

НАУКА

и глобальные вызовы
XXI века

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В РЕШЕНИИ АКТУАЛЬНЫХ СОЦИАЛЬНЫХ И ЭКОНОМИЧЕСКИХ ПРОБЛЕМ XXI ВЕКА

Сборник статей по материалам
Четвертой всероссийской научно-практической конференции,
проводимой в рамках Пермского естественнонаучного форума
«Математика и глобальные вызовы XXI века»

Часть II

г. Пермь, 21–23 мая 2019 г

Пермский государственный национальный исследовательский университет
Пермский национальный исследовательский политехнический университет
Национальный исследовательский университет «Высшая школа экономики»
Пермский государственный гуманитарно-педагогический университет
Пермский государственный медицинский университет им. академика Е. А. Вагнера
Пермское отделение Научного совета РАН по методологии искусственного интеллекта

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В РЕШЕНИИ АКТУАЛЬНЫХ СОЦИАЛЬНЫХ И ЭКОНОМИЧЕСКИХ ПРОБЛЕМ XXI ВЕКА

Сборник статей по материалам
Четвертой всероссийской научно-практической конференции,
проводимой в рамках Пермского естественнонаучного форума
«Математика и глобальные вызовы XXI века»
(г. Пермь, 21–23 мая 2019 г.)

Часть II



Пермь 2019

УДК 004.8: 3
ББК 32.813 + 6/8
И868

Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века: сб. ст. по материалам Четвертой всерос. науч.-практ. конф. (г. Пермь, 21–23 мая 2019 г.) Ч. II. / Перм. гос. нац. исслед. ун-т. – Пермь, 2019. – 179 с.

ISBN 978-5-7944-3291-6 (ч. 2)
ISBN 978-5-7944-3289-3

В сборнике представлены материалы Четвертой всероссийской научно-практической конференции «Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века», которая проводилась 21–23 мая 2019 г. в г. Перми в рамках Пермского естественнонаучного форума «Математика и глобальные вызовы XXI века».

Сборник предназначен для научных и педагогических работников, преподавателей, аспирантов, магистрантов, студентов и всех, кто интересуется проблемами и перспективами развития и применения методов искусственного интеллекта.

**УДК 004.8: 3
ББК 32.813 + 6/8**

*Печатается по решению кафедры прикладной математики и информатики
Пермского государственного национального исследовательского университета*

Редакционная коллегия

Кузнецов Андрей Геннадьевич, канд. техн. наук,
Левченко Елена Васильевна, д-р психол. наук,
Русаков Сергей Владимирович, д-р физ.-мат. наук,
Столбов Валерий Юрьевич, д-р техн. наук,
Хлынова Ольга Витальевна, д-р мед. наук, чл.-корр. РАН

Ответственный редактор

Ясницкий Леонид Нахимович, д-р техн. наук,

*Сборник подготовлен при финансовой поддержке РФФИ: грант № 19-010-00307
Статьи публикуются в авторской редакции, авторы несут ответственность
за содержание статей, за достоверность приведенных в статье фактов, цитат,
статистических и иных данных, имен, названий и прочих сведений*

ISBN 978-5-7944-3291-6 (ч. 2)
ISBN 978-5-7944-3289-3

© ПГНИУ, 2019

Оглавление

ЦИФРОВАЯ ПЕДАГОГИКА	5
ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПРИ ЗАКАЗЕ УЧЕБНОЙ ЛИТЕРАТУРЫ НА ОСНОВЕ АНАЛИЗА ДАННЫХ ВОСТРЕБОВАННОСТИ КНИЖНЫХ ИЗДАНИЙ	6
ИСПОЛЬЗОВАНИЕ НЕЙРОСЕТЕВОГО МОДЕЛИРОВАНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ КАРЬЕРНОГО РОСТА ПРОГРАММИСТА	13
ОНЛАЙН ЛАБОРАТОРИЯ ОНЛАЙН ДИСЦИПЛИНЫ «ВСТРОЕННЫЕ СИСТЕМЫ»	18
РАЗРАБОТКА ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ ДИАГНОСТИКИ И ПРОГНОЗИРОВАНИЯ ОЦЕНКИ ЗА ЭКЗАМЕН	23
ТЕХНОЛОГИИ АНАЛИЗА ДАННЫХ: КЛАССИФИКАЦИЯ, ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА, ЭКСПЕРИМЕНТЫ С МОДЕЛЯМИ	31
МОДЕЛЬ ФОРМИРОВАНИЯ ФУНКЦИОНАЛЬНЫХ СИСТЕМ ПРИ БОС-ТРЕНИНГЕ СТУДЕНТОВ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ	36
НЕЙРОСЕТЕВАЯ МОДЕЛЬ ПРОГНОЗИРОВАНИЯ УСПЕВАЕМОСТИ СТУДЕНТА- ПЕРВОКУРСНИКА	41
ЦИФРОВОЙ СПОРТ	48
ПРОГНОЗИРОВАНИЕ УСПЕШНОСТИ БОЙЦА ММА НА ОСНОВЕ ЕГО ФИЗИОЛОГИЧЕСКИХ ПОКАЗАТЕЛЕЙ	49
ПРОГНОЗИРОВАНИЕ ВЕРОЯТНОСТИ ПЕРЕХОДА ИГРОКА ИЗ КОНТИНЕНТАЛЬНОЙ ХОККЕЙНОЙ ЛИГИ В НАЦИОНАЛЬНУЮ ХОККЕЙНУЮ ЛИГУ	54
ВОЗМОЖНОСТЬ ПРОГНОЗИРОВАНИЯ КОЛИЧЕСТВА ГОЛОВ, ЗАБИТЫХ КОМАНДОЙ ЧЕЛСИ В ОДНОМ МАТЧЕ	60
ЦИФРОВАЯ ЛИНГВИСТИКА И ЦИФРОВОЕ ИСКУССТВО	67
СИНТЕЗ МУЗЫКАЛЬНОГО ПРОИЗВЕДЕНИЯ ПО ТЕСТОВОМУ ОПИСАНИЮ	68
ВЛИЯНИЕ ЭМОЦИОНАЛЬНОЙ НАГРУЗКИ НОВОСТНОГО ТЕКСТА НА ИЗМЕНЕНИЕ ДОХОДНОСТИ АКЦИЙ	73
О ПРИМЕНЕНИИ КЛАСТЕРНОГО АНАЛИЗА ДЛЯ ОПРЕДЕЛЕНИЯ СЕМАНТИКИ СОБЫТИЙ В СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ СИСТЕМАХ	76
АВТОМАТИЗИРОВАННЫЙ СИНТЕЗ МУЗЫКАЛЬНЫХ КОМПОЗИЦИЙ НА ОСНОВЕ РЕКУРРЕНТНЫХ НЕЙРОННЫХ СЕТЕЙ	80
МЕТОД ВЕКТОРНОГО ПРЕДСТАВЛЕНИЯ ТЕКСТА НА ОСНОВЕ ЭНТРОПИИ, ДЛЯ ЗАДАЧИ РАСПОЗНАВАНИЯ ИНТЕНТА В ДИАЛОГОВОЙ СИСТЕМЕ	85
ИДЕИ МЕТОДОВ ПРЕДСТАВЛЕНИЯ И ПОИСКА ОБЩИХ ПАТТЕРНОВ В ТРЕКАХ ВЗГЛЯДА, ПРИ ПРОСМОТРЕ ЛЮДЬМИ ОДНОГО И ТОГО ЖЕ ОБЪЕКТА	90
ПРИМЕРЫ РАЗРАБОТКИ ЧАТ БОТОВ	94
ПРИМЕНЕНИЕ ЧАТ-БОТОВ ДЛЯ РЕШЕНИЯ РАЗНОПЛАНОВЫХ ЗАДАЧ	102

ЦИФРОВАЯ ПСИХОЛОГИЯ И КРИМИНАЛИСТИКА	108
ИСПОЛЬЗОВАНИЕ МЕТОДОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ПРОГНОЗИРОВАНИИ ПРЕДРАСПОЛОЖЕННОСТИ ЧЕЛОВЕКА К АЛКОГОЛИЗМУ	109
ОПРЕДЕЛЕНИЕ СОВЕРШЕННОЛЕТΙΑ КЛИЕНТА МЕТОДОМ НЕЙРОСЕТЕВОГО МОДЕЛИРОВАНИЯ	115
ПРОГНОЗИРОВАНИЕ УРОВНЯ ПРЕСТУПНОСТИ В РЕГИОНАХ РОССИИ	119
ПРИМЕНЕНИЕ НЕЙРОСЕТЕВЫХ ТЕХНОЛОГИЙ В ИЗУЧЕНИИ ФАКТОРОВ, ВЛИЯЮЩИХ НА ПРЕСТУПНОСТЬ В ГОРОДАХ РОССИИ.....	126
ПРОГНОЗИРОВАНИЕ ТЯЖЕСТИ ДОРОЖНО-ТРАНСПОРТНЫХ ПРОИСШЕСТВИЙ	132
ЦИФРОВАЯ ГЕОФИЗИКА	137
РЕЗУЛЬТАТ РЕШЕНИЯ ОПТИМИЗАЦИОННЫХ ЗАДАЧ, СВЯЗАННЫХ С ФИЛЬТРАЦИЕЙ В НЕФТЯНОМ ПЛАСТЕ.....	138
МЕТОДЫ ЧИСЛЕННОГО РЕШЕНИЯ ЗАДАЧ ОПТИМИЗАЦИИ	141
НЕЙРОСЕТЕВОЕ МОДЕЛИРОВАНИЕ ПРОГНОЗИРОВАНИЯ ВЕРОЯТНОСТИ ВОЗНИКНОВЕНИЯ ПЕРКОЛЯЦИОННОГО КЛАСТЕРА	145
ПРОГНОЗИРОВАНИЕ КОНЦЕНТРАЦИИ К-МЕРОВ, НЕОБХОДИМОЙ ДЛЯ ДОСТИЖЕНИЯ ПОРОГА ПЕРКОЛЯЦИИ С ПОМОЩЬЮ НЕЙРОСЕТЕВОЙ СИСТЕМЫ	150
ФОРМАЛИЗАЦИЯ ЗАДАЧИ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПРИ ОБСЛУЖИВАНИИ ТЕХНОЛОГИЧЕСКОГО ОБОРУДОВАНИЯ НЕФТЕДОБЫЧИ.....	156
ЭМОЦИОНАЛЬНЫЕ РОБОТЫ	162
ОБЩИЕ ПОДХОДЫ К МАТЕМАТИЧЕСКОМУ МОДЕЛИРОВАНИЮ ЦИФРОВЫХ ДВОЙНИКОВ	163
ЦИФРОВАЯ ПРОМЫШЛЕННОСТЬ 2	166
МЕТОДЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ПРОГНОЗИРОВАНИЯ, УПРАВЛЕНИЯ И СОВЕРШЕНСТВОВАНИЯ КАЧЕСТВА ОТЛИВОК ЛОПАТОК АВИАЦИОННЫХ ДВИГАТЕЛЕЙ КАК ИНСТРУМЕНТ ПОВЫШЕНИЯ ЭКОНОМИЧЕСКОЙ ЭФФЕКТИВНОСТИ ПРЕДПРИЯТИЯ.....	167
ПРОГНОЗИРОВАНИЕ СТРУКТУРНЫХ СДВИГОВ НА ОСНОВЕ САМООРГАНИЗУЮЩИХСЯ КАРТ КОХОНЕНА.....	173

ЦИФРОВАЯ ПЕДАГОГИКА

ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПРИ ЗАКАЗЕ УЧЕБНОЙ ЛИТЕРАТУРЫ НА ОСНОВЕ АНАЛИЗА ДАННЫХ ВОСТРЕБОВАННОСТИ КНИЖНЫХ ИЗДАНИЙ¹

Болодурина Ирина Павловна, Дусакаева Слушаи Тугайбаевна

Оренбургский государственный университет,
460018, Россия, г. Оренбург, пр. Победы 13, slushashdusakaeva@rambler.ru

В статье рассмотрена актуальная проблема качественного комплектования библиотечного фонда вузовских библиотек, связанная с оптимальным заказом учебной литературы в условиях ограниченного финансирования закупки книжных изданий. Формирование окончательного списка заказа литературы основано на идее максимального удовлетворения информационных потребностей различных типов пользователей библиотечных ресурсов. Для оценки качества комплектования библиотечного фонда выбрана востребованность учебной литературы с учетом библиографических характеристик книжных изданий, приоритетности разделов тематического рубрикатора и информационных предпочтений читателей.

Ключевые слова: библиотечный фонд, востребованность учебной литературы, интегрированная автоматизированная информационная система (ИАИС).

Будем рассматривать деятельность библиотеки как предоставление информационных услуг населению. Организуем процесс обслуживания читателей таким образом, чтобы их информационные потребности были максимально удовлетворены. Для достижения поставленной цели следует выбрать грамотный подход к формированию библиотечного фонда, осуществляя его пополнение и обновление требуемыми источниками удовлетворения информационных потребностей пользователей библиотечных ресурсов и учитывая ограниченность в финансировании по приобретению новых изданий.

Модели оптимизации комплектования библиотечного фонда разрабатывались в трудах отечественных исследователей П.А. Болдырева, Е.В. Зеленцовой, Т.Н. Злобиной, В.В. Орловой, Г.В. Рудиной, Н.В. Соколовой, Р.Т. Усманова, Л.А. Кроминой [8].

Проведенный анализ исследований выявил отсутствие подхода, основанного на формализации процессов выражения информационных предпочтений пользователей. Согласно принятой в библиотековедении методике качество библиотечного фонда оценивают степенью

удовлетворения информационных потребностей пользователей через отношение

$$k_{к.н} = \frac{\text{количество книговыдач}}{\text{объем фонда}} .$$

Однако, ни уменьшение объёма фонда, ни увеличение количества книговыдач не способствует повышению уровня удовлетворения информационных потребностей читательской аудитории. Поэтому качество книжного фонда будем измерять коэффициентом востребованности литературы.

Востребованность библиотечных изданий является интегральной характеристикой, зависящей от многих переменных. К числу наиболее важных из них следует отнести: библиотечные характеристики книг (автор, год издания, издательство и др.); соответствие издания разделам библиотечного тематического рубрикатора информационных предпочтений читательской аудитории; отношение пользователей библиотечных ресурсов к проблеме удовлетворения информационных интересов[1].

Разработанная авторами комплексная модель реализации информационных потребностей пользователей[5] учитывает характеристики востребованных изданий[1,9], приоритетность разделов тематического рубрикатора[4,6,7] и читательские информационные предпочтения различных типов пользователей[3]. Использование комплексной модели при управлении комплектованием библиотечного фонда нуждается в создании системы поддержки принятия решений, разработанной на основе интеллектуальных методов анализа данных востребованной учебной литературы.

В рамках проекта ИАС ОГУ задачи автоматизации библиотечно-библиографических процессов выделены в отдельную функциональную подсистему «Библиотека». На рисунке 1 приведена схема взаимодействия ресурсов распределения автоматизированной информационной системы (АИС) вуза при обработке учебных и библиотечно-библиографических процессов.

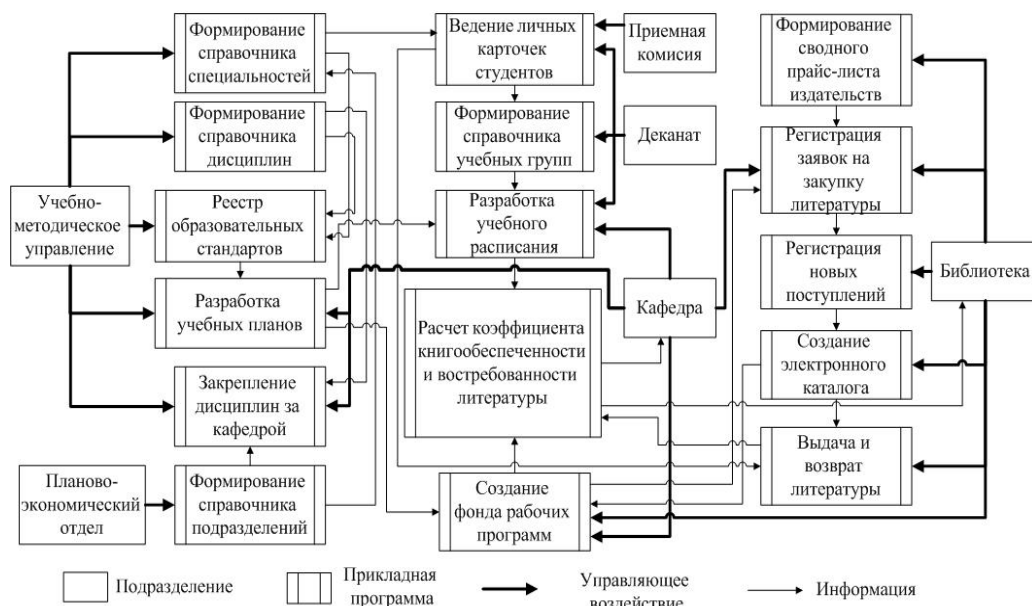


Рисунок 1. Схема взаимодействия ресурсов распределенной АИС вуза

Комплексы задач, решаемые в подсистеме «Библиотека», приведены на рисунке 2. Разработанный для получения актуальных данных построенной модели комплектования библиотечного фонда программный модуль «Востребованность учебной литературы» внедрен в АИС ОГУ и используется для решения комплекса задач «Анализ библиотечно-библиографических процессов».

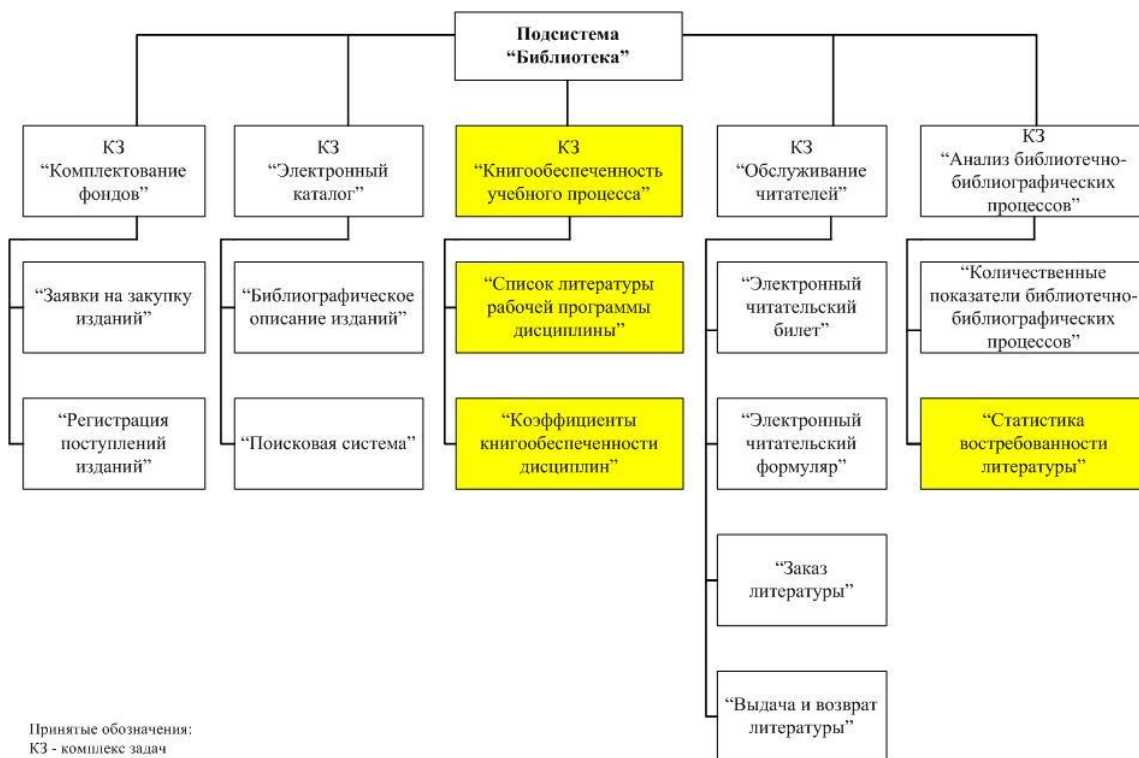


Рисунок 2. Подсистема «Библиотека» ИАС ОГУ

В подсистеме «Библиотека» ИАС ОГУ реализованы следующие комплексы задач:

1. «Электронный каталог»;
2. «Комплектование фондов»;
3. «Книгообеспеченность учебного процесса»;
4. «Обслуживание читателей»;
5. «Анализ библиотечно-библиографических процессов».

Для комплектования библиотечного фонда в научной библиотеке ОГУ успешно реализована программная система «Учет и обработка заявок на закупку литературы», проектирование которой началась в 2007г. Целью создания системы послужила необходимость автоматизации учета заявок, формирования сводных заявок, формирование общего заказа. Основным критерием при заказе учебной литературы в используемой ранее ИСППР являлось удовлетворение показателям книгообеспеченности.

Для комплектования библиотечного фонда востребованными печатными изданиями взаимодействие библиотеки и кафедр университета можно представить схемой, представленной на рисунке 3.



Рисунок 3. Схема взаимодействия кафедр и библиотеки при комплектовании библиотечного фонда

Для комплектования библиотечного фонда востребованными печатными изданиями при решении задачи оптимизации заказа учебной литературы, разработан комплекс алгоритмов, позволяющий производить генерацию и отбор формируемых заказов до тех пор, пока не будет построен оптимальный список заказа учебной литературы. Структура комплекса алгоритмов формирования заказа учебной литературы, обеспечивающего

максимум целевой и функции и удовлетворяющего требованиям ограниченности финансирования закупок, приведена на рисунке 4.

1. «Обработка первичной информации о заказе учебной литературы на востребованность» - процедура, объединяющая все заявки, поступившие с кафедр университета в отдел комплектования библиотечного фонда, для анализа востребованности заказываемых печатных изданий с позиций:

а) «Ранжирование по значимости библиотечно-библиографических характеристик печатных изданий»;

б) «Ранжирование по приоритетности разделов тематического рубрикатора»;

в) «Ранжирование по типам пользователей библиотечных ресурсов».

2. «Генерация общего ранжированного списка заказа учебной литературы» упорядочивает все поступившие заявки по критерию востребованности в порядке убывания (наименее востребованные печатные издания окажутся в конце списка).

3. «Расчет стоимости сгенерированного общего ранжированного списка заказа учебной литературы» производит калькуляцию затрат на приобретение учебной литературы с учетом их стоимости и доставки.

4. «Генерация сформированного списка заказа учебной литературы» в соответствии с уровнем финансирования отсекает наименее востребованную заказываемую учебную литературу или увеличивает количество наиболее востребованной.

5. «Вывод сформированного списка заказа учебной литературы» представляет собой окончательный вариант заказа учебной литературы.

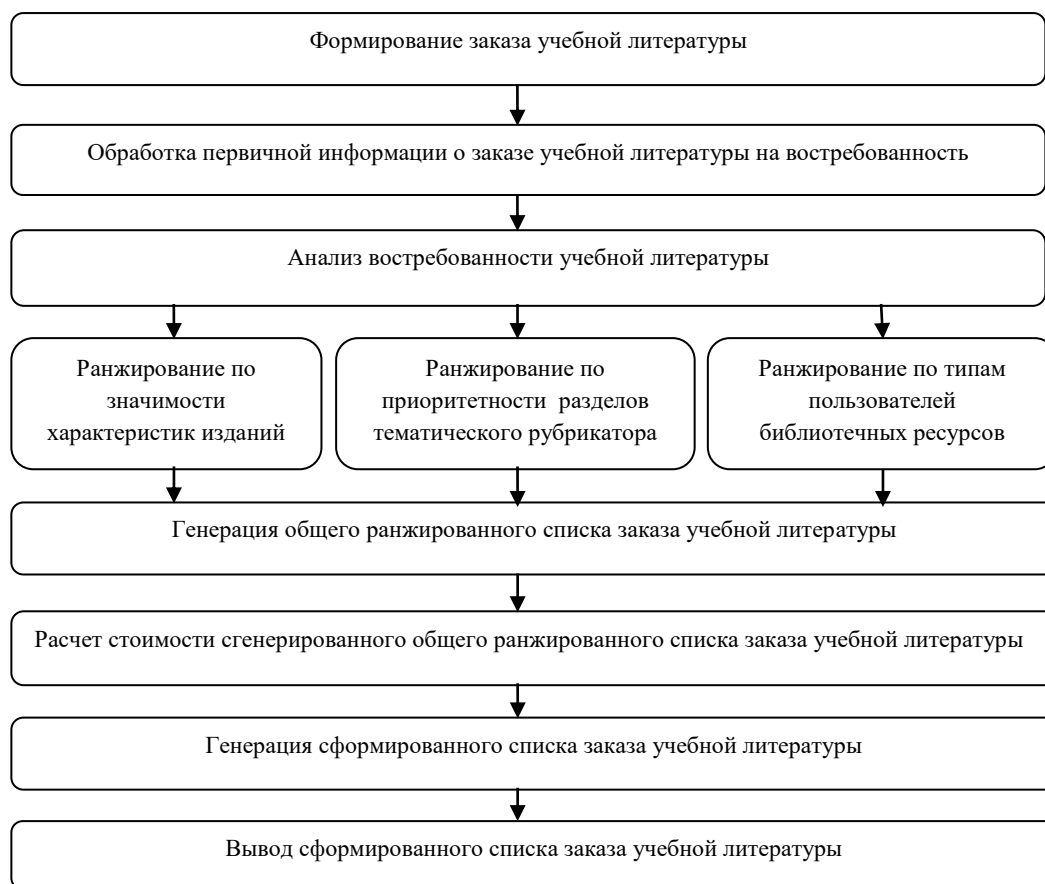


Рисунок 4. Структура комплекса алгоритмов формирования заказа учебной литературы

Предложенная интеллектуальная система поддержки принятия решений при комплектовании библиотечного фонда позволяет рационально использовать экономические ресурсы библиотеки – время на обработку заявок на приобретение учебной литературы и материально-техническую базу. Эффективность разработанной методики подтверждена улучшением показателей деятельности вузовской библиотеки, выражающихся в повышении степени удовлетворения информационных потребностей студентов: качество фонда возросло на 12%, посещаемость на 41%, обращаемость на 72%, читаемость на 80%, обновляемость фонда на 12%, книгообеспеченность на 2%.

Библиографический список

1. Болодурина, И.П. Анализ качества комплектования библиотечного фонда учебной литературы средствами DATA MINING. // И.П. Болодурина, П.А. Болдырев, С.Т. Дусакаева. Вестник Волжского университета имени В.Н. Татищева Серия «Экономика». Выпуск 1 (30). Тольятти: ВУиТ, 2014. – С.101-109.
2. Болодурина И.П. Информационно-аналитическое обеспечение востребованности учебной литературы // И.П. Болодурина, П.А.

- Болдырев, С.Т. Дусакаева. Вестник Оренбургского государственного университета Выпуск 9. – Оренбург. – 2014 - С.168-173.
3. Болодурина И.П. Влияние востребованности учебной литературы на стратегию комплектования библиотечного фонда // И.П. Болодурина, П.А. Болдырев, С.Т. Дусакаева. Интеллект. Инновации. Инвестиции. Выпуск 1 . – Оренбург. – 2015 - С. 8-14.
 4. Болодурина И.П. Методы нечеткой логики при исследовании востребованности учебной литературы [Электронный ресурс] / / И. П. Болодурина, П. А. Болдырев, С. Т. Дусакаева. // Научное обозрение, 2015. - №8470; 14. - С. 224-231.
 5. Болодурина, И.П. Комплексная модель реализации информационных потребностей пользователей в процессе комплектования библиотечного фонда // Болодурина, И.П., Дусакаева, С.Т. Вестник Южно-уральского государственного университета. Том 18, номер 3. – Челябинск. – 2018 – С. 59-67.
 6. Дусакаева, С.Т. Использование методов нечеткой логики при исследовании востребованности учебной литературы // С.Т. Дусакаева. Интернет-журнал «Наукоеведение». Том 7, номер 1 (выпуск 1(26) январь-февраль). Режим доступа: <http://naukovedenie.ru/PDF/114PVN115.pdf>
 7. Friedrich A.: - Logik und Fuzzy-Logik. – Expert Verlag, Stuttgart, 319 S. (2006).
 8. Кромина, Л.А. Автоматизированная информационная система поддержки принятия решений при заказе литературы для вуза на основе ранжирования изданий по уровню потребности: монография / Л.А. Кромина, Р.А. Ярцев; Уфимский государственный авиационный технический университет. – Уфа: Уфимский государственный авиационный технический университет, 2015. – 126 с.
 9. Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям: 2-е изд., перераб. и доп., СПб.: Питер, 2010. –624 с. - с. 445-448.

INTELLIGENT DECISION SUPPORT IN THE ORDERING OF TEXTBOOKS ON THE BASIS OF THE ANALYSIS OF DEMAND FOR BOOKS

Bolodurina Irina Pavlovna, Duskaeva Slushash Tugaibaeva

Orenburg state University, 460018, Russia, Orenburg, Pobedy Ave. 13,
slushashdusakaeva@rambler.ru

The article deals with the actual problem of high-quality acquisition of the library Fund of University libraries associated with the optimal order of educational literature in the conditions of limited funding for the purchase of books. The formation of the final list of references is based on the idea of maximum satisfaction of information needs of different types of users of library

resources. Assessment of the quality of acquisition of the library Fund selected the demand for educational literature, taking into account the bibliographic characteristics of books, the priority sections of the thematic categories and information preferences of readers.

Key words: library Fund, demand for educational literature, integrated automated information system (IAIS).

УДК 004.855.5

ИСПОЛЬЗОВАНИЕ НЕЙРОСЕТЕВОГО МОДЕЛИРОВАНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ КАРЬЕРНОГО РОСТА ПРОГРАММИСТА

Власова Ольга Владимировна

Национальный исследовательский университет «Высшая школа экономики»
614070, Россия, г. Пермь, ул. Студенческая, д. 38,
vlasova.olga22@gmail.com

В статье описан опыт разработки и применения нейросетевой модели для оценки и прогнозирования должности разработчика программного обеспечения. Рассматриваются социальные характеристики, личные предпочтения, знание языков программирования и технологий разработки. Выделены наиболее значимые параметры, выполнено прогнозирование изменения карьерного уровня при изучении ряда технологий.

Ключевые слова: прогнозирование, карьерный рост, программист, поиск персонала, нейронная сеть.

На сегодняшний день совокупный объем мирового рынка ИТ превышает два триллиона долларов США, это один из самых перспективных сегментов экономики, а также источник огромного числа высокооплачиваемых рабочих мест.

Разнообразие сфер, в которых применяются информационные технологии, рост числа направлений разработки и высокая скорость развития технологий создают определенный барьер для людей, желающих развиваться и работать в этой области. Многие студенты математических специальностей и компьютерных наук теряются при выборе технологий и языков программирования, которые бы обеспечили им быстрый карьерный рост и финансовое благополучие.

Другие аспекты с высоким уровнем неопределенности – это требуемый опыт работы и уровень заработной платы. В сфере ИТ нередки случаи, когда должности высокого уровня занимают сотрудники с десятилетним опытом, и

со стажем 1-2 года. Величина заработной платы разработчика может варьироваться от 10-15 тысяч до нескольких миллионов рублей в месяц [1].

Специалистам по подбору персонала бывает сложно быстро оценить, подойдет ли человек для определённой должности, какой величины оклад ему предложить и т.п. Многие компании, занимающиеся поиском и наймом разработчиков, проводят свои исследования с целью оценить состояние рынка труда ИТ-специалистов. К примеру, компания Stack Overflow, развивающая популярную систему вопросов и ответов о программировании, ежегодно проводит опрос пользователей своей платформы. В 2017 году в опросе участвовало более 64 000 респондентов из 213 стран [1]. Список вопросов учитывал такие аспекты, как образование, опыт работы, демографические характеристики, используемые технологии, должность и роль в компании, заработную плату, удовлетворение от работы и другие. По результатам опроса был составлен отчёт, отражающий тренды и наиболее популярные ответы респондентов. К примеру, в США наиболее востребованными и высокооплачиваемыми специалистами являются специалисты по машинному обучению, а в мире в целом – специалисты DevOps.

Подобные исследования носят в основном описательный характер, то есть показывают текущее состояние рынка, и не предлагают каких-либо прогнозов или рекомендаций по отдельно взятым параметрам. По-видимому, единственным исключением на сегодняшний день является нейронная сеть [5], прогнозирующая карьерный рост выпускников Пермского государственного гуманитарно-педагогического университета и позволяющая студентам подбирать оптимальную траекторию своего обучения. Нейронная сеть выложена в свободном доступе в разделе «Проекты» сайта www.PermAi.ru.

В настоящей работе для обучения нейронной сети использованы данные опроса пользователей сайта HackerRank – разработчиков и специалистов по найму. Они были получены из открытой платформы для специалистов, занимающихся машинным обучением, Kaggle [2, 3].

Общий объем исходного набора данных состоит из 25 тысяч строк и 250 параметров, он включает в себя пять файлов: с данными в формате текста, с данными в числовом формате, с расшифровкой названия полей, с кодами стран и с указателями по кодированию исходных данных. Для данного исследования из исходного датасета были выделены только результаты опроса разработчиков. Окончательный объем набора данных составил 6934 примера.

В качестве входных параметров используются ответы на 40 вопросов:

X1 – Код страны,

X2 – Возраст, с которого респондент занимается программированием,

X3 – Возраст на момент опроса,

X4 – Пол,

X5 – Образование,

X6 – Направление обучения,

X7 – Учился ли программировать в университете,
X8 – Учился ли программировать самостоятельно,
X9 – Является ли студентом на момент опроса,
X10 – Текущая роль,
X11 – Индустрия, сфера работы,
X12 – X20 – Условия, приоритетные при выборе работы (профессиональный рост, интересные задачи, размер заработной платы и др.),
X21 – X30 – Знание языков программирования (C, C++, Java, Python, Ruby, C#, JavaScript и др.),
X31 – X40 – Знание фреймворков и технологий разработки (AngularJS, React, Spring, Django, ASP и др.).

Выходной параметр модели **Y** соответствует карьерному уровню кандидата (Студент, Выпускник, Фрилансер, Младший разработчик (junior), Старший разработчик (senior), Ведущий инженер, Архитектор, Главный инженер, Директор или Вице-президент по разработке, Основатель или Генеральный директор).

Набор данных был поделен на обучающее, тестовое и подтверждающее множества. Объем обучающего множества составил 5200 примеров (75%), тестового – 1040 примеров (15%) и подтверждающего – 694 (10%).

Проектирование нейронной сети, дальнейшее обучение, тестирование и исследование проводились с использованием программного продукта «Нейросимулятор 5.0» на основе методики «Пермской научной школы искусственного интеллекта» [4].

Оптимальная структура нейронной сети, полученная в результате экспериментов, представляла собой персептрон с 40 нейронами на входном слое, 20 нейронами на скрытом слое и одним нейроном на выходном слое. В качестве активационной функции для входного и промежуточных слоев использовалась сигмоидная функция, для выходного – тангенс гиперболический. В качестве алгоритма обучения использовался алгоритм быстрого распространения, всего было произведено 3000 итераций.

В результате обучения среднеквадратичная ошибка обучающего множества составила 7,1%, среднеквадратичная ошибка тестирования 7,08%, на валидирующем множестве – 7,67%.

Таким образом, нейронная сеть верно определяла результат для более чем 90% входных данных. На рисунке 1 отражен результат определения значимости параметров, то есть степень их влияния на полученный результат – карьерный уровень специалиста.



Рисунок 1. Значимость параметров нейросети

Проведенный анализ позволил выявить параметры модели, оказывающие наибольшее влияние. Так например, наиболее весомым (около 19%) оказался параметр X9 – является ли кандидат на текущий момент студентом. Такая значимость параметра может быть объяснена тем, что платформой HackerRank пользуется в образовательных целях большое число студентов, еще не трудоустроенных.

Следующие по значимости: X3 – текущий возраст (10%), X2 – возраст, с которого специалист занимается программированием. Среди группы параметров X12-X20 – факторов, приоритетных при выборе работы, наиболее значимыми оказались X15 – баланс рабочего и свободного времени (work/life balance, 4%) и X14 – корпоративная культура (3,2%). В группе X21 – X30 приоритетными оказались параметры X21 – язык С и X26 – язык JavaScript (около 3,4% каждый). В группе X31-X40, отвечающей за знание фреймворков программирование, наиболее значимыми оказались X34 – Django (4,9%) и X32 – React (4,3%).

Далее было проведено прогнозирование карьерного уровня на основе данных студента 4 курса. При прогнозировании значения входных параметров варьировались в соответствии с реально возможными в перспективе изменениями. Результаты прогнозирования приведены в таблице 1.

Таблица 1. Результаты прогнозирования

№	Входные данные	Прогноз	Интерпретация
1	Студент 4 курса, 21 год. Изучен язык C#, технология ASP.NET	1,0432	1 – Студент
2	Выпускник университета, 22 года, веб-разработчик. Языки: C#, JavaScript, Python. Технологии: ASP.NET, React, Node.js	3,759	3 – Фрилансер, 4 – Junior-разработчик
3	Выпускник магистратуры, 24 года, full-stack разработчик. Языки: C#, JavaScript, Python, Java. Технологии: ASP.NET, React, Node.js, Django, Spring.	4,6541	5 – Senior-разработчик

Таким образом, разработанная модель позволяет предсказать изменение карьерного уровня специалиста при изменении параметров модели – предпочтений, навыков, таких как владение языками программирования и фреймворками, и социальных признаков, таких как страна проживания, возраст или уровень образования.

Результаты данного исследования могут быть особенно полезны студентам и начинающим специалистам, выбирающим направление своего профессионального развития, преподавателям, как их консультантам, а также специалистам по управлению персоналом при оценке кандидатов.

Библиографический список

1. Stack Overflow Developer Survey 2017 [Электронный ресурс] URL: <https://insights.stackoverflow.com/survey/2017> (дата обращения: 10.03.2018)
2. HackerRank – 2018 Developer Skills Report [Электронный ресурс] URL: <https://research.hackerrank.com/developer-skills/2018/> (дата обращения: 21.03.2018)
3. Kaggle – HackerRank Developer Survey 2018 [Электронный ресурс] URL: <https://www.kaggle.com/hackerrank/developer-survey-2018/data> (дата обращения: 23.03.2018)
4. Ясницкий Л.Н. Интеллектуальные системы. М.: Лаборатория знаний, 2016. с. 115-149.
5. Ясницкий Л.Н., Кузнецов А.Г., Селезнева С.М., Солохина А.Д., Тюлькина Д.В., Черепанов Ф.М. Применение нейросетевых технологий в изучении акмеологического потенциала студентов вуза // Вестник Пермского университета. Серия: Математика. Механика. Информатика. 2014. № 4. С.120-126.

USING NEURAL NETWORK MODELING IN SOFTWARE DEVELOPER CAREER GROWTH FORECASTING

Vlasova Olga V.

National Research University Higher School of Economics, 38, Studenchskaya st.,
Perm, 614070, Russia, vlasova.olga22@gmail.com

The article covers article the experience of a neural network model development for software developer career level evaluation and forecasting. Social characteristics, personal preferences, knowledge of programming languages and frameworks are examined. The most valuable parameters are highlighted with the completed forecasting of a career level change when a person adds to a stack of known technologies.

Key words: forecasting, career growth, programmer, recruiting, neural network.

УДК 007.52

ОНЛАЙН ЛАБОРАТОРИЯ ОНЛАЙН ДИСЦИПЛИНЫ «ВСТРОЕННЫЕ СИСТЕМЫ»

Гончаровский Олег Владленович

Пермский национальный исследовательский политехнический университет,
614990, Россия, г. Пермь, Комсомольский проспект, д. 29а,
35911953@mail.ru

В статье представлен результаты создания онлайн лаборатории. Работа в такой лаборатории позволяет получить навыки разработки программного обеспечения прототипов подсистем управления (встроенных систем) в режиме онлайн.

Ключевые слова: онлайн лаборатория, встроенная система, быстрое прототипирование технических систем.

Онлайн курс по дисциплине «Встроенные системы» позволяет узнать, как быстро создать прототип подсистемы управления технической системы (рисунок1):

- аппаратную часть прототипа на основе микроконтроллера;
- программу для микроконтроллера.

Онлайн курс позволяет получить навыки разработки программного обеспечения прототипа системы управления используя как концепцию «Сверхцикла», так и концепцию «Многозадачности».

Пусть необходимо разработать новое устройство, которое должно определять состояние объекта управления, воздействовать на него и иметь пользовательский интерфейс. Как убедиться, что концепция устройства работоспособна. Для этого необходимо создать прототип (макет), разработав, изготовив и собрав (например, на контрактном производстве) собственные узлы или же использовать близкие по характеристикам готовые узлы.

Рассматривается второй путь, известный как быстрое прототипирование. Одним из вариантов быстрого прототипирования является "Обратимый взлом" ("reversible hacking") – это модификация серийного устройства, которую можно отменить для возвращения устройства в исходное состояние.

Во многих университетах для освоения компетенций, относящихся к разработке встроенных систем, используется, например, модификация серийного робота-пылесоса с открытым интерфейсом компании iRobot, известного как iRobot Create 2. iRobot Create 2 рассматривается как прототип автоматической самоходной тележки – базовый элемент мобильных роботов.

Если прототип доказывает работоспособность, то разработчик создает проект для его серийного производства, т.е. рассматривает вопросы размеров, цены, энергопотребления, процесса производства, наличия комплектующих, надежности и контроля качества. На рисунке 2 приведен пример прототипа робота-официанта на основе робота-пылесоса.

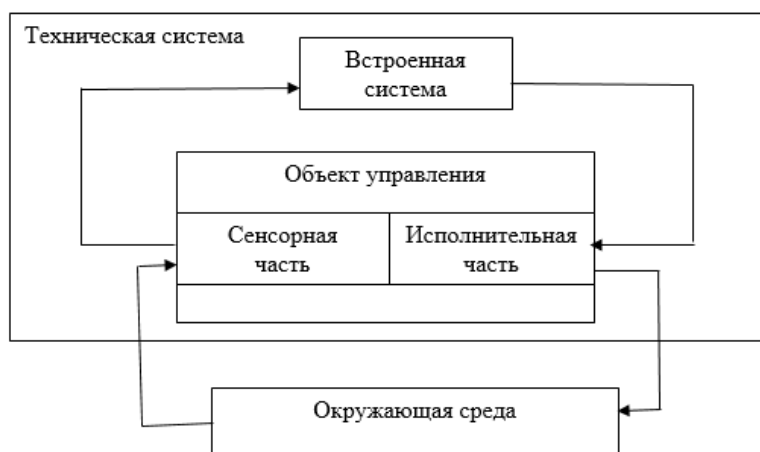


Рисунок 1. Техническая система

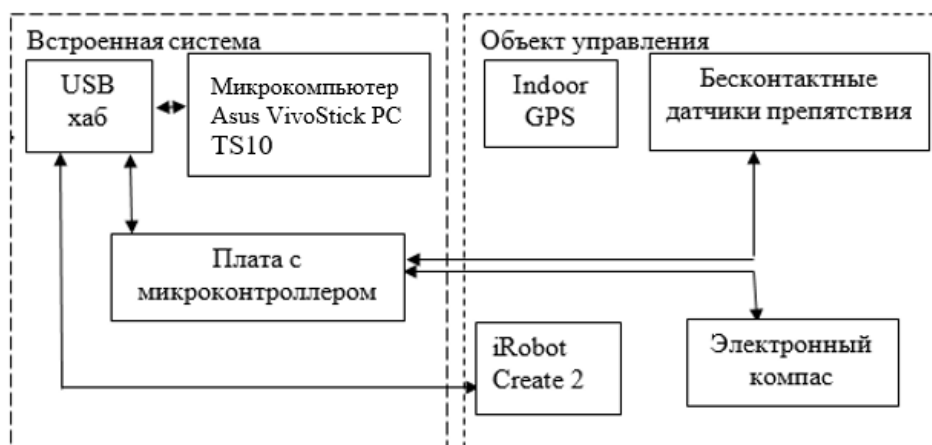


Рисунок 2. Структурная схема робота-официанта на базе робота-пылесоса iRobot Create 2 и микрокомпьютер Asus VivoStick PC TS10

Разработка онлайн лекций дисциплины «Встроенные системы» как и любой другой дисциплины не вызывает особых проблем. Выполнение онлайн лабораторного практикума по разработке программного обеспечения требует создания онлайн лаборатории, которая позволит загружать и тестировать через сеть Интернет разработанное пользователем программное обеспечения для заданного прототипов технической системы, находящейся от автора на значительном расстоянии. Прототип может быть, как статическим, так и мобильным.

Из примера видно, что прототип содержат плату с микроконтроллером. Она выполняет функции сбора информации от сенсорной части и формирование управляющих воздействий для исполнительской части объекта управления. В ряде вариантов она одновременно выполняет роль центрального элемента управления. Наиболее популярной для разработки платы с микроконтроллером является платформа mbed – аппаратно-программная платформа с открытым исходным кодом и одноимённая операционная система (Mbed OS) для устройств на базе 32-разрядных микроконтроллеров семейства ARM Cortex-M. IDE платформы работает онлайн. Аппаратная составляющая платформы содержит платы семейства FRDM от NXP Semiconductors, семейство Nucleo от STMicroelectronics, семейство EFM32 от Silicon Labs и многие другие. Аппаратная составляющая содержит также модули беспроводной связи, платы расширения и другие компоненты. Рассмотрим для примера некоторые характеристики платы NUCLEO-F401RE платформы Mbed:

Микроконтроллер STM32F401RET6 в корпусе LQFP64:

- Процессорное ядро ARM®32-бит Cortex®-M4 CPU с сопроцессором плавающей арифметически (FPU), максимальная частота ЦПУ 84 МГц
- Флэш память (Flash) 512 KB, ОЗУ(SRAM) 96 KB
- Количество портов общего назначения (GPIO) –50, с возможностью внешнего прерывания
- 12-бит АЦП (ADC) с16 каналами
- Часы реального времени (RTC)

- Таймер (Timer) с расширенным управлением
- Таймеры общего назначения – 7
- Сторожевые таймеры (Watchdog Timers) – 2
- Интерфейс USART/UART –4
- Интерфейс I2C –3
- Интерфейс SPI –3
- Интерфейс SDIO
- Интерфейс USB 2.0 OTG FS
- Отладчик/программатор ST-LINK/V2-1

Разработчик регистрируется на сайте mbed.com получает доступ к необходимым для создания программы ресурсам платформы и приступает к работе руководствуясь методическим указаниям.

Тестирования встроенной системы производится изменением параметры окружающей среды и фиксацией реакций. Наблюдение за реакцией прототипа технической системы выполняется разработчиком через веб-камеру. Для изменения параметров окружающей среды используются различные дистанционно-управляемые механизмы – имитаторы.

На рисунке 3а приведен пример онлайн лаборатории для прототипа системы освещения, управляемой жестами. Манипулирование руками заменяется линейными движителями, которые то удаляют, то приближают пластины к датчикам приближения прототипа.

На рисунке 3б приведен пример онлайн лаборатории для разработки функции движения вдоль стены автоматической самоходной тележкой. На микроконтроллере установлена плата расширения с тремя датчиками приближения VL53L0X. Отклонение от стены имитируется изменением положением качалки, обозначенной «L R» на основе сервопривода с ШИМ-управлением. Повороты у стены имитирует шаговый двигатель.

Разработчик управляет подобными механизмами через специально разработанное сетевое приложение. С помощью этого приложения выполняется также загрузка программ, откомпилированных в mbed.

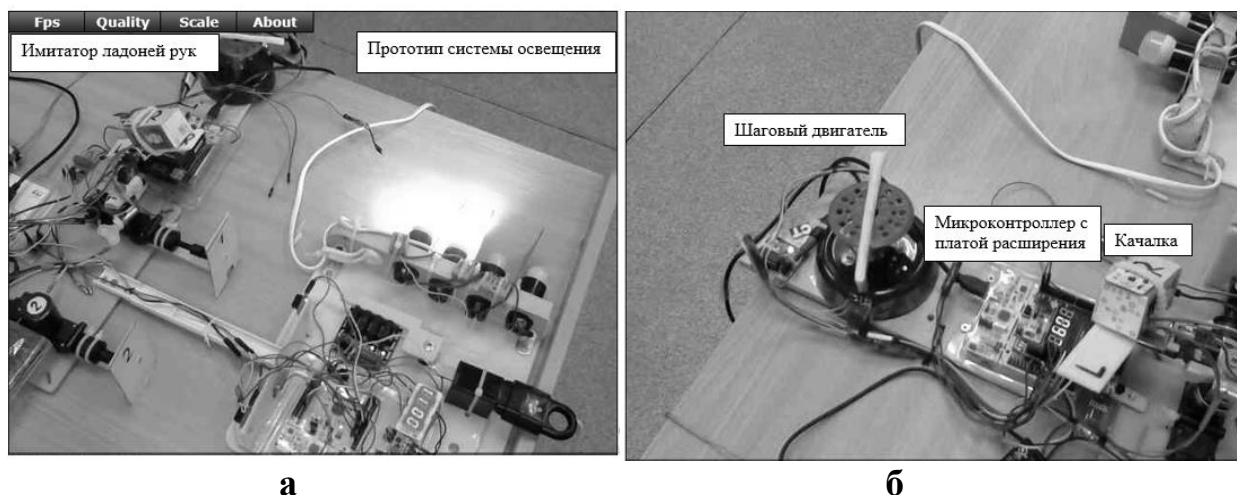


Рисунок 3. Вид через веб-камеру на прототип системы освещения и управление движением вдоль стены с соответствующими элементами онлайн лаборатории

Онлайн лаборатория содержит и прототипы подвижных систем — автоматическую самоходную тележку (рисунок 4а) и сервисного робота PROMOBOT V.4 (рисунок 4б).

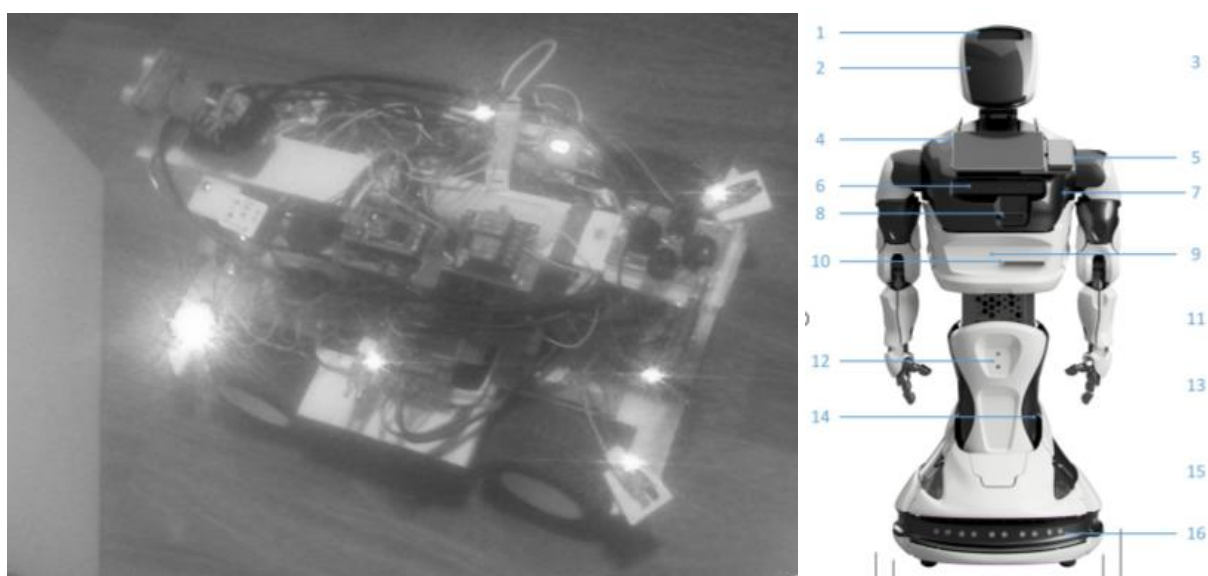


Рисунок 4. Мобильные компоненты удаленной лаборатории: автоматическая самоходная тележка и PROMOBOT V.4

Закключение. Приведены результаты разработки онлайн лаборатории для онлайн дисциплины «Встроенные системы».

ONLINE LABORATORY OF ONLINE DISCIPLINE “EMBEDDED SYSTEM”

Goncharovskiy Oleg V.

Perm National Research Polytechnic University

614990, Russia, Perm, Komsomolsky prospekt, d. 29, 35911953@mail.ru

The article presents the results of creating the online laboratory. Working in such a laboratory allows you to get the skills to develop software online prototypes of control subsystems (embedded systems) of technical systems using both the concept of "super Cycle" and the concept of "Multitasking".

Key words: online laboratory, embedded system, rapid prototyping.

УДК 004.89; 616.34

РАЗРАБОТКА ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ ДИАГНОСТИКИ И ПРОГНОЗИРОВАНИЯ ОЦЕНКИ ЗА ЭКЗАМЕН

Кузнецов Андрей Геннадьевич, Федорук Мария Николаевна,

Митрофанов Игорь Андреевич

Пермский государственный национальный исследовательский университет,
ПМИ. 614990, Россия, г. Пермь, ул. Букирева, 15, sp_9914@mail.ru

В статье описано создание математической модели прогнозирования оценки за экзамен. В основе модели лежит нейронная сеть, обученная на данных о сдаче экзаменов студентами пермских высших учебных заведений, полученных с помощью проведенного в социальных сетях опроса. С помощью построенной модели были выполнены исследования предметной области путем «замораживания» - изменения значения одного из входных параметров в то время, как значения остальных оставались прежними. Было установлено, что улучшение отношений с преподавателем положительно влияют на оценку. Также исследование показало, что возраст преподавателя оказывает влияние на оценку, причем чем возраст преподавателя ближе к среднему, тем больше вероятность получить плохую оценку. Склонность модели к прогнозированию более высокой оценки за экзамен при увеличении количества посещений занятий и подготовке шпаргалок подтверждает адекватность построенной модели. Практическая ценность исследования состоит в том, что построенная модель может быть использована для повышения будущих экзаменационных оценок.

Ключевые слова: образование, педагогика, студенты, экзамен, нейронные сети, сценарное прогнозирование.

Введение

Применение нейронных сетей в педагогике исследовано довольно плохо. «Первооткрывателями» можно назвать [1], которые разработали программу, предназначенную для прогнозирования вероятности трудоустройства по специальности уровня успешности студента в его будущей карьере, описали ключевые проблемы и показали методику создания программного продукта.

Создание интеллектуальной системы прогнозирования оценки за экзамен на сегодняшний день актуально по нескольким причинам. Во-первых, многие студенты заинтересованы в повышении будущей оценки за экзамен, следовательно, необходимо знать, от чего она зависит. Во-вторых, преподавателям также будет полезна данная интеллектуальная система, так как они заинтересованы в повышении успеваемости студентов.

Перспективным инструментом для выявления скрытых в статистической информации закономерностей и построения на их основе математических моделей являются нейронные сети. Хорошо спроектированные и правильно обученные нейронные сети способны самостоятельно выявлять закономерности практически любых предметных областей и строить адекватные математические модели во многих технических, естественнонаучных и других сферах [2].

Целью работы является создание интеллектуальной системы прогнозирования оценки за экзамен на основе данных, полученных с помощью опроса, проведенного в социальных сетях среди пермских студентов.

Постановка задачи и формирование обучающего множества

При постановке задачи в качестве входных параметров нейросетевой математической модели были выбраны критерии, которые, по мнению авторов статьи, являются наиболее важными для получения оценки за экзамен. Далее перечислены вопросы, которые задавались студентам ПГНИУ и варианты ответов (кодирование критериев):

X1 – День Вашего рождения

X2 – Месяц Вашего рождения

X3 – Год Вашего рождения

X4 – Ваш пол

1. Мужской

2. Женский

X5 – Полное наименование ВУЗа, в котором Вы учились на время сдачи экзамена.

