

Югорский научно-исследовательский институт информационных технологий
Федеральный исследовательский центр «Информатика и управление» РАН
Челябинский государственный университет
Уфимский государственный авиационный технический университет
Уральский федеральный университет

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И СИСТЕМЫ

**Труды Восьмой Всероссийской научной конференции
с международным участием**

Ханты-Мансийск, Россия,
17–21 марта 2020 года

Научное электронное издание

ОРГАНИЗАТОРЫ КОНФЕРЕНЦИИ:

Югорский научно-исследовательский институт
информационных технологий

Челябинский государственный университет

Федеральный исследовательский центр «Информатика и управление» РАН

Уфимский государственный авиационный технический университет

Уральский федеральный университет

Подготовлено по решению редакционно-издательского совета

Информационные технологии и системы [Электронный ресурс]: тр. Восьмой Междунар. науч. конф., Ханты-Мансийск, Россия, 17-21 марта 2020 г. (ИТиС - 2020): науч. электрон. изд. (1 файл 12 Мб)/ отв. ред. Ю.С. Попков, А.В. Мельников. Ханты-Мансийск 2020.
ISBN 978-5-6044083-8-4

Тематика конференции ориентирована на прикладные исследования в области информационных технологий и систем с использованием математического моделирования, теории управления, компьютерной лингвистики, интеллектуальных интернет-технологий и семантического web, инженерии знаний, онтологии, управления знаниями, цифровой обработки изображений и сигналов, разработки интеллектуальных и мультиагентных систем, разработки систем поддержки принятия решений, распределенных и облачных вычислений, системного анализа и моделирования бизнес-процессов.

Предназначено для ученых, специалистов и преподавателей, работающих в области информатики и вычислительной техники.

Мероприятие поддержано Российским фондом фундаментальных исследований, проект № 20-07-20029.

МЕЖДУНАРОДНЫЙ ПРОГРАММНЫЙ КОМИТЕТ

Председатель (сопредседатели) программного комитета:

Юрий Соломонович Попков, академик РАН (ФИЦ «Информатика и Управление», Институт системного анализа РАН, Россия).

Заместитель председателя программного комитета:

Андрей Витальевич Мельников, доктор технических наук, профессор (Югорский научно-исследовательский институт информационных технологий, Россия).

Программный комитет:

В. Арлазаров, член-корреспондент Российской академии наук

(Институт системного анализа РАН);

Б. Блессинг (Университет прикладных наук, Австрия);

В. Кобер (Центр научных исследований и высшего образования Энсенады, Мексика);

О. Христодуло (Уфимский государственный авиационный технический университет, Россия);

В. Грибова (Институт автоматизации и процессов управления Дальневосточного научного центра Сибирского отделения РАН, Россия);

Р. Гимранов (Сургутнефтегаз, Россия);

С. Матвеев, академик Российской академии наук

(Челябинский государственный университет, Россия);

Л. Массель (Институт энергетических исследований Российской академии наук, Россия);

Т. Михеева (Самарский государственный университет, Россия);

А. Олейник (Кольский научный центр, Россия);

Г. Осипов (Институт системного анализа РАН, Россия);

Ж. Овчарова (Технологический институт Карлсруэ, Германия);

Ю. Петриченко (Челябинский государственный университет, Россия) **А. Петунин** (Уральский федеральный университет, Россия);

С. Петунин (Федеральное государственное унитарное предприятие Всероссийский научно-исследовательский институт автоматизации им. Н. Л. Духова, Россия);

Ю. Полищук (Югорский научно-исследовательский институт информационных технологий, Россия);

С. Пятков (Югорский Государственный Университет, Россия);

Г. Ройзензон (ФИЦ «Информатика и управление», Институт системного анализа РАН, Россия);

С. Смирнов (Институт проблем управления сложными системами РАН, Россия);

О. Сметанина (Уфимский государственный авиационный технический университет, Россия);

В. Васильев (Уфимский государственный авиационный технический университет, Россия);

А. Вохминцев (Челябинский государственный университет, Россия);

В. Вольфенгаген (Национальный исследовательский ядерный университет «МИФИ», Россия);

В. Венцель (Университет прикладных наук, Австрия);

Н. Юсупова (Уфимский государственный авиационный технический университет).

СОДЕРЖАНИЕ

СЕКЦИЯ 1. МЕТОДОЛОГИЯ СИСТЕМНОГО АНАЛИЗА. МАТЕМАТИЧЕСКИЕ МОДЕЛИ И АЛГОРИТМЫ.....6

Попков Ю.С. Метод рандомизированного прогнозирования.....6

Гимранов Р.Д. Фактор времени в онтологических моделях бизнес-процессов.....13

Полищук В.Ю., Полищук Ю.М. Моделирование запасов метана в термокарстовых озерах на основе геоимитационного подхода и спутниковых снимков.....16

Пятков С.Г., Ротко В.В. Обратные задачи в теории фильтрации и тепломассопереноса....22

Куркина М.В., Самарин В.А., Самарина О.В., Славский В.В. Применение интегральных характеристик круговой три-ткани в решении задач регистрации цифрового изображения...28

Попков Ю.С., Мельников А.В., Полищук Ю.М., Сокол Е.С., Полищук В.Ю. Новый подход к восстановлению пропущенных данных о площади термокарстовых озер Арктики...32

СЕКЦИЯ 2. МЕТОДЫ АВТОМАТИЗАЦИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ЭНЕРГЕТИКЕ.....40

Грибова В.В., Шалфеева Е.А. Системы и технологии искусственного интеллекта в медицине. Какой набор средств повысит качество решений?.....40

Вохминцев А.В. Решение вариационной задачи ICP в замкнутой форме для аффинных преобразований для идентификации трещиноватостей на основе 3D-сейсморазведочной информации.....44

Куприянов М.А. Оценка погрешности измерения геометрических характеристик объектов после сглаживания их границ.....49

СЕКЦИЯ 3. МЕТОДЫ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ПОДДЕРЖКИ ПРИНЯТИЙ РЕШЕНИЙ53

Маслов С.Г., Бельтюков А.П. Проблемы научной археологии в ИТ-пространстве.....53

Гартман К.А. Разработка системы сбора и анализа персонифицированных данных в онтологии.....58

Горбачева О.М., Боровский А.С. К вопросу разработки системы очистки отработанного масла на основе интеллектуальных методов управления62

Крупицкий В.В., Полищук Ю.М. Геоинформационная методология выявления активных зон трещиноватости на основе сочетания сейсморазведочной и ландшафтной информации по космическим снимкам68

Михеева Т.И., Михеев С.В., Чугунов А.И. Методология системного анализа формирования инфраструктуры интеллектуального управления транспортными процессами с научным обоснованием стратегических решений73

Неволин О.В., Петриченко Ю.В. Оценка технологического потенциала развития программы лояльности продуктовой ритейл-сети80

Пупынин К.В. Веб-ориентированный инструментальный микроскопического моделирования транспортных потоков85

СЕКЦИЯ 4. МАШИННОЕ ОБУЧЕНИЕ И ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ.....	90
Бакановская Л.Н., Глебова Н.С. Проектирование автоматизированной системы критериальной оценки научно-исследовательской работы магистрантов с использованием методологии SADT и DFD.....	90
Егоров Ю.А., Захарова И.Г., Гасанов А.Р., Филицин А.А. Стохастический метод распознавания действий человека на базе скелетной модели	96
Максименко З.В., Бахитова Р.Х. Методика оценки экономической нагрузки для информационно-аналитической системы мониторинга и планирования расходов по выхаживанию и лечению новорожденных, родившихся с ОНМТ и ЭНМТ	103
Мещеряков Р.В., Исхаков С.Ю. О проблемах анализа данных в системах управления инцидентами безопасности роботов	108
Муратов И.Н., Байсалимова О.А. Дистанционные измерения параметров термокарстовых озер в Арктике для задач их моделирования	115
Брекоткина Е.С., Павлов А.С., Павлов С.В., Христовуло О.И., В.Д. Трубин Анализ и синтез структуры пространственных данных для управления сложными распределенными системами.....	120
Перевозчиков Н.Д., Головнин О.К. Информационная поддержка принятия решений для надзорного контроля над некредитными финансовыми организациями	126
Пьянзина Е.А., Мельников В.А. Применение нейронных сетей в прогнозировании спроса на новую продукцию приборостроительного предприятия	134
Ромазанов А.Р., Захаров А.А., Захарова И.Г. Прогнозирование температуры помещений общественного здания с помощью искусственной нейронной сети.....	138
Скрипов С.А. Реализация искусственной нейронной сети на основе микроконтроллера с использованием целочисленной арифметики.....	144
Якимчук А.В. Модель оценки и прогнозирования рисков лесных рубок на территории нефтедобывающего региона	149

Метод рандомизированного прогнозирования

Попков Ю.С.

Федеральный исследовательский центр «Информатика и управление» РАН, Москва, Россия,

Институт проблем управления РАН, Москва, Россия,

Department of Engineering, ORT Braude College, Karmiel, Israel.

popkov@isa.ru

Аннотация. Предлагается новые процедуры машинного обучения и прогнозирования, основанные на использовании рандомизированных иерархических моделей динамической регрессии, содержащих случайные параметры, измерительные шумы и случайный вход. Развита технология энтропийно-рандомизированного машинного обучения, включающая оценивание характеристик динамической регрессионной модели и ее тестирование путем генерации ансамблей прогнозных траекторий на основе сэмплирования энтропийно-оптимальных плотностей распределения вероятностей параметров модели и измерительных шумов. Указанные плотности определяются на этапе обучения решением задачи условной максимизации функционала информационной энтропии с учетом балансов с реальными данными.

Ключевые слова: прогнозирование, рандомизация, динамическая регрессия, информационная энтропия, эмпирический баланс, рандомизированное машинное обучение.

ВВЕДЕНИЕ

Ресурсы и вычислительные возможности компьютеров неуклонно растут. Это позволяет накапливать и хранить огромное количество данных как в естественных, так и в оцифрованных форматах. Естественно возникает вопрос: что можно с ними делать, кроме их сохранения. Весьма привлекательным кажется предложение, попытаться извлекать из данных новые знания. Так была сформулированы и начали развиваться концепции DataMining (DM) [1, 2], BigData (BD) [3], DataScience (DS) [4].

Весьма заманчивая цель - извлечение из данных новых знаний - неизбежно приводит к вербальному или формальному (математическому) моделированию «ожидаемых» знаний. Поэтому любая модель обладает в некотором смысле предсказательными свойствами, реализация которых возможна, если известны количественные значения ее характеристик (параметров). Данные, которые лежат в основе всех трех концепций, дают возможность количественно оценивать характеристики модели с помощью процедур *машинного обучения* (МО), и тем самым, извлекать новые знания из данных.

В отличие от указанных выше DM, BD, DS, концепция МО имеет более чем 70-ти летнюю историю и опыт решения многочисленных задач. Первая публикация из этой области относится к 1957 г. [5]. Понятие эмпирического риска, ключевое для *ML*-процедур, было введено в 1970 г. в монографии [6]. Метод потенциальных функций в задачах классификации и распознавания был опубликован в 1970 г. в монографии [7]. Современная *ML*-

концепция базируется на *детерминированной* параметризации моделей и оценках, использующих пакеты данных с постулируемыми свойствами. Качество оценивания характеризуется функциями эмпирического риска, минимизация которого дает оптимальные оценки [8, 9].

Реальные задачи, для решения которых привлекаются МО-процедуры, как правило, погружены в некоторую неопределенную среду. Если речь идет о данных, то они получены с какими-то ошибками, пропусками, сомнительной достоверностью. Формирование моделей и их параметризация - процесс неформализуемый и субъективный, зависящий от индивидуальных знаний исследователя. Поэтому в массовом применении *ML*-процедур уровень неопределенности достаточно высок.

Все эти обстоятельства говорят о том, что необходимо как-то компенсировать неопределенность. Неким общим трендом является применение ее стохастической имитации, а именно к параметризованным моделям и данным. Это означает, что параметры моделей предполагаются случайными (рандомизированными подходящим образом), а данные - имеющими случайные ошибки. Процедуры машинного обучения с указанными свойствами относятся к классу процедур *рандомизированного машинного обучения* (РМО). Их отличие от существующих МО-процедур в том, что в рамках РМО-процедур строятся оптимальные оценки не параметров, а функций плотности распределения вероятностей (ПРВ) случайных параметров и ПРВ «наихудших» случайных ошибок в данных. В энтропийных процедурах РМО в качестве критерия оптимальности оценок используется функционал обобщенной информационной энтропии [10].

Ядром РМО-процедуры является параметризованная предсказательная модель, предназначенная для имитации развития во времени (или во времени и пространстве) исследуемого процесса, и поэтому относящаяся к классу динамических моделей. Наиболее распространенными являются *параметрические модели динамической регрессии* (ПМДР), в которых текущее состояние модели определяется ее состояниями на некотором временном отрезке в прошлом [11, 12]. Формальный образ ПМДР - это разностные уравнения, в общем случае p -го порядка [13]. Наиболее распространены в прикладных задачах линейные ПМДР. В частности, они являются базовыми во многих проблемах макроэкономического моделирования и прогнозирования, таких, как моделирование временных рядов

Ю.С. Попков

Метод рандомизированного прогнозирования

экономических индикаторов [14], адекватность ПМДР [15], прогнозирование обменных курсов [16]. Линейные ПМДР оказываются достаточно эффективными для краткосрочного прогнозирования. Но для больших горизонтов прогноза они приводят к значительным погрешностям. Поэтому естественными являются попытки улучшения прогнозов путем введения в ПМДР различных нелинейностей. Общему подходу к формированию и использованию нелинейных ПМДР посвящена монография [17]. Однако, прикладные задачи требуют более “персонафицированного” подхода к выбору наиболее полезной и эффективной нелинейности. На этом пути кажется полезным, например, для прогнозирования обменных курсов применение логистической и экспоненциальной нелинейностей [18], периодических авторегрессионных моделей [19], или многомерных временных рядов [20] для прогнозирования суточной электрической нагрузки энергетической системы.

Поскольку прогнозирование осуществляется в условиях неопределенности, то погрешности прогноза, возникающие как следствие каких-то неучтенных факторов, пытаются по возможности компенсировать приданием прогнозу некоторой вероятности [21, 22]. Наиболее распространенным на этом пути является использование формулы Байеса об апостериорной вероятности. Если задать некоторую структуру параметризованной условной плотности распределения вероятностей данных и априорную плотность распределения параметров, то их произведение после нормировки определяет апостериорную плотность параметров при фиксированных данных. Принципиальные проблемы в этом подходе связаны с выбором структуры условной ПРВ и априорной ПРВ. Обычно выбираются нормальные ПРВ или их смесь, причем веса смеси оцениваются по ретроспективным данным [23-25]. Аналогичный подход используется в прикладных исследованиях: популяционной генетике [26], где развивается метод численной аппроксимации апостериорных ПРВ, взаимодействии финансового сектора экономики и трудового ранка [27], где для оценивания параметров указанных выше ПРВ используется алгоритм Метрополиса-Гастингса, динамики населения [28], где использовалась иерархия байесовых моделей для факторов рождаемости, смертности и миграционной активности. Одна из проблем, возникающая при применении этих подходов, состоит в значительной размерности соответствующих задач. Поэтому важное значение приобретают методы снижения размерности, весьма информативный обзор которых изложен в [29]. Вероятностные прогнозы формируются и другими методами с учетом специфики прикладных задач. В частности, в метеорологии путем накопления ретроспективных прогнозов погоды определяются оценки функций ПРВ, которые затем используются для кратко временного прогнозирования [30-33]. Довольно любопытным является формирование вероятностного прогноза путем смешивания прогнозов, полученных разными методами [34].

В данной статье предлагается принципиально иной метод формирования прогнозов, названный *энтропийно-рандомизированным прогнозированием*

(ЭРП). Он основан на генерации ансамбля случайных прогнозов *предсказательной моделью динамической регрессии (ПМДР)* со случайными входом и параметрами. Соответствующие вероятностные характеристики, а именно, функции ПРВ определяются с помощью процедуры энтропийно-рандомизированного машинного обучения. Ансамбли прогнозных траекторий строятся путем сэмпирования энтропийно-оптимальных ПРВ.

ТЕХНОЛОГИЯ ЭНТРОПИЙНО-РАНДОМИЗИРОВАННОГО ПРОГНОЗИРОВАНИЯ

Рандомизация как средство придания искусственных, целесообразно организованных случайных свойств событиям, индикаторам, методам, по своей природе неслучайным, является довольно распространенным приемом достижения позитивного эффекта. Существует много примеров в самых разных областях науки, управлении, экономики, таких, например, как рандомизированные численные методы оптимизации [35, 36], смешанные (случайные) стратегии в биржевой торговле [37], рандомизированное прогнозирование динамики населения [38], вибрационное управление технологическими процессами [39]. При этом рандомизация предполагает придание неслучайным объектам искусственных стохастических свойств с оптимальными, в выбранном смысле, вероятностными характеристиками. Вопрос о выборе количественных характеристик оптимальности всегда оказывается дискуссионным и неоднозначным. Он требует аргументов, которые бы как-то отражали важную специфику рандомизированного объекта. В частности, принципиальной особенностью процедур прогнозирования является сопровождающая их *неопределенность* в используемых данных, предсказывающих моделях, методах генерации прогнозов и др.

Здесь будет использоваться в качестве характеристики неопределенности *информационная энтропия* [40]. В работах [41-43] показано, что энтропия, согласно первому закону термодинамики, является естественным функционалом, характеризующим процессы всеобщей эволюции. Во-вторых, согласно второму закону термодинамики ее максимизация определяет наилучшее состояние эволюционного процесса при наихудших воздействиях на него (максимальной неопределенности). Стоит упомянуть еще одно качество информационной энтропии, связанное с измерительными и иными погрешностями, являющимися важными характеристиками данных. Использование информационной энтропии для учета влияния указанных погрешностей позволяет оценить вероятностные характеристики шумов, оказывающих наихудшее воздействие на процедуры прогнозирования.

Технология энтропийно-рандомизированного прогнозирования складывается из следующих этапов. В начале формируется *предсказательная рандомизированная модель (ПРМ)* исследуемого объекта и синтезируются определяющие ее параметры. ПРМ преобразует реальные данные в модельный выход. В общем случае эти преобразования предполагаются динамическими, т.е. модельный выход, наблюдаемый в

момент времени n , зависит от состояний, наблюдаемых на некотором историческом интервале. Параметры ПРМ предполагаются интервального типа и случайными, вероятностные свойства которых характеризуются функциями ПРВ.

Оценивание функций ПРВ есть задача второго этапа рассматриваемой технологии - рандомизированного машинного обучения, а точнее его энтропийной версии. На этом этапе, используя обучающие коллекции данных и алгоритм обучения в форме задачи функционального энтропийно-линейного программирования, определяются оценки функций ПРВ.

На третьем этапе производится тестирование оптимизированной ПРМ (с энтропийно-оптимальными ПРВ) с использованием тестовой коллекции данных и принятых количественных характеристик качества обучения. Оптимизированная ПРМ представляет собой генератор ансамбля случайных траекторий, векторов, событий, параметры которого энтропийно-оптимальные.

Обученные и протестированные ПРМ используются для прогнозирования. При этом энтропийно-оптимальные ПРМ генерируют ансамбли случайных прогнозных траекторий, по которым вычисляются их числовые характеристики: средние траектории, дисперсионные трубки, медианные траектории, эволюция ПРВ прогнозных траекторий и др.

РАНДОМИЗИРОВАННЫЕ МОДЕЛИ ДИНАМИЧЕСКОЙ РЕГРЕССИИ СО СЛУЧАЙНЫМ ВХОДОМ И ПАРАМЕТРАМИ

Рандомизированные модели динамической регрессии (РМДР) образуют класс динамических моделей со случайными параметрами, описывающим параметризованную зависимость состояния объекта в данный момент времени от внешних факторов и его состояний в какие-то предыдущие моменты времени.

Структуры моделей формируются на базе существующих знаний и гипотез о свойствах объекта, которые довольно часто оказываются весьма неточными. Кроме того, сами внешние факторы могут меняться во времени, и поэтому необходимо прогнозировать их для моделирования динамики объекта. Надежная информация о реально измеряемых воздействиях, приводящих к временной эволюции внешних факторов, часто бывает неизвестна. Все отмеченное в совокупности свидетельствует о наличии *неопределенности* как при формировании, так и при использовании моделей. В [10] был предложен метод снижения ее влияния, основанный на рандомизации используемых моделей, в том числе и РМДР. Специфика его применения в данном случае состоит в распространении идеи рандомизации на моделирование эволюции внешних факторов.

Структура, предлагаемой РМДР, изображена на рис. 1. Она состоит из модели основного объекта РМДР-О со случайными параметрами $a \in R^p$, и модели внешних факторов РМДР-Ф со случайными параметрами $b \in R^s$ и случайным входом $\zeta \in R^q$. Состояния объекта и его модели принадлежат векторному пространству R^m , в котором $x[n]$ — вектора состояний объекта, $x[n] \in R^m$ — вектора состояний РМДР-О. Внешние факторы характеризуются вектором $y[n] \in R^q$, и изменение во времени состояния

РМДР-Ф — вектором $y[n] \in R^q$. Переменная n — целочисленная, обозначает дискретное время в интервале $L = [n^-, n^+]$.



Рис. 1.

Рассмотрим *линейную версию РМДР-О*. Состояние модели $x[n]$ в момент времени n меняются под влиянием p ретроспективных состояний $x[n-1], \dots, x[n-p]$ и измеряемых в каждый момент времени внешних факторов $z[n] \in R^q$. Соответствующее уравнение имеет вид:

$$x[n] = X^{(n,p)} A^{(p)} + A_{(p+1)} z[n], \quad (1)$$

где:

- блочный вектор-столбец параметров

$$A^{(p)} = [A_1, \dots, A_p]; \quad (2)$$

A_i — $(m \times m)$ -матрица со случайными элементами интервального типа

$$A_i \in \mathcal{A}_i = [A_i^-, A_i^+], \quad i = \overline{1, p}; \quad (3)$$

- $A_{(p+1)}$ — $(m \times q)$ -матрица со случайными элементами интервального типа

$$A_{(p+1)} \in \mathcal{A}_{(p+1)} = [A_{(p+1)}^-, A_{(p+1)}^+]; \quad (4)$$

- блочный вектор-строка p ретроспективных состояний

$$X^{(n,p)} = [x[n-1], \dots, x[n-p]]. \quad (5)$$

Вероятностные свойства блочного вектора $A^{(p)}$ характеризуются функцией совместной ПРВ $P(A^{(p)})$, а матрицы $A_{(p+1)}$ — функцией ПРВ $F(A_{(p+1)})$.

Предполагается, что измеряемое в каждый момент времени n состояние РМДР-О содержит аддитивный шум $\mu[n]$:

$$v[n] = x[n] + \mu[n]. \quad (6)$$

Случайные векторы $\mu[n]$ — интервального типа:

$$\mu[n] \in \mathcal{M}_n = [\mu_n^-, \mu_n^+], \quad (7)$$

с функцией ПРВ $M_n(\mu[n])$. Случайные векторы, измеряемые в разные моменты времени предполагаются статистически независимыми.

Рассмотрим *линейную версию РМДР-Ф*, которая имеет аналогичную РМДР-О структуру. Будем иметь:

$$y[n] = Y^{(n,s)} B^{(s)} + \zeta[n], \quad (8)$$

где

- блочный вектор - столбец

$$B^{(s)} = [B_1, \dots, B_s]^T, \quad (9)$$

- B_i — $(q \times q)$ -матрица со случайными элементами интервального типа

$$B_i \in \mathcal{B}_i = [B_i^-, B_i^+], \quad i = \overline{1, s}, \quad (10)$$

- блочный вектор - строка

$$Y^{(n,p)} = [y[n-1], \dots, y[n-s]]. \quad (11)$$

Вероятностные свойства параметров характеризуются непрерывно-дифференцируемой функцией ПРВ $W(B^{(s)})$.

Случайный вектор $\zeta[n]$ — интервального типа:

$$\zeta[n] \in \mathcal{E}_n = [\zeta_n^-, \zeta_n^+] \quad (12)$$

с функцией ПРВ $Q_n(\zeta[n])$, также непрерывно-дифференцируемой. Векторы $\zeta[n]$ для разных моментов времени статистически независимы.

По аналогии с РМДР-О, измеряемое в каждый момент времени n состояние РМДР-Ф $z[n]$ содержит аддитивный шум $\xi[n]$:

$$z[n] = y[n] + \xi[n]. \quad (13)$$

Случайный вектор $\xi[n]$ - интервального типа:

$$\xi[n] \in \Xi_n = [\xi_n^-, \xi_n^+], \quad (14)$$

с непрерывно-дифференцируемой функцией ПРВ $G_n(\xi[n])$. Случайные векторы, измеряемые в разные моменты времени предполагаются статистически независимыми.

Итак, в общей РМДР (РМДР-О и РМДР-Ф) неизвестными являются функции ПРВ параметров $P(A^{(p)})$, $F(A_{(p+1)})$, $W(B^{(s)})$, и шумов $M_n(\mu[n])$, $Q_n(\zeta[n])$, $G_n(\xi[n])$, $n \in \mathcal{L}$.

МОДЕЛИ ДАННЫХ ИЗ ОБУЧАЮЩЕЙ КОЛЛЕКЦИИ

Оценивание указанных в предыдущем параграфе ПРВ производится с использованием обучающих коллекций данных, полученных на интервале обучения $n \in \mathcal{L} = [n^-, n^+]$ и согласованных с РМДР.

Рассмотрим РМДР-О. На интервале обучения будем иметь:

$$\begin{aligned} x[n^-] &= X^{(n^-, p)} A^{(p)} + A_{(p+1)} z[n^-], \\ x[n^- + 1] &= X^{(n^- + 1, p)} A^{(p)} + A_{(p+1)} z[n^- + 1], \\ &\dots = \dots \dots \dots, \\ x[n^+] &= X^{(n^+, p)} A^{(p)} + A_{(p+1)} z[n^+]. \end{aligned} \quad (15)$$

Тогда наблюдаемые состояния РМДР-О на интервале обучения \mathcal{L} представляют собой набор векторов:

$$v[n] = x[n] + \mu[n], \quad n = \overline{n^-, n^+}. \quad (16)$$

Отсюда следует, что коллекция обучающих данных состоит из данных о ретроспективных состояниях объекта $X^{(n^-, p)}, X^{(n^- + 1, p)}, \dots, X^{(n^+, p)}$,

$$(17)$$

и данных о наблюдаемых текущих состояниях:

$$v[n^-], \dots, v[n^+], \quad z[n^-], \dots, z[n^+]. \quad (18)$$

Рассмотрим РМДР-Ф. На интервале обучения будем иметь:

$$\begin{aligned} y[n^-] &= Y^{(n^-, s)} B^{(s)} + \zeta[n^-], \\ y[n^- + 1] &= Y^{(n^- + 1, s)} B^{(s)} + \zeta[n^- + 1], \\ &\dots = \dots \dots \dots, \\ y[n^+] &= Y^{(n^+, s)} B^{(s)} + \zeta[n^+]. \end{aligned} \quad (19)$$

Наблюдаемые состояния РМДР-Ф на интервале обучения \mathcal{L} представляют собой набор векторов:

$$z[n] = y[n] + \xi[n], \quad n = \overline{n^-, n^+}. \quad (20)$$

Отсюда следует, что коллекция обучающих данных состоит из данных о ретроспективных состояниях объекта $\hat{Y}^{(n^-, s)}, \hat{Y}^{(n^- + 1, s)}, \dots, \hat{Y}^{(n^+, s)}$,

$$(21)$$

и данных о наблюдаемых текущих состояниях:

$$z[n^-], \dots, z[n^+]. \quad (22)$$

Итак, в процедуре обучения РМДР участвуют коллекции данных (18), (21), (22).

АЛГОРИТМ РАНДОМИЗИРОВАННОГО МАШИННОГО ОБУЧЕНИЯ (РМО)

Энтропийная версия [10] алгоритмов РМО используется для оценивания функций ПРВ параметров и шумов для РМДР-О и РМДР-Ф. Применительно к РМДР-О соответствующий алгоритм имеет вид:

$$\begin{aligned} \mathcal{H}_O = & - \int_{\mathcal{A}} P(A^{(p)}) \ln P(A^{(p)}) dA^{(p)} \\ & - \int_{\mathcal{A}_{(p+1)}} F(A_{(p+1)}) \ln F(A_{(p+1)}) dA_{(p+1)} - \\ & - \sum_{n=n^-}^{n^+} \int_{\mathcal{M}_n} M_n(\mu[n]) \ln M_n(\mu[n]) dM_n(\mu[n]) \Rightarrow \max, \end{aligned} \quad (23)$$

при ограничениях:

- нормировки ПРВ

$$\begin{aligned} \int_{\mathcal{A}} P(A^{(p)}) dA^{(p)} &= 1, \quad \int_{\mathcal{A}_{(p+1)}} F(A_{(p+1)}) dA_{(p+1)} = 1, \\ \int_{\mathcal{M}_n} M_n(\mu[n]) dM_n(\mu[n]) &= 1, \quad n = \overline{n^-, n^+}, \end{aligned}$$

- эмпирических балансов

$$\begin{aligned} \int_{\mathcal{A}} P(A^{(p)}) X^{(n, p)} A^{(p)} dA^{(p)} \\ + \int_{\mathcal{A}_{(p+1)}} F(A_{(p+1)}) A_{(p+1)} z[n] dA_{(p+1)} + \\ + \int_{\mathcal{M}_n} M_n(\mu[n]) \mu[n] d\mu[n] = v[n], \quad n = \overline{n^-, n^+}. \end{aligned} \quad (24)$$

Заметим, что эмпирические балансы представляют собой систему $(n^+ - n^-)$ блоков из m уравнений. Каждому такому блоку соответствует вектор множителей Лагранжа $\theta^{(n)}$ размерности m . Эта задача относится к классу функциональных задач энтропийно-линейного программирования ляпуновского типа [44]. Она имеет аналитическое решение, параметризованное множителями Лагранжа:

$$\begin{aligned} P^*(A^{(p)}) &= \frac{\exp(-\sum_{n=n^-}^{n^+} \langle \theta^{(n)}, X^{(n, p)} A^{(p)} \rangle)}{\mathcal{P}(\theta)}, \\ F^*(A_{(p+1)}) &= \frac{\exp(-\sum_{n=n^-}^{n^+} \langle \theta^{(n)}, A_{(p+1)} z[n] \rangle)}{\mathcal{F}(\theta)}, \\ M_n^*(\mu[n]) &= \frac{\exp(-\langle \theta^{(n)}, \mu[n] \rangle)}{\mathcal{M}_n(\theta^{(n)})}, \quad n = \overline{n^-, n^+}. \end{aligned} \quad (25)$$

В этих равенствах

$$\begin{aligned} \mathcal{P}(\theta) &= \int_{\mathcal{A}} \exp\left(-\sum_{n=n^-}^{n^+} \langle \theta^{(n)}, X^{(n, p)} A^{(p)} \rangle\right) dA^{(p)}, \\ \mathcal{F}(\theta) &= \int_{\mathcal{A}_{(p+1)}} \exp\left(-\sum_{n=n^-}^{n^+} \langle \theta^{(n)}, A_{(p+1)} z[n] \rangle\right) dA_{(p+1)}, \\ \mathcal{M}_n(\theta^{(n)}) &= \int_{\mathcal{M}_n} \exp(-\langle \theta^{(n)}, \mu[n] \rangle) d\mu[n], \quad n = \overline{n^-, n^+}. \end{aligned} \quad (26)$$

Матрица множителей Лагранжа $\theta = [\theta^{(n^-)}, \dots, \theta^{(n^+)}]$ определяется решением следующих балансовых уравнений:

$$\begin{aligned} & \mathcal{P}^{(-1)}(\theta) \int_{\mathcal{A}} \exp(-\sum_{n=n^-}^{n^+} \langle \theta^{(n)}, X^{(n,p)} A^{(p)} \rangle) X^{(n,p)} A^{(p)} dA^{(p)} + \\ & + \mathcal{F}^{(-1)}(\theta) \int_{\mathcal{A}_{(p+1)}} \exp(-\sum_{n=n^-}^{n^+} \langle \theta^{(n)}, A_{(p+1)} z[n] \rangle) A_{(p+1)} z[n] dA_{(p+1)} + \\ & + \mathcal{M}_n^{(-1)}(\theta^{(n)}) \int_{\mathcal{M}_n} \exp(-\langle \theta^{(n)}, \mu[n] \rangle) \mu[n] d\mu[n] = \\ & x[n], \quad n = \overline{n^-, n^+}. \end{aligned} \quad (27)$$

Функции ПРВ $P^*(A^{(p)}), F^*(A_{(p+1)})$ параметров РМДР-О и измерительных шумов $M_n^*(\mu[n]), n = \overline{n^-, n^+}$ определяются по обучающим коллекциям ретроспективных данных $X(n^-, p), X(n^- + 1, p), \dots, X(n^+, p)$, данных о текущих состояниях $x[n^-], \dots, x[n^+]$ и данных $z[n^-], \dots, z[n^+]$, которые генерируются РМДР-Ф.

Для получения последних применим алгоритм РМО к оцениванию функций ПРВ параметров и шумов для РМДР-Ф. Согласно [10] будем иметь:

$$\begin{aligned} & \mathcal{H}_F = - \int_B W(B) \ln W(B) dB - \\ & - \sum_{n=n^-}^{n^+} \int_{\varepsilon_n} Q_n(\zeta[n]) \ln Q_n(\zeta[n]) dQ_n(\zeta[n]) - \\ & - \sum_{n=n^-}^{n^+} \int_{\varepsilon_n} G_n(\xi[n]) \ln G_n(\xi[n]) dG_n(\xi[n]) \Rightarrow \max, \end{aligned} \quad (28)$$

при ограничениях:

- нормировки ПРВ

$$\begin{aligned} & \int_B W(B) dB = 1, \\ & \int_{\varepsilon_n} Q_n(\zeta[n]) dQ_n(\zeta[n]) = \\ & 1, \quad \int_{\varepsilon_n} G_n(\xi[n]) dG_n(\xi[n]) = 1, \quad n = \overline{n^-, n^+}; \end{aligned}$$

- эмпирических балансов

$$\begin{aligned} & \int_B W(B^{(s)} Y^{(n,s)} B^{(s)}) dB^{(s)} + \\ & \int_{\varepsilon_n} Q_n(\zeta[n]) \zeta[n] d\zeta[n] + \\ & + \int_{\varepsilon_n} G_n(\xi[n]) \xi[n] d\xi[n] = z[n], \quad n = \\ & \overline{n^-, n^+}. \end{aligned} \quad (29)$$

Эта задача имеет аналитическое решение в терминах множителей Лагранжа:

$$\begin{aligned} & W^*(B^{(s)}) = \frac{\exp(-\sum_{n=n^-}^{n^+} \langle \eta^{(n)}, Y^{(n,s)} B^{(s)} \rangle)}{\mathcal{W}(\eta)}, \quad (30) \\ & Q_n^*(\zeta[n]) = \frac{\exp(-\langle \zeta[n], \eta^{(n)} \rangle)}{\mathcal{Q}_n(\eta^{(n)})}, \\ & G_n^*(\xi[n]) = \frac{\exp(-\langle \xi[n], \eta^{(n)} \rangle)}{\mathcal{G}_n(\eta^{(n)})}, \quad n = \overline{n^-, n^+}. \\ & \eta = [\eta^{(n^-)}, \dots, \eta^{(n^+)}]. \end{aligned}$$

В этих равенствах

$$\mathcal{W}(\eta) = \int_B \exp(-\sum_{n=n^-}^{n^+} \langle \eta^{(n)}, Y^{(n,s)} B^{(s)} \rangle) dB^{(s)}, \quad (31)$$

$$\begin{aligned} & \mathcal{Q}_n(\eta^{(n)}) = \\ & \int_{\varepsilon_n} \exp(-\langle \zeta[n], \eta^{(n)} \rangle) d\zeta[n], \quad \mathcal{G}_n(\eta^{(n)}) = \\ & \int_{\varepsilon_n} \exp(-\langle \xi[n], \eta^{(n)} \rangle) d\xi[n], \\ & n = \overline{n^-, n^+}. \end{aligned}$$

Матрица множителей Лагранжа η определяется решением следующих балансовых уравнений:

$$\begin{aligned} & \mathcal{W}^{(-1)}(\eta) \int_B \exp(-\sum_{n=n^-}^{n^+} \langle \eta^{(n)}, Y^{(n,s)} B^{(s)} \rangle) Y^{(n,s)} B^{(s)} dB^{(s)} + \\ & + \mathcal{Q}_n^{(-1)}(\eta^{(n)}) \int_{\varepsilon_n} \exp(-\langle \zeta[n], \eta^{(n)} \rangle) \zeta[n] d\zeta[n] + \\ & + \mathcal{G}_n^{(-1)}(\eta^{(n)}) \int_{\varepsilon_n} \exp(-\langle \xi[n], \eta^{(n)} \rangle) \xi[n] d\xi[n] = z[n], \end{aligned} \quad (32)$$

БЛАГОДАРНОСТИ

Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект № 19-07-00282).

ЛИТЕРАТУРА

1. Piatetsky-Shapiro G., Frawley W., eds. Knowledge Discovery in Databases. AAAI/MIT Press, 1991.
2. Witten I.H., Frank E. Data Mining: Practical Learning Tools and Techniques (Second Edition). Morgan Kaufmann, 2005.
3. Editorial Community cleverness required. Nature, 2008, v.455(1).
4. Dhar V. Data Sciences and Prediction. Communication of the ACM, 2013, v.58, № 12, p.64-73.
5. Rosenblatt M. The Perceptron - perceiving and recognizing automaton. Report 85-460-1, 1957.
6. Цыпкин Я.З. Основы теории обучающихся систем. М., Наука, 1970.
7. Айзерман М.А., Браверман Э.М., Розоноэр Л.И. Метод потенциальных функций в теории обучения машин. М., Наука, 1970.
8. Vapnik V.N. Statistical Learning Theory. NY, John Wiley & Sons Inc., 1998.
9. Bishop C.M. Pattern Recognition and Machine Learning. Springer, Series: Information Theory and Statistics, 2006.
10. Попков Ю.С., Попков А.Ю., Дубнов Ю.А. Рандомизированное машинное обучение. УРСС, 2018.
11. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. Springer, 2001. <http://www-stat.stanford.edu/tibs/ElemStatLearn>.
12. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Исследование зависимостей. М., Финансы и статистика, 1985.
13. Tarasov A. Forecasting US money growth using economic uncertainty measures and regularization techniques. International Journal of Forecasting, 2019, v.35, 443-457.
14. Marcellino M., Stock J.H., Watson M. A comparison of direct and iterated multisteps AR methods for forecasting macroeconomic time series. Journal of Econometrics, 2006, v.135(1-2), 499-526.

15. Eitrheim O., Terasirta T. Testing the adequacy of smooth transition autoregressive models. *Journal of Econometrics*, 1996, v.74(1), 59-75.
16. Molodtsova T., Papell D. Out-of-sample exchange rate predictability with TR fundamentals. *Journal of International Economy*, 2009, v.77, 167-180.
17. Granger C.W. Terasirta T. *Modelling Nonlinear Economic Relationship*. 1993, Oxford: Oxford University Press.
18. Wang R., Motley B., Stamatogiannis M. Forecasting the exchange rate using nonlinear Taylor rule based model. *International Journal of Forecasting*, 2019, v.35, 429-442.
19. Bessec M., Fuaguau J. Short-run electricity load forecasting with combinations of stationary wavelet transforms. *EJOR*, 2018, v.264, p. 149-163.
20. Clements A.E., Hurn A.S., Li Z. Forecasting day-ahead electricity load using a multiple equation time series approach. *EJOR*, 2016, v.251, p.522-530.
21. Hong T., Fan S. Probabilistic electric load forecasting: a tutorial review. *International Journal of Forecasting*, 2016, v. 32, p.914-938.
22. Wheatcroft E. Interpreting the skill score from of forecast performance metrics. *International Journal of Forecasting*, 2019, v.35, 573-579.
23. Canale A., Ruggiero M. Bayesian nonparametric forecasting of nonmonotonic functional time series. *Electronic Journal of Statistics*. 2016, v.10(2), 3265-3288.
24. Дубнов Ю.А., Булычев А.В. Байесовская идентификация параметров смеси нормальных распределений. *Информационные технологии и вычислительные системы*, 2017, № 1, 101-114.
25. Frazier D.T., Maneeosonthern W., Martin G.M., McCabe B.P.M. Approximate Bayesian forecasting. *International Journal of Forecasting*, 2019, v.35, 521-539.
26. Beaumont M.A., Zhang W., Baldung D.J. Approximate Bayesian computation in population genetics. *Genetics*, 2001, v.162(4), 2025-2035.
27. McAdam P., Warne A. Euro-area real time density forecasting with financial or labour market friction. *International Journal of forecasting*, 2019, v.35, 580-600.
28. Alkema L., Gerland P., Raftery A., Wilmoth T. The UN probabilistic population projection: an introduction to demographic forecasting with uncertainty. *Forsight*, 2015, v.37, 19-24.
29. Blum M.G.B., Nunes M.A., Prangle D., Sisson S.A. A comparison review of dimension reduction methods in approximate Bayesian computation. *Statistical Science*, 2013, v.28(2), 189-208.
30. Brier C.W. Verification of forecast expressed in term probability. *Monthly Weather Review*, 1950, v.78, 1-3.
31. Brier J., Smith L.A. From ensembles forecasts to predictive distribution functions. *Tellus A*, 2008, v.60(4), 663-678.
32. Christensen H.M., Moroz I.M., Palmer T.N. Evaluation of ensemble forecast uncertainty using a new proper score: application to medium-range and seasonal forecast. *Quarterly Journal of the Royal Meteorological Science*, 2015, v.141(687), 538-549.
33. Gneiting T., Katzfuss M. Probabilistic Forecasting. *Annual Review of Statistics and Its Applications*, 2014, v.1, p.125-151.
34. Lahiri K., Wang J.C. Evaluating probability forecasts for GDP declines using alternative methodological. *International Journal of Forecasting*, 2013, v.29(1), 175-190.
35. Vidyasagar M. *Statistical Learning Theory and Randomized Algorithms for Control*. *IEEE Control System Magazine*, 1998, v.1/17, 69-88.
36. Граничин О.Н., Поляк Б.Т. Рандомизированные алгоритмы оценивания и оптимизации при почти произвольных помехах. М., Наука, 2002, 291 с.
37. Biondo A.E., Pluchino A., Rapisarda A., Helbing D. Are random trading strategies more successful than technical ones? *PLoS ONE*, 2013, v.6(7), e68344.
38. Lutz W., Sandersen S., Scherbov S. The end of world population growth. *Nature*, 2001, v.412(6846), 543-545.
39. Цирлин А.М. Методы усредненной оптимизации и их применение. М., Физматлит, 1997, 304 с.
40. Shannon C. *Communication Theory of Secrecy Systems*. *Bell System Technical Journal*, 1949, v.28(4), p. 656-715.
41. Jaynes E.T. *Information Theory and Statistical Mechanics*. *Physics Review*, 1957, v.106, pp.620-630.
42. Jaynes E.T. *Papers on probability, statistics and statistical physics*. Dordrecht, Kluwer Academic Publisher, 1989.
43. Jaynes E.T. *Probability Theory. The logic and science*. Cambridge University Press, 2003.
44. Иоффе А.Д., Тихомиров В.М. Теория экстремальных задач, М., Наука, 1974, 481 с.

Method of Randomized Forecasting

Popkov Y.S.

Federal Research Center “Computer Science and Control” of Russian Academy of Sciences, Moscow, Russia,
Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia,
Department of Engineering, ORT Braude College, Karmiel, Israel.
popkov@isa.ru

Abstract. We propose a new forecasting procedure that involve randomized hierarchical dynamic regression models with random parameters, measurement noises and random input. We developed the technology of entropy-randomized machine learning, which includes the estimation of characteristics of a dynamic regression model and its testing by generating ensembles of predicted trajectories through the sampling of the entropy-optimal probability density functions of the model parameters and measurement noises. The density functions are determined at the

learning stage by solving the constrained maximization problem of an information entropy functional subject to the empirical balances with real data.

Keywords: forecasting, randomization, dynamic regression, information entropy, empirical balance, randomized machine learning.

Фактор времени в онтологических моделях бизнес-процессов.

Гимранов Р.Д.
ПАО «Сургутнефтегаз»
Сургут, Россия
Gimranov_rd@mail.ru

Аннотация. Рано или поздно онтологии уйдут «под капот» цифровых платформ, которые, благодаря онтологическим моделям, смогут в режиме реального времени «понимать» особенности и изменения как реального мира, так и цифровых сущностей. Но, текущее состояние исследований и разработок пока не позволяет получить зрелый стек готовых решений для крупных промышленных предприятий, содержащий взаимосвязанную совокупность теоретических основ, соглашения о моделировании, методик и программной среды реализации. Одной из проблем, над которой необходима работа, является включение времени в онтологические модели и соответственное изменение функциональности программных решений.

Ключевые слова: онтологическое моделирование, цифровая платформа, темпоральная онтология, время в онтологиях

«Экзистенциально-онтологическое устройство целостности присутствия основывается во временности. Поэтому исходный способ временения самой экзистенциальной временности должен делать возможным экзистенциальный набросок бытия «вообще». Как интерпретировать этот модус временения временности? От исходного времени ведет путь к смыслу бытия? Время само открывается как горизонт бытия?»

Мартин Хайдеггер «Бытие и время».

В настоящее время уже не надо приводить доказательства необходимости и важности онтологических моделей для современных цифровых платформ. Ведется значительное количество исследовательских работ, разрабатываются онтологические модели архитектуры для крупных предприятий, развиваются программные решения. В числе значимых примеров можно упомянуть работы по архитектуре предприятия Самарина и Захарчука [2], [3], разработки Rosemann and Green [4], развитие различных решений в технологическом стеке Semantic Web, подход к архитектуре предприятия Шведина [5], онтологическая архитектура Диеца [6], модели онтологической архитектуры цифровой платформы [7]. Практически все они так или иначе опираются на мета модель BWB (Bunge-Wand-Weber) [8]

В рамках проекта по реализации цифрового ТОРО (техобслуживание и ремонт оборудования), мы попытались обобщить проблемы, которые возникли при внедрении ТОРО в различных подразделениях в течение 2000-

2015 г.г. Основная проблематика была связана с тем, что каждая служба по-разному видит объект для ТОРО. Механики рассматривают особенности конструкции, электрики – принципиальную схему, нагрузки и подключения, технологи – результат деятельности и необходимые действия персонала, бухгалтеры – основное средство на балансе, плановики – статьи затрат, участие в прибыли... Все эти взгляды были реализованы в различных модулях информационной системы SAP ERP и, зачастую, в совершенно различных перечнях объектов. Для учета всех взглядов в одной модели объекта, была разработана онтологическая модель насоса. Реализовать эту модель средствами информационной системы и построить на этом платформу для цифрового ТОРО пока нет возможности [9]. На стадии моделирования была также поставлена задача отразить в модели состояние объекта на любом этапе жизненного цикла, от приобретения или создания до ликвидации. Это оказалась очень непростой задачей, работа над которой идёт до сих пор. Подходы к моделированию хорошо отображают срез состояния, но не его динамику. Есть возможность заложить динамику определённых показателей, свойств, и привязать их к этапу жизненного цикла или к конкретным выполняемым работам, но и этап и работы также должны быть заранее структурированы и отражены в модели как объекты. При попытке поиска адекватных способов отражения течения времени в онтологических моделях было выяснено, что на сегодня это проблема, которая не имеет однозначного решения.

В рамках цифровизации говорится о необходимости создания цифровых двойников [10] и о реализации в цифровых платформах «цифровых отражений» или «цифрового следа». Однако большинство поставщиков просто «перекрасили» имеющиеся решения и выдают их за истинных цифровых двойников и цифровые отражения. Оборудование, описанное в базе данных несколькими записями во взаимосвязанных таблицах, называют цифровым двойником, а историю изменения некоторых параметров – цифровым следом. Этого явно недостаточно для получения какого-либо нового качества от цифровизации. Более того, тиражирование такого подхода приведёт к огромным затратам, так как интернет вещей (а по-старому - АСУ ТП) сегодня может порождать огромную массу данных по значительному числу показателей. Складывать это в один «цифровой след» и затем обрабатывать его традиционным способом, значит существенно увеличивать совокупную стоимость владения не получив при этом какого-либо внятного бизнес-результата от такого потока.

Большинство разработчиков онтологических моделей архитектуры предприятия предпочитают не акцентировать эту проблему и закладывать время как один из пара-

метров или как характеристику при формировании факта. А события рассматриваются также как объекты со своим набором свойств, одним из которых является время, когда событие произошло. Например, в ЕРС событие – это состояние, которое определено на входе или выходе функции, заранее определённой, набором параметров, заранее определённых, в момент времени. Кроме того, что таковое определение не онтологично, оно не позволяет сформировать полноценный цифровой след, так как необходимы специальные средства для формирования ориентированного по времени потока взаимосвязанных событий над изменяющимся объектом разными субъектами. И эти средства как правило отсутствуют, потому что для каждого отдельного запроса их нужно создавать специально и на их исполнение в традиционной реляционной СУБД будет потрачено значительное количество вычислительных ресурсов. Поэтому все, имеющиеся в распоряжении крупных предприятий, платформы предлагают «срезовые» картинки на конкретный момент времени и аналитику по какому-либо параметру или набору параметров. Усложнив постановку тем, что составные части объекта и его параметры могут появляться и исчезать в течение жизненного цикла, при этом заранее весь набор составных частей и параметров неизвестен, мы сразу делаем задачу не выполнимой с помощью существующих программных платформ.

При этом онтологические подходы на верхнем уровне – в онтологии менеджмента – содержат фактор времени и обозначают это необходимым элементом [11], однако до реализации не доходят.

Другой очевидной крайностью онтологического подхода, в противовес «срезовому» объектному, является темпоральная онтология. Темпоральная онтология рассматривает мир как поток событий. Этот поток определяет объекты. Вещи «разложимы на события, представимы, как структуры событий, как нечто фиксируемое, оформляемое, конфигурируемое (и снаружи, и изнутри) событиями» (цитата из [12], стр 4). Таким образом появляется возможность рассматривать цифровой след на временной оси, но возникает другая проблема – из этого цифрового следа событий надо специальным образом получать объекты на конкретный момент или промежуток времени. Ведь события на промышленном предприятии всегда происходят на конкретном объекте с конкретным наблюдаемым состоянием.

Итак, мы имеем две крайности, одна из которых находится целиком в исследовательской зоне, другая является массовой парадигмой доцифрового прошлого. И ни одна из них не содержит цельного решения проблемы. Сделаю предположение, что одним из перспективных направлений для поиска решений является соединение двух подходов путём контекстуализации онтологических моделей как на конкретный состав объектов, так и на конкретный период времени. При чём темпоральная онтология больше подходит для реализации такого «окошка времени».

Таким образом, мы имеем значительное поле для прикладных исследований и разработок в данном направлении – реализации фактора времени в онтологических моделях процессов для архитектуры цифрового предприятия.

Работа выполнена в рамках проекта «Новые компьютерные модели в задачах бережливого производства цифровой экономики», поддержанного грантом Российского фонда фундаментальных исследований №18-47-860008p_a

ЛИТЕРАТУРА

1. Хайдеггер М. Бытие и время. СПб., 2006. N.Y., 1980.
2. Захарчук О.Т., Самарин А.В.. Построение исполняемой архитектуры предприятия реального времени // Инжиниринг предприятий и управление знаниями (ИП&УЗ-2017): сборник научных трудов XX юбилейной Всероссийской научной конференции. 26–28 апреля 2017 г. / под науч. ред. Ю. Ф. Тельнова: в 2 т. – Москва: ФГБОУ ВО «РЭУ им. Г. В. Плеханова», 2017. - с.236-244.
3. Захарчук О.Т. Фундаментальная модель описания организационно-технических систем.// Материалы международной научно-практической конференции «Теория активных систем» ИПУ РАН, Москва, 17-19 ноября 2014 г. – с. 249-255.
4. Green P. and Rosemann M. (2000) Integrated Process Modeling: An Ontological Evaluation, Information Systems, 25(2), 73-87.
5. Шведин Б.Я. Онтология предприятия. Экспериментологический подход., - Ленанд, 2010. – 240с.
6. Jan L.G. Dietz, Enterprise Ontology. Theory and Methodology, Springer, 2006.
7. Гимранов Р.Д., Холкин И.Н.. Изобретая информационные системы будущего. Теория и практика. – Сургут, 2017. – 192 с.
8. Wand, Y. and Weber, R. (1990b) Mario Bunge's Ontology as a formal foundation for information systems concepts, In Studies on Mario Bunge's Treatise, eds. P. Weingartner, G. J. W. Dorn, Rodopi. Atlanta, 123-149.
9. Чалей И.В., Каратаева Г.Е., Гимранов Р.Д. Математические и информационные модели цифрового предприятия на основе онтологий.//Инновационное развитие экономики. 2018. № 6-3 (48). С. 69-72.
10. Боровков А.И., Рябов Ю.А. О дорожной карте «Технет» (передовые производственные технологии) Национальной технологической инициативы. Трампин к успеху, 2017, № 10, с. 8-11.
11. Неаполитанский С.М. Онтологический менеджмент: От онтологии управления до управления онтологиями. Издательские решения, 2019. — 38 с.
12. Болдачев А. Введение в темпоральную онтологию//Кто сегодня делает философию в России. Том III / Автор составитель Нилогов А.С.. - М.: ООО «Сам Полиграфист», 2015.

Time factor in ontological business process models.

Rinat Gimranov
IT Division
PJSC Surgutneftegas
Surgut, Russia

Abstract. Sooner or later, ontologies will go “under the hood” of digital platforms, which, thanks to ontological models, will be able to “understand” the features and changes of both the real world and digital entities in real time. But, the current state of research and development does not yet allow a mature stack of ready-made solutions for enterprises to contain an interconnected set of theoretical foundations, modeling agreements, techniques and a software implementation environment. One of the

problems that requires work is the inclusion of time in ontological models and the corresponding change in the functionality of software solutions.

Keywords: ontological modeling, digital platform, temporal ontology, time in ontologies

Моделирование запасов метана в термокарстовых озерах на основе геоимитационного подхода и спутниковых снимков

Полищук Владимир Юрьевич
Институт мониторинга климатических и
экологических систем СО РАН,
пр. Академический 10/3, Томск, 634055,
Россия
liquid_metal@mail.ru

Полищук Юрий Михайлович
Югорский НИИ Информационных Технологий,
ул. Мира, 151, Ханты-Мансийск, 628011,
Россия
yupolishchuk@gmail.com

Аннотация. Рассмотрены вопросы моделирования полей термокарстовых озер в зоне многолетней мерзлоты в стратегии геоимитационного подхода. Кратко изложены особенности получения экспериментальных данных о свойствах реальных полей термокарстовых озер по спутниковым снимкам на территории арктической зоны Западной Сибири. Приведены данные о распределении случайных координат центров озер и распределении озер по их размерам. Рассмотрены алгоритмические вопросы геоимитационного моделирования полей озер с учетом равномерного распределения координат центров озер и логнормального распределения их размеров. Описана структура программного комплекса моделирования полей термокарстовых озер, представляющего собой совокупность программных модулей, которые обеспечивают ввод параметров модели, формирование псевдослучайных числовых последовательностей и вывод результатов моделирования. Программный комплекс реализован на языке высокого уровня Python 3.8, позволяющем отображать результаты моделирования полей озер с использованием средств современных геоинформационных систем. Представлен фрагмент модельного поля термокарстовых озер как результат вычислительного эксперимента с моделью с использованием программного комплекса геоимитационного моделирования.

Ключевые слова: геоимитационное моделирование, программный комплекс, многолетняя мерзлота, термокарстовые озера, метан, глобальное потепление, логнормальный закон распределения.

ВВЕДЕНИЕ

Одной из главных проблем современности стало потепление климата, создающие глобальные вызовы мировому сообществу в последние десятилетия. Проведенный в связи с этим Мировой Климатический саммит (Париж, 2015) принял рекомендации всем странам мира по разработке мер, способных не допустить повышение среднегодовой температуры Земли более чем на 2 °C до 2050 года. Однако разработка мер по недопущению повышения температуры более чем на 2 °C до 2050 года на территории России, более 60% которой располагается в зоне мерзлоты, невозможна без формирования обоснованных прогнозов изменения объемов эмиссии метана и углекислого газа, для разработки которых необходимы знания о динамике полей термокарстовых озер на территории мерзлоты. В связи с этим актуальность разработки геоинформационной техно-

логии оценки и прогноза динамики полей термокарстовых озер в зонах мерзлоты не вызывает сомнения.

Вопросы моделирования термокарстовых озер рассматривались в работах Веста и Плуга, Субина и Райли, Дюге, Хостетлера и др., обзор которых приведен в [1], где показано, что использование аналитических моделей позволяет достаточно эффективно исследовать различные процессы в одиночных термокарстовых озерах. Однако такие модели не пригодны для исследования пространственно-временных изменений полей термокарстовых озер на территории многолетней мерзлоты. С точки зрения моделирования полей термокарстовых озер интерес вызывают разрабатываемые А.С. Викторовым [2, 3] методы математической морфологии ландшафта, ориентированные на использование аналитических моделей для моделирования динамики территорий. Методы и модели математической морфологии ландшафта могут использоваться для оценки геоэкологических рисков при развитии термокарстовых процессов на территориях в зоне мерзлоты. Однако эти методы также не предназначены для исследования пространственно-временных изменений полей термокарстовых озер в условиях изменяющегося климата. В связи с этим возникает необходимость в разработке методов и средств математического моделирования динамики термокарстовых процессов, позволяющих путем проведения компьютерных экспериментов прогнозировать изменения состояния полей термокарстовых озер с учетом климатических изменений.

Поэтому перспективным рассматривается использование имитационных моделей [4], позволяющих исследовать динамику полей термокарстовых озер в условиях современного глобального потепления. В последнее время в рамках идеологии имитационного моделирования сформировалось одно из новых направлений компьютерного моделирования – геоимитационное моделирование [5, 6], понимаемое как имитационное моделирование сложных объектов с пространственной структурой и реализуемое с использованием методов и средств геоинформатики [7]. В [8, 9] предложена геоимитационная модель поля термокарстовых озер в виде совокупности случайных окружностей, свойства которой учитывают основные статистические характеристики реальных полей термокарстовых озер, определенные по экспериментальным данным спутниковых измерений. В частности, в этих исследованиях было установлено, что распределение озер по размерам подчи-

В.Ю. Полищук, Ю.М. Полищук

няется экспоненциальному закону, которое и было учтено при разработке указанной модели.

Однако разработанная в [8, 9] модель была основана на результатах дистанционных исследований свойств полей термокарстовых озер по спутниковым снимкам среднего разрешения (30 м) Landsat, на которых видны озера сравнительно больших размеров (от 0,5 га). В более поздних наших исследованиях термокарстовых озер [10 - 13] с использованием снимков высокого (1-10 м) и сверхвысокого разрешения (менее 1 м) было показано наличие в реальных полях термокарстовых озер в зоне мерзлоты (на примере Западной Сибири) огромного числа озер малых размеров, которые не обнаруживаются на снимках среднего разрешения. Разработанная в [10, 11] методика интеграции данных со снимков среднего и высокого/сверхвысокого разрешения позволила построить синтезированные гистограммы распределения озер по площадям в очень широком диапазоне размеров (от единиц м² до десятков км²). Анализ этих гистограмм показал [12, 13], что распределение реальных озер в зоне мерзлоты удовлетворяет не экспоненциальному, а логнормальному закону с высокой вероятностью (0,99).

В связи с изложенным выше возникла необходимость модификации модели и совершенствования системы геоимитационного моделирования полей термокарстовых озер, направленных на учет логнормального закона распределения их по размерам. Однако вопросы программно-алгоритмической реализации системы геоимитационного моделирования полей термокарстовых озер с учетом логнормального закона в настоящее время рассмотрены недостаточно, что и определило цель настоящей работы.

1. ЭКСПЕРИМЕНТАЛЬНЫЕ ДАННЫЕ О ЗАКОНАХ РАСПРЕДЕЛЕНИЯ ОЗЕР

Информационной основой для экспериментального изучения свойств полей термокарстовых озер являются данные дистанционного измерения площадей и периметров озер по спутниковым снимкам исследуемой территории. Исследования проводились на территории всех трех зон вечной мерзлоты (островной, прерывистой и сплошной) в Западной Сибири на основе спутниковых снимков среднего и высокого разрешения. Все снимки были отобраны в довольно короткий период летнего сезона (конец июня - август), чтобы минимизировать влияние сезонных колебаний уровня воды в озерах. В этот период полностью исчезает ледяной покров на озерах, не позволяющий выделять их при дешифрировании снимков.

Создание геоимитационной модели полей термокарстовых озер требует знания основных свойств этих полей, которые можно получить экспериментально по спутниковым снимкам. Анализ гистограмм распределения координат расположения центров озер по данным со спутника Landsat показал [8,9], что эмпирические гистограммы распределения координат центров озер соответствуют, согласно критерию χ^2 , закону равномерной плотности с вероятностью 95% [14]. Проведено экспериментальное исследование свойств извилистости береговых линий реальных озер.

Одной из важных для создания геоимитационной модели полей термокарстовых озер является задача построения гистограмм распределения озер по размерам, которые бы учитывали все озера в широком диапазоне размеров - от десятков метров до десятков километров. Вопросы постро-

ения такой гистограммы рассмотрены в [11,12], где было предложено выбрать частичные интервалы гистограммы с нерегулярным шагом (согласно логарифмическому закону), а именно: 20-50 м², 50-100 м², 100-200 м² и т. д. до 200 км². Это позволило достаточно компактно представить данные о распределении озер по интервалам их размеров в очень широком диапазоне изменений площадей озер.

С использованием разработанной в [11] методики интеграции данных о числе и площадях водоемов, полученных по спутниковым снимкам как среднего, так и высокого разрешения на основе объединения (синтеза) данных о числе и площадях озер, полученных по снимкам различного пространственного разрешения, была построена синтезированная гистограмма распределения озер по площадям в очень широком диапазоне их размеров. В соответствии с этой методикой синтезированная гистограмма распределения озер по площадям была получена путем «сшивания» двух исходных гистограмм, первая из которых основана на данных со снимков среднего разрешения (30 м) Landsat-8 и представляет большие озера (размером от 0,5 до 20 000 га). Вторая исходная гистограмма, полученная по данным со снимков высокого разрешения (2 м) Канопус-В на 78 тестовых участках в зоне вечной мерзлоты, включает малые озера. Синтезированная гистограмма распределения озер по размерам, полученная выше описанным образом по результатам проведенных исследований в зоне мерзлоты Западной Сибири, представлена в [13].

Согласно [14], плотность вероятности для логнормального закона распределения площади озер (s) определяется уравнением:

$$f(s) = \frac{1}{\sqrt{2\pi}\sigma s} \exp\left(-\frac{(\ln s - \mu)^2}{2\sigma^2}\right), \quad (1)$$

где μ - математическое ожидание, σ - среднеквадратическое отклонение.

Оценки параметров μ и σ теоретического логнормального закона распределения озер по размерам, полученные в [13] из выборочных данных для зоны мерзлоты Западной Сибири, принимают следующие значения: 6,88 и 3,24 для математического ожидания и дисперсии соответственно. Выполненная в [13] проверка соответствия эмпирического и теоретического распределений с использованием критерия Пирсона показала, что величина расчетного χ^2 равна 0,23 и, следовательно, синтезированная гистограмма распределения площадей озер в широком диапазоне их размеров может быть принята соответствующей логнормальному закону с высокой вероятностью 0,99.

Проведен корреляционный анализ многомерного массива экспериментальных данных об изменениях координат центров, степени извилистости и площадей термокарстовых озер, полученных по спутниковым снимкам. Анализ показал, что коэффициенты взаимной корреляции между всеми указанными выше показателями достаточно малы, что позволило сделать вывод [9] о статистической независимости перечисленных выше показателей полей термокарстовых озер на исследуемой территории мерзлоты Западной Сибири.

2. АЛГОРИТМЫ МОДЕЛИРОВАНИЯ

Реализация математической модели полей термокарстовых озер предполагает создание генератора псевдослу-

чайных числовых последовательностей - троек, первые два из которых распределены по равномерному закону, а третьи - по логнормальному закону. В общем случае математическая модель полей термокарстовых озер может быть представлена в следующем виде:

$$f(x, y, s), \quad (2)$$

где x и y координаты центров озер, s – площадь озер.

Следовательно, совокупность окружностей представляется совокупностью троек случайных величин (x, y, s) .

Для разработки алгоритма необходимо учесть отсутствие статистической связи между изменениями координат озер и их площадями. Анализ, проведенный в [18], показывает, что случайные пространственные изменения площадей и координат центров озер, так же, как и координат центров озер между собой, статистически независимы. Поэтому для моделирования полей термокарстовых озер совместная плотность вероятностей (3) может быть представлена в виде:

$$f(x, y, s) = f(x) \times f(y) \times f(s), \quad (3)$$

где $f(x)$ и $f(y)$ - плотности вероятности равномерного распределения, $f(s)$ - плотность вероятности для логнормального закона распределения.

Согласно [14], плотность вероятности для логнормального закона распределения площади озер (s) определяется уравнением (1).

В нашем случае оценки математического ожидания (M) и дисперсии (D) для логнормального распределения площадей озер Западной Сибири могут быть получены из выборочных данных в виде:

$$M = \sum_{i=1}^m (\ln s_i \times w_i), \quad (4)$$

$$D = \sum_{i=1}^m ((\ln s_i)^2 \times w_i) - M^2, \quad (5)$$

где m – число частичных интервалов эмпирической гистограммы, i - номер интервала.

С учетом уравнения (3) генерирование последовательности случайных чисел, определяющих характеристики местоположения центров окружностей, проводится с использованием датчика псевдослучайных чисел, распределенных по закону равномерной плотности. Необходимо отметить, что для моделирования озер со случайными размерами, площади которых распределены по закону, соответствующему уравнению (1), будем использовать генерирование последовательности случайных чисел, распределенных по логнормальному закону согласно уравнению:

$$s_i = \exp(\mu + \sigma \times r_i), \quad (6)$$

где r_i – случайное число, распределенное по нормальному закону и рассчитываемое по формуле:

$$r_i = \sum_{j=1}^{12} q_{ij} - 6, \quad (7)$$

где q_{ij} - случайная величина, равномерно распределенная на отрезке $[0,1]$

В соответствии с выше изложенным, алгоритм численного моделирования полей термокарстовых озер может быть представлен в виде следующей последовательности:

шаг 1. задается количество моделируемых окружностей N ;

шаг 2. задается значение параметра логнормального распределения μ ;

шаг 3. задается значение параметра логнормального распределения σ ;

шаг 4. генерируется псевдослучайная величина, распределенная по нормальному закону (7);

шаг 5. с использованием заданных параметров и рассчитанной псевдослучайной величины, по формуле рассчитывается псевдослучайная величина площади окружности, распределенная по логнормальному закону (6);

шаг 6. генерируется пара псевдослучайных чисел, распределенных по равномерному закону, определяющих местоположение центра окружности на экране;

шаг 7. расчет объема накопления метана в термокарстовых озерах;

шаг 8. с использованием тройки чисел, полученных на предыдущих шагах 5 и 6, формируется окружность;

шаг 9. если количество полученных окружностей меньше определенного на шаге 1 числа, то алгоритм повторяется с шага 4, иначе алгоритм завершён.

Также алгоритм численного моделирования полей термокарстовых озер приведен в виде блок-схемы (рис. 1).

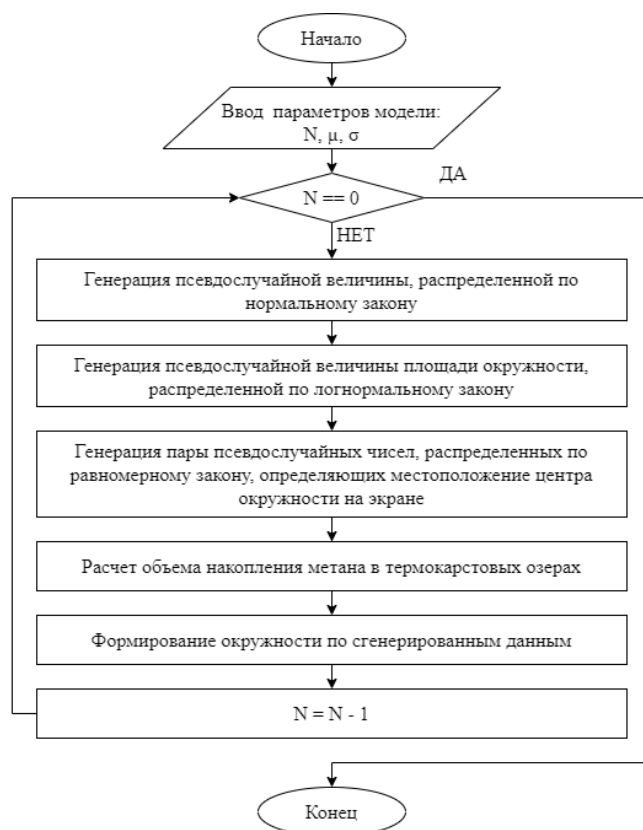


Рис. 1 Схема алгоритма численного моделирования полей термокарстовых озер

РАСЧЕТ ОБЪЕМА НАКОПЛЕНИЯ МЕТАНА В
ТЕРМОКАРСТОВЫХ ОЗЕРАХ

В.Ю. Полищук, Ю.М. Полищук

Моделирование запасов метана в термокарстовых озерах на основе геоимитационного подхода
и спутниковых снимков

Для определения объемов накопления метана в термокарстовых озерах требуется знать его концентрацию в воде каждого озера и его площадь, и глубину. Результаты наземных биогеохимических исследований концентрации метана в термокарстовых озерах, которые проводились в зоне мерзлоты Западной Сибири в период 2008–2018 гг., опубликованы в работах О.С. Покровского с соавторами [18].

Важными для решения поставленной задачи являются вопросы исследования взаимосвязи между концентрацией метана в воде термокарстовых озер и их площадями. В [18,19] на основе анализа результатов биогеохимических исследований состава воды термокарстовых озер установлено, что концентрация метана существенно зависит от размеров озер: в малых озерах с площадью менее нескольких сотен квадратных метров она более чем на порядок превышает ее величину в больших озерах с размерами, превышающими указанные выше.

По данным [20], глубина малых и больших озер в среднем составляет 0,34 и 0,54 м соответственно. Эти величины могут быть использованы при оценке запасов метана в термокарстовых озерах на основе моделирования. Масса метана, накопленного в n озерах, расположенных на рассматриваемой территории мерзлоты, может быть рассчитана в виде:

$$P = \sum_{i=1}^n p_i, \quad i = \overline{1, n}, \quad (8)$$

где p_i – масса метана в i -м озере.

Масса накопленного метана в каждом i -м озере может быть определена по формуле:

$$p_i = C_i \times h_i \times S_i, \quad (9)$$

где C_i – концентрация метана в воде i -го озера; h_i – глубина i -го озера; S_i – площадь i -го озера.

Процедура оценки объемов накопленного метана в водных термокарстовых объектах на заданной территории мерзлоты на основе геоимитационного моделирования заключается в следующем. В процессе моделирования полей этих объектов находится площадь каждого i -го модельного объекта. В зависимости от величины площади рассчитывается в соответствии с (9) масса метана в данном i -м модельном озере. Затем по формуле (8) рассчитывается общая масса накопленного метана во всех модельных озерах на заданной территории мерзлоты.

3 ПРОГРАММНАЯ РЕАЛИЗАЦИЯ СИСТЕМЫ МОДЕЛИРОВАНИЯ

Структурная схема программного комплекса геоимитационного моделирования представлена на рис. 2. Разработанный программный комплекс геоимитационного моделирования полей термокарстовых озер, представляет собой совокупность программных модулей, обеспечивающих ввод параметров модели, формирование псевдослучайных числовых последовательностей и вывод результатов моделирования. На рис. 2 схема программного комплекса представлена в виде следующих составных частей:

- Графический интерфейс предназначен для выбора пользователем параметров модели, закона распределения и вида представления результатов моделирования;

- Блок модулей генерации псевдослучайных величин является основной составной частью программного комплекса имитационного моделирования и предназначен для формирования случайных числовых последовательностей при реализации алгоритмов численного моделирования полей термокарстовых озер;

- Модуль представления результатов моделирования предназначен для преобразования результатов моделирования в один из следующих форматов: Microsoft Excel (*.xlsx), векторный формат (*.shp), растровый формат (*.jpeg).

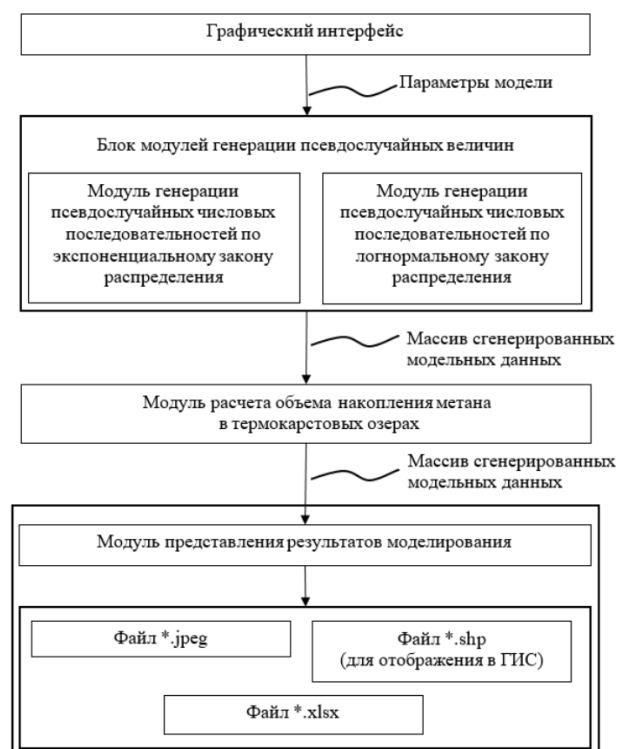


Рис. 2. Обобщенная схема программного комплекса геоимитационного моделирования полей термокарстовых озер

Представленный в виде схемы (рис. 2) программный комплекс реализован на языке высокого уровня – Python [15]. Python – интерпретируемый язык программирования. Это означает, что исходный код частями преобразуется в машинный в процессе его чтения специальной программой – интерпретатором. Python является полноценным, универсальным и используемым в различных сферах языком программирования. Основная, но не единственная, поддерживаемая им парадигма – объектно-ориентированное программирование. Фактически круг ролей, которые может играть Python как многоцелевой язык программирования, практически не ограничен в возможностях. Он может использоваться для реализации, в том числе, и программных комплексов. Интерпретаторы Python распространяются свободно на основании лицензии, подобной GNU General Public License [16].

Для демонстрации работы созданного программного комплекса на рис. 3 приведен фрагмент модельного поля термокарстовых озер. Численные значения параметров модели, использованные для моделирования: $M = 6,88$, $D = 3,42$ и $N = 3000$.

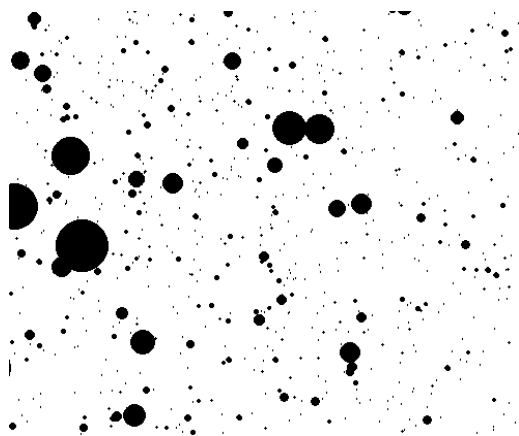


Рис. 3. Фрагмент модельного поля термокарстовых озер, распределенных по логнормальному закону.

