой точности позиционирования. Далее производится проверка принадлежности указанной точки рабочей зоне манипулятора. Если точка не входит в рабочую зону, то выводится информационное сообщение и алгоритм возвращается к блоку ввода координат. В противном случае производится определение приближенного решения с помощью нейросетевой модели. В случае если заданная пользователем точность не была достигнута, полученное решение уточняется с помощью метода Гаусса-Ньютона. На следующем шаге полученные значения элементов вектора обобщенных

координат передаются на соответствующие приводящие механизмы МРМ.

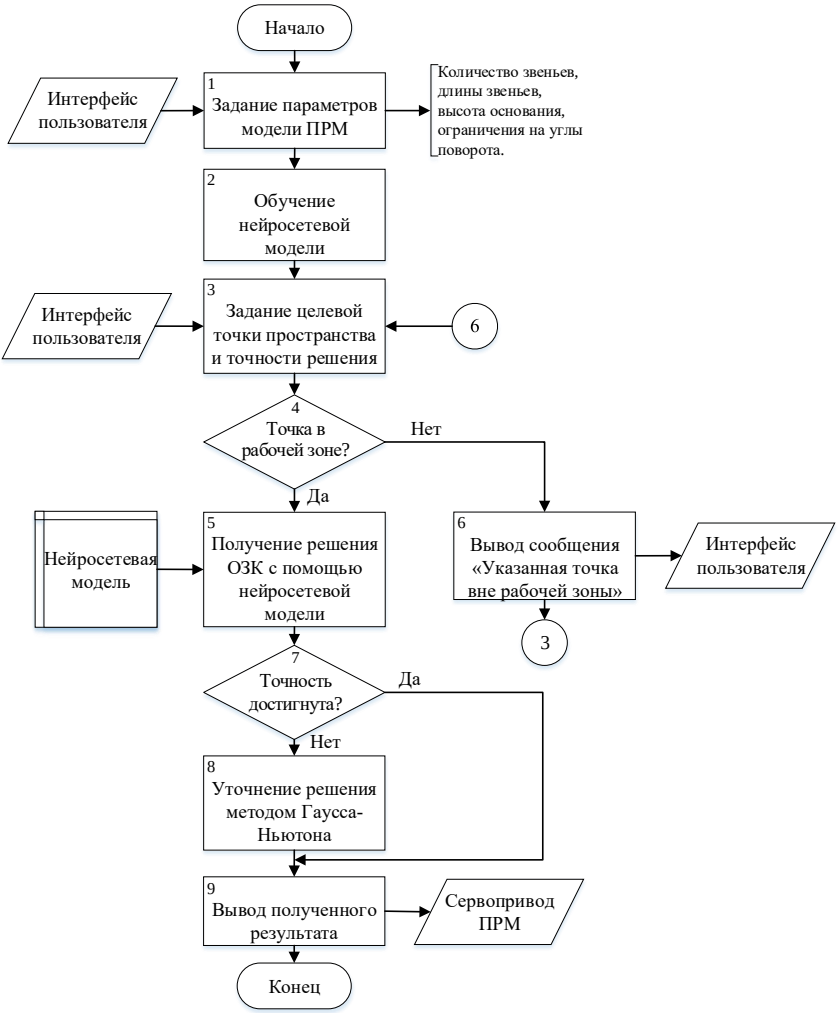


Рис.1. Блок-схема алгоритма реализации комбинированного метода

Программа предусматривает возможность трехмерной визуализации процесса функционирования МРМ, вид которой приведен на рис.2.

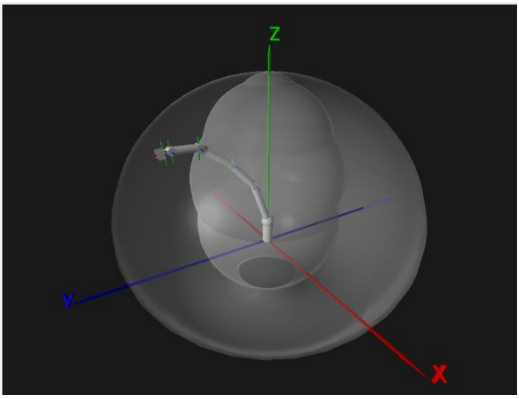


Рис.2. Визуализация процесса функционирования ПРМ с использованием предложенного метода

4. РЕАЛИЗАЦИЯ НЕЧЕТКОГО ЧИСЛЕННОГО МЕТОДА

На практике могут возникать ситуации, при которых невозможно точно идентифицировать положение целевого объекта захвата из-за наличия факторов внешней среды. К таким факторам в первую очередь могут быть отнесены воздействия, влияющие на системы технического зрения. Данная проблема особенно актуальна для мобильных роботов специального назначения, работающих в агрессивных условиях внешне среды.

В случае неопределенности при идентификации точного пространственного положения целевого объекта, область его возможного местоположения может быть описана вектором, компоненты которого являются нечеткими числами. Функции принадлежности нечетких чисел представляют собой выпуклые унимодальные кривые, которые в общем случае могут иметь любую форму. Конкретная форма определяется на основе экспертной оценки путем анализа полученных для конкретных факторов неопределенности статистических данных. Для примера в работе будет рассматриваться представление координат целевого объекта в виде нечетких чисел с функциями принадлежности треугольной формы, вид которых приведен на рис.3.

В заданной постановке прямая задача кинематики (ПЗК) может быть записана в виде:

$$\tilde{\boldsymbol{k}} = \tilde{\Phi}(\tilde{\boldsymbol{q}}),$$

где $\tilde{\Phi}(\tilde{\boldsymbol{q}})$ – матрица преобразования.

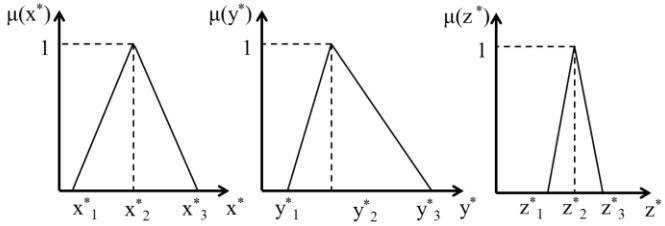


Рис.3. Представление координат положения целевого объекта в виде нечетки чисел с треугольной функцией принадлежности

В общем виде система, описывающая кинематику МРМ с угловой системой координат может быть представлена в виде системы нелинейных уравнений:

$$\begin{cases} \tilde{x} = \left[\sum_{i=1}^{N} L_i \cdot \sin\left(\sum_{j=1}^{i} \tilde{q}_j\right)\right] \cdot \cos(\tilde{q}_0); \\ \tilde{y} = \left[\sum_{i=1}^{N} L_i \cdot \sin\left(\sum_{j=1}^{i} \tilde{q}_j\right)\right] \cdot \sin(\tilde{q}_0); \\ \tilde{z} = L_0 + \sum_{i=1}^{N} L_i \cdot \cos\left(\sum_{j=1}^{i} \tilde{q}_j\right). \end{cases}$$

В указанной системе уравнений независимые переменные также будут представлять собой нечеткие числа с треугольной функцией принадлежности. При этом четкие значения координат положения схвата МРМ в начальный момент времени могут быть представлены в виде вырожденной треугольной функции принадлежности, у которой граничные значения совпадают с модальным.

Для вычисления значений косинуса и синуса от углов, представленных нечеткими числами, соответствующие тригонометрические функции представляются в виде разложения в ряд Тейлора с необходимым для достижения требуемой точности расчета числом членов:

$$\sin(\tilde{q}) = \sum_{n=0}^{\infty} \frac{(-1)^n \tilde{q}^{2n+1}}{(2n+1)!},$$

$$\cos(\tilde{q}) = \sum_{n=0}^{\infty} \frac{(-1)^n \tilde{q}^{2n}}{(2n)!}.$$

В общем виде ОЗК может быть сформулирована в виде следующего выражения:

$$\tilde{\boldsymbol{q}} = \tilde{\Phi}^{-1}(\tilde{\boldsymbol{k}}),$$

где $\tilde{\Phi}^{-1}(\tilde{\boldsymbol{k}})$ – матрица, обратна матрице преобразований.

В условиях нечетко заданного положения целевого объекта предлагается использование нечеткого метода Гаусса-Ньютона.

Для нахождения численного решения может быть определена функция:

$$\tilde{F}(\tilde{\boldsymbol{q}}) = \left\|\tilde{\Phi}(\tilde{\boldsymbol{q}}) - \tilde{\boldsymbol{k}}\right\|^2 \to \min|_{q\subset R^n},$$

где квадрат нормы представляет собой нечеткое расстояние между текущим позиционированием схвата МРМ и заданным положением целевого объекта.

Матрица Якоби:

$$\tilde{J}(\tilde{\boldsymbol{q}}) = \left\{\frac{\partial \tilde{F}_j(\tilde{\boldsymbol{q}})}{\partial q_i}, j = 0 \dots 2, i = 0 \dots N\right\}.$$

Правая и левая псевдообратные матрицы Якоби:

$$\tilde{J}^+ = (\tilde{J}^T \cdot \tilde{J})^{-1} \cdot \tilde{J}^T,$$
$$\tilde{J}^+ = \tilde{J}^T \cdot (\tilde{J}^T \cdot \tilde{J})^{-1}.$$

Формула для вычисления очередного значения алгоритма имеет следующий вид:

$$\tilde{\boldsymbol{q}}^{[i+1]} = \tilde{\boldsymbol{q}}^{[i]} - \alpha_i \cdot \tilde{J}^+\left(\tilde{\boldsymbol{q}}^{[i]}\right) \cdot \tilde{F}(\tilde{\boldsymbol{q}}^{[i]}),$$

где $\alpha_i \leq 1$ – величина шага, которая может иметь постоянное или переменное значение.

При выполнении итерационных операций с нечеткими числами возникает эффект «накопления» неопределенности, который заключается в увеличении на каждом очередном шаге алгоритма нечеткости получаемого результата, что может негативно сказываться на качестве последнего. Для минимизации эффекта накопления неопределенности предлагается использовать подход, изложенный в работе [6], который заключается в использовании взаимодействия нечетких чисел на основе их модальных значений.

Критерий останова может быть записан в виде:

$$\left|\tilde{\boldsymbol{q}}^{[i+1]} - \tilde{\boldsymbol{q}}^{[i]}\right| > \varepsilon, \qquad (2)$$

где ε – четкое значение, задающее точность вычисления.

В формуле (2) расстояние между соответствующими компонентами вектора нечетких углов на прошлом и предыдущем шагах может быть вычислено как псевдометрика Евклида:

$$d_{\mathrm{E}}(A,B) = \sqrt{\sum_{i=1}^{n} (\mu_A(x_i) - \mu_B(x_i))^2}, \quad x_i \in X.$$

Для осуществления операций с нечеткими числами предлагается использовать нечеткий интервальный метод, представленный в работе [7].

5. ЗАКЛЮЧЕНИЕ

Анализ специфики применения ПРМ двойного и специального назначения показали, что для достижения высокой скорости нахождения решения ОЗК целесообразно применение двухэтапной вычислительной

процедуры. Указанная двухэтапная процедура предполагает последовательное применение нейронной сети для получения приближенного решения и последующего использования численного метода для его уточнения.

Выбор рационального варианта структуры и параметров нейронной сети, а также типа итерационного метода был основан на том, что критерием управления многозвенным избыточным ПРМ является обеспечение высокого быстродействия при соблюдении допустимой точности позиционирования рабочего органа, а также простота программной реализации, допускающая возможность снижения требований к вычислительным устройствам, входящих в состав рассматриваемых робототехнических систем. В результате в качестве архитектуры нейронной сети был выбран многослойный персептрон, а в качетстве численного метода – метод Гаусса-Ньютона.

Предложен нечеткий метод Гаусса-Ньютона, позволяющий осуществлять решение ОЗК в случае, когда невозможно точно определить пространственное положение захватываемого объекта.

Литература

1. Борисов В.В., Мисник А.Е., Комбинированный нейросетевой способ моделирования для оперативного управления сложными системами // Информационные технологии. – 2012 – № 7 – С. 69-72.
2. Obychaiko D., Shikhin V. Development of an algorithm to maintain the performance of the cyber-physical system by introducing an updated indicator of operational reliability // Journal of applied informatics – 2018 – V. 13 –№.6(78) – P. 60-70.
3. Ганин П. Е., Кобрин А.И. Методика построения гибридной нейросистемы реального времени для решения обратной задачи кинематики избыточного манипулятора. // Вестник МЭИ. – 2018 – № 4. – С. 128-137.
4. Ganin P.E., Kobrin A.I. Hierarchical adaptive control system of a manipulator based on the synthesis of a neural network of fuzzy inference and an iterative refinement algorithm // Herald of the Bauman Moscow State Technical University. Series Instrument Engineering – 2019. – № 4 (127) – P. 18-31.
5. Vladimirov G., Koceski S. Inverse Kinematics Solution of a Robot Arm based on Adaptive Neuro Fuzzy Interface System // International Journal of Computer Applications – 2019 – №178(39). – P. 10-14
6. Бобков В.И., Борисов В.В., Дли М.И. Подход к исследованию теплопроводности нечеткими численными методами в условиях неопределенности теплофизических характеристик // Системы управления, связи и безопасности. 2017. № 3. С. 73-83.
7. Федулов А.С. Вид взаимодействия нечетких чисел, ограничивающий возрастание неопределенности при выполнении операций нечеткой арифметики // Вестник Московского энергетического института. Вестник МЭИ. 2006. № 1. С. 101-109.

УДК 007.52+004.896: 656.052.48: 519.876.5

АДАПТИВНОЕ УПРАВЛЕНИЕ ПЕРЕДВИЖЕНИЕМ КОЛЛЕКТИВА МОБИЛЬНЫХ РОБОТОВ РАЗВЕРНУТЫХ В ЛИНИЮ*

Веселов Г.Е., *д.т.н.*
e-mail: gev@sfedu.ru
Лебедев Б.К.*, д.т.н.*
e-mail: lebedev.b.k@gmail.com
Лебедев О.Б. *к.т.н.*
e-mail: lebedev.ob@mail.ru
Южный федеральный университет

1. ВВЕДЕНИЕ

Данная работа посвящена актуальной проблеме математического моделирования и теории управления: задаче децентрализованного управления мультиагентной системой, состоящей из агентов, моделирующих автономных роботов, с целью обеспечения движения группы роботов, развернутых в линию [1].

Группу дронов, развернутых в линию, можно задействовать в поисково-спасательных операциях для быстрого обследования гигантских территорий. Другой пример использования: скоординированная атака роя беспилотных летающих устройств − у них могут быть разные задачи, например, кто-то обеспечивает связь с базой, кто-то ставит помехи, а часть роя является барражирующими боеприпасами. Возможны группы роботов, используемые при вспашке или убора урожая на больших площадях [2].

Конструктивно, робот включает в себя движители, манипуляторы и их схваты, двигатели, различные датчики, устройства связи и т.п. Тип движителя робота определяется реализуемым им способом передвижения в пространстве, который, в свою очередь обусловлен средой, в которой должен функционировать робот. В работе предполагается использование движителей для перемещения по воздуху и по земле.

Для равномерного распределения агентов в зоне выполнения миссии, поддержания устойчивой связи внутри группы и *исключения*

* Работа выполнена при финансовой поддержке РФФИ, проект №19-07-00645

столкновений роботы должны при движении соблюдать некоторую геометрическую структуру (определённое расположение относительно друг друга внутри строя или относительно центра масс группы, образующее определённую геометрическую фигуру). Возникает необходимость проработать математическую модель, описывающую движение группы роботов, и разработать децентрализованное правило управления и алгоритм, которые бы позволили осуществлять эффективное управление движением агентов с сохранением геометрической формы строя [3].

Целью данной работы является разработка полностью децентрализованного правила управления группой агентов, моделирующих мобильных роботов, которое бы обеспечивало движение группы агентов с соблюдением определённой геометрической структуры строя (определённых взаимных расстояний относительно друг друга), при условиях полной автономности агента и возможности получения информации только от своих ближайших соседей. При этом искомое правило управления должно быть работоспособно в условиях возникновения внештатных ситуаций, описанных в [4, 5, 6].

2. УПРАВЛЕНИЕ ДВИЖЕНИЕМ ГРУППЫ РОБОТОВ РАЗВЕРНУТЫХ В ЛИНИЮ

При совместном движении двух и более робототехнических комплексов (РТК) важно знать их взаимное расположение с высокой точностью [7, 8] Каждому из роботов необходим набор датчиков, позволяющих ему получать информацию об окружающей среде. Поэтому, для полноценной работы мобильного РТК роботы оснащены радиочастотными датчиками определения взаимного местоположения. В состав каждого интеллектуального агента (мобильного робота), входит решающее устройство, получающее данные от других устройств по беспроводным каналам. Производя их обработку, устройство формирует команды автопилоту. Обрабатывая считанную сенсорами информацию, система управления движением решает задачу построения пути перемещения [9].

Роботы, развернутые в линию, пронумерованы от R_0 − ведущего робота до R_n − замыкающего линейку роботов. Этот тип движения требует, чтобы каждый робот R_i (i=1,2, …, n), кроме первого (ведущего), двигался по траектории, параллельно той по которой движется робот R_{i-1}, с одинаковой скоростью $V_i=V_0$ с соблюдением постоянного расстояния D между каждой парой соседних роботов R_i и R_{i-1}, при этом робот R_i выравнивает траекторию по траектории робота R_{i-1}.

Ведущий робот R_0 может быть оснащен сенсорной системой, позволяющей ему планировать траекторию в непредсказуемой обстановке, в то время как остальные роботы должны просто повторять траекторию своего ближайшего соседа. На рис.1 представлена схема движения коллектива мобильных роботов, выстроенных в шеренгу, относительно базовой линии [9-10].

В работе рассматриваются системы, в которых агенты перемещаются в обычном евклидовом пространстве с дискретным временем.

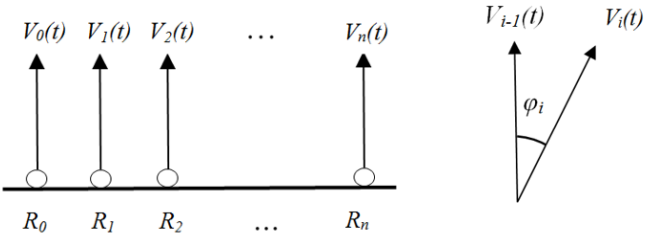


Рис.1. Схема движения роя мобильных роботов, развернутых в линию

Известен вектор $W_i=(x_i, y_i, \alpha_i, h_i, V_i)$ состояния каждого из роботов в абсолютной системе координат, где (x_i, y_i) − координаты место расположения робота, V_i − скорость передвижения робота, α_i − угол между вектором V_i и базовой линией Line, h_i – модуль вектора V_i. Заданы α_0 − для ведущего робота и D − расстояние в линии между соседними роботами R_i и R_{i-1}.

Рассмотрим теперь задачу управления ведомыми роботами. Главная идея состоит в том, что целью управления ведомыми роботами является выполнение (достижение R_i с курсом робота R_{i-1}) равенств:

1. $V_{i-1}(t)=V_i(t)$
2. $h_{i-1}(t)=h_i(t)$
3. $x_i(t)-x_{i-1}(t)=D$
4. $y_i(t)=y_{i-1}(t)$
5. $\alpha_i(t)=\alpha_{i-1}(t)=\alpha_0$

$\varphi_i(t)=\alpha_i(t)-\alpha_{i-1}(t)$ − угол между $V_{i-1}(t)$ и $V_i(t)$.

Способ управления заключается в том, чтобы в процессе движения в каждый момент времени t производится перепланирование параметров движения каждого робота R_i, движущегося параллельно с соседним роботом R_{i-1}. Ошибкой робота R_i является:

$$\delta_i=k_1(|D-|(x_i(t)-x_{i-1}(t)||+k_2|(y_i(t)-y_{i-1}(t)|+k_3|(h_i(t)-h_{i-1}(t)|)+k_4|(\alpha_i(t)-\alpha_0)|).$$

Ошибка перемещения роботов: $\Delta=\sum_i \delta_i$.

Программная траектория каждого ведомого робота формируется следующим образом. Закон управления, ограничивается рассмотрением кинематической модели. Здесь не учитываются инерционные и конструктивные параметры робота.

Максимальные значения отклонения скорости $\delta V=(V_i(t+1)-V_i(t))$ − и угла «рыскания» $\varphi_{i,i-1}(t+1)$ $\varphi_i(t)=\alpha_i(t)-\alpha_{i-1}(t)$ устанавливаются системой управления в допустимых пределах.

Правило 1. (Устранение расхождения курса робота R_i с курсом робота R_{i-1}). Если $\varphi_i(t)>0$, то выполняется коррекция угла $\alpha_i(t)$ между V_i и базовой линией Line с целью уменьшения $\varphi_i(t)$, в соответствии с выражением (5). Цель коррекции − минимизация величины $\varphi_i(t)$.

Правило 2. Если существует отклонение $\varepsilon_h=h_i(t)-h_{i-1}(t)$ и время отработки отклонения не превышает наперед заданных пороговых значений, то выполняется коррекция скорости $h_i(t)$ движения по текущему отклонению от заданной траектории следования псевдоцели, в соответствии с выражениями $h_i(t+1)=h_i(t)+\xi_v$. Цель коррекции − минимизация величины ξ_v.

Правило 3. Если $y_i(t)-y_{i-1}(t)=\varepsilon_y$ и $|\varepsilon_y)|>0$, то выполняется коррекция $y_i(t)=y_i(t)+\xi_y$. Цель коррекции − минимизация значения $|\varepsilon_y|$.

Правило 4. Если $x_i(t)-x_{i-1}(t)=\varepsilon_x$ и $|\varepsilon_x)|>D$, то выполняется коррекция $x_i(t+1)=x_i(t)+\xi_x$. Цель коррекции − минимизация разности $(|\varepsilon_x| - D)$.

Предложена структура маневра, выполняемого роботом для коррекции отклонений параметров, заключающаяся в следующем. Сначала робот R_i изменяет значение параметра $\alpha_i(t)$ на величину δ_i, затем в течении времени δ_t робот передвигается с новым значением параметра $\alpha^*_i(t)=\alpha_i(t)+\delta_i$. По истечении времени δ_t возвращается исходное значение пararараметра, с которым робот продолжает движение. Маневр, выполняемый роботом R_i, представлен на рис.2.

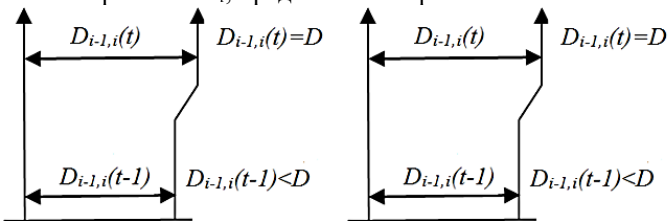


Рис.2. Схема маневра, выполняемого роботом для коррекции параметра $x_i(t)$

Первое правило предназначено для решения задачи управления при условии оказания управляющего воздействия на направление

перемещения робота, а второе правило − при условии оказания управляющего воздействия на скорость агента. Для каждого правила управления представлен его дискретный аналог.

При рассмотрении пары роботов R_{i-1} и R_i параметры: линейные скорости $V_{i-1}(t)$, $V_i(t)$, углы $\alpha_i(t)$, $\alpha_{i-1}(t)$, $\varphi_i(t)$, координаты (x_i, y_i) и (x_i, y_i), модули $h_i(t)$, $h_{i-1}(t)$ являются входными. При этом $V_{i-1}(t)$, $\alpha_{i-1}(t)$, (x_{i-1}, y_{i-1}), $h_{i-1}(t)$, являются возмущениями неопределенного характера, а параметры робота R_i: $V_i(t)$, $\alpha_i(t)$, $(x_i, y_i)(t)$, $h_i(t)$ используются в качестве параметров управления передвижением робота. На следующем шаге параметры робота R_i приобретают новые значения $V_i(t+1)$, $\alpha_i(t+1)$, $\alpha_i(t)$, $(x_i, y_i)(t+1)$, $h_i(t+1)$.

3. АЛГОРИТМ ДВИЖЕНИЯ ГРУППЫ РОБОТОВ РАЗВЕРНУТЫХ В ЛИНИЮ НА ОСНОВЕ КОЛЛЕКТИВНОЙ АДАПТАЦИИ

Для управления движением группы роботов, развернутых в линию разработан адаптивный алгоритм с обратной связью, базирующийся на методе самообучения и самоорганизации.

Подход к решению этой задачи состоит в формализации понятия отклонения текущего положения ведомого робота R_i от требуемого положения. Это достигается на основе визуальной информации и поиске такого управления с обратной связью, которое обеспечивает сведение к нулю нормы этого отклонения. Для управления движением робота параллельно траектории ведущего (предыдущего) робота R_{i-1} использованы кинематические уравнения в переменных состояния, характеризующих выполнение цели управления. Решена задача локализации роботов с использованием относительного положения.

Управление движением группы роботов, развернутых в линию осуществляется с помощью алгоритма альтернативной коллективной адаптации, основанного на идеях коллективного поведения объектов адаптации [10].

Объектами адаптации ведомого робота R_i являются рассматриваемые в качестве управлений параметры вектора W_i − положение и ориентация робота в абсолютной системе координат, и линейная скорость робота соответственно.

Коллектив объектов адаптации (их совокупность) группы роботов, развернутых в линию, соответствует объекту оптимизации (ОО).

Для реализации механизма адаптации каждому объекту (параметру p_{ij} вектора $W_i = (x_i, y_i, \alpha_i, h_i, V_i)$) сопоставляется автомат адаптации АА, моделирующий поведение объекта адаптации в среде.

Автомат адаптации (рис.3) имеет две группы состояний: $C_1=\{c_{1l}|l=1,2,\ldots,g\}$ и $C_2=\{c_{2l}|l=1,2,\ldots,g\}$, соответствующие двум альтернативам A_1 и A_2 поведения объекта адаптации в среде: A_1 – изменить размер параметра, A_2 – оставить без изменения. Таким образом, выходной алфавит автомата адаптации: $A=\{A_1, A_2\}$. Число состояний в группе задается параметром g, называемым глубиной памяти. Входной алфавит $Q=\{+,-\}$ включает возможные отклики среды: «поощрение» (+) и «наказание» (-). Граф-схема переходов АА представлена на рис.3. На вход автомата адаптации подается сигнал «поощрение» или «наказание» в зависимости от состояния объекта адаптации (соответствующего параметра вектора W_i) в среде. Знаком (+) помечены переходы в АА под действием сигнала «поощрение», знаком (-) помечены переходы под действием сигнала «наказание».

Не нарушая общности, рассмотрим принципы функционирования одного АА. Первоначально АА находится в одном из начальных состояний.

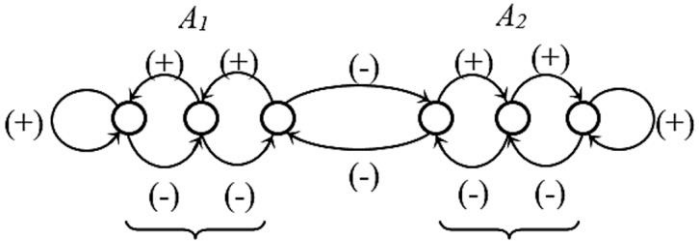


Рис.3. Структура автомата адаптации

Локальная цель объекта адаптации – достижение такого состояния, при котором его оценка равна 0.

Глобальная цель коллектива объектов адаптации заключается в достижении такого состояния S (т.е. таких значений параметров роботов) при которых «Ошибка перемещения» роботов $\Delta=\sum_i \delta_i \rightarrow min$.

Процесс альтернативной коллективной адаптации, под управлением которого осуществляется передвижение группы роботов, развернутых в линию, осуществляется за четыре такта для каждого робота на каждом шаге t (рис.4).

Рис.4. Структура алгоритма управления движением коллектива роботов, развернутых в линию

Перепланирование параметров движения роботов R_i, производится в каждый момент времени t.

На первом такте коллективной адаптации для каждого R_i рассчитываются значения параметров вектора W_i.

На втором такте для каждого автомата адаптации вырабатывается отклики среды: «поощрение» или «наказание».

Для каждого параметра p_{ij} вектора W_i рассчитывается отклонение $\mu_{ij}(t)$ от его значения в векторе W_{i-1}.

Если $sgn(\mu_{ij}(t))=0$, то для соответствующего АА вырабатывается сигнал «поощрение» (+).

Если $sgn(\mu_{ij}(t))\neq 0$, то для соответствующего АА вырабатывается сигнал «наказание» (-).

На третьем такте в каждом автомате адаптации под действием подаваемого на его вход отклика осуществляется переход в новое состояние.

На четвертом такте для каждого объекта адаптации реализуется альтернатива в соответствии с выходами АА:

1.Если АА находится в одном из состояний группы C^1ij, то значение параметра p_{ij} не изменяется.

2.Если АА находится в одном из состояний группы C^2ij, то значение параметра p_{ij} изменяется на величину пропорциональную отклонению $\mu_{ij}(t)$.

4. ЗАКЛЮЧЕНИЕ

В работе разработаны новые математические методы моделирования движения роботов. Рассматривается задача управления движением коллектива роботов, развернутых в линию. Способ управления заключается в том, чтобы в процессе движения в каждый момент времени t производится перепланирование параметров движения каждого робота R_i, движущегося параллельно с соседним роботом R_{i-1}. Управление осуществляется с помощью алгоритма альтернативной коллективной адаптации, основанного на идеях коллективного поведения объектов адаптации. Рассмотрены принципы функционирования одного автомата адаптации. Целью управления ведомыми роботами является минимизация отклонений. Для реализации механизма адаптации параметрам вектора сопоставляются автоматы адаптации, моделирующие поведение объектов адаптации в среде. Разработана структура процесса альтернативной коллективной адаптации, под управлением которого осуществляется передвижение группы роботов, развернутых в линию.

Разработаны оригинальные правила управления параметрами, обладающие рядом преимуществ по сравнению с другими методами: полная децентрализация управления в сочетании с динамической коррекцией параметров роботов, задающих положение и ориентация робота в абсолютной системе координат, и линейную скорость робота соответственно. Предложена структура маневра, выполняемого роботом для коррекции отклонений параметров. Управление осуществляется с помощью алгоритма альтернативной коллективной адаптации. В его основу положены идеи коллективного поведения объектов адаптации, что позволяет эффективно обрабатывать внештатные ситуации, такие как выход агентов из строя, изменения числа агентов вследствие выхода из строя или внезапного приобретения связи с очередным агентом, а также в условиях наличия ошибок измерений и шумов, удовлетворяющих определённым ограничениям.

Литература

1. Карпов В.Э. Коллективное поведение роботов. Желаемое и действительное // Современная мехатроника. Сб. научн. трудов Всероссийской научной школы. – 2011. – С. 35-51.
2. Карпенко А.П. Современные алгоритмы поисковой оптимизации. Алгоритмы, вдохновленные природой: учебное пособие. – М: Издательство МГТУ им. Н.Э. Баумана. – 2014. – 448 с.
3. Каляев И.А., Гайдук А.Р., Капустян С.Г. Методы и модели коллективного управления в группах роботов. – М.: Физматлит. – 2009. – 280 с.
4. Морозова Н.С. Децентрализованное управление движением строя роботов при динамически изменяющихся условиях // Искусственный интеллект и принятие решений. – 2015. – Т.1. – С. 65-74.
5. Бурдаков С.Ф., Сизов П.А. Алгоритмы управления движением мобильного робота в задаче преследования // Информатика, телекоммуникации и управление. – 2014. – №6 (210). – С. 49-58.
6. Зенкевич С.Л., Чжу Х. Управление движением группы роботов в строю типа «конвой» // Мехатроника, автоматизация, управление. – 2017. – Т.18 (1). – С. 30-34.
7. Лебедев Б.К., Лебедев О.Б., Лебедева Е.М. Гибридный алгоритм ситуационного планирования траектории на плоскости в условиях частичной неопределенности // Вестник МГТУ им. Н.Э. Баумана. Сер. Приборостроение. – 2018. – № 1. – С. 76-93.
8. Veselov G.E., Lebedev B.K., Lebedev O.B. Management of Behavior of a Swarm of Robots Applicable to the Tasks of Monitoring a Some Territory // Artificial Intelligence Methods in Intelligent Algorithms. Proceedings of the 9th Computer Science On-line Conference. Springer, Czech Republic. – 2020. – Vol. 2, – pp. 324-332.
9. Кузнецов А.В. Модель движения и взаимодействия системы интеллектуальных агентов // Вестник ВГУ. Серия: Системный анализ и информационные технологии. – 2018. – № 2. – С. 130-138.
10. Курейчик В.М., Лебедев Б.К., Лебедев О.Б. Поисковая адаптация: Теория и практика. – М.: Физматлит. – 2006. – 272 с.

УДК 681.5, 621.85-8, 007.52

О РАЗРАБОТКЕ ИНФРАСТРУКТУРЫ ДЛЯ СОЗДАНИЯ БИОПОДОБНОГО ПОДВОДНОГО РОБОТА*

Ханбекян А.Р.
e-mail: artemkhan1@yandex.ru
НИУ МФТИ
Карпов В.В.
e-mail: kvv96@mail.ru
Воробьев В.В.
e-mail: vorobev_vv@nrcki.ru
НИЦ «Курчатовский иснтитут»

1. ВВЕДЕНИЕ

Одно из наиболее активно развивающихся направлений в области робототехники является исследование и разработка биоинспирированных робототехнических систем. Адаптация различных биологических механизмов движения, обучения, паттернов поведения животных и т.д. позволяет роботам решать качественно новые задачи [1, 2], например, совместного движения [3], построения формаций [4, 5] и т.д.

С другой стороны, биоиснпирированный подход – это не только попытки использования множества форм и моделей поведения животных в системе управления робота. Имеет место и морфологическая адаптация, т.е. конечное техническое устройство обычно «похоже» на биологический оригинал. При этом речь идет не о сходстве в узком смысле этого слова, а в более широком, где оно является необходимым условием для решения поставленной задачи.

Особенно это важно для ряда задач, которые решаются в рамках подводной робототехники, например, для наблюдения за целевой акваторией с целью обнаружения в ней нежелательной активности или разведки акватории противника. В любом случае, разведывательный аппарат должен быть максимально незаметен для вражеских средств обнаружения. В связи с этим маскировка

* Работа выполнена при поддержке НИЦ "Курчатовский институт" (приказ № 1057 от 02.07.2020)

разведывательного/патрульного робота под характерных представителей флоры данной акватории, например, под рыб, является одним из путей снижения его заметности. При этом сосредоточение только на внешнем сходстве робота с рыбой не всегда достаточно, т.к. средства эхолокации легко смогут отличить робота от настоящей рыбы по издаваемым шумам его эффекторов, гидродинамике движения и т.д.

Сходство робота, в широком смысле этого слова, с оригинальным биологическим объектом, например, с рыбой, вместе с аутентичным «рыбьим поведением», снижает общую заметность аппарата для средств наблюдения. Более того, в ряде случаев даже сами рыбы могут не отличать его от своих сородичей [6].

Так как отличить искусственно созданный аппарат от животного особенно легко в момент его движения, необходимо, чтобы характер этого движения был идентичен характеру движения рыбы. У рыб выделяется два крайних типа движения: угреобразное и кузовкообразное. Они отличаются относительной длиной части тела, по которой проходит локомоторная волна – волнообразное изгибание части туловища и хвостового стебля, при котором рыба отталкивается от толщи воды телом и хвостовым плавником и числом пиков на этой волне. При этом особенно интересен, с точки зрения задачи скрытного подводного наблюдения, промежуточный ставридообразный тип движения, т.к. он наиболее распространен у большинства видов рыб.

Для создания локомоторной волны необходимо изгибать части туловища и хвостового стебля. Исходя из этого, самым простым, с конструкционной точки зрения, механизмом будет являться неподвижная основная часть, где располагаются все «жизненно важные органы», а также хвостовая часть, состоящая из движителя и хвостового плавника. Один из таких аппаратов, который использует рассматриваемый метод движения, был разработан учеными из MSU [7] (рис.1) для мониторинга экосистем. Используя механизм с только одной степенью свободы в хвосте, они добились высокой маневренности при относительной простоте конструкции. В создании локомоторной волны может использоваться не только механические движители, но и гибкая обтягивающая оболочка, как в работах [8, 9].

Одна степень свободы позволяет создать локомоторную волну, однако этого оказывается недостаточным для реализации нужной степени биоподобия волнообразного движения. Так, хвост BAUV в работе [10], приводится в движение двумя сервомоторами Maxon EC-i, колебательные движения которых должным образом синхронизированы для создания волнообразного движения. Данный

подводный аппарат не оснащен датчиками окружающей среды и управляется оператором.

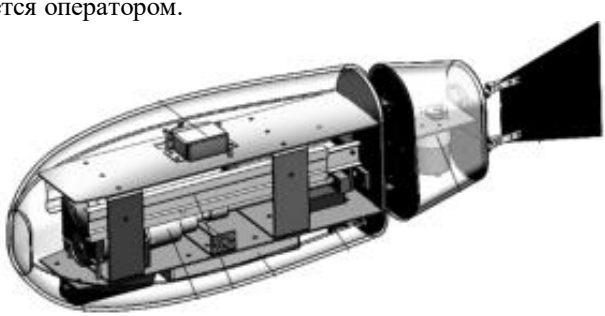


Рис.1. Робот BAUV

Количество степеней свободы в хвостовом плавнике важный параметр, однако, когда всё тело одевается в гибкую оболочку, влияние количества уходит на второй план. Более важными характеристиками становятся характер взаимодействия оболочки со средой и метод передачи изгиба механического «позвоночника».

Одним из таких роботов является робот-тунец [11] (рис.2). Он обладает только одним хвостовым сервомотором, однако данная конструкция снабжена развитым многозвенным механизмом, способным изменять направление основного движения, а также создавать подобие волнообразного изгибания тела. На данном ППА отрабатывался только принцип движения, и он не ориентируется в окружающем пространстве и не обладает как таковой системой управления.

Рис.2. Однозвенный ППА с гибкой оболочкой хвоста

Одним из самых известных прототипов робота-рыбы является американский проект «SoFi» [12] (рис.3). При крайне небольших размерах разработчикам удалось создать подводный плавательный

аппарат, способный управляться на небольшом удалении в условиях открытого моря. Движение гибкого хвоста создается циклическим потоком, исходящим из объёмного насоса, а регулировка относительного количества жидкости, закачиваемой в каждую сторону хвоста, может создать движущую силу. Глубина погружения регулируется с помощью изменения положения боковых плавников. Данный ППА управляется с пульта дайвером-оператором при помощи акустических волн. На ней установлена видеокамера, изображение с которой передается дайверу-оператору. Также робот обладает датчиком давления, что позволяет ему погружаться и всплывать в автономном режиме на заданную оператором глубину.

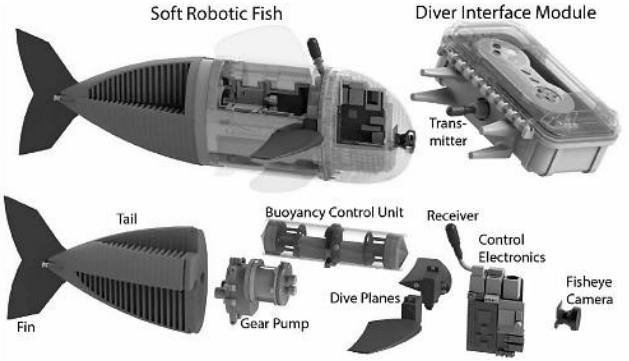


Рис.3. Робот SoFi

Существует еще множество различных технических решений [13-16], исследований в области гидродинамики движения рыб [17] и т.д. Общий анализ позволяет сделать вывод, что разработка биоподобного подводного робота, в т.ч. решающего задачи скрытного наблюдения, имеет длительную историю, которая берет свое начало как минимум с 90-х годов прошлого века [18], но пока конечный продукт в виде серийного робота отсутствует.

Во многом такой результат обусловлен междисциплинарностью проекта по разработке биоподобного аппарата. Необходима развитая сенсорная система, которая также не должна демаскировывать аппарат. Последнее замечание ставит определенные вопросы касательно использования ультразвука, например, для локализации робота в пространстве. Необходимы движители, способные передвигать аппарат в толще воды, не издавая при этом лишних шумов. Сенсорная и эффекторная часть должны быть достаточно энергоэффективны, т.к. предполагается как можно более длительная автономная работа. При

этом техническое исполнение каждого блока должно быть достаточно миниатюрно, чтобы поместиться в робота-рыбу, который размерами не должен превышать рыбу-оригинал.

2. ОПИСАНИЕ ИНФРАСТРУКТУРНОГО КОМПЛЕКСА

Целью проекта является разработка автономного подводного аппарата для задачи скрытного наблюдения за заданной акваторией. Скрытность во многом определяет использование биоподобной парадигмы разработки, т.к. аппарат маскируется под реального представителя морской фауны, что должно значительно снизить его показатели заметности для средств наблюдения. Так как этот параметр зависит от многих факторов: внешнего сходства аппарата с биологическим оригиналом, наблюдаемого характера его движения, издаваемого эффекторами шума, гидродинамической картины движения, использования активных датчиков, реакции обитателей акватории на аппарат и наоборот и др., то становится очевидной междисциплинарность подхода к ее решению. В связи с этим была поставлена задача создания инфраструктуры для тестирования и отработки различных программных, аппаратных, программно-аппаратных комплексов, технических и технологических решений и т.д.. в естественной среде обитания.

Комплекс состоит из четырех объектов: двух малых подводных роботов БПР-М-21, одного подводного робота БПР-С-21 и малой подводной лаборатории ПЛП-М-21.

Малая подвижная лабораторная платформа ПЛП-М-21

Данный ППА в режиме телеуправления должен передвигаться в водоеме и регистрировать движение роботов-мишеней с помощью целевого оборудования. Также он является базой для установки биоэнергетической установки (рис.4).

Описание конструкции. ППА имеет внешний силовой каркас, на котором установлены модули управления и целевое оборудование (см. Рис.). Имеется шесть бесщеточных двигателей: четыре из них для погружения, два – для перемещения по горизонтали. Аппарат имеет незначительную положительную плавучесть. Основное оборудование находится в трех герметичных модулях, связанных друг с другом герморазъемами: силовом, драйверном и логическом.

Система управления. Вычислительный блок построен на основе микрокомпьютера Raspberry Pi 4, с установленной на нем надстройкой над ОС – ROS. В логическом модуле находятся также периферийные устройства для удобного подключения дополнительных модулей по протоколу I2C. Управление осуществляется либо по Ethernet, либо по

Wi-Fi, через выностную антенну. Драйверы двителей находятся в отдельном модуле.

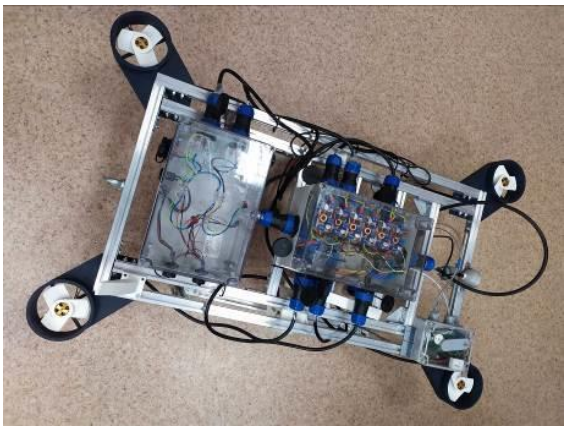


Рис.4. Фото подводной лаборатории ПЛП-М-21

Малый беспилотный подводный робот БПР-М-21

Данный ППА должен в автономном режиме плавать в водоёме, периодически погружаясь на глубину на определенное время, а затем всплывая на поверхность. Применяется в качестве «мишени», за которой наблюдает оператор, используя различные технические средства подводного обнаружения. Аппарат необходим для первоначальной отработки и тестирования различного рода эффекторов, подводной системы технического зрения и т.д.

Описание конструкции. ППА собран по принципу двухстепенного хвоста рыбы, управляемых двумя герметичными сервоприводами. Движение робота по прямой осуществляется сервоприводом №1. Поворот осуществляется при помощи сервопривода №2. Погружение/всплытие контролируется двумя грудными плавниками, которые приводятся в действие при помощи сервопривода №3. Схематичное устройство рыбы-мишени приведено на рис.5.

Система управления. Главный контроллер построен на базе AT-Mega328. Для управления сервоприводами используется контроллер Pololu Maestro. На борту имеется датчик влажности, который сигнализирует о положении робота – на поверхности воды или под водой. Так же в передней части БПР-М-21 находятся два датчика столкновения с препятствиями.