X6 – Где Вы жили на время сдачи экзамена??

1. В общежитие

2. В съёмной квартире один

3. В съёмной квартире не один

4. В своей квартире один

5. В своей квартире не один

6. С родителями в квартире

- X7 – Ваше семейное положение на время сдачи экзамена
1. Один
 2. Есть девушка/молодой человек
 3. Женат/замужем
- X8 – Ваш доход на время сдачи экзамена
1. Живу на свою стипендию
 2. Живу за счёт родителей
 3. Работаю и сам зарабатываю себе на жизнь
 4. Работаю и получаю стипендию
- X9 – По Вашему мнению, каким был предмет?
1. Сложный, т.к. приходилось уделять много времени данному предмету в учебный период;
 2. Средней сложности: выполнял все задания в учебный период без особых трудностей;
 3. Лёгкий, т.к. очень редко или вообще не приходилось делать задания.
- X10 – Был ли Вам полезен данный предмет?
1. Был полезен
 2. Не был полезен
- X11 – Был ли преподаватель увлечён предметом?
1. Да, был увлечён
 2. Нет, не был увлечён
- X12 – Были ли Вы увлечены данным предметом?
1. Да, был увлечён
 2. Нет, не был увлечён
- X13 – Готовили ли Вы шпаргалки?
1. Да;
 2. Нет.
- X14 – Сами ли Вы готовили шпаргалки?
1. Да;
 2. Нет;
 3. Не готовил.
- X15 – Вы списывали?
1. Да;
 2. Нет.
- X16 – За какой период Вы подготовились к экзамену?
1. Больше 2 недель;
 2. От 1 до 2 недель;
 3. Меньше, чем 1 неделя.
- X17 – Склонны ли Вы к прокрастинации (прокрастина́ция — в склонность к постоянному откладыванию какого-либо дела)?
1. Да;
 2. Нет.
- X18 – Внимательно ли Вы слушали преподавателя во время лекции?
1. Да, конспектировал всё, что он говорит;

2. Когда как;
3. «В одно ухо влетало, в другое вылетало».

X19 – Было ли несколько удачных опытов в списывании до этого экзамена?

1. Да;
2. Нет.

X20 – Как Вы посещали пары по этому предмету?

1. Ходил на все пары;
2. Был почти на всех парах;
3. Очень часто прогуливал.

X21 – Какой формы был экзамен?

1. Устной;
2. Письменной.

X22 – Какой по характеру Ваш преподаватель?

1. Добрый и отзывчивый;
2. Строгий;
3. По настроению.

X23 – Пол преподавателя

1. Мужской
2. Женский

X24 – Приблизительный возраст преподавателя

1. До 30 лет
2. С 30 до 50 лет
3. Старше 50 лет

X25 – Преподаватель замечал, что Вы списываете?

1. Да и позволял это делать;
2. Да и делал замечания;
3. Да и выгонял с экзамена;
4. Нет, не замечал.

X26 – Как вёл пары преподаватель?

1. Весь материал был понятен;
2. Материал был понятен не весь;
3. Материал был вообще не понятен.

X27 – Ходил ли преподаватель во время экзамена?

1. Да;
2. Нет.

X28 – В каких отношениях Вы были с преподавателем?

1. В очень хороших отношениях;
2. Скорее всего, в хороших, чем в плохих
3. В нейтральных
4. Скорее всего, в плохих, чем в хороших
5. В плохих, преподаватель постоянно цеплялся ко мне.

у – Какую оценку Вы в итоге получили?

В результате проведенного в социальных сетях опроса были собраны ответы 452 человек. Полученное множество было разделено на обучающее и

тестирующее в соотношении 70:30. Таким образом, количество примеров в обучающем множестве составило 317, в тестирующем – 135.

Обучение и тестирование нейронной сети

Проектирование и обучение нейронной сети проводилось с помощью пакета Statistica. Поскольку среди признаков присутствуют как непрерывные, так и категориальные, для категориальных было применено кодирование One Hot Encoding. В результате количество входных параметров стало равно 92. На выходном слое располагаются 4 нейрона, выходом каждого из которых является вероятность получения соответствующей нейрону оценки. Для поиска наилучшей сети перебирались такие параметры как количество нейронов на скрытом слое, и функции активации скрытого и выходного слоев. В качестве алгоритма обучения использовался BFGS.

В результате обучения была получена сеть, имеющая 1 нейрон на скрытом слое, тангенс гиперболический в качестве функции активации скрытого слоя, softmax в качестве функции активации выходного слоя. Точность сети была проверена на тестирующем множестве, элементы которого не входили в обучающее, и составила 61,5%. На рис. 1 представлен график, отражающий адекватность полученной сети, проверенную на некоторых элементах тестирующего множества.

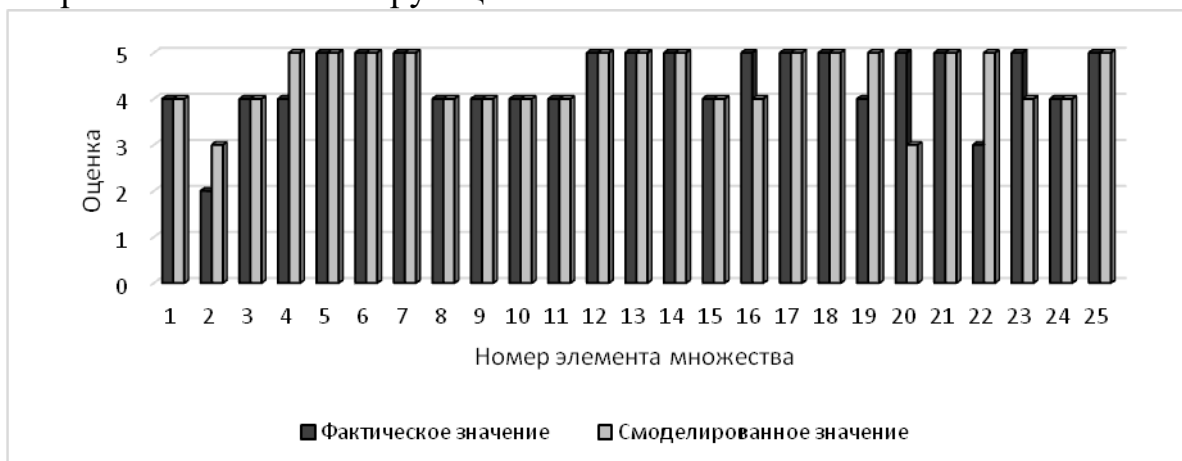


Рисунок. 1. Разность между фактическими и прогнозируемыми значениями оценок за экзамен

Исследуя разницу между фактическими и смоделированными значениями на всем тестирующем множестве, было установлено, что среднее отклонение составляет 1 балл, что является относительно неплохим результатом.

С помощью встроенной в Statistica функции была определена значимость параметров. Результаты анализа представлены на рис. 2.

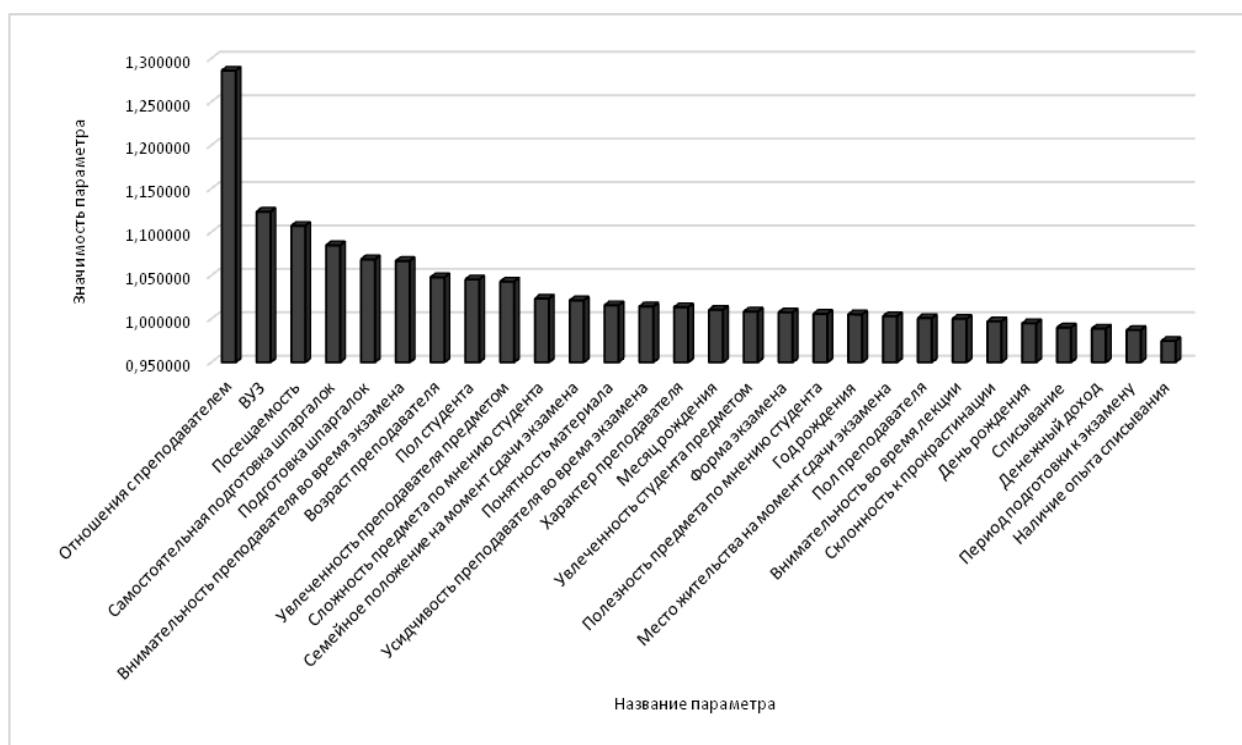


Рисунок 2. Значимость параметров

Те параметры, значимость которых меньше 1, не оказывают положительного влияния на качество модели. Таких параметров оказалось 6. Исключив данные параметры и обучив сеть заново, повышения точности замечено не было. Таким образом, все параметры остаются в модели для дальнейшего исследования.

Исследование полученной модели

Исследование модели производилось путем изменения значения одного из параметров в то время, как значения остальных оставались неизменными. Таким образом был исследован параметр «Отношения с преподавателем». Результаты представлены на рис. 3.

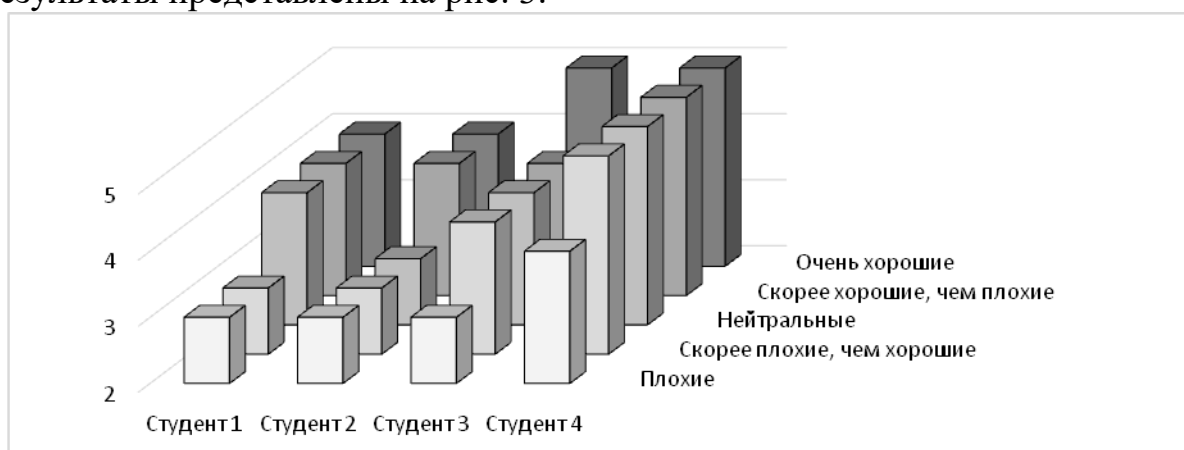


Рисунок 3. Зависимость оценки за экзамен от отношений с преподавателем

Как видно из графика, оценка за экзамен снижается при ухудшении отношений с преподавателем. По-видимому это вызвано тем, что отношения с преподавателем отражают отношение студента к предмету.

Следующим исследованным параметром является посещаемость. На графике, изображенном на рис. 4 видно, что оценка снижается при ухудшении посещаемости, что вполне логично, поскольку чем ниже посещаемость, тем меньше студент получает знаний, необходимых при сдаче экзамена.

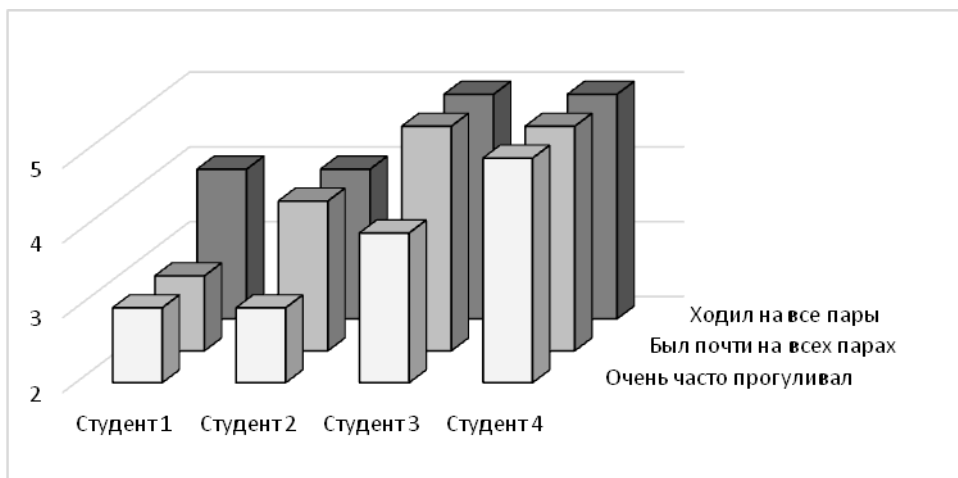


Рисунок 4. Зависимость оценки за экзамен от посещаемости

На рис. 5 изображен график, отображающий зависимость оценки за экзамен от подготовки шпаргалок. Можно увидеть, что студент №3 мог бы повысить свою оценку на 1 балл, если бы к экзамену приготовил шпаргалки. Возможно, это помогло бы ему лучше запомнить материал.

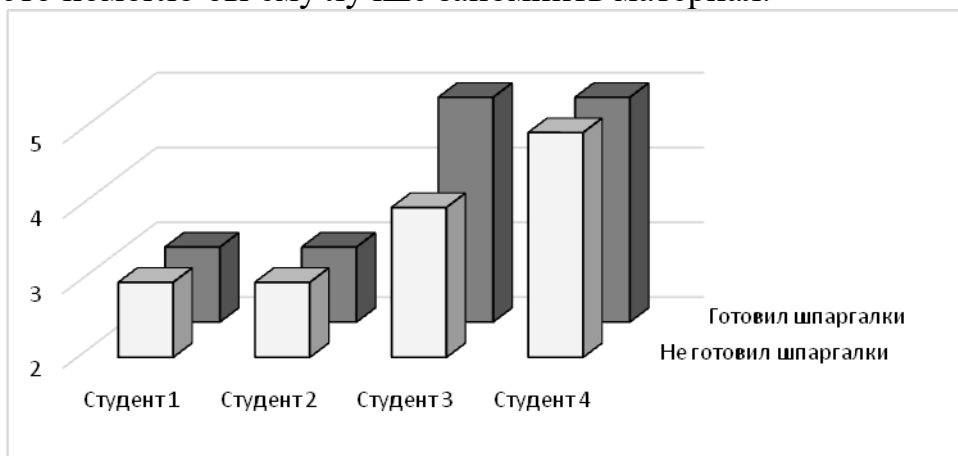


Рисунок 5. Зависимость оценки за экзамен от подготовки шпаргалок

Последним исследованным фактором является возраст преподавателя. Результаты исследования представлены на рис. 6.

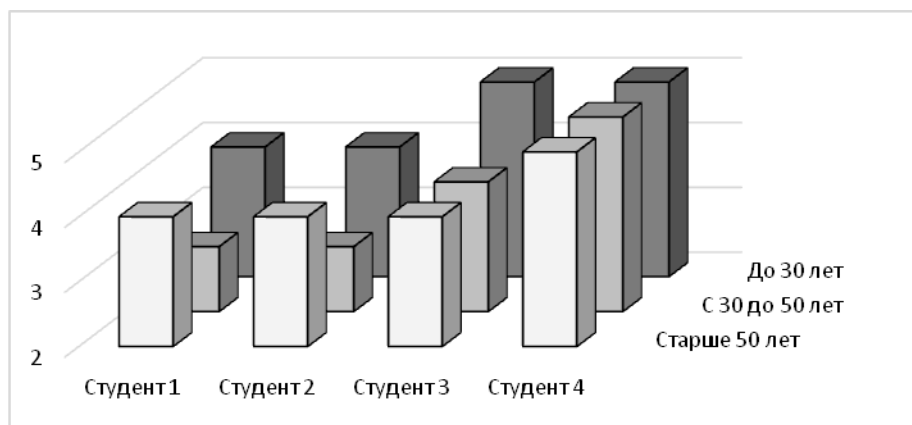


Рис. 6. Зависимость оценки за экзамен от возраста преподавателя

Как видно из графика, зависимость имеет нелинейный характер. Исследование параметра показывает, что молодые преподаватели и преподаватели в возрасте склонны ставить более высокие оценки, чем преподаватели среднего возраста.

Таким образом, исследование показывает, что построенная нейросетевая модель адекватно реагирует на изменение входных параметров, и может быть применена для повышения будущей оценки за экзамен.

Заключение

В результате проведенного исследования было проделано следующее:

1. Построена нейросетевая модель, выявляющая закономерности между оценкой за экзамен и её входными параметрами.
2. Установлено, что улучшение отношений с преподавателем способно повысить будущую оценку за экзамен.
3. Установлено, что оценка за экзамен зависит от возраста преподавателя. Однако эта зависимость имеет нелинейный характер, а именно вероятность получить более низкую оценку за экзамен у преподавателя, возраст которого 30-50 лет, выше, чем у преподавателя, чей возраст находится вне этого диапазона.
4. Несмотря на то, что модель имеет относительно небольшую точность, она может быть использована для оптимизации учебного процесса, нацеленного на получение более высоких оценок.

Библиографический список

1. Ясницкий Л.Н., Кузнецов А.Г., Селезнева С.М., Солохина А.Д., Тюлькина Д.В., Черепанов Ф.М. Применение нейросетевых технологий в изучении акмеологического потенциала студентов вуза // Вестник Пермского университета. Серия: Математика. Механика. Информатика. 2014. № 4. С.120-126.
2. Ясницкий Л.Н., Данилевич Т.В. Современные проблемы науки. М.:БИНОМ. Лаборатория знаний, 2008. 294с.

NEURAL NETWORK MODELING TO DIAGNOSE AND PREDICT THE DEVELOPMENT OF CARDIOVASCULAR DISEASES

*Kuznecov Andrey G.,
Fedoruk Mariya N., Mitrofanov Igor A.*
Perm State University

614990, Russia, Perm, street Bukireva, 15, mf1808@mail.ru

This article describes the creation of a mathematical model of prediction and assessment for the exam. The models are based on neural networks that are trained on the basis of survey results. With the help of the constructed model, studies were carried out using the "freezing". It was found that improving relations with the teacher has a positive effect on the assessment. In addition, the likelihood that the teacher will receive a bad grade. The propensity of models to predict a higher score for the exam with an increase in the number of visits and preparation of evidence of the adequacy of the constructed model. The practical value lies in the fact that the constructed model can be used to improve the future grade for the exam.

Keywords: education, pedagogy, students, exam, neural networks, scenario forecasting.

УДК 004.9

ТЕХНОЛОГИИ АНАЛИЗА ДАННЫХ: КЛАССИФИКАЦИЯ, ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА, ЭКСПЕРИМЕНТЫ С МОДЕЛЯМИ

Пащенко Данил Эдуардович, Шатров Анатолий Викторович
Вятский государственный университет, 610000, Россия, г. Киров,
Московская, 36, avshatrov1@yandex.ru

Анализ данных востребован для различных областей науки и большого числа прикладных исследований в промышленности, медицине, образовании и экономике. Для решения задач анализа данных в настоящее время применяется множество разнообразных методов и алгоритмов. Важным фактором является их эффективная компьютерная реализация. В планах подготовки специалистов по современным курсам анализа данных сложилась иерархия методик глубокой обработки данных и получения прогноза на основании проведенного анализа. Для квалифицированного применения методов анализа данных в преподаваемых курсах необходимо использовать доступные и достоверные базы данных. Такую возможность предоставляет платформа Kaggle. Учитывая, что используемые базы данных содержат

разнообразные как по форме, так и по содержанию данные, особое внимание следует уделять предварительной подготовке, построению моделей и последующим вычислительным экспериментам с моделями.

Ключевые слова: анализ данных, машинное обучение, классификация данных, кросс-валидация, бустинг, нейронные сети

Анализ данных (*Data Mining*) (DM) или вычислительная статистика является сравнительно новой научной областью. В отличие от теоретической статистики, вычислительная статистика ориентирована на обработку и анализ реальных данных. Соответственно, центральным вопросом является выбор оценивающего критерия, согласно которому производится сравнение различных методов и алгоритмов. Исследования и эксперименты в нашей работе базируются на данных, которые использовались при проведении международных соревнований по анализу данных на платформе Kaggle⁵.

При решении практических исследовательских задач возникают проблемы несбалансированности данных: интересующий признак может быть неравномерно представлен в обучающих выборках или составляет малую часть, которая игнорируется при использовании стандартных методов. Преодоление этих трудностей осуществляется в процессе так называемой предварительной обработки данных и в результате вычислительных экспериментов.

Эксперименты с данными Credit. Данные «Credit» использовались при проведении соревнования «Give me some credit» на платформе Kaggle. В рамках этого соревнования было необходимо определить кредитоспособность клиента в ближайшие 90 дней. Цель соревнования состояла в том, чтобы построить на основании исторических данных наиболее эффективный классификатор кредитоспособности клиентов. Соревнование привлекло около тысячи участников-команд со всего мира. Объясняющие переменные включают всю информацию о клиентах: общий остаток по кредитным картам (v_1), возраст (v_2), месячные выплаты долгов (v_3), доход (v_4), кредитная история и кредитные линии (v_5), количество нарушений выплат на 30-59 дней (v_6), количество нарушений выплат на 60-89 дней (v_7), количество нарушений выплат на 90 дней и более (v_8), количество ипотечных кредитов и кредитов на объекты недвижимости (v_9), состав семьи (v_{10}). Объясняемая (целевая) переменная является бинарной и принимает два значения – 0 для «хороших» клиентов и 1 для «плохих», имеющих определенный срок задолженности. Особенностью данной задачи является её несбалансированность: проблемные клиенты составляют ~6,5% от общей выборки. Анонимизированные данные соответствуют 250000 клиентам, из которых 100000 с неизвестными метками были отведены для тестирования и подведения итогов соревнования. Данные предоставляются в виде списков

⁵ www.kaggle.com

по клиентам и имеют представленную выше структуру, формализованную от переменных v_1-v_{10} .

Все данные (250000 наблюдений) разделены на 2 части: 150000 тысяч для тренировки и 100000 для тестирования. В данном случае первый набор разбит десять раз случайным образом на три части (триплет):

- Тренировочный набор – значения целевой переменной известны. Необходим для обучения модели.
- Валидационный набор – значения целевой переменной также известны. Необходим для предварительной проверки качества модели.
- Тестирующий набор – значения целевой переменной необходимо предсказать, используя полученную ранее модель.

Учитывая, что набор данных в этой задаче является несбалансированным, необходимо предварительно провести классификацию тренировочной и валидационной выборок. Одним из самых распространённых способов оценки качества бинарной классификации является нахождение площади под ROC-кривой. ROC-кривая (Receiver Operator Characteristic) – график, позволяющий оценить качество классификации в задаче с двумя классами. Она отображает зависимость между долей, правильно классифицированных и долей ошибочно классифицированных объектов [1]. Численной характеристикой данного метода является AUC (Area Under Curve) – площадь под ROC-кривой. Её значения лежат в интервале $[0,1]$. Именно этот показатель чаще всего используется для сравнительного анализа нескольких классификаторов. Чем больше значение AUC, тем лучше качество классификации; при значении 0.5 («случайное угадывание») модель можно расценивать как плохую. Графики этой величины позволяют визуально оценивать эффективность выбранных методов. При сравнении результатов линейной, логистической регрессии, алгоритма решающих деревьев, оказалось, что наилучшие результаты достигаются в методе решающих деревьев. Максимальное значение $AUC = 0.843$ получается при построении решающего дерева с максимальной глубиной дерева = 7, максимальным числом наблюдений в узле = 2 или 4. Дальнейшее улучшение результатов было достигнуто с помощью бустинга над решающими деревьями [2]. Бустинг⁶ (улучшение) – одна из наиболее эффективных техник машинного обучения, разработанных за последние двадцать лет. Если посмотреть на результаты обучения модели, то можно убедиться в том, что после 180-200 итераций AUC не имеет значительного улучшения. Значение AUC данной модели для 500 деревьев - 0.8605, для 190 - 0.861, что говорит о необходимости использования оптимального числа итераций при построении прогноза.

Проблема несбалансированности данных проявляется при использовании нейронных сетей, так как обученный на выборке, где плохие клиенты в подавляющем меньшинстве, алгоритм нейронной сети игнорирует

⁶ Бустинг (boosting) процедура объединения множества «слабых» базовых классификаторов (base-leaners) с целью получения одного более мощного «ансамбля».

малочисленный класс. В частности, если построить нейронную сеть на всей обучающей выборке, получим тривиальное нулевое решение с соответствующим результатом $AUC = 0.5$. Для того, чтобы избежать такого результата, необходимо выполнить кросс-валидацию данных обучающей выборки, то есть выполнить перебор различных подвыборок с целью создания сбалансированной выборки. При этом сбалансированная выборка (подмножество) получается в результате смешивания случайных выборок с весовыми коэффициентами α и β .

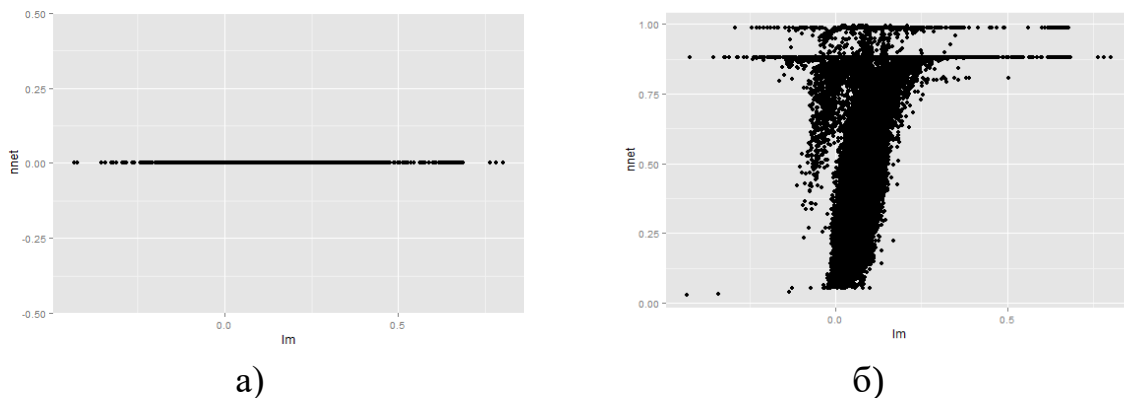


Рисунок 1. Соотношение прогнозов линейной регрессии (ось X) и нейронной сети (ось Y). (а) – при использовании несбалансированных данных, (б) – при использовании сбалансированного подмножества

На графике (рисунок 1, б) видно, что в этом случае (для сбалансированной выборки) полученные вероятности ненулевые, а AUC равен 0.8285. Чтобы найти оптимальное число клиентов обоих классов в составе сбалансированного подмножества, необходимо провести несколько экспериментов, используя различные пары параметров α и β . На этой основе построен метод случайных сбалансированных подмножеств [3]. Суть этого метода такова:

- 1) строится большое число локальных классификаторов на основе случайно сформированных подмножеств;
- 2) в силу своей стохастичности каждое из них не дает объективной оценки качества классификатора;
- 3) для получения объективной оценки используем технику ассемблирования – вычисляем решающую функцию как среднее арифметическое большого числа решений, полученных на локальных подмножествах.

В результате вычислительных экспериментов сбалансированное подмножество получено при $\alpha = 0.85$, $\beta = 0.06$, число сбалансированных подмножеств $n=20$, а оптимальная структура нейронной сети достигается при семи нейронах скрытого слоя. При этом максимальное значение AUC по всем итерациям составило 0.859

Библиографический список

1. Hastie T., Tibshirani R., Friedman J.H. The Elements of Statistical Learning, 2nd edition. - Springer, 2009. — 533 p.
2. Никулин В.Н., Палешева С.А., Зубарева В.С. Об однородных ансамблях при использовании метода бустинга в приложении к несбалансированным данным// Вестник Пермского университета. Сер. Экономика. 2012. №1. С. 7-14.
3. Nikulin V. On the Evaluation of the Homogeneous Ensembles with CV-passports // PAKDD 2013.- LNCS 7867 Springer 7867, J. Li et.al. (Eds.). - 109-120 pp.

TECHNOLOGIES OF DATA MINING: QUALIFICATION, CROSS-VALIDATION, EXPERIMENTS WITH MODELS

Pashchenko Danil E.
Vyatka State University,
610000, Russia, Kirov, Moskovskaya str, 36, smile_dan@mail.ru

Shatrov Anatoly V.
Vyatka State University
610000, Russia, Kirov, Moskovskaya str, 36, avshatrov1@yandex.ru

Data analysis is in demand for various fields of science and for a large number of applied research in industry, medicine, education and economics. A variety of methods and algorithms is currently used to solve the problems of data analysis. An important factor is their effective computer implementation. In training specialist plans on modern courses of data analysis there is a hierarchy of methods of deep data processing and forecasting based on the analysis. It is necessary to use available and reliable databases for qualified practical application of data analysis methods in courses. This feature is provided by the Kaggle platform. Having in mind that the using databases contain a variety of data, both in form and content, special attention should be paid to preprocessing, model building and subsequent computational experiments with models.

Key words: data mining, machine learning, data classification, cross-validation, boosting, neural network.

МОДЕЛЬ ФОРМИРОВАНИЯ ФУНКЦИОНАЛЬНЫХ СИСТЕМ ПРИ БОС-ТРЕНИНГЕ СТУДЕНТОВ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ

Петрунина Елена Валерьевна

Московский государственный гуманитарно-экономический университет,
111674, Россия, г. Москва, ул. Льва Яшина, д.7, кв. 14, petrunina@mggeu.ru

Трифонов Андрей Андреевич, Шаталова Ольга Владимировна

Юго-Западный государственный университет,
305040, Россия, г. Курск, ул. 50 лет Октября, 94, Shatolg@mail.ru

Рассмотрена модель функциональных систем, формируемых с помощью биологической обратной связи. Предложена функциональная модель гиперграфа, описывающего взаимодействие структурных элементов модели. Предложенные модели предназначены для моделирования функциональных систем БОС-тренинга студентов с ограниченными возможностями здоровья.

Ключевые слова: модель функциональной системы, биологическая обратная связь, нейронная сеть, дефицит внимания.

В работах П.К. Анохина выдвинуты основные постулаты, лежащие в основе теории моделирования функциональных систем (ФС) [1]. Однако в его работах не дан ответ на вопросы: «Как сформировать ФС с заданными целевыми характеристиками? Как ускорить процесс формирования «нужной» ФС? Как выделить функциональные резервы для формирования «нужных» ФС?». Ответы на эти вопросы актуальны для работников экстремальных профессий, студентов, лиц с ограниченными возможностями здоровья и т.д.

Для ответа на поставленные вопросы необходима модель формирования ФС. Модель формирования ФС может быть построена на основе понятия структурной функциональной единицы (СФЕ), введенного в [2]. Согласно выдвинутой в [2] теории, ФС строится на основе установки устойчивых связей между СФЕ, находящимися в резерве, то есть «активными» СФЕ. При этом организм человека с точки зрения теории СФЕ рассматривается как закрытая система с ограниченным ресурсом СФЕ. СФЕ могут находиться в «активном» состоянии, то есть быть доступны для «Менеджера» для построения очередной ФС. «Активные» СФЕ представляют функциональный резерв организма, который может быть пополнен из функционального ресурса. Ресурс СФЕ системы - число «не активированных» СФЕ в процессе поддержания системой своего номинального или текущего режима (состояния). Резерв СФЕ системы -

минимальное количество СФЕ, которые может включать/отключать система, находясь в номинальном или текущем режиме (функциональном состоянии). Такие СФЕ могут рассматриваться и как «оперативный резерв» системы, а его использование - как средство «перераспределения» релевантности между ФС [2].

Таким образом, ФС – это взаимосвязь активных элементов, которая направлена на достижение полезного приспособительного результата. Однако для построения ФС недостаточно одного ресурса СФЕ. Для формирования ФС необходим Менеджер – интеллектуальный агент, собирающий ее из СФЕ на основе имеющегося у него запаса знаний, который, при необходимости, может быть пополнен. В качестве такого Менеджера в модели используем нейронную сеть, занимающую промежуточное положение между обучающейся и самоорганизующейся нейросетевой структурой. Модель формирования ФС на основе такого подхода представлена на рисунке 1.

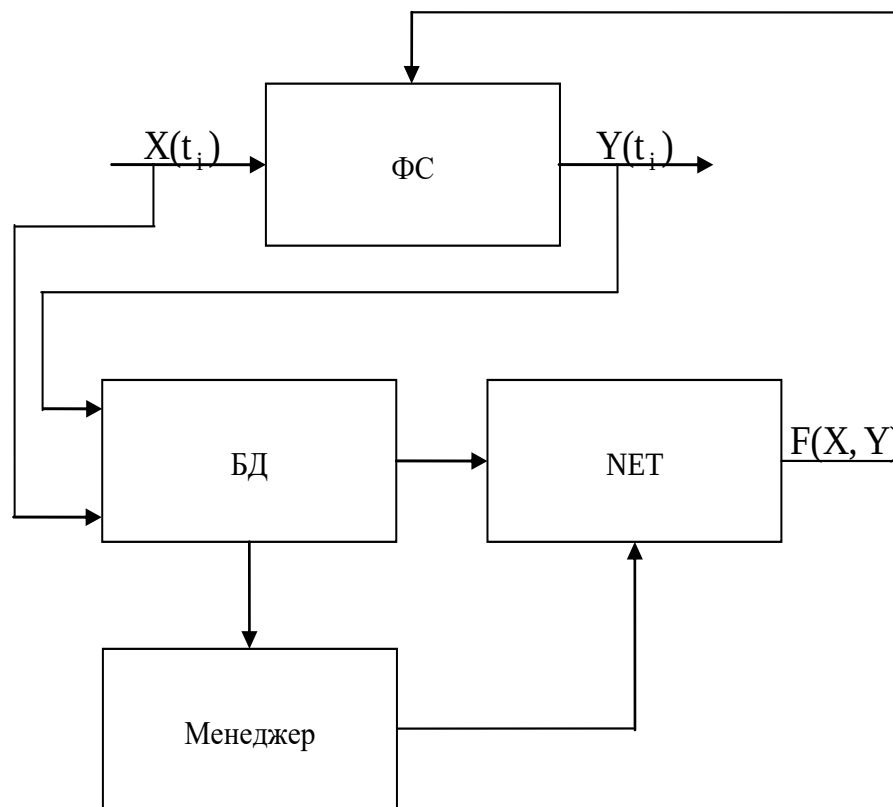


Рисунок 1. Схема формирования функциональной системы

Задача ФС заключается в том, чтобы на входной стимул X , поступивший в момент времени t_i , выдать реакцию Y заданного класса, которая определяется оператором A :

$$Y(t_i) = A\{X(t_i)\}. \quad (1)$$

Для формирования оператора A используется нейронная сеть NET , работающая на основе гибридного алгоритма [3]. Для настройки нейронной

сети NET в начале стимуляции нет обучающей выборки. На вход ФС поступают стимулы X , которые нарушают гомеостазис организма. На выходе ФС необходимо сформировать реакцию Y , которая снижает деструктивное влияние стимула X . Сущность гибридной настройки нейросети NET заключается в том, что при ее настройке используется как обучающая выборка, формируемая в базе данных БД, так и системная память – база знаний (БЗ), которую в модели рисунок 1 представляет Менеджер.

Особенность гибридного алгоритма состоит в том, что он осуществляет последовательное переключение настройки нейросетевой структуры NET с экспертного оценивания на обучающую выборку и, наоборот. Алгоритм является бесконечным, так как он может «уснуть», если прекращают поступать данные в БД (отсутствуют стимулы).

Так как в начальный момент обучающая выборка отсутствует, то Менеджер на основе памяти о предыдущих ситуациях формирует ФС и настраивает NET. Затем, им же, определяется апертура наблюдения $N1$, после чего в БД записывается множество пар $\{X(t_i), Y(t_i)\}$ и определяется ошибка ϵ , которая выражает «удовлетворенность» Менеджера работой ФС.

Если Менеджер признает работу ФС удовлетворительной, то осуществляется выбор новой апертуры наблюдения $N1$. Если ошибка велика, то Менеджер приступает к анализу данных в БД. Здесь могут иметь место три варианта действий. По первому варианту Менеджер приходит к выводу о смене настройки NET на основе своего опыта. Во втором случае Менеджер находит нерегулярность в данных и решает переформатировать обучающую выборку, после чего переобучить нейронную сеть. Если данные, по решению Менеджера, удовлетворительны, то осуществляется переход на блок обучение нейронной сети. Для моделирования ФС используются нейросетевые модели [3].

При обучении студентов с ограниченными возможностями здоровья (ОВЗ) их функциональный ресурс ограничен, так как значительная часть его уходит на построение ФС, компенсирующих ограничения здоровья. Поэтому ставится вопрос об оптимизации формирования ФС как по быстрдействию, так и по эффективности использования СФЕ.

Эта проблема может быть решена путем введения дополнительного контура обратной связи в схеме рисунок 1, что предполагает включение дополнительных ФС, но в целом, приводит к повышению эффективности использования ресурса СФЕ за счет того, что для построения обратной связи используются не только ресурсы организма, но и технические средства, а также интеллектуальные ресурсы как ЛПР, так и искусственного интеллекта. С этой целью необходимо включить в процесс формирования ФС биотехническую систему. Обратная связь организуется на основе суррогатных маркеров – биологических сигналов, получаемых от человека в процессе внешних стимулирующих воздействий, поэтому она называется биологической (БОС).

В БОС используется некая форма технологии, обеспечивающая дополнительную информацию, выходящую за пределы обычного восприятия

физических (физиологических) функций. Человек впервые рассматривает информацию как обратную связь для того, чтобы увеличить понимание или осознать изменения в функционировании тела и разума. На рисунке 2 представлена схема формирования ФС с использованием БОС-тренинга.

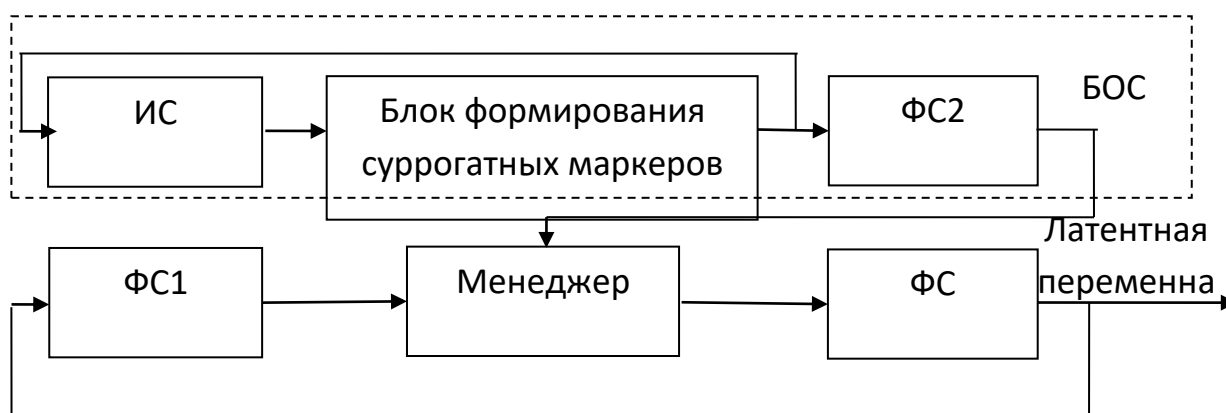


Рисунок 2. Структурная схема формирования функциональной системы с БОС-тренингом

Особенностью представленной модели является наличие трех обратных связей. Первые две обратных связи замкнуты на Менеджера: первая обратная связь формируется биотехнической системой через функциональную систему ФС2, а вторая обратная связь аналогична связи БД-менеджер на рисунке 1. Третья обратная связь реализована исключительно технической системой и обеспечивает контроль успешности обучения интеллектуальной системой (ИС) или ЛПР.

Менеджер воспринимает информацию по эффективности ФС по двум каналам, один из которых формируется живой системой (ФС1), а второй является гибридным каналом, в формировании которого участвует и техническая подсистема биотехнической системы (ФС2).

Для адекватного отображения рассмотренных особенностей системы с БОС предлагается воспользоваться интегрированной моделью в форме функционального гиперграфа F -граф или FG

$$F = \{V, R\} \text{ или } FG = \{V, R\},$$

где V и R - множества функциональных вершин и функциональных ребер (или гиперребер), соответственно.

Для графа справедливы следующие уравнения

$$V = V_1 \cup V_2 \cup V_3 \cup V_4, \text{ а } R = R_1 \cup R_2 \cup R_3 \quad (2)$$

V_1 - подмножество вершин, интерпретирующих воздействия ФС на отдельные элементы (ОЭ);

V_2 - подмножество вершин, интерпретирующих воздействия ФС1 на ОЭ;

V_3 - подмножество вершин, интерпретирующих воздействия ФС2 на ОЭ;

V_4 - подмножество вершин, интерпретирующих воздействия Менеджера на ОЭ ФС.

Множество гиперребер R состоит из следующих подмножеств:

R_1 - подмножество гиперребер, интерпретирующих процессы в ФС;

R_2 - подмножество гиперребер, интерпретирующих процессы в ФС1;

R_3 - подмножество гиперребер, интерпретирующих процессы в ФС2.

При реализации иерархического подхода в моделировании каждое из рассматриваемых подмножеств вершин V_i и гиперребер R_i может быть представлено в виде объединения подмножеств вершин более низкого уровня.

Наиболее актуальна для студентов с ОВЗ проблема дефицита внимания. При обучении БОС при наличии синдрома дефицита внимания был использован электроэнцефалографический бета-стимулирующий тренинг, который приводит к улучшению метаболизма головного мозга. Полагают, что чем больше мощность бета-ритма, тем выше церебральный кровоток, потребление глюкозы и кислорода. Чтобы выбрать между бета- и бета/тета тренингом, необходимо иметь общие сведения об энцефалографических показателях. Данные по ним получают на выходе блока формирования суррогатных маркеров, например с помощью программного обеспечения системы «Кинезис». В этой системе БОС-тренинга обучающемуся предлагают повышать бета-ритм, решая появляющиеся на экране математические примеры разной сложности. Общая длительность курса БОС при дефиците внимания значительно варьируется в зависимости от выраженности исходных нарушений.

Библиографический список

1. Анохин П.К. Избранные труды: Кибернетика функциональных систем / П.К. Анохин — М.: Медицина, 1998. — 400 с.
2. Филист, С.А. Структурно-функциональная модель для мониторинга влияния управляющих воздействий на функциональное состояние самоорганизующихся систем / С.А. Филист, А.Н. Шуткин, П.С. Кудрявцев [и др.] // Прикаспийский журнал: управление и высокие технологии. Научно-технический журнал. — 2015. № 2 (30).-С.105-119.
3. Филист, С.А. Использование гибридных нейросетевых моделей для многоагентных систем классификации в гетерогенном пространстве информативных признаков / С.А. Филист, А.Г. Курочкин, В.В. Жилин и др.//Прикаспийский журнал: управление и высокие технологии. Научно-технический журнал. — 2015. № 3 (31).-С.85-95.

MODEL OF FORMATION OF FUNCTIONAL SYSTEMS IN BOS- TRAINING OF STUDENTS WITH RESTRICTED HEALTH OPPORTUNITIES

Petrulina Elena V.

Moscow State University of Humanities and Economics, Southwestern State
University

111674, Russia, Moscow, st. Lev Yashin, 7, apt. 14, petrunina @ mggeu.ru

Trifonov Andrey A., Shatalova Olga V.

Southwestern State University,

305040, Russia, Kursk, ul. 50 years of October, 94, Shatolg@mail.ru

The model of functional systems formed using biologic feedback is considered. A functional model of a hypergraph describing the interaction of the structural elements of the model is proposed. The proposed models are designed to simulate the functional BOS-training of students with disabilities.

Key words: functional system model, biological feedback, neural network, attention deficit.

УДК 004.8; 378

НЕЙРОСЕТЕВАЯ МОДЕЛЬ ПРОГНОЗИРОВАНИЯ УСПЕВАЕМОСТИ СТУДЕНТА-ПЕРВОКУРСНИКА

Чингаева Елена Станиславовна, Русаков Сергей Владимирович

Пермский государственный национальный исследовательский университет,

ПМИ. 614990, Россия, г. Пермь, ул. Букирева, 15,

chingaeva.elena@gmail.com, rusakov-eduv@mail.ru

В статье описывается анализ факторов, влияющих на успеваемость студента, а также результат построения нейросетевой модели, с помощью которой возможно прогнозирование группы риска среди студентов 1-го курса механико-математического факультета.

Ключевые слова: успеваемость студентов, нейронная сеть

В последние годы многие высшие учебные заведения сталкиваются с проблемой низкой успеваемости студентов. При этом высокие результаты, полученные абитуриентом на Едином государственном экзамене (ЕГЭ), и окончание специализированной школы вовсе не гарантируют успешность обучения на первом курсе. Высокий уровень студенческих отчислений особенно характерен для специальностей, связанных с точными науками, в

число которых входят ИТ-направления. Поэтому на сегодняшний день актуальной задачей для исследований в области образования является изучение факторов, оказывающих влияние на академическую успеваемость студентов, а также разработка рекомендаций для ее повышения.

Во многих работах, посвященных прогнозированию успеваемости студентов, исследователи в первую очередь изучают влияние на успеваемость академических факторов, например, успехов студента во время обучения в школе, результатов ЕГЭ, участия в предметных олимпиадах. Так, в исследовании [1] было установлено, что суммарный балл ЕГЭ имеет достаточно высокую предсказательную валидность: по результатам проведенного регрессионного анализа, баллами ЕГЭ объясняется от 13 до 30% вариации успеваемости студента (в зависимости от направления подготовки). В ходе исследования [2] была оценена связь между успеваемостью студентов по всем дисциплинам первого курса и средним баллом ЕГЭ путем расчета оценки коэффициента корреляции. Для всех студентов оценка коэффициента корреляции составила 0,51, а для отдельных факультетов её значения входят в диапазон от 0,4 до 0,7. Таким образом, полученные результаты позволяют говорить об объективности ЕГЭ как системы для отбора абитуриентов. С другой стороны, во многом успеваемость студента зависит и от его личностных качеств. Авторы работы [3] выделяют следующие личностные детерминанты высокой успеваемости:

- Определенная иерархия ценностных ориентаций студента, главными в которой являются ценности достижения и самостоятельности;
- Наличие высокой внутренней мотивации к процессу обучения;
- Развитая волевая сфера личности, позволяющая преодолевать трудности и упорно идти к поставленным целям.

Авторы отмечают следующую закономерность: студенты с высокой успеваемостью ориентированы на достижения и самостоятельность, они более ответственные, упорны и последовательны в выборе приоритетов, среди данных студентов присутствует высокий уровень самоконтроля и организованности; ориентиром же студентов с низкой успеваемостью является гедонизм и развлечения, они менее самостоятельны и менее последовательны в выборе приоритетов.

При прогнозировании успеваемости необходимо учитывать и социальные факторы. В исследовании [4] было установлено, что студенты, обучающиеся на бюджетной основе, отличаются более высокими показателями успеваемости, чем студенты-договорники. Данный результат авторы объяснили тем, что многие студенты для обучения на коммерческой основе вынуждены совмещать учебу с работой. В то же время студенты-бюджетники с низкой успеваемостью в основном проживали в съемных квартирах, что говорит о взаимосвязи условий проживания и академических успехов студента. Авторы [5] решили проверить, оказывает ли влияние уровень дохода семьи на баллы, получаемые абитуриентом на ЕГЭ. В процессе исследования была обнаружена положительная зависимость между

результатами ЕГЭ и уровнем обеспеченности семьи. Объясняется это тем, что более обеспеченные родители затрачивают больше средств на подготовку к поступлению (здесь имеются в виду занятия с репетиторами, подготовительные курсы), что приводит к более высоким результатам ЕГЭ, чем в семьях со средним и низким уровнем дохода. Отсюда можно сделать предположение, что и в процессе обучения в вузе успеваемость студента будет связана с материальной обеспеченностью его семьи.

Немалое число исследований направлено на установление зависимости между уровнем физической подготовки учащихся и их академической успеваемостью. Так, в работе [6] авторы приходят к выводу, что студенты-спортсмены, которые отличаются хорошим состоянием здоровья, функциональной готовностью основных систем организма и целеустремленностью, имеют более высокие результаты учебной деятельности в отличие от студентов, не занимающихся спортом.

Рассмотренные выше факторы можно классифицировать следующим образом:

- Факторы, связанные с учебной деятельностью студента до поступления в вуз (результаты ЕГЭ; участие в предметных олимпиадах; характеристики учебного заведения, в котором обучался студент);
- Психологические факторы (способность адаптироваться к новой среде, внутренняя мотивация к учебе, умение организовать учебный процесс, тип темперамента и др.);
- Состояние здоровья и уровень физической подготовленности;
- Социальные факторы (социальный состав семьи, материальное положение, условия проживания, совмещение учёбы с работой);
- Факторы, связанные с обучением в вузе (наличие платы за обучение, методика преподавания, психологическая обстановка в коллективе, посещаемость студентом занятий).

Настоящее исследование было проведено в рамках направления «Прикладная математика и информатика» механико-математического факультета ПГНИУ. Данное направление характеризуется высоким уровнем отчислений, о котором можно судить по таблице 1, где приведена статистика по первокурсникам по итогам первого триместра за последние 4 года.

Таблица 1. Количество отчисленных студентов

Год	Количество поступивших студентов	Количество отчисленных студентов	Процент отчисленных студентов
2014	99	23	23,23%
2015	86	27	31,39%
2016	89	27	30,33%
2017	86	17	19,77%

В качестве выборки для обучения нейронной сети были взяты данные из единой телеинформационной системы (ЕТИС) ПГНИУ о студентах первого курса 2014-2017 годов поступления. Используемый при этом набор входных параметров представлен в таблице 2. Выходной параметр Y принимает два значения: 1, если студент был отчислен по итогам 1-го триместра, и 0 в противном случае.

Таблица 2. Входные параметры нейронной сети

№	Расшифровка	Значения
X1	ЕГЭ по математике	Количество баллов, полученных за ЕГЭ по дисциплине «Математика»
X2	ЕГЭ по русскому языку	Количество баллов, полученных за ЕГЭ по дисциплине «Русский язык»
X3	ЕГЭ по информатике	Количество баллов, полученных за ЕГЭ по дисциплине «Информатика»
X4	Город и тип образовательного учреждения	1 – пермские лицеи, гимназии, СОШ 9, СОШ 146, 2 – лицеи и гимназии за пределами Перми, 3 – пермские СОШ, 4 – СОШ за пределами Перми
X5	Язык, изучаемый в школе	0 – другой, 1 – английский
X6	Повторное поступление	0 – поступление на направление «Прикладная математика и информатика» осуществляется первый раз, 1 – поступление на направление «Прикладная математика и информатика» уже было в предыдущие года
X7	Пол	0 – женский, 1 – мужской
X8	Место проживания	1 – проживание дома, по прописке, 2 – проживание дома, не по прописке, 3 – проживание в общежитии
X9	Получение соц. стипендии	0 – студент не получает соц. стипендию, 1 – студент получает соц. стипендию

Перед обучением нейронной сети было решено определить, каким образом связаны результаты ЕГЭ и успеваемость студентов в каждый из рассматриваемых четырех годов. Для этого тремя способами вычислялась функция $f(x_1, x_2, x_3)$, параметрами которой были баллы ЕГЭ, полученные студентом по трем предметам:

$$f(x_1, x_2, x_3) = x_1 + x_2 + x_3, \quad (1)$$

$$f(x_1, x_2, x_3) = (x_1 - 40)(x_2 - 40)(x_3 - 40), \quad (2)$$

$$f(x_1, x_2, x_3) = (x_1 - 40)^2 + (x_2 - 40)^2 + (x_3 - 40)^2. \quad (3)$$

Формула (1) представляет из себя стандартную линейную свертку. Формулы (2)-(3) соответственно критерии Нэша и Купера с учетом того, что для всех трех предметов минимальный проходной балл был равен 40.

Далее студенты одного года поступления ранжировались в порядке убывания полученного значения функции и разбивались на две группы, границей которых была медиана данных значений, после чего каждая группа снова разбивалась на две таким же образом. В итоге студенты были поделены на четыре примерно равные по объему группы. В каждой группе было посчитано число отчисленных студентов, что позволило вычислить вероятность отчисления студента при условии, что он попадает в определенную группу. В таблице 3 представлены итоги анализа, где числами 1-4 обозначены номера групп. Нетрудно видеть, что ситуация в выборках 2014-2015 и 2016-2017 годов существенно различается. Так, в 2014-2015 годах все студенты из первой группы успешно справились с учебным планом, а доля отчисленных первокурсников возрастает от первой группы к четвертой. Выборки 2016-2017 годов явно не подчиняются данному закону. Поэтому было решено обучить нейросеть сначала на всем множестве примеров, а затем – отдельно для первой и второй пары годов поступления.

Таблица 3. Вероятность отчисления студентов

Год	Способ №1				Способ №2				Способ №3			
	1	2	3	4	1	2	3	4	1	2	3	4
2014	0,000	0,200	0,250	0,500	0,000	0,200	0,200	0,520	0,000	0,240	0,280	0,400
2015	0,000	0,227	0,304	0,750	0,000	0,182	0,364	0,714	0,000	0,273	0,273	0,714
2016	0,136	0,227	0,435	0,409	0,136	0,182	0,391	0,500	0,136	0,227	0,435	0,409
2017	0,045	0,333	0,143	0,273	0,091	0,286	0,190	0,227	0,045	0,286	0,190	0,273

В качестве инструмента для нейросетевого моделирования был выбран пакет Deductor Studio Academic, представляющий собой платформу со встроенными методами извлечения, анализа и визуализации данных.

Процесс создания нейросетевой модели состоял из следующих этапов:

1. *Формирование примеров.* Проверялось наличие выбросов, противоречащих или повторяющихся примеров, которые исключались из рассмотрения. Множество примеров разбивалось на обучающее и тестовое в отношении 4:1;
2. *Проектирование сети.* Выбиралась структура сети, активационная функция нейронов и другие параметры;
3. *Обучение сети;*

4. Проверка сети на тестовом множестве примеров.

Информация о результатах обучения для построенных моделей приведена в таблице 4.

Таблица 4. Итоги моделирования

Выборка	2014-2017	2014-2015	2016-2017
Количество примеров	359	185	175
% верно классифицированных примеров обучающего множества	98.26	98.65	97.86
% верно классифицированных примеров тестового множества	75	81.08	85.71

Полученные модели могут быть использованы для выявления группы риска среди первокурсников последующих годов поступления. Данная информация может быть полезна тьюторам и преподавателям, ведущим занятия, для индивидуальной работы со студентами, с целью снижения доли отчисленных.