ЗАКЛЮЧЕНИЕ

Известно [17], что термокарстовые озера арктических регионов рассматриваются в качестве одного из основных источников эмиссии природного метана. Из них наиболее активными генераторами метана являются малые озера с размерами менее 500-1000 м², концентрация метана в которых, по данным [18, 19], более чем на порядок превышает концентрацию метана в более крупных озерах. Поэтому использование в модели логнормального закона распределения, позволяющего учесть при моделировании малые озера, делают разработанную систему геоимитационного моделирования удобным инструментом для получения оценки запасов и эмиссии метана в термокарстовых озерах. Система геоимитационного моделирования может также использоваться для получения прогнозных оценок динамики эмиссии метана из термокарстовых озер на ближайшие десятилетия.

БЛАГОДАРНОСТИ

Работа проводилась частично при поддержке грантов по проектам РФФИ № 18-47-700001 и 18-45-860002.

ЛИТЕРАТУРА

1. Modeling thermokarst lakes dynamics and carbon flux – URL: http://www.docstoc.com/docs/35351317/Methods_of_the_rmokarst_lakes_modeling, свободный (дата обращения: 02.09.2012).
2. Викторов А.С. Основные проблемы математической морфологии ландшафта: монография / А.С. Викторов. – М.: Наука, 2006. – 252 с.
3. Викторов А.С. Математическая морфология ландшафтов криолитозоны: монография / А.С. Викторов, Т.В. Орлов, В.Н. Капралова, О.Н. Трапезникова, М.В. Архипов, С.А. Садков, А.В. Зверев, Е.Г. Панченко, П.В. Березин. – М.: РУДН, 2016. – 232 с.
4. Моисеев Н.Н. Системный анализ динамических процессов биосферы. Концептуальная модель биосферы / Н.Н. Моисеев, Ю.М. Свирижев // Вестник АН СССР. – 1979. – № 2. – С. 47-54.
5. Полищук Ю.М. Геоимитационное моделирование зон атмосферного загрязнения в результате сжигания газа на нефтяных месторождениях / Ю.М. Полищук, О.С. Токарева // Информационные системы и технологии, 2010. – № 2. – С. 39-46.

6. Zhao Y. Urban dynamics analysis using spatial metrics geosimulation. In Book: Spatial analysis and modeling in geographical transformation process / Y. Zhao, Y. Murayama. – Dordrecht-Heidelberg-New York-London: Springer, The Geo-Journal Library (Eds. Y. Murayama, R. Thapa), 2011. – Vol. 100. – P. 153-168.
7. Капралов Е.Г. Основы геоинформатики: учеб. пособие для вузов в 2 кн. / Е.Г. Капралов, А.В. Кошкарёв, В.С. Тикунов. – М.: Изд. центр «Академия», 2004. – 352 с.
8. Полищук В.Ю. Геоимитационное моделирование полей термокарстовых озер в зонах мерзлоты: монография / В.Ю. Полищук, Ю.М. Полищук. – Ханты-Мансийск: УИП ЮГУ, 2013. – 129 с.
9. Polishchuk V.Y. Modeling of thermokarst lake dynamics in West-Siberian permafrost. In book: Permafrost: Distribution, Composition and Impacts on Infrastructure and Ecosystem. – Ed. O.S. Pokrovsky / V.Y. Polishchuk, Y.M. Polishchuk. – New York: Nova Science Publishers, 2014. – Is. 6. – P. 205-234.
10. Полищук Ю.М. Методические вопросы построения обобщенных гистограмм распределения площадей озер в зоне мерзлоты на основе космических снимков среднего и высокого разрешения / Ю.М. Полищук, А.Н. Богданов, И.Н. Муратов // Современные проблемы дистанционного зондирования Земли из космоса. – 2016. – том 13. – № 6. – С. 224-232.
11. Полищук Ю.М. Интеграция космических снимков сверхвысокого и среднего разрешения для построения гистограмм распределения площадей термокарстовых озёр в расширенном диапазоне их размеров / Ю.М. Полищук, А.Н. Богданов, Н.А. Брыксина, В.Ю. Полищук, И.Н. Муратов // Современные проблемы дистанционного зондирования Земли из космоса. – 2018. – том 15. – № 3. – С. 38-46.
12. Полищук Ю.М. Опыт и результаты дистанционного исследования озёр криолитозоны Западной Сибири по космическим снимкам различного разрешения за 50-летний период / Ю.М. Полищук, А.Н. Богданов, Н.А. Брыксина, И.Н. Муратов, В.Ю. Полищук, М.А. Куприянов, О.А. Байсаямова, В.П. Днепровская // Современные проблемы дистанционного зондирования Земли из космоса. – 2017. – том 14. – № 6. – С. 42-55.
13. Полищук Ю.М. Распределение площадей озёр криолитозоны в широком диапазоне их размеров по космическим снимкам среднего и высокого разрешения / Ю.М. Полищук, А.Н. Богданов, В.Ю. Полищук // Известия Томского политехнического университета. Инжиниринг георесурсов. – 2018. – том 329. – № 3. – С. 16-25.
14. Кремер Н.Ш. Теория вероятностей и математическая статистика: учебник для вузов / Н.Ш. Кремер. – М.: ЮНИТИ-ДАНА, 2004. – 573 с.
15. Лутц М. Изучаем Python. – 4-е издание, пер. с англ. - М.– СПб.: Символ-Плюс, 2011. – 1280 с.
16. GNU General Public License – URL: https://ru.wikipedia.org/wiki/GNU_General_Public_License (дата обращения: 15.02.2020).
17. Walter K.M. Methane bubbling from northern lakes: present and future contributions to the global methane budget / K.M. Walter, L.C. Smith, F.S. Chapin // Philosophical Transactions of the Royal Society. – 2007. – vol. 365. – P. 1657-1676.
18. Pokrovsky O.S. Effect of permafrost thawing on the organic carbon and metal speciation in thermokarst lakes of Western Siberia / O.S. Pokrovsky, L.S. Shirokova, S.N. Kirpotin, S.

В.Ю. Полищук, Ю.М. Полищук

Моделирование запасов метана в термокарстовых озерах на основе геоимитационного подхода и спутниковых снимков

Audry, J. Viers, B. Dupre // Biogeosciences. – 2011. – vol. 8. – P. 565-583.
19. Holgerson M.A. Large contribution to inland water CO₂ and CH₄ emissions from very small ponds / M.A. Holgerson, P.A. Raymond // Nature Geoscience Letters. – 2016. – vol. 9. – P. 222-226.

20. Полищук Ю.М. Методические вопросы оценки запасов метана в малых термокарстовых озерах криолитозоны Западной Сибири / Ю.М. Полищук, В.Ю. Полищук, Н.А. Брыксина, О.С. Покровский, С.Н. Кирпотин, Л.С. Широкова // Известия Томского политехнического университета, 2015. – Т. 326. – № 2. – С. 127-135.

Modeling methane reserves in thermokarst lakes based on a geo-simulation approach and satellite imagery

Vladimir Y. Polishchuk

Institute of Monitoring of Climatic and Ecological Systems,
10/3 Akademicheskoy Avenue, Tomsk 634055, Russia,
liquid_metal@mail.ru

Yury M. Polishchuk

Ugra Research Institute of Information Technologies,
st. Mira, 151, Khanty-Mansiysk, 628011, Russia
yupolishchuk@gmail.com

Abstract. The issues of modeling the fields of thermokarst lakes in the permafrost zone in the strategy of the geo-simulation approach are considered. The features of obtaining experimental data on the properties of real fields of thermokarst lakes from satellite images in the Arctic zone of Western Siberia are briefly described. Data on the distribution of random coordinates of the lakes centers and the distribution of lakes by their size are presented. Algorithmic issues of the geo-simulation modeling of lake fields are considered, taking into account the uniform distribution of the coordinates of the lakes centers and the lognormal distribution of their sizes. The structure of the software

package for modeling fields of thermokarst lakes is described, which is a set of software modules that provide input of model parameters, formation of pseudorandom numerical sequences and output of simulation results. The software package is implemented in Python 3.8, which allows displaying the results of modeling lake fields using modern geographic information systems. A fragment of the model field of thermokarst lakes is presented as a result of a computational experiment using a software package for geo-simulation modeling of lakes fields.

Keywords: geo-simulation, modeling, software package, permafrost, thermokarst lakes, global warming, methane, lognormal distribution law.

Обратные задачи в теории фильтрации и тепломассопереноса

С.Г. Пятков

Югорский государственный университет
Ханты-Мансийск, Россия

s_pyatkov@ugrasu.ru, v_rotko@ugrasu.ru

В. В. Ротко

Югорский государственный университет
Ханты-Мансийск, Россия

s_pyatkov@ugrasu.ru, v_rotko@ugrasu.ru

Аннотация. Данная работа является обзорной. Мы описываем результаты о корректности в пространствах Соболева обратных задач с точечными условиями переопределения для математических моделей конвекции-диффузии и фильтрации, описываемых при помощи параболических уравнений и систем. Неизвестные функции, зависящие от времени, входят в функцию источника и в сам оператор в качестве коэффициентов. В качестве условий переопределения рассматриваются значения решения в некотором наборе внутренних точек области. Приводятся условия, гарантирующие локальную по времени корректность задачи в классах Соболева. Мы описываем известные результаты и, в частности, приводим последние результаты авторов. Приведенные результаты являются во многих случаях точными как в смысле гладкости данных, так и в смысле минимальности используемых условий.

Ключевые слова: параболическая система, обратная задача, функция источника, конвекция-диффузия, тепломассоперенос, фильтрация.

ВВЕДЕНИЕ

Мы рассматриваем вопрос о определении вместе с решением функции источника и коэффициентов, характеризующих параметры среды в квазилинейных математических моделях конвекции-диффузии. Пусть G – область в \mathbb{R}^n с границей Γ класса C^2 и $Q = (0, T) \times G$. Соответствующая параболическая система имеет вид

$$Lu = u_t + A(t, x, D)u - f(t, x, u, \nabla u) = \sum_{i=1}^r f_i(t, x)q_i(t) + f_0(t, x), \quad (t, x) \in Q, \quad (1)$$

где A – матричный эллиптический оператор вида $A(t, x, D)u = -\sum_{i,j=1}^n a_{ij}(t, x)u_{x_j x_i} + \sum_{i=1}^n a_i(t, x)u_{x_i} + a_0(t, x)u$, a_{ij}, a_i – $h \times h$ матрицы, u – вектор длины h и $\vec{q}(t) = (q_1(t), q_2(t), \dots, q_m(t))$ – неизвестные функции подлежащие определению вместе с решением u , которые входят как в правую часть, так и в сам оператор A как коэффициенты. Система (1) дополняется начальными и граничными условиями

$$u|_{t=0} = u_0, \quad Bu|_S = g, \quad (2)$$

где $S = (0, T) \times \Gamma$, $Bu = \sum_{i=1}^n \gamma_i(t, x)u_{x_i} + \sigma(t, x)u$ или $Bu = u$ и $\gamma_i(t, x), \sigma(t, x)$ матрицы-функции размерности $h \times h$. Условия переопределения записываются в виде

$$u(x_i, t) = \psi_i(t), \quad i = 1, 2, \dots, s, m = hs. \quad (3)$$

Таким образом, дополнительные условия являются данными замеров решения (например, концентрации переносимого вещества) в определенных точках области. По этим данным, возможно определение как самих источников (например, источников загрязнения в водоеме или воздухе), так и параметров среды. Таким образом, обратная задача состоит в нахождении решения u уравнения (1) и функций $q_i(t)$ ($i = 1, 2, \dots, m$), входящих в правую часть или сам оператор A по данным (2), (3). В задачах тепломассопереноса и фильтрации правая часть в (1) характеризует распределение источников и их интенсивности. В случае точечных источников, т.е. $f_i = \delta(x - x_i)$ (δ – дельта-функция Дирака), q_i – интенсивность i -го источника в задачах тепломассопереноса (см. [1]), и в задачах фильтрации, например, в задачах нефтедобычи q_i – дебит i -й скважины, в этом случае u – давление (см. [2]). В различных практических задачах используются как распределенные источники, так и точечные источники.

Опишем вначале некоторые результаты, посвященные задачам с пространственно распределенными источниками. Большое количество результатов было получено в случае линейной по своим аргументам функции f . Можно отметить работу [3], где получена теорема существования и единственности решений задачи (1)-(3) об определении источника в пространствах Гельдера в случае $h = 1, r = 1$. Аналогичные результаты в случае задачи об определении функции источника и в случае более общих коэффициентных задач получены в монографии [4] но в одномерной ситуации ($n=1$). В работе [5] рассмотрена задача об определении младшего коэффициента в параболическом уравнении, а в работе [6] определению подлежат младший коэффициент и правая часть вида $q(t)f(t, x)$. В обоих случаях доказана корректность соответствующей обратной задачи. Имеется много работ посвященных модельным уравнениям вида (1) в основном в одномерной ситуации (см., например, [7,8]). Первая наиболее существенная работа посвященная квазилинейному случаю – это работа [9], где были приведены условия на нелинейную функцию гарантирующую глобальную по времени разрешимость задачи (1)-(3) в пространствах Гельдера в случае $r=1$. Это условие не более чем линейного роста функции f по своим аргументам $u, \nabla u$. Аналогичный результат уже в случае параболической системы и в пространствах Соболева был

получен в работе авторов [10]. Вопрос локальной корректности задачи (1)-(3) в пространствах Соболева были далее рассмотрены в работах авторов [11]-[13]. Далее, мы отметим монографию [14], где рассмотрены обратные задачи для квазилинейных уравнений вида (1) в абстрактном случае, т.е. для операторно-дифференциального уравнения первого порядка. Результаты достаточно интересные, однако во многих случаях их применение не дает оптимальных условий на данные на данные и они не всегда применимы. Некоторые уточнения этих результатов в случае пространств Соболева были получены в работе [15].

В монографии [16] рассмотрены вопросы корректности более общих классов обратных задач, когда данные переопределения задаются на сечениях пространственной области или на некоторых пространственных многообразиях [17], [18]. Задача (1)-(3) часто является частным случаем в таких постановках.

Огромное количество работ посвящено численному решению задач вида (1)-(3). Отметим, например, работы [19]-[21]. Имеется большое количество монографий, посвященных численным методам решения подобных обратных задач. Практически все постановки и большое количество методов рассмотрены в монографии [22] в случае $n=1$. Более общей ситуации посвящены монографии [23], [24], ряд интересных постановок и задач (в том числе задачи конвективного теплообмена) рассматриваются в [25,26].

Перейдем к описанию некоторых результатов в случае точечных источников. Как уже было отмечено эти задачи не являются корректными в классах конечной гладкости и результатов, посвященных теоремам существования решений, практически нет (см. [27]).

Имеется огромное количество работ, посвященных численному решению задачи определения точечных источников, однако, как правило, эти работы не содержат никаких теоретических обоснований и очень часто в соответствующих постановках при определенных значениях параметров может иметь место как несуществование решений так и их неединственность. Примерами таких работ являются, например, работы [28]-[30]. Выделим работы, где все же имеется некоторое теоретическое обоснование и обоснованные алгоритмы нахождения решений [31-35]. Наиболее интересная идея построения источников изложена в работе [36]. Она впоследствии была использована в [34]. Отметим, что задачи определения точечных источников являются нелинейными в отличие от случая распределенных источников. Здесь необходимо определять число источников, их местоположение и интенсивности.

Опишем содержание работы. Вначале мы вводим некоторые обозначения. Затем описываем основные результаты по определению распределенных источников. Далее приводим некоторые описания алгоритмов нахождения точечных источников в простейшем случае. Обозначения функциональных пространств стандартные (см., например, [12]).

ОПРЕДЕЛЕНИЯ И ОБОЗНАЧЕНИЯ

Пусть E – банахово пространство. Символом $L_p(G; E)$

(G – область в \mathbb{R}^n) обозначаем пространство сильно измеримых функций определенных на G со значениями в E наделенное нормой $\| \|u(x)\|_E \|_{L_p(G)}$ [37]. Мы также используем пространства Гельдера $C^\alpha(\bar{G})$. Обозначения $W_p^s(G; E)$, $W_p^s(Q; E)$ пространств Соболева являются стандартными (см. определения в [37]). Если $E = \mathbb{C}$ ($E = \mathbb{R}$) или $E = \mathbb{C}^n$ ($E = \mathbb{R}^n$) тогда последнее пространство обозначается через $W_p^s(Q)$. Аналогично, используем обозначения $W_p^s(G)$ или $C^\alpha(\bar{G})$ вместо $W_p^s(G; E)$ или $C^\alpha(\bar{G}; E)$. Таким образом, включение $u \in W_p^s(G)$ (или $u \in C^\alpha(\bar{G})$) для данной вектор-функции $u = (u_1, u_2, \dots, u_k)$ означает, что каждая из ее компонент u_i принадлежит $W_p^s(G)$ (или $C^\alpha(\bar{G})$). В этом случае норма вектора есть просто сумма норм координат. То же самое соглашение принимаем для матриц-функций. Для интервала $J = (0, T)$ положим $W_p^{s,r}(Q) = W_p^r(J; L_p(G)) \cap L_p(J; W_p^s(G))$. Соответственно, $W_p^{s,r}(S) = W_p^r(J; L_p(\Gamma)) \cap L_p(J; W_p^s(G))$. Аналогично определяем пространство Гельдера $C^{r,s}(\bar{Q})$.

Пусть $B_\delta(x_i)$ – шар радиуса δ с центром в точке x_i . Далее мы используем следующие обозначения: $Q^\tau = (0, \tau) \times G$, $S^\gamma = (0, \gamma) \times \Gamma$. Дан набор точек $\{x_j\}$ из (3), параметр $\delta > 0$ назовем допустимым, если $\bar{B}_\delta(x_i) \subset G$, $\bar{B}_\delta(x_i) \cap \bar{B}_\delta(x_j) = \emptyset$ для $i \neq j$, $i, j = 1, 2, \dots, r$. Пусть $G_\delta = \cup_i B_\delta(x_i)$, $Q_\delta = (0, T) \times G_\delta$, $Q_\delta^\tau = (0, \tau) \times G_\delta$.

РАСПРЕДЕЛЕННЫЕ ИСТОЧНИКИ

Мы будем предполагать, что у нас выполнены условия

Условия на коэффициенты:

$$a_{ij} \in C(\bar{Q}), a_k \in L_p(Q), \quad p > n + 2; \quad (4)$$

$$\begin{aligned} a_{ij} &\in L_\infty(0, T; W_p^s(G_\delta)), \\ a_k &\in L_p(0, T; W_p^s(G_\delta)), i, j = 1, 2, \dots, n, k \\ &= 0, 1, \dots, n, \end{aligned} \quad (5)$$

для некоторого допустимого $\delta > 0$ и $s \in (n/p, 1]$. Опишем условия параболичности оператора L . Рассмотрим матрицу $A_0(t, x, \xi) = \sum_{i,j=1}^n a_{ij}(t, x) \xi_i \xi_j$ ($\xi \in \mathbb{R}^n$), и предположим, что найдется постоянная $\delta_1 > 0$ такая, что корни p полинома $\det(A_0(t, x, i\xi) + pE) = 0$ (E – единичная матрица) удовлетворяют условию

$$\begin{aligned} \operatorname{Re} p &\leq -\delta_1 |\xi|^2 \quad \forall \xi \in \mathbb{R}^n, \\ &\forall (x, t) \in Q. \end{aligned} \quad (6)$$

Пусть $B_0 u = u$ в случае условий Дирихле в (2) и $B_0 u = \sum_{j=1}^n \gamma_j \partial_{x_j} u$ в противном случае. Условие Лопатинского может быть записано в виде: для любой точки $(t_0, x_0) \in S$, и операторов $A(x, t, D)$ и $B_0(x, t, D)$, записанных в локальной системе координат y в этой точке (ось y_n направлена по нормали к S и оси y_1, \dots, y_{n-1} лежат в касательной плоскости в точке (x_0, t_0)), система

$$\begin{aligned} (\lambda E + A_0(x_0, t_0, i\xi', \partial_{y_n}))v(z) &= 0, \\ B_0(x_0, t_0, i\xi', \partial_{y_n})v(0) &= h_j, \end{aligned} \quad (7)$$

где $\xi' = (\xi_1, \dots, \xi_{n-1})$, $y_n \in \mathbb{R}^+$, имеет единственное решение из $C(\mathbb{R}^+)$ ограниченное на бесконечности при всех $\xi' \in \mathbb{R}^{n-1}$, $|\arg \lambda| \leq \pi/2$, и $h_j \in \mathbb{C}$ таких, что $|\xi'| + |\lambda| \neq 0$. Мы также предполагаем, что

$$u_0(x) \in W_p^{2-\frac{2}{p}}(G), \quad g \in W_p^{2k_0, k_0}(S), \quad (8)$$

$$B(x, 0)u_0(x)|_\Gamma = g(x, 0) \forall x \in \Gamma,$$

где $k_0 = 1 - 1/2p$ в случае условия Дирихле и $k_0 = 1/2 - 1/2p$ в противном случае,

$$u_0(x) \in W_p^{2+s-\frac{2}{p}}(G_\delta), s \in (n/p, 1]. \quad (9)$$

для некоторого допустимого $\delta > 0$. Даны постоянные $\delta_1 < \delta_2 < \delta$. Построим вспомогательную функцию $\varphi \in C_0^\infty(G_\delta)$ такую, что $\varphi \equiv 1$ в области G_{δ_1} и $\varphi \equiv 0$ в $G_\delta \setminus G_{\delta_2}$. Мы предполагаем, что

$$f_i \in L_\infty(0, T; L_p(G)) \cap L_\infty(0, T; W_p^s(G_\delta)), \quad (10)$$

где $i = 1, \dots, m, m = r$,

$$\psi_j \in W_p^1(0, T), p > n + 2 \quad (11)$$

$$u_0(x_j) = \psi_j(0) \quad (j = 1, 2, \dots, s), \quad (12)$$

$$\gamma_i, \sigma \in C^{\frac{1}{2} - \frac{1}{2p} + \varepsilon_0, 1 - \frac{1}{p} + 2\varepsilon_0}(\bar{S}), \varepsilon_0 > 0. \quad (13)$$

Строим матрицу $B(t)$ размерности $m \times m$, строки которой с номерами от $(j-1)h+1$ до jh занимают вектора столбцы

$$(f_1(x_j, t), f_2(x_j, t), \dots, f_m(x_j, t)). \quad (14)$$

Условие (А). Существует число $\delta_0 > 0$ такое что $|\det B(t)| \geq \delta_0$ п.в. на $(0, T)$.

Теоремы 1-3 ниже были получены в работах [11]-[13]. В следующей теореме, считаем, что функция f в (1) равна нулю.

Теорема 1. Пусть выполнено условие (А) и условия (4)-(13) для некоторого допустимого $\delta > 0$ и $s \in (n/p, 1]$, $f_0 \in L_p(Q)$, $f_0\varphi \in L_p(0, T; W_p^s(G_\delta))$. Тогда существует единственное решение $u \in W_p^{2,1}(Q)$, $q_i(t) \in L_p(0, T)$, $i = 1, \dots, m$ задачи (1)-(3) такое, что $\varphi u \in L_p(0, T; W_p^{2+s}(G_\delta))$, $\varphi u_t \in L_p(0, T; W_p^s(G_\delta))$.

Сейчас мы приведем теорему о разрешимости задачи (1)-(3), в случае когда неизвестные функции q_i входят в оператор A как младшие коэффициенты. Как и ранее предполагаем, что $f=0$. В этом случае оператор A имеет вид

$$A = L_0 - \sum_{i=r+1}^m q_i(t)l_i(t, x),$$

$$L_0 u = - \sum_{i,j=1}^n a_{ij}(t, x)u_{x_j x_j} + \sum_{i=1}^n a_i u_{x_i} + a_0 u,$$

$$l_k u = \sum_{i=1}^n a_i^k(t, x)u_{x_i} + b_k(t, x)u$$

где

$$a_i^k, b_k \in L_\infty(0, T; L_p(G)) \cap L_\infty(0, T; W_p^s(G_\delta)) \quad (15)$$

Строим матрицу $B(t)$ размерности $m \times m$, строки которой с номерами от $(j-1)h+1$ до jh занимают вектора столбцы

$$(f_1(x_j, t), f_2(x_j, t), \dots, f_r(x_j, t),$$

$$lu_0(x_j), \dots, lm u_0(x_j)).$$

Условие (В). Существует число $\delta_0 > 0$ такое что $|\det B(t)| \geq \delta_0$ п.в. на $(0, T)$.

Теорема 2. Пусть выполнено условие (В) и условия (4)-(15) для некоторого допустимого $\delta > 0$ и $s \in (n/p, 1]$, $f_0 \in L_p(Q)$, $f_0\varphi \in L_p(0, T; W_p^s(G_\delta))$. Тогда найдется число $\tau_0 \in (0, T]$ такое, что на промежутке $[0, \tau_0]$

существует единственное решение $(u, q_1, q_2, \dots, q_m)$ задачи (1)-(3) такое что $u \in L_p(0, \tau_0; W_p^2(G))$, $u_t \in L_p(Q^{\tau_0})$, $q_i(t) \in L_p(0, \tau_0)$, $i = 1, \dots, m$. Кроме того, $\varphi u \in L_p(0, \tau_0; W_p^{2+s}(G))$, $\varphi u_t \in L_p(0, \tau_0; W_p^s(G))$.

Перейдем к общему случаю. Мы рассматриваем задачу (1)-(3) о восстановлении правой части уравнения вида $\sum_{i=1}^r f_i(x, t)q_i(t) + f_0(x, t)$ и коэффициентов, в частности, входящих в главную части уравнения (1). Предположим, что оператор A имеет вид

$$A = L_0 - \sum_{k=r+1}^m q_k(t)L_k,$$

$$L_k u = - \sum_{i,j=1}^n a_{ij}^k(t, x)u_{x_i x_j} + \sum_{i=1}^n a_i^k(t, x)u_{x_i} + a_0^k(t, x)u,$$

где $k = 0, r+1, r+2, \dots, m$. Поскольку неизвестные могут входить в главную части уравнения, мы будем искать их в классе $C([0, T])$. Построим матрицу $B(t)$ размерности $m \times m$ чьи строки с номерами от $(j-1)h+1$ до jh занимают вектор столбцы

$$(f_1(x_j, t), f_2(x_j, t), \dots, f_r(x_j, t), \dots,$$

$$L_{r+1}u_0(x_j), \dots, L_mu_0(x_j)).$$

Мы предполагаем, что

$$a_{ij}^k \in C(\bar{Q}) \cap L_\infty(0, T; W_p^s(G_\delta)), \quad (16)$$

$$a_i^k \in L_p(Q) \cap L_p(0, T; W_p^s(G_\delta)) \quad (17)$$

$$(i, j = 1, \dots, n, \quad l = 0, 1, \dots, n),$$

$$f_i \in L_p(Q) \cap L_p(0, T; W_p^s(G_\delta)) \quad (i = 1, \dots, m), \quad (18)$$

для некоторого допустимого $\delta > 0$, $s > (2+n)/p$, и $k = 0, r+1, \dots, m$;

$$a_i^k(x_l, t), f_a(x_l, t) \in C([0, T]) \quad (19)$$

для все возможных значений i, j, k, l, a . Также нам необходимо, чтобы было выполнено условие:

Условие (С). Существует число $\delta_0 > 0$ такое что $|\det B(t)| \geq \delta_0$ на $(0, T)$.

Заметим, что элементы матрицы B принадлежат $C([0, T])$. Рассмотрим систему

$$\psi_{jt}(0) + L_0 u_0(x_j, 0) - f(0, x_j, u_0(x_j), \nabla u_0(x_j)) =$$

$$= \sum_{k=1}^r q_{0k} f_k(x_j, 0) + \sum_{k=r+1}^m q_{0k} L_k u_0(x_j)$$

относительно вектора

$\vec{q}_0 = (q_{01}, q_{02}, \dots, q_{0m})$. При выполнении условия (С) система имеет единственное решение. Обозначим $A_0 = L_0 - \sum_{k=r+1}^m q_{0k} L_k$. Пусть B_R – шар радиуса R с центром в нуле в $\mathbb{R}^{(n+1)h}$.

Условие (D). Функция $f(t, x, u, p)$ непрерывна по совокупности переменных $(u, p \in \mathbb{R}^{(n+1)h})$, для любого $R > 0$, найдется постоянная $M_0 > 0$ такая, что

$$|f(t, x, u^1, p^1) - f(t, x, u^2, p^2)|$$

$$\leq M_0(|u_1 - u_2| + |p_1 - p_2|),$$

для всех $(u^1, p^1), (u^2, p^2) \in B_R$, $f(t, x, u, p) \in C([0, T] \times \bar{B}_R; W_p^s(G_\delta))$.

Условие (Е). Функция $f(t, x, u, p)$ дифференцируема по параметрам $(u, p) \in \mathbb{R}^{(n+1)h}$ при п.в. $(t, x) \in Q_\delta$ и $f_u(t, x, u, p), f_{p_i}(t, x, u, p) \in C([0, T] \times \bar{B}_R; W_p^s(G_\delta))$ для любого $R > 0$, найдутся функции

$\Phi_1(t, x), \Phi_2(t, x) \in C([0, T]; L_p(G))$ такие, что

$$\|f_u(t, x, u^1, p^1) - f_u(t, x, u^2, p^2)\|_{L(\mathbb{R}^h)} \leq |\Phi_1(t, x)| |u_1 - u_2| + |\Phi_2(t, x)| |p^1 - p^2|,$$

$$\|f_{p_i}(t, x, u^1, p^1) - f_{p_i}(t, x, u^2, p^2)\|_{L(\mathbb{R}^h)} \leq |\Phi_1(t, x)| |u_1 - u_2| + |\Phi_2(t, x)| |p^1 - p^2|$$

для всех $(u^1, p^1), (u^2, p^2) \in B_R$ и любого $R > 0$. Здесь величины f_u, f_{p_i} ($i = 1, 2, \dots, n$) – соответствующие матрицы Якоби и $\|\cdot\|_{L(\mathbb{R}^h)}$ – норма в пространстве линейных непрерывных отображений из \mathbb{R}^h в \mathbb{R}^h .

Как и ранее предполагаем, что

$$f_0 \in L_p(Q), f_0 \varphi \in L_p(0, T; W_p^s(G_\delta)). \quad (20)$$

При указанных условиях теорема существования примет следующий вид.

Теорема 3. Пусть условия (C)-(E), (8), (9), (11)-(13), (16)-(20) выполнены. Предположим также, что оператор $M_0 = \partial_t + A_0$ параболический и выполнено условие Лопатинского, т.е., условия (6), (7) выполнены. Тогда найдется число $\tau_0 \in (0, T]$ такое, что на промежутке $[0, \tau_0]$ существует единственное решение $(u, q_1, q_2, \dots, q_m)$ задачи (1)-(3) такое что $u \in L_p(0, \tau_0; W_p^2(G))$, $u_t \in L_p(Q^{\tau_0})$, $q_i(t) \in C([0, \tau_0])$, $i = 1, \dots, m$. Кроме того, $\varphi u \in L_p(0, \tau_0; W_p^{2+s}(G))$, $\varphi u_t \in L_p(0, \tau_0; W_p^s(G))$.

ТОЧЕЧНЫЕ ИСТОЧНИКИ

Рассмотрим случай $n=1$.

$$Lu = u_t + A(t, x, D)u = \sum_{i=1}^r \delta(x - x_i) q_i(t) + f_0(t, x), \quad (t, x) \in Q, \quad (21)$$

где A – эллиптический оператор вида $A(t, x, D)u = -a(t, x)u_{xx} + b(t, x)u_x + a_0(t, x)u$ ($x \in (a, b)$).

Начальные и краевые условия имеют вид

$u(0, x) = u_0(x)$, $B_1 u(t, a) = \varphi_1(t)$, $B_2 u(t, b) = \varphi_2(t)$, где $B_1 u = u$ или $B_1 u = u_x + \alpha_1 u$, $B_2 u = u$ или $B_2 u = u_x + \alpha_2 u$. Условия переопределения записываются в прежнем виде

$$u(y_i, t) = \psi_i(t), \quad i = 1, 2, \dots, s. \quad (22)$$

Требуется найти решение этой начально-краевой задачи, точки x_i , число s и интенсивности $q_i(t)$, используя данные (22). Первое что необходимо отметить, это тот факт, что любое число замеров s не позволяет определить эти неизвестные величины однозначно, если не задано условие на взаимное расположение точек x_i (см. [27]) (грубо говоря замеры и неизвестные точки должны чередоваться). Некоторое условие такого типа возникает в многомерном случае (при $n=2$ никакие три точки не лежат на одной прямой, при $n=3$ никакие четыре точки не лежат на одной плоскости см. [34]). При выполнении условий такого типа и некоторых дополнительных условий решение обратной задачи можно выписать и рассмотреть вопрос о его единственности и о представлении решения. Наиболее интересные представления получаются, если использовать асимптотические представления функции Грина соответствующей эллиптической задачи с параметром. В частности, при $r=1$, если коэффициенты оператора A не зависят от t , то два замера (т.е. $s=2$) позволяют определить точку x_1 и интенсивность q_1

однозначно при условии, что $y_1 < x_1 < y_2$.

Чтобы упростить формулу, мы возьмем $a = a^2 = \text{const}$, $f_0 = 0, u_0 = 0$. В этом случае точка x_1 может быть определена из равенства [35].

$$x_1 = \frac{y_1 + y_2}{2} - \frac{\alpha}{2\sqrt{\lambda}} \ln \frac{\Omega_1(\lambda)}{\Omega_2(\lambda)} - \frac{1}{2\alpha\sqrt{\lambda}} \int_{y_1}^{y_2} b(\eta) d\eta + O\left(\frac{1}{\lambda}\right),$$

где $\Omega_i(\lambda) = \int_0^\infty \exp(-\lambda t) \psi_i(t) dt$.

БЛАГОДАРНОСТИ

Работа проводилась при поддержке РФФИ № 18-01-00620

ЛИТЕРАТУРА

1. Marchuk, G.I. Mathematical Models in Environmental Problems / G.I. Marchuk. – Studies in Mathematics and its Applications. V. 16. – Amsterdam: Elsevier Science Publishers. – 1986.
2. Vabishchevich P.N. Numerical solution of an inverse filtration problem / P. N. Vabishchevich, V.I. Vasil'ev, M.V. Vasil'eva, and D.Ya. Nikiforov // Lobachevskii Journal of Mathematics. – 2016. – Vol. 37, No. 6. – pp. 777–786.
3. Прилепко, А.И. Теоремы разрешимости и метод Ротэ в обратных задачах для уравнения параболического типа / А.И. Прилепко, В.В. Соловьев // Дифференц. уравнения. – 1987. Т.23, 10. – С. 1791-1799.
4. Ivanchov M. Inverse Problems for Equation of Parabolic Type / M. Ivanchov // Math. Studies. Monograph Series. {bf ~10}, Lviv: WNTL Publishers, 2003.
5. Прилепко, А.И. О разрешимости обратных краевых задач определения коэффициента перед младшей производной в параболическом уравнении / А.И. Прилепко, В.В. Соловьев // Дифференц. уравнения. – 1987. – Т. 23, 1. – С. 136-143.
6. Кулиев, М.А. Многомерная обратная задача для параболического уравнения в ограниченной области / М.А. Кулиев // Нелинейные граничные задачи. – 2004. – В. 14. – С. 138-145.
7. Fan Y. and Li D.G. Identifying the heat source for the heat equation with convection term / Y. Fan and D.G. Li // Int. J. Math. Anal. – 2009. – V. 3, № 27. – P.1317--1323.
8. Белов, Ю.Я. О задаче идентификации функции источника для уравнения типа Бюргерса / Ю.Я. Белов, К.В. Коршун // J. of Siberian Federal University. Mathematics & Physics. – 2012. – 5(4), P. 497-506.
9. Соловьев В. В., Существование решения в “целом” обратной задачи определения источника в квазилинейном уравнении параболического типа / В.В. Соловьев // Дифференц. уравнения // 1996, Т. 32, № 4, С. 536–544.
10. Пятков С.Г., Ротко В.В. Обратные задачи для некоторых квазилинейных параболических систем с точечными условиями переопределения / С.Г. Пятков, В.В. Ротко // Матем. Труды. – 2019. – Т. 22, № 1, С. 178–204.
11. Pyatkov S.G., Rotko V.V. On some parabolic inverse problems with the pointwise overdetermination // AIP Conference Proceedings. 2017. V. 1907, 020008.
12. Пятков С.Г., Ротко В.В. Об определении функции источника в квазилинейных параболических задачах с точечными условиями переопределения / С.Г. Пятков, В.В. Ротко // Вестник ЮУрГУ. Серия "Математика. Механика. Физика". 2017. Т. 9, №4. С.19-26.

13. Ротко В.В. Обратные задачи для математических моделей конвекции-диффузии с точечными условиями переопределения / В.В. Ротко // Вестник ЮГУ. – 2018. Вып. 3(50). – С. 57-66.
14. Prilepko, A. I. Methods for solving inverse problems in Mathematical Physics / A.I. Prilepko, D.G. Orlovsky, I.A. Vasin. – New-York: Marcel Dekker, Inc. – 1999.
15. Пятков С.Г. О некоторых обратных задачах для операторно-дифференциальных уравнений первого порядка / Пятков С.Г. // Сибирский математический журнал. 2019. Т. 60, № 1. с. 183-193.
16. Belov Yu.Ya. Inverse problems for parabolic equations / Yu.Ya. Belov. -- Utrecht: VSP, 2002.
17. Пятков, С. Г. О некоторых классах коэффициентных обратных задач для параболических систем уравнений / С.Г. Пятков, М.Л. Самков // Матем. тр. – 2012. – Т. 15, 1. – С. 155–177.
18. Пятков С.Г. О некоторых классах обратных задач с данными переопределения на пространственных многообразиях. / С.Г. Пятков // Сибирский математический журнал. 2016. Т. 57, № 5, с. 1114-1126.
19. Вабишевич П. Н. Вычислительная идентификация младшего коэффициента параболического уравнения / П. Н. Вабишевич, В. И. Васильев // Доклады РАН. – 2014, Т. 455, № 3, с. 258–260.
20. Dehghan M. Numerical computation of a control function in a partial differential equation // Applied mathematics and computation. – 2004. – V. 147. – P. 397-408.
21. Mamonov, A. V. Point source identification in nonlinear advection-diffusion-reaction systems / A.V. Mamonov, Y-H. R. Tsai // Inverse Problems. – 2013. – V. 29, 3. – 26 p.
22. Ozisik, M.N. Inverse Heat Transfer / M.N. Ozisik, H.R.B. Orlande. – New York: Taylor & Francis. – 2000.
23. Samarskii A.A. Numerical Methods for Solving Inverse Problems of Mathematical Physics / A.A. Samarskii, P.N. Vabishchevich. -- Berlin/Boston: De Gruyter., 2007.
24. Кабанихин С.И. Обратные и некорректные задачи / С.И. Кабанихин // Новосибирск: Сибирское научное издательство, 2009.
25. Alifanov O.M. Inverse Heat Transfer Problems. Berlin, Heidelberg: Springer-Verlag, 1994.
26. Алифанов О.М. Обратные задачи сложного теплообмена / О.М. Алифанов, Е.А. Артюхов, А.В. Ненароком // М.: Янус-К, 2009.
27. Пятков С.Г. О некоторых классах обратных задач об определении функции источников / С.Г., Пятков, Е.И. Сафонов // Математические труды. 2016, том 19, № 1, 178–198.
28. Панасенко Е.А., Старченко А.В. Численное решение некоторых обратных задач с различными типами источников атмосферного загрязнения / Е. А. Панасенко, А. В. Старченко // Вестн. Томск. гос. ун-та. Матем. и мех. – 2008 – № 2(3), С. 47-55.
29. Пененко В.В. Вариационные методы усвоения данных и обратные задачи для изучения атмосферы, океана и окружающей среды / В. В. Пененко // Сиб. журн. вычисл. матем. – 2009. – Т. 12:4. – С. 421-434
30. Murray-Bruce J., Dragotti P.L. Estimating localized sources of diffusion fields using spatiotemporal sensor measurements/ J. Murray-Bruce and P.L. Dragotti // IEEE Trans. on Signal Processing. – 2015. – V. 63, № 12. P.3018--3031.
31. Badia A.El, Hamdi A. Inverse source problem in an advection-dispersion- reaction system: application to water pollution / A.~El Badia, A. Hamdi // Inverse Problems 2007. V. 23. -P.2103-2120.
32. Badia A.El, Ha-Duong T. Inverse source problem for the heat equation. Application to a pollution detection problem / A.El Badia, T. Ha-Duong // J. Inv. Ill-Posed Problems. 2002. – V.10, № 6. P. 585-599.
33. Badia A.El, Ha-Duong T., Hamdi A. Identification of a point source in a linear advection-dispersion-reaction equation: application to a pollution source problem / A.El Badia, T. Ha-Duong // Inverse Problems. 2005. V. 21, № 3. P. 1121-1136.
34. Ling L. Point sources identification problems for heat equations / L. Ling, T. Takeuchi // Commun. Comput. Phys. – 2009. – V. 5, №5. P. 897--913.
35. Pyatkov S.G., Safonov E.I. Point sources recovering problems for the one-dimensional heat equation / S.G. Pyatkov, E.I. Safonov // Journal of Advanced Research in Dynamical and Control Systems. V. 11, Iss. 01. P: 496-510.
36. Badia A.El, Ha-Duong T. An inverse source problem in potential analysis / A El Badia and T Ha-Duong // Inverse Problems. – 2000. V. 16. – P. 651–663.
37. Triebel H. Interpolation Theory. Function Spaces. Differential Operators. / H. Triebel – Berlin: VEB Deutscher Verlag der Wissenschaften, 1978.

Inverse problems in theory and heat and mass transfer

С.Г. Пятков

Югорский государственный университет
Ханты-Мансийск, Россия

s_pyatkov@ugrasu.ru, v_rotko@ugrasu.ru

В. В. Ротко

Югорский государственный университет
Ханты-Мансийск, Россия

s_pyatkov@ugrasu.ru, v_rotko@ugrasu.ru

Abstract. This article is survey devoted to the well-posedness questions in the Sobolev spaces of inverse problems with the pointwise overdetermination for convection-diffusion and filtration mathematical models, which are described by quasilinear parabolic systems of the second order. The unknowns depending on time occur in the right-hand side and the differential operator itself as coefficients. The overdetermination conditions are the values of a solution at some interior points. We expose the conditions ensuring local (in time) well-posedness of the problems in the Sobolev classes and describe latest results including those by the authors. These results are sharp. They are obtained under the minimal smoothness assumptions for the data and other conditions as well.

Keywords: parabolic system, inverse problem, heat-and-mass transfer, source function, convection-diffusion, filtration

Применение интегральных характеристик круговой три-ткани в решении задач регистрации цифрового изображения

М.В. Куркина, В.А. Самарин, О.В. Самарина, В.В. Славский
Югорский государственный университет
г. Ханты-Мансийск, Россия
samarina_ov@mail.ru, slavsky2004mail.ru

Аннотация. В работе представлен подход к обработке и анализу цифровых трехканальных изображений, основанный на теории круговой три-ткани. Определяются и исследуются интегральные инварианты относительно группы аффинных преобразований и группы преобразований Мёбиуса. Свойства данных инвариантов позволяют эффективно использовать их в качестве «меток» при регистрации изображений.

Ключевые слова: круговые три-ткани, инварианты, трехканальные изображения.

ВВЕДЕНИЕ

КРУГОВАЯ ТРИ-ТКАНЬ RGB-ИЗОБРАЖЕНИЯ И “ПОЛНАЯ КРИВИЗНА” ТРИ-ТКАНИ КАК ДЕСКРИПТОР ИЗОБРАЖЕНИЯ

Пусть имеются два уравнения $g_1(x, y) = 0$, $g_2(x, y) = 0$ задающие окружности на плоскости:

$$g_1(x, y) = a_{1,1}(x^2 + y^2) + a_{1,2}x + a_{1,3}y + a_{1,4}, \quad (1)$$

$$g_2(x, y) = b_{1,1}(x^2 + y^2) + b_{1,2}x + b_{1,3}y + b_{1,4}. \quad (2)$$

Уравнение вида $u = f \cdot g_1 + (1-f)g_2 = 0$ назовем уравнением линейного пучка окружностей с вершинами g_1 , g_2 . Разрешая равенство $u = 0$ относительно f , получим функцию $f(x, y)$ все линии уровня которой есть окружности.

Три-ткань $W(u_1, u_2, u_3)$, образованная тремя линейными пучками окружностей, называется круговой тканью [1].

В работах [2, 3] дана полная классификация регулярных круговых тканей. Ниже приведены условия на функцию $f(x, y)$ при которых линии уровня прямые (или окружности).

Теорема 1 (линии уровня). Пусть $f(x, y)$ функция класса C^3 регулярная на плоскости, т.е. $\nabla f \neq 0$. Тогда, если $C_1(f) \equiv 0$, то все линии уровня прямые, а если $C_2(f) \equiv 0$, то все линии уровня окружности, где:

$$C_1(f) = f^{(0,2)}p^2 - 2pqf^{(1,1)} + f^{(2,0)}q^2, \quad (3)$$

$$C_2(f) = (p^2 + q^2)(p^3 f^{(0,3)} - 3p^2 q f^{(1,2)} + 3p q^2 f^{(2,1)} - q^3 f^{(3,0)}) + 3f^{(1,1)}(q^4 f^{(2,0)} - p^4 f^{(0,2)}) + 6p q \cdot (f^{(1,1)})^2 (p^2 - q^2) - 3p q \cdot (f^{(0,2)} - f^{(2,0)}) \cdot (p^2 f^{(0,2)} - 3p q f^{(1,1)} - q^2 f^{(2,0)}), \quad (4)$$

здесь $f^{(i,j)} = f^{(i,j)}(x, y) = \frac{\partial^{i+j} f(x, y)}{\partial^i x \partial^j y}$, $i, j = 0, 1, \dots$;
 $p = f^{(1,0)}$, $q = f^{(0,1)}$.

Доказательство. Пусть $\{x = x(t), y = y(t)\}$ натуральная параметризация линии $f(x, y) = \text{const}$, где t длина дуги. Единичный касательный вектор и единичный вектор нормали в этом случае имеют вид $\tau = \{\dot{x}, \dot{y}\}$, $n = \{-\dot{y}, \dot{x}\}$ соответственно (здесь точка – производная по t). Из формул Френе-Серре следует:

$$\ddot{x} = -\kappa \dot{y}, \quad \ddot{y} = \kappa \dot{x} - \kappa^2 \dot{y}, \quad \ddot{\kappa} = \kappa \dot{\kappa} - \kappa^2 \dot{y}, \quad (5)$$

где κ – кривизна линии. Из четырех равенств

$$\dot{x}^2 + \dot{y}^2 \equiv 1, \quad \frac{d}{dt} f(x(t), y(t)) \equiv 0, \quad (6)$$

$\frac{d^2}{dt^2} f(x(t), y(t)) \equiv 0$, $\frac{d^3}{dt^3} f(x(t), y(t)) \equiv 0$ с учетом (3) в пакете Mathematica вычисляем $\dot{x}, \dot{y}, \kappa, \dot{\kappa}$:

$$\kappa = \frac{C_1(f)}{(p^2 + q^2)^{3/2}}, \quad \dot{\kappa} = \frac{C_2(f)}{(p^2 + q^2)^3} \quad (7)$$

Условию $\kappa \equiv 0$ соответствует равенство $C_1(f) \equiv 0$, а условию $\dot{\kappa} \equiv 0$ соответствует равенство $C_2(f) \equiv 0$.

Замечание 1. При исследовании цифровых изображений возникает важная задача улучшения цифрового изображения путем удаления “шума”. Естественный способ это сделать путем минимизации следующих интегральных характеристик:

$$IC_1 = \iint |C_1(f)(x, y)| dx dy, \quad IC_2 = \iint |C_2(f)(x, y)| dx dy. \quad (8)$$

Замечание 2. Свойство C_1 обращаться в нуль в той области изображения, где линии уровня прямолинейны, отражает инвариантность относительно группы аффинных преобразований изображения. Свойство C_2 обращаться в

нуль в той области изображения где линии уровня – дуги окружностей, отражает инвариантность относительно Мёбиусовых преобразований (необходимо отметить, что группа Мёбиусовых преобразований включает в себя группу аффинных). Эти свойства позволяют использовать величины C_1 и C_2 в качестве "меток" в решении задач регистрации цифровых изображений.

ПОСТРОЕНИЕ КРУГОВОЙ ТРИ-ТКАНИ И ВЫЧИСЛЕНИЕ ЕЕ ИНВАРИАНТНЫХ ХАРАКТЕРИСТИК

Рассмотрим практический пример построения круговой три-ткани в математическом пакете Mathematica. Круговая три-ткань в рассматриваемом примере будет образована тремя функциями u_1, u_2, u_3 линии уровней которых три пучка окружностей с центрами в точках $C_1 = (0,0)$, $C_2 = (0,1)$, $C_3 = (1,0)$ на плоскости. Значения функций меняются в пределах от 0 до 1 на квадрате $[0,2] \times [0,2]$. Функция яркости RGB-изображения в данном случае имеет вид:

$$\left\{ u_1 = \frac{\sqrt{x^2 + y^2}}{2\sqrt{2}}, u_2 = \frac{\sqrt{x^2 + (-1+y)^2}}{\sqrt{5}}, u_3 = \frac{\sqrt{(-1+x)^2 + y^2}}{\sqrt{5}} \right\} \quad (9)$$

Исключая из данных равенств координаты $\{x, y\}$ получим функцию данной три-ткани:

$$W(u_1, u_2, u_3) = 2 + 128u_1^4 - 10u_2^2 + 25u_2^4 - 10u_3^2 + 25u_3^4 - 80u_1^2(u_2^2 + u_3^2). \quad (10)$$

На рис. 1 представлена соответствующая три-ткань и иллюстрация к формуле (16.6) из книги В. Бляшке "Введение в геометрию тканей" [1].

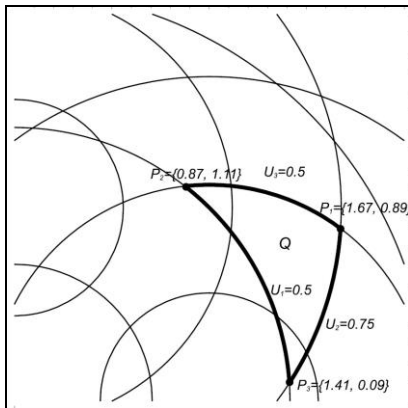


Рис. 1. Три-ткань RGB-изображения

Далее рассмотрим пример построения представленной выше круговой три-ткани, и вычислению ее интегральных характеристик C_1 , C_2 в системе Matlab.

Круговую ткань построим с помощью следующих команд:

```
h=0.1; % задаем величину шага
r=200; % задаем размер изображения
[X,Y]=meshgrid(1:r,1:r); % строим сетку изображения
```

```
f=sqrt(h*(X.^2+Y.^2)); % задаем значения функции в узлах сетки.
```

Получившееся изображение представлено на рис. 2.

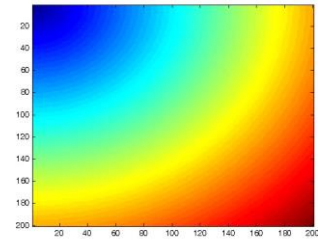


Рис. 2. Круговая три-ткань в системе Matlab

Интегральные характеристики C_1 , C_2 для функции изображения f легко вычисляются по формулам (3),(4).

Полученные характеристики, представлены на рис. 3.

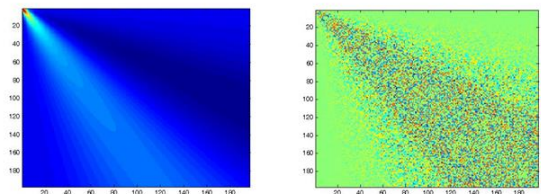


Рис. 3. Характеристики C_1 C_2

Проведенные в системе Matlab практические исследования круговой три-ткани изображения, показали устойчивость характеристики C_1 к аффинным преобразованиям и характеристики C_2 к Мёбиусовым преобразованиям.

ЗАКЛЮЧЕНИЕ

Подход к анализу изображений, основанный на теории круговой ткани В. Бляшке является достаточно перспективным направлением в области обработки цифрового изображения [4, 5].

Представленные в работе интегральные инварианты относительно группы аффинных преобразований и группы преобразований Мёбиуса являются эффективными характеристиками цифрового изображения и могут найти применение в решении широкого класса задач обработки, анализа и классификации изображений, поиска соответствия по образцу.

ЛИТЕРАТУРА

1. Бляшке В. Введение в геометрию тканей. Пер. с нем. – М.: Физмат, 1959. – 144 с.
2. Лазарева В. Б. Классификация регулярных круговых три-тканей с точностью до круговых преобразований // Фундамент. и прикл. матем. 16:1.2010. – С. 95–107.
3. Шелехов А. М. Криволинейные три-ткани: монография. /А. М. Шелехов, В. Б. Лазарева, А. А. Уткин. – Тверь: Твер. гос. ун-т, 2013. – 232 с.
4. Самарина О.В. Геометрические методы в решении задач обработки изображений/ О.В. Самарина, В.В. Слав-

Application of the circular three-webs integral characteristics in solving digital image registration problems

M.V. Kurkina, V.A. Samarin, O.V. Samarina, V.V. Slavsky

Yugra State University

Khanty-Mansiysk, Russia

samarina_ov@mail.ru, slavsky2004@mail.ru

Abstract. The paper presents an approach to digital three-channel images processing and analysis, based on the circular three-web theory. Integral invariants concerning a group of affine transformations and a group of Möbius transformations are determined and investigated. The properties of these invariants

allow them to be effectively used as "labels" when registering images.

Keywords: *circular three-web, invariants, three-channel image.*

Новый подход к восстановлению пропущенных данных о площади термокарстовых озер Арктики

Попков Ю.С.

Федеральный исследовательский центр "Информатика и управление" РАН, Москва, Россия
popkov.yuri@gmail.com

Мельников А.В.

АУ «Югорский НИИ информационных технологий»,
Ханты-Мансийск, Россия
melnikovav@uriit.ru

Полищук Ю.М.

АУ «Югорский НИИ информационных технологий»,
Ханты-Мансийск, Россия
yupolishchuk@gmail.com

Сокол Е.С.

АУ «Югорский НИИ информационных технологий»,
Ханты-Мансийск, Россия
sokoles@uriit.ru

Полищук В.Ю.

Институт мониторинга климатических и экологических систем СО РАН,
Томск, Россия
liquid_metal@mail.ru

Аннотация. В задачах прогнозирования объемов метана как одного из важных парниковых газов, рассматриваемых в перечне причин современного глобального потепления, необходимо использовать наряду с климатическими характеристиками и данные о динамике площадей термокарстовых озер в зонах мерзлоты, которые обычно получают с помощью космических снимков. Ввиду большого количества пасмурных дней на северных территориях удастся получить лишь малое количество безоблачных снимков, что приводит к значительным пропускам во временных рядах площадей озер. Для восстановления пропущенных значений площадей озер в работе предлагается использовать новый подход к восстановлению пропусков, основанный на методах и алгоритмах энтропийно-рандомизированного машинного обучения. Этот подход предполагает проводить восстановление пропущенных значений в экспериментальных данных о площадях термокарстовых озер с использованием временных рядов среднегодовой температуры и годовой суммы осадков. В качестве экспериментальных данных о площадях термокарстовых озер и климатических параметрах (температура и сумма осадков) использованы результаты исследований, проведенных в арктической зоне Западной Сибири в период с 1973 по 2007 г. Исследования проведены на девяти тестовых участках, выбранных в разных зонах мерзлоты (сплошной, прерывистой и островной). Данные о среднегодовой температуре и годовой сумме осадков для каждого тестового участка получены методом реанализа. Разработанный алгоритм восстановления пропущенных значений в рамках указанного подхода реализован с использованием инструментов MATLAB R2019a. Рассчитаны пропущенные значения для выбранных девяти тестовых участков. Для иллюстрации приведены временные ряды значений площади озер, температуры и осадков на одном из тестовых участков. Проведен анализ погрешностей восстановления пропусков, который показал, что разработанный алгоритм позволяет восстанавливать пропущенные значения площадей озер по данным об изменении температуры и осадков с практически приемлемой точностью.

Ключевые слова: рандомизация, машинное обучение, термокарстовые озера, климатические параметры, моделирование, восстановление пропущенных данных.

ВВЕДЕНИЕ

Особую важность рандомизированный подход представляет для решения задач прогнозирования динамики накопления парниковых газов в термокарстовых озерах арктической зоны в связи с их влиянием на глобальные климатические изменения, что может явиться основой разработки и функционирования систем адаптации к меняющимся условиям среды обитания на различных управленческих уровнях. С грядущим глобальным потеплением климата в ближайшие десятилетия будут ускоряться процессы таяния мерзлых пород, приводя к дополнительному высвобождению углекислого газа, а также метана как продукта жизнедеятельности микроорганизмов, перерабатывающих оттаявшее органическое вещество. Это способно внести дополнительный ощутимый вклад в потепление климата, что вызывает озабоченность мировой общественности. Осознание этого послужило поводом для принятия Парижского климатического соглашения (2016), предполагающего разработку в разных странах мер, способных не допустить повышение среднегодовой температуры Земли более чем на 1,5 °C до 2050 года. Разработка таких мер на региональном уровне для арктических регионов невозможна без формирования обоснованных прогнозов объемов эмиссии метана и углекислого газа, которые могут быть основаны на знаниях о пространственно-временной динамике полей озер на территории регионов.

Задачи прогнозирования динамики накопления парниковых газов в термокарстовых озерах на ближайшие десятилетия

тилетия требуют использования данных о временных рядах площадей озер и климатических параметров (температуры, осадков).

Ввиду большой степени заболоченности арктических территорий данные о площадях озер могут быть получены только с использованием спутниковых снимков. Ввиду большого количества пасмурных дней на северных территориях удастся получить безоблачные снимки лишь в отдельные годы. В результате этого полученные временные ряды площадей озер обладают значительным числом пропущенных значений. Вопросы восстановления пропусков во временных рядах, применительно к данным о площадях озер, в настоящее время разработаны недостаточно. Наиболее перспективным подходом к восстановлению пропусков в наших условиях рассматривается использование энтропийно-рандомизированных методов, показавших, согласно [1], высокую эффективность в решении задач глобальной экономики, демографии и др. Однако методические вопросы восстановления пропусков во временных рядах площадей озер в рамках энтропийно-рандомизированного подхода не разработаны, что и явилось целью настоящей работы.

ДАННЫЕ

На рис. 1 приведена схема размещения тестовых участков (ТУ) для проведения исследований, направленных на получение данных о временных рядах площадей термокарстовых озер, среднегодовой температуры и годовой суммы осадков.

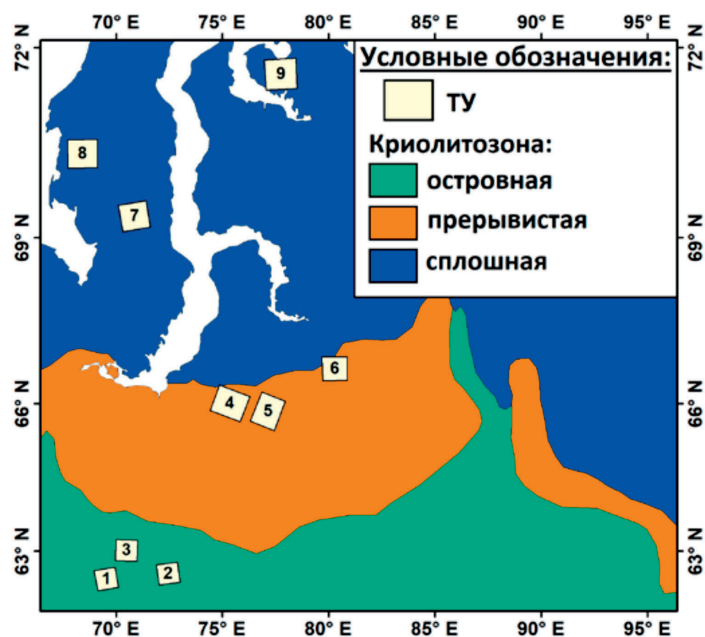


Рис. 1. Схема расположения тестовых участков на исследуемой территории Западной Сибири

Для получения данных о площадях озер использованы космические снимки среднего разрешения (30 м) Landsat, полученные в период с 1973 по 2007 г. Методические вопросы проведения этих исследований изложены в [2]. Данные о климатических характеристиках получены с использованием процедур реанализа данных [2].

Для примера в табл. 1 приведены данные о средней

площади термокарстовых озер $\bar{S}_{\text{изм}}$, среднегодовой температуре \bar{T} , годовой сумме осадков \bar{R} на ТУ - 5.

Таблица 1.

Данные о площади, температуре и осадках на ТУ-5.

Годы \ Данные	$\bar{S}_{\text{изм}}$, га	\bar{T} , °C	\bar{R} , мм	$\bar{S}_{\text{восст}}$, га
1	2	3	4	5
1973	67,90	-1,00	650,30	67,90
1974	-	-3,50	494,25	65,43
1975	-	-2,00	684,80	67,12
1976	-	-2,50	487,30	65,06
1977	-	-2,50	486,50	65,18
1978	-	-3,50	706,40	68,32
1979	-	-2,00	700,10	67,49
1980	-	-2,50	556,50	65,87
1981	-	-1,00	534,60	64,85
1982	-	-2,50	489,95	65,06
1983	-	-1,00	578,80	65,58
1984	68,89	-2,50	465,05	68,89
1985	-	-3,50	482,20	65,40
1986	-	-3,50	577,30	66,68
1987	-	-2,50	467,50	64,78
1988	64,27	-1,00	501,20	64,27
1989	-	-1,00	487,70	64,37
1990	-	-2,50	462,10	64,85
1991	-	-1,00	551,50	65,44
1992	-	-3,50	585,30	66,78
1993	-	-1,00	471,50	64,30
1994	-	-2,50	562,50	65,95
1995	-	0,50	578,90	64,87
1996	-	-1,00	571,35	65,68
1997	-	-2,50	521,10	65,66
1998	-	-3,50	544,80	66,11
1999	-	-2,50	563,15	66,10
2000	-	-1,00	561,30	65,17
2001	62,36	-2,00	430,20	62,36
2002	-	-2,06	520,15	65,21
2003	62,31	-0,48	429,20	62,31
2004	-	-1,80	338,40	62,90
2005	-	0,44	343,90	62,07
2006	-	-2,78	222,00	61,86
2007	66,96	0,14	325,40	66,96

Как видно из табл. 1 (столбец 2), данные о площадях озер имеют большое число пропущенных значений, что является характерным для всех исследованных тестовых участках. Заметим, что в столбце 5 (табл. 1) представлены

восстановленные (модельные) данные о средней площади озер. В табл. 1 вместо модельных данных которых сохранены значения реальных измерений.

Рассмотрим вопросы подготовки данных к расчетам. Данные по площади \tilde{S} , температуре \tilde{T} , осадкам \tilde{R} собраны по каждому из исследованных девяти тестовых участков. Для расчётов преобразуем данные к стандартному (нормализованному) виду по следующим формулам:

$$\begin{aligned} S &= \frac{\tilde{S} - \tilde{S}_{min}}{\tilde{S}_{max} - \tilde{S}_{min}}, \\ T &= \frac{\tilde{T} - \tilde{T}_{min}}{\tilde{T}_{max} - \tilde{T}_{min}}, \\ R &= \frac{\tilde{R} - \tilde{R}_{min}}{\tilde{R}_{max} - \tilde{R}_{min}}. \end{aligned} \quad (1)$$

где \tilde{S} , \tilde{T} , \tilde{R} – средняя площадь, среднегодовая температура и сумма годовых осадков в натуральных величинах (га, °C, мм, соответственно), S , T , R – нормализованные значения площади, температуры и осадков (отображенные на промежутке $[0,1]$), нижний индекс у всех показателей означает минимальное и максимальное значение выборки.

МОДЕЛИ И АЛГОРИТМЫ

Технология энтропийно-рандомизированного прогнозирования реализуется в следующей последовательности этапов [1]. Сначала формируется предсказательная рандомизированная параметрическая модель (РПМ), синтезируются ее определяющие параметры и согласуется с ней необходимое информационное обеспечение. РПМ преобразует массив реальных входных данных $X = [x^1, \dots, x^s]$, где $x(j) \in \mathbf{R}^n$ в модельный выход, характеризуемый матрицей $Z = [z(1), \dots, z(s)]$, где $z(j) \in \mathbf{R}^m$. В общем случае это преобразование предполагается динамическим, т.е. модельный выход, наблюдаемый в момент времени j , зависит от входа, наблюдаемого на некотором историческом интервале $j - \varrho, \dots, j$, т.е. от матрицы $X_{\varrho}(j) = [x^{j-\varrho}, \dots, x^j]$. Математическим выражением этой связи является векторный функционал $\hat{\Omega}(X_{\varrho}(j)|\alpha, P(\alpha))$ со случайными параметрами α интервального типа

$$\alpha \in \mathcal{A} = [\alpha^-, \alpha^+]. \quad (2)$$

Вероятностные свойства параметров характеризуются функцией плотности распределения вероятности (ПРВ) $P(\alpha)$, которая предполагается непрерывно-дифференцируемой.

Выход РПМ в j -й момент времени (момент измерения) представляет собой ансамбль $\hat{Z}(j | P(\alpha))$ случайных векторов

$$\hat{z}(j|\alpha) = \hat{\Omega}(X_{\varrho}(j)|\alpha, P(\alpha)), \quad j = \overline{1, s}. \quad (3)$$

Для имитации влияния измерительных ошибок на выходе объекта вводятся случайные шумы $\xi \in \mathbf{R}^m$ интервального типа:

$$\xi^j \in \Xi_j = [\xi_-^j, \xi_+^j], \quad j = \overline{1, s} \quad (4)$$

с непрерывно-дифференцируемыми функциями ПРВ $Q_j(\xi^j)$, $j = \overline{1, s}$, в соответствии с которыми генерируется для каждого момента измерения выхода объекта ансамбль $\mathcal{F}(j|Q_j(\xi^j))$. Набор случайных векторов (шумов измерений) за весь интервал измерений описывается матрицей

$$K = [\xi^{(j)}, j = \overline{1, s}], \quad (5)$$

которая характеризуется совместной функцией ПРВ $Q(K)$. Если шумы в измерениях статистически независимые, то

$$Q(K) = \prod_{j=1}^s Q_j(\xi^{(j)}). \quad (6)$$

Наблюдаемый выход РПМ представим в виде:

$$v(j|\alpha, \xi^{(j)}) = \hat{\Omega}(X_{\varrho}(j)|\alpha, P(\alpha)), \quad j = \overline{1, s}. \quad (7)$$

Случайные векторы (6) образуют ансамбли, математическое ожидание которых имеет вид:

$$\begin{aligned} \mathcal{M}\{v(j|\alpha, \xi^{(j)})\} &= \int_{\mathcal{A}} \hat{z}(j|\alpha) P(\alpha) d\alpha + \\ &+ \int_{\Xi_j} Q_j(\xi^{(j)}) \xi^{(j)} d\xi^{(j)} = \\ &= \mathcal{W}[P(\alpha), Q_j(\xi^{(j)})], \quad j = \overline{1, s}. \end{aligned} \quad (8)$$

Второй этап технологии рандомизированного прогнозирования связан с обучением РПМ. т.е. с оцениванием ПРВ ее параметров и шумов измерений. Он реализуется с помощью следующего алгоритма [3]:

$$[P^*(\alpha), Q^*(K)] = \arg \max \mathcal{H}[P^*(\alpha), Q^*(K)] \quad (9)$$

на множестве нормированных функций $P^*(\alpha), Q^*(K)$, для которых выполняются условия эмпирических балансов:

$$\mathcal{W}[P(\alpha), Q_j(\xi^{(j)})] = y^{(j)}, \quad j = \overline{1, s} \quad (10)$$

где $y^{(j)} \in \mathbf{R}^m$ - вектор реальных измерений выхода объекта.

Задача (9-10) относится к классу функциональных энтропийно-линейных уравнений ляпуновского типа [4], которые имеют аналитическое решение, получаемое с помощью множителей Лагранжа $\Theta = [\theta^j, j = \overline{1, s}]$ (векторы $\theta^j \in \mathbf{R}^m$):

$$\begin{aligned} P^*(\alpha) &= \frac{\exp(-\sum_{j=1}^s \langle \theta^j, \hat{z}(j|\alpha) \rangle)}{\mathcal{P}(\Theta)}, \\ Q_j^*(\xi^{(j)}) &= \frac{\exp(-\langle \theta^j, \xi^{(j)} \rangle)}{Q_j(\theta^j)}, \quad j = \overline{1, s}; \\ Q_j^*(\xi^{(j)}) &= \frac{\exp(-\langle \theta^j, \xi^{(j)} \rangle)}{Q_j(\theta^j)}, \quad j = \overline{1, s}; \end{aligned}$$

$$Q(K) = \prod_{j=1}^s Q_j^*(\xi^{(j)}). \quad (11)$$

В знаменателях этих выражений стоят нормировочные константы

$$\mathcal{P}(\theta) = \int_{\mathcal{A}} \exp(-\sum_{j=1}^s \langle \theta^j, \hat{z}(j|\alpha) \rangle) d\alpha, \\ Q_j(\theta^j) = \int_{\Xi_j} \exp(-\langle \theta^j, \xi^{(j)} \rangle) d\xi^{(j)}, \quad j = \overline{1, s}. \quad (12)$$

Оптимальные ПРВ и нормировочные константы параметризованы множителями Лагранжа, которые определяются решением следующих балансовых уравнений:

$$\frac{\mathcal{E}(\theta)}{\mathcal{Q}(\theta)} + \frac{\mathcal{T}_j(\theta^j)}{\mathcal{Q}_j(\theta^j)} = y^{(j)}, \quad j = \overline{1, s} \quad (13)$$

где

$$\mathcal{E}(\theta) = \int_{\mathcal{A}} \hat{z}(j|\alpha) \exp\left(-\sum_{j=1}^s \langle \theta^j, \hat{z}(j|\alpha) \rangle\right) d\alpha, \quad (14) \\ \mathcal{T}_j(\theta^j) = \int_{\Xi_j} \xi^{(j)} \exp(-\langle \theta^j, \xi^{(j)} \rangle) d\xi^{(j)}, \quad j = \overline{1, s}.$$

В нашем случае измеренные данные о площади озер, обладают большим числом пропущенных значений. Для восстановления пропущенных данных воспользуемся принципом энтропийной рандомизации по площади, используя имеющиеся данные по температуре и осадкам. Известно, что на площадь озер влияет температура и осадки, и в первом приближении это влияние можно описать линейной зависимостью с шумом в виде:

$$S[n] = \alpha T[n] + \beta R[n] + \xi[n]. \quad (15)$$

Коэффициенты α, β - случайные, интервальные:

$$\alpha \in \mathcal{A} = [\alpha^-, \alpha^+], \quad \beta \in \mathcal{B} = [\beta^-, \beta^+]. \quad (16)$$

Обозначим ПРВ параметров $P(\alpha), F(\beta)$.

Шум также стандартизованный и интервальный:

$$\xi[n] \in \Xi_j = [\xi^-, \xi^+]. \quad (17)$$

Обозначим ПРВ шума $Q_n(\xi[n])$.

Далее, применяя алгоритм рандомизированного машинного обучения, получим:

$$\mathcal{H} = - \int_{\mathcal{A}} P(\alpha) \ln P(\alpha) d\alpha - \int_{\mathcal{B}} F(\beta) \ln F(\beta) d\beta - \\ - \sum_{m=1}^k \int_{\Xi_m} Q_m(\xi[m]) \ln Q_m(\xi[m]) d\xi[m] \Rightarrow \max \quad (18)$$

при условии нормировки:

$$\int_{\mathcal{A}} P(\alpha) d\alpha = 1, \\ \int_{\mathcal{B}} F(\beta) d\beta = 1, \\ \int_{\Xi_m} Q_m(\xi[m]) d\xi[m] = 1, \quad m = \overline{1, k} \quad (19)$$

и эмпирических балансов:

$$\int_{\mathcal{A}} P(\alpha) \alpha T[m] d\alpha + \int_{\mathcal{B}} F(\beta) \beta R[m] d\beta + \\ + \int_{\Xi_m} Q_m(\xi[m]) \xi[m] d\xi[m] = S[m], \quad m = \overline{1, k}. \quad (20)$$

Решение задачи (18) имеет вид:

$$P^*(\alpha, \theta) = \frac{\exp(-\alpha l_r(\theta))}{\mathcal{P}(\theta)}, \\ F^*(\beta, \theta) = \frac{\exp(-\beta h_r(\theta))}{\mathcal{F}(\theta)}, \quad (21) \\ Q_m^*(\xi[m], \theta) = \frac{\exp(-\theta_m \xi[m])}{Q_j(\theta_m)}, \quad m = \overline{1, k}$$

где $\theta = \{\theta_1, \dots, \theta_k\}$ - множители Лагранжа;
- нормировочные коэффициенты

$$\mathcal{P}(\theta) = \int_{\mathcal{A}} \exp(-\alpha l_r(\theta)) d\alpha; \\ \mathcal{F}(\theta) = \int_{\mathcal{B}} \exp(-\beta h_r(\theta)) d\beta \quad (22) \\ Q_j(\theta_m) = \int_{\Xi} \exp(-\theta_m \xi[m]) d\xi[m], \quad m = \overline{1, k} \\ l_r(\theta) = \sum_{m=1}^k \theta_m T[m], \quad h_r(\theta) = \sum_{m=1}^k \theta_m R[m].$$

Для определения значений множителей Лагранжа нужно решить следующую систему уравнений:

$$L(\theta)T(m) + K(\theta)R(m) + Gm(\theta) = S(m), \quad m = \overline{1, k} \quad (23)$$

$$L(\theta) = \frac{\exp(-\alpha^- l_r(\theta))(\alpha^- l_r(\theta) + 1)}{\exp(-\alpha^- l_r(\theta)) - \exp(-\alpha^+ l_r(\theta))} + \\ + \frac{\exp(\alpha^+ l_r(\theta))(-\alpha^+ l_r(\theta) + 1)}{\exp(\alpha^+ l_r(\theta)) - \exp(-\alpha^+ l_r(\theta) + 1)} \\ K(\theta) = \frac{\exp(-\beta^- h_r(\theta))(\beta^- h_r(\theta) + 1) + \exp(\beta^+ h_r(\theta))(-\beta^+ h_r(\theta) + 1)}{\exp(-\beta^- h_r(\theta)) - \exp(-\beta^+ h_r(\theta))} + \\ + \frac{\exp(\beta^+ h_r(\theta))(-\beta^+ h_r(\theta) + 1)}{\exp(-\beta^- h_r(\theta)) - \exp(-\beta^+ h_r(\theta))} \\ Gm(\theta) = \frac{\exp(-\xi^-[m]\theta_m)(\xi^-[m]\theta_m + 1)}{\exp(-\xi^-[m]\theta_m) - \exp(-\xi^+[m]\theta_m)} + \\ + \frac{\exp(-\xi^+[m]\theta_m)(\xi^+[m]\theta_m + 1)}{\exp(-\xi^-[m]\theta_m) - \exp(-\xi^+[m]\theta_m)}. \quad (24)$$

Используя модель для расчета всех пропущенных данных и сэмплируя ПРВ (24), построим ансамбль траекторий $S[n]$. Вычислим среднюю траекторию и по ней заполним недостающие данные.

После этого преобразуем данные от нормализованных величин к натуральным (га). Для этого произведем действие, обратное к (1):

$$\tilde{S}_{\text{восст}} = S * (\tilde{S}_{\text{max}} - \tilde{S}_{\text{min}}) + \tilde{S}_{\text{min}}. \quad (25)$$

РЕЗУЛЬТАТЫ

В соответствии с алгоритмом были восстановлены данные на всех тестовых участках. Для иллюстрации на рис. 2. представлены результаты восстановления данных на тестовом участке № 8.

На рис. 2 (а) изображен график временного хода восстановленных значений площади озер. Сплошными чер-

ными точками отображены реальные (измеренные) значения. Также изображены графики временного хода данных температуры и осадков на рис. 2 (b) и рис. 2 (c) соответственно.

