



УНИВЕРСИТЕТ ИТМО

СБОРНИК ТРУДОВ

VIII

КОНГРЕССА  
МОЛОДЫХ  
УЧЕНЫХ



ТОМ 4



КОНГРЕСС МОЛОДЫХ УЧЕНЫХ

Санкт-Петербург

2019

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ**

**УНИВЕРСИТЕТ ИТМО**

**Сборник трудов  
VIII Конгресса молодых ученых**

**Том 4**



**УНИВЕРСИТЕТ ИТМО**

**Санкт-Петербург**

**2019**

Сборник трудов VIII Конгресса молодых ученых. Том 4. – СПб.: Университет ИТМО, 2019. – 116 с.

Издание содержит результаты научных работ молодых ученых, доложенные на VIII Конгрессе молодых ученых Университета ИТМО по тематикам: информационные технологий и программирование; инфокоммуникационные технологии; дизайн и урбанистика.

#### РЕДАКЦИОННАЯ КОЛЛЕГИЯ

##### **Председатель редколлегии:**

*Бухановский Александр Валерьевич*

доктор технических наук, директор мегафакультета трансляционных информационных технологий Университета ИТМО.

##### **Члены редколлегии:**

*Матвеев Юрий Николаевич*

доктор технических наук, профессор факультета информационных технологий и программирования

*Парфенов Владимир Глебович*

доктор технических наук, декан факультета информационных технологий и программирования

*Хоружников Сергей Эдуардович*

кандидат физико-математических наук, декан факультета инфокоммуникационных технологий

*Карсаков Андрей Сергеевич*

кандидат технических наук, доцент института дизайна и урбанистики

*Митягин Сергей Александрович*

кандидат технических наук, директор института дизайна и урбанистики

ISBN 978-5-7577-0603-0

ISBN 978-5-7577-0615-3 (Том 4)



**УНИВЕРСИТЕТ ИТМО**

Университет ИТМО – ведущий вуз России в области информационных и фотонных технологий, один из немногих российских вузов, получивших в 2009 году статус национального исследовательского университета. С 2013 года Университет ИТМО – участник программы повышения конкурентоспособности российских университетов среди ведущих мировых научно-образовательных центров, известной как проект «5 в 100». Цель Университета ИТМО – становление исследовательского университета мирового уровня, предпринимательского по типу, ориентированного на интернационализацию всех направлений деятельности.

© Университет ИТМО, 2019

© Авторы, 2019

## **ВВЕДЕНИЕ**

Издание содержит результаты научных работ молодых ученых, доложенные 15–19 апреля 2019 года на VIII Конгрессе молодых ученых Университета ИТМО по тематикам: информационные технологии и программирование; инфокоммуникационные технологии; дизайн и урбанистика.

Конгресс проводится в целях реализации программы Национального исследовательского университета, программы повышения конкурентоспособности Университета ИТМО среди ведущих мировых научно-образовательных центров на 2013–2020 гг., а также с целью стимулирования научно-технической деятельности молодых ученых, приобретения ими опыта публичных выступлений, повышения научного уровня и апробации выпускных квалификационных работ магистрантов и бакалавров за 2018/2019 учебный год.



**НАПРАВЛЕНИЕ  
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ  
И ПРОГРАММИРОВАНИЯ**

**Аксенов Юрий Владиславович**

Год рождения: 1996

Университет ИТМО, факультет информационных технологий и программирования, студент группы № М3405

Направление подготовки: 09.03.02 – Информационные системы и технологии

e-mail: Yura290610@yandex.ru

**Возисов Никита Сергеевич**

Год рождения: 1997

Университет ИТМО, факультет информационных технологий и программирования, студент группы № М3405

Направление подготовки: 09.03.02 – Информационные системы и технологии

e-mail: nikitavozisov2108@gmail.com

**Шорохов Станислав Андреевич**

Год рождения: 1997

Университет ИТМО, факультет информационных технологий и программирования, студент группы № М3405

Направление подготовки: 09.03.02 – Информационные системы и технологии

e-mail: stashor16@gmail.com

**Гусарова Наталия Федоровна**

Университет ИТМО, факультет инфокоммуникационных технологий,

к.т.н., ст.н.с.

e-mail: natfed@list.ru

**УДК 004.04****РАЗРАБОТКА СИСТЕМЫ ДЛЯ АВТОМАТИЗАЦИИ  
ПЕРСОНАЛИЗИРОВАННОЙ ПОДДЕРЖКИ МЕДИЦИНСКИХ  
ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ С ПРИМЕНЕНИЕМ ПАТТЕРНОВ  
ПРОЕКТИРОВАНИЯ****Аксенов Ю.В., Возисов Н.С., Шорохов С.А.****Научный руководитель – к.т.н. Гусарова Н.Ф.**

В данной работе рассмотрено применение различных паттернов программирования в медицинской области, их анализ и сопоставление с задачами работы. Изучены возможные решения и представлены выбранные методики. Предложен разбор технологий на основе спроектированной и реализованной архитектуры и исходного кода.

**Ключевые слова:** состояние пациента, паттерны проектирования, автоматное программирование, фабрика, состояние конечного автомата.