Библиографический список

1. Хавенсон Т.Е., Соловьева А.А. Связь результатов Единого государственного экзамена и успеваемости в вузе // Вопросы образования. – 2014. - № 1. – С. 176-199.
2. Щеголева Л.В., Суровцова Т.Г. Результаты ЕГЭ и успеваемость студентов первого курса // Непрерывное образование: XXI век. – 2015. - № 4 (12). – С. 33-41.
3. Балина Т.Н., Дерипаско Д.А. Личностные детерминанты успеваемости студентов // Вестник Таганрогского института управления и экономики. – 2008. - № 2. – С. 78-83.
4. Байжанова Н.С., Рослякова Е.М., Хасенова К.Х. Зависимость успеваемости студентов от некоторых физиологических показателей и ряда социальных факторов // Международный журнал экспериментального образования. – 2015. - № 4. – С. 424-425.
5. Прахов И.А., Юдкевич М.М. Влияние дохода домохозяйств на результаты ЕГЭ и выбор вуза // Вопросы образования. – 2012. - № 1. – С. 126-147.
6. Радакина Д.С., Мухамбетова А.С. Влияние физической культуры на успеваемость студента // Современная наука: актуальные проблемы и пути их решения. – 2015. - № 4 (17). – С. 198-202.

NEURAL NETWORK MODEL FORECASTING THE PERFORMANCE OF A FIRST-YEAR STUDENT

Chingaeva Elena S., Rusakov Sergey V.

Perm State University 614990, Russia, Perm, street Bukireva, 15,
chingaeva.elena@gmail.com, Rusakov-eduv@mail.ru

The article describes the analysis of factors affecting student performance, as well as the result of building a neural network model, which predicts a risk group among first-year students of the Faculty of Mechanics and Mathematics.
Keywords: student performance, neural network.

ЦИФРОВОЙ СПОРТ

ПРОГНОЗИРОВАНИЕ УСПЕШНОСТИ БОЙЦА ММА НА ОСНОВЕ ЕГО ФИЗИОЛОГИЧЕСКИХ ПОКАЗАТЕЛЕЙ

Бакулин Никита Константинович

Национальный исследовательский университет «Высшая школа экономики»,
Пермский филиал, Россия, 614070, Пермь, ул. Студенческая, 38,
bakulin.nikita@bk.ru

В статье рассмотрена задача прогнозирования успешности бойцов боёв без правил (ММА) на основе их физиологических показателей. Методика нейросетевого моделирования может быть использована для создания аналогичных компьютерных программ, позволяющих получать рекомендации по изменению физиологических показателей бойца с целью увеличения шанса его победы в поединке.

Ключевые слова: бои без правил, ММА, UFC, прогноз, нейронная сеть, процент побед в поединке, эффективность бойца.

В современном мире всё большую популярность набирают бои без правил (ММА). Эти бои отличаются «отсутствием» правил, увеличивающим их зрелищность и реалистичность, что ведёт к постоянному росту зрительской аудитории таких мероприятий. Призовые фонды таких турниров исчисляются миллионами долларов США, бойцы являются кумирами миллионов человек, про них снимают фильмы, делают видеоигры, выпускают специальную символику и атрибутику.

В современной научной литературе уже имеются работы, посвященные применению нейросетевых технологий для прогнозирования успешности как спортивных команд, так и отдельных спортсменов, а также для выдачи на этой основе полезных рекомендаций [1-4]. В настоящей работе, предложенная в [1-4] методика нейросетевого моделирования применена для построения компьютерной программы, прогнозирующей процент побед бойца, участвующего в спортивной игре «Бои без правил».

В работе использованы следующие параметры бойца:

- X1 - Рост (см) – рост бойца.
- X2 - Вес (кг) – средний вес бойца по всем его боям.
- X3 - Размах рук (см) – расстояние от кончиков пальцев одной руки до кончиков пальцев другой руки, руки разведены в стороны.
- X4 - Размах ног (см) - максимальное расстояние между ног, разведённых в стороны, аналог растяжки.
- X5 - Возраст (лет) – возраст бойца.

- X6 - Позиция (правша - 1, левша - 2, меняемая позиция (амбидекстр) - 3) – боец выступает как правша, левша или амбидекстр.
- X7 - Весовая категория (9 весовых категорий) – в зависимости от веса бойцы делятся на весовые категории, внутри которых и проводятся соревнования. Мы учитываем категории: наилегчайший (женщины) - 1, наилегчайший (мужчины) - 2, легчайший - 3, полулёгкий - 4, лёгкий - 5, полусредний - 6, средний - 7, полутяжёлый - 8, тяжёлый - 9.
- X8 - Знак зодиака.
- X9 - Пол (мужской, женский).

Далее, была спроектирована изображенная на рисунке 1 нейросеть с девятью нейронами на входном слое, с тремя нейронами на скрытом слое и одним нейроном на выходном.

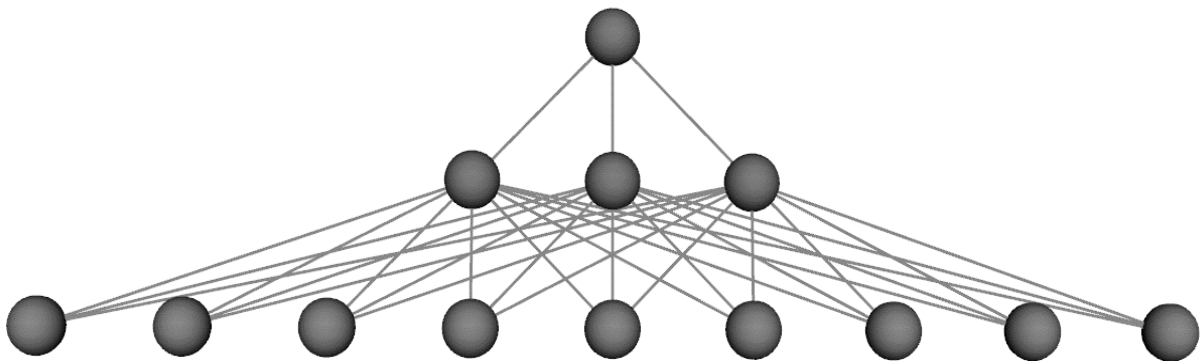


Рисунок 1. Спроектированная нейросеть

Для обучения и тестирования была собрана выборка из 150 бывших и действующих бойцов ММА. На рисунке 2 представлены прогнозируемый нейросетью и реальный проценты побед на тестовом множестве: 18 бойцов. Обученная нейросеть демонстрирует среднюю относительную погрешность, равную 7%.

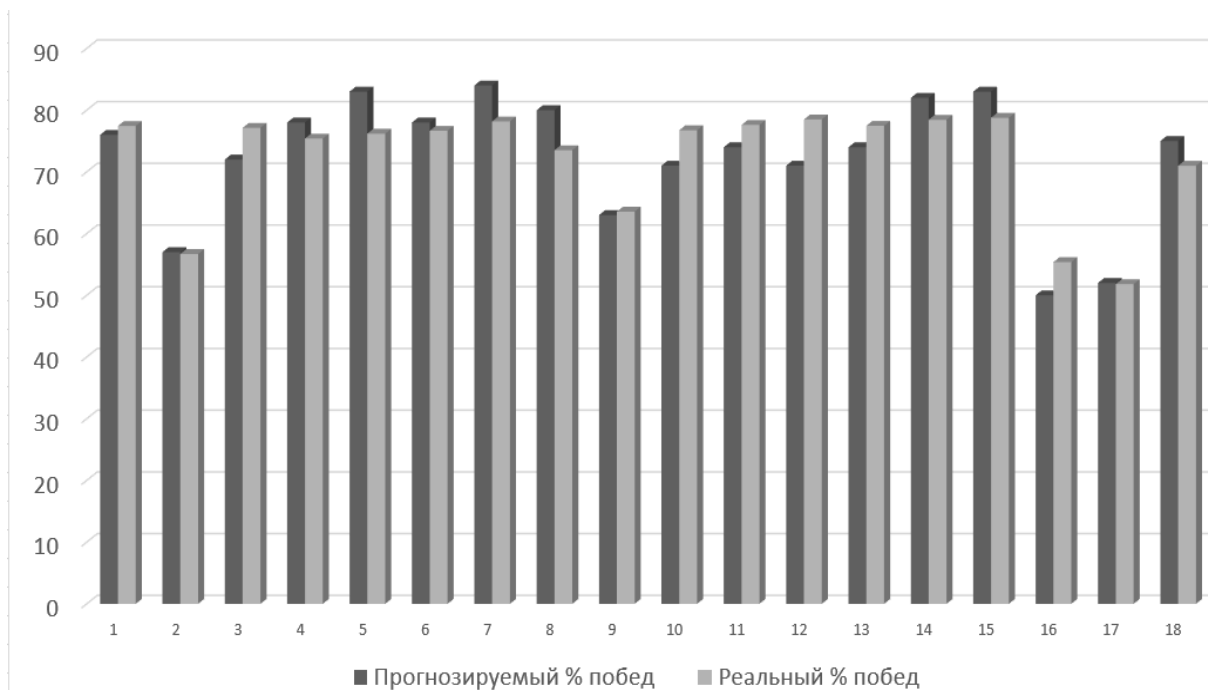


Рисунок 2. Прогнозируемый нейросетью и реальный проценты побед

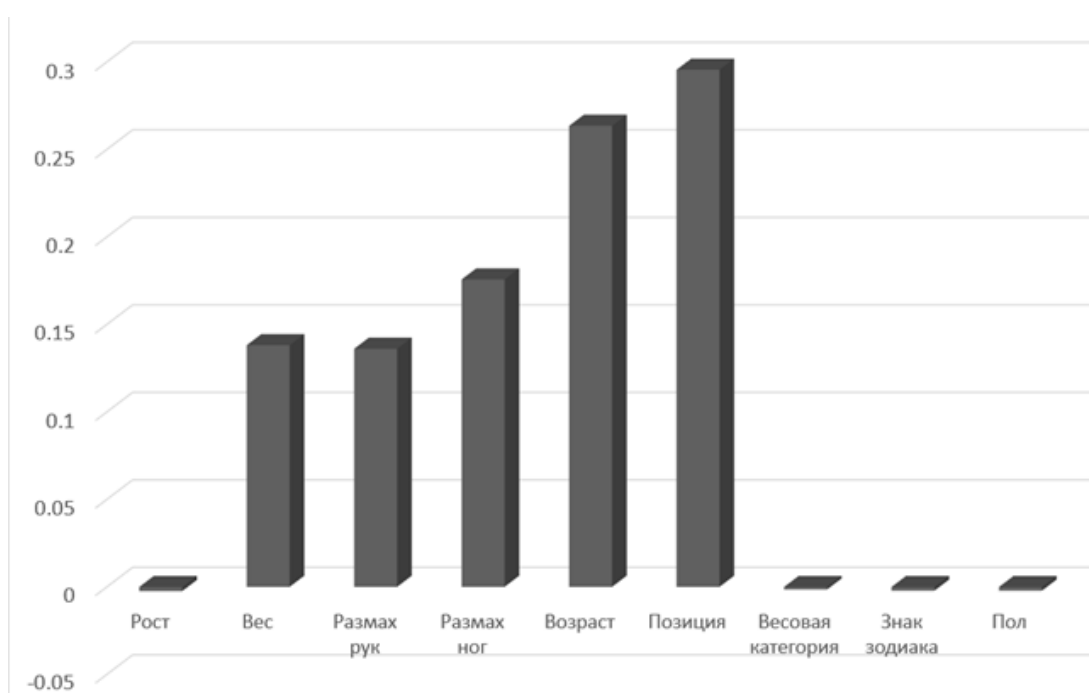


Рисунок 3. Значимость параметров

На основе данных значимости параметров (рис. 3), вычисленных нейросетью, можно выделить несколько основных параметров:

- Позиция (бойцы должны тренировать свою универсальность и разнообразить боевые стойки)
- Возраст – на возраст повлиять нельзя
- Размах ног и рук – врождённые физиологические параметры, которые практически невозможно изменить.

- Вес – самый легко меняемый параметр, чем ближе вес к границам весовой категории – тем боец эффективнее.

Возьмём бойца полусредней категории – Александра Яковлева. Александр имеет следующие параметры:

- Рост – 190 см
- Вес – 77 кг
- Размах рук – 188 см
- Размах ног – 107 см
- Возраст – 33 года
- Весовая категория – полусредняя
- Позиция - левша
- Знак зодиака - рак
- Пол – мужской
- Реальный процент побед – 75% (23 победы и 8 поражений)
- Спрогнозированный процент побед – 77%

Далее рассмотрим, как виртуальное изменение веса бойца (и вместе с тем его переход в другую весовую категорию) увеличит повлияет на процент его побед.

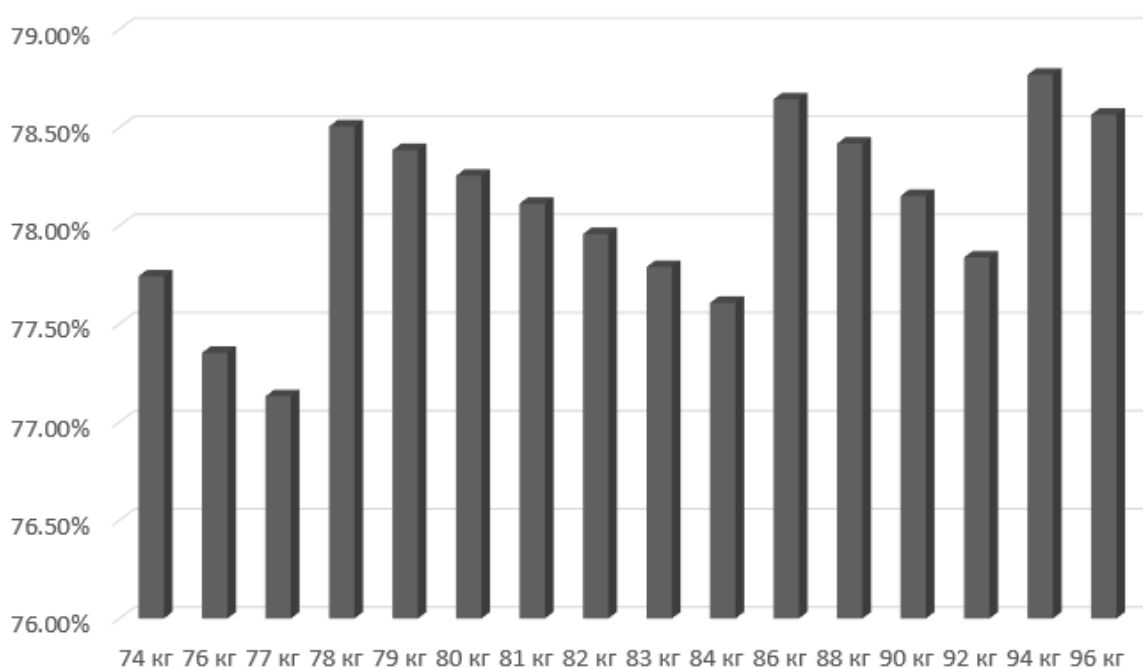


Рисунок 4. Прогнозируемый процент побед бойца в зависимости от виртуального изменения его веса

Как видно из рисунка 4, процент побед бойца резко увеличивается на минимальном весе каждой весовой категории. Так, наибольший процент побед бойца достигается переходом в полутяжёлую весовую категорию (большой рост бойца ему это позволяет) с весом 94 кг. В таком случае процент будет равен 78,77%, что на 1,77% больше изначального прогноза.

Закключение. Разработанная нейронная сеть позволяет не только прогнозировать результаты поединков, но и давать полезные рекомендации по улучшению результативности отдельных бойцов.

Библиографический список

1. Ясницкий Л.Н., Vnukova O.V. Прогноз результатов олимпиады-2014 в мужском одиночном фигурном катании методами искусственного интеллекта // Современные проблемы науки и образования. 2014. № 1. С. 189.
2. Ясницкий Л.Н., Кирасова А.В., Ратегова А.В., Черепанов Ф.М. Методика нейросетевого прогнозирования результатов спортивных состязаний на примере чемпионата мира-2015 по легкой атлетике // Вестник Пермского университета. Серия: Математика. Механика. Информатика. 2014. № 3 (26). С. 90-97.
3. Ясницкий Л.Н., Павлов И.В., Черепанов Ф.М. Прогнозирование результатов олимпийских игр 2014 года в неофициальном командном зачете методами искусственного интеллекта // Современные проблемы науки и образования. – 2013. – № 6.
4. Ясницкий Л.Н., Абрамова Ю.С., Бабушкина С.Д. Возможности получения рекомендаций по улучшению результативности сборных команд, готовящихся к участию в чемпионате Европы по футболу евро-2016 методом нейросетевого моделирования // Вестник спортивной науки. 2015. № 5. С. 15-20.

FORECASTING SUCCESS OF MMA FIGHTERS BASED ON THEIR PHYSIOLOGICAL PARAMETERS

Bakulin Nikita K.

Perm Campus of National Research University Higher School of Economics
614070, Russia, Perm, Studencheskaya, d. 38, bakulin.nikita@bk.ru

The article describes the task of predicting the success of hostilities without rules (MMA) based on their physiological indicators.

Methods of neural network modeling can be used to create similar computer programs that allow you to get recommendations on changes in physiological parameters.

Keywords: fights without rules, MMA, UFC, forecast, neural network, percentage of victories in a duel, effectiveness of a fighter.

Keywords: fights without rules, MMA, UFC, forecast, neural network, percentage of victories in a fight, effectiveness of a fighters.

ПРОГНОЗИРОВАНИЕ ВЕРОЯТНОСТИ ПЕРЕХОДА ИГРОКА ИЗ КОНТИНЕНТАЛЬНОЙ ХОККЕЙНОЙ ЛИГИ В НАЦИОНАЛЬНУЮ ХОККЕЙНУЮ ЛИГУ

Павлова Татьяна Викторовна

Пермский государственный национальный исследовательский университет
614990, Россия, г. Пермь, ул. Букирева, 15, PavlovaTV97@yandex.ru

В статье описан опыт разработки и применения компьютерной программы, предназначенная для прогнозирования и получения рекомендаций по улучшению результативности отдельных игроков Континентальной Хоккейной Лиги с учетом их индивидуальных особенностей и статистики. Работа программы продемонстрирована на примере игроков Континентальной Хоккей. Для игрока выявлены и исследованы зависимости вероятности прогноза от его возраста, роста, веса, индивидуальных показателей, достигнутых за прошедших сезон. Программа может быть использована для получения различных рекомендаций любого игрока Континентальной Хоккейной Лиги. Методика нейросетевого моделирования может быть использована и для создания аналогичных компьютерных программ, позволяющих получать рекомендации по улучшению результативности как отдельных игроков, так и спортивных команд, участвующих в любых соревнованиях в любых видах спорта.

Ключевые слова: хоккей, Континентальная Хоккейная Лига, прогноз, результативность, нейронная сеть.

1. Введение

В последнее время в спортивной науке наряду с классическими методами регрессионного анализа находят успешное применение нейросетевые технологии. Нейронные сети используются для предсказания спортивных состязаний, для прогнозирования численных значений мировых рекордов, для выявления перспективных спортсменов при формировании спортивных команд, для выявления наиболее важных факторов, влияющих на исход игры и др.

Следует отметить, что спортивные прогнозы, полученные ранее методом нейросетевого моделирования представителями Пермской научной школы искусственного интеллекта, оказались верны. Так, в работе [9], опубликованной в 2013 году, был выполнен подтверждавшийся впоследствии¹⁰ прогноз места в неофициальном командном зачете

© Павлова Т.В., 2019

¹⁰ В указанной работе российским спортсменам в их традиционных видах спорта прогнозировалось 4-е место на Олимпиаде-2014, тогда как по факту в медальном зачете Россия заняла первую строчку. Однако, если из 13-ти медалей, завоеванных российской сборной вычесть пять золотых медалей, принесенных

российских спортсменов на Зимней Олимпиаде-2014. В работах [6, 7], опубликованных в 2014 году, выполнен подтвердившийся прогноз результатов Чемпионата мира-2015 по легкой атлетике.

В мировой спортивной литературе практически отсутствуют публикации, посвященные применению еще одной замечательной возможности нейронных сетей: возможности не только предсказывать будущие события, но и получать рекомендации, позволяющие влиять на эти события, т.е. давать рекомендации спортсменам и спортивным командам – каким образом можно улучшить результаты и занять более высокие места, изменив прогнозы в лучшую сторону.

Впервые такие попытки были приняты в 2013 году и описаны в [5], тогда методом нейросетевого моделирования удалось получить несколько полезных рекомендаций для некоторых спортсменов, готовящихся к участию в Олимпиаде-2014. Следует отметить, что в тот момент, одна из рекомендаций была принята к сведению и, по-видимому, учтена тренером А. Мишиным при подготовке фигуриста Е. Плющенко¹¹.

В 2015 году в работе [10] с помощью той же получены рекомендации по улучшению результативности сборных команд, готовящихся к участию в чемпионате Европы по футболу евро-2016.

Настоящая статья посвящена прогнозированию вероятности переходов отдельных игроков Континентальной Хоккейной Лиги в Национальную Хоккейную Лигу.

2. Методика построения математической модели

При построении нейросетевой математической модели, было сформировано множество примеров, основанных на известных переходах игроков Континентальной Хоккейной Лиги в Национальную Хоккейную Лигу. Входные параметры модели, по которым оценивались футбольные команды, были взяты исходя из доступности информации в сети Интернет. Это: «Дата рождения» (день, месяц, год), «Возраст игрока», «Рост», «Вес», «Гражданство» (Россия, Финляндия, Другое: Казахстан, Чехия, Швеция и т.д.), «Амплуа игрока» (защитник, нападающий), данные о последнем сезоне в КХЛ: «Участие в плей-офф», «Количество сыгранных матчей», «Гол», «Пас», «Штрафное время», «Коэффициент полезности игрока», «Время игры» (минуты, секунды), «Участие в мировых турнирах», «Участие в чемпионатах мира», «Участие в Олимпиаде», «Участие в «Турнире 4-х»» (Кубок Первого Канала, Кубок Карьяла, Шведские игры, Чешские игры), «Участие в другие мировых турнирах» (Кубок Мира, Евротур и т.д.), таким образом, всего 24 параметра.

России приглашенными иностранными спортсменами Виком Уайльдом и Виктором Аном (в нетрадиционных для России видах спорта), то как раз и получается прогнозируемое 4-е место.

¹¹ В указанной работе [5] фигуристу Е. Плющенко прогнозировалось 2-е место на Олимпиаде-2014. Однако, в этой же работе ему было рекомендовано сбавить вес, с чем, незадолго до олимпиады, согласился его тренер А.Мишин (см. [2]). Напомним, что в командных соревнованиях на Олимпиаде-2014 Е.Плющенко показал блестящий результат, благодаря чему команда получила золотую медаль.

Выходной параметр модели кодирует результат следующим образом: 0 – осуществлен переход в Национальную Хоккейную Лигу по результатам прошлого сезона, 1 – игрок перешел в Национальную Хоккейную Лигу.

Всё множество собранных примеров состоит из 105 элементов, которое было разбито в процентном соотношении 70% : 30%. 70% - обучающее множество, которое использовалось для обучения сети, 30% - тестирующее множество, предназначенное для проверки ее прогностических свойств. Естественно, что примеры тестирующего множества при обучении сети не использовались. Проектирование, оптимизация, обучение, тестирование нейронной сети и эксперименты над нейросетевой математической моделью выполнялись по методике [4, 6] с помощью нейросимулятора [3]. Среднеквадратичная ошибка на тестирующих примерах, в т.ч. полученная методом многократной перекрестной проверки, составила 39%.

3. Результаты компьютерных экспериментов

Объективная оценка степени влияния входных параметров на результат математического моделирования была получена путем поочередного исключения входных параметров и наблюдением за погрешностью ее тестирования. Чем выше оказывается погрешность тестирования, тем значимее соответствующий исключенный входной параметр. Построенная таким способом гистограмма, представлена на рис. 1, показывает, что наибольшее влияние на результат перехода игрока имеют следующие параметры: «Штрафное время», «Коэффициент полезности», «Участие в различных мировых турнирах», а также «Время игры».

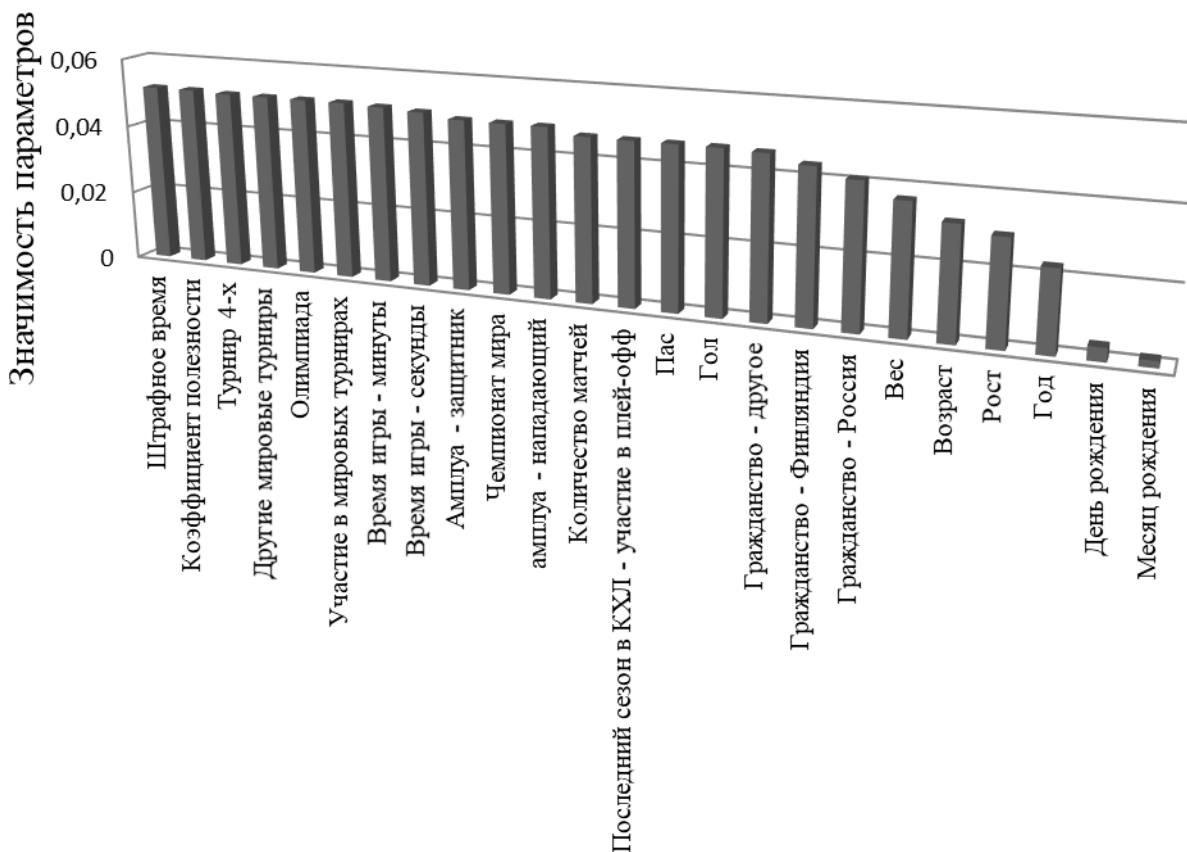


Рисунок 1. Оценка степени влияния параметров игроков на вероятность перехода игрока из КХЛ в НХЛ

Дальнейшая цель исследований состояла в выяснении характера влияния наиболее значимых входных параметров модели на результат моделирования. Для проведения виртуальных компьютерных экспериментов был выбран один из хоккеистов. Цель исследований состояла в подборе параметров таким образом, чтобы данный игрок смог совершить переход из Континентальной Хоккейной Лиги в Национальную Хоккейную Лигу.

Исследования проводились по следующим параметрам: «Штраф», «Коэффициент полезности», «Участие в мировых турнирах», «Участие в Олимпиаде», «Участие в турнире 4-х наций», а также участие в других мировых турнирах. Исследования выполнены путем фиксации одних входных параметров и изменения других входных параметров при одновременном наблюдении за значением выходного сигнала нейросети, показывающего прогнозируемый результат.

На рис. 2 представлены результаты прогнозирования прохождения игрока из одной лиги в другую при увеличении штрафного времени. Это возможно, если у игрока достаточно высокие другие показатели за прошедший сезон.

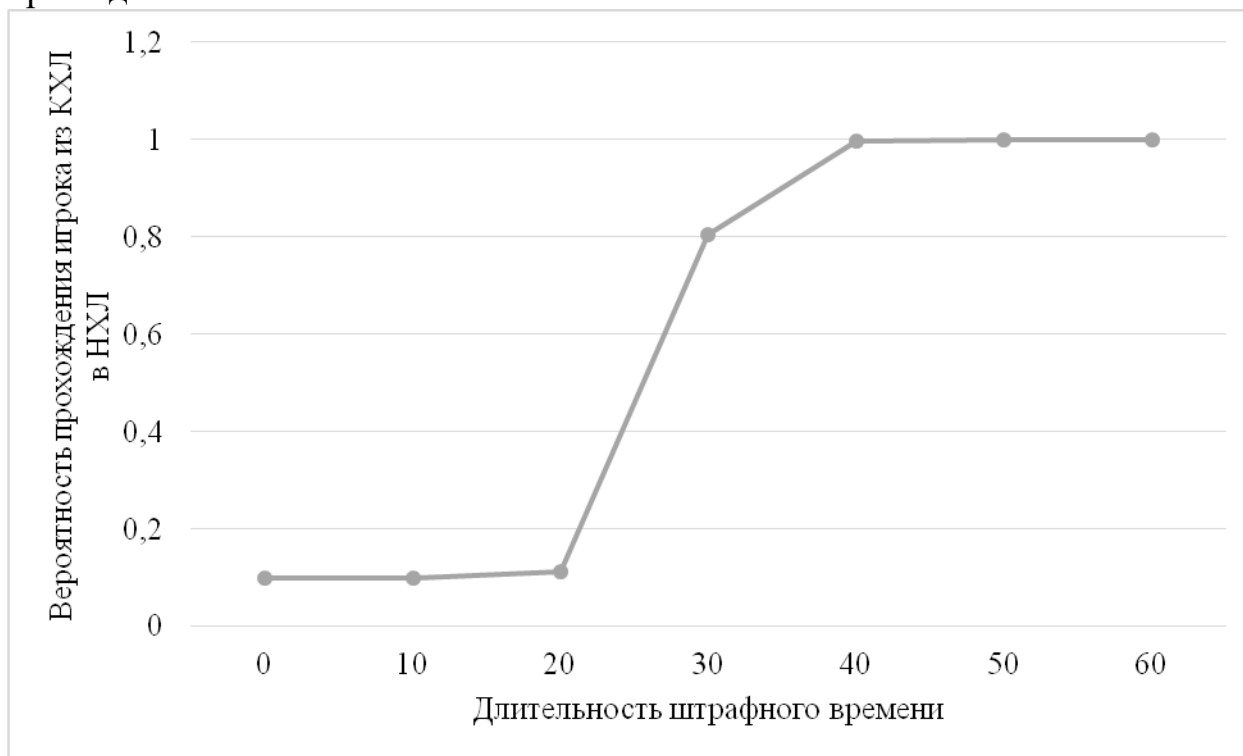


Рисунок 2. Зависимость вероятности прохождения игрока из КХЛ в НХЛ от длительности штрафного времени.

На рисунке 3 изображены зависимости прогнозируемых результатов игроков, полученные при изменении коэффициента полезности. Анализируя полученные данные, можно сказать, что вероятность прохождения игрока из одной лиги в другую значительно повышается, если его коэффициент полезности увеличивается за весь сезон.

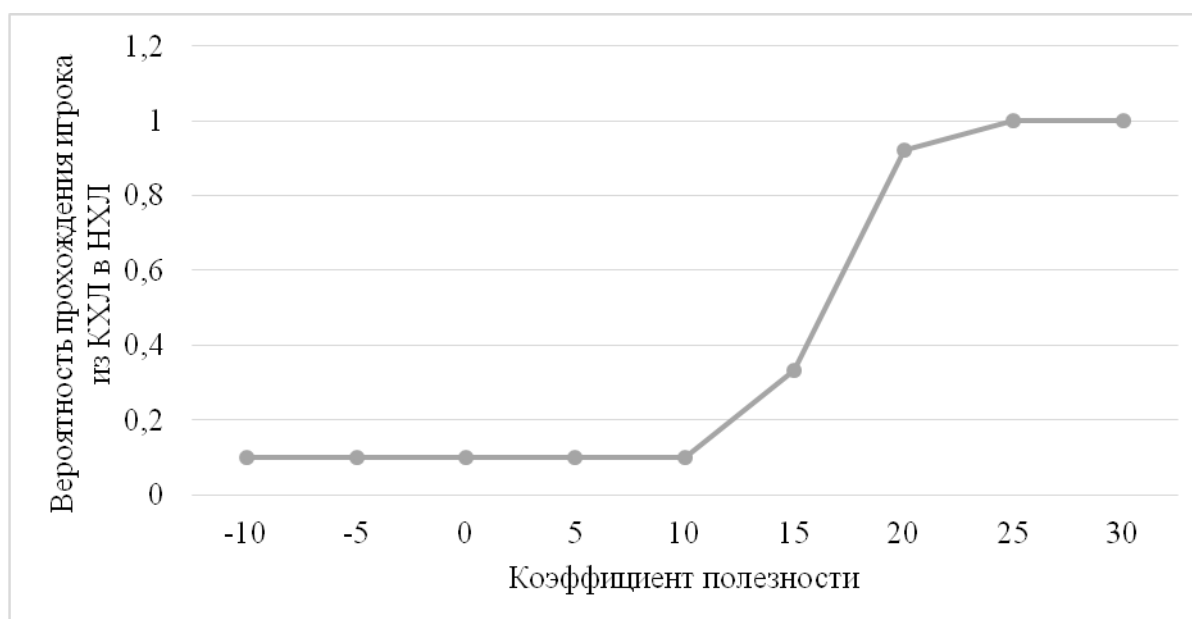


Рисунок 3. Зависимость вероятности прохождения игрока из КХЛ в НХЛ при изменении коэффициента полезности

Как видно из рис. 4, полученного аналогичным способом, шансы игрока на переход из Континентальной Хоккейной Лиги в Национальную Хоккейную Лигу значительно повышается, если повышается его время проведённого на поле.

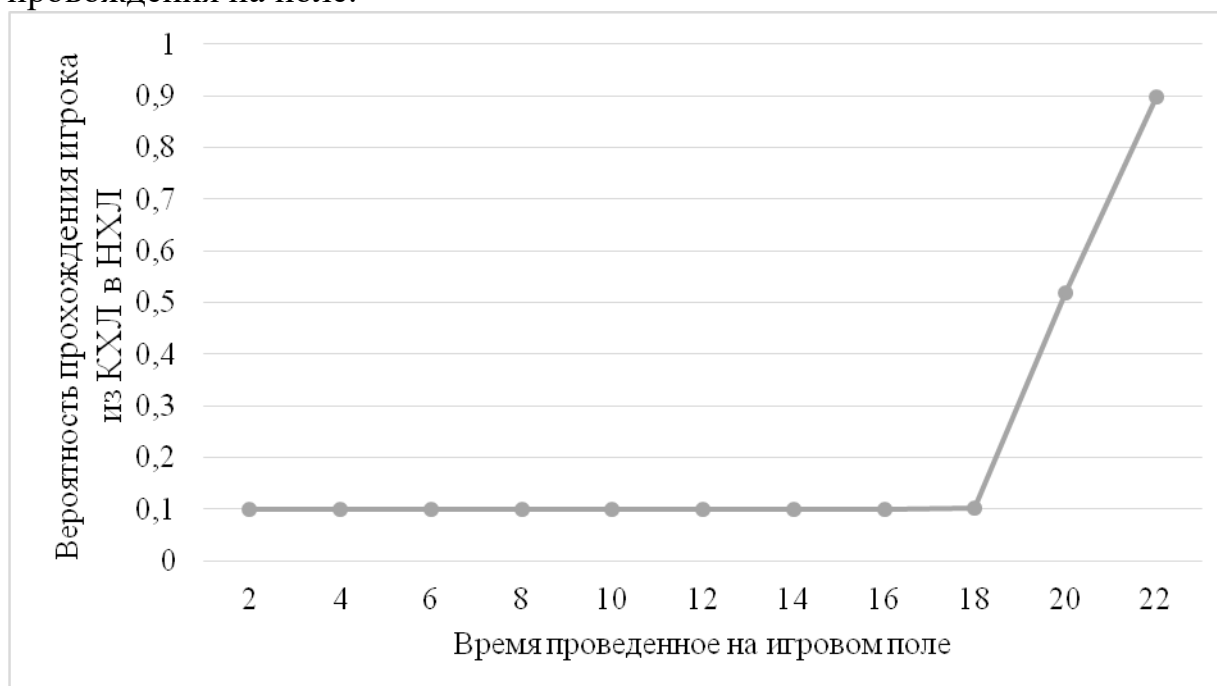


Рисунок 4. Зависимость прогнозируемого результата от времени, проведенного игроком на игровом поле за весь сезон

4. Заключение

Создана нейросетевая математическая модель, позволяющая прогнозировать вероятность перехода игрока из Континентальной Хоккейной Лиги в Национальную Хоккейную Лигу. Нейросетевая модель

реализована в виде компьютерной программы, работа которой продемонстрирована.

Разработанная компьютерная программа, может быть использована для прогнозирования и получения полезных рекомендаций для всех игроков Континентальной Хоккейной Лиги. Методика нейросетевого моделирования может быть использована для создания аналогичных компьютерных программ, позволяющих получать рекомендации по улучшению результативности как спортивных команд, так и отдельных спортсменов, готовящихся к участию в любых соревнованиях в любых видах спорта.

Библиографический список

1. Касюк С.Т., Вахтомова Е.М. Использование нейронных сетей для анализа и прогнозирования данных в физической культуре и спорте // Ученые записки университета имени П.Ф. Лесгафта. - 2013. - № 12 (106). - С. 72-77.
2. Семченко О. Пермские ученые занялись спортивными прогнозами // Поиск. Еженедельная газета Научного сообщества. 2014. № 7.
3. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о государственной регистрации программы для ЭВМ № 2014612546. Заявка Роспатент № 2014610341. Зарегистрировано в Реестре программ для ЭВМ 28 февраля 2014г.
4. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176с.
5. Ясницкий Л.Н., Внукова О.В., Черепанов Ф.М. Прогноз результатов олимпиады-2014 в мужском одиночном фигурном катании методами искусственного интеллекта // Современные проблемы науки и образования. – 2014. – № 1; URL: <http://www.science-education.ru/115-11339> (дата обращения: 25.12.2013).
6. Ясницкий Л.Н., Киросова А.В., Ратегова А.В., Черепанов Ф.М. Методика прогнозирования результатов спортивных состязаний на примере Чемпионата мира-2015 по легкой атлетике // Вестник Пермского университета. Серия: Математика. Механика. Информатика. 2014. № 3. С. 90-97.
7. Ясницкий Л.Н., Кировоса А.В., Ратегова А.В., Черепанов Ф.М. Прогноз результатов Чемпионата мира-2015 по легкой атлетике методами искусственного интеллекта // Современные проблемы науки и образования. – 2014. – № 4; URL: www.science-education.ru/118-14423 (дата обращения: 28.08.2014).
8. Ясницкий Л.Н. Нейронные сети – инструмент для получения новых знаний: успехи, проблемы, перспективы // Нейрокомпьютеры: разработка, применение. 2015. № 5. С. 48-56.
9. Ясницкий Л.Н., Павлов И.В., Черепанов Ф.М. Прогнозирование результатов олимпийских игр 2014 года в неофициальном командном зачете методами искусственного интеллекта // Современные проблемы науки и образования. – 2013. – № 6; URL: <http://www.science-education.ru/113-11206> (дата обращения: 25.12.2013).

10. Ясницкий Л.Н., Абрамова Ю.С., Бабушкина С.Д. Возможности получения рекомендаций по улучшению результативности сборных команд, готовящихся к участию в чемпионате Европы по футболу евро-2016 методом нейросетевого моделирования // Вестник спортивной науки. 2015. № 5. С. 15-20.

FORECASTING THE PROBABILITY OF TRANSITING A PLAYER FROM THE CONTINENTAL HOCKEY LEAGUE TO THE NATIONAL HOCKEY LEAGUE

Pavlova Tat'yana V.

- Perm State University

614990, Russia, Perm, street Bukireva, 15, PavlovaTV97@yandex.ru

The article describes the experience of developing and using a computer program designed to predict and obtain recommendations for improving the performance of individual players of the Continental Hockey League, taking into account their individual characteristics and statistics. The work of the program is demonstrated by the example of Continental Hockey players. For the player, dependences of the forecast probability on his age, height, weight, and individual indicators achieved during the past season have been identified and investigated. The program can be used to receive various recommendations of any player of the Continental Hockey League. The neural network modeling technique can also be used to create similar computer programs that allow you to receive recommendations for improving the performance of both individual players and sports teams participating in any competitions in any kind of sports.

Key words: hockey, Continental Hockey League, forecast, performance, neural network.

УДК 004.032.26, 796.332

ВОЗМОЖНОСТЬ ПРОГНОЗИРОВАНИЯ КОЛИЧЕСТВА ГОЛОВ, ЗАБИТЫХ КОМАНДОЙ ЧЕЛСИ В ОДНОМ МАТЧЕ

Шестаков Егор Сергеевич

Пермский государственный национальный исследовательский университет,
Механико-математический ф-т. КМБ

614990, Россия, г. Пермь, ул. Букирева, 15, esshestakov95@gmail.com

На базе нейронной сети разработана программа, предназначенная для получения результатов футбольного матча до его начала. В качестве примера рассмотрена команда Челси. Данные для обучения собраны на основе

предыдущих матчей Челси в рамках английского чемпионата. Программа позволяет получить прогнозное количество голов забитых в матче.

Ключевые слова: искусственный интеллект, нейронная сеть, прогноз, количество голов, тотал, футбол, Челси.

Введение. В современном мире спорт перестал быть зрелищем для простых людей, он трансформировался в прибыльный бизнес. Многие известные бизнесмены приобретают доли в футбольных клубах, чтобы увеличить свое состояние путем продажи молодых игроков и присвоения призовых, которые выплачивают футбольные ассоциации. Параллельно с этим появился и другой способ заработать на достижении других людей – букмекерское дело, то есть возможность заключить пари со специальной организацией, предметом которого является результат спортивного события.

Проблема предсказания спортивных результатов является актуальной для многих людей так, как, разрешив ее хотя бы отчасти, можно заработать очень большие деньги. В сети Интернет существует бесчисленное количество групп и сообществ, предоставляющих различные услуги в данной области, вплоть до продажи прогнозов. Игроки тратят много времени на анализ спортивных событий, чтобы дать максимально точный прогноз, стоит отметить, что это удается единицам, настоящим профессионалам, которым не без везения и интуиции удается спрогнозировать часть возможных исходов.

Так как человеку сложно учесть все возможные факторы и однозначно определить значимость каждого из них, для прогнозирования исходов спортивных состязаний начали использовать нейронные сети. Так, в [1] предлагается методика нейросетевого прогнозирования результатов спортивных состязаний на примере чемпионата мира-2015 по легкой атлетике. В [2] выполнен прогноз результатов олимпиады-2014 в мужском одиночном фигурном катании методами искусственного интеллекта. Работа [3] посвящена описанию возможностей получения рекомендаций по улучшению результативности сборных команд по футболу методом нейросетевого моделирования.

В настоящей работе инструментарий [1-3] применен для создания нейросетевой модели, способной прогнозировать количество голов, забитых командой Челси в одном матче.

Методика построения математической модели. Для создания нейросетевой математической модели, было сформировано множество примеров, основанных на результатах матчей команды Челси последних трех лет в чемпионате Англии. Данные были взяты из [5-7]. Входные параметры, по которым оценивалась команды, были взяты исходя из доступной информации в сети Интернет. Это “домашний матч” (играет команда на своем стадионе), “позиция” (положение в турнирной таблице перед матчем), “позиция соперника”, “первый сезон тренера”, “играет лучший бомбардир”, “играет основной вратарь”, “месяц” (когда проводился матч), “перед матчем в Европе”, “после матча в Европе”, “матч в Лондоне”, “следующий соперник

топ уровня” (топ соперником считается одна из команд: Арсенал, Ливерпуль, Манчестер Сити, Манчестер Юнайтед, Тоттенхем), ”соперник топ”, “2 матча в неделю” (если с момента последнего матча прошло меньше 4-х дней), “играет в Лиге Чемпионов”, “ играет в Лиге Европы”, “не играет в еврокубках”, всего 15 параметров.

Выходной параметр показывал количество голов, забитых в матче.

Все множество собранных примеров было разбито в соотношении 90:10 на обучающее множество, которое использовалось для обучения сети, и тестирующее множество, предназначенное для проверки ее прогностических свойств. Примеры тестирующего множества при обучении сети не использовались. Проектирование, оптимизация, обучение, тестирование нейронной сети и эксперименты над нейросетевой математической моделью выполнялись с помощью нейросимулятора [4]. Среднеквадратичная ошибка на тестирующих примерах составила 18%.

Результаты компьютерных экспериментов. Так как футбол стремительно развивается, результаты прошлых лет быстро устаревают. Поэтому было принято решение оценивать последние три сезона. Благодаря этому В результате получилась нейронная сеть, с относительно низкой погрешностью тестирования 18%. В результате тестирования были получены результаты, представленные на рисунке 1, которые показали, что нейронная сеть не может предсказать результат с высокой точностью, отклонение от действительного значения составляет чуть более 1 гола.

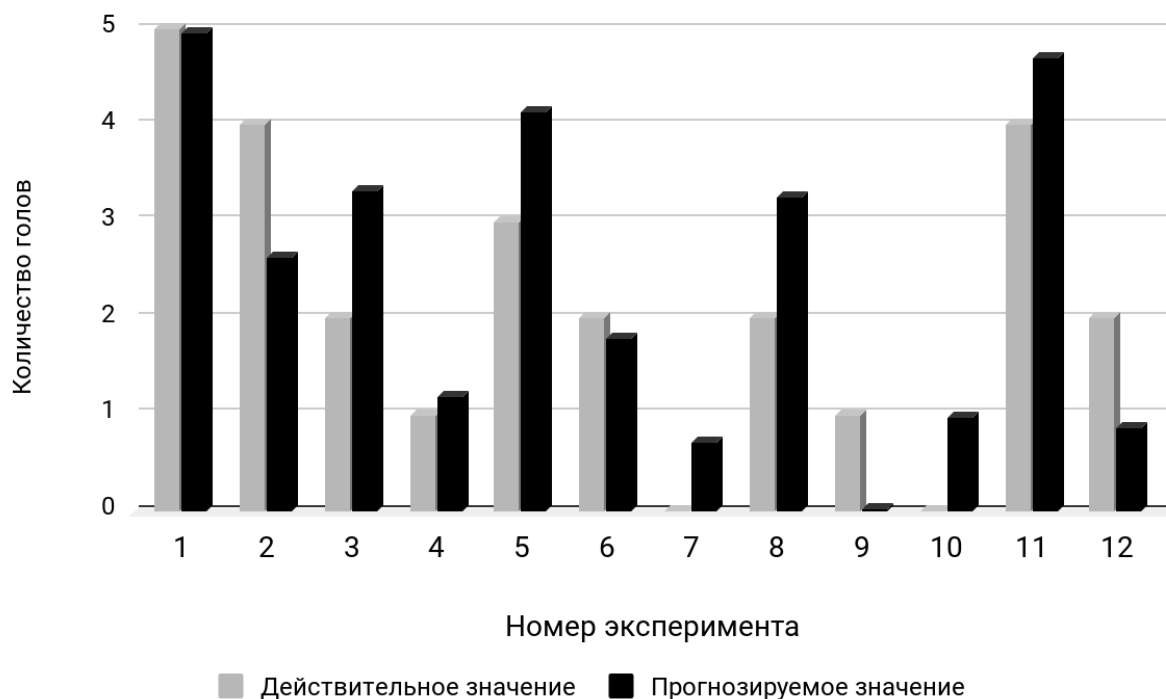


Рисунок 1. Результаты тестирования

Значимыми параметрами для прогнозирования оказались: «следующий матч в Европе» и «предыдущий матч в Европе». Это можно объяснить тем, что матч в Европе означает участие команды в другом турнире, который многие тренеры ценят больше чем национальный чемпионат, то есть они

могут выставлять на игру не оптимальный состав. Также команде требуется совершить длительный перелет, что негативно влияет на форму футболистов. Распределение значимостей параметров представлена на рисунке 2.

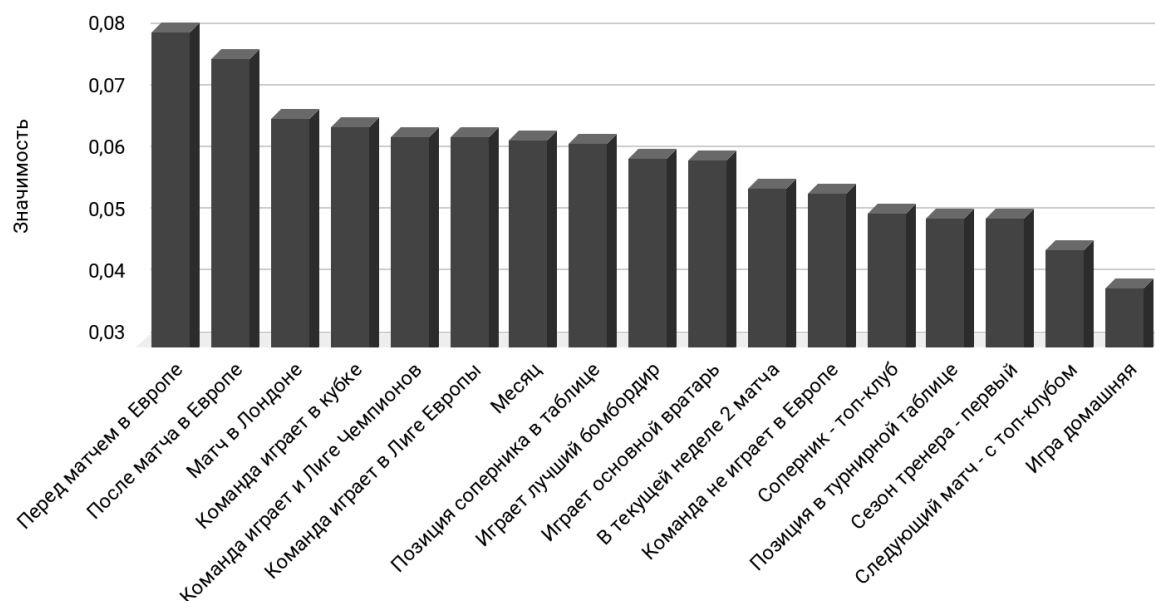


Рисунок 2. Значимость параметров

Для определения влияния параметров на результат был проведен эксперимент: изменены значения параметров («перед матчем в Европе», «после матча в Европе», «месяц») одного матча и спрогнозировано количество забитых голов. В качестве базового матча был взят матч со следующими параметрами: Челси играл дома, позиция Челси в таблице перед матчем - 15, позиция соперника в таблице перед матчем - 7, был первый сезон тренера, играл лучший бомбардир, играл основной вратарь, матч был в декабре, следующий матч был не в Европе, предыдущий матч был не в Европе, матч был проведен в Лондоне, Челси продолжал борьбу за кубок Англии, соперник не был топ-клубом, следующий соперник был топ-клубом, 1 матч в неделю, Челси играл в Лиге чемпионов, Челси не играл в Лиге Европы, Челси играл в Европе.

На рисунках 3 и 4 приведены результаты эксперимента: сравнение результатов при равных значениях всех параметров за исключением исследуемых. Серым цветом отмечены реальные значения, черным цветом - экспериментальные.

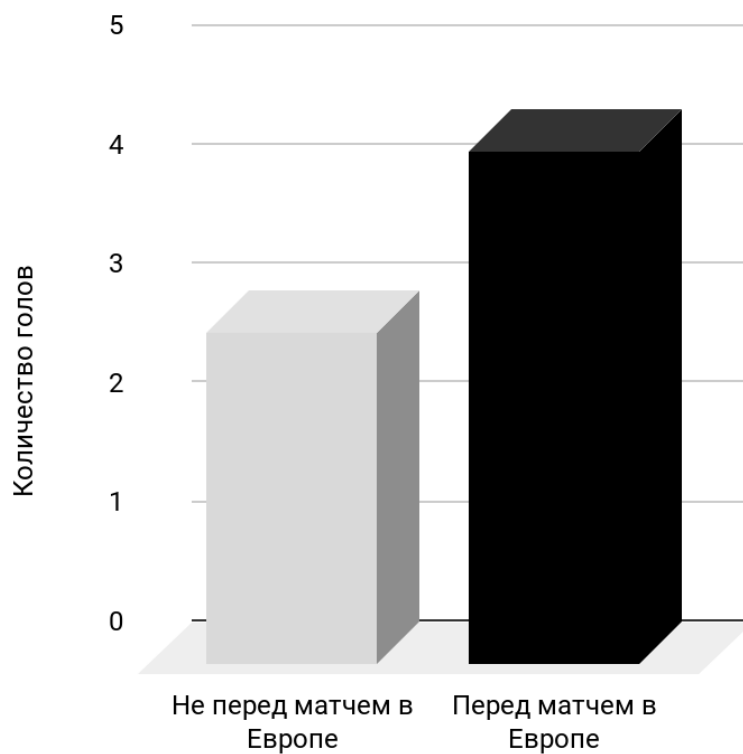


Рисунок 3. Исследование влияния параметра “Перед матчем в Европе”

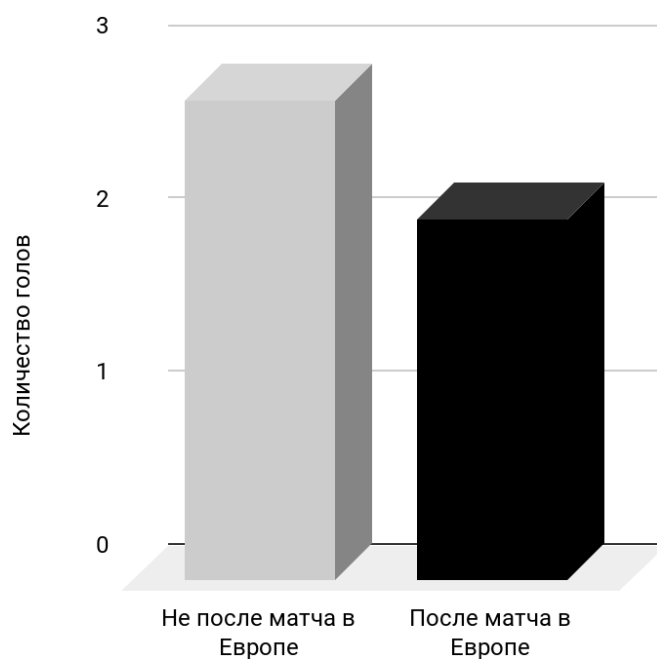


Рисунок 4. Исследование влияния параметра “После матча в Европе”

На рисунке 5 приведено исследование влияния месяца игры на результат.