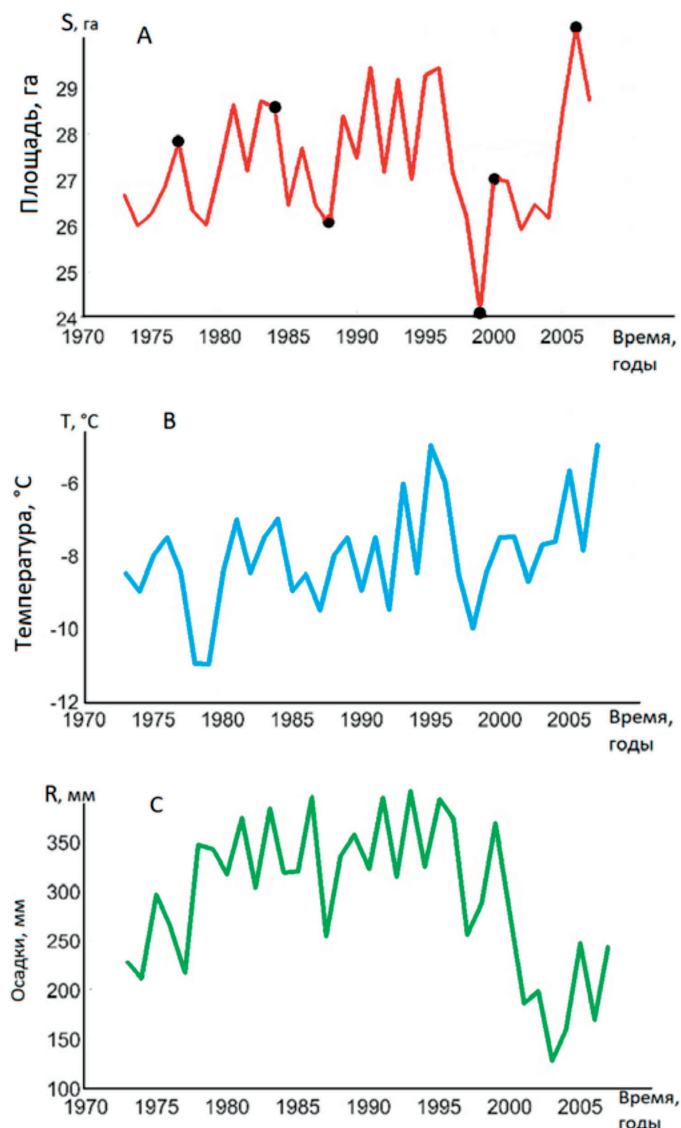


Рис 2. Временные ряды значений восстановленных данных о средней площади (а), результаты измерений средней температуры озер (b) и годовой суммы осадков (с) на ТУ-8

ОЦЕНКА АДЕКВАТНОСТИ МОДЕЛИ

Оценка погрешности моделирования (восстановления) проводилась как оценка среднего отклонения восстановленных значений от измеренных данных.

В табл. 2 приведены на всех тестовых участках измеренные и восстановленные (моделированные) значения в те годы, в которых были проведены измерения площадей озер.

Для каждого тестового участка рассчитаем среднее отклонение восстановленных данных о площади от измеренных значений по следующей формуле:

$$\Delta_s = \frac{1}{m} \sum_{i=1}^m \frac{|\tilde{S}_{\text{восст}i} - \tilde{S}_{\text{изм}i}|}{\tilde{S}_{\text{изм}i}}, \quad (27)$$

где $\tilde{S}_{\text{изм}i}$ – значение средней площади, полученное в результате измерений, $\tilde{S}_{\text{восст}i}$ – восстановленные значения средней площади, m – количество моделируемых точек на конкретном ТУ.

Таблица 2.

Восстановленные и измеренные данные на всех ТУ.

ТУ	Годы	$\tilde{S}_{\text{изм}}$, га	$\tilde{S}_{\text{восст}}$, га
5	1973	67,90	66,61
	1984	68,89	64,86
	1988	64,27	64,77
	2001	62,36	64,11
	2003	62,31	63,5
	2007	66,96	62,12
6	1973	47,60	42,95
	1987	37,05	36,68
	1988	39,43	34,47
	2001	33,51	37,46
	2003	34,30	35,79
	2007	48,66	33,03
7	1973	19,62	19,53
	1984	15,64	17,78
	1988	19,69	18,33
	2001	18,52	17,22
	2003	16,41	17,47
	2007	19,68	17,59
14	1984	23,04	21,7
	1989	23,07	21,51
	1999	21,37	22,48
	2001	20,94	21,71
	2006	21,21	20,5
	2007	20,54	20,85
15	1973	28,65	30,53
	1987	32,25	30,32
	2001	29,93	29,65
	2006	26,92	26,69
16	1973	17,45	16,58
	1987	16,96	15,36
	2001	14,58	15,85
	2006	16,09	15,26
	2007	16,53	16,1
20	1988	18,38	17,49
	1999	16,08	17,25
	2000	17,02	17,32

	2001	17,08	16,97
	2006	17,78	16,85
22	1977	27,82	26,36
	1984	28,59	28,32
	1988	26,02	26,66
	1999	24,14	27,8
	2000	27,04	27,7
	2006	30,32	26,34
24	1977	25,62	22,94
	1981	22,59	22,86
	1988	24,24	23,73
	1999	23,81	23,48
	2001	22,96	23,24
	2006	23,76	22,86

Полученные данные приведены в табл. 3.

Таблица 3.
Погрешности восстановления данных

ТУ	Δ_s
1	0,04
2	0,15
3	0,08
4	0,04
5	0,04
6	0,06
7	0,04
8	0,07
9	0,04
Среднее значение	0,05

Для оценки погрешности восстановления были рассчитаны величины среднего отклонения модельных значений площади от измеренных величин на всех тестовых участках по формуле:

$$\Delta = \frac{1}{k} \sum_{i=1}^k \frac{|\tilde{S}_{\text{восст}i} - \tilde{S}_{\text{изм}i}|}{\tilde{S}_{\text{изм}i}}, \quad (28)$$

где $\tilde{S}_{\text{изм}i}$ – измеренные значения площади, $\tilde{S}_{\text{восст}i}$ – восстановленные значения средней площади, k – количество всех измеренных значений средней площади по всем тестовым участкам.

В нашем случае было получено следующее значение: $\Delta = 0,06$, при этом среднеквадратичное отклонение, рассчитанное по стандартной формуле, равно 0,09. Следовательно, погрешность восстановления пропусков по разработанному алгоритму не превышает в среднеквадратическом 9%, что можно рассматривать как практически приемлемый результат.

ЗАКЛЮЧЕНИЕ

В статье рассмотрены методические вопросы нового подхода к восстановлению пропущенных значений в экспериментальных данных о площадях термокарстовых озер с использованием временных рядов среднегодовой температуры и годовой суммы осадков. В качестве экспериментальных данных о площадях термокарстовых озер и климатических параметрах (температура и сумма осадков) использованы результаты исследований, проведенных в арктической зоне Западной Сибири в период с 1973 по 2007 г. Для проведения исследований выбраны девять тестовых участков, по три в каждой из зон мерзлоты (сплошной, прерывистой и островной). Данные о среднегодовой температуре и годовой сумме осадков для каждого тестового участка получены методом реанализа. Данные о площадях озер являются результатами дистанционных измерений по космическим снимкам. Ввиду большого количества пасмурных дней на севере Западной Сибири в указанный период удалось получить не более восьми значений, полученных в те годы, когда имелись безоблачные снимки. Поэтому для восстановления пропущенных значений временного ряда, число которых значительно превышает число имеющихся измеренных значений, наиболее перспективным оказался подход, основанный на энтропийно-рандомизированном моделировании, методология которого разработана в [1, 3].

Разработан алгоритм восстановления пропущенных значений в рамках указанного подхода, реализованный с использованием инструментов MATLAB R2019a. Рассчитанные пропущенные значения для указанных девяти тестовых участков представлены в табличном виде. Для иллюстрации приведены временные ряды значений площади озер, температуры и осадков на одном из тестовых участков. Проведенный анализ погрешностей восстановления пропусков показал, что разработанный алгоритм позволяет восстанавливать пропущенные значения площадей озер по данным об изменении температуры и осадков с практически приемлемой точностью.

БЛАГОДАРНОСТИ

Работа выполнена при финансовой поддержке гранта РФФИ по проекту № 19-07-00282.

ЛИТЕРАТУРА

1. Попков Ю.С. Рандомизированное машинное обучение при ограниченных объемах данных. / А.Ю. Попков, Ю.А. Дубнов. – Москва: УРСС, 2019. – 310 с.
2. Полищук В.Ю. Геоимитационное моделирование полей термокарстовых озер в зонах мерзлоты / В.Ю. Полищук, Ю.М. Полищук. – Ханты-Мансийск: УИП ЮГУ, 2013. – 129 с.
3. Popkov Y.S. New Method of Entropy-Robust Estimation for Randomized Models under Limited Data / Y.S. Popkov, A.Y. Popkov // Entropy. – 2014. – vol. 16. – P. 675-698.
4. Иoffee А.Д., Теория экстремальных задач / А.Д. Иoffee, В. Тихомиров. – Москва: Наука, 1974. – 435 с.
5. Hastie T. The Elements of Statistical Learning / T. Hastie, R. Tibshirani, J. Friedman // Springer. – 2001. – 533 p.
6. Bishop C.M. Pattern Recognition and Machine Learning / C.M. Bishop, M. Jordan, J. Kleinberg // Springer, Series: Information Theory and Statistics. – 2006. – 345 p.

7. Tarasov A. Forecasting US money growth using economic uncertainty measures and regularization techniques // *International Journal of Forecasting*. – 2019. – vol. 35. – P. 443-457.
8. Marcellino M. A comparison of direct and iterated multisteps AR methods for forecasting macroeconomic time series / M. Marcellino, J.H. Stock, M. Watson // *Journal of Econometrics*. – 2006. – vol. 135, No. 1-2. – P. 499-526.
9. Eitrheim O. Testing the adequacy of smooth transition autoregressive models / O. Eitrheim, T. Terasirta // *Journal of Econometrics*. – 1996. – vol. 74, No. 1-2. – P. 59-75.
10. Molodtsova T. Out-of-sample exchange rate predictability with TR fundamentals / T. Molodtsova, D. Papell // *Journal of International Economy*. – 2009. – vol. 77. – P. 167-180.
11. Wang R. Forecasting the exchange rate using nonlinear Taylor rule based model / R. Wang, B. Motley, M. Stamatogiannis // *International Journal of Forecasting*. – 2019. – vol. 35. – P. 429-442.
12. Canale A., Bayesian nonparametric forecasting of nonmonotonic functional time series / A. Canale, M. Ruggiero // *Electronic Journal of Statistics*. – 2016. – vol. 10, No. 2. – P. 3265-3288.
13. Дубнов Ю.А., Байесовская идентификация параметров смеси нормальных распределений / Ю.А. Дубнов, А.В. Булычев // *Информационные технологии и вычислительные системы*. – 2017. – № 1. – С. 101-114.
14. Frazier D.T. Approximate Bayesian forecasting / D.T. Frazier, W. Maneesoonthorn, G.M. Martin, B.P.M. McCabe // *International Journal of Forecasting*. – 2019. – vol. 35. – P. 521-539.
15. Beaumont M.A. Approximate Bayesian computation in population genetics / M.A. Beaumont, W. Zhang, D.J. Baldung // *Genetics*. – 2001. – vol. 162, No. 4. – P. 2025-2035.
16. McAdam P. Euro-area real time density forecasting with financial or labour market friction. / P. McAdam, A. Warne // *International Journal of forecasting*. – 2019. – vol. 35. – P. 580-600.
17. Alkema L. The UN probabilistic population projection: an introduction to demographic forecasting with uncertainty / Alkema L., Geriand P., Raftery A. Wilmoth T. // *Forecasting*. – 2015. – vol. 37. – P. 19-24.
18. Blum M.G.B. A comparison review of dimension reduction methods in approximate Bayesian computation / M.G.B. Blum, M.A. Nunes, D. Prangle, S.A. Sisson // *Statistical Science*. – 2013. – vol. 28, No. 2. – P. 189-208.
19. Воронцов К. В. Математические методы обучения по прецедентам. – Курс лекций. – Москва: МФТИ, 2013. – 245 с.
20. Цыпкин Я.З. Основы теории обучающихся систем. – Москва: Наука, 1970. – 398 с.
21. Вапник В.Н. Теория распознавания образов. / В.Н. Вапник, А.Я. Червоненкис. – Москва: Наука, 1974. – 237 с.
22. Witten J.H. Data Mining: Practical Machine Learning Tools and Techniques / J.H. Witten, E. Frank // *Morgan Kaufmann*. – 2005. – 458 p.
23. Gaillard P. Additive models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting / P. Gaillard, K. Goude, R. Nedellec // *International Journal of Forecasting*. – 2016. – vol. 32. – P. 1038-1050.
24. Mangalova E. Wind power forecasting using the k-nearest neighbours algorithm / E. Mangalova, E. Agafonov // *International Journal of Forecasting*. – 2014. – vol. 30. – P. 402-406.
25. Zhang K. K-nearest neighbors and a kernel density estimator for GEFCom2014 probabilistic wind power forecasting / K. Zhang, J. Wang // *International Journal of Forecasting*. – 2016. – vol. 32. – P. 1074-1080.
26. Armstrong J. Long-range forecasting. – New York: Wiley, 1985. – 432 p.
27. Wheatcroft E. Interpreting the skill score from forecast performance metrics // *International Journal of Forecasting*. – 2019. – vol. 35. – P. 573-579.
28. Brier C. W. Verification of forecast expressed in term probability // *Monthly Weather Review*. – 1950. – vol. 78, No. 1-3. – P. 25-34.
29. Brocker J. From ensembles forecasts to predictive distribution functions / J. Brocker, L.A. Smith // *Tellus A*. – 2008. – vol. 60, No. 4. – P. 663-678.
30. Christensen H.M. Evaluation of ensemble forecast uncertainty using a new proper score: application to medium-range and seasonal forecast / H.M. Christensen, I.M. Moroz, T.N. Palmer // *Quarterly Journal of the Royal Meteorological Science*. – 2015. – vol. 141, No. 687. – P. 538-549.
31. Lahiri K. Evaluating probability forecasts for GDP declines using alternative methodological / K. Lahiri, J.C. Wang // *International Journal of Forecasting*. – 2013. – vol. 29, No. 1. – P. 175-190.
32. Gneiting T. Probabilistic Forecasting / T. Gneiting, M. Katzfuss // *Annual Review of Statistics and Its Applications*. – 2014. – vol. 1. – P. 125-151.
33. Vidyasagar M. Statistical Learning Theory and Randomized Algorithms for Control // *IEEE Control System Magazine*. – 1998. – vol. 1, No. 17. – P. 69-88.
34. Гранич О.Н. Рандомизированные алгоритмы оценивания и оптимизации при почти произвольных помехах. / О.Н. Гранич, Б.Т. Поляк // *Москва: Наука*, 2002. – 291 с.
35. Biondo A.E. Are random trading strategies more successful than technical ones? / A.E. Biondo, A. Pluchino, A. Rapisarda, D. Helbing // *PLoS ONE*. – 2013. – vol. 6, No. 7, P. 34-45.
36. Lutz W. The end of world population growth. / W. Lutz, S. Sandersen, S. Scherbov // *Nature*. – 2001. – vol. 412, No. 6846. – P. 543-545.
37. Цирлин А.М. Методы усредненной оптимизации и их применение. – Москва: Физматлит, 1997. – 304 с.
38. Shannon C. Communication Theory of Secrecy Systems // *Bell System Technical Journal*. – 1949. – vol. 28, No 4. – P. 656-715.
39. Jaynes E.T. Information Theory and Statistical Mechanics // *Physics Review*. – 1957. – vol. 106. – P. 620-630.
40. Jaynes E.T. Papers on probability, statistics and statistical physics. – Dordrecht: Kluwer Academic Publisher, 1989, 436 p.
41. Jaynes E.T. Probability Theory. The logic and science // *Cambridge: Cambridge University Press*, 2003. – 341 p.

A new approach to recovering missing data on the area of thermokarst lakes in the Arctic

Yuri S. Popkov

Federal Research Center "Informatics and Control" RAS,
Moscow, Russia
popkov.yuri@gmail.com

Andrey V. Melnikov

Ugra Research Institute of Information Technologies,
Khanty-Mansiysk, Russia
melnikovav@uriit.ru

Yury M. Polishchuk

Ugra Research Institute of Information Technologies,
Khanty-Mansiysk, Russia
yupolishchuk@gmail.com

Evgeny S. Sokol

Ugra Research Institute of Information Technologies,
Khanty-Mansiysk, Russia
eugen137@gmail.com

Vladimir Y. Polishchuk

Institute of Monitoring of Climatic and Ecological Systems,
Tomsk, Russia,
liquid_metal@mail.ru

Abstract. In the tasks of forecasting the volumes of methane as one of the important greenhouse gases considered in the list of causes of global warming today, it is necessary to use, along with climatic characteristics, data on the dynamics of the areas of thermokarst lakes in permafrost, which are usually obtained using satellite imagery. Because of a large number of cloudy days in the northern territories allows you to get only a small number of cloudless images, which leads to significant number of missing data in the time series of lakes areas. To restore missing values, it is necessary to use a new approach to recovering gaps based on methods and algorithms of entropy-randomized machine learning. This approach involves the restoration of missing values in the experimental data on the sizes of thermokarst lakes using time series of average annual temperature and annual precipitation. As experimental data on the thermokarst lakes areas and climatic parameters (temperature and amount of precipitation), we used the results of studies conducted in the Arctic zone of Western Siberia from 1973 to 2007. Studies were conducted in

nine test sites selected in different permafrost zones (continuous, discontinuous and insular). Data on the average annual temperature and annual precipitation for each test site were obtained by reanalysis. The developed algorithm for recovering missing values within the framework of this approach is implemented using the MATLAB R2019a tools. The missing values are calculated for the selected nine test sites. To illustrate, the time series of the values of the area of lakes, temperature and precipitation in one of the test sites are shown. An analysis of the omissions recovery errors was carried out, which showed that the developed algorithm allows us to restore the missing values of the lake areas from the data on changes in temperature and precipitation with practically acceptable accuracy.

Keywords: *randomization, machine learning, thermokarst lakes, climatic parameters, modeling, restoration of missing data.*

Системы и технологии искусственного интеллекта в медицине. Какой набор средств повысит качество решений?

Грибова В.В.

Институт автоматики и процессов управления ДВО
РАН
Владивосток, Россия
gribova@iacp.dvo.ru

Шалфеева Е.А.

Институт автоматики и процессов управления ДВО
РАН
Владивосток, Россия
shalfeeva@iacp.dvo.ru

Аннотация. Информационные технологии востребованы во всех областях, медицина не исключение. Несмотря на их уже активное использование, потенциал информационных технологий в медицине раскрыт недостаточно. Растет число врачебных ошибок, доля которых доходит в некоторых странах до 30%. Решением указанных проблем являются основанные на методах искусственного интеллекта системы поддержки принятия решений, однако их количество в медицине ничтожно мало и не покрывает потребности практикующих врачей и студентов медицинских вузов. Что мешает активному внедрению искусственного интеллекта в медицину, какие системы искусственного интеллекта в медицине должны быть, на каких принципах и моделях основанные? Ответы на эти вопросы дают авторы статьи.

Ключевые слова: искусственный интеллект, база знаний, система поддержки принятия решений, машинное обучение, онтологии.

ВВЕДЕНИЕ

Современный этап развития информатизации здравоохранения характеризуется внедрением в деятельность медицинских организаций информационных систем, охватывающих управление ресурсами, лечебным процессом, организацию взаимодействия пациентов с сотрудниками медучреждений. Развитие Интернета дало возможность пополнять набор удобных справочников и сервисов для врачей (типа калькуляторов индекса инвалидности), что стало важным направлением для Российского здравоохранения.

С начала 60-х в рамках научно-технологического направления "искусственный интеллект" (ИИ) создавались модели представления знаний, методы и технологии решения интеллектуальных задач, было создано множество прототипов экспертных систем, использующих в своем функционировании знания, заложенные экспертами, и позволяющие выполнять функции, присущие этим экспертам. В последние несколько лет на рынок вышли многочисленные фреймворки и библиотеки, поддерживающие нейросетевой подход к созданию систем такого класса.

В области медицины все разработанные методы, модели и технологии активно использовались и используются для решения задач диагностики заболеваний, лечения и распознавания изображений. Так, типичными примерами систем на основе баз знаний являются MICIN, ONCOCIN, Ada, Qapsula, DOC+, Диаген, ESTER, Caduceus и др., на

основе методов машинного обучения - Deepmind Health, Watson, Pathai, Quantitative Insights, Concerto Health AI и др.[1-11].

Несмотря на наличие различных подходов, технологии, использующие нейросети, в последнее время часто называют будущим искусственного интеллекта. Так ли это? Позволят ли они создавать и внедрять системы искусственного интеллекта в здравоохранении на новом уровне? Какими должны быть технологии ИИ, какими свойствами обладать? Авторы работы дают ответ на эти вопросы в данной работе.

ИДЕАЛЬНАЯ СИСТЕМА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ МЕДИЦИНЫ. КАКАЯ ОНА?

Искусственный интеллект — это раздел информатики, который моделирует функции естественного интеллекта с помощью компьютеров. Его цель — решение интеллектуальных задач не хуже человека. Какими функциями или интеллектуальными способностями обладает интеллект естественный? Ответ на этот вопрос в работе [12] дал В.К. Финн. Он выделил тринадцать таких способностей. Коротко их можно перечислить следующим образом: обнаружение существенного в данных; порождение последовательностей «цель—план—действие»; поиск посылок, релевантных цели рассуждений; способность к рассуждению: выводы, как следствия из посылок, выводы, которые непосредственно из посылок не следуют: выводы по аналогии, индуктивные обобщения данных, выводы с помощью объяснения данных (абдукция), решение задач при помощи различных познавательных процедур; способность к рефлексии — оценка знаний и действий; способность к объяснению — ответ на вопрос «почему?»; аргументация при принятии решения; способность к распознаванию — ответ на вопрос «что это такое?»; способность к обучению и использованию памяти; способность к интеграции знания для образования концепций и теорий; способность уточнения неясных идей и преобразования их в точные понятия; способность к изменению системы знаний при получении новых знаний и изменениях ситуации.

Таким образом, идеальная система ИИ должна обладать всеми способностями естественного интеллекта, при этом очевидно, что не все функции естественного интеллекта можно автоматизировать, некоторые из них возможно реализовать только посредством взаимодействия "человек-компьютер". Таким образом, если мы хотим создать идеальную систему ИИ, то она должна обладать

В.В. Грибова, Е.А. Шалфеева

всеми тринадцатью интеллектуальными способностями. В любом случае чем больше интеллектуальных способностей система ИИ реализует, тем выше уровень ее интеллектуальности. Важно отметить: в естественном интеллекте результат одного (или нескольких) видов познавательной деятельности всегда интегрируется с другими ее видами.

Так, например, если мы обладаем медицинскими знаниями и имеется пациент или больной, то мы, на основе своей системы знаний, можем сделать какой-либо логический вывод: оценить необходимость оказания помощи, выработать варианты решений по воздействию на организм и выбрать максимально адекватный; далее следить за состоянием организма; выявлять признаки ухудшения ситуации.

Если признаки пациента не вписываются в существующую систему знаний врача (ни одному из диагнозов картина не соответствует или ни одно лечение не подходит организму) - имеем прецедент. Последовательность пробных решений и их результат документируются (возможно, подобные случаи будут повторяться и имеющийся прецедент позволит принять решение рассуждением по аналогии). Несколько таких прецедентов могут инициировать получение новых теоретических знаний. Иногда знания могут быть добыты в результате машинного обучения. Накопление прецедентов с известным результатом (обучение) создает ситуацию, когда необходимо скорректировать систему знаний и т.д.. Новые знания должны интегрироваться с существующей базой знаний. База знаний будет, таким образом, непрерывно совершенствоваться, а сами знания будут доступны как системе ИИ, так и человеку, который на их основе сможет получить новое знание способом, недоступным для компьютерной системы.

Учитывая, что не все функции естественного интеллекта могут быть автоматизированы, требуется участие человека, поэтому важно, чтобы и знания (в компьютерной системе это база знаний), и прецеденты (данные) были понятны человеку.

Должна быть унифицирована терминология, которая понятна и принята сообществом. Если процедуры обработки информации (по использованию и по формированию медицинских знаний, документов и данных) работают в единых понятиях, видах отношений, законах, в единой классификации терминов, то связь между ними будет "бесшовной".

Для использования информации о состоянии пациента с технологичных аппаратов и приборов, система ИИ должна "понимать и принимать" такую информацию с внешних устройств. Она должна "комплексировать" требуемые подмножества познавательных процедур: вывод следствий из посылок - работа с имеющейся системой знаний, рассуждение по аналогии (в реальной жизни, естественном интеллекте, она играет существенную роль), обучение, как способ корректировки системы знаний, способность к объяснению и обоснованию своих действий.

Комплекс таких автоматизированных и автоматических познавательных процедур обработки информации, работающих в единых терминах, может претендовать на то, чтобы называться системой ИИ в медицине.

Создание технологий искусственного интеллекта в медицине должно охватывать не только системы поддержки принятия решений разного типа, но также и тренажеры для обучения студентов и повышения квалификации врачей.

ТРЕБОВАНИЯ К ИНСТРУМЕНТАМ СОЗДАНИЯ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В МЕДИЦИНЕ

Можно выделить следующие основные требования к инструментам для создания ИИ в медицине.

1. Поддержка разработки нескольких видов интеллектуальной деятельности с "бесшовной" интеграцией.

Данное требование рассмотрено в п. выше.

2. Наличие средств поддержки жизнеспособности создаваемых систем и средств их разработки.

Данное требование является очень емким, состоит из совокупности требований. Его выполнение, как и предыдущее требование, говорит о степени зрелости разработки. Этому требованию необходимо уделять ключевое внимание как при создании инструментов, так и создании систем ИИ на их основе. К сожалению, обычной практикой для большинства реализаций являются декларации о функциональных возможностях некоего программного средства с демонстрацией результатов. Вместе с тем разработчики затрудняются перечислить механизмы, заложенные в обеспечение жизнеспособности созданной системы.

3. Интеграция с существующими средствами выполнения практической работы.

Система ИИ должна иметь возможность взаимодействовать со сторонними программными системами на основе различных моделей взаимодействия.

Также важно мягко «встроиться» в существующие традиции работы специалистов (существующий документооборот), добавляя опциональные возможности, улучшающие практику, и, по мере их принятия, наращивая функциональность системы.

4. Коллективная разработка.

Учитывая, что современная система ИИ требует участия многих специалистов, инструментарий должен обладать средствами, обеспечивающими коллективную разработку компонентов системы. При этом коллективность должна поддерживаться как на уровне различных компонентов, так и давать возможность коллективу соавторов разрабатывать один компонент (например, базу знаний). Для этого должна обеспечиваться иерархическая система управления правами доступа к компонентам, мониторинга изменений, согласования между компонентами через личные кабинеты пользователей и облачные средства создания компонентов и управления ими.

ИМЕЮЩИЙСЯ НАУЧНЫЙ И ТЕХНОЛОГИЧЕСКИЙ ЗАДЕЛ У КОЛЛЕКТИВА АВТОРОВ

Коллектив, в котором работают авторы, имеют следующий научный и технологический задел [13-17]:

- разработан и апробирован онтологический подход к формированию баз знаний, классификаторов и других информационных ресурсов, основанный на семантическом представлении;

- разработан классификатор медицинских понятий (бо-

лее 27 тысяч вершин), необходимых для формирования знаний (по диагностике, планированию лечения, прогнозу, отслеживанию критических состояний) и других документов;

- разработаны базы знаний по диагностике и лечению ряда заболеваний (более 100000 вершин);

- разработаны программные сервисы по диагностике и лечению заболеваний произвольной направленности;

- разработан тренажер по отработке навыков постановки диагноза.

Все информационные и программные ресурсы реализованы на платформе IACPaaS (iacpaas.dvo.ru)[18]. На ней создана разнообразная инструментальная поддержка для редактирования баз знаний, классификаторов и других информационных ресурсов, а также создания программных сервисов по трем технологиям.

БЛАГОДАРНОСТИ

Работа выполнена при частичной поддержке РФФИ, гранты 18-29-03131 и 19-29-01077.

ЛИТЕРАТУРА

1. Рыбина Г.В. Интеллектуальные системы: От А до Я. Серия монографий в 3 книгах. Книга 3. Проблемно-специализированные интеллектуальные системы. Инструментальные средства построения интеллектуальных систем. 2015. М.: Издательство "Научтехлитиздат". 180 с

2. D.K.Srekantha, T.M. Girish and Dr. R.H. Fatterpur Applications of Experts Systems in Medical Sciences – A Review // Proc. of 18th International Conference on Fuzzy Set theory and Applications. Paris, France, pp - 2063-2066, 2016

3. IBM Watson has graduated from school and got a job [Internet]. Available at: <http://habrahabr.ru/company/ibm/blog/169067/>. Accessed July 23, 2018.

4. B. Prasad, H. Wood, J. Greer, G. McCalla, "A knowledge-based system for tutoring bronchial asthma diagnosis" proceedings of Second Annual IEEE Symposium on Computer-Based Medical Systems 1989 IEEE.

5. E. Bursuk, M. Ozkan, B. Llerigelen, "A medical expert system in cardiological diseases" proceedings of Engineering in Medicine and Biology twenty first Annual Conference and the 1999 Annual Fall Meeting of the Biomedical Engineering Society 1999 IEEE.

6. Neshat, M, Yaghobi, M, Naghibi, M.B, Esmaelzadeh, A, "Fuzzy Expert System Design for Diagnosis of Liver Disorders" proceedings of Knowledge Acquisition and Modeling, KAM eighth International Symposium 2008 IEEE.

7. Samy S. Abu Naser, Abu Zaiter A. Ola, "An Expert System for Diagnosing Eye Diseases Using CLIPS", Journal of Theoretical and Applied Information Technology, pp. 923-930, 2005-2008 JATIT.

8. Azian Azamimi Abdullah, Zulkarnay Zakaria, Nur Farahiyah Mohammad, "Design and Development of Fuzzy Expert System for Diagnosis of Hypertension", Second International Conference on Intelligent Systems, Modelling and Simulation 2011 IEEE.

9. Maitri Patel, Atul Patel, Paresh Virparia, "Rule Based Expert System for Viral Infection Diagnosis", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 3, Issue 5, May 2013.

10. Jimmy Singla, "The Diagnosis of Some Lung Diseases in a PROLOG Expert System", International Journal of Computer Applications, vol. 78, no. 15, pp. 37-40, September 2013.

11. Noura Ajam, "Heart Disease Diagnosis using Artificial Neural Network", IISTE Network and Complex Systems, Vol. 5, No. 4, 2015.

12. Ю. Арский, В. Финн Принципы конструирования интеллектуальных систем// Информационные технологии и вычислительные системы №4. 2008. с. 4-37

13. Грибова В.В., Петряева М.В., Окунь Д.Б., Шалфеева Е.А. Онтология медицинской диагностики для интеллектуальных систем поддержки принятия решений // Онтология проектирования. 2018. Т. 8. №1(27). С. 58-73. DOI:10.18287/2223-9537-2018-8-1-58-73

14. Грибова В.В., Окунь Д.Б. Онтологии для формирования баз знаний и реализации лечебных мероприятий в медицинских интеллектуальных системах // Информатика и системы управления. 2018. № 3(57). С. 71-80. DOI: 10.22250/isu.2018.57.71-80

15. Valeria V. Gribova, Margaret V. Petryaeva, Dmitry B. Okun, Alexey V. Tarasov, Software Toolkit for Creating Intelligent Systems in Practical and Educational Medicine // IEEE Xplore. 2018. DOI: 10.1109/RPC.2018.8482130

16. Грибова В.В., Петряева М.В., Шалфеева Е.А. Облачный сервис поддержки принятия диагностических решений в гастроэнтерологии // Врач и информационные технологии. 2019. № 3. С. 65-71.

17. Gribova V.V., Ostrovskii G.E. An Intelligent Cloud Service for Training Diagnostics Skills // Biomedical Engineering. 2018. Vol. 51. N 6. Pp. 416-421. DOI 10.1007/s10527-018-9761-8.

18. Valeria Gribova, Alexander Kleshev, Philip Moskalenko, Vadim Timchenko, Leonid Fedorischev, Elena Shalfeeva. The IACPaaS Cloud Platform: Features and Perspectives // 2017 Second Russia and Pacific Conference on Computer Technology and Applications (RPC) (Vladivostok, Russia, 25-29 sept. 2017). IEEE. 2017. P. 80-84. DOI: 10.1109/RPC.2017.8168076.

Artificial intelligence systems and technologies in medicine. What set of tools will improve the quality of solutions?

V. Gribova

Institute of Automation and Control Processes
Far Eastern Branch of Russian Academy
of Sciences
Vladivostok, Russia
e-mail: gribova@iacp.dvo.ru

E. Shalfeeva

Institute of Automation and Control Processes
Far Eastern Branch of Russian Academy
of Sciences
Vladivostok, Russia
e-mail: shalf@iacp.dvo.ru

Abstract. Information technologies are in demand in all areas, and medicine is no exception. Despite their already active use, the potential of information technologies in medicine is not disclosed enough. There is a growing number of medical errors, the number of which reaches up to 30% in some countries. The solution to these problems is based on artificial intelligence decision support systems, but their number in medicine is small and does not cover the needs of practicing doctors and medical students. What prevents the active implementation of artificial intelligence in

medicine, what systems of artificial intelligence in medicine should be? What principles and models should they be based on? The authors of the article give answers to these questions.

Keywords: artificial intelligence, knowledge base, decision support system, machine learning, ontologies.

Решение вариационной задачи ICP в замкнутой форме для аффинных преобразований для идентификации трещиноватостей на основе 3D-сейсморазведочной информации

Вохминцев А.В.

Челябинский государственный университет
Челябинск, Россия
vav@csu.ru

Аннотация. Нефтяные ресурсы Ямало-Ненецкого автономного округа составляют 300 млн. тонн, большая часть которых относится к трудноизвлекаемым. Единственным на данный момент экономически рентабельным способом вовлечения в разработку глубокопогруженных запасов является добыча нефти и газа из зон трещиноватости. В работе предлагается информационная технология, которая направлена на выявление и картографирование зон трещиноватости с целью прогнозирования высокопродуктивных нефтегазоносных участков в залежах юрского и доюрского комплексов. Математическую основу данной технологии составляют методы машинного обучения и искусственного интеллекта.

Ключевые слова: зоны трещиноватости, картирование трещинных систем, трудноизвлекаемые запасы, сейсморазведка, методы машинного обучения и искусственного интеллекта, сопоставление изображений, регистрация облаков точек.

ВВЕДЕНИЕ

Известно, что мировые запасы легкодоступных нефтей практически исчерпаны, и необходимый прирост добычи нефти в будущем возможен в основном за счет трудноизвлекаемых запасов глубокопогруженных залежей, коллектора которых характеризуются низкими и ультранизкими фильтрационными свойствами. Единственным на данный момент экономически рентабельным способом вовлечения в разработку глубокопогруженных запасов является добыча нефти и газа из зон трещиноватости, локально

Нефтяные ресурсы Ямало-Ненецкого автономного округа составляют 300 млн. тонн, большая часть которых относится к трудноизвлекаемым. Наиболее эффективным на данный момент экономически рентабельным способом вовлечения в разработку трудноизвлекаемых глубокопогруженных запасов является добыча нефти из зон трещиноватости, локально развитых по этим низкопроницаемым комплексам пород. Локализация трещин в узкие протяженные зоны следует из закономерностей постатийного развития разрывных нарушений, которые на ранних этапах развития представляют собой зоны трещиноватости (Шерман, 1977). Для платформенных разрывов характерно образование сеточных структур нарушений в складча-

том основании коренных пород [1]. Закономерности развития трещинных систем в породах осадочного чехла Западной Сибири рассмотрены в работах [2, 3]. Разработана методика выявления зон трещиноватости, подтвержденная результатами бурения на ряде месторождений Западной Сибири. Анализ данных об интенсивности тепловых потоков и физико-химических свойствах нефти позволяет локализовать наиболее перспективные территории нефтегазоносных бассейнов. Разработанная методика ввиду необходимости использования больших объемов сейсморазведочных данных предполагает большой объем ручного труда для интерпретации массивов 3D-сейсморазведочной информации. Однако, современные достижения в области машинного обучения могут позволить автоматизировать наиболее трудоемкие этапы выявления и картографирования зон трещиноватости на обширных территориях нефтегазоносных районов. При решении задачи обнаружения и классификации объектов на сейсморазведочных данных, искаженных помехами различной природы можно выделить три основных подхода: поиск специальных характеристик объекта распознавания на наблюдаемой сцене; статистические методы распознавания; с использованием обучаемых нейронных сетей [4]. Методы, основанные на поиске характеристик объекта, выделяют из изображения объекта инвариантные к аффинным преобразованиям и яркостным изменениям характеристики объекта распознавания, такие как гистограммы ориентированных градиентов [5], инвариантные к масштабированию особые точки [6], затем, размерность пространства характеристик максимально уменьшается, например, с помощью метода главных компонент, и полученные вектора характеристик сравниваются между собой. Процесс выделения и сравнения характеристик в лучших по качеству обнаружения алгоритмах этого типа обладают высокой вычислительной сложностью, алгоритмы этого типа плохо формализуются, не контролируют качество распознавания (например, точность локализации) и дают большие ошибки обнаружения объектов на сценах, содержащих сложный пространственно неоднородный фон, искаженных неравномерным освещением [7-9]. Распознавание объектов с помощью линейной или нелинейной (на основе порядковых статистик) фильтрации [10, 11] является хорошей альтернативой методам, осно-

А.В. Вохминцев

ванным на поиске характеристик объекта, и примером статистического распознавания образов. Обнаружению монохромных и многокомпонентных объектов с помощью линейной фильтрации посвящено огромное количество работ, однако, большинство методов сводится к вычислению различных вариантов функции корреляции заданного объекта с наблюдаемым изображением и последующим сравнением с порогом [12]. Теоретически и экспериментально установлено, что на достаточно сложных изображениях корреляционный обнаружитель имеет большую вероятность ошибочного отождествления искомого объекта с посторонними объектами фона. Для того чтобы увеличить надежность обнаружения предложены многочисленные эвристические усовершенствования: пространственное дифференцирование, предсказание формы объекта, оптимальный линейный обнаружитель, адаптивные и локально-адаптивные линейные и нелинейные алгоритмы надежного обнаружения объектов на изображениях [13, 14]. Методы, использующие обучаемые нейронные сети, как правило, дают хороший результат при наличии небольшого количества объектов распознавания на неискаженных контекстуально-простых сценах. В настоящий момент глубокое машинное обучение уже применяется для анализа трехмерных данных. Например, они позволяют осуществлять классификацию трехмерных объектов по точкам [15, 16], графам [17, 18] и другим структурам данных [19]. Для решения проблемы высокой вычислительной ресурсоемкости для анализа сейсморазведочных данных возможно использование моделей нейросетей на разреженных структурах данных [20, 21] и вариационных автокодировщиков [22]. Более того, существуют определенные достижения в уменьшении объема обучающей выборки: например, нейронные сети, робастные к изменению пространственного положения трехмерных объектов [23]; модели, обучаемые с частичной (semi-supervised) разметкой [24] и т.п. Все вышеупомянутые наработки могут быть использованы в задачах проекта.

В работе [25] сравниваются результаты прогноза трещиноватости по данным сейсморазведки при использовании различных методов интерпретации. Основным критерием оценки достоверности является соответствие сейсмического прогноза скважинным данным. На ряде примеров зон трещиноватости, достоверно установленных по разносторонним скважинным данным, показана точность выявления сейсмических аномалий с помощью различных методик: расчёта кубов когерентности, различных модификаций геометрических атрибутов, азимутальной анизотропии и миграции дуплексных волн (Рис. 1.). Однако, до сих пор в достаточной мере не разработаны методические вопросы использования методов машинного обучения с геолого-геофизической информацией и данными о физико-химических свойствах нефтей.

ПОСТАНОВКА ЗАДАЧИ

Целью данного исследования является разработка информационной технологии выявления и картографирования зон трещиноватости юрского и доюрского комплексов на основе данных сейсморазведки и методов машинного обучения и искусственного интеллекта. Для локализации зон трещиноватости можно использовать методы распознавания и выбора признаков на основе картографирования и техник глубокого машинного обучения, например, автокодировщик

(autoencoder). Применительно к задаче распознавания зон трещиноватости автокодировщики позволяют выделять информативные и характерные особенности локализации узких протяженных зон. Альтернативным направлением при решении задачи локализации зон трещиноватостей является поиск по подобию в базах данных сейсморазведочной информации с положением ранее закартированных зон трещиноватости. Для решения этой задачи в работе предлагается использовать методы трехмерной регистрации данных на основе итеративного метода ближайших точек (Iterative Closest Point, ICP). Ключевым элементом алгоритма является поиск ортогонального или аффинного преобразования, наилучшим образом, в смысле квадратичной метрики совмещающим два облака точек с заданным соответствием между точками (вариационная подзадача алгоритма).

В данной работе рассматриваются следующие задачи:

1. Разработка метода решения вариационной задачи в замкнутой форме для аффинных преобразований для идентификации трещиноватостей на основе 3D-сейсморазведочной информации.
2. Создание базы данных сейсморазведочной информации с положением ранее закартированных зон трещиноватости.

РЕШЕНИЕ ВАРИАЦИОННОЙ ЗАДАЧИ ICP В ЗАМКНУТОЙ ФОРМЕ ДЛЯ АФФИННЫХ ПРЕОБРАЗОВАНИЙ

Пусть $X = \{x_1, \dots, x_n\}$ первое облако точек и через $Y = \{y_1, \dots, y_m\}$ второе облако точек в \mathbb{R}^3 . Предположим, что отношения между точками в облаках X и Y такие, что для каждой точки в x_i можно вычислить соответствующую точку в y_i . Тогда итеративный алгоритм ближайших точек можно рассмотреть как геометрическое преобразование ригидных объектов из X в Y следующего вида:

$$Rx_i + T, \quad (1)$$

где R - матрица поворота, T вектор параллельного переноса, $i = 1, \dots, n$.

Аналогично можно ввести определение масштабированного алгоритма (S-ICP) как следующее геометрическое преобразование:

$$RSx_i + T, \quad (2)$$

где S - матрица масштаба.

Можно представить функцию для определения пар совпадающих точек в следующем виде:

$$E(R, S, T) = \frac{1}{N} \sum_{i=1}^N \| (RSx_i + T) - y_i \|^2. \quad (3)$$

Пусть lm - количество соответствующих пар точек. Давайте найдем минимум новой функции сопоставления кадров. Если выражение (3) имеет минимум, то он удовлетворит следующему уравнению:

$$\frac{dE(T)}{dT} = 0, \quad (4)$$

$$\frac{dE(T)}{dT} = \frac{2S}{N} \sum_{i=1}^N ((RSx_i + T) - y_i) = 0. \quad (5)$$

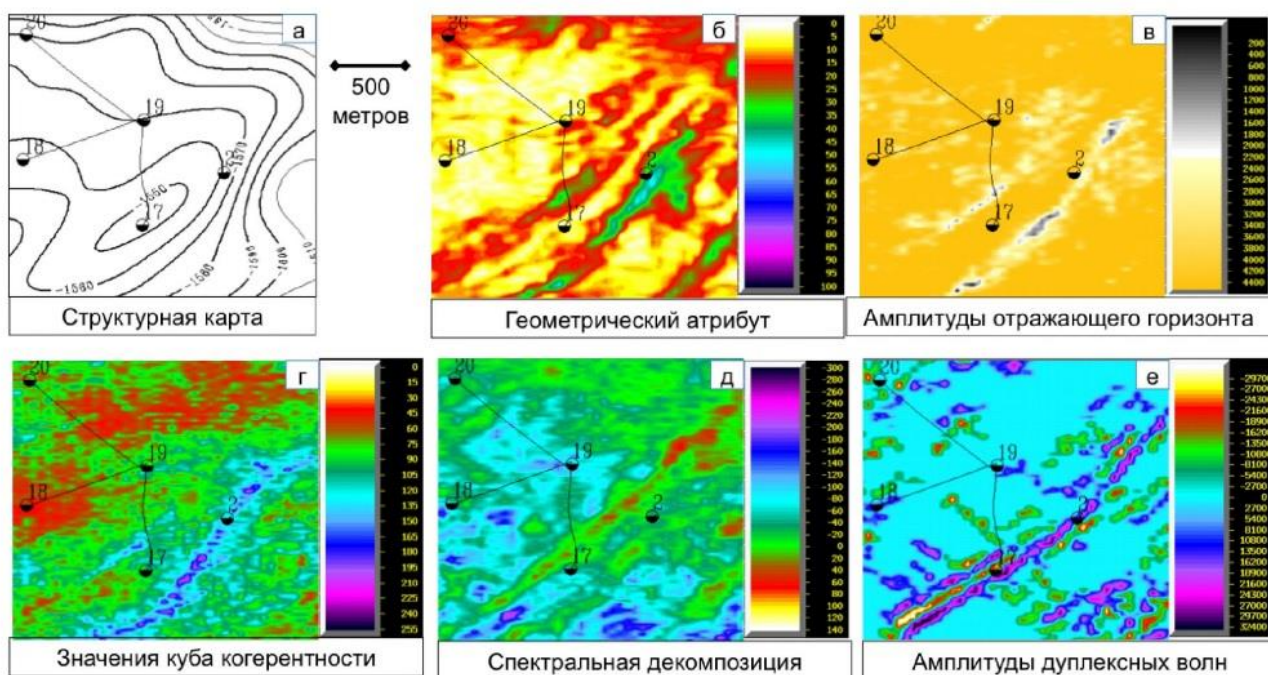


Рис. 1. Фрагменты карт различных сейсмических атрибутов, направленных на выделение тектонических деструкций и соответствующих одной интенсивной зоне трещиноватости

Следовательно, мы имеем

$$T = \frac{1}{N} \sum_{i=1}^N y_i - \frac{1}{N} \sum_{i=1}^N RSx_i.$$
 Таким образом, функция сопоставления для того чтобы достигнуть минимума должна быть представлена в следующем виде:

$$E(R, S) = \frac{1}{N} \sum_{j=1}^N \left\| RS \left(x_i - \frac{1}{N} \sum_{i=1}^N x_i \right) - \left(y_i - \frac{1}{N} \sum_{i=1}^N y_i \right) \right\|^2. \quad (6)$$

Пусть $a_i = \left(x_i - \frac{1}{N} \sum_{i=1}^N x_i \right)$, $b_i = \left(y_i - \frac{1}{N} \sum_{i=1}^N y_i \right)$, тогда функция сопоставления может быть представлена как:

$$E(R, S) = \frac{1}{N} \sum_{i=1}^N \|RSa_i - b_i\|^2. \quad (7)$$

Затем

$$E(R, S) = \frac{1}{l_m} \left(\sum_{i=1}^{l_m} a_i^T S^2 a_i - 2 \sum_{i=1}^{l_m} b_i^T RSa_i + \sum_{i=1}^{l_m} b_i^T b_i \right). \quad (8)$$

Для того чтобы найти минимум данной функции сопоставления, необходимо найти решение следующего дифференциального уравнения:

$$\frac{dE(R, S)}{dR} = 0. \quad (9)$$

Чтобы получить матрицу масштаба из уравнения (2.7) сделаем следующие преобразования:

$$\frac{dE(R, S)}{dS} = 2 \sum_{i=1}^{l_m} a_i^T SE_j a_i - 2 \sum_{i=1}^{l_m} b_i^T RSE_j a_i = 0 \quad (10)$$

где E_j есть диагональная матрица, $j = 1, 2, 3$ есть скаляр, который определяет размерность матрицы.

БАЗА ДАННЫХ СЕЙСМОРАЗВЕДОЧНОЙ ИНФОРМАЦИИ С ПОЛОЖЕНИЕМ РАНЕЕ ЗАКАРТИРОВАННЫХ ЗОН ТРЕЩИНОВАТОСТИ

В ходе ранее выполняемых исследований по выявлению зон трещиноватости и выработке рекомендаций по заложению поисково-разведочных и эксплуатационных скважин на основе данных сейсморазведки была сформирована база данных (БД-1), включающая геолого-геофизическую и скважинную информацию. База данных (БД-2) по физико-химическим свойствам нефти мира была создана более десяти лет назад. БД-2 зарегистрирована в Государственном реестре баз данных и в Роспатенте (Комитет Российской Федерации по патентам и товарным знакам, свидетельство № 2001620067). В настоящее время в базе данных БД-2 содержится более 30 000 образцов нефти 6000 месторождений в 191 нефтеносных бассейнах, расположенных в 94 странах.

ЗАКЛЮЧЕНИЕ

В работе предложена концепция, которая направлена на выявление и картографирование зон трещиноватости с

А.В. Вохминцев

целью прогнозирования высокопродуктивных нефтегазовых участков в залежах юрского и дюрского комплексов. Дан обзор современного состояния проблемы по российским и зарубежным источникам. Математическую основу данной технологии составляют методы машинного обучения и искусственного интеллекта. В работе рассмотрен один из таких методов, который использует решение вариационной задачи ICP в замкнутой форме для аффинных преобразований для идентификации трещиноватостей на основе 3D-сейсморазведочной информации.

БЛАГОДАРНОСТИ

Работа выполнена при поддержке гранта Российского фонда фундаментальных исследований и Ямало-Ненецкого Автономного округа в рамках научного проекта № 19-47-890001.

ЛИТЕРАТУРА

- Гзовский М.В. Основы тектонофизики. М., Наука, 1975, 536 С.
- Глухманчук Е.Д. Трещинно-блоковая структура Салымского месторождения как ведущий фактор в эффективной выработке запасов баженовской свиты / Крупницкий В., Леонтьевский А.В., Буков О.В. // Пути реализации нефтегазового и рудного потенциала ХМАО – Югра. Ханты-Мансийск. – 2013. – С. 337–342.
- Глухманчук Е.Д. Характеристика зон трещиноватости по неоднородности структуры поля деформаций отражающих горизонтов / Василевский А.Н. // Геология и геофизика. – 2013, Т. 54, № 1. – С. 106–112.
- Duda R.O., Hart P.E., D.G. Stork Pattern Classification (2nd Edition), Wiley-Interscience, 2000.
- Dalal N. Histograms of Oriented Gradients for Human Detection / Triggs B. // Computer Vision and Pattern Recognition. – 2005. – vol. 1. – P. 886–893.
- Lowe D.G. Object recognition from local scale-invariant features // Proc. Int. Conference on Computer Vision. – 1999. – vol. 2. – P. 1150–1157.
- Bay H. SURF: Speeded Up Robust Features / Ess A., Tuytelaars T., Van Gool L. // Computer Vision and Image Understanding. – 2008. – vol. 110 (3). – P. 346–359.
- Calonder M. BRIEF: Binary robust independent elementary features / Lepetit V., Strecha C., Fua P. // In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS 6314, P. 778–792.
- Zalesky B.A. Scale Invariant Algorithm to Match Regions on Aero or Satellite Images / Lukashevich P.V. // Proc. Pattern Recognition and Information Processing. – 2011. – P. 25–30.
- Manzurv T. Optical correlator based target detection, recognition, classification, and tracking / Zeller J., Serati S. // Applied Optics. – 2012. – vol. 51, P. 4976–4983.
- Ouerhani Y. Optimized preprocessing input plane GPU implementation of an optical face recognition technique using a segmented phase only composite filter / Jridi M., Alfalou A., Brosseau C. // Optics Communications. – 2013. – vol. 289. – P. 33–44.
- Yaroslavsky L.P. The theory of optimal methods for localization of objects in pictures // Progress in Optics, E. Wolf, ed., Elsevier, pp. 145–201, 1993.
- Kumar B.V.-K.V. Correlation pattern recognition / Mahalanobis A., Juday R.D., Cambridge Univ Press, 2005.
- Martinez-Diaz S. Nonlinear synthetic discriminant function filters for illumination-invariant pattern recognition/ Kober V. // Optical Engineering. – 2008. – vol. 47 (6), pp. 067201_067201_9.
- Charles R. Qi, Hao Su, Kaichun Mo.: PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, 2016; [http://arxiv.org/abs/1612.00593 arXiv:1612.00593].
- Komarichev A., Zhong Z.: A-CNN: Annularly Convolutional Neural Networks on Point Clouds, 2019; [http://arxiv.org/abs/1904.08017 arXiv:1904.08017].
- Edward J. Fujimoto S., Romero A.: GEOMETRICS: Exploiting Geometric Structure for Graph-Encoded Objects, 2019; [http://arxiv.org/abs/1901.11461 arXiv:1901.11461].
- Lerer A., Wu L., Shen J., Lacroix T., Wehrstedt L., Bose A.: PyTorch-BigGraph: A Large-scale Graph Embedding System, 2019, Proceedings of The Conference on Systems and Machine Learning, 2019; [http://arxiv.org/abs/1903.12287 arXiv:1903.12287].
- Zhou Y.: VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection, 2017; [http://arxiv.org/abs/1711.06396 arXiv:1711.06396].
- Riegler G., Ulusoy A.O.: OctNet: Learning Deep 3D Representations at High Resolutions, 2016; [http://arxiv.org/abs/1611.05009 arXiv:1611.05009].
- Wang P.-S., Sun C.-Y., Liu Y.: Adaptive O-CNN: A Patch-based Deep Representation of 3D Shapes, 2018, ACM Transactions on Graphics, 2018; [http://arxiv.org/abs/1809.07917 arXiv:1809.07917].
- Brock A., Lim T., Ritchie J.-M.: Generative and Discriminative Voxel Modeling with Convolutional Neural Networks, 2016; [http://arxiv.org/abs/1608.04236 arXiv:1608.04236].
- You Y., Lou Y., Liu Q., Tai Y.-W., Wang W., Ma L.: PRIN: Pointwise Rotation-Invariant Network, 2018; [http://arxiv.org/abs/1811.09361 arXiv:1811.09361].
- Mondal A.-K., Dolz J.: Few-shot 3D Multi-modal Medical Image Segmentation using Generative Adversarial Learning, 2018; [http://arxiv.org/abs/1810.12241 arXiv:1810.12241].
- Хромова И.Ю. Практическое сравнение методик прогноза трещиноватости по сейсмическим данным // Технологии сейсморазведки. – 2010. , Т.38, № 2. – С. 62–69.

The closed-form solutions of the variational problem ICP of affine transformations for fracture identification based on 3D seismic data

A.Vokhmintcev
Chelyabinsk State University,
Chelyabinsk, Russia
vav@csu.ru

Abstract. The oil resources of the Yamal-Nenets Autonomous district amount to 300 million tons, most of which are hard to recover. Nowadays the only economically viable way to engage in the development of deep reserves is the extraction of oil and gas from fractured zones. In the paper the information technology was proposed which aim at identifying and mapping fracture zones in order to forecast highly productive oil and gas-bearing areas in the Jurassic and pre-Jurassic complexes. The mathemat-

ical basis of this technology is based on machine learning and artificial intelligence methods.

Keywords: fracture zones, mapping of fractured systems, hard-to-recover reserves, seismic exploration, machine learning and artificial intelligence methods, image mapping, point cloud registration.

Оценка погрешности измерения геометрических характеристик объектов после сглаживания их границ

Куприянов М.А.

Югорский научно-исследовательский институт информационных технологий
Ханты-Мансийск, Россия

Аннотация. Проведено модельное исследование влияния сглаживания контура объекта, векторизованного по его растровому изображению, на погрешность измерения геометрических характеристик объекта. Установлено, что использование метода сглаживания Савицкого-Голея с предложенными в работе параметрами позволяет в среднем значительно уменьшить относительную погрешность измерения периметра и коэффициента формы объекта при незначительном влиянии на относительную погрешность измерения площади.

Ключевые слова: сглаживание, цифровые изображения, погрешность измерения, периметр, коэффициент формы.

ВВЕДЕНИЕ

Как было установлено в [1, 2] дискретная природа растровых изображений, которые состоят из пикселей, приводит к значительным погрешностям при измерении геометрических характеристик объекта – периметра, площади, а также производных от этих характеристик, например, коэффициентов формы. Для снижения погрешности измерения геометрических характеристик оцифрованных по растровым изображениям объектов, например, по комическим снимкам, можно использовать различные методы сглаживания. Однако, сравнение различных методов сглаживания представляет отдельную довольно трудную задачу. Поэтому нами был выбран один широко известный метод сглаживания – Савицкого-Голея, описанный в [3], суть которого заключается в построении аппроксимирующего полинома заданного порядка методом наименьших квадратов в окрестности точек заданной ширины.

В связи с этим целью данной работы является модельное исследование влияния сглаживания границ объектов, векторизованных по растровому изображению, методом Савицкого-Голея на погрешность измерения площади, периметра и коэффициента формы объекта.

В задачах моделирования и прогнозирования динамики парниковых газов из термокарстовых озёр, связанных климатическими изменениями, возникает необходимость измерения геометрических характеристик озёр (площади, извилистости береговых границ и др.) по космическим снимкам. При использовании снимков высокого и сверхвысокого пространственного разрешения извилистость береговых границ может оказывать влияние на измерение площадей озёр, поэтому в практике таких измерений используют различные методы сглаживания. Однако вопросы оценки погрешностей измерения параметров озёр возникающих при использовании процедур сглаживания до

сих пор не изучены, в связи с чем актуальность настоящей работы не вызывает сомнения.

МЕТОДИКА ИССЛЕДОВАНИЯ

Для создания наборов модельных данных была использована разработанная ранее методика формирования модельных векторных объектов (далее эталонов), подробно описанная в [2]. С использованием указанной методики было сгенерировано 10 выборочных последовательностей эталонов. В каждой выборке содержалось по 1 000 эталонов, площадь каждого из них равнялась 10 000 м². Эталоны внутри одной выборки имели одинаковую величину коэффициента формы, которая определялась по формуле [4, с. 68]:

$$F = \frac{L^2}{4\pi S}, \quad (1)$$

где F – коэффициент формы объекта (степень извилистости);

L – периметр объекта;

S – площадь объекта.

Коэффициенты формы F эталонов принимали следующие значения: 1; 1,25; 1,5; 1,75; 2; 2,5; 3; 3,5; 4; 5. Для моделирования процесса измерения по растровому изображению геометрических характеристик реального объекта, имеющего гладкую и непрерывную границу, применялась процедура растеризации, т.е. преобразования векторного типа данных в растровый. После растеризации происходило обратное конвертирование в векторный вид данных с сохранением искажений, внесённых при растеризации.

Растеризация проводилась с использованием программного обеспечения ArcGIS Desktop одним из наиболее простых методов растеризации "maximum combined area method" подробно описанным в [5]. Основная идея данного метода растеризации заключается в том, что пиксель растрового изображения заполняется (рисуеться) в том случае, если суммарная площадь контура полигона накрывает по крайней мере 50 % площади пикселя, в противном случае пиксель не заполняется.

Растеризация проводилась таким образом, чтобы после преобразования расчётное количество заполненных пикселей равнялось бы заданной величине N (1, 2, 5, 10, 20, 50, и т.д. до 100 000), которая определялась по формуле:

$$N = \frac{S}{R^2}, \quad (2)$$

где N – расчётное количество пикселей в растеризованном эталоне при фиксированном пространственном разрешении изображения;

S – площадь эталона;

R – пространственное разрешение изображения.

Для определения величины параметров сглаживания методом Савицкого-Голея из всех модельных данных была создана ограниченная выборка, в которую вошли 10 случайных эталонов (и соответствующие им растеризованные представления в количестве 160 шт.) с различной величиной F . Далее с использованием языка программирования Python, библиотеки SciPy [6], производилось сглаживание растеризованных эталонов методом Савицкого-Голея. При этом параметры метода путём полного перебора (с учётом ограничений данного метода) варьировались в следующих пределах: порядок полинома (p) от 1 до 11; размер окна фильтрации (w) от 5 до 49. После проведения сглаживания растеризованных эталонов определялись их геометрические характеристики (периметр, площадь, коэффициент формы), которые записывались в таблицу для последующего анализа.

АНАЛИЗ И ОЦЕНКА РЕЗУЛЬТАТОВ

Для подбора параметров сглаживания p , w было рассчитано среднее относительное отклонение погрешности измерения периметра после сглаживания по формуле:

$$\omega_{p,w} = \sum_{i=1}^m \left| \frac{L'_i - L_i^0}{L_i^0} \right|, \quad (3)$$

где $\omega_{p,w}$ – среднее относительное отклонение погрешности измерения периметра после сглаживания растеризованных эталонов в ограниченной выборке m с параметрами сглаживания p , w ;

m – количество объектов в выборке;

L'_i – периметр i -го растеризованного эталона после проведения сглаживания;

L_i^0 – периметр i -го эталона.

Результаты расчёта $\omega_{p,w}$ показали, что при некоторых сочетаниях параметров сглаживания Савицкого-Голея погрешность измерения периметра значительно выше, чем при других сочетаниях параметров. При этом несколько сочетаний параметров дают практически одинаково низкую величину $\omega_{p,w}$ равную около 10 %. Дальнейшее сглаживание проводилось с использованием следующих параметров: $p = 6$, $w = 9$.

Для оценки результатов сглаживания рассчитывалась относительная погрешность измерения площади, периметра и извилистости каждого объекта до сглаживания по формуле:

$$\delta_X = \left| \frac{X - X^0}{X^0} \right|, \quad (4)$$

где δ_X – относительное отклонение погрешности измерения площади, периметра или извилистости до сглаживания;

X – площадь, периметр или извилистость объекта до сглаживания;

X^0 – площадь, периметр или извилистость эталона.

Также рассчитывалась относительная погрешность измерения площади, периметра и извилистости каждого объекта после проведения сглаживания по формуле:

$$\delta'_X = \left| \frac{X' - X^0}{X^0} \right|, \quad (5)$$

где δ'_X – относительное отклонение погрешности измерения площади, периметра или извилистости после сглаживания;

X' – площадь, периметр или извилистость объекта после сглаживания;

X^0 – площадь, периметр или извилистость эталона.

Далее отдельно для каждой величины N рассчитывалось среднее абсолютное отклонение погрешности измерения площади (Δ_S), периметра (Δ_L) или извилистости (Δ_F) после процедуры сглаживания по формуле:

$$\Delta_X = \frac{1}{m} \times \sum_{i=1}^m \delta_{Xi} - \delta'_{Xi}, \quad (6)$$

где Δ_X – среднее абсолютное отклонение погрешности измерения площади, периметра или извилистости после сглаживания;

m – количество объектов;

δ_{Xi} – относительная погрешность измерения площади, периметра или извилистости до сглаживания i -го объекта;

δ'_{Xi} – относительная погрешность измерения площади, периметра или извилистости после сглаживания i -го объекта.

Данная величина показывает целесообразность проведения сглаживания границ. Если Δ_X выше нуля – значит сглаживание в среднем уменьшило относительную погрешность измерения площади, периметра или извилистости. Если Δ_X ниже нуля – значит сглаживание в среднем внесло ещё большую погрешность в измерения и проводить сглаживание было нецелесообразно.

На рис. 2 приведены результаты расчёта величин Δ_X . Как видно из рисунка, при расчётном количестве пикселей в измеряемом объекте $N \geq 10$ среднее абсолютное отклонение погрешности измерения извилистости (Δ_F) и периметра (Δ_L) становится положительным, при этом абсолютное отклонение погрешности измерения площади (Δ_S) колеблется около нуля. Это свидетельствует о положительном влиянии сглаживания методом Савицкого-Голея при выбранных параметрах ($p = 6$; $w = 9$) – об уменьшении в среднем погрешности измерения периметра и извилистости объектов после сглаживания.

ЗАКЛЮЧЕНИЕ

В результате модельного исследования показано, что применение метода Савицкого-Голея (с параметрами: порядок полинома – 6 и размером окна фильтрации – 9) для сглаживания границы объекта, растровое представление которого состоит из 10 и более пикселей, в среднем уменьшает относительную погрешность измерения периметра на величину от 10 до 20 %, а коэффициента формы (степени извилистости) на величину от 25 до 55 %. При этом относительная погрешность измерения площади в среднем незначительно увеличивается на величину до 2 %.

Таким образом, можно констатировать необходимость применения сглаживания контуров объектов, измеряемых по растровым изображениям, при проведении исследований, в основе которых лежат измерения геометрических характеристик объектов (периметра, коэффициентов формы).

БЛАГОДАРНОСТИ

Работа выполнена в рамках государственного задания Департамента информационных технологий и цифрового развития Ханты-Мансийского автономного округа – Югры на 2020 г.

ЛИТЕРАТУРА

1. Куприянов М.А. Оценка погрешности измерения периметра площадного объекта по его цифровому изображению / М.А. Куприянов // Информационные технологии и системы: Труды Седьмой Всероссийской научной конференции с международным участием. – Ханты-Мансийск: ЮНИИИТ, 2019. – С. 36-38.
2. Куприянов М.А. Анализ погрешности дистанционного измерения площади пространственных объектов на основе моделирования / М.А. Куприянов, Г.А. Кочер-

гин, Ю.М. Полищук // Вестник Югорского государственного университета. – 2018. – №. 3 (50). С. 25–34.

3. Savitzky A. Smoothing and differentiation of data by simplified least squares procedures / A. Savitzky, M.J.E. Golay // Analytical chemistry. – 1964. – vol. 36. – №. 8. – P. 1627-1639. doi:10.1021/ac60214a047

4. Берлянт А.М. Картографический метод исследования [Текст] / А.М. Берлянт. – Москва: Изд-во Моск. ун-та, 1978. – 257 с.

5. How Polygon To Raster works [сайт]: // ESRI: <http://desktop.arcgis.com/en/arcmap/10.3/tools/conversion-toolbox/how-polygon-to-raster-works.htm> (дата обращения: 27.02.2020).

6. Описание вызова функции `scipy.signal.savgol_filter` в библиотеке SciPy [сайт]: The Scipy community // https://docs.scipy.org/doc/scipy-0.16.1/reference/generated/scipy.signal.savgol_filter.html (дата обращения 28.02.2020).

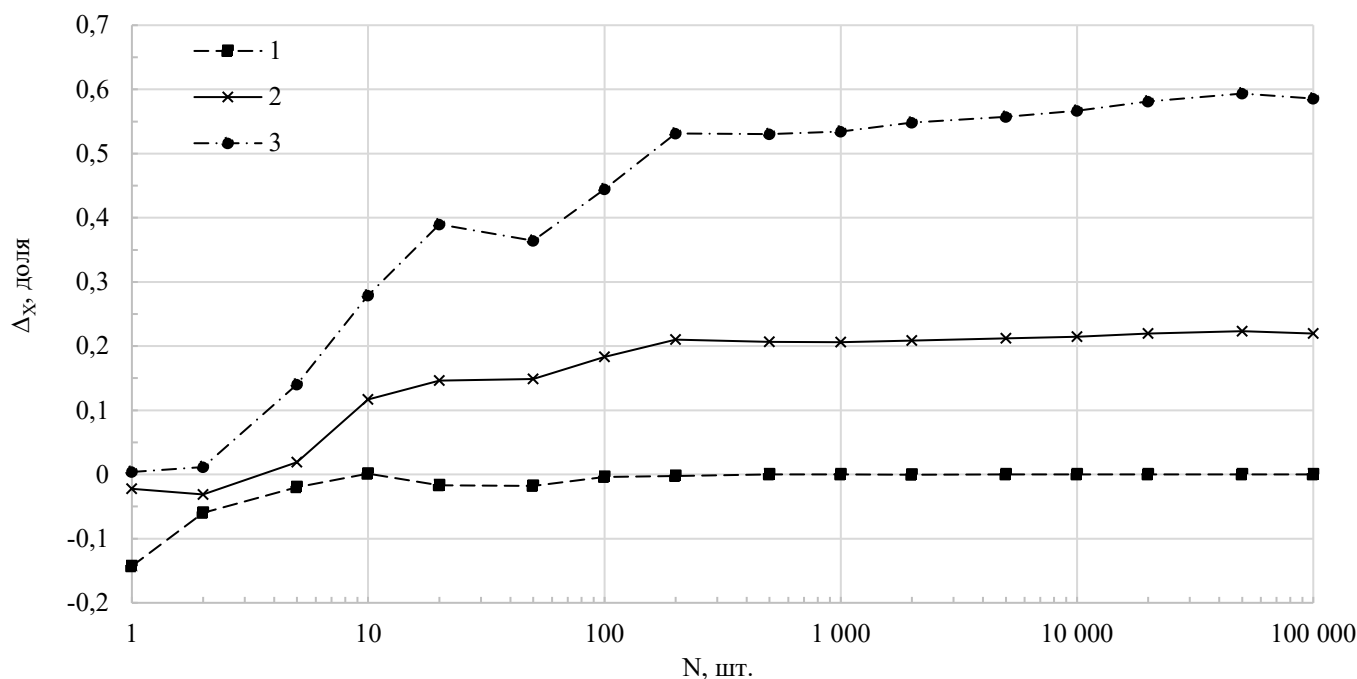


Рис. 1. Абсолютное отклонение погрешности измерения площади (1), периметра (2) или извилистости (3) после процедуры сглаживания в зависимости от рассчитанного количества пикселей в объекте (N)

Error estimation of measuring geometric characteristics of objects after smoothing their boundaries

Kupriyanov M.A.

Ugra Research Institute of Information Technologies

Khanty-Mansiysk, Russia

Abstract. A model study of the effect of smoothing the contour of an object, vectorized by its raster image, by the Savitzky-Golay method on the geometric characteristics of the object is carried out. It has been established that using the Savitzky-Golay smoothing method with the parameters proposed in the work allows, on average, to significantly reduce the relative measure-

ment error of the perimeter and the shape factor of the object with a slight effect on the relative measurement error of the area.

Keywords: smoothing, raster images, measurement error, perimeter, shape factor.

Проблемы научной археологии в ИТ-пространстве

Маслов С. Г.

ФГБОУ ВО "Удмуртский государственный университет"
Ижевск, Россия
msh.sci@mail.ru

Бельтюков А. П.

ФГБОУ ВО "Удмуртский государственный университет"
Ижевск, Россия
belt.udsu@mail.ru

Аннотация. Рассматриваются проблемы и методология системного анализа для современной информационной действительности с исторической и эволюционной точек зрения. Эта область – научная археология в ИТ-пространстве. “Бешеный” рост объема информационных ресурсов создает иллюзию всемогущества человека в управлении природой и все более полного удовлетворения его потребностей. Однако, следуя классическому пониманию информационных технологий и методологии системного анализа, важно раскрыть формы и содержательную сторону информации, информационных потоков во влиянии на жизненные процессы, раскрыть роль личности и коллектива в создании и освоении информационного пространства в исторической и возможной перспективе. В работе предлагается постановка задачи для построения системной модели информационной реальности, которая используется для выявления истоков и инициаторов информации, потоков и процессов ее трансформации и влияния на жизненную среду, смены ее активного, пассивного и скрытого состояний. Предлагаемый подход призван выявлять фундаментальные структуры информационной реальности, которые существенно сократят дублирование информации и ее “фейковые” или “раковые” структуры и образования.

Ключевые слова: археология науки, археология знаний, ИТ-пространство, жизненный процесс, дублирование, истоки, системная модель.

ВВЕДЕНИЕ

Человечеством накоплен колоссальный объем информации и знаний, которые сохраняются на различных носителях и остаются в пассивной форме, исчезают по техническим причинам, а иногда и варварски уничтожаются. В двадцатом веке мы вошли в новую информационную реальность, в которой возникла иллюзия того, что развитые средства сбора информации и совершенные методы ее обработки решат все наши проблемы. Однако, в результате происходит усиление бюрократических методов и бесконечное дублирование информации. Высокая скорость изменений жизненных ситуаций порождает дефицит времени выделяемого на осмысленную обработку информации и принятие адекватных решений, а также способствует порождению “фейковых” новостей или просто дезинформации. В этой ситуации можно констатировать, что *интервал наблюдения не соответствует интервалу проявления некоторого явления или факта*. Это приводит к необходимости перейти к точке зрения, которая раскрывает источники информации, ее контекст возникновения и эволюционные формы становления, а также связанных с этим людей. Таким образом, возникает потребность применить методы научной археологии в ИТ-пространстве,

особенности, которой мы и попытаемся раскрыть в этой статье.

К ИСТОРИИ ВОПРОСА И НЕКОТОРЫМ ТОЧКАМ ЗРЕНИЯ НА АРХЕОЛОГИЮ ЗНАНИЙ

Информационная проекция любого объекта или процесса несет в себе то, что позволяет нам формировать знания для понимания и средства для преобразования и разумного использования нашей жизненной среды. Идеальные или абстрактные образы, формируемые в сознательном и бессознательном виде, требуют, как правило, на порядок меньше энергии для преобразования и использования, чем материальные объекты. Это дает возможность рассматривать ИТ-пространство как эффективное средство накопления жизненного опыта и трансформировать его в адекватную форму, обеспечивающую разумную жизненную траекторию конкретного человека, общества и природы в целом, используя исторический и эволюционный подходы к анализу, оценке и созданию дескриптивно-конструктивной среды знаний. Однако, такая идеализация становится реальной только при включении всего арсенала научного и интуитивного мировоззрения, а также теорий, методов, средств и технологий познания и преобразования мира. Для этого целесообразно восстановить историю возникновения и эволюцию идей в лицах, контекстах, затраченных ресурсах с выявлением вреда или пользы от их реализации и применения, с конкретизацией пространства существования и с выявлением его границ.

Традиционные формы “складирования” информации и знаний закрепляют их пассивную форму, а при гигантских объемах провоцируют на использование неэффективных и неадекватных методов поиска и преобразования, надеясь на совершенство носителей информации и высокопроизводительные вычисления, а главное, они становятся уничтожителем реального и самого ценного ресурса человека – его жизненного времени.

Естественно возникают вопросы: чем же, здесь, помогут методы научной археологии знаний и что это такое?

Сегодня научная археология знаний проявляется в таких понятиях как “археология знаний” [1,2], “археология науки (Archaeology of Science)” [3], “историческая информатика” [4], “цифровая история (Digital History)” [5], “добыча и раскопки данных (Data Mining)”.

Уже в 60-е годы Мишель Фуко ввел термин “археология знаний” [1]. Под археологией знаний он понимал то, что опирается на бессознательное и высказывания, на устойчивые механизмы общественного функционирования и на связанные ансамбли синхронных закономерностей, заменяя эпистемологическую триаду [2]: (*сознание – по-*

знание – наука) на (дискурсивная практика – знание – наука).

Ключевым здесь являются коммуникации и коллективная память живых носителей знаний.

Под исторической информатикой (или цифровой историей) в основном понимают применение методов искусственного интеллекта, цифровых медиа и информационно-технологических методов и инструментов в исторических исследованиях, то есть использование новых цифровых форм документов и методов их сбора, анализа и обработки. Во многом этому способствуют и методы анализа больших данных.

Археология науки делает акцент на взаимодействии людей с артефактами в деятельности, используя для этого:

- описания, категории и классификации, эмпирические обобщения, экспериментальные законы, рецепты, теории и модели;
- исследования истории жизни, процессов, деятельности, взаимодействия как простых, так и неординарных людей в профессиональной и обычной жизни.

Такие исследования приводят к лучшему или новому пониманию видов знаний, обобщению процессов открытия и к созданию новых направлений в науке, а также выявлению нового и полезного знания.

Во всех перечисленных направлениях, при рассмотрении объекта исследования, основное внимание уделяется объективной стороне дела, фактам и выводам, а также инструментально-технологическим средствам. Человеческий фактор тоже учитывается, но в ограниченной форме, которая принимает во внимание в основном психологические и потребительские особенности.

ЦЕЛИ И МЕТОД НАУЧНОЙ АРХЕОЛОГИИ В ИТ-ПРОСТРАНСТВЕ

Можно выделить ряд важных аспектов исследований, которые требуют дальнейшего развития:

- уточнение ролей, функций и ответственности индивида и коллектива в исследованиях с учетом пространственно-временного взаимодействия активных, пассивных и нейтральных, отрицательных и положительных, полезных и вредных проявлений индивидуальной и совместной деятельности;
- динамика и прояснение эволюционных особенностей объектов, субъектов и процессов;
- определение границ для очерчивания среды активности, полезности или вредного проявления знаний и деятельности; это проблемы безопасного использования знаний;
- создание методов выявления ценных идей прошлых исследований и переноса их в новую теоретико-технологическую среду и практику.