Автоматизация процессов в наши дни затрагивает все больше областей жизни. Снятие с сотрудника лишней ответственности, переложение на информационные системы работы, как правило, положительно сказывается как на продуктивность компании в целом, так и на увеличении времени, которое человек может уделить работе, которую автоматизировать невозможно.

Целью проекта было создание рекомендательной автоматизированной системы по постановке диагнозов пациенту и рекомендациям по дальнейшему лечению его. Таким образом, разрабатываемую систему можно назвать уникальной.

**Анализ смежных решений.** Проанализировав существующие способы применения различных паттернов программирования в медицинской области, были выявлены несколько наиболее интересных для анализа.

Первая рассмотренная система реализует технологию машинного обучения (Multiclass Support Vector Machine, MSVM). Система на основе use case и предобученной MSVM причисляет его к тому или иному классу (в нашем случае болезни) и выдает советы по лечению пациента [1]. Использование машинного обучения может быть вполне обоснованным решением, так как поддерживает адаптивность системы. В любой момент в обучающую выборку можно добавить новый объект. В данном методе классом является болезнь, а объектом вектор признаков, которыми являются симптомы. Продукт, основанный на MSVM, имеет возможность дообучаться во время использования. Но у данного решения есть ряд недостатков, которые по итогу нивелируют все достоинства. Первой проблемой является переобучение метода (малая окрестность объектов из класса меньшинства классифицируется верно, а остальные объекты относятся к классу большинства). Также этот метод хорошо работает, когда граница между классами явно видна, однако в реальности обучающий датасет будет крайне зашумленным, а границы между классами не настолько явными. Это можно решить с помощью метода Лагранжиана, но итоговая точность определения болезней все еще будет недостаточно точной.

Параллельно был проанализирован один из паттернов организации хранения данных [2]. Данный паттерн реализован в HIS (Healthcare Information System), основной целью которой является поддержание медицинских данных в актуальном состоянии. Основная идея была предоставить врачу возможность самому формировать базу данных (БД), вплоть до настройки зависимости между сущностями. Универсальность этого метода его главный плюс и недостаток. Во-первых, невозможно быть уверенным в квалификации врачей в области проектирования БД. Каждому врачу потребуются, как минимум, на первое время консультант, чтобы сформированная БД соответствовала минимальным требованиям нормализации. Также неясно, как будет организовываться логика обработки данных из БД. Для анализа данных из каждой таблицы требуется написать отдельный сервис, который будет знать, что он получает из БД, что не представляется возможным. Для решения задачи данной работы это является критичным требованием.

Перед началом создания продукта на основе автоматного программирования требовалось проанализировать другое медицинское программное обеспечение (ПО) [3], основанное на автоматном методе программирования. Для реализации был использован язык программирования ДРАКОН. Плюсом языка является визуальное описание алгоритма работы автоматов, однако надо понимать, что автоматическая генерация кода на данный момент представляет формирование одного класса для всего решения, в котором собраны состояния, функции. Данное уплотнение программного кода в едином месте приводит к сложности в понимании работы кода, особенно если будет много вложенных if-ов. Генерируемый язык – JS, который не предоставляет объектно-ориентированное программирование парадигмы для структурированной организации кода.

Возможно использование множества взаимосвязанных автоматов. В ПО [4] присутствует несколько слоев автоматов, импульс каждого влияет на состояние другого автомата, система становится сложно анализируемой, что плохо для решения задачи, так как врач должен четко понимать причины определенных переходов, ведь в конечном итоге решение по лечению остается за врачом.

**Функциональность разрабатываемого продукта.** В качестве болезни была взята бронхиальная астма. Для формализации данных был проведен анализ болезни с выделением характерных признаков. На данный момент используется четыре показателя больного, которые определяют 10 различных состояний конечного автомата. Состояния, в которых может находиться пациент, были сформированы на

основе данных, полученных из описания медицинского технического процесса лечения на примере лечения астмы (рис. 1). Дальнейшее увеличение наполненности системы предполагает следование предложенному шаблону формирования состояний пациента.

Рис. 1. Общий вид Web-интерфейса

При добавлении нового пациента выпадает форма, в которой заносится основную личную информацию (имя, фамилию, пол и дату рождения). На момент создания в цепочке состояний пациента можно увидеть лишь начальное техническое состояние. Для получения первых советов по лечению пациента необходимо добавить его новое (актуальное на данный момент) состояние.

Затем по прошествии определенного промежутка времени требовалось актуализировать состояние пациента в системе с помощью добавлений новых состояний. На основании полученных с Web-интерфейса данных StateMachine, расположенная на сервере, принимает решение о переходе в определенное следующее состояние автомата или остается в текущем состоянии. Каждому состоянию соответствуют определенные советы по лекарствам, которые с точки зрения системы наиболее эффективны на данный момент. Сами советы для каждого состояния были определены экспертным способом.

Работа с конкретным пациентом заканчивается, когда автомат переходит в состояние EndState, состояние из которого никуда нельзя перейти.