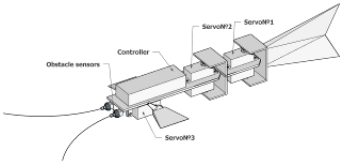


Рис.5. Условная схема БПР-М-21 и робот БПР-М-21-1

Средний беспилотный подводный робот БПР-С-21

Данный ППА выполняет функцию активного стенда для отработки и тестирования решений в области создания эффекторов нового типа, сенсорной системы, например, искусственной кожи с интегрированными датчиками давления, новых силовых материалов и др., а также алгоритмы движения и управления роботом. При этом он должен проводить тесты в естественной среде обитания.

Описание конструкции. За основу конструкции был взят окунь – пресноводная рыба-хищник, наиболее распространенная в Средней полосе России. Центральным опорно-двигательным элементом рыбы, как роботизированной, так и живой, является скелет, в центре которого находится позвоночник, к которому крепятся отростки позвонков сверху и ребра снизу. Такое строение скелета обеспечивает рыбе как подвижность и гибкость, необходимую для движения, так и защиту внутренних органов от внешних опасностей. Схема скелета БПР-С-21 и сам робот представлены на рис.6.

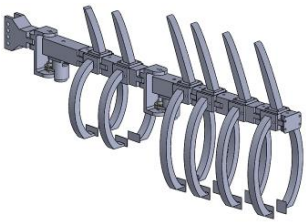


Рис.6. Схема скелета БПР-С-21 и робот БПР-С-21

Для обеспечения подвижности робота позвоночник обладает двумя степенями свободы. Первую обеспечивает подвижный хвост, который совершает колебательный движения, вторую – сам позвоночник, имеющий место сочленения, позволяющее делать изгиб. Оба места изгиба – и хвост, и позвоночник, – обеспечены ременным редуктором с соотношением 1:2 для уменьшения нагрузки на

движители. Погружение и всплытие осуществляется за счет грудных плавников.

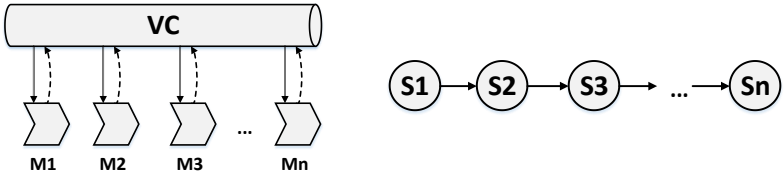
Система управления. Управление роботом осуществляется с использованием контроллера Atmega 2560, который через контроллер PololuMaestro задаёт положения сервоприводов. Всего сервоприводов 4: 1 приводит в движение хвостовой плавник, 1 обеспечивает изгиб позвоночника, еще 2 – поворот боковых плавников.

Также к контроллеру подключен набор датчиков: датчики протечки для обеспечения аварийного всплытия, датчик положения на 10 степеней свободы, включающий в себя гироскоп, акселерометр и компас, и вольтметр. Для обеспечения связи с оператором используются контроллер интерфейса RS485, а также BlueTooth-модуль.

Испытания всех изделий проводились в бассейне лаборатории робототехники НИЦ Курчатовский Институт, г. Москва, а также в бассейнах военного инновационного технополиса Эра, г. Анапа в 2021 году.

Сихронизация работы эффекторов

У животных локомоторная волна формируется за счет передачи сигналов на сокращение от позвоночного столба миомерам – составным частям мускулатуры туловища рыбы (см. рис.7а). При этом каждый миомер активируется определенным сигналом, а тот генерирует сигнал обратной связи.



а) б)

Рис.7. Синхронизация: а) синхронизация миомеров Mi; Пунктирные линии – сигналы обратной связи, б) последовательная синхронизация сервоприводов Si

Таким образом, для отработки движения важен порядок активации миомеров, т.е. управляющие сигналы должны быть синхронизированы между собой по времени. С точки зрения принципов управления, такая иерархическая система может быть сведена к схеме с последовательным запуском сервоприводов, когда каждый

сервопривод в определенный момент отправляет запускающий сигнал соседнему, см. рис.7б.

Для роботов БПР-М-21 и БПР-С-21 роль миомеров играют сервоприводы. Программа управления каждым сервоприводом задается в виде упорядоченной шестерки: $T = <A1, A2, V12, V21, T1, T2>$, где $A1$, $A2$ – минимальный и максимальный углы поворота; $V12$, $V21$ – скорости движения в прямом и обратном направлении; $T1$, $T2$ – задержки в начальной и средней фазах движения, т.е. время паузы в тактах автомата.

Эта программа реализуется конечным автоматом Мили (рис.8).

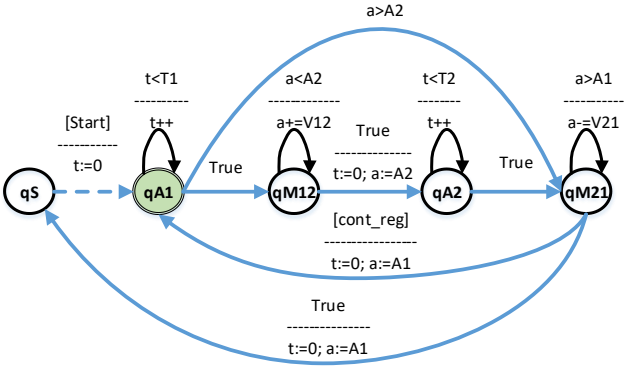


Рис.8. Управляющий автомат Мили для одного сервопривода

В таком случае синхронизация движения сервоприводов может осуществляться активизацией программ движения в моменты времени выполнения автоматом процедур на переходах $qA1 \rightarrow qM12$ и $qA2 \rightarrow qM21$. Это связано с тем, что процедуры на этих переходах выполняются одномоментно, за 1 такт. На рис.9 показан пример синхронизации таких программ. Пусть сервопривод $S1$ является лидером. В момент $P1$ (выполнение процедуры на переходе $qA1 \rightarrow qM12$) осуществляется активизация программы сервопривода $S2$, а момент $P2$ (переход $qA2 \rightarrow qM21$) – активизация программы сервопривода $S3$.

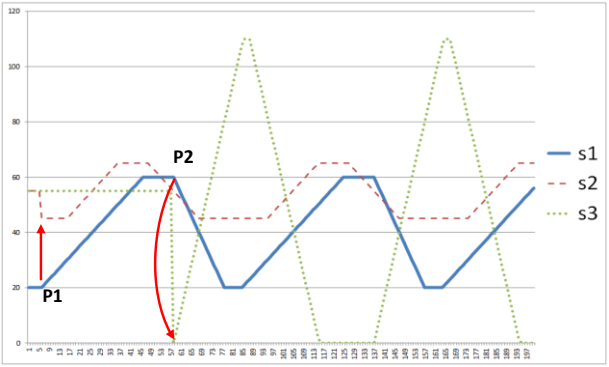
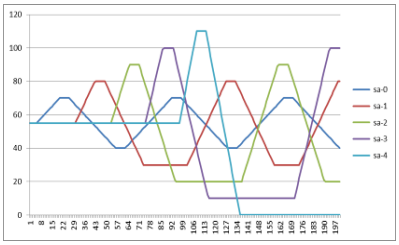
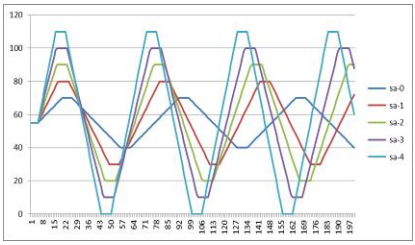


Рис.9. Пример синхронизации программ движения

На рис.10а представлена сихронизированная работа сервоприводов. Каждый сервопривод осуществляет запуск следующего за ним сервопривода. Лидер (основной инициирующий) сервопривод – «*sa*-0». На рис.10б показана независимая работа этих же сервоприводов.



а)



б)

Рис.10. Синхронизированная а) и независимая б) работа сервоприводов

3. ЗАКЛЮЧЕНИЕ

Состояние дел вокруг создания биоподобного подводного робота говорит о необходимости междисциплинарного подхода к решению этой задачи. В связи с этим возникла необходимость в создании базисного инфраструктурного комплекса, на котором можно тестировать и отрабатывать технологии в области биоэнергетики, новых материалов, методов управления, создания новых типов движителей и др. Представленная в работе инфраструктура в виде четырех аппаратов различного назначения показала свою практическую значимость в сфере разработки технологий по созданию биоподобного

подводного робота Алгоритм синхронизации эффекторов абстрагирован от объекта управления и может применяться для управления искусственными мышцами на реальном биоподобном роботе.

Литература

1. Карпов В.Э., Карпова И.П., Кулинич А.А. Социальные сообщества роботов. – 2018. – 352 с.
2. Dorigo M. et al. Swarmanoid : a novel concept for the study of heterogeneous robotic swarms // IEEE Robot. Autom. Mag. 2013. – Vol. 20, – № 4. – P. 60-71.
3. Mondada F. et al. The cooperation of swarm-bots: Physical interactions in collective robotics // IEEE Robot. Autom. Mag. 2005. – Vol. 12, – № 2. – P. 21-28.
4. Trianni V. et al. SWARM-BOT: Pattern Formation in a Swarm of Self-Assembling Mobile Robots // Proc.\ IEEE Int.\ Conf.\ Syst. Man Cybern. – 2002. – № January 2002.
5. Rubenstein M., Ahler C., Nagpal R. Kilobot : A Low Cost Scalable Robot System for Collective Behaviors.
6. Marras S., Porfiri M. Fish and robots swimming together: Attraction towards the robot demands biomimetic locomotion // J. R. Soc. Interface. – 2012.
7. Castaño M.L., Tan X. Model Predictive Control-Based Path-Following for Tail-Actuated Robotic Fish // J. Dyn. Syst. Meas. Control. Trans. ASME. – 2019. – Vol. 141. – № 7.
8. Shibata M., Sakagami N. Fabrication of a fish-like underwater robot with flexible plastic film body // Adv. Robot. – 2015. – Vol. 29. – № 1.
9. Shibata M. Fish-Like Robot Encapsulated by a Plastic Film // Recent Advances in Robotic Systems. – 2016.
10. Лушников Б.В. et al. Бионический плавающий робот для мониторинга природных и техногенных объектов в гидросфере // Электронный журнал Cloud Sci. – 2014. – Vol. 1. – № 1. – P. 61-77.
11. Masoomi S.F. et al. Design, fabrication, and swimming performance of a free-swimming tuna-mimetic robot // J. Robot. Vol. – 2014.
12. Katzschmann R.K. et al. Exploration of underwater life with an acoustically controlled soft robotic fish // Sci. Robot. – 2018. – Vol. 3. – № 16. – P. 1-12.
13. Zhong Y., Du R. Design and implementation of a novel robot fish with active and compliant propulsion mechanism // Robotics: Science and Systems. – 2016. – Vol. 12.
14. Scaradozzi D. et al. BCF swimming locomotion for autonomous underwater robots: a review and a novel solution to improve control and efficiency // Ocean Eng. – 2017. – Vol. 130.
15. Marcin M. et al. Fish-like shaped robot for underwater surveillance and reconnaissance – Hull design and study of drag and noise // Ocean Eng. – 2020. –Vol. – 217.
16. Li L. et al. Exploring the backward swimming ability of a robotic fish: Combining modelling and experiments // Int. J. Adv. Robot. Syst. 2016. – Vol. – 13 – № 5.
17. Романенко Е.В. Теория плавания рыб и дельфинов. Москва: Наука, –1986.
18. Allison Marsh. Meet Catfish Charlie, the CIA's Robotic Spy [Electronic resource]. – 2021.

СЕКЦИЯ 9 | **ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ В МЕДИЦИНЕ**

УДК 004.89

ОБЛАЧНАЯ ПЛАТФОРМА ДЛЯ ПЛАНИРОВАНИЯ МЕДИКАМЕНТОЗНОЙ ТЕРАПИИ*

***Грибова В.В.**, д.т.н*
e-mail: gribova@iacp.dvo.ru
ИАПУ ДВО РАН, ДВФУ
***Окунь Д.Б.**, к.м.н.*
ИАПУ ДВО РАН
Ковалев Р.И.
ИАПУ ДВО РАН, ДВФУ

1. ВВЕДЕНИЕ

По определению ВОЗ, рациональное использование лекарственных средств (ЛС) – это такое их применение, когда больные получают препараты в соответствии с клинической необходимостью, в дозах, отвечающих индивидуальным потребностям, на протяжении адекватного периода времени и с наименьшими затратами для себя и общества [1]. В современной медицине принято считать, что каждая болезнь нуждается в комплексном подходе. В настоящее время грань между комплексным подходом и нерациональным использованием ЛС очень тонка.

Проблема выбора оптимальной медикаментозной терапии очевидна, но в то же самое время, она связана с усложнением процесса принятия верных решений [2], что в контексте стремительно

* Работа выполнена при финансовой поддержке РФФИ, проекты № 19-29-01077 и № 18-29-03131

развивающейся, как научной, так и практической медицины, является ограничивающим фактором для достижения точности назначаемой терапии. Эффективным решением, которое может существенным образом улучшить данную ситуацию, является внедрение в систему здравоохранения и клиническую медицину систем [3], использующих в своей работе технологии искусственного интеллекта.

Построение таких систем представляет собой наукоемкий процесс, требующий больших затрат как временных, интеллектуальных, так и финансовых ресурсов, однако это создает оптимальные пациент ориентированные решения [4].

Целью данной работы является описание специализированной облачной платформы для построения интеллектуальных медицинских систем назначения медикаментозного лечения, требований к ней и алгоритма ее работы.

2. АНАЛИЗ И КЛАССИФИКАЦИЯ СУЩЕСТВУЮЩИХ СИСТЕМ ПОДДЕРЖКИ ВРАЧА ПРИ НАЗНАЧЕНИИ МЕДИКАМЕНТОЗНОЙ ТЕРАПИИ

Системы, помогающие врачам в назначении лечения, можно разделить на три основных класса: экспертные системы, основанные на знаниях, системы на основе методов машинного обучения и медицинские информационные системы.

Наиболее известными решениями, которые считаются прообразом последующих экспертных систем в медицине, являются системы MYCIN [5] и ONCOCIN [6]. MYCIN предназначена для помощи при постановке диагноза некоторых инфекционных заболеваний и рекомендации необходимого лечения, ONCOCIN решает задачу поддержки принятия решений при лечении пациентов, получающих химиотерапию.

Среди современных разработок можно выделить систему SPELTA [7] для лечения речевых и языковых расстройств; систему RTP-DSS [8] для выбора режима лучевой терапии; IndiGO (Archimedes) для формирования индивидуализированных протоколов диагностики и лечения заболеваний в кардиологии, эндокринологии, а также ряд других систем [9-11].

Известными преимуществами таких систем являются помощь врачам в подборе терапии и возможность генерации объяснений, недостатками – ориентация на определенную группу заболеваний, что ограничивает их использование при сочетанной патологии. Еще одним недостатком подобных систем является особенность реализации их баз

знаний и данных, которая не позволяет специалистам предметной медицинской области осуществлять сопровождение без IT-специалиста.

Типичным представителем второго класса систем (на основе методов машинного обучения) является система IBM Watson [12]. Для ее реализации использовалась большая обучающая выборка, которая позволила назначать лечение для онкологических заболеваний. В настоящее время данное направление исследований активно развивается и к настоящему времени предложены модели, методы и системы, помогающие врачам назначить лечение для различных заболеваний [13-17].

Основными недостатками систем подобного типа является то, что для их успешного обучения необходимо большое количество хорошо формализованных историй болезни, что в реальных условиях обеспечить практически невозможно (процесс поиска обучающей выборки, формализации историй болезни и подготовки данных очень длительный, дорогой и трудоемкий).

К третьей категории относятся всевозможные информационно-справочные системы. Представителями данного класса являются: Aarogya, Caresoft Hospital Information System, Medstar HIS, MMI Mediface, Clinical Rules, Droice Labs, ISABEL, Litmusdx [18, 19]. Такие системы основываются на различных базах данных и медицинских справочниках. Они обладают ограниченными функциональным возможностями и по сути являются справочной системой, помогающей проверить совместимость лекарственных средств, возможные дозировки, нозологии, при которых лекарственный препарат используется, подсказать врачу протокол лечения согласно справочнику, не учитывая персональные данные пациента, отражающие реальную клиническую картину.

Анализ существующих систем в области поддержки назначения лечения показал, что на текущий момент не существует программных решений, ориентированных на широкий спектр заболеваний, помогающих врачу назначить лечение с учетом персональных особенностей пациента, что делает работу по созданию программной оболочки для создания систем поддержки принятия решений, ориентированной на широкий спектр заболеваний важной и актуальной.

3. АРХИТЕКТУРА ОБЛАЧНОЙ ПЛАТФОРМЫ

В этом разделе описывается архитектура облачной платформы для назначения лечения. Платформа основана на онтологическом подходе и состоит из информационных и программных компонентов (рис.1).

3.1. Основные требования и принципы создания

Выделим следующие ключевые требования для создания специализированной облачной платформы (далее системы) для назначения медикаментозного лечения:

1. Система должна быть расширяемой, не зависеть от конкретного заболевания (либо их группы), раздела медицины, поскольку совершенно очевидно, что невозможно разработать и сопровождать множество разрозненных систем.
2. Знания в системе должны формироваться и модифицироваться в процессе эксплуатации экспертом предметной области (либо группами экспертов) на основе собственного опыта лечения заболеваний (авторских методик, если это допустимо в медицинском учреждении), либо в соответствии с принятыми клиническими рекомендациями.
3. Система должна назначать лечение с учетом персональных особенностей пациента, с детализированным обоснованием принятых решений.

Одним из решений выполнения указанных требований является использование онтологического подхода. При наличии соответствующих редакторов (управляемых онтологией) он обеспечит понятность формирования знаний, возможность реализации единственного (унифицированного) онтолого-ориентированного решателя, а также формирование детализированного объяснения [20].

3.2. Концептуальная архитектура

В состав интеллектуальной облачной платформы для назначения медикаментозного лечения входят информационные и программные компоненты (рис.1).

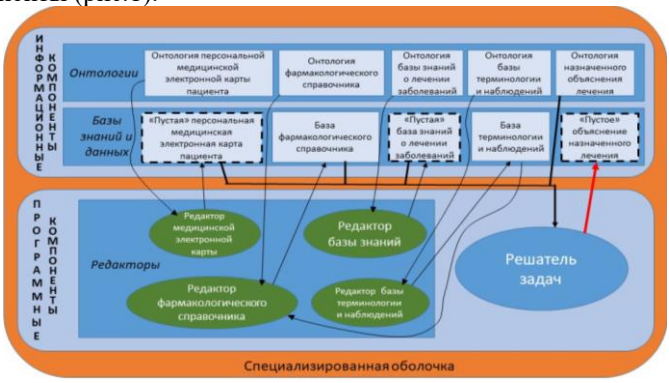


Рис.1. Концептуальная архитектура облачной платформы для назначения лечения

К информационным компонентам относятся онтологии и порожденные на их основе базы данных и знаний. К программным компонентам относятся средства редактирования знаний и данных, а также решатель задач.

3.3. Информационные компоненты

Для достижения поставленной цели был разработан ряд онтологий [21, 22]. На основе данных онтологий созданы информационные ресурсы, позволяющие формировать современные знания о лечении заболевания независимо от клинического раздела медицины.

Онтология знаний о лечении заболеваний позволяет формировать знания лечения конкретного заболевания или изложить, согласно определенного документа, знания о лечении некоторой группы заболеваний, которые имеют общие патогенетические принципы, этиологическую компоненту или клинически важные симптоматические проявления.

Каждый элемент онтологии включает сложно структурированный блок условий, позволяющий описать в формальном представлении необходимые клинические критерии, которые определяют условия и его использования в лечении данного заболевания.

Действующее вещество – перечень ЛС, из которого выбирается одно ЛС с наилучшими характеристиками для конкретного клинического случая. Каждое ЛС описывается группой терминов определяющих его клиническую необходимость (рис.2).

Рис.2. Фрагмент онтологии базы знаний о лечении заболеваний, группа альтернативно используемых ЛС

Онтология объяснения назначенного лечения описывает структуру объяснения результата работы системы. Объяснение, генерируемое на основе онтологии, формирует три основных блока:

- назначенное лечение (множество действующих веществ с указанием формы, дозировки – суточной и курсовой, а также объяснение

того, какие цели реализует назначение каждого из действующих веществ);

- не назначенное лечение – множество отклоненных системой действующих веществ с указанием того, почему не назначено то или иное действующее вещество для данного пациента (на основе анализа его электронной карты);
- множество действующих веществ, для назначения которых не хватило тех или иных исходных данных из электронной карты, например, информации о заболеваниях и наличии аллергий на лекарства.

Онтология фармакологического справочника включает элементы, позволяющие описать непосредственно лекарственное средство в виде международного непатентованного наименования (МНН), а также раздел специфических фармакологических терминов.

Онтология базы медицинской терминологии и наблюдений описывает структуру используемых на практике медицинских терминов и наблюдений, а также структуру вспомогательной терминологии необходимую для целостности описания медицинских знаний. Они могут иметь качественные, количественные или составные значения. Если значение является составным, то оно состоит из множества характеристик. В настоящее время база медицинской терминологии содержит более 27 тысяч различных значений.

Онтология персоняльной электронной медицинской карты (ПЭМК) задает структуру описания здоровья человека на всем его жизненном пути. Созданный на основе данной онтологии ресурс позволит быстро находить существующую и добавлять новую информацию обо всех случаях оказания пациенту медицинской помощи, а также в автоматизированном режиме формировать медицинские документы.

3.4. Программные компоненты

Программные компоненты оболочки включают в себя решатель задач и средства редактирования знаний и данных.

Решатель задач представляет собой онтолого-ориентированный алгоритм (ontological reasoner), который для назначения персонифицированного лечения производит обход декларативной базы знаний о лечении некоторого заболевания в соответствии с онтологией и сопоставляет информацию из базы знаний входной информации, которой является ПЭМК. Задача такого «обхода» состоит в поиске элементов (наблюдений) в ПЭМК, влияющих на назначение лечения и

сравнение их значений с диапазонами значений соответствующих наблюдений в базе знаний с целью построения персонифицированного назначения ДВ.

Решатель считывает значение термина <Заключительный диагноз>, производит поиск этого значения в базе знаний о лечении заболеваний. В соответствии с базой знаний о заболевании анализирует модели терапии, их виды, цели, схемы приема на каждом шаге проверяя соответствующие условия, значения которых анализируются по медицинской электронной карте. В результате работы решателя формируется информационный ресурс с детализированным объяснением назначенного лечения для конкретного пациента, либо формируется объяснение, почему то или иное ДВ не назначено: каких признаков (из анамнеза жизни, лабораторных, инструментальных, объективных методов исследования и др.) не хватает для назначения лечения.

Средства редактирования знаний и данных представлены *рядом редакторов* (рис.3). Все редакторы управляются соответствующими онтологиями. Интерфейсы редакторов ориентированы на экспертов предметной области. Они позволяют просматривать с разной степенью детализации и в разной форме фрагменты базы знаний, контролировать правильность вводимой информации с учетом ограничений, определенных в онтологии, полноту формирования базы знаний или данных.

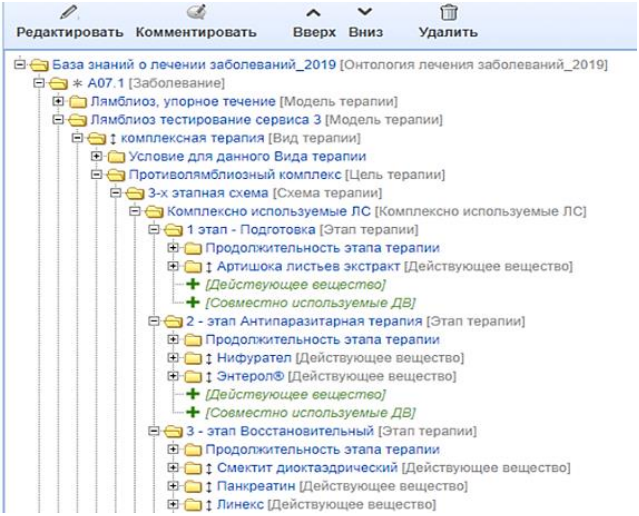


Рис.3. Интерфейс редактора базы знаний о лечении лямблиоза

4. РАБОТА СИСТЕМЫ

Специализированная оболочка для назначения персонифицированного медикаментозного лечения заболеваний реализована и доступна для использования на облачной платформе IACPaaS (https://iacpaas.dvo.ru) [25].

С использованием данной оболочки были разработаны системы для лечения заболеваний из различных разделов медицины. Учитывая, что создание системы сводится к созданию базы знаний, были разработаны следующие базы: инфекционные болезни – «База знаний по лечению лямблиоза» (рис.3), неврология – «База знаний по симптоматической терапии когнитивных нарушений», гастроэнтерология – «База знаний по медикаментозной терапии язвенной болезни желудка и двенадцатиперстной кишки» [21], гематология – «База знаний по лечению анемии у детей».

На рис.4 и рис.5 представлены ПЭМК пациента с диагнозом анемия и сформированное лечение с объяснением, согласно которому, было назначено лечение, соответствующее цели терапии «Нормализация биоценоза кишечника», так как было выполнено только ее условие.

Рис.4. ПЭМК пациента

Кроме того, была разработана база знаний лечения осложнений коронавирусной инфекции COVID-19 методами традиционной китайской медицины. База знаний была разработана на основе Diagnosis and Treatment Protocols for Patients with Novel Coronavirus Pneumonia

(Trial Version 6, Revised), подготовленной The General Office of National Health Commission of China, от 18 февраля.

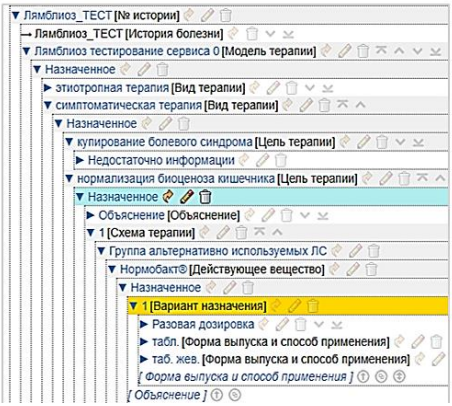


Рис.5. Лечение, назначенное системой

Система назначения лечения методами китайской терапии, реализованная на основе разработанной оболочки, использовалась в феврале 2020 в Китае для назначения лечения коронавирусной инфекции методами китайской традиционной медицины [24].

5. ЗАКЛЮЧЕНИЕ

Особенностями разработанной облачной платформы, принципиально отличающими ее от других медицинских систем, является:

- ориентация на широкий спектр заболеваний; для реализации конкретной системы на основе оболочки необходимо добавить в нее соответствующие знания о лечении заболевания;
- все информационные ресурсы формируются с использованием редакторов знаний/данных, управляемых соответствующими онтологиями, что соответствует современному подходу к разработке интеллектуальных систем и обеспечивает контроль целостности сформированной базы знаний; при ее формировании также учитываются семантические и синтаксические ограничения;
- база знаний отделена от программного кода, что позволяет экспертам предметной области обеспечивать непрерывное ее совершенствование без участия программистов (без изменения онтолого-ориентированного решателя);
- результатом работы системы являются не только возможные варианты лечения, соответствующие персональным особенностям пациента (на каждом шаге выбора терапии проверяются условия назначения), но и

детализированное объяснение, почему-то или иное ЛС назначено/не назначено, либо подсказка врачу о необходимости провести дополнительные исследования для назначения терапии;

- наличие единой онтологии, по которой формируются базы знаний о планировании лечения и онтологического решателя, позволяет назначать лечение при сочетанной патологии (нескольких заболеваниях одновременно); при этом осуществляется контроль возможного дублирования назначений;
- онтология позволяет описать несколько моделей терапии для лечения одного заболевания что, при накапливании прецедентов, позволяет не только верифицировать базу знаний, но и обеспечить анализ эффективности терапии;
- платформа является облачной, что позволяет увеличить аудиторию пользователей, а также разрабатывать базы знаний о лечении экспертам предметной области независимо от их географического положения.

Литература

1. World Health Organization et al. The pursuit of responsible use of medicines: sharing and learning from country experiences. – World Health Organization, 2012. – №. WHO/EMP/MAR/2012.3.
2. Halford, G. S., Baker, R., McCredden, J. E., & Bain, J. D. (2005). How many variables can humans process? Psychological science, 16(1), 70-76.
3. Lysaght, T., Lim, HY, Xafis, V. et al. ABR (2019) 11: 299. https://doi.org/10.1007/s41649-019-00096-0.
4. Walsh, Seán, et al. "Decision support systems in oncology." JCO clinical cancer informatics 3 (2019): 1-9. DOI: 10.1200/CCI.18.00001.
5. William van Melle MYCIN: a knowledge-based consultation program for infectious disease diagnosis, International Journal of Man-Machine Studies, 1978
6. Edward H. Shortliffe, A. Carlisl e Scott, Miriam B. Bischoff Oncocin: an expert system for oncology protocol management. – 1981
7. Vladimir E. Robles-Bykbaev, Martín López-Nores, José J. Pazos-Arias, DaysiArévalo-Lucero “SPELTA: An expert system to generate therapy plans for speech and language disorders”.
8. Watkins, W., et al. "A Radiation Therapy Treatment Planning Decision Support System (RTP-DSS) for Selecting Patient-Specific Optimal Treatment." International Journal of Radiation Oncology• Biology Physics 96.2 (2016): S82. DOI: https://doi.org/10.1016/j.ijrobp.2016.06.207.
9. Khozeimeh, Fahime, etal. "An expert system for selecting wart treatment method." Computers in biology and medicine 81 (2017): 167-175, DOI: 10.1016/j.compbiomed.2017.01.001.
10. Jiang, Xia, et al. "A clinical decision support system learned from data to personalize treatment recommendations towards preventing breast cancer metastasis." PloS one 14.3 (2019). https://doi.org/10.1371/journal.pone.0213292.

11. Abu-Naser, Samy& Al-Dahdooh, Rami. (2016). Lower Back Pain Expert System Diagnosis And Treatment. Lower Back Pain Expert System Diagnosis And Treatment.
12. IBM Watson has graduated from school and got a gob [Internet]. Available at: http://habrahabr.ru/company/ibm/blog/169067/ Accessed Apr06, 2020.
13. Rajkomar, A., Dean, J., &Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347-1358. DOI: 10.1056/NEJMra1814259.
14. Yahyaoui, A., Jamil, A., Rasheed, J., &Yesiltepe, M. (2019, November). A Decision Support System for Diabetes Prediction Using Machine Learning and Deep Learning Techniques. In 2019 1st International Informatics and Software Engineering Conference (UBMYK) (pp. 1-4). IEEE. DOI: 10.1109/UBMYK48245.2019.8965556.
15. Yuan, Y., Huang, J., & Yan, K. (2019, August). Virtual Reality Therapy and Machine Learning Techniques in Drug Addiction Treatment. In 2019 10th International Conference on Information Technology in Medicine and Education (ITME) (pp. 241-245). IEEE. DOI: 10.1109/ITME.2019.00062.
16. Anakal, S., & Sandhya, P. (2017, December). Clinical decision support system for chronic obstructive pulmonary disease using machine learning techniques. In 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT) (pp. 1-5). IEEE. DOI:10.1109/iceeccot.2017.8284601.
17. Ray, M., &Qidwai, U. (2003, April). Machine learning in medicine: calculating the minimum dose of haemodialysis using neural networks. In Annual Technical Conference IEEE Region 5, 2003 (pp. 23-27). IEEE. DOI: 10.1109/REG5.2003.1199705.
18. Top 10 Hospital Information System You Need to Know About, 2019, [cited 2020 Apr.07]/ Available from: https://www.softwaresuggest.com/blog/top-hospital-information-system/.
19. Shahsavarani, A. M., Azad MarzAbadi, E., HakimiKalkhoran, M., Jafari, S., &Qaranli, S. (2015). Clinical decision support systems (CDSSs): state of the art review of literature. International Journal of Medical Reviews, 2(4), 299-308.
20. Sanfilippo, E. M., Schwarz, U., & Schneider, L. (2012, October). The Health Data Ontology Trunk (HDOT). Towards an ontological representation of cancer-related knowledge. In Proceedings of the 2012 5th International Advanced Research Workshop on In Silico Oncology and Cancer Investigation (pp. 1-4). IEEE.
21. 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. – С. 71-80.
22. Gribova, Valeria V., et al. "Software Toolkit for Creating Intelligent Systems in Practical and Educational Medicine." 2018 3rd Russian-Pacific Conference on Computer Technology and Applications (RPC). IEEE, – 2018. – DOI: 10.1109/RPC.2018.8482130.
23. Грибова В. В. и др. Облачная платформа IACPaaS для разработки оболочек интеллектуальных сервисов: состояние и перспективы развития //Программные продукты и системы. – 2018. – Т. 31. – №. 3.
24. http://www.cankaoxiaoxi.com/science/20200206/2401404.shtml Accessed Apr. 03, 2020.

УДК 04.852:07.76

МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ РИСКА РАЗВИТИЯ ФИБРИЛЛЯЦИИ ПРЕДСЕРДИЙ ПОСЛЕ КОРОНАРНОГО ШУНТИРОВАНИЯ*

***Шахгельдян К.И.**, д.т.н.*
e-mail: carinash@vvsu.ru
Институт Информационных технологий ВГУЭС, ДВФУ
Рублев В.Ю.
e-mail: dr.rublev.v@gmail.com
***Гельцер Б.И.**, д.м.н.*
e-mail: boris.geltser@vvsu.ru
Щеглов Б.О.
e-mail: b.shcheglov@mail.ru
ДВФУ
***Щеглова С.Н.**, к.пед.н.*
e-mail: schegloff@mail.ru
СВГУ

1. ВВЕДЕНИЕ

Внедрение технологий искусственного интеллекта (ИИ) в различные области клинической медицины позволяет повысить качество медицинской помощи за счет расширения диагностических возможностей и улучшения прогностических оценок. Так, например, в кардиохирургии с помощью алгоритмов машинного обучения и других методов интеллектуального анализа данных разрабатываются прогностические шкалы оценки индивидуального риска пациента, которые позволяют оптимизировать тактику ведения лечения и минимизировать вероятность послеоперационных осложнений.

К одному из наиболее частых осложнений кардиохирургической практики относится послеоперационная фибрилляция предсердий (ПоФП), которая, в зависимости от типа оперативного

* Работа выполнена при финансовой поддержке РФФИ, проекты №18-29-03131 и №19-29-01077

вмешательства отмечается у 25-40% пациентов [1, 2]. Анализ публикаций по прогнозированию фибрилляции предсердий (ФП) у больных ишемической болезнью сердца (ИБС) с помощью методов машинного обучения показал эффективность применения искусственных нейронных сетей с целью разработки точных прогностических моделей. Так, в исследовании Sidrah Liaqat и соавт. [3] разработали сверточную нейронную сеть с метрикой качества площадь под ROC-кривой (AUC), равной 0,81, LSTM (AUC – 0,83) и сверточную LSTM (AUC – 0,80). Исследователями Hill N.R. и соавт. [4] на основе данных электрокардиографии (ЭКГ) была разработана модель нейронной сети с AUC 0,83. Для прогноза развития ФП были получены и более точные модели. В исследовании Yong Xia и соавт. [5] получена модель сверточной нейронной сети с чувствительностью 98,34%, специфичностью 98,24% и точностью 98,29%. Lown M. и соавт. [6] была создана модель SVM с чувствительностью 99,2% и специфичностью 99,5%.

В кардиохирургии важнейшее значение играет риск развития ПоФП. Ее возникновение после коронарного шунтирования (КШ) в 4 раза повышает риск развития инсульта, послеоперационных кровотечений и острой почечной недостаточности и в 2 раза увеличивает вероятность смертности на 30-дневном и 6-месячном горизонтах наблюдения [1, 7]. Несмотря на большое количество исследований, до настоящего времени не существует единой патофизиологической концепции, подробно описывающей механизмы развития ПоФП и унифицированной шкалы для расчета риска ее возникновения после КШ. Однако, в ряде работ с помощью алгоритмов, применяемых для бинарной классификации (логистической регрессии (ЛР), метода случайного леса (СЛ), деревьев решений и многослойного перцептрона) были выделены ряд предикторов и разработаны прогностические модели [8, 9]. Например, в исследовании Mathew J.P. и соавт. [10] показатель точности модели ЛР по метрике AUC составил 0,7. Выборка включала 1503 пациентов с ИБС после КШ, а в качестве предикторов были определены возраст больных, ФП у пациентов в дооперационном периоде, наличие операций на клапанах сердца в анамнезе, послеоперационная отмена β-адреноблокаторов и ингибиторов ангиотензинпревращающего фермента. В ретроспективном исследовании Güngör и соавт. [11] на данных 125 пациентов была получена модель ЛР, включающая в качестве предикторов возраст пациентов, тромбоцитарно-лимфоцитарный индекс. Точность модели по метрике AUC составила 0,634, а чувствительность и специфичность – 64% и 56%, соответственно. В

исследовании Magee M.J. и соавт. [12] разработанная ЛР-модель включала 14 факторов (AUC 0,72). К наиболее значимым предикторам отнесились дооперационные и интраоперационные факторы: возраст пациентов, потребность в длительной искусственной вентиляции (24 часа и более), использование искусственного кровообращения и наджелудочковые аритмии в анамнезе. Zaman A.G. и соавт. [13] в ходе анализа данных 326 пациентов (среди которых ПоФП наблюдалась у 92) с помощью ЛР выявили следующие предикторы в модели при прогнозировании ФП с чувствительностью и специфичностью в 49% и 84% случаев соответственно: возраст пациентов (>75 лет), увеличение длительности зубца Р (>155мс) и мужской пол. В исследовании David Amaret и соавт. [14] с помощью пошаговой ЛР разработали модель с AUC 0,69, которая включала следующие факторы: сумма отфильтрованных сигналов от 3 ортогональных отведений ЭКГ до операции, пожилой возраст пациентов и мужской пол. В исследовании Todorov и соавт. [15] на выборке из 999 пациентов была разработана многофакторная модель ЛР с AUC 0,68 и выделены следующие до- и послеоперационные предикторы: длительность нахождения на искусственной вентиляции легких и пребывания в отделении интенсивной терапии, пожилой возраст, повышение концентрации С-реактивного белка и креатинина в плазме крови. Для прогнозирования ПоФП после КШ многие исследователи применяли прогностические шкалы EuroSCORE [16], Kolec [17], PAFAC (Predictors of AF After CABG) [18], CHA2DS2-VASc и HAS-BLED, рассчитанные с помощью ЛР, показывающие точность по метрике AUC от 0,65 до 0,68 [19, 20].

Таким образом, данные литературы указывают на высокую точность прогноза ФП у больных ИБС, но недостаточную точность разработанных ранее моделей для прогноза впервые выявленной ПоФП после КШ, что побудило авторов к более детальному анализу данной проблемы. Цель исследования состояла в разработке в совершенствовании алгоритма отбора предикторов в прогнозировании ПоФП у больных ИБС после КШ.

2. МАТЕРИАЛЫ И МЕТОДЫ

Разработка моделей выполнялась на основе данных ретроспективного исследования, содержащих электронные истории болезни (ЭИБ) 886 больных ИБС (181 женщина и 685 мужчин) в возрасте от 35 до 81 года с медианой (МЕ) 63 года и 95% доверительным интервалом (ДИ) [63; 64], которым выполнялось изолированное КШ в условиях искусственного кровообращения в период с 2008 по 2019 гг.

на базе кардиохирургического отделения ГБУЗ «Приморская краевая клиническая больница №1» г. Владивостока. Из исследования были исключены больные с любой формой ФП в анамнезе. Общее количество таких больных составило 85 человек. Таким образом, датасет был представлен результатами обследования 801 больного ИБС. Верификация ПоФП осуществлялась по результатам мониторинга ЭКГ в течение не менее 96 часов после КШ. Среди обследованной когорты было выделено 2 группы лиц. В первую из них вошли 153 (19,1%) больных, у которых в послеоперационном периоде были зарегистрированы пароксизмы ФП, во вторую – 648 (80,9%) больных без нарушений сердечного ритма. В первой группе госпитальная летальность составила 9,8% (15 пациентов), а во второй – 4,6% (30 пациентов).

Конечная точка исследования представляла собой ПоФП в форме категориального бинарного признака («отсутствие» или «развитие»). Входные признаки – подгруппа 100 потенциальных предикторов выражалась в форме непрерывных и категориальных переменных. Для обработки и анализа данных использовали методы статистического анализа и машинного обучения. Первые из них были представлены тестами Хи-квадрат, Фишера, Манна-Уитни и однофакторной ЛР с расчетом весовых коэффициентов. Вторые – многофакторной ЛР, стохастическим градиентным бустингом (СГБ) и искусственной нейронной сетью (ИНС). Архитектура ИНС, представляющей многослойный персептрон, подбиралась путем максимизации площади под ROC-кривой (AUC) и в окончательном виде состояла из двух скрытых слоев по 90 и 80 нейронов в каждом. В качестве функции активации ИНС использовались «sigmoid» и «relu». В окончательные модели с учетом лучшей точности ИНС вошла только функция «sigmoid». Точность моделей оценивали по 4 метрикам качества: AUC, чувствительность (Sen), специфичность (Spec) и точность (Acc). Разработка моделей включала процедуру кросс-валидации методом k-блоков. Модели были построены на обучающей выборке (9/10) пациентов и верифицированы на тестовой (1/10). Все значения метрик качества, приведенные в работе, получены путем усреднения оценок прогноза на тестовых выборках.

Дизайн исследования включал 4 этапа. На первом из них применяли статистический анализ, с помощью которого проводили межгрупповые сравнения потенциальных предикторов ПоФП. Для непрерывных переменных использовали тест Манна-Уитни, так как предварительная оценка близости данных к нормальному распределению методом Шапиро-Уилки показала отрицательный

результат. Для сравнения категориальных переменных использовали критерий хи-квадрат, а для оценки отношения шансов (ОШ) и их ДИ – тест Фишера. На втором этапе с помощью данных методов определяли границы значений анализируемых факторов с наилучшим предиктивным потенциалом. Данная процедура включала проверку гипотез о равенстве распределений признаков в группах сравнения. Выделение прогностически значимых диапазонов осуществляли с шагом тестирования 0,05-0,1 усл. ед. для различных показателей. Критериям отбора соответствовали границы значений факторов, p-value которых имели минимальную, а ОШ – максимальную величину. На третьем этапе исследования по нормированным признакам с помощью однофакторной ЛР определяли весовые коэффициенты, соответствующие значимости влияния отдельных признаков на развитие ПоФП. На четвертом этапе были разработаны многофакторные модели на основе ЛР, СГБ и ИНС, структура которых пошагово дополнялась потенциальными предикторами ПоФП с оценкой метрик качества. При увеличении значений последних считали, что включенный в модель показатель может рассматриваться в качестве предиктора ПоФП. Обработка и анализ данных выполнялись на языке R в среде R-studio и языке Python с помощью пакетов keras, tensorflow и xgboost.

3. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Сравнительный межгрупповой (пациенты с и без ПоФП) анализ факторов, характеризующих дооперационный клинико-функциональное состояние пациентов после КШ, показал, что достоверно значимые межгрупповые различия фиксировались только по 10 показателям: возраст; составляющие части электрокардиограмм – длительность сегмента PQ, комплекса деполяризации желудочков QRS, интервала RR, электрической систолы сердца QT; эхокардиографические параметры сердца – верхне-нижний (LA3) и поперечный (LA2) размеры левого предсердия (ЛП), продольный (RA1) и поперечный (RA2) размеры правого предсердия (ПП), конечный систолический размер (КСР) левого желудочка (ЛЖ); наличие недостаточности трикуспидального клапана (НТК) в анамнезе; фракция выброса (ФВ) крови ЛЖ во время систолы (таблица 1).

Наибольшая статистически значимая разница в группах пациентов с наличием или отсутствием ПоФП наблюдалась у следующих факторов: RA1, RA2 и длительности интервала QT (p-value < 0,0001). Относительно пациентов без аритмии у больных ПоФП показатели ЛП (LA2, LA3) были достоверно значимо выше, а

длительность QRS статистически значимо меньше. Дальнейший анализ показал, что попарное произведение показателей LA2 и LA3, но и RA1 и RA2 имели более высокий уровень достоверной значимости межгрупповых различий, чем оценка влияния отдельных геометрических параметров камер сердца. Менее заметные, но статистически значимые отличия пациентов с послеоперационной аритмией ассоциировались с более старшим возрастом, увеличением КСР ЛЖ, продолжительностью интервала QRS и наличием НТК.