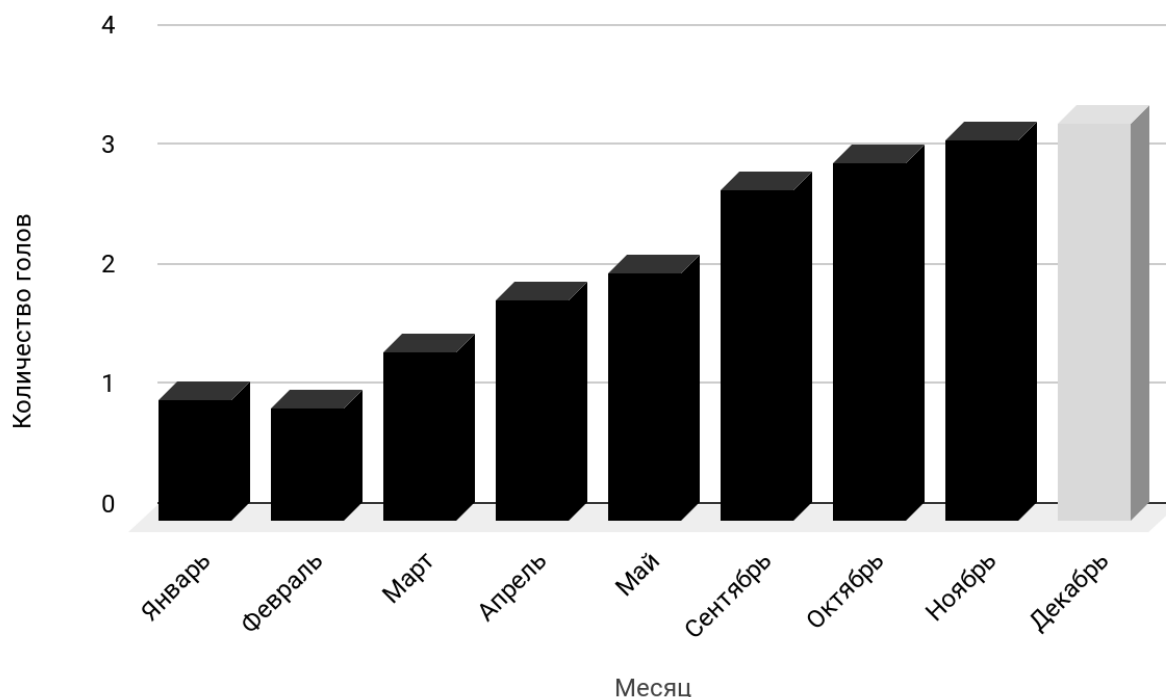


Рисунок 5. Исследования влияния месяца

Данный результат можно объяснить тем, что, когда начинается чемпионат, команды только обретают свои лучшие кондиции, идет подъем до нового года. В это время в Англии начинается “boxing day”, то есть матчи проводятся очень часто, футболисты не успевают восстанавливаться и, как следствие, снижается результативность. После чего календарь нормализуется и форма восстанавливается.

Заключение. В результате проведенного исследований была разработана программа, которая позволяет предсказать количество забитых голов в одном матче командой Челси в рамках Английского чемпионата с точностью до одного гола.

Библиографический список

1. Ясницкий Л.Н., Киросова А.В., Ратегова А.В., Черепанов Ф.М. Методика нейросетевого прогнозирования результатов спортивных состязаний на примере чемпионата мира-2015 по легкой атлетике // Вестник Пермского университета. Серия: Математика. Механика. Информатика. 2014. № 3 (26). С. 90-97.
2. Ясницкий Л.Н., Vnukova O.V. Прогноз результатов олимпиады-2014 в мужском одиночном фигурном катании методами искусственного интеллекта // Современные проблемы науки и образования. 2014. № 1. С. 189.
3. Ясницкий Л.Н., Абрамова Ю.С., Бабушкина С.Д. Возможности получения рекомендаций по улучшению результативности сборных команд, готовящихся к участию в чемпионате Европы по футболу

Евро-2016 методом нейросетевого моделирования // Вестник спортивной науки. 2015. № 5. С. 15-20

4. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о государственной регистрации программы для ЭВМ № 2014612546. Заявка Роспатент № 2014610341. Зарегистрировано в Реестре программ для ЭВМ 28 февраля 2014г.
5. Электронная база данных: [Электронный ресурс] [Режим доступа <https://www.myscore.ru/>] [Проверено: 05.04.2019].
6. Электронная база данных: [Электронный ресурс] [Режим доступа <https://sport.ua>] [Проверено: 05.04.2019].
7. Электронная статья: [Электронный ресурс] [Режим доступа <https://www.sports.ru>] [Проверено: 05.04.2019].

POSSIBILITY OF FORECASTING THE NUMBER OF GOALS BY THE TEAM OF CHELSEA IN ONE MATCH

Shestakov Egor S.
Perm State University
614990, Russia, Perm, street Bukireva, 15, esshestakov95@gmail.com

On the basis of the neural network, a program was designed to obtain the results of a football match before it starts. As an example, the Chelsea team was considered. Training data was collected based on previous Chelsea matches in the APL. The program allows you to get the exact number of goals scored in the match.

Key words: artificial intelligence, neural network, forecast, number of goals, total.

ЦИФРОВАЯ ЛИНГВИСТИКА И ЦИФРОВОЕ ИСКУССТВО

СИНТЕЗ МУЗЫКАЛЬНОГО ПРОИЗВЕДЕНИЯ ПО ТЕКСТОВОМУ ОПИСАНИЮ

*Волощук Мария Александровна, Акишев Константин Андреевич,
Розалиев Владимир Леонидович, Орлова Юлия Александровна*
Волгоградский государственный технический университет,
400005, Россия, г. Волгоград, пр. им. Ленина, 28,
vladimir.rozaliev@gmail.com

В статье описан метод синтеза музыкального произведения по текстовому описанию на основе наборов разнородных мелодий одного жанра. Выделены основные этапы по анализу текстовой и музыкальной составляющей музыкального произведения.

Ключевые слова: синтез музыки, анализ текста, генерация музыкального произведения, методы алгоритмической музыкальной композиции

В настоящее время наблюдается высокая потребность в разного рода музыкальных композициях различных жанров. С начала 20 века и с внедрением в кинофильмы звуковых дорожек музыка в фильмах стала неотъемлемой частью. С каждым годом потребность в однотипных фоновых звуковых дорожках росла. И на 2018 год в любом фильме используются в среднем около 20-25 фоновых музыкальных композиций с простыми музыкальными фразами и размерами музыкального ритма. Создание любых музыкальных произведений производится композиторами вручную. Даже над созданием фоновых композиций к фильмам, рекламе, tv-передачам и компьютерным играм работают композиторы, что увеличивает временные и финансовые расходы таких компаний.

Отечественными и зарубежными разработчиками и композиторами было создано множество программ для различного рода синтезирования музыкальных произведений. Однако все они базируются на методах, которые не позволяют создавать большого количества мелодий из определенной выборки существующих наборов звуковых дорожек [1].

Таким образом, было решено разработать метод генерации музыкального произведения по текстовому описанию на основе наборов разнородных мелодий. Целью разрабатываемого метода является повышение эффективности создания однотипных мелодий при помощи метода синтеза музыкального произведения на основе мелодий одного жанра с использованием методов алгоритмической музыкальной композиции, а также

методов сравнения высот мелодий, тональности и их частот. Эффективность измеряется временными затратами на генерацию музыкального произведения [3].

Разрабатываемый метод синтеза музыкальных произведений состоит из нескольких частей.

Первая часть заключается в подготовке данных. На сайтах, где музыкальные произведения находятся в свободном доступе, таких как Яндекс.Музыка, Spotify и др., выбираются музыкальные произведения одного жанра. Так как для синтеза музыки необходимы короткие мелодии, берется не всё произведение, а только припев, поскольку он является центральной частью песни, её кульминацией, и в первую очередь слушатель запоминает именно его.

Чтобы определить, какие слова характеризуют данное произведение, проводится анализ текста припева и определяется частота встречаемости различных слов. Как правило, значимые слова, несущие основную мысль песни, повторяются несколько раз. Таким образом, отбираются слова с наибольшей частотой встречаемости и назначаются как ключевые для данного музыкального отрывка.

В связи с тем, что при соединении текстов разным музыкальных произведений есть большая вероятность получить бессмысленный набор слов, производится поиск готовой минусовой фонограммы припева произведения, в которой отсутствует партия вокала, либо минусовой фонограммы целого произведения, из которой вырезается и оставляется отрывок, содержащий припев.

Выделенные ключевые слова и минусовая фонограмма припева музыкального произведения заносятся в базу, которая в дальнейшем будет использоваться при синтезе музыки.

Во второй части происходит обработка текстового описания и его сопоставление с мелодиями. Алгоритм обработки текста включает в себя несколько этапов. В первую очередь исходный текст, поданный в программу, разбивается на отдельные токены - слова. Разбивка происходит по пробельным символам и знакам препинания, таким как точка, запятая, восклицательный и вопросительный знаки, двоеточие, точка с запятой, открывающие и закрывающие кавычки и скобки. На выходе получается набор слов для дальнейшей работы.

Затем для этих слов с помощью готовых библиотек определяются части речи и отбрасываются незначительные части речи, такие как предлоги, союзы, частицы и прочие.

Наконец, завершающий этап обработки текста представляет собой постановку слов в начальную форму. Для глагола начальной формой является инфинитив, для существительного – единственное число и именительный падеж, для прилагательного, в отличие от существительного, добавляется ещё мужской род.

После того, как текст обработан, выбранные слова сравниваются с ключевыми словами из базы, и им в соответствие ставятся мелодии, которые

в дальнейшем будут синтезированы в одно музыкальное произведение. Если несколько слов из текста являются ключевыми к одному и тому же музыкальному отрывку, то данный музыкальный отрывок будет использован только один раз.

Третья часть заключается в синтезе музыкального произведения. Сутью разрабатываемого метода является разбивка нескольких музыкальных композиций на n -секундные звуковые дорожки, анализ данных частей произведения и нахождения соответствий в частоте, тональности и высоте. При этом следует учитывать то, что не все звуки и мотивы будут подходить друг другу при выравнивании частоты, высоты и тональности. Из этого следует, что необходим также анализ на совмещение октав музыкальной нарезки, а также на выстраивание гамм в различных октавах [2].

В ходе разработки метода был составлен алгоритм анализа различных разнородных звуковых последовательностей на предмет совпадения с образцом, выбранным для совмещения композиций (рис.1).

Также был выделен бизнес-процесс данного метода, в котором отражены все возможные обращения и работа пользователя с данным методом. В ходе разработки бизнес-модели были учтены как внешние ресурсы, из которых будут взяты новые музыкальные композиции, так и уже ранее задействованные в других проектах композиции, благодаря чему наращивается библиотека композитора для дальнейшей быстрой и эффективной работы (рис.2).

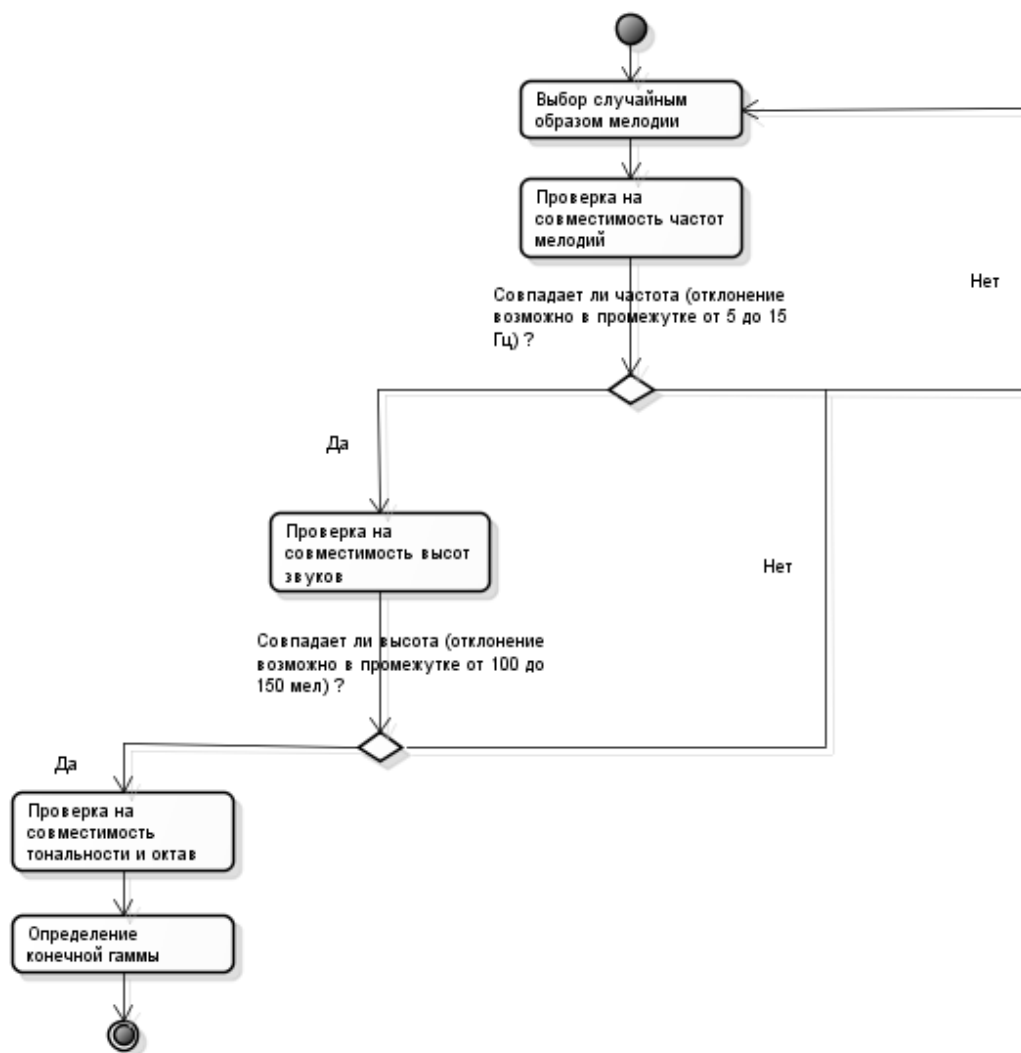


Рисунок 1. Алгоритм поиска соответствий между образцом и случайно выбранной мелодией

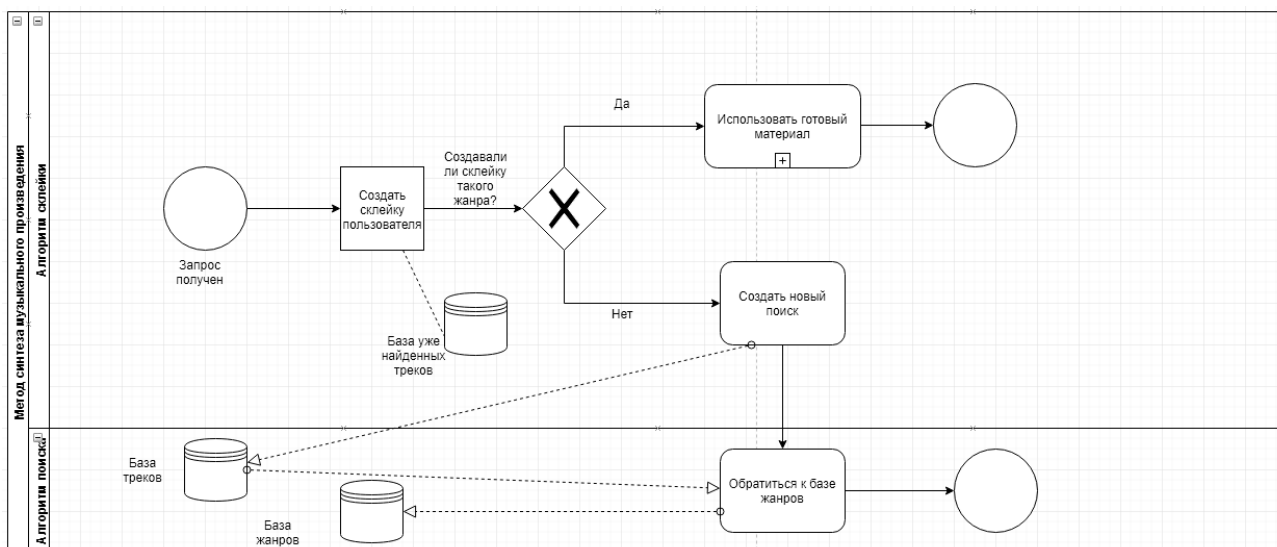


Рисунок 2. Диаграмма бизнес-процесса разрабатываемого метода

В ходе данной разработки были проанализированы существующие программы генерации звуковых последовательностей, найдены и

проанализированы некоторые методы алгоритмической музыкальной композиции, а также составлены алгоритмы для анализа и создания новых музыкальных композиций из набора готовых музыкальных произведений одного жанра.

Данный метод может способствовать эффективному и быстрому созданию фоновых разнородных мелодий одного жанра, что удешевит составление новых фоновых музыкальных произведений у композиторов, поскольку позволит без минимальных затрат разработать большое количество мелодий в короткий срок.

Работа выполнена при частичной поддержке РФФИ (гранты 17-07-01601, 18-47-342002, 18-07-00220, 19-07-00020).

Библиографический список

1. Чубарьян, А.С. Инкрементная генерация музыкальных произведений на основе динамических паттернов : дис. ... магистра прикладной математики и информатики: 010400 / Чубарьян Артем Суменович; Южный Федеральный Университет. - Ростов н/Д., 2013. – 65с.
2. Фазылова Э.Ф. Системы генерации музыки или как автоматизировать искусство? // Молодежный научно-технический вестник МГТУ им. Н.Э. Баумана. – 2014. - № 3.
3. Nierhaus G. Algorithmic Composition Paradigms of Automated Music Generation / G. Nierhaus. – Springer Wien New York, 2009. – 287 p.

SYNTHESIS OF MUSICAL WORK ON TEXT DESCRIPTION

Voloshchuk M., Akishev K., Rozaliev V., Orlova Yu.

Volgograd State Technical University

400005, Russia, Volgograd, Lenin avenue, 28, vladimir.rozaliev@gmail.com

The article describes a method for synthesizing a musical work according to a textual description based on sets of heterogeneous melodies of the same genre. The main stages of analyzing the textual and musical component of a musical work are highlighted.

Key words: music synthesis, text analysis, music generation, algorithmic music composition methods.

ВЛИЯНИЕ ЭМОЦИОНАЛЬНОЙ НАГРУЗКИ НОВОСТНОГО ТЕКСТА НА ИЗМЕНЕНИЕ ДОХОДНОСТИ АКЦИЙ

Ершов Михаил Олегович, Лозинская Агата Максимовна

Национальный исследовательский университет «Высшая школа экономики»
614070, Россия, г. Пермь, ул. Студенческая, д. 38,
mo.ershov@mail.ru, alozinskaia@hse.ru

В работе исследуется влияние эмоциональной тональности новостного текста о компании на изменение доходности ее акций. Для определения частоты слов в новостном тексте, имеющих определённую эмоциональную тональность, используются элементы машинного обучения, в частности, сентимент анализ (sentiment analysis), основанный на модели «мешка слов». Для решения задачи разделения новостей на классы «плохие» и «хорошие» по отношению к изменению доходности используется классификатор на основе метода опорных векторов (Support Vector Machine, SVM). Эмпирический анализ проводился на дневных данных о котировках акций компании Apple за период 01.01.2017 г. — 08.02.2019 г. Полученные результаты указывают на большую объясняющую способность построенного SVM классификатора в сравнении с наивным классификатором. Также авторами обнаружен статистически значимый обратный эффект частотности «негативных» слов в тексте новости на доходность акций компании, что свидетельствует о наличии краткосрочной неэффективности американского фондового рынка.

Ключевые слова: американский фондовый рынок, доходность акций, анализ тональности текста, модель «мешка слов», метод опорных векторов

Исследования, посвященные изучению эмоциональной нагрузки новостей, позволяют определить, какими поведенческими факторами руководствуются инвесторы при принятии инвестиционных решений. В данной работе для оценки эмоциональной нагрузки новостей применяется анализ тональности текста. Для определения тональности текста используется модель «мешка слов», основанная на словаре МакДональда и Лоугрэна (Loughran, McDonald, 2011). Данный словарь представляет собой крупный эмпирический сборник слов, которым присвоены эмоциональные тональности. Сначала с помощью данного словаря выделяется количество «положительных», «негативных» и «неопределенных» слов в тексте. Затем определяются частотность трех вышеуказанных типов слов в новостном тексте о компании, под которой понимается отношение числа слов с вышеуказанными эмоциональными тональностями к общему числу слов. В

случае публикации нескольких новостей в один день, вышеуказанные показатели усредняются.

Для исследования влияния тональности текста новости на будущее изменение доходности акции применяется линейная регрессионная модель. Для решения задачи классификации новостей используется классификатор, основанный на методе опорных векторов (Support Vector Machine, SVM). В качестве регрессоров в SVM модели и линейной регрессионной модели выступают показатели частотности слов, несущих «положительную», «негативную» или «неопределённую» эмоциональную нагрузку. Исследуемая зависимая переменная является бинарной и равна «1» — при положительном дневном изменении доходности акции, «0» — при отрицательном. Доходность акции измеряется как дневное процентное изменение ее рыночного курса, дивидендная доходность при этом не учитывается. Поскольку в исследуемой выборке не наблюдалось нулевое изменение доходностей акций, то это обеспечивает полное разделение выборки на два класса и применимость методов машинного обучения. При нахождении оптимальных параметров SVM модели (гиперпараметра ядра σ и размера штрафа за неверно классифицированный объект C) проводится десятикратная кросс-валидация. Наилучшая модель определяется на обучающей выборке по максимальному показателю точности (accuracy), т.е. доли правильно классифицированных объектов. SVM классификатор оценивается ядерной аппроксимирующей функцией Гаусса. Выборка разделена на обучающую и тестирующую часть. Модель обучается на 75% от всей выборки, а тестируется на оставшихся 25%. Такое соотношение тестирующей и обучающей выборки обеспечивает большие значения площади под ROC кривой (Area Under Curve, AUC) на тестирующей выборке в сравнении с альтернативными вариантами деления выборки.

Исследуемая выборка включает в себя ежедневные данные о ценах акций компании Apple, источником которых является база investing.com, и новостных публикациях о компании из информационной базы Thomson Reuters за период 01.01.2017 г. — 08.02.2019 г. Общее число новостных публикаций составляет 170, которые включают в себя следующие категории новостей: слияние и поглощения, выход компании на новые рынки, интервью или комментарии генеральных директоров или иных лиц, занимающих руководящие посты в компаниях, аналитические комментарии к годовым отчетам, и аналитические обзоры сделанные внутренними авторами Thomson Reuters, решения общего собрания акционеров компании, публикации отчетов рейтинговых агентств по изменению кредитного рейтинга долговых бумаг компании.

В линейной регрессионной модели был выявлен обратный статистический значимый эффект частоты «негативных» слов в новостном тексте на изменение доходности акции (на уровне значимости 10%). По результатам десятикратной кросс-валидации на тестирующей выборке для классификатора SVM значение гиперпараметра ядра (σ) составило 1.26, а размер штрафа за неверно классифицированный объект (C) 0.5.

Данный классификатор на тестирующей выборке показал площадь под ROC кривой равную 0.59, что немногим больше, чем показатель AUC у «наивного» классификатора, делящего выборку случайным образом. AUC для такого классификатора равен 0.5.

Несмотря на невысокую объясняющую способность, которую показал SVM классификатор, полученные результаты позволяют заключить, что инвесторы в краткосрочном периоде склонны принимать решения, основываясь на эмоциональной тональности новостных публикаций, что противоречит гипотезе об эффективности рынка, согласно которой вся публичная информация находится в цене и принятия инвестиционного решения на основе новостей неэффективны (Володин, Головченко, 2014). Малая объясняющая способность классификатора может быть объяснена тем, что американский фондовый рынок является одним из самых старых во всем мире. Согласно ряду эмпирических статей, фондовые рынки с возрастом становятся более эффективными. Это означает, что на таких рынках сложнее получать доходность выше среднерыночной (Chong, Cheng, 2010).

Библиографический список

1. Chong T. T. L., Cheng S. H. S., Wong E. N. Y. A comparison of stock market efficiency of the BRIC countries // Technology and Investment. 2010. Vol. 1. No 04. Pp. 235-238. DOI: 10.4236/ti.2010.14029.
2. Loughran T., McDonald B. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks // The Journal of Finance. 2011. Vol. 66. No. 1. Pp. 35-65. DOI: 10.1111/j.1540-6261.2010.01625.x.
3. Володин С. Н., Головченко А. Э. Применимость технического анализа в различных рыночных ситуациях // Аудит и финансовый анализ. 2014. № 5. С. 99-105.

THE INFLUENCE OF THE STOCK'S NEWS EMOTIONAL CONTENT ON STOCK RETURNS

Ershov Mikhail O.

Research University Higher School of Economics, 614070, Russia, Perm,
Studencheskaya str. 38, mo.ershov@mail.ru

Lozinskaia Agata M.

Higher School of Economics, 614070, Russia, Perm, Studencheskaya str. 38,
alozinskaia@hse.ru

The article examines the effect of the text emotional content relating to the company's news on the change in its stock returns. To determine the frequency of words in a news text with certain emotional content, elements of machine learning, in particular, sentiment analysis, based on the «bag of words» model are applied.

To divide news into «bad» and «good» classes in relation to the change in stock returns the Support Vector Machine (SVM) classifier is used. The empirical analysis is conducted on the daily Apple stock price data for the time period 01.01.2017 — 08.02.2019. The obtained results show that the SVM classifier has higher an explanatory power compared with the naive. The authors also have found a statistically significant inverse effect of the frequency of «negative» words in the text related to the company's news company's stock returns. This finding indicates the presence of a short-term inefficiency of the American stock market.

Keywords: American stock market, stock returns, sentiment analysis, support vector machine, «bag of words».

УДК 004.942

О ПРИМЕНЕНИИ КЛАСТЕРНОГО АНАЛИЗА ДЛЯ ОПРЕДЕЛЕНИЯ СЕМАНТИКИ СОБЫТИЙ В СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ СИСТЕМАХ

Заякин Виктор Сергеевич

Национальный исследовательский университет «Высшая школа экономики»,
Пермский филиал, 614070, Россия, г. Пермь, ул. Студенческая, 38,
vszayakin@yandex.ru

В работе описывается возможность применения методов Process Mining для анализа социально-экономических процессов с использованием текстов новостных сообщений в качестве данных о событиях. Для математического представления текстов применяется модель векторного пространства из теории информационного поиска. Обсуждается использование методов кластерного анализа для присвоения событиям типов на основе гипотезы распределения.

Ключевые слова: кластеризация, модель векторного пространства, гипотеза распределения, Process Mining, анализ процессов, анализ новостных сообщений.

Введение

Построение моделей социально-экономических процессов представляет большую ценность как для социологов, так и для аналитиков (экспертов) компаний. Помимо выявления теоретических закономерностей между различными событиями, модели могут быть использованы для прогнозирования поведения социально-экономических систем. Существуют немногие попытки разработки аналитических моделей [1] и использования

нейронных сетей [2] и аппарата цепей Маркова [3-4] для решения этой задачи.

С точки зрения системного анализа построение моделей осложняется тем, что структура социально-экономических систем, а также законы, которым подчиняются их процессы, априорно неизвестны. В связи с этим, изучение теоретических зависимостей процессов возможно только при использовании эмпирических данных. В данной работе предлагается построение моделей процессов методами Process Mining на основе текстов новостных сообщений.

Модель векторного пространства

Использование текстов новостных сообщений в качестве данных для построения моделей глобальных процессов, очевидно, обладает следующими преимуществами:

Доступность. Доступ к текстам осуществим с помощью существующих новостных агрегаторов (например, Yandex News и Google News).

Охват. Тексты отражают информацию о **событиях**, происходящих в различных социально-экономических системах, то есть возможен одновременный анализ большого числа процессов.

Возможность автоматизации. Сбор, лексическая обработка [5] и анализ текстов осуществим с помощью программных средств.

Модель векторного пространства [6] может быть использована для описания текстов новостных сообщений в виде векторов $x_i = (w_{i1}, w_{i2}, \dots, w_{im})$, координатами w_{ij} которых являются частоты слов, входящих в документ. Таким образом, каждому сообщению соответствует точка некоторого многомерного пространства $\mathbb{R}^m$, где $\mathbb{R}$ - поле вещественных чисел.

Данная модель получила достаточно широкое применение в работах отечественных авторов (в т. ч. при использовании с текстами новостных сообщений). Так, например, В. Порицкий и О. Волчек [7] в своей работе обучают модели векторного пространства на художественных и газетных текстах для решения задачи классификации пар синонимов, а А. Зевайкин и В. Корнеев [8] используют модель для оптимизации составления новостных сюжетов с помощью анализа потоков сообщений. Кроме этого, А. Сигов и др. [1] используют модель и тексты сообщений для прогнозирования событий в информационном пространстве.

Построение моделей процессов

Методы Process Mining возможно применять для моделирования процессов в виде ориентированных графов. Они используют для построения протокол работы системы, к которому предъявляются определенные требования [9, с. 391]. Применительно к анализу текстов новостных сообщений протоколом можно считать множество текстов сообщений, упорядоченных по времени (предполагается, что время публикации сообщений известно).

В соответствии с требованиями к протоколу каждое событие (новостное сообщение) должно определяться своим типом. Текст сообщения при этом в общем случае не содержит в явном виде информацию о типе события, поскольку тип – это неизвестная *метаинформация* о событии, запечатленном

в текстовом документе. Таким образом, определение типа события на основе его текстового описания представляет собой задачу, решение которой должно предшествовать построению моделей процессов методами Process Mining.

Применение кластерного анализа для определения типов событий с использованием текстов новостных сообщений

Введем в рассмотрение множество $E = \{e_j \mid j = 1, 2, \dots, l\}$, состоящее из l элементов, определяющих *типы событий*, которые могут соответствовать событиям в некоторой социально-экономической системе. В таком случае, сюръективное отображение $f_E: \mathbb{R}^m \rightarrow E$ будет являться возможным решением сформулированной задачи определения типов событий.

Определим множество всех отображений (функций, правил, алгоритмов) векторного пространства $\mathbb{R}^m$ на множество типов событий E как $\mathcal{F} = \{f_E \mid f_E: \mathbb{R}^m \rightarrow E\}$, при этом необязательно, чтобы множество E (или его кардинальность) было известно заранее. Все тексты сообщений с помощью модели векторного пространства опишем как векторы x_i , тогда $X = \{x_i\} \subset \mathbb{R}^m$ – множество известных наблюдений. Очевидно, что для того, чтобы отображение a было решением задачи *достаточно*, чтобы нашлась такая функция $f_E \in \mathcal{F}$, что $a \equiv f_E$ на $X \subset \mathbb{R}^m$ (оставим этот факт без строгого доказательства). Также легко видеть, что любой алгоритм кластеризации является решением поставленной задачи, поскольку во введенных обозначениях любой алгоритм кластеризации принадлежит множеству $\mathcal{F}$, в свою очередь любой элемент этого множества является решением задачи.

Стоит отметить, что несмотря на формальную возможность применения алгоритмов кластеризации для определения типов событий, состоятельность такой процедуры с точки зрения семантической схожести новостных сообщений в пределах одного кластера (и различия между разными кластерами) остается под вопросом. Однако согласно гипотезе распределения [10] статистические зависимости распределения слов могут быть использованы для оценки схожести смысла, заложенного в текст. При построении векторных пространств в качестве меры схожести обычно выбирается косинус угла между соответствующими векторами (текстами).

Заключение

В данной работе предложен подход к изучению социально-экономических процессов путем построения их графовых моделей методами Process Mining с использованием текстов новостных сообщений. Кроме этого, продемонстрирована возможность применения методов кластерного анализа для разбиения текстов на смысловые кластеры. Планируется исследование влияния выбора алгоритма и меры векторного пространства на качество кластеризации с точки зрения семантического различия текстов в разных кластерах.

Библиографический список

1. Сигов А. С., Жуков Д. О., Новикова О. А. Моделирование процессов реализации памяти и самоорганизации информации при прогнозировании новостных событий с использованием массивов

- естественно-языковых текстов // Современные информационные технологии и ИТ-образование. – 2016. – Т. 12, № 1. С. 42-55.
2. Богданова Т. К., Шевгунов Т. Я., Уварова О. М. Применение нейронных сетей для прогнозирования платёжеспособности российских предприятий обрабатывающих отраслей // Бизнес-информатика. – 2013. – № 2 (24). – С. 40-48.
 3. Huo D. Cluster Analysis of market Potential in Emerging Markets: A Dynamic Research Based on Markov Chain // Romanian Journal of Economics Forecasting. – 2013. Vol. 16, № 4. – P. 218-231.
 4. Netzer O., Lattin J. M., Srinivasan V. A Hidden Markov Model of Customer Relationship Dynamics // Marketing Science. – 2008. Vol. 27, № 2. – P. 185-204.
 5. Маннинг К., Рагхаван П., Шютце Х. Введение в информационный поиск: Пер. с англ. – М.: ООО «И.Д. Вильямс», 2011. – 528 с.
 6. Salton G., Wong A., Yang C. S. A vector Space Model for Automatic Indexing // Communications of the ACM. – 1975. – Vol. 18, № 11. – P. 613-620.
 7. Порицкий В. В., Волчек О. А. Построение векторной семантической модели на основе русскоязычных текстов: первые эксперименты // Тр. конф. «Информационные технологии и системы-2013» (Калининград, 1–6 сент. 2013 г.). Калининград, 2013.
 8. Зазевайкин А. Н., Корнеев В. В. Формирование выпуска новостей на основе автоматического анализа новостных сообщений // Сборник работ научных стипендиатов Яндекс «Интернет-математика 2005», Ярославль, 2005.
 9. Анализ данных и процессов: учеб. пособие / А. А. Барсегян, М. С. Куприянов, И. И. Холод и др. – СПб.: БХВ-Петербург, 2009. – 512 с.
 10. Sahlgren M. The Distributional Hypothesis // Italian Journal of Disability Studies. – 2008. – Vol. 20, № 1. – P. 33-53.

ON THE APPLICATION OF CLUSTER ANALYSIS FOR SEMANTICS SPECIFICATION OF EVENTS IN SOCIO-ECONOMIC SYSTEMS

Zayakin Viktor S.

National Research University Higher School of Economics,

38, Studencheskaya st., Perm, 614070, Russia,

vszayakin@yandex.ru

The paper describes the feasibility of the application of Process Mining techniques for socio-economic processes analysis using texts of news as event data. Information retrieval's vector space model is used for mathematical representation of texts. The use of cluster analysis techniques for assigning types to events based on the distributional hypothesis is described.

Keywords: clustering, vector space model, distributional hypothesis, Process Mining, process analysis, news analysis.

УДК 004.89; 616.34

АВТОМАТИЗИРОВАННЫЙ СИНТЕЗ МУЗЫКАЛЬНЫХ КОМПОЗИЦИЙ НА ОСНОВЕ РЕКУРРЕНТНЫХ НЕЙРОННЫХ СЕТЕЙ

Никитин Никита Андреевич, Розалиев Владимир Леонидович,

Орлова Юлия Александровна

Волгоградский государственный технический университет

400005, Россия, Волгоград, пр. им. Ленина, 28, set.enter@mail.ru

В данной работе описан метод для автоматизации процесса создания музыки, путём автоматизированной генерации звуков по изображению. Разработанный метод автоматической генерации звуков по изображению основывается на совместном использовании нейронных сетей и светомузыкальной теории, что позволяет повысить качество выходной музыкальной композиции и снизить роль пользователя. Описана программа для генерации звуков по изображению, для подтверждения эффективности предложенного метода. Также описано тестирование программы.

Ключевые слова: рекуррентная нейронная сеть, светомузыкальная теория, Keras, автоматизированная генерация музыки, схемы соотнесения цветов и нот, анализ изображений

С тех пор как музыку стали записывать на бумаге в виде нотных знаков, стали появляться оригинальные «способы» ее сочинения. Одним из самых первых методов алгоритмической композиции стал способ сочинения

музыки, придуманный Моцартом – «Музыкальная игра в кости» [1]. Первое компьютерное музыкальное произведение – «Illiac Suite for String Quartet» – было создано в 1956 году пионерами применения компьютеров в музыке – Лежарен Хиллер и Леонард Айзексон [2]. В этом произведении использованы почти все главные методы алгоритмической музыкальной композиции: теория вероятностей, марковские цепи и генеративная грамматика.

Наибольших успехов автоматизация процесса написания и создания музыки достигла сравнительно недавно (в последние десятилетия), однако по большей части связана с изучением и повторением различных музыкальных стилей [3]. Поскольку процесс создания музыки сложно формализуем, то для программного (автоматизированного) создания композиций лучше всего подходят искусственные нейронные сети, так как они позволяют выявить связи, которые не видит человек [4]. Целью данной работы является увеличение гармоничности и мелодичности программной генерации звуков по цветовой гамме изображений посредством использования нейронных сетей.

Для снижения роли пользователя-композитора в генерации звуков, часть характеристик музыкального произведения получается путём анализа цветовой гаммы изображения. Таким образом, характер полученной музыкальной композиции будет соответствовать входному изображению. Данная особенность делает возможным применение данного подхода для создания фоновых музыкальных произведений в компьютерных играх, рекламе и фильмах.

Ключевыми характеристиками музыкального произведения является его тональность и темп. Именно эти параметры определяются путём анализа цветовой гаммы изображения. Для начала определим соотношение цветовых и музыкальных характеристик [5]: оттенок цвета соотносится с высотой ноты, цветовая группа с музыкальным ладом, яркость с октавой ноты, насыщенность с длительностью.

Затем, необходимо определить схему соотнесения названия цвета и ноты. На данный момент существует большое количество подобных схем, однако в данной работе была выбрана схема Ньютона.

Алгоритм определения тональности опирается на анализ изображения и таблицу 1, состоит из 3 шагов и описан ниже.

Шаг 1. Преобразуем входное изображение из цветового пространства RGB в HSV. Данный шаг позволяет преобразовать изображение к более удобному виду, поскольку HSV пространство уже содержит необходимые характеристики – название цвета (определяется по параметру hue), насыщенность (параметр saturation) и яркость (параметр brightness).

Шаг 2. Анализируя в целом изображение, определяем преимущественный цвет.

Шаг 3. Определяем название и цветовую группу преимущественного цвета.

Шаг 4. Согласно таблице 1 и схеме Ньютона определяем тональность произведения (нота и музыкальный лад).

Для определения темпа произведения, необходимо получить яркость и насыщенность (по параметрам saturation и brightness) преимущественного цвета, и рассчитать темп, согласно данным параметрам.

В данной работе предполагается следующий алгоритм получения композиции по изображению:

- 1) согласно методу соотнесения цветовых и музыкальных характеристик получаем тональность произведения и последовательность первых 20% нот, считанных с изображения;

- 2) далее по полученной последовательности нот предсказываем продолжение произведения с помощью обученной модели и нейронной сети;

- 3) по итоговой последовательности нот и тональности, согласно методу соотнесения цветовых и музыкальных характеристик, строим гармоническую часть произведения.

Для обработки любой условно бесконечной последовательности, в которой важно не только содержание, но и порядок, в котором следует информация (например, текст или музыка) необходимо использовать нейронные сети с обратными связями – рекуррентные нейронные сети (RNN). В рекуррентных нейросетях нейроны обмениваются информацией между собой: например, вдобавок к новому «кусочку» входящих данных нейрон также получает некоторую информацию о предыдущем состоянии сети. Таким образом, в сети реализуется «память», что принципиально меняет характер ее работы и позволяет анализировать любые последовательности данных, в которых важно, в каком порядке идут значения [6].

Однако большой сложностью сетей RNN является проблема исчезающего (или взрывного) градиента – она заключается в быстрой потере информации с течением времени. Конечно, это влияет лишь на веса для нейронов, а не на их состояния. Однако ведь именно в весах нейронов накапливается информация. Сети с долгой краткосрочной памятью (long short term memory, LSTM) стараются решить вышеупомянутую проблему потери информации, используя фильтры и явно заданную клетку памяти. У каждого нейрона есть клетка памяти и три фильтра: входной, выходной и забывающий. Целью этих фильтров является защита информации. Входной фильтр определяет, сколько

информации из предыдущего слоя будет храниться в клетке. Выходной фильтр определяет, сколько информации получают следующие слои. Такие сети способны научиться создавать сложные структуры, например, сочинять тексты в стиле определённого автора или сочинять простую музыку. Однако при этом потребляется большое количество ресурсов [7].

Таким образом, для реализации программы автоматизированной генерации музыкальных композиций необходимо использовать именно рекуррентные нейронные сети с долгой краткосрочной памятью – RNN LSTM (долгая краткосрочная память – разновидность архитектуры рекуррентных нейронных сетей). Именно данный вид нейронных сетей используется для генерации музыкальных композиций в программе Magenta – это музыкальный проект с открытым исходным кодом от Google; RNN LSTM используется в программе сочинения композиций в стиле И.С. Баха – BachBot; DeepJaz – система позволяет генерировать джазовые композиции на основе анализа midi файлов [8].

Для оценки композиций были использованы специалисты-эксперты, которые оценивали композиции по следующим критериям:

- соответствие характеру изображения (по пяти бальной шкале);
- реалистичность звучания инструмента (фортепьяно или гитара);
- мелодичность композиции;
- качество гармонии (аккомпанемента);
- приятность мелодии для восприятия;
- цельность композиции;
- реалистичность/искусственность композиции.

Проанализировав оценки всех экспертов и высчитав средние по каждому критерию, можно сделать вывод о том, что фортепьяно на слух экспертов звучит реалистичнее, чем гитара. Также можно сделать вывод о том, что композиция, сгенерированная по абстрактным изображениям, более приятна на слух, чем генерация по пейзажам. В целом общее впечатление от сгенерированных звуков у экспертов положительное. Среди минусов некоторые эксперты выделяют однотипность гармонии, иногда рваность и недостаточную реалистичность произведения, и не достаточную реалистичность гитары.

Делая вывод по каждому критерию, можно сказать, что все эксперты оценили на высокий бал соответствие произведения характеру изображения, по второму критерию – инструмент фортепьяно звучит довольно реалистично. Мелодичность композиций разделилась пополам, то есть половина композиций эксперты оценили на высший бал, другую половину на 4, в целом неплохой результат. Качество гармонии также было оценено экспертами на высший бал. Приятность мелодий для восприятия получил 60% высших баллов и 40% четвёрок, что говорит о том, что некоторые произведения звучат не вполне реалистично. Реалистичность и цельность композиций в среднем оценено на 4, что является естественным результатом для компьютерной генерации звуков.

Работа выполнена при частичной поддержке РФФИ (гранты 17-07-01601, 18-47-342002, 18-07-00220, 19-07-00020).

Библиографический список

1. Фазылова, Э.Ф. Системы генерации музыки или как автоматизировать искусство? // Молодёжный научно-технический вестник. – 2014. URL: <http://sntbul.bmstu.ru/doc/723360.html>. (Дата обращения: 20.03.2019).
2. Ariza, C. Two Pioneering Projects from the Early History of Computer-Aided Algorithmic Composition / C. Ariza // Computer Music Journal. – MIT Press, 2012. – №3. – pp. 40-56
3. D. Cope, Computer Models of Musical Creativity, MIT Press, Cambridge, Mass., 2005.
4. Mazurowski, L. Computer models for algorithmic music composition / L. Mazurowski // Proceedings of the Feder-ated Conference on Computer Science and Information Systems. – Szczecin, Poland, 2012. – pp. 733–737.
5. Caivano, J. L., Colour and sound: Physical and Psychophysical Relations, Colour Research and Application, 12(2), pp. 126-132, 1994.
6. Sak, H., Senior, A., Beaufays, F. Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition. ArXiv e-prints, 2014.
7. Doornbusch, P. Gerhard Nierhaus: Algorithmic Composition: Paradigms of Automated Music Generation Computer Music Journal, Volume: 34, Issue: 3, 2014.
8. Brinkkemper, F. Analyzing Six Deep Learning Tools for Music Generation (2015). Available at: <http://www.asimovinstitute.org/analyzing-deep-learning-tools-music/>

AUTOMATED SYNTHESIS OF MUSICAL COMPOSITIONS BASED ON RECURRENT NEURAL NETWORKS

Nikitin, Nikita A., Rozaliev, Vladimir L., Orlova Yulia A.
Volgograd state technical university, 400005, Russia, Volgograd,
28 Lenin Ave.,
set.enter@mail.ru

This paper describes a method for automating the process of creating music by means of automated generation of sounds from an image. The developed method of automatically generating sounds by image is based on the sharing of neural networks and light-music theory, which allows improving the quality of the output musical composition and reducing the role of the user. A program is

described for generating sounds by image to confirm the effectiveness of the proposed method. Testing of the program was also described.

Key words: recurrent neural network, color music theory, Keras, automated music generation, color and note matching schemes, image analysis.

УДК 004.89; 519.237.8

МЕТОД ВЕКТОРНОГО ПРЕДСТАВЛЕНИЯ ТЕКСТА НА ОСНОВЕ ЭНТРОПИИ, ДЛЯ ЗАДАЧИ РАСПОЗНАВАНИЯ ИНТЕНТА В ДИАЛОГОВОЙ СИСТЕМЕ¹⁶

Перевалов Александр Максимович, Курушин Даниил Сергеевич

Пермский национальный исследовательский политехнический университет,
614990, Россия, г. Пермь, Комсомольский пр., 29,
perevalovproduction@gmail.com, dan973@yandex.ru

В данной работе предложен подход векторного представления текстовых вопросов на основе вычисления энтропии Шеннона для каждого из слов вопроса. Целью данного подхода является максимизировать критерий качества классификации при условии малого количества обучающих данных. Оценка предложенного проводится в сравнении с существующими методами: Bag of Words, TF-IDF, word2vec, FastText на текстовом корпусе, состоящем из вопросов абитуриентов и студентов ПНИПУ, размеченном по 6-ти классам. Критерием качества классификации выступает F-мера.

Ключевые слова: классификация текста, векторное представление слов, энтропия Шеннона, диалоговые системы, обработка естественного языка

Разработка доменно-специфичной диалоговой системы включает в себя решение ряда задач по пониманию естественного языка. Одной из этих задач является классификация интента пользователя, которая в большинстве случаев решается методами машинного обучения. Очевидно, что методы машинного обучения, являющиеся математическими моделями, и не могут работать с текстовыми данными, поэтому необходимо задать некоторое отображение единицы текста (слово, предложение) к векторному виду.

Большинство современных методов векторного представления текста основаны на дистрибутивной гипотезе, которая гласит: «Слова, встречающиеся в схожих контекстах, имеют схожий семантический смысл» (*Words that occur in the same contexts tend to have similar meanings* [1]). Такие методы векторного представления слов, как vector space model [2], word2vec [3], FastText [4] являются формализацией дистрибутивной гипотезы.

Все рассмотренные методы используют нейронные сети, обучающиеся с критерием максимального правдоподобия между текущим и следующим словом в контексте. Нейронные методы иногда называют data-hungry или методами, голодными до данных, это происходит потому, что на практике, исследователи часто встречаются с недостатком обучающих данных, и, следовательно, такие методы оказываются менее эффективными по сравнению с классическими. Использование нейронных методов с небольшой обучающей выборкой для задачи классификации интенгов также является нетривиальным и требует модернизации методов для достижения оптимального решения [5].

Резюмируя вышесказанное, целью данной работы является разработка простого, не требующего много обучающих данных метода векторного представления текстовых вопросов. Метод должен генерировать векторы с фиксированной размерностью, задающейся в виде параметра. Критерием успешности метода будет являться его превосходство по определенной метрике над уже имеющимися аналогами в условиях идентичности обучающих данных и модели классификатора.

Сбор данных производился путем парсинга открытых источников в интернете, а именно сайты: pstu.ru, vk.com/politehperm, abiturient.ru. В результате, было собрано порядка 7300 вопросов. На основании анализа собранных данных, была разработана таксономия вопросов, состоящая из 6-ти классов:

- DOC – вопросы, касающиеся документов;
- ENTER – вопросы, касающиеся процесса поступления;
- ORG – организационные вопросы;
- PRIV – вопросы о привилегиях при поступлении;
- RANG – вопросы, связанные с проходными баллами и списками зачисленных;
- HOST – вопросы о студенческих общежитиях.

Пример данных представлен в таблице 1.

Таблица 1. Фрагмент размеченных данных

Вопрос	Класс
Можно ли подать документы в субботу?	DOC
Когда день открытых дверей?	ORG
Когда публикуются списки зачисленных?	RANG
Какие документы нужны для заселения в общежитие?	HOST

Разметка данных проводилась с помощью инструмента ipyannotate вручную. В результате было размечено 1300 вопросов (каждый вопрос был отнесен к одному из классов таксономии).

В данной работе предлагается метод векторного представления документов на основе расчёта энтропии Шеннона [6] для каждого слова в

документах корпуса. Одна из интерпретаций энтропии Шеннона - мера информации. В этом смысле, мы измеряем количество информации в каждом слове текстового документа.

Опишем алгоритм векторного представления документов: прежде всего составим словарь слов, находящихся в обучающих данных. После чего, рассчитываем энтропию каждого слова в словаре для всех документов по формуле (1)

$$e_{ij} = \begin{cases} -p_{ij} \log_2(p_{ij}), p_{ij} = w_{ij} / n_{ij}, w_{ij} > 0 \\ -0.0001, w_{ij} = 0 \end{cases} \quad (1)$$

Говоря терминами машинного обучения, мы рассчитали матрицу, в которой строки соответствуют документам, а столбцы - словам. Полученная матрица содержит 1300 строк (документов) и 1212 столбцов (признаков или уникальных слов в тренировочных данных). Однако, нам необходимо получить векторы слов и из них составлять векторы документов. Для этого транспонируем полученную матрицу и теперь строки соответствуют словам, а столбцы документам (2). Таким образом, каждое слово определяется распределением энтропии на документах в текстовом корпусе.

$$M^T = \begin{vmatrix} e_{11} & e_{21} & \dots & e_{m1} \\ e_{11} & e_{11} & \dots & e_{m1} \\ \dots & \dots & \dots & \dots \\ e_{1n} & e_{1n} & \dots & e_{m,n} \end{vmatrix} \quad (2)$$

Полученная транспонированная матрица является разреженной - каждая вектор-строка имеет размерность 1300 компонент, большинство из которых равны нулю. Для того, чтобы снизить размерность матрицы применим метод усеченного сингулярного разложения, количество компонент, которое оставим после усечения будет равно 200.

Для того, чтобы из векторов слов получить вектор документа, выберем только те вектора слов, которые встречаются в данном документе, а затем усредним их (3). Заново составленная матрица документов в векторном виде и является набором данных для проведения экспериментов в задаче классификации интента.

$$Q_i = \frac{\sum w_j}{count(W)}, w_j \in W \quad (3)$$

Для экспериментального тестирования предлагаемого подхода используется линейный классификатор (логистическая регрессия). Поскольку набор данных разделен на несколько классов (больше 2-х), используется метод классификации one-vs-rest.

Предлагаемый подход к векторному представлению документов экспериментально сравнивается с TF-IDF, word2vec и FastText. Векторы слов документа, полученные с помощью word2vec и FastText усредняются для получения вектора документа. В ходе экспериментов, набор данных был разделен на 5 частей для перекрестной валидации. В качестве оценочной метрики используется F1-мера (4).

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Результаты экспериментов на наборе данных с вопросами абитуриентов представлены в таблице 2.

Таблица 2. Результаты экспериментов

	F1-мера			
	TF-IDF	Word2vec	FastText	Энтропия
DOC	0.77	0.76	0.66	0.76
ENTER	0.62	0.59	0.58	0.65
ORG	0.73	0.66	0.65	0.74
PRIV	0.24	0.25	0.31	0.32
RANG	0.74	0.65	0.65	0.75
HOST	0.91	0.85	0.75	0.91
Average	0.72	0.67	0.63	0.74

В данной работе был предложен метод векторного представления документов на основе энтропии Шеннона. В ходе работы был собран набор данных, состоящий из 7300 вопросов студентов и абитуриентов. Набор данных был размечен на 6 классов, каждый из которых относит вопрос к той или иной тематике. Предложенный метод был реализован программно и протестирован в сравнении с существующими методами векторного представления текста.

По результатам экспериментов, на собранном наборе данных предложенный подход показал наилучший результат по F-мере, опередив TF-IDF на 2%.

В рамках последующей работы планируется редизайн имеющейся таксономии вопросов для сбалансированности выборки, а также ряд улучшений по взвешиванию слов при получении вектора документа. Полученные в ходе работы данные и модели используются в диалоговой системе для консультации студентов и абитуриентов ПНИПУ и размещены по адресу https://github.com/Perevalov/intent_classifier.

Библиографический список

1. Harris, Zellig, "Distributional structure," In: Word, S. , no. 23, pp.146-162, 1954.
2. Turney, Peter D, Pantel, Patrick and others, "From frequency to meaning: Vector space models of semantics," Journal of artificial intelligence research 37 , no. 1, pp. 141-188, 2010.
3. Mikolov, Tomas, Chen, Kai, Corrado, Greg and Dean, Jeffrey. "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, 2013.
4. Joulin, Armand, Grave, Edouard, Bojanowski, Piotr, Douze, Matthijs, Jégou, Hervé and Mikolov, Tomas, "FastText.zip: Compressing text classification models," arXiv preprint arXiv:1612.03651, 2016.
5. K. Shridhar, A. Dash, A. Sahu, G. Grund Pihlgren, P. Alonso, V. Pondenkandath, G. Kovacs, F. Simistira, M. Liwicki, "Subword Semantic Hashing for Intent Classification on Small Datasets," arXiv preprint arXiv:1810.07150, 2018.
6. Vajapeyam, Sriram, "Understanding Shannon's Entropy metric for Information," arXiv preprint arXiv:1405.2061, 2014.

METHOD OF WORD-TO-VECTOR REPRESENTATION BASED ON ENTROPY FOR INTENT CLASSIFICATION TASK IN DIALOGUE SYSTEM.

Perevalov Aleksandr M., Kurushin Daniil S.
Perm National Research Polytechnic University
614990, Russia, Perm, Komsomolsky pr., 29
perevalovproduction@gmail.com, dan973@yandex.ru

In this paper, we propose an approach of the questions embeddings representation based on calculation of Shannon entropy. The goal of the approach is to produce low dimensional question vectors as neural approaches do and to outperform related methods, described above in condition of small dataset. We evaluate and compare our model with existing ones using logistic regression and dataset that contains questions asked by students and enrollees. The data is labeled into six classes. Experimental comparison of proposed approach and other models revealed that proposed model performed better in the given task.

Key words: Text Classification, Word Embeddings, Shannon Entropy, Intent Classification, Natural Language Processing, Dialogue Systems, Word2vec, FastText.

ИДЕИ МЕТОДОВ ПРЕДСТАВЛЕНИЯ И ПОИСКА ОБЩИХ ПАТТЕРНОВ В ТРЕКАХ ВЗГЛЯДА, ПРИ ПРОСМОТРЕ ЛЮДЬМИ ОДНОГО И ТОГО ЖЕ ОБЪЕКТА

*Поликарпова Анна Игоревна, Розалиев Владимир Леонидович,
Орлова Юлия Александровна*

Волгоградский государственный технический университет,
400005, Россия, г. Волгоград, пр. им. Ленина, 28, vladimir.rozaliev@gmail.com

В статье описана гипотеза о том, что все люди, смотря на один и тот же предмет, руководствуются одним и тем же визуальным поведением. Основываясь на этой гипотезе в статье представлены некоторые идеи о том, как можно представлять общие паттерны и как их выявлять из треков взгляда. Так как основой таких подходов является выделение общих подстрок, то основной проблемой в этой работе является так называемая проблема наибольшей общей подстроки, которая пока не имеет четкого и оптимизированного решения для множества последовательностей.

Ключевые слова: айтрекер, трек взгляда, паттерн, визуальное поведение, разностный алгоритм Майреса.

Зрение – важная человеческая способность, благодаря которой мы получаем более 80% всей информации об окружающем нас мире. Существует гипотеза Джаста и Карпентера [1], которая говорит о взаимосвязи визуальной системы человека и его сознания, наблюдение и анализ визуального поведения позволит получить представление о когнитивных процессах, в общем смысле, лежащих в основе человеческого поведения. И, в частности, происходящих во время визуального сканирования человеком тех или иных объектов. Это говорит о том, что наше зрение имеет очень большое влияние на то, как мы себя ведем.