### Общая структура системы

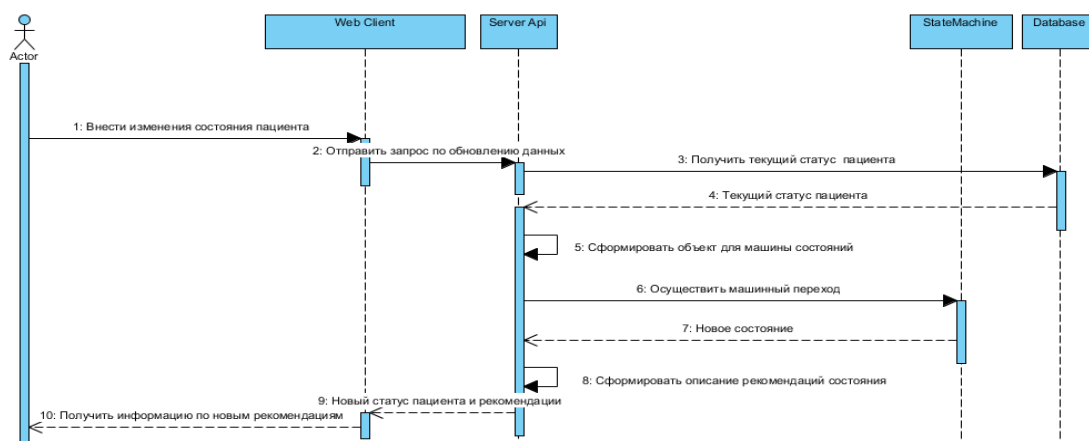


Рис. 2. Диаграмма последовательности продукта

На изображенной Sequence диаграмме (рис. 2) представлены основные компоненты продукта и методы их взаимодействия между собой.

В продукте были использованы порождающие, структурные и поведенческие паттерны. Рассмотрим несколько из них, которые позволяют показать применение для организации выполнения основной задачи (предназначения) разрабатываемого продукта.

**Паттерн конечный автомат.** Основной задачей было организовать логичные взаимосвязанные переходы от одного состояния пациента к другому. Было решено выбрать паттерн автоматного программирования, как наилучший для данной задачи [5]. Он позволяет обеспечить прозрачность принятия решений по переходу от одного состояния к другому, а также состояния автомата явно выделены за счет использования объектно-ориентированного языка.

Ниже показана диаграмма со всеми переходами состояний автомата в данном продукте. InitialState – начальное состояние, State\*number\* – промежуточные состояния, EndState – конечное состояние.

Весь путь начинается в InitialState, заканчивается в EndState. Количество переходов из одного состояния в другое может быть разное от 1 до  $n$ , что определяется спецификой диагноза и выбранных основных теоретических состояний. При этом InitialState и EndState – это технические (не теоретические) состояния, которые позволяют характеризовать конечные точки всего графа. На рис. 3 представлен граф состояний реализованного автомата для бронхиальной астмы.

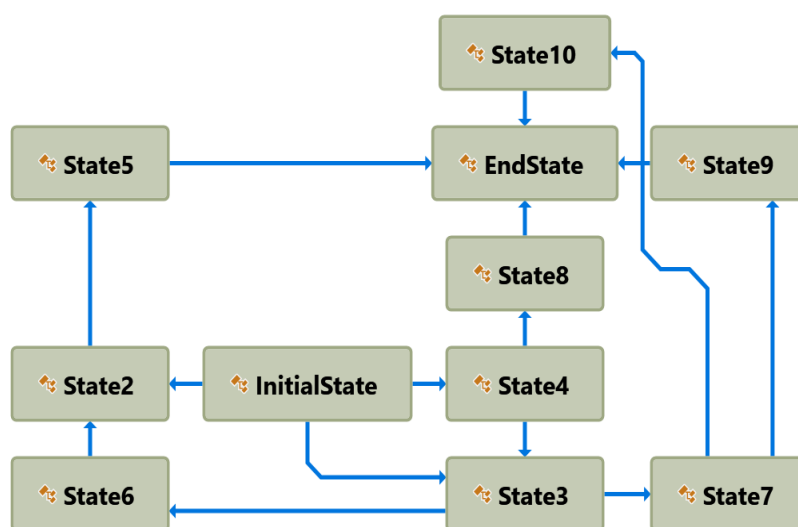


Рис. 3. Граф состояний

В системе существует базовый абстрактный класс состояния, от которого наследуются остальные состояния. Существование базового класса обосновывается тем, что это позволяет единообразно создавать цепочки состояний, а также позволяет содержать единообразную логику работы основных методов.

Основной метод базового класса – NextState, который дочерние классы должны в обязательном порядке переопределить. Данный метод включает в себе логику перехода в следующее состояние. Соответственно логика (форма метода) изменяется в зависимости от того, какое состояние имеется сейчас.