Системная модель научной археологии в ИТ-пространстве, включает:

- активную терминологическую систему (плоский нижний уровень) [6];
- конструктивную онтологию (вышележащий слой, связующие терминологические слои), конструктивные уровни терминологии [7];
- дескриптивно-конструктивный процесс взаимодействия людей, выполняющих различные роли (исследователь, конструктор, организатор, наставник) [8-10];

- жизненные оценки человека, его мотивационные и мировоззренческие концепции;
- временные и причинные модальности (как было, как должно было быть, как стало и почему) в связи с терминологической системой;
- образные лингвистические представления с прямым и обратным преобразованием и взаимодействием.

Такое построение системы знаний позволит связывать факты и субъектов в пространственно-временных отношениях и эффективно и своевременно находить адекватные решения возникающих проблем.

Приведём некоторые интерпретации используемой модели. В частности, можно разобраться с вопросом приоритетности (или конкуренции) и сотрудничества в исследованиях и определить, какой вред и какую пользу они могут принести в различных ситуациях. Из истории науки ниже приведём четыре примера. Казалось бы тождественные результаты были получены в следующих направлениях:

- спор о приоритете открытия дифференциального и интегрального исчисления между Исааком Ньютоном (1665–1666) и Готфридом Вильгельмом Лейбницем (1675), который так и остался неразрешенным [11];
- спор о приоритете в изобретении метода наименьших квадратов между Карлом Фридрихом Гауссом (1795) и Адриеном Мари Лежандром (1806), который также остался без удовлетворительного разрешения [12];
- смешанные вычисления А.П. Ершов (1977) и проекции Ё. Футамуры (1971) – “какое разочарование должен был испытать Андрей Петрович, когда отыскал статью японского ученого Ёсихико Футамуры, опубликовавшего эти соотношения еще в 1971 году. Примером профессиональной этики является тот факт, что Андрей Петрович назвал независимо открытые им соотношения проекциями Футамуры, несмотря на то, что последней проекции в той статье не было.” [13];
- пример, связанный с одним из авторов настоящей статьи – разрешимость универсальной теории натуральных чисел со сложением и делимостью А. П. Бельтюков – Л. Липшиц [14-17] (результаты были получены одновременно – в сентябре 1975, что обнаружилось почти сразу же через научных руководителей авторов – Юрия Матиясевича и Джулии Робинсон соответственно – и привело к переписке между авторами, при которой выявился некоторый недочёт в первоначальном варианте работы Леонарда Липшица, который был им успешно устранён, работы обоих авторов были опубликованы, так как подходы имели отличия; это можно считать примером сотрудничества в подобном спорном случае).

Сопоставление результатов разных авторов, работающих над одной темой позволяет глубже проникнуть в исследуемую проблему и получить адекватные оценки результатов исследований, а в дальнейшем усовершенствовать исследовательский процесс. Само первенство даёт основания для ресурсного стимулирования авторов, а со-

трудничество – получить более обширную область распространения результатов исследований.

Исторические исследования позволяют более адекватно определять области эффективного использования научных моделей, а с учётом скорости и объёмом накапливаемой информации это становится востребованным.

Для установления источников, ресурсоемкости и условий получения новых результатов требуется изучать различные отношения между исследователями в рамках среды и целей исследования:

- *эмоциональные и стимулирующие связи* – дружба (друзья), любовь (любовные связи), покровители, наставники, сочувствующие;
- *организационные и управляющие связи* – начальники, подчиненные; личности, оказавшие влияние; личности, подвергшиеся влиянию; личности, ценившие субъекта; личности, ценимые субъектом;
- *политические, финансовые, деловые связи* – инвесторы, стейкхолдеры, благотворители;
- *творческие и конструктивные связи* – соперники (оппоненты), союзники (соратники), единомышленники (личности близкие по духу и мировоззрению, сторонники), коллеги (художники, космонавты, физики, монархи,...), сподвижники;
- *конфликтные или разрушающие связи* – противники (враги, враждебно настроенные личности);
- *пространственно-временные и идейные связи* – предшественники, современники, последователи (наследники, ученики).

Богатый материал по этим вопросам можно найти в работах [17-21].

Научный фонд современных журналов содержит массу полезного материала, который необходимо активно перерабатывать. Для быстрой обработки сейчас выделяют актуальную терминологию, терминологические аннотации, определяя тренды развития. К сожалению, эти тренды часто срезают действительно *инновационные ростки* системы знаний. В качестве примера можно привести исследования исторических тенденций эволюции тематики публикаций за последние 50 лет в Международном журнале исследований человека и компьютера (IJHCS – International Journal of Human-Computer Studies и конференции оп изучению взаимодействия человека и компьютера (HCI – Human-Computer Interaction). Были получены некоторые интересные результаты [22]:

- преобладание междисциплинарного разнообразия ссылок в журнальных статьях в сравнении со статьями для конференций;
- более равномерно распределение цитат по годам в журналах в противопоставлении в основном на недавние статьи в статьях на конференциях;
- обнаружено несколько тенденций исследований, появившихся в последние десять лет, которые включают следующие научные направления: мобильные вычисления, машинное обучение, компьютерная безопасность, конфиденциальность данных, робототехника, компьютерные игры, системы для пользователей с ограниченными возможностями, виртуальная реальность и анализ социальных сетей.

ЗАКЛЮЧЕНИЕ

В настоящей работе рассмотрены проблемы и постановка задачи создания системной модели для автоматизации и развития научной археологии в пространстве информационных технологий и систем. В частности, к наиболее важным можно отнести следующие проблемы:

- разработка структуры и формы накопления многоаспектных знаний в виде фактов, моделей, полисистемных представлений знаний, альтернативных решений,
- обычного человеческого мышления, даже оснащённого современными компьютерными технологиями, уже становится недостаточно для своевременного получения качественных знаний и их эффективного использования – мы приходим к необходимости использования эргатических (человеко-машинных) систем, способных обрабатывать информацию в синхронном и асинхронном виде (некоторые считают асинхронный режим аналогом “сна”);
- выявление границ адекватного применения знаний, которые обеспечивают их безопасное, надежное, полезное и своевременное применение;
- сохранение потенциала и целостности системы знаний в эволюционных процессах, которые учитывают как тенденции развития знаний, так и скрытые потенциалы уникальных, но неочевидных знаний.

Исторические знания неразрывно связаны с реальной жизнью, чрезвычайно ценны для целостного и общего понимания предмета науки.

ЛИТЕРАТУРА

1. Фуко Мишель Археология знания – СПб.: ИЦ «Гуманитарная Академия»; Университетская книга, 2004. – 416 с.
2. Автономова Н.С. Археология знания // Новая философская энциклопедия – URL: https://iphlib.ru/library/collection/newphilenc/document/HAS_H01f70f90c475ac376b0f0df6 (дата обращения 30.01.2020).
3. Schiffer, Michael Brian. The Archaeology of Science. Studying the Creation of Useful Knowledge. The Archaeology of Science // Series: Manuals in Archaeological Method, Theory and Technique, Springer, Heidelberg, 2013, Vol. 9
4. Володин А.Ю. Историческая информатика // Большая российская энциклопедия - электронная версия - URL: https://bigenc.ru/world_history/text/3271466 (дата обращения 30.01.2020).
5. Бородкин Л.И. Digital history: применение цифровых медиа в сохранении историко-культурного наследия? // Историческая информатика. 2012. Т. 1. № 1. – С. 14–21.
6. Маслов С. Г. Активная расслоенная терминологическая система / С.Г. Маслов, А.П. Бельтюков // Устойчивое инновационное развитие: проектирование и управление. – Дубна, том 7, 3 (12), 2011, – 103–113 с.
7. Beltukov, Anatolii. Constructive Ontologies // 21st International Scientific Workshop on Computer Science and Information Technologies (CSIT 2019) – Atlantis Highlights in Computer Sciences, vol. 3, pp.83-88.
8. Кузнецов О.Л. Синтез междисциплинарных знаний и устойчивое развитие в системе Пространство – Время / О.Л. Кузнецов, П.Г. Кузнецов, Б.Е. Большаков. – М.: Из-

дательство Русского Космического Общества, 2018. – 276с.

9. Маслов С.Г. Системный анализ в формировании физико-антропо-технических систем / С.Г. Маслов, А.П. Бельтюков // Информационные технологии и системы: тр. Седьмой Всероссийской научной конференции с международным участием (Ханты-Мансийск, 12–16 марта 2019 г.) – Ханты-Мансийск, 2019. – С.16-19.

10. Маслов С.Г. 2010. Системный подход к формированию среды программирования: Интеллектуальные системы управления / под ред. С.Н. Васильева – М.: Машиностроение, С. 41–47.

11. Спор Ньютона и Лейбница о приоритете // Материал из Википедии — свободной энциклопедии – URL: https://ru.wikipedia.org/w/index.php?title=Спор_Ньютона_и_Лейбница_о_приоритете&oldid=104180367. (дата обращения 30.01.2020).

12. Степанов О.А. Фильтр Калмана: история и современность // Гирокоспия и навигация, № 2 (69), 2010. – С.107-121.

13. Бульонков М.А. Смешанные вычисления в Новосибирске // Виртуальный компьютерный музей – URL: https://www.computer-museum.ru/books/n_ershov/2_ershov_sv.htm (дата обращения 30.01.2020).

14. Бельтюков А.П. Разрешимость универсальной теории натуральных чисел со сложением и делимостью // Записки научных семинаров ЛОМИ. 1975. Т. 60. С. 15–28.

15. Lipshitz L. The Diophantine problem for addition and divisibility // Transactions of the American Mathematical Society. 1976. Vol. 235. P. 271–283.

16. Lipshitz L. Some remarks on the Diophantine problem for addition and divisibility // Bull. Soc. Math. Belg. Ser. B. 1981. Vol. 33, no. 1. P. 41–52.

17. 100 человек, которые изменили ход истории // Ежедневное издание - Москва: 000 "Де Агостини", 2008-2009. Выпуски №1-100.

18. Архив академика А.П. Ершова при Институте систем информатики им. А.П. Ершова Сибирского отделения РАН – URL: <http://ershov.iis.nsk.su/> (дата обращения 30.01.2020).

19. История информатики и кибернетики в Санкт-Петербурге (Ленинграде). Вып.1. Яркие фрагменты истории // Под общ. ред. чл.-кор. РАН Р.М. Юсупова; составитель М.А. Вус; Ин-т информатики и автоматизации РАН. – СПб.: Наука, 2008. – 356 с.

20. История информатики и кибернетики в Санкт-Петербурге (Ленинграде). Выпуск 2. // Под общ. ред. чл.-кор. РАН Р. М. Юсупова. Составитель М. А. Вус. Санкт-Петербургский институт информатики и автоматизации РАН. – СПб.: Наука, Изд-во ООО "Анатолия", 2010. – 152с.

21. Виртуальный компьютерный музей – URL: <https://computer-museum.ru/aboutmus/> (дата обращения 30.01.2020).

22. Andrea Mannocci, Francesco Osborne, Enrico Motta The evolution of IJHCS and CHI: A quantitative analysis // International Journal of Human-Computer Studies 131 (2019) 23–40.

Problems of scientific archaeology in the IT space

S. G. Maslov

Department of Theoretical Foundations of
Computer Science
Udmurt State University, UdsU
Izhevsk, Russia
msh.sci@mail.ru

A. P. Beltiukov

Department of Theoretical Foundations of
Computer Science
Udmurt State University, UdsU
Izhevsk, Russia
belt.udsu@mail.ru

Abstract. The problems and methodology of system analysis for modern informational reality from a historical and evolutionary point of view are considered. It is scientific archeology in the IT space. The "frantic" growth in the volume of information resources creates the illusion of the omnipotence of man in the management of nature and the more complete satisfaction of his needs. However, following the classical understanding of information technology and the system analysis methodology, it is important to disclose the forms and the content side of information, information flows in influencing life processes, to reveal the role of the individual and the collective in creating and mastering the information space in a historical and possible perspective. The work proposes proposes a statement of the problem for

constructing a system model of information reality, which is used to identify the sources and initiators of information, flows and processes of its transformation and impact on the living environment, change of its active, passive and hidden states. The proposed approach is designed to identify the fundamental structures of informational reality, which will significantly reduce the duplication of information and its "fake" or "cancerous" structures and formations.

Keywords: archaeology of science, archeology of knowledge , IT space, life process, duplication, origins, system model.

Разработка системы сбора и анализа персонифицированных данных в онтологии

Гартман К.А.

Тюменский индустриальный университет

Тюмень, Российская Федерация

k1rusja@mail.ru

Аннотация. В статье рассмотрено появление онтологических технологий и их использование в медицинских информационных системах. Особое внимание автором уделено изучению медицинских баз данных связанных с персонифицированным подходом, которые нацелены на обработку непрерывно растущего объема различной информации. Проведённые исследования открывают возможность разработки системы автоматизированной (с участием человека) обработки данных с использованием персонифицированного подхода при проведении исследований сердечно-сосудистых заболеваний, в целях количественного и качественного анализа эффективности медицинской деятельности врача определенного направления, что приведет не только к изменениям технологии работы медицинских учреждений, но и документооборота.

Ключевые слова: *онтология, персонифицированные данные, сердечно-сосудистые заболевания, медицинские базы данных, нормативно-техническая документация, программное обеспечение.*

ВВЕДЕНИЕ

Онтология активно используется такими науками как гуманитарными, так и естественными, и техническими. Всякий раз, когда в любой науке возникает необходимость выстроить терминологический аппарат, она обращается к онтологии. Именно онтология в ограниченном смысле определяет понятия и термины, описывает их и организует таксономию. За правила вывода одного термина из другого тоже отвечает онтология. Это используется не только в науке, но и в компьютерных играх при создании виртуальной реальности. Для того чтобы продумать и создать бытие виртуального мира, без онтологии не обойтись. Многочисленные исследования в нашей стране и зарубежных исследователей [1–5] показывают, что сердечно-сосудистые заболевания (ССЗ) остаются проблемой № 1 в большинстве стран современного мира. В частности, каждый год от ССЗ в Европе умирают 4 миллиона человек, из которых 1 миллион приходится на Россию. Одним из важных показателей в структуре смертности населения от всех причин является доля умерших от ССЗ, которая в Российской Федерации составляла 55,9%, тогда как в Европе - 47%.

В трудоспособном возрасте, на который ложится основное бремя выработки внутреннего валового продукта, от величины которого зависит благосостояние нации, болезни системы кровообращения составляют более 1/3 всех случаев смерти. Причем уровень смертности мужчин превышает уровень смертности женщин в целом от БСК в 4,7 раза, от ишемической болезни сердца (ИБС) в 7,2 раза, от

инфаркта миокарда - в 9,1 раза и от cerebrovasкулярных болезней - в 3,4 раза [6]. Несмотря на то что в развитых странах, а в последние годы и в России, наблюдается снижение смертности, ситуация в нашей стране остается чрезвычайно серьезной, что подчеркивается в многочисленных отечественных и зарубежных публикациях [7–9].

МЕДИЦИНСКИЕ БАЗЫ ДАННЫХ

Разработка структуры данных, используемых в работе медицинских информационных систем, важный и трудоемкий процесс. Для поддержания фундаментальных основ работоспособности медицинской организации в лечебно-диагностической деятельности, необходимо иметь возможность оперативно внести изменения как в структуру связей базы данных, так и в табличные составляющие данных медицинской информационной системы.

Для обработки непрерывно растущего объема различной информации, связанной с персонифицированным подходом, используются базы данных. Для более эффективной обработки информации в медицинских базах данных проводится мониторинг по всем передвижениям пациента, начиная от поступления, заканчивая его реабилитацией. Кроме того для медицинских работников, проще работать с структурированной информацией. Поэтому центральное место в медицинских информационных системах принадлежит базам данных.

Базы данных - совокупность данных, хранимых согласно структуре данных, манипулирование которыми выполняют в соответствии с правилами моделирования данных. В ходе анализа учебно-методической литературы [7–13] к современным базам данных и к СУБД, на которых они строятся, выявлены следующие основные требования:

Высокое быстродействие (малое время отклика на запрос). Время отклика - промежуток времени от момента запроса к БД до фактического получения данных;

- Простота обновления данных;
- Независимость данных - возможность изменения логической и физической структуры БД без изменения представлений пользователей;
- Совместное использование данных многими пользователями;
- Безопасность данных - защита данных от преднамеренного или непреднамеренного нарушения секретности, искажения или разрушения;
- Стандартизация построения и эксплуатации БД (фактически СУБД);
- Адекватность отображения данных соответствующей предметной области;
- Простой интерфейс пользователя.

К.А. Гартман

Важнейшими являются первые два противоречивых требования: повышение быстродействия требует упрощения структуры БД, что, в свою очередь, затрудняет процедуру обновления данных, увеличивает их избыточность.

Литературный анализ показывает, что в последние годы, в связи с резким увеличением обрабатываемых объемов данных и их сложности, началось широкое развитие технологии хранения данных в нереляционных БД (НБД), которые предоставляют более высокую гибкость при работе в сложных средах, а также простоту модификации на самом нижнем уровне - уровне хранения данных по сравнению с традиционными реляционными базами данных (РБД) [14-16].

Так же в ходе информационного анализа выявлено широкое применение баз данных в медицинских исследованиях. Например, авторы Борискин Б.Б., Кулакова И.М. [17], в своей статье «Выбор программного обеспечения создания базы данных медицинской информационной системы» рассматривают современные методы и инструменты проектирования и формирования базы данных медицинской информационной системы.

В статье «Обработка медицинских изображений в графической базе данных и идентификация пациентов» Горюновой В.В., Горюновой Т.И., Кухтевич И.И., Поляковой Т.Т. [18], рассмотрены вопросы обработки и анализа медицинских изображений. Представлен метод изменения контрастности изображения, который анализирует гистограмму изображения. Суть метода состоит в преобразовании яркостей исходного изображения таким образом, чтобы гистограмма яркостей приобрела нужную форму.

В статье «Принципы работы с графической базой данных медицинских изображений» Горюнова В.В., Горюнова Т.И., Шубина И.И. [19] описаны основные функции выбора баз данных медицинских изображений GDB (Graphical Data Base). Рассмотрены функции администрирования пользователей и их ролей, определения всех настроек GDB и место физического размещения по сетевым ресурсам. Перечислены виды администрирования, выбираемые в соответствующих таблицах.

В статье Богданова Н.В., Кузьмина Е.Ю. «Особенности проектирования базы данных программного комплекса мониторинга медицинской статистики» [20], рассмотрены создание программного продукта максимальной комфортности для специалистов медицинской статистики с использованием средств «безбумажных» технологий при экономии времени.

ПЕРСониФИЦИРОВАННЫЙ ПОДХОД

Сердечно - сосудистые заболевания - ежегодно причина смерти около 17 миллионов человек во всем мире, что обуславливает их высокую социальную значимость и необходимость разработки новых подходов к лечению и ранней диагностике [21]. Лица с симптомами заболевания ИБС и АГ образуют группу наивысшего риска развития осложнений и смерти от ССЗ. Сердечно - сосудистая патология включает гипертоническую болезнь (ГБ), ишемическую болезнь сердца (ИБС), инфаркт миокарда (ИМ) и ряд других заболеваний. ГБ представляет собой полиэтиологическое заболевание и служит причиной высокой смертности среди лиц среднего и пожилого возраста. ГБ,

несмотря на наличие множества антигипертензивных средств, поддается успешному лечению только в 27% случаев. Сложность молекулярно-генетического анализа ГБ заключается в большом количестве потенциальных генов-кандидатов, способных участвовать в формировании предрасположенности к этому заболеванию. Известно более 150 таких генов, включающих в себя компоненты ренин-ангиотензиновой системы, метаболизма липидов, продукты, обеспечивающие поддержание сосудистого тонуса и структуры сосудов. В настоящее время в рамках ПМ существует два подхода к лечению и диагностике ГБ:

- поиск оптимального сочетания антигипертензивных средств;
- изучение популяционных особенностей полиморфизмов генов, участвующих в развитии ГБ. Применение молекулярной диагностики позволяет увеличить эффективность проводимого лечения ГБ.

В последние годы ученые многих стран активно обсуждают концепцию персонифицированной или предикативной (предупредительной) медицины [22]. По прогнозам, она способна существенно повысить качество лечения. Раннее выявление рисков возникновения того или иного заболевания у конкретного пациента, оптимальная для него врачебная помощь уже в ближайшей перспективе могут стать реальностью.

Понятие «персонифицированная медицина» (ПМ, другое название предикативная медицина) впервые было введено в научный лексикон в 1998 года для обозначения инновационного подхода к диагностике, профилактике и лечению социально значимых заболеваний [23]. Базовой основой для развития ПМ является фармакология, современной важнейшей задачей которой стало создание эффективных лекарственных препаратов таргетного (целенаправленного) действия.

Если ранее считалось, что лекарственные препараты могут быть универсальными при лечении какого-либо заболевания, то современные биомедицинские исследования свидетельствуют о том, что эффективность применения того или иного лекарственного средства зависит от генетических и эпигенетических факторов, среди которых ключевое место занимают сигнальные молекулы, осуществляющие широкие межклеточные взаимодействия; нарушения последних лежат в основе патогенеза многих заболеваний.

Таким образом, исследование индивидуальных особенностей экспрессии сигнальных молекул в органах и тканях является перспективным и многообещающим направлением для разработки эффективных методов профилактики, диагностики и лечения различной патологии [24, 25, 26]. ПМ включает 6 основных разделов: молекулярную диагностику, интеграцию диагностики и лечения, мониторинг лечения, фармакогеномику, фармакогенетику и фармакопротеомику [27]. Верификация и изучение экспрессии сигнальных молекул (биомаркеров) у конкретного пациента при той или иной патологии можно рассматривать как основной методологический подход к своевременной диагностике заболевания, выбору адекватной схемы лечения и оценке эффективности проводимой терапии.

ЗАКЛЮЧЕНИЕ

В мировом информационно-аналитическом пространстве на базе современных информационных технологий созданы и функционируют различные информационные системы, связанные с деятельностью медицинских учреждений. Для улучшения работы медицинских учреждений и, как следствие, повышения эффективности развития медицины в масштабах государства необходимо перманентно анализировать информацию о результатах деятельности врачей различного профиля и коллективов.

Основными результатами деятельности медицинских организаций, входящих в медицинское сообщество, как правило, считаются публикации сотрудников, результаты патентных исследований, участие в конференциях, руководство курсовыми, дипломными и диссертационными работами, чтение лекций и ряд других. При этом, как показывает практика, далеко не все результаты такой деятельности представлены в открытом доступе в Интернет. Зачастую единственным источником подобной информации могут служить лишь годовые отчеты сотрудников медицинских организаций, представленные с той или иной степенью подробности.

Разработка и внедрение информационных систем в медицинском учреждении должна рассматриваться не как перевод традиционного медицинского учета и отчетности на компьютерную платформу (т.е. не на автоматизацию), а как разработка и внедрение системы информационного обеспечения деятельности медицинского учреждения и, в первую очередь - лечебной и управленческой деятельности. При этом реорганизация бизнес-процессов и пересмотр традиционного документооборота должны рассматриваться и восприниматься как неперенные элементы процесса информатизации.

ЛИТЕРАТУРА

1. Чудинов И.Л. Базы данных: учеб. - метод. пособие / И.Л. Чудинов, В.В. Осипова; Томский политехнический университет, 2012. - 140 с.
2. Артёмов В.Ю. Онтология компетентностного подхода в системе высшего профессионального образования / В.Ю. Артёмов. // *univer-sum: психология и образование: электрон. научн. журн.* - 2014.- №12(11).
3. Nichols M. European Cardiovascular Disease Statistics (2012): European Heart Network, Brussels and European Society of Cardiology/ M. Nichols, N. Townsend, P. Scarborough, R. Luengo-Fernandez, J. Leal, A. Gray, M. Rayner - Direct text // *Sophia: Antipolis.*-2012.
4. Медико-демографические показатели Российской Федерации в 2011 г. 2013: Стат. справочник: [сайт]. URL: <http://www.mednet.ru>. Текст электронный.
5. Наймушина А.Г. Системный анализ и обработка экспертной информации для определения степени значимости факторов риска / А.Г. Наймушина, И.А. Чекардовская, Л.Н. Бакановская // *Информационные технологии и системы: Труды Седьмой Всероссийской научной конф. с междунар. учас.* - Ханты-Мансийск, 2019.- С:20-25.
6. Наймушина А.Г. Системный анализ в оценке состояния здоровья мужчин активного работоспособного возраста / А.Г. Наймушина, И.А. Чекардовская, Л.Н. Бакановская // *Водные ресурсы - основа устойчивого развития поселений Сибири и Арктики в XXI веке: Сборник докладов XXI Междунар. Науч.- практ. конф.* -Тюмень, 2019 С. 450-454.
7. Соловьев В.С. Стратификация факторов риска стресс-индуцированных состояний у жителей города Тюмени / В.С. Соловьев, А.Г. Наймушина, С.В. Соловьева // *Вестник Тюменского государственного университета. Экология и природопользование.*- 2013. - № 6. - С. 96 -103.
8. The Demographic Yearbook of Russia. 2010. Federal State Statistical Service (Rosstat). Moscow 2010. (Демографический ежегодник России, 2010).
9. Шальнова С.А. Тенденции смертности в России в начале XXI века (по данным официальной статистики) / С.А. Шальнова, А.Д. Деев // *Кардиоваск тер. и проф.* - 2011. №10. №6. С. 5-10.
10. Оганов Р.Г. Демографические тенденции в Российской Федерации: вклад болезней системы кровообращения / Р.Г. Оганов, Г.Я. Масленникова // *Кардиоваск тер. и проф.*-2012; №11,2.С. 5-10.
11. Nichols M. Cardiovascular disease in Europe: epidemiological update./ Nichols M., Townsend N., Scarborough P, Rayner M. // *Eur Heart J.*- 2013.- №34. - P.3028-3034.
12. Илюшечкин В. М. Основы использования и проектирования баз данных / В. М. Илюшечкин. - М.: Юрайт, - 2015. С.-516.
13. Голицына, О. Л. Базы данных/О.Л. Голицына, Н.В. Максимов, И.И.Попов. - М.: Форум, - 2015.- С. 400.
14. L.N. Bakanovskaya Application of artificial neural networks in the heart electrical axis position conclusion modeling / Bakanovskaya L.N. // *IOP conference series: materials science and engineering : electronic edition. National Research Tomsk Polytechnic University. VII International scientific practical conference "Innovative technologies in engineering"* // *Yurga.* - 2016. - P. 12-10.
15. Илюшечкин, В. М. Основы использования и проектирования баз данных / В.М. Илюшечкин. М.: Юрайт, 2013. - С. 24.
16. Илюшечкин, В. М. Основы использования и проектирования баз данных. Учебник / В.М. Илюшечкин, - М.: Юрайт, 2014. - С. 214.
17. Малыхина М.П. Интеллектуальная система для проектирования баз данных эффективной структуры / М.П. Малыхина, М.В. Буянов // *Научный журнал КубГАУ*, - 2013.- № 89 (05). С. 1-3.
18. Борискин Б.Б. Выбор программного обеспечения создания базы данных медицинской информационной системы ангарский государственный технический университет / Б.Б. Борискин, И.М. Кулакова // *Современные технологии и научно-технический прогресс*, - 2019. С. 128-129.
19. Горюнова В.В. Обработка медицинских изображений в графической базе данных и идентификация пациентов / В.В. Горюнова, Т.И. Горюнова, И.И. Кухтевич, Т.Т. Полякова // *ФГОУ ВО «Пензенский государственный технологический университет»*, науч. ст. - 2017. - С. 1390-1393.
20. Горюнова В.В. Принципы работы с графической базой данных медицинских изображений/ В.В. Горюнова,

Т.И. Горюнова, И.И. Шубин // ФГОУ ВО «Пензенский государственный технологический университет» науч. стат. - 2017. - С. 1397-1400.

21. Богданова Н.В. Особенности проектирования базы данных программного комплекса мониторинга медицинской статистики / Н.В. Богданова, Е.Ю. Кузьмина // Рязанский государственный университет имени С.А. Есенина, науч. стат. - 2017. - С. 30-36.

22. Белоусов С.С. Роль молекулярно - иммунологических нарушений в патогенезе инфаркта миокарда и их модуляции в клинической практике / С.С. Белоусов, И.И. Новиков // Медицинский альманах. - 2010. № 2. - С. 66.

23. Наймушина А.Г. Персонифицированная медицина. Иллюзии и реальность / А.Г. Наймушина // Новые информационные технологии в нефтегазовой отрасли и об-

разовании: материалы VII Междунар. науч.-технич. конф. - Тюмень, - 2017. - С. 138-144.

24. Kearney P. / Global burden of hypertension: analysis of worldwide data/ P. Kearney, M. Whelton, K. Reynolds // Lancet. - 2005. - № 365 (9455). - P. 217, 23.

25. Ong F. Personalized edicine and pharmacogenetic biomarkers: progress in molecular oncology testing/ F.Ong, J. Das K., Wang // Expert. Rev. Mol. Diagn. - 2012; - № 12 (6). - P.593-602.

26. Jain K. Personalized medicine / K. Jain // Curr. Opin. Mol. Ther. - 2002. - № 4 (6). - P. 48-58.

27. Jain K. Role of biological therapies in the development of personalized medicine/ K. Jain // Expert. Opin. Biol. Ther. - 2012.- № 12 (1). P.1-5.

Development of a collection and analysis system personalized data in the ontology

Gartman Kira

Tyumen industrial University
Tyumen, Russian Federation
k1rusja@mail.ru

Abstract. The article discusses the emergence of ontological technologies and their use in medical information systems. The author pays special attention to the study of medical databases related to the personalized approach, which are aimed at processing a continuously growing volume of various information. The conducted research opens up the possibility of developing a

system of automated (with human participation) data processing using a personalized approach when conducting research on cardiovascular diseases, in order to quantitatively and qualitatively analyze the effectiveness of medical activity of a doctor of a certain direction, which will lead not only to changes in the technology of medical institutions, but also to document circulation.

Keywords: ontology, personalized data, cardiovascular diseases, medical databases, regulatory and technical documentation, software..

К вопросу разработки системы очистки отработанного масла на основе интеллектуальных методов управления

Горбачева О.М.
ФГБОУ ВО Оренбургский государственный
университет
г. Оренбург, Россия
Ol.gorba4ewa2017@yandex.ru

Боровский А.С.
ФГБОУ ВО Оренбургский государственный
университет
г. Оренбург, Россия
borovski@mail.ru

Аннотация. Статья посвящена актуальной на сегодняшний день проблеме применения интеллектуальных методов в системе управления технологическим процессом очистки отработанных масел. Используя методы системного анализа, с помощью диаграмм IDFO, были представлены функциональные модели автоматизированных систем очистки отработанного масла. Проведя анализ достоинств и недостатков представленных систем, была предложена оптимальная автоматизированная система очистки отработанного масла по заданным параметрам.

Ключевые слова: отработанное масло, автоматизированная система управления, центрифугирование.

ВВЕДЕНИЕ

Согласно национальной стратегии развития искусственного интеллекта на период до 2030 г., утвержденной Президентом РФ, искусственный интеллект необходимо использовать для автоматизации рутинных производственных операций и на его основе нужно создавать автономное интеллектуальное оборудование [1].

В связи с этим, проблема разработки оптимальной автоматизированной системы управления процессом очистки отработанного масла с использованием интеллектуальных методов управления является актуальной.

Для определения оптимальных параметров автоматизации системы очистки отработанного масла и пред

ложения ее дальнейшего усовершенствования (внедрение интеллектуальных методов управления), необходимо провести анализ существующих автоматизированных систем очистки отработанного масла.

Поиску способов реализации поставленных задач и посвящена данная статья.

ОСНОВНАЯ ЧАСТЬ

Исходя из перспектив контроля и управления надежностью систем очистки отработанного масла в [2] предложен алгоритм (рис.1) подконтрольного использования силовых агрегатов с замкнутой системой смазки, включающий комплекс устройств и аппаратов, обеспечивающих долгосрочное и рациональное использование жидких смазочных масел за счет их автоматизированного контроля, комбинированной очистки, физико-химического модифицирования и утилизации. Эти процедуры могут применяться как в сочетании, так и в отдельности и осуществляться в стационарных условиях предприятия и непосредственно на объекте. Для практической реализации результатов исследований автор исходил из той предпосылки, что достаточно легко реализуемым и относительно малозатратным способом восстановления эксплуатационных свойств моторного масла может служить их глубокая комплексная очистка в гравитационном и центробежном поле, а также фильтрация объемными маслоочистителями.

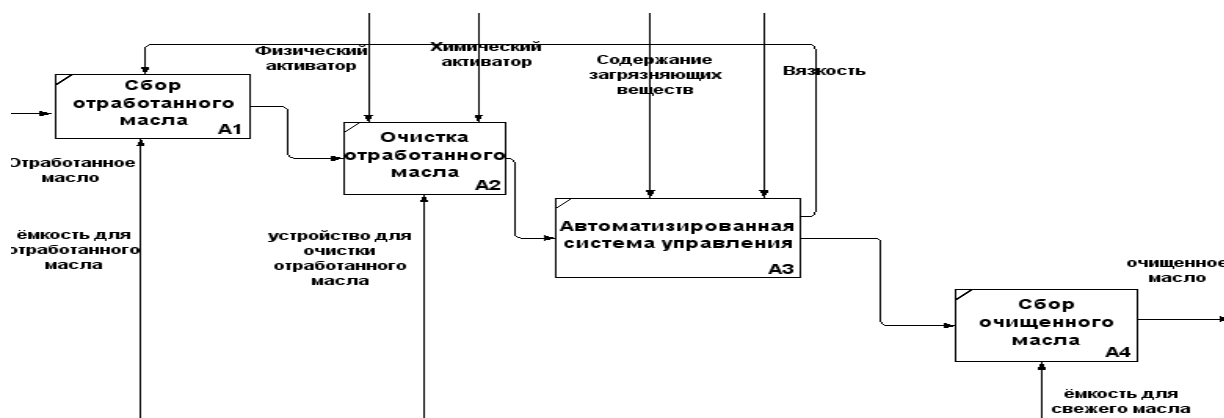


Рис.1. Декомпозиция автоматизированного процесса очистки согласно В.А. Аметову

Предлагаемая автоматизированная система состоит из ёмкости для сбора отработанного масла, устройства для очистки отработанного масла, включающее в себя

действия на физические параметры системы, такие как давление и температура.

В [3] рассматривается система очистки отработан-

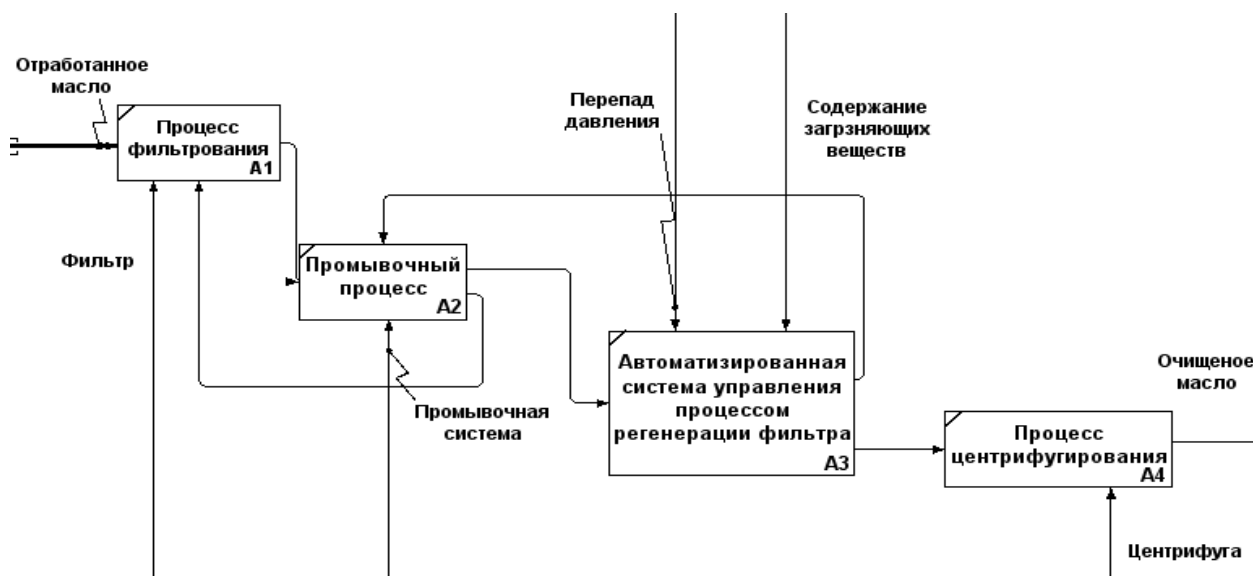


Рис 2. Декомпозиция автоматизированного процесса очистки согласно С.П. Бойко

фильтр и центрифугу, емкость для очищенного масла, а также краны для отбора проб очищенного масла. Пробы очищенного масла отбираются и происходит автоматическое определение количественного значения содержания загрязняющих веществ, уровня pH и вязкости масла.

В случае определения отклонений сравниваемых показателей качества очищаемого масла, в системе происходит возврат масла в цикл очистки).

В качестве **достоинства** данного способа автоматизации системы очистки отработанного масла можно отметить автоматизированный контроль качества очищаемого масла без остановки процесса очистки. В качестве **недостатка**, можно отметить невозможность воз-

ного масла, состоящая из фильтра и включенной в систему очистки центрифугу.

В процессе фильтрации загрязняющие частицы осаждаются на стенках фильтра, в результате чего пропускная способность снижается. Особенностью фильтра с противоточной регенерацией является возможность противоточной промывки.

Удаление осажденных загрязняющих веществ, происходит в результате автоматизированного управления работой фильтра до полного восстановления начального перепада давлений (Рис.2).

В [4] представлена центрифуга для очистки отработанных масел на дизельном двигателе с автоматическим определением толщины слоя образованного осадка (Рис.3).

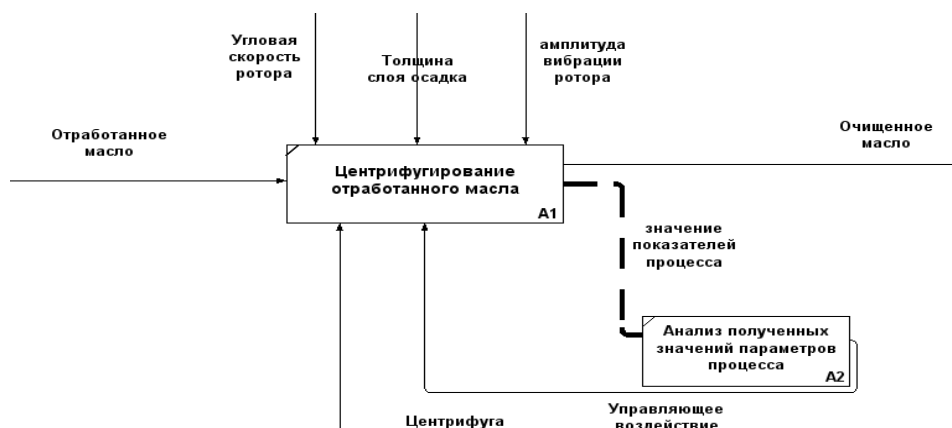


Рис.3. Декомпозиция процесса функционирования автоматизированной системы очистки согласно Патенту №1554982

Центрифуга снабжена датчиком для определения амплитуды вибрации корпуса и датчиком для определения частоты вращения ротора. Выполнение датчика устройства для определения толщины осадка в роторе в виде цилиндрического патрона с образованием внутри него герметичной полости, заполненной маслом, позволяет создавать дисбаланс ротора при его заполнении осадком.

Снабжение центрифуги датчиком для определения амплитуды вибрации корпуса, вызываемой разбалансировкой ротора, и датчиком для определения угловой скорости ротора дает возможность расчетным путем с высокой степенью точности [4].

ет снизить трудоемкость обслуживания за счет исключения ее преждевременной разборки для очистки ротора от осадка.

В [5] представлена система, состоящая из центрифуги (Рис.4), которую можно использовать для очистки отработанного масла, электродвигателя насоса, насоса, редуктора, электродвигателя шнека, датчиков определяемых параметров и персонального компьютера

Система организована таким образом, что происходит автоматический отбор проб масла после очистки и в соответствии с определенными параметрами, после проведения в ПК анализа данных, при необходимости передается управляющий сигнал на электро-

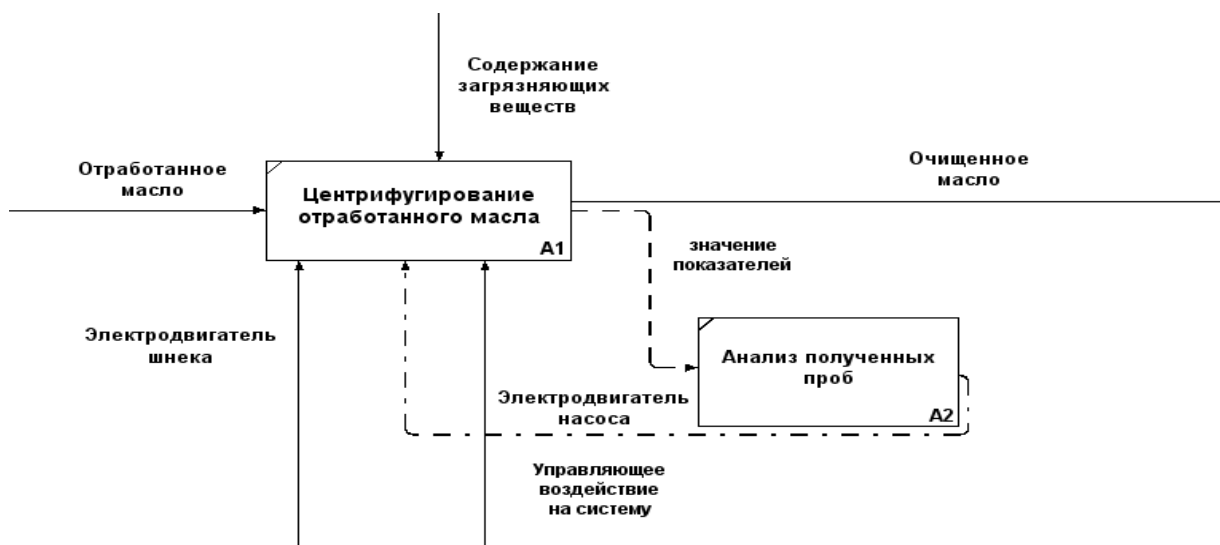


Рис. 4. Декомпозиция процесса функционирования автоматизированной системы очистки согласно Патенту №2690440

Недостатком данного способа автоматизации центрифугирования отработанного масла является невозможность определения качества очищенного мала, а также необходимость механической выгрузки образованного осадка. **Достоинством** данного способа можно считать возможность определять толщину осадка в роторе без остановки и разборки центрифуги, что позволя-

ет двигать насоса или электродвигатель шнека, в результате управляющего воздействия скорость движения ротора центрифуги увеличивается или уменьшается в соответствии с поставленными задачами. При необходимости можно задавать временные периоды отбора проб.

В качестве **недостатка** данного способа автоматизированной очистки отработанного масла можно отме-

титель не проводимый контроль за важными технологическими параметрами системы - давление и температура.

К неоспоримым **достоинствам** предлагаемого способа можно отнести высокий уровень автоматизации технологического процесса.

Проведя системный анализ существующих автоматизированных систем управления технологическим процессом очистки отработанного масла, можно выделить ряд недостатков в представленных системах:

1. Отсутствуют автоматизированные системы в которых происходит контроль важных технологических параметров – давление и температура;

2. В системах, включающих в себя одну центрифугу недостаточно высокая степень очистки отработанного масла;

В рассмотренных системах недостаточно описано влияние непостоянных внешних воздействий на систему.

Таким образом, встает вопрос о необходимости разработки автоматизированной системы очистки отработанного масла, которая будет учитывать вышеперечисленные недостатки, что позволит сделать процесс очистки масла более эффективным и добиться основной цели процесса очистки – снижения основного параметра очищенного масла – вязкости до заданных величин.

Предполагаемая автоматизированная система управления процессом очистки отработанного масла на двигателе тепловоза включает в себя (Рис.5):

- дизельный двигатель;
- фильтр грубой очистки;
- центрифугу;
- датчик температуры;
- датчик давления;
- ПИД – регулятор;
- устройство сравнения;

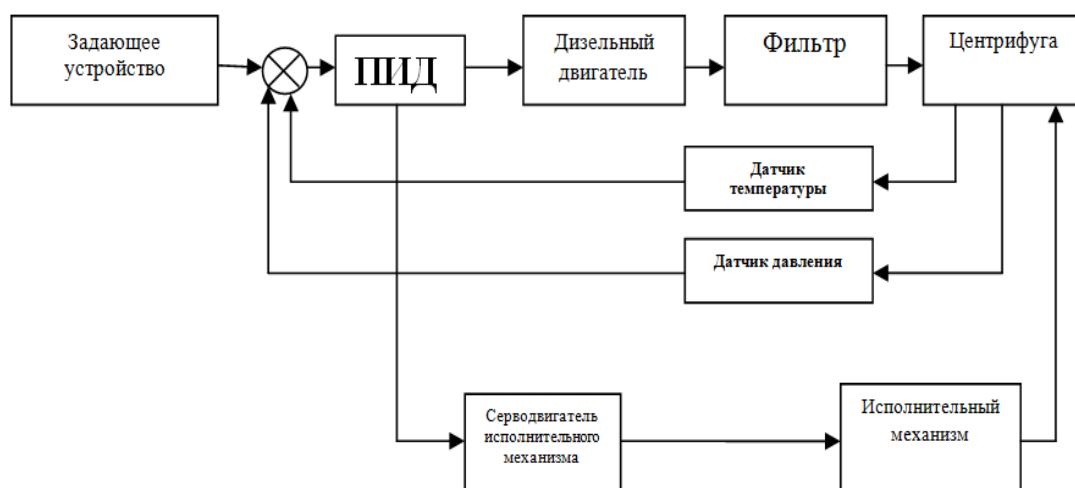


Рис. 5. Функциональная схема автоматической системы очистки отработанного масла

- серводвигатель исполнительного механизма;
- исполнительный механизм [6].

Объектом управления системы очистки отработанного масла является центрифуга. Центрифуга – это главное звено в рассматриваемой технической системе. От эффективности ее работы зависят качественные по-

казатели очищенного масла, которое в дальнейшем поступает в дизельный двигатель.

Очистка отработанного масла и получение на выходе масла с необходимыми характеристиками, для дальнейшего его использование в работе двигателя, является важной характеристикой системы управления процессом очистки отработанного масла.

Важными параметрами технологического процесса очистки являются температура и давление масла в центрифуге. Поэтому встает вопрос о постоянном контроле данных параметров.

В результате изменения давления в системе очистки масла происходит изменение важного параметра – вязкости масла, таким образом, исполнительный механизм, получая сигнал с датчика давления и сравнивая его значение с допустимыми значениями изменяет расход масла в системе, воздействуя тем самым на давление под-держивает вязкость масла в требуемых пределах.

Используя методику выбора настроечных параметров T_d, k_r, T_e ПИД – регулятора, при неизвестной модели системы управления экспериментальным методом, проведем расчет значений настроечных параметров в соответствующих блоках, и таким образом, получим снижение значения времени регулирования системы очистки отработанного масла с 83 с. до 52 с.

ЗАКЛЮЧЕНИЕ

Проведя системный анализ существующих автоматизированных систем управления процессом очистки отработанного масла можно выделить ряд недостатков в исследуемых системах.

Учитывая выше обозначенные недостатки, была предложена автоматизированная система очистки масла, в которой учитываются параметры давления и тем-

пературы, в зависимости от изменения которых, происходит воздействие на расход масла с помощью изменения сечения отверстия, через которое масло поступает в центрифугу, тем самым оказывается влияние на вязкость масла.

Разработка такой системы управления, предполагает существование неполноты или нечеткости внешнего возмущающего воздействия, вследствие чего, для большей устойчивости предполагаемой системы, необходимо использовать нетрадиционные способы управления, основанные на использовании интеллектуальных методов управления процессом очистки отработанного масла.

ЛИТЕРАТУРА

1. Указ Президента РФ от 10 октября 2019 г. № 490 "О развитии искусственного интеллекта в Российской Федерации"//Информационно-правовой портал Гарант
<https://www.garant.ru/products/ipo/prime/doc/72738946/>
(Дата обращения 12.01.2020 г.)
2. Аметов В. А. Повышение эксплуатационной надежности агрегатов автотранспортных средств путем

контроля и модифицирования смазочного масла: автореф. дис. доктора техн.наук.- Тюмень, 2006 год – 44 с.

3. Бойко С.П. Моделирование и повышение эффективности функционирования саморегулирующихся фильтров в смазочных системах судовых дизелей: автореф.канд.техн.наук.-Владивосток,2018 год – 28 с.

4. Удановский Э. А., Алексеев В. Г., Станиславский Л. В. Центрифуга для очистки масла//Патент России №1554982. 07.04.90.Бюл.№13.

5. Швек Майкл Дж., Деррик Брэдли Т. Центрифуга с автоматическим отбором проб и управлением//Патент России № 2690440. 03.06.2019.Бюл.№16.

6. Горбачева О.М., Боровский А.С. Исследование свойств системы автоматического управления процессом очистки отработанного масла на двигателе тепловоза /О.М. Горбачева, А.С. Боровский//Научно-технический вестник Поволжья.-2019.-№6.-С.- 43-46.

On the issue of developing a cleaning system waste oil based smart management practices

Gorbacheva O.M.
Orenburg State university
Orenburg, Russia
Ol.gorba4ewa2017@yandex.ru

Borovsky A.S.
Orenburg State university
Orenburg, Russia
borovski@mail.ru

Annotation. The article is devoted to the current problem of the use of intelligent methods in the process control system for the cleaning used oil were presented. After analyzing the advantages and disadvantages of the systems presented, an optimal automated system for cleaning used oil by the specified parameters was proposed treatment of waste oils. Using methods of system analysis, using IDF0 diagrams, functional models of automated systems for cleaning used oil were presented. After analyzing the advantages and disadvantages of the systems presented, an optimal automated system for cleaning used oil by the specified parameters was proposed

Keywords: waste oil, automated control system, centrifugation.

Геоинформационная методология выявления активных зон трещиноватости на основе сочетания сейсморазведочной и ландшафтной информации по космическим снимкам

Крупницкий В. В.

Югорский НИИ информационных технологий
Ул. Мира, 151, Ханты-Мансийск, 628011, Россия
krupitskiyv@uriit.ru

Полищук Ю. М.

Югорский НИИ информационных технологий
Ул. Мира, 151, Ханты-Мансийск, 628011, Россия
yupolishchuk@gmail.com

Аннотация. Статья посвящена рассмотрению методических вопросов выявления активных зон трещиноватости как зон повышенной нефтеотдачи с использованием информационно-космических технологий. В связи с недостаточностью сейсморазведочных методов для выявления активных зон трещиноватости предложено использовать ландшафтные индикаторы активных зон трещиноватости, которые определяются по результатам совместного анализа данных сейсморазведки и космических снимков. Для определения ландшафтных индикаторов использован летний космический снимок Landsat. Предложено для выявления ландшафтных индикаторов зон активной трещиноватости использовать ландшафтные особенности территории месторождения, в частности, направления течения рек и других водотоков, прямоугольные формы границ перехода заболоченных участков к лесным комплексам и др. В качестве подтверждения выявления активных зон трещиноватости посредством ландшафтных индикаторов приведено сопоставление положения ландшафтных индикаторов вблизи высокодебитной скважины на космическом снимке. Приведены примеры дешифрирования активных зон трещиноватости на основе выявленных ландшафтных индикаторов.

Ключевые слова: сейсмические данные, космические снимки, активные зоны трещиноватости, ландшафтные индикаторы.

ВВЕДЕНИЕ

В связи с естественным ухудшением ресурсной базы нефти Западной Сибири основного нефтедобывающего региона России, большой интерес представляют трудноизвлекаемые запасы. Наибольший их объем сконцентрирован в залежах, характеризующихся низкими и ультранизкими фильтрационно-емкостными свойствами (проницаемость, пористость пород коллекторов). Наибольший интерес представляют нижнемеловые и юрские пласты, развитые как на территории ЯНАО, так и ХМАО. В связи с их низкой проницаемостью и низкими дебитами основным подходом к их разработке в настоящее время является дорогостоящая технология бурения горизонтальных скважин с гидроразрывом пласта. Однако наряду с низкодебитными скважинами в Западной Сибири имеются и высокодебитные скважины, наличие которых специалисты связывают с активными зонами трещиноватости пород. Однако существующие геофизические методы поиска зон трещиноватости по

данным сейсморазведки до сих пор развиты крайне недостаточно. Так разработанная ранее [1, 2] технология выявления зон трещиноватости по данным сейсморазведки, ориентированная на их картографирование, не позволяет отличать активные от неактивных зоны трещиноватости.

Применение аэрофотосъемки и космических снимков в дешифрировании зон разломов имеет большую историю и описано в большом количестве работ [3-18]. Однако о работах по выявлению зон активной трещиноватости посредством сопоставления сейсморазведочных данных и космических снимков авторам статьи не известно. В связи с этим предлагается новый подход к решению задачи, основанный на сочетании геофизических и ландшафтно-индикационных методов. Проявления активных зон трещиноватости на земной поверхности должны отображаться с помощью ландшафтных индикаторов, в качестве которых рассматриваются особенности ландшафтных обстановок. Насколько нам известно, методические вопросы определения таких индикаторов по космическим снимкам не разработаны.

В связи с выше описанным целью статьи явилась разработка методических вопросов определения ландшафтных индикаторов активных зон трещиноватости на основе космических снимков и их использование совместно с результатами картографирования сети разломов по данным сейсморазведки.

ОПРЕДЕЛЕНИЕ ЛАНДШАФТНЫХ ИНДИКАТОРОВ АКТИВНЫХ ЗОН ТРЕЩИНОВАТОСТИ ПО КОСМИЧЕСКИМ СНИМКАМ

Как указано выше, современные геофизические методы не позволяют выявлять активные зоны трещиноватости. Перспективным для их выявления можно рассматривать подход, основанный на использовании ландшафтных индикаторов активных зон трещиноватости, которые могут быть определены по космическим снимкам. Известно [5, 9, 13], что активные зоны трещиноватости «пробиваются» на дневную поверхность как проявления геологических разломов в виде положительных или отрицательных форм рельефа. Вопросы применения методов дистанционного зондирования в дешифрировании зон разломов рассматривается в большом количестве работ [3-18]. Высокая степень заболоченности территории Западной Сибири приводит к нивелированию указанных особенностей рельефа, что затрудняет их выявление на космических снимках. В связи с

В.В. Крупницкий, Ю.М. Полищук

этим предлагается использовать ландшафтные индикаторы, учитывающие не только особенности рельефа, но и наличие и границы перехода природных комплексов (лесной или болотной растительности, водотоков и др.).

Для определения ландшафтных индикаторов активных зон трещиноватости по космическим снимкам проведем анализ связи ландшафтных изменений на земной поверхности и картографированных зон трещиноватости по данным сейсморазведки – сети разломов. Для проведения этих исследований был выбран летний космический снимок Landsat (пространственное разрешение 30м). Обработка космоснимка выполнена средствами геоинформационной системы ArcGis 10.3.

На рис. 1 представлен фрагмент космического снимка (КС) участка территории Средне-Назымского месторождения, расположенного в северо-западной части ХМАО. В центральной части фрагмента снимка показано местоположение скважины 219, которая характеризуется аномально

высокими дебитами нефти из пород баженовско-абалакского комплекса [19] и будет нами использоваться как показатель активности зон трещиноватости пород в непосредственной близости от скважины. Голубыми линиями на рисунке показана сеть разломов, картографированная по данным сейсморазведки на глубине 2400-2500 метров. На снимке зеленым и темно-зеленым цветами показаны темнохвойные леса, желтым цветом – смешанные и мелколиственные леса, розовым - заболоченные участки, черным - участки чистой воды на заболоченных территориях. Анализируя ландшафтные изменения, можно заметить, что заболоченные русла водотоков, как правило, «адаптируются» к активным зонам трещиноватости, картографированным по данным сейсморазведки [19]. Отметим, что резкие границы переходов заболоченных участков к лесным комплексам, а также мелких рек и водотоков к заболоченным участкам и т.п., будем рассматривать тоже как ландшафтные индикаторы проявления активных зон трещиноватости.

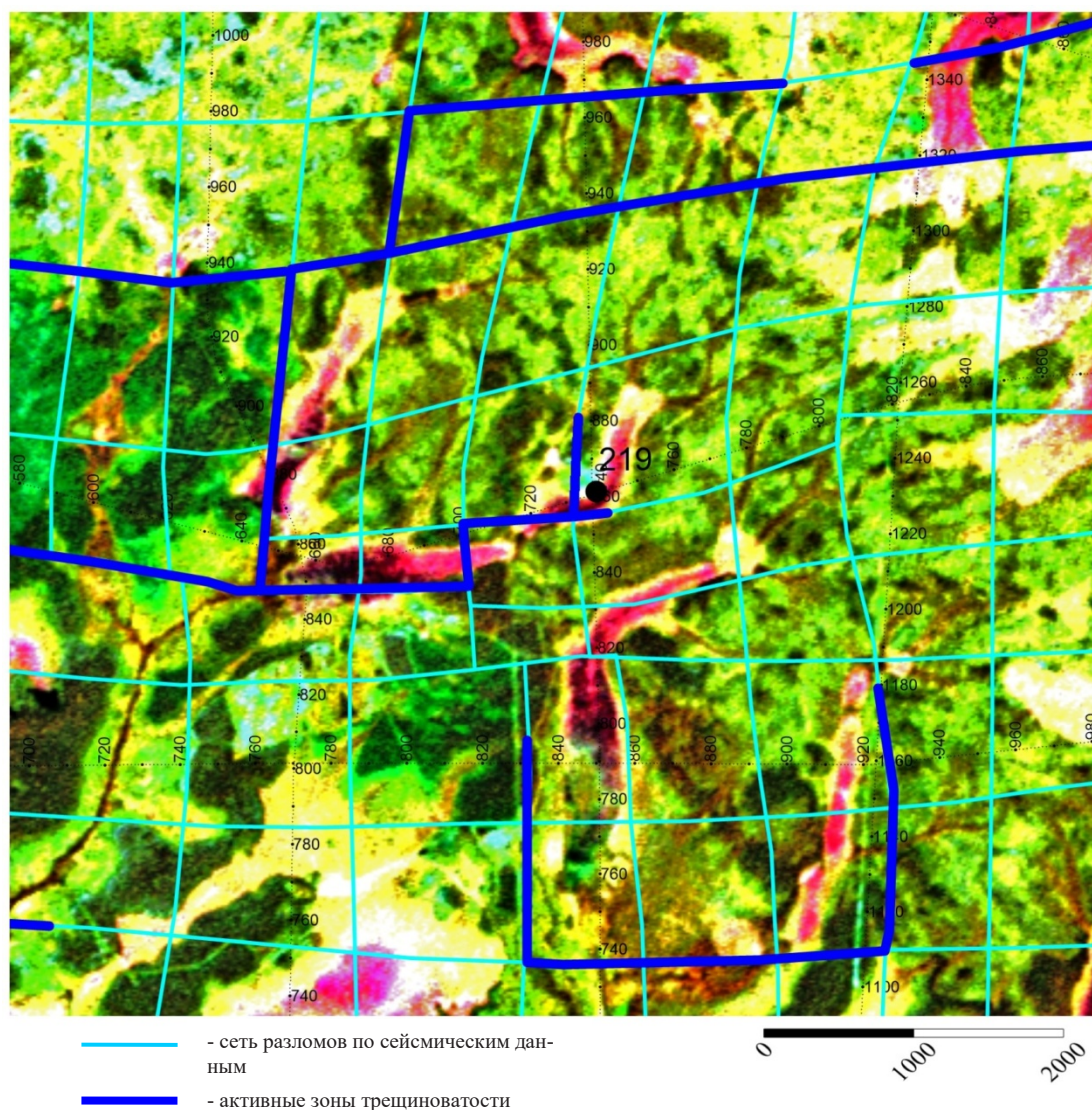


Рис. 1. Фрагмент космического снимка с нанесенными на него сетью геологических разломов и местоположением высокодебитной скважины

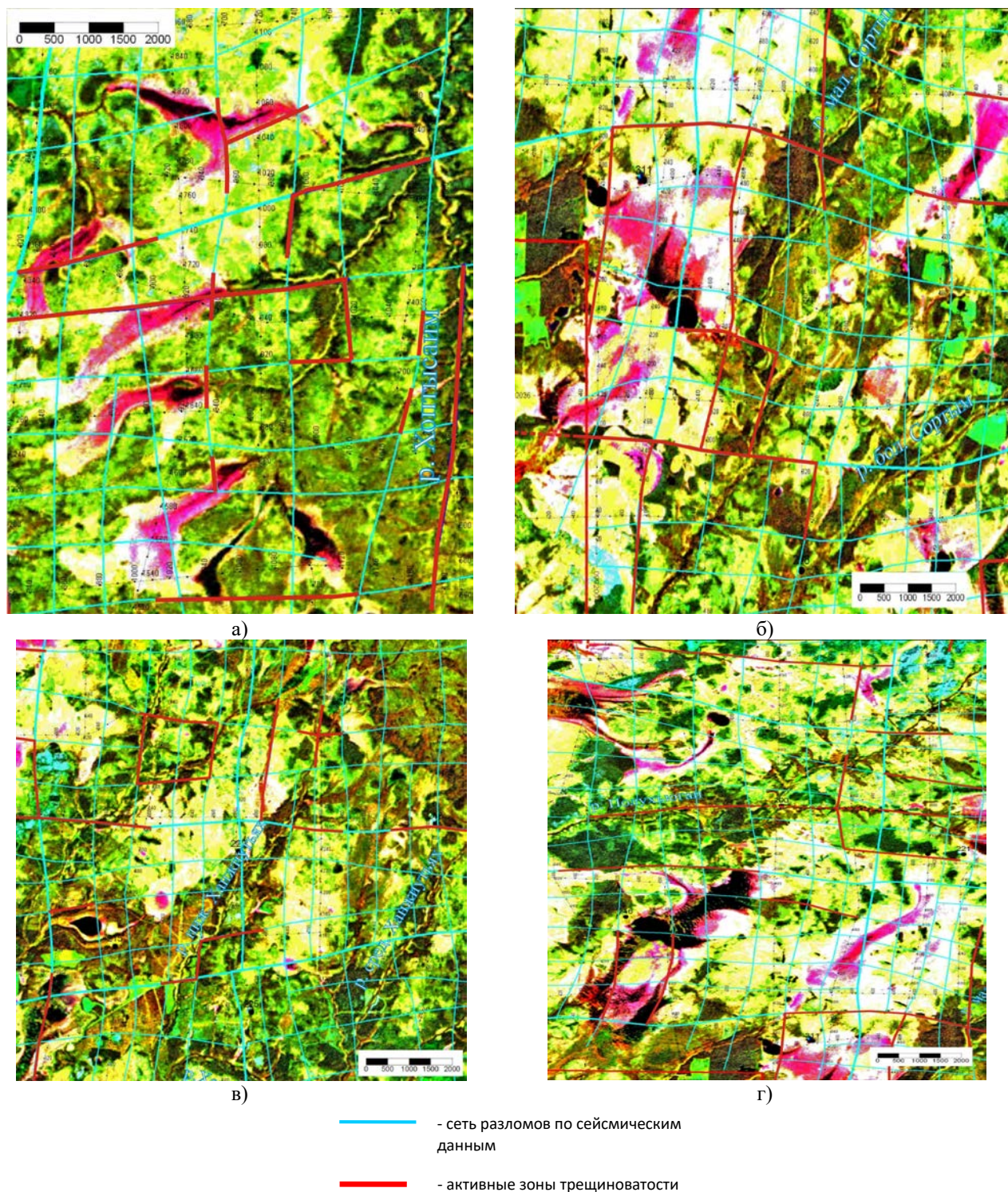


Рис.2 Примеры выявления активных зон трещиноватости по результатам комплексного анализа данных сейсморазведки и космических снимков

Используя эти ландшафтные индикаторы, на рис. 1 синими толстыми линиями показаны отрезки ребер разломной сети, которые расположены вблизи выделенных на снимке ландшафтных индикаторов. Следовательно, обозначенные синим цветом отрезки разломной сети будем рассматривать как места расположения активных зон трещиноватости. В качестве подтверждения этого может рассматриваться расположение высокодебитной скважины 219 на снимке вблизи синих линий (на расстоянии приблизительно 125 метров).

ПРИМЕНЕНИЕ ЛАНДШАФТНЫХ ИНДИКАТОРОВ В ЗАДАЧАХ ВЫЯВЛЕНИЯ АКТИВНЫХ ЗОН ТРЕЩИНОВАТОСТИ

С использованием рассмотренного выше подхода покажем возможность выявления активных зон трещиноватости на ряде других участков данного месторождения с использованием ландшафтных индикаторов. Примерами взаимосвязи между расположением русел рек и простираением зон трещиноватости на территории месторождения являются участки русел рек, в частности, реки Хоптысаим (Рис.

В.В. Крупницкий, Ю.М. Полищук

2-а), которые можно рассматривать как ландшафтные индикаторы активных зон трещиноватости. Анализ фрагмента космоснимка на рис. 2-б, показал, что ландшафтным индикатором активных зон трещиноватости, так же как и на предыдущем фрагменте, являются резкие изменения направления русел рек. На рис. 2-в и 2-г приведены фрагменты космоснимков, на которых показано, что активные зоны трещиноватости контролируют как границы перехода заболоченных участков к лесным комплексам, так и русла рек.

ЗАКЛЮЧЕНИЕ

Предложен новый подход к выявлению зон активной трещиноватости как участков повышенной нефтеотдачи, основанный на результатах совместного анализа данных сейсморазведки и космических снимков. Проведенные исследования на нефтедобывающей территории в условиях существенно заболоченных обстановок, характерных для Западной Сибири, позволили определить ландшафтные индикаторы активных зон трещиноватости, в качестве которых использованы следующие характерные признаки: направление течения рек и водотоков, а также прямоугольные формы границ перехода заболоченных участков к лесным комплексам, которые, по мнению авторов, могут являться признаками активности глубинных разломов, контролирующих зоны трещиноватости.

БЛАГОДАРНОСТИ

Работа выполнена при поддержке гранта Российского фонда фундаментальных исследований и Ямало-Ненецкого Автономного округа в рамках научного проекта № 19-47-890001.

ЛИТЕРАТУРА

1. Глухманчук Е.Д. Закономерности развития зон трещиноватости в породах осадочного чехла Западной Сибири по результатам применения технологии «OilRiver», ГИС горизонтальных скважин и данных ГРП. / Е.Д. Глухманчук, В.В. Крупицкий, А.В. Леонтьевский // Георесурсы. - 2018. - Вып. 3. - С. 222-227. DOI: <https://doi.org/10.18599/grs.2018.3.222-227>
2. Глухманчук Е.Д. Причины несоответствия геологических моделей месторождений результатам их разработки. / Е.Д. Глухманчук, В.В. Крупицкий, А.В. Леонтьевский // Геология нефти и газа. - 2016. - Вып. 1. - С. 43-49.
3. Аксенов А.А. Задачи аэрокосмогеологических исследований по структурному обеспечению нефтегазопроисковых работ. / А.А. Аксенов, В.Г. Можаяева // Исследования Земли из космоса. - 1990. - № 4. - С. 26-32.
4. Алексеев А.С. Томографический подход к выделению линеаментов на аэрокосмических изображениях. / А.С. Алексеев, И.Г. Казанцев, В.П. Пяткин // Исследования Земли из космоса. - 1988. - № 5. - С. 99-103.
5. Амурский Г.И. Использование космических снимков Земли при изучении строения зон дегазации нефтегазонасыщенных бассейнов. / Г.И. Амурский, М.С. Бондарева // Исследования Земли из космоса. - 1981. - № 3. - С. 5-10.
6. Астапенко В.Н. О природе регионального линеамента, выделенного по космическим снимкам в центральной Беларуси. / В.Н. Астапенко, А.В. Матвеев, Л.А. Нечипоренко, Н.А. Шишонков // Исследования Земли из космоса. - 1999. - № 6. - С. 64-69.

7. Бажинский Ю. Линеаменты и кольцевые образования территории Польской Народной Республики. / Ю. Бажинский, Б. Данель-Данельская, М. Граничный, М. Вильчинский // Исследования Земли из космоса. - 1982. - № 5. - С. 20-22.
8. Брюханов В.Н. О некоторых основных параметрах космических снимков с позиции их геологической информативности. // Исследования Земли из космоса. - 1983. - № 4. - С. 39-46.
9. Бурлешин М.И. Использование космических снимков для изучения особенностей развития новейших тектонических структур на территории Туранской плиты. // Исследования Земли из космоса. - 1991. - № 1. - С. 98-102.
10. Гоникберг В.Е. Использование космических снимков для реконструкции новейшего поля тектонических напряжений. // Исследования Земли из космоса. - 1983. - № 6. - С. 39-51.
11. Гушин Б.М. Геоиндикационное дешифрирование аэрокосмических снимков при нефтегазопроисковых работах в широтном Приобье и на сопредельных площадях Западной Сибири. // Исследования Земли из космоса. - 1986. - № 2. - С. 41-49.
12. Зверев А.Т. Дешифрирование кольцевых структур на космических снимках и их корреляция с геофизическими полями и строением земной коры территории СССР. / А.Т. Зверев, Я.Г. Кац // Исследования Земли из космоса. - 1986. - № 1. - С. 43-50.
13. Кроткова О.Т. Кольцевые структуры закрытых платформенных территории и оценка их тектонической активности по геоморфологическим данным. // Исследования Земли из космоса. - 1988. - № 6. - С. 57-65.
14. Лопатин Д.В. Линеamentная тектоника и месторождения-гиганты Северной Евразии. // Исследования Земли из космоса. - 2002. - № 2. - С. 77-91.
15. Татевян С.К. Роль спутниковых локационных измерений в изучении современной геодинамики. // Исследования Земли из космоса. - 1999. - № 1. - С. 87-96.
16. Трифонов В.Г. Вопросы методики геологического дешифрирования космических изображений Земли. / В.Г. Трифонов, С.Л. Бызова, Л.А. Ведешин, О.С. Деревянко, Т.П. Иванова, М.Л. Копп, Н.Н. Курдин, В.И. Макаров, С.Ф. Скобелев, П.В. Флоренский // Исследование природной среды космическими средствами. Геология и геоморфология. Москва.: ВИНТИ, - 1973. - С. 11-77.
17. Щепин М.В. Выявление кольцевых структур по результатам обработки изображений космических снимков. / М.В. Щепин, С.В. Евдокимов, Ю.В. Головченко // Исследования Земли из космоса. - 2007. № 4. - С. 74-87.
18. Аэрометоды геологических исследований // Под ред. В. К. Ерёмкина Л.: Недра. - 1971. - 704 с.
19. Отчет о результатах НИР «Прогнозирование зон развития трещинных коллекторов на основе структурно-деформационного анализа на Средне-Назымском лицензионном участке» (ответственный исполнитель Е.Д. Глухманчук). Ханты-Мансийск. - 2007. - 63 с.

Geoinformation methodology for identifying active fracture zones based on a combination of seismic and landscape information from satellite images

Vladimir V. Krupitskiy
Geology modeling Center
Ugra Research Institute of Information Technology
Khanty-Mansiysk, 628011, Russia
krupitskiyv@uriit.ru

Space service Center
Ugra Research Institute of Information Technology
Khanty-Mansiysk, 628011, Russia
yupolishchuk@gmail.com

Keywords: seismic data, satellite images, active fracture zones, landscape indicators

Abstract. The article is devoted to the consideration of methodological issues in identifying active fracture zones as zones of enhanced oil recovery using information and space technologies. Due to the insufficient seismic methods for identifying active fracture zones, it is proposed to use landscape indicators of active fracture zones, which are determined by the results of a joint analysis of seismic data and satellite images. The Landsat summer satellite image was used to determine landscape indicators. It is proposed to use landscape features of the deposit area to identify landscape indicators of active fracture zones, in particular, the direction of rivers and other watercourses, the rectangular shapes of the boundaries of the transition of marshy areas to forest complexes, etc. As a confirmation of the identification of active zones of fracture through landscape indicators, a comparison is made the position of landscape indicators near a high-debit well in a satellite image. Examples of interpretation of active fracture zones based on the identified landscape indicators are given.

Yuri M. Polishchuk

Методология системного анализа формирования инфраструктуры интеллектуального управления транспортными процессами с научным обоснованием стратегических решений

Михеева Т.И.

Самарский национальный исследовательский университет имени академика С.П. Королёва,
Самарский государственный технический университет Самара, Группа компаний ИнтелТранс
Mikheevati@gmail.com

Михеев С.В.,

Самарский национальный исследовательский университет имени академика С.П. Королёва, Самара,
Группа компаний ИнтелТранс
Mikheevati@gmail.com

Чугунов А.И.

Самарский национальный исследовательский университет имени академика С.П. Королёва, Самара,
Группа компаний ИнтелТранс

Аннотация. В статье рассматривается методология системного анализа формирования инфраструктуры интеллектуального управления транспортными процессами с научным обоснованием стратегических решений в интеллектуальной транспортной геоинформационной системе ИТСГИС («ITSGIS»). Описан системный подход сетецентрического управления.

Ключевые слова: интеллектуальная транспортная геоинформационная система, «ITSGIS», инфраструктура, сетецентрическое управление, стратегические решения управления

ВВЕДЕНИЕ

Успехи автомобилестроения последних десятилетий, приведшие к «взрывному росту» автомобильного парка России, в значительной степени опережают темпы дорожного строительства, в связи с чем предъявляются все более высокие требования к качеству проектирования улично-дорожных сетей (УДС) и организации дорожного движения. В этой связи, серьезную проблему представляет создание систем комплексного автоматизированного управления, как всей транспортной инфраструктурой урбанизированной территории, включающей в себя улично-дорожную сеть, технические средства организации дорожного движения, транспортные потоки, так и ее отдельными объектами. Требуется принципиально новые подходы к созданию таких интегрированных систем, одновременно охватывающих массивы разнородных данных и обеспечивающих многоуровневое взаимодействие множества подчиненных сложных подсистем, строительство городских скоростных дорог, развитие методов управления транспортными потоками, включая применение интеллектуальных транспортных систем [1, 2, 3].

Создание интегрированных интеллектуальных транспортных систем (ИТС), как систем управления дорожно-транспортной инфраструктурой, позволит повысить уро-

вень организации дорожного движения: улучшить характеристики улично-дорожной сети, усовершенствовать дислокацию технических средств организации дорожного движения, оптимизировать процесс управления транспортными потоками на всех фазах движения, уменьшая транспортные задержки, повышая безопасность движения.

В работе рассматривается системный анализ формирования инфраструктуры интеллектуального управления транспортными процессами с научным обоснованием стратегических решений [4, 5, 6]. Анализ российского и зарубежного опыта проектирования и развития ИТС, принципов их интеграции, использования при осуществлении грузовых и пассажирских перевозок, при управлении ТП в различных транспортных ситуациях при возникновении ДТП позволяет сделать вывод о целесообразности использования ИТС на автомобильном транспорте.

РАЗВИТИЕ МЕТОДИЧЕСКОГО ОБЕСПЕЧЕНИЯ В ОБЛАСТИ ИТС

Разрабатываемая многоуровневая, сложноорганизованная ИТС представляет собой гибридную систему, состоящую из множества разнородных систем, сложным образом взаимодействующих друг с другом – управляющих, классифицирующих, прогнозирующих, экспертных, принимающих решения или поддерживающих эти процессы, объединенных для достижения единой цели, с визуализацией и дислокацией объектов транспортной инфраструктуры на тематических слоях интерактивной электронной карты в интеллектуальной транспортной геоинформационной системе ИТСГИС («ITSGIS») [7, 8].

Интеллектуальная транспортная система является классическим примером сложной системы с присущими ей свойствами [9, 10, 11, 12]:

- многомерностью и иерархичностью системы, обусловленной большим числом взаимосвязанных между собой элементов;

- эмерджентностью системы;
- многофункциональностью элементов системы;
- многокритериальностью, обусловливаемой имманентностью (несовпадением) целей отдельных элементов системы;
- сложным (вероятностным и динамическим) поведением, проявляющимся во взаимосвязи подсистем и требующим обратной связи при управлении;
- сложностью информационных процессов;
- необходимостью высокой автоматизации управления.