Таблица 1. Клинико-функциональная характеристика больных

Показатели	Объем выборки	1 группа, n=153	2 группа, n= 648	p-value
Возраст, лет	801	64 [63; 66]	63 [62; 64]	0,00076
ФВ, %	783	59 [57;60]	60 [60; 60]	0,039
LA2, мм	734	41 [40; 42]	39 [39; 40]	0,026
LA3, мм	734	38 [37; 39]	37 [36; 37]	0,013
LA2*LA3, мм^2	734	160 [147; 168]	144 [141; 148]	0,011
RA1, мм	734	39,5 [3,9; 40]	37 [36; 37]	0,00007
RA2, мм	734	43 [41; 43]	39 [38; 40]	0,00003
RA1 * RA2, мм^2	734	164 [160; 176]	144 [140;148]	0,000012
PQ, мс	801	160 [150; 160]	150 [140; 150]	0,1
QRS, мс	801	80 [80; 100]	100 [80; 100]	0,0019
RR, мс	761	936,5 [909; 1000]	920 [882,4; 950]	0,22
QT, мс	761	400 [400; 410]	400 [380; 400]	0,00012
КСР ЛЖ, мм	733	350 [330; 360]	340 [330; 350]	0,037
НТК, абс (%)	801	34 (22,2%)	79 (12,2%)	0,002

На втором этапе исследования среди показателей, имеющих достоверные различия в группах сравнения, верифицировали диапазоны их значений, обладающие наибольшим предиктивным потенциалом (таблица 2). Были выделены следующие границы непрерывных факторов, имеющие статистически значимое влияние на развитие ПоФП: возраст мужчин старше 55 и младше 75 лет и женщины старше 60 и младше 79 лет (ОШ = 3,4, p<0,0001); длительность сегмента PQ больше 160 и меньше 200 мс (ОШ=2,2, p=0,0004), комплекса деполяризации желудочков QRS больше 50 и меньше 100 мс (ОШ=1,5, p=0,021), интервал RR в диапазоне от 1000 до 1100 мс, электрической систолы сердца QT больше 420 мс; ЭхоКГ параметры сердца – произведение LA3 на LA2 больше 150 мм^2, произведение RA1 на RA2 больше 160 мм^2, КСР ЛЖ больше 4,9 см с повышением риска в 2,9 раза; ФВ более 45 и менее 60 % увеличивала риск возникновения осложнения в виде мерцательной аритмии в 1,7 раза (p=0,0058). Процедура поиска границ непрерывных переменных позволила выделить не только

значимые диапазоны предикторов и таким образом определить факторы риска, но и выделить дополнительные предикторы, которые ранее не были таковыми верифицированы (PQ и RR).

Таблица 2. Границы значений факторов риска ПоФП с наилучшим предиктивным потенциалом

Предиктор и его граница	1 группа, n=153	2 группа, n= 648	ОШ, 95% ДИ	p-value
Возраст, лет М 55-74 Ж 60-78	142 (92,8 %)	511 (78,9%)	3,4 [1,9; 6,9]	0,0001
ФВ,% 45-60%	81 (52,9%)	268 (41,4%)	1,68 [1,17; 2,42]	0,0058
КСР > 490 мм	7 (4,6%)	11 (1,7%)	2,9 [1,05; 7,7]]	0,049
LA2*LA3 >160 мм^2	67 (43,8%)	191 (29,5%)	2,1 [1,4; 3]	0,0002
RA1 * RA2 >150 мм^2	92 (60%)	269 (41,5%)	2,5 [1,7; 3,8]	<0,0001
PQ 170-210 мс	38 (24,8%)	47 (7,3%)	2,2 [2,6; 6,8]	0,0004
QRS 50-80 мс	88 (57,5%)	303 (46,8%)	1,53 [2,2; 14,4]	0,021
RR 1000-1100 мс	70 (45,8 %)	175 (27%)	2 [1,4; 2,9]	0,00033
QT, мс М>420 мс Ж>440 мс	45 (30,8%)	100 (16,3%)	2,3 [1,5; 3,5]	0,00013
НТК, абс (%)	34 (22,2%)	79 (12,2%)	2 [1,3; 3,2]	0,002

Для верификации возможных взаимосвязей факторов риска с вероятностью развития ПоФП на третьем этапе исследования нами были построены однофакторные модели ЛР с расчетом весовых коэффициентов, характеризующих предсказательную ценность анализируемых показателей (таблица 3). Данный подход позволяет более детально оценить влияние потенциальных предикторов на результирующую переменную. Наибольшее влияние на ПоФП оказывали возраст, QRS и PQ.

Таблица 3. Весовые коэффициенты однофакторных моделей ЛР для оценки риска ПоФП

Показатель	Коэффициент	p-value
Возраст, лет М 55- 74 Ж 60-78	1,24	0,00015
ФВ: 45 – 60%	0,42	0,03
LA2*LA3 >160 мм^2	0,73	0,00016
КСР ЛЖ > 490 мм	1,07	0,03
RA1 * RA2 >150 мм^2	0,94	<0,0001

PQ 170 – 210 мс	1,44	<0,0001
QRS 50-80 мс	1,63	0,0005
RR 1000- 1100 мс	0,87	<0,0001
QT, мс М>420 мс Ж>440 мс	0,78	0,00016
НТК, абс (%)	0,72	0,0016

На четвертом этапе исследования, используя пошаговое включение предикторов в непрерывной или в категориальной форме, были разработаны модели ЛР (таблица 3), СГБ (таблица 4) и ИНС (таблица 5). Одной из задач нашего исследования являлась оценка эффективности использования предикторов в категориальной форме по сравнению с их непрерывными аналогами. Повышение точности многофакторных моделей на основе ЛР фиксировалось в моделях 2-7 (таблица 4). Максимальная точность по метрике Sen (0,67) была достигнута в моделях 6 и 7 при AUC 0,71 и 0,7 соответственно. Уровень Spec составил 0,66. Дальнейшее расширение возможных предикторов снижало метрику Sen, сохраняя уровни других метрик. Данный факт свидетельствовал о том, что такие параметры как КСР, ФВ и НТК связаны, возможно линейно, с показателями размеров предсердий, данных ЭКГ или возраста. Анализ моделей на основе ЛР, использующих категориальные переменные (модели 11-20 в таблице 4), показал, что точность моделей повышалось аналогично непрерывному случаю, но она могла достигать лучших значений. Например, сравнение моделей 7 и 17, показывает, что последняя имеет лучшую точность по всем 4 метрикам качества (0,69 vs 0,67, 0,71 vs 0,66, 0,74 vs 0,7 и 0,71 vs 0,67). Дальнейшее расширение категориальных предикторов моделей 18-20 существенно не меняло точность прогноза.

Таблица 4. Оценка точности прогностических моделей ПоФП на тестовых выборках для многофакторной ЛР

N	Предикторы	Граница предиктора	Метрики точности			
			Sen	Spec	AUC	ACC
1	Возраст	-	0,46	0,63	0,58	0,59
2	Возраст, PQ	-	0,51	0,62	0,59	0,6
3	Возраст, PQ, QRS	-	0,6	0,6	0,62	0,6
4	Возраст, PQ, QRS, RR	-	0,56	0,61	0,63	0,6
5	Возраст, PQ, QRS, RR, QT	-	0,62	0,61	0,67	0,61
6	Возраст, PQ, QRS, RR, QT, RA1*RA2	-	0,67	0,66	0,71	0,66
7	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3	-	0,67	0,66	0,7	0,67

8	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР	-	0,64	0,68	0,71	0,66
9	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР, ФВ	-	0,62	0,68	0,7	0,66
10	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР, ФВ, НТК	-	0,58	0,7	0,7	0,68
11	Возраст	М 55- 74 Ж 60-78	0,93	0,21	0,57	0,35
12	Возраст + PQ	170-210 мс	0,25	0,93	0,64	0,8
13	Возраст, PQ + QRS	50-80 мс	0,72	0,46	0,66	0,5
14	Возраст, PQ, QRS + RR	1000-1100 мс	0,56	0,75	0,7	0,71
15	Возраст, PQ, QRS, RR + QT	М>420 мс Ж>440 мс	0,67	0,69	0,72	0,69
16	Возраст, PQ, QRS, RR, QT + RA1*RA2	> 160 мм2	0,63	0,74	0,75	0,72
17	Возраст, PQ, QRS, RR, QT, RA1*RA2 + LA1*LA2	> 150 мм2	0,69	0,71	0,74	0,71
18	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3 + КСР	>490 мм	0,68	0,71	0,75	0,71
19	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР + ФВ	45-60%	0,68	0,72	0,75	0,71
20	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР, ФВ, НТК	-	0,69	0,71	0,75	0,71

Модели на основе СГБ разрабатывали аналогичным образом (таблица 5). Увеличение точности прогностической оценки фиксировалось во 2-5 моделях СГБ. Полученные в 5-ой модели 4 метрики качества являются лучшими на текущий момент прогностическими оценками по сравнению со всеми ранее опубликованными результатами: 0,82, 0,76, 0,77 и 0,76 для AUC, ACC, Sen и Spec соответственно. Дальнейшее включение в модели на основе СГБ статистически значимых предикторов в непрерывной форме снижало качество прогноза (модели 6-10). Модели СГБ 11-20, разработанные с использованием предикторов в категориальной форме, достигали пика точности при том же наборе показателей что и модели ЛР (модели 17 таблицы 4 и 5). Дальнейшее расширение спектра

категориальных предикторов не приводило к изменению качества прогноза. Сопоставимыми по уровню точности были модели на основе ЛР и СГБ.

Таблица 5. Оценка точности прогностических моделей ПоФП на тестовых выборках для СГБ

N	Предикторы	Граница предиктора	Метрики точности			
			Sen	Spec	AUC	ACC
1	Возраст	-	*0,5*	*0,59*	*0,54*	*0,57*
2	Возраст, PQ	-	*0,58*	*0,66*	*0,65*	*0,64*
3	Возраст, PQ, QRS	-	*0,66*	*0,63*	*0,71*	*0,64*
4	Возраст, PQ, QRS, RR	-	*0,76*	*0,76*	*0,8*	*0,76*
5	Возраст, PQ, QRS, RR, QT	-	*0,77*	*0,76*	*0,82*	*0,76*
6	Возраст, PQ, QRS, RR, QT, RA1*RA2	-	*0,72*	*0,76*	*0,8*	*0,75*
7	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3	-	*0,68*	*0,78*	*0,79*	*0,76*
8	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3,КСР	-	*0,69*	*0,75*	*0,78*	*0,74*
9	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР, ФВ	-	*0,68*	*0,76*	*0,78*	*0,75*
10	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР, ФВ, НТК	-	*0,68*	*0,76*	*0,78*	*0,75*
11	Возраст	*М 55-74 Ж 60-78*	*0,93*	*0,21*	*0,57*	*0,35*
12	Возраст + PQ	*170-210 мс*	*0,25*	*0,93*	*0,64*	*0,8*
13	Возраст, PQ + QRS	*50-80 мс*	*0,72*	*0,46*	*0,66*	*0,5*
14	Возраст, PQ, QRS + RR	*1000-1100 мс*	*0,56*	*0,75*	*0,7*	*0,71*
15	Возраст, PQ, QRS, RR + QT	*М>420 мс Ж>440 мс*	*0,67*	*0,69*	*0,72*	*0,69*
16	Возраст, PQ, QRS, RR, QT + RA1*RA2	*> 160 мм2*	*0,63*	*0,74*	*0,75*	*0,72*
17	Возраст, PQ, QRS, RR, QT, RA1*RA2 + LA1*LA2	*> 150 мм2*	*0,69*	*0,71*	*0,74*	*0,71*
18	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3 + КСР	*>490 мм*	*0,68*	*0,71*	*0,75*	*0,71*
19	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР + ФВ	*45-60%*	*0,68*	*0,72*	*0,75*	*0,71*

20	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР, ФВ + НТК	-	*0,69*	*0,71*	*0,75*	*0,71*

Для моделей на основе ИНС фиксировалось увеличение точности при расширении спектра используемых предикторов как в непрерывной, так и в категориальной форме (таблица 6). Лучшая модель ИНС получена при включении всех значимых предикторов в категориальной форме (Sen – 0,74, Spec – 0,73, AUC – 0,75, ACC – 0,73). Точность последней уступает точности лучшей модели на основе СГБ, но превышает точность модели ЛР, при этом использование ИНС позволяет подтвердить значимость влияние рассматриваемых предикторов.

Таблица 6. Оценка точности прогностических моделей ПоФП на тестовых выборках для ИНС

N	Предикторы	Граница предиктора	Метрики точности			
			Sen	Spec	AUC	ACC
1	Возраст	-	*0,16*	*0,86*	*0,56*	*0,7*
2	Возраст, PQ	-	*0,2*	*0,9*	*0,59*	*0,73*
3	Возраст, PQ, QRS	-	*0,5*	*0,77*	*0,71*	*0,6*
4	Возраст, PQ, QRS, RR	-	*0,6*	*0,68*	*0,63*	*0,64*
5	Возраст, PQ, QRS, RR, QT	-	*0,63*	*0,68*	*0,68*	*0,65*
6	Возраст, PQ, QRS, RR, QT, RA1*RA2	-	*0,65*	*0,68*	*0,71*	*0,67*
7	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3	-	*0,68*	*0,7*	*0,69*	*0,69*
8	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР	-	*0,69*	*0,72*	*0,71*	*0,7*
9	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР, ФВ	-	*0,7*	*0,71*	*0,71*	*0,71*
10	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР, ФВ, НТК	-	*0,71*	*0,72*	*0,72*	*0,71*
11	Возраст	*М 55-74 Ж 60-78*	*0*	*1,0*	*0,55*	*0,81*
12	Возраст + PQ	*170-210 мс*	*0,24*	*0,94*	*0,62*	*0,8*
13	Возраст, PQ + QRS	*50-80 мс*	*0,38*	*0,8*	*0,67*	*0,73*
14	Возраст, PQ, QRS + RR	*1000-1100 мс*	*0,58*	*0,74*	*0,71*	*0,71*
15	Возраст, PQ, QRS, RR + QT	*М>420 мс Ж>440 мс*	*0,68*	*0,68*	*0,72*	*0,68*

16	Возраст, PQ, QRS, RR, QT + RA1*RA2	*> 160 мм2*	*0,63*	*0,75*	*0,75*	*0,73*
17	Возраст, PQ, QRS, RR, QT, RA1*RA2 + LA1*LA2	*> 150 мм2*	*0,65*	*0,74*	*0,75*	*0,72*
18	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР	*>490 мм*	*0,65*	*0,74*	*0,74*	*0,73*
19	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР + ФВ	*45-60%*	*0,7*	*0,71*	*0,74*	*0,71*
20	Возраст, PQ, QRS, RR, QT, RA1*RA2, LA2*LA3, КСР, ФВ + НТК	-	*0,74*	*0,73*	*0,75*	*0,73*

4. ЗАКЛЮЧЕНИЕ

В данной работе представлены модели прогнозирования развития ПоФП после КШ, разработанные на основе методов статистического анализа и машинного обучения: межгруппового сравнения, поиска значимых диапазонов, однофакторной ЛР, многофакторных ЛР, СГБ и ИНС. Поиск предикторов выполнялся из набора 100 доперационных характеристик пациентов с ИБС. Использование статистических методов позволило выделить наиболее значимые факторы. При этом удалось выдвинуть гипотезы о границах диапазонов, формирующих факторы риска. К факторам с наибольшим предиктивным потенциалом относились возраст больных (55-74 года для мужчин, 60-78 лет – для женщин), RR 1000-1100 мс, QRS 50-80 мс, PQ 170-210 мс, QT (>420 мс для мужчин; >440 мс – для женщин), произведение линейных размеров ЛП > 160 мм2 и ПП > 150 мм2, наличие НТК, ФВ 45-60%, КСР ЛЖ > 490 мм. Эти гипотезы были подтверждены однофакторными на основе ЛР и многофакторными ЛР и ИНС моделями. Часть предикторов (ФВ, НТК, КСР) имела нелинейные взаимосвязи с конечной точкой и не повышала точность моделей ЛР. Подтверждением влияния этих факторов на ПоФП являлось пошаговое увеличение метрик качества моделей ИНС. Использование показателей в категориальной форме не вело к снижению точности моделей во всех рассматриваемых случаях.

В тоже время, наилучшую модель авторы получили при использовании СГБ с непрерывными наиболее значимыми факторами. Данная модель для прогнозирования впервые выявленной ПоФП после КШ по 4 метрикам качества (AUC – 0,82 , ACC – 0,76, Sen -0,77 и Spec – 0,76) является наилучшей по сравнению с ранее опубликованными. Расширение предикторов моделей СГБ снижало

точность прогноза. Использование показателей в категориальной форме для СГБ также не улучшило качество моделей по сравнению с непрерывными аналогами.

Полученные в ходе исследования предикторы объясняются современными представлениями о патофизиологии ИБС и ФП. Так, изначально (до операции) измененный размер предсердия, нарушенная сократительная функция ЛЖ и отклонения показателей ЭКГ усугубляется на фоне операционного стресса представляются субстратом, способствующим возникновению патологической аритмии. Неоспоримым фактом является сложность клинической интерпретация СГБ и ИНС, но в рассматриваемом эксперименте, некоторые, применяемые для моделирования предикторы, являются в основе своей широко распространенными маркерами функционального состояния больных и их предиктивное значение показано в ряде работ [21]. Однако, были выявлены и ранее не используемые предикторы, такие как, конечный систолический размер, недостаточность трикуспидального клапана, размеры ПП, что является задачей дальнейшего исследования.

Литература

1. Kirchhof, P. Разработано при особом участии Европейской ассоциации ритма сердца (EHRA). Одобрено Европейской организацией по изучению инсульта (ESO). Рекомендации ESC по лечению пациентов с фибрилляцией предсердий, разработанные совместно с EACTS / P. Kirchhof, S. Benussi, D. Kotecha, A. Ahlsson. Научное редактирование перевода выполнено Лебедевым Д. С., заведующим НИО аритмологии НМИЦ им. В.А. Алмазова, профессором РАН. Российский кардиологический журнал. – 2017. – 7 (147). – С.7-86 http://dx.doi.org/10.15829/1560-4071-2017-7-7-86.
2. Российское кардиологическое общество Национальное общество по изучению атеросклероза Национальное общество по атеротромбозу / "Клинические рекомендации "Стабильная ишемическая болезнь сердца" (утв. Минздравом России) // [Электронный ресурс] / Режим доступа: https://legalacts.ru/doc/klinicheskie-rekomendatsii-stabilnaja-ishemicheskaja-bolezn-serdtsa-utv-minzdravom-rossii/.
3. Liaqat S, Dashtipour K, Zahid A, Assaleh K, Arshad K, Ramzan N. Detection of Atrial Fibrillation Using a Machine Learning Approach. Information. 2020; 11(12):549. https://doi.org/10.3390/info11120549.
4. Hill NR, Ayoubkhani D, McEwan P, Sugrue DM, Farooqui U, Lister S, et al. (2019) Predicting atrial fibrillation in primary care using machine learning. PLoS ONE 14(11): e0224582. https://doi.org/10.1371/journal.pone.0224582.
5. Yong Xia, Naren Wulan, Kuanquan Wang, Henggui Zhang, Detecting atrial fibrillation by deep convolutional neural networks, Computers in Biology and Medicine, Volume 93, 2018, Pages 84-92, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2017.12.007.

6. Lown M, Brown M, Brown C, Yue AM, Shah BN, Corbett SJ, et al. (2020) Machine learning detection of Atrial Fibrillation using wearable technology. PLoS ONE 15(1): e0227401. https://doi.org/10.1371/journal.pone.0227401.
7. Skaria, R., Parvaneh, S., Zhou, S., Kim, J., Wanjiru, S., Devers, G., Konhilas, J., & Khalpey, Z. (2020). Path to precision: prevention of post-operative atrial fibrillation. Journal of thoracic disease, 12(5), pp. 2735-2746. https://doi.org/10.21037/jtd-19-3875.
8. Christalin, Beulah & Jeeva, S.. (2019). Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. Informatics in Medicine Unlocked. 16. 100203. 10.1016/j.imu.2019.100203.
9. Шахгельдян К.И., Рублев В.Ю., Гельцер Б.И., Щеглов Б.О. и др. Оценка предиктивного потенциала дооперационных факторов риска фибрилляции предсердий у больных ишемической болезнью сердца после коронарного шунтирования. Сибирский журнал клинической и экспериментальной медицины. – 2020. – 35(4). – С.128-136. https://doi.org/10.29001/2073-8552-2020-35-4-128-136.
10. Mathew JP, Fontes ML, Tudor IC, et al. A Multicenter Risk Index for Atrial Fibrillation After Cardiac Surgery. JAMA. 2004;291(14): pp. 1720-1729. doi:10.1001/jama.291.14.1720.
11. Güngör, Hasan & Babu, Abraham & Zencir, Cemil & Akpek, Mahmut & Selvi, Mithat & Erkan, Muhammet & Durmaz, Selim. (2017). Association of Preoperative Platelet-to-Lymphocyte Ratio with Atrial Fibrillation after Coronary Artery Bypass Graft Surgery. Medical Principles and Practice. 26. Pp. 164-168. 10.1159/000453614.
12. Magee, M. J., Herbert, M. A., Dewey, T. M., Edgerton, J. R., Ryan, W. H., Prince, S., & Mack, M. J. (2007). Atrial fibrillation after coronary artery bypass grafting surgery: development of a predictive risk algorithm. The Annals of thoracic surgery, 83(5), 1707–1712. https://doi.org/10.1016/j.athoracsur.2006.12.032.
13. Zaman AG, Archbold RA, Helft G, Paul EA, Curzen NP, Mills PG. Atrial fibrillation after coronary artery bypass surgery: a model for preoperative risk stratification. Circulation. 2000;101(12): pp. 1403-1408. doi:10.1161/01.cir.101.12.1403.
14. David Amar, Weiji Shi, Charles W. Hogue, Hao Zhang, Rod S. Passman, Betsy Thomas, Peter B. Bach, Ralph Damiano, Howard T. Thaler, Clinical prediction rule for atrial fibrillation after coronary artery bypass grafting, Journal of the American College of Cardiology, Volume 44, Issue 6, 2004, Pp. 1248-1253, ISSN 0735-1097, https://doi.org/10.1016/j.jacc.2004.05.078.
15. Todorov, Hristo & Janssen, Inka & Honndorf, Stefanie & Bause, Daniela & Gottschalk, Antje & Baasner, Silke & Volkert, Thomas & Färber, Valentin & Stover, John & Westphal, Martin & Ellger, Björn. (2017). Clinical significance and risk factors for new onset and recurring atrial fibrillation following cardiac surgery – a retrospective data analysis. BMC Anesthesiology. 17. 10.1186/s12871-017-0455-7.
16. Ad N., Holmes S.D., Patel J., Pritchard G., Shuman D.J., Halpin L. Comparison of EuroSCORE II, Original EuroSCORE, and The Society of Thoracic Surgeons

Risk Score in Cardiac Surgery Patients. Ann Thorac Surg. 2016;102(2): pp. 573-579. doi:10.1016/j.athoracsur.2016.01.105.
17. Kolek M.J., Muehlschlegel J.D., Bush W.S. et al. Genetic and clinical risk prediction model for postoperative atrial fibrillation. Circ Arrhythm Electrophysiol. 2015;8(1):25-31. doi:10.1161/CIRCEP.114.002300.
18. Lin S.Z., Crawford T.C., Suarez-Pierre A. et al. A Novel Risk Score to Predict New Onset Atrial Fibrillation in Patients Undergoing Isolated Coronary Artery Bypass Grafting. Heart Surg Forum. 2018;21(6):E489-E496. Published 2018 Dec 13. doi:10.1532/hsf.2151.
19. Mariscalco G., Biancari F., Zanobini M., et al. Bedside tool for predicting the risk of postoperative atrial fibrillation after cardiac surgery: the POAF score. J Am Heart Assoc. 2014;3(2):e000752. Published 2014 Mar 24. doi:10.1161/JAHA.113.000752.
20. Burgos L.M., Seoane L., Parodi J.B. et al. Postoperative atrial fibrillation is associated with higher scores on predictive indices. J Thorac Cardiovasc Surg. 2019;157(6):2279-2286. doi:10.1016/j.jtcvs.2018.10.091.
21. Ревишвили А.Ш., Попов В.А., Коростелев А.Н., Плотников Г.П., Малышенко Е.С., Анищенко М.М. Предикторы развития фибрилляции предсердий после операции аортокоронарного шунтирования. Вестник аритмологии. 2018;(94):11-16; Revishvili A.S., Popov V.A., Korostelev A.N., Plotnikov G.P., Malyshenko E.S., Anishchenko M.M. PREDICTORS OF NEW ONSET OF ATRIAL FIBRILLATION AFTER CORONARY ARTERY BYPASS GRAFTING SURGERY. Journal of Arrhythmology. 2018;(94): pp. 11-16.(In Russ.) DOI:10.25760/VA-2018-94-11-16.

УДК 004.8(075.8)

ОСОБЕННОСТИ РЕАЛИЗАЦИИ ПРОТОТИПА ИНТЕГРИРОВАННОЙ ЭКСПЕРТНОЙ СИСТЕМЫ ДЛЯ КОМПЛЕКСНОЙ ДИАГНОСТИКИ ЗАБОЛЕВАНИЙ МОЛОЧНОЙ ЖЕЛЕЗЫ: ИНТЕГРАЦИЯ С ОБРАБОТКОЙ ИЗОБРАЖЕНИЙ*

***Рыбина Г.В.**, д.т.н.*
e-mail: gvrybina@mephi.ru
Григорьев А.А.
НИЯУ МИФИ

1. ВВЕДЕНИЕ

Интерес к созданию прикладных интеллектуальных систем, предназначенных для такой социально-значимой проблемной области (ПрО), как медицина, неизменно возрастает и постоянно актуализируется в соответствии с появлением новых медицинских технологий и совершенствованием используемых программно-аппаратных средств и платформ, что существенно расширяет арсенал применяемых моделей, методов и средств искусственного интеллекта (ИИ), в частности, для задач медицинской диагностики.

Следует отметить, что традиционный взгляд на возможности эффективного использования методов и технологий ИИ связан, как правило, с описанием целей и задач, стоящих перед разнообразными системами, основанными на знаниях, в частности, экспертными системами (ЭС). Основной акцент здесь делается на поддержку принятия решений для повышения качества и точности работы врачей по выявлению наиболее сложных и труднодиагностируемых заболеваний, обнаружению признаков заболеваний на ранних стадиях и назначению лечебных мероприятий, для обучения медицинского персонала и т.д. [1-6]. Важным требованием в медицинских ЭС всегда является как наличие эксплицитного описания используемых знаний, так и реализация удобных для врачей режимов консультации с ЭС с объяснением результатов диагностики.

В последние годы стал активно развиваться онтологический подход, в рамках которого создаются различные прикладные

* Работа выполнена при финансовой поддержке РФФИ, проект №18-01-00457

онтологии, например, широко известна онтология «Медицинская диагностика» [7] и др, а также обсуждаются вопросы построения некоторой универсальной онтологии знаний о диагностических процессах в различных ПрО, что в целом может быть полезно для разработчиков интеллектуальных систем [8].

Однако, существенный «прорыв» в области медицинских технологий, ориентированных на использование широкого спектра новых типов оборудования для диагностики заболеваний, показал необходимость интеграции методов и средств традиционных ЭС с системами обработки изображений (ультразвуковых, рентгеновских и др.), системами лучевой и компьютерной томографии, обработкой больших данных (Data Mining, Text Mining) и др. в процессе комплексной диагностики.

Исследования в области интеграционных процессов, возникающих при совместном использовании семантически разнородных объектов, компонентов, инструментов и т.п., а также наличие целостной методологии и технологии проектирования и разработки интегрированных интеллектуальных систем, особенно важны для преодоления автономности программно-аппаратных технологий и отсутствия их влияния на процессы принятия решений. Конечно, на данном этапе речь идет не о преодолении аппаратной зависимости от типов электроники и оборудования, а только о развитии комплексных подходов к процессам диагностирования заболеваний, как за счет инженерии знаний, так и путем уменьшения влияния человеческого фактора при восприятии и интерпретации визуальной информации на снимках.

В настоящее время наиболее развитыми в контексте проблемы интеграции и хорошо апробированными на практике для решения задач реальной сложности и значимости являются архитектуры интегрированных ЭС (ИЭС) с гибкой расширяемой функциональностью и масштабируемой архитектурой, разработанные на основе задачно-ориентированной методологии [9] и интеллектуальной программной среды комплекса АТ- ТЕХНОЛОГИЯ [9-11].

Целью данной работы является расширение онтологии типовых архитектур ИЭС за счет разработки моделей и методов интеграции базовых компонентов ИЭС с новыми функциональными компонентами, реализующими некоторые методы обработки результатов ультразвуковых и рентгеновских исследований. В качестве экспериментальной ПрО для реализации процессов прототипирования ИЭС рассматривалась комплексная диагностика заболеваний молочной

железы, включающая клиническую, рентгенологическую (маммографическую) и ультразвуковую диагностику.

2. ОСОБЕННОСТИ ТЕХНОЛОГИИ РАЗРАБОТКИ ПРИКЛАДНЫХ ИЭС НА ОСНОВЕ СРЕДСТВ БАЗОВОЙ ВЕРСИИ КОМПЛЕКСА АТ-ТЕХНОЛОГИЯ

Методы и способы реализации интеграционных процессов в ИЭС определяются базовыми принципами задачно-ориентированной методологии [9], в основе которой лежит многоуровневая модель интеграции, моделирование конкретных типов задач, релевантных технологии традиционных ЭС, методы и способы построения программной архитектуры ИЭС и ее компонентов на каждом уровне интеграции с комплексным учетом специфики решаемых задач для данной ПрО.

Современный опыт применения задачно-ориентированной методологии и различных версий постоянно развивающегося инструментария для автоматизированной поддержки этой методологии – интеллектуальной программной среды комплекса АТ-ТЕХНОЛОГИЯ [9-11] уже насчитывает несколько десятков прикладных ИЭС различной архитектурной типологии, разработанных для задач диагностики, проектирования, планирования, управления и обучения, описание которых приводится в монографиях [9,10] и различных публикациях последних лет.

Соответственно, сложилась и постоянно усовершенствуется технология прототипирования прикладных ИЭС с использованием базовых функциональных средств комплекса АТ-ТЕХНОЛОГИЯ, для которых в отдельных случаях проводится рефакторинг, а также активно развиваются и экспериментально исследуются новые и модифицируемые компоненты интеллектуальной программной среды [10,11] такие, как интеллектуальный планировщик, онтология типовых архитектур ИЭС, повторно используемые компоненты, типовые проектные процедуры и др.

С учетом контекста данной работы рассмотрим некоторые особенности процесса прототипирования прикладных ИЭС, который традиционно осуществляется в два этапа: разработка комплекса моделей, предусмотренных задачно-ориентированной методологией (например, для статических ПрО – это модель архитектуры ИЭС, модель ПрО, модель диалога, модель объяснений); проектирование и программная реализация прототипа ИЭС на основе использования средств базовой версии комплекса АТ-ТЕХНОЛОГИЯ.

На этапе анализа системных требований строится модель архитектуры проектируемого прототипа ИЭС, исходя из идей глубинной интеграции компонентов [9], что связано с возможным расширением функциональности традиционных ЭС путем включения функций, реализуемых другими компонентами. Средний уровень интеграции в модели архитектуры каждой конкретной ИЭС определяется спецификациями наборов функций как простой ЭС (неформализованная операция), так и функций, не свойственных ЭС (формализованная операция), что отражает состав и структуру всех компонентов ИЭС и их информационные и управляющие связи. Исходя из этого модель архитектуры представляет собой иерархию расширенных диаграмм потоков данных (РДПД).

Следующая важная задача связана с построением модели ПрО. Данный процесс подразумевает автоматизированное приобретение знаний с помощью базовых средств поддержки комбинированного метода приобретения знаний (КМПЗ) из различных источников знаний [9,10], в результате чего поэтапно осуществляется построение промежуточного представления (поля знаний), а затем после верификации и объединения фрагментов поля знаний (в случае наличия нескольких источников знаний) производится их конвертация в базу знаний (БЗ) на языке представления знаний, предусмотренном задачно-ориентированной методологией.

С точки зрения анализа особенностей этапов проектирования и программной реализации отдельных компонентов прототипа ИЭС, здесь следует выделить два основных аспекта технологии прототипирования прикладных ИЭС на основе средств комплекса АТ-ТЕХНОЛОГИЯ. С одной стороны, в зависимости от модели архитектуры ИЭС осуществляется конфигурирование программного обеспечения ИЭС на основе выбора для использования некоторых базовых компонентов (АТ-РЕШАТЕЛЬ, объяснительный компонент и др.), а с другой – производится расширение прототипа ИЭС за счет разработанных специализированных компонентов, например, компонента интеграции с внешними программными средствами, отражающими специфику конкретной ПрО.

В этом случае, функционирование подсистемы общения в разработанной ИЭС будет являться результатом интерпретации построенной модели диалога с пользователем (режим консультации), представленной на специальном языке описания сценариев диалога [9,10], предусматривающем описание и отправку сообщений другим компонентам ИЭС (внутренним и внешним).

Теперь непосредственно рассмотрим результаты экспериментального программного моделирования в контексте реализации интеграционных процессов при совместном использовании технологии построения ИЭС и методов обработки и интерпретации изображений, полученных в результате рентгенографических (маммографических) и ультразвуковых медицинских исследований (УЗИ) заболеваний молочной железы (МЖ).

3. ОБЩАЯ ХАРАКТЕРИСТИКА ПРОБЛЕМНОЙ ОБЛАСТИ «КОМПЛЕКСНАЯ ДИАГНОСТИКА ЗАБОЛЕВАНИЙ МОЛОЧНОЙ ЖЕЛЕЗЫ» И ЗАДАЧ, СВЯЗАННЫХ С ОБРАБОТКОЙ ИЗОБРАЖЕНИЙ

Системный анализ (обследование) данной ПрО при создании демонстрационного прототипа ИЭС проводился с участием экспертов – Ковынева А.В. (врач-эксперт по УЗИ-аппаратуре, научный сотрудник на кафедре ультразвуковой диагностики и хирургии РУДН) и Мазо М.Л. (к.м.н., старший научный сотрудник МНИОИ им. П.А. Герцена, филиал ФГБУ «НМИРЦ радиологии» Минздрава России).

Для первого этапа прототипирования ИЭС по согласованию с экспертами и с учетом [12, 13] был выделен круг неформализованных задач, в которых логика принятия решений на основе БЗ (модели ПрО) включает использование результатов, полученных при обработке изображений на различных снимках. Ниже приводятся их краткие постановки.

Задача клинико-рентгенографической диагностики МЖ (маммография)

Дано: Набор маммограмм МЖ в различных проекциях. Клинические симптомы МЖ.

Требуется: Оценить общее состояние МЖ, обнаружить и при наличии первично диагностировать образования.

Задача ультразвуковой диагностики МЖ в В-Режиме (сонография)

Дано: УЗИ-изображение МЖ (сонограмма). Результаты маммографии.

Требуется: Повторно оценить общее состояние МЖ, обнаружить и при наличии повторно (зависимо и независимо от маммографии) диагностировать образования.

Задача ультразвуковой диагностики МЖ в доплер-режиме (доплерсонография (ЦДК))

Дано: УЗИ-изображение МЖ с цветным отображением мест кровотока и его уровня. Результаты маммографии и сонографии.

Требуется: При наличии образования дополнительно оценить риск злокачественности.

Кроме того, была выделена формализованная задача интерпретации начальных признаков МЖ по изображению:

Дано: УЗИ или РГ-изображение МЖ, полученное в результате обследования пациента.

Требуется: При наличии образования, выявить его признаки для диагностики, а также выявить признаки для диагностики общего состояния МЖ и диффузных изменений.

Общая схема процесса решения задачи комплексной диагностики МЖ представлена на рис.1.

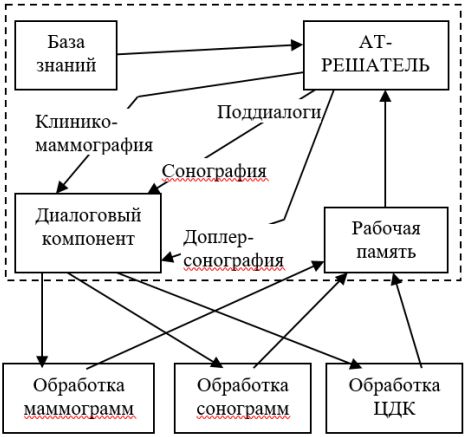


Рис.1. Схема диагностики

Рассмотрим кратко особенности выбранных и реализованных методов обработки, интерпретации и описания изображений для выявления диагностических признаков.

В настоящее время существует достаточно обширное множество методов обработки изображений, однако отметим при этом, что медицинские изображения МЖ – маммограммы и сонограммы, являются монохромными изображениями, поэтому их обработка сводится к обработке уровня яркости. Основными целями, решаемыми врачом в процессе анализа маммограмм и эхограмм, являются: описание общего состояния железистой и жировой тканей, поиск и определение характера известковых включений – диффузный анализ маммограмм и сонограмм в целом; локализация образований и выявление их диагностических признаков – локальный анализ участков маммограмм и сонограмм.

Для описания состояния железистой и жировой тканей ставятся задачи определения степени выраженности и соотношения этих тканей на маммограмме и эхограмме. С точки зрения цифровой обработки изображений для решения данных задач необходимо сначала снизить видимые эффекты различных шумов и дефектов изображений, а после – классифицировать состояние ткани. Здесь, как показано в [15], для улучшения визуализации изображения можно использовать следующие методы коррекции яркости: логарифмическое преобразование; гамма-коррекция (степенное преобразование); кусочно-линейные преобразования; преобразования гистограммы изображения; различные методы пространственной фильтрации изображений.

Классификация состояния тканей МЖ проводится по показателям различных текстурных характеристик изображения, среди которых наиболее универсальными, в соответствии с [15], являются: матрица смежности уровней серого тона; энергия (второй угловой момент); контрастность (инерция); максимальная вероятность; локальная однородность (гомогенность); энтропия (неоднородность текстуры); след матрицы смежности уровней серого тона; математическое ожидание серого тона; среднеквадратичное отклонение серого тона; гистограмма яркости изображения.

Далее необходимо решать задачи, связанные с локализацией образования и выявлением его диагностических, в основном, геометрических и текстурных признаков [12-14]. Процесс локализации на данной стадии удобнее всего возложить на пользователя-врача, предоставив необходимые для этого средства по обрезке изображения, поскольку врач обладает более четкой информацией о расположении образования (при его наличии) за счет клинического обследования МЖ. Тогда для интерпретации диагностических признаков образования необходимо сначала осуществить выделение контуров образования, а затем произвести описание и классификацию выделенного контура, для чего применяются различные градиентные методы [15] (например, алгоритм Марра–Хилдрета, алгоритм Кэнни, оператор Прюитт). При решении задачи описания и интерпретации найденных контуров чаще всего используются такие характеристики как кривизна границы, а также ее статистические характеристики.

Следует отметить, что для выбора наиболее пригодных из всех описанных методов обработки изображения были проведены серии экспериментов на языке Python с использованием библиотек Numpy, Python Image Library (PIL), Open-CV в специализированной среде Jupyter Notebook, на основе чего сделаны следующие выводы:

- наиболее пригодным для данной проблематики методом по коррекции яркости изображения является гамма-коррекция;
- все описанные текстурные и контурные характеристики информативны, поскольку имеют достаточно ощутимый отклик;
- метод выделения контура с помощью оператора Прюитт выбран как наиболее эффективный для данной задачи;
- для устранения шума при контурной обработке необходимо предварительно локализовать узловое образование (при его наличии);
- для обеспечения большей точности выделения контура необходимо производить пороговую бинаризацию изображения, причем порог бинаризации подбирается индивидуально для каждого изображения.

В качестве инструмента, выполняющего построение классификатора изображений, использовался специализированный программный пакет See 5. Выбранные методы и подходы к обработке изображений послужили основой для разработки и программной реализации алгоритмов выявления диагностических признаков поражений МЖ по маммограммам и эхограммам, а именно: алгоритм формирования классификаторов изображений, алгоритм интерпретации медицинских изображений МЖ с использованием классификаторов и др.

4. ПРИМЕР ФУНКЦИОНИРОВАНИЯ ДЕМОНСТРАЦИОННОГО ПРОТОТИПА ИЭС ДЛЯ КОМПЛЕКСНОЙ ДИАГНОСТИКИ ЗАБОЛЕВАНИЙ МОЛОЧНОЙ ЖЕЛЕЗЫ

В настоящее время текущий объем БЗ для данной ПрО составляет около 120 правил, среди которых в том числе имеются правила, содержащие различные типы НЕ-факторов знаний, такие как неопределенность, неточность и нечеткость [9, 10].

Значительно расширен функционал уточняющих поддиалогов АТ-РЕШАТЕЛЯ с целью поддержки обращений не только к пользователю, но и для организации взаимодействия компонентов ИЭС с внешними программами с помощью компонента интеграции Scripter, обеспечивающего вызов функций и процедур на языках JavaScript и VBScript. Для этого было разработано и включено в состав архитектуры специализированное программное обеспечение (СПО) для обработки медицинских изображений молочной железы, взаимодействие с которым обеспечивает интеграционный компонент.

Рассмотрим кратко пример функционирования прототипа в режиме консультации.

После запуска режима консультации с ИЭС на экране отобразится диалоговое окно, в котором необходимо зарегистрировать пациента и указать такие характеристики, как: наследственные заболевания; наличие/отсутствие перенесенных онкологических заболеваний, перенесенных травм, аллергических заболеваний; количество беременностей, родов и абортов; наличие/отсутствие непериодических лактаций.

Далее осуществляется переход к клиническому обследованию пациента, где выясняются такие симптомы, как болезненность, наличие пальпируемых образований и т.д. Клиническое обследование пациента проходит в несколько этапов в зависимости от введенных на каждом шагу симптомов, при этом врач имеет возможность вводить симптомы с учетом своей степени уверенности.

При выявлении необходимости более глубокой диагностики на основе клинических обследований, начинается этап консультации, связанный с обработкой медицинских изображений МЖ (маммограмм и сонограмм (эхограмм)). Рассмотрим этот этап на примере обработки маммограммы (рентгенографического изображения МЖ).

В данном случае врачу предоставляется возможность выбора одной из маммографических проекций (рис.2.), на основе которой производится анализ изображения и выделение требуемых диффузных признаков МЖ, таких как состояние тканей, наличие и тип известковых включений и др.

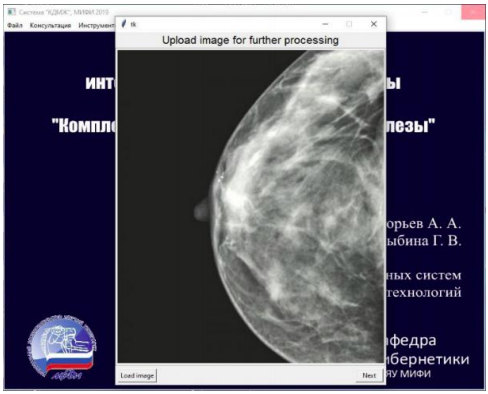


Рис.2. Загрузка маммограммы

Далее врачу предлагается локализовать образование путем выделения прямоугольной области изображения, содержащей образование (рис.3). Если образование отсутствует, то данный этап

пропускается, а в случае локализации образования будет производиться анализ выделенной части изображения и определение всех необходимых характеристик образования (форма, вид и четкость контуров).

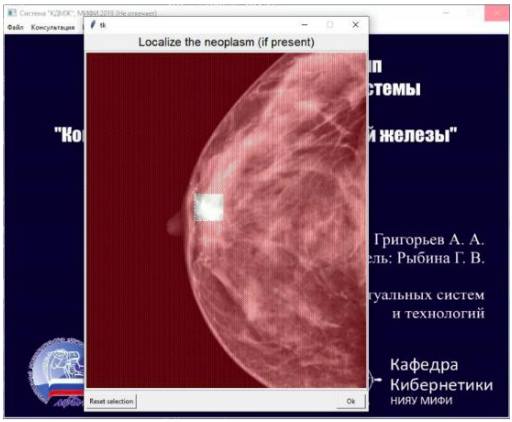


Рис.3. Локализация образования

По итогам консультации в диалоговом окне будут отображены полученные результаты диагностики общего состояния молочной железы и найденных образований, а также даны возможные рекомендации.