Далее рассмотрен пример начального состояния машины (InitialState). Как уже говорилось ранее, метод NextState имеет разную логику переходов (рис. 4). В текущем примере имеет переход в три других состояния из данного. При этом перехода может не случиться, если не было изменений или все условия перехода были ложные.

```

public override BaseState NextState(Status status)
{
    if (!status.IsAnyChanged()) return new InitialState();

    var pef = status.Parameters.Pef;
    var isWheezing = status.Parameters.IsWheezing;

    if (pef > 0.8 && !isWheezing && status.ElapsedTime().TotalHours >= 5) return new State2();

    if (pef < 0.6 && isWheezing) return new State3();

    if (pef < 0.6) return new State4();

    return new InitialState();
}

```

Рис. 4. Переопределение метода NextState

**Паттерн фабричный метод.** Для программного продукта была выбран один диагноз («Бронхиальная астма»), чтобы проанализировать применяемый подход, алгоритмы, выявить базовые структуры и поведения, которые в дальнейшем можно будет масштабировать. Поэтому на данный момент используется фабричный метод, что позволяет отработать базовую логику создания состояния. При увеличении количества диагнозов предполагается использовать абстрактную фабрику, в которую будет подаваться конкретная стратегия переходов (машины-состояния, заточенные под конкретный диагноз).

В текущей реализации используется два фабричных метода: фабрика состояние и фабричный метод статусов.

Фабрика состояний представляет собой процессный метод, по которому формируется «заготовка», которая в дальнейшем помещается в машину состояний. На основе данной заготовки производится анализ параметров пациента, а также формируется логика переходов в другие состояния. Дополнительно необходимо отметить, что состояния и статусы пациентов хранятся в реляционной БД, поэтому фабрика состояний выполняет еще функцию восстановления «упакованных» данных – восстановление звена в графе состояний.

В текущей реализации фабрики состояний имеет метод ProcessPatient, принимающий на вход текущее состояние в виде перечисления (полученное из БД) и текущий статус (получен из БД). Далее восстанавливает состояние уже в виде объекта и осуществляет вызов метода NextState для осуществления перехода в новое состояние.

В силу специфики реализации хранения и связности данных необходимо было создать фабричный метод для статусов пациента. Он представляет собой создание нового статуса, на основе полученных данных от клиента и текущего статуса пациента (полученные от БД). Осуществляется логика по упаковке нового статуса в неизменяемый объект, формируется связный список статусов путем присваивания необходимых идентификаторов объектов.

Как было упомянуто ранее, каждый статус представляет собой неизменяемый объект, т.е. при каждом внесении новых данных в систему он создается заново. В статусе есть StartDate и EndDate, на основе которого вычисляет пятый временной показатель пациента, который характеризует длительность данного статуса (рис. 5).

```

int StatusId { get; set; }
Parameters Parameters { get; private set; }
Status PreviousStatus { get; private set; }
DateTime StartDate { get; private set; }
DateTime EndDate { get; private set; }

```

Рис. 5. Исходный код класса Status

**Хранение данных.** Для хранения данных был выбор между хранением сериализованных данных в формате JSON и хранением данных в БД. Выбор был сделан в пользу хранения в БД, так как это позволяет обеспечить требования ACID средствами БД.

Все данные хранятся на MS SQL Server 2017. Данная структура хранения позволяет восстановить историю состояния пациента в течение его лечения. Метод хранения данных налагает специфические характеристики по восстановлению данных из БД. Для того чтобы данные были восстановлены из БД и продолжена с ними работа, необходимо было описать строителей объектов из «упакованных» данных. Свойства системы сохранять и восстанавливать состояния объектов являются реализацией паттерна Хранитель (Memento), что позволяет выносить внутреннее состояние объекта за его пределы для последующего возможного восстановления объекта без нарушения принципа инкапсуляции (рис. 6).

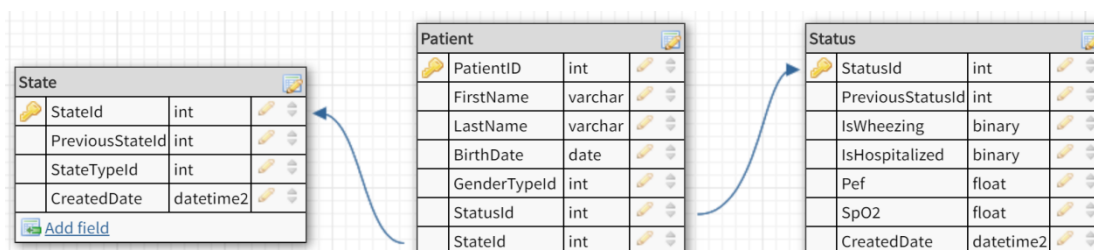


Рис. 6. Таблицы хранения информации в БД продукта

**Состояние разработки на сегодняшний день и перспективы ее развития.** В результате было реализовано Web-приложение, которое является ассистентом врача при лечении бронхиальной астмы у пациента. В дальнейшем концепт, используемый для данной болезни, можно распространить на больший список известных науке болезней.

Также при дальнейшей разработке требуется анализ возможных вводимых значений параметров, чтобы избежать закливание на одном состоянии автомата.