С позиций структурно-функционального подхода ИТС можно определить, как единство структуры, функций и целостности. Структура характеризует элементы ИТС и их взаимодействие. Функции определяют природу связей между элементами и поведение ИТС. Целостность выражает взаимозависимость структуры, функций и целей ИТС и проявляется в наличии у реальной системы ряда новых, т.е. эмерджентных свойств, не присущих отдельным ее элементам, и не выводимых из свойств этих элементов, и способов их соединения.

Попытка формирования инфраструктуры интеллектуального управления транспортными процессами с научным обоснованием стратегических решений на основе ИТС реализована созданием Центра организации дорожного движения. Центр выступает координатором по циклу работ, связанных с реализацией мероприятий:

- подготовка проектов законодательных актов и нормативов в рамках юрисдикции Администрации города, методического обеспечения создания безопасной и эффективной системы управления дорожным движением, организации дорожного движения и парковок;
- контроль за соблюдением законодательных актов и нормативов, разработанных Департаментом организации дорожного движения федерального и регионального уровней;
- координацию деятельности по разработке и внедрению стратегий, планов, мероприятий по обеспечению организации дорожного движения;
- рассмотрение и согласование проектов и схем организации дорожного движения;
- внедрение современных инновационных технологий и методов организации дорожного движения: интеллектуальных транспортных систем, автоматизированных систем управления дорожным движением [13, 14];
- рассмотрение и согласование проектов строительства дорог и улиц, утверждение сроков строительства, заключение контрактов на строительные работы, связанные с мероприятиями по организации и безопасности дорожного движения, назначение экспертиз проектов строительства и реконструкции дорожной сети, маршрутов и схем движения пассажирского транспорта, проектов размещения всех видов наружной рекламы и т.д., разработанных другими организациями и влияющих на условия дорожного движения;
- участие в работе градостроительных и технических советов, комиссий по приемке в эксплуатацию дорог, дорожных сооружений, железнодорожных переездов, технических средств организации дорожного движения и иных объектов;
- осуществление контроля над выполнением работ по организации дорожного движения, в соответствии с действующими нормативами, тарифами, правилами и утвержденной проектно-сметной документацией;
- подготовку и оформление распорядительной и технической документации по закрытию движения транспорта на участках дорог при их строительстве, реконструкции и ремонте; рассмотрение и согласование предложений по временному ограничению или запрещению дорожного движения на УДС;
- осуществление мониторинга и сбора данных о дорожном движении на УДС, анализ результатов мониторинга, оценка схем и стратегий организации дорожного движения, пропускной способности улично-дорожной сети, разработка рекомендаций по повышению эффективности и безопасности дорожного движения;
- организацию информирования, маршрутизации и ориентирования участников движения;
- анализ причин возникновения очагов концентрации дорожно-транспортных происшествий и подготовка предложений по их ликвидации;
- осуществление мероприятий по повышению привлекательности и эффективности перевозок общественным транспортом с обеспечением приоритета движения общественного транспорта;
- разработка концепции развития уличных и общественных внеуличных парковок, включая их дислокацию на карте в ИТСГИС, количество машиномест, тарифную политику с методикой парковочных норм для применения в жилых и нежилых районах, осуществление контроля за соблюдением принятых норм и порядка организации парковок автотранспорта.

СИСТЕМНЫЙ АНАЛИЗ, КЛАССИФИКАЦИЯ ЗАДАЧ ИТСГИС

Классификация задач, решаемых в рамках функционирования транспортной инфраструктуры, определяет стратегию и тактику синтеза интеллектуальных транспортных геоинформационных систем.

1. Задачи мониторинга

1.1. Мониторинг транспортных потоков [14]:

- мониторинг характеристик транспортных потоков (скорость, интенсивность, плотность, состав и др.);
- сбор данных об условиях движения с помощью контрольных автомобилей;
- управление движением на скоростных дорогах.

1.2. Мониторинг характеристик улично-дорожной сети:

- паспортизация УДС, многоуровневых транспортных развязок и тоннелей;
- паспортизация надземных и подземных пешеходных переходов;
- паспортизация железнодорожных переездов;
- оценка текущего состояния УДС;
- мониторинг аварийно-восстановительных работ.

1.3. Мониторинг технических средств управления движением

- реестр технических средств организации дорожного движения (дорожные знаки, светофорные объекты дорожная разметка и т.д.);
- магистральное и сетевое управление светофорной сигнализацией;

- автоматическая электронная плата за проезд и парковку.

1.4. Мониторинг загрязнения окружающей среды

2. Задачи управления

2.1. Управление транспортными потоками [15]:

- координированное управление транспортными потоками;
- оценка качества функционирования транспортной сети;
- управление движением в чрезвычайных ситуациях;
- обнаружение дорожно-транспортных происшествий;
- мониторинг заторовых ситуаций для оценки динамики их развития;
- разработка стратегии управления дорожным движением в условиях заторовой ситуации;
- интеграция систем управления дорожным движением;

2.2. Управление перевозочным процессом

- обеспечение дотранспортной информацией, информирование клиентов о маршрутной сети, планирование поездов;
- бронирование транспортных услуг;
- оценка спроса на перевозки;
- маршрутное ориентирование, on-line мониторинг прохождения маршрута;
- выработка стратегии управления в конкретных ситуациях;
- оперативное изменение схем организации дорожного движения;
- управление приоритетным движением маршрутного транспорта;
- маршрутная навигация и предоставление приоритета специальным одиночным и колоннам транспортных средств;
- мониторинг перевозки опасных и крупногабаритных грузов;
- оптимизация маршрутной сети;
- интеграция систем управления перевозками;

3. Задачи информационного обеспечения участников движения

- передача информации по каналам связи;
- сегментация информационных потоков;
- интеграция систем управления БД о дорожном движении.

4. Задачи визуализации геообъектов транспортной инфраструктуры [16-20]

- разработка тематических слоев интерактивной электронной карты;
- визуализация комплексной дислокации геообъектов (автодорог, тротуаров, технических средств организации дорожного движения и т.д.);
- экспертный анализ комплексной дислокации геообъектов;
- моделирование транспортных процессов с интеграцией методов управления.

Анализ ИТСГИС показывает, что концепция ее функционирования заключается в реализации функций управления движением и перевозками, в оценке степени влияния различных плагинов на развитие транспортной инфраструктуры, в создании архитектуры системы управле-

ния транспортной инфраструктурой и согласовании стандартов для развития ИТСГИС, как интегрированных систем. Технологии ИТСГИС имеют в настоящее время более 40 направлений применения, однако, в силу имманентности целей каждого плагина ИТСГИС в отдельности, его потенциальные возможности, как системной единицы, реализуются решением специализированного функционала.

Синергетический эффект при проектировании ИТСГИС проявляется в форме организационно обусловленного перехода от имманентности к синергии за счет расширенной системной и функциональной интеграции (см. рис. 1):

- постановка проблем организации перевозок и движения;
- разработка концепции функционирования;
- обеспечение функциональной, институциональной, информационной интеграции;
- развитие плагинов каждой функциональной группы;
- интеграция информационных технологий;
- интеграция информационных потоков между плагином.

Уровень системности тем выше, чем выше интенсивность взаимодействия плагинов ИТСГИС друг с другом, чем сильнее отличаются свойства системы от свойств входящих в нее элементов.

Функциональная интеграция ИТСГИС обеспечивается полнотой и эффективностью сбора, обработки и передачи информации в реальном масштабе времени о характеристиках транспортных потоков, дорожно-транспортных происшествиях, заторах, движении общественного и грузового транспорта и т.д.

Временная интеграция обусловлена необходимостью оперативного принятия решения в процессе управления транспортными потоками, выработки и реализации управляющих воздействий в реальном масштабе времени.

Институциональная интеграция необходима для примирения интересов частных фирм, муниципальных властей, органов государственного управления, которые, пытаясь сохранить свои интересы, ограничивают доступ к информации ИТСГИС.

Современные тенденции развития ИТСГИС показывают, что одной из основных целей их функционирования является предоставление мультимодальной информации не только управляющим структурам, но и непосредственно участникам движения. Одним из наиболее интенсивно развивающихся направлений ИТСГИС является маршрутная навигация. Расширенная концепция навигации в ИТСГИС предусматривает обязательное выполнение таких функций как *on-line* мониторинг транспортных потоков и показателей качества функционирования улично-дорожной сети, динамический выбор маршрута движения и информационное обеспечение водителей при прохождении маршрута.

Новейшие достижения в области информационных технологий, компьютерной техники, современных видов связи, эффективных навигационных систем, технических средств сбора, обработки информации и регулирования дорожным движением нашли широкое применение в «ИТСГИС» при управлении движением и автомобильными

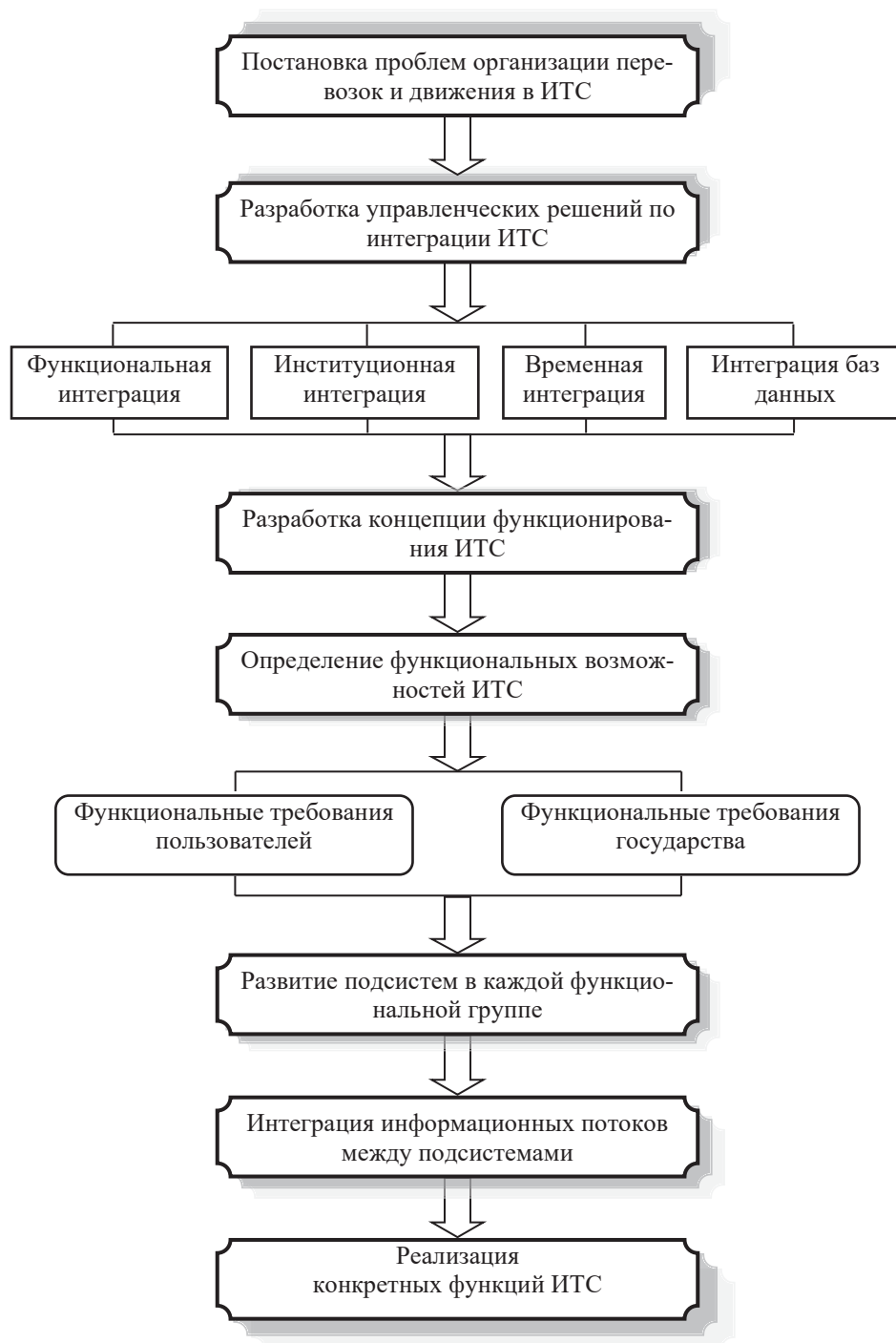


Рис. 1. Интеграция интеллектуальных транспортных систем

перевозками из-за разрабатываемых и реализуемых научных знаний для использования всего спектра функциональных возможностей указанных разработок.

Реализуются разработка и развитие методов, моделей, алгоритмов и программного обеспечения для комплексного решения задач управления транспортной инфраструктурой, оптимизации управления транспортными потоками на урбанизированной территории в условиях применения «ITSGIS». Интеграция разнородных информационных моделей системы управления транспортной инфраструктурой осуществляется на базе формального аппарата формирования знаний о предметной области, основанного на концепции объектно-ориентированного подхода.

Сетецентрическая модель системы управления транспортной инфраструктурой в ITSGIS основана на принципах распределенности объектов и зональности управления. Модель сетецентрического управления является основой системы поддержки процессов принятия решений при сетецентрическом управлении объектами, зонами и самой транспортной инфраструктурой – модель формализованного синтеза сетецентрического управления. Семантика свойств сетецентрической модели: самоорганизация, открытость, слабая иерархия в контуре принятия решений и способность порождать цели внутри себя, мультиагентность. Модели объектов транспортной инфраструктуры взаимодействуют в едином информационном пространстве. Сетецентрическим принципом управления является построение зон управления, предназначенных для принятия управленческих решений и обмена информацией между зональными центрами управления.

Модель сетецентрического управления при движении транспортных средств на регулируемых перекрестках на основе мониторинга исходных данных:

$$f(t) = \begin{cases} 0, & \text{если } t < t_{\min}, \\ \alpha \lambda e^{-\lambda(t-t_{\min})}, & \text{если } t \geq t_{\min}, \end{cases} \quad (1)$$

где $f(t)$ – плотность распределения интервалов в транспортном потоке;

α – доля свободной части транспортного потока, определяемый как $\alpha = 1 - \theta$,

θ – доля автомобилей в группах; λ – параметр распределения;

t_{\min} – минимальный интервал между транспортными средствами в потоке главного направления, с;

λ – параметр распределения интервалов в главном потоке;

$$\lambda = \frac{(1-\theta)I_p}{1-t_{\min}I_p} = \frac{\alpha I_p}{1-t_{\min}I_p}. \quad (2)$$

где I_p – интенсивность движения на главном направлении проезжей части, авт/с;

$$\alpha = e^{-\beta I_p}; \quad (3)$$

где β – параметр, определяемый экспериментально и имеющий значения от 6 до 9.

Пропускная способность

$$Q_e = \frac{3600(1-\theta)I_p e^{-\lambda(t_c-t_{\min})}}{1-e^{-\lambda t_f}} = \frac{3600 \alpha I_p e^{-\lambda(t_c-t_{\min})}}{1-e^{-\lambda t_f}} \quad (4)$$

где Q_e – пропускная способность второстепенного направления на пересечении, авт/ч;

t_c – критический интервал, с;

t_f – интервал следования из очереди второстепенного потока, с.

Критический интервал – минимальный интервал в потоке проезжей части, при котором возможен въезд на кольцо со второстепенного направления $t_c = 4,8$ с.

По результатам моделирования пропускная способность пересечений достаточно высока и их можно применять при суммарной интенсивности движения до 2100 – 2300 авт/ч. В указанном диапазоне интенсивности нерегулируемые перекрестки имеют лучшие показатели суммарной пропускной способности и суммарных задержек, только в случае если на одну из улиц в составе перекрестка приходится более 80% от общей интенсивности движения.

ЗАКЛЮЧЕНИЕ

В качестве структурно-параметрической модельной проблемы задача сетецентрического управления транспортной инфраструктурой (объектами и процессами) реализована в разработке методов управления транспортными объектами и процессами и методики проведения имитационного эксперимента в рамках спроектированной системы поддержки принятия решения при управлении транспортной инфраструктурой в локальных зонах координированного управления. Описываемая интеллектуальная транспортная геоинформационная система «ITSGIS» отличается наличием развитых средств поддержки имитационной среды, обеспечивающих простоту модификации и расширение спектра исследовательских задач на основе паттернов, нейронных сетей.

В рамках сетецентрического управления решены задачи стратифицированного зонального управления транспортными процессами: локальное управление на перегоне и перекрестке, координированное управление на магистралях с выработкой управляющих воздействий для различных типов зонирования транспортной инфраструктуры. Для больших зон, с точки зрения территории района управления, координированное управление разделено на задачи:

- раздел района на особые зоны и оперативное формирование программ координации для каждой из них;
- синхронизация программ координации (при условии равенства или кратности циклов управления).

Задачи системного управления объектами и процессами транспортной инфраструктуры взаимосвязаны между собой.

ЛИТЕРАТУРА

1. Brilon W. Roundabouts: A State of the Art in Germany. National Roundabout Conference, Vail, Colorado; May 22 – 25, 2005. – 16 p.
2. Mikheeva T.I. Hardware-software complex of management of interactive content / T.I. Mikheeva, S.V. Mikheev,

- O.K. Golovnin // Proceedings of the 14th international workshop on computer science and information technologies CSIT'2012, Vol. 2. Ufa – Hamburg – Norwegian Fjords, 2012. Pp. 63-65.
3. Бетелин, В.Б. Основные понятия контролируемого выполнения сложных систем / В.Б. Бетелин, В.А. Галатенко, К.А. Костюхин // Информационные технологии. 2013. № 3. Приложение. 32 с.
4. Михеева Т.И. Параметризация управляющих объектов урбанизированной территории / С.В. Михеев, О.К. Головнин // Известия Самарского научного центра Российской академии наук. – 2015. – Т. 17, № 2 (5). – С. 1058–1062.
5. Capacity and Level of Service at Finish Unsignalized Intersections// Finnra Reports, 2004. – 214 p.
6. Cascetta, E. Transportation systems analysis : models and applications // NY : Springer, 2009. – 752 p.
7. Mikheev, S.V. Disposition of objects of a transport infrastructure in the GIS «ITSGIS» / S.V. Mikheev, T.I. Mikheeva, K.A. Malykova, D.A. Mikhaylov / Computer Science and Information Technologies // Proceedings of the 14th international workshop on computer science and information technologies CSIT'2012, Vol. 2. Ufa – Hamburg – Norwegian Fjords, 2012. – Pp. 74-76.
8. Михеева Т.И. Модель пространственных данных оценки состояния объектов транспортной инфраструктуры в интеллектуальной ГИС «ITSGIS» / Т.И. Михеева, А.А. Федосеев, О.А. Япрынцева, О.К. Головнин // Геоинформационные технологии в проектировании и создании корпоративных информационных систем. – Уфа: Уфимский гос. авиационный технический ун-т, 2013. – С. 68-72.
9. Михеева Т.И. Программная таксономия – основа для создания гипермедийных обучающих систем / Т.И. Михеева, И.Е. Михеенков // Информационные технологии, 1998. №8. С. 40-43.
10. Рокицкий Р.Б. Объектно-ориентированные базы данных с использованием реляционных СУБД // Кибернетика и системный анализ. 2000. № 6. С. 27 – 38.
11. Кудинов, А.В. Геоинформационные технологии в задачах управления пространственными сетями / А.В. Кудинов // Геоинформатика-2000 : труды междунауч.-практ. конф. – Томск : Изд-во Томского ун-та, 2000. – С. 224 – 229.
12. Михеев, С.В. Проектирование систем управления дорожным движением на основе отношений наследования свойств // Труды 6 международной научно-практ. конф. «Организация и безопасность дорожного движения в крупных городах» – СПб: СПбАДИ, – 2004. С. 241–245.
13. Михеев, С.В. Управление транспортной инфраструктурой / Т.И. Михеева, С.В. Михеев, О.Н. Сапрыкин. – Самара: Интелтранс, 2015. – 173 с.
14. Михеева Т.И. Интеллектуальная транспортная система. Дислокация дорожных знаков / // Вестник Самарского гос. техн. ун-та. Серия «Технические науки» №32. Самара: СамГТУ, 2005. С.53-63.
15. Михеева Т.И. Метод синтеза системы зонального сетевидного управления транспортными процессами / Т.И. Михеева, С.В. Михеев, О.К. Головнин // Известия Самарского научного центра Российской академии наук. – 2016. – Т. 18, № 4 (4). – С. 799–807.
16. Михеева Т.И. Информационная технология автоматической дислокации геообъектов транспортной инфраструктуры на улично-дорожной сети / Т.И. Михеева, А.В. Сидоров, О.К. Головнин // Перспективные информационные технологии (ПИТ-2013) //Труды междунауч.-техн. конф. – Самара: Изд-во Самарск. науч. центра РАН , 2013. – С.236-241.
17. Куржанский, А.Б. Роль макромоделирования в активном управлении транспортной сетью / А.Б. Куржанский, А.А. Куржанский, П. Варайя // Труды МФТИ. – 2010. – Т. 2, № 4. – С. 100 – 118.
18. Михеева Т.И. Паттерны поддержки принятия решений по дислокации технических средств организации дорожного движения / Т.И. Михеева, О.К. Головнин // Перспективные информационные технологии (ПИТ-2013) //Труды междунауч.-техн. конф. – Самара: Изд-во Самарск. науч. центра РАН , 2013. – С.267-273.
19. Михеева Т.И. Исследование методов локального управления транспортными потоками / Т.И. Михеева, С.В. Михеев, // Вестник Самарского государственного аэрокосмического университета. Сер. «Актуальные проблемы радиоэлектроники» - Самара: СГАУ, - 2003. С. 24-30.
20. Михеева Т.И. Система моделирования «Транспортная инфраструктура города» / Т.И. Михеева, И.А. Рудаков, И.А. Чугунов // Вестник Самарск. гос. техн. ун-та. Серия «Технические науки». 2008, №1. С. 28–38.

The methodology of system analysis of the formation of the infrastructure of intelligent control of transport processes with the scientific substantiation of strategic decisions

Mikheeva T.I.

Doctor of Technical Sciences, Professor

Samara National Research

University named after academician S.P. Koroleva,

Samara State Technical University

Samara, IntelTrans Group

Mikheevati@gmail.com

Mikheev S.V.,

Doctor of Technical Sciences, Associate Professor

Samara National Research

University named after academician S.P. Koroleva,

Samara, IntelTrans Group of Companies

Mikheevati@gmail.com

Chugunov A.I.

graduate student

Samara National Research

University named after academician S.P. Koroleva,

Samara, IntelTrans Group of Companies

Mikheevati@gmail.com

Annotation. The article discusses the methodology of a system analysis of the formation of the infrastructure of intelligent management of transport processes with the scientific substantiation of strategic decisions in the intelligent transport geographic information system ITSGIS ("ITSGIS"). The system approach of network-centric management is described.

Keywords: Intelligent transport geographic information system, "ITSGIS", infrastructure, network-centric management, strategic management solutions.

Оценка технологического потенциала развития программы лояльности продуктовой ритейл-сети

Неволин О.В., Петриченко Ю.В.

Челябинский государственный университет

Челябинск, Россия

nevoleg96@gmail.ru, yulia-c@yandex.ru

Аннотация. Программы лояльности продуктовых ритейл-сетей в настоящее время – это один из важнейших рычагов управления уровнем удовлетворенности покупателей, что подтвердили ранее проведенные исследования рынка. В статье рассмотрены возможные пути развития существующей программы лояльности с технической, управленческой и экономической точек зрения на примере продуктовой ритейл-сети, ведущей деятельность на рынке Челябинской области. Это позволило предложить описанный в статье переход к модели динамической мотивации потребителя.

Ключевые слова: продуктовой ритейлинг, программа лояльности, промо-акция, потребители, торговая сеть, динамическая мотивация.

ВВЕДЕНИЕ

Современный потребитель не просто совершает выбор между продуктами разных торговых марок, отличающимися качеством и стоимостью, – на первом шаге происходит выбор «поставщика», т.е. торговой сети в целом или конкретного магазина, где можно выгоднее или удобнее купить «здесь и сейчас». В этом заключается суть актуальной в наши дни индивидуализации подхода к покупателю. Именно к такому выводу привел нас предшествующий далее описанному исследованию анализ рынка на предмет поиска тенденций современного сетевого продуктового ритейлинга в разрезе удовлетворенности потребителей.

Выделенные тенденции снижения индекса потребительского доверия и эффективности традиционного подхода к промо-акциям побудили к дальнейшему изучению темы совершенствования индивидуализации предложений для конкретного покупателя, и в качестве объекта исследования выступает одна из торговых сетей, работающих в сфере реализации продовольствия, а также других повседневных потребительских товаров.

ОБЪЕКТ ИССЛЕДОВАНИЯ В КОНКУРЕНТНОЙ СРЕДЕ

Ритейлер начал свою деятельность на территории Челябинской области более 25 лет назад. Сегодня сеть располагает более чем семьюдесятью магазинами разных форматов – от экспресс- до гипермаркетов, – по собственной оценке ежедневно обслуживающих порядка ста тысяч покупателей. Будучи региональным игроком рынка продуктового ритейла, у компании есть возможность как опираться на опыт крупных федеральных ритейлеров в различных сферах (в том

числе, вопросах программ лояльности), так и экспериментировать, пробуя неординарные решения. Но прежде чем говорить о собственном пути развития, попробуем рассмотреть примерную позицию рассматриваемой сети на рынке относительно конкурентов.

Выручка от реализации продукции (на основании данных, находящихся в открытом доступе) за 2018 год составила 11 873 512 тысяч рублей, при этом в том периоде функционировали 63 разноформатных магазина. Таким образом, средняя выручка за год с одной точки продаж составила 188 468 444 рубля. Попробуем сопоставить эти данные с информацией по другим торговым сетям за аналогичный период (таблица 1). Данные по рассматриваемой компании взяты из внутрикорпоративной документации, а также из информации в открытых источниках (как и относительно сетей-конкурентов).

Таблица 1

Усредненные значения оборачиваемости одной торговой точки некоторых продуктовых ритейл-сетей

Ритейлер	Количество торговых точек по состоянию на 2018 год, шт.	Совокупная выручка по состоянию на 2018 год, руб.	Средний оборот по торговой точке, руб.
Исследуемая сеть	63 ²	11 873 512 000	188 468 444
Пятерочка	13 522	1 198 000 000 000	88 596 361
Перекресток	760	230 800 000 000	303 684 210
Магнит	18 399	1 230 000 000 000	66 581 459
Лента	244	392 100 000	1 606 967

Безусловно, свою роль в полученных показателях играет и территориальная распределенность магазинов (то есть, другими словами, регионы, в которых ритейлер имеет наиболее влиятельное присутствие). Получить подобную информацию в разрезе региона Челябинской области по федеральным сетям не представляется возможным, ввиду чего попробуем оперировать полученными значениями.

Из результатов видно, что средний оборот по торговой точке исследуемой сети превышает аналогичные

О.В. Неволин, Ю.В. Петриченко

показатели сетей “Пятерочка” и “Магнит”. Следует полагать, что более низкое значение обусловлено колоссальным количеством торговых точек формата экспресс-магазинов у дома (так, например, “Пятерочка” не работает в формате гипермаркетов, а у “Магнита” 244 магазина площадью свыше 2 тыс. м², что составляет 1,2% всех магазинов) [1]. В то же время как у исследуемой сети магазины формата “гипер-” составляют 9,5% от всего числа торговых точек.

Если говорить о количественном соотношении торговых точек на территории Челябинского городского округа в сравнении с другими сетями (по данным картографических сервисов), можно получить следующую картину: количество торговых точек уступает федеральным игрокам, работающим в большей степени в формате экспресс-магазинов, но сопоставимо с показателями “Дикси” (рисунок 1).

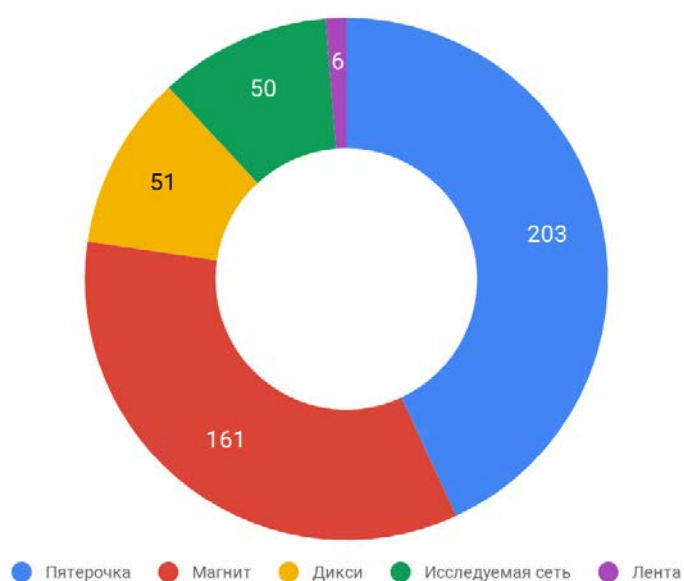


Рис. 1. Соотношение торговых точек на территории Челябинского городского округа и г. Копейска

Таким образом, можно сказать, что сейчас положение компании на рынке Челябинска достаточно стабильное и конкурентоспособное. Теперь, когда рассматриваемая сеть рассмотрена в разрезе конкурентов, можно перейти к рассмотрению основного инструмента конкурентной борьбы и достижения высокого уровня потребительского доверия – программам лояльности.

ТЕКУЩАЯ СИТУАЦИЯ И ПРЕДПОСЫЛКИ РАЗВИТИЯ

В исследуемой торговой сети действует бонусная программа лояльности на следующих условиях: при совершении покупки с использованием карты покупателя клиент получает бонусы по курсу, которыми впоследствии может оплатить до 100% последующих покупок. Исключение по возможности начисления и списания баллов составляют подакцизные товары; по возможности начисления баллов – товары из акционного ассортимента.

Актуальность программ лояльности накопительного (бонусного) типа (как и сравнение базовых условий

программ лояльности основных игроков рынка) была обозначена в предшествующем исследовании: рассмотрение и сравнение типов программ в динамике показало, что доля карт скидочного формата снизилась за последние 6 лет, в то время как получение и использование бонусных карт стало гораздо более распространенным явлением [2]. Стоимость карты в рассматриваемой сети фиксированная, приобрести ее можно на кассе в любом магазине сети. После приобретения покупатель может зарегистрироваться в программе лояльности несколькими способами:

- через форму личного кабинета на сайте ритейлера;
- через SMS-сообщение по шаблону;
- через call-центр.

На предприятии выделены несколько сотрудников, в сферу ответственности которых входят процессы принятия обращений от покупателей, регистрации бонусных карт и ответов на вопросы о работе бонусной программы.

При регистрации покупателю необходимо указать свое полное имя, пол, номер телефона и адрес электронной почты. При этом при каждом из способов регистрации он может опционально подключить возможность получения информации по СМС или на электронную почту. Каналы валидируются клиентом посредством смс-оповещения с кодом подтверждения. Вся эта информация аккумулируется на серверах предприятия и проходит через информационную систему поддержки программ лояльности, а в последствии может быть использована в маркетинговых целях, к чему мы вернемся далее. При этом административно (правилами программы лояльности) и, соответственно, технически установлено ограничение, побуждающее новоиспеченного владельца бонусной карты пройти регистрацию в программе – неактивированные бонусные карты заблокированы для списания баллов.

В настоящий момент эти персональные данные не применяются в целях маркетинга, то есть персонализации предложений, которую развивают федеральные игроки, нет. До сих пор свою роль в продвижении акционного ассортимента играют лишь печатные каталоги, доставляемые в почтовые ящики и доступные в торговых точках. Такое положение дел на фоне игроков вроде “Ленты”, давно предлагающей владельцам карт программ лояльности индивидуальные бонусы, например, перед днем рождения, или “Пятерочки”, экспериментирующей с форматом программ лояльности, соединяя бонусный и скидочный форматы, является достаточно застывшим по ряду причин, хотя технологически и управленчески подобные возможности есть.

Прежде всего, стоит отметить, что ритейлером реализован и поддерживается личный кабинет покупателя на сайте компании, где доступна полная история покупок (когда была совершена покупка и на какую сумму, в том числе и в разрезе каждой номенклатурной единицы), начисления и списания баллов. Кроме того, у предприятия есть контактная информация клиента. Таким образом, мы знаем, что покупает наш лояльный клиент, кто он и как установить с ним контакт.

Помимо прочего, все товары в используемой ERP-системе поделены на группы, категории и подкатегории, а

если говорить о политике ценообразования, то акции также подразделяются на некоторые совокупности.

СУТЬ МОДЕЛИ ДИНАМИЧЕСКОЙ МОТИВАЦИИ

Все это при совместном рассмотрении позволяет представить переход существующей программы лояльности с урезанным функционалом к модели динамической мотивации покупателя.

Ее суть можно обозначить следующим образом: благодаря тому, что каждый товар привязан к своей ассортиментной группе, а также наличию уже собранных и ежедневно пополняемых массивов транзакционных данных, при совершении очередной покупки клиентом с использованием бонусной карты становится возможным однозначно определить категорию и подкатегорию товара, и «привязав» эту информацию к клиенту, проанализировав частоту его покупок, содержимое его продуктовой корзины и корзин «похожих» покупателей, в дальнейшем предложить ему релевантный акционный товар из этой категории или смежной ей.

Данные категории и подкатегории, а также товары в них можно автоматически приоритезировать на уровне ERP-системы, чтобы выполнять рассылку наиболее потенциально покупаемых товаров (например, по размеру скидки, общей частоте покупок товара и так далее). К условиям автоматической установки приоритетов можно отнести, например, среднестатистическую частоту покупок данного товара, частоту покупки товаров этой категории конкретно этим покупателем, размер скидки и другие показатели.

Динамичность метода заключается в его регулярном повторении, что позволяет с каждой покупкой предлагать все более точные и наиболее интересные для покупателя варианты.

Если формализовать описанный процесс, то его можно представить следующими шагами:

1. Лояльный клиент совершает покупку с применением карты;
2. История покупок этого клиента пополняется;
3. Анализируем историю покупок на предмет частоты приобретения того или иного товара, изменения величины среднего чека, частоты посещения нашего магазина;
4. Соотносим наиболее релевантные товары с категориями товаров, историей покупок и текущим акционным ассортиментом;
5. Предсказываем дату следующего посещения;
6. В подходящий момент высылаем релевантное запросам клиента индивидуальное предложение.

ТЕХНОЛОГИЧЕСКАЯ СТОРОНА

В основу метода предсказания обозначенных аспектов (например, предполагаемой даты следующего посещения) могут лечь широко признанные в наши дни методы машинного обучения, например, нейронные сети. Обучать их можно на основе имеющихся данных по истории покупок как по конкретному (целевому) покупателю, чью дату следующего посещения мы попытаемся предсказать, так и по другим покупателям со схожим «портретом» (сочетание возраста, пола и индивидуальной потребительской корзины) для получения большей точности предсказания.

В целях поиска наиболее релевантного для клиента товара можно использовать метод коллаборативной фильтрации, то есть находить и анализировать максимально похожих по потребительскому поведению на нашего «целевого» покупателя (адресата рассылки) клиентов и предложить ему то, что обычно покупает эта группа [3]. Кроме того, полезным будет применение метода ассоциативных правил для обнаружения искомых связей в выше указанных данных (например, при отталкивании от предыдущей истории покупок нашего «целевого» покупателя) [4, 5].

Компания располагает огромными массивами данных по историям покупок и информации о клиентах, предоставляя достаточно широкие возможности для успешного применения алгоритмов и методов из областей статистики и машинного обучения.

Выполнять саму рассылку возможно посредством SMS-сообщений на номер, привязанный к программе лояльности, либо через бизнес-чаты в приложении Viber. Последним решением сейчас пользуются многие ритейлеры (например, и «Пятерочка»), в том числе не только из сферы продовольствия (например, «Связной»).

ЭКОНОМИЧЕСКАЯ СТОРОНА

С экономической точки зрения для измерения эффективности предложенной модели можно рассматривать следующие элементы взаимодействия лояльного покупателя с розничной сетью: средний чек, количество посещений за период, процентное соотношение акционного и неакционного ассортимента в корзине и другие. Ожидаемым эффектом применения модели служит повышение уровня лояльности покупателей, заключающееся в росте перечисленных показателей как на уровне индивидуального покупателя, так и на уровне групп потребителей с симильными характеристиками.

В плане экономической составляющей уже сейчас можно рассмотреть потенциальный эффект. Сейчас смена акционного ассортимента происходит каждые 2 недели, что сопровождается версткой, печатью и распространением бумажных каталогов. Перейдя к персональным предложениям, присылаемым лояльным клиентам в электронном виде, развивая модель динамической мотивации покупателя, можно будет сократить или полностью отказаться от бумажных каталогов, распространяемых в торговых точках и поквартирно. По данным, полученным от отдела маркетинга, в среднем в месяц на полный цикл создания каталога (тираж порядка 300 тысяч копий) и его распространение (три четверти тиража разносится по почтовым ящикам) уходит порядка 3,5 миллионов рублей в месяц.

Так, предложенная к реализации модель динамической мотивации может послужить поводом к рассмотрению возможности отказа от печатных каталогов в пользу индивидуального подхода к предоставлению информации об акционном ассортименте. В свете популяризации экологического движения в наши дни подобный шаг и соответствующий этому инфоповод может послужить еще одним маркетинговым ходом и конкурентным

преимуществом относительно более успешных игроков, занимающих большую долю рынка.

ЗАКЛЮЧЕНИЕ

Таким образом, описанная модель действительно включает в себе принцип динамичности – постоянное, регулярное применение аналитических методов и алгоритмов, затронутых в работе, способно стать основой новой концепции отношения ритейлинговой сети и ее лояльного покупателя, увеличивая вероятность его скорого возвращения за счет своевременного предоставления релевантного для него предложения.

Резюмируя, считаем нужным отметить, что у рассмотренной ритейл-сети есть все технологические предпосылки к скорой реализации описанной модели динамической мотивации лояльных покупателей, а также к последующему тестированию, исследованию и анализу ее эффективности. Отсутствие решений по индивидуализации предложений тормозит развитие компании, что может привести к большому отставанию от более крупных конкурентов; в свою очередь подобный кейс в случае успешной реализации может стать

прорывным не только как для регионального игрока, но и стать заметным на федеральном уровне.

ЛИТЕРАТУРА

1. “Магнит” откажется от гипермаркетов в пользу супермаркетов формата “Семейный+”. Retailer – URL: <https://retailer.ru/magnit-otkazhetsja-ot-formata-gipermarketov/> (дата обращения 11.01.2020).
2. Карточная лояльность. Исследовательский холдинг Romir – URL: <https://romir.ru/studies/kartochnaya-loyalnost> (дата обращения 12.01.2020).
3. Коллаборативная фильтрация. Хабр – URL: <https://habr.com/ru/post/150399/> (дата обращения 14.01.2020).
4. Ассоциативные правила. BaseGroup Labs – URL: <https://basegroup.ru/deductor/function/algorithm/association-rules> (дата обращения 14.01.2020).
5. Анализ рыночной корзины (Market Basket Analysis). Loginom Wiki – URL: <https://wiki.loginom.ru/articles/market-basket-analysis.html> (дата обращения 14.01.2020).

The assessment of the technological potential of improvement of the consumer goods retailer' loyalty program

O.V. Nevolin, Y.V. Petrichenko

Chelyabinsk State University

Chelyabinsk, Russia

nevoleg96@gmail.ru, yulia-c@yandex.ru

Abstract. Nowadays, consumer goods retailers' loyalty programs are one of the main tools of customer satisfaction' managing. It was confirmed by our previous research. This article describes possible way of existing loyalty program' improvement from technical, managerial and economic points of view using local retailer as an example. It allowed us to propose the transition to the dynamic consumer motivation model, which is also described in this article.

Keywords: consumer goods retail, loyalty program, promotion, consumers, trading network, dynamic motivation.

Веб-ориентированный инструментарий микроскопического моделирования транспортных потоков

Пупынин К.В., Головнин О.К.

Кафедра информационных систем и технологий
Самарский университет
Самара, Россия
kir730@yandex.ru

Аннотация. В работе представлен программный инструментарий для моделирования дорожного движения. Преимущества инструментария определяются использованием программной реализации, основанной на веб-ориентированном подходе, который не требует установки на конечный пользовательский компьютер каких-либо специализированных программ и утилит, а также монтажа высокопроизводительного оборудования. Для повышения эффективности моделирования в инструментарии используются слои: на первом неизменном статическом слое находится дорожная сеть; на втором динамическом слое находятся транспортные средства. Транспортные средства перерисовываются на каждом этапе моделирования; дорожная сеть перерисовывается по требованию. Программный инструментарий основан на симуляторе транспортных потоков SUMO с использованием протокола TraCI для передачи данных. В статье описаны компонентные и структурные схемы архитектуры инструментария и приведены разработанные алгоритмы обработки данных. Техника проектирования, используемая для повышения эффективности инструментария, необходимого для моделирования трафика, реализуется с помощью паттернов. Представлен программный интерфейс инструментария с возможностью изменения параметров моделирования во время симуляции. Выполнено моделирование на участке дорожной сети г.о. Самара, Россия. На основании результатов проведенного моделирования выявлены ограничения, накладываемые на размер моделируемой улично-дорожной сети и количество транспортных средств: общая площадь дорожной сети не должна превышать 3 кв. км; количество элементов сети не должно превышать 10 тыс., общее количество транспортных средств не должно превышать 5 тыс.

Ключевые слова: микроскопическое моделирование, транспортный поток, движение, веб-ориентированный подход, SUMO.

ВВЕДЕНИЕ

С ростом числа автомобилей на улично-дорожной сети растет количество и протяженность транспортных заторов, кроме того, снижается безопасность дорожного движения. Для избавления городов от пробок и улучшения организации дорожного движения существует немало эффективных способов [1]. Например, интеллектуальные транспортные системы (ИТС) предоставляют конечным пользователям большую информативность и безопасность за счет использования инновационных разработок в

регулировании транспортных потоков и моделировании транспортных систем [2]. Такие системы позволяют значительно повысить уровень взаимодействия между участниками движения [3], а также проводить онлайн моделирование дорожного движения на участках улично-дорожной сети с целью повышения качества управления [4]. Ключевой частью ИТС являются VANET-сети, которые используют технологию беспроводной связи, с помощью которой объединяются в группы участники движения [5]. Кроме того, современные системы используют электронные карты местности и географические информационные системы [6]. Однако до внедрения ИТС и VANET сетей на дорогах необходимо проведение реалистичных компьютерных симуляций с использованием комбинации имитации городского трафика и моделирования сети [7].

На данный момент разработано довольно большое количество мощных систем моделирования, которые поддерживают моделирование дорожных сетей больших размеров: MATSim [8], VisSim [9], MOTUS [10].

Каждый из этих продуктов во время моделирования основывается на одной из групп имитационных моделей. Как правило, выделяется 3 группы имитационных моделей, которые обычно различаются между собой степенью детализации проводимого анализа: макроскопические, мезоскопические и микроскопические. Все типы имитационных моделей хорошо описаны в [11].

Основная цель работы состоит в повышении доступности систем моделирования для конечных пользователей. В связи с этим, реализуется веб-ориентированный инструментарий для проведения транспортного моделирования, который не требует установки какого-либо программного обеспечения на конечный компьютер и не зависит от операционной системы. Однако наиболее важным является то, что пользователю не потребуется мощное и дорогостоящее оборудование, так как моделирование будет происходить на удаленном сервере, с которым будет взаимодействовать клиентская часть инструментария. Таким образом, пользователю непосредственно в браузере доступно моделирование движения трафика с настройкой параметров моделирования как до его начала, так и непосредственно в процессе.

При разработке инструментария выбрана микроскопическая имитационная модель, поскольку она позволяет добиться наибольшей детализации при моделировании за счет того, что каждое транспортное средство в транспортном потоке рассматривается отдельно и моделируется собственным уравнением, функцией его положения, скорости и ускорения [12]. В качестве программы моделирования выбран микроскопический симулятор с открытым исходным кодом SUMO, который позволяет моделировать различные виды транспорта, движение пешеходов, а также поддерживает симуляцию дорожных сетей большого размера [13]. Инструментарий реализован с использованием подхода, описанного в [14].

МИКРОСКОПИЧЕСКАЯ ИМИТАЦИОННАЯ МОДЕЛЬ ТРАНСПОРТНОГО ПОТОКА

Рассматриваемый инструментарий будет использовать микроскопическую модель транспортного потока, который движется по различным полосам движения $\tilde{l}_i \in \tilde{L}$. Согласно микроскопическому подходу, каждый участник движения $\tilde{v}_i \in \tilde{V}$ рассматривается отдельно.

Объекты класса «Vehicle» $\tilde{v}_i^{l_i} = \{V_j^{l_i}\}$ описываются следующими атрибутами: уникальный идентификатор $id^V \in N = \{1, 2, \dots, n\}$, тип транспортного средства $vType^V \in VType^V$, маршрут $route^V \in Lanes = \{l_1, l_2, \dots, l_n\}$, $l_i \in Lane$, скорость $speed^V \in R^+$, текущая линия маршрута $lane^V \in Lane$, текущая полоса движения $laneIndex^V \in N = \{1, 2, \dots, n\}$, текущая позиция $position^V = \{x: double, y: double\}$, пройденная дистанция $distance^V \in R^+$.

«Vehicle Type» представлен следующими атрибутами: уникальный идентификатор $id^{vType} \in N = \{1, 2, \dots, n\}$, класс транспортного средства $vClass^{vType} \in VClass^{vType} = \{passenger, evehicle, truck, bus, ebus, etc\}$, цвет $color^{vType} \in Color [R, G, B]$, геометрия $shape^{vType} \in Polygon = \{p_1, p_2, \dots, p_n\}$, где $p_i = \{x: double, y: double\}$, длина $length^{vType}$, модель смены полосы движения $laneChangeModel^{vType} \in string$, модель следования за лидером $carFollowModel^{vType} \in string$.

«Car Following Model» описывается параметрами: $\sigma^{vType} \in R^+ = [0, 1]$, $\tau^{vType} \in R^+$.

«Lane» является набором следующих атрибутов: уникальный идентификатор $id^l \in N = \{1, 2, \dots, n\}$, полоса движения $index^l \in N = \{1, 2, \dots, n\}$, максимально разрешенная скорость $speed^l \in R^+$, длина $length^l \in R^+$, геометрия $shape^l \in Polygon = \{p_1, p_2, \dots, p_n\}$, где $p_i = \{x: double, y: double\}$.

АРХИТЕКТУРА ИНСТРУМЕНТАРИЯ

Рисунок 1 демонстрирует архитектуру инструментария. Конечный пользователь может загружать дорожные сети из открытых источников, таких как Open Street Map или выбрать из предложенных. После того как карта была передана системе моделирования SUMO и обработана, пользователь может нажать кнопку «Start» для запуска симуляции. После чего будет запущено моделирование и сервер приложения будет использовать TraCI протокол для обмена необходимой информацией между системой моделирования и клиентской частью приложения в

браузере. Пользователь может приостанавливать моделирование или изменять его параметры в реальном времени, сервер приложений будет реагировать на действия пользователя и отправлять соответствующие команды SUMO, которая, в свою очередь, будет изменять параметры моделирования. Таким образом, пользователю не требуется установки сторонних программ на своей компьютер.

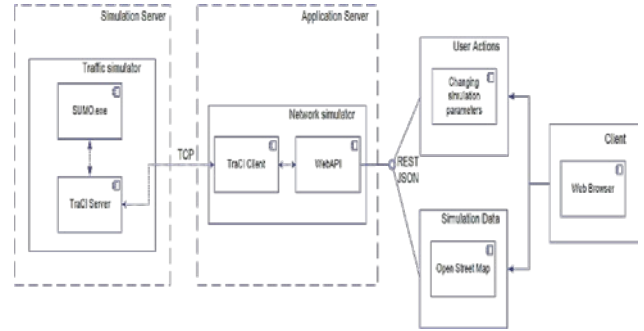


Рис. 1. Диаграмма компонентов инструментария

Во время разработки инструментария обнаружена проблема чрезмерного потребления памяти для хранения большого количества однотипных объектов во время моделирования. Более того, веб-приложение не сильно зависит от идентичности объекта, и большая часть внутреннего состояния некоторых объектов может быть заменена внешним состоянием. Таким образом, чтобы уменьшить нагрузку на оперативную память компьютера, используется паттерн проектирования «Легковес». Суть использования данного паттерна состоит в том, чтобы вынести повторяющееся состояние из объекта и заменить его ссылкой на объект, в котором хранится внешнее состояние, которое может быть одинаковым для большого количества объектов.

Рассмотрим класс Vehicle: такие его параметры как vClass, color, sprite, length можно вынести как внешнее состояние. Перенесём эти параметры из класса Vehicle в другой класс, который будет хранителем повторяющегося внешнего состояния, назовём его VehicleType. Процесс переноса показан на рисунке 2.

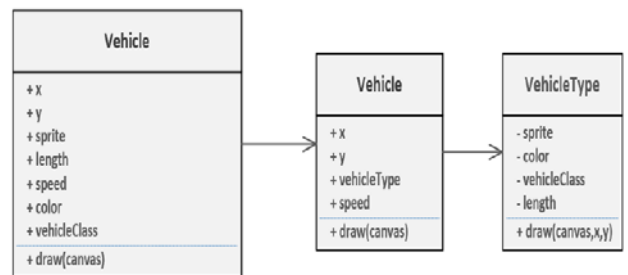


Рис. 2. Перенос внешнего состояния класса Vehicle

Схема создания нового экземпляра класса Vehicle представлена на рисунке 3.

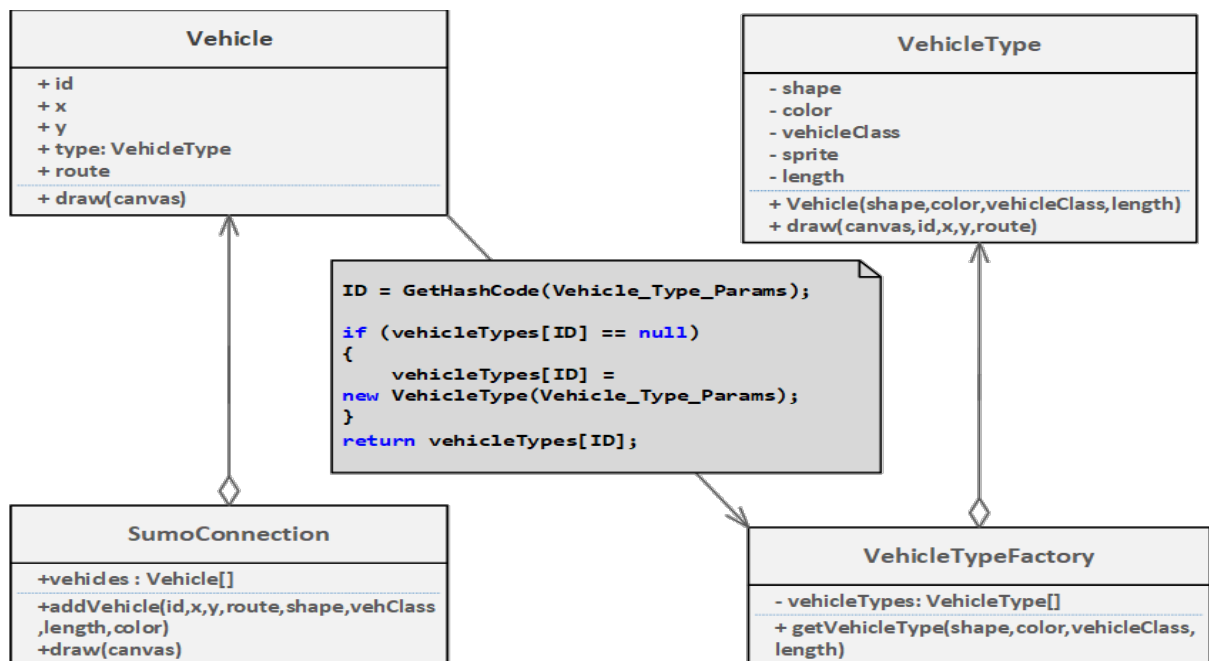


Рис. 3. Структурная схема паттерна «Легковес»

Благодаря этому подходу нагрузка на оперативную память была значительно уменьшена (с 3.1 GB до 2 MB).

Vehicle:

ID: 4B, length: 4B, color: 4B, speed: 4B, vClass: 4B, position: 8B, sprite: 30KB

Всего: $\sim 31KB * 100000 \approx 3.1 GB$.

Vehicle:

ID: 4B, position: 8B, speed: 4B, vehicleTypeID: 4B

VehicleType:

vehicleTypeID: 4B, color: 4B, length: 4B, vClass: 4B, sprite: 30 KB

Всего: $46KB + (20B * 100000) \approx 2 MB$.

РЕЗУЛЬТАТЫ

Разработан веб-ориентированный инструмент для проведения моделирования транспортных потоков. Интерфейс инструментария представлен на рисунке 4.

На рисунке 5 представлено меню, которое позволяет контролировать моделирование в реальном режиме времени. Используя кнопки на панели, можно: открыть симуляцию;

- открыть дорожную сеть;
- запустить симуляцию загруженной дорожной сети;
- остановить симуляцию;

- произвести один шаг симуляции;
- найти на карте любой из объектов, такой как: дуги, полигоны, транспортные средства, пересечения и т.д.;
- редактировать визуализацию и окно просмотра.

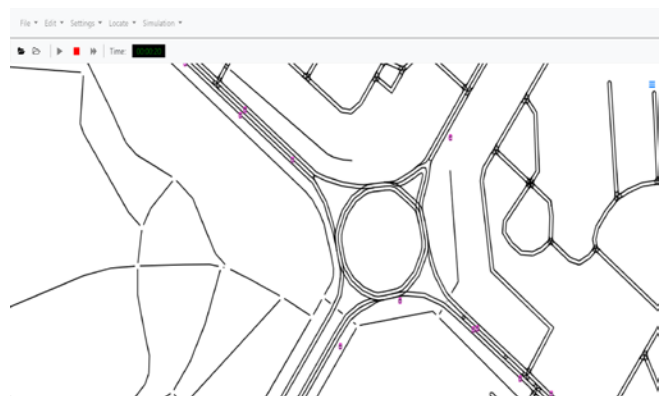


Рис. 4. Интерфейс инструментария

Проведено моделирование с шагом 500 мс на участке улично-дорожной сети г.о. Самара, Россия. Моделирование показало эффективность инструментария при работе на площади до 3 кв. км с числом транспортных средств не более 5000, при этом количество элементов сети не должно превышать 10 тыс.

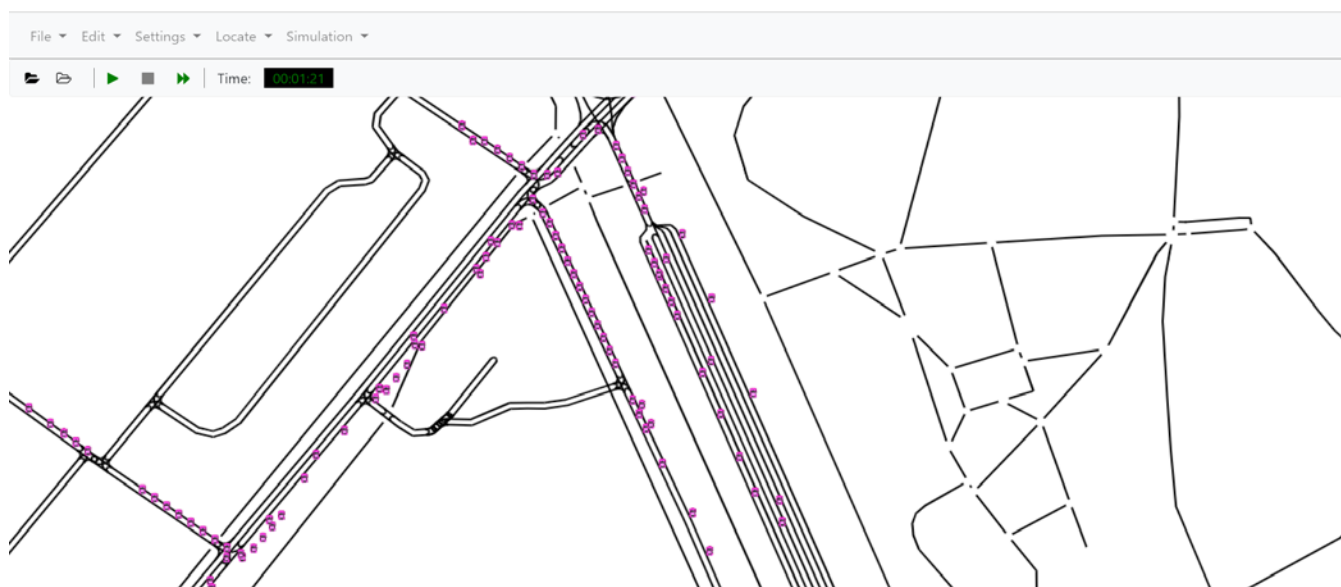


Рис. 5. Моделирование трафика

ЗАКЛЮЧЕНИЕ

В целях повышения эффективности моделирования трафика разработан инструментальный для моделирования транспортных потоков на основе микроскопической имитационной модели и системы мультимодального моделирования SUMO. Созданный инструментальный позволяет моделировать дорожные сети с высокой степенью детализации без производительного оборудования и специального программного обеспечения на конечном компьютере пользователя – моделирование в инструментальной контролируется из веб-браузера.

Достигнутые результаты показали способность моделировать участок дорожной сети, состоящий из нескольких перекрестков. Дальнейшая работа будет направлена на увеличение размеров моделируемой улично-дорожной сети и улучшение качества визуализации.

ЛИТЕРАТУРА

1. Sumalee A. Smarter and more connected: Future intelligent transportation system / A. Sumalee, H. W. Ho // IATSS Research. – 2018. – Vol. 42(2). – P. 67-71.
2. Lin Y. Intelligent transportation system (ITS): Concept, challenge and opportunity / Y. Lin, P. Wang, M. Ma // In 2017 IEEE 3rd International Conference on Big Data Security on Cloud (BigDataSecurity). – 2017. – P. 167-172.
3. Mathew E. Intelligent Transport Systems and Its Challenges / E. Mathew // International Conference on Advanced Intelligent Systems and Informatics. – 2019. – P. 663-672.
4. Hoogendoorn S.P. State-of-the-art of vehicular traffic flow modelling / S.P. Hoogendoorn, P. H. Bovy // Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering. – 2001. – Vol. 215(4). – P. 283-303.
5. Cherkaoui B. Road traffic congestion detection in VANET networks / B. Cherkaoui, A. Beni-Hssane, M. El

Fissaoui, M. Erritali // Procedia Computer Science. – 2019. – Vol. 151. – P. 1158-1163.

6. Fedoseev A.A. An approach for GIS-based transport infrastructure model synthesis on the basis of hyperspectral information / A.A. Fedoseev, O.K. Golovnin, T.I. Mikhееva // Procedia Engineering. – 2017. – Vol. 201. – P. 363-371.

7. Song D. Multi-vehicle tracking using microscopic traffic models / D. Song, R. Tharmarasa, G. Zhou, M. C. Florea, N. Duclos-Hindie, T. Kirubarajan // IEEE Transactions on Intelligent Transportation Systems. – 2018. – Vol. 99. – P. 1-13.

8. Horni A. Introducing matsim / A. Horni, K. Nagel, K. W. Axhausen // The multi-agent transport simulation MATSim. Ubiquity Press. – 2016. – P. 3-7.

9. Yang Z. H. Research on Microscopic Traffic Simulation Model in VISSIM / Z. H. Yang, L. I. Ge, X. Zhang, and Peng, Wang // DEStech Transactions on Computer Science and Engineering. – 2019.

10. Schakel W. A modular approach for exchangeable driving task models in a microscopic simulation framework / W. Schakel, B. van Arem, H. van Lint, G. Tamminga // 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013). – 2013. – P. 565-571.

11. Azlan N.N.N. Overview of application of traffic simulation model / N.N.N. Azlan, M.M. Rohani // MATEC Web of Conferences. – Vol. 150. – 2018. – P. 03006.

12. Wang C. A combined use of microscopic traffic simulation and extreme value methods for traffic safety evaluation / C. Wang, C. Xu, J. Xia, Qian, L. Lu // Transportation Research Part C: Emerging Technologies. – 2018. – Vol. 90. – P. 281-291.

13. Lopez P.A. Microscopic traffic simulation using Sumo / P.A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.P. Flötteröd, R. Hilbrich, E. Wießner // 2018 21st International Conference on Intelligent Transportation Systems (ITSC). – 2018. – P. 2575-2582.

A Microscopic Traffic Simulation Web Toolkit

K.V. Pupynin, O.K. Golovnin
Department of Information Systems and Technologies
Samara University
Samara, Russia
e-mail: kir730@yandex.ru

Abstract. There is presented a software toolkit for urban traffic simulation. The advantages of the toolkit are defined by using software implementation based on a web-oriented approach that does not require the installation of any programs as well as high-performance hardware. In order to increase the modeling efficiency, the toolkit uses layers: on the first unchanged static layer is a road network; on the second dynamic layer are vehicles. Vehicles are redrawn at each modeling step; the road network is redrawn on demand. The software toolkit is based on SUMO traffic flow simulator by TraCI protocol utilization for data transfer. The paper describes a component and structural diagrams of toolkit architecture and contains developed data processing algorithms. The technique used to optimize hardware required for traffic modeling are implemented through the use of

design patterns. The program interface of the toolkit for changing modeling parameters during the simulation is presented. Modeling on a section of the road network of Samara, Russia was carried out. Based on the results of this modeling, restrictions are imposed on the size of the simulated road network and on the number of vehicles: total area of road network should not exceed 3 square km, while the number of road network elements should not exceed 10000, the total number of vehicles should not exceed 5000.

Keywords: microscopic simulation; traffic flow; mobility; web-oriented approach; SUMO.

Проектирование автоматизированной системы критериальной оценки научно-исследовательской работы магистрантов с использованием методологии SADT и DFD

Бакановская Л. Н., Глебова Н. С.

ФГБОУ ВО "Тюменский индустриальный университет"
город Тюмень, Российская Федерация -
e mail: l.bakanovskaya@gmail.com

Аннотация. В работе рассмотрен структурный анализ автоматизированной системы критериальной оценки научно-исследовательской работы магистрантов. Особое внимание уделено движению информационных потоков между магистрантом, научным руководителем, руководителем направления и заведующим кафедрой. Разработанные критерии оценки, используемые в автоматизированной системе, позволят повысить качество выполняемых научных исследований и магистерских диссертаций, а также максимально приблизить будущих выпускников к решению реальных практических задач.

Ключевые слова: методология SADT, автоматизированная система, структурный анализ, критериальная оценка, научно-исследовательская деятельность.

ВВЕДЕНИЕ

Активизация научно-исследовательской деятельности (НИД) студентов-магистрантов выдвигает ряд требований перед учебным заведением, которые положительно влияют на повышение качества образовательного процесса. Исследования в области совершенствования и развития мотивационной составляющей к научной деятельности у магистрантов, с различных точек зрения, преподавателями вузов ведутся постоянно. Актуальность и необходимость данной работы подтверждают многочисленные публикации в отечественных и зарубежных изданиях [1-23].

Для проведения структурного анализа в НИД магистранта выделены следующие особенности:

1. работа над индивидуальной темой диссертации ведётся на протяжении всего периода обучения (4 семестра);
2. мотивация магистрантов с первого семестра к публикации статей и докладов;
3. апробация темы исследования на конференциях;
4. обсуждение магистрантами результатов работы на научных семинарах, круглых и дискуссионных столах кафедры;
5. проработка методик исследования;
6. апробация полученных результатов на практике;
7. приобщение к процессу познания и получения новых практических навыков и собственных результатов

исследований, а не к потребительскому использованию широко известных методов;

8. ощущение ответственности за принятые самостоятельные решения;

9. возможность работы в составе научного коллектива преподавателей кафедры, участие в грантах и конкурсах;

10. постоянный контроль научного руководителя за работой магистранта и полученными им результатами, начиная с первого семестра обучения.

Содержание НИД, выполняемой магистрантом, для разработки автоматизированной системы удовлетворяет следующим требованиям [24]:

- а) является формализуемым;
- б) имеет качественные и количественные критерии оценки;
- в) разработана математическая модель суммарной оценки по каждому критерию;
- г) ориентирована на объективные (воспроизводимые) методы контроля по всем критериальным показателям.

МЕТОДОЛОГИЯ ИССЛЕДОВАНИЯ

Создание автоматизированной системы критериальной оценки НИД магистранта требует четких представлений о предметной области, действиях, которые необходимо предпринять руководителю и магистранту, замечания и способы их устранения, требующиеся для этого ресурсы и т.д., которыми будет оперировать система. Таким образом, формируется целостная картина деятельности магистранта: от моделей организации работы в семестре до работы над диссертацией в целом.

Важным этапом проектирования автоматизированной системы является структурный анализ. Аналитическая модель, созданная с использованием методологии SADT (Structured Analysis and Design Technique), [25, 26] позволяет провести структурный анализ построения функциональной модели автоматизированной системы критериальной оценки научно-исследовательской работы магистранта. На рис. 1 представлена модель верхнего уровня, где показаны основные функции системы.

Модель первого уровня показывает:

1. управляющую информацию, которая заложена в систему (стрелки сверху);
2. входящую информацию (блок слева) которая подвергается обработке;
3. механизмы, выполняющие управление системой (блок снизу);

Л.Н. Бакановская, Н.С. Глебова

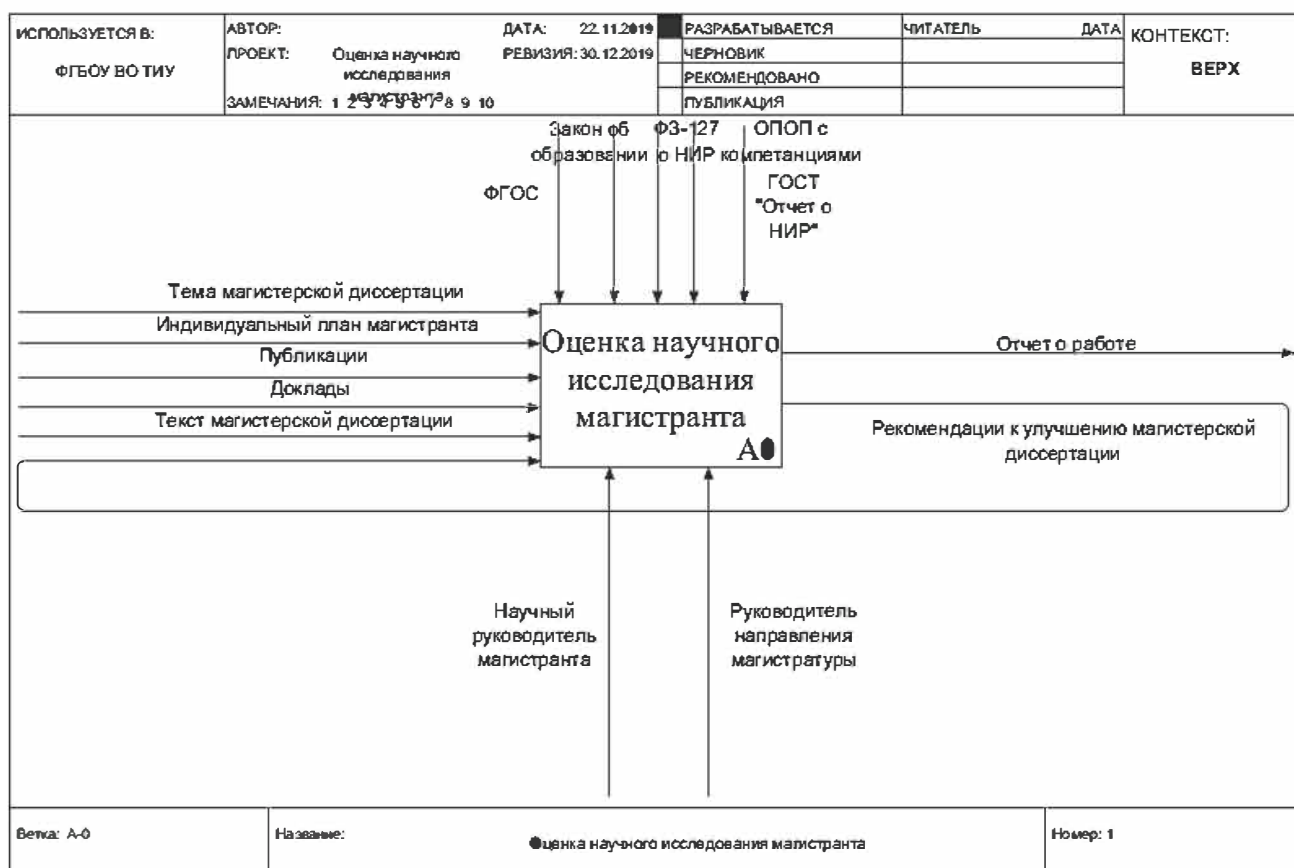


Рис. 1 – Контекстная диаграмма A0
«Критериальная оценка НИД магистранта»

4. результаты обработки, выходная информация (стрелки справа).

В процессе детализации контекстной диаграммы выделены локальные функции системы (рис. 2).

Каждая из указанных функций получает входящую информацию:

- а) тема магистерской диссертации;
- б) индивидуальный план магистранта;
- в) количество и наименование публикаций и очных выступлений на различных мероприятиях.

Каждый вид работ из входящей информации оценивается научным руководителем по критериям [24]:

1. Актуальность темы;
2. Уровень прогрессивности решаемой проблемы;
3. Степень разработанности проблемы;
4. Цель работы;
5. Уровень формирования задач;
6. Практическая значимость работы;
7. Теоретическая значимость работы;
8. Уровень выводов;
9. Анализ литературы по требованию ГОСТ;
10. Использование научно-технической документации;
11. Использование ГОСТов.

Все представленные в модели функции связаны информационными потоками. Научный руководитель формирует отчет для руководителя направления (рис. 3). Руководитель направления формирует общий отчет по всем научным руководителям для заведующего кафедрой. Для анализа информационных потоков с точки зрения

получения, анализа, обработки и хранения информации, в работе разработана диаграмма потоков данных (DFD – data flow diagrams).

На основе построенной DFD-модели видно, какие меры по повышению эффективности НИД магистранта необходимо предпринять, возможные методы оптимизации процесса контроля за подготовкой диссертации [27-29]. Каждый семестр магистрант получает от научного руководителя замечания, которые в следующем семестре он может устранить.

ЗАКЛЮЧЕНИЕ

Разработанная методика критериальной оценки позволяет повысить качество выполняемых НИД и магистерских диссертаций [24], а также максимально приблизить будущих выпускников к решению реальных практических задач. Необходимо подчеркнуть, что реализация требуемых профессиональных навыков имеет огромное значение для целостности специалиста как профессионала на примере научной деятельности. Так, в публикационной активности магистрантов проявляются компетенции и стимулирующая составляющая (получение более высоких баллов). Работа с автоматизированной системой в интерактивном режиме позволит магистрантам своевременно отслеживать свои результаты, а руководителям предоставит удобную платформу оценивания НИД в единой системе автоматизированного управления университетом.