5. ЗАКЛЮЧЕНИЕ

Проведенные исследования и разработки показали достаточно высокую эффективность технологии прототипирования прикладных ИЭС с использованием базовых функциональных средств комплекса АТ-ТЕХНОЛОГИЯ и разработанного СПО для обработки изображений. Планируется развитие демонстрационного прототипа ИЭС до уровня исследовательского и его экспериментальная апробация для данной ПрО, а также рассматриваются возможности создания версии в контексте архитектуры обучающей ИЭС для использования в учебных целях.

Литература

1. Кобринский Б.А. Ретроспективный анализ медицинских экспертных систем. // Новости искусственного интеллекта №2, 2005. – С. 6-17.
2. Кобринский Б.А. Триединство факторов уверенности в задачах медицинской диагностики. // Искусственный интеллект и принятие решений. – 2018. – №2. – С. 62-72.

3. Singla J., Grover D., Bhandari A. Medical expert systems for diagnosis of various diseases. // International journal of computer applications. 2014. Vol. 93 – No.7, – PP. 36-43.
4. Mohammed Abbas Kadhim, M. Asfar Alam, Harleen Kaur. Design and implementation of Fuzzy expert system of back pain diagnosis. // International journal of innovative technology & creative engineering. 2011 Vol. 1, No. 9, – PP. 16-22.
5. Singla J. The diagnoses of some lung diseases in a PROLOG expert system. // International journal of computer applications. 2013. Vol. 78 – No.15, – PP. 37-40.
6. Грибова В. В., Островский Г.Е. Интеллектуальная обучающая среда для диагностики острых и хронических заболеваний // Пятнадцатая национальная конференция по искусственному интеллекту с международным участием КИИ-2016 (3-7 октября 2016 г., г. Смоленск). Труды конференции. Т.3. – Смоленск: Универсум. – 2016. – С. 171-179.
7. Клещев А.С., Москаленко Ф.М., Черняховская М.Ю. Модель онтологии предметной области «Медицинская диагностика». Часть 1. Неформальное описание и определение базовых терминов. // НТИ. Серия 2. 2005. №2. С. 1-7.
8. В.В. Грибова, Е.А. Шалфеева. Онтология диагностики процессов. // Онтология проектирования. – 2019. – №4 (34). – С. 449-461.
9. Рыбина Г.В. Теория и технология построения интегрированных экспертных систем. – М.: Научтехлитиздат, – 2008. – 482 с.
10. Рыбина Г.В. Интеллектуальные системы: от А до Я. Серия монографий в трех книгах. Книга 1. Системы, основанные на знаниях. Интегрированные экспертные системы. М.: Научтехлитиздат, – 2014. – 224 с.
11. Рыбина Г.В., Блохин Ю.М. Методы и программные средства интеллектуального планирования для построения интегрированных экспертных систем. // Искусственный интеллект и принятие решений. – 2018. – №1. – С. 12-28.
12. Dominique A. Lobar Approach to Breast Ultrasound. – M.: Springer International Publishing AG. Part of Springer Nature – 2018. – 346 c
13. Якобс О.Э., Рожкова Н.И., Мазо М.Л., Микушин С.Ю. Опыт использования виртуальной сонографии молочной железы. // Вестник рентгенологии и радиологии. Лучевая диагностика. – 2014. No 1. – С. 23-33.
14. Рожкова Н.И., Горшков В.А., Меских Е.В., Мазо М.Л., Прокопенко С.П., Киреева М.Н. и др. Цифровая маммологическая клиника: технологии визуализации. М.: СИМК; 2012.
15. Рафаэл С. Гонсалес, Ричард Е. Вудс. Цифровая обработка изображений. М.: «ТЕХНОСФЕРА». 2012. – 1105 с.

УДК 004.891.3

РЕАЛИЗАЦИЯ НЕЧЕТКОСТИ КЛИНИЧЕСКОЙ КАРТИНЫ НАСЛЕДСТВЕННЫХ БОЛЕЗНЕЙ В ЭКСПЕРТНОЙ ДИАГНОСТИЧЕСКОЙ СИСТЕМЕ

Благосклонов Н.А.
e-mail: nblagosklonov@frccsc.ru
ФИЦ ИУ РАН

1. ВВЕДЕНИЕ

Орфанным болезням в последнее время уделяется всё большее внимание. В Российской Федерации это находит отражение в изменениях, внесенных за последний год в федеральный закон №323-ФЗ (ред. от 22.12.2020) "Об основах охраны здоровья граждан в Российской Федерации" и постановление правительства РФ от 26.04.2012 N 403 (ред. от 05.06.2020) "О порядке ведения Федерального регистра лиц, страдающих жизнеугрожающими и хроническими прогрессирующими редкими (орфанными) заболеваниями, приводящими к сокращению продолжительности жизни граждан или их инвалидности, и его регионального сегмента". Отдельно стоит отметить указ президента РФ №16 от 5 января 2021 г. "О создании Фонда поддержки детей с тяжелыми жизнеугрожающими и хроническими заболеваниями, в том числе редкими (орфанными) заболеваниями, «Круг добра»", который отражает приоритет правительства в отношении помощи лицам, чаще всего детям, больным наследственными болезнями. По определению ФЗ №323-ФЗ редкими (орфанными) заболеваниями являются заболевания, которые имеют распространенность не более 10 случаев на 100 тысяч населения.

Повысить эффективность диагностики таких заболеваний может использование компьютерных систем поддержки принятия врачебных решений. Ранее такие решения уже создавались и эксплуатировались в России [1], Франции [2], Англии [3], Австралии [4]. Однако, в настоящее время, отечественных функционирующих систем не существует.

Зарубежные системы, ассистирующие врачу на долабораторном этапе постановки диагноза, при всех своих очевидных плюсах, имеют ряд существенных недостатков, среди которых можно выделить следующие:

- Отсутствие поддержки русского языка, что помимо неудобства в переводе понятий влечет за собой также неполную сопоставимость отдельных терминов. Это обусловлено различным подходом к описанию клинических проявлений, для которых в медицине присуща метафоричность наименования.
- Отсутствие объяснения предлагаемого компьютерной программой решения в классическом его понимании для искусственного интеллекта. Особенно это актуально для системы Face2Gene [3], работающей на основе принципов нейронных сетей, в которой в качестве объяснения строится тепловая карта при работе с лицами пациентов или выводится некий процент уверенности без объяснения того, как он был получен, когда в систему введены только признаки пациента. Не все системы предоставляют врачу перечни совпавших и не совпавших признаков (иначе «контрпризнаков»).
- Ориентированность на web-приложение, то есть для работы с такими системами необходимо постоянное подключение к интернету и передача данных пациента на сторонний зарубежный сервер, что идет в разрез с политикой работы с персональными данными на территории Российской Федерации. И зачастую в медицинских учреждениях ограничена возможность выхода во внешнюю сеть, что связано с чувствительностью персональной медицинской информации и обеспечением безопасности инфраструктуры.

Отсутствие отечественной интеллектуальной системы поддержки принятия решений обуславливает актуальность настоящей работы. При разработке системы ГенДиЭС особое внимание уделено проблеме нечеткости клинической картины различной выраженности наследственных заболеваний, особенно проявляющихся в раннем детском возрасте.

2. ОСОБЕННОСТИ ОРФАННЫХ ЗАБОЛЕВАНИЙ

В качестве прототипа системы ГенДиЭС рассматриваются лизосомные наследственные болезни обмена веществ, относящиеся к орфанным [5]. Данная группа включает в себя около 50 заболеваний (по различным классификациям), наиболее крупными подгруппами являются мукополисахаридозы (15 клинических форм), муколипидозы (8 клинических форм) и ганглиозидозы (7 клинических форм) [6]. Как и другие подобные наследственные заболевания они характеризуются рядом особенностей, которые необходимо учитывать при построении системы. Наиболее существенными являются:

- Возрастная динамика изменений клинической картины, что проявляется прогрессированием отдельных признаков, но также возможный переход одних проявлений в противоположные. Примером прогрессирования выраженности является нарастающее «огрубление» черт лица пациента по мере взросления от момента первичного проявления данного признака (манифестации). Примером перехода является нарушение функции крупных суставов, а именно «разболтанность», или повышенная мобильность, сменяющаяся на «скованность», вплоть до полной неподвижности.
- Нечеткость вербальных характеристик клинических признаков, которая выражается в лингвистических сравнениях или сопоставлениях, без уточнения конкретики, например – увеличенная голова (макроцефалия).
- Сложность воспроизведения в лингвистической форме трудных для восприятия образных характеристик. Довольно часто для описания внешних проявлений наследственных болезней используется устоявшееся метафорическое выражение. Например, «череп в форме трилистника», «гаргоилоподобные черты лица», «клювовидный нос» и др. Разложить подобные описания на отдельные составляющие, то есть более мелкие проявления, во многих случаях не представляется возможным. Лицевые признаки были достаточно подробно описаны врачами-исследователями и каждый элемент лица представлен своим описанием. Но когда образ больного раскладывается на отдельные составляющие, то теряется комплексность внешних проявлений, или иначе «портрет заболевания».
- Возможное отсутствие в клинической картине релевантных признаков, что обусловлено очень сильным вкладом индивидуальных особенностей каждого пациента, начиная от этнической принадлежности и заканчивая сочетанием диагностируемого заболевания с особенностями развития индивида (обусловленного наследственностью) и проявлениями других болезней (возможно иной генетической природы). В клинической практике довольно часто встречаются отклонения от полного описания болезней в литературе.

Данные особенности проявления наследственных болезней приводят к необходимости разработки компьютерной системы поддержки принятия решений, которая могла бы функционировать в условиях неполной, а иногда и противоречивой информации без существенного снижения точности выдаваемых гипотез. Решение такой

задачи возможно с помощью применения подходов нечеткой логики, предложенной Лотфи Заде [7].

3. МЕРЫ ДОВЕРИЯ К ПРИЗНАКАМ

При создании системы, учитывающей возможность работы с нечеткостью признаков, были использованы факторы уверенности [8] в модификации Б.А. Кобринского [9] для характеристики манифестации и выраженности признаков, а также шкала, разработанная в процессе исследования для оценки модальности.

Извлечение знаний происходило в два этапа. На первом этапе работа велась с литературными данными. Когнитологом были проработаны монографии, статьи, онлайн базы данных по наследственным болезням с акцентированием внимания на внешних проявлениях болезни. Знания из литературы были представлены в виде специально разработанной формы – текстологической карты [10], учитывавшей возраст проявления признака, его уровень выраженности и частоту встречаемости в популяции. На втором этапе проводилось извлечение знаний у экспертов. Эксперты проанализировали текстологические карты и опираясь на собственный опыт при орфанных заболеваниях выделили ограниченный набор признаков для каждой группы заболеваний. Окончательный набор признаков с экспертными оценками по четырем возрастным группам, включавшим характеристику модальности с учетом шкалы, факторы уверенности манифестации и выраженности признаков, составил структурированное формализованное описание лизосомных болезней накопления (матрица болезни – признаки). Данное представление знаний включало для мукополисахаридозов, ганглиозидозов и муколипидозов 22, 21, 27 признаков и 3960, 1764, 2592 экспертных оценок соответственно.

Таким образом, в основе разрабатываемой системы лежат экспертные знания, отражающие модальность признака, доверие к его манифестации и уровню выраженности в данном возрасте. Комплексный подход к представлению знаний о внешних проявлениях заболевания в разных возрастных периодах необходим для преодоления проблемы нечеткости клинической картины при орфанных заболеваниях и помощи врачу на долаборатоном этапе постановки предварительного диагноза.

4. ПОСТРОЕНИЕ МОДЕЛИ

С целью учета возможной нечеткости проявлений орфанных болезней была разработана модель заболевания на основе экспертных оценок. В общем виде данная модель является эталонным клиническим

описанием заболевания, которое учитывает вклад каждого признака в формирование клинической картины. Это достигается за счет того, что оценивается не только факт наличия признака, но и его «вес», который является произведением экспертных оценок. Такой подход позволил комплексно интегрировать экспертные оценки модальности, манифестации и выраженности в пределах каждой возрастной группы и рассматривать заболевание как синтез различных атрибутов каждого признака. Поэтому модель заболевания получила название интегральной оценки заболевания.

Интегральная оценка заболевания есть сумма комплексных оценок признаков в пределах возрастной группы:

$$I = \sum_{i=1}^{n} P_i ,$$

где:

- I – интегрированная оценка признаков болезни;
- P_i – комплексная оценка признака;
- i – количество признаков;
- n – совокупность признаков заболевания (группы заболеваний).

Комплексная оценка признака рассчитывалась по формуле:

$$P_i = M_i \cdot m_i \cdot s_i ,$$

где:

- P_i –признак (симптом);
- M_i – модальность признака, характеризующая его частоту;
- m_i – фактор уверенности манифестации признака;
- s_i – фактор уверенности выраженности признака.

Модель заболевания нашла применение в двух возможных модификациях: как эталонная модель заболевания, которая складывалась из всех описанных экспертами признаков, и как модель случая, получаемая как сумма только тех признаков, которые присутствовали у пациента.

5. ПРОТОТИП СИСТЕМЫ

Для получения прототипа системы был реализован алгоритм сопоставительного анализа. Задачей, которую решает данный алгоритм, является сравнение случая с эталоном, для определения перечня наиболее соответствующих клинической картине пациента гипотез. Однако, ввиду необходимости обработки нечеткости, была сделана надстройка. Для учета ситуаций, когда процент совпадения случая с эталоном может быть очень высоким, но при это у пациента имеется большое количество «лишних» признаков, не описанных экспертами в

эталонном варианте заболевания, был введен динамический порог для «признаков, не относящихся к гипотезе». Реализация такого подхода представлена на рис.1.

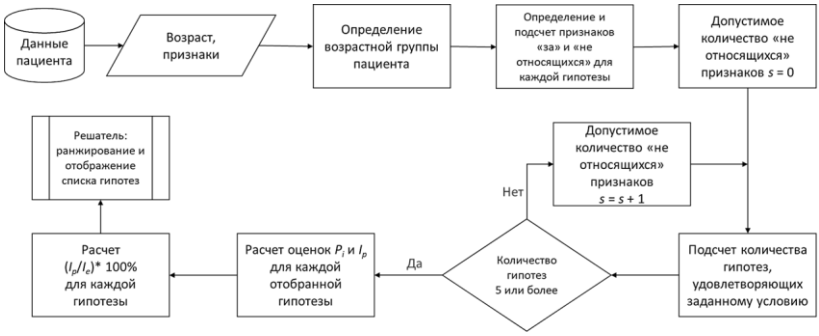


Рис.1. Схема работы системы

В основе надстройки с динамическим порогом лежит представление о том, что все признаки, которые описал врач, осматривая пациента, относятся к проявлениям одного наследственного заболевания. Далее алгоритм проверяет возможно ли отобрать необходимое число гипотез для последующего ранжирования. В данном случае число минимально необходимых гипотез было определено в количестве пяти. Это обусловлено оптимальным числом предварительных диагнозов, которые врач может проверить с помощью направления на молекулярно-генетическое тестирование для подтверждения окончательного диагноза и, в тоже время, согласуется с рядами, выдаваемыми в большинстве зарубежных систем и ранее функционировавшей российской системе. Если невозможно сформировать минимальный дифференциально-диагностический ряд гипотез с фиксированным условием, то количество допустимых признаков увеличивается на единицу за счет признаков эталонного описания. Затем производится проверка. Такой цикл продолжается до тех пор, пока не появится возможность формирования дифференциально-диагностического ряда, состоящего из требуемого количества гипотез.

В отношении учета нечеткости такой подход позволяет на этапе до сопоставления случая с эталонными описаниями отсеивать гипотезы, которые содержат избыточное количество ложноположительных признаков, при этом имея более низкий процент общего совпадения с эталонным описанием.

Целесообразным является предоставление врачу возможности выбора величины списка гипотез, например, в количестве трех, пяти, десяти или без ограничений. Однако, необходимо отметить, что чем меньше перечень дифференциально-диагностических гипотез, выдаваемых системой, тем меньше вероятность того, что верный диагноз будет включен в него. Увеличивая список предварительных диагнозов, будет увеличиваться вероятность вхождения в него правильного диагноза, но при этом он может находиться на более низкой позиции, чем неточные.

Для отобранных гипотез производится расчет интегрированных оценок случаев, которые впоследствии сравниваются с эталонными описаниями и ранжируются по проценту совпадения от большего к меньшему. Ранжированный перечень выводится врачу с указанием на:

- признаки пациента, которые совпали с эталоном;
- признаки пациента, которые отсутствовали в эталоне (при их наличии);
- признаки из эталона, которые отсутствовали у пациента, с предложением провести дополнительный осмотр на наличие таких признаков.

Данный подход был протестирован на двадцати описаниях клинических случаев мукополисахаридозов из журнальных статей. Публикации для проведения эксперимента отбирались как отечественные, так и зарубежные. Описания клинических проявлений приводились к единой терминологии, используемой в прототипе системы ГенДиЭС.

Критерием успешности работы прототипа было вхождение верифицированного диагноза (на основе молекулярно-генетического анализа по данным публикации) в ранжированный перечень из пяти гипотез, выданный системой. По результатам эксперимента точность диагностики с использованием прототипа системы ГенДиЭС составила 90%.

6. ЗАКЛЮЧЕНИЕ

Учет вариабельности внешних проявлений, включая непосредственно факт манифестации и силу выраженности признаков при редких наследственных заболеваниях играет определяющую роль при разработке системы.

Использование экспертных знаний в отношении не только фактов наличия или отсутствия признаков при заболеваниях, а на основе комплексных оценок модальности, манифестации и выраженности позволяют получить интегрированную оценку заболевания. Реализованная модель может функционировать в

условиях ограниченной, а порой и неточной информации о клинических проявлениях наследственных болезней.

Результат тестирования системы на примере описаний случаев мукополисахаридозов в литературе, учитывающий нечеткость фенотипических проявлений на основе экспертных оценок, показал эффективность на уровне 90%.

Литература

1. Кобринский Б.А. Ретроспективный анализ медицинских экспертных систем. // Новости искусственного интеллекта. – 2005. – №2. – С. 6-17.
2. Ayme S., Caraboenf M., Gouvernet J. GENDIAG: A computer assisted syndrome identification system // Clinical Genetics. – 1985. – Vol. 28. – No.5. – P. 410-411.
3. Allanson J.E., Cunniff C., Hoyme H.E., McGaughran J., Muenke M., Neri G. Elements of morphology: standard terminology for the head and face // American Journal of Medical Genetics. Part A. – 2009. – Vol. 149A. – No.1. – Pp. 6-28.
4. Fryer A. POSSUM (Pictures of Standard Syndromes and Undiagnosed Malformations) // Journal of Medical Genetics. – 1991. – Vol.28 – No.1. – Pp. 66-67.
5. Orphanet [Электронный ресурс] URL: https://www.orpha.net/consor/cgi-bin/index.php [Дата обращения: 15.03.2021].
6. La Marca, G. Lysosomals // Physician's Guide to the Diagnosis, Treatment, and Follow-Up of Inherited Metabolic Diseases / Blau, N., Duran, M., Gibson, K.M., Vici, C.D. (eds.) – Berlin Heidelberg: Springer, 2014. – Pp. 785-793
7. Zadeh L.A. Toward a Theory of Fuzzy Information Granulation and its Centrality in Human Reasoning and Fuzzy Logic // Fuzzy Sets and Systems. – 1997. – Vol. 90. – P. 111-127.
8. Shortliffe E.H., Buchanan B.G. A Model of Inexact Reasoning in Medicine // Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project / B.G. Buchanan and E.H. Shortliffe (Eds.). – Reading, London, Amsterdam, Sydney: Addison-Wesley Publishing Company, 1984. – Pp. 233-262.
9. Кобринский Б.А. Триединство факторов уверенности в задачах медицинской диагностики // Искусственный интеллект и принятие решений. – 2018. – №2. – С. 62-72.
10. Kobrinskii B.A., Blagosklonov N.A. Hybrid approach to knowledge extraction: textual analysis and evaluations of experts // Открытые семантические технологии проектирования интеллектуальных систем. – 2018. – Т.2. – №8. – С. 191-195.

УДК 004.822

РАЗРАБОТКА БАЗЫ ЗНАНИЙ ЭКСПЕРТНОЙ СИСТЕМЫ ДИАГНОСТИКИ ПАЦИЕНТОВ С ЖЕЛУДОЧНО-КИШЕЧНЫМИ КРОВОТЕЧЕНИЯМИ

Будыкина А.В.
e-mail: budykina.rsmu@yandex.ru
ФГАОУ ВО РНИМУ им. Н.И. Пирогова Минздрава России

1. ВВЕДЕНИЕ

Одним из самых тяжелых осложнений заболеваний желудочно-кишечного тракта (ЖКТ) является желудочно-кишечное кровотечение (ЖКК). Диагностика больных с подозрением на желудочно-кишечные кровотечения, в том числе тонкокишечные, имеет ряд трудностей, которые увеличивают время постановки диагноза, начала лечения и, соответственно, отрицательно влияют на качество оказываемой медицинской помощи пациентам данного профиля [5, 11]. С активным внедрением в настоящее время в медицинские организации информационных систем (МИС) повышается востребованность использования систем поддержки принятия клинических решений (СППКР) с целью улучшения качества оказания медицинской помощи [4]. На сегодняшний день в отечественной практике врача-эндоскописта, врача-хирурга или врача-гастроэнтеролога не встречено примеров использования систем, поддерживающих решения в вопросе диагностики больных желудочно-кишечными кровотечениями. В отечественной и зарубежной литературе описаны прототипы подобных систем, но, к сожалению, они не имеют практического применения [8, 10, 13].

2. МАТЕРИАЛЫ И МЕТОДЫ

Данные клинических рекомендаций по диагностике и лечению пациентов с ЖКК [6, 9, 12], а также знания ведущих экспертов в области клинической эндоскопии кафедры госпитальной хирургии № 2 с научно-исследовательской лабораторией хирургической гастроэнтерологии и эндоскопии Российского национального исследовательского медицинского университета имени Н.И. Пирогова послужили материалом для построения базы знаний по диагностике желудочно-кишечных кровотечений у пациентов с заболеваниями ЖКТ.

Методологической основой для проведения текстологических процедур извлечения знаний с целью передачи его смысла явилось формализованное представление инженером по знаниям текста клинических рекомендаций в виде набора ключевых слов с использованием Microsoft Office Excel. Согласованные с экспертами знания переносились в пространство графовой системы управления базами данных с открытым исходным кодом – Neo4j [14].

На основе формализованных знаний в области диагностики и лечения желудочно-кишечных кровотечений неясного генеза разработаны анкеты для оценки значимости клинических и диагностических признаков, а также выбора диагностических исследований при ЖКК [3]. Первичная структура и наполнение анкет произведено в программе Microsoft Office Word. Результаты согласования экспертных решений оформлены в среде MS Excel.

3. РЕЗУЛЬТАТЫ

Структура разработанной базы знаний построена на основе этапов ее использования в составе СППКР:

1. Постановка предварительного диагноза:

1.1 По набору клинических признаков формируется диагностическая гипотеза о предварительном диагнозе, а также перечень необходимых и достаточных исследований для его подтверждения.

2. Оптимизация назначения дополнительных методов исследования:

2.1 Выявление дополнительных методов диагностики – формирование набора диагностических признаков, необходимых для подтверждения гипотезы, выдвинутой в процессе постановки предварительного диагноза, а также формирование перечня диагностических методов, выявляющих сформированные признаки.

2.2 Определение признаков, при наличии которых противопоказано то или иное диагностическое исследование.

Для наполнения базы знаний из источников знаний извлечены ключевые понятия и согласованы с экспертами. Под понятием подразумевается совокупность признаков, которые все вместе достаточны, а каждый необходим для того, чтобы отделить эту совокупность от других [7].

Для сохранения смысла понятий извлечены не только лингвистические синонимы, например, «пурпура Геноха-Шенлейна», «геморрагический васкулит» и «ревматическая пурпура», но и общепринятые варианты написания понятий, например, «неспецифический язвенный колит» и «НЯК». Весь динамический

тезаурус предметной области представлен в виде онтологии. С учетом сценариев использования онтологии в составе СППКР разработана база знаний диагностики и лечения желудочно-кишечного кровотечения неясного генеза в пространстве Neo4j [2]. Главными достоинствами данной графовой системы управления базами данных являются гибкость модели данных, простой поиск, язык запросов Cypher, команды которого представлены в удобочитаемом формате и просты в освоении. Кроме этого, Neo4j работает с такими языками программирования, как Java, Spring, Scala, что упрощает процесс реализации СППКР.

Архитектура базы знаний включает в себя восемь групп (анатомическая локализация, диагноз, диагностический признак, диагностическое исследование, клинический признак, лабораторное исследование, медикаментозное лечение и немедикаментозное лечение). Группа представляет собой совокупность понятий, имеющих общее назначение. Отношения между понятиями определяют типы связей (выявление, лечение, локализация, осложнение, показание, признак, противопоказание, род-вид, синоним, сопутствие, фактор риска, часть-целое), которые могут быть одно- и двусторонними.

Совместно с ведущими экспертами в области клинической эндоскопии кафедры госпитальной хирургии № 2 с научно-исследовательской лабораторией хирургической гастроэнтерологии и эндоскопии Российского национального исследовательского медицинского университета имени Н.И. Пирогова согласовано 605 уникальных ключевых понятий (263 из которых – синонимы), выделенных исходно из клинических рекомендаций (таблица 1).

Таблица 1. Фрагмент номенклатуры понятий

Уникальный идентификатор понятия	Наименование понятия	Синоним понятия	Группа
51	Кровавая рвота	Рвота кровью	Клинический признак
110	Профузное кровотечение	Массивное кровотечение	Клинический признак
204	Синдром Мэллори-Вейсса	Рвота алой кровью	Клинический признак
235	Гемоглобин	Hb	Лабораторное исследование
286	Баллоно-ассистированная энтероскопия	БАЭ	Диагностическое исследование
540	Спайки кишки	Сращения кишки	Диагностический признак

Для построения правил вывода базы знаний экспертной системы диагностики ЖКК рассмотрены различные классификации желудочно-кишечных кровотечений. Экспертным сообществом принято решение о практической значимости трех классификаций, необходимых и достаточных для диагностики пациентов с ЖКК – клиническая картина, анатомическая локализация, степень тяжести (таблица 2).

Таблица 2. Классификации желудочно-кишечных кровотечений

По клинической картине	По локализации	По степени тяжести кровотечения
1. Явное (острое)	1. Верхние отделы ЖКТ	1. Легкая
2. Скрытое (оккультное)	2. Средние отделы ЖКТ	2. Средняя
	3. Нижние отделы ЖКТ	3. Тяжелая

В соответствии с комбинаторным правилом умножения [1]:

$$n_1 \cdot n_2 \cdot n_3 \cdot \ldots \cdot n_k \quad (1)$$

где n – количество элементов в каждой классификации, k – количество классификаций, получено 18 заключений. По формуле (1) комбинации классификаций определяют следующие заключения:

1. ЖКК явное, локализованное в верхних отделах ЖКТ, легкой степени тяжести.
2. ЖКК явное, локализованное в верхних отделах ЖКТ, средней степени тяжести.
3. ЖКК явное, локализованное в верхних отделах ЖКТ, тяжелой степени.
4. ЖКК явное, локализованное в средних отделах ЖКТ, легкой степени тяжести.
5. ЖКК явное, локализованное в средних отделах ЖКТ, средней степени тяжести.
6. ЖКК к явное, локализованное в средних отделах ЖКТ, тяжелой степени.
7. ЖКК явное, локализованное в нижних отделах ЖКТ, легкой степени тяжести.
8. ЖКК явное, локализованное в нижних отделах ЖКТ, средней степени тяжести.
9. ЖКК явное, локализованное в нижних отделах ЖКТ, тяжелой степени.
10. ЖКК скрытое, локализованное в верхних отделах ЖКТ, легкой степени тяжести.
11. ЖКК скрытое, локализованное в верхних отделах ЖКТ, средней степени тяжести.
12. ЖКК скрытое, локализованное в верхних отделах ЖКТ, тяжелой степени.

13. ЖКК скрытое, локализованное в средних отделах ЖКТ, легкой степени тяжести.
14. ЖКК скрытое, локализованное в средних отделах ЖКТ, средней степени тяжести.
15. ЖКК скрытое, локализованное в средних отделах ЖКТ, тяжелой степени.
16. ЖКК скрытое, локализованное в нижних отделах ЖКТ, легкой степени тяжести.
17. ЖКК скрытое, локализованное в нижних отделах ЖКТ, средней степени тяжести.
18. ЖКК скрытое, локализованное в нижних отделах ЖКТ, тяжелой степени.

На основе формализованного описания предметной области все оценочные признаки были сгруппированы по следующим направлениям [3]:

1. Клинические признаки, полученные при сборе анамнеза и осмотре пациента;
2. Диагностические исследования, осуществляемые для диагностики ЖКК;
3. Диагностические признаки, выявленные при проведении диагностических мероприятий.

Характерность проявления каждого признака, а также необходимость проведения конкретного диагностического исследования для каждого заключения из таблицы 1 оценивалось по 10-бальной шкале, где «0» – полное отсутствие проявления признака или необходимости назначения диагностического метода, а 10 – обязательное наличие признака или направление на диагностическое исследование.

Для работы экспертов были составлены три оригинальные анкеты. В столбцах анкеты указаны заключения, в строках – признаки заболевания или диагностические методы. На пересечении каждый эксперт проставил оценку от 0 до 10 (рис.1).

На рис.1 представлен фрагмент одной из заполненных экспертом анкет. По горизонтали указаны заключения, по вертикали – признаки. На их пересечении указаны баллы, проставленные экспертами самостоятельно.

На первом этапе каждым экспертом независимо заполнены все анкеты после предварительной инструкции когнитологом. Всего в процессе анкетирования каждый эксперт ответил на 2160 вопросов.

Признак заключения	ЖКК явное, локализованное в СРЕДНИХ отделах ЖКТ, ЛЕГКОЙ степени тяжести.	ЖКК явное, локализованное в СРЕДНИХ отделах ЖКТ, СРЕДНЕЙ степени тяжести.	ЖКК явное, локализованное в СРЕДНИХ отделах ЖКТ, ТЯЖЕЛОЙ степени.
Рвота "кофейной гущей"	0	0	0
Рвота алой кровью	0	0	0
Рвота темной кровью	0	0	0
Рвота малоизменённой кровью	0	0	0
Мелена	0	0	0
Гематохезия	2	7	10
Кровянистая слизь	0	0	1

Рис.1. Фрагмент анкеты «Клинические признаки»

В результате анализа 15 анкет было установлено полное совпадение ответов лишь в 435 вопросах, что составляет 20% от общего числа вопросов. Такую разницу можно объяснить особенностями специализации экспертов, наличием субъективности в ответах при заполнении анкет.

Для устранения разногласий было принято решение о необходимости совместного обсуждения анкет с привлечением инженера по знаниям и согласования ответов.

На рис.2 изображен фрагмент одной из анкет. В столбцах указаны заключения, в строках – наименования диагностических исследований, назначаемых пациентам при подозрении на желудочно-кишечные кровотечения. На их пересечении указаны баллы, проставленные экспертами самостоятельно (цветные строки), а также итог согласования (белые строки).

Наименование исследования	ЖКК СКРЫТОЕ, локализованное в верхних отделах ЖКТ, ЛЕГКОЙ степени тяжести.	ЖКК СКРЫТОЕ, локализованное в верхних отделах ЖКТ, СРЕДНЕЙ степени тяжести.	ЖКК СКРЫТОЕ, локализованное в верхних отделах ЖКТ, ТЯЖЕЛОЙ степени.
МР-энтерография	0	0	0
МР-энтерография	0	0	0
МР-энтерография	0	0	0
МР-энтерография	0	0	0
МР-энтерография	0	0	0
	0	0	0
Общий анализ крови	10	10	10
Общий анализ крови	10	10	10
Общий анализ крови	10	10	10
Общий анализ крови	10	10	10
Общий анализ крови	10	10	10
	10	10	10
Пассаж бария по кишке	0	0	0
Пассаж бария по кишке	0	0	0
Пассаж бария по кишке	0	0	0
Пассаж бария по кишке	0	0	0
Пассаж бария по кишке	0	0	0
	0	0	0
ЭРХПГ	0	0	1
ЭРХПГ	7	7	7
ЭРХПГ	0	0	0
ЭРХПГ	0	0	0
ЭРХПГ	6	5	0
	1	1	2

Рис.2. Фрагмент анкеты «Диагностические исследования»

4. ЗАКЛЮЧЕНИЕ

В результате проделанной работы получена номенклатура понятий в области диагностики больных с ЖКК, а также экспертные оценки, которые, наряду с ретроспективными данными о пациентах и данными клинических рекомендаций в рассматриваемой предметной области, будут использованы при построении решающего правила диагностики пациентов с подозрением на желудочно-кишечные кровотечения.

Литература

1. Беккенбах, Э. Прикладная комбинаторная математика. Сборник статей / Э. Беккенбах. – М., 2015. – 528 с.
2. Будыкина А.В., Киселёв К.В., Раузина С.Е., Зарубина Т.В., Федоров Е.Д., Юдин О.И., Тихомирова Е.В. Онтология диагностики и лечения желудочно-кишечных кровотечений с неустановленным источником. Материалы VII Международной конференции «Знания-Онтологии-Теории» (ЗОНТ-19), 7-11 октября 2019 г., Новосибирск. – 2019. – С. 88-93.

3. Будыкина А.В., Тихомирова Е.В., Киселев К.В., Зарубина Т.В., Раузина С.Е., Федоров Е.Д., Юдин О.И. Формализация знаний о желудочно-кишечном кровотечении неясного генеза для использования в интеллектуальных системах поддержки принятия врачебных решений // Вестник новых медицинских технологий. – 2020. – №4. – С. 98–101. DOI: 10.24411/1609-2163-2020-16741.
4. Гусев А.В., Зарубина Т.В. Поддержка принятия врачебных решений в медицинских информационных системах медицинской организации // Врач и информационные технологии. – 2017. – № 2. – С. 60-72.
5. Домарев Л.В. Капсульная эндоскопия в диагностике заболеваний тонкой кишки. Москва: ФГУ Институт Хирургии им. А.В. Вишневского, 2007.
6. Руководство по скорой медицинской помощи при желудочно-кишечном кровотечении. Москва: Минздрав РФ. – 2014.
7. Стожок Е.В. Термин, понятие и значение // ОНВ. – 2011. – №1. – С. 74-81.
8. Chu A., Ahn H., Halwan B., Kalmin B., Artifon E.L.A., Barkun A., Kumar A. A decision support system to facilitate management of patients with acute gastrointestinal bleeding // Artificial Intelligence in Medicine. 2008. Vol. 42, N3. P. 247-259. DOI: 10.1016/j.artmed.2007.10.003.
9. Gerson L.B., Fidler J.L., Cave D.R., Leighton J.A. ACG Clinical Guideline: Diagnosis and Management of Small Bowel Bleeding // The American Journal of Gastroenterology. 2015. Vol. 110, N9. P. 1265-1287. DOI: 10.1038/ajg.2015.246.
10. Le Berre C., Sandborn W.J., Aridhi S., Devignes M.-D., Four-nier L., Smaïl-Tabbone M., Danese S., Peyrin-Biroulet L. Application of Artificial Intelligence to Gastroenterology and Hepatology, Gastroenter-ology, 2019. DOI: https://doi.org/10.1053/j.gastro.2019.08.058.
11. Nawaz A. Chronic gastrointestinal bleeding of obscure origin: diagnosis and management // Hospital Physician. 2001. P. 48-53.
12. Rondonotti E., Spada C., Adler S., May A., Despott E., Koulaouzidis A., Pennazio M. Small-bowel capsule endoscopy and device-as-sisted enteroscopy for diagnosis and treatment of small-bowel disorders: European Society of Gastrointestinal Endoscopy (ESGE) Technical Review // Endoscopy. 2018. Vol. 50, N04. P. 423-446. DOI: 10.1055/a-0576-0566.
13. Thomson M.A., Leton N., Belsha D. Acute Upper Gastrointes-tinal Bleeding in Childhood // Journal of Pediatric Gastroenterology and Nutrition. 2015. Vol. 60, N. 5. P. 632-636. DOI: 10.1097/mpg.0000000000000680.
14. Neo4j. Graph database management system [Electronic re-source], [accessed may 03, 2020]. Available at: https://neo4j.com/.

УДК 57.081

ПЛАТФОРМА ДЛЯ ИССЛЕДОВАНИЯ АНГИОГЕНЕЗА IN VITRO

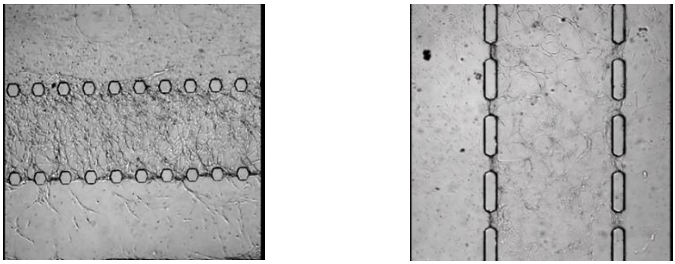
***Глотов В.А.**, д.м.н.*
e-mail: forssma@yandex.ru
Демидов А.Л.
Жилкин В.В.
***Фёдоров Г.Н.**, к.м.н.*
***Прудников И.М.**, к.ф.-м.н.*
ФГБОУ ВО «Смоленский государственный медицинский университет» Минздрава России

Постановка экспериментальных исследований ангиогенеза in vitro – сложнейшая научно-техническая задача, которая по силам только высокотехнологичным хорошо финансируемым исследовательским группам в лабораториях клеточных технологий, обладающих соответствующим оборудованием, штатами и компетенциями. Лабораторное оборудование, посуда, реактивы, питательные среды, первичные клеточные культуры эндотелиальных клеток, микрофлюидные чипы (МФЧ) и прочее, как правило, не производятся на территории РФ и приобретаются через отечественные компании-посредники, являющиеся дилерами крупных зарубежных компаний Cell Applications, Inc.; Dolomite Microfluidics и др. [1, 2, 3], финансируя их дальнейшее развитие. РФ находится в полной технологической зависимости от развитых биотехнологических держав и это отставание стремительно нарастает, РАН не владеет ситуацией в этой области. В России проведение таких исследований могут, к сожалению, позволить себе только единичные крупные высокобюджетные научно-исследовательские центры и лаборатории, имеющие современную научную инфраструктуру. При этом они практически всегда находятся в положении догоняющих, повторяющих, дополняющих, ведомых, но только не ведущих.

В тоже время, в высокоразвитых биотехнологических державах, таких как Южная Корея, США, Израиль, Китай, Япония, экспериментальные исследования ангиогенеза in vitro развиваются всё возрастающими темпами. Это начинает напоминать ядерную гонку, которая началась после открытия Лизой Мейтнер и Отто Ганом нового

типа ядерной реакции деления ядра урана с помощью нейтронов в конце 30-х годов прошлого века.

Исследователям и инженерам под руководством профессора Ноо Ли Чона из Сеульского национального университета (Южная Корея) удалось разработать и создать экспериментальную платформу, в которой был воспроизведён феномен развития в культуре эндотелиальных клеток in vitro, совмещённой с микропотоками питательной среды, 3-D функционирующих саморазвивающихся эндотелиальных капиллярных сетей (сокращённо СЭКС), способных к массопереносу. Решена ключевая проблема экспериментального ангиогенеза (рис.1) [4-8].



а) б)

Рис.1. Так выглядят искусственные функционирующие СЭКС, способные к массопереносу in vitro [6]

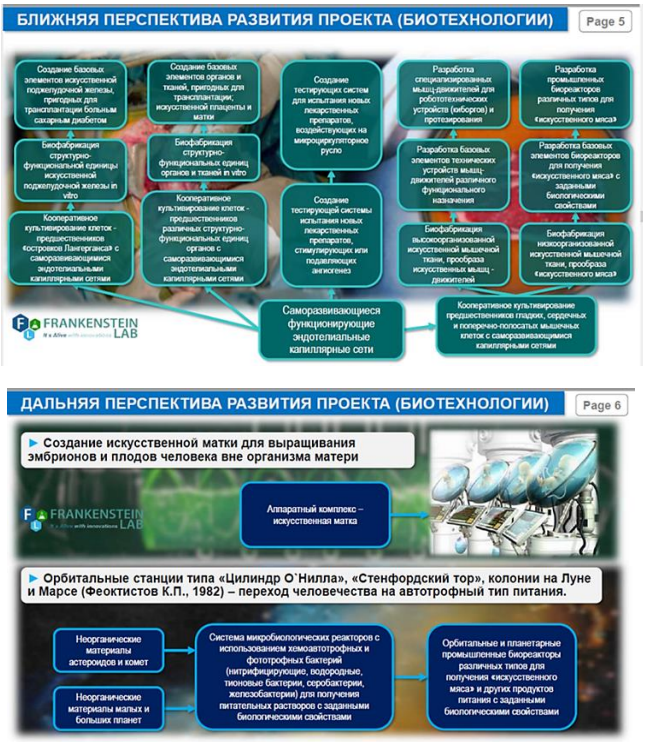
Полные видеозаписи воспроизведения феномена функционирующих СЭКС, способных к массопереносу in vitro представлены по ссылкам:

- рис.1а) – https://yadi.sk/i/bP2wtwvU6Lyjzw [6];
- рис.1б) – https://yadi.sk/i/-ruxrxZv76J4xQ [6].

Совершен колоссальный биотехнологический прорыв в области клеточной и тканевой инженерии эндотелия. Появилась реальная перспектива на основе технологии функционирующих СЭКС приступить к экспериментам по конструированию in vitro искусственных тканеподобных и органоподобных образований с заданными проектируемыми биологическими свойствами. В основе таких экспериментов лежит кооперативное культивирование СЭКС и первичных культур различных типов клеток (ГМК, мышечных клеток, нервных клеток, β-клетки островков Лангерганса и др.).

Технологии функционирующих СЭКС открывают перед человечеством фантастические перспективы: создание мясоподобной биомассы с заданными биологическими свойствами, как полуфабриката для пищевой промышленности и предпосылок, в перспективе, перехода человечества на псевдоавтотрофный тип питания; искусственных мышц –

движителей для роботизированных комплексов природоподобного типа и киборгов; искусственной плаценты и матки – способных обеспечить эмбриональный и плодный периоды развитие человека in vitro от зиготы до новорожденного ребёнка – шаг к управляемой демографии, возможности быстрого наращивание численности населения в отдельно взятой стране либо во внеземных колониях; создание in vitro искусственных органов природоподобного типа пригодных к трансплантации без иммунологических проблем, одним из первых таких органов скорее всего будут искусственные островки Лангерганса; исследовательские лабораторные комплексы для изучения действия и тестирования лекарственных препаратов на природоподобных моделях и многое-многое другое (рис.2).



а)

б)

Рис.2. Ближняя (а) и дальняя (б) перспектива технологии функционирующих СЭКС (Проект Фонда Сколково «Универсальная платформа «Франкенштейн» для биофабрикации искусственных тканей и органов» (Научный руководитель – В.А. Глотов)

Такие исследования требуют создания специальных экспериментальных платформ с инженерным сопровождением высочайшего уровня. Проектирование и создание таких платформ в настоящее время из сферы научной перешло в сферу инженерно-научную.

Разрабатываемая платформа представляет собой сложную структуру с многими параметрами и представляет собой новую область исследований. Нахождение оптимальных параметров давления, температуры, влажности, скорости потока питательной жидкости, при которых образуется капиллярная сеть, – это самостоятельная сложная математическая и инженерная задача, так как значения параметров сильно зависят от конструкции платформы. Решением проблемы может быть создание саморегулируемых, адаптивных систем с возможным интерактивным вмешательством математика-программиста. Платформа в общем случае описывается негладкими липшицевыми функциями, поэтому для нахождения оптимальных требуется применять новые методы негладкой оптимизации, разрабатываемые математиками и инженерами Смоленского государственного медицинского университета.

В 2020 году на кафедре анатомии человека Смоленского государственного медицинского университета была завершена совершенно оригинальная НИОКР при поддержке Фонда содействия развитию малых форм предприятий в научно-технической сфере: Проект НТИ-50323 «Разработка платформы на базе микрофлюидного чипа для исследования ангиогенеза ex vivo» (Научный руководитель – д.м.н., проф. В. А. Глотов, главный конструктор – В.В. Жилкин) (рис.3-7) [9].