## Литература

1. Lalitha R., Latha B., Sumathi G. A software design technique for developing medical expert systems through use case analysis. [Электронный ресурс]. – Режим доступа: <http://www.alliedacademies.org/articles/a-software-design-technique-for-developing-medical-expert-systems-through-use-case-analysis.html> (дата обращения: 06.03.2019).
2. Muji M., Ciupa R.V., Dobru D., Bica C., Olah P., Bacarea V., Marusteri M. Database Design Patterns for Healthcare Information Systems [Электронный ресурс]. – Режим доступа: [https://www.researchgate.net/publication/227168121\\_Database\\_Design\\_Patterns\\_for\\_Healthcare\\_Information\\_Systems](https://www.researchgate.net/publication/227168121_Database_Design_Patterns_for_Healthcare_Information_Systems) (дата обращения: 06.03.2019).
3. Митькин С.Б. Автоматное программирование на языке ДРАКОН // Программная инженерия. – 2019. – Т. 10. – № 1. – С. 3–13.
4. Афанасьева И.В., Новиков Ф.А. Архитектура программного обеспечения систем оптической регистрации [Электронный ресурс]. – Режим доступа: <https://cyberleninka.ru/article/v/arhitektura-programmnogo-obespecheniya-sistem-opticheskoy-registratsii> (дата обращения: 06.03.2019).
5. Полицарпова Н.И., Шалыто А.А. Автоматное программирование [Электронный ресурс]. – Режим доступа: [http://is.ifmo.ru/books/\\_book.pdf](http://is.ifmo.ru/books/_book.pdf) (дата обращения: 06.03.2019).

**Алсуфьев Антон Анатольевич**

Год рождения: 1985

Университет ИТМО, факультет информационных технологий и программирования,  
e-mail: alsufyev@gmail.com**Старикова Екатерина Георгиевна**

Год рождения: 1986

Университет ИТМО, факультет информационных технологий и программирования,  
e-mail: katty.zhukova@gmail.com**Былецкий Дмитрий Александрович**

Год рождения: 1983

Университет ИТМО, корпоративная лаборатория технологий человеко-машинного взаимодействия, вед. инженер  
e-mail: byletskyi@yandex.ru**Матвеев Юрий Николаевич**

Год рождения: 1955

Университет ИТМО, факультет информационных технологий и программирования,  
д.т.н.  
e-mail: matveev@mail.ifmo.ru**УДК 65.01****ПРИМЕНЕНИЕ ГИБРИДНЫХ ПОДХОДОВ ПРИ УПРАВЛЕНИИ  
СОДЕРЖАНИЕМ В ПРОЕКТАХ В РАМКАХ ФЦП «ИССЛЕДОВАНИЯ  
И РАЗРАБОТКИ»****Алсуфьев А.А., Старикова Е.Г., Былецкий Д.А.  
Научный руководитель – д.т.н. Матвеев Ю.Н.**

В работе рассмотрено применение гибридных подходов при управлении содержанием в проектах в рамках федеральной целевой программы «Исследования и разработки», выделены ключевые заинтересованные стороны таких проектов, определены их ожидания, рассмотрены особенности управления проектами с учетом требования заинтересованных сторон, обоснована целесообразность использования гибридных подходов при управлении проектами с учетом требований заинтересованных сторон.

**Ключевые слова:** управление проектами, гибридные методологии, водопадная модель, гибкое управление проектами, федеральная целевая программа, Waterfall, Agile.

Целью данной работы являлось рассмотрение особенностей управления научно-исследовательскими проектами в рамках федеральной целевой программы «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014–2020 годы» (ФЦП ИР) с учетом требований трех сторон: Минобрнауки России, получателя субсидии и индустриального партнера. Рассмотрение проводилось на примере опыта реализации проектов корпоративной лаборатории технологий человеко-машинного взаимодействия Университета ИТМО.

Задачи работы:

- выделить ключевые заинтересованные стороны (ЗС) проектов, выполняемых в рамках ФЦП ИР;
- определить ожидания каждой из ЗС проектов;
- рассмотреть управление содержанием проектов с учетом требований ключевых ЗС;
- рассмотреть и обосновать целесообразность используемых методик управления проектами ФЦП ИР.



Схема работы в рамках проектов по ФЦП ИР выглядит следующим образом: Минобрнауки России на конкурсной основе предоставляет научным организациям субсидию на выполнение прикладных научных исследований и экспериментальных разработок (ПНИ или ПНИЭР). Дополнительное софинансирование предоставляет заранее определенный индустриальный партнер (предприятие реального сектора экономики). Получатель субсидии принимает на себя обязательство по достижению показателей результативности предоставления субсидии (индикаторов), определенных конкурсной документацией и соотносящихся со Стратегией научно-технологического развития Российской Федерации [1], предоставлению отчетности по результатам выполнения этапов проекта и проекта в целом, разработке технологии, которая необходима индустриальному партнеру. Результаты проекта в виде получивших правовую охрану результатов интеллектуальной деятельности передаются индустриальному партнеру для дальнейшей коммерциализации. В свою очередь, индустриальный партнер принимает на себя обязательство ежегодно в течение пяти лет отчитываться перед Минобрнауки России о коммерциализации результатов проекта. Схема взаимодействия ЗС приведена на рис. 1.