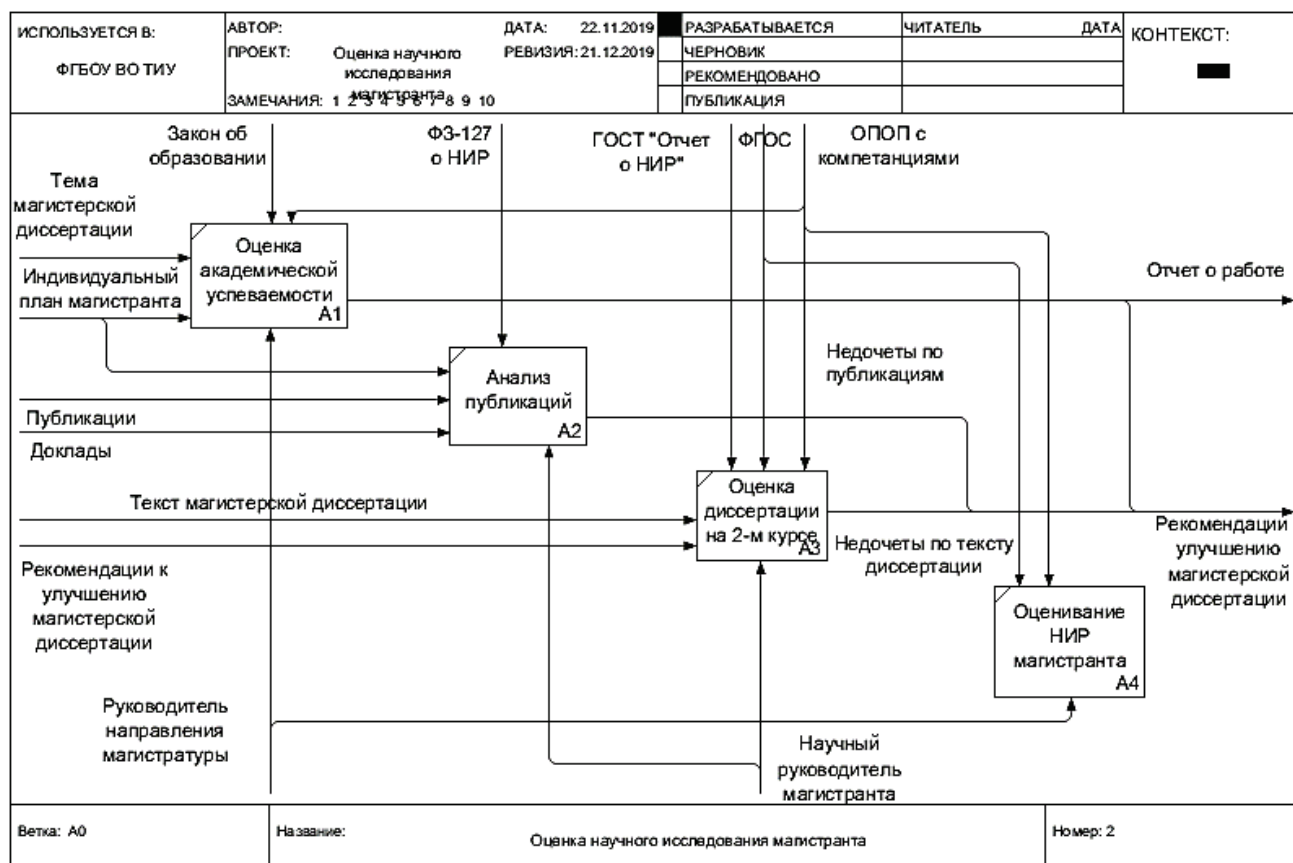


Рис. 2 – Диаграмма первого уровня декомпозиции

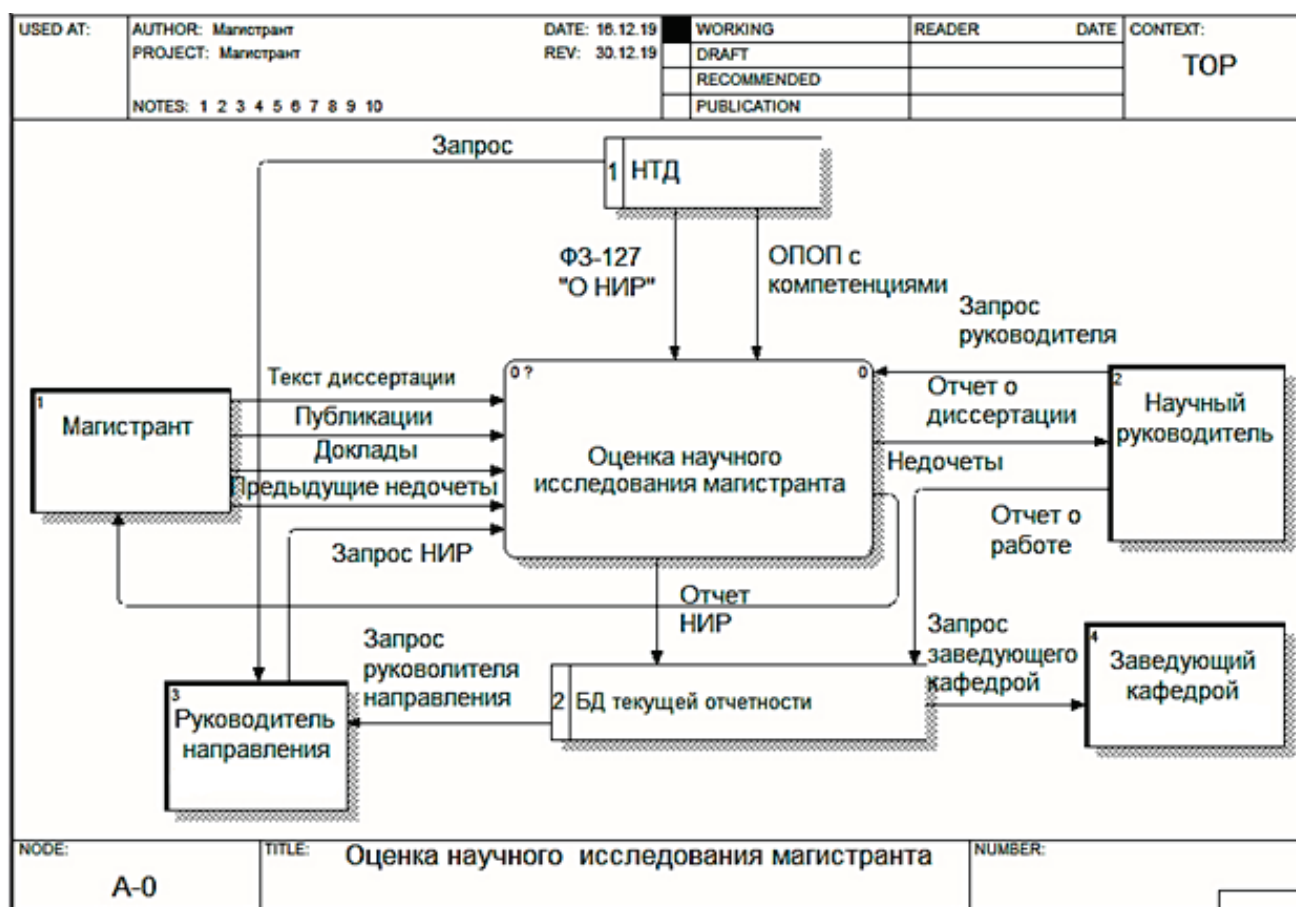


Рис. 3 – Диаграмма потоков данных автоматизированной системы

Л.Н. Бакановская, Н.С. Глебова

Проектирование автоматизированной системы критериальной оценки научно-исследовательской работы магистрантов с использованием методологии SADT и DFD

ЛИТЕРАТУРА

1. Панова Е. А. Организация научно-исследовательской работы студента при обучении в магистратуре / Е. А. Панова, О. М. Гуртовенко, Г. Г. Левкин. – Текст : непосредственный // Инновационная экономика и общество. – 2017. – № 4(18). – С. 104-112;
2. Вовлеченность студентов в научно-исследовательскую работу и ее перспективы как элемента самостоятельной работы студента-магистранта / Д. А. Хашхожева, А. Ю. Паритов, М. Т. Гелястанова [и др.]. – Текст : непосредственный // Современные проблемы науки и образования. – 2017. – № 3. – С. 114-121;
3. Запорожцев А. В. Системный подход к научно-исследовательской работе студентов магистратуры / А. В. Запорожцев. – Текст : непосредственный // Инновационные технологии в образовательной деятельности: материалы Всероссийской научно-методической конференции. – Нижний Новгород, 2019. – С. 77-82;
4. Панюкова А. А. Организация научно-исследовательской работы студента в вузе на основе системного подхода / А. А. Панюкова, Дударева В. И. – Текст : непосредственный // Современные информационные технологии и ИТ-образование: сборник научных трудов II Международной научной конференции и XII Международной научно-практической конференции. Под редакцией В.А. Сухомлина. – Москва, 2017. – С. 97-102;
5. Рудакова Е. В. Роль научно-исследовательской работы при подготовке программиста-практика / Е. В. Рудакова. – Текст : непосредственный // Актуальные проблемы социально-гуманитарного и научно-технического знания. – 2015. – № 2(5). – С. 3-4;
6. Сафина Т. А. Научно-исследовательская работа как основополагающий фактор подготовки ИТ-специалиста / Т. А. Сафина. – Текст : непосредственный // Вестник межрегионального открытого социального института. – 2017. – № 1(5). – С. 30-35;
7. Бирюкова И. В. Научно-исследовательская работа студентов как первый шаг создания инновационного научного продукта / И. В. Бирюкова. – Текст : непосредственный // Наука вчера, сегодня, завтра: теория и практика: материалы Международного электронного Симпозиума. – Махачкала, 2015. – С. 48-52;
8. Китаев А. Я. Из опыта деятельности выпускающей кафедры по организации научно-исследовательской работы студентов / А. Я. Китаев, А. Б. Храмцов. – Текст : непосредственный // Проблемы управления речными бассейнами при освоении Сибири и Арктики в контексте глобального изменения климата планеты в XXI веке: сборник докладов XIX Международной научно-практической конференции. – Тюмень, 2017. – С. 139-148;
9. Багаев И. З. О некоторых современных проблемах научно-исследовательской работы студентов в России / И. З. Багаев. – Текст : непосредственный // Устойчивое развитие науки и образования. – 2019. – № 3. – С. 148-157;
10. Апанасенок А. В. Научно-исследовательская работа студентов в современном университете: актуальные вызовы / А. В. Апанасенок, Н. П. Шульгина, Р. К. Боженкова. – Текст : непосредственный // Известия Юго-Западного государственного университета. Серия: Лингвистика и педагогика. – 2016. – № 2(19). – С. 123-130;
11. Гладковский В. И. Научно-исследовательская работа студентов как фактор повышения качества университетского образования / В. И. Гладковский, А. И. Пинчук, Т. Л. Кушнер. – Текст : непосредственный // Личность и общество. – 2019. – № 2(2). – С. 28-30;
12. Муратова Г. К. Организация научно-исследовательской работы студента в вузе / Г. К. Муратова. – Текст : непосредственный // Вопросы педагогики. – 2019. – № 4-1. – С. 110-113;
13. Маметьева О. С. Научно-исследовательская работа студентов вуза: результативность и проблемы организации / О. С. Маметьева, Н. Г. Супрун, Д. А. Халикова. – Текст : непосредственный // Современные проблемы науки и образования. – 2018. – № 1. – С. 8-17;
14. Любимова Г. А. Научно-исследовательская работа студентов как основной компонент профессиональной подготовки в вузе / Г. А. Любимова, Н. А. Громцева. – Текст : непосредственный // World science: problems and innovations: сборник статей XIX Международной научно-практической конференции. : в 3 ч., Пенза, 2018. – С. 248-250;
15. Арсентьева М. В. Формирование исследовательской компетенции студентов вуза / М. В. Арсентьева, М. С. Воротилин. – Текст : непосредственный // Известия тульского государственного университета. Технические науки. – 2018. – № 11. – С. 473-477;
16. Ворожко Т. В. Роль научно-исследовательской деятельности студентов в процессе обучения в вузе / Т. В. Ворожко. – Текст : непосредственный // Наука и образование: новое время. – 2017. – № 5(22). – С. 395-401;
17. Рифицкая И. И. Научно-исследовательская деятельность как компонент подготовки будущих специалистов / И. И. Рифицкая. – Текст : непосредственный // Научные труды Республиканского института высшей школы. – 2018. – № 18-3. – С. 275-281;
18. Hui W.G. Strategy and practice of undergraduate research competence building in research universities in China / W.G. Hui – Direct text // Tsinghua J. of Education. – 2008. – Vol. 3. – P. 10;
19. Wynne H.C. Criteria for evaluating systems for student assessment / H. Wynne. – Direct text // Studies in Educational Evaluation. – 2007. – Vol. 33, Issue 1. – P. 15–28. – DOI: 10.1016/j.stueduc.2007.01.003;
20. Denicolo P. Assessing the PhD: a constructive view of criteria / P. Denicolo. – Direct text // Quality Assurance in Education. – 2003. – Vol. 11, Issue 2. – P. 84–91;
21. Evaluating research: A multidisciplinary approach to assessing research practice and quality / P. Mårtensson, U. Fors, S.-B. Wallin [et al.]. – Direct text // Research Policy. – 2016. – Vol. 45, Issue– 3. P. 593 603. – DOI: 10.1016/j.respol.2015.11.009;
22. Rear D. One size fits all? The limitations of standardised assessment in critical thinking / Rear D. – Direct text // Assessment and Evaluation in Higher Education. – 2019. – Vol. 44, Issue – 5. P. 664 675. – DOI: 10.1080/02602938.2018.1526255;
23. Mathematical model of successful research activities for technical university students / M.A. Fedorova, M.V. Tsyguleva, T.A. Vinnikova, V.Y. Kirnosov. – Direct text // J. of Physics: Conference Series. – 2018. – Vol. 1141, Issue–1. DOI:10.1088/1742- 6596/1141/1/012012;

24. Чекардовская И. А. Программа автоматизации методики критериальной оценки научной работы обучающихся по направлению магистратуры "Нефтегазовое дело" / И. А. Чекардовская, Л. Н. Бакановская. – Текст : непосредственный // Автоматизация, телемеханизация и связь в нефтяной промышленности. –2019. – № 8(553). – С. 44-48;

25. Затонский А. В. Информационные технологии: разработка информационных моделей и систем : учебное пособие/А.В. Затонский. – Москва : НИЦ ИНФРА-М, 2014. – 344 с. – Текст : непосредственный;

26. Syzrantseva, K Modeling of the Cement Unloading Process with the Queuing Theory and Optimization of its Parameters / K. Syzrantseva, and A. Tarasov.– Direct text // Journal of Engineering and Applied Sciences. – 2018. – 13: 6754- 6759. DOI: 10.3923/jeasci.2018.6754.6759;

27. Bakanovskaya L. N. Improvement of information processing quality in preproduction engineering efficiency

assessment / L. N. Bakanovskaya, I. A. Chekardovskaya. – Direct text // IOP conference series: Materials science and engineering .– 2016. – P. 012018;

28. Сызранцева К. В. Оптимизация управления потоками газа в системе транспорта газа на основе теории графов / К. В. Сызранцева, Л. Н. Бакановская, В. О. Румянцев. – Текст : непосредственный // Автоматизация, телемеханизация и связь в нефтяной промышленности. –2019. – № 6(551). – С. 44-49;

29. Наймушина А. Г. Системный анализ и обработка экспертной информации для определения степени значимости факторов риска / А. Г. Наймушина, И. А. Чекардовская, Л. Н. Бакановская. – Текст : непосредственный // Информационные технологии и системы: Труды Седьмой Всероссийской научной конференции с международным участием. Ответственные редакторы Ю.С. Попков, А.В. Мельников. – Ханты-Мансийск, 2019. – С. 20-25.

Designing the automated system of criterial evaluation of the scientific research work of students in the magistracy using SADT and DFD methodologies

Bakanovskaya L.N., Glebova N. S.

Federal State Budget Educational Institution of Higher Education « Industrial University of Tyumen»/ IUT
Tyumen, Russian Federation -
e mail: l.bakanovskaya@gmail.com

Abstract. The paper considers a structural analysis of an automated system of criteria-based assessment of the research work of undergraduates. Particular attention is paid to the movement of information flows between the undergraduate, supervisor, supervisor and head of the department.

The developed assessment criteria used in the automated system will improve the quality of research and master's

theses and bring future graduates as close as possible to solving real practical problems.

Keywords: SADT methodology, automated system, structural analysis, criteria-based assessment, research activities.

Стохастический метод распознавания действий человека на базе скелетной модели

Егоров Ю. А.^{1,2}, Захарова И.Г.^{1,2}, Гасанов А.Р.¹, Филицин А.А.¹

¹ФГАОУ ВО «Тюменский государственный университет», Тюмень, Россия, {y.a.egorov, i.g.zakharova}@utmn.ru

²Научно-технический университет «Сириус», Сочи, Россия

Аннотация. Моделирование движений и паттернов действий человека для распознавания или обнаружения особых событий на видеозаписях в последние годы вызывает большой интерес среди исследователей. В работе представлен подход к распознаванию аномальных действий людей в виде последовательностей промежуточных состояний. Предлагается разложить каждое действие в последовательность дискретных промежуточных состояний и представить переходы между состояниями как стохастический процесс. Каждое состояние описывается положением суставов человека. Действия описываются с помощью скрытой Марковской модели, основанной на найденных состояниях и переходах между ними. Полученная в результате модель является комбинацией стохастической модели действий человека и скелетной модели, описывающей промежуточные состояния. Для распознавания промежуточных состояний применяется сверточная нейронная сеть. Для поиска параметров Марковской модели используется алгоритм Витерби. Предложенный метод был реализован и протестирован на двух выборках: MPII Human Pose Dataset и на видеозаписях проведения экзаменов.

Ключевые слова: моделирование поведения, распознавание действий людей, скрытая Марковская модель, последовательность промежуточных состояний, скелетная модель действий человека.

ВВЕДЕНИЕ

Интерес исследователей к решению задач распознавания динамических объектов обусловлен практической значимостью соответствующих результатов для разработки систем компьютерного зрения. В первую очередь это системы отслеживания движущихся объектов, анализа загруженности объектов инфраструктуры, системы для описания динамических объектов, моделирования движений. Особое место занимает исследование проблем распознавания и диагностики действий определенных объектов в кадре [1-3], непосредственно связанных с созданием интерфейсов для взаимодействия человека и компьютера, систем автоматизированного распознавания жестов, интеллектуальных систем видео наблюдения. В частности, такие системы должны выявлять аномальные ситуации.

Материалом настоящего исследования служат данные системы видеонаблюдения, отражающие ситуации в ходе проведения экзамена. Аномальными ситуациями в данном случае понимаются нарушения правил проведения экза-

мена: списывание, использование запрещенных предметов, нарушения со стороны организаторов экзамена.

Для распознавания аномальных ситуаций обычно используется подход, заключающийся в построении моделей ожидаемых ситуаций с помощью вероятностных методов [4, 5] или алгоритмов машинного обучения [6, 7]. При этом аномальными считаются те ситуации, которые не вписываются в рамки разработанных моделей. Наше решение задачи основано на гипотезе о том, что типы ситуаций, нарушающих регламент проведения экзамена, известны, и необходимо распознать эти заданные ситуации на видеозаписях. Таким образом, распознавание сводится к решению задачи классификации.

Целью работы является исследование результативности комбинирования скелетной модели объекта и стохастического подхода для описания действий, совершаемых некоторым объектом, с точки зрения точности распознавания аномальных ситуаций.

ОБЗОР ЛИТЕРАТУРЫ

Исследования в области диагностики аномальных действий определяют два основных подхода к решению данной проблемы [1-3]. В основе первого подхода лежит гипотеза о том, что аномальные действия относятся к одному из конкретных классов, заданных экспертами, и сводит диагностику аномальных действий к решению задачи многоклассовой классификации [8, 9]. Второй подход основан на гипотезе о том, что аномальными являются действия, отличающиеся от действий, относящихся к классу нормальных, и сводит диагностику аномальных действий к задаче одноклассовой классификации [6, 7], в том числе, задаче исследования распределения признаков нормальных действий [5].

Среди моделей, описывающих действия объектов, широкое распространение получили низкоуровневые модели, основанные на оптическом потоке. Wang и Snoussi [6] предлагают использовать гистограммы направленных градиентов оптического потока для кодирования движений, а Yuan et al [10] используют выборочные гистограммы оптического потока для вычисления разницы движений между объектами в скоплениях людей. В [11] используется интегральный оптический поток для построения региональных индикаторов движения, описывающих поведение динамических объектов. Кроме того, используют

ся другие низкоуровневые модели, такие как мешок визуальных слов [4]; модели, описывающие связи между траекториями движения соседних точек на кадрах [12], автокодировщики движений, которые находятся с помощью Motion DeepNet [7] и вариационный автокодировщик Гаусса (Gaussian Mixture Variational Autoencoder) [5].

Также существует ряд моделей, описывающих действия с помощью высокоуровневых признаков: ключевых точек и их характеристик. В частности, Chelli и Pätzold [13] используют ускорение и угловую скорость ключевых точек для обнаружения падения объектов. Wang et al [8] представили пространственно-временную модель, в которой действие объекта описывается с помощью мешка визуальных слов, где каждое слово отражает взаимосвязи между перемещением ключевых точек – суставов объектов в кадре. В [14] Niebles et al также выделяют опорные точки, описывающих действие объекта, при этом они предлагают разделять действия на простые фрагменты. Идея представления действий в виде последовательности состояний используется также при стохастическом моделировании и моделировании с помощью рекуррентных нейронных сетей. В частности, в работах [15, 16], применяются скрытые Марковские модели, в которых промежуточные состояния кодируются мультимодальными разложимыми моделями (Multimodal Decomposable Models). Du et al [17] разработали модель действий с помощью иерархической рекуррентной нейронной сети.

Для решения задачи диагностики действий используются различные методы. Первая группа методов – это методы классического машинного обучения, такие как, метод опорных векторов (SVM) [8, 6], деревья решений, метод ближайших соседей [13]. Вторая группа – методы глубокого обучения. В частности, Simonyan и Zisserman в [9] предложили использовать двухпоточковую сверточную сеть, в которой один поток анализирует траектории движений объектов между кадрами, второй – статичные изображения, извлекая контекст, в котором находятся объекты. Feng et al [18] предложил использовать PCANet для извлечения признаков действий и смешанную модель Гаусса для классификации. Также используются комбинированные методы, например, Xu et al [7] извлекают признаки действий с помощью Motion DeepNet и использует ансамбль SVM для распознавания аномальных действий.

На основе анализа методов решения задачи диагностики действий, мы предлагаем использовать комбинированный подход, включающий методы глубокого машинного обучения для извлечения ключевых точек [19, 20], описывающих промежуточные состояния, и скрытые марковские модели для моделирования действий человека и решения задачи диагностики.

А. Постановка задачи

Диагностика действий объекта решается как задача классификации.

Дано множество $X = \{x: x - \text{описание состояния объекта}\}$, множество $Y = \{y: y - \text{состояние объекта}\}$, при этом состояние $y_i \in Y$ соответствует описанию $x_j \in X$; множество $Z = \{z: z - \text{действие, совершаемое объектом}\}$, действие $z_i \in Z$ характеризуется последовательностью состояний $y_i = \{y_i^k\}^K$, где K – количество последовательных состояний, определяющих действие, задано априорно.

Необходимо построить классификатор $F_1: X \rightarrow Y$, который каждому описанию состояния объекта $x \in X$ сопоставляет верное состояние объекта $y \in Y$, построить классификатор $F_2: Y^K \rightarrow Z$, который последовательности $y_i = \{y_i^k\}^K$ сопоставляет действие объекта $z \in Z$.

В. Модель промежуточных состояний

Промежуточное состояние объекта описывается скелетной моделью. Скелетная модель представляет собой вектор $x_i = (x_i^1, \dots, x_i^j, \dots, x_i^M)$, где x_i^j – координаты опорной точки x_i , M – количество опорных точек, задано априорно для $\forall x \in X$. В такой форме модель не отражает связи между точками, так как они являются фиксированными и одинаковыми для $\forall x \in X$. Для распознавания используется графическая форма скелетной модели, то есть изображение, пример которого показан на рис. 1.

Распознавание промежуточных состояний осуществляется с помощью классификатора $F_1: X \rightarrow Y$, который представляет собой сверточную нейронную сеть. Предложенная сеть включает в себя 4 блока: три сверточных блока и один полносвязный блок.

Каждый сверточный блок состоит из последовательности слоев $C - BN - C - BN - P$, где C – сверточный слой, BN – слой нормализации, P – слой пулинга. После каждого сверточного слоя применяется функция активации **LeakyReLU**. Каждый сверточный слой имеет по 9 фильтров размера 3×3 пикселя, пулинг осуществляется методом выбора максимального элемента с ядром пулинга 2×2 пикселя.

Полносвязный слой состоит из последовательности слоев $FL_1 - BN - FL_2$, где FL_1 – скрытый слой из двадцати нейронов, FL_2 – выходной слой. В качестве функции активации используется **ReLU**.

Параметры для обучения содержатся в слоях C, BN, FL_1 и FL_2 .

С. Действие объекта как последовательность состояний

Процесс совершения действия объектом представлен как стохастический процесс, в ходе которого наблюдае-

мый объект с некоторой вероятностью переходит из текущего состояния в следующее.

Тогда можно представить действие с помощью Марковской модели F_2 , заданной элементами $(Y, Z, \mathcal{R}, \mathcal{P}, \mathcal{P}_0, F_1)$:

- Y – множество состояний объекта;
- Z – множество действий объекта;
- \mathcal{R} – распределение вероятности того, что объект совершает действие $z^{j+1} \in Z$, находясь в состоянии y^{j+1} ;
- \mathcal{P} – распределение вероятности перехода в состояние y^{j+1} из текущего состояния y^j ;
- \mathcal{P}_0 – распределение вероятностей начального состояния y^0 ;
- $F_1: X \rightarrow Y$ – классификатор, который каждому описанию состояния объекта $x \in X$ сопоставляет верное состояние объекта $y \in Y$.

Для того, чтобы найти F_2 , необходимо при заданных Y, Z найти следующие параметры модели: $\mathcal{R}, \mathcal{P}, \mathcal{P}_0, F_1$.

Примечание. Такой подход к моделированию предполагает, что:

- для описания различных действий достаточно использовать одинаковый и фиксированный шаг дискретизации действий;
- не рассматриваются ситуации, в которых наблюдаемые объекты взаимодействуют друг с другом;
- в каждый момент времени имеется полное описание, достаточное для распознавания состояния объекта, достаточность описания определяется на основе экспертной оценки.

D. Алгоритм распознавания действия

Комбинированный метод, использующий модели, описанные выше, реализуется следующим итерационным алгоритмом.

- Создается классификатор промежуточных состояний F_1 .
- На шаге $j = 0$ создается классификатор F_2 с параметрами: F_1 – классификатор, $y^0 \sim \mathcal{P}_0$ – начальное состояние.
- На шаге $j = 1, K - 1$:
 - по описанию x^j классификатор F_1 находит $Y^j \subset Y$ – множество, из которого будет выбрано следующее состояние y^{j+1} ;
 - классификатор выбирает следующее состояние $y^{j+1} \sim \mathcal{P}(\cdot | y^j, Y^j)$;
 - классификатор выбирает действие, которое совершает объект $z^{j+1} \sim \mathcal{R}(\cdot | y^{j+1})$;
 - классификатор возвращает класс действия z^{j+1} .

На рис. 1-2 приведены примеры ситуаций, которые должна различать предложенная модель. Согласно экс-

пертным правилам, при обнаружении списывания важно зафиксировать, что перед тем, как опустить руку за шпаргалкой, студент поднимает голову для того, чтобы оглядеться по сторонам и проверить, наблюдают ли за ним.

ЭКСПЕРИМЕНТЫ

Для оценки практической применимости предложенного метода были проведены два эксперимента. В первом было проведено сравнение точности классификации промежуточных состояний объектов на основе исходных изображений и с использованием различных форм скелетной модели. Эксперимент проводился на наборе данных MPII Human Pose Dataset [21], состоящем из изображений людей, выполняющих различные действия. Во втором эксперименте был осуществлен поиск параметров скрытой марковской модели для представления последовательности состояний точки на видеозаписях. Эксперимент проводился на видеозаписях экзаменов, из которых были извлечены данные о перемещении опорной точки объекта во время выполнения различных действий.

A. Сравнение представлений исходных данных для классификации состояний объектов

В ходе эксперимента сравнивались три представления исходных данных для решения задачи классификации промежуточных состояний объектов:

- 1) изображения объектов без предобработки;
- 2) изображения объектов совместно со скелетной моделью в графической форме;
- 3) скелетная модель в графической форме.

В качестве классификатора промежуточных состояний F_1 использовалась сверточная нейронная сеть. Критерием для сравнения эффективности представления исходных данных была точность классификации.

Эксперимент был выполнен на выборке MPII Human Pose Dataset [21]. Все изображения разделены на категории действий, каждая категория включает в себя несколько видов действий. Были выбраны пять категорий: охота и рыбалка, работа в саду, активность на воде, танцы, игра на музыкальных инструментах, где каждая категория обозначала состояние объекта в кадре. Всего для эксперимента было использовано около 4200 изображений из них 70% были отнесены к обучающей выборке, 15% – к проверочной, 15% – к тестовой.

Результаты эксперимента представлены в табл. 1, из которой видно, что использование скелетной модели в графической форме всегда дает наименьшую точность распознавания. При этом для категорий «охота и рыбалка» и «игра на музыкальных инструментах» использование исходных изображений со скелетной моделью дает ухудшение результата. В то же время, использование скелетной модели позволяет улучшить точность классификации для категории «активность на воде».

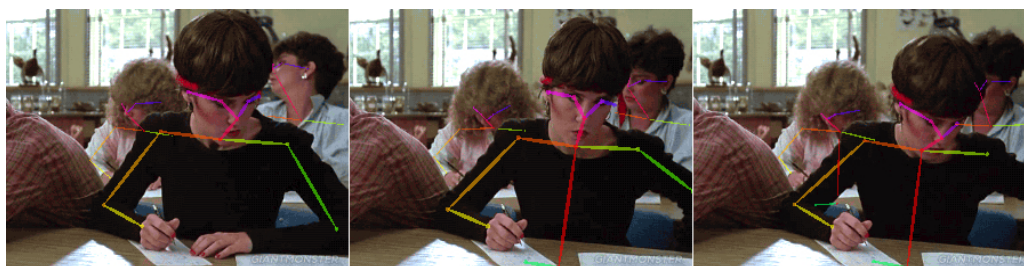


Рис. 1. Раскадровка первой ситуации: объект пытается списать. Слева направо: кадр 1 – голова опущена, руки лежат на парте; кадр 2 – голова поднята, левая рука находится под партой; кадр 3 – голова опущена, левая рука находится под партой

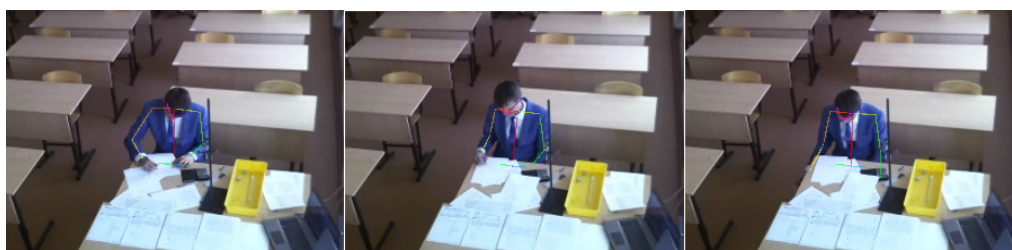


Рис. 2. Раскадровка второй ситуации: объект ведет себя похожим образом, но не списывает. Слева направо: кадр 1 – голова опущена, руки находятся на парте; кадр 2 – голова опущена, левая рука находится под партой; кадр 3 – голова опущена, и левая, и правая руки находятся под партой

По матрицам ошибок, представленных на рис. 3, можно увидеть зависимость между точностью классификации с помощью скелетных моделей и изменением точности классификации при добавлении скелетной модели к исходным изображениям. В частности, из матрицы третьего эксперимента видно, что скелетная модель не позволяет точно классифицировать категории “охота и рыбалка” и “игра на музыкальных инструментах”, так как для этих действий важен контекст. Поэтому при смещении акцента на положение объекта с помощью скелетной модели, точность классификации снижается в данном случае. Но для категории “активность на воде” использование скелетной модели позволяет более точно отличить действия данной категории от действий категории “охота и рыбалка”, так как часть изображений в этих категориях имеют общий

контекст – активность на фоне воды с использованием плавательных средств.

Таблица 1.

Точность классификации промежуточных состояний

Тип состояния	Исходные изображения	Изображения + скелетная модель	Скелетная модель
Охота и рыбалка	0,62	0,543	0,161
Работа в саду	0,887	0,909	0,381
Активность на воде	0,659	0,818	0,506
Танцы	0,843	0,885	0,5
Игра на музыкальных инструментах	0,809	0,607	0,36
Общее	0,751	0,808	0,414

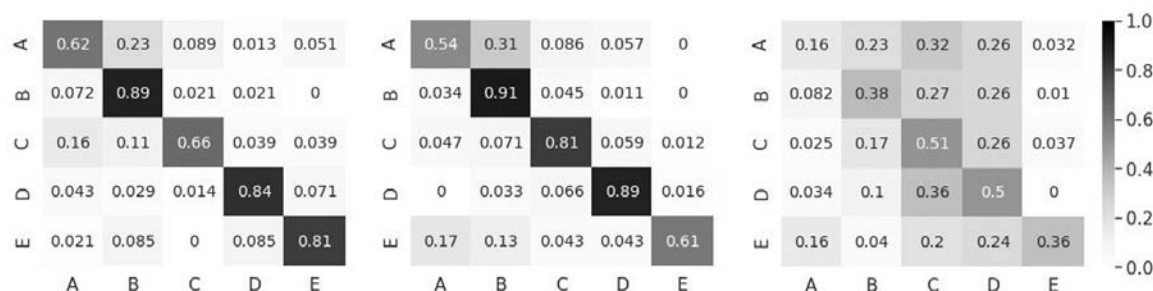


Рис. 3. Матрицы ошибок для классификации промежуточных состояний слева направо: эксперимент 1 с использованием исходных изображений; эксперимент 2 с использованием исходных изображений и скелетной модели в графической форме; эксперимент 3 с использованием скелетной модели в графической форме. Категории дей-

В. Поиск параметров модели действия

В ходе данного эксперимента были найдены параметры модели действия объекта F_2 – матрица переходов между состояниями объекта.

Для поиска параметров были использованы видеозаписи экзаменов. Из исходных данных была извлечена информация о перемещении опорной точки объекта следующим образом:

- 1) была выбрана одна опорная точка, принадлежащая наблюдаемому объекту;
- 2) на основании положений всех опорных точек объекта определен условный центр тяжести Q – точка, относительно которой необходимо следить за траекториями опорной точки;
- 3) зафиксированы траектории точки во время совершения объектом некоторых действий;
- 4) выбраны ключевые положения точки, относящиеся к одному из пяти классов, обозначающих состояние точки: 0 – точка выше объекта, 1 – точка левее объекта, 2 – точка ниже объекта, 3 – точка правее объекта, 4 – точка в области объекта.

Входными данными для поиска параметров модели действия являлись последовательности состояний объекта с точки зрения наблюдаемой точки. Поиск параметров осуществлялся с помощью алгоритма Витерби [22]. На рисунке 4 представлена матрица переходов обученной модели, где на пересечении строки i и столбца j указана вероятность перехода состояния точки из класса i в класс j .

В ходе эксперимента было выдвинуто предположение о том, что важен выбор точки Q , и данный вопрос требует дополнительного исследования.

ЗАКЛЮЧЕНИЕ

Исследование стохастического метода распознавания аномальных действий людей, описанных с помощью скелетных моделей, позволило выявить определенные закономерности:

0	0.532	0.189	0.019	0.226	0.033
1	0.334	0.277	0.07	0.19	0.129
2	0.009	0.021	0.775	0.005	0.19
3	0.247	0.12	0.008	0.587	0.038
4	0.033	0.071	0.215	0.051	0.63
	0	1	2	3	4

Рис. 4. Матрица переходов обученной модели

- невысокая точность классификации при использовании только скелетной модели объясняется тем, что в используемой выборке большое значение имеет контекст, в котором находятся объекты. Под контекстом здесь подразумевается окружение и инструменты, используемые объектами. Можно предположить, что, если объекты будут совершать различные действия в одном контексте, точность распознавания улучшится. В рассматриваемой задаче распознавания нарушений во время проведения экзамена контекст не является существенным, поэтому допустимо использовать скелетную модель для распознавания состояний объектов;
- использование сверточной нейронной сети для распознавания промежуточных состояний позволяет не нормировать координаты опорных точек;
- положение условного центра тяжести объекта зависит от особенностей распознаваемых аномальных действий, поэтому необходимо дополнительно исследовать методы поиска данной точки для получения репрезентативных траекторий опорных точек.

Предложенный метод опирается на комбинированную модель, объединяющую скелетный и стохастический подходы, что определяет ее гибкость. В частности, допускается использование другого вида классификатора промежуточных состояний или иного представления исходных данных. В то же время, нельзя сказать, что модель является универсальной, и ее использование для конкретных случаев требует предварительных исследований. При этом особый интерес представляет развитие предложенного метода для выявления аномальных ситуаций, связанных с взаимодействием объектов.

БЛАГОДАРНОСТИ

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-37-51028

ЛИТЕРАТУРА

1. Popoola O.P. Video-based abnormal human behavior recognition – A review / O.P. Popoola, K. Wang // IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). – 2012. – vol. 42, is. 6. – P. 865-878.
2. Mabrouk A.B. Abnormal behavior recognition for intelligent video surveillance systems: A review / A.B. Mabrouk, E. Zagrouba // Expert Systems with Applications. – 2018. vol. 91. – P. 480-491.
3. Dhiman C. A review of state-of-the-art techniques for abnormal human activity recognition / Dhiman C., Vishwakarma D.K. // Engineering Applications of Artificial Intelligence. – 2019. – vol. 77. – P. 21-45.
4. Roshtkhari M.J. An on-line, real-time learning method for detecting anomalies in videos using spatio-temporal compositions / M.J. Roshtkhari, M.D. Levine // Computer Vision and Image Understanding. – 2013. – vol. 177, is. 10. – P. 1436-1452.
5. Video Anomaly Detection and Localization via Gaussian Mixture Fully Convolutional Variational Autoencoder – URL: <https://arxiv.org/abs/1805.11223> (дата обращения 23.01.2020).

Ю.А. Егоров, И.Г. Захарова, А.Р. Гасанов, А.А. Филицин

Стохастический метод распознавания действий человека на базе скелетной модели

6. Wang T. Detection of Abnormal Visual Events via Global Optical Flow Orientation Histogram / T. Wang, H. Snoussi // *IEEE Transactions on Information Forensics and Security*. – 2014. – vol. 9, is. 6. – P. 988-998.
7. Xu D. Detecting anomalous events in videos by learning deep representations of appearance and motion / D. Xu, Y. Yan, E. Ricci, N. Sebe // *Computer Vision and Image Understanding*. – 2017. – vol. 156. – P. 117-127.
8. Wang C. An approach to pose-based action recognition / C. Wang, Y. Wang, A.L. Yuille // *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (Portland, Oregon, June 25-27, 2013). – Portland, Oregon. – P. 915-922.
9. Simonyan K. Two-Stream Convolutional Networks for Action Recognition in Videos / K. Simonyan, A. Zisserman // *Neural Information Processing Systems (Montreal, Quebec, December 8-13, 2014)*. Montreal, Quebec. – P. 568-576.
10. Yuan Y. Online Anomaly Detection in Crowd Scenes via Structure Analysis / Y. Yuan, J. Fang, Q. Wang // *IEEE Transactions on Cybernetics*. – 2014. – vol. 45, is. 3. – P. 548-561.
11. Chen H. Motion Maps and Their Applications for Dynamic Object Monitoring / H. Chen, Sh. Ye, O. Nedzvedz, S. Ablameyko, Zh. Bai // *Pattern Recognition and Image Analysis*. – 2019. – vol. 29, is. 2. – P. 131-143.
12. Garzón G. A Fast Action Recognition Strategy Based on Motion Trajectory Occurrences / G. Garzón, F. Martínez // *Pattern Recognition and Image Analysis*. – 2019. – vol. 29, is. 2. – P. 477-456.
13. Chelli A. A Machine Learning Approach for Fall Detection and Daily Living Activity Recognition / A. Chelli, M. Pätzold // *IEEE Access*. – 2019. – vol. 7. – P. 38670-38687.
14. Niebles J.C. Modeling Temporal Structure of Decomposable Motion Segments for Activity Classification / J.C. Niebles, Ch. Chen, L. Fei-Fei // *11th European Conference on Computer Vision, Proceedings, Part II (Heraklion, Crete, September 5-11, 2010)*. – Heraklion, Crete. – P. 392-405.
15. Zhang D. Semi-supervised adapted HMMs for unusual event detection / D. Zhang, D. Gatica-Perez, S. Bengio, I. McCowan // *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. – vol. 2. – (San Diego, CA, June 20-25, 2005). – San Diego, CA. – P. 302-309.
16. Hendryli J. Classifying abnormal activities in exam using multi-class Markov chain LDA based on MODEC features / J. Hendryli, M.I. Fanany // *2016 4th International Conference on Information and Communication Technology (ICoICT)* (Bandung, Indonesia, May 25-24, 2016). – Bandung, Indonesia. – P. 1-6.
17. Du Y. Hierarchical Recurrent Neural Network for Skeleton Based Action Recognition / Y. Du, W. Wang, L. Wang // *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Boston, Massachusetts, June 8-10, 2015). – Boston, Massachusetts. – P. 1110-1118.
18. Feng Y. Learning deep event models for crowd anomaly detection / Y. Feng, Y. Yuan, X. Lu // *Neurocomputing*. – 2017. – vol. 219. – P. 548-556.
19. Wei S. Convolutional Pose Machines / S. Wei, V. Ramakrishna, T. Kanade, Y. Sheikh // *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, Nevada, June, 26 – July, 1, 2016). – Las Vegas, Nevada. – P. 4724-4732.
20. Cao Z. Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields / Z. Cao, T. Simon, S. Wei, Y. Sheikh // *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Honolulu, Hawaii, July, 22-25, 2017). – Honolulu, Hawaii. – P. 7291-7299.
21. MPII Human Pose Dataset. – URL: <http://human-pose.mpi-inf.mpg.de/> (дата обращения 23.01.2020).
22. Viterbi A. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm / A. Viterbi // *IEEE Transactions on Information Theory*. – 1967. – vol. 13, is. 2. – P. 260-269.

Stochastic Method for Skeleton Based Human Action Diagnostics

Yury A. Egorov^{1,2}, Irina G. Zakharova^{1,2}, Alexandr R. Gasanov¹, Andrew A. Filitsin¹

¹University of Tyumen,
Tyumen, Russia,
{y.a.egorov, i.g.zakharova}@utmn.ru

²Sirius University of Science and Technology,
Sochi, Russia

Abstract. In recent years, modeling of human actions and activity patterns for recognition or detection of the special event has attracted a significant research interest. We present our approach for abnormal human action recognition as a sequence of intermediate states. We propose to decompose each action into a sequence of discrete intermediate states and to present state tran-features for intermediate states recognition. Viterbi algorithm is employed to find model parameters. We implemented proposed methods in a framework for human abnormal action recognition and tested our approach on two benchmarks: MPII Human Pose Dataset and exam footages.

sitions as a stochastic process. Each state is described with the joint locations of a human skeleton. Actions are described with Hidden Markov Model based on the found states and its inter-connections. As a result, we combine our stochastic model of human action with intermediate states described via skeleton joints. Convolutional Neural Network is employed to learn skeleton *Keywords:* behavior modeling, human action recognition, Hidden Markov Models, intermediate state sequences, skeleton based actions detection.

Методика оценки экономической нагрузки для информационно-аналитической системы мониторинга и планирования расходов по выхаживанию и лечению новорожденных, родившихся с ОНМТ и ЭНМТ

Максименко З.В., Бахитова Р.Х.

ФГБОУ ВО Башкирский государственный университет
Уфа, Россия
zubazzz@maill.ru

Брюханова О.А.

ФГБОУ ВО Башкирский государственный медицинский университет
Уфа, Россия
bryukhanova.olga22@yandex.ru

Аннотация. В статье предложена концепция комплексной оценки расходов на выхаживание и реабилитацию и поддержание в течение всей жизни недоношенных детей, родившихся с очень низкой и экстремально низкой массой тела для информационно-аналитической системы медицинского мониторинга. Сформирована целостная структура прямых и косвенных затрат, определены методы их оценки.

Ключевые слова: социально-экономическая нагрузка, недоношенные дети, очень низкая масса тела, экстремально низкая масса тела, затраты на выхаживание.

ВВЕДЕНИЕ

В России в настоящее время все расходы на выхаживание, реабилитацию и поддержания здоровья и качества жизни недоношенных детей с очень низкой (ОНМТ) и экстремально низкой массой тела (ЭНМТ) в основном лежат на государстве. Затраты по более или менее «высокому» тарифу предусмотрены только на неонатальный период в возрасте до 28 дней жизни за счет выделяемого объема по высокотехнологичной медицинской помощи. В последующем при развитии заболеваний такие дети лечатся по ОМС, где тарифные соглашения очень низкие [1,2].

Существующий подход к определению затрат на выхаживание и реабилитацию недоношенных детей не рассматривает ни дифференциацию по массе тела, ни по периоду необходимой интенсивной терапии. Также не учитываются риски возникновения инвалидности, ожидаемое качество жизни детей в будущем и расходы, связанные с этим. Структура затрат в основном включает зарплату, оплату ЖКХ и других расходов, на медицинскую составляющую (медикаменты, технологии, медицинское оборудование) приходится лишь малая часть. Реальные расходы на лечение, по оценкам различных медицинских учреждений, занимающихся выхаживанием и реабилитацией недоношенных детей, таких как ГБУЗ РБ «Городская детская клиническая больница № 17», г. Уфа, гораздо выше установленного тарифа [2,7].

Таким образом, актуальной становится разработка методики оценки суммарных затрат на выхаживание и

реабилитацию детей, родившихся с ОНМТ и ЭНМТ, а также на поддержание здоровья и качества жизни таких детей в детском, а затем и взрослом возрасте, и их мониторинга с помощью информационно-аналитической системы [4,9].

ОПИСАНИЕ МЕТОДИКИ

Предлагаемый в данной статье подход позволяет всесторонне оценить проблему и учесть дифференциацию по массе тела и риски возникновения инвалидности [2,3]. Помимо расходов непосредственно на выхаживание и реабилитацию на первом году жизни оцениваются также возникающие в течение всей жизни ребенка:

- прямые (затраты на лечение заболеваний и осложнений, в том числе препараты, консультации врачей, операционные вмешательства и пр.);

- косвенные затраты (экономический ущерб, который понесет государство из-за инвалидизации и нетрудоспособности ребенка в будущем и нетрудоспособности одного из родителей или опекунов).

Прямые медицинские затраты, производимые органами здравоохранения и членами семей детей для проведения соответствующего лечения, рассчитывались как сумма двух составляющих:

1) Затрат на стационарное лечение при рождении и в период выхаживания в отделении реанимации и интенсивной терапии (ОРИТН) и отделении патологии новорожденных (ОПН):

$$\text{СПЗ} = \text{Смед} + \text{Сиссл} + \text{Сопер} + \text{Спит} + \text{Сгем} + \text{Спец} + \\ + (\text{Сгост} + \text{Синв} + \text{Срм} + \text{Скисл} + \text{Скдз}) \times N_{\text{ОРИТН}} + \\ + (\text{Сгост} + \text{Синв} + \text{Срм} + \text{Скисл} + \text{Скдз}) \times N_{\text{ОПН}},$$

где Смед – стоимость медикаментозной терапии, Сиссл – затраты на проведение лабораторных и инструментальных исследований, Сопер – затраты на проведение оперативных вмешательств, Спит – стоимость питательных смесей, Сгем – стоимость гемотрансфузий, Спец – стоимость консультаций узких специалистов, в том числе из других стационаров, Сгост – стоимость гостиничных услуг на пациента в сутки, Синв – стоимость мягкого инвентаря на 1 койко-день, Срм – затраты на расходные материалы на 1 койко-день, Скисл – затраты на

обеспечение кислородом 1 койко-день, Склад – затраты на выплату заработной платы сотрудникам отделения за 1 койко-день, проведенный пациентом в отделении, $N_{оритн}$ и $N_{опн}$ – средняя длительность пребывания пациента в отделениях ОРИТН и ОПН соответственно, дней. Расчет затрат на «гостиничные услуги» проводится на основании данных оборотно-сальдовой ведомости стационара и включает следующие виды расходов: услуги связи, транспортные услуги, коммунальные услуги, работы и услуги по содержанию имущества, расходы на противопожарную безопасность, амортизация основных средств, продукты питания, хозяйственные и прочие расходы [3,7].

Для определения стоимости лечения по данным собранной базы были рассчитаны следующие показатели: средняя длительность пребывания в ОРИТН, ОПН для каждой группы детей (500-749 грамм, 750-999 грамм, 1000-1499 грамм), среднее количество дней применения препарата для каждой группы, среднее количество дней применения питательной смеси для каждой группы, среднее количество проведенных лабораторных и инструментальных исследований, оперативных вмешательств для каждой группы, среднее количество переливаний СЗП и эритроцитарной массы для каждой группы.

2) Затрат на стационарное лечение диагностированных на этапе выхаживания заболеваний и их осложнений.

Для оценки данных затрат на основании клинических рекомендаций, утвержденных соответствующими профессиональными ассоциациями и научным советом Минздрава РФ, приказов Минздрава РФ (например, «Об утверждении стандарта специализированной медицинской помощи при детском церебральном параличе (фаза медицинской реабилитации)» [5]), а также анализа медицинской литературы и практики, сформированы типовые ежегодные программы реабилитации в зависимости от поставленного диагноза и проведен расчет расходов на такую программу для наиболее распространенных заболеваний глубоко недоношенных детей: бронхолегочная дисплазия, детский церебральный паралич, ретинопатия, задержка развития и пр. В расчет данных затрат включены расходы на проведение необходимых лабораторных и инструментальных обследований, консультаций врачей, медикаментозной терапии, в т.ч. дорогостоящей, санаторно-курортного лечения, ЛФК, массажа, госпитализаций и оперативных вмешательств по соответствующему диагнозу и пр.

Косвенные (социально-экономические) затраты на самого ребенка и на родителей/опекунов включают: выплаты по инвалидности, потери ВВП/ВРП, налоговых сборов, доходов детей и их родителей, связанные с временной нетрудоспособностью как минимум одного из родителей, потерей ребенком трудоспособности в будущем и пр. Для оценки потерь ВВП/ВРП по причине нетрудоспособности применяется метод «человеческого капитала», при этом учитываются потери ВВП/ВРП за весь период нетрудоспособности потенциально работоспособного гражданина. Оценка социальных затрат и потерь основывается на данных о числе дней нетрудоспособности детей и одного из родителей/опекунов, на потерях консолидированного

бюджета РФ вследствие снижения налоговых сборов из-за нетрудоспособности и преждевременной смерти. При оценке учитываются следующие показатели: расчетное число человеко-лет временной нетрудоспособности (для родителей/опекунов), расчетное число человеко-лет потерянной трудоспособности, и годы по инвалидности на основании ожидаемой длительности жизни (для детей), размер налоговой нагрузки, доход работающего за рабочий день, ежемесячной средней пенсии по инвалидности согласно материалам официальной статистики, общее число новорожденных, доля детей, родившихся с ОНМТ и ЭНМТ согласно данным медицинской статистики, расчетная доля детей-инвалидов из числа, родившихся с ОНМТ и ЭНМТ.

РЕЗУЛЬТАТЫ

В качестве исходных данных для разработки методики оценки экономической нагрузки на выхаживание, реабилитацию, поддержание здоровья и качества жизни недоношенных детей, родившихся с очень низкой и экстремально низкой массой тела, использовалась база данных обследования и лечения детей, родившихся с ОНМТ и ЭНМТ, находившихся на стационарном лечении в 2013-2019 гг. в Детской городской клинической больнице №17 и Республиканском клиническом перинатальном центре г. Уфа. Выборка включала 281 выживших детей, разделенных по трём категориям недоношенности по весу (до 749 гр., 750-999 гр., 1000-1500 гр.).

Расчеты прямых затрат производились на данных Детской ГКБ №17 г. Уфа. Стоимость стационарного лечения при рождении и в период выхаживания в ОРИТН и ОПН, рассчитанная по разработанной методике, составила для детей, рожденных:

- с весом 500-749 гр. – 1 388 839,32 руб.,
- с весом 750 - 999 гр. – 1 101 434,44 руб.,
- с весом 1000-1500 гр. – 921 237,07 руб.

Распределение по статьям прямых затрат на стационарное лечение при рождении и в период выхаживания в ОРИТН и ОПН в рублях в разрезе категорий недоношенности представлено в таблице 1.

Таблица 1

Распределение прямых затрат по статьям, руб.

Статья затрат	500-749 гр.	750-999 гр.	1000-1500 гр.
Пребывание ОПН	609 300	271 488	118 856
Пребывание ОРИТН	249 767	243 562	203 432
Питание	8 598,47	10 091,67	4 273,03
Лабораторные исследования	80 245,45	48 796,25	39 200,53
Консультации врачей	7 272,73	4 630,00	3 284,66
ИВЛ	85 600	24 726	9 930
Переливание крови	5 067,27	3 511,50	1 054,71
Лекарственная терапия	342 988,73	494 628,76	541 206,84
Итого	1 388 839,32	1 101 434,44	921 237,07

Соответствующие тарифы ОМС по Республике Башкортостан на выхаживание глубоко недоношенных детей на 2019 год [8]:

- при оплате медицинской помощи по коэффициенту относительной затратоемкости КСГ/КПГ, скорректированные на коэффициент уровня медицинского учреждения (1,11 для Детской ГКБ №17 г. Уфа),

З.В. Максименко, Р.Х. Бахитова, О.А. Брюханова

составляли 108 418,21 руб. и 412 555,75 руб. для родившихся с ОНМТ и ЭНМТ соответственно;

- при оплате высокотехнологичной медицинской помощи по профилю «Неонатология» составляли 246 915,00 руб. и 360 728,00 руб., для родившихся с ОНМТ и ЭНМТ соответственно.

Предусмотренные ОМС тарифы существенно ниже рассчитанных на эмпирической базе данных прямых затрат на выхаживание и реабилитацию недоношенных детей с ОНМТ и ЭНМТ.

Стоимость сформированных ежегодных типовых программ реабилитации рассчитана для наиболее распространенных заболеваний глубоко недоношенных детей:

- бронхолегочная дисплазия – 225 384,44 руб., срок реабилитации 3 года;

- детский церебральный паралич – 250 377,78 руб., срок реабилитации 18 лет;

- ретинопатия – 255 786,11 руб., срок реабилитации 1 год.

Прямые затраты на ежегодную реабилитацию до достижения ребенком 18 лет в среднем составляют около 2 млн. руб. на 1 ребенка.

Косвенные (социально-экономические) затраты на самого ребенка и на родителей/опекунов рассчитывались по данным официальной статистики (Пенсионный фонд РФ, Росстат, Башстат) на конец 2019 г. [5]:

1) Выплаты пенсии по инвалидности детям-инвалидам до 18 лет – 12 681,09 руб. в месяц.

2) Выплаты компенсаций родителям/опекунам по инвалидности до достижения ребенком 18 лет – 10 000 руб. в месяц;

3) Потери консолидированного бюджета РФ, связанные с временной нетрудоспособностью как минимум одного из родителей до 18-летия ребенка:

- потери налоговых сборов от налога на доходы физических лиц – 13% от средней заработной платы работников организаций в Республике Башкортостан (35 756 руб. в месяц), составляют 4 648,30 руб. в месяц;

- потери ВРП на душу населения по Республике Башкортостан (значение на 2017 г., скорректированное на индекс-дефлятор 2019 г.) составляют 374,4 тыс. руб.

Итоговые прямые затраты до достижения ребенком 18 лет в среднем составляют более 4 млн. руб., косвенные – более 12 млн. руб.

ЗАКЛЮЧЕНИЕ

По результатам разработки методики оценки экономической нагрузки для информационно-аналитической системы мониторинга и планирования расходов по выхаживанию и лечению новорожденных, родившихся с ОНМТ и ЭНМТ, проанализирована и оценена общая структура прямых и косвенных затрат в среднем на 1 ребенка. Выявлено, что наибольший объем прямых затрат приходится на недоношенных детей, родившихся с весом 500-749 гр. Анализ результатов расчета показал, что фактические расходы на выхаживание недоношенных детей существенно выше, чем предусмотренные тарифом ОМС. Также очевидна необходимость разделения тарифа ОМС на 3 группы в соответствии с весом новорожденных, вместо

существующего разделения на ОНМТ (1000-1500 гр.) и ЭНМТ (до 1000 гр.). Кроме того, выявлено, что расходы на последующую реабилитацию и косвенные расходы значительно выше расходов на выхаживание и первоначальную реабилитацию.

Необходимо пересмотреть тарифы ОМС и оценить целесообразность перераспределения финансирования на наиболее важные этапы выхаживания и ранней реабилитации, чтобы снизить риск инвалидизации и затраты, которые понесет как государство, так и члены общества (семьи) на поддержание здоровья и качества жизни таких пациентов [10].

БЛАГОДАРНОСТЬ

Исследование выполнено при финансовой поддержке гранта РФФИ (проект №18-010-00872 «Разработка методологии, математического и программного инструментария интеллектуальной оценки и планирования расходов на выхаживание, реабилитацию, поддержание здоровья и качества жизни недоношенных детей, родившихся с очень низкой и экстремально низкой массой тела»).

ЛИТЕРАТУРА

1. Lakman I.A. Statistical Models of Economic Burden: A Case Study in Medicine / Lakman I.A., Maksimenko Z.V., Shangareeva R.Kh., Gindullin R.V. // International Journal of Economics and Business Administration. – 2019. – vol. VII, Special Issue 2. – P. 63-73. DOI: 10.35808/ijeba/371.

2. Бахитова Р.Х. Подход к оценке и планированию расходов на выхаживание глубоко недоношенных детей / Бахитова Р.Х., Максименко З.В., Лакман И.А. // Экономика и управление: научно-практический журнал. – 2018. – № 5 (143). – С. 144-148.

3. Максименко З.В. Структура затрат на выхаживание, реабилитацию и поддержание здоровья и качества жизни детей, родившихся с ОНМТ и ЭНМТ / Максименко З.В., Бахитова Р.Х., Хафизова Р.К., Прудников В.Б., Брюханова О.А., Лакман И.А. // Региональные перспективы развития экономики здоровья: сборник докладов I Всероссийской научно-практической конференции. Ответственный редактор К.Е. Гришин. – Уфа, 2019. – С. 206-210.

4. Максименко З.В. Информационно-аналитическая поддержка оценки и планирования расходов на выхаживание и реабилитацию недоношенных детей, родившихся с очень низкой и экстремально низкой массой тела / Максименко З.В., Лакман И.А., Бахитова Р.Х. // Информационные технологии и системы: труды Седьмой Всероссийской научной конференции с международным участием. Ответственные редакторы Ю.С. Попков, А.В. Мельников. – Ханты-Мансийск, 2019. – С. 69-73.

5. О внесении изменения в Указ Президента Российской Федерации от 26 февраля 2013 г. № 175 «О ежемесячных выплатах лицам, осуществляющим уход за детьми-инвалидами и инвалидами с детства I группы // Указ Президента РФ от 07 марта 2019 г. Сборник законодательств. – 2019.

6. Об утверждении стандарта специализированной медицинской помощи при детском церебральном параличе (фаза медицинской реабилитации) // Приказ Министерства здравоохранения РФ №349 от 16 июня 2015 г. Сборник Росминздрав. – 2015.

7. Солдатова И.Г., Затраты на стационарное лечение детей с неонатальными инфекциями в условиях реальной практики / Солдатова И.Г., Панкратьева Л.Л., Дегтярева М.В., Омеляновский В.В., Авксентьева М.В., Свешникова Н.Д., Цфасман Ф.М., Кулинчик Т.В., Реброва О.Ю., Деркач Е.В., Ивахненко О.И. // Медицинские технологии. Оценка и выбор. – 2011. № 4 (6). – С. 64-69.

8. Территориальный фонд обязательного медицинского страхования по Республике Башкортостан. – URL: <https://tfoms-rb.ru/ru/node/6085> (Дата обращения 10.01.2020).

9. Шангареева Р.Х. Факторы риска преждевременных родов и ближайшие исходы у детей, родившихся с очень низкой и экстремально низкой массой тела /

Шангареева Р.Х., Брюханова О.А., Фатыхова А.И., Бахитова Р.Х., Максименко З.В., Лакман И.А., Гиндуллин Р.В. // Профилактическая медицина. – 2019. – Т. 22, № 6. – С. 63-72. DOI: 10.17116/profmed20192206163.

10. Шангареева Р.Х.. Основные причины преждевременных родов и ближайшие исходы у глубоко недоношенных детей / Шангареева Р.Х., Брюханова О.А., Бахитова Р.Х., Максименко З.В., Лакман И.А., Гиндуллин Р.В., Зиннурова Н.Р., Трибунская Л.З. // Региональные перспективы развития экономики здоровья: сборник докладов I Всероссийской научно-практической конференции. Ответственный редактор К.Е. Гришин. – Уфа, 2019. – С. 109-113.

Methodology for assessing the economic burden for information analytical system for monitoring and planning costs for nursing and treatment of infants born with VLBW and ELBW

Z.V. Maksimenko, R.Kh. Bakhitova
Bashkir State University
Ufa, Russia
zubazzz@maill.ru

O.A. Bryukhanova
Bashkir State Medical University
Ufa, Russia
bryukhanova.olga22@yandex.ru

Abstract. The paper proposes the concept of a complex assessment of the costs of nursing and rehabilitation and maintenance throughout life of premature babies born with very low and extremely low body weight for the information analytical medical monitoring system. An integral structure of direct and indirect costs has been formed, methods for their assessment have been defined.

Keywords: social-economic burden, premature newborns, very low birth weight, extremely low birth weight, nursing costs.

О проблемах анализа данных в системах управления инцидентами безопасности роботов

Мещеряков Р.В.
ИПУ РАН,
Москва, Россия
mrvt@ipu.ru

Исхаков С.Ю.
ТУСУР,
Томск, Россия
iskhakov.sy@gmail.com

Аннотация. Вмешательство злоумышленников в управление роботами может не только воспрепятствовать выполнению возложенных на них задач и, как следствие, привести к финансовым потерям, но и создать угрозы жизни и здоровью людей. В связи с этим актуальной является проблема формирования механизмов своевременного обнаружения инцидентов в работе робототехнических систем различных масштабов. Одним из актуальных векторов развития научного знания в этой отрасли является совершенствование и разработка новых методов выявления и реагирования на инциденты информационной безопасности в гетерогенных информационных системах посредством анализа потоков дискретных сообщений. Несмотря на значительное количество научно-исследовательских работ в данной области, в большинстве представленных на рынке коммерческих продуктов применяются методы корреляции на основе правил, характеризующиеся низким уровнем работоспособности и необходимостью неавтоматизированной доработки на каждом конкретном объекте защиты. Данные обстоятельства обусловлены такими проблемами как потеря данных при нормализации, связанная с трансформацией моделей; отсутствие методического аппарата нормализации событий в робототехнических комплексах; отсутствие поддержки изменений модели объекта защиты и методологии написания правил корреляции.

В работе предложена методика визуализации сервисных данных от систем аудита робототехнических комплексов как средство совершенствования методов управления инцидентами и представлены результаты ее апробации. Рассмотрены варианты использования современных инструментов для поиска и визуализации больших данных.

Ключевые слова: *робототехника, нормализация, агрегация, корреляция, система управления событиями и инцидентами безопасности.*

ВВЕДЕНИЕ

Современные робототехнические комплексы (РТК) зачастую представляют собой распределенные децентрализованные системы, для успешного функционирования которых необходимо обеспечивать беспроводную связь между их отдельными элементами. Связь между элементами РТК может быть реализована через межмашинное взаимодействие, в том числе в рамках концепции Интернета вещей (Internet of Things, IoT). Однако использование инфраструктуры IoT как платформы для взаимодействия интеллектуальных элементов РТК помимо очевидного преимущества в виде получения эффективного инструмента для управления множеством гетерогенных устройств приводит к наследованию ряда существенных уязвимостей, характерных для данной концепции.

Для обеспечения безопасного функционирования подобных комплексов эффективным решением является применение систем интеллектуального выявления инцидентов безопасности в подсистемах журналирования. Технологические ограничения в большинстве случаев не позволяют хранить журнал событий непосредственно на робототехническом устройстве, в связи с этим актуальным становятся системы агрегации и комплексирования журналов агентов робототехнических групп. Одной из основных идей в данном случае является применение современных технологий сбора информации о состоянии IoT-устройств и анализ полученных данных для выявления инцидентов безопасности. Гетерогенность устройств и отсутствие стандартов к реализации систем логирования событий приводят к тому, что каждый производитель обеспечивает собственные форматы и протоколы для отправки данных об их состоянии.

Необходимо корректно и оперативно обработать получаемые данные для выявления инцидентов информационной безопасности и выделения тех областей инфраструктуры, наиболее подверженных риску реализации угроз. Объемы и гетерогенность подлежащих анализу данных обуславливают задачу их структурированного, консолидированного и визуального представления в рамках единого интерфейса для принятия решений в области безопасного управления группой РТК. В настоящее время на рынке наблюдается переход от средств управления журналами регистрации (Log Management) к SIEM-системам [1,2], обеспечивающим возможность аналитического анализа событий с точки зрения информационной безопасности.

ПОЛУЧЕНИЕ ДАННЫХ ОТ ИСТОЧНИКОВ СОБЫТИЙ

В основе функционирования подобных систем лежит процесс, обеспечивающий непрерывный мониторинг информационной безопасности инфраструктуры, в том числе элементов РТК, за счет объединения информации об активности всех объектов в единый набор данных [3]. При этом основными задачами являются:

- сбор данных и приведение их к единому нормализованному виду
- группирование данных по определенным признакам и атрибутам
- выявление инцидентов на основе обнаружения корреляций и оповещение персонала служб безопасности
- визуализация обрабатываемых данных как инструмент анализа и проведения расследования инцидентов

Р.В. Мещеряков, С.Ю. Исхаков

- создание отчетов о состоянии активов защищаемой системы.

Существующие SIEM-системы критичны к обработке слабо структурированных и неструктурированных данных высокой размерности, порождаемых большим количеством разнообразных устройств. Таким образом, актуальным является разработка методов и алгоритмов определения форматов данных, генерируемых устройствами, извлечения из них значимых параметров и приведения их к единому формату для дальнейшего анализа состояния безопасности защищаемых объектов.

На этапе проектирования РТК не предъявляются комплексные требования к структуре журналов работы всех подсистем, поэтому разработчики аппаратных и программных продуктов разрабатывают системы логирования, исходя из собственных принципов. В одних случаях сообщения записываются в текстовый файл, в других отправляются на syslog-сервер [4,6] или помещаются в базу данных. Форматы данных также могут быть различными, даже если основной механизм транспорта лог-файлов совпадает. Для событий, собранных с разных источников, могут отличаться формат (TXT, XML, JSON) [4] и стандарт записи. Для анализа потока событий требуется преобразовать все события к единому виду, т.е. провести процедуру анализа исходного события в соответствии с заранее заданной для источника и типа события формулой нормализации. Нормализованное событие представляет собой совокупность некоторых полей, состав которых определен в рамках некоторой таксономии, заполненных данными из необработанного события согласно правилам, указанным в формуле нормализации. При этом нормализация полученных событий подразумевает парсинг (анализ данных исходного события) и маппинг (сопоставление изъятых данных таксономическим полям). Поэтому каждый коннектор между источником данных и SIEM-системой представляет собой скрипт или приложение, которое позволяет получать поток событий с источника и проводить процедуру нормализации, включая парсинг и маппинг.

Ниже представлен пример события, полученного от источника на syslog-сервер.

```
<118>Jan 23 11:20:26 NGIPSV NGIPS: Protocol: TCP,
SrcIP: 195.38.41.42, OriginalClientIP: ::, DstIP: 192.168.7.16,
SrcPort: 26449, DstPort: 80, TCPFlags: 0x0, IngressInterface:
eth2, EgressInterface: eth1, IngressZone: untrust, EgressZone:
dmz, DE: Primary Detection Engine (23737be0-de45-11e7-
818a-92ab33c876b8), Policy: Access_Control_Policy_DMZ,
ConnectType: Start, AccessControlRuleName: Ac-
cess_to_DMZ, AccessControlRuleAction: Allow, Prefilter
Policy: Unknown, UserName: No Authentication Required,
UserAgent: Mozilla/5.0 (Windows NT 10.0; WOW64; Tri-
dent/7.0; rv:11.0) like Gecko, Client: Internet Explorer, Cli-
entVersion: rv:11.0, ApplicationProtocol: HTTP, Initia-
torPackets: 3, ResponderPackets: 1, InitiatorBytes: 656, Re-
sponderBytes: 66, NAPPolicy: Balanced Security and Connec-
tivity, DNSResponseType: No Error, Sinkhole: Unknown,
HTTPReferer: http://www.company.ru/production/complex/,
ReferencedHost: www.company.ru, URLCategory: Unknown,
URLReputation: Risk unknown
```

В результате применения правил нормализации SIEM-системы MaxPatrol SIEM [2], было получено нормализованное событие. В данной системе используется внутрен-

ний стандарт, разработанный на основе стандарта Common Event Expression (CVE) [2,5].

```
"action": "allow",
"object": "connection",
"status": "success",
"datafield1": "Access_Control_Policy_DMZ",
"datafield2": "HTTP",
"dst.ip": "192.168.7.16",
"dst.port": 80,
"event_src.category": "IDS/IPS",
"event_src.hostname": "NGIPSV",
"event_src.title": "asa firepower",
"event_src.vendor": "cisco",
"importance": "info",
"protocol": "TCP",
"src.ip": "195.38.41.42",
"src.port": 26449,
"time": "2020-01-23T11:20:26Z"
```

Учитывая большое количество источников данных, генерируемых отдельными подсистемами различных элементов РТК необходимо осуществлять фильтрацию и укрупнение выборок событий, подлежащих анализу, применяя методы агрегации. Поток данных с источника может содержать однотипные события, отличающиеся значением одного или нескольких полей (например, временем регистрации). В подобных случаях целесообразно организовать процесс отбора таких событий, которые удовлетворяют условию заранее настроенного правила агрегации, и объединения их в одно агрегированное событие. Имея в распоряжении нормализованные и агрегированные данные о состоянии информационной безопасности инфраструктуры РТК, можно осуществлять действия, направленные на анализ происходящих в системе процессов и выявление инцидентов. Основным механизмом в решении данной задачи является применение методов корреляции.

ПРОБЛЕМЫ НОРМАЛИЗАЦИИ И КОРРЕЛЯЦИИ СОБЫТИЙ

Выявление корреляции событий – процесс обнаружения инцидентов информационной безопасности путем анализа потока нормализованных событий. При обнаружении в потоке событий такой их последовательности, которая указана в условии одного из заранее настроенных правил корреляции, регистрируется корреляционное событие.

Развитие рынка систем управления событиями и инцидентами безопасности невозможно без наукоемких разработок и развития методического аппарата. Так, в работах [4, 6] предложен подход к формированию экспертных знаний для разработки систем управления событиями безопасности. Данный подход основан на приоритизации определенных признаков объектов системы и эвристическом выявлении известных разновидностей скрытых несовместимостей между компонентами защиты устройств. Подход базируется на корреляции событий на основе шаблонов, за счет чего обеспечивается достаточная точность обнаружения аномалий и возможность расширяемости списка шаблонов, однако недостаточно учитывается взаимовлияние устройств друг на друга и большое число факторов, влияющих на все устройства. В свою очередь в [6] представлена система признаков для постро-

ения модели события в SIEM-системе. В [7] авторами предлагается методика визуализации метрик кибербезопасности как инструмент поддержки принятия решений в процессе анализа рисков. Схожее исследование с элементами визуализации данных приведено также в [8]. Основными же направлениями научных исследований в этой отрасли являются методы и механизмы нормализации, агрегации и корреляции событий.

В [6, 10] представлена классификация, которая различает сигнатурные и бессигнатурные методы корреляции. Сигнатурные методы подразумевают создание человеком неких правил определения инцидентов. Бессигнатурные основаны на обнаружении аномалий по принципу черного ящика, среди которых выделяются подходы, основанные на спецификации, и базирующиеся на интеллектуальном анализе данных (data-miningbased). Анализ рынка SIEM [9, 10] систем, позволил сформировать список наиболее распространенных к применению на практике методов:

Statistical – бессигнатурный метод корреляции событий, основанный на измерении двух или более переменных и вычислении степени статистической связи между ними.

RBR Rule-based (pattern based) – метод, в котором взаимосвязи между событиями определяются аналитиками в заранее заданных специфических правилах.

CBR Codebook (case based). Корреляция производится по подходящим векторам из предварительно заданной матрицы событий.

MBR (model based reasoning) — метод основан на абстракции объектов и наблюдения за ними в рамках модели.

Graph based. Корреляция определяется на основе поиска зависимостей между системными компонентами в графическом представлении и построении на их основе графа, который в случае обнаружения корреляции используется для поиска причины инцидента.

Несмотря на множество существующих методов выявления корреляций, производители SIEM-систем зачастую используют именно сигнатурные методы, поскольку они обладают большей гибкостью и на практике имеют большую эффективность. В большинстве случаев выявление причины необходимо для адекватной реакции на инцидент или осуществляется в контексте инцидента (например, пользователь изменил параметры запуска служб, необходимо выяснить причину). Наступлению каждого инцидента предшествуют различные события S (рис.1): сканирование сетевых ресурсов, попытки установить соединение по закрытым портам, подозрительные вложения в почтовое трафике и т.д.

Правило корреляции позволяет объединить и сгруппировать подобные предшествующие события для определения момента возникновения инцидента. Правило имеет триггер, срабатывающий по условию, счетчик и сценарий реакции. Часть систем включают интуитивно понятные правила в графическом режиме. Счетчики предназначены для подсчета количества совпадений по одному и тому же правилу. Правила корреляции могут включать условия различной сложности. Одним из наиболее практичных подходов является применение правил, ограничивающих глубину корреляции и разделение базы событий на онлайн и архивную части. Например, события, произошедшие за последние сутки, хранятся в онлайн базе, по истечению таймера помещаются в архивную часть. Для работы с

большим объемом данных применяются различные специализированные поисковые движки и инструменты визуализации.

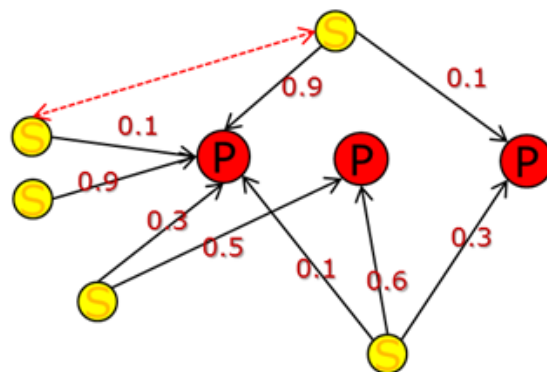


Рис. 1. Использование вероятностной модели

Одному инциденту могут предшествовать одно или несколько симптомов, которые возникают в результате события. Из одного события зачастую можно получить несколько симптомов. Симптомы одной проблемы могут быть как составные, так и взаимоисключающие.

Один и тот же симптом может приводить к одной или нескольким проблемам. При этом могут существовать вероятностные связи между проблемой и симптомом. Предположим, что в системном журнале ОС РТК было обнаружено событие о том, что был отключен один из компонентов сенсорики. Это событие потенциально может привести к нескольким инцидентам:

- если отключение конкретного компонента сенсорики влияет на работу других подсистем может привести к ошибкам навигации, позиционирования и выполнения оперативных задач РТК в целом.
- остановка работы компонента может привести к нарушению стабильности работы ОС.
- если изменение параметров запуска служб данного компонента разрешено только определенным администраторам или служба должна находиться в определенном состоянии, то возможно, что такое событие является предпосылкой нарушения политик доступа.

При использовании вероятностных моделей будет зафиксирован только один инцидент с максимальной вероятностью. Описав необходимое правило корреляции, можно контролировать, например, службы выявления вредоносных запросов. В случае применения бессигнатурных методов выявления корреляции между подобными единичными событиями будет затруднено. Например, метод graph based сможет их обнаружить, если возникнет событие, связанное со сбоем соединения между какой-либо службой.

Таким образом, можно сделать вывод, что основным механизмом корреляции событий, используемом на практике, является корреляция на основе правил, что обеспечивает достаточную точность выявления инцидентов, однако не позволяет обнаруживать неизвестные атаки.

Основным недостатком большинства перечисленных выше научно-исследовательских работ и коммерческих реализаций продуктов является актуальность проблемы потери деталей исходного объекта при моделировании.

Например, во время исполнения определенного процесса операционная система может передавать данные о его протекании в локальную систему логирования событий на данном устройстве. При этом будет передана только та часть информации о ходе выполнения процедуры, которая предусмотрена разработчиками, что приводит к потере некоторых деталей и упрощению исходной модели. После этого наступает второй этап, и система логирования осуществляет запись полученной информации в локальное хранилище либо передает его агенту-коллектору для отправки в SIEM-систему. Здесь в качестве ограничителей могут выступать выбранный формат хранения записей или протокол, используемый для передачи информации, где четко определены размеры полей данных. На этом этапе имеет место значительное упрощение модели и потеря части данных. В результате для обработки в SIEM поступает существенно многократно упрощенная модель реального события в виде дискретной записи (одного события в журнале). Во время нормализации событий полученная информация соотносится с определенными таксономическими полями. При этом количество полей в схеме не может отражать все возможные семантики всех событий от всех источников, что приводит к очередному упрощению модели из-за исключения некоторой информации. Если потерянная информация является актуальной для расследования конкретного инцидента, то эксперту или офицеру безопасности, проводящему расследование, необходимо либо обратиться к исходному событию, либо моделировать самому. В то же время большое количество полей, доступных при нормализации, приведет к ситуации дублирования смысловой нагрузки отдельных данных, когда одни и те же данные могут быть определены в различные поля нормализованного события.

МЕТОДИКА ВИЗУАЛИЗАЦИИ ИНЦИДЕНТОВ БЕЗОПАСНОСТИ

На сегодняшний день применение технологий визуализации стало эффективным средством выявления и решения возникающих проблем в управлении информационной безопасностью. Однако в отличие от визуализации физических явлений или зарегистрированных данных с различных датчиков визуальное представление сведений об информационной безопасности РТК подразумевает появление неких абстрактных данных. Визуализация информации о событиях и инцидентах в инфраструктуре интернета вещей предназначена для отображения данных и однозначной интерпретации, последующего принятия решений в отношении необходимости оперативной корректировки деятельности по реагированию на инциденты.

Для решения задачи визуального представления гетерогенных данных существует необходимость поиска и разработки методического обеспечения, которое позволяло бы организовать все этапы этого процесса. В [6, 7] представлены подходы к формализации данной задачи, а в [8] предложена методика для визуализации данных о состоянии топологии сетей различного масштаба. Тем не менее, основным недостатком данного методического обеспечения является его направленность в первую очередь на визуальное представление топологии сети и состояния сетевых объектов (хостов). Учитывая же представленные выше способы нормализации и корреляции событий, очевидно, что данные методики не подходят для

решаемой задачи, поскольку SIEM-системы предназначены для работы с различными типами элементов и входным материалом для анализа защищенности служат именно сообщения, поступающие от них по различным протоколам. Далее представлена методика визуализации данных, обрабатываемых системами управления событиями информационной безопасности с учетом вышеизложенных методов нормализации, агрегации и корреляции событий.

Шаг 1. Определение и выбор элементов РТК, являющихся источниками данных в рамках конкретной решаемой задачи. Для каждого элемента, используемого в конкретной задаче необходимо определить механизмы транспорта лог-сообщений (интерфейсы и протоколы взаимодействия) и организовать получение событий SIEM-системой.

Шаг 2. Провести анализ получаемых событий с целью выделения тех сообщений, которые имеют ценность для обнаружения инцидентов и подлежат нормализации в ходе обработки.

Шаг 3. В соответствии с рассмотренными выше методами разработать правила нормализации событий, определенных на шаге 2.

Шаг 4. В соответствии с рассмотренными выше методами определить необходимость и разработать способы агрегации сообщений, определенных на шаге 3. В случае отсутствия необходимости агрегации для конкретного типа сообщений данный шаг может быть пропущен.

Шаг 5. Основываясь на рассмотренных выше методах разработать алгоритмы и правила корреляций нормализуемых событий. Поскольку именно в результате обнаружения корреляций создаются записи об инцидентах, то именно они и будут являться теми самыми абстрактными данными, подлежащими визуализации. При этом необходимо определить принципиальные схемы событий, описывающих инциденты (определить субъекты, объекты, и производимые действия).