Главная идея всей работы – люди одинакового смотрят на одни и те же предметы. И говоря «одинакового», я не имею ввиду, что траектория движения нашего зрачка совпадает у всех людей. Под этим я подразумеваю, что, сканируя один и тот же объект, люди неосознанно придерживаются одного и того же алгоритма сканирования. Иначе говоря, паттерна. Это называется визуальное поведение [2].

Обычно все задачи, связанные с визуальным поведением человека, направлены на изучение того, как люди сидят в интернете и на улучшение веб-страниц, чтобы пользователям было наиболее комфортно работать с ними. И для этого на изображении выделяют так называемые области интереса (AOI), обычно обозначая их буквами. И сам эксперимент довольно прост – при помощи айтрекера считывается трек взгляда испытуемого, а

затем координаты точек траектории движения зрачка соотносятся с АОІ, в которые они попали и получается строка, представляющая собой трек взгляда [3].

В нашем случае задача несколько усложняется тем, что интересующий нас объект на изображении и есть область интереса. Мы не можем более детально выделить АОІ на объекте, так как мы точно не знаем, что именно интересует взгляд человека, а что нет. Всё что у нас есть это область на изображении и треки взгляда людей, смотрящих на него. И нашей задачей является выявление общего паттерна, которому «подчиняются» разные люди, смотря на один и тот же объект. По мере осмысления этой задачи я задалась вопросом «А что именно можно считать общим паттерном в треках?». То, как этот паттерн будет представляться значительно влияет на метод его поиска среди множества треков.

В данной статье я представляю и пишу о треках и паттернах как о строках, подразумевая, что координаты траектории можно соотнести с определенными символами, несмотря на то, что было сказано выше. Это возможно если на всю область интереса растянуть равномерную сетку и каждой секции назначить свой символ. При этом стоит сливать несколько близко друг к другу расположенных и последовательно идущих друг за другом точек, если они при этом все находятся в одной секции, тем самым сокращая длину трека. Далее о треках будет говориться как о строках, т.е. последовательностях символов, которыми можно представить сами треки.

Наиболее распространённый способ представления общего паттерна звучит следующим образом: общим паттерном для объекта можно считать, как одну последовательность набора символов, повторяющаяся во всех треках взгляда, она же LCS (Longest Common Subsequence). Идея такого представления хорошо описана в работе [4], как и алгоритм его выявления. Но мой алгоритм выделения общего паттерна несколько иной, хоть и имеет с ним сходство.

Ниже представлен псевдокод алгоритма выявления общего паттерна, представленного как LCS, из множества треков при помощи разностного алгоритма Майерса [5]:

Входные данные: лист треков взгляда.

Выходные данные: общий паттерн.

Алгоритм:

1. пропустить первые две строки из множества через разностный алгоритм Майерса;
2. среди полученных разностных последовательностей найти самую короткую и выделить из неё последовательность повторяющихся символов (те, которые не были ни добавлены, ни удалены). Полученная последовательность будет так называемым очередным паттерном;
3. для каждой строки, кроме первых двух:

- 3.1.получить последовательность повторяющихся символов для очередного паттерна и очередной строки из начального множества при помощи алгоритма Майерса;
- 3.2.если количество символов в последовательности не нулевая и не больше и не сильно меньше очередного паттерна:
 - 3.2.1. заменить очередной паттерн на полученную последовательность;
4. вернуть общий паттерн.

При использовании такого алгоритма могут возникать некоторые проблемы. Одна из них: в очередной строке найдено слишком маленькое по сравнению с очередным паттерном количество повторяющихся символом. В таком случае можно поступить по-разному. Либо никак не реагировать на такую ситуацию и тогда на выходе мы можем получить паттерн из трех символов, двух, одного или даже нулевую строку. Любое сильное расхождение в каком-либо треке взгляда может привести к тому, что паттерн не будет выявлен. Поэтому такие строки будут пропускаться, произведя проверку перед тем, как обновить очередной паттерн.

Так же хочу сказать пару слов о моем выборе алгоритма для поиска LCS – разностном алгоритме Майерса, а не, например, алгоритма Ханта и МакИлроя [6], который делает практически тоже самое. Главным критерием выбора стала занимаемая память. Разностный алгоритм Майерса выигрывает у алгоритма Ханта и МакИлроя, потому что он не запоминает пути во время работы, и ему не нужно «предвидеть», куда идти, и при этом он может сконцентрироваться в любое время только на самых дальних позициях, которые он может достичь с помощью разностной последовательности наименьшей сложности. Эта «наименьшая сложность» гарантирует, что найденная строка является LCS.

Другой идеей представления общего паттерна является представления его не как целой последовательности символов, как это обычно привыкли делать, а как множество характерно повторяющихся последовательностей во всех треках определенного объекта.

Множество последовательностей, общих для всех треков, относящихся к определенному объекту можно считать паттерном. Эта идея имеет место быть так как человек может применять общий паттерн частично и в разных промежутках времени. Как если бы одну большую последовательность разбили и собрали заново, но не в том порядке.

Это идея может быть применима к случаям с цветным изображением, где помимо тона, есть ещё и цвет. Представьте, что идёте по улице по своим делам, а где-то на другой стороне дороги стоит обычная ничем не примечательная серая машина. Вы, может быть даже бы, и не заметили её и пошли бы мимо, но если вдруг на ней будет, допустим, ярко-красное пятно, то вы сразу её заметите. Но сначала, конечно же, вы будете рассматривать пятно, а уже потом взгляд может зацепиться и за машину. Так различные акценты на предметах могут «перебивать» наше визуальное поведение, как

бы разламывая наш привычный паттерн, после чего мы сами попытаемся собрать его воедино.

Допустим, что для нашего объекта были собраны следующие треки:

aabcdjihdeab
abcddejiheab
abcdhjidddeab

И для этого набора строк общими подстроками будут:

- | | | |
|---------|--------|-------|
| 1. ab | 4. de | 7. ji |
| 2. abc | 5. ea | |
| 3. abcd | 6. eab | |

Заметьте, что при этом не учитывается ни порядок вхождение одних и тех же общих подстрок, ни их пересечение, ни единичные символы.

Конечно, на практике таких треков будет не три, а много больше. В этот момент мы плавно утыкаемся к так называемую проблему наибольшей общей последовательности. Проблема эта известна всем и её решение имеет множество различных реализаций. Но, стоит заметить, что почти все алгоритмы на поиск общих подстрок в строках, сделаны для двух исходных строк и многие из них предполагают наличие искомой общей подстроки. У нас же условия задачи таковы, что по сути мы не знаем, что именно мы ищем в наших строках, а также наш начальный набор строк состоит как минимум из трех. По мимо этого стоит учитывать, что нам так же не известна длина строк и их максимальное количество, а значит искомый алгоритм должен быть хорошо оптимизирован для больших наборов данных.

Работа выполнена при частичной поддержке РФФИ (гранты 18-07-00220, 18-47-342002, 19-07-00020).

Библиографический список

1. Heiner Deubel, Werner X. Schneider. Saccade target selection and object recognition: Evidence for a common attentional mechanism; Vision Research Volume 36, Issue 12, June 1996, Pages 1827-1837.
2. Michael Argyle, Mark Cook. Gaze and Mutual Gaze, Cambridge University Press, New York, 1976, xii, 210 pp.
3. Haruhiko Takeuchi and Yoshiko Habuchi. A quantitative method for analyzing scan path data obtained by eye tracker. In CIDM '07, pages 283 – 286, 1 2007-april 5 2007.
4. Yeliz Yesilada, Sukru Eraslan. Patterns in eyetracking scanpaths and the affecting factors.
5. Eugene W. Myers. AnO(ND)DifferenceAlgorithmandItsVariations, Department of Computer Science, University of Arizona, Tucson, AZ85721, USA.
6. J. W. Hunt, M. D. McIlroy, An Algorithm for Differential File Comparison.

IDEAS OF THE METHODS OF SUBMISSION AND SEARCH OF COMMON PATTERNS IN GAZE PLOT FOR SAME OBJECT

Polikarpova Anna I., Rozaliev Vladimir L.,
Volgograd State Technical University,
400005, Russia, Volgograd, prosp. Lenina, 28, vladimir.rozaliev@gmail.com

The article describes the hypothesis that all people, looking at the same subject, are guided by the same visual behavior. This article presents some ideas on how you can present common patterns and how to identify them from gaze plot. Basis of such approaches is the allocation of common substrings and main problem in this article is the so-called largest common subsequence problem, which does not yet have a clear and optimized solution for a set of sequences.

Key words: eyetracker, gaze plot, pattern, visual behavior, Myers diff algorithm.

УДК 004.8

ПРИМЕРЫ РАЗРАБОТКИ ЧАТ БОТОВ

Тен Анастасия Евгеньевна, Кузьмин Александр Юрьевич
Волгоградский государственный технический университет, 400005, Россия,
г. Волгоград, проспект Ленина, д. 28
ten_anastasia@mail.ru, sasha-kuz77@yandex.ru

В статье описаны, актуальность чат-ботов на сегодняшний день, представлен примеры разработки чат-бота, интегрируемого в платформу Telegram.

Ключевые слова: чат-бот, интернет магазин.

В настоящее время чат-бот является инструментом коммуникаций, который широко используется в повседневной жизни человека. Пользователь ведет беседу с виртуальным собеседником или программой-собеседником с развлекательной или познавательной целью. Например, чат-боты могут выполнять функцию личного помощника, который записывает планы и составляет расписание; бизнес-партнера, который поможет сделать заказ в кулинарии; собеседника, который расскажет о погоде за окном.

Преимущества использования чат-ботов заключаются в том, что он способен откликаться в кратчайшие сроки 24 часа в сутки, экономия трудовых ресурсов и автоматизация бизнес-процессов, дружелюбное общение. Чат-боты быстро совершенствуются и в некоторых областях способны полностью заменить человека [1].

К недостаткам можно отнести несовершенный интерфейс: люди по-разному реагируют и воспринимают информацию.

Первый виртуальный собеседник Элиза был создан в 1966 году американским учёным, специалистом в области искусственного интеллекта, Джозефом Вейценбаумом [2].

Элиза пародирует диалог с психотерапевтом, реализуя технику активного слушания. Программа по большей части просто перефразировала высказывания пользователя. Например: «У меня болит голова» — «Почему вы говорите, что у вас болит голова?»; «Мой отец меня ненавидит» — «Кто ещё из семьи вас ненавидит?» [3].

Элиза выделяет во входной фразе значимые слова, которые подставляет в шаблонную фразу. Программа интерпретирует слово как отдельные символы, в отличие от человека.

На сегодняшний день существует несколько видов чат-ботов:

Консультанты — это, например, менеджер в интернет-магазине. Их обязанности — это общаться с потенциальным и реальным клиентом, отвечая на всевозможные вопросы, связанные с работой компании.

Помощники — можно встретить в мессенджерах и поисковиках. Помогают при поиске информации, также проводят ее первичный анализ.

Развлечения — часто виртуальный собеседник, отвечающий на вопросы. Программы обучаются на оставленных сообщениях, и со временем могут заменить реального собеседника.

Бизнес чат-боты — задача этих программ оптимизировать работу, сделав ее более эффективной. Бизнес чат-боты — задача этих программ оптимизировать работу, сделав ее более эффективной.

Функция чат-ботов лежит в плоскости оперативных услуг, когда надо быстро и в любой момент ответить на вопрос или выполнить анализ ситуации. Позволяет экономить на веб-персонале, делая работу автоматизированной [4].

Для разработки чат-ботов существует множество платформ, таких как ВКонтакте, WhatsApp, Telegram, Viber, Facebook, Twitter и другие.

Для реализации рассмотренных ниже примеров использовался язык программирования Python, платформа Telegram, так как:

Telegram имеет возможность простого создания и администрирования бота, удобное клиентское приложение, возможность создания интерактивных кнопок в чате, готовые библиотеки для реализации, популярность платформы. Именно Telegram в 2016-2017 годах начал активное использование чат-ботов.

Python имеет большое количество документаций и библиотек для реализации чат-ботов.

Рассмотрим пример бизнес чат-бота: чат-бот интернет магазин печенья на платформе Telegram.

Для оптимизации времени клиента и производителя существует возможность производить заказ онлайн через чат-бот на платформе Telegram.

Программа предназначена для работы с каталогом продукции, формирования и оформления заказа в интернет-магазине.

Описанная ниже реализация чат-бота интернет-магазина будет рационально использоваться для магазинов с небольшим ассортиментом. В таком случае каталог выглядит наиболее эстетично и удобно в применении.

Работа с программой осуществляется в несколько этапов:

Чат-бот интернет магазин реагирует на определенные команды в виде текста. Например, для того, чтобы инициировать беседу нужно нажать «Старт» или вручную ввести «/start», что изображено на рисунке 1.

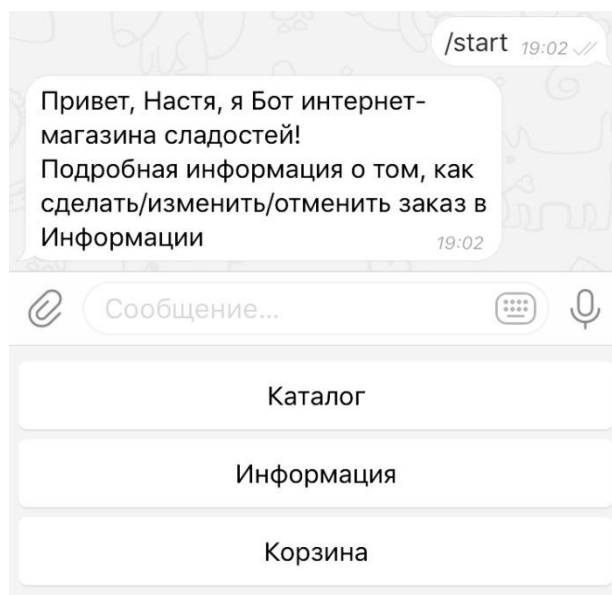


Рисунок 1. Инициирование беседы пользователем

Для заказа необходимо перейти в каталог, выбрать нужные товары, добавить некоторое количество в корзину и перейти к оформлению заказа, что изображено на рисунках 2 и 3. При добавлении товара в корзину программа чат-бот присваивает данному пользователю идентификационный номер, под которым в дальнейшем будут записаны все выбранные им товары. При просмотре корзины из базы читается информация о пользователе при помощи присвоенного ранее идентификационного номера.



Рисунок 2. Просмотр конкретного товара из каталога

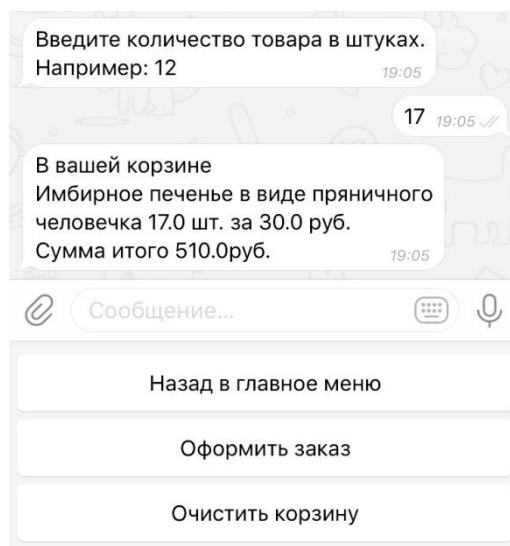


Рисунок 3. Ввод некоторого количество конкретного товара из каталога

При оформлении заказа пользователем требуются его контактные данные для связи, что изображено на рисунке 4.

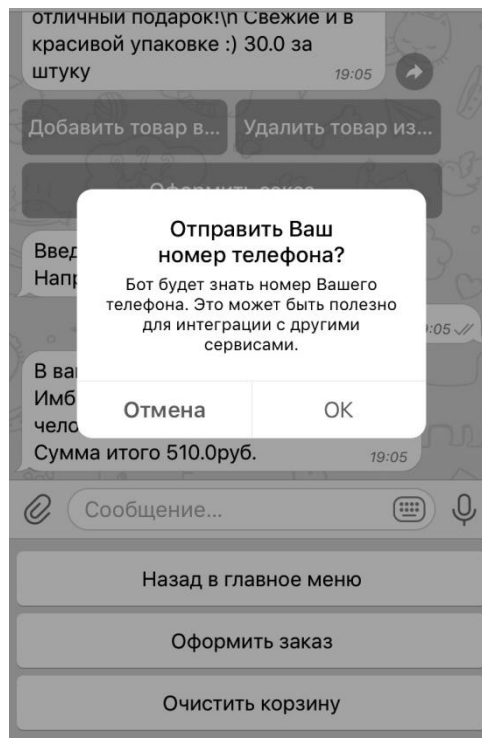


Рисунок 4. Запрос на работу с контактными данными

Пользователь должен выбрать способ связи с администратором: позвонить или написать в Telegram. Только после этого действия будут отправлены данные о заказе администратору, что изображено на рисунке 5. Выделенное сообщение может просматриваться только администратором.

Далее, после обработки заказа, администратор свяжется с пользователем, выбравшим способ с помощью личных данных, таких как телефон, имя и другие.



Рисунок 5. Оформление заказа

Данные линейные действия можно производить в любом порядке.

В результате работы был создан чат-бот, выполняющий функции интернет-магазина, имеющий возможность просматривать каталог, формирование и оформление заказа, работающий на платформе Telegram.

В дальнейшем можно увеличить функционал, интегрировать с другими сервисами, перенести чат-бота в другие мессенджеры и социальные сети.

Рассмотрим пример помощника чат-бота: чат-бот планировщик для составления расписания на платформе Telegram.

Проблема в необходимости ведения расписания и учета времени решается с помощью такого инструмента, как чат-бота.

Основными функциями чат-бота планировщика являются: создание заметок, записей о событиях, напоминаний, автоматизированное формирование и заполнение личного расписания на период времени, редактирование созданного расписания, оповещение при наступлении времени напоминания, возможность голосового ввода по средствам отправки голосовых сообщений [5].

Для взаимодействия с ботом, пользователь должен отправить сообщение в чат. Мессенджер выступает в роли интерфейса, посредника между пользователем и приложением. Так как основной акцент сделан на удобстве использования чат-бота, была предусмотрена функция обработки голосовых сообщений, т.е. помимо текстовых сообщений пользователь может озвучить свой запрос голосом. Для распознавания голоса и перевода речи в текст был применён сторонний сервис Wit.ai [6], который предоставляет возможность загрузить звуковой файл с речью и получить ответ в виде текста.

Следующим этапом обработки сообщений пользователя является выделение ключевой информации о событии из текста сообщения. Есть два подхода к обработке текста: заранее заготовленные команды и шаблоны входящих сообщений и обработка естественного языка с использованием технологий машинного обучения.

Первый подход удобен с точки зрения обработки, имеет детерминированный характер, однако вводит множество ограничений для пользователя. Любое отклонение от шаблона приведёт к ошибке или не правильным результатам.

Второй подход удобен для пользователя, так как не нужно задумываться над форматом сообщений, можно просто написать или озвучить сообщение в удобном для него виде. Этот подход использует алгоритмы машинного обучения и имеет вероятностный характер.

В работе используется комбинированный метод обработки. На первом этапе используется сервис Wit.ai, который помимо перевода речи в текст позволяет создать и обучить модель для выделения и классификации сущностей в тексте, что изображено на рисунке 6. На втором этапе используются регулярные выражения и шаблоны для поиска и уточнения даты и времени, что изображено на рисунке 7.

встреча завтра в 8 00

☐ event

встреча

1.000

☐ date

завтра

1.000

☐ time

в 8 00

0.965

Рисунок 6. Результат работы модели на сервисе Wit.ai

Фраза	Распознанная дата
25 марта 2019 г	2019-03-25 18:37
двадцать пятого марта 2020 г	2020-03-25 18:37
на следующей неделе	2019-04-16 18:37
через 20 минут	2019-04-09 18:57
в эту пятницу	2019-04-12 18:37
через 3 часа	2019-04-09 21:37
через 4 дня	2019-04-13 18:37
в этот понедельник	2019-04-15 18:37
в эту субботу	2019-04-13 18:37
завтра	2019-04-10 18:37
послезавтра	2019-04-11 18:37
сегодня	2019-04-09 18:37

Рисунок 7. Результат анализа с использование регулярных выражений

После того как вся важная информация была получена из текста, необходимо проверить, есть ли возможность записать новое событие в выбранны период времени. Если же время занято другим событием, пользователю будет предложено изменить время нового события или перенести существующее.

Далее все необходимое записывается в базу данных для дальнейшей работы.

Пользователь имеет возможность просматривать и редактировать ранее созданный список событий, что изображено на рисунке 8. Так же предусматривается интеграция чат-бота с сервисом работы с календарями для более наглядного представления информации.

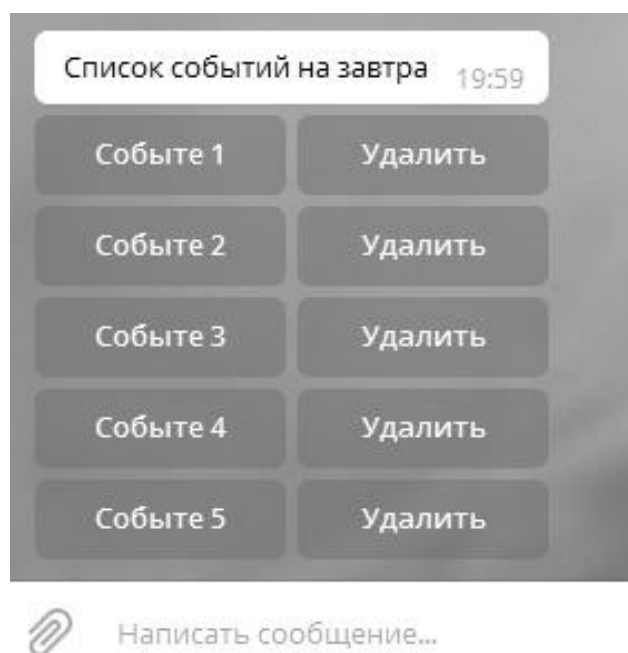


Рисунок 8. Просмотр и редактирование списка событий

При наступлении даты и времени, указанных в сообщении, чат-бот пришлёт пользователю оповещение с текстом из сообщения. Так же предусмотрена функция настройки времени оповещения, что позволит пользователю всегда быть в курсе предстоящих событий. Например, «...напомнить за 15 минут».

В результате работы был создан чат-бот, выполняющий функции персонального помощника ассистента, имеющий возможность голосового и текстового ввода и работающий на платформе мессенджера Telegram.

В дальнейшем можно увеличивать функционал, расширять возможности чат-бота, добавить интеграции с другими сервисами планирования расписания и перенести чат-бота в другие мессенджеры и социальные сети.

Заключение. Чат-боты способны решать многие задачи, связанные не только с развлечениями, но и с осуществлением коммуникаций с клиентами, направленных на привлечение новой аудитории, продвижение рекламы, стимулирование продаж, что способствует увеличению эффективности.

Заклучение. Чат-боты способны решать многие задачи, связанные не только с развлечениями, но и с осуществлением коммуникаций с клиентами, направленных на привлечение новой аудитории, продвижение рекламы, стимулирование продаж, что способствует увеличению эффективности.

Библиографический список

1. Виртуальный собеседник (https://ru.wikipedia.org/wiki/Виртуальный_собеседник)
2. Психология роботов: Книга о чувствах людей и машин. Левшенко Федор Валерьевич PNG press, 21 февр. 2019 г. - Всего страниц: 153
3. Элиза (программа) ([https://ru.wikipedia.org/wiki/Элиза_\(программа\)](https://ru.wikipedia.org/wiki/Элиза_(программа)))

4. Чат-боты: настоящее и будущее искусственного интеллекта (<https://promdevelop.ru/chat-boty-nastoyashhee-budushhee-iskusstvennogo-intellekta/>)
5. Что такое чат-боты и зачем они нужны? (<https://informburo.kz/cards/chto-takoe-chat-boty-i-zachem-oni-nuzhny.html>)
6. Natural Language for Developers Wit.ai (<https://wit.ai>)

EXAMPLES OF DEVELOPMENT CHAT-BOTS

Ten A., Kuzmin A.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

400005, Russia, Volgograd, Lenin avenue, 28, ten_anastasia@mail.ru, sasha-kuz77@yandex.ru

The article describes the relevance of chat bots to date, presents examples of the development of a chat bot that is integrated into the Telegram platform.

Key words: chatbot, smartbot, scheduler.

УДК 004.8

ПРИМЕНЕНИЕ ЧАТ-БОТОВ ДЛЯ РЕШЕНИЯ РАЗНОПЛАНОВЫХ ЗАДАЧ

*Тен Анастасия Евгеньевна, Кузьмин Александр Юрьевич,
Розалиев Владимир Леонидович, Орлова Юлия Александровна*
Волгоградский государственный технический университет,
400005, Россия, г. Волгоград, проспект Ленина, д. 28,
vladimir.rozaliev@gmail.com

В статье описаны, актуальность чат-ботов на сегодняшний день, представлен примеры разработки чат-бота, интегрируемого в платформу Telegram.

Ключевые слова: чат-бот, интернет магазин.

В настоящее время чат-бот является инструментом коммуникаций, который широко используется в повседневной жизни человека. Пользователь ведет беседу с виртуальным собеседником или программой-собеседником с развлекательной или познавательной целью. Например, чат-боты могут выполнять функцию личного помощника, который записывает планы и составляет расписание; бизнес-партнера, который поможет сделать заказ в кулинарии; собеседника, который расскажет о погоде за окном.

Преимущества использования чат-ботов заключаются в том, что они способны откликаться в кратчайшие сроки 24 часа в сутки, экономия

трудовых ресурсов и автоматизация бизнес-процессов, дружелюбное общение. Чат-боты быстро совершенствуются и в некоторых областях способны полностью заменить человека. К недостаткам можно отнести несовершенный интерфейс: люди по-разному реагируют и воспринимают информацию.

Первый виртуальный собеседник Элиза был создан в 1966 году американским учёным, специалистом в области искусственного интеллекта, Джозефом Вейценбаумом [1].

На сегодняшний день существует несколько видов чат-ботов: консультанты; помощники; развлечения; бизнес чат-боты. Функция чат-ботов лежит в плоскости оперативных услуг, когда надо быстро и в любой момент ответить на вопрос или выполнить анализ ситуации. Позволяет экономить на веб-персонале, делая работу автоматизированной [2].

Для разработки чат-ботов существует множество платформ, таких как ВКонтакте, WhatsApp, Telegram, Viber, Facebook, Twitter и другие.

Для реализации рассмотренных ниже примеров использовался язык программирования Python, платформа Telegram, так как: Telegram имеет возможность простого создания и администрирования бота, удобное клиентское приложение, возможность создания интерактивных кнопок в чате, готовые библиотеки для реализации, популярность платформы. Python имеет большое количество документаций и библиотек для реализации чат-ботов.

Рассмотрим пример бизнес чат-бота: чат-бот интернет магазин печенья на платформе Telegram.

Для оптимизации времени клиента и производителя существует возможность производить заказ онлайн через чат-бот на платформе Telegram.

Программа предназначена для работы с каталогом продукции, формирования и оформления заказа в интернет-магазине.

Описанная ниже реализация чат-бота интернет-магазина будет рационально использоваться для магазинов с небольшим ассортиментом. В таком случае каталог выглядит наиболее эстетично и удобно в применении.

Работа с программой осуществляется в несколько этапов:

1) Чат-бот интернет магазин реагирует на определенные команды в виде текста. Например, для того, чтобы инициировать беседу нужно нажать «Старт» или вручную ввести «/start», что изображено на рисунке 1.

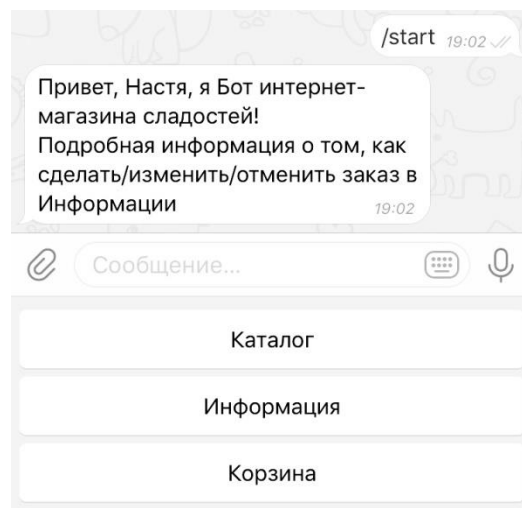


Рисунок 1. Инициирование беседы пользователем

2) Для заказа необходимо перейти в каталог, выбрать нужные товары, добавить некоторое количество в корзину и перейти к оформлению заказа. При добавлении товара в корзину программа чат-бот присваивает данному пользователю идентификационный номер, под которым в дальнейшем будут записаны все выбранные им товары. При просмотре корзины из базы читается информация о пользователе при помощи присвоенного ранее идентификационного номера.

3) При оформлении заказа пользователем требуются его контактные данные для связи. Пользователь должен выбрать способ связи с администратором: позвонить или написать в Telegram. Только после этого действия будут отправлены данные о заказе администратору. Выделенное сообщение может просматриваться только администратором.

4) Далее, после обработки заказа, администратор свяжется с пользователем, выбравшим способ с помощью личных данных, таких как телефон, имя и другие.

Данные линейные действия можно производить в любом порядке.

В результате работы был создан чат-бот, выполняющий функции интернет-магазина, имеющий возможность просматривать каталог, формирование и оформление заказа, работающий на платформе Telegram.

В дальнейшем можно увеличить функционал, интегрировать с другими сервисами, перенести чат-бота в другие мессенджеры и социальные сети.

Рассмотрим пример помощника чат-бота: чат-бот планировщик для составления расписания на платформе Telegram.

Основными функциями чат-бота планировщика являются: создание заметок, записей о событиях, напоминаний, автоматизированное формирование и заполнение личного расписания на период времени, редактирование созданного расписания, оповещение при наступлении времени напоминания, возможность голосового ввода по средствам отправки голосовых сообщений [3].

Для взаимодействия с ботом, пользователь должен отправить сообщение в чат. Мессенджер выступает в роли интерфейса, посредника

между пользователем и приложением. Так как основной акцент сделан на удобстве использования чат-бота, была предусмотрена функция обработки голосовых сообщений, т.е. помимо текстовых сообщений пользователь может озвучить свой запрос голосом. Для распознавания голоса и перевода речи в текст был применён сторонний сервис Wit.ai [4], который предоставляет возможность загрузить звуковой файл с речью и получить ответ в виде текста.

Следующим этапом обработки сообщений пользователя является выделение ключевой информации о событии из текста сообщения. Есть два подхода к обработке текста: заранее заготовленные команды и шаблоны входящих сообщений и обработка естественного языка с использованием технологий машинного обучения.

Первый подход удобен с точки зрения обработки, имеет детерминированный характер, однако вводит множество ограничений для пользователя. Любое отклонение от шаблона приведёт к ошибке или не правильным результатам.

Второй подход удобен для пользователя, так как не нужно задумываться над форматом сообщений, можно просто написать или озвучить сообщение в удобном для него виде. Этот подход использует алгоритмы машинного обучения и имеет вероятностный характер.

В работе используется комбинированный метод обработки. На первом этапе используется сервис Wit.ai, который помимо перевода речи в текст позволяет создать и обучить модель для выделения и классификации сущностей в тексте, что изображено на рисунке 2. На втором этапе используются регулярные выражения и шаблоны для поиска и уточнения даты и времени, что изображено на рисунке 3.

Entity Type	Entity Value	Confidence Score
event	встреча	1.000
date	завтра	1.000
time	в 8 00	0.965

Рисунок 2. Результат работы модели на сервисе Wit.ai

Фраза	Распознанная дата
25 марта 2019 г	2019-03-25 18:37
двадцать пятого марта 2020 г	2020-03-25 18:37
на следующей неделе	2019-04-16 18:37
через 20 минут	2019-04-09 18:57
в эту пятницу	2019-04-12 18:37
через 3 часа	2019-04-09 21:37
через 4 дня	2019-04-13 18:37
в этот понедельник	2019-04-15 18:37
в эту субботу	2019-04-13 18:37
завтра	2019-04-10 18:37
послезавтра	2019-04-11 18:37
сегодня	2019-04-09 18:37

Рисунок 3. Результат анализа с использование регулярных выражений

После того как вся важная информация была получена из текста, необходимо проверить, есть ли возможность записать новое событие в выбранны период времени. Если же время занято другим событием, пользователю будет предложено изменить время нового события или перенести существующее.

Далее все необходимое записывается в базу данных для дальнейшей работы.

Пользователь имеет возможность просматривать и редактировать ранее созданный список событий, что изображено на рисунке 4. Так же предусматривается интеграция чат-бота с сервисом работы с календарями для более наглядного представления информации.

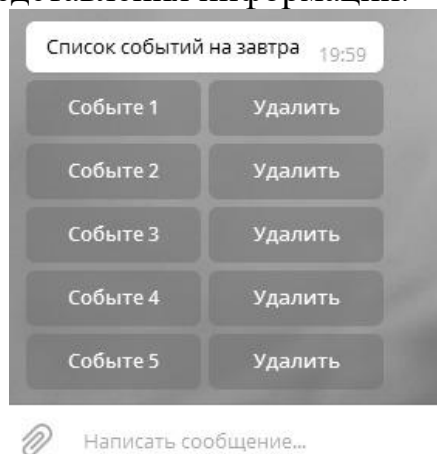


Рисунок 4. Просмотр и редактирование списка событий

При наступлении даты и времени, указанных в сообщении, чат-бот пришлёт пользователю оповещение с текстом из сообщения. Так же предусмотрена функция настройки времени оповещения, что позволит пользователю всегда быть в курсе предстоящих событий. Например, «...напомнить за 15 минут».

В результате работы был создан чат-бот, выполняющий функции персонального помощника ассистента, имеющий возможность голосового и текстового ввода и работающий на платформе мессенджера Telegram.

В дальнейшем можно увеличивать функционал, расширять возможности чат-бота, добавить интеграции с другими сервисами планирования расписания и перенести чат-бота в другие мессенджеры и социальные сети.

Заключение. Чат-боты способны решать многие задачи, связанные не только с развлечениями, но и с осуществлением коммуникаций с клиентами, направленных на привлечение новой аудитории, продвижение рекламы, стимулирование продаж, что способствует увеличению эффективности.

Работа выполнена при частичной поддержке РФФИ (гранты 17-29-07021, 18-47-343007, 18-07-00220, 19-07-00020).

Библиографический список

1. Психология роботов: Книга о чувствах людей и машин. Левшенко Федор Валерьевич PNG press, 21 февр. 2019 г. - Всего страниц: 153
2. Чат-боты: настоящее и будущее искусственного интеллекта (<https://promdevelop.ru/chat-boty-nastoyashhee-budushhee-iskusstvennogo-intellekta/>)
3. Что такое чат-боты и зачем они нужны? (<https://informburo.kz/cards/chto-takoe-chat-boty-i-zachem-oni-nuzhny.html>)
4. Natural Language for Developers Wit.ai (<https://wit.ai>)

APPLICATION OF CHAT BOTS FOR SOLVING DIFFERENT TASKS

Ten A., Kuzmin A., Rozaliev V., Orlova Yu.

VOLGOGRAD STATE TECHNICAL UNIVERSITY

400005, Russia, Volgograd, Lenin avenue, 28, vladimir.rozaliev@gmail.com

The article describes the relevance of chat bots to date, presents examples of the development of a chat bot that is integrated into the Telegram platform.

Key words: chatbot, smartbot, scheduler.

ЦИФРОВАЯ ПСИХОЛОГИЯ И КРИМИНАЛИСТИКА

ИСПОЛЬЗОВАНИЕ МЕТОДОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ПРОГНОЗИРОВАНИИ ПРЕДРАСПОЛОЖЕННОСТИ ЧЕЛОВЕКА К АЛКОГОЛИЗМУ

Жигалова Светлана Андреевна

Национальный исследовательский университет «Высшая школа экономики»,
Пермский филиал, Россия, 614070, Пермь, ул. Студенческая, 38,
sazhigalova@edu.hse.ru, jsa.zh@yandex.ru

В статье описан процесс проектирования и разработки нейросетевой математической модели, предназначенной для прогнозирования предрасположенности людей к алкогольной зависимости. Было выполнено обучение и тестирование системы на библиографических данных, полученных от людей как предрасположенных к алкоголизму, так и не имеющих склонности к нему. На основе полученных результатов были выявлены закономерности, которые определяют наиболее значимые параметры, сильнее других влияющие на предрасположенность человека к алкоголизму, а также дают рекомендации по избавлению от зависимости.

Ключевые слова: алкоголизм, склонность, зависимость, искусственный интеллект, нейронная сеть, математическое моделирование.

Анализ литературных источников [1-7] показывает, что их авторы обращают внимание на несколько факторов, влияющих на злоупотребление спиртных напитков. Однако на данный момент не существует единого мнения ученых о том, какой фактор наиболее значим в выявление алкогольной зависимости, отсутствует оценка степени влияния отдельных факторов на алкогольную зависимость. Аналогичная ситуация наблюдается и в других родственных областях психодиагностики. Для решения этой проблемы при определении предрасположенности людей к наркомании [8] и к насилию [9] успешно используется аппарат нейросетевого моделирования. Для построения аналогичной системы определения предрасположенности к алкогольной зависимости впервые была предпринята попытка применения нейронных сетей авторами работы [10].

Настоящая работа выполнена с применением методики [8-10] и является продолжением этих исследований.

Для построения нейросетевой модели было выбрано 17 входных параметров: пол; месяц рождения; день рождения; год рождения; имели ли предрасположенность к алкоголизму родители; полная ли семья; наличие насилия в семье (рукоприкладство); текущее семейное положение; наличие

наркотической зависимости (употребление наркотиков); наличие никотиновой зависимости (курение); склонность к депрессиям, психозам, неврозам; наличие пережитых эмоциональных потрясений (потеря близкого человека, жертва насилия); наличие черепно-мозговых травм; тип темперамента; уровень образования; финансовое положение; тип профессии в зависимости от предмета труда. Выходной параметр нейросети кодирует значение 1, если человек имеет предрасположенность к алкоголизму, и 0 – если нет.

Для обучения нейронной сети было сформировано множество примеров, основанных на анализе результатов проведенного анонимного опроса среди друзей, родственников и знакомых, а также пациентов в Пермском краевом наркологическом диспансере в городе Кунгур (Краевой наркологический диспансер № 3, ГУЗ). Все собранные данные были разделены на обучающее множество, состоящее из 80 человек и тестирующее, состоящее из 15 человек.

Оптимальная структура нейронной сети, обеспечивающая минимальные ошибки обучения и тестирования, представляла собой персептрон, имеющий 17 входных нейронов, три скрытых слоя с двумя, одним и тремя нейронами и один выходной нейрон. После обучения, нейронная сеть была проверена на данных из тестирующего множества. Результат тестирования представлен на рис. 1 в виде гистограммы, позволяющей сопоставить фактические (заданные анкетами) и действительные (вычисленные нейронной сетью) значения на тестовом множестве примеров. Среднеквадратическая ошибка тестирования составила 1,8%.

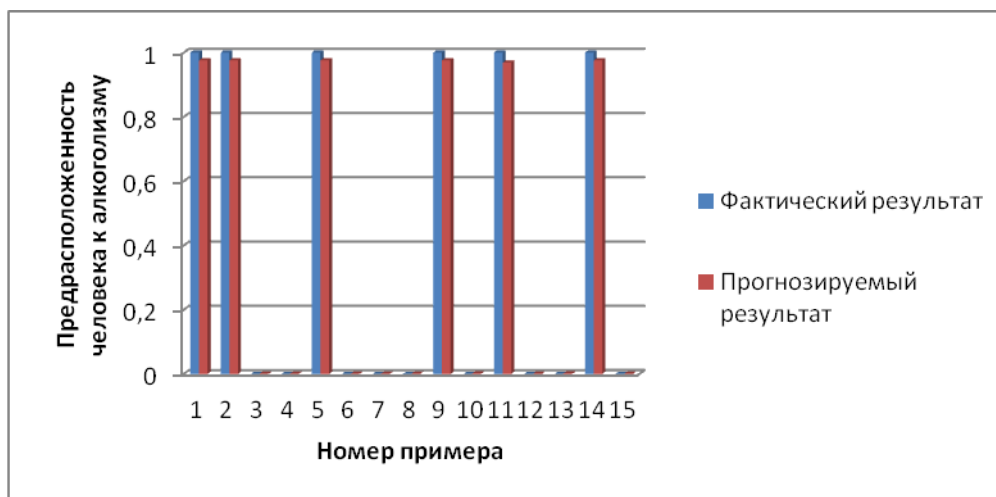


Рисунок 1. Результаты проверки нейронной сети на тестирующем множестве

После того, как работа нейронной сети проверена на тестовых примерах и, таким образом, подтверждена адекватность нейросетевой математической модели, можно приступать к ее исследованию.

Из гистограммы, представленной на рисунке 2 видно, что наибольший вес имеет параметр 15 - уровень образования, а также большую роль играют

параметры 12 - наличие пережитых эмоциональных потрясений и 13 – наличие черепно-мозговых травм.

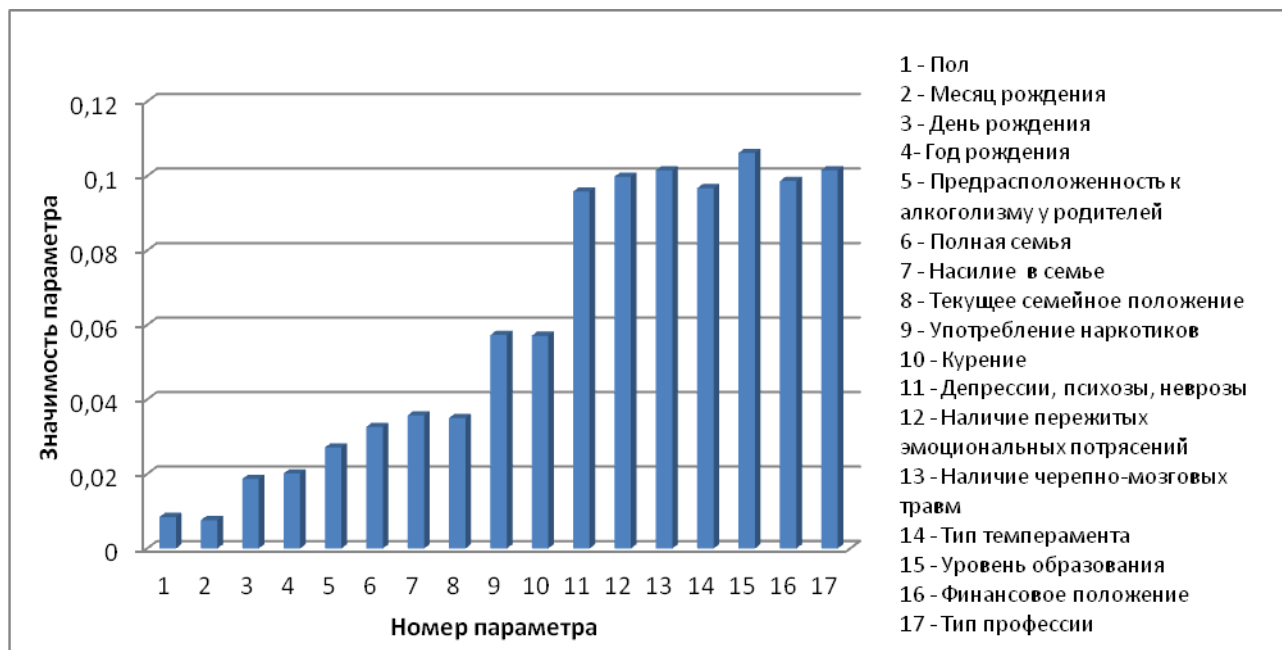


Рисунок 2. Значимость параметров для обучения нейронной сети

Дальнейшая цель исследований состояла в выяснении характера влияния различных входных параметров модели на результат моделирования, т.е. – на предрасположенность человека к алкоголизму. Для проведения виртуальных компьютерных экспериментов были выбраны три человека с предрасположенностью к алкоголизму:

1. Женщина; черепно-мозговой травмы не было; склонна к депрессиям; не переживала сильных эмоциональных потрясений (не теряла близкого человека, не была жертвой насилия); пил отец; употребляет наркотики; не курит; тип профессии – человек-художественный образ.

2. Мужчина; черепно-мозговой травмы не было; склонен к депрессиям; не переживал сильных эмоциональных потрясений (не терял близкого человека, не был жертвой насилия); пила мать; не употребляет наркотики; курит; тип профессии – человек-техника.

3. Мужчина; была черепно-мозговой травма; склонен к депрессиям; переживал сильных эмоциональных потрясений (терял близкого человека или был жертвой насилия); родители не пьющие; не употребляет наркотики; курит; тип профессии – человек-техника.

Результаты виртуальных компьютерных экспериментов над wybranymi людьми представлены на рисунках 3 – 4. Исследования выполнены путем изменения одних входных параметров и сохранении других при одновременном наблюдении за значением выходного сигнала нейросети.

На рисунке 3 представлены полученные таким способом результаты вычисления предрасположенности к алкоголизму каждого из испытуемых при виртуальном изменении их предрасположенности к депрессиям и

сохранении всех остальных параметров неизменными. У всех троих есть предрасположенность к депрессиям. Красным цветом на гистограмме закрашены столбцы, показывающие реальное значение предрасположенности к алкоголизму людей, выбранных для исследования, а синим – прогнозируемого системой при новом значении параметра. Согласно полученным данным, у всех трех человек при виртуальном уменьшении исходной предрасположенности к депрессиям, предрасположенность к алкоголизму снижается практически до нулевой отметки.

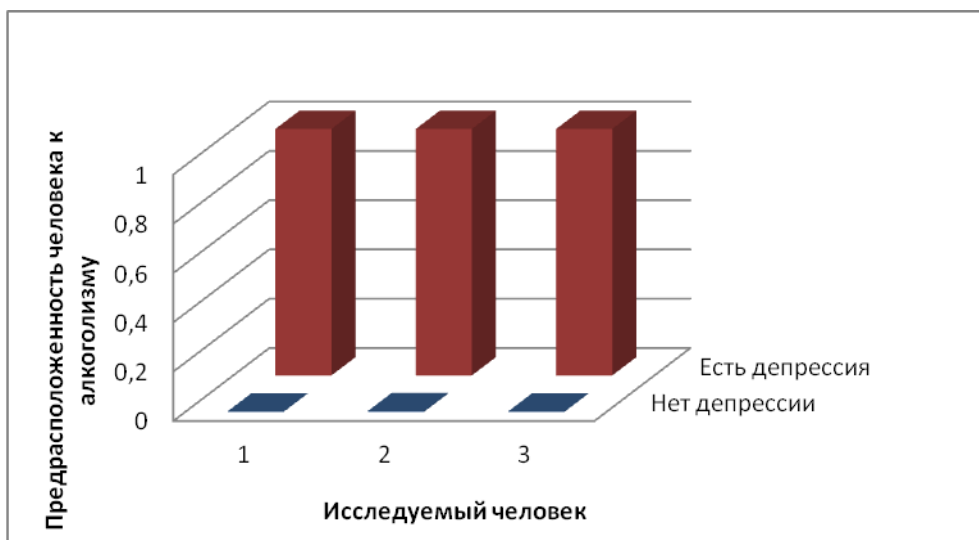


Рисунок 3. Предрасположенность к алкоголизму в зависимости от наличия у человека депрессии

На рисунке 4 изображена гистограмма, представляющая результаты вычисления предрасположенности к алкоголизму каждого из испытуемых при виртуальном изменении двух их вредных привычек (наркотиков и курения) и сохранении всех остальных параметров неизменными. Как мы помним из приведенной выше информации об испытуемых, женщина (на гистограмме обозначена цифрой 1) употребляет наркотики и не курит, а мужчины (на гистограмме – цифры 2 и 3 соответственно) не употребляют наркотики, но курят. Как и на предыдущем графике, красным цветом закрашены столбцы, показывающие реальное значение предрасположенности к алкоголизму людей, выбранных для исследования, а синим – прогнозируемое системой. Как видно из гистограммы, для исследуемого человека № 1 предрасположенность к алкоголизму значительно снижается, если он будет курить, но не будет употреблять наркотики или откажется от вредных привычек совсем. Для двух других исследуемых человек есть незначительные колебания в результатах, однако они все также сильно предрасположены к алкоголизму.

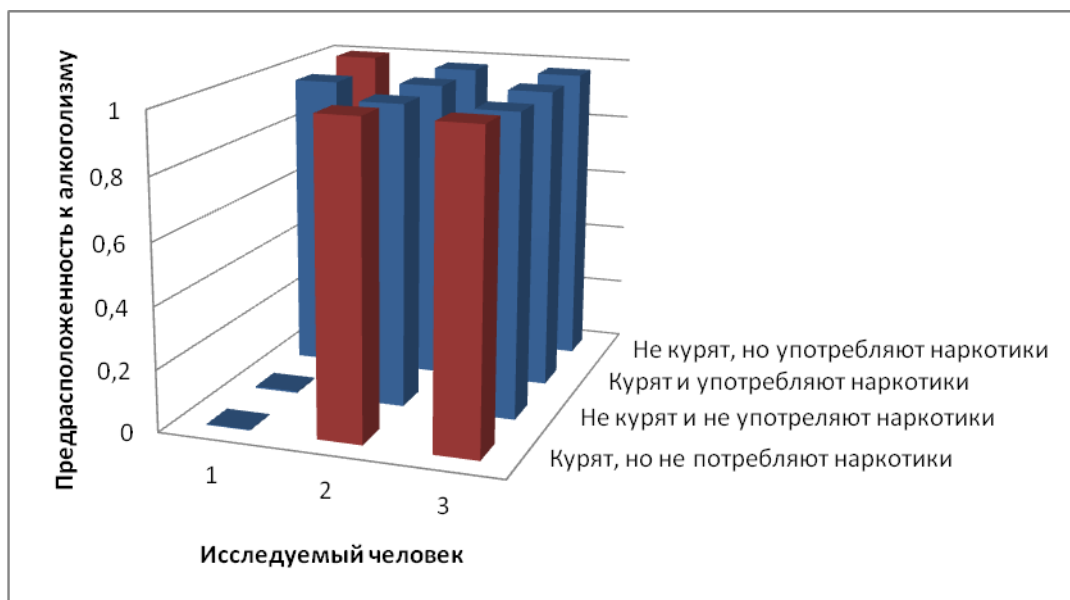


Рисунок 4. Предрасположенность к алкоголизму в зависимости от наличия у человека вредных привычек

Таким образом, можно говорить о том, что из выбранных для вычислительных экспериментов параметров наиболее значимым является предрасположенность человека к депрессивным расстройствам, психозам и неврозам, так как при изменении этого параметра у всех исследуемых людей предрасположенность к алкоголизму падает практически до нулевой отметки.

В итоге можно заключить что, данная модель имеет достаточно обобщённый характер, но, всё-таки, позволяет сделать определённые выводы при оценке предрасположенности человека к алкоголизму, даёт представление об общей ситуации, складывающейся в современном обществе. В нее достаточно легко вносить корректировки, добавлять новые входные параметры, уточнять их, чтобы сделать возможным более точное определение наличия предрасположенности у человека алкогольной зависимости.

Библиографический список

1. «Алкоголизм & Здоровый образ жизни». URL: <http://alcoholismhls.ru/vladimir-cygankov/> (дата обращения: 7.10.2018).
2. Лечебный центр доктора Демченко «Благо». URL: <http://www.blagoperm.ru/index.php/alcoholism/alcoholism> (дата обращения: 5.11.2018).
3. «Личность и алкоголизм». URL: <http://www.medical-enc.ru/alcoholism/lichnost/issledovanie.shtml> (дата обращения: 5.11.2018).
4. «Предрасположенность к алкоголизму и его причины». URL: http://www.doctorsan.ru/prichiny_alkogolizma (дата обращения: 5.11.2018).
5. «Психологический центр Р.Р. Гарифуллина». URL: <http://www.garifullin.net/> (дата обращения 6.11.2011).

6. «Психология алкоголизма: признаки и проявления». URL: http://narcofree.ru/materiali_ob_alkogolizme/osobaya-psikhologiya-alkogolizma (дата обращения: 6.07.2015).
7. «Пьянство и алкоголизм». URL: <http://petrsu.ru/Chairs/Psychiatry/alcohol.htm> (дата обращения: 5.11.2018).
8. Ясницкий Л.Н., Грацилёв В.И., Куляшова Ю.С., Черепанов Ф.М. Возможности моделирования предрасположенности к наркозависимости методами искусственного интеллекта // Вестник Пермского университета. Философия. Психология. Социология. 2015. № 1 (21). С. 61-73.
9. Ясницкий Л.Н., Ваулева С.В., Сафонова Д.Н., Черепанов Ф.М. Использование методов искусственного интеллекта в изучении личности серийных убийц // Криминологический журнал Байкальского государственного университета экономики и права. 2015. Т. 9. № 3. С. 423-430.
10. Ясницкий Л.Н., Гилевский А. С., Радостев Д. А., Рычина Е. В., Черепанов Ф.М. Нейросетевая система диагностики и нивелирования предрасположенности к алкогольной зависимости // Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века, 2016 г. – с.200-206.

THE USE OF ARTIFICIAL INTELLIGENCE METHODS IN PREDICTING A PERSON'S PREDISPOSITION TO ALCOHOLISM

Zhigalova Svetlana Andreevna

National Research University Higher School of Economics, Russia, 614070,
Perm, 38 Studenchskaya st., sazhigalova@edu.hse.ru

The article describes the process of designing and developing a neural network mathematical model designed to predict the predisposition of people to alcohol dependence. Training and testing of the system was carried out on bibliographic data obtained from people who are predisposed to alcoholism and do not have a tendency to it. On the basis of the obtained results, the regularities that determine the most significant parameters that affect the predisposition of a person to alcoholism more than others, as well as give recommendations for getting rid of dependence were revealed.

Key words: alcoholism, addiction, dependence, artificial intelligence, neural network, mathematical modeling.

ОПРЕДЕЛЕНИЕ СОВЕРШЕННОЛЕТИЯ КЛИЕНТА МЕТОДОМ НЕЙРОСЕТЕВОГО МОДЕЛИРОВАНИЯ

Кетов Дмитрий Константинович

Национальный исследовательский университет «Высшая школа экономики»,
Пермский филиал, Россия, 614070, Пермь, ул. Студенческая, 38
ketovdk@rambler.ru

Разработана нейро-экспертная система, предназначенная для определения совершеннолетия человека по фотографии на основе сверточной нейросети. В результате компьютерных экспериментов, выполненных с помощью предлагаемой системы, была достигнута точность определения совершеннолетия в 3% ошибок первого рода и 10% ошибок второго рода. Была высчитана предполагаемая прибыль и необходимая для бизнеса вероятность определения совершеннолетия клиента, на основе построенной математической модели.

Ключевые слова: нейронная сеть, сверточные нейронные сети, обработка изображений, обработка лиц, возраст.

Введение. Существует множество областей бизнеса, в которых запрещено предоставлять услуги несовершеннолетним людям. Зачастую, для удостоверения совершеннолетия проверяется паспорт, но часто бывает так, что человек не может предоставить его в данный момент. В такой ситуации при отказе от предоставления услуги владелец бизнеса теряет деньги, рискуя выплачивать штраф иначе. Эту проблему можно решить, используя нейросетевые технологии.

В настоящее время в научной литературе можно найти множество обзоров таких методов и сообщений об успешном опыте разработки систем, работающих с лицом человека на основе нейросетей [3-7]. Более того, были разработаны специальные сервисы на основе сверточных нейронных сетей для определения возраста человека [8-9].

Стоит отметить, что работы обладают недостаточной точностью. Но можно добиться ее повышения за счет уточнения задачи определения именно совершеннолетия, а не возраста непосредственно.