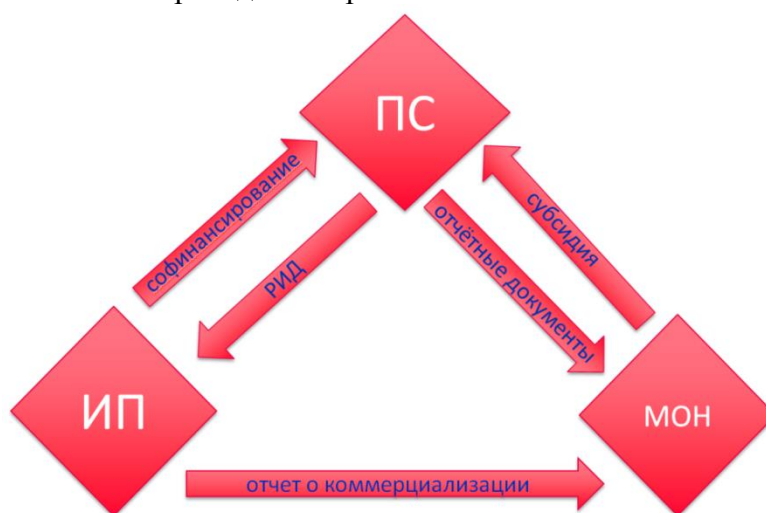


Рис. 1. Схема взаимодействия ЗС в проектах по ФЦП ИР: ИП – индустриальный партнер; ПС – получатель субсидии; МОИ – Минобрнауки России; РИД – результат интеллектуальной деятельности

Исходя из вышесказанного, можно выделить три основные заинтересованные стороны проектов в рамках ФЦП ИР:

1. Минобрнауки России;
2. получатель субсидии;
3. индустриальный партнер.

Ожидания каждой из заинтересованных сторон вытекают из логики ФЦП ИР.

Минобрнауки России, являясь распорядителем бюджетных средств, заинтересовано в том, чтобы данные средства расходовались целевым образом, реализуя при этом Стратегию научно-технологического развития России. Предоставляя средства субсидии на конкурсной основе, Минобрнауки России требует от получателя субсидии, во-первых, подробного отчета о расходовании бюджетных средств, во-вторых, выполнения взятых на себя обязательств по достижению количественных индикаторов проекта, закрепленных в договоре: научные публикации, индексируемые в базах данных Scopus и Web of Science; получившие правовую охрану результаты интеллектуальной деятельности; защищенные по результатам проекта диссертации; участие в мероприятиях по популяризации достижений науки, привлечение молодых ученых и т.п.

Индустриальный партнер, предоставляющий внебюджетное софинансирование, в первую очередь заинтересован в получении технологии, готовой к коммерциализации посредством создания и продажи инновационного продукта. Соответственно, основное ожидание данной заинтересованной стороны – получение технологии, отвечающей его коммерческим потребностям и стратегии развития.

Получатель субсидии (как правило, университет, научное учреждение) заинтересован в развитии собственных компетенций в предметной области, наращивании научно-технического задела, а также в получении вознаграждения в виде лицензионных платежей от индустриального партнера.

Для удовлетворения ожиданий каждой из заинтересованных сторон руководителю проекта необходимо выбрать ту методологию управления проектами, которая позволит получить желаемый для каждой из сторон результат в требуемые сроки и в рамках ограниченного бюджета. Таким образом, перед руководителями проекта появляется задача выбрать подходящие методологии, объединить их и эффективно применить в проектах ФЦП ИР.

При выборе методологии управления проектами можно руководствоваться следующими критериями:

- стратегия развития и ценности организации;
- имеющиеся ограничения в проекте;
- заинтересованные лица;
- риски;
- сложность проекта;
- масштаб и стоимость проекта [2].

На кафедре РИС Университета ИТМО с 2014 года в проектах ФЦП ИР применяется водопадная модель управления, позволяющая учитывать проектные риски, оптимизировать использование трудовых ресурсов, контролировать исполнение составленного плана, накапливать, анализировать и использовать в дальнейшем опыт реализованных проектов. Водопадная модель управления проектами имеет последовательную логику реализации процесса разработки, используется во многих отраслях, в том числе при разработке программного обеспечения. Водопадная модель включает в себя такие этапы, как анализ требований, проектирование, тестирование, реализация [3], которые выполняются в определенном порядке. В соответствии с ГОСТ 15.101-98 [4], на основе которого составляется план-график проектов по ФЦП ИР, проекты включают этапы выбора направления исследований, теоретические исследования (на этом этапе происходит сама разработка), экспериментальные исследования, обобщение и оценку результатов работ. Выполнение проекта в соответствии с водопадной моделью обосновано требованиями Минобрнауки России по строгому следованию плану-графику и поэтапному предоставлению отчетности. Водопадная модель имеет ряд преимуществ: например, формализованный этап планирования обеспечивает полную и согласованную документацию на каждом этапе; также такая модель позволяет легко определять и контролировать сроки и затраты на проект.