Методика основывается на положении, что в каждом событии можно выделить следующие элементы:

- субъект: инициатор взаимодействия, о котором повествует событие;
- объект: основная сущность, описываемая событием;
- источник: наблюдатель события, фактически подсистема определенного элемента конкретного РТК, сформировавшая и приславшая это сообщение.

Исходя из этого можно выделить следующие схемы событий с точки зрения взаимодействия субъекта и объекта.

- 1) В событии не содержится информации о каком-либо взаимодействии – источник передает информацию о состоянии объекта.
- 2) В событии есть информации только об объекте, но субъект подразумевается контекстом события.
- 3) В событии объект выполняет действие над самим собой.
- 4) В событии имеется информация о субъекте и объекте, а также их взаимодействии/

Шаг 6. Обеспечить накопление событий и записей об инцидентах в хранилище, обеспечивающем возможность

работы с большими данными и предоставляющим функции поиска, в том числе поиска с нечеткими условиями.

Шаг 7. Определить возможные для использования графические модели [6] и скорректировать их с учетом сценария работы SIEM-системы. При определении моделей возможно использование критериев эффективности восприятия подсистемы визуализации, предложенными в [6].

Шаг 8. В соответствии с определенными на шаге 7 моделями реализовать визуализацию данных об инцидентах, извлекаемых из хранилища с помощью средств поиска, посредством использования программных компонентов или отдельных продуктов.

АПРОБАЦИЯ МЕТОДИКИ

Провести сравнение предложенного методического аппарата с исследованиями других ученых [3-6] на основе количественных характеристик не представляется возможным ввиду их различия в части основных положений методик. В связи с этим был проведен эксперимент по практическому применению предложенной методики для решения задачи визуализации событий информационной безопасности в действующем центре интеллектуальных робототехнических систем ИПУ РАН.

На рис.2 представлена архитектура взаимодействия данных компонентов для организации возможности поиска, анализа и визуализации больших данных при анализе журналов событий с роя наземных БТС.

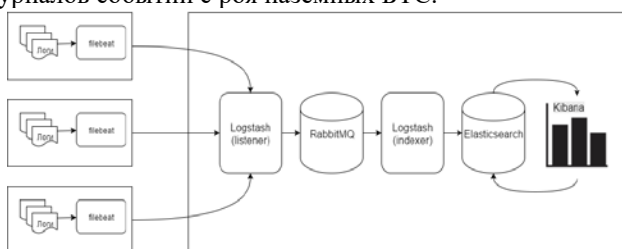


Рис. 2. Архитектура взаимосвязи компонентов поиска и визуализации

На базе стационарного компьютера был развернут стенд, имеющий следующие характеристики: 4-х ядерный процессор с тактовой частотой 3,1 ГГц; 16 Гб ОЗУ, на котором установлен стек ELK в составе: Elasticsearch 6.2.0, Logstash 6.2.0, Kibana 6.2.0. Стенд был развернут на базе операционной системы CentOS 7.

После настройки процесса отправки данных с тестовых элементов группы РТК было установлено, что средняя скорость индексирования в Elasticsearch составила 15000 пакетов в секунду при средней загрузке каждого из ядер процессора 50%. Однако, поскольку основной целью эксперимента была оценка времени поиска среди индексированных сообщений, то были предприняты попытки отправки запросов в Elasticsearch для поиска среди сообщений слова из 6 букв. При данном эксперименте среднее время поиска составило 1,02 мс.

В некоторых публикациях [5, 6] представлены данные смежных экспериментов. Ниже рассмотрены попытки сравнить полученные данные.

1. Аппаратные ресурсы. В большинстве рассмотренных примеров [3-6] используется распределение нагрузки на потоки в кластере из нескольких узлов (серверов). В данном случае использовались виртуальные машины на

базе одного физического стенда, характеристики которого указаны выше.

2. Скорость индексирования и анализа используемых данных. В связи с невозможностью проведения экспериментов при равных условиях на идентичном оборудовании проведение какой-либо количественной оценки по данному параметру не представляется возможным. Однако, такие сравнения вероятно будут проведены на следующих этапах исследования.

3. Механизм обработки данных. В данном эксперименте, применялась потоковая обработка, как и в большинстве исследований.

4. Пул задач. Все рассмотренные задачи в смежных исследованиях имеют конкретную постановку и получить результаты их сравнения в численном виде не представляется возможным. Однако, на следующем этапе исследования будет возможно применение предлагаемого методического обеспечения для решения различных задач и получения количественных оценок будущих результатов.

Помимо поисковых функций, стек ELK позволяет использовать проиндексированные данные для быстрой и качественной визуализации и представления пользователю.

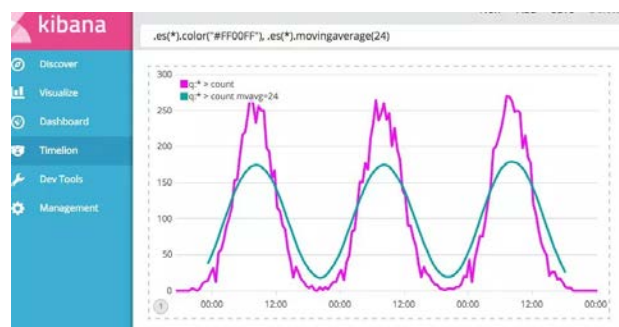


Рис. 3. Пример визуализации с помощью Kibana

На рис. 3 представлен пример визуализации данных с помощью Kibana. Визуализация информации увеличивает возможности ее восприятия оператором управления РТК, способствует улучшению качества обработки инцидентов операторами SIEM-систем и всех интеллектуальных процессов управления инцидентами ИБ.

БЛАГОДАРНОСТИ

Работа выполнена при частичной финансовой поддержке гранта РФФИ 19-08-00331 А.

ЗАКЛЮЧЕНИЕ

РТК и IoT относятся к числу важнейших концепций современного мира [11,12]. И хотя обе эти технологии имеют самостоятельное значение, в полной мере их потенциал раскрывается при их совмещении. Интернет вещей в своем развитии тесно переплетается с аналитикой больших данных. Задача состоит в том, чтобы выявить значимые элементы данных, получить на их основе новые знания и применить их для развития РТК и процессов управления ими [13-16]. Рассмотренные выше недостатки современных научных работ и коммерческих продуктов в области управления событиями безопасности приводят к низкому уровню работоспособности правил корреляции на практике и необходимости трудоемкой неавтоматизированной доработки подобных комплексов на каждом конкретном объекте. Совокупность ограничивающих фак-

Р.В. Мещеряков, С.Ю. Исхаков

торов может быть охарактеризована как отсутствие методологических основ для разработки правил корреляции.

Этот же недостаток в значительной степени ограничивает развитие такого перспективного направления как развитие методов поведенческого анализа объектов гетерогенных информационных инфраструктур, для которых SIEM-системы являются наиболее подходящим источником данных. Тем не менее, сформулированные выше недостатки свидетельствуют о невозможности формирования подобных методологий по разработке правил корреляции событий в отрыве от основных проблем в методическом и алгоритмическом обеспечении процесса управления инцидентами.

Таким образом, можно выделить следующие актуальные проблемы в исследуемой предметной области: 1) Потеря данных при нормализации, связанная с трансформациями моделей; 2) Отсутствие методического аппарата нормализации событий; 3) Отсутствие поддержки изменений модели объекта защиты; 4) Отсутствие методологии написания правил корреляции.

Проведенный анализ применения SIEM-систем в анализе событий, генерируемых РТК, показал, что визуализация информации в них – реальный инструмент повышения качества обеспечения защищенности инфраструктуры. Тем не менее в связи с большим объемом данных, генерируемых множеством устройств, сам компоненты и библиотеки визуализации данных не будут работоспособны без использования механизмов нормализации и аналитики событий с помощью методов корреляции.

ЛИТЕРАТУРА

1. Abomhara M. Cyber security and the internet of things: vulnerabilities, threats, intruders and attacks / M. Abomhara, G.M. Kien // *Journal of Cyber Security*. – 2015. – Vol. 4. – P. 65–88.
2. <https://www.ptsecurity.com/> (дата обращения 03.12.2019) MaxPatrol SIEM – URL: <https://www.ptsecurity.com/ru-ru/products/mpsiem/> (дата обращения: 03.12.2019).
3. Chen J. A Parallel Random Forest Algorithm for Big Data in a Spark Cloud Computing Environment / J. Chen, L. Kenli, Tang Z., Bilal K., Yu S., Weng C., Li K. // *IEEE Transactions on Parallel and Distributed Systems* – 2017. – vol. 28. – №. 4. – P. 919–933.
4. Dumitras T. Toward a Standard Benchmark for Computer Security Research: the Worldwide Intelligence NetworkEnvironment (WINE) / T. Dumitras, D. Shou // *Proceedings of the First Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS'11)* – 2011. – P. 89–96.
5. Shu X. Massive Distributed and Parallel Log Analysis For Organizational Security / X. Shu, J. Smiy, D. Yao, H. Lin // *IEEE Globecom Workshops*. – 2013. – P. 194–199.
6. Котенко И.В. Система сбора, хранения и обработки информации и событий безопасности на основе средств

Elastic Stack / И. В. Котенко, А. А. Кулешов, И. А. Ушаков // *Труды СПИИРАН*. – 2017. – № 54. – С. 5–34.

7. Милославская Н. Визуализация информации при управлении информационной безопасностью информационной инфраструктуры организации / Н. Милославская, А. Толстой, А. Бирюков // *Научная визуализация*. – 2014. – Т. 6, №2. – С. 74–91.

8. Семёнов Д.П. Визуализация процессов информационной безопасности // *Актуальные проблемы авиации и космонавтики*. – 2017. – Т. 2, №13. – С.230–232

9. <http://new.groteck.ru> (дата обращения: 23.07.2019). Решения для задач управления и мониторинга событий – URL: <http://new.groteck.ru/images/catalog/40353/6d54d68b1ef91ca522278b4e5b0791d0.pdf> (дата обращения: 23.07.2019).

10. <https://www.securitylab.ru> (дата обращения: 23.07.2019). Корреляция SIEM. – URL: <https://www.securitylab.ru/analytics/431459.php> (дата обращения: 23.07.2019).

11. Исхаков А.Ю. Концепция мультиагентной системы мониторинга «умного города» / А.Ю. Исхаков, С.А.К. Диане, А.О. Исхакова // *Известия Кабардино-Балкарского научного центра РАН*. – 2018. – № 6 (86), Ч. II. – С. 196–202.

12. Шумская О.О. Проблемы маскирования управляющих сигналов агентов мобильных робототехнических групп / О.О. Шумская, А.О. Исхакова // *XIII всероссийское совещание по проблемам управления ВСПУ-2019 : труды (Москва, 17-20 июня 2019 г.)*. – Москва, 2019. – С. 3086–3091.

13. Iskhakov A.Yu. Application of User Behavior Thermal Maps for Identification of Information Security Incident / A.Yu. Iskhakov, A.O. Iskhakova, R.V. Meshcheryakov, R. Bendraou, O. Melekhova // *SPIIRAS Proceedings*. – 2018. – № 6(61). – P. 147–171.

14. Кутахов В.П. Принципы формирования модели оптимизации системы роботизированных авиационных средств / В.П. Кутахов, Р.В. Мещеряков // *XIII всероссийское совещание по проблемам управления ВСПУ-2019 : труды (Москва, 17-20 июня 2019 г.)*. – Москва, 2019. – С. 1211–1214.

15. Исхаков А.Ю. Разработка мультиагентной технологии экологического картографирования мегаполиса / А.Ю. Исхаков, С.А.К. Диане, А.О. Исхакова // *Модели мышления и интеграция информационно-управляющих систем (ММИУС-2018) : материалы второй Международной научной конференции, посвящённой 25-летию юбилею Кабардино-Балкарского научного центра Российской академии наук (Нальчик, 04-09 декабря 2018 г.)*. – Нальчик, 2018. – С. 173–179.

16. Евсютин О.О. Обзор методов встраивания информации в цифровые объекты для обеспечения безопасности в «Интернете вещей» / О.О. Евсютин, А.С. Кокурина, Р.В. Мещеряков // *Компьютерная оптика*. – 2019. – № 1. – С. 137–154.

About the problems of data analysis in the robot security incident management system

Roman Meshcheryakov
ICS RAS
Moscow, Russia
mrv@ipu.ru

Sergey Iskhakov
TUSUR
Tomsk, Russia
iskhakov.sy@gmail.com

Abstract. The interference of intruders in the management of robots could not only impede the tasks assigned to them and, as a result, lead to financial losses, but also pose threats to human life and health. The problem of establishing mechanisms for timely incidents detection in group of robotics systems of various proportions was therefore urgent. One of the relevant vectors for the development of scientific knowledge in this industry is the improvement and development of new methods for detecting and responding to information security incidents in heterogeneous information systems through the analysis of discrete message flows. Despite a significant number of research activities in the field, most commercial products on the market adopted rule-based correlation techniques characterized by poor performance and the need for non-automated improvement at each concretely protected facility. These circumstances are due to problems such

as data loss in normalization associated with model transformation; absence of a methods for normalizing events in robotic complexes; support lack for changes in the protection scope model and the methodology for writing correlation rules. The paper proposed a technique for visualizing service data from robotics audit systems as a means of improving incident management methods and presented the results of its testing. Options for using modern tools for searching and visualization of big data are considered.

Keywords: robotics, normalization, aggregation, correlation, security information and event management system.

Дистанционные измерения параметров термокарстовых озер в Арктике для задач их моделирования

Муратов И.Н.

Югорский НИИ информационных технологий
г. Ханты-Мансийск, Россия
ildarmur@gmail.com

Байсаямова О.А.

Югорский НИИ информационных технологий
г. Ханты-Мансийск, Россия
oksanaabaysalyamova@gmail.com

Аннотация. Статья посвящена вопросам использования снимков высокого разрешения для исследования параметров термокарстовых озер в Арктической зоне. Дистанционные исследования проведены на 20 тестовых участках, расположенных достаточно равномерно на территории криолитозоны Большеземельской тундры. Для описания извилистости использован показатель формы пространственных объектов, принятый в картоведении, который рассчитывается по результатам измерения площади и периметра озер по спутниковым снимкам. Построены графики распределения числа озер и извилистости в зависимости от размеров озер в широком диапазоне размеров. Показано, что степень извилистости озерных границ проявляет в среднем положительный линейный тренд с ростом площади озер. Построена гистограмма распределения степени извилистости береговых границ озер в широком диапазоне их размеров от 50 м² до 20 га. Показано, что эмпирическое распределение озер по площадям может соответствовать логнормальному закону распределения.

Ключевые слова: дистанционные методы, спутниковые снимки, геоинформационные системы, вечная мерзлота, Большеземельская тундра, термокарстовые озера.

ВВЕДЕНИЕ

Озера и другие поверхностные водоемы как важные источники эмиссии парниковых газов в атмосферу в зонах мерзлоты являются объектами дистанционных исследований ввиду труднодоступности и высокой степени заболоченности северных территорий. Потепление климата последних десятилетий привело к значительному ускорению термокарстовых процессов в зоне многолетней мерзлоты, вызывающему рост концентрации метана в термокарстовых озерах. Проведенные в последнее время многочисленные дистанционные исследования динамики численности и размеров термокарстовых озер в Северной Евразии и Северной Америке [1-4] позволили создать геоимитационную модель динамики полей термокарстовых озер [5] пригодную для формирования прогнозных оценок динамики накопления озерного метана.

Пространственная структура этой модели определяется совокупностью окружностей со случайными размерами и случайным расположением [5]. Дистанционные исследования параметров реальных озер, проведенные в различных зонах мерзлоты (сплошная, прерывистая и островная) Западной Сибири, показали [6], что поля термокарстовых озер можно рассматривать как совокупности

случайных плоских фигур с существенно различающимися размерами (площадями) и с разной степенью извилистости границ.

Экспериментальные исследования возможности аппроксимации контуров озер окружностями были проведены на трех тестовых участках (ТУ), выбранных в разных зонах мерзлоты на территории Западной Сибири [7]. Результаты статистического анализа данных этих исследований подтверждают возможность замены контуров реальных озер окружностями с погрешностью около 5 %. В этих исследованиях использовались космические снимки среднего разрешения Landsat (разрешение 30 м), на которых не обнаруживаются озера малых размеров. Полученные по ним данные о параметрах озер относятся к озерам с размерами более 1-10 га. Поэтому представляет интерес изучить параметры озер с размерами от десятков квадратных метров до десятков га, для чего следует применить снимки высокого разрешения.

Актуальной является проблема расчета объемов эмиссии парниковых газов с территории Арктической зоны, как района сосредоточения термокарстовых озер. Известно, что масса накопленного в озерах метана, а следовательно и его эмиссия в атмосферу, определяется пространственными параметрами озер. Для формирования прогнозных оценок динамики накопления озерного метана необходимо уточнение геоимитационной модели динамики полей термокарстовых озер, с использованием данных о параметрах озер в Арктической зоне. Пока подобные данные для озер в широком диапазоне размеров имеются только для территории Арктической зоны в районе Западной Сибири [8]. В связи с этим целью работы явилось изучение пространственных параметров озер на территории Арктической зоны в районе Большеземельской тундры с использованием снимков высокого разрешения.

ДАННЫЕ И МЕТОДЫ ИССЛЕДОВАНИЯ

Исследования озер в настоящей работе проводились в Арктической зоне Большеземельской тундры с использованием космических снимков высокого разрешения Канопус-В (разрешение 2,1 м). Все снимки выбирались в достаточно короткий период летнего сезона (конец июня - август) для минимизации влияния сезонных колебаний

уровня воды в озерах. В этот период полностью исчезает ледовый покров на озерах, мешающий их выделению при дешифрировании снимков. Обработка космических снимков, проведенная с использованием стандартных средств геоинформационной системы ArcGIS 10.3, была направлена на получение данных о площадях и периметрах озер с целью определения закона распределения озер по размерам и по степени извилистости границ, изучения закономерностей изменения извилистости в зависимости от размеров озер.

Тестовые участки выбирались нами, как принято в подобных исследованиях, в местах скопления термокарстовых озер, т.е. в зонах, где существуют условия для формирования очагов термокарстовых процессов [9]. Схема размещения ТУ на исследуемой территории Большеземельской тундры представлена на рис. 1. Выбранные тестовые участки имели одинаковые размеры, площадь каждого участка составляла 3111 га. Суммарная площадь всех ТУ составила 62 220 га, что соответствует 0,42 % от всей площади исследованной территории Большеземельской тундры.

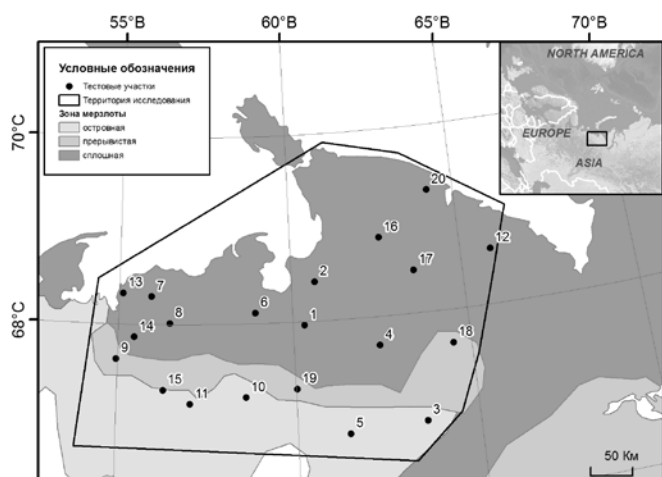


Рис. 1. Схема расположения тестовых участков

В нашем исследовании для оценки степени извилистости границ термокарстовых озёр был использован показатель формы z , введенный в [10] в следующем виде:

$$z = \frac{p^2}{4\pi S} \quad (1)$$

где p – периметр; S – площадь озера.

Исследование изменений показателя z было проведено на 20 тестовых участках, расположенных в трех разных зонах мерзлоты, на каждом из которых определяется несколько сотен или тысяч термокарстовых озёр. Для каждого озера были рассчитаны значения z , использованные в дальнейшем для определения закона распределения степени извилистости. Кроме этого на основе рассчитанных значений степени извилистости были определены для каждого ТУ средние значения (z_{cp}), использованные далее для анализа зависимости степени извилистости от размеров озер.

РЕЗУЛЬТАТЫ

На рис. 2 представлен график зависимости числа озер, полученных для каждого ТУ, от размеров озер. Заметим, что в связи с необходимостью графического представления числа озер в зависимости от размеров озер в очень широком их диапазоне по оси абсцисс на рис. 2 использована логарифмическая шкала для интервалов площадей озер, возможность чего была обоснована в [11].

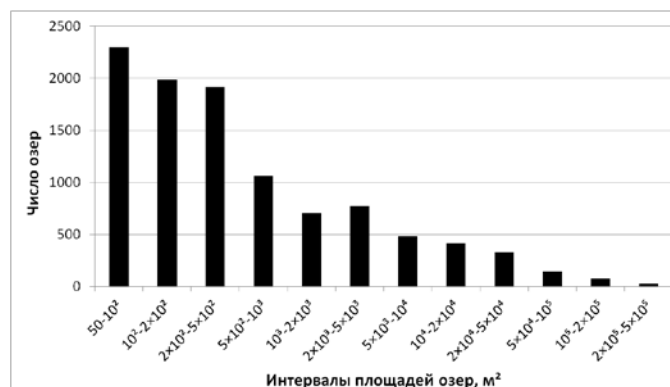


Рис. 2. Гистограмма распределения числа озер по размерам

График на рис. 2 демонстрирует плавное сокращение количества озер при увеличении их площадей. Однако при этом, наблюдается резкий спад количества озер при увеличении их размеров до 500 м². Также наблюдается небольшой рост количества озер в интервале 2000 – 5000 м², по сравнению с предыдущим интервалом. Резкий спад количества озер можно объяснить слиянием их между собой при достижении размеров более 500 м², в результате чего появляются новые озера большей площади, при этом общее количество озер значительно сокращается. Данная зависимость может быть обусловлена ландшафтными особенностями исследуемой территории, выраженной моренным рельефом. Рис. 3 демонстрирует расположение термокарстовых озер относительно друг друга на территории Большеземельской тундры – хорошо заметны крупные озера, образовавшиеся в результате слияния нескольких более мелких.

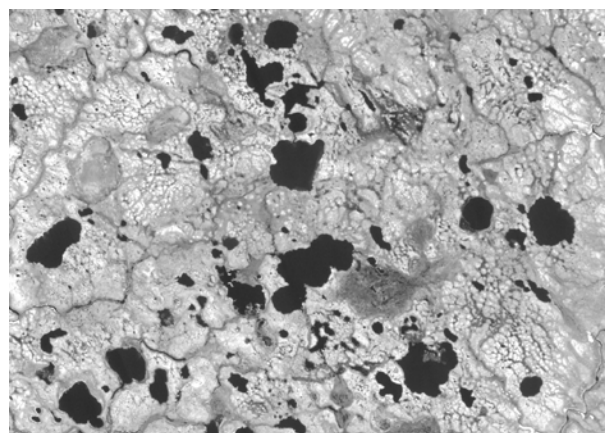


Рис. 3. Фрагмент космического снимка Канопус-В

На рис. 4 представлен график зависимости средних значений степени извилистости границ озер, полученных

для каждого ТУ, от размеров озер. Рис. 4 показывает, что зависимость степени извилистости границ озер от площади озер проявляет заметный линейный тренд с положительным коэффициентом тренда. Следовательно, степень извилистости береговых линий термокарстовых озер Западной Сибири увеличивается почти в 2 раза с ростом площади озер в интервале их размеров от 0,005 до 100 га.

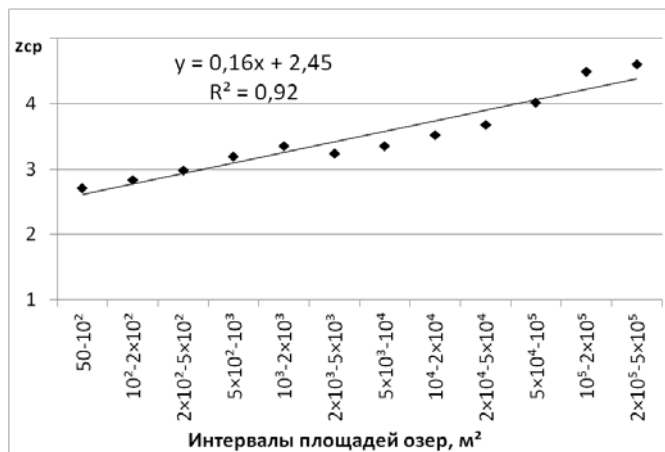


Рис. 4. Зависимость средней степени извилистости границ озер от их площади

Для задач моделирования полей термокарстовых озер и прогноза эмиссии озерного метана важное значение имеет знание законов распределения степени извилистости береговых границ озер. На рис. 5 приведена эмпирическая гистограмма распределения степени извилистости границ озер на основе данных о величине z , рассчитанных в соответствии с (1) по результатам дистанционных измерений на 20 ТУ на исследуемой территории Большеземельской тундры.

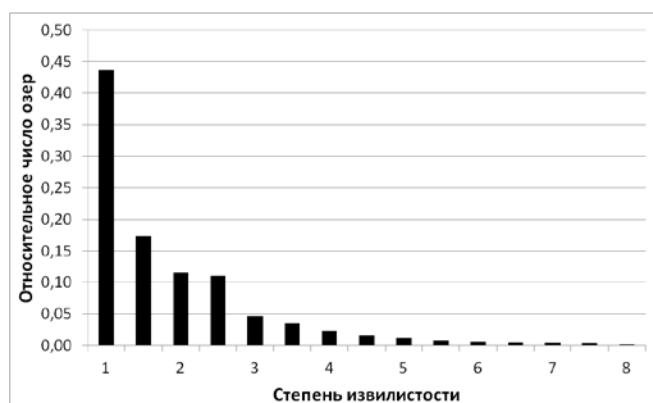


Рис. 5. Гистограмма распределения степени извилистости границ озер

По виду графика гистограммы на рис. 5 можно предполагать, что распределение степени извилистости будет соответствовать теоретическому логнормальному закону. Согласно [12] плотность вероятности для логнормального закона распределения степени извилистости (z) определяется уравнением:

$$f(z) = \frac{1}{\sqrt{2\pi}z\sigma} \exp\left(-\frac{(\ln z - \ln a)^2}{2\sigma^2}\right) \quad (2)$$

где a и σ – параметры логнормального закона распределения.

Оценка параметров логнормального закона распределения извилистости озерных границ Западной Сибири может быть произведена на основе выборочных данных. Используя полученные данные о степени извилистости в широком диапазоне размеров озер, можно определить параметры логнормального распределения площадей озер Большеземельской тундры. Соответствие экспериментально полученного по снимкам высокого пространственного разрешения распределения озер теоретическому логнормальному закону определяется с помощью стандартной процедуры проверки статистических гипотез о виде закона распределения [12].

ЗАКЛЮЧЕНИЕ

Рассмотрены методические вопросы исследования параметров термокарстовых озер в Арктической зоне Большеземельской тундры в достаточно широком диапазоне размеров озер на основе космических снимков высокого пространственного разрешения. Для проведения исследований в работе использованы спутниковые снимки высокого разрешения Канопус-В. Исследования проведены на 20 тестовых участках, расположенных достаточно равномерно во всех зонах многолетней мерзлоты Большеземельской тундры. Обработка снимков проведена в автоматическом режиме с использованием средств геоинформационной системы ArcGIS 10.3. Приведена расчетная формула для степени извилистости по результатам дистанционного измерения периметров и площадей озер.

Полученные экспериментальные данные позволили изучить зависимость извилистости от размеров озер. Показано, что в среднем извилистость озерных границ проявляет положительный линейный тренд с ростом площади озер.

Рассмотрены вопросы исследования распределения извилистости контуров озер в зоне многолетней мерзлоты Западной Сибири в достаточно широком диапазоне размеров озер на основе космических снимков высокого пространственного разрешения. В процессе обработки определялись периметры и площади термокарстовых озер. Была построена гистограмма распределения извилистости границ озер, представляющая озера с площадями от 50 м² до 20 га на территории криолитозоны Большеземельской тундры.

Выдвинуто предположение о том, что полученная эмпирическим путем гистограмма распределения соответствует теоретическому логнормальному закону. При выявлении соответствия этого закона эмпирической гистограмме появляется возможность использовать логнормальный закон распределения озер по площадям для моделирования полей термокарстовых озер в Арктической зоне и в других задачах.

Заметим, что учета влияния географических и климатических факторов на полученные по спутниковым снимкам данные не проводилось. Поэтому авторы исходят из

допущения о том, что полученные данные и определенные по ним значения параметров логнормального закона являются показательными для данного региона.

БЛАГОДАРНОСТИ

Работа выполнена в рамках государственного задания Департамента информационных технологий и цифрового развития Ханты-Мансийского автономного округа – Югры на 2020 г.

ЛИТЕРАТУРА

1. Kirpotin S. Abrupt changes of thermokarst lakes in Western Siberia: impacts of climatic warming on permafrost melting / S. Kirpotin, Y. Polishchuk, N. Bryksina // *International Journal of Environmental Studies*. – 2009. Vol. 66. No. 4. – P. 423–431.
2. Luoto M. Thermokarst ponds as indicator of the former distribution of palsas in Finnish Lapland / M. Luoto, M. Seppala // *Permafrost and Periglacial Processes*. – 2003. – Vol. 14. – P. 19–27.
3. Riordan B. Shrinking ponds in subarctic Alaska based on 1950-2002 remotely sensed images / B. Riordan, D. Verbyla, A.D. McGuire // *J. Geophys. Res.* – 2006. – Vol. 111.
4. Zuidhoff F.S. Changes in palsa distribution in relation to climate change in Laivadalén, Northern Sweden, especially 1960-1997 / F.S. Zuidhoff, E. Kolstrup // *Permafrost and Periglacial Processes*. – 2000. – Vol. 11. – P. 55–69.
5. Polishchuk V.Y. Modeling of thermokarst lake dynamics in West-Siberian permafrost. Ch. 6 / V.Y. Polishchuk, Y.M. Polishchuk // In: *Permafrost: Distribution, Composition and Impacts on Infrastructure and Ecosystems* / Edited by O. Pokrovsky. – NY: Nova Science Publishers, 2014. – P. 205–234.
6. Полищук В.Ю. Геоимитационное моделирование полей термокарстовых озер в зонах мерзлоты. / В.Ю. Полищук, Ю.М. Полищук. – Ханты-Мансийск: УИП ЮГУ, 2013. – 129 с.
7. Полищук Ю.М. Дистанционные исследования изменчивости формы береговых границ термокарстовых озер на территории многолетней мерзлоты Западной Сибири / Ю.М. Полищук, В.Ю. Полищук // *Исследования Земли из космоса*. – 2012. – № 1. – С. 61–64.
8. Полищук Ю.М. Исследование извилистости береговых границ термокарстовых озёр Западной Сибири по снимкам высокого разрешения «Канопус-В» / Ю.М. Полищук, И.Н. Муратов, В.Ю. Полищук // *Современные проблемы дистанционного зондирования Земли из космоса*. – 2019. – Т. 16. – № 5. – С. 130-137
9. Викторов А.С. Основные проблемы математической морфологии ландшафта. М.: Наука, 2006. – 252 с.
10. Берлянт А.М. Картоведение / А.М. Берлянт, А.В. Востокова, В.И. Кравцова, И.К. Лурье, К.А. Салищев. – М.: Аспект Пресс, 2003. – 478 с.
11. Полищук Ю.М. Методические вопросы построения обобщенных гистограмм распределения площадей озер в зоне мерзлоты на основе космических снимков среднего и высокого разрешения / Ю.М. Полищук, А.Н. Богданов, И.Н. Муратов // *Современные проблемы дистанционного зондирования Земли из космоса*. – 2016. – Т. 13. – № 6. – С. 224–232
12. Кремер Н.Ш. Теория вероятностей и математическая статистика. – М.: ЮНИТИ-ДАНА, 2004. – 573 с.

Remote measurements of thermokarst lakes in the Arctic for modeling purposes

Muratov I.N.
Space Services Centre
URIIT
Khanty-Mansiysk, Russia
ildarmur@gmail.com

Baysalamova O.A.
Space Services Centre
URIIT
Khanty-Mansiysk, Russia
oksanabaysalyamova@gmail.com

Abstract. The article is devoted to the issues of using high resolution images to study parameters of thermokarst lakes in the Arctic zone. Remote sensing was carried out on 20 test sites, located quite evenly on the territory of the Bolshezemelskaya tundra. To describe the curvature we used the spatial object shape index, adopted in cartography, which is calculated by measuring the area and perimeter of the lakes using satellite images. Graphs of lake number and curvature distribution depending on lake sizes in a wide range of sizes are built. It is shown that the degree of curvature of lake boundaries shows an average positive linear trend with the growth of lake area. A histogram of the distribution of the degree of curvature of lake shore borders in a wide

range of their sizes from 50 m² to 20 ha is constructed. It is shown that the empirical distribution of lakes by areas can correspond to the log-normal law of distribution.

Keywords: remote sensing, satellite imagery, geoinformation systems, permafrost, Bolshezemelskaya tundra, thermokarst lakes.

Анализ и синтез структуры пространственных данных для управления сложными распределенными системами

Е.С. Брекоткина

Институт экономики и управления
Уфимский государственный авиационный технический университет
Уфа, Россия
e-mail: brekotkina@mail.ru

О.И. Христодуло

Факультет информатики и робототехники
Уфимский государственный авиационный технический университет
Уфа, Россия
e-mail: o-hristodulo@mail.ru

А.С. Павлов

Факультет информатики и робототехники
Уфимский государственный авиационный технический университет
Уфа, Россия
e-mail: pavlov.alex@hotmail.com

В.Д. Трубин

Факультет информатики и робототехники
Уфимский государственный авиационный технический университет
Уфа, Россия
e-mail: refracto@yandex.ru

С.В. Павлов

Факультет информатики и робототехники
Уфимский государственный авиационный технический университет
Уфа, Россия
e-mail: psvgis@mail.ru

Аннотация. В связи с появлением новых источников и технологий получения данных о местоположении объектов и систем в целом, а также их интеграции в методы и алгоритмы управления в реальном времени – сформировались условия для более современного и качественного использования компьютерных технологий в управлении сложными распределенными системами, которые характеризуются существенной удаленностью составляющих компонент друг от друга и существенным и быстрым во времени изменением их характеристик. Одним из наиболее распространённых классов сложных распределённых систем являются большие предприятия и организации, структурные и/или производственные подразделения которых размещаются на значительном (от нескольких сот метров до десятков и сотен километров) расстоянии друг от друга и от центра принятия решений (центра управления). Без нарушения общности рассуждений, а так же для наглядности и конкретности примеров, в настоящей статье в качестве сложной распределённой системы рассматривается Уфимский государственный авиационный технический университет (далее Университет). Для эффективного получения и использования всей необходимой информации предлагается организовать ее хранение в одной распределённой базе данных (РБД) Университета, в свою очередь для синтеза структуры РБД в соответствии с общепринятой технологией проектирования информационных систем необходимо провести анализ требований ко всей информации Университета в целом и по отдельным подразделениям.

Предложен формальный метод синтеза структуры пространственных данных в сложных распределенных системах,

который позволяет создавать распределенную базу пространственных данных и организовывать доступ к ней различных пользователей, включая алгоритмы и методы управления.

Ключевые слова: сложные распределенные системы, информационная поддержка управления, модели и методы управления, пространственные данные, базы данных, модели данных, синтез структуры данных.

ВВЕДЕНИЕ

Современное развитие мировой экономики (в том числе в России и её регионах) во многом определяется таким глобальным научным и технологическим направлением как четвертая промышленная революция (4 IR), которое в Европе называют индустрия 4.0, а в России – цифровая экономика [1]. Это глобальное явление, которое, по мнению многих экспертов, в корне изменит образ жизни, стиль работы, принципы и методы взаимодействия людей и других объектов и систем реального мира на основе повсеместного и комплексного использования современных аппаратно-программных комплексов. Новые научные направления, технологии и инновации объединяют физический, биологический и цифровой миры таким образом, чтобы постоянно получать данные о состоянии объектов этих миров и использовать их для дальнейшей оптимизации процессов управления. Это объединение можно рассматривать как сложную взаимосвязанную совокупность объектов различной сложности и физической природы. Успех создания новых и использования существующих и вновь создаваемых объектов и их совокупностей (систем) во многом зависит

от умения формально описывать различные процессы, происходящие в них, что позволит повысить эффективность управления такими сложными системами.

В связи с появлением новых источников и технологий получения данных о местоположении объектов и систем в целом, а также их интеграции в методы и алгоритмы управления в реальном времени – сформировались условия для более современного и качественного использования компьютерных технологий в управлении сложными распределенными системами, которые характеризуются существенной удаленностью составляющих компонент друг от друга и существенным и быстрым во времени изменением их характеристик. В процессе управления такими системами возникает двойственность ситуации: с одной стороны модели и методы управления используют большие объемы разнородных данных, на обработку которых требуется много времени, с другой стороны – имеются существенные ограничения на время принятия управленческих решений на основе этих данных. Эти особенности изменяют уязвимость этих систем, так как приводят к возникновению новых угроз функционированию и использованию сложных распределенных систем, которые определяются включением в эти системы новых более сложных компонентов (подсистем), нарушение нормального функционирования

которых может привести к нарушению целостности системы и ее неправильному функционированию и использованию.

ПРОСТРАНСТВЕННАЯ ИНФОРМАЦИЯ В СЛОЖНЫХ РАСПРЕДЕЛЕННЫХ СИСТЕМАХ СИНТЕЗ СТРУКТУРЫ ПРОСТРАНСТВЕННЫХ ДАННЫХ

Одним из важных направлений фундаментальной научной проблемы обеспечения создания и использования сложных распределенных систем является определение соответствия желаемых характеристик объектов их характеристикам в настоящем и будущем. Сложность этой проблемы определяется сложностью составляющих компонентов и подсистем, большим количеством определяющих и управляемых параметров, нечеткостью требований и изменчивостью целей управления со стороны различных пользователей результатов функционирования систем и участников процесса управления.

Одним из наиболее распространенных классов сложных распределенных систем являются большие предприятия и организации, структурные и/или производственные подразделения которых размещаются на значительном (от нескольких сот метров до десятков и сотен километров) расстоянии друг от друга и от центра принятия решений (центра управления). Без нарушения общности рассуждений, а также для наглядности и конкретности примеров, в

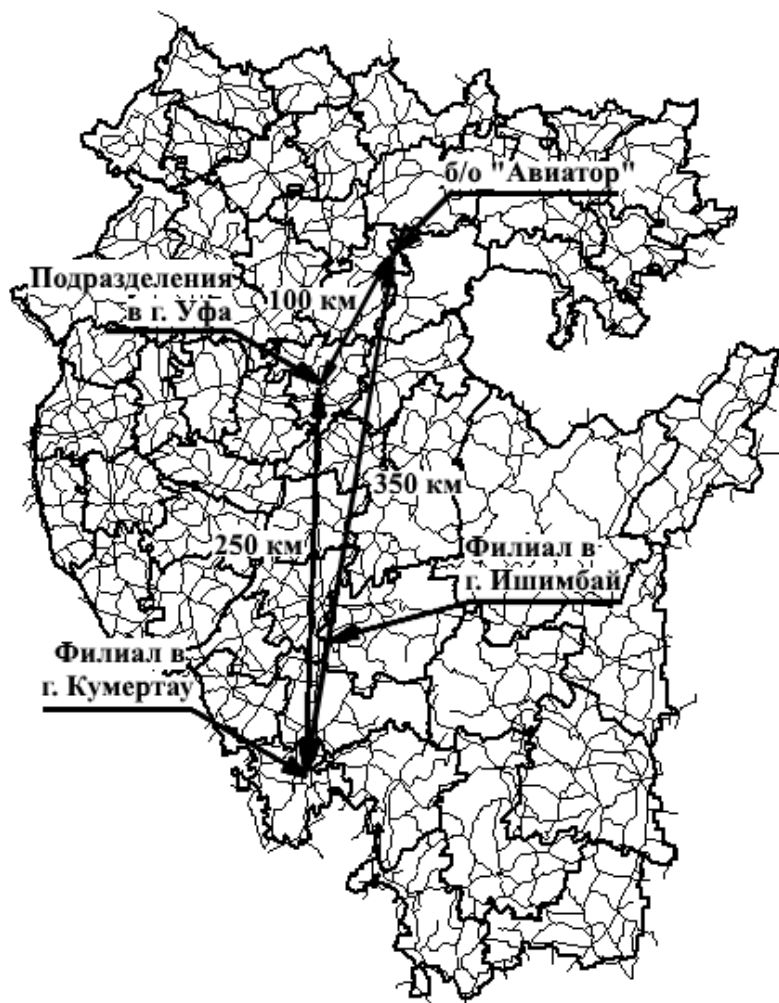


Рис. 1. Схема размещения подразделений Университета на территории Республики Башкортостан

настоящей статье в качестве сложной распределённой системы рассматривается Уфимский государственный авиационный технический университет (далее Университет), схема размещения крупных подразделений которого на территории республики Башкортостан приведена на рис. 1.

Расстояние между наиболее удалёнными подразделениями университета (филиал в г. Кумертау и база отдыха «Авиатор») составляет по прямой примерно 350 км а по автомобильной дороге приблизительно 400 км. Максимальное удаление подразделения (филиал в г. Кумертау) от центра принятия решений (ректорат в г. Уфа) составляет 250 и 300 км соответственно.

Качественное управление всеми видами деятельности университета (учебной, научной, воспитательной, хозяйственной, финансово-экономической и др.) в современных

условиях может быть организована только с использованием программно-технических комплексов получения и обработки всей необходимой для целей управления информации. При этом необходимо учитывать, что для управления используется как информация об Университете так и об окружающей его среде: экономической, демографической, природной, транспортной и др. В данной статье авторы рассматривают специфические аспекты анализа и структурирования преимущественно пространственных данных для целей управления сложными распределёнными системами, поэтому обобщённая схема возникновения и использования пространственной информации для Университета представлена на рис. 2.

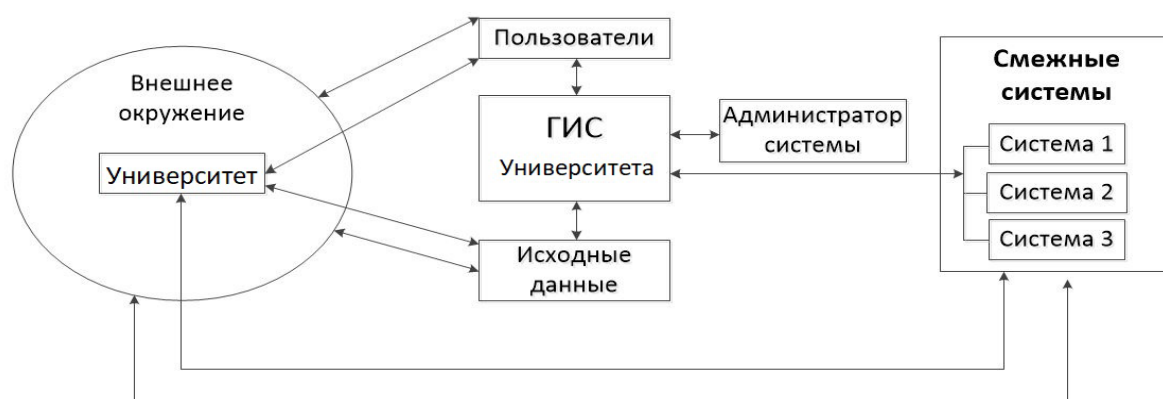


Рис. 2.Обобщённая схема возникновения и использования пространственной информации для Университета

В работах [2,3,4] авторами показано, что для управления сложными распределёнными системами большое количество разнородной пространственной информации создается и используется большим количеством специалистов различного профиля. При этом информация хранится и используется (см. рис. 1) в различных, территориально удалённых и организационно самостоятельных подразделениях (подсистемах), но в тоже время необходимо обеспечить доступ ко всей этой информации большого количества пользователей, в том числе и для её использования в формализованных методах и алгоритмах управления. Для достижения этой цели в настоящей работе предлагается один из возможных подходов к анализу и синтезу структуры пространственных данных.

СИНТЕЗ СТРУКТУРЫ ПРОСТРАНСТВЕННЫХ ДАННЫХ

Вся возникающая и используемая информация для целей управления в различных подразделениях Университета может быть условно подразделена на два класса: специализированные данные (СД) - описывающие объекты и процессы, связанные с деятельностью одного конкретного

подразделения; общие данные (ОД) - используемый двумя и более (в предельном значении – всеми) подразделениями Университета.

Для эффективного получения и использования всей необходимой информации предлагается организовать ее хранение в одной распределённой базе данных (РБД) Университета, в свою очередь для синтеза структуры РБД в соответствии с общепринятой технологией проектирования информационных систем необходимо провести анализ требований ко всей информации Университета в целом и по отдельным подразделениям. Для формализованного описания всех частей информации и связей между ними введём следующие обозначения: базу данных i -ого подразделения обозначим как $БД_i$, $i = \overline{1, n}$, где n - общее число подразделений для которых создаются базы данных. В силу введённых обозначений

$$РБД = \bigcup_{i=1}^n БД_i. \quad (1)$$

Учитывая разделение всей информации на два класса -

ОД и СД, все специализированные данные i -ого подразделения предлагается хранить в отдельной базе данных - БСД $_i$, а ОД хранить в общей для всех подразделений БОД. В силу введенных обозначений

$$\begin{aligned} \text{РБД} &= \text{БОД} \bigcup \text{БСД}_1 \bigcup \text{БСД}_2 \bigcup \dots \bigcup \text{БСД}_n \\ &= \text{БОД} \bigcup \left(\bigcup_{i=1}^n \text{БСД}_i \right). \end{aligned} \quad (2)$$

Учитывая (1) справедливо соотношение

$$\text{РБД} = \bigcup_{i=1}^n \text{БД}_i = \text{БОД} \bigcup \left(\bigcup_{i=1}^n \text{БСД}_i \right). \quad (3)$$

Каждая из принятых к рассмотрению баз данных разрабатывается на основе соответствующей модели данных (МД), поэтому для разработки структуры РБД предлагается метод синтеза структуры или модели базы данных на основе анализа и взаимно-однозначного (функционального) преобразования моделей входящих в неё баз пространственных данных.

Обозначим модель данных каждой из БД i через МД $_i$, $i = \overline{1, n}$. Каждая база пространственных данных содержит тематические слои пространственных (плано-картографических) данных, атрибутивные данные объектов этих слоев и связи между отношениями (таблицами) и слоями, что можно отразить в модели данных следующим образом:

$$\text{МД}_i = \{(R_1^i, R_2^i \dots R_{k_i}^i, S_1^i, S_2^i \dots S_{m_i}^i, C_1^i, C_2^i \dots C_{t_i}^i)\}, i = \overline{1, n}, \quad (4)$$

где k_i – количество отношений R в модели данных i -ой БД $_i$, m_i – количество слоев S в модели данных i -ой БД $_i$, t_i – количество связей C между отношениями и слоями модели данных i -ой БД $_i$.

Представим модель данных i -ой БД $_i$ как совокупность двух частей, описывающих атрибутивную - МД $_i^A$ и пространственную - МД $_i^S$ информацию:

$$\text{МД}_i = \text{МД}_i^A \cup \text{МД}_i^S. \quad (5)$$

В каждой i -ой МД $_i^A$ выделим две части: МД $_i^{1A}$ – часть МД $_i^A$, описывающая информацию, входящую в состав базы общих данных, МД $_i^{2A}$ – часть МД $_i^A$, описывающая информацию, входящую в состав базы специализированных данных. Тогда

$$\text{МД}_i^A = \text{МД}_i^{1A} \cup \text{МД}_i^{2A}. \quad (6)$$

Введем в рассмотренное множество N_i^A номеров различных отношений из МД $_i^A$:

$$N_i^A = \{1, 2, 3, \dots, k_i\}. \quad (7)$$

По аналогии с (6) в N_i^A выделим две части

$$N_i^A = N_i^{1A} \cup N_i^{2A}, \quad (8)$$

где N_i^{1A} – определим как множество номеров отношений из МД $_i^A$, входящих только в эту модель данных

$$N_i^{1A} = \{j: R_j^i \in \text{МД}_i \text{ и } R_j^i \notin \text{МД}_k \text{ при } k \neq i\}. \quad (9)$$

N_i^{2A} определим как множество номеров отношений из МД $_i^A$, входящих еще хотя бы в одну другую модель данных.

$$N_i^{2A} = \{j: R_j^i \in \text{МД}_i \text{ и } \exists l: l \neq i, R_j^i \in \text{МД}_l\}. \quad (10)$$

Учитывая введенные обозначения:

$$\begin{aligned} \text{МД}_i^{2A} &= \{R_j^i\}, j \in N_i^{1A}, \\ \text{МД}_i^{1A} &= \{R_j^i\}, j \in N_i^{2A}. \end{aligned} \quad (11)$$

Тогда модель атрибутивных данных БОД можно представить как совокупность всех отношений, входящих хотя бы в две МД $_i$, $i = \overline{1, n}$

$$\text{МОД}^A = \{R_j^i\}, j \in \bigcup_{i=1}^n N_i^{2A}, i = \overline{1, n}. \quad (12)$$

В свою очередь модель специализированных атрибутивных данных i -го подразделения

$$\text{МСД}_i^A = \{R_j^i\}, j \in N_i^{1A}, i = \overline{1, n}. \quad (13)$$

По аналогии с МД $_i^A$ в каждой МД $_i^S$ можно провести подобные преобразования и получить для МОД S и МСД $_i^S$ соотношения, аналогичные (12) и (13). Тогда с учетом соотношения (3) модель данных всей РБД Университета может быть представлена следующим образом:

$$\text{МД}(\text{РБД}) = \text{МОД} \bigcup \left(\bigcup_{i=1}^n \text{МСД}_i \right), \quad (14)$$

а МОД и МСД $_i$ определяются преобразованиями (4) – (13).

Построение модели пространственных данных всей РБД Университета позволяет создать базу пространственных данных, удовлетворяющую потребностям и требованиям всех подразделений и внешних пользователей. Поскольку при создании БОД производится ряд преобразований всех БД $_i$ путем переноса общих данных в БОД необходимо установить соответствие (естественно, прозрачное для конечных пользователей) между хранением пространственных данных в БД $_i$ до преобразования и новым местом хранения этих данных в БОД. Для этого предлагается построить набор отображений F_i , таких что

$$F_i: \text{МСД}_i \rightarrow \text{МОД}, i = \overline{1, n}. \quad (15)$$

На практике (в реальной информационной системе), эти отображения задаются в виде отдельных таблиц базы метаданных, содержащих фактические места (адреса) хранения слоев и отношений всех пространственных данных.

ЗАКЛЮЧЕНИЕ

Реализация предложенного подхода организации хранения пространственных данных Университета позволит организовать доступ различных пользователей (как внутренних, так и внешних) ко всей необходимой для управления информацией, причем пользователями могут быть как люди, так и формальные методы и алгоритмы управления.

Результаты исследований, представленные в статье, поддержаны грантом РФФИ №20-08-00301\20 – А «Мето-

дологические и теоретические основы управления уязвимостью на основе парирования угроз в сложных распределенных системах»

ЛИТЕРАТУРА

1. Шваб К. Четвертая промышленная революция / К. Шваб – «Эксмо» (Top Business Awards), 2016 г. – 137с.
2. Бахтизин Р.Н, Павлов С.В., Павлов А.С., Сайфутдинова Г.М. Создание инфраструктуры пространственных данных Республики Башкортостан на основе геоинформационных технологий. Уфа: Нефтегазовое дело, 2008 г - 103 с.
3. Брекоткина Е.С., Павлов С.В., Павлов А.С. Метод организации больших массивов данных региона в рамках пере-

хода к цифровой экономике // Управление экономикой: методы, модели, технологии: материалы XVIII Международной научной конференции / Уфимск. гос. авиац. техн. ун-т. – Уфа : РИК УГАТУ, 2018. – С. 34 – 37.

4. Брекоткина Е.С., Брекоткин В.Е., Павлов С.В., Павлов А.С. Организация межведомственной системы электронного документооборота как компоненты больших данных цифровой экономики региона // Информационные технологии интеллектуальной поддержки принятия решений: труды VII Всероссийской научной конференции (с приглашением зарубежных ученых). май 28-30, 2019, Уфа, Россия: том 1. – Уфа, 2019 – С. – 130-134.

Analysis and synthesis of spatial data structure for managing complex distributed systems

E. S. Brekotkina
Institute of Economics and management
Ufa state aviation technical University
Ufa, Russia
e-mail: brekotkina@mail.ru

A. S. Pavlov
Faculty of computer science and robotics
Ufa state aviation technical University
Ufa, Russia
e-mail: pavlov.alex@hotmail.com

S. V. Pavlov
Faculty of computer science and robotics
Ufa state aviation technical University
Ufa, Russia
e-mail: psvgis@mail.ru

O. I. Hristodulo
Faculty of computer science and robotics
Ufa state aviation technical University
Ufa, Russia
e-mail: o-hristodulo@mail.ru

V. D. Trubin
Faculty of computer science and robotics
Ufa state aviation technical University
Ufa, Russia
e-mail: refracto@yandex.ru

Annotation. Due to the emergence of new sources and technologies for obtaining data on the location of objects and systems in General, as well as their integration into real-time management methods and algorithms, conditions have been created for a more modern and high-quality use of computer technologies in the management of complex distributed systems, which are characterized by a significant distance of the components from each other and a significant and rapid change in their characteristics over time. One of the most common classes of complex distributed systems is large enterprises and organizations whose structural and / or production units are located at a significant distance (from several hundred meters to tens or hundreds of kilometers) from each other and from the decision-making center (control center). In this article, the Ufa state aviation technical University (hereinafter referred to as the University) is considered as a complex distributed system without violating the generality of judgments, as well as for clarity and concreteness of the application. In order to effectively

obtain and use all the necessary information, it is proposed to organize its storage in a single distributed database (RBD) of the University. In turn, to synthesize the structure of the RBD in accordance with the generally accepted technology for designing information systems, it is necessary to analyze the requirements for the Entire University's information as a whole and for individual subdivisions.

A formal method for synthesizing the structure of spatial data in complex distributed systems is proposed. This method allows creating a distributed database of spatial data and organizing access to it for different users, including algorithms and management methods.

Keywords: complex distributed systems, management information support, management models and methods, spatial data, databases, data models, data structure synthesis..

Информационная поддержка принятия решений для надзорного контроля над некредитными финансовыми организациями

Перевозчиков Н.Д., Головин О.К.

Самарский университет

Самара, Россия

nikiperevozchikov@yandex.ru

Аннотация. С изменением регуляторного ландшафта в финансовой сфере к организации деятельности некредитных финансовых организаций (НФО) предъявляется все больше и больше требований со стороны государственных контролирующих органов. Для повышения эффективности надзора за НФО применяется системы, автоматизирующие надзорную деятельность контролирующих органов, и, в частности, системы поддержки принятия решений. В данной работе представлена система поддержки принятия решений, которая позволяет планировать надзорные мероприятия над НФО в соответствии с принципом риск-ориентированного надзора, при котором осуществляется концентрация ресурсов на НФО с максимальным объемом рисков или масштабом деятельности. Разработанная система позволяет отслеживать и предотвращать нарушения, а также корректировать правильность действий в отношении НФО. Кроме этого, система позволяет планировать надзорные мероприятия, изменять режимы надзора за НФО и контролировать сроки. Применение разработанной системы позволяет повысить эффективность деятельности контролирующих органов в части контроля и надзора как за отдельными НФО, так и за финансовым рынком в целом.

Ключевые слова: система надзора, планирование мероприятий, режимы надзора, некредитные организации, риск-ориентированный надзор.

ВВЕДЕНИЕ

Некредитные финансовые организации (НФО) уже несколько лет находятся в довольно сложной ситуации [1]. С изменением регуляторного ландшафта в финансовой сфере к организации деятельности НФО предъявляется все больше и больше требований со стороны контролирующих органов [2, 3].

Для повышения эффективности надзора за НФО применяется автоматизация надзорной деятельности контролирующих органов, например, путем разработки систем мониторинга, сбора отчетности, планирования, прогнозирования и поддержки принятия решений. Такие системы становятся всё более интеллектуальными и дружелюбными пользователю [4], что позволяет их эффективно внедрять в различных сферах деятельности. Допустимо их применение в реальном масштабе времени для принятия ответственных решений [5, 6].

Важным критерием при организации эффективного процесса надзора за НФО является обеспечение устойчи-

вости, при этом могут выделяться краткосрочные (например, влияние на клиента НФО) и долгосрочные цели (например, влияние деятельности НФО на финансовый рынок). На основе системного анализа информации разрабатывается поэтапный план мероприятий, служащий для достижения глобальных целей [7]. В процессе системного анализа выделяются паттерны типичных ситуаций и типичные реакции на них [8], нужные для поддержки принятия решений. Помимо анализа информации о деятельности НФО, для эффективной поддержки принятия решений требуется всесторонний обзор характеристик организационного и системного уровня [9], при этом требуется учитывать аспекты безопасности [10].

В последнее время активно развиваются технологии искусственного интеллекта, которые в будущем позволят применять новые технологии для поддержки принятия решений или для полной замены лиц, принимающих решения [11], но к официальному регулированию финансового рынка такие технологии не могут быть применены в силу недостаточной проработанности.

В данной работе представлена система информационной поддержки принятия решений, которая позволяет планировать надзорные мероприятия над НФО в соответствии с принципом риск-ориентированного надзора, при котором осуществляется концентрация ресурсов на НФО с максимальным объемом рисков или масштабом деятельности.

РИСК-ОРИЕНТИРОВАННЫЙ НАДЗОР

В соответствии с требованиями нормативно-правовой документации, за НФО предусмотрена различная периодичность и интенсивность надзорных мероприятий с концентрацией на выявленных рисках в деятельности. Другими направлениями риск-ориентированного надзора являются различный объем и периодичность представления отчетности как одного из основных источников информации для куратора, а также различная периодичность проверки. Под режимом надзора понимается порядок осуществления надзорных мероприятий с заданной периодичностью, интенсивностью и концентрацией на выявленных рисках в деятельности НФО, который устанавливается в зависимости от категории «размера и значимости» и риск-профиля поднадзорной НФО на основании матрицы режимов надзора.

В зависимости от воздействия НФО на финансовую систему, финансовый сектор и потребителей финансовых

Н.Д. Перевозчиков, О.К. Головин

услуг, профильное надзорное подразделение, исходя из размера компании, оборотов, величины клиентской базы, социальной значимости и других факторов, относит организацию к следующим категориям “размер и значимость”: крупные и системно значимые, средние, малые. Риск-профиль НФО определяется исходя из вероятности наступления и значимости неблагоприятных событий для финансовой устойчивости НФО или осуществления профильной деятельности на финансовом рынке. Риск-профиль НФО – это общее описание или матрица основных рисков, с которыми сталкивается НФО, с указанием уровня влияния (например, высокий, средний, низкий) и вероятности наступления событий. Профильное надзорное подразделение устанавливает методику определения риск-профиля, содержащую порядок определения риск-профиля НФО, включая перечень оцениваемых показателей и критерии отнесения, периодичность и условия пересмотра риск-профиля. Методика определения риск-профиля должна предусматривать комплексную оценку рисков в зависимости от вида НФО с учетом специфики бизнес-модели НФО. Отчетность является одним из основных источников информации для куратора. В то же время при снижении интенсивности надзора за НФО в зависимости от ее категории, целесообразно снижать объем и периодичность отчетности в целях оптимизации нагрузки на поднадзорные НФО.

МОДЕЛЬ ПРЕДМЕТНОЙ ОБЛАСТИ

Для описания основных сущностей предметной области разработана модель (“рис. 1”).

Используя теоретико-множественный подход, опишем модель предметной области для поддержки принятия решений.

Представим процесс надзора в виде последовательности мероприятий, представленных в дискретные моменты времени t :

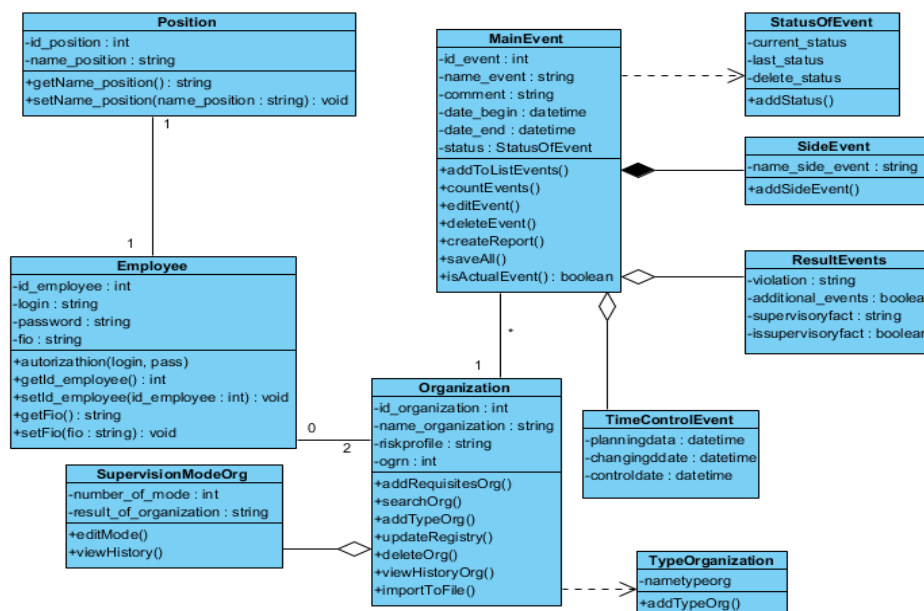


Рис. 1. Модель предметной области

$$\{y^t\}, t = \overline{1, m}. \quad (1)$$

Множество непротиворечивых оценок мероприятий представим:

$$\{w^l\}, l = \overline{1, L}. \quad (2)$$

w^l может быть сложной оценкой, которая состоит из нескольких простых.

Выбрана функция, определенная в моменты времени $t = \overline{1, m}$:

$$F(t, \theta) = (\theta, \varphi(t)), \quad (3)$$

где θ – вектор параметров $\theta = (\theta_1, \dots, \theta_k)$,

φ – векторная функция $\varphi = (\varphi_1, \dots, \varphi_k)$, компонентами которой являются неизвестные функции.

Задача поддержки принятия решения заключается в отыскании такой прогнозной последовательности, которая будет наиболее согласованной с составом мероприятий и их оценкой, т.е.:

$$y_t = (\theta, \varphi(t)), t = \overline{m+1, m+n}. \quad (4)$$

В качестве меры близости возьмем:

$$D(\theta) = \sum_{t=1}^m R(\tilde{y}_t - (\theta, \varphi(t))), \quad (5)$$

где R – строго выпуклая функция.

В результате получим задачу: $D(\theta) \rightarrow \min$, которая может быть сведена к задаче линейного динамического программирования:

$$D(\theta) = \sum_{t=1}^m |\tilde{y}_t - (\theta, \varphi(t))|. \quad (6)$$

АЛГОРИТМ ПЛАНИРОВАНИЯ МЕРОПРИЯТИЙ

Разработан алгоритм планирования мероприятий с использованием риск-ориентированного подхода (“рис. 2”).



Рис. 2. Алгоритм планирования мероприятий

Суть данного алгоритма заключается в том, что каждой организации назначается определенный режим надзора, в соответствии с которым формируется перечень мероприятий над этой организацией. У каждого мероприятия устанавливается период назначения мероприятия и период планового действия мероприятия.

Шаг 1. Выбор организации. На этом шаге определяется НФО, для которой требуется изменение перечня мероприятий. НФО определяется на основе анализа имеющихся сведений о НФО: размер и обороты, влияние на финансовый сектор, количество потребителей финансовых услуг.

Шаг 2. Назначение режима надзора. Определяется риск-профиль НФО и в соответствии с настраиваемой матрицей рисков назначается режим надзора.

Шаг 3. Определение перечня мероприятий. В соответствии с режимом надзора формируется перечень мероприятий в отношении НФО.

Шаг 4. Установка периода назначения мероприятий. Для каждого мероприятия в перечне назначается периодичность его проведения. Периодичность учитывает возможные выходные и праздничные дни, а также наличие внеплановых мероприятий в отношении НФО.

Шаг 5. Установка планового периода действия мероприятий. Для каждого мероприятия в перечне назначается плановый период действия.

ОСОБЕННОСТИ И ВОЗМОЖНОСТИ СИСТЕМЫ

В системе предусмотрены три роли пользователей: администратор, куратор, руководитель. Для работы в системе сотрудник должен авторизоваться в ней, вводя логин и пароль. В зависимости от роли пользователю будут доступны различные функции и возможности, описанные ниже.

АДМИНИСТРАТОР

Сотрудникам с ролью «Администратор» должны быть доступны следующие функции:

- назначение основных и резервных организаций сотрудникам;
- ведение внутреннего реестра организаций.

Назначение основных и резервных организаций сотрудникам. В базу данных должны быть загружены реестры субъектов рынка НФО, из которых каждому куратору назначаются основные организации. Одну организацию

можно назначить в качестве основной только одному сотруднику. Для удобства весь перечень организаций можно разбить на виды. Также предусмотрена возможность поиска конкретной организации по части государственного регистрационного номера или наименования. В качестве резервных куратору можно назначить те организации, которые уже назначены в качестве основных одному из других кураторов. Одна организация может быть назначена в качестве резервной нескольким кураторам. Организация не может быть одновременно и резервной, и основной у одного куратора.

Ведение внутреннего реестра организаций. Внесение реквизитов поднадзорных организаций, по каким-либо причинам отсутствующих в загруженных в реестр субъектов рынка микрофинансирования и саморегулируемые организации производится «Администратором» в окне «Внутренний реестр организаций». После заполнения всех предусмотренных полей во внутреннем реестре организация появится в окне «Назначение организаций», и ее можно будет назначить куратору в качестве основной или резервной организации.

КУРАТОР

Сотрудникам с ролью «Куратор» должны быть доступны следующие функции:

- планирование мероприятий в отношении поднадзорных НФО;
- просмотр истории изменения режимов надзора;
- настройка параметров работы куратора;
- импорт мероприятий из файлов различных форматов.

Планирование мероприятий в отношении поднадзорных НФО. В зависимости от режима надзора, категории, вида организации, риск-профиля, исходя из анализа этих параметров меняется дата назначения мероприятия и плановый период действия мероприятия, его вида и названия.

Просмотр истории изменения режимов надзора. Каждая смена режима надзора фиксируется, вследствие чего меняется список проводимых надзорных мероприятий. В результате подготавливается список с изменением статуса режима надзора по датам, регулирующим надзорным документом, видом объекта, категорией и названием режима надзора.

Настройка параметров работы куратора. Куратору нужно выдать специальные права, для экспорта и импорта организаций и мероприятий. Куратор должен быть ознакомлен со списком всех организаций и надзорных мероприятий, изменениями в режимах надзора, для исключения последующих ошибок, которые будут выявляться руководителем.

Импорт мероприятий из файлов различных форматов. Происходит загрузка отчетов, с их семантическими харак-

теристиками с информацией по поднадзорной организации, по организации и ее мероприятиям, отдельно по надзорным мероприятиям, по истории изменения режима надзора в отношении каждой из организаций. Импорт производится из файлов различных форматов.

РУКОВОДИТЕЛЬ

Сотрудникам с ролью «Руководитель» должны быть доступны следующие функции:

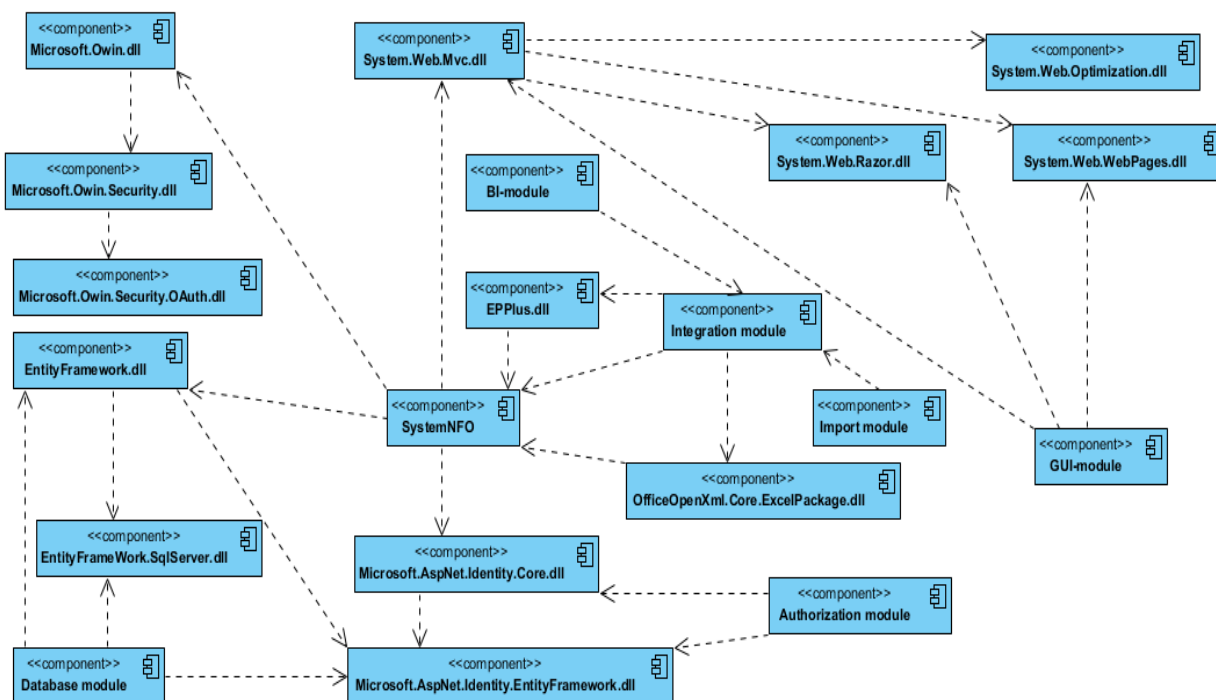
- поддержка принятия решений по планированию мероприятий в отношении НФО;
- просмотр перечня кураторов и назначенных им НФО;
- формирование, просмотр и печать сводных отчетов.

Поддержка принятия решений по планированию мероприятий в отношении НФО. Система по результатам предыдущих мероприятий, учитывая режим надзора, позволяет осуществлять поддержку принятия решений руководителю по назначению мероприятий и анализе их эффективности.

Просмотр перечня кураторов и назначенных им НФО. Доступен просмотр основных и резервных организаций для контроля исполнения своих обязанностей с возможностью выбора куратора из списка.

Формирование, просмотр и печать сводного отчета. Отчет формируется по всем ответственным сотрудникам. В отчете, среди прочего, отражаются просроченные мероприятия и не распланированные мероприятия. Кроме

Модуль аналитики предоставляет возможности по формированию как predetermined отчетных документов, так и настраиваемых с помощью мастера. Отчеты строятся с журналированием по временным периодам (месяц, квартал, полгода, год), по НФО, по поднадзорным регионам.



МОДУЛЬ РАБОТЫ С БАЗОЙ ДАННЫХ

Модуль работы с базой данных имеет в своем распоряжении Entity Framework. Сущность представляет набор данных, ассоциированных с определенным объектом. Происходит работа с объектами и их наборами. Взаимодействие с базой данных происходит по принципу Code First, создается класс модели данных, которые будут храниться в базе данных, а затем Entity Framework по этой модели генерирует базу данных и ее таблицы.

МОДУЛЬ ИМПОРТА

Модуль импорта предназначен для загрузки отчетов, их чтению и записи в базу данных системы. Модуль импорта позволяет в режиме Wizard настроить правила импорта. Импорт позволяет внедрять систему в организации, где деятельность ранее велась с использованием другого программного обеспечения.

МОДУЛЬ ПОЛЬЗОВАТЕЛЬСКОГО ИНТЕРФЕЙСА

Модуль пользовательского интерфейса строится в основном из компонентов Mvc, Razor и WebPages. В представлении используя возможности Razor и HTML для построения интерфейса пользователя.

ДРУГИЕ КОМПОНЕНТЫ

Owin используется для абстрагирования веб-сервера и компонентов фреймворка. ASP.NET Identity в системе используется для управления доступом пользователей в приложение. Entity Framework в системе используется для абстрагирования от самой базы данных и работы с данными независимо от типа хранилища. MVC используется для

разделения данных приложения, пользовательского интерфейса и управляющей логики. Razor используется для явного определения серверных блоков кода в HTML. Для построения страниц представления для пользователя используется компонент WebPages, а для оптимизации внутренних процессов компонент Optimization. Основной компонент SystemNFO в своем распоряжении имеет компоненты EPPlus и компонент OfficeOpenXml. Компонент EPPlus нужен для открытия Excel файла и сохранения отчетных данных в правильном формате из базы данных по организациям и надзорным мероприятиям. Компонент OfficeOpenXml нужен для чтения и записи файлов Excel.

РЕАЛИЗАЦИЯ

Разработана система поддержки принятия решений, позволяющая с помощью клиент-серверного приложения планировать мероприятия над организациями, а также формировать отчеты по организациям и мероприятиям. Система реализована на языке C# с использованием Entity Framework, шаблона проектирования MVC, и системы управления базами данных SQL Server. Система предназначена для регуляторов, которые проводят учет надзорных мероприятий в отношении некредитных финансовых организаций и выполняют контроль над выполнением мероприятий в срок.

В зависимости от должности сотрудника, имеется доступ к разному функционалу. На "рис. 4" представлен интерфейс системы в режиме руководителя. В нижней части экраны представлены мероприятия, проводимые в отношении организаций, которые разделены на: текущие, прошед-

Система надзора | Администратор | Куратор | Руководитель | Импорт отчета для кураторов | Импорт отчета для руководителей

Здравствуйте, curator@gmail.com! | Выйти

ОГРН: 343435722345 | Сокр. наименование: МФК Пример | Вид: Микрофинансовая комп. | Категория: Крупные и системноз. | Риск-профиль: Оранжевый | Дата изменения: 02.05.2019 | Режим: 2

Отчет

Мероприятия

Добавить мероприятие

Название мероприятия	Комментарий	Дата начала	Дата окончания	
Анализ информации из внутренних баз данных	Требуется	11.04.2019	10.07.2019	Редактировать Удалить
Анализ информации из открытых источников	Требуется	12.05.2019	11.08.2019	Редактировать Удалить
Мониторинг нефинансовых показателей	Требуется	18.05.2019	17.08.2019	Редактировать Удалить
Определение или проверка актуальности режима надзора	Требуется	23.04.2019	22.07.2019	Редактировать Удалить
Мониторинг раскрытия информации широкому кругу лиц	Требуется	03.04.2019	02.07.2019	Редактировать Удалить
Формирование индивидуального перечня мероприятий дистанционного надзора	Требуется	30.03.2019	29.07.2019	Редактировать Удалить
Иные мероприятия	Требуется	10.05.2019	09.08.2019	Редактировать Удалить

Рис. 4. Поддержка принятия решений при назначении мероприятий для НФО

шие и удаленные. Доступны функции по ручной корректровке состава мероприятий, т.е. мероприятие можно добавить, удалить или отредактировать. При этом ведется учет всех действий, проводимых над составом мероприятий, что позволит адаптировать.

На “рис. 5” представлен интерфейс системы при назначении кураторов для надзора за НФО. При назначении кураторов анализируется их предыдущий опыт по работе с поднадзорными организациями и достигнутые показатели КРП

Система позволяет повысить эффективность деятельности регулятора в части контроля и надзора как за отдельными некредитными финансовыми организациями, так и за рынком в целом. Система позволяет отслеживать и предотвращать нарушения, а также корректировать правильность действий в отношении организаций. Система позволяет планировать надзорные мероприятия, изменять режимы надзора и предупреждать об окончании срока выполнения мероприятия.