В ходе проекта проведена разработка платформы на базе микрофлюидного чипа для исследования ангиогенеза ex vivo. Платформа позволит исследовать комбинированное влияние индукторов и ингибиторов ангиогенеза на образование функционирующих СЭКС.

Разработанная платформа на базе микрофлюидного чипа для исследования ангиогенеза ex vivo имеет следующие параметры и обеспечивает следующий функционал:

- возможность культивации клеточных линий (как единственной, так и композиций из нескольких), в том числе воссоздание тканеподобной клеточной структуры на коллагеновой матрице биологического происхождения;
- возможность поддержания клеточной линии в течение не менее 20 суток;

- система поддержания жизнедеятельности клеточных линий, культивируемых в МФЧ, обеспечивать заданные параметры среды не менее 24-х часов без вмешательства оператора;
- температурный режим работы платформы +25°С до +60 °С, режим поддержания жизнедеятельности клеток – +25°С до +45°С;
- гидравлическая система оснастки платформы (система трубок, соединителей-разветвителей, запирающие клапаны) изготовлена из биосовместимых материалов, устойчивых к воздействию слабых кислот и щелочей;
- возможность дозированной подачи химических факторов в концентрациях, заданных требованиями эксперимента, в объёме от 1 мкл до 100 мкл;
- возможность оптической регистрации с разрешением не менее 600х600 пикселей, со скоростью не менее 10 кадров в секунду;
- возможность описать процесс образования сосудистого русла и количественно оценить следующие параметры: количество отростков, общая длина, участки и интенсивность ветвления, величина просвета в различных участках русла, а также зафиксировать временные параметры протекающих процессов.

Разработанная платформа (рис.3-7) соответствует Техническому заданию по проекту. Запланированные результаты проекта достигнуты.

Рис.3. Структурная схема платформы для исследования ангиогенеза in vitro

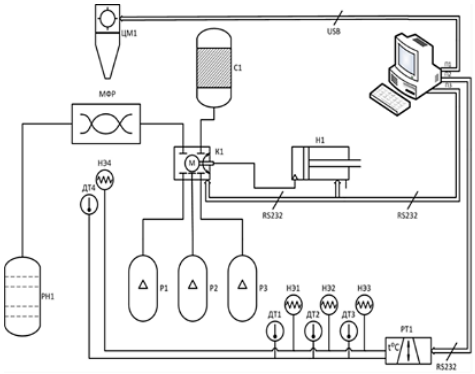


Рис.4. Функциональная схема платформы для исследования ангиогенеза in vitro

Рис.5. Внешний вид платформы для исследования ангиогенеза in vitro

Рис.6. Платформа в сборе без кейса

На рис. 6 изображены:
1. Управляющий компьютер.
2. Драйверы устройств.
3. GSM модем.
4. Микроконтроллер управления.

5, 6. Клапаны – селекторы.

7, 8. Поршневые насосы – микродозаторы.

9, 10. Рециркуляторы газовой смеси.

11. USB микроскоп.

12. Батарея емкостей с питательной жидкостью, индуктрами и ингибиторами ангиогенеза и др.

13. Электронный термостат с датчиком температуры.

14. Микрофлюидный чип в фокусе объектива микроскопа.

15. Монтажная панель.

16 Нагревательный мат.

17. Баллон с CO2, с редуктором и электомагнитным клапаном. Б/Н Патрубки циркуляции смеси.

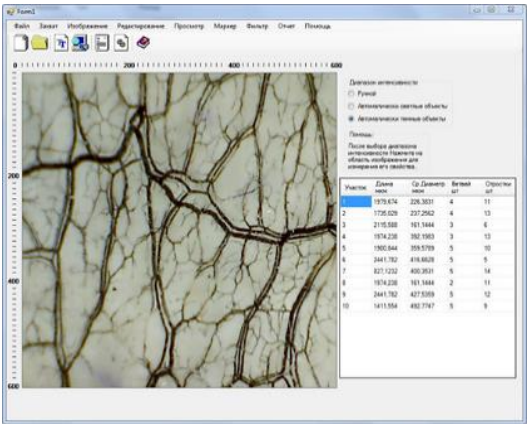


Рис.7. Графический интерфейс программы технического зрения платформы

Разработанная платформа представляет собой мобильную, автономную, кибернетическую биотехническую систему, позволяющую при соблюдении соответствующих протоколов исследования воспроизводить в культуре эндотелиальных клеток феномен функционирующих СЭКС и проводить на них широкий спектр тканеинженерных и биотехнологических исследований. Развитие этого направления исследований и НИОКР в РФ позволит быстро (за 2-3 года) преодолеть наше биотехнологическое отставание от ведущих биотехнологических держав, а при должной политической, организационной и финансовой поддержке стать лидерами в этой стратегически важной области.

Исследование выполнено в рамках проектов РФФИ №94-04-13544 «Структурный анализ микрососудистых бифуркаций» (Научный руководитель – Глотов В.А.), РФФИ №96-04-50991 «Клеточная и тканевая инженерия эндотелия (формирование в культуре эндотелия in vitro функционирующих саморазвивающихся капиллярных сетей)» (Научный руководитель – Глотов В.А.), Проекта Фонда Сколково «Универсальная платформа «Франкенштейн» для биофабрикации искусственных тканей и органов» (2015-2018) (Научный руководитель – Глотов В.А.), Проектов Фонда содействия развитию малых форм предприятий в научно-технической сфере. Проект НТИ-50323 «Разработка платформы на базе микрофлюидного чипа для исследования ангиогенеза ex vivo» (Научный руководитель – Глотов В.А., главный конструктор – Жилкин В.В.). Проект НТИ-84017 «Разработка мобильной, автономной, кибернетической, биотехнической платформы для контролируемого получения эффективных популяций свободных эндотелиальных клеток (первичных культур эндотелия) ex tempore из эндотелиальной выстилки кровеносных сосудов пупочного канатика человека (Генератор эндотелиальных клеток) для клеточных и тканеинженерных технологий (МАКБП-ГЭК)» (Научный руководитель – Глотов В.А., главный конструктор – Жилкин В.В.).

Литература

1. https://www.cellapplications.com/.
2. http://sputnic-group.ru/sputnik/.
3. Dolomite Microfluidics – URL: www.dolomite-microfluidics.com.
4. Noo Li Jeon, Ju Hun Yeon, Qing Ping Hu, Su Dong Kim, Hyun Jae Lee. Device and method of 3-dimensionally generating in vitro blood vessels – Patent US 20110244567Al – Oct. 6, 2011.
5. Ju Hun Yeon, Hyun Ryul Ryu, Minhwan Chung, Qing Ping Hu, Noo Li Jeon In vitro Formation and characterization of a perfusable three-dimensional tubular capillary network in microfluidic devices. – Lab on a Chip (Cover), 12, 2815-2822, May 24, 2012.
6. Sudong Kim, Hyunjae Lee, Minhwan Chung, Noo Li Jeon. Engineering of functional, perfusable 3D microvascular networks on a chip. – Lab on a chip, 13(8), 1489-1500 (2013).
7. Hyunjae Lee, Minhwan Chung, Noo Li Jeon. Microvasculature: An essential component for organ-on-chip systems. – MRS Bulletin, 39(1), 51-59 (2014).
8. Hyunjae Lee, Sudong Kim, Minhwan Chung, Jeong Hun Kim, Noo Li Jeon. A Bioengineered Array of 3D Microvessels for Vascular Permeability Assay. Microvascular Research, 91, 90-98 (2014)].
9. Отчёт по проекту Фонда содействия развитию малых форм предприятий в научно-технической сфере НТИ-50323 «Разработка платформы на базе микрофлюидного чипа для исследования ангиогенеза ex vivo» (Заключительный). – 20.11.2020.

УДК 004.8

МОДЕЛЬ ПРЕДИКЦИИ ОТВЕТА ЯИЧНИКОВ НА ГОРМОНАЛЬНУЮ ТЕРАПИЮ ПРИ ЛЕЧЕНИИ БЕСПЛОДИЯ НА ОСНОВЕ ГРАДИЕНТНОГО БУСТИНГА

***Тихаева К.Ю.**, к.м.н.*
e-mail: tikhaeva34@gmail.com
ООО «Аналитическая лаборатория Содас»,
ФГБОУ ВО ВолгГМУ МЗ РФ,
ООО «Клиника Семья»
Нестерова Н.С.
e-mail: sunflovertash@mail.ru
Томилов Е.В.
Соткин С.И.
ООО «Аналитическая лаборатория Содас»
***Сулима А.Н.**, д.м.н.*
e-mail: gsulima@yandex.ru
ФГАОУ ВО «КФУ имени В. И. Вернадского,
ООО «Медицинская клиника «Ваш Доктор»
***Балезин С.Л.**, к.м.н.*
e-mail: s.balezin@mail.ru
***Росюк Е.А.**, к.м.н.*
АО «Центр семейной медицины»

1. ВВЕДЕНИЕ

Количество получаемых ооцитов при лечении бесплодия методами вспомогательных репродуктивных технологий это количественная переменная, которая, является значимым предиктором шанса на наступление беременности. По количеству получаемых ооцитов ответ яичников может быть идентифицирован как оптимальный – 10–16 ооцитов, субоптимальный – 4-9 ооцитов, слабый – менее 4 ооцитов, а так же гиперответ более 16 ооцитов [1, 2].

Проблема слабого («бедного») ответа является актуальной проблемой репродуктивной медицины. Важным является тот факт, что слабый и субоптимальный ответ встречается не только у женщин со

сниженным овариальным резервом, но и в группе с нормальным и высоким овариальным резервом. И если в первом случае неудовлетворительный результат стимуляции яичников является прогнозируемым, то во втором случае – неожиданным. При этом известно, что в группе пациентов с неожиданно слабым ответом яичников на стимуляцию гонадотропинами, частота наступления беременности значительно ниже в сравнении с группой пациентов с ожидаемым слабым ответом яичников – 6-7% и 17-26% соответственно [3-5].

Таким образом, важно сравнивать прогноз количества ооцитов с фактическим количеством полученных ооцитов.

Помимо этого, важным является тот факт, что большинство систем прогноза количества ооцитов, направлено на расчет общего числа ооцитов, включая незрелые и некачественные или на расчет количества преовуляторных фолликулов. Необходима сложная система прогноза количества именно зрелых и качественных ооцитов, пригодных к оплодотворению. Эта система прогноза должна включать не только показатели овариального резерва, но и чувствительность фолликулов к гонадотропинам. Основой такой системы прогноза может являться предиктивная модель, включающая все многообразие влияющих факторов.

Таким образом, целью исследования являлась разработка базовой модели, которая была бы способна восстанавливать функциональную зависимость между переменными-предикторами и количеством зрелых ооцитов.

2. МАТЕРИАЛЫ И МЕТОДЫ

Для разработки модели были использованы ретроспективные данные о случаях стимуляции яичников и количестве полученных ооцитов. Всего включено 658 случаев стимуляции яичников гонадотрпинами.

В качестве метрики для оценки эффективности модели выбрана sMAPE (symmetric mean absolute percentage error, симметричная средняя абсолютная ошибка в процентах) [6]. Формула данной ошибки:

$$sMAPE = \frac{100\%}{n} \sum_{t=1}^{n} \frac{|F_t - A_t|}{(|A_t| - |F_t|)/2} ,$$

где:

- n – объём выборки;
- A_t – истинное значение отклика;
- F_t– значение отклика, возвращённое моделью.

Данная ошибка выбрана вследствие того, что:

- в имеющихся данных целевая переменная строго положительна;
- целевая переменная измерена в абсолютной шкале;
- *sMAPE* является интуитивно интерпретируемой метрикой, что улучшает понимание качества модели экспертами;
- *sMAPE* является симметричной метрикой, она равноценно корректирует модель как в сторону слишком высоких предсказаний, так и слишком низких.

Таким образом, определяется целевую точность итоговой модели как 85%, что выражается через $1 - sMAPE$.

Для увеличения точности оценки применяется также средняя абсолютная ошибка (*MAE*) [7]. Количество зрелых ооцитов в данных изменяется в пределах от 1 до 20 и на данном этапе целевое значение *MAE* составляет 3 зрелых ооцита.

Формально средняя абсолютная ошибка описывается следующим образом:

$$MAE = \sum_{t=1}^{n} \frac{|y_i - x_i|}{n},$$

где:

- n – объём выборки;
- y_i – истинное значение отклика;
- x_i – значение отклика, возвращённое моделью.

Таким образом, формальные цели разработки заключаются в следующем:

- модель с *MAE* <= 3;
- модель с *sMAPE* <= 0.15.

Для технической реализации моделирования был выбран язык *R* 4.0.3.

После формирования подмножества наиболее информативных переменных методом *Boruta* в выборке присутствовал 31 предиктор и 1 целевая переменная.

В качестве целевой переменной выступало количество зрелых ооцитов.

В качестве потенциальных предикторов были использованы данные об общем анамнезе, репродуктивном анамнезе, наличии соматических и репродуктивных заболеваний, объективные данные антропометрии, лабораторные и инструментальные показатели овариального резерва, содержание в крови стероидных и нестероидных гормонов, показатели общего анализа крови, биохимического

исследования крови отражающего углеводный, жировой и белковый гомеостаз.

Переменные, которые использовались для разработки модели относились к категориальными или количественными.

Все категориальные переменные для моделей линейной регрессии, дерева решений, случайного леса, k-ближайших соседей преобразовывались в бинарный метод one-hot. Для catboost использовался встроенный метод target encoding.

Все пропущенные значения в переменных были восстановлены методом импутации вспомогательными моделями. Необходимо отметить, что предварительно из множества переменных удалялись те переменные, процент пропущенных значений которых более 15. Таким образом понижалась вероятность неправдоподобного восстановления пропущенных значений.

Все количественные переменные нормализовались путём Z-масштабирования.

$$Z = \frac{x - \mu}{\sigma} \ ,$$

где:

- x – значение переменной;
- μ – среднее значение данной переменной;
- σ – стандартное отклонение переменной.

Всего после очистки данных от невалидных объектов в выборке присутствует на данный момент 658 строк данных.

Базовую схему взаимосвязи можно представить следующим образом:

$$\text{количество зрелых ооцитов} \sim a + b_1 * x_1 + b_2 * x_2 + \cdots + b_n * \\ * x_n \ + \varepsilon,$$

где «количество зрелых ооцитов» – целевая переменная, a – константа, b_n – коэффициенты, x_n – соответствующие значения переменных, ε – «шум», случайная ошибка, которая неизбежно присутствует в данных.

Датасет был разделён на тренировочную и тестовую выборки в соотношении 80%/20% (529/129). На первом подмножестве подбирались гиперпараметры методом кроссвалидации, второе подмножество оставлено для тестирования модели.

Модели сравниваются по метрикам *MAE* и *sMAPE*, после чего выбиралась лучшая.

3. МОДЕЛИРОВАНИЕ

3.1. Модель линейной регрессии

Для базовой модели использовался метод шаговой регрессии по критерию Акаике: Акаике: $AIC = 2k - 2ln(L)$, где k – число параметров в статистической модели, L – максимизированное значение функции правдоподобия модели, после чего из полученных моделей выбирается та, в которой AIC имел наименьшее значение (таблица 1) [8].

Таблица 1. Метрики модели линейной регрессии

Метрика	Значение
MAE	2,40
sMAPE	38,5%
1 – *sMAPE*	61,5%

MAE базовой модели меньше 3. Ожидалось, что альтернативные модели должны иметь меньшую ошибку для того, чтобы считаться более эффективными, чем базовая наивная модель.

3.2. Модель дерева регрессии

Модель дерева регрессии не отличается значимо от базовой модели, имея даже больший *sMAPE* (таблица 2) [9].

Таблица 2. Метрики модели дерева регрессии

Метрика	Значение
MAE	2,40
sMAPE	38,7%
1 – *sMAPE*	61,3%

3.3. Модель случайного леса

Модель на основе случайного леса имеет низкую интерпретируемость, но позволяет оценивать значимость отдельных признаков в модели, а также соединяет в себе все преимущества деревьев решений, компенсируя их недостатки (таблица 3) [11].

Таблица 3. Метрики модели случайного леса

Метрика	Значение
MAE	1,7
SMAPE	29,4%
1 – SMAPE	70,6%

Как видно из таблицы результатов, полученная модель в среднем ошибается на 1,7 ооцита, что является существенным улучшением по сравнению с деревом регрессии.

sMAPE снизилась до 29,4%, то есть на 9,3%.

3.4. Модель k-ближайших соседей

В качестве метрики была выбрана стандартная евклидова метрика (таблица 4) [11].

Количество соседей K было выбрано равным 13.

Таблица 4. Метрики модели k-ближайших соседей

Метрика	Значение
MAE	2,3
sMAPE	37,4%
1 – *sMAPE*	62,6%

Метрики данной модели ближе к моделям линейной регрессии и дерева решений, что даёт основания отнести их в группу слабо пригодных для использования.

3.5. Модель catboost

Модель catboost представляет из себя разработанный компанией Яндекс алгоритм градиентного бустинга на основе деревьев решений (таблица 5) [12].

Таблица 5. Метрики модели catboost

Метрика	Значение
MAE	1,09
sMAPE	17,7%
1 – *sMAPE*	82,3%

Данная модель имеет наивысшее качество из всех ранее исследованных. *MAE* составляет лишь 1,08 ооцита, а *sMAPE* 17,7%. Таким образом мы принимаем catboost в качестве рабочей модели по следующим причинам:

Наиболее низкая ошибка, вплотную приближающаяся к целевому уровню;

Оптимизированная библиотека позволяет создавать наиболее быструю имплементацию модели для использования в реальной практике.

4. ЗАКЛЮЧЕНИЕ

В целях разработки модели предикции ответа яичников на гормональую стимуляцию при лечении бесплодия методами вспомогательных репродуктивных технологий были использованы данные 700 случаев лечения бесплодия методами ВРТ, по каждому из случаев в анонимизированном и обезличеном виде собраны данные касающиеся анамнеза, статуса пациента, овариального резерва, метрики

лечения и результат стимуляции. Построение системы прогнозирования осуществлено с помощью языка *R* 4.0.3. Моделирование осуществлялось на основе алгоритмов линейной регрессии, регрессионных деревьев, случайного леса, метода k-случайных соседей, градиентного бустинга.

Модель предикции, разработанная на основе методов случайного леса и варианта градиентного бустинга «catboost» позволяет прогнозировать количество зрелых ооцитов со средней ошибкой (*MAE*) в 1,09 зрелого ооцита и с точностью в 82,3% (*sMAPE*).

Прогностическое моделирование может решать задачу прогнозирования ответа яичников при лечении бесплодия методами ВРТ. Наиболее эффективным для данной задачи является вариант метода градиентного бустинга «catboost» в силу механизма его работы, заключающегося в построении множества последовательно корректирующих друг друга деревьев решений.

Литература

1. Li Y, Li X, Yang X, Cai S, Lu G, Lin G, Humaidan P, Gong F. Cumulative Live Birth Rates in Low Prognosis Patients According to the POSEIDON Criteria: An Analysis of 26,697 Cycles of in vitro Fertilization/Intracytoplasmic Sperm Injection. Front Endocrinol (Lausanne). – 2019 Sep 19;10:642. doi: 10.3389/fendo.2019.00642. PMID: 31608011; PMCID: PMC6761219.
2. Loutradis, Dimitris et al. “FSH receptor gene polymorphisms have a role for different ovarian response to stimulation in patients entering IVF/ICSI-ET programs.” Journal of assisted reproduction and genetics vol. 23,4 (2006): pp. 177-84. doi:10.1007/s10815-005-9015-z.
3. van der Gaast MH, Eijkemans MJ, van der Net JB, de Boer EJ, Burger CW, van Leeuwen FE, Fauser BC, Macklon NS. Optimum number of oocytes for a successful first IVF treatment cycle. Reprod Biomed Online. – 2006 Oct;13(4):476-80. doi: 10.1016/s1472-6483(10)60633-5. PMID: 17007663.
4. Wu CX, Zhang T, Shu L, Huang J, Diao FY, Ding W, Gao Y, Wang W, Mao YD, Cui YG, Liu JY. [Cumulative live birth rates per oocytes retrieved cycle: evaluation of clinical outcomes of IVF/ICSI]. Zhonghua Fu Chan Ke Za Zhi. 2018 Mar 25;53(3):160-166. Chinese. doi: 10.3760/cma.j.issn.0529-567X.2018.03.004. PMID: 29609229.
5. Drakopoulos P, Blockeel C, Stoop D, Camus M, de Vos M, Tournaye H, Polyzos NP. Conventional ovarian stimulation and single embryo transfer for IVF/ICSI. How many oocytes do we need to maximize cumulative live birth rates after utilization of all fresh and frozen embryos? Hum Reprod. 2016 Feb;31(2):370-6. doi: 10.1093/humrep/dev316. Epub 2016 Jan 2. PMID: 26724797.
6. Flores, B. E. (1986) "A pragmatic view of accuracy measurement in forecasting", Omega (Oxford), 14(2), 93–98. doi:10.1016/0305-0483(86)90013-7.
7. Willmott, Cort J.; Matsuura, Kenji (December 19, 2005). "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing

average model performance". Climate Research. 30: pp. 79-82. doi:10.3354/cr030079.
8. Rencher, Alvin C.; Christensen, William F. (2012), "Chapter 10, Multivariate regression – Section 10.1, Introduction", Methods of Multivariate Analysis, Wiley Series in Probability and Statistics, 709 (3rd ed.), John Wiley & Sons, p. 19, ISBN 9781118391679.
9. Quinlan, J. R. (1986). "Induction of decision trees" (PDF). Machine Learning. 1: pp. 81-106. doi:10.1007/BF00116251. S2CID 189902138.
10. Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome (2008). The Elements of Statistical Learning (2nd ed.). Springer. ISBN 0-387-95284-5.
11. Altman, Naomi S. (1992). "An introduction to kernel and nearest-neighbor nonparametric regression" (PDF). The American Statistician. 46 (3): pp. 175-185. doi:10.1080/00031305.1992.10475879. hdl:1813/31637.
12. Dorogush, Anna Veronika; Ershov, Vasily; Gulin, Andrey (2018-10-24). "CatBoost: gradient boosting with categorical features support". arXiv:1810.11363 [cs.LG].

СЕКЦИЯ 10 **ИНТЕЛЛЕКТУАЛЬНЫЕ ПРОИЗВОДСТВА, УПРАВЛЕНИЕ ЖИЗНЕННЫМ ЦИКЛОМ СИСТЕМ, ТЕХНОЛОГИИ «ИНДУСТРИИ 4.0»**

УДК 004.8

ИНТЕГРАЦИЯ ИНТЕРНЕТА ЗНАНИЙ И ИНТЕРНЕТА ВЕЩЕЙ ДЛЯ СОЗДАНИЯ ИНДУСТРИИ 5.0

***Евгенев Г.Б.**, д.т.н.*
e-mail: g.evgenev@mail.ru
МГТУ им. Н.Э. Баумана

1. ВВЕДЕНИЕ

В настоящее время происходит четвертая промышленная революция (4ПР), для которой используется термин «Индустрия 4.0». В развитии производительных сил человечества принято выделять ряд этапов. Этапы 1.0 и 2.0 были связаны с развитием сельского хозяйства. Третий этап - это промышленная революция, начало массового использования промышленной продукции. Этап 4.0 охватывает период общей компьютеризации, включая технологическое оборудование и этапы производства продукции. Сущность новой революции заключается в том, что материальный мир связан с виртуальным, в результате чего рождаются новые киберфизические комплексы, объединенные в одну цифровую экосистему. Фаза 5.0 должна охватывать все стадии жизненного цикла продукта (ЖЦ). В статье [1] приводится интервью с представителем японской компании о развитии

цифрового общества. Отмечается, что Индустрия 4.0 является лишь частью процесса перехода к Обществу 5.0, который станет этапом, следующим за информационным обществом [2]. Общество 5.0 представляет собой оптимизацию ресурсов общества в целом за счет интеграции физического и киберпространства.

В материале [3] рассматривается концепция интеграции Индустрии 5.0 в трех основных областях: управление проектами и цифровая экономика, развитие личности и компетенций, промышленный и домашний Интернет. Индустрия 5.0 описывается как информационная система, обеспечивающая планирование, регистрацию, обработку и отражение информации, по крайней мере, по трем измерениям элементов влияния: в областях, которые поддерживают и создают продукт, на этапах реализации деятельности и в составе элементов управленческой деятельности.

В основе 4ПР лежит Интернет вещей, IoT - концепция вычислительной сети физических объектов («вещей»), оснащенных встроенными технологиями для взаимодействия друг с другом или с внешней средой, учитывающая организацию таких объектов в виде сети. Интернет вещей – это технология, которая может развивать экономические и социальные процессы, исключающие из части действий и операций потребность в человеческом участии. Индустрия 5.0 должна охватывать автоматизацию абсолютно всех этапов и процессов, включая цифровое проектирование изделия, создание его виртуальной копии, сотрудничество инженеров-конструкторов и технологов в едином цифровом пространстве и т.д., когда изделия еще не вещи, но существуют в виртуальном мире в виде информационных моделей.

Цифровая революция должна позволить непрограммируемому носителю знаний вводить их в компьютер без посредников. Это стало возможным с использованием методологии экспертного программирования [4]. В этой методологии знания описываются языком деловой прозы, который максимально приближен к литературному, но настолько формализован, что появляется возможность автоматически генерировать программное обеспечение, соответствующее исходным текстам.

Интеграцию промышленных и цифровых революций следует рассматривать в совокупности двух миров: виртуального мира, реализуемого Интернетом знаний (IoK), и реального мира, реализуемого Интернетом вещей (IoT). Интернет знаний построен на онтологической основе, корневым объектом которой является метаонтология. С точки зрения искусственного интеллекта (ИИ)

онтология – это явная спецификация концептуализации знания. Метаонтология оперирует общими концепциями и отношениями, которые не зависят от конкретной предметной области. Она должна содержать классы и отношения, необходимые для описания онтологий нижнего уровня: предметной онтологии, онтологии задач и онтологии оптимизации (рис.1).

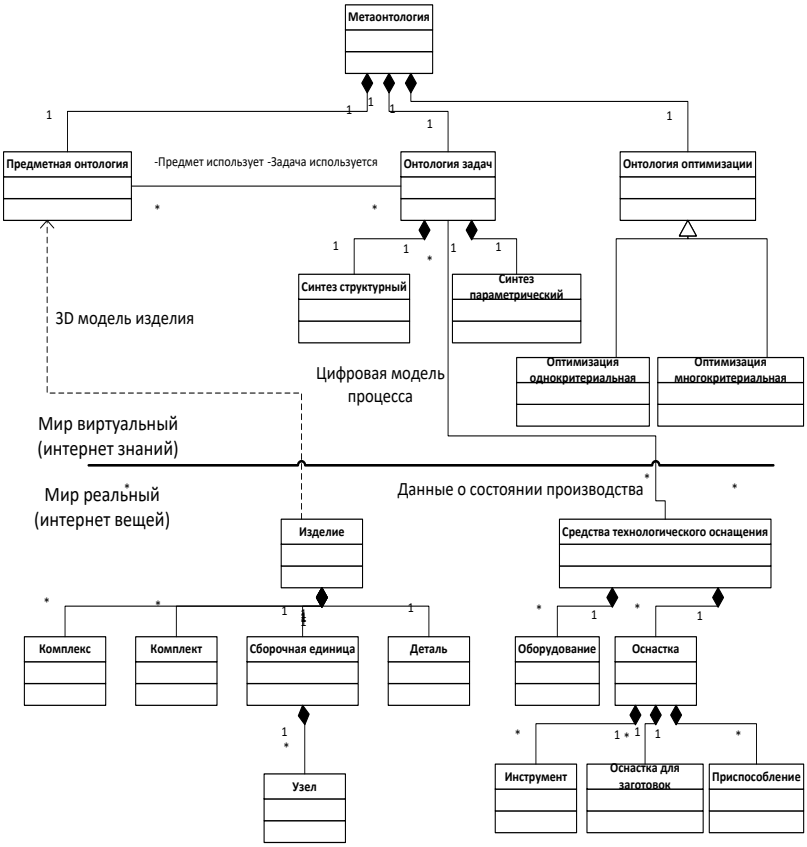


Рис. 1. Интегрированная структура Интернета знаний и Интернета вещей

Предметная онтология состоит из иерархии понятий, их определений и атрибутов, а также связанных аксиом и правил вывода. Предметная онтология, основанная на использовании задач, обеспечивает на основе технических спецификаций создание трехмерных моделей изделий, адекватно представляющих изделия в виртуальном мире и

удовлетворяющих требованиям задач. В онтологию задач входят задачи структурного и параметрического синтеза моделей изделий и процессов. С его помощью обеспечивается генерация цифровых моделей процессов и производств [4]. Онтология оптимизации включает компоненты однокритериальной и многокритериальной оптимизации.

В описанном подходе мир вещей состоит из изделий, изготовленных с использованием 3D-принтеров или станков и роботов, а также оборудования как части технологического средств конкретного производства. Производственный процесс регулируется цифровыми моделями, созданными с использованием соответствующих систем, включающих необходимые задачи.

Проектирование процессов осуществляется с использованием производственных данных из Интернета вещей.

2. ЖИЗНЕННЫЙ ЦИКЛ МАШИНОСТРОИТЕЛЬНЫХ ИЗДЕЛИЙ

На рис.2 показаны жизненный цикл (ЖЦ) изделий и интегрированная структура Интернета знаний и Интернета вещей.

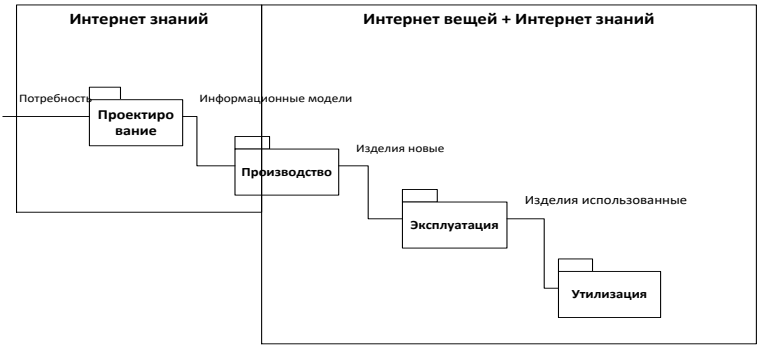


Рис.2. Жизненный цикл продукции и интегрированная структура Интернета знаний и Интернета вещей

ЖЦ изделия содержит основные этапы: проектирование, производство, эксплуатация и утилизация. На этапах инженерного и технологического проектирования изделия и процессы существуют в виртуальном мире, в связи, с чем используется Интернет знаний. В процессе производства изделия предстают как вещи, поэтому существует возможность совместного использования Интернета вещей и Интернета знаний. На этапе проектирования изделие представляется в виде формальной системы с оформлением соответствующей проектной документации.

3. КОНЦЕПТУАЛЬНЫЕ ОСНОВЫ ИНТЕРНЕТА ЗНАНИЙ

Семантика Интернета знаний определяется мета-онтологией (см. рис.3).

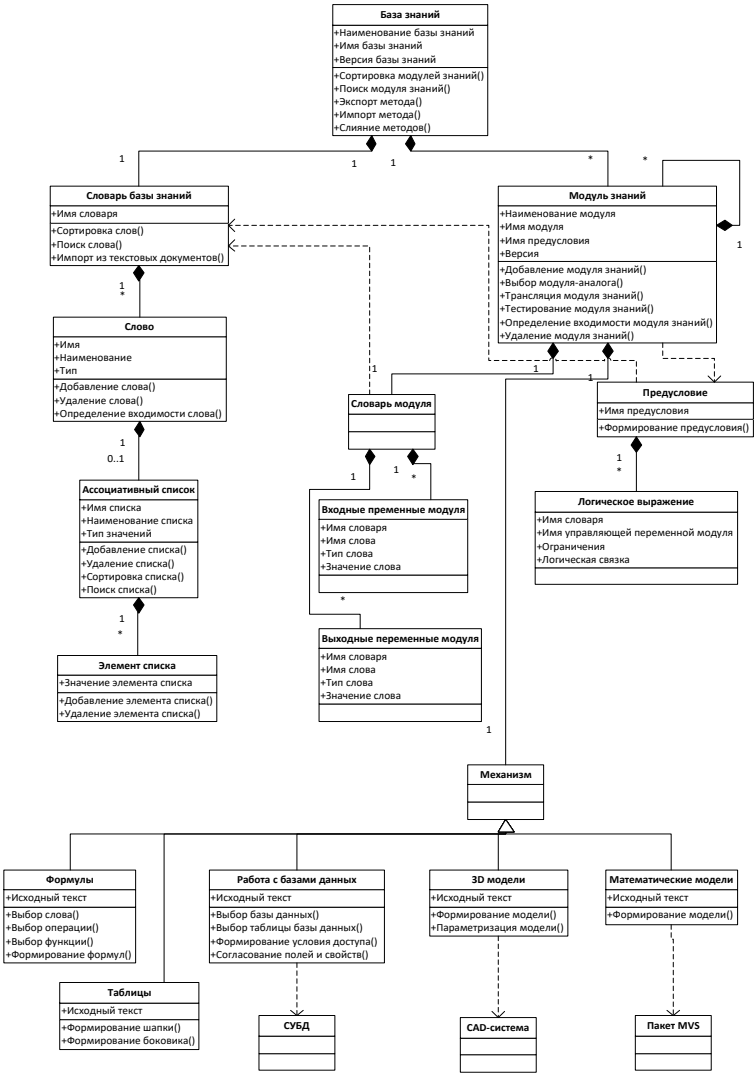


Рис3. Метаонтология Интернета знаний

Корневым метаобъектом в данном случае является База знаний (БЗ), у которой есть собственное имя, имя-идентификатор и версия. Методы, подключенные к базе знаний, позволяют сортировать и искать модули знаний, из которых она состоит. Подмножество модулей знаний, объединенных в семантическую сеть, представляет собой метод, который можно экспортировать в другие базы знаний, а также импортировать и объединить в единую семантическую сеть.

Согласно технологии Wiki, БЗ – это страница. В этом случае БЗ можно рассматривать как модуль внутри интегрированной БЗ, которая в данном случае состоит из набора страниц. Структурными компонентами БЗ являются словарь БЗ и множество модулей.

Возможность замены словаря БЗ позволяет удовлетворить концептуальное требование вики-систем: использование любых языков мира.

Второй структурный компонент БЗ - это набор модулей знаний. Модули знаний выполняют основное функциональное назначение БЗ – преобразование текущего состояния данных с целью получения новых объектов, соответствующих целям проектирования.

4. МНОГОАГЕНТНЫЕ МЕТОДЫ СОЗДАНИЯ БЗ

Важнейшей методологической основой реализации концепций интеграции, интеллектуализации и индивидуализации является теория многоагентных систем (МАС).

Рассмотрим применение МАС на примере проектирования и управления интеллектуальными системами (ИС) [4]. Агентами в машиностроении агентами выступают системы, комплексы, сборочные единицы, детали и их элементы, а также технологические процессы (ТП) их изготовления.

Обобщенная модель класса инженерных агентов показана на рис.4.

Любой агент – это открытая система, помещенная в определенную среду. В случае САПР эта среда представляет собой проект, созданный в базах данных (БД). При этом рекомендуется использовать БД объектного типа для представления модели изделия (внутренняя среда) и реляционные БД для поиска стандартных и приобретаемых изделий, свойств материала и другой информации (внешняя среда). Внешняя среда обычно является сетевой.

Процессор агента формируется его методами, обеспечивающими интеграцию и обработку разнородных данных, выработку соответствующих ответов на информацию о состоянии среды (проекта) и принятия решений о выполнении определенных действий. Как правило, процессор определяет поведение агента, которое можно

наблюдать, например, в графическом окне, в котором отображаются сгенерированные объекты.

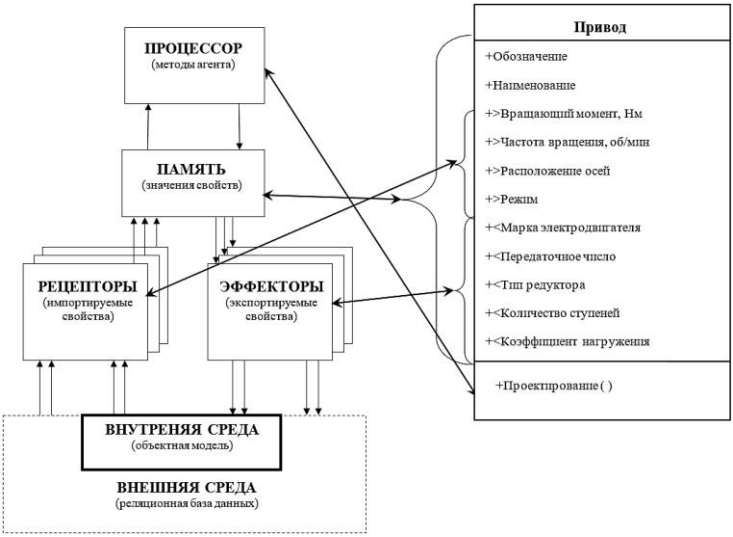


Рис.4. Обобщенная модель класса инженерных агентов

При создании ИС первым этапом проектирования является формирование номенклатуры и организационной структуры агентов. Номенклатура агентов определена в системном словаре. Организационная структура агентов формируется в виде построения диаграммы классов UML. Диаграммы классов эквивалентны чертежам сборки для разработки программного обеспечения.

Предметная онтология состоит из иерархии понятий, их определений и атрибутов, а также связанных аксиом и правил вывода. Предметная онтология, основанная на использовании задач, обеспечивает на основе технических спецификаций создание трехмерных моделей изделий, адекватно представляющих изделия в виртуальном мире и удовлетворяющих требованиям задач. В онтологию задач входят задачи структурного и параметрического синтеза моделей изделий и процессов. С его помощью обеспечивается генерация цифровых моделей процессов и производств [4].

В описанном подходе мир вещей состоит из изделий, изготовленных с использованием 3D-принтеров или станков и роботов, а также оборудования как части технологического средств конкретного производства. Производственный процесс регулируется цифровыми

моделями, созданными с использованием средств имитации, включающих необходимые задачи.

Проектирование процессов осуществляется с использованием производственных данных из Интернета вещей.

5. МАШИНЫ ПОЛУАВТОМАТИЧЕСКОГО КОНСТРУИРОВАНИЯ ИЗДЕЛИЙ

Есть два основных способа использования технологий искусственного интеллекта в автоматизации проектирования: встраивание и развитие. Первый подход используется в ряде САПР, например, SolidWorks. Недостатком этого подхода является то, что инженеры, не являющиеся проектировщиками, например, технологи, не могут использовать интеллектуальные возможности таких систем. В статье предлагается использовать подход надстройки, который позволяет использовать методы искусственного интеллекта на всех этапах создания продуктов ЖЦ.

Основная статическая диаграмма UML - это диаграмма классов. Представление объектного класса в нем не имеет функциональности, необходимой для описания MAS.

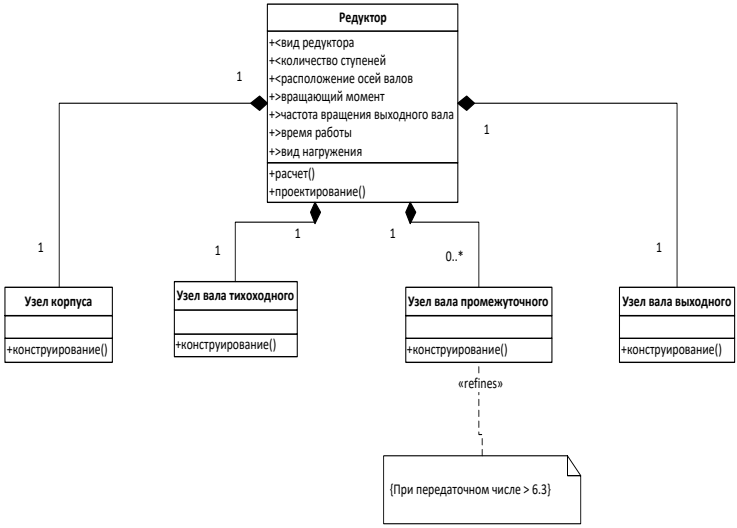


Рис.5. Модифицированная диаграмма классов для MAC

На рис.5 показана модифицированная диаграмма классов для MAS. На ней на примере описания редуктора входные свойства обозначены символом>, а исходящие, полученные в результате проектирования,

символом <. Предварительное условие для узла промежуточного вала представлено в виде ограничения, в котором написано, что этот узел появляется, когда передаточное число больше 6,3.

Что касается динамических свойств, методы классов объектов представлены в виде диаграмм состояний, в которых предусловия включены внутри модулей.

На рис.6 дана схема функционирования полуавтоматической системы в процессе проектирования одноступенчатой коробки передач.

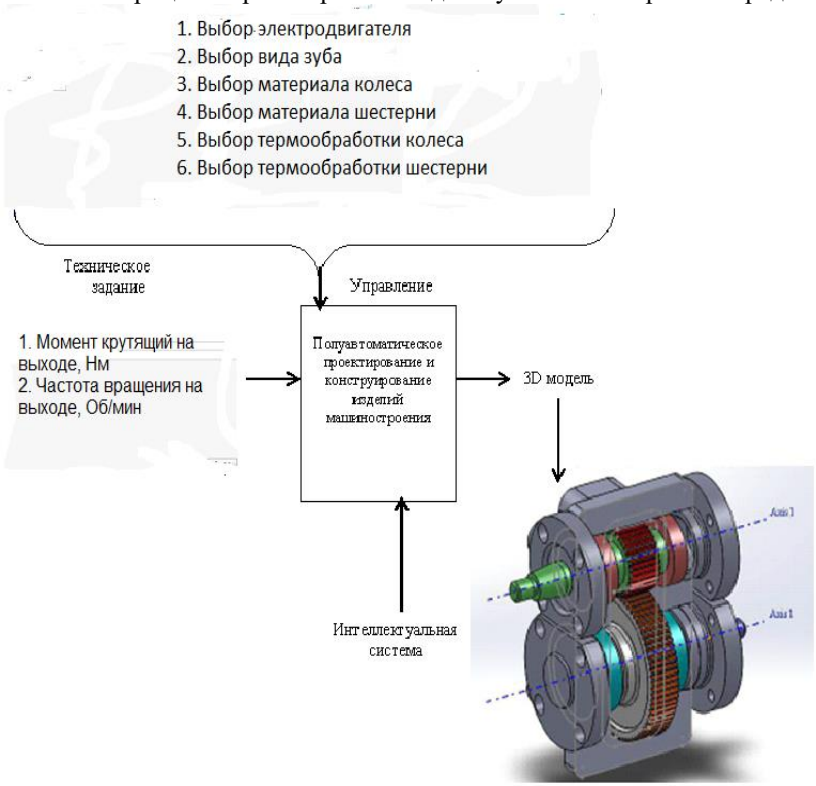


Рис.6. Схема функционирования полуавтоматической системы

На входе в систему имеется техническое задание (ТЗ), которое формируется с помощью соответствующего окна. Есть возможность формирования ТЗ разного состава. На рис.6 показан вариант ТЗ, включающий значения крутящего момента на выходном валу в Нм, выходную скорость в об/мин, ресурс изделия в часах и количество изделий.

Процесс проектирования контролируется ответами на вопросы с заданной номенклатурой вариантов значений, отображаемых БЗ. Ответив на все вопросы, ИС производит все необходимые расчеты и формирует документацию и 3D-модели. Если в результате ответов на вопросы нарушаются какие-либо ограничения, заложенные в БЗ, ИС выдает отказ с объяснением причин.

6. ЗАКЛЮЧЕНИЕ

В основе 4ПР лежит Интернет вещей. Однако Индустрия 5.0 должна охватывать автоматизацию всех этапов и процессов, когда продукты еще не являются вещами, а существуют в виртуальном мире в виде информационных моделей.

Накануне выхода Индустрии 5.0 необходимо рассмотреть в совокупности два мира: виртуальный мир, реализованный посредством Интернета знаний, и реальный мир, реализованный посредством Интернета вещей. Целесообразно построить Интернет знаний на основе онтологической методологии, которая основана на мире виртуальных агентов, представляющих функции объекта, способных генерировать новые данные на основе доступной информации. Попытка такой интеграции, охватывающей все этапы продуктов ЖЦ, представлена в этой статье.