Однако данная методология имеет ряд недостатков в сложившейся структуре процесса выполнения проектов ФЦП ИР как раз из-за того, что индустриальному партнеру как одной из ключевых заинтересованных сторон необходима конкурентоспособная технология, способная к коммерциализации. Так как длительность проекта ФЦП ИР составляет 2–3 года, то к моменту начала самой разработки изначально прописанные в техническом задании требования могут устареть, к тому же в техническом задании прописаны только общие требования к техническому результату (технологии или системе). В такой ситуации индустриальный партнер вынужден уточнять и детализировать требования для того, чтобы в результате у него была технология, отвечающая требованиям рынка. Применение

формализованного подхода при постоянно уточняющихся требованиях приводит к качественному росту затрат из-за переработок, отставанию от графика и вынуждает каждый раз переходить к начальной фазе выполнения проекта.

В связи с этим было принято решение рассмотреть другой подход к управлению проектами, а именно, обратить внимание на Agile-методологию или гибкую методологию управления проектами. Изначально методология разрабатывалась для проектов, которым необходима высокая гибкость и быстрая реализация [5]. Проектирование разбивается на короткие итерации – спринты (Scrum-метод). Agile подходит для проектов, требующих интенсивного взаимодействия между исполнителями, не нуждающимися в дополнительном контроле, благодаря чему появляется возможность быстро подстраиваться под изменяющиеся требования. Одно из главных преимуществ такой методологии состоит в том, что можно быстро выявлять проблемы и вносить необходимые изменения на ранней стадии разработки, не дожидаясь завершения тестирования. Agile-проектирование обеспечивает применение повторяющихся процессов, снижение рисков, оперативную обратную связь. Данная методология подходит для ведения процесса разработки для удовлетворения ожиданий индустриального партнера. Однако для такой заинтересованной стороны, как Минобрнауки России, данная методология не может быть эффективной при выполнении индикаторов, таких как публикации, РИДы, отчеты.

Недостатки и преимущества описанных выше подходов привели к необходимости использовать гибридную методологию, когда этапы планирования и определения начальных требований выполняются согласно методологии водопада, а этапы проектирования, разработки и тестирования соответствуют гибкому подходу.

Таким образом, в ходе работы над проектом ФЦП ИР совмещается несколько методик управления проектами: предиктивный цикл (водопадная модель) и адаптивный цикл (гибкий подход или Agile) [3]. Это объясняется следующим: с одной стороны, в договоре между получателем субсидии и Минобрнауки России строго прописаны сроки работ, этапы, контрольные точки, стоимость работ, из-за чего целесообразным становится использование водопадной модели – в рамках проекта в целом, включающего научные исследования и разработку, она оказывается более эффективной; с другой стороны, требования к конечному прикладному результату работы (в первую очередь к разрабатываемому в ходе проекта экспериментальному образцу программного комплекса) прописываются в техническом задании в общем виде и на этапе разработки экспериментального образца могут уточняться индустриальным партнером, выступающим в качестве функционального заказчика. Таким образом, на этапе разработки программного обеспечения целесообразным оказывается использование гибких методологий (инкрементный или итеративный подход).

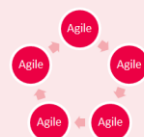
Waterfall	Agile
Выбор направления исследований	Нет
Теоретические исследования	
Экспериментальные исследования	Нет
Обобщение и оценка результатов работ	Нет

Рис. 2. Схема выполнения проекта в рамках ФЦП ИР в соответствии с ГОСТ 15.101-98 с применением методологий Waterfall и Agile

На рис. 2 представлена схема выполнения проекта в рамках ФЦП ИР в соответствии с ГОСТ 15.101-98, а также место в данной схеме методологий Waterfall и Agile.

Можно сделать вывод, что у каждой методологии управления проектами имеются свои сильные и слабые стороны, поэтому можно использовать сразу несколько методологий, исходя из уникальной природы проекта, его целей и организационной структуры.

В данной работе были рассмотрены такие особенности управления научно-исследовательскими проектами в рамках ФЦП ИР, как наличие трех основных заинтересованных сторон проектов (Минобрнауки России, получатель субсидии, индустриальный партнер), подходы к управлению содержанием проекта с учетом ожиданий каждой из заинтересованных сторон. Выбранные методологии управления проектами необходимо корректировать с учетом развития самих методологий, внешних условий и меняющихся потребностей заинтересованных лиц. Управление проектами – это один из шагов по отладке механизма коммерциализации разработок, основной инструмент эффективного общения научных учреждений с коммерческими организациями с целью создания продуктов и технологий, востребованных рынком, а применение гибридных подходов к управлению проектами позволяет качественно управлять ожиданиями заинтересованных сторон.

### Литература

1. Стратегия научно-технологического развития Российской Федерации: утв. указом Президента Рос. Федерации от 1 дек. 2016 г. № 642 // Собр. законодательства Рос. Федерации, 2016. № 49, ст. 6887.
2. Александр М. Как выбрать наилучшую методологию управления проектами [Электронный ресурс]. – Режим доступа: <https://www.osp.ru/cio/2017/08/13053143/> (дата обращения 25.03.2019).
3. Руководство к своду знаний по управлению проектом (руководство РМВОК). – Шестое изд. – Pennsylvania: Project Management Institute, Inc., 2017. – 131 с.
4. ГОСТ 15.101-98. Система разработки и постановки продукции на производство. Порядок выполнения научно-исследовательских работ. – Минск: ИПК Издательство стандартов, 2003. – 11 с.
5. Алджанов В. ИТ-архитектура. Практическое руководство от А до Я. – Изд-во Litres, 2018. – 1400 с.