Кураторы			Выбор организации, где куратор является основным			Куратор является основным		
ФИО	Выбор организации		ОГРН	Сокращенное название		ОГРН	Сокращенное название	
Петров Петр Петрович	Выбор организации		343435722345	МФК Пример	Редактировать Выбор Удалить	343435722345	МФК Пример	Редактировать Удалить
Уфимов Андрей Евгеньевич	Выбор организации		677983673234	КПК КООП	Редактировать Выбор Удалить	102420172019	КПК ИСХОД	Редактировать Удалить
Иванов Иван Иванович	Выбор организации		6453536723234	СКПК СОДРУЖЕСТВО	Редактировать Выбор Удалить			
Андреев Александр Вадимович	Выбор организации		875435677654	ЖНЧ ЕДИНСТВО	Редактировать Выбор Удалить			
Иванов Иван Алексеевич	Выбор организации		109754113156	СРО ЖИЗНЬ	Редактировать Выбор Удалить			
Рудин Андрей Анатольевич	Выбор организации		120757568986	Ломбард ДОИНФО	Редактировать Выбор Удалить			
Ефимов Иван Иванович	Выбор организации		102420172019	КПК ИСХОД	Редактировать Выбор Удалить			

Выбор организации, где куратор является резервным			Куратор является резервным		
ОГРН	Сокращенное название		ОГРН	Сокращенное название	
343435722345	МФК Пример	Редактировать Выбор Удалить	343435722345	МФК Пример	Редактировать Удалить
677983673234	КПК КООП	Редактировать Выбор Удалить	102420172019	КПК ИСХОД	Редактировать Удалить

Рис. 5. Назначение НФО для кураторов

ЗАКЛЮЧЕНИЕ

Таким образом, в этой статье проведена декомпозиция предметной области с использованием принципов системного анализа, разработана модель предметной области, разработаны алгоритмы функционирования и описана архитектура системы. Разработана система информационной поддержки принятия решений, которая реализует следующие функции:

- поддержка принятия решений по планированию надзорных мероприятий;
- планирование надзорных мероприятий над организациями;
- контроль выполнения мероприятий;
- распределение задач между сотрудниками на основе их загрузки, предыдущего опыта и КРП;
- формирование отчетов по организациям и мероприятиям.

ЛИТЕРАТУРА

1. Lalon R.M. An Analysis of Financial Performance on Non-bank Financial Institutions (NBFI) in Bangladesh: A Study on Lanka-Bangla Finance Limited / Lalon R.M., Hussain S. // International Journal of Economics, Finance and Management Sciences. – 2017. – №5. – P. 251.
2. Shanmugam K.R. Non-banking Financial Intermediaries: International Experiences / K. R. Shanmugam, R. Kannan, S. Bhaduri // Non-Banking Financial Companies Role in India's Development. – 2019. – №1. – P. 103-124.
3. Mungai K. The Impact of Big Data on the South African Banking Industry / K. Mungai, A. Bayat // ICICKM 2018 15th International Conference on Intellectual Capital Knowledge Management & Organisational Learning, 2018. – 225 p.
4. Coury B. G. Supervisory control and the design of intelligent user interfaces / B. G. Coury, R. D. Semmel // Automation and human performance. – 2018. – P. 221-242.
5. Favarò F. M. Application of temporal logic for safety supervisory control and model-based hazard monitoring / F. M. Favarò, J. H. Saleh // Reliability Engineering & System Safety. – 2018. – №169. – P. 166-178.

6. Şensoy A. A. Developing a decision support framework for real-time flood management using integrated models / A. Şensoy, G. Uysal, A. A. Şorman // *Journal of Flood Risk Management*. – 2018. – №11. – P. 866-883.
7. Moradi-Aliabadi M. Decision support for enhancement of manufacturing sustainability: a hierarchical control approach / M. Moradi-Aliabadi, Y. Huang // *ACS Sustainable Chemistry & Engineering*. – 2018. – T. 6, №4. – P. 4809-4820.
8. Ivaschenko A. Spatial Clustering Based on Analysis of Big Data in Digital Marketing / A. Ivaschenko, A. Stolbova, O. Golovnin // *Communications in Computer and Information Science*, Springer, Cham, 2019. – T. 1093 – P. 335–347.
9. Scholl I. Organizational-and system-level characteristics that influence implementation of shared decision-making and strategies to address them—a scoping review / I. Scholl // *Implementation Science*. – 2018. – T. 13, №1. – P. 40.
10. Li X. A dynamic decision-making approach for intrusion response in industrial control systems / X. Li // *IEEE Transactions on Industrial Informatics*. – 2018. – T. 15, №5. – P. 2544-2554.
11. Duan Y. Artificial intelligence for decision making in the era of Big Data—evolution, challenges and research agenda / Y. Duan, J. S. Edwards, Y. K. Dwivedi // *International Journal of Information Management*. – 2019. – T. 48. – P. 63-71.

Decision-making Support for Supervisory Control of Non-banking Financial Institutions

Perevozchikov N., Golovnin O.

Department of Information Systems and Technologies

Samara University

Samara, Russia

nikiperevozchikov@yandex.ru

Abstract. With the change in the regulatory landscape in the financial sector, the organization of the activities of non-banking financial institutions (NBFI) is making more and more demands on the part of state regulatory authorities. To increase the effectiveness of supervision of NBFIs, regulatory bodies use systems that automate the supervisory activities, and, in particular, decision support systems. This paper presents a decision support system to plan supervisory activities over NBFIs in accordance with the principle of risk-based supervision, in which resources are concentrated on NBFIs with the maximum amount of risk or scope of activity. The developed system tracks and prevents violations, adjusts the correctness of actions in relation to NBFIs. In addition,

the system plans supervisory activities, changes the supervision regimes for NBFIs and controls the timing. The application of the developed system increases the effectiveness of the regulatory bodies in terms of control and supervision of both individual NBFIs and the financial market as a whole.

Keywords: supervision system, event planning, supervision regimes, non-credit organizations, risk-based supervision.

Применение нейронных сетей в прогнозировании спроса на новую продукцию приборостроительного предприятия

Пьянзина Е. А.

Челябинский государственный университет
Челябинск, Россия
pyanzina_evgeniia@mail.ru

Мельников В.А.

Челябинский государственный университет
Челябинск, Россия
protei300@gmail.com

Аннотация. В данной статье рассмотрен алгоритм прогнозирования спроса на новую продукцию предприятия приборостроительной отрасли методом искусственных нейронных сетей.

Ключевые слова: Искусственные нейронные сети, прогнозирование спроса, новая продукция, спрос на новую продукцию, подготовка данных.

ВВЕДЕНИЕ

В условиях развитого рынка, и высокой конкурентной среды основным преимуществом обладают те представители отрасли приборостроения, которые не почуют на лаврах, а развиваются, совершенствуя свою продукцию, и изобретая новые продукты, двигая рынок и отрасль вперед. Мало кто из потребителей может представить, что перед выпуском новинки на рынок предприятие проводит масштабную и трудоемкую работу, при которой задействовано большое количество сотрудников из разных сфер и отделов. Как правило, процесс разработки нового прибора требует колоссальных капитальных вложений, что, в условиях неопределенности, связанной с отсутствием опыта предложения потребителю подобного товара, подразумевает под собой риск потери всех вложенных средств в продукт, не вызвавший интереса у потребителей. Таким образом, предприятия, которые хотят остаться на рынке и получить конкурентное преимущество должны учитывать предполагаемый спрос на новую продукцию ещё до выпуска её на рынок, а именно в момент образования идеи создания новинки. Это значит, что прогнозирование спроса на новые продукты, становится необходимым уже на стадии разработки идеи, что обуславливает актуальность данного исследования.

АЛГОРИТМ СБОРА И ОБРАБОТКИ ДАННЫХ

Результатом прогнозирования спроса станет определение объемов продукции в количественном выражении. В общепринятом понимании, при прогнозировании спроса исследователь опирается на реальные данные, которые являются результатом спроса на продукцию в предыдущих периодах. Исходя из этого возникает вопрос, как прогнозировать спрос на новую продукцию, которой еще не было? В данном случае продукт (датчик), предлагающийся к разработке, будет усовершенствован и изменен таким образом, что он не будет полностью аналогичен своему прототипу, и на его спрос будут влиять, в том числе, другие факторы, отличные от тех, которые учитываются при прогнозировании спроса на прототип. При выборе метода оценки и предсказания спроса можно столкнуться с боль-

шим количеством методик, имеющих свои достоинства и недостатки. [4]

Опираясь на все рассмотренные варианты, было принято решение, для предсказания спроса на новые продукты приборостроительного предприятия, в исследовании использовать аппарат искусственных нейронных сетей (ИНС), которые, обрабатывая данные из прошлых периодов, имеют способность к обучению и являются одним из наиболее эффективных методов прогнозирования на данный момент. [4]

При прогнозировании спроса, используя ИНС, исследователь проходит ряд этапов, описанных ниже:

1. Подготовка исходных данных к анализу. На данном этапе были собраны и обработаны данные по продажам, заказам и ценам на продукцию измерительных приборов и сопутствующих товаров по семи направлениям (Давление, Метрология, Термометрия, Расходомерия, Измерение уровня, Беспроводные технологии учета, Прочее) от двух выпускаемых марок по 12 финансовым периодам за предыдущие 5 лет.

После того, как были определены критерии, которые влияют на спрос, нужно собрать и подготовить данные. Основная задача, которая ставится перед исследователем на этом этапе – создание векторного пространства признаков через приведение данных в вид, который будет пригоден для обучения модели. А именно, необходимо привести данные в числовую форму, которую сможет распознать нейронная сеть. Для этого можно воспользоваться основными методами:

- Нормализовать данные путем достижения значений дисперсии равной единице и приведения среднего значения по всем имеющимся данным к нулю.
- Изменить размерность векторного пространства путем повышения или понижения, в зависимости от количества имеющихся признаков.

Для увеличения размерности базы данных в нее, помимо исторических данных о заказах, продажах и ценах были внесены основные технические характеристики продукции, которые влияют на выбор потребителем какого-либо измерительного прибора или сопутствующего товара.

2. Для обучения и проверки построенной нейронной сети необходимо создать обучающую, контрольную и тестовую выборки, которые автоматически задаются в ана-

Е.А. Пьянзина, В.А. Мельников

литической платформе, использованной для обучения нейронной сети.

3. Формируем тип и архитектуру нейронной сети. На этом этапе выбираем количество слоев и количество нейронов в каждом слое. Выбор наилучшего типа нейросети зависит от вида поставленной перед нами задачи. Из большого разнообразия видов нейросетей, для решения нашей задачи, была выбрана Рекуррентная нейронная сеть, особенностью которой является способность нейронов получать информацию не только от предыдущего слоя, но и от самих себя из предыдущего прохода, что можно увидеть на схеме рекуррентной нейронной сети на Рисунке 1.

Если же выбрать сеть самостоятельно не представля-

На практике алгоритм, разобранный выше, был применен при помощи аналитической платформы Deductor.

Исходим из того факта, что база данных состоит из продуктов, которые относятся к разным маркам и направлениям. Было принято решение разбить так же базу данных на 2 части, в зависимости от марки, под которой выпускалась продукция и на 7 частей, в зависимости от направления измерений, к которому относятся приборы учета и сопутствующие продукты. Таким образом, всего было образовано 10 групп данных, по которым были построены и обучены 66 нейронных сетей, из которых 8 ИНС были обучены на всей базе данных.

АНАЛИЗ ДАННЫХ

Изучив, полученные в ходе обучения нейронных сетей,

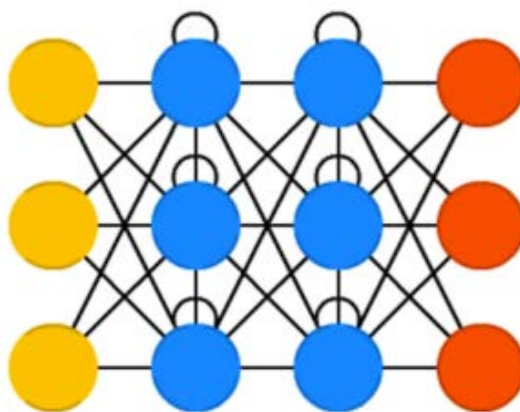


Рис. 1. Архитектура рекуррентной нейронной сети

ется возможным, то следует выбрать несколько типов, и после получения результатов отобрать лучшую из всех.

4. Выбираем алгоритм обучения нейронной сети. Существуют разные алгоритмы обучения нейронных сетей. Опираясь на многочисленные исследования, можно сделать вывод, что для данной задачи следует выбрать алгоритм обратного распространения ошибки.

Процесс обучения в данном случае проходит от конца к началу, то есть, сеть выдвигает предположение, а исследователь должен оценить, насколько сеть ошибается и скорректировать веса входных значений. Таким образом, происходит продвижение назад с попутным изменением весов входных значений, пока не будет достигнут ввод.

5. Применение нейронной сети. Запускаем нашу сеть, чтобы обучить её.

6. Рассчитываем ошибку. На данном этапе необходимо подогнать веса нейронов, для соответствия обучающей и контрольной выборки.

7. Выбираем лучший вариант построенной нейронной сети. Из всех обученных нейросетей мы должны выбрать сеть, которая будет иметь наименьшую ошибку. Проводя кросс-проверку, мы должны убедиться, что прогноз в будущем будет приближен к реальным данным. Поэтому данный этап очень важен при проведении исследования.[1].

прогнозы и сопоставив их с реальными данными были выведены среднеквадратическая и средняя абсолютная ошибки прогноза (1), (2). Ошибки прогноза были рассчитаны по формулам:

$$MSE = \frac{1}{n} \times \sum_{i=1}^n (Y_i - Y'_i)^2 \quad (1)$$

где, MSE – это среднеквадратическая ошибки прогноза, n – это количество наблюдений, Y_i – это фактическое значение переменной, Y'_i – это прогнозное значение переменной;

$$MAPE = \frac{1}{n} \times \sum_{i=1}^n \frac{|Y_i - Y'_i|}{Y_i} \quad (2)$$

где, MAPE – это средняя абсолютная ошибки прогноза, n – это количество наблюдений, Y_i – это фактическое значение переменной, Y'_i – это прогнозное значение переменной;

Исходя из полученных прогнозных значений количества продаж приборов по каждой группе, были рассчитаны значения ошибок прогнозирования нейронных сетей, исходя из которых были отобраны лучшие нейронные сети, которые обладают наименьшим значением ошибок.

Лучшие нейронные сети и значения их ошибок (MAPE) по каждой группе можно увидеть в Таблице 1, представленной ниже.

Таблица 1

ИНС с минимальной ошибкой прогноза по каждой группе.

Группа	Сеть	Ошибка MAPE
Все датчики	86x2x1	662%
Датчики Марка 1	78x20x1	21%
Датчики Марка 2	72x72x1	89%
Датчики Давление	66x10x1	442%
Датчики Метрология	24x20x1	215%
Датчики Прочие беспроводные продукты	5x2x1	374%
Датчики Расход	54x2x1	120%
Датчики Температура	51x2x1	248%
Датчики Уровень	26x15x1	361%
Датчики ФиВА	32x5x1	194%

Проанализировав показатели лучших ИНС, которые представлены в Таблице 1, можно увидеть, что все они имеют большую ошибку прогноза. Особенно большой ошибкой обладают сети, построенные на всей базе данных. Сети этой группы совершенно непригодны для прогнозирования. Наиболее приемлемым вариантом построения прогноза можно считать разбиение базы на группы, которые будут содержать приборы с большим числом схожих характеристик. К такому разбиению, относятся группы, выделенные ранее и представленные в Таблице 1. Можно отметить, что ИНС, построенные на данных из представленных групп, имеют процент ошибки значительно меньше, чем в случае с ИНС, построенной по всей базе данных. Исходя из показателей MSE и MAPE, были выделены 2 группы данных, делящие основную выборку в зависимости от марки, под которой производятся приборы. Однако, следует заметить, что даже эта ошибка является большой и говорит о непригодности моделей. На диаграмме рассеяния, показанной на Рисунке 2 можно увидеть, что прогнозные значения находятся достаточно далеко от линии целевой функции, что подтверждает сделанный ранее вывод о непригодности моделей к прогнозированию.

ЗАКЛЮЧЕНИЕ

В ходе проведения исследования, путем прохождения алгоритма построения и обучения нейронной сети для прогнозирования спроса на новую продукцию приборостроительного завода было выявлено, что метод ИНС непригоден для поставленной цели, так как ошибка прогноза является большой.

Таким образом, для выполнения поставленной задачи следует обратить внимание на другие методы, которые могут быть применимы в прогнозировании спроса на новую продукцию.

ЛИТЕРАТУРА

1. Круглов, В.В. Искусственные нейронные сети. Теория и практика /В.В. Круглов, В.В. Борисов. – М. : Горячая линия-Телеком, 2002. – 382 с.
2. Фролова, Т.А. Прогнозирование спроса на химическую продукцию с применением аппарата временных рядов / Т.А. Фролова, Д.С. Туляков // Вопр. соврем. науки и практики. Ун-т им. В.И. Вернадского. – 2009. – № 5(19). – С. 92–97.
3. Хайкин, С. Нейронные сети: полный курс : пер. с

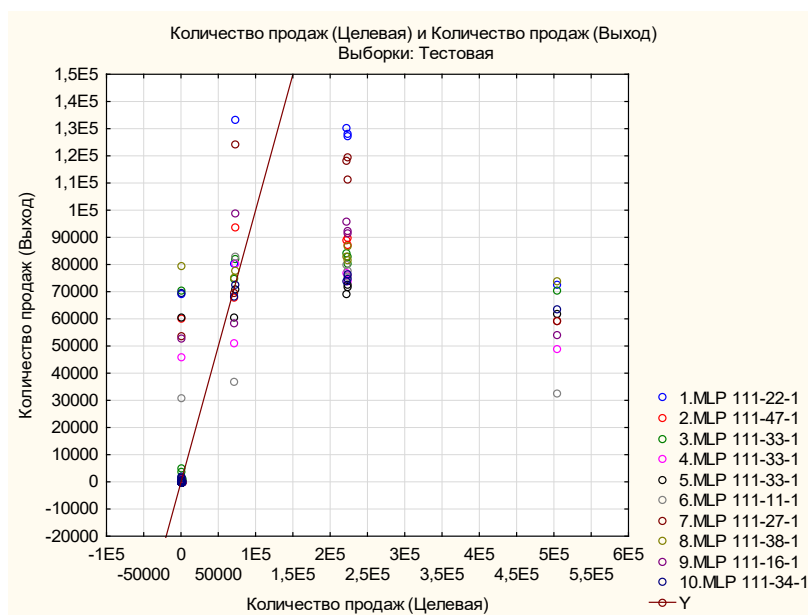


Рис. 2. Диаграмма рассеяния

англ. / С. Хайкин. – 2-е изд. – М. : Вильямс, 2006. – 1104 с.

4. Афанасьев, В.Н. Анализ временных рядов и прогнозирование: Учебник / В.Н. Афанасьев. - М.: Финансы и статистика, 2012. - 320 с.

5. Muller A. Introduction to Machine Learning with Python / Sarah Guido , Andreas Müller. – Gravenstein Highway North, Sebastopol: O'Reilly Media, 2017/ 392 с.

Application of neural networks in forecasting demand for new products of an instrument-making enterprise.

E.A. Pianzina

Chelyabinsk State University
Chelyabinsk, Russia
pyanzina_evgeniia@mail.ru

V.A. Melnikov

Chelyabinsk State University
Chelyabinsk, Russia
protei300@gmail.com

Abstract. In this article it is considered and offered the algorithm of forecasting the demand for a new-created products of the instrument-making industry using method of artificial neural networks.

Keywords: Artificial neural networks, demand forecasting, new products, demand for new products, data preparation.

Прогнозирование температуры помещений общественного здания с помощью искусственной нейронной сети

Ромазанов А.Р.

Захаров А.А.

Захарова И.Г.

ФГАОУ ВО «Тюменский государственный университет»

Тюмень, Россия

a.r.romazanov@utmn.ru

Аннотация. Предложен подход к прогнозированию температуры в помещениях общественного здания. Модель здания описывается средними температурами в его помещениях, характеристиками внешних стен и элементов отопления. Внешние условия определяются температурой, скоростью и направлением ветра. Состояние теплового узла описывается температурой теплоносителя на входе и выходе, а также расходом. Для построения прогнозной модели выявляется нелинейная зависимость температуры внутри помещения от указанных параметров. Данная задача решается с помощью рекуррентной искусственной нейронной сети.

Ключевые слова: прогнозирование, искусственная нейронная сеть, моделирование теплового режима.

ВВЕДЕНИЕ

Прогнозирование температуры помещений в общественном здании в зависимости от режима отопления и внешних условий (температуры воздуха, скорости и направления ветра) представляет интерес как с точки зрения экономии затрат тепловой энергии, так и в плане обеспечения комфортной рабочей среды. Традиционный подход к решению этой задачи опирается на исследование локальных (на уровне помещения) и глобальных (для всего здания) теплопотерь. Получение модели процессов теплообмена в здании позволяет более точно регулировать системы отопления для соблюдения санитарных норм и снижения потребления теплоносителя [1, 2].

Теплоэнергетический баланс зданий определяется, первую очередь, интенсивностью теплопереноса за счет механизма теплопроводности. При этом в зимний период изменение внешних условий существенно влияет на значения параметров теплопередачи наружных стен зданий [3, 4]. Дополнительные сложности возникают в случае использования экономных режимов отопления – прерывистого отопления. В этом случае при периодических продолжительных снижениях температуры помещений могут изменяться также параметры теплопередачи для внутренних стен и перекрытий. Наконец, не для всяких зданий можно установить структуру стен и перекрытий с точки зрения числа слоев и материалов, что приводит к использованию некоторых условных значений параметров.

Таким образом, можно предположить, что классический подход, который основан на полноценном описании процесса теплопереноса и соответствующих балансовых соотношениях, позволяет выполнить общую оценку со-

стояния здания с точки зрения тепловых потерь. Но для прогнозирования температуры помещений при нестационарных режимах отопления и изменяющихся внешних условиях необходимо вводить множество поправок для параметров теплопередачи, тепловой инерции помещений и всего здания в целом или использовать иной подход.

В условиях, когда имеется возможность обеспечить мониторинг температуры воздуха в здании, условий во внешней среде, а также параметров отопительной системы, задача прогнозирования температуры помещений может быть поставлена как классическая задача многопараметрической нелинейной регрессии. Решение может быть получено с помощью методов глубокого обучения – искусственной нейронной сети.

ОБЗОР ЛИТЕРАТУРЫ

Методы машинного и глубокого обучения достаточно широко используются для исследования тепловых характеристик зданий и управления Heating Ventilation and Air Conditioning (HVAC) системами. Так, Ferreira P. M. et al [5] используют нейронные сети в сочетании с генетическими алгоритмами для управления HVAC системой с целью достижение желаемого уровня комфортной теплоты и энергосбережения в здании университетского кампуса. При этом, как отмечают в своих обзорных статьях Enescu [6], Moon J. W. et al. [7], Afram A. et al. [8], внимание исследователей сосредоточено на вопросах экономической эффективности теплоснабжения и на поддержке комфортных условий преимущественно для жилых помещений, эксплуатация которых существенно отличается от общественных зданий.

В работе Liang and Du [9] нейронная сеть используется для управления HVAC системой с целью поддержания комфортной температуры. В работах Moon J. W. et al. [10], Castilla M. et al. [11] рассматриваются подходы к управлению тепловым режимом с использованием искусственных нейронных сетей в отдельном помещении, однако не рассматривается возможность масштабирования решения для всего здания. Исследование Sablani S. S. et al [12] направлено на получение значений параметров теплопередачи конструкций здания с помощью ИНС.

Соответственно, применение искусственных нейронных сетей, базируясь на полносвязной архитектуре сети, которая не будет иметь практического применения в случае рассмотрения большого здания. При этом в работе

А.Р. Ромазанов, А.А. Захаров, И.Г. Захарова

Moretti M. et al. [13] показано, что для построения эффективных прогнозных моделей существенную роль играют и архитектура сети, и длина временного ряда измеряемых параметров, которые будут использоваться при обучении модели. Именно этот результат является принципиально важным для настоящего исследования.

КЛАССИЧЕСКИЙ МЕТОД РАСЧЕТА ПРОГНОЗНОЙ ТЕМПЕРАТУРЫ В ПОМЕЩЕНИИ

Тепловые характеристики здания

Состояние здания с точки зрения его тепловых характеристик определяется соответствующими параметрами его перекрытий. Для описания модели здания выделяются следующие значимые характеристики с учетом множества слоев из различных материалов для стен [14, 15], а также других существенных участников процесса теплообмена:

- коэффициент теплопроводности;
- коэффициент теплоотдачи;
- теплоемкость материала;
- площадь поверхности;
- толщина слоев;
- площадь отопительных элементов;
- площадь оконных проемов.

Для установления характера теплообмена между помещениями и внешней средой выделим единый коэффициент K , отражающий описанные ранее характеристики.

Для оценки тепловых характеристик перекрытий используется за основу математическая модель теплового режима здания, описанная в работе [14].

Помимо температуры воздуха снаружи здания также необходимо учитывать и ветер. Влияние ветра на микроклимат применимо как к конкретному помещению, так и ко всему зданию [16, 17] и должно быть учтено в модели.

Допущения

Получаемое и поглощаемое тепло для каждого помещения не конкретизируется по его источнику, а принимается как суммарное.

Элементы отопительной системы для каждого помещения не разделяются на отдельные элементы конструкции и не определяются конкретными параметрами.

Не учитывается теплообмен между помещениями в рамках здания.

Модель теплового режима здания

В качестве исходных данных выделяются следующие параметры помещения:

- температура окружающей среды и направление ветра;
- средняя температура в помещении;
- объем получаемого помещением тепла.

Для контроля баланса энергии используется суммарный объем теплоты, который получило здание за счет работы системы отопления. На основании разности температур на входе и выходе теплоносителя в тепловом узле выделяется количество теплоты для отдельных помещений:

$$Q_c = \sum Q_i, i = 1, n \quad (1)$$

где
 Q_c – суммарное количество теплоты, полученное зданием;
 Q_i – количество теплоты, для i -го помещения;

n – количество помещений в здании.

Показатели собираются с датчиков температур, размещенных внутри помещений и снаружи здания с заданных интервалом.

Метод расчета

Таким образом, задача установления границ изменения тепловых характеристик сводится к задаче линейной регрессии:

$$Q_c = \sum K_{ij}(T_i - T_j), i = 1, n, j = 1, m_i \quad (2)$$

где

Q_c – суммарное количество теплоты, полученное зданием;

K_{ij} – тепловая характеристика перекрытия, соединяющего пространства i и j ;

T_i – средняя температура в i -ом помещении;

T_j – наружная температура, соответствующая внешней стене с учетом стороны света;

m_i – количество внешних стен i -го помещения;

n – количество помещений в здании.

ПРОГНОЗИРОВАНИЕ ТЕМПЕРАТУРЫ С ИСПОЛЬЗОВАНИЕМ РЕКУРРЕНТНОЙ ИНС

Очевидно, что классический подход позволяет сделать лишь приблизительную оценку температуры в помещении без учета изменений основных динамических параметров. При этом именно динамика погодных условий, режима отопления существенно влияют на характер изменения температуры в помещении.

Предлагаемый подход предполагает учет динамики изменения состояния стен, отопительных элементов, теплоносителя и воздействие этих изменений на температуру в помещении.

Мы исходим из того, что выполняется мониторинг температуры воздуха в здании и отопительных элементов, температуры воздуха во внешней среде, скорости и направления ветра, температуры и расхода теплоносителя. В этом случае прогнозирование температуры в помещениях здания можно обеспечить на основе решения задачи многопараметрической нелинейной регрессии. Исходя из особенностей исходных данных, мы полагаем, что соответствующий прогноз может быть построен с помощью искусственной нейронной сети (ИНС).

Архитектура ИНС

В данной работе использована архитектура ИНС Gated Recurrent Unit (GRU) [18, 19], которая является модификацией Long-Short-Term Memory (LSTM) сети. Особенности, отличающими GRU от LSTM, является объединение модулей, отвечающих за забывание и обработку входа в единый вентиль обновления. Кроме того, в GRU отсутствует скрытое состояние. Исходя из этого, вычисления в GRU происходят быстрее, поскольку задействуют меньше параметров и имеют меньше переменных в работе.

Архитектура сети позволяет естественным образом учитывать изменения в состоянии объекта во времени. Типовая архитектура GRU-блока представлена на рис. 1 [20], где:

- x_t – вектор входных значений;
- \tilde{h}_t – вектор-кандидат состояния модуля t ;
- h_t – текущее состояние модуля t ;
- σ, \tanh – функции активации.

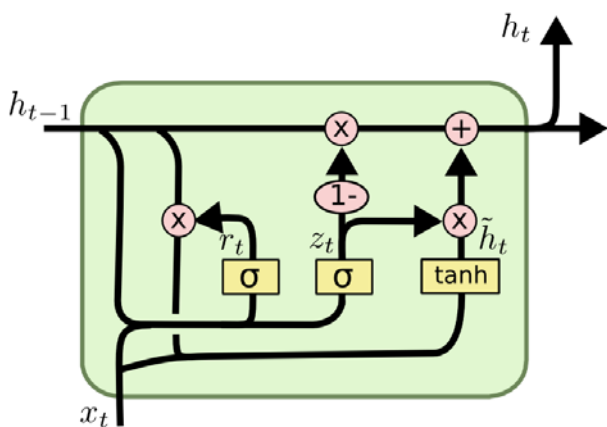


Рис. 1. Архитектура GRU-блока

Вектора z_t , r_t , \tilde{h}_t , h_t вычисляются следующим образом:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r)$$

$$\tilde{h}_t = \tanh(W_h \cdot [r_t * h_{t-1}, x_t] + b_h)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

где W , U , b – параметры модуля.

Вентиль обновления (вектор r_t) отвечает за получение информации из прошлого состояния. Вентиль сброса (вектор z_t) управляет заменой состояния модуля на новое или его модификацией.

Адаптация архитектуры

Особенности работы данной ИНС позволяют использовать для получения результата, как историю наблюдений за состоянием объекта, так и зависимости прогнозной температуры от дополнительных факторов, которые не закладываются в модель в явном виде:

- эффективная температура относительно направления и скорости ветра для каждой стены;
- изменение характеристик теплопередачи в зависимости от температуры стены.

Использование GRU модификации LSTM сети обусловлено тем, что внутреннее состояние допустимо считать целевым значением работы сети.

Входными данными являются кортежи, описывающие состояние помещения, отопительной системы и внешней среды в определенный момент времени.

В составе сети используются GRU-блоки с входными и выходными вентилями. Количество блоков зависит от требуемого количества шагов по времени из расчета один блок на один шаг.

Состоянием сети является текущее прогнозное значение, полученное на основании уже обработанных данных. Данное значение обновляется при каждой последующей обработке входных данных по очередному шагу по времени.

Вентиль обновления в данной задаче опускает ситуации, при которых произошли незначительные изменения в состоянии модели помещения, в то время как вентиль сброса реагирует на резкие изменения, например, погодных условий или значительного снижения температуры в помещении ввиду открытия окна или неполадок в отопительной системе.

Прогнозное значение температуры в соответствующем помещении определяется последним состоянием ИНС и является выходом сети.

Параметры входных данных ИНС

Для получения прогнозной температуры в момент времени t , при обработке k шагов по времени, учитываются параметры:

- T_{inside} – температура воздуха в помещении;
- T_{bat} – температура батареи;
- $T_{outside}$ – температура воздуха на улице;
- $Wind_s$ – скорость ветра;
- $Wind_d$ – направление ветра;
- T_{in} – температура теплоносителя на входе;
- T_{out} – температура теплоносителя на выходе;
- Q – расход теплоносителя;
- $Wall_d$ – направление нормали к стене;
- S_{bat} – площадь отопительных элементов;
- V_r – объем комнаты;
- τ – шаг измерений по времени;
- k – длина ряда.

Параметры $Wall_d$, S_{bat} , V_r являются постоянными и не меняются с очередным входом x_t . Остальные параметры являются динамическими и отражают изменение состояния помещения во времени.

Обучение ИНС

Процесс обучения нейронной сети происходит по методу обратного распространения ошибки. При обучении ИНС на вход поочередно подается состояние на момент очередного шага по времени. Значение температуры в помещении на последнем шаге по времени является целевым. Оно используется для оценки точности работы ИНС в процессе обучения.

Для установления начального состояния ИНС при обучении модели для очередного помещения можно использовать значения, которые были получены при работе с помещением того же класса согласно классификации помещений по отклонению от температурных норм [15, 21, 22].

Вычислительный эксперимент

Исходные данные

Для исследования результативности предлагаемого подхода основной акцент был сделан на изучении влияния внешних условий на динамику изменения температуры помещений. Важной особенностью является то, что при генерации температурного режима учитывается положение помещения. Для генерации температур обучающей выборки использовались показания датчиков в отдельном помещении с поправками для соответствующего класса. Для увеличения выборки были сгенерированы наборы псевдослучайных значений температуры воздуха в помещениях, основанные на показаниях датчиков. В эти наборы были включены случайные выбросы, отражающие события, ведущие к кратковременному снижению температуры (открытые окна и/или двери). Данные о скорости и направлении ветра для каждого шага по времени были получены из архива метеоусловий. Периодом генерации данных является интервал в 1 час. На рис. 2 приведен график изменений температуры в помещении и окружающей среды.

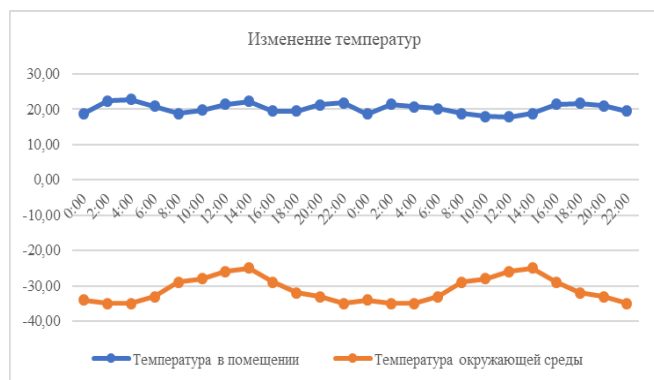


Рис. 2. График изменений температур

Для проведения вычислительного эксперимента были сгенерированы данные, соответствующие ежедневным замерам в декабре 2019 года с указанным периодом, всего 744 записи. Температура теплоносителя за указанный период изменялась незначительно и составляла соответственно на входе от 78 °С до 89 °С и на выходе от 39 °С до 53 °С для указанного ранее периода.

Параметры модулей GRU-блоков

Для построения решения выбрана структура, в которой количество внутренних блоков равно количеству шагов по времени, используемых для прогноза. Внутренние параметры в блоках инициализируются псевдослучайными числами с равномерным распределением на отрезке $[0, 1]$.

Обучение ИНС

С целью определения наиболее результативной модели прогнозной, при формировании обучающей и тестирующей выборки выделялись последовательности длиной в 6, 12, 24, 48, 72, 96, 120 шагов. Исходные данные были нормированы на отрезке $[0, 1]$.

Для обучения прогнозной модели были зафиксированы следующие параметры обучения:

- количество эпох – 10000;
- коэффициент скорости обучения – 0,001;
- количество параметров – 14;
- скрытых слоев сети – k;
- размерность выхода – 1.

Прогнозирование

В табл. 1 приведены среднее значения среднеквадратического отклонения прогнозных данных от фактических замеров температуры для каждой из моделей.

Таблица 1

Среднеквадратическое отклонение прогнозных значений от реальных замеров температуры

Длина временного ряда	Среднее значение температуры, °С	Среднеквадратическое отклонение
6	21,2	3,48
12	21,4	3,04
24	21,3	4,96
48	20,9	5,09
72	21,5	5,43
96	20,1	8,08
120	21,3	2,76

Вычислительный эксперимент показал, что наиболее точные результаты были получены для модели, основанной на данных с длиной временного ряда равной 120. Однако, краткосрочный прогноз на основании лишь 12 часов дает сопоставимые по точности результаты, что может найти свое применение в определенных сценариях. Существенный разброс прогнозных значений для $k=96$ связан с кратковременным резким изменением внешних условий. Важно отметить, что модель показала устойчивость относительно случайных выбросов температуры воздуха в помещениях.

ЗАКЛЮЧЕНИЕ

В результате исследования показана результативность метода прогнозирования температуры воздуха в помещении общественного здания с помощью ИНС. Решение этой задачи позволяет формировать экономичные режимы прерывистого отопления при различных погодных условиях. Предложенный подход может быть развит с учетом влияния присутствия людей в помещении на температуру и влажность воздуха.

ЛИТЕРАТУРА

1. Круковский П. Г. и др. Некоторые пути снижения энергозатрат зданий путем регулирования теплопотребления //Thermophysics and Thermal Power Engineering. – 2016. – Т. 38. – №. 1. – С. 62-67.
2. Тарасов В. А. и др. Математическое моделирование режимов прогнозного и дежурного отопления //Известия высших учебных заведений. ПРОБЛЕМЫ ЭНЕРГЕТИКИ. – 2019. – Т. 21. – №. 3. – С. 73-85.
3. Коршунов О. В., Зуев В. И. Измерение термического сопротивления наружных стен зданий //Энергобезопасность и энергосбережение. – 2011. – Т. 38. №. 2. С. 40-45.
4. Коршунов О. В., Зуев В. И. Время тепловой инерции и термическое сопротивление слоистых стен //Энергобезопасность и энергосбережение. – 2011. – №. 4. С. 23-26.
5. Ferreira P. M. et al. Neural networks based predictive control for thermal comfort and energy savings in public buildings //Energy and buildings. – 2012. – Т. 55. – С. 238-251.
6. Enescu D. A review of thermal comfort models and indicators for indoor environments //Renewable and Sustainable Energy Reviews. – 2017. – Т. 79. – С. 1353-1379.
7. Moon J. W. et al. Comparative study of artificial intelligence-based building thermal control methods–Application of fuzzy, adaptive neuro-fuzzy inference system, and artificial neural network //Applied Thermal Engineering. – 2011. – Т. 31. – №. 14-15. – С. 2422-2429.
8. Afram A. et al. Artificial neural network (ANN) based model predictive control (MPC) and optimization of HVAC systems: A state of the art review and case study of a residential HVAC system //Energy and Buildings. – 2017. – Т. 141. – С. 96-113.
9. Liang J., Du R. Thermal comfort control based on neural network for HVAC application //Proceedings of 2005 IEEE Conference on Control Applications, 2005. CCA 2005. – IEEE, 2005. – С. 819-824.
10. Moon J. W., Jung S. K., Kim J. J. Application of ANN (artificial-neural-network) in residential thermal control

- //Proceedings of the Eleventh International IBPSA Conference, Glasgow, Scotland. – 2009. – С. 27-30.
11. Castilla M. et al. Neural network and polynomial approximated thermal comfort models for HVAC systems //Building and Environment. – 2013. – Т. 59. – С. 107-115.
 12. Sablani S. S. et al. Non-iterative estimation of heat transfer coefficients using artificial neural network models //International Journal of Heat and Mass Transfer. – 2005. – Т. 48. – №. 3-4. – С. 665-679.
 13. Macas M. et al. The role of data sample size and dimensionality in neural network-based forecasting of building heating related variables //Energy and Buildings. – 2016. – Т. 111. – С. 299-310.
 14. Круковский П. Г. и др. Определение параметров теплотермических помещений при нестационарном тепловом режиме //Промышленная теплотехника. – 2013. – Т. 35. – №. 6. – С. 47-56.
 15. Захаров А. А. Моделирование теплового режима и управление теплоснабжением помещений умного здания / А. А. Захаров, И. Г. Захарова, А. Р. Ромазанов, А. В. Широких // Вестник Тюменского государственного университета. Физико-математическое моделирование. Нефть, газ, энергетика. 2018. Том 4. № 2. С. 105-119.
 16. Дрозд Д. В., Елистратова Ю. В., Семенов А. С. Влияние ветра на микроклимат в помещении //Современные наукоемкие технологии. – 2013. – №. 8-1. – С. 37-39.
 17. Табунщиков Ю. А., Бродач М. М. Математическое моделирование и оптимизация тепловой эффективности зданий. – М.: АВОК-Пресс, 2002.
 18. Cho K. et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation //arXiv preprint arXiv:1406.1078. – 2014.
 19. Chung J. et al. Empirical evaluation of gated recurrent neural networks on sequence modeling //arXiv preprint arXiv:1412.3555. – 2014.
 20. Understanding LSTM Networks – URL: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/index.html> (дата обращения 23.01.2020)
 21. Zakharov A. et al. Intellectual Data Analysis System of Building Temperature Mode Monitoring //2019 International Russian Automation Conference (RusAutoCon). – IEEE, 2019. – С. 1-6.
 22. Ромазанов А. Р., Захарова И. Г. Мониторинг и моделирование теплового режима здания //Тезисы XIX Всероссийской конференции молодых учёных по математическому моделированию и информационным технологиям. – 2018. – С. 76-76.

Temperature prediction in a public building using artificial neural network

Artur Romazanov
Irina Zakharova
Software Department
University of Tyumen
Tyumen, Russia
a.r.romazanov@utmn.ru

Alexander Zakharov
Secure Smart City's Information
Technologies Department
University of Tyumen
Tyumen, Russia

Abstract. The paper proposes an approach to predict the temperature in the rooms of a public building. The model of the building is described by the average temperatures in its rooms, the characteristics of external walls and heating elements. Weather conditions are determined by the temperature, speed and direction of the wind. The state of the thermal unit is described by the temperature of heat agent at the inlet and outlet of a heat supply system, as well as the flow rate. To build a predic-

tive model, it is necessary to identify a nonlinear dependence of the temperature inside the room on these parameters. This problem is solved using a recurrent artificial neural network.

Keywords: prediction, artificial neural network, temperature mode modeling.

Реализация искусственной нейронной сети на основе микроконтроллера с использованием целочисленной арифметики

Скрипов С.А.

Челябинский государственный университет
Челябинск, Россия
skripov@csu.ru

Аннотация. В статье рассматриваются возможности машинного обучения с использованием сравнительно малопроизводительных аппаратных средств, таких как микроконтроллеры, платформы для построения простых систем автоматики, одноплатные компьютеры. Экспериментально изучается вопрос развертывания ранее обученной искусственной нейронной сети на базе малопроизводительного микроконтроллера. Обучение предлагается производить с помощью высокопроизводительной ЭВМ с использованием стандартных инструментов и библиотек. Такой подход позволит сделать предложенный метод универсальным, а также упростит использование уже существующих разработок. В статье производится измерение производительности и энергопотребления для микроконтроллера с классификатором на основе ранее обученной нейронной сети.

Ключевые слова: искусственная нейронная сеть, микроконтроллер, сенсорная сеть, интернет вещей.

ВВЕДЕНИЕ

В настоящее время машинное обучение может на практике использоваться для решения широкого круга задач, таких как распознавание и классификация изображений, распознавание речи, прогнозирование в финансовой сфере, прогнозирование поведения клиентов, прогнозирование природных явлений и т.д.

Применение машинного обучения возможно в системах реального времени, например, для реализации роботов [1] и беспилотных транспортных средств [2]. Подобные решения могут применяться для получения информации из окружающей среды с помощью множества датчиков [3]. Эти решения могут быть распределенными, таким образом для обработки данных используется множество узлов, которые могут связываться, например, по беспроводным каналам [4].

В статье рассмотрены нейронные сети, реализованные на устройствах с низкой производительностью, таких как одноплатные компьютеры [5], датчики и микроконтроллеры [6] или платформы для автоматизации [7]. Особый интерес представляют распределенные решения и сенсорные сети, где обработка данных производится датчиком, который эти данные извлекает из окружающей среды. Такой подход имеет несколько преимуществ, например, уменьшится количество данных, передаваемых на базовую станцию, а также периодичность передачи. Это позволит значительно продлить работу автономного устройства от батареи [3].

РЕАЛИЗАЦИЯ НЕЙРОННЫХ СЕТЕЙ НА ОСНОВЕ ОДНОПЛАТНЫХ КОМПЬЮТЕРОВ

Существует большое количество различных разработок на основе одноплатных компьютеров. Многие модели позволяют использовать операционную систему Linux, что дает возможность использовать готовое программное обеспечение и библиотеки. По своим вычислительным возможностям одноплатный компьютер мало чем уступает офисному персональному компьютеру, причем быстродействие новых моделей постоянно увеличивается [8]. Такие устройства можно использовать для решения даже достаточно сложных задач, например, анализа изображений. Для обработки и анализа изображений можно использовать, например, OpenCV, а для анализа данных с помощью нейронных сетей – TensorFlow. При этом вполне успешно решаются такие задачи как классификация, локализация, обнаружение объектов, выделение областей [9]. Анализом изображений применение одноплатных компьютеров не ограничивается, например, в [10] описана система для выращивания томатов с помощью гидропонике, где нейронная сеть использует температуру, влажность, освещенность, кислотность, уровни жидкостей и т.д. для управления процессом выращивания. Существенным недостатком одноплатных компьютеров является их относительно высокое энергопотребление [8]. Это сильно усложняет или делает полностью невозможным вариант с их использованием как автономных устройств с питанием от батареи.

РЕАЛИЗАЦИЯ НЕЙРОННЫХ СЕТЕЙ НА ОСНОВЕ МИКРОКОНТРОЛЛЕРОВ

Другой тип устройств, на основе которых возможно развертывание нейронных сетей – это микроконтроллеры. Их главные преимущества – относительно невысокая цена, малые размеры, малое энергопотребление. Например, для PIC контроллера цена может начинаться от \$0.36. В режиме нормальной работы такой микроконтроллер может потреблять <2мА, в режиме с уменьшенной частотой – 15мкА, а в режиме энергосбережения (сна) – <1мкА [6]. Это делает возможным использовать как отдельные устройства, так и распределенные системы или сенсорные сети, которые собирают и обрабатывают данные.

Недостатком микроконтроллеров являются их ограниченные вычислительные ресурсы. Обычно такие устройства имеют несколько десятков килобайт флеш памяти, 1-2 килобайта оперативной памяти, несколько сотен байт EEPROM памяти. Тем не менее, в настоящее время есть

С.А. Скрипов

примеры практического использования микроконтроллеров для решения задач с использованием машинного обучения. Так в [11] описан комплекс для определения прозрачности воды в водоемах на основе нейронной сети. Для этого используется датчик, определяющий ослабление солнечного света для пяти спектральных составляющих. Нейронная сеть реализована на микроконтроллере PIC18F4550. Преимуществом предложенного метода является независимость результатов измерения от субъективных факторов, а также наличия в воде водорослей, планктона, и т.д.

Подобные подходы могут применяться в медицине. Так, в [12] описан метод лечения для пациентов с тяжелой эпилепсией, которые не могут лечиться традиционными средствами. Основа метода – имплантируемый автономный датчик с микроконтроллером MSP430FR5994. Для предсказания эпилептических припадков используется сверточная нейронная сеть, использующая данные электроэнцефалографии.

Также необходимо отдельно выделить микроконтроллер ESP8266 с встроенным модулем Wi-Fi [16]. Для этого микроконтроллера существует программное обеспечение NodeMCU, представляющее из себя платформу для реализации Интернета вещей. Для NodeMCU имеется множество полезных библиотек для коммуникации с другими устройствами по беспроводной сети или через проводные соединения [17]. Новые приложения можно разрабатывать на языке Lua. Также существует реализация MicroPython для ESP8266, позволяющая разрабатывать программное обеспечение на языке Python3 с применением подмножества стандартных библиотек [18]. Возможности ESP8266 позволяют использовать машинное обучение для решения широкого класса задач

Для быстрого создания прототипов нейронных сетей на основе микроконтроллера можно использовать платформы автоматизации, такие как Arduino [19]. Эти платформы также можно использовать и для практической реализации. Так в [20] описана система мониторинга на основе нейронной сети, в [21] – нейронная сеть для управления роботом

Реализация нейронной сети на основе микроконтроллера осложняется малым количеством памяти, невысокой производительностью процессора, в недорогих устройствах отсутствуют команды для обработки чисел с плавающей точкой. В [13-15] рассматриваются вопросы, связанные с реализацией нейронных сетей на недорогих малопроизводительных микроконтроллерах, предлагаются решения, позволяющие развертывать нейронные сети для различных задач без привязки к конкретным моделям микроконтроллеров.

ПОРТИРОВАНИЕ НЕЙРОННОЙ СЕТИ НА МИКРОКОНТРОЛЛЕР

В рамках данной работы был проведен эксперимент, показывающий возможности и ограничения переноса обученной нейронной сети на микроконтроллер. В качестве микроконтроллера был выбран ESP8266, который позволяет быстро создавать прототипы и легко перемещать данные из компьютера в микроконтроллер и обратно. Использовалось программное обеспечение NodeMCU. Исходная нейронная сеть была обучена на обычном персональном компьютере с использованием языка Python3,

платформы TensorFlow и библиотеки Keras. Для эксперимента был использован датасет Fashion MNIST, содержащий изображения размером 28x28. Назначение нейронной сети состоит в классификации предметов одежды в соответствии с 10 категориями. Таким образом, входной слой состоит из 28x28 нейронов, выходной – из 10 нейронов. Нейронная сеть имеет также один скрытый слой, количество нейронов в котором изменялось в процессе эксперимента для изучения возможной производительности. Использовались функции активации “relu” и “softmax”.

Для портирования нейронной сети необходимо перенести на микроконтроллер изображения датасета, веса синапсов и смещения для выходного и скрытого слоев. В рамках эксперимента разработаны две реализации нейронной сети: с использованием арифметики с плавающей точкой, а также с использованием целочисленной арифметики. Если микроконтроллер не имеет аппаратной поддержки для работы с вещественными числами, использование типов с плавающей точкой существенно увеличивает расход памяти и уменьшает производительность, так как в этом случае основные операции с числами будут реализованы программно.

Были рассмотрены три типа данных для экспорта значений обученной нейросети. Первый вариант предполагает перенос двоичных данных в формате вещественных чисел одинарной точности. Вторым вариантом повторяется способ, описанный в [13], где предлагается использовать собственный тип с фиксированной точкой. Однако реализация операций сложения и умножения для этого типа на языке lua для NodeMCU показала крайне низкую производительность. Можно предположить, что этот способ будет хорошо работать при реализации на языке C, где доступны различные базовые типы и встроенные побитовые операции с числами. Было принято решение в рамках эксперимента этот тип данных не использовать. Третий вариант предполагал использование целого 32-битного типа, причем для передачи части знаков после запятой значение умножается на 10000. Такой способ представления вещественного числа должен учитываться программным обеспечением микроконтроллера.

Обученная нейронная сеть экспортируется в виде 4 файлов, содержащих веса и смещения для двух уровней. Структура самой нейронной сети была реализована в коде для микроконтроллера с возможностью выбора количества нейронов скрытого слоя, 16, 32 или 64. Таким образом в рамках эксперимента перенос структуры нейронной сети не требуется.

РЕАЛИЗАЦИЯ И ОПТИМИЗАЦИЯ НЕЙРОННОЙ СЕТИ НА МИКРОКОНТРОЛЛЕРЕ

В языке программирования lua поддерживается единственный тип для хранения числовых значений. Поэтому для эксперимента были подготовлены две сборки NodeMCU, использующие в качестве базовых типов соответственно вещественный тип с двойной точностью и целочисленный 32-битный тип.

Файлы, содержащие значения для обученной нейронной сети были загружены на флеш-память ESP8266. Для извлечения данных из файлов в сборки были добавлены модули file и struct, позволяющие читать данные из файлов и записывать значения во встроенные типы.

Программный код классификатора для каждого базового типа данных представлен двумя файлами: nn.lua и activation.lua для вещественного типа, nn_int.lua и activation_int.lua для целочисленного типа. Здесь файлы вида nn.lua содержат реализацию классификатора, а activation.lua представляет собой дополнительный модуль, где реализованы функции активации.

В процессе тестирования реализации нейронной сети на микроконтроллере было замечено, что узким местом является флеш-память. Параметры нейронной сети не помещаются целиком в оперативную память микроконтроллера, поэтому данные читаются из файла во время работы классификатора. Первые реализации считывали значения последовательно по 4 байта. Однако экспериментальным путем было установлено, что оптимальный размер блока данных для считывания составляет 256 байт. После оптимизации программного кода производительность классификатора увеличилась в 5 раз.

Реализация классификатора для целых типов оптимизирована для максимальной производительности, однако итоговые значения имеют погрешность. Сложение, умножение и деление реализованы с помощью целочисленных команд процессора, для вычисления экспоненты использована таблица степеней числа e с шагом 1. Итоговые значения выходного слоя, полученные с помощью функции активации “softmax”, имеют погрешность начиная со второго знака после запятой, однако для задачи классификации точности вполне достаточно. Пример вычисления значений выходного слоя с помощью целочисленной арифметики приведен в табл. 1.

Реализация классификатора для вещественных типов позволяет получить значения, максимально приближенные к истинным.

ИЗМЕРЕНИЕ ПРОИЗВОДИТЕЛЬНОСТИ И ЭНЕРГОПОТРЕБЛЕНИЯ

ESP 8266 имеет относительно производительный процессор, который может работать на частотах 80 МГц и 160 МГц. Ток потребления для частоты 80 МГц составил 20.4 мА, для 160 МГц – 24.9 мА. В процессе эксперимента на устройство подавалось рекомендованное напряжение 3.3 В. Процессор поддерживает только целочисленные операции, поэтому ток потребления для целочисленной и вещественной реализаций не отличается.

Также было измерено время работы классификатора в различных режимах и с различным количеством нейронов скрытого слоя. Результаты измерения приведены в табл. 2

Таблица 1

Пример полученных значений для выходного слоя

Истинные значения для выходного слоя	Значения, полученные классификатором с использованием целочисленных вычислений
2,5663330e-10	0
9,0482961e-16	0
1,3790408e-10	0
3,0633094e-09	0
1,4281262e-10	0
1,6861045e-04	0,0003
6,0989542e-09	0
7,6196929e-03	0,0066

Истинные значения для выходного слоя	Значения, полученные классификатором с использованием целочисленных вычислений
1,4039239e-07	0
9,9221158e-01	0,9929

Таблица 2

Результаты измерения времени работы классификатора

Кол-во нейронов скрытого слоя	Точность для вещественной реализации	Размер данных для экспорта, байт	Время работы классификатора, секунд			
			вещественная реализация, 80 МГц	вещественная реализация, 160 МГц	целочисленная реализация, 80 МГц	целочисленная реализация, 160 МГц
64	0,8848	203560	9,04	5,48	7,22	3,76
32	0,8785	101800	4,67	2,74	3,7	1,92
16	0,8619	50920	2,41	1,45	1,95	1,01

ЗАКЛЮЧЕНИЕ

Проведенное исследование показывает, что современные микроконтроллеры обладают вполне приемлемой производительностью для реализации простых нейронных сетей. В рамках эксперимента нейронная сеть была реализована на языке lua с использованием дополнительных модулей NodeMCU, что также повлияло на производительность. Таким образом существует возможность для дальнейшей оптимизации производительности и энергопотребления. Существенной проблемой является малый объем памяти микроконтроллера, а также особенности доступа к ней.

Также в рамках работы был предложен метод, предполагающий обучение нейронной сети на высокопроизводительной ЭВМ с использованием стандартных инструментов и библиотек и дальнейший перенос ее на автономное или портативное устройство.

ЛИТЕРАТУРА

1. Farooq Umar Design and Implementation of Neural Network Based Controller for Mobile Robot Navigation in Unknown Environments / Umar Farooq, Muhammad Amar, Muhammad Usman Asad, Athar Hanif, and Syed Omar Saleh. // International Journal of Computer and Electrical Engineering. – 2014. – vol. 6, No. 2. – P. 83-89.
2. Farooq Umar Microcontroller Based Neural Network Controlled Low Cost Autonomous Vehicle / Umar Farooq, Muhammad Amar, Eitzaz ul Haq, Muhammad Usman Asad, Hafiz Muhammad Atiq // Second International Conference on Machine Learning and Computing. – 2010. – P. 96-100.
3. Gutierrez-GalanJuan D. Embedded neural network for real-time animal behavior classification / D. Gutierrez-GalanJuan, P. Dominguez-Morales, E. Cerezuela-Escudero, A. Rios-Navarro, R. Tapiador-Morales, M. Rivas-Perez, M. Dominguez-Morales, A. Jimenez-Fernandez, A. Linares-Barranco // Neurocomputing. – 2017. – is. 272(2018). – P. 17–26.
4. Susnea Ioan Distributed Neural Networks Microcontroller Implementation and Applications // Studies in Informatics and Control. – 2012. – vol. 21, No. 2. – P. 165-172.

С.А. Скрипов

5. What's Orange Pi ? – URL: <http://www.orangepi.org/Orange%20Pi%203/> (дата обращения 05.02.2020)
6. New/Popular 8-bit PIC MCU Products – URL: <https://www.microchip.com/ParamChartSearch/Chart.aspx?branchID=30048> (дата обращения 05.02.2020)
7. Arduino Products – URL: <https://www.arduino.cc/en/Main/Products> (дата обращения 05.02.2020)
8. Deep Neural Network Benchmark with Raspberry Pi 2, 3 and 3+ – URL: <https://www.element14.com/community/thread/63591/1/deep-neural-network-benchmark-with-raspberry-pi-2-3-and-3> (дата обращения 05.02.2020)
9. How to easily Detect Objects with Deep Learning on Raspberry Pi – URL: <https://medium.com/nanonets/how-to-easily-detect-objects-with-deep-learning-on-raspberrypi-225f29635c74> (дата обращения 05.02.2020)
10. Mehra Manav IoT based hydroponics system using Deep Neural Networks / Manav Mehra, Sameer Saxena, Suresh Sankaranarayanan, Rijo Jackson Tom, M. Veeramankandan // *Computers and Electronics in Agriculture*. – 2018. – is. 155(2018). – P. 473–486.
11. Civera Javier Ibáñez Artificial neural network onto eight bit microcontroller for Secchi depth calculation / Javier Ibáñez Civera, Eduardo Garcia Breijo, Nicolás Laguarda Miró, Luis Gil Sánchez, José Garrigues Baixauli, Inmaculada Romero Gil, Rafael Masot Peris, Miguel Alcaniz Fillol // *Sensors and Actuators B*. – 2011. – is. 156(2011). – P. 132–139
12. Hügler Maria Early Seizure Detection with an Energy-Efficient Convolutional Neural Network on an Implantable Microcontroller / Maria Hügler, Simon Heller, Manuel Watter, Manuel Blum, Farrokh Manzouri, Matthias Dümpelmann, Andreas Schulze-Bonhage, Peter Woias, Joschka Boedecker // *International Joint Conference on Neural Networks*. – 2018
13. Tripathy Jnana Ranjan Artificial Neural Network Implementation in Microchip PIC 18F45J10 8-Bit Microcontroller / Jnana Ranjan Tripathy, Hrudaya Kumar Tripathy, S.S.Nayak // *International Journal of Engineering and Advanced Technology*. – 2014. – vol. 3, is. 5. – P. 131-135.
14. Cotton Nicholas J. A Neural Network Implementation on an Inexpensive Eight Bit Microcontroller / Nicholas J. Cotton, Bogdan M. Wilamowski, Günhan Dündar // *12th International Conference on Intelligent Engineering Systems*. – 2008. – P. 109-114.
15. Cotton Nicholas Jay A Neural Network Implementation on Embedded Systems. – Auburn, Alabama. – 2010. – P. 102
16. WiFi модуль ESP8266 ESP07 для домашней автоматизации – URL: <https://mysku.ru/blog/ebay/30626.html> (дата обращения 05.02.2020)
17. NodeMCU custom builds. Date Views – URL: <https://nodemcu-build.com/> (дата обращения 05.02.2020)
18. Getting started with MicroPython on the ESP8266 – URL: <https://docs.micropython.org/en/latest/esp8266/tutorial/intro.html> (дата обращения 05.02.2020)
19. A Neural Network for Arduino. Date Views – URL: <http://robotics.hobbizine.com/arduinoann.html> (дата обращения 05.02.2020)
20. Abubakar I. Development of Arduino Microcontroller Based NonIntrusive Appliances Monitoring System using Artificial Neural Network / I. Abubakar, S. N. Khalid1, M. W. Mustafa, M. Mustapha and Hussain Shareef // *Journal of Computational and Theoretical Nanoscience*. – 2018. – is. 24(6). – P. 4483-4488.
21. Робот управляемый нейронной сетью. Дата обращения – URL: <http://robocraft.ru/blog/3298.html> (дата обращения 05.02.2020)

Implementation of an artificial neural network based on a microcontroller using integer arithmetic

Skipov S.A.

Institute of Information Technology
Chelyabinsk state University
Chelyabinsk, the Russian Federation
skipov@csu.ru

Abstract. The article discusses the possibilities of machine learning using low-performance hardware, such as microcontrollers, platforms for building simple automation systems, single-board computers. The issue of deploying a previously trained artificial neural network based on a low-performance microcontroller is experimentally studied. Training is proposed to be carried out using high-performance computers using standard tools and libraries. This approach will make the proposed method universal, as well as simplify the use of existing developments.

For a microcontroller with a classifier based on a previously trained neural network, performance and energy consumption measurements were performed.

Keywords: artificial neural network, microcontroller, sensor network, Internet of things.

Модель оценки и прогнозирования рисков лесных рубок на территории нефтедобывающего региона

Якимчук А.В.

Югорский государственный университет, г. Ханты-Мансийск, Россия

AV_Yakimchuk@mail.ru

Аннотация. В статье рассматривается применение нейронных сетей для реализации модели оценки и прогнозирования лесных рубок на примере Ханты-Мансийского автономного округа-Югры. В качестве сведений используется накопленный массив данных о ранее случившихся инцидентах на территории округа в период с 2010–2019 гг. В основе реализованной модели лежат методы машинного обучения, которые позволяют получить вещественное число, являющееся степенью риска возникновения повторного инцидента на территории определенного участка. Полученные результаты выводятся на тепловой карте, построенной с использованием геоинформационной системы.

Ключевые слова: машинное обучение, прогнозирование рисков, нейронные сети, данные по лесным рубкам, риск-ориентированный подход, ГИС, надстройка Keras.

ВВЕДЕНИЕ

Одной из серьёзных экологических проблем многих стран, в том числе России, являются незаконные рубки лесных насаждений [1]. Больше всего страдают районы с большой площадью лесов и малозаселёнными территориями, к которым и относится Ханты-Мансийский автономный округ-Югра. Общая площадь земель округа, на которых расположены леса, составляет 50 402,7 га (4% от лесного фонда Российской Федерации). В течение 2018–2019 годов было зарегистрировано более 500 нарушений, которые нанесли не только экологический, но и существенный экономический ущерб [2].

Большая и труднодоступная территория, а также слабо развитая коммуникационная инфраструктура округа существенно затрудняют своевременный контроль за экологическим состоянием окружающей среды. В связи с этим необходимо внедрение принципиально новых подходов к осуществлению контрольно-надзорной деятельности, которые позволят прогнозировать и оперативно контролировать лесные участки, наиболее подверженные незаконным лесным рубкам.

В последние годы ведется активная работа над внедрением дифференцированного или риск-ориентированного подхода [3] при организации и осуществлении контрольно-надзорных мероприятий, что позволит существенно повысить эффективность расходования ресурсов на функционирование контрольно-надзорных органов путем сосредоточения усилий инспекторского состава государственного надзора на наиболее значимых направлениях.

Однако на уровне субъектов Российской Федерации до настоящего времени не сложилось устойчивых практик применения и тем более автоматизации методов риск-ориентированного подхода в деятельности контрольно-надзорных органов [4, 5].

В настоящее время активно ведётся работа по обобщению накопленного опыта управления рисками и разработке различных подходов к решению задач мониторинга, оценки и прогнозирования.

В [6] предложен геосистемный подход к классификации и геоинформационному картированию экологического риска. Проанализировано пространственно-временное состояние геосистем в Байкальском регионе. Проведено зонирование области исследования в соответствии со степенью склонности к экологическому риску. Разработана интерпретирующая информационно-картографическая система природной среды Байкальского региона для поддержки управления экологическими рисками.

В [7] анализируется роль моделирования в оценке риска, рассматриваются понятия экологического риска, риска здоровью человека, раскрываются проблемы и методология оценки экологического риска.

В [8] рассмотрена методика расчета экологических рисков, которая позволяет учесть возможные последствия наступления риска. Проанализирована информационная основа оценки экологических рисков и необходимость классификации факторов экологической опасности. Приведена структура полного экологического ущерба. Рассмотрены методы оценки ущерба, как для окружающей природной среды, так и для здоровья человека. Показана связь экологического ущерба с социальной политикой. Показана зависимость выбора метода количественной оценки ущерба для здоровья человека от продолжительности и уровня интенсивности воздействия. Рассмотрены экономические оценки экологических воздействий. Приведены основные элементы процесса управления рисками.

В [9] представлены результаты системного анализа нечетких продукционных систем поддержки принятия решений в контексте геоинформационного мониторинга экологических рисков. Рассмотрен способ представления правил нечетких продукций в нечеткой системе анализа и сокращения экологических рисков, основанный на применении нечетких сетей Петри. Предложен способ представления правил нечетких продукций при решении при-

кладных задач нечеткого моделирования по экологическим рискам.

В [10] описан подход к использованию нейросетевых технологий для решения аналитических задач в ГИС. Дано изложение некоторых из таких задач и показаны области их применения. Представлен макет программной системы для анализа географических данных на основе обучаемых нейронных сетей.

Приведённый выше анализ исследований позволяет увидеть, что используются различные способы определения вероятности возникновения рисков, каждый из которых направлен на решение определенной задачи. В связи с этим для решения задачи оценки рисков незаконных лесных рубок предлагается модель, реализуемая алгоритмическими методами, использование которых позволяет построить тепловую карту, которая позволяет отобразить риски незаконных лесных рубок на участках лесничеств округа. Для реализации модели используются методы машинного обучения и методы пространственного анализа данных. Методы машинного обучения основаны на создании, обучении и последующем использовании нейронной сети, которая позволяет обрабатывать различные объемы разнородной информации об объектах исследования. Методы геоинформационного анализа данных используются для учета пространственных свойств объектов исследования.

ДАННЫЕ И МЕТОДИКА

Построение модели прогнозирования с итоговым выводом на тепловую карту подразумевает решение следующих подзадач:

1. Сбор и проведение препроцессинга данных;
2. Пространственный анализ данных;
3. Построение модели машинного обучения и выбор основных признаков;
4. Представление результатов на тепловой карте.

В качестве данных для реализации модели предлагается использование данных, накопленных Службой по контролю и надзору в сфере охраны окружающей среды, объектов животного мира и лесных отношений Ханты-Мансийского автономного округа – Югры (Природнадзор Югры) с 2010 года по настоящее время, а также данные, полученные в результате дешифрования космических снимков, на основе которых можно получить информацию о ранее случившихся инцидентах.

В качестве признаков выступает информация о местоположении лесных рубок, близость к техногенным объектам, в том числе к населенным пунктам и к транспортной инфраструктуре, а также информация о близости к водным объектам, о породах деревьев и т.п. Всего используется семьдесят пять признаков.

Собранный массив данных требует проведения предварительной обработки информации, в котором необходимо преобразование некорректных координат и восстановление пространственных характеристик, а также исключение дубликатов. Также необходимо приведение признаков к числовому виду, таким образом, все признаки категорий (тип и вид леса и др.) преобразованы в вектора с использованием техники OneHotEncoding [11].

На следующем этапе все записи сопоставляются в соответствии с их пространственным расположением на

основе координат или информации о выделах. Далеко не все записи о лесных рубках обладают координатной привязкой и для того, чтобы определить местоположение, используется информация о лесном делении (лесничество, участковое лесничество, кварталы и выделы).

Таким образом, решение задачи сводится к реализации такой модели для прогнозирования, которая позволит определить участки, наиболее подверженные риску лесной вырубki. В соответствии с [12] принято выделять следующие пять категорий риска для регионального государственного экологического надзора: высокий, значительный, средний, умеренный и низкий, для каждого из которых устанавливается своя очередность плановых проверок.

В качестве модели для прогнозирования выбраны нейронные сети с многослойной архитектурой, которые являются одной из наиболее применимых в настоящее время [13-16].

Реализованная нейросетевая модель имеет 75 нейронов на входном слое (по количеству признаков для каждого участка участкового лесничества (категорийные признаки векторизованы)), три скрытых слоя, в каждом из которых в качестве функции активации используется сигмоид, и выходной слой, возвращающий вещественное число в интервале от 0 до 1. Это число, характеризующее степень негативного воздействия на лесной фонд в границах участкового лесничества, будем трактовать как оценку экологического риска изменения состояния лесного фонда по аналогии с понятием отнесения производственных объектов, оказывающих негативное воздействие на окружающую среду, к определенной категории риска. В качестве метки для прогнозирования использовались аварии за 2019 год. Архитектура реализованной нейронной сети представлена на рисунке 1.

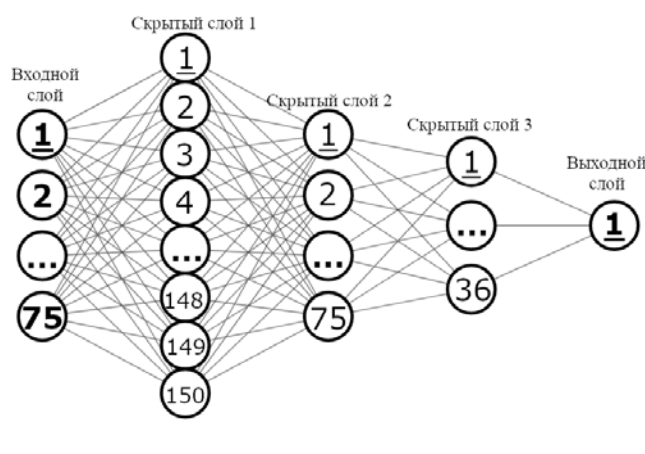


Рисунок 1- Архитектура нейронной сети по прогнозированию лесных вырубок

В качестве ответа получаем число в диапазоне от 0 до 1, которое позволяет категоризовать степень негативного воздействия на лесной фонд на пять категорий риска в зависимости от значения:

$$\left\{ \begin{array}{l} y = 0 - \text{низкий риск,} \\ 0 < y \leq 0,16 - \text{умеренный риск,} \\ 0,16 < y \leq 0,39 - \text{средний риск,} \\ 0,39 < y \leq 0,71 - \text{значительный риск,} \\ 0,71 < y \leq 1 - \text{высокий риск.} \end{array} \right.$$

Для валидации реализованной модели используется метод перекрестной проверки (K-fold cross-validation) [17]. Набор данных для обучения был разбит на десять одинаковых по размеру блоков. Всего было проведено десять численных экспериментов, в каждом из которых один блок из десяти использовался для тестирования модели, а оставшиеся девять блоков использовались в качестве тренировочного набора. По результатам численных экспериментов с различными комбинациями блоков набора данных были рассчитаны промежуточные оценки качества модели. Результаты оценки представлены в таблице 1.

Таблица 1
Результаты оценки качества нейросетевой модели по прогнозированию лесных вырубок

Номер блока набора данных	Рассчитанная оценка, %
1	83,7
2	84,1
3	84,2
4	83,5
5	85,6
6	84,6
7	81,8
8	84,7
9	83,9
10	82,9

Средняя оценка качества реализованной многослойной нейронной сети составила 83,9%, что позволяет говорить о высоком качестве и возможности её практического применения для решения данного класса задач.

На основе реализованной модели построен прогноз рисков незаконных рубок на территории Ханты-Мансийского автономного округа-Югра на 2020 год, данные прогноза представлены в виде тепловой карты на рисунке 2, где видно, что наиболее подверженной территорией является Кондинский район.

ЗАКЛЮЧЕНИЕ

В статье представлен подход к реализации контрольно-надзорной деятельности, касающейся рисков незаконных рубок лесных насаждений на территории Ханты-Мансийского автономного округа-Югры. Данный подход основан на использовании ретроспективных данных, которые используются для обучения нейронной сети, а также на основе пространственного анализа. Достаточно высокая оценка качества позволяет говорить о практической применимости модели.

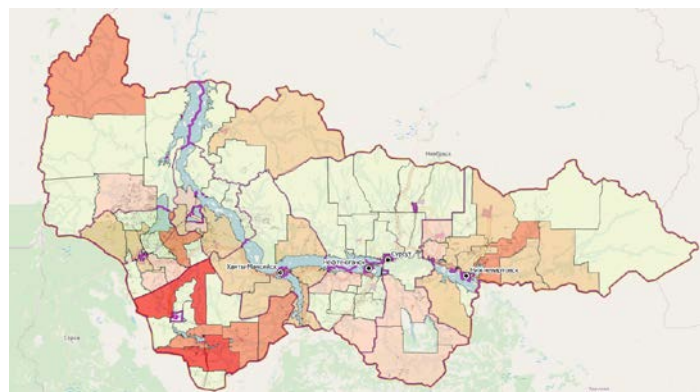


Рис. 2. Тепловая карта рисков возникновения лесных рубок на территории Ханты-Мансийского автономного округа-Югра

БЛАГОДАРНОСТИ

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 18-45-860003 «Модель оценки рисков состояния лесного фонда на территории Ханты-Мансийского автономного округа – Югры».

ЛИТЕРАТУРА

1. Черных Ю. В. Актуальные проблемы выявления и пресечения незаконной рубки лесных насаждений на примере Иркутской области // Молодой ученый. — 2018. — №25. — С. 239-242. — URL <https://moluch.ru/archive/211/51554/>
2. Доклад об экологической ситуации в Ханты-Мансийском автономном округе-Югра. URL <https://prirodnadzor.admhmao.ru/doklady-i-otchyety/doklad-ob-ekologicheskoy-situatsii-v-khanty-mansiyskom-avtonomnom-okruge-yugre/>
3. Распоряжение Правительства Российской Федерации от 1 апреля 2016 г. № 559-р «План мероприятий по совершенствованию контрольно-надзорной деятельности в российской федерации на 2016 - 2017 годы».
4. Комов В.Э., Кабалинский А.И. Обоснование повышения эффективности деятельности контрольно-надзорной деятельности на региональном уровне // Известия Тульского государственного университета. Экономические и юридические науки. — 2017. — №. 1-1. — С. 159–164.
5. Кузнецова Т.И., Плюснин В.М. Геоинформационное ландшафтное картографирование для обеспечения управления экологическим риском Байкальского региона // Материалы Международной конференции «ИнтерКарто/ИнтерГИС». — 2016. — Т. 1. — №. 20. — С. 318-327.
6. Чаплинский А.В., Плаксин С.М. Управление рисками при осуществлении государственного контроля в России // Вопросы государственного и муниципального управления. 2016. № 2. С. 7-29
7. Дмитриев В. Г. Оценка экологического риска. Аналитический обзор публикаций // Арктика и север. — 2014. — №. 14. — С. 126–148.
8. Бельская Е.Н., Бразговка О.В., Сугак Е.В. Методика расчета экологических рисков // Современные проблемы науки и образования. — 2014. — № 6.; URL: <http://www.science-education.ru/ru/article/view?id=15755>
9. Акинина Н.В. и др. Теория и практика применения нечетких сетей Петри для мониторинга экологических

рисков // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. – 2017. – №. 41. – С. 4–11. DOI: 10.17223/19988605/41/1

10. Питенко А.А. Использование нейросетевых технологий при решении аналитических задач в ГИС // Методы нейроинформатики: сборник научных трудов / Под ред. А.Н. Горбаня. – 1998. – С. 152–163.

11. Kaggle. Using Categorical Data with One Hot Encoding. Дата обращения: 02.02.2019.<https://www.kaggle.com/dansbecker/using-categorical-data-with-one-hot-encoding>

12. Постановление Правительства Российской Федерации от 22 ноября 2017 г. № 1410 "О критериях отнесения производственных объектов, используемых юридическими лицами и индивидуальными предпринимателями, оказывающих негативное воздействие на окружающую среду, к определенной категории риска для регионального государственного экологического надзора и об особенностях осуществления указанного надзора".

13. Rojas R. Neural networks: a systematic introduction. – Springer Science & Business Media, 2013.

14. Haykin S. Neural networks. – New York : Prentice hall, 1994. – Т. 2.

15. Riedmiller M., Braun H. A Direct Adaptive Method for Faster Backpropagation Learning: the RPROP Algorithm // International Conference of Neural Networks. San Francisco, USA, 1993. Vol. 1, pp. 586-591.

16. Zhang G., Patuwo B. E., Hu M. Y. Forecasting with artificial neural networks:: The state of the art //International journal of forecasting. – 1998. – Т. 14. – №. 1. – С. 35-62.

17. Friedman J., Hastie T., Tibshirani R. The elements of statistical learning. – New York : Springer series in statistics, 2001. – Т. 1. – №. 10.