Литература

1. Noritsugu W. Society 5.0: Mitsubishi Electric's View// Economic Strategies. – 2017. –№4. – pp. 2-11.
2. Ржевский Г.А. Самоорганизация в социальных системах// Онтология проектирования. – 2014. – №4 (14). – С.8-17.
3. Concept of Industry 5.0 - http://industry5.ru/koncept.
4. Евгенев, Г. Основы автоматизации технологических процессов и производств. Т.1: Информационные модели. Т.2: Методы проектирования и управления. – М.: Изд-во МГТУ им.Н.Э.Баумана. 2015. – Т.1: – 441 с. Т.2: – 479 с.

УДК 621.941.2

ПРИНЦИПЫ СОЗДАНИЯ АВТОМАТИЗИРОВАННЫХ СИСТЕМ УПРАВЛЕНИЯ ПРОЦЕССАМИ В МАШИНОСТРОЕНИИ

Бурдо Г.Б., *д.т.н.*
e-mail: gbtms@yandex.ru
ТвГТУ

1. ВВЕДЕНИЕ

Машиностроение и приборостроение относятся к высокотехнологичной сфере производства, и обладает рядом особенностей, обусловленных следующими причинами:

а) особыми (температура, давление, агрессивные и абразивные среды, работа на глубине) и изменяющимися условиями эксплуатации приборов;
б) сложностью конструкции механики изделий вследствие их ограниченных габаритов;
в) сложностью электроники изделий и алгоритмов её работы;
г) не вполне формальными критериями качества изделия и отсутствием детального технического задания;
д) желанием заказчиков иметь адаптированные под конкретные условия эксплуатации изделия;
е) сжатыми сроками на разработку и изготовление;
ж) необходимостью выполнения поисковых и опытно-конструкторских работ;
з) необходимостью реализации при изготовлении новых технологий;
и) небольшие партии выпуска изделий по неизменным чертежам.

Все эти факторы, к сожалению, приводят к большим временным и денежным затратам при реализации проектов по созданию новых приборах, при этом не всегда обеспечивая их необходимое качество. Следовательно, страдает и их (приборов) конкурентоспособность.

Поэтому, можно констатировать, что разработка подходов к созданию автоматизированных систем систем управления процессами в автоматизированных производственных системах в указанных отраслях востребовано временем.

2. КРИТЕРИИ РАБОТЫ ПРОИЗВОДСТВЕННОЙ СИСТЕМЫ

Дадим некоторые пояснения относительно термина «производственная система» (далее ПС). Наиболее часто в технической литературе под ним подразумевают производственные участки (цеха) машиностроительных или приборостроительных предприятий, или их совокупность. На наш взгляд такой подход не вполне приемлем для рассматриваемых предприятий, относящихся к многономенклатурному производству ввиду того, что искусственно разрывает цикл проекта по созданию нового изделия.

Поэтому, подход, включающий в ПС конструкторские и технологические службы, производственные мощности, испытательные, научные, а также управляющие (в т.ч. и самой производственной системой) подразделения, представляется более продуктивным.

В этом случае появляется возможность более оперативного реагирования на изменение внешней ситуации, организовывать итерационные процедуры по выработке приемлемого решения с представителями различных служб. Очевидно, что основные функции предприятия (функции, связанные с выпуском изделий), реализуются именно в производственной системе, которую в этом случае можно охарактеризовать как организационно-технологическую.

Рассмотрим критерии, в соответствии с которыми должна работать современная ПС. К сожалению, в последнее время акцент в производственных организационно-технологических системах сместился в сторону экономики, в сторону максимального извлечения прибыли, причем вопреки здравому смыслу, и не вполне отвечает требованию конкурентоспособности.

Если говорить о цели работы современных развивающихся организационно-технологических производственных систем применительно к нашей проблематике – это выпуск конкурентоспособных изделий. Конкурентоспособность достигается обеспечением требуемого (заказчиком, тенденцией развития отрасли) качества изделий (но не превышения его), их выпуском в заданные сроки, и при заданном уровне производственных затрат (временных и денежных). Именно требуемого качества, а не максимально возможного. В таком ракурсе, управление качеством становится не довеском к какой-либо системе управления организацией, а именно основополагающий процесс организации. Весьма эффективным инструментарием такого управления могут явиться автоматизированные системы.

3. ЖИЗНЕННЫЙ ЦИКЛ ИЗДЕЛИЯ И УПРАВЛЕНИЕ ПРОЦЕССАМИ В ПРОИЗВОДСТВЕННЫХ СИСТЕМАХ

Анализ жизненного цикла изделия (ЖЦИ) составляет методологическую основу [1, 2] анализа процессов деятельности организации, так как представляет собой инструмент управления сведениями об изделии, его состоянии, уровне качества.

Зачастую, в наукоемком машиностроении предприятия являются одновременно разработчиками и изготовителями своих приборов, а этапы эксплуатации, ремонта и утилизации выполняется силами потребителей изделий. Поэтому, для нашего рассмотрения наиболее важны этапы проектирования и конструирования, изготовления и испытания. Однако, и последующие этапы интересны для нас с точки получения в виде обратной связи информации об эксплуатации изделий. Наукоемкая продукция имеет некую специфичность жизненного цикла. В первую очередь, это обязательное наличие стадии научно-исследовательских работ (НИР) в этапе конструирования и проектирования, наличие большого числа итерационных процедур поиска технического решения, значительный объем испытаний. При этом, от последующих этапов к предыдущим реализуются обратные связи.

Следующим шагом может явиться определение процессов, задействованных в проектах по созданию изделий. Первый вопрос, который требуется определить, чем мы будем управлять? Мы выяснили, что для современных производственных систем – это управление качеством, поэтому, выполняя этапы жизненного цикла, их завершение либо продолжение мы должны определять исходя из предполагаемого достигнутого уровня.

Рассмотрим модель процессов обеспечения качества (рис.1).

На рисунке использованы следующие обозначения: АСУП – автоматизированная система управления организацией (предприятием); АСУ ТППиИ - автоматизированная системы управления технической подготовкой производства, изготовления и испытания опытного изделия, то есть на этапах жизненного цикла изделия; А - множество критериев показателей качества изделия и (или) качества выполнения работ на этапах; D_0 – блок задания, формирует множество технических заданий ТЗ; T_1 – техническое задание на проведение НИОКР; T_2 – техническое задание на проведение конструкторской подготовки производства (КПП); T_3 – техническое задание на проведение технологической подготовки производства (ТхПП); T_4 – техническое задание на проведение планирования производства (ПП) по реализации технологических процессов в производственной системе; T_5 –

техническое задание на проведение испытаний изделия; R_0 – оператор, управляет работой на всех этапах создания образца наукоемкой продукции;R_0^1 – управление оператором P_1 (НИОКР); R_0^2 – управление оператором P_2 (КПП); R_0^3 – управление оператором P_3 (ТхПП); R_0^4 – управление оператором P_4 (ПП); R_0^5 – управление оператором P_5 (испытания изделия, И); R_0^6 – управление работой оператора Q_2 (контроль качества работ, КК); R_0^7– оператор по координация управления с АСУП; U_1– множество параметров, содержащихся в отчетной документации по НИОКР, необходимых для проведения КПП; U_2^1 – множество параметров, содержащихся в конструкторской документации и необходимых для ТхПП; ΔU_2^1 – множество параметров, определяющих требования оператора P_4 по повышению уровня технологичности изделия; U_2^2 – множество параметров, содержащихся в конструкторской документации и необходимых для изготовления изделий в технологической системе, контроля и испытаний; U_3^1 – множество параметров технологических процессов по трудозатратам и станкозатратам (трудозатратам); оператор Q_1 – изготовление; U_3^2 – множество параметров, определяющих реализацию технологических процессов на рабочих местах (технологические процессы, инструкции, управляющие программы для станков с ЧПУ); S_1^2 – множество параметров, характеризующих состояние производственной системы (загрузка рабочих мест, заказы в исполнении, ход выполнения планов и т.д.); U_4– множество параметров, характеризующих планы различных уровней по выпуску образца продукции; ΔS_1^1– множество параметров – сведения об изделиях, не прошедших контроль и нуждающихся в доработке (переделке);ΔU_4– множество параметров, характеризующих диспетчирование работ по данному изделию; S_2 – материальная связь – изделие (образец продукции); Оператор Q_1 – изготовление опытного образца изделия; M – материалы и покупные изделия; S_1^1 – изготовленные образцы продукции и сведения о их готовности; оператор Q_2 – контроль деталей, узлов и изделий целиком. Модель может быть использована для выявления и описания функциональных процессов управления, что может быть сделано в следующей последовательности.

В первую очередь, следует выявить, какие действия, выполняемые на протяжении ЖЦИ в рассматриваемой ПС влияют на качество предполагаемого к выпуску изделия (обозначим их как действия группы 1). На рис.1 – это этапы - научно-исследовательская работа, конструкторская подготовка производства, технологическая подготовка производства, изготовление (с планированием, как управление изготовлением и контролем), и испытания. Понятно, что

этими процессами надо управлять, появляется верхний уровень управления (определение технического задания, ТЗ, и собственно управление этапами ЖЦИ). Реализация технологических процессов идет в производственных подразделениях, поэтому появляется низший уровень – уровень АСУ ТП. Таким образом, процессы первой группы реализуются на 4 уровнях. Первый уровень – автоматизированная система управления организацией (предприятием), АСУП, координация всех работ в рамках предприятия. Второй уровень – уровень управления качеством на этапах ЖЦИ. Третий уровень – уровень реализации процессов на этапах ЖЦИ. Четвертый уровень – уровень управления технологическими процессами. Процессы на этих 4 уровнях являются базовыми для предприятия.

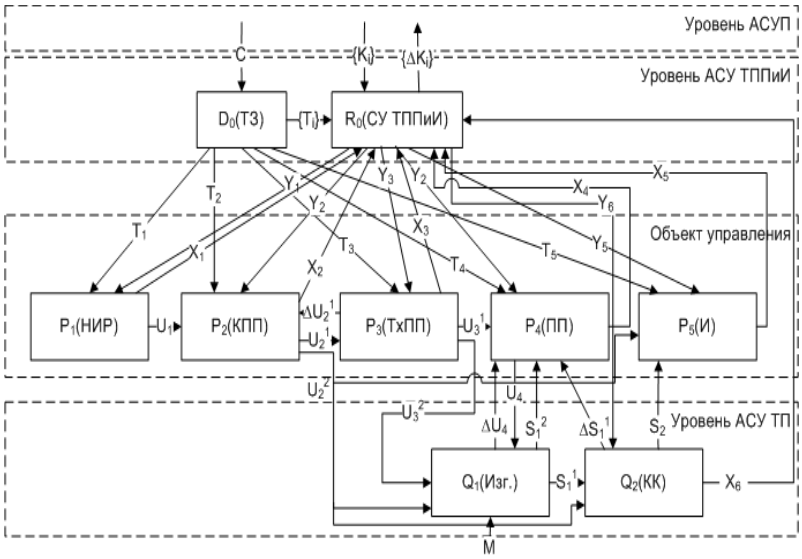


Рис.1. Структурно-функциональная модель автоматизированных процессов обеспечения качества

Процессы 1 вида не могут существовать сами по себе, поэтому следует выявить действия (действия группы 2), непосредственно не влияющие на качество изделия, но необходимые для выполнения действий вида 1. К примеру, это транспортные работы, документооборот, управление финансами, заключение договоров и контрактов и т.д. Следовательно, совокупность этих действий и будет составлять суть процессов, реализуемых на предприятии. Далее необходимо будет продетализировать процессы, разбив их на операции.

Каждая операция должна иметь законченный функциональный смысл, включать выполнение одной функции в рамках процесса. Так же важно найти «золотую середину» между концентрацией и дифференциацией действий в операции, исходя из квалификации исполнителей и возможностью принятия ими самостоятельного решения. Далее определяют контрольные места в процессе, в которых выполняется оценка качества.

Затем следует привязать процессы вида 2 к процессам вида 1 по времени. Привязка по времени процессов вида 1 идет друг к другу.

Немаловажным является и определение ответственного за тот либо иной процесс на каждом их уровне, так как в них включены участники из разных (по административной структуре) подразделений. Сложность в том, что участвующие в нем (процессе) разные подразделения организации уже имеют своих руководителей. Здесь получается логичное, но, к сожалению, не всегда принимаемое, распределение руководства по функциональному (не административному) признаку. Руководитель процесса низшего уровня функционально подчиняется руководителю процесса, в который входит данный.

4. ОЦЕНКИ КАЧЕСТВА ТЕХНИЧЕСКИХ РЕШЕНИЙ

Следует понимать, что контроль качества решений, направленных на обеспечение качества изделия будет выполняться между операциями процесса. Чтобы был понятны действия при оценке решений, следует сопоставить операции процесса этапам ЖЦИ, которые определяют общепринятый смысл решений (НИР, КПП, ТхПП, изготовление (включая управление технологическими процессами), испытания).

Единственным «эталоном», служащим для оценки решений с позиций качества изделия, является техническое задание на изготовление прибора. Следовательно, оно должно быть структурированным, конкретным, определять не только общие требования к техническому уровню изделия, но и отразить его в виде точного значения параметров (или их интервалов), а также требования к ремонтным циклам, условиям эксплуатации, эргономики, безопасности, хранения и т.д. Другими словами, в ТЗ должна быть показана модель качества изделия

Естественно, что эти требования относятся к готовому изделию, а на ряде этапов ЖЦИ, по сути, будет во многом осуществляться прогнозная оценка соответствия качества изделия ТЗ.

Критерии качества любого процесса деятельности организации могут определяться не только моделью качества изделия, но и дополняться критериями качества самого процесса.

Так, процессы КПП зачастую имеют непосредственную связь с параметрами модели качества изделия (например, производительность, мощность, коэффициент унификации и т.д.).

А, скажем, процессы ТхПП влияют опосредованно. Они должны обеспечить разработку технологических процессов изготовления изделия, обеспечивающие его получение в соответствии с конструкторской документацией. Учитывая, что многие оценки будут носить неявный, качественный вид, важное значение приобретает подбор экспертов из числа работников организации (как правило) для оценки решений. Более подробно требования к системе критериев приведены в [4, 5].

5. СОВЕРШЕНСТВОВАНИЕ ПРОЦЕДУР ПРИНЯТИЯ РЕШЕНИЙ ПРИ ВЫПОЛНЕНИИ ПРОЦЕССОВ В ПРОИЗВОДСТВЕННОЙ СИСТЕМЕ

Как уже говорилось выше, одним из важнейших параметров, определяющих эффективность работы производственных систем, являются временные затраты на выполнение этапов жизненного цикла изделия, реализуемых в рамках данного предприятия. Рассмотрим мероприятия, позволяющие сократить время выпуска образца нового изделия.

1. Автоматизация процедур управления процессами с точки зрения их иерархии и временного фактора. Должна быть реализована в виде системы с сервис-ориентированной архитектурой, с определенными правами доступа каждого пользователя.

2. Автоматизация инженерных работ при выполнении процессов, связанных с КПП, с ТхПП, с изготовлением и испытаниями. Разговор идет о внедрении и применении соответствующих инструментальных средств. В чем их особенность для условий рассматриваемых нами производственных систем? Очевидно, в мобильности, и в возможности с их помощью оперативно выполнять корректировку решений. Для этого программные средства на уровне пользователей должны быть информационно интегрированы между собой [6].

3. Параллельное выполнение работ, включенных в разные этапы ЖЦИ (или в разные стадии одного этапа). Способ реализации данного мероприятия достаточно прост. С этой целью необходимо, учитывая специфику конкретного предприятия, сделать соответствующим образом временную увязку процессов, входящих в разные этапы ЖЦИ

(или в разные стадии одного этапа), и приглашать для обсуждения технических решений работников, ответственных за соответствующие операции процессов, включая и процессы 2 группы.

4. Разработка интеллектуальных систем поддержки принятия решений при реализации процессов [7], в первую очередь о решениях, принимаемых и реализуемых представителями различных функциональных подразделений.

5. Создание предпосылок эффективного принятия решений на последующих этапах и стадиях ЖЦИ.

Например, закрепление операций обработки за наименее загруженным оборудованием (проектирование на основе текущей производственной ситуации) уменьшает вероятность появления «узких» мест и упрощает управление технологическими процессами.

6. Построение всех процессов организации (не только производственных) в соответствии с идеологией «бережливого производства», что означает делать то, что надо именно сейчас и уменьшать непроизводительные траты труда и средств [8, 11].

6. ОСНОВНЫЕ ТРЕБОВАНИЯ К СТРУКТУРЕ ПРОИЗВОДСТВЕННОЙ СИСТЕМЫ

Очевидно, что совершенство производственных (технологических) подразделений (ТП) во многом определяет и эффективность работы всей производственной системы. Специфика многономенклатурного производства вносит определенную специфику в перечень требований, реализуемых при создании ТП. Рассмотрим наиболее важные, на наш взгляд, из них.

1. Обеспечение резервов производственных мощностей. Как показывает опыт, плановыми работами оборудование каждой группы должно быть загружено не более, чем на 80%. В этом случае уменьшается (или исключается) время ожидания следующей операции.

2. Пропорциональность производственных мощностей по видам оборудования. Пропорциональность мощностей достигается на основе структуры станкоемкостей предполагаемых к выпуску изделий, и методик, изложенных в [9].

3. Производственная гибкость. Вопрос ставится таким образом, что оборудование должно иметь широкие технологические возможности. Надо учитывать, что технологические возможности должны перекрываться не только в пределах типов оборудования одной группы, но и в пределах разных групп.

4. Технологическая гибкость оборудования. Под параметром, характеризующим технологическую гибкость, понимается время

переналадки оборудования на выполнение другой операции. Данный параметр определяет технико-экономические показатели эксплуатации оборудования.

5. Это понятие (технологическая гибкость) - комплексное и охватывает разные аспекты (технический, технологический, организационный).

6. Расстановка оборудования. От рациональности расстановки оборудования в производственных подразделениях многономенклатурных производств напрямую не зависят времена технологических циклов, но создаются предпосылки для негативного влияния на них из-за увеличения времени межоперационного транспортирования изделий.

7. ЗАКЛЮЧЕНИЕ

Указанные выше мероприятия невозможно успешно реализовать без соответствующего кадрового обеспечения. Перечислим основные моменты.

1. Выработка у исполнителей чувства взаимопомощи и коллективной ответственности за результат. Следует понимать, что разделение функций в рамках общего процесса не означает разделение коллектива, что все действия должны быть направлены на общий успех – выпуск качественного изделия.

2. Повышение квалификации исполнителей должно вестись постоянно, как силами работников самой фирмы (курсы, самообучение), так и во внешних организациях (семинары, обмен опытом и т.п.). Весьма важна кадровая подготовка цехового управленческого персонала, особенно мастеров и руководителей диспетчерских служб. Задача – уметь вырабатывать управленческие решения по управлению технологическими циклами и производственными мощностями.

3. Не обойти стороной и вопросы готовности руководства организацией на серьезные технические и кадровые инновации. Но надо понимать, что качественное изделие не бывает дешевым, в том числе и в силу высоких затрат на подготовительном этапе.

Работа носит во многом принципиальный характер, в некоторых моментах, возможно, и спорный. Целью ставилось определение ключевых моментов и принципов, в соответствии с которыми, на взгляд авторов, должно происходить развитие систем управления современными производственными системами машиностроения и приборостроения [10-14].

Литература

1. Колчин А.Ф., Овсянников М.В., Стрекалов А.Ф., Сумароков С.В. Управление жизненным циклом продукции. М.: Анахарсис. 2002. – 304 с.
2. Никифоров А.Д., Бакиев А.В. Процессы жизненного цикла продукции в машиностроении. М.:, Абрис. 2011. 688 с.
3. Бурдо Г.Б., Сорокин А.Ю. Автоматизация процессов управления качеством при создании наукоемких геофизических изделий. Научно-технический вестник "Каротажник". 2016, № 9 (267). С.185-199.
4. Гличев А.В. Основы управления качеством продукции. М.: Стандарты и качество, 2001. 424 с.
5. Бурдо Г.Б., Стоянова О.В. Автоматизированная система управления процессами создания наукоемких машиностроительных изделий. Программные продукты и системы. – 2014. – №2(106). С.164-170.
6. Бурдо Г.Б., Стоянова О.В., Исаев А.А., Чаплыгин С.А. Гибридные интегрированные автоматизированные системы программирования и управления технологическими процессами. В сборнике: Труды конгресса по интеллектуальным системам и информационным технологиям «IS&IT'17». 2-9 сентября 2017, Дивноморское, Россия. С. 94-101.
7. Частиков А.П., Гаврилова Т.А., Белов Д.Л. Разработка экспертных систем. Среда CLIPS. СПб.: БХВ-Петербург. 2003. 608 с.
8. Смирнова Н.В. Как не наступить на грабли при переходе на процессное управление. Управлении производством. -2009, №3. С.26-32.
9. Вороненко В.П, Егорова А.В., Косов М.Г и др. Проектирование автоматизированных участков и цехов / Под. ред. Ю.М. Соломенцева. М.: Высшая школа. 2000. 272 с.
10. Бурдо Г.Б. Совершенствование технологической подготовки машиностроительного производства. Станки и инструмент (СТИН), № 7, 2016. – С. 2-8.
11. Вумек Джеймс П., Джонс Даниел Т. Бережливое производство. Как избавиться от потерь и добиться процветания вашей компании. М.: Альпина Бизнес Букс. 2008. – 473 с.
12. Womack, J.P., Jones D.T.: Lean Thinking: Banish Waste and Create Wealth in Your Cor-poration. Productivity Press, 1996.
13. Womack, J.P., Jones D.T.: Lean Thinking: Banish Waste and Create Wealth in Your Cor-poration. 2nd ed., Free Press, New York, 2003.
14. Shingo Sh.: Kaizen and the Art of Creative Thinking, Enna Product Corp., PCS Inc, 2007.

УДК 621.391

МЕТОД РЕШЕНИЯ ЗАДАЧ УПРАВЛЕНИЯ ЖИЗНЕННЫМ ЦИКЛОМ ТЕХНИЧЕСКИХ СИСТЕМ

Овсянников М.В.
e-mail: mvo50@mail.ru
НИУ «МЭИ»
Буханов С.А.
e-mail: bukhanov@yandex.ru
МГТУ им. Н. Э. Баумана

1. ВВЕДЕНИЕ

Большинство задач управления процессами и поддержки технических решений на этапах ЖЦП представляют собой задачи структурного синтеза сложного объекта

Структуру класса объектов, имеющих одинаковое функциональное назначение, называют обобщенной. Элементы этого класса являются аналогами и прототипами реализуемого проекта. Обобщенная структура представляет собой «комбинаторное пространство», в котором выстраиваются различные сочетания элементов, образующие структуры разрабатываемого технического объекта. В качестве средств описания обобщенных структур используются табличные, алгебраические, логические и сетевые модели.

2. ЗАДАЧА УПРАВЛЕНИЯ КОНФИГУРАЦИЕЙ

Примером такой задачи является задача управления конфигурацией, направленная на установление и поддержание соответствия эксплуатационных, функциональных и физических свойств изделия заданным требованиям процессе создания и преобразования этого изделия в течение его ЖЦ [11].

Под управлением конфигурацией понимается совокупность мероприятий, проводимых на протяжении ЖЦ изделия, и направленных на установление и поддержание соответствия между требованиями, реализованными характеристиками изделия и документацией по конфигурации.

В частности, такие задачи возникают при производстве изделий для требований конкретного заказчика, создания компьютерных,

энергетических систем, при построении процедур обслуживания и др. [16, 11].

Базовые элементы технологии управления конфигурацией для применения в условиях машиностроительного предприятия сосредоточены в основном на решении задач учета конфигурации, хранения и обеспечения доступа к ней на протяжении всего ЖЦ изделия [16].

Среди всех подходов к решению задачи структурного синтеза наибольшее распространение в системах автоматизированного проектирования получили различные методы, принадлежащие к классу комбинаторно-логических. В основе этого подхода лежит хорошо организованные перебор в массиве решений, которые являются аналогами и прототипами.

Наибольшее распространение среди этих моделей получили морфологические таблицы и А-деревья (И–ИЛИ-деревья – в исследованиях по искусственному интеллекту и в программировании) [4].

Среди прикладных методов решения задач комбинаторной оптимизации (ЗКО) в последнее время формируется класс моделе-ориентированных (model-based) алгоритмов – в отличии от большинства традиционных алгоритмов, которые относят к классу задаче-ориентированных (instance-based) [3, 18].

Задаче-ориентированные алгоритмы генерируют новые варианты решений только на основе одного или нескольких текущих вариантов решений. К числу таких методов относятся генетические алгоритмы, повторяемый локальный поиск и другие [13].

Большинство практически важных задач КО относится к числу NP-трудных, что наряду с возможными погрешностями в задании исходных данных и существенным количеством локальных минимумов целевой функции делает нецелесообразным использование точных алгоритмов решения ввиду значительных вычислительных затрат [3].

В приближенных методах, прежде всего, будем различать по наличию рандомизации два принципа принятия решений – детерминированный и стохастический [7].

Примером стохастического метода является простой локальный поиск – алгоритм, представляющий собой итеративный процесс, который итерационно изменяет некоторое начальное решение (найденное каким-то другим алгоритмом или выбранное случайно) и пытается, в конечном итоге, заменить его лучшим [1, 7, 11]. На каждом шаге локального поиска происходит переход от текущего решения к соседнему решению из окрестности с лучшим значением целевой

функции до тех пор, пока не будет достигнут локальный оптимум. Соседнее решение, при таком подходе, не обязательно должно быть наилучшим в окрестности, а вот критерий оценки решения не должен меняться по ходу итеративного процесса. Эффективность этого алгоритма может быть увеличена за счет использования эвристик, учитывающих особенности решаемой задачи, такие как количество позиций конфигурации (управляемых комбинаторных переменных), объем справочника и специфики системы ограничений задачи.

В целом, на основе анализа литературы можно сделать вывод, что не существует единого метода, который позволяет эффективно решать все задачи управления конфигурацией [5].

Предлагаемый подход позволит задачи управления конфигурацией решать с помощью единого метода, включающего 3 этапа:

- построение формальной прикладной онтологии проблемной области;
- настройка и корректировка ограничений и требований к конфигурации;
- применение алгоритма решения комбинаторной задачи методом удовлетворения ограничений с использование онтологии проблемной области.

В основу предлагаемой онтологии ЖЦП положена онтология отношений, которая представляет свойства сущностей в виде бинарных отношений, что позволяет создать метаонтологию как совокупность классов объектов и отношений, а онтологию верхнего уровня в форме классификаторов объектов и их свойств. Этот подход обеспечивает возможность применения универсальных алгоритмов на основе метода программирования в ограничениях для широкого класса задач структурного синтеза и управления конфигурациями.

Общей чертой подходов, использующих бинарные модели – это представление элементов посредством реализации бинарных ассоциаций, т. е. предложений, в которых участвуют только два терма. Основные понятия атомарные, что позволяет явно задавать ограничения. Изменения в проблемной области отражаются добавлением или удалением типов бинарных отношений, сущностей или имен сущностей, а также изменением ограничений. Это приводит к большей стабильности по сравнению с подходами, где ограничения жестко связываются с понятиями.

В основу данного подхода положены работы Овсянникова и Буханова по созданию алгоритмов и построения онтологий различных задач ЖЦП [2, 6, 14].

В качестве прототипа разрабатываемой метаонтологии жизненного цикла использованы схемы и язык описания RDF и OWL, и основные сущности и положения, заложенные в Модель данных «Системы промышленной автоматизации и интеграция. Интеграция данных жизненного цикла для перерабатывающих предприятий, включая нефтяные и газовые производственные предприятия» (ISO 15926).

Основное требование к разработанной модели – возможность применения универсальных алгоритмов поиска решений методами удовлетворения ограничений и другими алгоритмами ИИ.

В данной работе используется метод описания интерпретаций, предназначенный для **использования алгоритмов удовлетворения ограничений (Constrain method).** Для решения задачи синтеза конфигурации Бухановым С.А. был разработан специализированный язык описания ограничений [2].

Прикладная онтология задачи синтеза конфигурации, построенная на основе предложенной метаонтологии включает классификаторы объектов и свойств и онтологию задачи (синтеза или определении степени соответствия конфигурации).

Использование предложенной онтологии позволит решать задачи методом удовлетворения ограничений и обеспечить взаимодействие с внешними источниками данных для наполнения классификаторов и зависимостей через стандарт ISO 15926.

Основная проблема существующих подходов к автоматизации решения данного класса задач – это необходимость в трудоемкой программной реализации алгоритмов поиска решений и поддержке разработанного программного средства для учета всех изменений в методике расчета.

Для решения подобных задач был разработан метод управления конфигурацией на основе метода удовлетворения ограничений (CSP), позволяющий инженеру описывать и решать задачи, не прибегая к программированию сложных алгоритмов [2].

В задачах CSP вся система утверждений формулируется в виде конъюнкции отдельных выражений, называемых ограничениями. При формулировании задачи стараются разбить исходную систему утверждений на максимальное число элементарных ограничений, выполнимость каждого из которых может быть проверена независимо от других. Каждое ограничение связывает между собой одну или несколько переменных задачи, что позволяет получить дополнительную информацию о конфликтности присваиваемых

значений переменных задачи. Решение считается допустимым, если выполняются все ограничения задачи.

Рассматривая некоторую задачу в виде задачи CSP, можно достичь нескольких важных преимуществ. Представление состояний в задаче CSP соответствует некоторому стандартному шаблону (семантический интерфейс) (т.е. выражается в виде множества переменных с присвоенными значениями), поэтому функцию определения преемника и проверку цели можно записать в универсальной форме, применимой ко всем задачам CSP.

Задачи CSP простейшего вида характеризуются тем, что в них используются переменные, которые являются дискретными и имеют конечные области определения.

Более того, могут быть разработаны эффективные, универсальные эвристические функции, для создания которых не требуются дополнительные знания о конкретной проблемной области.

Наконец, для упрощения процесса решения может использоваться сама структура графа ограничений, что позволяет в некоторых случаях добиться экспоненциального уменьшения сложности.

Для задач CSP существуют методы распространения ограничений, позволяющие редуцировать домены переменных, что может значительно снизить время поиска.

Применяемая парадигма, называемая программированием в ограничениях, требует только описания задачи, но не заставляет разработчика определять алгоритм ее решения [9]. Задача и модель представляются в этом случае как неупорядоченная совокупность отношений, которые соответствуют связям, существующим между переменными задачи. Эти отношения, называемые общим термином "ограничения", могут иметь вид уравнений, неравенств, логических выражений и т.п. [8].

Определение 1.

Задачу управления конфигурации в терминах CSP можно определить следующей структурой:

$$(X, D, S, C, F),$$

где:

- множество «позиций» («переменных» в терминологии CSP) $X = \{X_1, \ldots, X_n\}$ соответствуют терминальным объектам функциональной конфигурации (ОФК);
- множество доменов компонентов, реализующих функции позиции $D = \{D_1, \ldots, D_n\}$, где D – функция (механизм), связывающая

позицию X_i с ее доменом $D(X_i)$. Домен переменной представляет собой конечный набор объектов справочника;

- каждый домен содержит множество объектов $D_i = \{d_1, \ldots, d_m\}$;
- множество свойств позиции и компонентов $S = \{S_1, \ldots, S_r\}$;
- множество ограничений $C = \{C_1\{S(X)\}, \ldots, C_o\{S(X)\}, \}$, где ограничение C_j определено на подмножестве свойств позиций $S(X)$.

Целевая функция (функция предпочтения) $F = f\{S(X)\}$.

Определение 2.

Конфигурацией (в терминах ISO проектной конфигурацией) называется совокупность P назначений экземпляров d_{mj} из доменов D_J компонентов для позиций x_i

$$P = \{p_1 = d_{mi} \partial D_i, \ldots p_I = d_{mI} \partial D_I\}$$

Конфигурация является согласованной, если назначения в P согласуются с ограничениями в C. Кроме того, конфигурация является полной, если для всех позиций в X назначены конкретные экземпляры компонентов d_{mi}. Наконец, конфигурация допустима, если она является одновременно согласованной и полной.

Для управления конфигурацией можно выделить три класса задач для разных этапов ЖЦП:

- синтез ПК на основе объектов и ограничений ФК и требований к конкретному проекту;
- оценка степени соответствия значений характеристик изделия требованиям заказчика на основе объектов и ограничений ФК для всех этапов ЖЦ;
- определение значений характеристик объектов ПК, при которых выполняются ограничения ФК и требований к изделию.

Определение 3.

Решением задачи синтеза конфигурации называют нахождение такого кортежа назначений компонентов, который удовлетворяет всем ограничениям.

Тогда задачу синтеза можно сформулировать:

Найти такое назначения компонентов P, из доменов $D = \{D_1, \ldots, D_n\}$, для позиций $X = \{X_1, \ldots, X_n\}$, при котором все ограничения $C = \{C_1, \ldots, C_o\}$, определенные при значении свойств назначенных компонентов d_{mj} справедливы, а функция предпочтения имеет наилучшее значение.

$$F = f\{S(P)\} \rightarrow extr,$$

$$Ÿ\{C = \{C_1\{S(P_1)\}, \ldots, C_o\{S(P_o)\} = true\},$$

где: *S(P)* – значения свойств компонентов, назначенных на соответствующие позиции.

Для задач CSP существует специальный класс модифицированных методов разного рода, имеющих повышенную эффективность за счет особенностей работы с ограничениями. При решении задач методами программирования в ограничениях характерно систематическое использование особенностей решаемой задачи для определения набора эвристик.

Основной особенностью реализованного алгоритма стохастического поиска является стратегия поиска в двух подпространствах: подпространство недопустимых и подпространство допустимых решений. Что позволяет начинать поиск с любого случайного решения, в том числе недопустимого (см. рис.1).

Для недопустимых конфигураций значение оценочной функции определяется на основе комплексного показателя, называемого степенью расхождения, которая выражается в оценке отклонения текущей конфигурации от допустимой.

В качестве оценочной функции допустимых решений используется вычисленное значение целевой функции в соответствии с заданным критерием оптимальности (минимизация или максимизация).

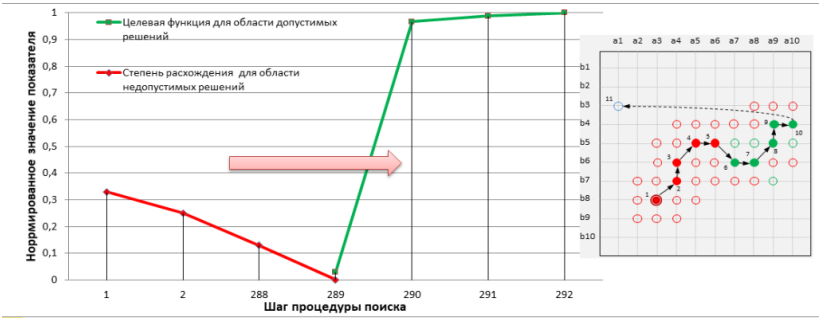


Рис.1. Выход из подпространства недопустимых решений

Предложенные методы и алгоритмы были **программно реализованы** в виде ПО решателя задач синтеза конфигурации CMSolver [6].

Решатель позволяет выполнять все этапы решения задач от ее описания до получения и анализа результатов.

Важной частью описания задачи является система ограничений. Ограничения связывают параметры объектов конфигурации. При постановке задачи синтеза конфигурации как задачи удовлетворения ограничений структура взаимосвязей параметров объектов

конфигурации может быть представлена в виде графа, называемого в литературе графом ограничений [15].

Предлагаемая методика и алгоритмы были исследованы на простом примере формирования конфигурации изделия «автомобиль» (в примере использованы характеристики автомобилей марки VW). Процесс основан на формировании конфигурации путем перебора вариантов с предварительным ограничением пространства поиска.

В ходе исследования эффективности разработанных алгоритмов была проведена серия прогонов на нескольких модификациях задачи и различных объемах справочников.

3. ЗАКЛЮЧЕНИЕ

Предлагаемый подход позволит задачи управления конфигурацией решать с помощью единого метода, включающего 3 этапа:

- построение формальной прикладной онтологии проблемной области,
- настройка и корректировка ограничений и требований к конфигурации;
- применение алгоритма решения комбинаторной задачи методом удовлетворения ограничений с использование онтологии проблемной области.

В основе предлагаемого в работе подхода лежит расширение процедуры идентификации функциональная конфигурация построением онтологии предметной области с определением закономерностей и ограничений для формирования структуры изделия: выбора возможных компонентов из числа альтернативных и определения требований для проектирования недостающих компонентов.

Предлагаемая методика позволяет решать важные инженерные задачи определения состава объектов конструируемых изделий или формируемых процессов, что может найти применение на различных этапах ЖЦ изделий.

Литература

1. Бочкарев С.В., Петроченков А.Б., Буханов С.А., Овсянников М.В. Структурный синтез сложного электротехнического оборудования на основе метода удовлетворения ограничений// Электротехника, 2015, № 6. С. 63-69.
2. Буханов С.А., Овсянников М.В. Управление конфигурацией с использование метода программирования в ограничениях // Эффективные

методы автоматизации подготовки и планирования производства: сборник научных трудов. М.: Издательский дом «Спектр», 2011, С. 150-156.
3. Гуляницкий Л.Ф., Сиренко С.И. Гибридная метаэвристика, основанная на оптимизации муравьиными колониями и н-методе // Компьютерная математика, 2009, № 1, С. 142-151.
4. Евгенев Г.Б. Системология инженерных знаний: Учеб. пособие для вузов. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2001, - 376 с.
5. Левин А.И., Судов Е.В. Методы и технологии управления конфигурацией сложных изделий -Технологии приборостроения, 2003, №4.
6. Овсянников М.В., Буханов С.А. Управление конфигурацией методом программирования в ограничениях// Известия высших учебных заведений. Машиностроение. 2012, № 14, С, 70-75.
7. Сергиенко И.В., Гуляницкий Л.Ф., Сиренко С.И. Классификация прикладных методов комбинаторной оптимизации // Кибернетика и системный анализ, 2009, №5, С. 71-83.
8. Щербина О.А. Удовлетворение ограничений и программирование в ограничениях // Интеллектуальные системы. 2011, № 15, С. 53-170.
9. Apt K. R. Principles of Constraint Programming. Cambridge University Press, 2003.
10. Dechter R. Constrain Processing. Morgan Kaufmann Publishers, 2003, 481 p.
11. Hajria I., Goknil A., Brianda L.C., Stephany T. (2019). Change impact analysis for evolving configuration decisions in product line use case models. Journal of Systems and Software. Volume 139, pp. 211-237.
12. Hass A.M. Configuration Management Principles and Practice. - Addison-Wesley Pub Co, 1st edition 2002, p.432.
13. Hoos H.H., Stutzle T. Stochastic Local Search: Foundations and Applications. – San Francisco: Morgan Kaufmann Publ., 2005. 658 p.
14. Construction of an ontological model of the life cycle of electrotechnical equipment // Petrochenkov A.B.,Ovsyannikov M.V., Bochkarev S.V., Bukhanov S.A. -Russian Electrical Engineering, 2015, Vol. 86, № 6, pp. 320-325.
15. Rossi F., van Beek P., Walsh T. (eds.) Handbook Of Constraint Programming. Elsevier, 2006, 978 p.
16. Schäffer, E., Mayr, A., Huber, T., Höflinger, T., Einecke, M., & Franke, J. (2019). Gradual tool-based optimization of engineering processes aiming at a knowledge-based configuration of robot-based automation solutions. Procedia CIRP, 81, pp. 736-741.
17. Watts F.B. Engineering Documentation Control Handbook : Configuration Management for Industry. - Noyes Publications; 2-nd edition, 2000, p.265.
18. Zlochin M, Birattari M, Meuleau N, Dorigo M, Model-based search for combinatorial optimization: A critical survey. Annals of Operations Research, 2004, №131, pp. 373-395

УДК 629.488

ИНТЕЛЛЕКТУАЛЬНОЕ УПРАВЛЕНИЕ ТЕХНИЧЕСКИМ СОСТОЯНИЕМ ЛОКОМОТИВА

Федотов М.В.
e-mail: fedotov-mv@vnikti.com
Клименко Ю.И., *к.т.н.*
e-mail: klimenko-ui@vnikti.com
АО «ВНИКТИ»
Грачев В.В., *д.т.н.*
e-mail: v_grach@mail.ru
ФГБОУ ВО «ПГУПС»

1. ВВЕДЕНИЕ

Для управления техническим состоянием локомотива на разных стадиях его жизненного цикла используются различные методы: от улучшения свойств применяемых материалов и конструкции оборудования на стадиях разработки, производства и испытаний до выбора оптимальных режимов его эксплуатации, периодичности и объема технических обслуживаний и ремонтов (ТОиР) на стадии использования (владения). Основным критерием, определяющим выбор системы ТОиР в конкретных условиях эксплуатации, является получение максимальной экономической выгоды (прибыли) от владения локомотивом, необходимое условие которого – снижение стоимости его жизненного цикла на стадии владения.

Одним из наиболее перспективных способов снижения расходов на ТОиР является переход от существующей планово-предупредительной системы ремонтов локомотивов к системе ремонтов по фактическому состоянию их оборудования. Научно-технологической основой такой системы должны стать диагностирование и прогнозирование технического состояния оборудования локомотивов с использованием встроенных средств контроля и диагностики в качестве источника объективной диагностической информации, по результатам обработки которой определяется текущее техническое состояние локомотива и принимаются решения по переводу его из этого исходного состояния в другое (заданное) состояние. Таким образом, управление техническим состоянием локомотива представляет собой процесс преобразования

диагностической информации в определенные целенаправленные действия, переводящие управляемую систему (локомотив) из исходного технического состояния в заданное.

Наличие больших объемов телеметрической информации, получаемой с борта локомотивов, в сочетании с интенсивным развитием технологий интеллектуального анализа данных позволяет использовать методы машинного обучения для оценки и прогнозирования изменения технического состояния узлов локомотивов. Однако примеров практического применения таких технологий для диагностики сложного локомотивного оборудования крайне мало. Это объясняется в первую очередь сложностью формирования обучающих выборок для диагностических моделей оборудования по классам состояния, отличных от эталонного. Если формирование таких выборок для класса состояния оборудования, принимаемого за эталонный в конкретных условиях эксплуатации, не представляет особой сложности, то накопить необходимый объем информации для состояний, соответствующих различным отказам оборудования, и напрямую связать их с процессом деградации его состояния крайне сложно.

2. МЕТОДИКА СИНТЕЗА ИСХОДНЫХ ДАННЫХ С ПОМОЩЬЮ АНАЛИТИЧЕСКИХ МОДЕЛЕЙ

Решением проблемы получения необходимой исходной информации может быть создание развернутых аналитических моделей диагностируемых объектов, способных адекватно отражать изменение параметров оборудования в зависимости от деградации его состояния (износа, загрязнения, и т.п.). При использовании такой аналитической модели может быть синтезирован необходимый объем исходных данных для обучения и валидации различных диагностических моделей.

Для отработки предлагаемой методики в качестве объекта диагностирования была выбрана система смазки тепловозного дизеля и входящие в нее узлы, которые напрямую влияют на надежность дизеля и тепловоза в целом. При разработке аналитической модели [1] были учтены структурные параметры системы смазки: степень загрязнения фильтра, изменение вязкости масла, изменение зазора в подшипниках коленчатого вала, которые наряду с частотой вращения коленчатого вала и температурой масла задаются в качестве входных параметров. Выходными параметрами модели являются расходы и напоры в узлах и элементах гидравлической сети.

При помощи аналитической модели были синтезированы три массива (выборки) векторов *F*, *W*, *Z*, представляющих собой совокупность известных (контролируемых на тепловозе) параметров системы смазки:

$$F, W, Z = \{\omega, P_{н2}, P_{Д}, dP, T\},$$

где ω – угловая скорость вращения коленчатого вала, $с^{-1}$; $P_{н2}$ – давление масла на выходе второго масляного насоса, МПа; $P_{д}$ – давление масла в масляном коллекторе дизеля, МПа; *dP* – перепад давления на фильтре, МПа; *T* – температура масла на входе в дизель, К.

Массив векторов *F* характеризует степень загрязненности полнопоточного фильтра масла, массив векторов *W* – изменение вязкости масла, массив *Z* – изменение средней величины зазора «на масло» в подшипниках коленчатого вала, которая является интегральной характеристикой их состояния. Выборки *F, W, Z* использовались для обучения эталонных диагностических моделей, выходными (объясняемыми) параметрами которых являются структурные параметры: *f* – оценка текущего состояния фильтра, *w* – оценка текущей вязкости масла, *z* – оценка текущей средней величины зазоров в подшипниках коленчатого вала.