Цель настоящей работы состоит в попытке применения математического аппарата нейросетевых технологий для решения реальной бизнес-проблемы определения совершеннолетия клиента.

1. Анализ задачи

Предположим, что средняя прибыль с клиента равняется A , вероятность того, что при предоставлении услуги несовершеннолетнему придется

выплачивать штраф за P , размер штрафа за S , а вероятность ошибки нейросети за $P2$.

Таким образом, матожидание прибыли за одного клиента равняется $A - P2 \cdot P \cdot S$. Таким образом, максимально возможная вероятность ошибки (чтобы матожидание превышало 0) равняется $A / (P \cdot S)$.

Исходя из предметной области, $S=10000$, $P=10\%$, а $A=100$. Тогда максимальная вероятность ошибки равняется 10%.

2. Создание приложения

В основе приложения возьмем и настроим почти готовую сверточную сеть [8]. Также были подобраны 100 не совершеннолетних людей и 100 совершеннолетних (от 15 до 22 лет). Будем рассматривать результаты нейросети следующим образом:

- 1) Человеку больше восемнадцати лет. Ответ: человеку больше 18 лет.
- 2) Человеку меньше восемнадцати лет. Ответ: человеку больше 18 лет (ошибка первого рода).
- 3) Человеку больше восемнадцати лет. Ответ: человеку меньше 18 лет (ошибка второго рода).
- 4) Человеку меньше восемнадцати лет. Ответ: человеку меньше 18 лет.

Результаты будут представлены в квадратах следующего формата, изображенного на рисунке 1:

1	2
3	4

Рисунок 1. Шаблон результатов

Результат начальной нейросети представлен на рисунке 2.

95	13
5	87

Рисунок 2. Начальный результат

Как видно по результату, количество ошибок первого рода достаточно высоко и сильно превосходит количество ошибок второго рода. Также было обнаружено, что в среднем, нейросеть показывает возраст на 3 года больше реального. Это связано с тем, что база лиц, использованная в сети основана на европейских лицах, что люди такого возраста часто преувеличивают свой возраст в социальных сетях (в отличии от нашей выборки, в которой используется исключительно достоверная информация, основанная на

данных школы), а также с тем, что фотография гарантировано выкладывается позже момента, в который она была сделана.

После нескольких экспериментов было решено использовать сдвиг возраста совершеннолетия. Зависимость количества ошибок от сдвига возраста изображена на рисунке 3.

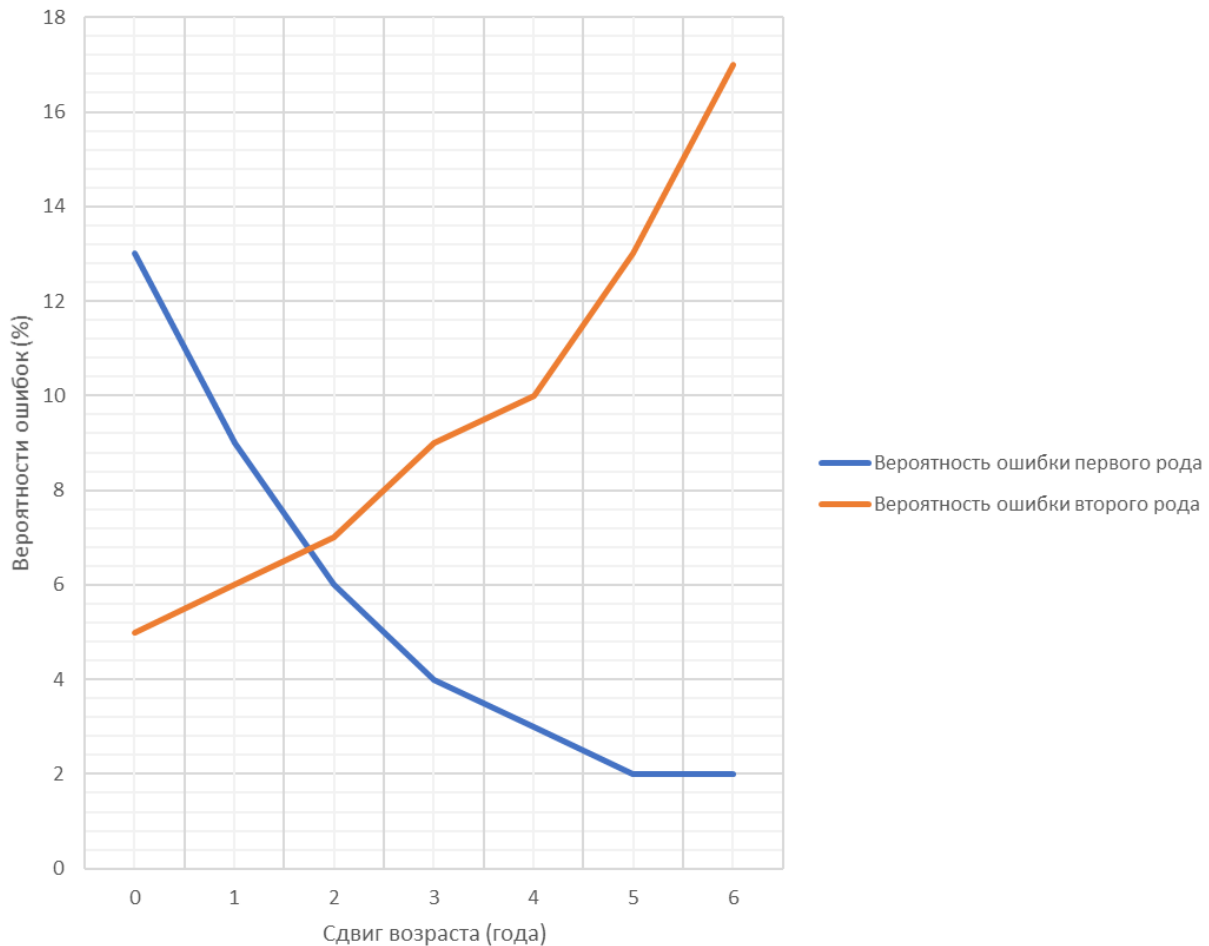


Рисунок 3. Зависимость количества ошибок от сдвига возраста

Таким образом был выделен идеальный сдвиг в 4 года, т.к. при дальнейшем увеличении сдвига вероятность ошибок первого рода почти перестает меняться, а вероятность ошибок второго рода сильно возрастает (природа возрастания похожа на экспоненциальную). Вероятности определения для сдвига в 4 года изображены на рисунке 4.

90	3
10	97

Рисунок 4. Итоговый результат

Теперь необходимо реализовать достаточно удобный для использования интерфейс. Было принято решение как основу интерфейса использовать чат-бота, принимающего фотографии и отвечающего, совершеннолетний человек (люди) или нет. Далее этого чат-бота можно встроить в любой мессенджер (используя Microsoft Bot Framework) или использовать локально.

Заключение. Создано приложение, позволяющее определять возраст человека по его фотографии с достаточной точностью. Приложение почти полностью готово к использованию бизнесом и не требует больших затрат по его внедрению в любые предприятия.

Вычисления производятся облачно, что дополнительно повышает удобность внедрения за счет легкости масштабирования и отсутствия необходимости владения техникой, серверами и т.д. (за хостинг ботов отвечает Microsoft Azure, а за вычисление возраста отвечают сервисы Face API на основе нейросетей проекта Microsoft Cognitive Services).

Таким образом, нейросетевые технологии могут помочь реальному бизнесу и повысить его прибыль без больших трудозатрат на производство и внедрение с легкой масштабируемостью и содержанием пропорциональным частоте использования технологии.

Библиографический список

1. Ясницкий Л.Н., Богданов К.В., Черепанов Ф.М. Технология нейросетевого моделирования и обзор работ Пермской научной школы искусственного интеллекта // Фундаментальные исследования. 2013. № 1 (часть 3). С. 736-740.
2. Ясницкий Л.Н. Введение в искусственный интеллект. Издание 3. М.: Издательский центр «Академия», 2010. С. 176.
3. Li Deng, Dong Yu. Deep Learning: Methods and Applications, 2014.
4. Alexey Dosovitskiy, Jost Tobias Springenberg, Maxim Tatarchenko, Thomas Brox. Learning to Generate Chairs, Tables and Cars with Convolutional Networks. С. 58.
5. Anders Boesen, Lindbo Larsen, Søren Kaae Sønderby. Generating Faces with Torch, 2015.
6. Amit Konar, Aruna Chakraborty. Emotion Recognition // A Pattern Analysis Approach. 1st Edition.
7. Christopher M. Bishop. Pattern Recognition and Machine Learning // Information Science and Statistics.
8. <https://docs.microsoft.com/ru-ru/azure/cognitive-services/face/> (дата обращения: 15.11.2018)
9. <https://docs.microsoft.com/ru-ru/azure/cognitive-services/welcome/> (дата обращения: 15.11.2018)

DETERMINATION THE AGE OF CUSTOMERS BY THE METHOD OF NEURAL NETWORK MODELING

Ketov Ditrui K.

National Research University Higher School of Economics,
38 Studenetskaya st., Perm, 614070, Russia,
ketovdk@rambler.ru

A neural network designed to calculate humans age by his photography. As a result of computer experiments, made by the offering system, 3 % of first kind mistakes and 10% of second kind mistakes had been achieved. The probable profit for business was calculated with created mathematical model.

Key words: neutral network, convolutional neural network, image processing, face processing, age.

УДК 004.823.3

ПРОГНОЗИРОВАНИЕ УРОВНЯ ПРЕСТУПНОСТИ В РЕГИОНАХ РОССИИ

Попонина Алёна Олеговна

Национальный исследовательский университет
«Высшая школа экономики», Пермский филиал,
Россия, 614070, Пермь, ул. Бульвар Гагарина, 37А,
poponina-alena@yandex.ru

В статье представлено описание разработки нейронной сети для прогнозирования уровня преступности в регионах России. В качестве входных параметров выступают 6 переменных: потребление алкоголя, уровень безработицы, число разводов, миграционный прирост, процент населения с денежными доходами ниже величины прожиточного минимума и среднедушевые доходы. В работе подробно описано проектирование, обучение и тестирование нейросети, после чего произведён прогноз и анализ влияния каждого из параметров в отдельности. Кроме того, в работе рассматривается построение и дальнейшее прогнозирование уровня преступности при помощи эконометрической модели. Однако, результаты прогноза нейросети оказались более точными.

Ключевые слова: искусственный интеллект, нейросеть, эконометрическая модель, уровень преступности.

В современном мире, где технологическое развитие достигло небывалых высот, возникает необходимость во всё более прогрессивном использовании имеющихся компьютерных ресурсов. На сегодняшний день всё большую популярность приобретает моделирование и прогнозирование

при помощи различных моделей, искусственного интеллекта или же нейросетей.

Область прогнозирования уровня преступности является актуальным и является объектом исследования как отечественных, так и зарубежных авторов. В одной из иностранных работ исследователи рассматривают более 100 факторов влияния, а также применяют метод генетического программирования [1]. Данный подход более подробно рассматривается в работе российских авторов, которые сравнивают качество полученного прогноза с помощью генетического программирования с другими математическими и эконометрическими моделями [2]. В некоторых работах рассматривается не только какие внешние факторы влияют на уровень преступности, но и исследуется психологический портрет преступников, в частности, серийных убийц [3].

В данной работе рассматривается уровень преступности, факторы, оказывающие на него влияние. Прогнозирование производится двумя методами, а именно с помощью нейросети и эконометрической модели. Целью данного исследования является обучение нейросети с помощью Пермского Нейросимулятора, построение наилучшей эконометрической модели, сравнение результатов прогнозирования и дальнейшее использование наилучшего варианта для расчета уровня преступности.

Для анализа были выбраны 80 регионов России, по которым были собраны все необходимые данные. В качестве источников были использованы сайты «Росстат» (gks.ru), «ЕМИСС» (fedstat.ru), а также «Портал правовой статистики» (crimestat.ru). Далее для того чтобы получить сопоставимые данные, все значения переменных, кроме unemployment, poverty и divorce, были поделены на общую численность населения региона.

Таким образом, для проведения обучения Нейросимулятора под обучающее множество можно выделить 60 наборов, а под тестовое – 20.

Таблица 5. Описание переменных

Обозначение параметра	Описание переменной	Единицы измерения
Выходной параметр D	Количество всего зарегистрированных преступлений в регионе поделенное на население региона	Преступлений на одного жителя региона
X1	Количество проданных декалитров алкогольных напитков в натуральном выражении поделённое на население региона	Тысяча декалитров на одного жителя региона
X2	Общая численность безработных, в процентах к экономически активному населению (уровень безработицы)	Проценты
X3	Число разводов за год поделённое на население региона	Разводов на одного жителя региона
X4	Арифметическая разница между числом прибывших и числом выбывших поделённое на население региона	Миграционный прирост на одного жителя региона
X5	Процент населения с денежными доходами ниже величины прожиточного минимума	Проценты
X6	Среднедушевые денежные доходы	Рубли

На данном этапе необходимо определить структуру персептрона. Количество входных нейронов должно равняться размерности входного вектора, в котором в рамках данного исследования содержится 6 параметров, то есть необходимо 6 входных нейронов. В свою очередь выходной нейрон будет всего один, так как с помощью Нейросимулятора будет прогнозироваться только количество преступлений, приходящееся на одного жителя региона. Наилучшие результаты были получены при 5 нейронах на скрытом слое.

На рисунке 1 видно, что самым значимым параметром является число разводов в регионе, однако стоит отметить, что четыре других параметра: уровень безработицы, миграционный прирост, процент населения за чертой бедности и средний доход, имеют почти такой же уровень значимости. В свою очередь уровень потребления алкоголя является наименее значимым из всего входного вектора X.

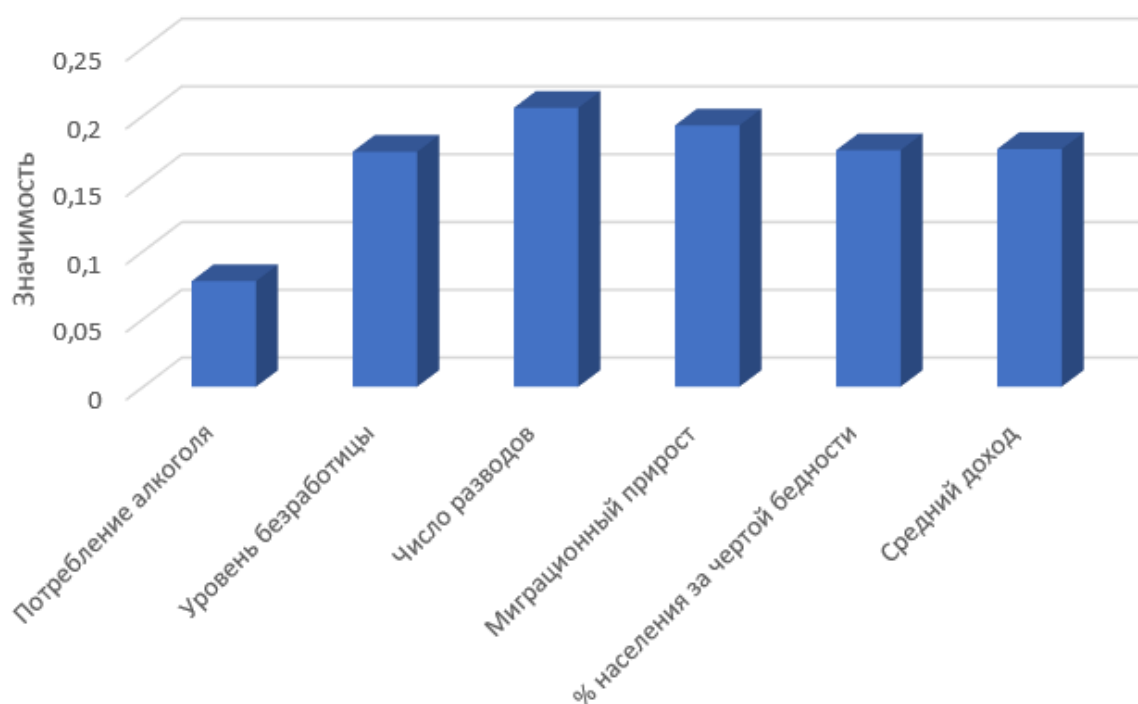


Рисунок 1. Значимость входных параметров нейросети

Рассмотрим влияние различных параметров на уровень преступности в трёх регионах России: Алтайский край, Пермский край и Хабаровский край. Для этого возьмём 15 примеров для каждого региона, в которых будет меняться значение только выбранного параметра:

1. Употребление алкоголя. Будем постепенно увеличивать уровень потребляемого алкоголя от 0,002 до 0,0132.

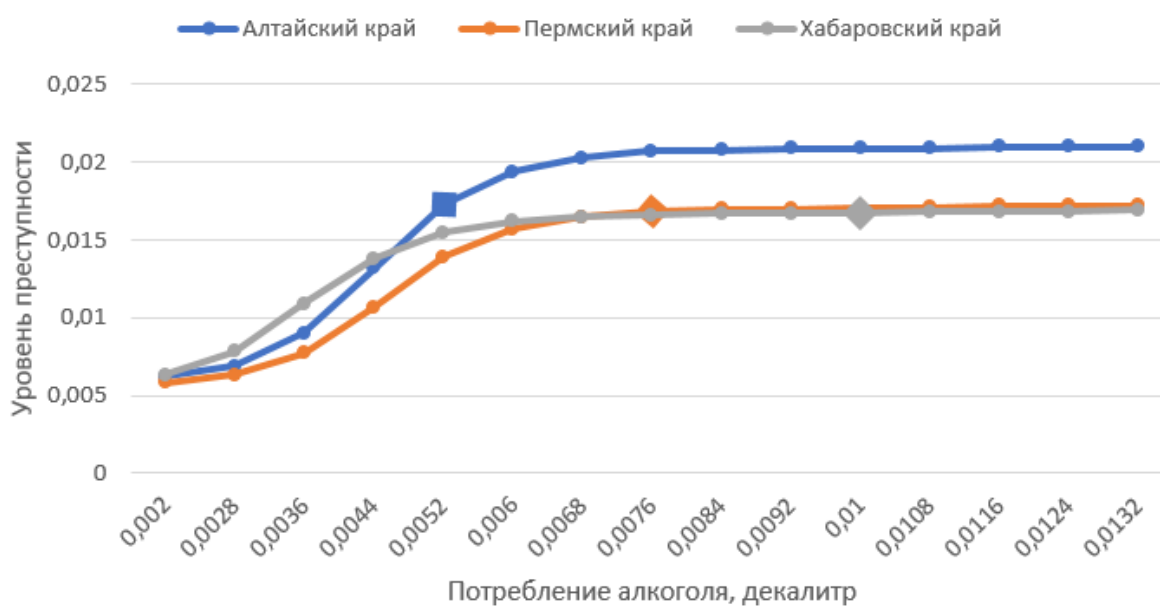


Рисунок 2. Зависимость уровня преступности от количества потребляемого алкоголя

Из диаграммы на рисунке 2 видно, что в среднем с увеличением количества потребляемого алкоголя растёт и уровень преступности. Текущие значения потребления алкоголя отмечены специальным маркером, ромбиком.

2. Уровень безработицы. Теперь попробуем постепенно снижать уровень безработицы в регионе с 0,12 до 0,05, зафиксировав значения всех остальных параметров.

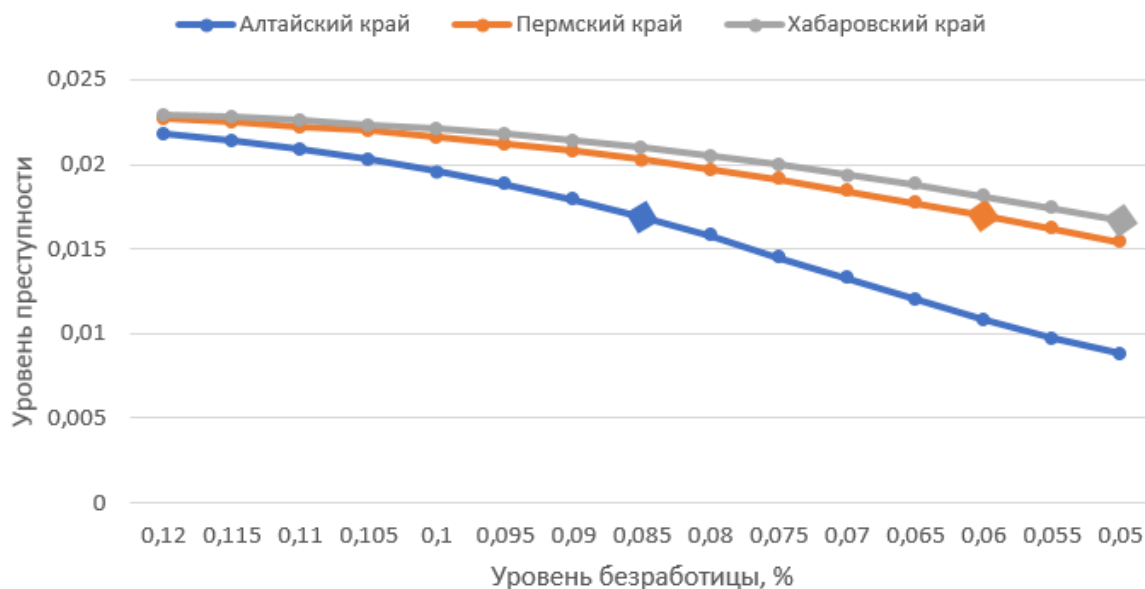


Рисунок 3. Зависимость уровня преступности от уровня безработицы

Как следует из диаграммы на рисунке 3, с постепенным снижением уровня безработицы также снижается и уровень преступности в регионе. Текущие значения уровня безработицы отмечены специальным маркером, ромбиком. Процент населения с доходами ниже минимального уровня. В данном примере будем изменять значение данного параметра от 0 до 0,12

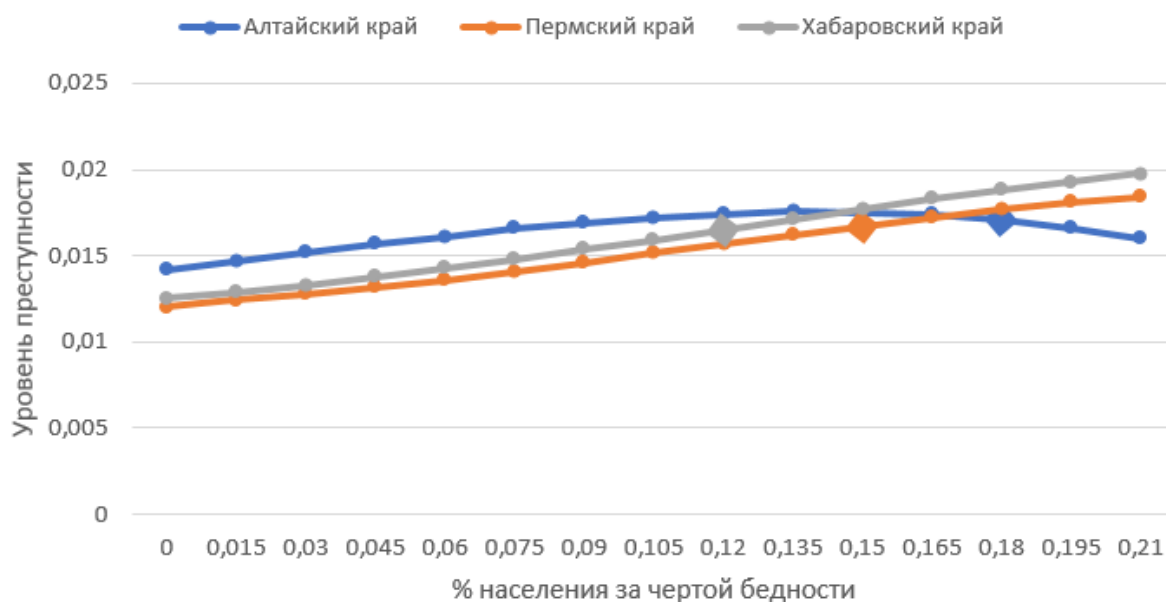


Рисунок 4. Зависимость уровня преступности от процента населения за чертой бедности

Для построения эконометрической модели будем использовать полностью идентичный набор входных данных, а моделирование будет производиться при помощи программного продукта «Gretl».

Итоговая эконометрическая модель имеет вид:

$$crime = -0,017 + 0,171 * unemployment + 0,008 * l_alcohol + 0,006 * l_income \quad (4)$$

В ходе тестирования получившейся модели было выявлено, что в ней нет мультиколлинеарности и избыточных переменных, а также ни одна из других рассматриваемых переменных, то есть *divorce*, *migration* и *poverty* не была пропущена. Также модель имеет правильную функциональную спецификацию.

Однако была выявлена гетероскедастичность. Для избавления от неё был применён взвешенный метод наименьших квадратов, а в качестве весовой переменной была выбрана *unemployment*. Таким образом, была получена новая уже гомоскедастичная модель.

Теперь, имея обученную нейросеть и построенную эконометрическую модель, необходимо сравнить точность прогноза и выбрать вариант, с помощью которого можно получить наиболее точные значения уровня преступности.

Для этого возьмём исходную выборку данных и в случайном порядке выберем 20 примеров для тестирования нейросети и эконометрической модели. Рассчитав 20 прогнозных значений уровня преступности двумя методами, необходимо определить можно сравнить качество прогноза. На рисунке 5 представлена диаграмма, с помощью которой можно визуальное оценить точность расчётного значения уровня преступности.

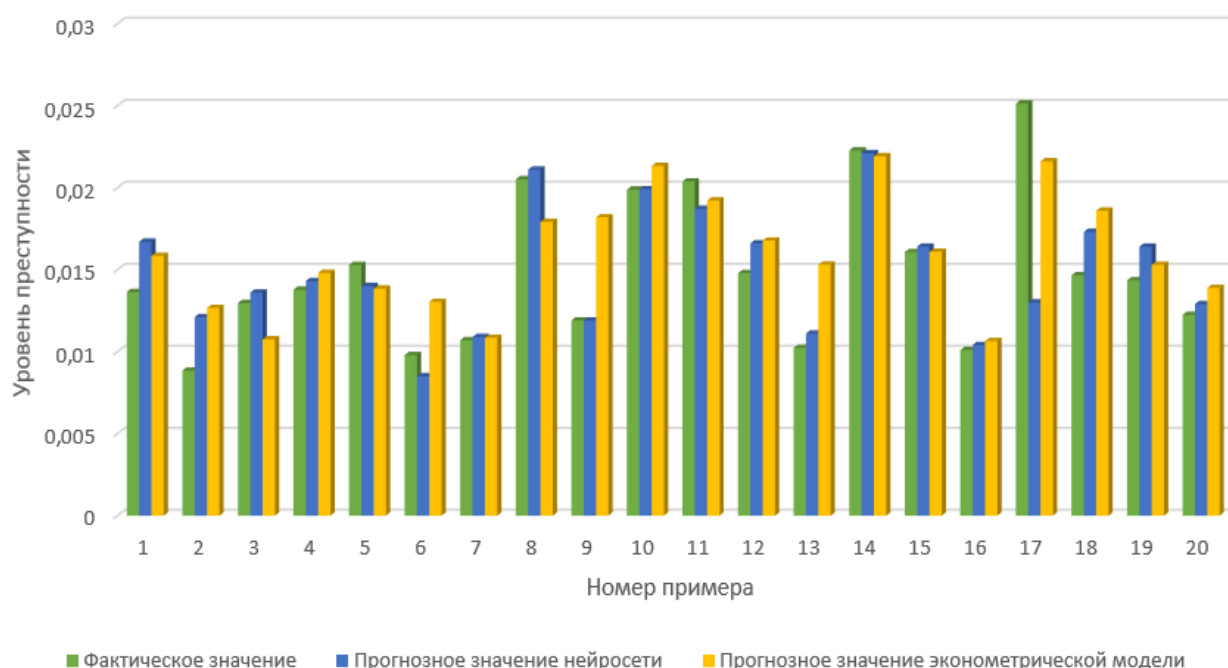


Рисунок 5. Сравнение прогнозных значений нейросети и эконометрической модели

Сумма ошибки прогноза при помощи нейросети получилась 0,0336, а при помощи эконометрической модели – 0,0436. Таким образом, в рамках данного исследования нейросеть даёт более качественный прогноз. Также стоит отметить, что нейросеть учитывает все рассмотренные факторы влияния, а в эконометрическую модель были включены только три из них.

В результате данной работы были рассмотрены нейросеть и эконометрическая модель, рассматривающие уровень преступности в качестве выходного параметра. Было выявлено, что с помощью нейросети можно получить более точные прогнозные значения.

Библиографический список

1. Castelli M., Sormani R., Trujillo L., Popovic A. Predicting per capita violent crimes in urban areas: an artificial intelligence approach // J. Ambient Intell. Humaniz. Comput. 2017. Vol. 8, № 1. P. 29–36.
2. Лановой А.Ф. Об одном подходе к построению прогнозирующих моделей временных рядов показателей преступности // Запобігання та ліквідація надзвичайних ситуацій 2008. С. 167-169.
3. Ясницкий Л.Н., Ваулева С.В., Сафонова Д.Н., Черепанов Ф.М. Использование методов искусственного интеллекта в изучении личности серийных убийц // Криминологический журнал Байкальского государственного университета экономики и права. 2015. Т. 9. № 3. С. 423-430.
4. Ясницкий Л. Н. Интеллектуальные системы. М.: Лаборатория знаний, 2016. 221 с.

FORECASTING CRIME RATE IN REGIONS OF RUSSIA

Poponina Alena O.

National Research University Higher School of Economics, 37A Gagarina Bulvar st., Perm, 614070, Russia, poponina-alena@yandex.ru

The article describes the development of neural network for forecasting crime rate in regions of Russia. There were 6 input parameters: alcohol consumption, unemployment rate, migration, percentage of the population below poverty level and average income. The article describes designing, training and testing the neural network. Furthermore, the article reviews the econometric model and compares results of forecasting with neural network and econometric model. Neural network made more accurate forecasts.

Key words: artificial intelligence, neural network, econometric model, crime rate.

УДК 343.9, 004.8

ПРИМЕНЕНИЕ НЕЙРОСЕТЕВЫХ ТЕХНОЛОГИЙ В ИЗУЧЕНИИ ФАКТОРОВ, ВЛИЯЮЩИХ НА ПРЕСТУПНОСТЬ В ГОРОДАХ РОССИИ

Пьянков Данила Дмитриевич, Малюгин Михаил Ильич.

Пермский государственный национальный исследовательский университет.

Юридический факультет.

614990, Россия, г. Пермь, ул. Букирева, 15, pyankoff.danila@yandex.ru

Ясницкий Леонид Нахимович

Пермский государственный национальный исследовательский университет

614990, Россия, г. Пермь, ул. Букирева, 15, yasn@psu.ru

Разработана компьютерная программа, предназначенная для прогнозирования преступности в городах России. В основе программы лежит нейронная сеть, обученная на примерах статистических данных по преступности в городах России. Программа дает возможность влиять на результаты прогнозов путем изменения различных факторов благоустройства города.

Ключевые слова: преступность, города, Россия, нейронная сеть, сценарный прогноз.

Первые попытки создания методики применения нейросетевых технологий в криминалистике, в частности, для создания интеллектуальных систем, предназначенных для выявления серийных убийц и наркоманов, а также для разработки инструментальных средств допроса преступников, описаны в работах [1-3]. В настоящей работе указанная методика применена для изучения факторов, влияющих на преступность в городах России.

1. Актуальность. Одна из основных проблем современных городов мира заключается в ежегодном повышении уровня преступности. Указанную проблему отмечают почти 90% городов мира. В результате этого, правительства городов мира прикладывают большие усилия для решения проблемы. По-видимому, одним из инструментов решения проблемы городской преступности может быть создание и исследование нейронной сети, моделирующей данную предметную область.

2. Построение нейросетевой модели. Для построения нейросетевой математической модели, предназначенной для прогнозирования возможной преступности городов России, было найдено множество структурных и описательных элементов городов России. Данная информация была взята из компетентных источников.

В качестве входных параметров были приняты:

X1 – Численность населения города.

X2 – Площадь города.

X3 – Плотность населения.

X4 – Год основания города.

X5 – Процентный состав русского населения.

X6 – Часовой пояс.

X7 – Рождаемость (на душу населения).

X8 – Смертность (на душу населения).

X9 – Количество вузов (на душу населения).

X10 – Количество театров (на душу населения).

Модель имела выходной параметр: D – Процент преступности на душу населения в городах России.

Всего было собрана и обработана описательная информация по 100 городам России, из них 85 городов были использованы для обучения нейросети и 15 городов – для ее тестирования. Нейронная сеть персептронного типа была спроектирована и оптимизирована по традиционной методике, сложившейся в Пермской научной школе искусственного интеллекта. В окончательном варианте персептрон имел 110 входных нейронов, 1 нейрон на скрытом слое и 1 выходной нейрон. После поиска и исключения нескольких на тестирующем множестве примеров была достигнута среднеквадратичная погрешность 20%, причем дополнительное тестирование по методу многократной перекрестной проверки дали аналогичные результаты.

3. Исследование модели. После того, как работа нейросети была проверена на тестовых примерах и, таким образом, проверена адекватность нейросетевой математической модели, можно приступить к ее исследованию.

Прежде всего, путем поочередного исключения входных нейронов и наблюдением за погрешностью сети была определена значимость входных параметров модели, т.е. – степень их влияния на результат моделирования. Как видно из рисунка 1, наиболее значимыми входными параметрами оказались: часовой пояс, рождаемость, процент русских смертность, количество театров и количество вузов.

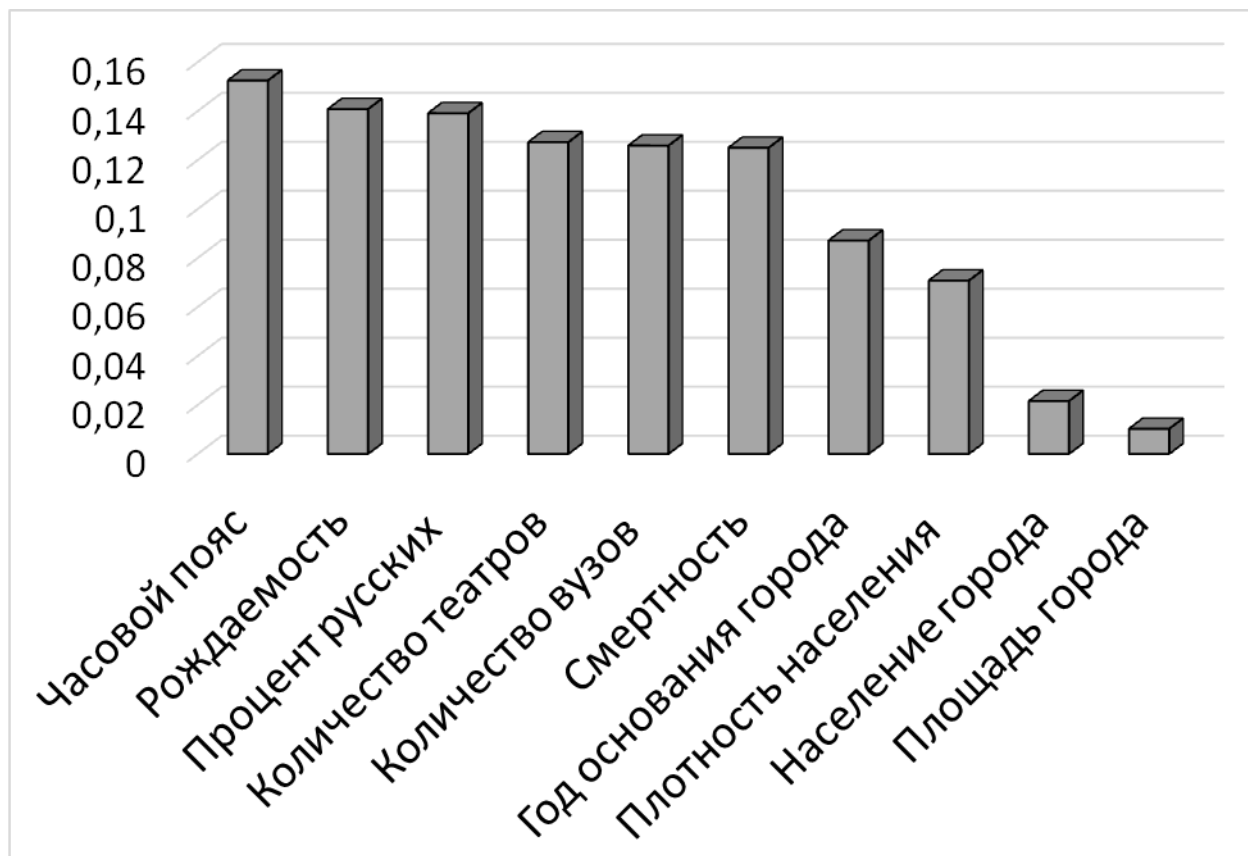


Рисунок 1. Наиболее значимые параметры, влияющие на уровень преступности городов России

С целью изучения предметной области были выполнены компьютерные эксперименты по прогнозированию процента преступности в городах Иркутск, Якутск и Москва.

Город Иркутск имеет следующие параметры: численность населения – 624 тыс.чел., площадь города – 227 км.кв., плотность населения – 2252 чел./км.кв., год основания города – 1661, процентный состав русского населения – 92%, часовой пояс – 8, рождаемость (на душу населения) - 0,0128чел., смертность (на душу населения) - 0,0124чел., количество вузов (на душу населения) - 0,000025шт., количество театров (на душу населения) - 0,0000032шт.

Город Якутск имеет следующие параметры: численность населения – 312 тыс.чел., площадь города – 122 км.кв., плотность населения – 2555 чел./км.кв., год основания города – 1632, процентный состав русского населения – 37%, часовой пояс – 9, рождаемость (на душу населения) -

0,0149чел., смертность (на душу населения) - 0,0083чел., количество вузов (на душу населения) - 0,000057шт., количество театров (на душу населения) - 0,0000064шт.

Город Москва имеет следующие параметры: численность населения – 12506 тыс.чел., площадь города – 2562 км.кв., плотность населения – 4931 чел./км.кв., год основания города – 1147, процентный состав русского населения – 92%, часовой пояс – 3, рождаемость (на душу населения) - 0,0107чел., смертность (на душу населения) - 0,0095чел., количество вузов (на душу населения) - 0,000012шт., количество театров (на душу населения) - 0,0000032шт.

В ходе компьютерных экспериментов поочередно изменялись два входных параметра: «Количество вузов (на душу населения)» и «Количество театров (на душу населения)». Остальные входные параметры оставались неизменными. В результате вычислений были построены диаграммы рисунков 2 и 3. Маркером увеличенного размера отмечена реальная преступность городов, мелкие маркеры – результат сценарного прогнозирования.

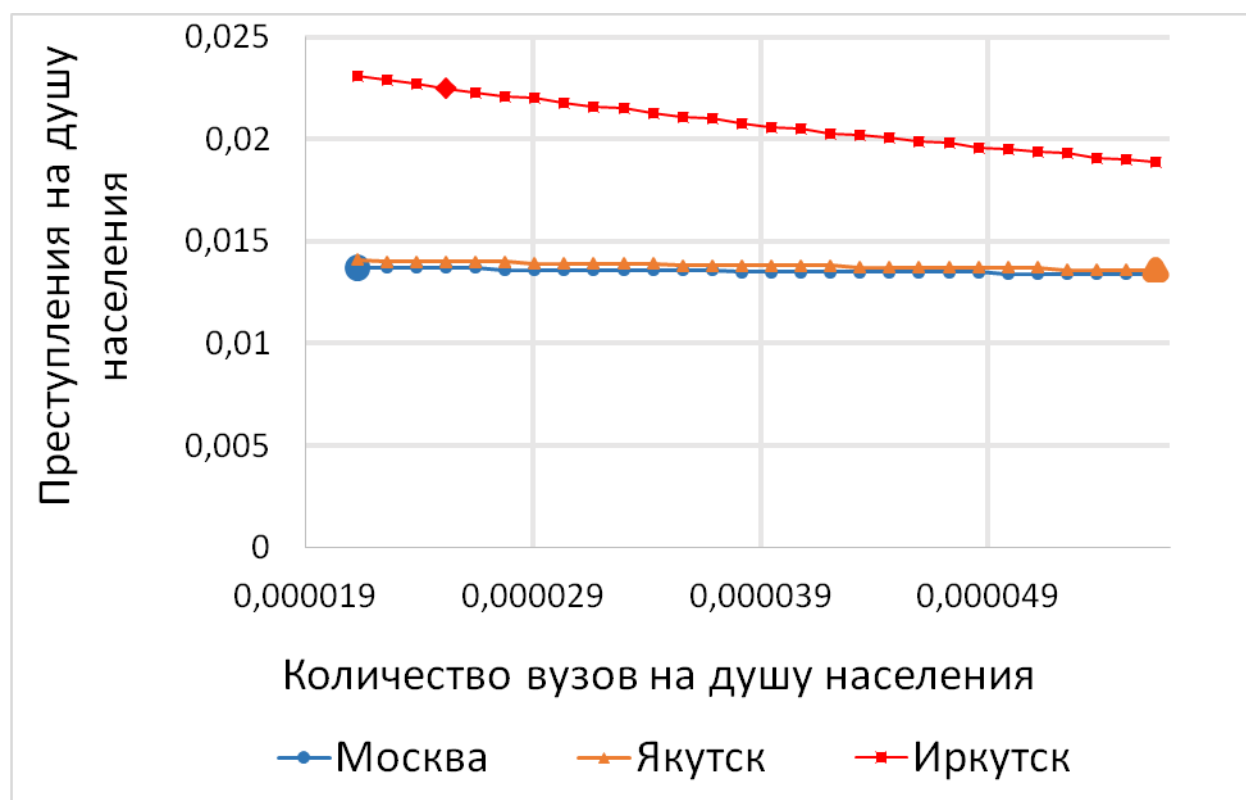


Рисунок 2. Зависимость Преступности в городах Иркутск, Москва и Якутск от количества вузов

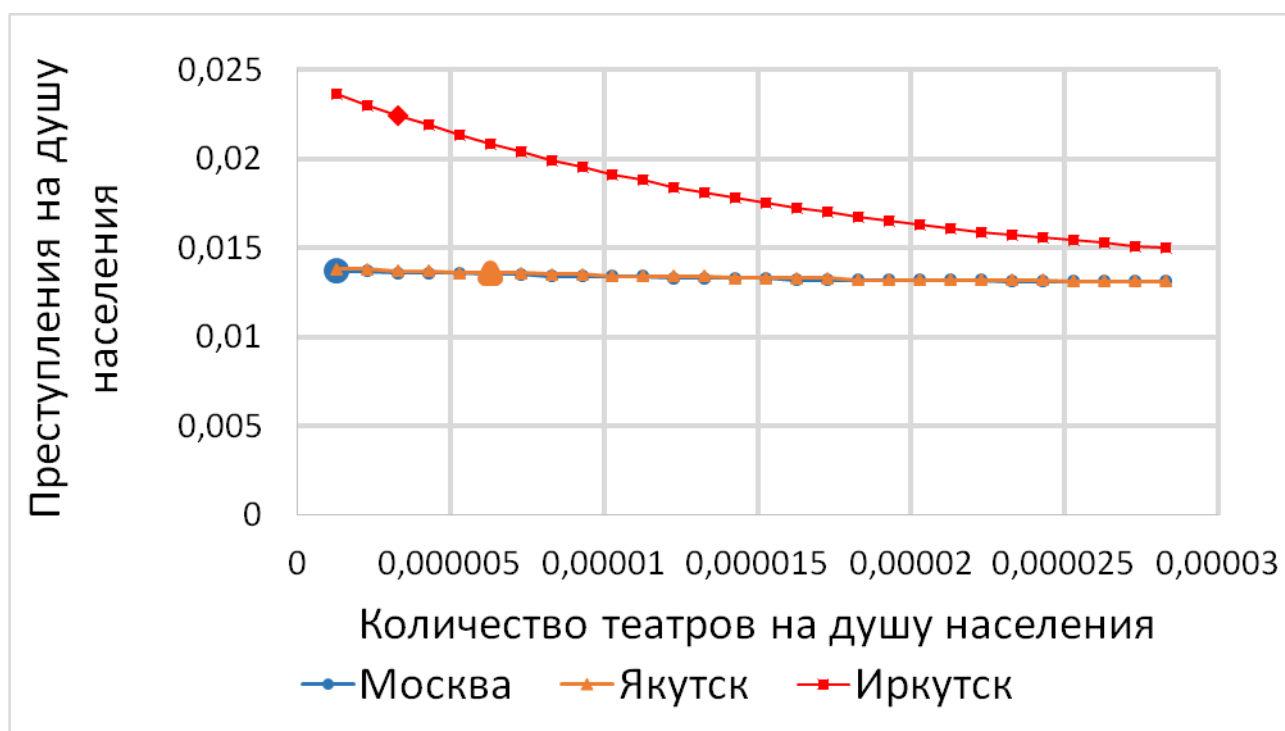


Рисунок 3. Зависимость преступности в городах Иркутск, Москва и Якутск от количества театров

Как видно из рисунков 2 и 3, количество вузов и количество театров в Москве и Якутске незначительно влияют на уровень преступности в этих городах, тогда как с повышением количества театров и вузов в Иркутске, уровень преступности заметно снижается. Так, на сегодняшний день в Иркутске имеется 2 театра, а население составляет 624 000 человек, т.е. на душу населения приходится 0,0000032. Преступность при этом составляет 0,023 на душу населения. При увеличении количества театров с 2-х до 3-х, на душу населения будет приходиться 0,000004 театров. Тогда, как следует из рисунка 3 уровень преступности в городе Иркутске упадет с 0,023 до 0,022.

Аналогичный результат достигается при увеличении количества вузов.

Результаты компьютерного эксперимента в ходе изменения входного параметра - процентный состав русского населения и сравнения его с этим же параметром других городов, было выявлено, что с повышением процентного состава русского населения уровень преступности в Иркутске, Москве и Якутске повышается, как показано на рисунке 4.

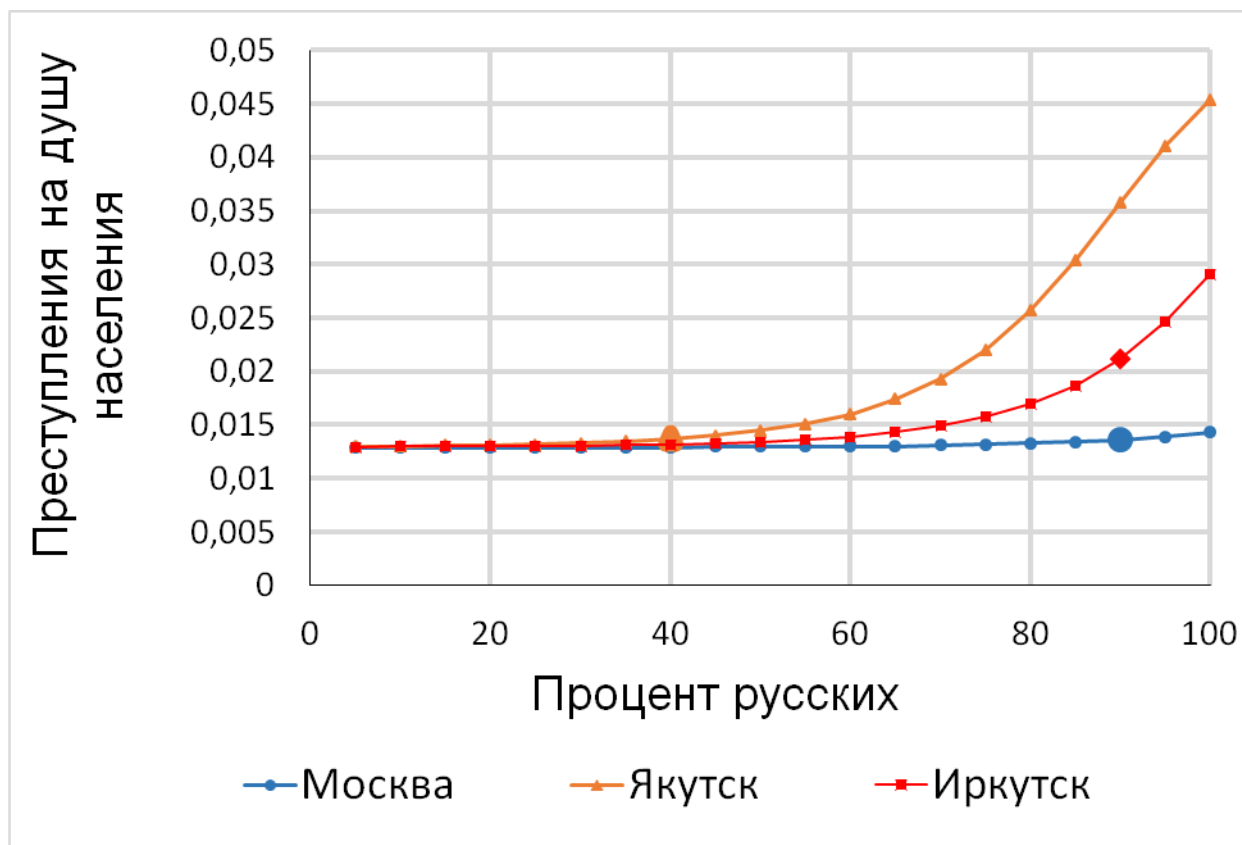


Рисунок 4. Зависимость преступности в городах Иркутск, Москва и Якутск от процентного состава русского населения

Обсуждение результатов и заключение. Подводя итог выполненным исследованиям, отметим, что приведенные здесь выводы, на первый взгляд, кажутся бесполезными и даже сомнительными. Но, во-первых, напомним, что они являются результатом компьютерных экспериментов над математической моделью, адекватность которой доказана сравнительно низкой погрешностью (20%), полученной на тестовых примерах, т.е. на примерах поведения реальных городов, данные о которых при создании модели не использовались. Именно такой погрешностью можно оценить результаты расчетов, приведенные на рисунках 2–4. Кроме того, нами были найдены наиболее значимые параметры, влияющие на уровень преступности (рисунок 1).

Во-вторых, эти результаты имеют некоторое практическое значение. Нами показано, как математически строго можно получать объективную оценку значимости параметров городов России, которые в наибольшей степени влияют на уровень преступности. В ходе исследования установлено, что уровень преступности в рассмотренных городах может быть снижен, например, за счет увеличения миграции жителей других стран в Россию. Кроме того, установлено, что способом для понижения уровня преступности в городе Иркутск может быть открытие новых вузов и театров, но эти же

действия практически не изменяют уровень преступности в Якутске и в Москве.

Библиографический список

1. Ясницкий Л.Н., Ваулева С.В., Сафонова Д.Н., Черепанов Ф.М. Использование методов искусственного интеллекта в изучении личности серийных убийц // Криминологический журнал Байкальского государственного университета экономики и права. 2015. Т. 9. № 3. С. 423-430.
2. Ясницкий Л.Н., Грацилёв В.И., Куляшова Ю.С., Черепанов Ф.М. Возможности моделирования предрасположенности к наркозависимости методами искусственного интеллекта // Вестник Пермского университета. Философия. Психология. Социология. 2015. № 1 (21). С. 61-73.
3. Ясницкий Л.Н., Петров А.М., Сичинава З.И. Технологии построения детектора лжи на основе аппарата искусственных нейронных сетей // Информационные технологии. 2010. № 11. С. 66-70.

THE USE OF NEURAL NETWORK TECHNOLOGIES IN THE STUDY OF THE FACTORS INFLUENCING CRIME

Pyankov Daniel D., Malugin Mikhail I., Yasnitsky Leonid N.
Perm state University, 15, Bukireva str., Perm, Russia, 614990,
yasn@psu.ru

A computer program designed to predict crime in Russian cities has been developed. The program is based on a neural network trained on the examples of crime statistics in Russian cities. The program makes it possible to influence the results of forecasts by changing various factors of improvement of the city.

Keywords: crime, cities, Russia, neural network, scenario forecast.

УДК 004.89

ПРОГНОЗИРОВАНИЕ ТЯЖЕСТИ ДОРОЖНО-ТРАНСПОРТНЫХ ПРОИСШЕСТВИЙ

Романова Марина Павловна

Пермский государственный национальный исследовательский университет,
614990, Россия, г. Пермь, ул. Букирева, 15,
romanovamp@psu.ru

В работе были использованы две реализации нейронных сетей для прогнозирования тяжести ДТП, произошедших в городе Пермь, – Нейросимулятор и библиотека *Keras*, а также проведен анализ адекватности

и сравнение результатов двух моделей. Были выявлены наиболее значимые входные параметры. Предсказанная двумя моделями тяжесть возможного ДТП иногда отличается от реальных данных, т.к. помимо учтенных шести параметров, на ДТП влияют и другие факторы, не принятые во внимание в текущих моделях.

Ключевые слова: ДТП, прогнозирование, нейронная сеть, нейросимулятор.

За 2017 год на территории Российской Федерации зафиксировано порядка 126 тыс. дорожно-транспортных происшествий, в которых были зарегистрированы пострадавшие и погибшие [1]. Суммарное количество погибших, по сравнению с 2016 годом, возросло достаточно существенно – на 16.6 тыс. человек. Суммарное количество пострадавших, но оставшихся в живых и получивших серьезные травмы, несколько снизилось, но в сравнении с другими странами остается чрезмерно высоким – 160 тыс. человек, поэтому проблема не перестанет быть актуальной еще долгое время.

Для прогнозирования тяжести нанесенного ущерба при возможном ДТП, произошедшем в городе Пермь, было решено применить две реализации нейронных сетей: пакет «Нейросимулятор5» [2], а также программа на языке программирования *Python* с использованием библиотеки *Keras*. Основным источником, раскрывающим теоретические основы построения и применения нейронных сетей, являются учебные пособия [3,4].

Для прогнозирования тяжести ДТП выбраны параметры, которые доступны обычному пользователю сети Интернет [5,6], это дата рождения и пол водителя – виновника ДТП, дата поездки (день, месяц, время суток) и район Перми. Для обучения нейронной сети были вручную собраны примеры реальных ДТП в городе Пермь за 2015-2018гг. Множество было разбито на обучающее, состоящее из 70 примеров, использованное для обучения сети, и тестирующее множество, состоящее из 12 примеров, предназначенное для проверки ее прогностических свойств. Оптимальная структура нейронной сети представляет собой персептрон, имеющий 6 входных нейронов, скрытый слой с 3 нейронами и один выходной нейрон.

После обучения прогностические свойства нейронной сети проверялись на примерах тестирующего множества. Результат тестирования двух реализаций нейронных сетей представлен на рисунке 1 в виде гистограммы, позволяющей сопоставить желаемые и действительные значения выходных параметров.

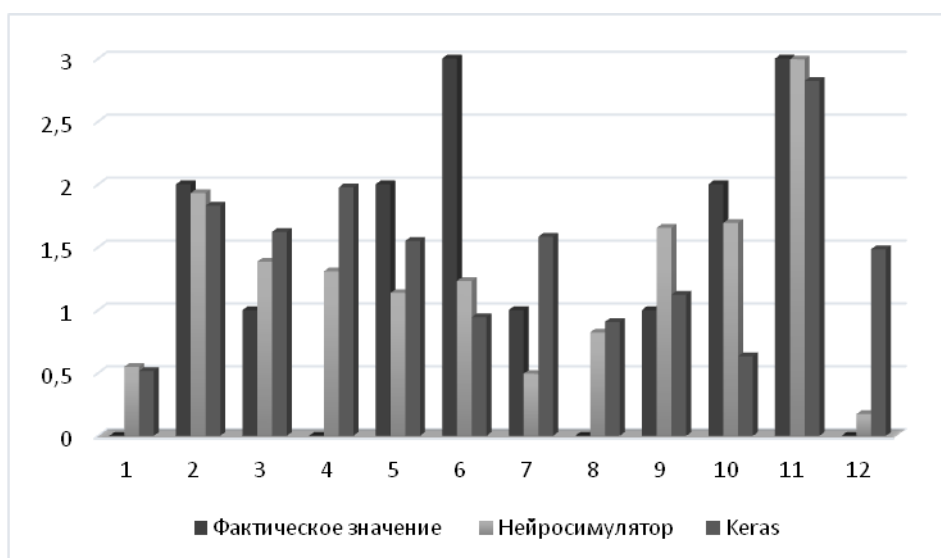


Рисунок 1. Гистограмма сравнения двух реализаций нейронных сетей

Ошибки двух реализаций нейронных сетей, которые оказались достаточно близки, были вычислены по формуле среднеквадратичной ошибки MSE. Ошибка Нейросимулятора составляет 9.2%, а ошибка библиотеки *Keras* – 9.07%.

Следующим важным этапом исследования являлось выявление значимости параметров (Рисунок 2).

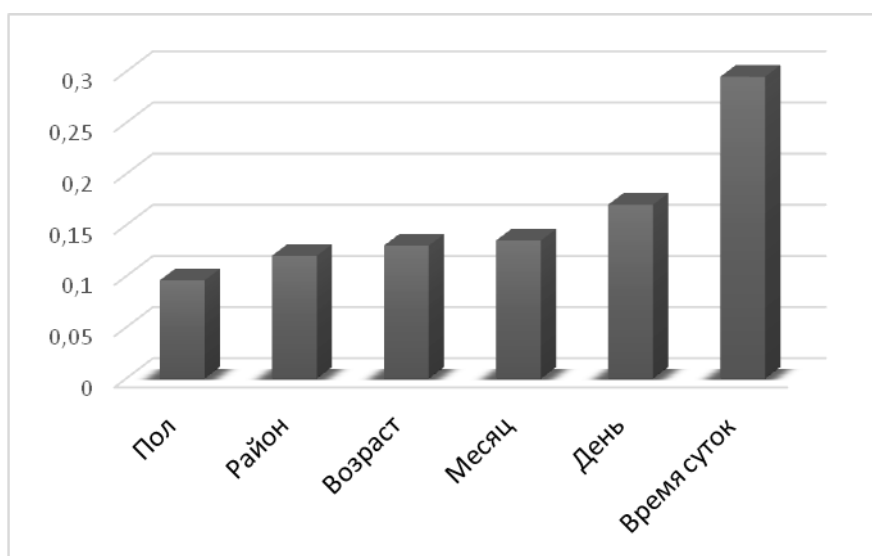


Рисунок 2. Значимость параметров

Из гистограммы видно, что наиболее важный параметр, влияющий на тяжесть ДТП – время суток. Кроме него, важна дата (день и месяц) ДТП. С другой стороны, такой параметр как пол виновника вообще практически не влияет на результат.

Исходя из выше описанных результатов, была рассмотрена зависимость тяжести ДТП от времени суток более подробно. Проанализировано ДТП со смертельным исходом, произошедшее 25 октября, примерно в 23:00 в Орджоникидзевском административном районе (Рисунок 3). На диаграмме показана зависимость тяжести ДТП от времени суток.

Видно, что около 12:00 была высока вероятность того, что в результате ДТП причинённый ущерб был бы наименьшим.

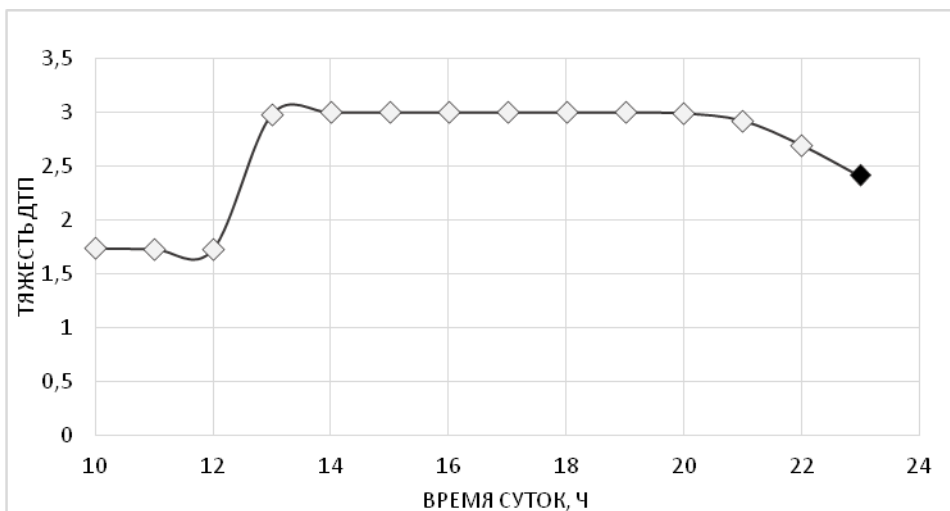


Рисунок 3. Зависимость тяжести ДТП от времени суток

Вторым по значимости параметром является дата, поэтому также была проанализирована зависимость тяжести ДТП от дня месяца (Рисунок 4) рассмотренного выше происшествия.

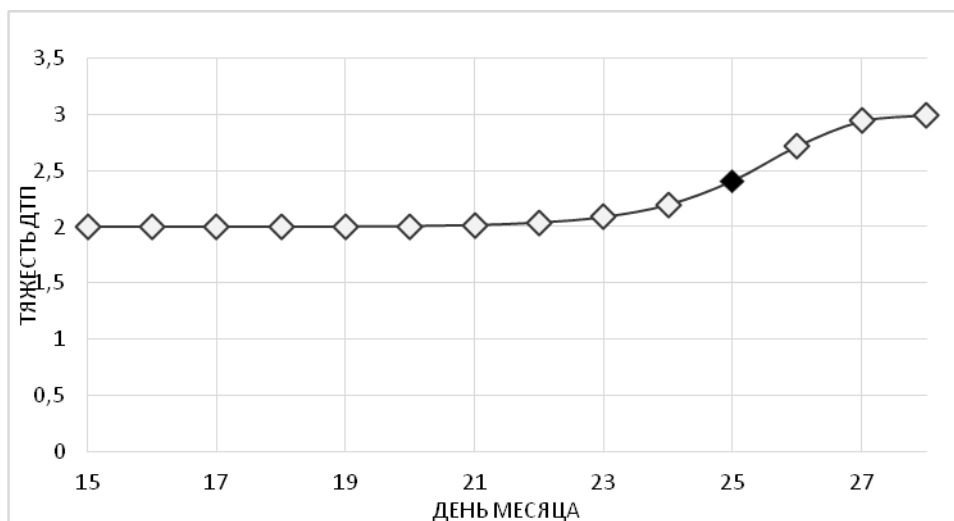


Рисунок 4. Зависимость тяжести ДТП от дня месяца

Можно заметить, что с 15 по 21 октября, ущерб при ДТП наименьший.

Заключение

В работе были использованы две реализации нейронных сетей Нейросимулятор и библиотека *Keras* для прогнозирования тяжести ДТП, а также проведен анализ адекватности и сравнение результатов двух моделей.

В данной работе показана разница ошибок двух реализаций нейронных сетей, а также выявлены наиболее значимые входные параметры – это время суток и дата ДТП.

Предсказанная тяжесть возможного ДТП на тестирующей выборке иногда отличается от реальных данных, отчего можно сделать вывод, что помимо учтенных шести параметров, на происшествие влияют и другие факторы, не принятые во внимание в текущих моделях.

Библиографический список

1. ДТП: [Электронный ресурс] // URL: <https://properm.ru/news/dtp/>. (Дата обращения: 21.02.2019).
2. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 5.10. Свидетельство о государственной регистрации программы для ЭВМ № 2014612546. Заявка Роспатент № 2014610341. Зарегистрировано в Реестре программ для ЭВМ 28 февраля 2014г.
3. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – С.176.
4. Ясницкий Л.Н. Интеллектуальные системы. М.: Лаборатория знаний, 2016. С.221.
5. Статистика автокатастроф в России: [Электронный ресурс] // URL: <http://avtopravozashita.ru/dtp/statistika-dtp-v-rossii-za-2016-god.html>. (Дата обращения: 25.02.2019).
6. Помощник водителя сегодня: [Электронный ресурс] // URL: <http://www.driver-helper.ru/dtp/r/perm/now>. (Дата обращения: 18.02.2019).

PREDICTING SEVERITY OF ROAD TRAFFIC ACCIDENTS

Romanova Marina P.
Perm State University
614990, Russia, Perm, street bukireva, 15,
romanovamp@psu.ru

In this work, two implementations of the neural networks Neurosimulator and the Keras library were used to predict the severity of accidents that occurred in the city of Perm, and an analysis of the adequacy and comparison of the results of the two models was carried out. The most significant input parameters were identified. The severity of possible accidents predicted by the two models sometimes differs from real data, since in addition to the six parameters taken into account, other factors that are not taken into account in current models also affect the crash.

Key words: accidents, forecasting, neural network, neurosimulator.

ЦИФРОВАЯ ГЕОФИЗИКА

РЕЗУЛЬТАТ РЕШЕНИЯ ОПТИМИЗАЦИОННЫХ ЗАДАЧ, СВЯЗАННЫХ С ФИЛЬТРАЦИЕЙ В НЕФТЯНОМ ПЛАСТЕ

Мирошниченко Сергей Сергеевич,

Ижевский государственный технический университет имени

М.Т. Калашникова,

426069, Россия, г. Ижевск, Студенческая, д. 7,

iite@list.ru

В статье представлены результаты по регулированию работы нефтяных и газовых скважин, определение местоположения добывающих и нагнетающих скважин приводит к необходимости решать задачи оптимизации различного типа. Мы будем рассматривать задачи, связанные с эксплуатацией отдельного нефтяного пласта. Процесс эксплуатации пласта является нестационарным, так как его характеристики изменяются по мере извлечения нефти и закачки воды.

Ключевые слова: нефтяной пласт, нагнетательная скважина, фильтрация

В качестве самого простого примера рассмотрим задачу о размещении нагнетательных скважин в стационарной постановке однофазной фильтрации [1-4].

Примеры приведены для прямоугольного пласта с размерами 2000x1000м и средней толщиной 10м. Конфигурация пласта, при возможности построения криволинейной сетки, не вносит принципиальной сложности в решение оптимизационной задачи [5-7].

Рассмотрим сравнительно краткосрочный плановый период $T=1.6$ года. Будем рассматривать 5 добывающих и 8 водонапорных скважин для пласта.

Сходимость итерационного процесса представлена на рис. 1.

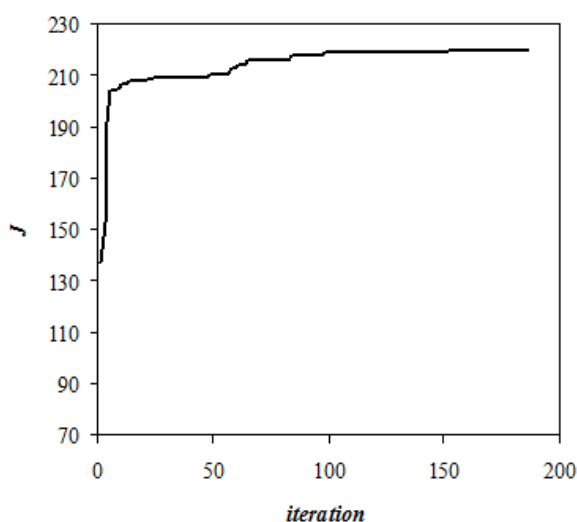


Рисунок 1. Сходимость итерационного процесса

Здесь объем добываемой нефти отнесен к времени T . Достигнутая при оптимизации величина $J/T = 219.8$ означает среднесуточную добычу в кубических метрах. По сравнению с первоначальным значением $J/T = 137$ объем добычи увеличился более, чем на 60%. Найденное расположение скважин, показано на следующем рисунке 2.

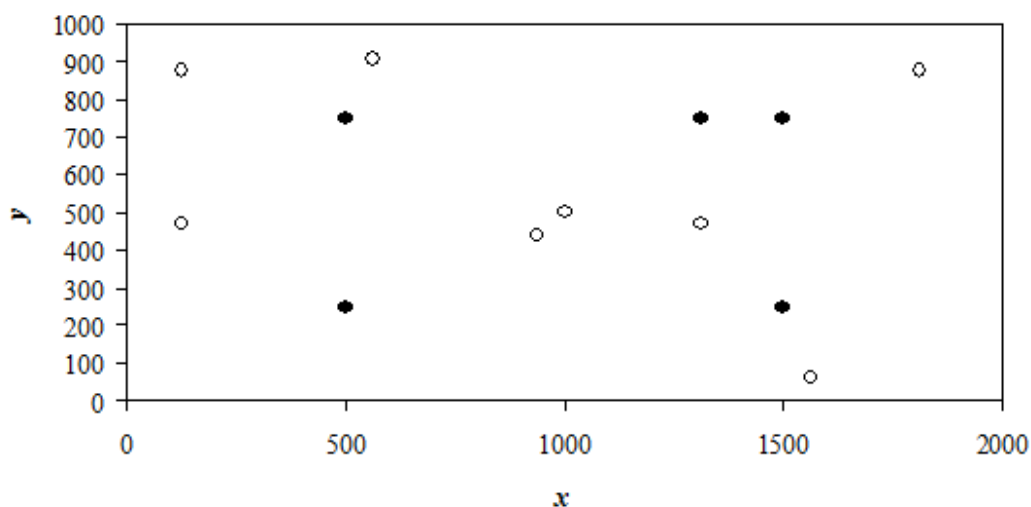


Рисунок 2. Оптимальное расположение скважин

Конечное распределение водонасыщенности представлено на рис. 3, а на рис. 4 показано течение воды в пласте.

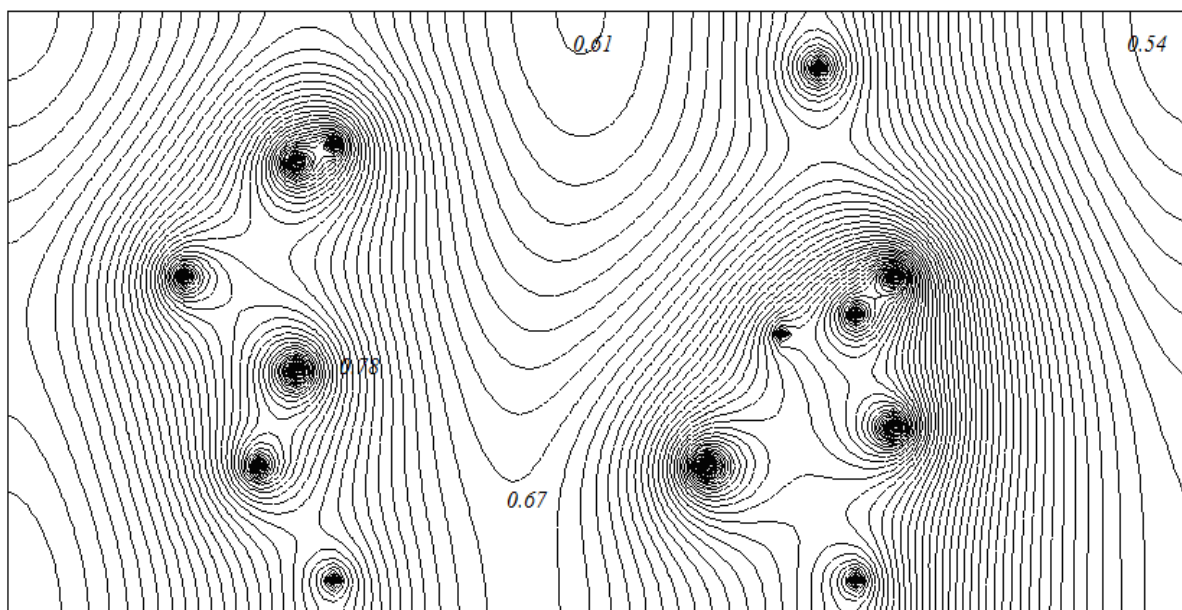


Рисунок 3. Распределение водонасыщенности в конечный момент времени $T=16$ лет

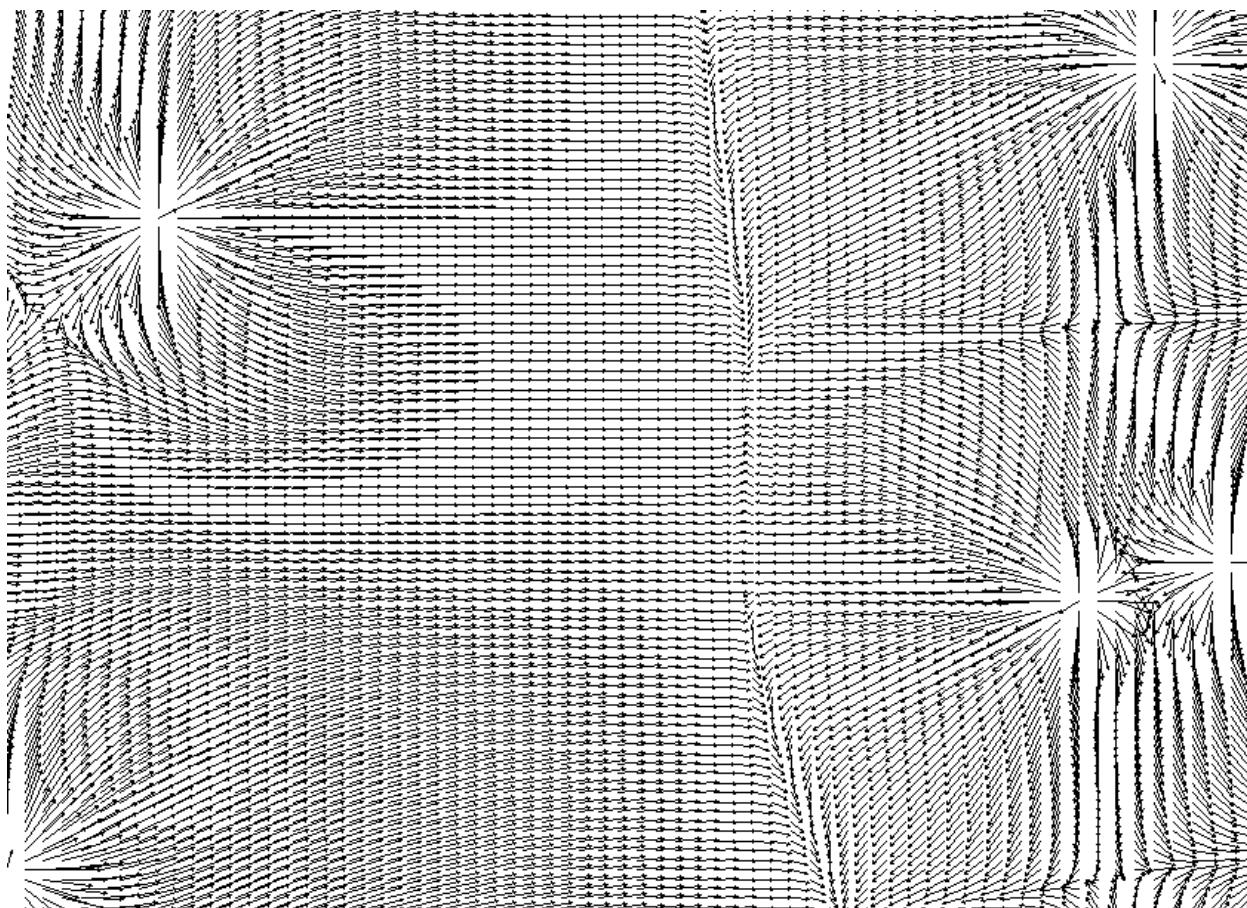


Рисунок 4. Течение воды в пласте (T=16 лет)

Результаты оптимизационных расчетов показывают, что решение задачи оптимального размещения и регулирования работы скважин в постановке нестационарной двухфазной фильтрации, существенно повышает эксплуатационные характеристики нефтяного пласта. Не составляет принципиальных сложностей задания других критериев оптимальности и дополнительных ограничений. Также реальной является и задача трехфазной фильтрации в двумерной и квазитрехмерной постановках.

Библиографический список

1. Азиз Х., Сеттари Э. Математическое моделирование пластовых систем: Пер. с англ. / Под ред. М.М. Максимова. – М.-Ижевск: Институт компьютерных исследований, 2004. – 416 с. Репринтное издание. Оригинальное издание: М.: «Недра», 1982 г.
2. Juanes, R., Displacement theory and multiscale numerical modeling of three-phase flow in porous media, Ph. D. Thesis, University of California, Berkeley, California, 2003.
3. Schiozer, D.J. Simultaneous simulation of reservoir and surface facilities, Ph.D Thesis, Stanford University, 1994. – 172 с.
4. Эрлагер Р мл. Гидродинамические методы исследования скважин. – Москва-Ижевск: Институт компьютерных исследований, 2007. – 512 с.
5. Пыхачев Г.Б., Исаев Р.Г. Подземная гидравлика: Учебное пособие. – М.: «Недра», 1972. – 360 с.

6. Thiele, M.R., Modeling multiphase flow in heterogeneous media using streamtubes, Ph. D. Thesis, Stanford University, Stanford, California, 1994. - 203 p.
7. Wesseling, P., Principles of computational fluid dynamics, Springer, Berlin, 2001. 644 p.

RESULTS OF THE OPTIMIZATION TASKS SOLUTION RELATED TO FILTRATION IN OIL PLATE

Miroshnichenko Sergey S.
Kalashnikov Izhevsk State Technical University
426069, Russia, Izhevsk, Studencheskaya, 7, iite@list.ru

The article presents the results of regulating the operation of oil and gas wells, determining the location of production and injection wells leads to the need to solve various types of optimization problems. We will consider the tasks associated with the operation of a separate oil reservoir. The process of reservoir operation is unsteady, since its characteristics change as oil is extracted and water is pumped.

Key words: oil reservoir, injection well, filtration

УДК 550.832

МЕТОДЫ ЧИСЛЕННОГО РЕШЕНИЯ ЗАДАЧ ОПТИМИЗАЦИИ

Мирошниченко Сергей Сергеевич, Лялин Вадим Евгеньевич
Ижевский государственный технический университет имени М.Т.
Калашникова, 426069, Россия, г. Ижевск, Студенческая, д. 7,
iite@list.ru

Шушков Андрей Александрович
Удмуртский федеральный исследовательский центр УрО РАН,
426067, Ижевск, ул. Татьяны Барамзиной, 34,
ligrim@mail.ru

В статье представлены методы численного решения задач оптимизации процесса эксплуатации пласта, которые является нестационарным, так как его характеристики изменяются по мере извлечения нефти и закачки воды. Основой для решения оптимизационных задач являются математические модели однофазной и двухфазной фильтрации, а также решатели для этих моделей. Поскольку применяемые математические модели содержат дифференциальные уравнения, то задача оптимизации в общем случае является бесконечномерной, то есть задачей оптимального управления.

Ключевые слова: нефтяной пласт, нагнетательная скважина, фильтрация

Выбор численных методов решения оптимизационных задач определяется видом задачи. Для простого класса оптимизационных задач с линейными целевыми функциями и линейными ограничениями (задачи линейного программирования) разработаны методы, позволяющие найти оптимальное решение (или установить неразрешимость задачи) за конечное количество шагов расчета. Введение условий дискретности переменных усложняет нахождение решения. Произвольный вид функций, ограничивает вычислительные возможности существующих методов оптимизации. Самой большой сложностью для решения задач оптимизации является наличие нескольких экстремумов. При получении решения, без дополнительных исследований поведения целевой функции, нельзя однозначно установить к какому экстремуму оно сошлось. Поэтому глобальная оптимизация является пока нерешенной проблемой. Результат применения классических методов оптимизации, как требующих вычисления производных, так и прямых методов сильно зависит от имеющегося начального допустимого приближения. Появившиеся в последние десятилетия методы оптимизации, основанные на генетических алгоритмах, в определенной степени уменьшают эти трудности. В работах [1, 2] достаточно полно и подробно описаны современные подходы и концепции к реализации генетических алгоритмов. Проведен анализ существующих операторов скрещивания, отбора и мутации и предложены некоторые новые. Основное внимание уделено алгоритмам решения задач безусловной оптимизации либо с ограничениями «коридорного» типа. Поэтому будем рассматривать дальнейшее развитие генетического алгоритма, приспособленного для решения задач с ограничениями произвольного вида.

Основой самых первых типов операторов генетического алгоритма было двоичное кодирование признаков. Двоичное представление хромосом влечет за собой определенные трудности при выполнении поиска в непрерывных пространствах, которые связаны с большой размерностью пространства поиска. Это происходит, когда фенотип объекта представляется множеством вещественных чисел, например, при оптимизации сложной функции. Точность представления переменной определяется количеством разрядов, используемых для кодирования одной хромосомы. Поэтому при увеличении разрядности и размерности вектора переменных пространство поиска расширяется и становится огромным. Для решения таких задач в непрерывных пространствах возник новый тип ГА – генетический алгоритм с вещественным кодированием (англ.: Real-Coded Genetic Algorithm, RCGA) [1, 3, 4]. Основная идея RCGA заключается в том, чтобы напрямую представлять гены в виде вещественных чисел, т.е. генотип объекта становится идентичным его фенотипу. Вектор хромосомы состоит из вектора вещественных чисел, и точность найденного решения будет определяться не количеством разрядов для кодирования битовой строки, а будет ограничена возможностями ЭВМ, на которой реализуется вещественный ГА.

Применение вещественного кодирования может повысить точность найденных решений и повысить скорость нахождения глобального минимума или максимума. Скорость повышается из-за отсутствия процессов кодирования и декодирования хромосом на каждом шаге алгоритма. Единственное преобразование, которое целесообразно проводить, это приведение переменных к безразмерному виду в диапазоне $[0;1]$.

Для RCGA стандартные операторы скрещивания и мутации не подходят, т.к. алгоритм работает только с вещественными числами.

Метод сопряженного градиента имеет высокую скорость сходимости к оптимальному решению, близкую к квадратичной. Кроме того, в этом методе не вычисляются матрицы и не хранятся в оперативной памяти, что позволяет применять его для решения оптимизационных задач большой размерности.

Гибридный алгоритм.

- 1) $k=0$. Формируется популяция, состоящая из m особей $\{C^s, s = \overline{1, m}\}^k$ по BGA или RGA методу. Первый номер принимает особь C^1 с лучшим показателем (минимальным значением функции). С помощью преобразования ψ^{-1} получаем вектор $\mathbf{x}_b^k$ и $\mathbf{x}^k = \mathbf{x}_b^k$.
- 2) $k=k+1$. С помощью метода МСГ вычисляется следующее приближение вектора $\mathbf{x}^k$. Генетическим алгоритмом создается следующая популяция $\{C^s, s = \overline{1, m}\}^k$ и находится лучшая особь, определяющая очередной вектор $\mathbf{x}_b^k$.
- 3) Если $F(\mathbf{x}_b^k) < F(\mathbf{x}^k)$, то $\mathbf{x}^k = \mathbf{x}_b^k$.
- 4) Если $F(\mathbf{x}_b^k) \geq F(\mathbf{x}^k)$, то $C^1 = [\mathbf{x}^k, \psi]$.
- 5) Если выполняется условие остановки, то *конец* иначе на пункт 2.

В качестве дополнительного метода целесообразно также применить генетический алгоритм, но в более узкой области поиска. Границы узкой области поиска определяются следующим образом. Если $\mathbf{x} = (x'_i), i = \overline{1, N}$ - лучшая в популяции особь, то границы области дополнительного поиска равны:

$$A'_i = x'_i - \Delta_i, B'_i = x'_i + \Delta_i; \quad \Delta_i = \delta(B_i - A_i).$$

Результаты тестов показали, что величина $\delta \in (10^{-4}, 10^{-3})$.

Для решения оптимизационных задач с ограничениями не существует какого-либо универсального метода, который бы давал одинаково хорошие результаты не зависимо от вида целевой функции и функций ограничений. Как правило, задачу условной минимизации сводят к задаче безусловной минимизаций путем применения штрафных функций, после чего находят решение, применяя генетические алгоритмы или градиентные методы. При использовании штрафных функций возникает проблема определения весовых коэффициентов, а сама функция часто имеет овражный характер. Разные целевые функции и ограничения требуют тщательного подбора коэффициентов, что далеко не всегда удается сделать успешно, а при

изменении условия исходной задачи возникает необходимость в новом подборе.

Достаточно универсальными являются два подхода, основанные на операциях селекции и формирования новых популяций. В одном подходе [2, 5] используется идея параллельных популяций с допустимыми и недопустимыми особями. В другом [6] также к скрещиванию привлекаются недопустимые особи с селекцией по доминированию. Будем применять для численного решения задачи с ограничениями базовый генетический алгоритм, описанный в [1] с дополнительной селекцией, изложенной в [6]. Генетический алгоритм из [1] для скрещивания применяет операторы BLX, B1, B2 и классический оператор Холланда. Эти операторы выбираются при каждой процедуре скрещивания случайным образом, что расширяет их возможности.

При селекции применяется турнирный метод отбора с выполнением следующих правил:

1. Если обе сравниваемые особи допустимые, то есть удовлетворяющие ограничениям, то выбирается с лучшим показателем целевой функции.
2. Если одна особь допустимая, а вторая нет, то выбирается допустимая.
3. Если обе особи недопустимые, то выбирается та, у которой меньшее количество невыполненных ограничений.
4. Если обе особи недопустимые и имеют одинаковое количество нарушенных ограничений, то выбирается та, у которой минимальное нарушение ограничений.

Эти правила легко алгоритмически реализуются и не требуют существенной переработки базового генетического алгоритма.

Результаты решения тестовых задач показали, что предлагаемый алгоритм справляется со всеми задачами, рассмотренными в работах [6, 7] при произвольном виде ограничений и целевых функций. Этот алгоритм применяется далее для решения задач по оптимальной эксплуатации нефтяных скважин.

Библиографический список

1. Thompson J. F., Thames F. C., Mastin C. W. Automatic numerical generation of body-fitted curvilinear coordinate system for field containing any number of arbitrary two-dimensional bodies // J. Comp. Phys.- 1974. V.15. P.299-319.
2. Кеннон С. Р., Дуликравич Дж. С. Построение сеток с помощью оптимизационного метода // Аэро-косм. техн.- 1987. N1. С.107-112.
3. Шридхоф, Дейвис. Метод Шварца-Кристоффеля для построения координатных сеток для двухмерных течений // Теор. осн. инж. расчетов.- 1985. N3. С.145-158.
4. Громадка П Т., Лей Ч. Комплексный метод граничных элементов в прикладных науках. -М: Мир, 1990. 303с.

5. Погорелов А.В. Дифференциальная геометрия. М.: Наука, 1974. с.178.
6. Корн Г.А., Корн Т.М. Справочник по математике для научных работников и инженеров. — М.: «Наука», 1974. — 832.
7. Голуб Дж., Ван Лоун Ч. Матричные вычисления: Пер. с англ. — М.: Мир, 1999. — 548 с.

METHODS OF OPTIMIZATION PROBLEMS NUMERICAL SOLUTION

Miroshnichenko Sergey S., Lyalin Vadim E.
Kalashnikov Izhevsk State Technical University
426069, Russia, Izhevsk, Studencheskaya, 7, iite@list.ru

The article presents methods for the numerical solution of problems of optimizing the process of reservoir operation, which is non-stationary, since its characteristics change as oil is extracted and water is pumped. The basis for solving optimization problems are mathematical models of single-phase and two-phase filtration, as well as solvers for these models. Since applied mathematical models contain differential equations, the optimization problem in the general case is infinite-dimensional, that is, the problem of optimal control.

Key words: oil reservoir, injection well, filtration.

УДК 004.89; 519.6; 004.421

НЕЙРОСЕТЕВОЕ МОДЕЛИРОВАНИЕ ПРОГНОЗИРОВАНИЯ ВЕРОЯТНОСТИ ВОЗНИКНОВЕНИЯ ПЕРКОЛЯЦИОННОГО КЛАСТЕРА

*Никитин Илья Владимирович, Решетникова Елизавета Алексеевна,
Бузмакова Мария Михайловна*

Пермский государственный национальный исследовательский университет,
614990, Россия, г. Пермь, ул. Букирева, 15,
nikitin.w.95@gmail.com

В настоящей работе авторами описан опыт разработки и применения нейросетевой модели для прогнозирования вероятности возникновения перколяционного кластера и выявления новых знаний в этой области. Найденные методом математического моделирования закономерности и сделанные из них выводы нацелены на то, чтобы помочь исследователям избежать неудачных экспериментов, в которых мала вероятность возникновения перколяционного кластера. Такая оптимизация ведёт к значительной экономии времени и денежных средств, так как наноматериалы имеют высокую стоимость.

Ключевые слова: математическое и компьютерное моделирование, полимер, углеродные нанотрубки, перколяционный кластер, прогнозирование, вероятность, нейронная сеть.

Введение

В последнее время всё больший интерес у учёных вызывают композиционные наноматериалы, такие как полимеры, модифицированные углеродными нанотрубками. Дело в том, что нанокомпозитам можно задать необходимые свойства путем введения различных концентраций нанонаполнителей и использования некоторых управляющих факторов. Таким образом, с помощью нанокомпозитов исследователи решают актуальные и важные задачи современной экономики и промышленности.

Для описания структуры и свойств полимерных нанокомпозитов как правило используются перколяционные модели [1], так как свойства наноматериала при изменении концентрации наполнителя меняются скачкообразно, а такое поведение принято называть перколяционным.

Прогнозирование вероятности возникновения перколяционного кластера необходимо для того, чтобы сделать вывод достигается ли порог перколяции, и имеет большую практическую значимость в плане выявления параметров структуры наноматериала, оказывающих сильное влияние на вероятность возникновения перколяционного кластера. Порог перколяции соответствует значению концентрации углеродных нанотрубок в полимере, при которой предполагается значительное изменение имеющихся или появление новых свойств наноматериала. Проанализировав полученные результаты, исследователь может избежать неудачных экспериментов, в которых перколяционный кластер не возник, что приведёт к значительной экономии времени и денежных средств, так как наноматериалы имеют высокую стоимость.

Методика прогнозирования

Для прогнозирования вероятности возникновения перколяционного кластера необходимо спроектировать нейросетевую модель, содержащую определенные входные и выходные параметры. После изучения и анализа источника [2], авторами настоящей работы было выделено некоторое множество входных параметров модели. Ниже приведены сами параметры и способ их кодировки:

x_1 – линейный размер куба, то есть размер полимера, который модифицируют нанотрубками (может быть любым, в настоящей работе рассматривался размер 1000 мм); x_2 – математическое ожидание длины нанотрубки; x_3 – дисперсия длины нанотрубки; x_4 – радиус нанотрубки; x_5 – концентрация нанотрубок в полимере; x_6 – максимальная дистанция, которая может быть между элементами для того чтобы их отнести к одному кластеру; x_7 – количество кластеров, возникших в испытании; x_8 – количество перколяционных кластеров, возникших в испытаниях; Выходным параметром – y является вероятности возникновения перколяционного кластера.

Для прогнозирования все имеющиеся данные были разделены на два множества: обучающее (559 примеров) и тестирующее (140 примеров), причём, примеры, входящие в тестирующее множество, не повторяются в обучающем.

Для проектирования, обучения и тестирования нейросети был выбран нейросимулятор Nsim5-10s, разработанный в Пермской научной школе искусственного интеллекта (www.PermAi.ru) [3-4].

Персептрон, построенный для прогнозирования вероятности возникновения перколяционного кластера, имеет следующую структуру: восемь входных параметров и один выходной. Так же введён один скрытый слой. Количество нейронов на скрытом слое было найдено экспериментально. Определяющим фактором успешно построенной модели служит ошибка тестирования нейросети [3]. Исходя из этого, после обучения и тестирования нейросети с использованием различного количества скрытых слоев и нейронов на скрытом слое был построен оптимальный персептрон с тремя нейронами на скрытом слое. Он изображен на рисунке 1. При такой структуре персептрона ошибка обучения составила 1,53%.

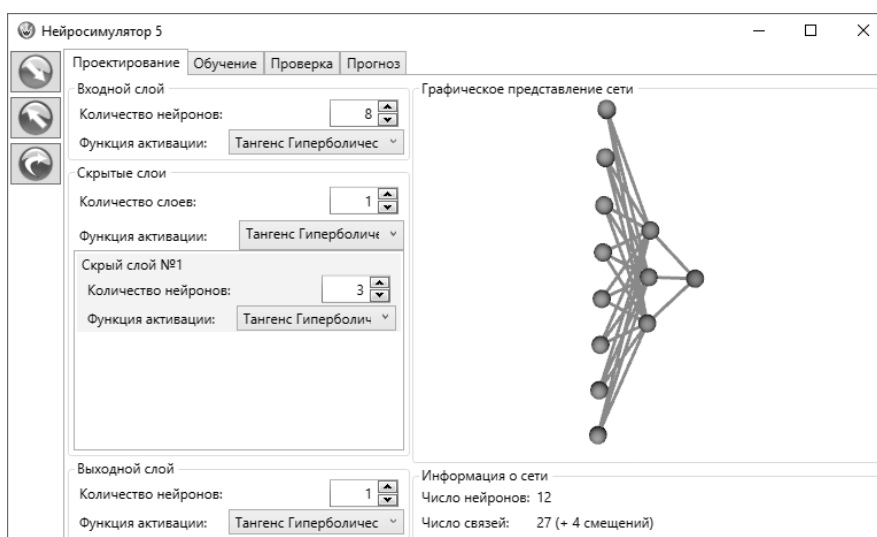


Рисунок 1. Структура персептрона

После обучения нейросети было проведено тестирование, которое служит для определения прогностических свойств обученной сети. Нейросеть была проверена на тестирующем множестве, примеры которого отличны от примеров обучающего множества. Ошибка тестирования составила 0,97%, что можно увидеть на графике (рисунок 2). Малый размер ошибки говорит об адекватности и закономерности полученных результатов.

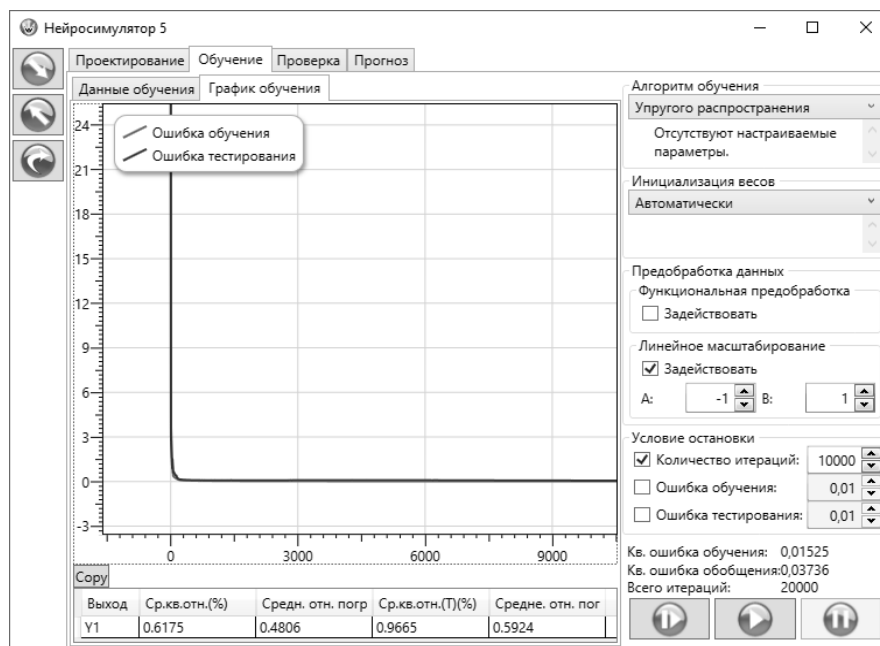


Рисунок 2. Графики обучения нейросети

Результаты вычислительных экспериментов и их обсуждение

Адекватность работы, обученной нейросети, была проверена с помощью тестирующего множества, в которое вошли примеры не дублирующиеся ни в одном из ранее упомянутых множеств. После получения результатов был построен график, из которого видно разницу между реальным и прогнозным значением вероятности (см. рисунок 3). По графику можно сделать вывод, что нейросетевая система прогнозирует предельно точно.

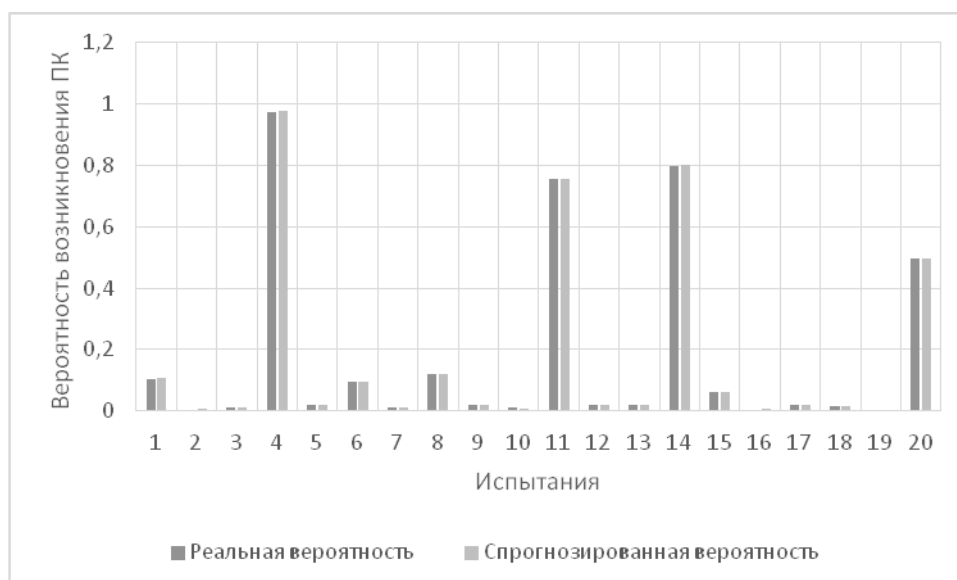


Рисунок 3. Реальное и прогнозное значение вероятности

С помощью нейросимулятора можно определить значимость параметров, вошедших в модель. На рисунке 4 изображена диаграмма, построенная нейросимулятором. Нетрудно заметить, что параметр

отвечающий за количество перколяционных кластеров, возникших в испытаниях, больше всего влияет на вероятность, что вполне обосновано. Вторым по значимости оказалось математическое ожидание длины нанотрубки. Из чего можно сделать вывод, что добиться появления в наноматериале перколяционного кластера проще всего путём изменения длины нанотрубок.

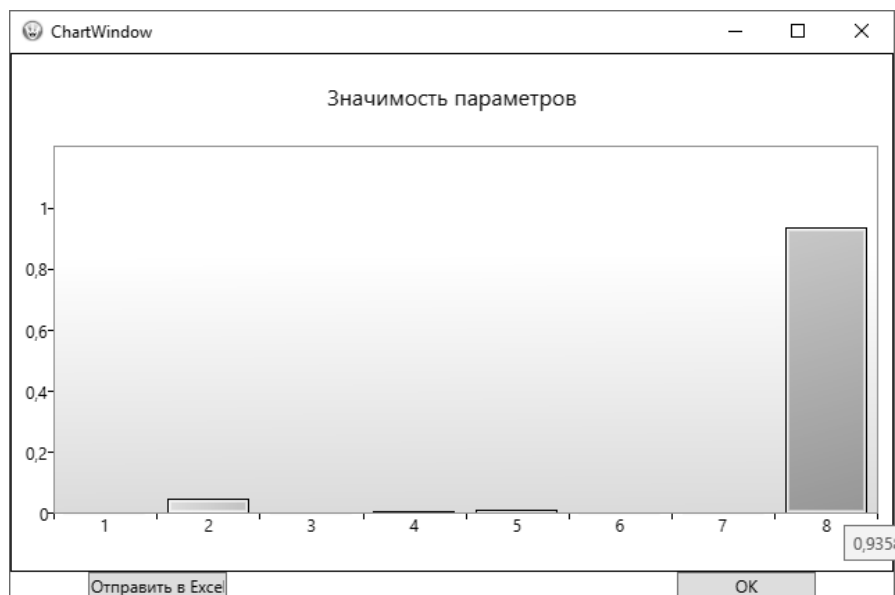


Рисунок 4. Оценка значимости параметров

В дальнейшем авторами планируется изменение количества входных параметров нейросетевой модели, оценивание значимости параметров без учёта восьмого параметра, вошедшего в рассматриваемую модель. Кроме того, планируется произвести более подробную оценку влияния наиболее значимых параметров на вероятность возникновения перколяционного кластера.

Библиографический список

1. Silva J., Ribeiro S., Lanceros-Mendez S., Simões R. The influence of matrix mediated hopping conductivity, filler concentration, aspect ratio and orientation on the electrical response of carbon nanotube/polymer nanocomposites // Composites Science and Technology. 2011. V. 71. P. 643646.
2. Бузмакова М.М. Компьютерное моделирование континуальной перколяции сфер и эллипсоидов с проницаемыми оболочками: дис. канд. ф.-м. наук. Астраханский гос. университет, Астрахань, 2013.
3. Ясницкий Л.Н., Бондарь В.В. и др. Пермская научная школа искусственного интеллекта и ее инновационные проекты. 2-е изд. М.; Ижевск: НИЦ «Регулярная и хаотическая динамика», 2008. 75 с.
4. Нейросимулятор 5.0. Свидетельство Роспатент о государственной регистрации программы для ЭВМ № 2014618208 от 12.07.2014 г. / Черепанов Ф.М., Ясницкий Л.Н. (РФ)

NEURAL NETWORK MODELING OF FORECASTING THE PROBABILITY OF THE OCCURANCE OF A PERCOLATION CLUSTER

Nikitin Ilya V., Reshetnikova Elizaveta A., Buzmakova Maria M
Perm State University
614990, Russia, Perm, street Bukireva, 15,
nikitin.w.95@gmail.com

In this paper, the authors described the experience of developing and applying a neural network model to predict the probability of a percolation cluster and identify new knowledge in this area. The regularities found by the method of mathematical modeling and the conclusions drawn from them are aimed at helping researchers avoid unsuccessful experiments in which the probability of a percolation cluster is low. This optimization leads to significant savings in time and money, since nanomaterials are expensive.

Key words: the mathematical and computer modeling, the polymer, the carbon nanotubes, the percolation cluster, the prediction, the probability, the neural network.

УДК 004.89; 519.6; 004.421

ПРОГНОЗИРОВАНИЕ КОНЦЕНТРАЦИИ К-МЕРОВ, НЕОБХОДИМОЙ ДЛЯ ДОСТИЖЕНИЯ ПОРОГА ПЕРКОЛЯЦИИ С ПОМОЩЬЮ НЕЙРОСЕТЕВОЙ СИСТЕМЫ

*Решетникова Елизавета Алексеевна, Никитин Илья Владимирович,
Бузмакова Мария Михайловна*

Пермский государственный национальный исследовательский университет
614990, Россия, г. Пермь, ул. Букирева, 15,
lizica1995@mail.ru

В статье описана методика разработки и применения нейросетевой системы прогнозирования концентрации k -меров, необходимой для достижения порога перколяции, и выявления новых знаний в области теории перколяции. k -мер представляет собой отрезок от начальной точки (x_i, y_i, z_i) до точки, расположенной на расстоянии l от неё в направлении, заданном углом наклона и углом поворота (φ_i, ψ_i) . Методом математического моделирования в рамках настоящей работы были найдены концентрации k -меров для моделей с различными параметрами, необходимые для достижения порога

перколяции. Сделанные в процессе исследования выводы позволяют проводить только испытания, представляющие интерес для науки.

Ключевые слова: математическое и компьютерное моделирование, k -мер, концентрация, порог перколяции, перколяционный кластер, прогнозирование, вероятность, нейронная сеть.

Теория перколяции используется в математике, физике, химии и других областях науки для описания образования связанных структур (кластеров), состоящих из отдельных элементов. В настоящей работе в качестве отдельных элементов используются k -меры. Перколяционным называется кластер, соединяющий две противоположные стороны системы [1]. В последнее время теория перколяции находится на стадии активного развития и актуальной задачей является изменение свойств структуры, путём добавления в неё такого количества элементов, которое необходимо для достижения порога перколяции.

В настоящей работе предложена нейросетевая модель для прогнозирования концентрации k -меров, необходимой для достижения порога перколяции. С помощью предложенной модели исследователь может найти необходимую концентрацию элементов, не проводя многочисленных экспериментов в лаборатории. Ему достаточно ввести параметры своего эксперимента в нейросетевую систему, и последняя выдаст ему точную концентрацию элементов, которые необходимо внедрить в структуру для изменения её свойств.

Чтобы спрогнозировать концентрацию k -меров, необходимую для достижения порога перколяции была спроектирована нейросетевая модель. Она содержит входные и выходные параметры, которые были выделены после изучения соответствующей литературы [1-2]. Параметры были закодированы следующим образом:

X_1 – линейный размер куба, в который исследователь внедряет k -меры.

X_2 – математическое ожидание длины k -мера.

X_3 – дисперсия длины k -мера (в кубе могут размещаться k -меры длины от $X_2 - X_3$ до $X_2 + X_3$).

X_4 – показатель, меньше которого должно быть расстояние между k -мерами чтобы отнести их к одному кластеру.

X_5 – количество кластеров, возникших в испытании.

X_6 – вероятность возникновения перколяционного кластера (от 0,50 до 1).

Y – концентрация k -меров, необходимая для возникновения перколяционного кластера (Выходной параметр).

Имеющаяся выборка данных об испытаниях, в которых достигался порог перколяции, была перемешана. Все данные были разделены по принципу Парето на два множества: обучающее (178 испытаний) и тестирующее (45 испытаний). Данные, входившие в тестирующее множество, не повторялись в обучающем для получения достоверных результатов исследования.

Проектирование, обучение и тестирование нейросети были проведены в нейросимуляторе Nsim5-10s, разработанном в Пермской научной школе искусственного интеллекта (www.PermAi.ru) [3-5].

Структура персептрона, построенного авторами настоящей работы имеет следующий вид: шесть входных параметров и один выходной. Кроме того, введён скрытый слой. Количество нейронов на скрытом слое было определено экспериментально.

Показателем успешно построенной модели является ошибка тестирования нейросети [3]. Поэтому, после обучения и тестирования нейросети с использованием различного количества скрытых слоев и нейронов на скрытом слое был построен оптимальный персептрон с четырьмя нейронами на скрытом слое. Структура персептрона приведена на рисунке 1. Ошибка обучения составила 1,9%.

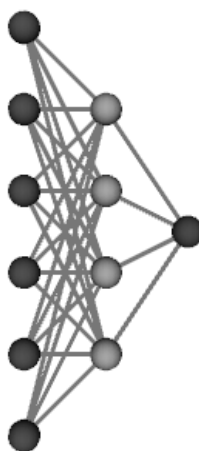


Рисунок 1. Структура персептрона

После обучения нейросети следует этап тестирования. С помощью тестирования можно определить прогностические свойства нейросети. Ошибка тестирования составила 2,8%, из этого был сделан вывод, что полученные с помощью нейросетевой модели результаты адекватны.

Для доказательства адекватности предложенной нейросетевой модели была произведена её проверка на специальном множестве. В него вошло 45 примеров, не дублирующих ни в одном из множеств, на которых производилось обучение и тестирование нейросети. Результаты проверки изображены на рисунке 2.

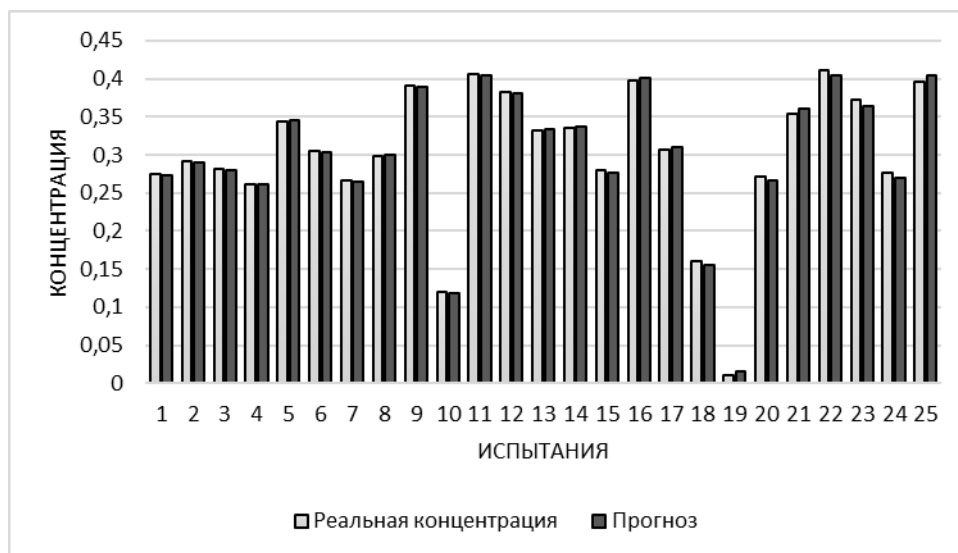


Рисунок 2. Разница между реальной концентрацией и спрогнозированной

Судя по графику, можно сделать вывод, что обученная нейросеть позволяет получать достаточно близкие к реальным данным значения. Этот факт свидетельствует об адекватности предложенной нейросетевой модели.

Для дальнейшего исследования был определен вклад, который вносит каждый параметр в прогнозирование интересующей нас концентрации к-меров. Такая возможность есть в нейросимуляторе Nsim5-10s. Результаты оценки значимости каждого из параметров предложенной модели приведены на рисунке 3.

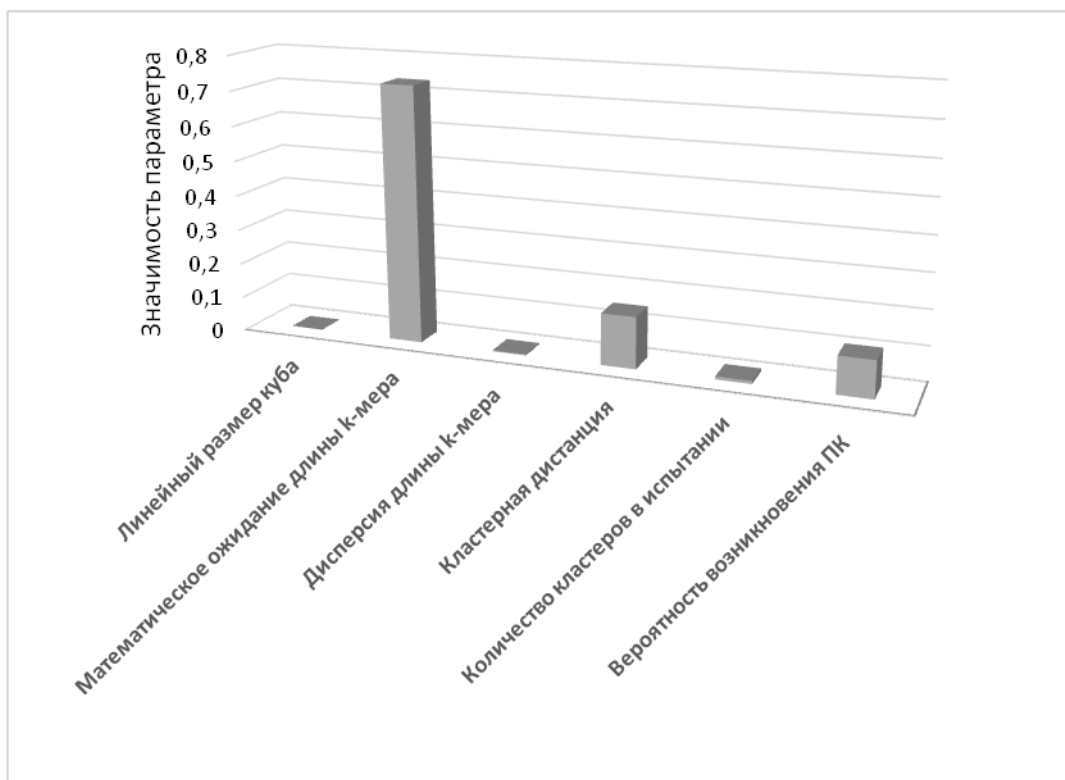


Рисунок 3. Оценка значимость параметров

Наиболее значимым оказался параметр, отвечающий за среднюю длину k -мера. То есть, концентрация k -меров будет значительно больше всего изменяться при изменении их длин при прочих одинаковых параметрах. Чтобы доказать это было выбрано четыре разных значения дисперсии длины k -мера: 0 (длина k -меров постоянна), 20, 40 и 60. Математическое ожидание длины k -мера при этом менялось в пределах от 100 до 350 с шагом 50. Был получен закономерный результат, свидетельствующий о том, что при увеличении средней длины k -меров концентрация увеличивается (см. рисунок 4). Самая значительная динамика в изменении концентрации была зафиксирована для случая, когда дисперсия равна 0.