3. ПОСТРОЕНИЕ ДИАГНОСТИЧЕСКИХ МОДЕЛЕЙ НА ОСНОВЕ КЛАССИФИКАТОРОВ С ИСПОЛЬЗОВАНИЕМ СИНТЕЗИРОВАННЫХ ВЫБОРОК

Предложенная методика синтеза векторов недостающих исходных данных дает возможность использовать для построения диагностических моделей сложного оборудования локомотивов всю палитру алгоритмов, применяемых в задачах машинного обучения [2].

Была изучена возможность использования в диагностических моделях методов мультиклассовой классификации и регрессии. Для мультиклассовой классификации был сформирован алфавит состояний системы смазки, состоящий из 20 классов, соответствующих различным состояниям фильтра масла и зазорам в подшипниках коленчатого вала (рис.1). Модели классификаторов реализовывались с помощью алгоритмов машинного обучения «Случайный лес» (Random Forest, RF) и «Градиентный бустинг» (Gradient Boosting Machines, GBM), которые построены с использованием методов бэггинга и бустинга на основе ансамблей решающих деревьев [3].

Полученная точность моделей классификации с использованием ансамблей решающих деревьев на тестовой выборке составила около

40%, чего явно недостаточно для практического применения данного подхода к диагностированию сложных объектов. Причиной является недостаточная чувствительность моделей классификаторов вследствие влияния точности аналитической модели. Поскольку структурные параметры диагностируемой системы являются медленно меняющимися, то для повышения чувствительности моделей может быть применен метод мультиклассовой статистической классификации. Метод заключается в классификации одного и того же состояния объекта по выборке входных векторов, отобранной в течение заданного промежутка времени, где состояние структурного параметра можно считать неизменным, и интерпретации полученных результатов по принципу WTA (Winner takes all – «Победитель получает все»). Применение данного метода позволило существенно (до 80%) повысить точность классификации.

Рис.1. Пример сформированного алфавита классов состояния системы смазки тепловозного дизеля и динамика изменения классов в эксплуатации

Для экспериментальных исследований диагностических моделей, построенных с использованием описанного метода мультиклассовой статистической классификации, использовались данные, полученные в процессе эксплуатации тепловозов ТЭП70БС. Вполне ожидаемо, что в процессе эксплуатации тепловоза распределение оценок классов текущего состояния системы смазки тепловозного дизеля смещается в область ухудшения состояния полнопоточного фильтра масла с последующим улучшением состояния

после цикла технического обслуживания при сохранении неизменным состояния подшипников коленчатого вала. Данная динамика состояния системы смазки соответствует изменениям, определенным с помощью диагностических моделей на основе статистической классификации (рис.2). Это соответствует изменению в пространстве алфавита классов (рис.1), а также сложившейся практике эксплуатации тепловозов и подтверждает достоверность выполненных оценок.

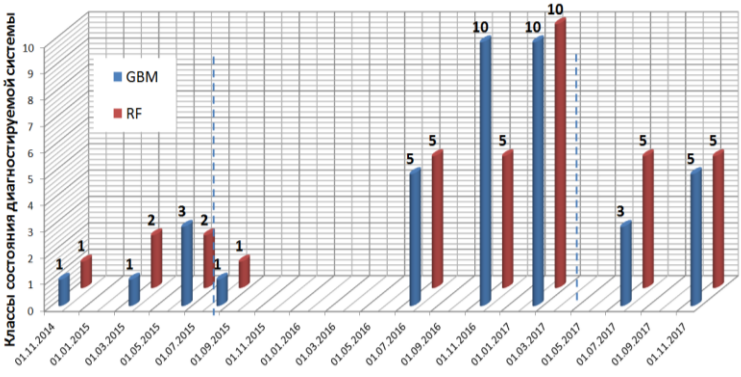


Рис.2. Динамика изменения технического состояния системы смазки дизеля тепловоза в процессе эксплуатации

4. ПОСТРОЕНИЕ РЕГРЕССИОННЫХ ДИАГНОСТИЧЕСКИХ МОДЕЛЕЙ С ИСПОЛЬЗОВАНИЕМ НЕЙРОННЫХ СЕТЕЙ

Другим подходом к построению диагностических моделей сложных технических объектов является применение методов регрессии.

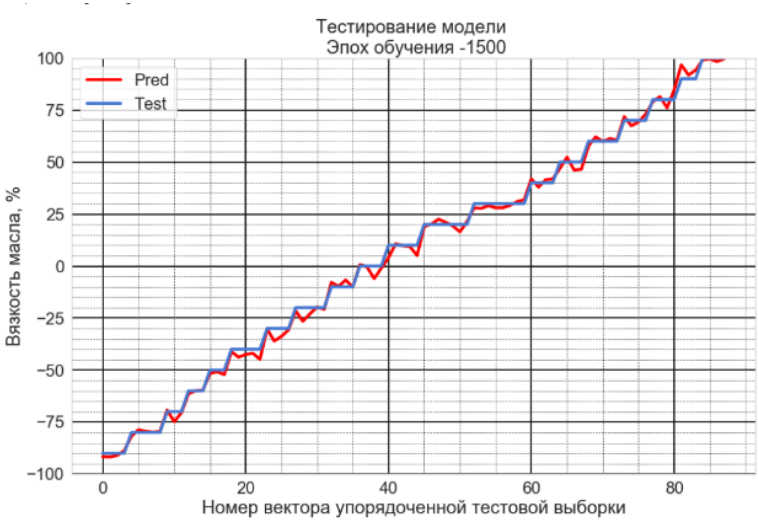
Проведенные исследования [4] показали, что для построения таких моделей хорошо подходят искусственные нейронные сети прямого распространения [5]. В качестве отклика моделей при этом могут использоваться структурные параметры *f*, *w*, *z*, характеризующие техническое состояние диагностируемого оборудования. Для всех трех моделей опытным путем была выбрана трехслойная архитектура нейронных сетей с общим числом нейронов не более 400. Распределение нейронов по слоям и количество эпох обучения подбиралось для каждой сети индивидуально с учетом минимизации ошибки обучения и валидации. Были сформированы, обучены на расчетной выборке и исследованы три нейронные сети для диагностирования технического состояния элементов системы смазки:

нейронная сеть для определения степени загрязнения фильтра, для оценки величины среднего зазора в подшипниках коленчатого вала и для определения вязкости масла. На тестовых выборках среднеквадратичные ошибки откликов моделей не превысили 3% при коэффициенте детерминации не менее 0,92 (рис.3).

Результаты исследования регрессионных нейросетевых диагностических моделей на реальных данных показали, что большая вариативность откликов моделей обуславливает актуальность использования в качестве диагностических параметров их статистически усредненные оценки. Условием достоверности таких оценок является нормальность распределения выборок откликов. Для проверки нормальности распределения выборок применялся метод Е.И. Пустыльника [6], не требующий использования громоздких таблиц для хранения граничных значений критериев и наиболее удобный для программной реализации.

На рис.4 приведены результаты оценки изменения состояния фильтра и вязкости масла в процессе эксплуатации, полученные с использованием нейросетевых диагностических моделей. Изменение отклика модели на рис.4а соответствует процессу загрязнения фильтра. Результаты на рис.4б получены для тепловоза, длительное время эксплуатировавшегося с пониженной вязкостью масла и периодической его заменой по причине неисправности топливной аппаратуры и попаданию топлива в масляную систему, приведшей в итоге к повреждению подшипников и шеек коленчатого вала дизеля.

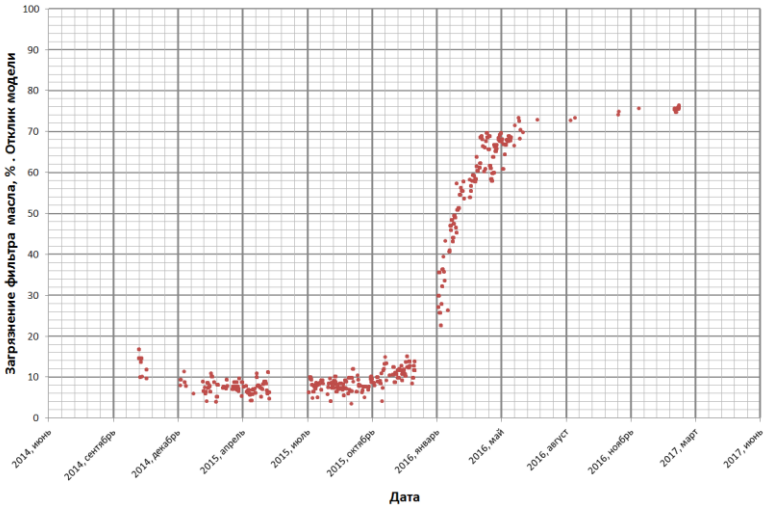
Отклики регрессионных диагностических моделей могут использоваться в качестве входных параметров для классификаторов состояния оборудования. Высокая точность нейросетевых регрессионных моделей в сочетании с предложенным статистическим способом классификации позволяет повысить точность классификации состояния системы смазки до 80%.



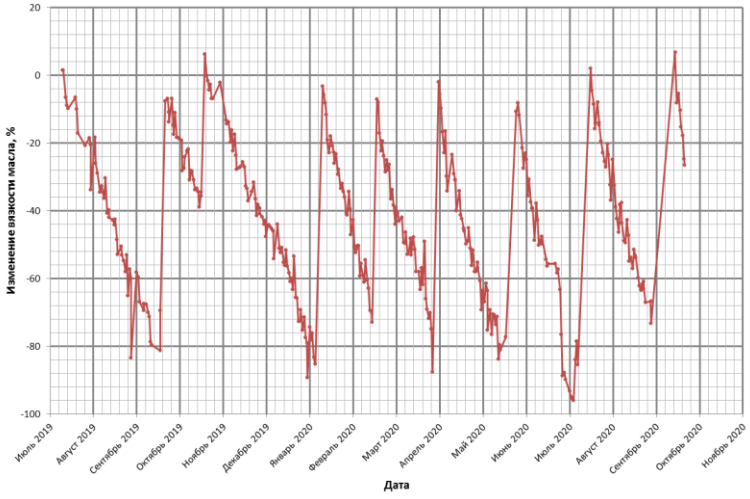
а)

б)

Рис. 3. Результаты тестирования нейросетевых регрессионных диагностических моделей: а) вязкость масла; б) зазор в подшипниках коленчатого вала



а)



б)

Рис.4. Изменение состояния в процессе эксплуатации по данным нейросетевой регрессионной диагностической модели: а) фильтра масла; б) вязкости масла

5. ИСПОЛЬЗОВАНИЕ РЕКУРРЕНТНЫХ НЕЙРОННЫХ СЕТЕЙ С LSTM-СЛОЯМИ ДЛЯ ПРОГНОЗИРОВАНИЯ ТЕХНИЧЕСКОГО СОСТОЯНИЯ ОБОРУДОВАНИЯ

Одной из основных задач системы управления техническим состоянием локомотива является задача прогнозирования его изменения в процессе эксплуатации. Следует отметить, что оценки, полученные с использованием регрессионных моделей, в отличие от классификаторов, менее вариативны и более удобны при организации предиктивного технического обслуживания локомотива.

Учитывая непрерывный и упорядоченный во времени характер диагностической информации, а также тот факт, что диагностируемый объект эксплуатируется в неизменных условиях, прогнозирование изменения технического состояния может осуществляться с использованием методов и алгоритмов анализа временных рядов. Учитывая высокую (как правило) размерность пространства признаков сложных технических объектов, эффективным аппаратом для анализа временного тренда изменения их состояния является аппарат рекуррентных нейронных сетей с LSTM-слоями [7].

Для исследования прогнозных моделей, построенных на основе рекуррентных нейронных сетей с LSTM-слоями, использовались упорядоченные во времени посуточные оценки средних значений откликов диагностических моделей *f*, *w* и *z*, преобразованные во временные ряды. Временные ряды, в свою очередь, были дополнительно сглажены (отфильтрованы) по методу скользящего среднего с размером окна фильтрации от 10 до 30 суток. Затем ряды были преобразованы в формат фреймов данных размерности $j \times N - j$ для подачи на вход рекуррентной нейронной сети. Фреймы данных имеют следующий вид:

$$W = \begin{vmatrix} W_0 W_1 \ldots\ldots\ldots\ldots W_j \\ W_1 \ldots\ldots\ldots\ldots W_{j+1} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ W_{N-j} \ldots\ldots\ldots\ldots W_N \end{vmatrix},$$

$$F = \begin{vmatrix} F_0 F_1 \ldots\ldots\ldots\ldots F_j \\ F_2 \ldots\ldots\ldots\ldots F_{j+1} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ F_{N-j} \ldots\ldots\ldots\ldots F_N \end{vmatrix},$$

$$Z = \begin{vmatrix} Z_0 Z_1 \ldots\ldots\ldots\ldots Z_j \\ Z_1 \ldots\ldots\ldots\ldots Z_{j+1} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ Z_{N-j} \ldots\ldots\ldots\ldots Z_N \end{vmatrix},$$

где N – размер обучающей выборки.

На рис.5 приведены результаты обучения прогнозной модели и ее прогноз для состояния фильтра масла тепловоза.

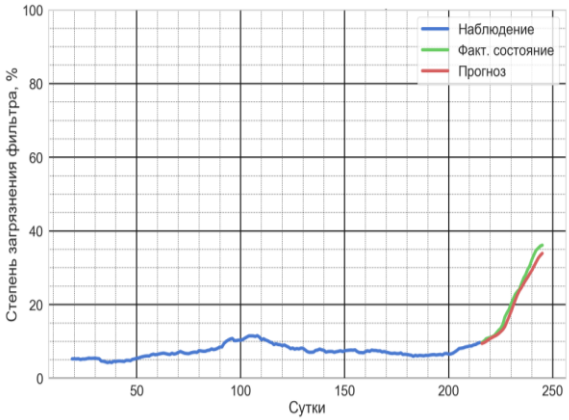


Рис.5. Изменение состояния фильтра масла и прогноз его изменения по данным диагностической и прогнозной моделей

Исследования построенных прогнозных моделей с использованием реальных данных эксплуатации локомотивов показали возможность построения прогноза с ошибкой не более 3,5% на периоде упреждения до 30 суток, чего вполне достаточно для планирования объемов регламентах работ по циклу обслуживания локомотивов ТО-3.

6. ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ СИСТЕМЫ ИНТЕЛЛЕКТУАЛЬНОГО УПРАВЛЕНИЯ ТЕХНИЧЕСКИМ СОСТОЯНИЕМ ЛОКОМОТИВА

Предлагаемая структура системы интеллектуального управления техническим состоянием локомотивов (рис.6) содержит аналитическую цифровую модель диагностируемой системы, сформированную с ее помощью базу данных для обучения и валидации эталонных

диагностических моделей и сами модели функционально обособленных групп оборудования.

В процессе эксплуатации локомотива оперативные данные по контролируемым параметрам непрерывно поступают в базу данных. Диагностические модели периодически (ежесуточно или несколько раз за сутки) производят обработку этих данных и сохраняют результаты обработки. Таким образом, база данных результатов диагностики ежесуточно пополняется новыми данными. Также ежесуточно производится переобучение прогнозных моделей с учетом новых результатов диагностирования. Переобученные модели осуществляют прогнозирование с интервалом упреждения до очередного технического обслуживания (ТО) или технического ремонта (ТР).

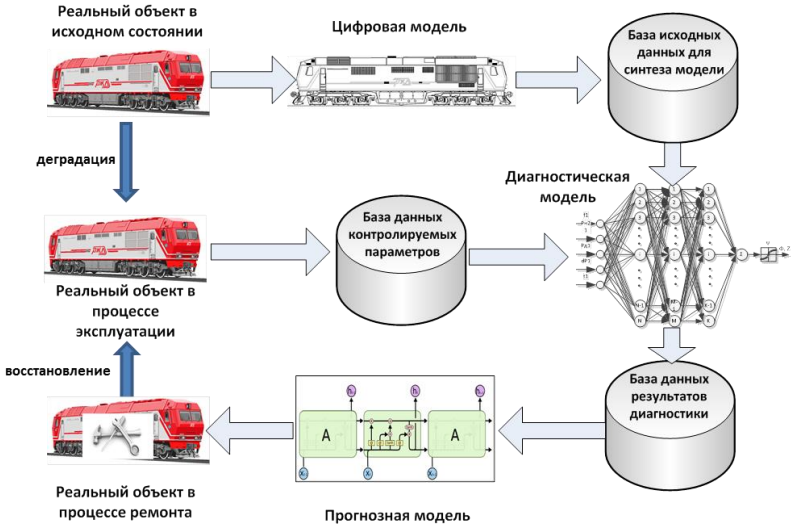


Рис.6. Структура системы интеллектуального управления техническим состоянием локомотивов

Проведение этого прогноза необходимо для понимания ситуации, что диагностируемые системы или узлы имеют достаточный ресурс до очередного ТО (ТР) и не приведут в этот период к отказам в пути следования. В случае выявления подобного негативного прогноза необходимо принятие мер по устранению либо временной локализации неисправности на промежуточном виде ТО, а если это невозможно, то требуется постановка локомотива на неплановый ремонт. При приближении даты проведения ТО (ТР) интервал упреждения должен достигать очередного ТО (ТР), следующего за текущим (рис.7). Это

позволит оценить и скорректировать (сократить) объемы предстоящих регламентных работ в тех случаях, когда спрогнозированного ресурса оборудования окажется достаточно для безаварийной эксплуатации до очередного ТО (ТР).

Рис.7. Определение границ принятия решения для проведения ремонтных работ

Это позволит оценить и скорректировать (сократить) объемы предстоящих регламентных работ в тех случаях, когда спрогнозированного ресурса оборудования окажется достаточно для безаварийной эксплуатации до очередного ТО (ТР).

7. ЗАКЛЮЧЕНИЕ

Для построения интеллектуальной системы управления техническим состоянием с использованием моделей машинного обучения требуются дополнительные объемы исходных данных, соответствующие различному состоянию диагностируемого объекта, которые могут быть получены с использованием специально построенных аналитических моделей диагностируемого оборудования.

При наличии достаточного объема исходной информации для построения диагностических моделей могут быть применены различные алгоритмы машинного обучения.

Прогнозирование изменения технического состояния диагностируемых объектов с износовым характером отказа (загрязнение фильтра, изменение зазоров в подшипниках, изменение свойств масла) может осуществляться с использованием методов, используемых для прогнозирования временных рядов, таких как рекуррентные нейронные сети с ячейками долгой краткосрочной памяти.

Литература

1. Федотов М.В., Грачев В.В., Грищенко А.В., Кручек В.А., Будюкин А.М., Кондратенко В.Г. Математическая модель системы смазки тепловозного дизеля // Вестник Ростовского государственного университета путей сообщения. – 2020. – № 4. – С. 64-80.
2. Бринк Х., Ричардс Дж., Феверолф М. Машинное обучение. – СПб.: Питер, 2017. – 336 с.
3. Жерон О. Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем. – СПб.: Альфа-книга, 2018. – 688 с.
4. Грачев В.В. Прескриптивный контроль энергоэффективности тепловоза с использованием интеллектуальных методов обработки измерительной информации. – СПб.: ПГУПС, 2019. – 134 с.
5. Осовский С. Нейронные сети для обработки информации / пер. с польского Рудинского И.Д. – М.: Финансы и статистика, 2002. – 344 с.
6. Пустыльник Е.И. Статистические методы анализа и обработки наблюдений. – М.: Наука, 1968. – 288 с.
7. Джулли А., Пал С. Библиотека Keras – инструмент глубокого обучения. Реализация нейронных сетей с помощью библиотек Theano и TensorFlow. – М.: ДМК Пресс, 2018. – 294 с.

УДК 004.8; 621.791

ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА КОНТРОЛЯ СВАРОЧНЫХ РАБОТ

Болотов С.В., *к.т.н.*
e-mail: s.v.bolotov@mail.ru
Захарченков К.В., *к.т.н.*
Макаров Е.В.
Фурманов В.А.
Белорусско-Российский университет

1. ВВЕДЕНИЕ

Интеллектуальные системы управления процессами сварки начинают выходить за рамки исследовательских организаций и широко применяться в промышленности, что способствует переходу к Индустрии 4,0, внедрению в сварочном производстве киберфизических систем [1].

Системы контроля сварочных работ наиболее востребованы на объектах трубопроводного транспорта [2]. Производители сварочного техники с мировым именем поставляют в комплекте с поставляемым оборудованием автоматизированные системы контроля сварочных работ: ESAB WeldQAS [3], Kemppi Weld Eye [4], Fronius Weld Cube [5], Ewm Xnet [6], Lorch Q-Data [7], отдельные отечественные производители выпускают регистраторы сварочных процессов, например, НПФ «ИТС» [8], которые позволяют собирать данные о параметрах режима сварки и передавать их на сервер.

Актуальной задачей является разработка интеллектуальных алгоритмов обработки полученных с регистраторов данных [9-11].

2. ОПИСАНИЕ АППАРАТНОЙ ЧАСТИ И ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ СИСТЕМЫ

Интеллектуальная система контроля сварочных работ (рис.1) включает аппаратную и программную часть.

Аппаратная часть системы содержит регистратор сварочных процессов РСП-БРУ 20-01, блок датчиков и пульт сварщика. Регистратор осуществляет идентификацию сварочного оборудования, сварщиков и руководителей сварочных работ, определение GPS координаты, считывание с сервера задания на сварку, запись и передачу

на сервер параметров режима сварки. Компенсационные датчики фирмы LEM, работа которых основана на эффекте Холла, обеспечивают измерение мгновенных значений сварочного тока и напряжения на дуге с частотой 10 кГц. До передачи в программный модуль обработки данных полученные с датчиков значения усредняются по 1000 точек. Значения температуры и влажности окружающего воздуха, температуры предварительного подогрева и межслойной температуры поступают в регистратор с пульта сварщика. При выходе параметров режима сварки за пределы допустимых значений, установленных инструкцией на технологический процесс сварки, на пульт сварщика поступают соответствующие сигналы, каждый из которых соответствует определенному звуку, генерируемому пультом.

Рис.1. Структура интеллектуальной системы контроля сварочных работ

Алгоритм работы аппаратной части системы (рис.2) включает три основных блока:

1) Блок идентификации. В данном блоке на основе RFID-меток выполняется идентификация руководителей сварочных работ, сварщиков и сварочного оборудования. Определение местонахождения сварочного оборудования осуществляется на основе GPS-координаты.
2) Блок подготовки регистратора к работе. В данном блоке формируется перечень заданий сварщику с указанием допустимых значений параметров сварочного процесса, выполняется считывание температуры и влажности воздуха, температуры предварительного подогрева и межслойной температуры с пульта сварщика.

3) Блок регистрации параметров процесса сварки. Регистрация параметров процесса сварки осуществляется с частотой 10 кГц. При выходе параметров за пределы допустимых значений на пульте сварщика генерируется звуковой сигнал, отличающийся в зависимости от параметра, вышедшего за пределы допустимых значений. До передачи в программную часть системы значения параметров сварочного процесса усредняются по 1000 точек.

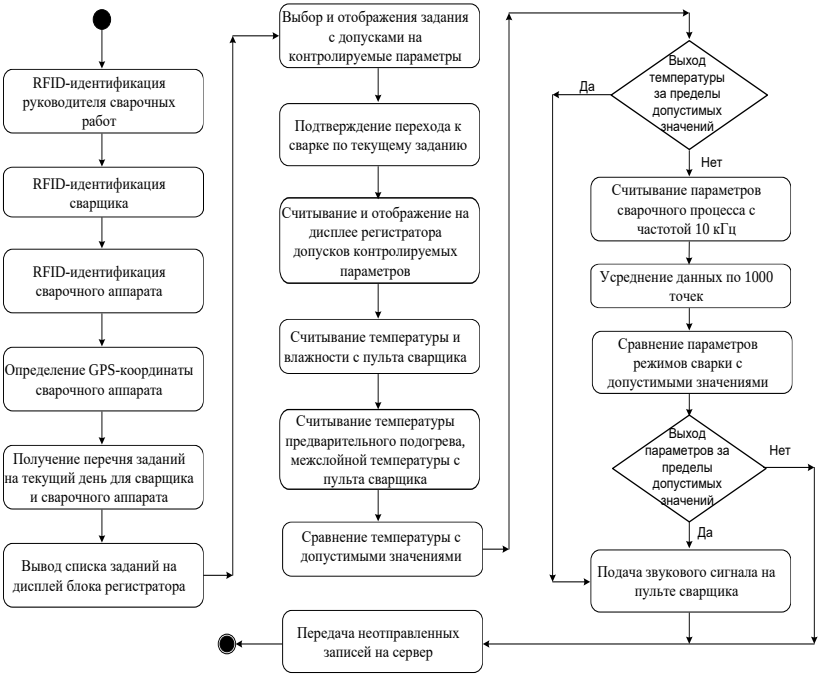


Рис.2. Алгоритм работы аппаратной части регистратора

Программная часть интеллектуальной системы контроля сварочных работ включает следующие основные модули (рис. 1):

1) Модуль обработки данных с регистраторов. В данный модуль поступает информация с регистраторов и с трех других модулей программной части системы. Модуль предназначен для организации обмена предварительно обрабатываемыми данными между аппаратной и программной частью системы. Для контроля сварочных работ на основе параметров режимов сварки осуществляется предварительная обработка данных с регистраторов с последующей передачей полученных значений в модуль

интеллектуального контроля сварочных работ. На основании информации о сварочных процессах, сварщиках и оборудовании из технологических инструкций ежедневно формируются задания на выполнение сварки на конкретных объектах. В данном модуле осуществляется формирование основных документов и отчетов по результатам выполнения сварочных работ (паспорта сварного шва, отчетов о работе сварщиков и сварочного оборудования).

2) Модуль обработки данных о сотрудниках обеспечивает ввод, изменение, удаление и поиск данных о руководителях сварочных работ или мастерах, бригадах и сварщиках. Результаты обработки информации в данном модуле предназначены для оценки соответствия квалификации сварщика технологическим процессам сварки. С учетом результатов обработки данных о сотрудниках осуществляется формирование заданий на выполнение сварочных работ.
3) Модуль обработки данных о технологических процессах сварки обеспечивает ввод, изменение, удаление и поиск данных о технологических инструкциях. Результаты обработки информации в данном модуле учитываются при формировании заданий на выполнение сварочных работ.
4) Модуль интеллектуального контроля сварочных работ предназначен для оценки работы сварщиков и сварочного оборудования. В данном модуле реализована методика интеллектуального контроля сварочных работ на основе результатов обработки данных о режимах сварки. По результатам работы модуля результаты оценки качества сварочного процесса каждый 0,5 с передаются на пульт сварщика.

При реализации программной системы использованы React как основная библиотека пользовательского интерфейса; ASP. Net Core – для реализации бизнес-логики системы; PostgreSQL для создания базы данных; EntityFramework для эффективного взаимодействие объектов базы данных с серверной частью приложения; LabView для сбора и обработки данных с регистратора сварочных процессов.

3. РЕЗУЛЬТАТЫ АППРОБАЦИИ СИСТЕМЫ

Интеллектуальная система контроля сварочных работ может применяется при ручной дуговой (MMA) и полуавтоматической (MIG/MAG) сварке [12]. На рис.3 представлены зафиксированные регистратором сварочных процессов осциллограммы тока и напряжения с частотой оцифровки 10 кГц. Сварка (наплавка) производилась на пластины 100х200х10 мм из стали Ст.3 проволокой

Св08Г2С диаметром 1,2 мм в среде защитного газа CO_2 при сварочном токе 200 А, напряжении 21,3 В, скорости сварки 28,8 м/ч.

Анализ представленных осциллограмм показывает, что перенос электродного металла в сварочную ванну осуществляется короткими замыканиями дугового промежутка [13-15].

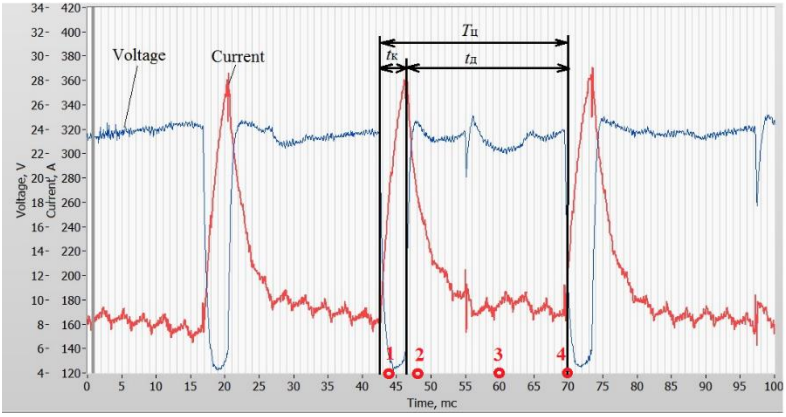


Рис.3. Осциллограммы процесса сварки

В ходе экспериментальных исследований установлено, что программную обработку осциллограмм следует производить в двух частотных диапазонах [16, 17]:

1) Среднечастотный (до 10 кГц) характеризует процессы переноса капель электродного металла в сварочную ванну;
2) Низкочастотный (до 10 Гц) характеризует процесс формирования сварочной ванны (сварного шва).

Можно выделить следующие стадии процесса переноса (рис.3):

- горения дуги (образования капли) с длительностью $t_д$;
- короткое замыкание в сварочную ванну (отрыв капли) длительностью $t_к$.

Полный цикл процесса переноса: $T_ц = t_д + t_к$.

Для исследования процессов переноса электродного металла применяют специальные информационно-измерительные комплексы [18, 19]. На рис.4 представлены картины образования капли за цикл переноса её в сварочную ванну, соответствующие моментам времени обозначенными точками 1..4 на рис.3. Кадры получены высокоскоростной съёмкой и совмещены с осциллограммами тока и напряжения с помощью информационно-измерительного комплекса, разработанного в Белорусско-Российском университете [20].

Стабильность процесса сварки обеспечивается формированием капель одинаковых размеров. Авторами работ [13, 18, 21] исследовались различные критерии стабильности процесса переноса электродного металла. В работах [22, 23] установлено, что наиболее тесную связь с размером (диаметром) капли имеет энергия дуги на этапе её формирования (за время $t_д$), которая определяется:

$$q_д = \sum (i_д \times u_д \times \Delta t),$$

где $i_д \times u_д$ – мгновенные значения сварочного тока и напряжения на дуге; Δt – шаг дискретизации сигнала с датчиков тока и напряжения.

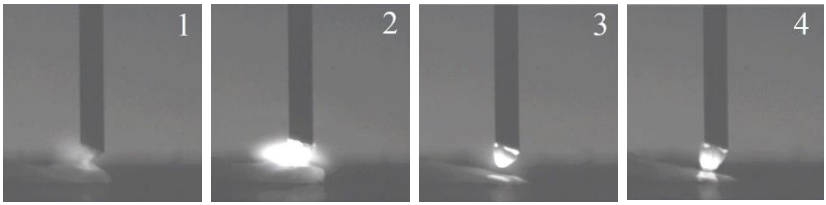


Рис.4. Картины формирования капли расплавленного металла

Энергия дуги на этапе формирования капли электродного металла зависит от параметров режима сварки и изменяется под действием внешних возмущений на систему источник питания – сварочная дуга. Образцовые значения $q_д$ определяются при сварке контрольных образцов определённого материала на фиксированных параметрах режима сварки: сварочного тока, напряжения на дуге, скорости сварки, вылете электрода, расходе защитного газа и поступают в модуль обработки данных с регистраторов.

Для исследования процесса формирования сварочной ванны или сварного шва производится усреднение мгновенных значений тока и напряжения, снимаемых регистратором сварочных процессов с частотой 10 кГц. При усреднении 1000 точек получаем данные о параметрах режима сварки с частотой 10 Гц (рис.5). Отклонения сварочного тока и напряжения от установленных технологическим процессом сварки значений приводят к изменению геометрических размеров сварного шва, к образованию дефектов, таких как: непровар, прожог, поры, трещины и т.д.

Методика интеллектуального контроля сварочных работ включает следующие основные шаги.

1. При обнаружении момента зажигания дуги (резкий рост напряжения практически от нуля до рабочего) идёт расчёт энергии дуги $q_д$ до момента короткого замыкания капли в сварочную ванну (ток

возрастает от рабочего до тока короткого замыкания). Значения энергии дуги $q_д$ сохраняются в память регистратора.

Результаты регистрации мгновенных значений энергии дуги представлены в таблице 1.

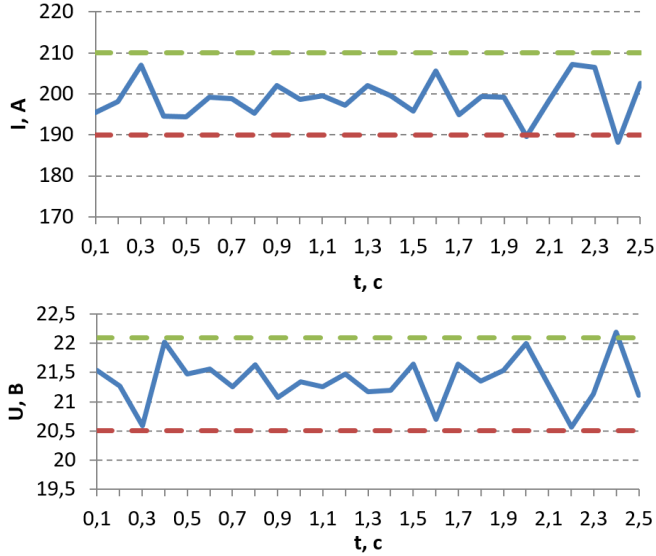


Рис.5. Средние значения тока и напряжения, получаемые с регистратора с частотой 10 Гц

2. За интервал времени 0,5 с (постоянная времени кристаллизации сварочной ванны) происходит расчёт коэффициента вариации энергии дуги $KVq_д$:

$$KVq_д = \frac{\sigma_{q_д}}{\overline{q_д}},$$

где $\sigma_{q_д}$ – среднеквадратичное отклонение значений энергии дуги $q_д$; среднее значение энергии дуги $q_д$, полученных в течение 0,5с.

Результаты расчета коэффициента вариации энергии дуги представлены в таблице 2.

Таблица 1. Результаты регистрации мгновенных значение энергии дуги в течение 2,5 с

№ п/п	Время, с				
	0,5	1	1,5	2	2,5
	Энергия дуги qд				
1	5,069	5,666	8,051	8,197	5,023
2	5,292	5,823	8,475	7,729	5,593
3	6,857	6,86	5,241	8,468	6,254

4	5,598	6,006	6,24	6,39	6,566
5	6,65	5,97	6,017	7,129	6,635
6	6,844	6,479	7,595	6,658	6,542
7	5,49	6,709	6,83	5,451	5,443
8	6,425	5,975	7,828	6,75	5,374
9	5,58	6,685	5,856	3,998	4,686
10	6,304	6,242	5,434	5,48	7,332
11	3,489	4,692	6,535	8,164	8,448
12	1,286	2,224	5,405	6,142	9,39
13	7,441	3,656	5,506	9,965	8,455
14	10,99	13,18	2,87	10,64	5,623
15	9,41	10,35	10,6	8,7	6,338

Таблица 2. Результаты расчета коэффициента вариации энергии дуги

Время, с	0,5	1	1,5	2	2,5
KVqд	0,34	0,38	0,26	0,23	0,20

3. Мгновенные значения сварочного тока и напряжения усредняются на регистраторе и передаются в модуль обработки данных с частотой 10 Гц. (см. рис.6).

4. Общая оценка работы сварщика и оборудования складывается из трех компонентов (оценки сварочного тока *I*о, оценки напряжения *U*о и оценки стабильности процесса сварки на основе коэффициента вариации энергии дуги $KVq_д$) выставляется каждые 0,5 с. При выходе параметра за границы заданного допуска в течение 0,5 с и более – оценка по параметру выставляется равной нулю. В рассматриваемом случае максимальное значение сварочного тока составляет 22,1 А, минимальное значение сварочного тока составляет 20,5 А; максимальное значение напряжения составляет 210 В, минимальное значение напряжения составляет 190 В. Значение коэффициента вариации энергии дуги $KVq_д$ не должно превышать 0,3 [23-26].

Оценка сварочного тока выполняется по следующей формуле:

$$I_o = \frac{\left|\overline{I_{0,5}} - I_{mid}\right|}{(I_{\max} - I_{mid})},$$

где $\overline{I_{0,5}}$ – среднее значение сварочного тока в течение 0,5 с; I_{mid} – среднее между максимальным и минимальным значением сварочного тока (в рассматриваемом случае I_{mid} = 21,3 А); $I_{\max}$ – максимальное значение сварочного тока.

Оценка напряжения выполняется по следующей формуле:

$$U_o = \frac{\left|\overline{U_{0,5}} - U_{mid}\right|}{(U_{\max} - U_{mid})},$$

где $\overline{U_{0,5}}$ – среднее значение напряжения в течение 0,5 с; U_{mid} – среднее между максимальным и минимальным значением напряжением (в рассматриваемом случае U_{mid} = 200 В); $U_{\max}$ – максимальное значение напряжения.

На основании экспертных оценок определены уровни значимости для оценки сварочного тока ($\delta I_o = 3$), напряжения ($\delta U_o = 3$) и коэффициента вариации энергии дуги ($\delta KVq_д = 4$). Соответственно, оценка сварочных работ по 10-бальной системе рассчитывается по формуле:

$$Mw = \delta I_o \times I_o + \delta U_o \times U_o + \delta KVq_o \times KVq_o .$$

Результаты оценки работы сварщиков и сварочного оборудования для рассматриваемого случая представлены в таблице 3.

Таблица 3. Результаты оценки работы сварщиков и сварочного оборудования

Время, с	Uo	Io	$KVq_д$	Оценка (Mw)
0,5	197,95	21,38	0,34	5,1
1	197,99	21,40	0,38	5,0
1,5	199,94	21,26	0,26	6,3
2	199,61	21,27	0,23	6,7
2,5	196,37	21,56	0,20	5,3

4. ЗАКЛЮЧЕНИЕ

В статье описана интеллектуальная система контроля сварочных работ, содержащая аппаратную и программную часть. Аппаратная часть в виде регистратора обеспечивает получение в реальном времени параметров сварочных процессов для оценки качества сварочных работ. Программная часть реализует обработку данных и выставление оценок в зависимости от параметров сварочных процессов.

Представленная в статье интеллектуальная система контроля позволяет в реальном времени (с интервалом 0,5 с) получать оценки качества сварочных работ в зависимости от стабильности процесса сварки и отклонений параметров режима сварки от оптимальных значений, а также предупреждать сварщика и руководителя сварочных работ о выходе параметров режима сварки за пределы допустимых значений посредством генерации звукового сигнала на пульте.

Предложенная методика интеллектуального контроля сварочных работ отличается использованием аддитивного показателя

нормированных значений отклонений тока сварки и напряжения от оптимальных значений, использованием коэффициента вариации энергии дуги на этапе образования капель электродного металла. Результатом применения предложенной методики в разработанной интеллектуальной системе является оценка сварочных работ по 10-бальной шкале, которая передается на пульт сварщика с интервалом 0,5 с и на сервер. Полученная информация может использоваться сварщиками и руководителями сварочных работ для принятия мер по устранению отклонений параметров сварочных процессов от оптимальных, что способствует повышению качества сварки.

Литература

1. Baicun Wanga , S. Jack Hub , Lei Suna , Theodor Freiheita. Intelligent welding system technologies: State-of-the-art review and perspectives. // Journal of Manufacturing Systems. Volume 56, July 2020, pp. 373-391.
2. Колесников О. И., Юшин А. А., Гончаров Н. Г. Анализ применения автоматизированных систем контроля выполнения сварочных работ на объектах трубопроводного транспорта // Наука и технологии трубопроводного транспорта нефти и нефтепродуктов. 2018. Т. 8. № 6. С. 686-691.
3. ESAB: официальный сайт компании / WELDQAS. https://www.esab.ru/ru/ru/products/esab-digital-solutions-eds/quality-assurance/weldqas.cfm.
4. KEMPPI: официальный сайт / Программное обеспечение для управления сварочным производством. https://www.kemppi.com/ru/offering/category/programmnoe-obespecenie-2/programmnoe-obespecenie-dla-upravlenia-svarocnym-proizvodstvom-2/.
5. Fronius: официальный сайт / WELDCUBE. https://www.fronius.com/en/welding-technology/innovative-solutions/weldcube
6. EWM: официальный сайт / Ewm Xnet – система управления сварочными процессами согласно концепции Welding 4.0. https://www.ewm-group.com/ru/produkte/software/ewm-xnet.html.
7. LORCH: официальный сайт / Система документирования сварочных данных Q-DATA. https://lorch.ru/qdatalorch/.
8. ИТС: официальный сайт / Регистратор параметров сварочных процессов РСП-102Д. http://www.npfets.ru/catalog/ets/prochee/registratoru/registrator_parametrov_svarochnuh_processov_rsp_102d/.
9. Казаков С. И. Информационно-компьютерные технологии в сварочном производстве. Курган : Изд-во Курганского гос. университета, 2013. 114 с.
10. Сас А. В., Чернов А. В. Информационно-измерительные системы в управлении сварочным производством : монография. Новочеркасск : Изд-во Южно-Российского гос. политехнического университета, 2008. 148 с.

11. Оценка сварочных свойств оборудования с инверторными источниками питания по энергетическим характеристикам / Э. А. Гладков [и др.] // Сварка и диагностика. 2011. № 1. С. 31-35.
12. Куликов, В. П. Технология и оборудование сварки плавлением и термической резки / В. П. Куликов. – Минск : Экоперспектива, 2003. – 416 с.
13. Потапьевский А.Г. Сварка в защитных газах плавящимся электродом. Ч.1 / А.Г. Потапьевский.– К.: Экотехнология, 2007. – 192 с.
14. Choi S.K. Dynamic Simulation of Metal Transfer in GMAW - Part 1: Globular and Spray Transfer Modes / S.K Choi., C.D. Yoo, Y.S Kim // Welding Journal. 1998, № 11, p. 38-44.
15. Choi S.K. Dynamic Simulation of Metal Transfer in GMAW - Part 2: Short-Circuit Transfer Mode / S.K Choi., C.D. Yoo, Y.S Kim// Welding Journal. 1998. № 12. p. 45-51.
16. Юшин А.А. Разработка критериев оценки сварочных свойств установок для дуговой сварки с управляемым каплепереносом: автореф. дис. канд. техн. наук (15.03.2012) / А.А. Юшин (МГТУ); – Москва, 2012. – 16 с.
17. Оценка сварочных свойств оборудования с инверторными источниками питания по энергетическим характеристикам / Э.А. Гладков, А.А. Юшин, Р.А. Перковский и др. // Сварка и диагностика. 2011. № 1. – С. 31-35.
18. Ленивкин В.А. Технологические свойства сварочной дуги в защитных газах: 2-е изд. / В.А. Ленивкин, Н.Г. Дюргеров, Х.Н. Сагиров. – М.: Гефест, 2011. – 368 с.
19. Сараев, Ю.Н. Комплекс для исследования процессов дуговой сварки / Ю.Н. Сараев, А.Г. Лунев, А.С. Киселёв, А.С. Гордынец, М.В. Тригуб // Автоматическая сварка. – 2018. – № 8. – С. 15-24.
20. Болотов С.В., Хомченско А.В., Шульга А.В., Болотова Е.Л. Информационно-измерительный комплекс для исследования процессов плавления и переноса электродного металла при дуговой сварке // Вестник Брянского государственного технического университета / 2020. – Т. 2020, № 6 (91). – С. 4-11.
21. Pan J. Arc welding control, Jiluan Pan, Woodhead Publishing Limited, Cambridge. England. 2003. – 604 p.
22. Bolotov S.V. Investigation of the Criteria for Evaluating Electrode Metal Transfer in Short Circuit Gas-Shielded Arc Welding [Electronic resource] // IOP Conference Series: Materials Science and Engineering: IOP Conf. Ser.: Mater. Sci. Eng., 2021. – №1118. P.0120003.
23. Букаров В.А. Оценка стабильности дуговой сварки по осциллограммам процесса с использованием статистических методов / В.А. Букаров, С.С. Ермаков, Т.А. Дорина // Сварочное производство. 1990. № 12. с. 31-33.
24. Березовский Б.М. Математические модели дуговой сварки. Т. 2 / Б.М. Березовский. – Челябинск: Изд. ЮУрГУ, 2003. – 601 с.
25. Лебедев В.А. Управление проплавлением при механизированной сварке и наплавке / В.А. Лебедев // Автоматическая сварка. 2011. № 1. С.3-11.
26. Babkin A.S. Identification of Welding Parameters for Quality Welds in GMAW / A.S. Babkin, E.A. Gladkov// Welding Gournal, 2016, №1, pp. 37-48.

ABSTRACTS

Section 5

REINFORCEMENT LEARNING IN THE INTELLIGENT ROBOT CONTROL TASK

***D.A. Dobrynin**, PhD*
e-mail: rabota51@mail.ru
Russian State University for the Humanity,
Federal Research Center for Computer Science and Control RAS, Moscow

The paper describes an approach to reinforcement learning for the dynamic JSM method in the control problem of an intelligent robot. The simulation of training in a software simulator is carried out, and examples are given. It is shown that the training work can be carried out in real time.

Keywords: robots, machine learning, reinforcement learning, JSM-method.

MODELS OF ROBOTS TEAMS BEHAVIOR

***A.A. Kulinich,** PhD*
e-mail: alexkul@rambler.ru
Institute of Control Sciences RAS

The work considers command behavior models of reactive robots and intelligent robots with quality BDI architecture. For reactive robots, the principles of the formation of spontaneous teams based on the principles of social organization of people are determined. The architecture of an intelligent robot with quality BDI architecture is proposed. For intelligent robots, behavioral planning models are proposed, and the conditions for their cooperation are formulated. The proposed principles for the formation of teams of reactive robots and the conditions for cooperation of intelligent robots were investigated in the simulation of the mechanisms of formation and functioning of robot teams.

Keywords: reactive robot, flock, team, intelligent robot, behavior, society.

APPLICATION OF MORAL TANDARDS IN BEHAVIOR MODELLING OF ARTIFICIAL AGENTS

***V.E. Karpov**, PhD*
e-mail: karpov_ve@mail.ru
P.S. Sorokoumov
e-mail: petr.sorokoumov@gmail.com
NRC "Kurchatov Institute"

This article describes a model of a social agent, whose behavior can be stated in terms of basic moral mechanisms and norms. Morality is considered here as a flexible adaptive mechanism that allows agents to vary behavior depending on the environment conditions. The control system of the social agent is based on the "emotion-requirement" architecture. Together with the mechanisms of imitative behavior and the identification of other observable agents with the subjective "Me" concept, this architecture allows to interpret the agent's behavior in terms of empathy, sympathy, and friend-foe relationships. Experiments with this model are described, the main variable parameter of which was the tendency to sympathy. The objective of the experiments was to determine the dependence of the group "well-being" indicators on their altruism. The results obtained are quite consistent with the well-known sociological conclusions, which made it possible to say that the proposed behavioral models and architecture of agents are adequate to intuitive ideas about the role and essence of morality. Thus, the possibility of transition in this area from abstract humanitarian reasoning to constructive schemes and models of adaptive behavior of artificial agents was demonstrated.

Keywords: social behavior models; moral agent, animat, emotional-needs architecture, empathy, sympathy.

INTELLIGENT CONTROL SYSTEM FOR HAPTIC MULTIMODAL USER INTERFACE

E.I. Goluburdo
e-mail: vixryawa@yandex.ru
V.V. Leushina
e-mail: wandbpand@gmail.com
P.S. Sorokoumov
e-mail: petr.sorokoumov@gmail.com
NRC "Kurchatov Institute"

A hardware-software complex of haptic user interface, controlled by a system of fuzzy rules, is described. The complex includes devices for the formation of vibration, temperature, compression and electrical effects on the user, which can be used both together and separately. The rules describing the operation of the complex allow linking the triggering of its components in a certain mode and events in the mobile robotic complex. The experiments showed the efficiency of the system and made it possible to outline the directions for the development of the work.

Keywords: haptic interface, fuzzy logic, vibroactuators.

APPROACH TO MODELING THE BEHAVIOR OF ARTIFICIAL ENTITIES

***G.P. Vinogradov**, DSc*
e-mail: wgp272ng@mail
Tver State Technical University

The analysis of approaches to the construction of artificial entities has shown that there is a gap between the formal models of their behavior and the expectations of users. The approach to their development based on behavior patterns is justified. Ten components of the pattern are identified. It is shown that the coordination of ideas about the architecture of an artificial entity by the TOTHE method allows us to determine an effective pattern model.

Keywords: decision-making, fuzzy judgment, pattern, choice situation.

PATTERNS IN COGNITIVE MODELS OF BEHAVIOR OF AUTONOMOUS INTELLECTUAL SYSTEMS

D.A. Chapelle
e-mail: shapel@cps.tver.ru
Research Institute "Centerprogrammsystem"

An autonomous intelligent system should be considered as a cyber-physical object. We consider the logical and mathematical abstractions that are the basis for building their onboard control systems. Behavioral intellectualization of such systems based on pattern theory is relevant. It is shown that this ensures the transfer of effective experience and ensures the compatibility of the theological approach and the approach based on cause-and-effect relationships. Problems of identification and construction of pattern models are considered. It is proposed to use scheme of four information processing positions for this purpose, and a method of logical inference on patterns is developed.

Keywords: autonomous systems, behavior patterns, fuzzy logic, decision making.

Section 6

SITUATION MONITORING ON THE BASE OF TWOMODAL WORLD MODEL

***A.A. Kharlamov**, DSc*
e-mail: kharlamov@analyst.ru
Institute of Higher Nervous Activity and Neurophysiology of RAS
R.M. Zharkoy
e-mail: roman.jarkoi@iss.ru
ISS (International group of companies)

The paper describes the mechanism of a bi-modal representation of a situation as its textual description and its representation on a 2.5D model, a convenient tool to model a situation. The model of the world consisting of two parts (linguistic and extralinguistic) makes it possible to effectively use the capabilities of video analytics (including those based on artificial neural networks) and semantic analysis of textual information (using artificial neural networks as well). The co-use of these two parts of the model of the world also enables implementation of effective interaction of such a representation with user while maintaining the level of detail of representations related to the world.

Keywords: bi-modal model of the world, language model, 2.5D model, video analytics, semantic text analysis, artificial neural networks.

NEURAL NETWORK CRYPTOGRAPHIC OBFUSCATION FOR TRUSTED CLOUD COMPUTING

***V.L. Eliseev**, PhD*
e-mail: vlad-eliseev@mail.ru
JSC InfoTeCS,
National Research University "MPEI"
E.A. Miliukova
e-mail: k.milyush@bk.ru
National Research University "MPEI"
S.V. Kolpinskiy
e-mail: skolpinsky@bk.ru
JSC InfoTeCS,
National Research University "MPEI"

A problem of trusted computation in the paradigm of cloud processing system and the related adversary model are discussed. Obfuscation, traditional and homomorphic cryptography are reviewed as candidate technologies. An approach to create a trusted cloud computing system with the help of neural network obfuscation of encrypted data processing is introduced.

Keywords: artificial neural networks, obfuscation, encrypted data processing, cloud computing, homomorphic encryption.

APPLICATION OF AN APPROACH BASED ON HYBRID INTELLIGENT INFORMATION SYSTEMS IN GENERATING QUESTIONS FOR RUSSIAN-LANGUAGE TEXTS

M.A. Belyanova
e-mail: belyanovama@student.bmstu.ru
Bauman Moscow State Technical University

The research is dedicated to the description of the Hybrid Intelligent Information System (HIIS) application in question generation for texts in the Russian language. It contains the architecture of HIIS and the comparison of neural networks model quality trained on two different Russian language datasets.

Keywords: Hybrid Intelligent Information System (HIIS), Natural Language Processing (NLP), neural networks, sequence-to-sequence.

MULTILINGUAL NEURAL MACHINE TRANSLATION FOR 7 TURKISH-RUSSIAN LANGUAGE PAIRS

A.F. Khusainov*, PhD*
e-mail: khusainov.aidar@gmail.com
D.Sh. Suleymanov, *DSc*
e-mail: dvdt.slt@gmail.com
Tatarstan Academy of Sciences

This article presents the results of experiments on the use of various methods and algorithms in creating machine translation systems for 7 Russian-Turkic language pairs. We proposed a semi-automatic procedure and created parallel corpora with a total volume of about 7.9 million sentence pairs. As a basic algorithm, we used a neural network approach based on the Transformer architecture. For the first time experiments were conducted for the Turkic languages on the use of transfer learning based on united Russian-Turkic parallel corpus. Experiments show that the systems fine-tuned on the base multilingual system are superior in quality to the basic Russian-Turkic translators. As the results of our last experiment, we have shown that a single multilingual model trained on the united Russian-Turkic corpus with additional language tags can show results comparable to fine-tuned models.

Keywords: multilingual machine translation, transfer learning, Turkic languages.

SELF-SUPERVISED TRAINING FOR THE TATAR SPEECH RECOGNITION SYSTEM

***A.F. Khusainov**, PhD*
e-mail: khusainov.aidar@gmail.com
***D.Sh. Suleymanov**, DSc*
e-mail: dvdt.slt@gmail.com
***I.R. Mukhametzyanov**, PhD*
e-mail: iIlnur.mukhametzyanov@gmail.com
Tatarstan Academy of Sciences

In this paper, we show the results of experiments on the creation of the Tatar speech recognition system based on self-supervised learning. We constructed a 328-hour unlabeled and a 129-hour annotated corpora. As the base model, we used the multilingual XLSR model, pre-trained it on collected unlabeled data, and then fine-tuned it on an annotated corpus. The resulting accuracy on the Common Voice (read speech) test dataset is WER 6.48%, on the Tatar Corpus (read clean speech) is 6.46%, and for the spontaneous speech dataset collected from the TV shows is 28.73%, all of the results are the best-published results on these datasets.

Keywords: self-training, speech recognition, the Tatar language.

COMPARATIVE ANALYSIS OF ANYTIME SOLUTION SEARCH ALGORITHMS FOR INTELLIGENT REAL TIME SYSTEMS

***A.P. Eremeev**, DSc*
e-mail: eremeev@appmat.ru
I.A. Poliushkin
e-mail: poliushkinia@yandex.ru
M.D. Sergeev
e-mail: sergeevmd@mpei.ru
National Research University "MPEI"

The paper presents a comparative analysis of anytime algorithms for finding a solution, built as a result of modification (conversion) of algorithms for finding a solution in the state space. Computer tests of these algorithms have been carried out in order to use them as part of real-time intelligent systems such as intelligent decision support systems of real-time.

Keywords: intelligent system, intelligent decision support system, real time, anytime algorithm.

APPLICATION OF ARTIFICIAL NEURAL NETWORKING TECHNOLOGIES FOR DETECTING DENIAL OF SERVICE ATTACKS

***D.V. Soloviev**, PhD*
e-mail: 9218964588@mail.ru
SPbSUT
***Yu.A.Gatchin**, PhD*
e-mail: gatchin1952@mail.ru
***V.I. Polyakov**, PhD*
e-mail: v_i_polyakov@mail.ru
ITMO University

The main purpose of Distributed Denial of Service attacks (DDoS attacks) is to lead the information system equipment to a state when it cannot serve legitimate users due to its limited resources. It is implemented by a virus infection of ordinary users' networks or smart IoT devices by the subsequent inclusion of such computers and devices into a botnet, which will after become a source of massive distributed attack. This article describes a method for early detection of such attacks using artificial neural network algorithms. The structure of an automated neural network system for detecting attacks on an information system is given.

Keywords: DDoS attacks, artificial neural networks, DDoS attacks detection, automated neural network system for detecting attacks.

DEVELOPMENT AND TESTING OF THE PEDESTRIAN MOTION MONITORING SYSTEM ON THE CONVENTIONAL NEURAL NETWORK PLATFORM

***A.V. Mezhenin**, PhD*
e-mail: mejenin@maill.ru
***V.I. Polyakov**, PhD*
e-mail: v_i_polyakov@mail.ru
ITMO University
***V.V. Izvozchikova**, PhD*
e-mail: viza-8.11@mail.ru
Orenburg State University
***A.G. Zykov**, PhD*
e-mail: zykov_a_g@mail.ru
ITMO University

The paper discusses existing solutions for monitoring pedestrian traffic, their features. Open source solutions were tested and evaluated. The issues of using convolutional neural networks in monitoring systems are discussed in detail. The issues of testing monitoring systems based on special data sets (Crowd Dataset) are considered separately. It is noted that such data are sometimes insufficient to conduct full-scale studies. It is proposed to use synthetic datasets (virtually constructed datasets) for testing software for detection, motion monitoring and object classification. It is proposed to create test videos using special 3D modeling and animation software, the so-called "animatics". Comparison of the two testing methods showed that test results on synthetic data correlate with test results on real data. The author's algorithm is proposed, which made it possible to improve the accuracy of work, while maintaining an acceptable speed of work. Comparison with existing solutions was carried out on synthetic and real data. It was possible to improve the accuracy of recognition and further counting of pedestrians, ensuring an acceptable speed of work.

Keywords: object tracking systems, pedestrian tracking, human detection, object detection, video surveillance, object movement monitoring, pattern recognition, convolutional neural networks, artificial intelligence, deep learning, YOLO, test suites, crowd datasets, synthetic datasets.

THE CALCULATION OF SHIP SPEED ON THE BASIS OF A NEURAL NETWORK

V.V. Deryabin
e-mail: gmavitder@mail.ru
Admiral Makarov State University of Maritime and Inland Shipping

A neural network based system, calculating ship longitudinal speed, is developed. The system takes into account not only propeller revolution rate, but also rudder and wind effect. Training dataset is generated using vessel motion modelling. The results of the test of the neural network shows its high-precision ability to predict the speed of the ship.

Keywords: ship speed, dead reckoning, neural network, machine learning.

Section 7

THE MODEL OF EVOLUTION AND LEARNING OF KAUFFMAN NK-AUTOMATA

***V.G. Red'ko**, PhD*
e-mail: vgredko@gmail.com
Scientific Research Institute for System Analysis,
Russian Academy of Sciences

The model of evolution and learning of Kauffman NK-automata has been constructed. The initial simulations of the evolution of the population of agents, which genotypes are coded by NK-automata, demonstrated the important role of mutations in optimizing the structure of agents.

Keywords: Kauffman NK-automata, evolution, learning, number of attractors.

HYBRID ALGORITHM FOR SOLVING THE DESIGN OF ELECTRONIC COMPUTING EQUIPMENT

***L.A. Gladkov**, professor*
e-mail: leo_gladkov@mail.ru
***Y.M.J.**, postgraduate student*
e-mail: yasir_82@mail.ru
***N.V. Gladkova**, senior teacher*
e-mail: nadyusha.gladkova77@mail.ru
Southern Federal University

The article deals with the problem of placing different-sized elements. The structure of a hybrid algorithm for solving the problem of placing various-sized elements is proposed, modified genetic operators, as well as components of the hybrid algorithm are developed. The hybrid algorithm is implemented in the form of an application program, and a series of computational experiments have been carried out.

Keywords: design automation, placement problem, hybrid methods, bioinspired algorithms, genetic algorithms, fuzzy control.

INTEGRATED APPROACH TO SOLVING DESIGN PROBLEMS

S.N. Shcheglov, PhD
e-mail: srg_sch@mail.ru
ITA IKTIB SFU

The paper considers an integration approach to solving the problems of structural design by the example of solving the problem of placement. The IPA technology is presented, which allows applying bioinspired methods to solving design and management problems. An algorithm scheme based on natural systems has been developed. The results of computational experiments are shown.

Keywords: design, technology, algorithm, method, computational experiment.

COMPARATIVE EVALUATION OF ALGORITHMS FOR AUTOMATIC CONSTRUCTION OF NONLINEAR MODELS BASED ON METAHEURISTIC PROGRAMMING WITH GENE EXPRESSION

***O.G. Monakhov**, PhD*
e-mail: monakhov@rav.sscc.ru
***E.A. Monakhova**, PhD*
e-mail: emilia@rav.sscc.ru
Institute of Computational Mathematics and Mathematical

A new algorithm of metaheuristic programming with gene expression based on various nature-inspired algorithms as a search engine for solving the problem of automatic synthesis of nonlinear models in an analytical form is proposed. We show that the proposed algorithm is superior to the standard gene expression programming algorithm.

Keywords: metaheuristic programming, gene expression programming, nature-inspired algorithms, automatic construction of models.

A MULTIMEME MODIFICATION OF THE QUANTUM FRESHWATER HYDRA ALGORITHM FOR SOLVING THE OPTIMIZATION PROBLEM

D.V. Maikov
e-mail: maykovd@yandex.ru
Izhevsk Trade and Economic College

A multimeme modification of the quantum freshwater hydra algorithm. A Particle Swarm Optimization and freshwater hydra algorithms based on the Analytic Hierarchy Process and the Bayesian approach were used as memes. For all test functions, the memetic modification of the quantum freshwater hydra algorithm of using the original version of this algorithm as a meme showed the best results.

Keywords: optimization problem, meme, memetic algorithm, quantum freshwater hydra algorithm.

Section 8

THEORETICAL AND APPLIED BASIS OF CONSTRUCTING THE INTELLIGENT CONTROL SYSTEM FOR PROTECTION OF SPACECRAFTS FROM ANTHROPOGENIC (TECHNOGENIC) PARTICLES IMPACT

B.V. Palyukh
e-mail: pboris@tstu.tver.ru
V.K. Kemaykin
e-mail: vk-kem@mail.ru
Tver State Technical University
M.L. Tsurkov
e-mail: mlts69@yandex.ru

The paper discusses the problem of increasing the survivability of spacecraft (SC) with long-term orbital operation and areas of autonomous operation under conditions of possible impacts of anthropogenic (technogenic) particles (AP). Algorithms have been developed for predicting the conditions and results of AP strikes, assessing and choosing a protective SC barrier taking into account the current conditions and predicting possible SC damage based on the integrated use of artificial intelligence methods. The principles of construction and functioning of intelligent control systems for SC protection are formulated.

Keywords: intelligent system, spacecraft, machine learning, knowledge base.

INTELLIGENT ASSESSMENT OF IMPACT OF HYPERVELOCITY ANTHROPOGENIC PARTICLES ON SPACE APPARATUS

V.V. Meshkov, *DSc*
e-mail: msf-tgtu@yandex.ru
G.A. Lopin, *DSc*
e-mail: georgi-l@mail.ru
N.N. Filatova, *DSc*
e-mail: nfilatova99@mail.ru
Tver State Technical University

A methodology for intelligent evaluating the parameters of impacts of hypervelocity antropogenic particles on the elements of the Space Apparatus is proposed, combining data on complex physical processes obtained using two types of complementary methods of estimation-experimental and computer modeling. The structure has been developed and functions of an integrated intelligent information system that assess and predict the results of the impact of hypervelocity antropogenic particles on the elements of the Space Apparatus have been described. The proposed intelligent assessment of the impact of hypervelocity anthropogenic particles on the spacecraft can be used in the development of theoretical and applied foundations for building an intelligent control system for the protection of spacecraft from particle impacts, as well as for optimizing the composition and design of elements of promising spacecraft.

Keywords: space apparatus, hypervelocity antropogenic particles, impact, damage, integrated intelligent system, fuzzy logic, experiments, modeling.

MODELS TO COUNTER THREATS OF INFORMATION SECURITY VIOLATIONS THROUGH A VIBRATING CHANNEL IN OFFICE-TYPE PREMISES

***V.V. Alekseev**, PhD*
e-mail: vvalex1961@mail.ru
M.V. Moiseeva
e-mail: mariyamoiseeva@mail.ru
***A.V. Yakovlev**, PhD*
e-mail: yava73@bk.ru
Tver State Technical University

Addressing the challenge of protecting sensitive information, especially for small organizations, is important. It is proposed to implement a unified systemic approach to assessing the security of information based on the measurements carried out and the formation of control influences on the elements of the information protection system of such organizations. A stand has been created for the software and hardware complex "Means of protection of information from leakage through technical channels", special software has been developed to study the process of leaking information through technical channels and optimize the characteristics of the system of protection of corporate information. To set up the protection of information of acoustic and vibroacoustic channels, software is being developed to investigate and implement acoustic interferences based on speech synthesis, as well as to regulate the level of interference to meet the requirements of the security of the room and at the same time comfortable conversation in it.

Keywords: acoustic and speech-like interference, vibrating channels of voice information leakage, model, monitoring, software and hardware complex, countering information security violation, information security features.

INTELLIGENT CONTROL OF DUAL-PURPOSE ROBOT MANIPULATOR BASED ON A COMBINATION OF NEURAL NETWORK AND FUZZY METHODS

***V.V. Borisov**, DSc*
e-mail: vbor67@mail.ru
Branch of National Research University "MPEI" in Smolensk
A.M. Sokolov
e-mail: andreisokol98@gmail.com
V.S. Lyferov
e-mail: lyferov@yandex.ru
National Research University "MPEI"

The paper describes an approach to intelligent control of dual-purpose and special-purpose robot manipulators based on the use of a combination of a neural network model and a fuzzy iterative algorithm. The neural network model is used to find an approximate solution to the inverse kinematics problem. The proposed fuzzy iterative method is used to refine the obtained solution and allows the robotic arm to be controlled in conditions where it is impossible to accurately identify the position of the target object of capture.

Keywords: intelligent control, robot manipulator, artificial neural network, fuzzy iterative method.

ADAPTIVE MOVEMENT CONTROL OF A COLLECTIVE OF MOBILE ROBOTS DEPLOYED IN A LINE

***G.E. Veselov**, DSc*
e-mail: gev@sfedu.ru
***B.K. Lebedev**, DSc*
e-mail: lebedev.b.k@gmail.com
***O.B. Lebedev**, PhD*
e-mail: lebedev.ob@mail.ru
South Federal University

This work solves the problem of controlling the movement of a team of robots deployed in a line. The control method consists in the fact that in the process of movement at each moment of time *t*, the parameters of movement of each robot R_i, moving in parallel with the neighboring robot R_{i-1}, are rescheduled. Management is carried out using an alternative collective adaptation algorithm based on the ideas of the collective behavior of adaptation objects. The principles of functioning of one adaptation automaton are considered. The goal of controlling the slave robots is to minimize deviations. To implement the adaptation mechanism, the parameters of the vector are compared with adaptation automata that simulate the behavior of adaptation objects in the environment. The structure of the process of alternative collective adaptation has been developed, under the control of which the movement of a group of robots in a formation is carried out.

Keywords: swarm of mobile robots, movement of mobile robots deployed in a line, group control, alternative collective adaptation.

INFRASTRUCTURE DEVELOPMENT FOR DESIGNING A BIOINSPIRED UNDERWATER ROBOT

A.R. Khanbekyan
e-mail: artemkhan1@yandex.ru
MIPT
V.V. Karpov
e-mail: kvv96@mail.ru
V.V. Vorobiev
e-mail: vorobev_vv@nrcki.ru
NRC "Kurchatov Institute"

The paper describes an infrastructure kit for creating a bioinspired AUV, which must solve the problem of covert observation of the water area. The kit includes autonomous and remote-controlled robots, for testing and working out various software, hardware, technical and technological solutions, etc. in their natural habitat. Also a method for synchronizing the effectors of a bioinspired AUV, which was tested on this kit, is presented.

Keywords: underwater robot, synchronization, bioinspired robot.

Section 9

CLOUD-BASED PLATFORM FOR DRUG THERAPY PLANNING

***V.V. Gribova**, DSc*
e-mail: gribova@iacp.dvo.ru
***D.B. Okun**, PhD*
R.I. Kovalev
IACP FEB RAS

The paper describes development of a decision support system for prescribing medication based on a specialized shell. The system allows you to prescribe a personalized treatment in various fields of medicine. The shell is based on the ontological approach, in accordance with which all information resources – knowledge and data bases are formed on the basis of ontologies. The authors describe the architecture, components and implementation of this shell. A knowledge base editor for a knowledge base formation and a system ready for use by doctors are implemented as cloud services.

Keywords: ontology; knowledge base; artificial intelligence; decision support system; medical informatics

USING MACHINE LEARNING MODELS TO PREDICT THE RISK OF ATRIAL FIBRILLATION AFTER AORTOCORONARY BYPASSING

***K.I. Shakhgeldyan**, DSc*
e-mail: carinash@vvsu.ru
V.Yu. Rublev
e-mail: dr.rublev.v@gmail.com
***B.I. Geltser**, DSc*
e-mail: boris.geltser@vvsu.ru
***B.O. Shcheglov**,*
e-mail: b.shcheglov@mail.ru
Far Eastern Federal University
***S.N. Shcheglova**, PhD*
e-mail: schegloff@mail.ru
Natural Sciences of North-Eastern State University

In this paper authors perform a comparative analysis of the quality metrics of multivariate logistic regression, stochastic gradient boosting, and an artificial neural network models with continuous and boundary-binarized preoperational factors predicting the risk of atrial fibrillation after CABG in patients who do not have episodes of atrial fibrillation in the preoperative period.

Keywords: predictive models, stochastic-gradient boosting, artificial neural network, features selection, post-operational fibrillation.

IMPLEMENTATION FEATURES OF THE PROTOTYPE OF AN INTEGRATED EXPERT SYSTEM FOR THE COMPLEX DIAGNOSIS OF BREAST DISEASES: INTEGRATION WITH IMAGE PROCESSING

***G.V. Rybina**, DSc*
e-mail: gvrybina@mephi.ru
A.A. Grigoryev
National Research Nuclear University "MEPhI"

The paper is devoted to the further development of research and development in the field of integrated expert systems (IES), the construction of which is carried out on the basis of a problem-oriented methodology.The focus is on the integration of the basic components of the IES with new functional components that implement some methods of image processing obtained as a result of X-ray and ultrasound studies. Complex diagnostics of breast diseases was considered as an experimental problem area for the implementation of the IES prototyping processes.

Keywords: integrated expert systems, problem-oriented methodology, AT-TECHNOLOGY workbench, integration, image processing, medical diagnostics, mammary gland.

IMPLEMENTATION OF FUZZINESS OF THE HEREDITARY DISEASES CLINICAL PICTURE IN AN EXPERT DIAGNOSTIC SYSTEM

N.A. Blagosklonov
e-mail: nblagosklonov@frccsc.ru
Federal Research Center "Computer Science and Control" of RAS

The study explores an approach to operating with the fuzziness of clinical manifestations of hereditary diseases, based on a combination of three expert evaluations. The contribution to a portrait of diseases of such components as modality, manifestation and severity is considered. Simultaneous accounting of these attributes and comparison of the manifestations of the disease in a patient with a standard expert description make it possible to carry out differential diagnostics at the pre-laboratory stage of diagnosis.

Keywords: expert system, certainty factors, confidence measure, orphan diseases, expert knowledge, fuzziness of clinical manifestations.

DEVELOPMENT OF AN EXPERT MEDICAL DECISIONS SUPPORT SYSTEM KNOWLEDGE BASE FOR THE DIAGNOSIS OF PATIENTS WITH GASTROINTESTINAL BLEEDING

A.V. Budykina
e-mail: budykina.rsmu@yandex.ru
Pirogov Russian National Research Medical University

The scientific work presents the knowledge base of the intelligent system supporting clinical decision-making in the field of diagnosis of patients with gastrointestinal bleeding. The article describes the process of extracting knowledge in this clinical area, dividing into groups and linking them together, also this article describes the process of coordinating expert opinions on the clinical picture, anatomical localization and severity degree of bleeding using a questionnaire method. The aim of the study is to develop a medical system knowledge base for supporting clinical decision-making in the field of diagnosis of patients with suspected gastrointestinal bleeding to be implemented in clinical practice. Materials and methods of research. There is some formalized information from foreign and domestic clinical recommendations for the diagnosis of patients with gastrointestinal bleeding are presented in the graph database management system Neo4j, the coordination of expert decisions was carried out using Microsoft Excel. Results and discussion. There were 605 unique terms allocated, divided into 8 groups and connected by 1762 links, each link of which belongs to one of 12 types. Based on the classifications of gastrointestinal bleeding and the selected unique terms, questionnaires were developed on the base of the clinical picture, anatomical localization and severity of bleeding, which were filled out and approved by experts with the participation of a knowledge engineer.

Keywords: expert system for clinical decision support, knowledge base, clinical recommendations, gastrointestinal bleeding, classification, diagnosis, medical expert, questionnaire method, knowledge structuring.

MICROFLUIDIC CHIP-BASED PLATFORM FOR EX VIVO ANGIOGENESIS RESEARCH

V.A. Glotov, *DSc*
e-mail: forssma@yandex.ru
A.L. Demidov
V.V. Zhilkin
G.N. Fedorov, *PhD*
I.M. Prudnikov, *PhD*
Smolensk State Medical University

A microfluidic chip-based platform for in vitro angiogenesis research has been developed that allows reproducing the phenomenon of functioning self-developing endothelial capillary networks in endothelial cell culture and conducting a wide range of tissue engineering and biotechnological research on them.

Keywords: endothelium, endothelial capillary networks, in vitro angiogenesis, tissue engineering, biotechnologies, regenerative medicine.

GRADIENT BOOSTING PREDICTIVE MODEL OF OVARIAN RESPONSE FOR HORMONAL THERAPY IN INFERTILITY TREATMENT

***K.Yu. Tikhaeva**, PhD*
e-mail: tikhaeva34@gmail.com
Volgograd State Medical University, Ministry of Health of the Russian Federation,
LLC "Clinic Family"
N.S. Nesterova
e-mail: sunflovertash@mail.ru
E.V. Tomilov
S.I. Sotkin
Analytical Laboratory Sodas LLC
***A.N Sulima.**, DSc*
e-mail: gsulima@yandex.ru
FSAEI HE "Kazan Federal University named after V.I. Vernadsky, LLC"
Medical Clinic "Your Doctor"
***S.L. Balezin**, PhD*
e-mail: s.balezin@mail.ru
***E.A. Rosyuk**, PhD*
JSC "Center for Family Medicine"

To develop models for predicting the response to hormonal stimulation in the treatment of infertility, assisted reproductive technologies, data from 700 cases of infertility treatment with IVF were used. The forecasting system was built using the R 4.0.3 language. Modeling was carried out based on linear regression, regression trees, regression trees, k-nearest neighbors, gradient boosting.

Keywords: gradient boosting, regression trees, linear regression, k-nearest neighbors, predictive models.

Section 10

INTEGRATION OF THE INTERNET OF KNOWLEDGE AND THE INTERNET OF THINGS TO CREATE AN INDUSTRY 5.0

***G.B. Evgenev**, DSc*
e-mail: g.evgenev@mail.ru
Bauman Moscow State Technical University

A methodology for creating systems of the "Industry 5.0" class using artificial intelligence technologies is proposed. The methodology is based on multi-agent knowledge base creation methods and is suitable for the development of design and control systems for digital intelligent manufacturing. An integrated structure of the Internet of knowledge and the Internet of things has been developed. The life cycle of mechanical engineering products is analyzed and methods of using the Internet of knowledge and the Internet of things at various stages of this cycle are proposed. The functional decomposition of the main stages of the life cycle is presented. The conceptual foundations of the Internet of knowledge are given. Multi-agent methods for creating knowledge bases have been developed. A meta-ontology of engineering agents is proposed. The principles of building multi-agent systems for semi-automatic design of products are described.

Keywords: Industry 4.0, Industry 5.0, digital manufacturing, Internet of knowledge, Internet of things, intelligent systems.

PRINCIPLES OF CREATING AUTOMATED PROCESS CONTROL SYSTEMS IN MECHANICAL ENGINEERING

***G.B. Burdo**, DSc*
e-mail: gbtms@yandex.ru
Tver State Technical University

Mechanical engineering and instrument engineering belong to the high-tech sphere of production and have a number of features that determine a specific approach to process management in production systems. It is shown that automated process management in these production systems should be performed according to the quality parameter while simultaneously implementing measures to reduce time and material costs. It is proposed to apply the product life cycle theory as a methodological basis for the analysis and optimization of processes. One of the most important parameters that determine the efficiency of production systems is the time spent on the implementation of the product life cycle stages implemented within the framework of this enterprise. The measures that allow optimizing production processes are considered.

Keywords: process management, quality management, decision-making, mechanical engineering, instrumentation, production system, product life cycle, decision evaluation criteria.

METHOD FOR SOLVING THE PROBLEMS OF MANAGING THE LIFE CYCLE OF TECHNICAL SYSTEMS

M.V. Ovsyannikov
mvo50@mail.ru
S.A. Bukhanov
bukhanov@yandex.ru
Bauman Moscow State Technical University

Typical tasks of the product life cycle: integrated logistics support, process management, configuration management have a number of common properties, which makes it possible to use a unified approach proposed in this work to solve them. The proposed approach is based on the development of construction of applied ontologies of specific subject areas based on the ontology of the life cycle, describing the objects of the applied area and the interdependencies between them and an algorithm using the concepts of applied ontologies. A universal algorithm for solving structural synthesis problems by the method of satisfying constraints for a wide class of conditions allows the user to correct the problem conditions in the process of solving. The developed solutions have been tested for the tasks of synthesizing the configuration of products, creating production and energy systems, when planning maintenance procedures, etc. The proposed methods make it possible to solve important engineering problems of determining the composition of objects of designed products or formed processes, which can be used at various stages of the life cycle of products.

Keywords: life cycle of CIM products, ontological model, constraint satisfaction problems, CSP.

INTELLIGENT CONTROL OF LOCOMOTIVE TECHNICAL CONDITION

M.V. Fedotov
e-mail: fedotov-mv@vnikti.com
Y.I. Klimenko*, PhD*
e-mail: klimenko-ui@vnikti.com
JSC "VNIKTI"
V.V. Grachev*, DSc*
e-mail: v_grach@mail.ru
Emperor Alexander I St. Petersburg State Transport University

It is proposed to manage the locomotive maintenance and repair system taking into account the results of the actual state assessment. A method for the synthesis of locomotive equipment diagnostic models using analytical models is proposed, which allows creating diagnostic models based on machine learning methods. To predict changes in the state of an object under diagnostics, it is proposed to use recurrent neural networks.

Keywords: diagnostics, condition monitoring, forecast/prediction, locomotive, diagnostic model, computer-assisted learning, neural network, information processing.

INTELLIGENT WELDING CONTROL SYSTEM

***S.V. Bolotov**, PhD*
e-mail: s.v.bolotov@mail.ru
***K.V. Zakharchenkov**, PhD*
E.V. Makarov
V.A. Furmanov
Belarusian-Russian University

An intelligent welding control system, that allowing to receive an assessment of the quality of welding in real time, has been developed. Proposed intelligent control method differs in the use of the arc energy variation coefficient at the stage of formation of drops of electrode metal and additive indicator of normalized values of welding current and voltage deviations from optimal values.

Keywords: intelligent system, control of welding works, welding process recorder, welding mode parameters, arc energy, assessment of the quality of welding.

АВТОРСКИЙ УКАЗАТЕЛЬ

Алексеев В.В. 222
Балезин С.Л. 319
Белянова М.А. 88
Благосклонов Н.А. 295
Болотов С.В. 369
Борисов В.В. 229
Будыкина А.В. 303
Бурдо Г.Б. 337
Буханов С.А. 347
Веселов Г.Е. 238
Виноградов Г.П. 46
Воробьев В.В. 247
Гатчин Ю.А. 123
Гельцер Б.И. 269
Гладков Л.А. 159
Гладкова Н.В. 159
Глотов В.А. 311
Голубурдо Е.И. 37
Грачев В.В. 356
Грибова В.В. 258
Григорьев А.А. 284
Демидов А.Л. 311
Дерябин В.В. 142
Добрынин Д.А. 6
Евгенев Г.Б. 327
Елисеев В.Л. 77
Еремеев А.П. 115
Жаркой Р.М. 71
Жилкин В.В. 311
Захарченков К.В. 369
Зыков А.Г. 130
Извозчикова В.В. 130
Карпов В.В. 247
Карпов В.Э. 27
Кемайкин В.К. 199
Клименко Ю.И. 356
Ковалев Р.И. 258
Колпинский С.В. 77
Кулинич А.А. 16
Лебедев Б.К. 238
Лебедев О.Б. 238
Леушина В.В. 37
Лопин Г.А. 211
Луферов В.С. 229
Майков Д.В. 190
Макаров Е.В. 369
Меженин А.В. 130
Мешков В.В. 211
Милюкова Е.А. 77
Моисеева М.В. 222
Монахов О.Г. 179
Монахова Э.А. 179
Мухаметзянов И.Р. 106
Нестерова Н.С. 319
Овсянников М.В. 347
Окунь Д.Б. 258
Палюх Б.В. 199
Полюшкин И.А. 115
Поляков В.И. 123
Поляков В.И. 130
Прудников И.М. 311
Редько В.Г. 151
Росюк Е.А. 319
Рублев В.Ю. 269
Рыбина Г.В. 284
Сергеев М.Д. 115
Соколов А.М. 229
Соловьев Д.В. 123
Сорокоумов П.С. 27, 37
Соткин С.И. 319
Сулейманов Д.Ш. 96, 106
Сулима А.Н. 319
Тихаева К.Ю. 319
Томилов Е.В. 319
Фёдоров Г.Н. 311
Федотов М.В. 356
Филатова Н.Н. 211
Фурманов В.А. 369
Ханбекян А.Р. 247
Харламов А.А. 71
Хусаинов А.Ф. 96, 106
Цурков М.Л. 199
Шапель Д.А. 57
Шахгельдян К.И. 269
Щеглов Б.О. 269
Щеглов С.Н. 170
Щеглова С.Н. 269
Яковлев А.В. 222
Ясир М.Д. 159

СОДЕРЖАНИЕ

СЕКЦИЯ 5. **ИНТЕЛЛЕКТУАЛЬНЫЕ АГЕНТЫ, РОБОТЫ И КОЛЛЕКТИВЫ РОБОТОВ**

Добрынин Д.А. Обучение с подкреплением в задаче управления интеллектуальным роботом …… 6

Кулинич А.А. Модели командного поведения роботов ……. 16

Карпов В.Э., Сорокоумов П.С. Применение моральных норм для управления поведением искусственных агентов ……...…….. 27

Голубурдо Е.И., Леушина В.В., Сорокоумов П.С. Интеллектуальная система управления для тактильного мультимодального пользовательского интерфейса ……..... 37

Виноградов Г.П. Подход к моделированию поведения искусственных сущностей ……. 46

Шапель Д.А. Паттерны в когнитивных моделях поведения автономных интеллектуальных систем ……... 57

СЕКЦИЯ 6. **НЕЙРОСЕТЕВЫЕ ТЕХНОЛОГИИ И ИХ ПРИЛОЖЕНИЯ**

Харламов А.А., Жаркой Р.М. Мониторинг ситуации на основе двумодальной модели мира …… 71

Елисеев В.Л., Милюкова Е.А., Колпинский С.В. Нейросетевая криптографическая обфускация для доверенных облачных вычислений …….. 77

Белянова М.А. Применение подхода на основе гибридных интеллектуальных информационных систем для генерации вопросов к русскоязычным текстам …….. 88

Хусаинов А.Ф., Сулейманов Д.Ш. Разработка многоязычной нейросетевой системы машинного перевода для 7 тюркско-русских пар языков …… 96

Хусаинов А.Ф., Сулейманов Д.Ш., Мухаметзянов И.Р. Самообучение на неразмеченных данных для системы распознавания татарской речи 106

Еремеев А.П., Полюшкин И.А., Сергеев М.Д. Сравнительный анализ гибких алгоритмов поиска решения для интеллектуальных систем реального времени …….. 115

Соловьев Д.В., Гатчин Ю.А., Поляков В.И. Применение технологий искусственных нейронных сетей для обнаружения атак на отказ в обслуживании ……..... 123

Меженин А.В., Поляков В.И., Извозчикова В.В., Зыков А.Г. Разработка и тестирование системы мониторинга движения пешеходов на платформе сверточной нейронной сети …… 130
Дерябин В.В. Расчёт продольной скорости судна на основе нейронной сети …… 142

СЕКЦИЯ 7. **ЭВОЛЮЦИОННОЕ МОДЕЛИРОВАНИЕ И БИОНИЧЕСКИЕ АЛГОРИТМЫ**

Редько В.Г. Модель эволюции и обучения NK-автоматов Кауфмана …. 151
Гладков Л.А., Ясир М.Д., Гладкова Н.В. Гибридный алгоритм решения задач проектирования электронно вычислительной техники ……. 159
Щеглов С.Н. Интеграционный подход к решению задач проектирования (размещение) …… 170
Монахов О.Г., Монахова Э.А. Сравнительная оценка алгоритмов автоматического построения нелинейных моделей на основе метаэвристического программирования с экспрессией генов ……. 179
Майков Д.В. Мультимемная модификация квантового алгоритма пресноводных гидр для решения задачи оптимизации …… 190

СЕКЦИЯ 8. **ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ВОЕННОГО, ДВОЙНОГО И СПЕЦИАЛЬНОГО НАЗНАЧЕНИЯ**

Палюх Б.В., Кемайкин В.К., Цурков М.Л. Теоретические и прикладные основы построения интеллектуальной системы управления защитой космических аппаратов от ударов антропогенных (техногенных) частиц ……... 199
Мешков В.В., Лопин Г.А., Филатова Н.Н. Интеллектуальная оценка ударного воздействия гиперскоростных антропогенных частиц на космический аппарат …….. 211
Алексеев В.В., Моисеева М.В., Яковлев А.В. Модели противодействия угрозам нарушения информационной безопасности по виброакустическому каналу в помещениях офисного типа …… 222
Борисов В.В., Соколов А.М., Луферов В.С. Интеллектуальное управление роботами-манипуляторами двойного назначения на основе комбинирования нейросетевых и нечетких методов …… 229
Веселов Г.Е., Лебедев Б.К., Лебедев О.Б. Адаптивное управление передвижением коллектива мобильных роботов, развернутых в линию 238
Ханбекян А.Р., Карпов В.В., Воробьев В.В. О разработке инфраструктуры для создания биоподобного подводного робота …….. 247

СЕКЦИЯ 9. **ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ В МЕДИЦИНЕ**

Грибова В.В., Окунь Д.Б., Ковалев Р.И. Облачная платформа для планирования медикаментозной терапии ……. 258

Шахгельдян К.И., Рублев В.Ю., Гельцер Б.И., Щеглов Б.О., Щеглова С.Н. Методы машинного обучения для прогнозирования риска развития фибрилляции предсердий после коронарного шунтирования ……..... 269

Рыбина Г.В., Григорьев А.А. Особенности реализации прототипа интегрированной экспертной системы для комплексной диагностики заболеваний молочной железы: интеграция с обработкой изображений 284

Благосклонов Н.А. Реализация нечеткости клинической картины наследственных болезней в экспертной диагностической системе …… 295

Будыкина А.В. Разработка базы знаний экспертной системы диагностики пациентов с желудочно-кишечными кровотечениями ….. 303

Глотов В.А., Демидов А.Л., Жилкин В.В., Фёдоров Г.Н., Прудников И.М. Платформа для исследования ангиогенеза IN VITRO ……... 311

Тихаева К.Ю., Нестерова Н.С., Томилов Е.В., Соткин С.И., Сулима А.Н., Балезин С.Л., Росюк Е.А. Модель предикции ответа яичников на гормональную терапию при лечении бесплодия на основе градиентного бустинга …… 319

СЕКЦИЯ 10. **ИНТЕЛЛЕКТУАЛЬНЫЕ ПРОИЗВОДСТВА, УПРАВЛЕНИЕ ЖИЗНЕННЫМ ЦИКЛОМ СИСТЕМ, ТЕХНОЛОГИИ «ИНДУСТРИИ 4.0»**

Евгенев Г.Б. Интеграция интернета знаний и интернета вещей для создания Индустрии 5.0 ……..... 327

Бурдо Г.Б. Принципы создания автоматизированных систем управления процессами в машиностроении …… 337

Овсянников М.В., Буханов С.А. Метод решения задач управления жизненным циклом технических систем …….. 347

Федотов М.В., Клименко Ю.И., Грачев В.В. Интеллектуальное управление техническим состоянием локомотива ……... 356

Болотов С.В., Захарченков К.В., Макаров Е.В., Фурманов В.А. Интеллектуальная система контроля сварочных работ …… 369

Abstracts …… 380

Авторский указатель ……... 400

Научное издание

Сборник трудов
X Международной научно-практической конференции
«Интегрированные модели и мягкие вычисления в искусственном интеллекте» ИММВ-2021
Труды конференции в 2-х томах
Том 2

Подписано в печать 12.05.2021 г.
Формат 60x84 1/16. Тираж 300 экз. Усл. печ. л. 23,48.

Издательство «Универсум»

ISBN 978-5-91412-469-1