Таким образом, в рамках настоящей работы авторами была предложена нейросетевая модель прогнозирования концентрации k -меров, необходимой для достижения порога перколяции, построена, обучена и протестирована нейронная сеть, выявлен и подробно исследован самый значимый параметр нейросетевой модели, сделаны значимые для теории перколяции выводы. Наиболее важным результатом настоящей работы является нейросеть, которая позволит исследователям экономить временные и денежные ресурсы, проводя испытания с заранее известной предсказанной концентрацией k -меров.

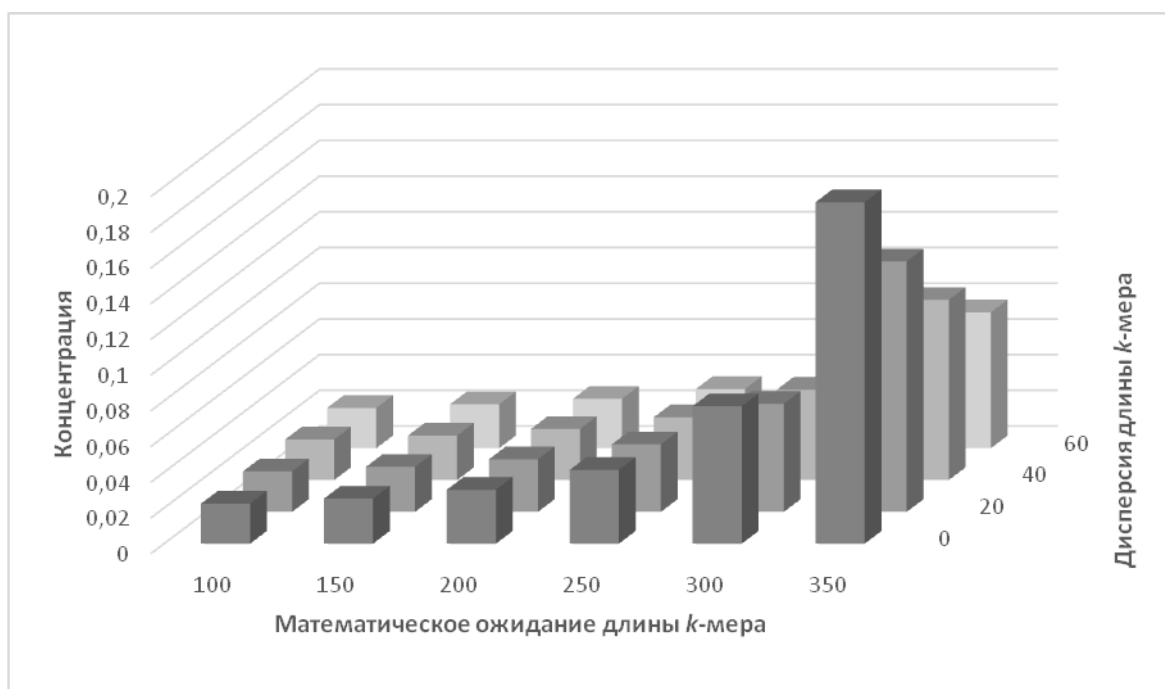


Рисунок 4. Зависимость концентрации от математического ожидания длины k -мера

Библиографический список

1. Тарасевич Ю.Ю. Перколяция: теория, приложения, алгоритмы. Учебное пособие. М.: Едиториал УРСС, 2002. - 112 с.
2. Бузмакова М.М. Компьютерное моделирование континуальной перколяции сфер и эллипсоидов с проницаемыми оболочками: дис. канд. ф.-м. наук. Астраханский гос. университет, Астрахань, 2013.

3. Ясницкий Л.Н., Бондарь В.В. и др. Пермская научная школа искусственного интеллекта и ее инновационные проекты. 2-е изд. М.: Ижевск: НИЦ «Регулярная и хаотическая динамика», 2008. - 75 с.
4. Ясницкий Л.Н., Данилевич Т.В. Современные проблемы науки. М.: БИНОМ. Лаборатория знаний, 2008. - 294 с.
5. Нейросимулятор 5.0 : Свидетельство Роспатент о государственной регистрации программы для ЭВМ № 2014618208 от 12.07.2014 г. / Черепанов Ф.М., Ясницкий Л.Н. (РФ).

FORECASTING THE K-MEARS CONCENTRATION FOR REACHING PERCOLATION THRESHOLD USING THE NEURAL NETWORK SYSTEM

Reshetnikova Elizaveta A., Nikitin Ilya V., Buzmakova Maria M
Perm State University
614990, Russia, Perm, street Bukireva, 15,
lizica1995@mail.ru

The article describes the methodology for the development and application of the neural network system for predicting the concentration of k -meres necessary to reach the percolation threshold and to reveal new knowledge in the field of percolation theory. A k -mer is a segment from the starting point (x_i, y_i, z_i) to a point located at a distance l from it in the direction given by the angle of inclination and angle of rotation (φ_i, ψ_i) . Using the method of mathematical modeling, the concentrations of k -meres for models with different parameters were found necessary to achieve the percolation threshold. The findings made in the research process allow only tests of interest to science.

Key words: the mathematical and computer modeling, the k -mer, the concentration, the percolation threshold, the percolation cluster, the prediction, the probability, the neural network.

ФОРМАЛИЗАЦИЯ ЗАДАЧИ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПРИ ОБСЛУЖИВАНИИ ТЕХНОЛОГИЧЕСКОГО ОБОРУДОВАНИЯ НЕФТЕДОБЫЧИ

Селетков Илья Павлович

Пермский государственный национальный исследовательский университет,
Физический факультет.
614990, Россия, г. Пермь, ул. Букирева, 15,
iseletkov@gmail.com

В статье описаны предметная область процесса обслуживания технологического оборудования добычи нефти, потоки данных, выбор модели и построение интеллектуальной системы поддержки принятия решений оператором при обслуживании скважин. Показано, что в существующих условиях одним из наиболее простых методов решения этой задачи является использование нечёткой логики.

Ключевые слова: обслуживание оборудования добычи нефти, система поддержки принятия решений, нечёткая логика, нейронные сети.

В настоящее время с целью реализации принятой в России концепции цифровизации экономики во многих отраслях промышленности внедряются современные системы автоматизированного управления производством и системы поддержки принятия решений (СППР) при управлении технологическими процессами. В сфере добычи углеводородного сырья такие системы высоко востребованы в процессах обслуживания и управления нефтегазодобывающими скважинами, нефтепроводами и другими технологическими объектами.

В процессе добычи нефти используется большое количество информационных систем, данные из которых надо учитывать при обслуживании технологического оборудования:

- гидродинамические модели месторождений и цифровые модели добывающих скважин позволяют рассчитать оптимальный режим работы скважин;
- системы автоматизированного управления (АСУТП) позволяют удалённо собирать текущие значения параметров работы оборудования и изменять режимы работы;
- специализированные хранилища данных позволяют хранить историю изменения значений параметров большого количества установок в

течении длительного промежутка времени, быстро обращаться к необходимым данным;

- системы планирования ремонтов (ТОРО) позволяют хранить историю ремонтов оборудования, планировать будущие ремонты и закупку необходимых комплектующих.

Другой особенностью процесса обслуживания нефтедобывающих скважин является высокая территориальная распределённость оборудования. Требование принятия решений по обслуживанию далёких от центральной инфраструктуры объектов, где зачастую отсутствуют все дешёвые виды связи, влечёт необходимость реализации СППР в виде мобильного приложения для операторов-обходчиков с возможностью работы без соединения с серверами. При этом вся работа со справочниками и интеграция с другими информационными системами должна осуществляться централизованно, поэтому для СППР требуется специальный сервер.

Таким образом, диаграмма укрупнённых потоков данных в такой СППР может выглядеть следующим образом:

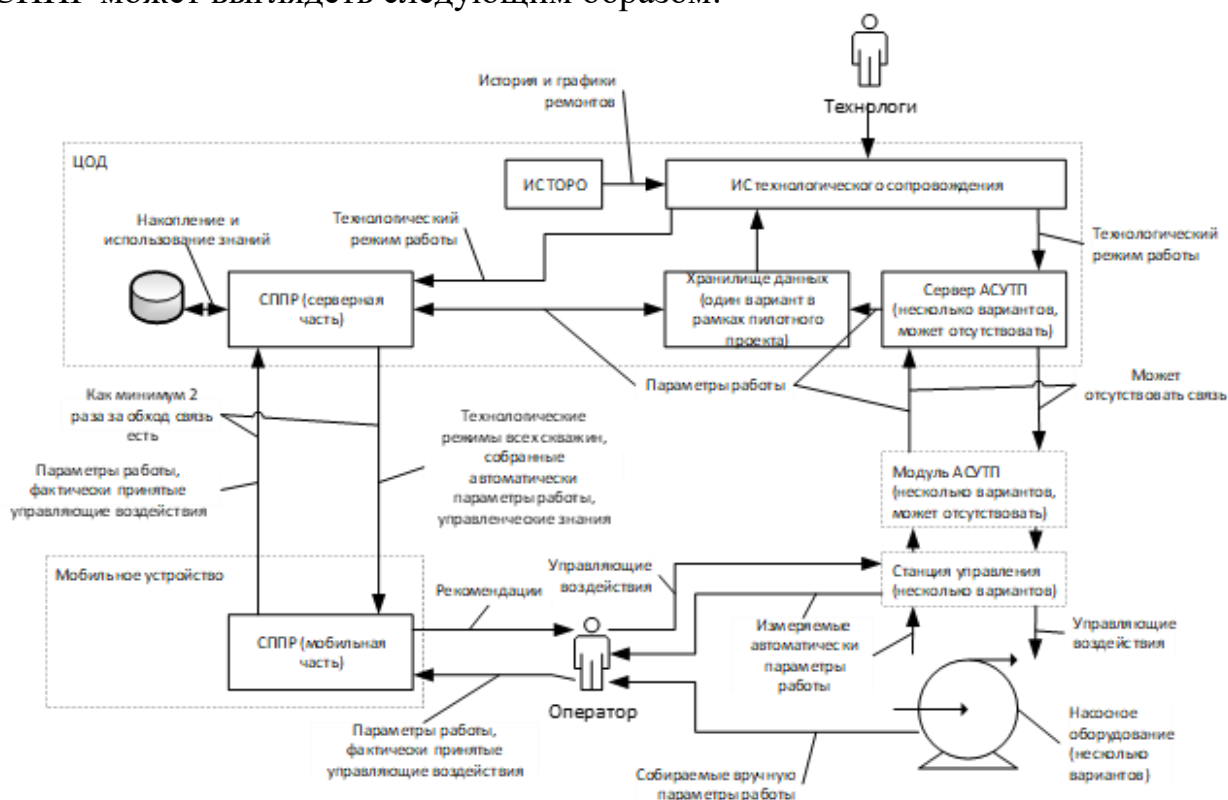


Рисунок 1. Ландшафт информационных систем и потоки данных в процессе обслуживания технологического оборудования добычи нефти

Наиболее массовыми видами применяемого насосного оборудования являются центробежный (ЭЦН) и шаговый (ШГН) насосы [1]. Основными параметрами работы этих насосов являются: нагрузка, загрузка, частота двигателя; буферное, линейное, затрубное давления; температура, сопротивление изоляции, вибрация, расход жидкости. Часть параметров может быть измерена автоматически и оцифрована, часть собирается вручную.

Наиболее частыми операциями по обслуживанию являются: скребкование, опрессовка труб, встряхивание изменением частоты и вращения, пропарка и ревизия нефтепровода и измерительных устройств, чистка и замена штуцера, поиск прорывов в нефтепроводе. Часть операций не может быть выполнена операторами лично и требует вызова специальных сервисных бригад.

Основные входные и выходные информационные потоки при расчёте рекомендаций показаны на схеме:

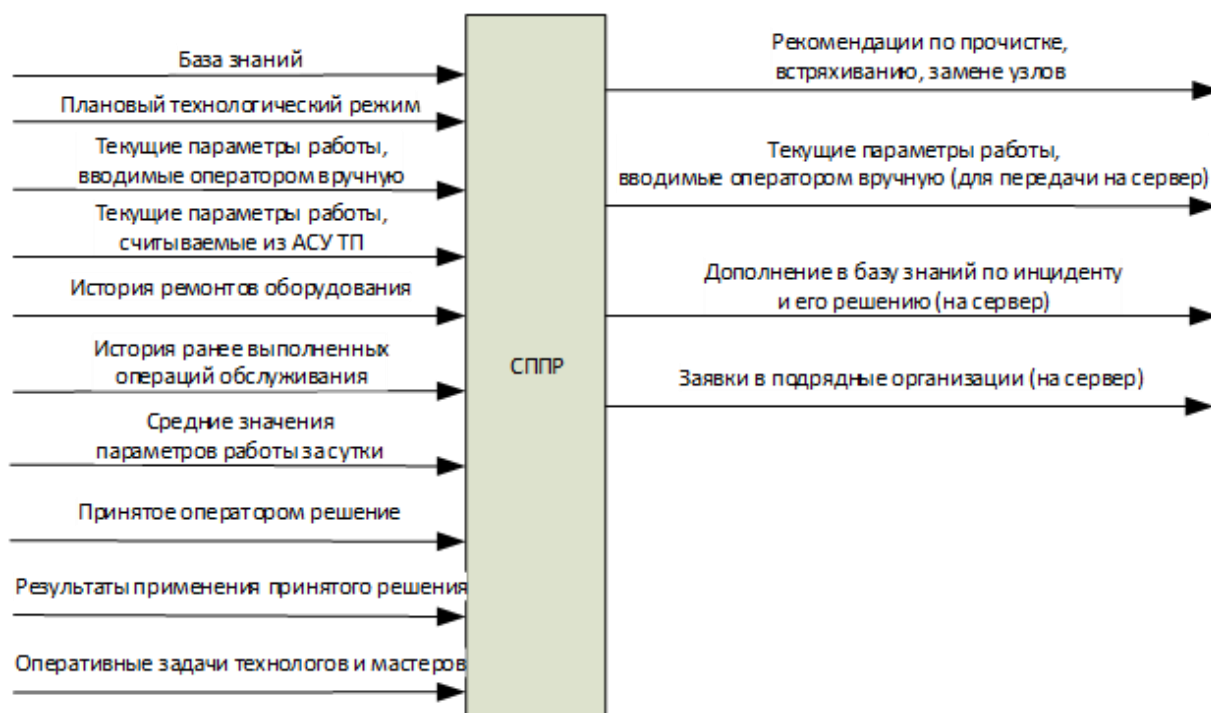


Рисунок 2. Описание СППР в нотации «чёрный ящик»

Построение автоматизированной интеллектуальной системы поддержки принятия решений с учётом этих требований может быть осуществлено, например, с использованием нейронных сетей, лесов деревьев решений, нечёткой логики, или алгоритмов классификации. Для обучения этих моделей необходимо в цифровом виде получить базу знаний или обучающую выборку, в которой должна быть зафиксирована связь между значениями параметров работы оборудования и совершёнными ранее или необходимыми действиями операторов по обслуживанию технологического оборудования. Источниками таких данных могут стать существующие информационные системы, при наличии доступа к их архивам:

- В системах ТОРО хранятся фактические данные по выполнению ремонтных работ, но не всех возможных операций обслуживания.
- Гидродинамические модели месторождений и модели скважин позволяют рассчитать оптимальный плановый режим работы (частоту двигателя, глубину спуска насоса), и получить его историю.

- Системы АСУТП и хранилища данных при наличии связи позволяют получить архив значений измеряемых и оцифрованных параметров работы скважин.

Но история значений части параметров работы, которые не поддаются оцифровке, или автоматический сбор которых на текущий момент является не целесообразным, а также фактически выполненные действия операторов, не связанные с ремонтами, фиксируются посуточно только в бумажных журналах.

Оцифровка достаточного количества журналов для подготовки репрезентативной обучающей выборки может оказаться весьма трудоёмкой задачей.

Генерация синтетической обучающей выборки потребует одновременно частичную оцифровку некоторых журналов, знания технологов, разработку вспомогательного генератора и верификацию полученных данных.

Таким образом, решение поставленной задачи с учётом обозначенных ограничений с помощью нейронных сетей или деревьев решений, которые требуют для обучения числовую оцифрованную выборку, может оказаться трудоёмким.

В случае, когда принятие решений по обслуживанию оборудования осуществляется людьми-экспертами из службы технологов, наименее ресурсозатратным способом получения знаний может быть интервью технологов. Результатом такого интервью является база знаний, состоящая из ограниченного набора правил виде «Если ..., то ...», которые легко могут быть обработаны с помощью аппарата нечёткой логики. Однако, полученная и этим способом база знаний тоже требует верификации.

Преимуществами использования нечёткой логики для решения поставленной задачи являются:

- малые трудозатраты на получение знаний;
- обоснование явным образом получаемых результатов (какие конкретно правила привели к получению результата);
- возможность проверки полноты БЗ на уровне лингвистических правил.

Трудностями использования нечёткой логики являются:

- построение объективных функций принадлежности/истинности;
- выбор подходящих моделей логических операций;
- проверка полноты и непротиворечивости БЗ на уровне функций истинности;
- автоматическая корректировка знаний в процессе эксплуатации системы.

Автоматизированная корректировка и объективизация функций истинности может быть осуществлена как на этапе построения системы, так и в процессе эксплуатации при условии однозначного обратного соответствия значений входных параметров получаемым результатам. Таким

свойством не обладают, например, операции минимума и максимума, которые очень часто используются при реализации нечётких алгоритмов.

Применение матричного аппарата нечёткой логики [2, 3] позволит решить часть трудностей. В частности, проблема выбора подходящих моделей нечётких логических операций исключена за счёт использования математически обоснованных операций. Проверка полноты и непротиворечивости базы знаний может быть произведена автоматически с использованием свойств нечёткого логического вывода. Объективизация (и дообучение) изначально субъективных функций истинности может быть осуществлена по мере поступления новых данных в процессе эксплуатации: все используемые модели логических операций и нечёткого вывода обладают свойством однозначного обратного соответствия значений входных параметров получаемым результатам.

В дальнейшем, по мере поступления оцифрованных данных в ходе работы системы, можно собирать обучающую выборку для тренировки и обучения нейронных сетей и деревьев решений. Когда данных окажется достаточно, для формирования более объективных рекомендаций оператору имеет смысл использовать результаты работы сразу всех интеллектуальных моделей – либо с помощью усреднения результата, либо с помощью классификатора второго уровня, который будет выбирать, результат какой из моделей использовать в каждом конкретном случае.

Одновременно с решением задачи формирования рекомендаций операторам, такая система позволяет оцифровать все заполняемые операторами бумажные журналы, что существенно сократит ежедневные трудозатраты на выполнение обходов.

Библиографический список

1. Камелетдинов Р. Механизированная добыча нефти в России [Электронный ресурс]. URL: http://pump-sovet.com/upload/statya_kamaletdinov_oil&gas_journal_%E2%84%966-7_2018.pdf (дата обращения: 10.04.2019)
2. Марценюк М.А. Матричная реализация алгоритмов нечёткого вывода [Текст] / М.А. Марценюк, В.Б. Поляков, И.П. Селетков // Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. Информатика. Телекоммуникации. Управление. – 2012. – Т. 6 (162). – С. 133-141.
3. Марценюк М.А. Приведение конечного нечёткого автомата к нечёткой комбинационной схеме с блоком памяти [Текст] / М.А. Марценюк, И.П. Селетков // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. СПб.: Изд-во Политехн. ун-та. – 2014. – № 6 (210). – С. 67-80.

TASK FORMALIZATION OF DECISION MAKING SUPPORT IN THE PROCESS OF OIL MINING EQUIPMENT SERVICE

Seletkov, Ilya P.

Perm State University

614990, Russia, Perm, street bukireva, 15, iseletkov@gmail.com

The article describes the subject area of process of oil mining equipment service, data flows, selection of appropriate model and development of intelligent decision support system for equipment operators. The paper shows that using of the fuzzy logic is one of the simplest approaches to solve this task.

Key words: oil mining equipment service, decision support system, fuzzy logic, neural networks.

ЭМОЦИОНАЛЬНЫЕ РОБОТЫ

ОБЩИЕ ПОДХОДЫ К МАТЕМАТИЧЕСКОМУ МОДЕЛИРОВАНИЮ ЦИФРОВЫХ ДВОЙНИКОВ

Ощепкова Наталья Владимировна

Пермский государственный национальный исследовательский университет,
614990, Россия, г. Пермь, ул. Букирева, 15,
nvo@psu.ru

В статье рассматриваются математические модели, описывающие воспитание роботов с неабсолютной памятью. Предложена схема воздействия на цифровых двойников сюжетами-стимулами для получения эффективного воспитания.

Ключевые слова: робот, воспитание, эмоции, память, математическое моделирование.

В настоящее время особый интерес вызывают роботы – цифровые аналоги человека. Под цифровым двойником понимается робот, с неабсолютной памятью, способный забывать прошлую информацию.

В работе [1] приведена математическая модель, позволяющая вычислять воспитание такого робота, получаемое им в результате непрерывного воздействия на него сюжетами и порождающимися в результате этого у него эмоциями:

$$R_i = r_i + \theta_i R_{i-1}, \quad (1)$$

где i – порядковый номер сюжета, воздействующего на робота и порождающего у него элементарное воспитание r_i , R_i – суммарное воспитание робота, полученное им в результате воздействия на него общего количества сюжетов, равных величине i , θ_i – коэффициент памяти, характеризующий долю предыдущего суммарного воспитания, которую помнит робот к моменту воздействия на него сюжетом с порядковым номером i , $\theta_i \in (0, 1 - \delta]$, $0 < \delta < 1$, $\delta = \text{const}$.

У равномерно забывчивых роботов, для которых все коэффициенты памяти, соответствующие конечному моменту времени каждой эмоции, равны и постоянны, непрерывное воспитание робота при положительных эмоциях обладает сходимостью, т.е. имеет пресыщение. Причем, с ростом порядкового номера сюжета, воздействующего на робота, рост суммарного воспитания замедляется.

Согласно гипотезе грузинского психолога Д.Н. Узнадзе [2] эффект от воспитания при неправильном проведении воспитательных мероприятий может мгновенно поменяться на противоположный по знаку. Для цифровых

двойников с математической моделью (1) гипотеза Д.Н. Узнадзе эквивалента тому, что положительный знак суммарного воспитания меняется на отрицательный знак тогда, когда роботу «надоело» оказываемое на него воспитательное воздействие. В работе [4] параметр «надоело» α описывается соотношением

$$\alpha = R_i - R_{i-1} = q\theta^{i-1}, \quad (2)$$

где q – значение элементарного воспитания.

Таким образом, зная величины параметра «надоело», элементарного воспитания и коэффициента памяти, можно согласно равенству (2) определить порядковый номер воспитательного воздействия, начиная с которого положительное отношение робота к обучению может поменяться на отрицательное, и, значит, воздействия нужно временно прекратить.

Во время перерыва в воспитании робот частично забывает эмоциональное состояние, которое возникло у него после последнего воспитательного воздействия.

В работе [3] предложена формула расчета воспитания цифрового двойника для фиктивных воспитательных тактов

$$R_j = \theta^j q \frac{1 - \theta^i}{1 - \theta}, \quad (3)$$

где j – количество временных пропусков в воспитании робота-двойника.

Обозначим через β величину ожидания обучения после перерыва:

$$\beta = R_j - R_{j+1} = q(1 - \theta^i)\theta^j. \quad (5)$$

Очевидно, что, чем больше величина ожидания β , тем с большим желанием робот воспримет начало воспитательного воздействия.

Предложенная математическая модель представляет собой гипотезу советского психолога Д.Н. Узнадзе о психологических установках человека [2].

Можно считать, что параметры α и β равны между собой, так как α – «надоело» оказываемое воспитательное воздействие, а β – «надоело» ждать отсутствия обучения. Приравняв правые части равенств (2) и (3) и разрешив полученное уравнение относительно j , найдем необходимое количество пропусков воспитательного воздействия на робота:

$$j = i - 1 - \log_{\theta}(1 - \theta^i) \quad (6)$$

При больших значениях i и в силу выполнения условия $\theta_i \in (0, 1 - \delta]$ количество временных пропусков приближенно должно быть на единицу меньше, чем количество воспитательных воздействий на робота.

Построенная модель позволяет определять план воспитательных воздействий на робота для получения постоянного эффекта от воспитания.

Библиографический список

1. Пенский О.Г., Черников К.В. Основы математической теории эмоциональных роботов. Пермь: изд-во Перм. гос. ун-та. – 2010. – 256с.
2. Узнадзе Д.Н. Общая психология: учеб. для вузов. – СПб.: Питер, 2004. – 413 с.
3. Черников К.В. Математические модели роботов с неабсолютной памятью: автореф. дис. на соиск. учен. степ. канд.физ.-мат.н. (05.13.18) ПНИПУ. Пермь, – 2013. – 16с.
4. Шарапов Ю.А. Модификация алгоритма Узнадзе в аспекте кратковременной и долговременной памяти робота // Вестник Пермского университета. Серия: Математика. Механика. Информатика. – 2013. – №1(13). – С.51 – 53.

GENERAL APPROACHES TO MATHEMATICAL MODELING DIGITAL DOUBLES

Oshchepkova Natalia

Perm State University

614990, Russia, Perm, street Bukireva, 15, nvo@psu.ru

The article discusses mathematical models that describe the accumulation of emotional information by robots with non-absolute memory. The scheme of influence on digital twins is shown to achieve the best perception of information.

Key words: robot, education, emotions, memory, mathematical modeling.

ЦИФРОВАЯ ПРОМЫШЛЕННОСТЬ 2

МЕТОДЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ПРОГНОЗИРОВАНИЯ, УПРАВЛЕНИЯ И СОВЕРШЕНСТВОВАНИЯ КАЧЕСТВА ОТЛИВОК ЛОПАТОК АВИАЦИОННЫХ ДВИГАТЕЛЕЙ КАК ИНСТРУМЕНТ ПОВЫШЕНИЯ ЭКОНОМИЧЕСКОЙ ЭФФЕКТИВНОСТИ ПРЕДПРИЯТИЯ

Миролюбова Татьяна Васильевна

Пермский государственный национальный исследовательский университет
614990, Россия, г. Пермь, ул. Букирева, 15, mirolubov@list.ru

*Голдобин Максим Алексеевич, Морозов Андрей Александрович
Ясницкий Леонид Нахимович*

Пермский государственный национальный исследовательский университет
614990, Россия, г. Пермь, ул. Букирева, 15, yasn@psu.ru

В статье дан обзор современного состояния, приведены примеры и проанализированы публикации, посвященные применению методов искусственного интеллекта для прогнозирования и управления качеством литейных изделий. Сделан вывод, что на современном этапе методы искусственного интеллекта для этих целей применяются весьма редко, но, практически во всех публикациях, посвященных этой теме, отмечается эффективность и перспективность применяемых методов.

Ключевые слова: брак, качество, литейный процесс, отливка, прогнозирование, управление, искусственный интеллект.

Как известно [1], техническое совершенство современного газотурбинного двигателя (ГТД), в первую очередь, определяют рабочие и сопловые лопатки турбин высокого и низкого давления. Наиболее перспективным методом, который позволяет в промышленном масштабе получать заготовки рабочих и сопловых лопаток турбины, не нуждающихся в последующей обработке, считается способ литья по выплавляемым моделям.

В технологическом процессе изготовления отливок охлаждаемых лопаток турбины выделяют [1] следующие основные этапы: изготовление керамического стержня; изготовление выплавляемой модели лопатки; сборка моделей лопаток в блоки (секции); приготовление огнеупорной суспензии, окраска и сушка секций лопаток; удаление модельной массы из оболочек; прокалка оболочек; плавка металла, заливка оболочек и кристаллизация отливок; разрезка секций; удаление керамических стержней; контроль отливок лопаток.

Каждый из этапов изготовления оказывает влияние на конечный результат и зачастую, несмотря на достигнутый прогресс в технологии производства литых лопаток ГТД, брак в литейных цехах достигает 50% и более. Это, в свою очередь, существенно снижает экономическую эффективность деятельности предприятия.

Типичными дефектами изготовления отливок являются: наличие пор и трещин; нарушение целостности каналов охлаждения; появление в полостях охлаждения фрагментов керамического стержня; неудовлетворительное качество кристаллизации; отклонения геометрических параметров отливки [2].

Проблема борьбы с браком осложняется тем, что качество литейных изделий зависит от весьма значительного количества технологических параметров, в связи с чем, применение классических детерминированных методов положительных результатов не дает.

В настоящее время во всем мире предпринимаются попытки поиска и применения самых различных методов, с целью снижения уровня брака литейных производств. Не последнее место среди таких методов отводится искусственному интеллекту (ИИ).

В работе [3] анализируются возможности и эффекты от применения сразу нескольких технологий (стратегий) ИИ. Возможности их применения авторы обсуждают в контексте решения следующих конкретных проблем:

- Контроль состояния параметров литейного процесса, влияющих на формирование дефектов отливки, и, в зависимости от полученных сигналов, принятие решений о необходимости вмешательства в литейный процесс.

- Автоматическое получение знаний (в виде правил) о явлениях и взаимосвязях между параметрами процесса.

- Формализация знаний предметной области и создание интеллектуальных систем поддержки принятия решений.

Для решения этих задач авторы анализируют несколько стратегий ИИ:

1. Создание деревьев решений - это процесс, требующий обширных наборов данных обучения. Основными задачами деревьев решений являются: (1) Классификация в случае дискретной зависимой переменной или (2) регрессия, когда зависимая переменная непрерывна. Алгоритмы построения деревьев являются весьма популярным способом обнаружения связей между параметрами, описывающими различные явления, поэтому с помощью них можно обнаружить корреляции, которые технолог может не заметить. Их популярность во многом обусловлена тем, что они могут одинаково хорошо работать с числовыми данными (такими как регрессионные модели, нейронные сети и метод опорных векторов) и качественными переменными, даже лингвистического характера. Деревья решений позволяют строить модели явлений, которые не могут быть описаны детерминированным образом, и природа явлений не является линейной. Результаты работы алгоритмов - это деревья, которые являются графической формой правил принятия решений и относительно легко интерпретируются человеком. Специфика алгоритмов создания деревьев

решений позволяет оценить влияние различных параметров на изучаемые явления. Манипулируя большими наборами данных, деревья выстраивают обобщения отношений, которые нечувствительны к отдельным выбросам и отклонениям (шуму). Наиболее популярными алгоритмами создания деревьев решений являются CART и CHAID, описанные в [7, 8] применительно для литейных производств. Они позволяют строить деревья, как классификации, так и регрессии в зависимости от того, является ли зависимая переменная непрерывной или дискретной.

Во всех случаях применение данного метода приводит к правилам, которые могут составлять базу знаний, необходимую для автоматической оценки и управления литейным процессом (рис.1).

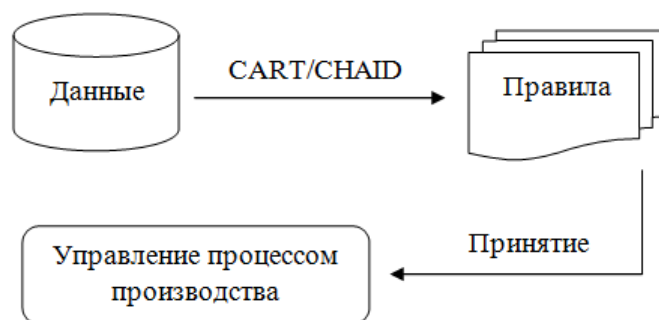


Рисунок 1. Схематическое представление использования деревьев решений

2. Нечеткая логика. Построение базы знаний на основе нечеткой логики подразумевает необходимость наличия экспертных знаний, помогающих в формализации правил, работающих на нечетких множествах. Однако существуют инструменты для машинного обучения на основе данных, такие как популярный ANFIS, который является расширением пакета Matlab. Нечеткая логика уже завоевала заслуженную популярность в вопросах управления технологическими процессами. Математический аппарат нечеткой логики используется для управления многих современных устройствах, таких как автомобили, стиральные машины, смартфоны и т.д. Нечеткие модели позволяют сократить количество правил и строить модели рассуждений, аналогичные рассуждениям, осуществляемым человеком. В области металлургии, нечеткие модели используются не только для управления технологическими процессами, но и, например в материаловедении для поддержки принятия решений.

3. Искусственные нейронные сети (ИНС). Возможные применения искусственных нейронных сетей для управления производственными параметрами литейных изделий рассматриваются через призму прогнозирования, что является одной из основных возможностей, предлагаемых данной группой методов ИИ. Их применение позволяет прогнозировать параметры качества готовой продукции в зависимости от значения отдельных производственных параметров. Для получения такого инструмента необходимо подвергнуть ИНС так называемому процессу обучения (рис. 2). Однако процесс обучения ИНС требует внимания и

контроля со стороны разработчика - слишком малый объем обучающих данных может способствовать неправильному толкованию, приводящему к очень серьезным ошибкам. Это основная проблема, возникающая при использовании нейронных сетей на производственных предприятиях, не имеющих больших ресурсов данных.

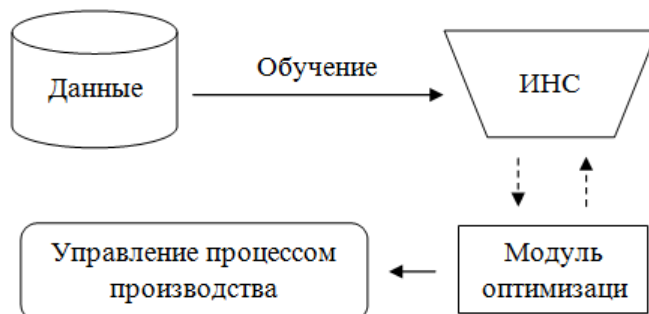


Рисунок 2. Схематическое представление типичного применения искусственной нейронной сети вместе с модулем оптимизации

Предполагая, однако, возможность обучения сети, позволяющей прогнозировать параметры качества литейных изделий, авторы работы [3] обращают внимание на возможность включения дополнительного модуля, задачей которого является оптимизация процесса управления. Модуль оптимизации по их мнению будет использовать прогнозные данные качества литья, выдаваемые нейронной сетью и вводить в систему управления технологическим процессом таких значений производственных параметров, для которых качество готовой продукции будет наилучшим. При проектировании оптимизационного модуля авторы работы [3] советуют применять дополнительные методы ИИ, например, генетические алгоритмы, что позволит определять наилучшее сочетание управляющих параметров с точки зрения конкретных критериев качества отливок. Более точное описание этой методики можно найти в [4, 5].

В работе [6] приводится пример достаточно успешного применения аппарата нейросетевого моделирования для прогнозирования значений будущих механических свойств материала на основе параметров технологического процесса его получения.

4. Рассуждения на основе прецедентов (Cased-Based Reasoning CBR). Основной парадигмой CBR (Cased-Based Reasoning) является вывод относительно решения текущей задачи путем (1) выбора аналогичного случая из базы данных, именуемой case base, и (2) адаптации решения выбранного случая к характеристикам задачи. Методология CBR также позволяет сохранять данные о текущей задаче наряду с реализованным решением в базе данных, что позволяет проводить обучение системы на основе указанных этой системой решений. Аргументация на основе прецедентов применялась, в частности, при поддержке производства медных сплавов, как система консультирования пользователя по выбору соответствующих технологических параметров производства сплава [7].

Схема, отображающая использование методологии CBR в контексте использования производственных данных и влияния на управление производственным процессом показана на рисунке 3.

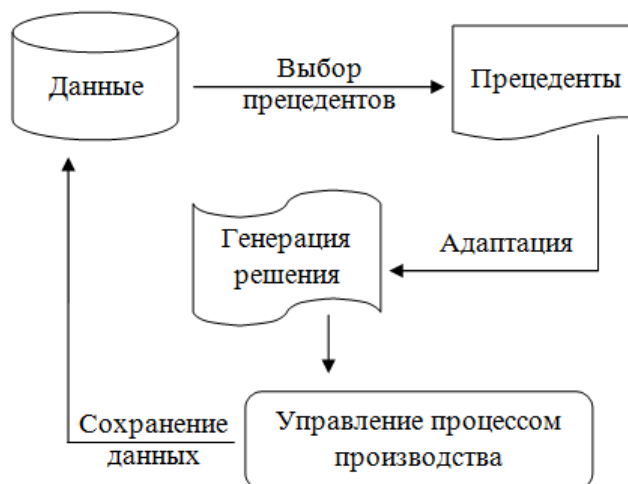


Рисунок 3. Схема использования методологии CBR для поддержки управления производственным процессом

Системы этого класса являются системами обучения во время использования, что означает, что база случаев постепенно растет с решением новых задач. Таким образом, методология CBR позволяет разрабатывать системы, поддерживающие управление производственным процессом, обеспечивающие заданные параметры качества готовых изделий при ограниченном количестве данных, при этом допуская добавление соответствующих данных при последующем использовании системы [8].

Закключение. Как показали приведенные выше исследования, методы искусственного интеллекта пока довольно слабо применяются для решения задач прогнозирования, оптимизации и управления качеством в процессах литья. Однако, имеются публикации [3-8], в которых отмечается перспективность и некоторые достигнутые успехи в этом направлении.

Исходя из опыта успешного применения методов искусственного интеллекта во множестве предметных областей [9], а также учитывая актуальность проблемы снижения литейного брака, можно ожидать, что в ближайшем будущем методы искусственного интеллекта станут основным и, возможно, единственным инструментом решения проблемы качества изделий литейных производств, в т.ч. отливок лопаток турбины авиационных двигателей. Это позволит значительно повысить экономическую эффективность деятельности предприятия.

Библиографический список

1. Испытания, обеспечение надежности и ремонт авиационных двигателей и энергетических установок: Учеб. пособие / Ю.С. Елисеев, В.В. Крымов, К.А. Малиновский, В.Г. Попов, Н.Л. Ярославцев. – М.: Изд-во МАИ, 2005. – 540 с.
2. Завалишин И.В., Финогеев А.Г. Особенности технологической подготовки производства деталей турбины газотурбинного двигателя // Труды МАИ. 2012. № 56. С. 1-11.
3. Rojek G, Regulski K., Wilk-Kołodziejczyk D., Kluska – Nawarecka S., Jaśkowiec K., Smolarek-Grzyb A. Methods of Computational Intelligence in the Context of Quality Assurance in Foundry Products // Archives of Foundry Engineering. 2016. Vol. 16, Issue 2. Pp. 11 – 16.
4. Sztangret Ł., Szeliga D., Kusiak J. & Pietrzyk M. Application of inverse analysis with metamodelling for identification of metal flow stress // Canadian Metallurgical Quarterly. 2012. Vol. 51. No 4. Pp. 440-446.
5. Rauch Ł., Sztangret Ł. & Pietrzyk M. Computer system for identification of material models on the basis of plastometric tests. // Archives of Metallurgy and Materials. 2013. Vol. 58. No 3. Pp. 737-743.
6. Mrzygłód B., Olejarczyk-Woźńska I. & Regulski K. The integration of knowledge about the manufacturing process of ADI with the use of artificial intelligence methods // Archives of Foundry Engineering. 2015. Vol. 15 No 2. Pp. 59-64.
7. Regulski K., Rojek G., Jaśkowiec K., Wilk-Kołodziejczyk D. & Kluska-Nawarecka S. Computer-Assisted Methods of the Design of New Materials in the Domain of Copper Alloys // Key Engineering Materials. 2016. Vol. 682 Pp. 143-150.
8. Rojek G., Regulski K., Szeliga D. & Kusiak J. Intelligent advisory system for support of production process design in the domain of metal forming // Key Engineering Materials. (2015). Vol. 651-653. Pp. 1375-1380. DOI: 10.4028/www.scientific.net/KEM.651-653.1375.
9. Ясницкий Л.Н. Развитие научной школы искусственного интеллекта в пермских университетах: история и научный приоритет (обзорная статья) // Прикладная математика и вопросы управления. 2018. № 4. С. 99-130. DOI: 10.15593/2499-9873/2018.4.06.

ARTIFICIAL INTELLIGENCE METHODS TO PREDICT, MANAGE AND IMPROVE THE QUALITY OF CASTING BLADES OF AIRCRAFT ENGINES AS A TOOL TO IMPROVE ECONOMIC EFFICIENCY OF ENTERPRISES

Mirolubova Tatyana V.

Perm state University, 15, Bukireva str., Perm, Russia, 614990,
mirolubov@list.ru

Goldobin, Maxim A., Morozov Andrei A.

JSC "UEC-Perm engines", 614990, Perm, GSP, Komsomolsky PR., 93.

Yasnitsky Leonid N.

Perm state University, 15, Bukireva str., Perm, Russia, 614990,
yasn@psu.ru

The article provides an overview of the current state, provides examples and analyzes publications on the use of artificial intelligence methods for forecasting and quality management of foundry products. It is concluded that at the present stage the methods of artificial intelligence for these purposes are used very rarely, but in almost all publications devoted to this topic, the effectiveness and prospects of the methods used are noted.

Key words: marriage, quality, casting process, casting, forecasting, management, artificial intelligence.

УДК 004.832

ПРОГНОЗИРОВАНИЕ СТРУКТУРНЫХ СДВИГОВ НА ОСНОВЕ САМООРГАНИЗУЮЩИХСЯ КАРТ КОХОНЕНА

Файзуллин Ринат Васильевич, Херинг Штефан

Ижевский государственный технический университет имени
М.Т.Калашникова, 426007, Россия, г. Пермь, ул.Студенческая, д. 7,
rf85@mail.ru

В статье рассматриваются киберфизические производственные системы как пример сложных технических систем. Описан опыт разработки и применения самоорганизующихся карт Кохонена для определения состояния системы и прогнозирования неисправностей или ухудшений работы системы. Проведена аналогия между изменением качества работы системы со структурными сдвигами во временных рядах. Предложено использование

самоорганизующихся карт Кохонена для прогнозирования структурных сдвигов.

Ключевые слова: прогнозирование, структурные сдвиги, самоорганизующиеся карты Кохонена, киберфизические производственные системы.

Современная действительность характеризуется все более сложными социально-экономическими и техническими системами. Одной из проблем формализации работы сложных активных систем является их динамичность (Активная система (элемент) - система (элемент), функционирование которой зависит от работы другой активной системы (элемента)). Понимая, что такая система постоянно приспосабливается ко всем изменениям, она может существенно изменять свои свойства и законы функционирования. Возникают ситуации, когда выявленные взаимозависимости могут хорошо описывать взаимодействия элементов системы или систем между собой в какой-то конкретный промежуток времени в прошлом, но со временем данные зависимости изменяются. При описании такой изменчивости зачастую используют термины «структурные сдвиги», «структурная нестабильность». Исследование структурной нестабильности во временных рядах имеет длинную историю [1,2]. Одним из наиболее популярных подходов является моделирование этой структурной нестабильности в виде так называемых структурных сдвигов, когда временной ряд до определенного момента времени порождается одним процессом, а после этого момента – другим. [3]

Следует различать два класса моделей. Если в связи с существенным изменением внешней среды, экзогенных факторов мы предполагаем структурные изменения, то это один случай, а другой когда изменения происходят из-за внутренних причин. Если в первом случае мы, понимая что произошли изменения можем сделать вывод что построенные ранее функциональные зависимости необходимо пересмотреть (после даты, когда изменения внешней среды произошли), то с эндогенными факторами мы зачастую не знаем не только даты изменения, но можем и какое-то время не знать наличия этих структурных изменений. Понятно, что нет никаких проблем с моделированием систем когда известно о наличии и времени появления причин существенного изменения структуры, поэтому основной интерес исследователей вызывает возможность выявления структурных сдвигов, когда нет информации о них из вне.

Перрон выделяет 3 типа структурных сдвигов [4]:

1. Скачком изменяется величина свободного члена

$$y_t = \mu + dD(TB)_t + y_{t-1} + e_t$$

2. Происходит излом наклона линии тренда

$$y_t = \mu_1 + y_{t-1} + (\mu_2 - \mu_1)DU_t + e_t$$

3. Обобщающий две предыдущих ситуации случай

$$y_t = \mu_1 + y_{t-1} + dD(TB)_t + (\mu_2 - \mu_1)DU_t + e_t,$$

Где
 $dD(TB)_t = 1$ если $t = T_B + 1$,
 $dD(TB)_t = 0$ если иначе;
 $DU_t = 1$ если $t > T_B$,
 $DU_t = 0$ если иначе

y_t – уровни временного ряда в момент времени t , μ_i – неизвестные коэффициенты, e_t – значение случайной составляющей в момент времени t . Обычно предполагается что случайная составляющая имеет нулевое мат. ожидание и постоянную дисперсию.

Рассмотрим современные киберфизические производственные системы (КФПС) как пример сложных технических систем. КФПС стремительно развиваются, становятся модульными, могут быть формализованы и содержать все больше датчиков. Это связано с увеличением разнообразия продукции, ее сложностью и необходимостью в конкурентной среде повышать эффективность [5]. Чем сложнее техническая система, тем труднее ее контролировать. Люди-операторы должны уметь диагностировать неисправности или аномальные поведения систем, которые приводят к их поломкам, вынужденным простоям или ухудшению качества продукции. Поддержание рабочего состояния системы связано с решением проблемы правильного определения срока службы компонентов и своевременного проведения технического обслуживания, ремонта оборудования. Данная практика имеет свои недостатки: срок службы компонентов, таких как двигатели и подшипники, часто является оценочным, а реальный срок службы компонента и, в конечном итоге, всей системы может зависеть от факторов окружающей среды, таких как изменения температуры, реальной нагрузки, качества сырья и т.п. Ранняя замена исправного компонента приводит к увеличению стоимости из-за необходимости приобретения нового компонента при рабочем старом. Гораздо худшим примером является то, что компонент достигает конца срока службы раньше, чем предполагалось. Система может пострадать от снижения эффективности, непредвиденных простоев, ухудшения качества продукции или производственных потерь, вызванных неожиданным отказом оборудования.

Мониторинг реального состояния или неисправностей системы может помочь специалистам в улучшении планирования времени обслуживания и избежать вышеупомянутых негативных эффектов, вызванных ухудшением состояния системы. Распространенный подход к решению данной задачи заключается в построении неких моделей для данной системы и сравнение предсказаний моделей с реальной системой [6]. Аномальное поведение обнаруживается, когда реальное поведение системы отклоняется от прогнозного. Данные модели могут быть созданы двумя способами. Экспертное построение моделей специалистами обычно затратно по времени и в финансовом плане. Экспертов с необходимыми знаниями трудно найти, и зачастую, некоторые из необходимых знаний вообще не доступны. Второй

вариант - построение моделей на основе анализа данных. Развитие автоматизированных интеллектуальных информационных систем призваны определить зависимость появления аномального поведения системы на основе анализа текущего состояния системы. Формализация взаимозависимостей обеспечило бы поддержку операторов и экспертов завода для поддержания работы системы в нормальном режиме.

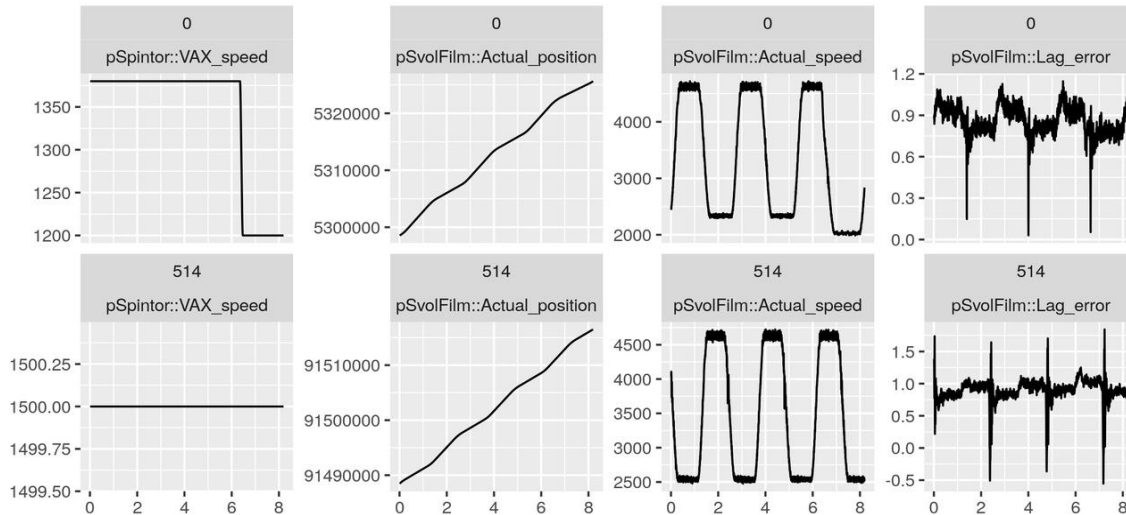


Рисунок 1. Графики изменения значений параметров ряда датчиков КФПС за 1 и 514 периоды наблюдений

Рассмотрим реальные данные о работе некоего оборудования (Рисунок 1) [7], можно заметить изменения поведения ряда параметров. Анализ показал что данные изменения не являются плавными. То есть существует период стабильной работы оборудования, а существует момент, после которого оборудование начинает работать хуже. Этот момент и является структурным сдвигом, потому что временной ряд до этого момента времени порождается одним процессом, а после этого момента – другим и скорее всего может быть описан уравнением вида $y_t = \mu_1 + y_{t-1} + dD(TB)_t + (\mu_2 - \mu_1)DU_t + e_t$. Данное уравнение является частным и самым простым случаем уравнения авторегрессии скользящего среднего ARMA(p, q) при необходимости могут быть взяты не только другие параметры (p, q), но и более сложные модификации, например, учитывающие сезонность и т.п.

Одним из подходов, основанных на анализе данных, мониторинге состояния, является использование самоорганизующихся систем [8] В основе модели лежит использование данных, когда система все еще находится в своем нормальном состоянии, когда ухудшения не произошло. Состояние системы нельзя определить с помощью одного датчика и должно быть получено из анализа множества сигналов, доступных системе. Информация, предоставленная моделью, может быть использована для того, чтобы сначала определить текущее состояние или изменения системы, а во-вторых, идентификация аномалии является помощью экспертам, чтобы они могли реагировать на ухудшения в системе, прогнозировать сбои и восстанавливать нормальное состояние системы прежде чем она выйдет из строя.

Существует успешный опыт разработки и применения самоорганизующихся карт Кохонена для определения состояния системы и прогнозирования работы системы [9], в том числе прогнозирования неисправностей или ухудшений в работе киберфизических производственных систем [10]. Самоорганизующиеся карты Кохонена - это нейронная сеть, которая без помощи учителя автоматически настраивает свои синаптические веса, решая задачу кластеризации входных векторов [11]. С точки зрения описания работы системы, кластеризация может помочь определять некоторые стандартные, типовые ситуации. Если часть времени система демонстрирует одно состояние, а потом другое, значит произошло какое-то изменение (структурный сдвиг). Если рассмотреть, что работа системы описана неким временным рядом, то помимо тех видов структурных сдвигов которые описаны выше, с точки зрения ухудшения работы можно рассматривать и другие, например, увеличение дисперсии случайной составляющей.

Таким образом, можно считать что задача идентификации или прогнозирования момента начала ухудшения работы системы является задачей, которая может быть решена с помощью самоорганизующихся карт Кохонена. Так как появление момента ухудшения работы системы означает, что временной ряд до этого момента времени порождался одним процессом, а после этого момента – другим, значит процесс идентификации момента ухудшения работы системы является процессом нахождения времени структурного сдвига. Следовательно, с помощью самоорганизующихся карт Кохонена можно идентифицировать структурные сдвиги при анализе временных рядов в целом, в том числе в других областях знаний (экономика, биология и т.п.)

Библиографический список

1. Perron, P. Dealing with Structural Breaks // Palgrave Handbooks of Econometrics: Vol. 1 Econometric Theory. Chapter 8. T.C. Mills and K. Patterson (eds.). 2006. Palgrave MacMillan, Basingstoke. Pp. 278–352
2. Casini, A., Perron, P. Structural Breaks in Time Series. 2018. Prepared for Oxford Research Encyclopedia of Economics and Finance.
3. О методах построения доверительных интервалов для дат структурных сдвигов / Соавт.: Скроботов А., Полбин А. // Экономическая политика. 2018. № 7. С.3-10
4. Pierre Perron The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis // Econometrica // Vol. 57, No. 6 (Nov., 1989), pp. 1361-1401
5. Factories of the Future: Multi-Annual Roadmap for the contractual PPP under HORIZON 2020. Luxembourg: European Union; 2013.
6. Niggemann, O., Lohweg, V.. On the diagnosis of cyber-physical production systems: State-of-the-art and research agenda. In: Twenty-Ninth AAAI Conference on Artificial Intelligence. 2015

7. One Year Industrial Component Degradation URL:
<https://www.kaggle.com/inIT-OWL/one-year-industrial-component-degradation>
8. Tian, J., Azarian, M.H., Pecht, M.. Anomaly detection using selforganizing maps-based k-nearest neighbor algorithm. Second European Conference of the Prognostics and Health Management Society 2014
9. Kohonen T. The Self-Organizing Map // Proceeding of the IEEE. 1990. Vol. 78. P. 1464-1480.
10. Von Birgelen, Alexander; Buratti, Davide; Mager, Jens; Niggemann, Oliver: Self-Organizing Maps for Anomaly Localization and Predictive Maintenance in Cyber-Physical Production Systems. In: 51st CIRP Conference on Manufacturing Systems (CIRP CMS 2018) CIRP-CMS, May 2018.
11. Ясницкий Л.Н. Интеллектуальные системы. М.: Лаборатория знаний, 2016. 221 с.

FORECASTING STRUCTURAL BREAKS BASES ON SELF - ORGANIZING MAPS

Fayzullin Rinat V., Hering Stefan

Kalashnikov Izhevsk State Technical University, 426007, Russia, Izhevsk,
 Studencheskaya str., 7, rf85@mail.ru

The article discusses cyber-physical production systems as an example of complex technical systems. The experience of the development and application of self-organizing maps is described for determining the state of the system and forecasting faults or deterioration of the system. The analogy between the changing in the quality of the system with structural shifts in the time series has been drawn. The using of self-organizing maps for forecasting structural changes is proposed.

Key words: Forecasting, Structural Breaks, Self-Organizing Maps, Cyber-Physical Production Systems .

Научное издание

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В РЕШЕНИИ АКТУАЛЬНЫХ СОЦИАЛЬНЫХ И ЭКОНОМИЧЕСКИХ ПРОБЛЕМ XXI ВЕКА

Сборник статей по материалам
Четвертой всероссийской научно-практической конференции,
проводимой в рамках Пермского естественнонаучного форума
«Математика и глобальные вызовы XXI века»
(г. Пермь, 21–23 мая 2019 г.)

Часть II

Издается в авторской редакции
Компьютерная верстка *М. Н. Федорук*

Подписано в печать 08.05.2019. Формат 60×84/16
Усл. печ. л. 9,53. Тираж 100 экз. Заказ 24

Издательский центр
Пермского государственного
национального исследовательского университета.
614990, г. Пермь, ул. Букирева, 15

ООО Учебный центр «ИНФОРМАТИКА».
614990, г. Пермь, ул. Букирева, 15