**Аусев Егор Владимирович**

Год рождения: 1996

Университет ИТМО, факультет информационных технологий и программирования, студент группы № М41211

Направление подготовки: 09.04.02 – Речевые информационные системы

e-mail: ausev5egor@gmail.com

**Новоселов Сергей Александрович**

Год рождения: 1985

ООО «ЦРТ»; Университет ИТМО, корпоративная лаборатория технологий человеко-машинного взаимодействия, к.т.н.

e-mail: emailname@email.ru

**УДК 004.93****ИССЛЕДОВАНИЕ ТЕХНОЛОГИЙ МАШИННОГО ПЕРЕВОДА  
ДЛЯ ВЕРИФИКАЦИИ ПРОИЗНЕСЕННОЙ ФРАЗЫ****Аусев Е.В.** (Университет ИТМО)**Научный руководитель – к.т.н. Новоселов С.А.** (ООО «ЦРТ»; Университет ИТМО)

Работа выполнена в рамках темы НИР № 618278 «Синтез моциональной речи на основе генеративных состязательных сетей».

Различные программные продукты реализуют верификацию фразы. Также существует множество разных методов для верификации. Для этого используются различные способы цифровой обработки сигнала. Однако данная задача в настоящее время не решается с высокой точностью. Целью работы являлось создание программного продукта, реализующего вышеописанный функционал, и исследование его свойств. В работе рассказано о представлении слов в векторном пространстве (embeddings), в котором векторное представление близких по смыслу слов находится в одной окрестности. Подробнее о векторном представлении можете прочесть в работах, посвященных данной тематике. Также продемонстрирована технология, используемая в онлайн-переводчиках, которая помогает преобразовать текст заданной длины в текст произвольной длины. В будущем данная технология будет использоваться, но вместо одного из языков будут подаваться звуковые дорожки. Планируется, что данная работа будет затрагивать тематику рекуррентных нейронных сетей.

**Ключевые слова:** машинный перевод, Google переводчик, Yandex переводчик, Promt, векторное представление слов.

В рамках работы рассмотрены различные вариации машинного перевода. Проблематики машинного перевода и верификации фразы имеют общие черты. Основная идея – представить звуковую дорожку в пространстве embedding, далее использовать технологию машинного перевода для задачи верификации [1, 2].

Машинный перевод – это автоматический перевод текста, используя произвольный программный продукт. В работе рассмотрены основные представители машинного перевода: Google переводчик, Yandex переводчик, Promt [3].

Для начала рассмотрим один из популярнейших переводчиков – Promt. Promt использует следующие технологии: машинный перевод на основе правил, статистический машинный перевод, а также двуязычное оценивание (BLUE). Остановимся на каждой технологии подробнее.

Машинный перевод на основе правил – общий термин, который обозначает системы машинного перевода на основе лингвистической информации об исходном и переводном языках. Они состоят из двуязычных словарей и грамматик, охватывающих основные семантические, морфологические, синтаксические закономерности каждого

языка. Такой подход к машинному переводу еще называют классическим. На основе этих данных исходный текст последовательно, по предложениям, преобразуется в текст перевода. Эти системы противопоставляют системам машинного перевода, которые основаны на примерах. Принцип работы таких систем – связь структуры входного и выходного предложения.

Статистический машинный перевод – разновидность машинного перевода, где перевод генерируется на основе статистических моделей, параметры которых являются производными от анализа двуязычных корпусов текста [4]. Но под понятием «статистического машинного перевода» подразумевается общий подход к решению проблемы перевода, который основан на поиске наиболее вероятного перевода предложения с использованием данных, полученных из двуязычной совокупности текстов.

Двуязычное оценивание (BLUE) – это алгоритм для оценки качества текста, который был переведен с одного языка на другой. Качество считается соответствием между переводом машины и человека: «чем ближе машинный перевод к человеческому переводу, тем лучше» – это основная идея BLEU. BLEU был одним из первых показателей, заявивших о высокой корреляции с человеческими суждениями о качестве, и остается одним из самых популярных автоматизированных и недорогих показателей. Оценки рассчитываются для отдельных переведенных сегментов – обычно предложений – путем сравнения их с набором справочных переводов хорошего качества. Эти оценки затем усредняются, чтобы получить оценку общего качества перевода. Разборчивость или грамматическая правильность не принимаются во внимание.

По итогам можно выделить сильные и слабые стороны данного переводчика. Из положительных вещей можно выделить: синтаксическая и морфологическая точность, стабильность и предсказуемость результата, возможность настройки на предметную область. Из недостатков можно выделить: трудоемкость и длительность разработки, необходимость поддерживать и актуализировать лингвистические базы данных.

Следующим в рассмотрении был Google переводчик. Данный переводчик имеет следующий набор технологий: нейронный машинный перевод, долгая краткосрочная память, двуязычное оценивание (BLUE). Рассмотрим повнимательнее данные технологии.

Модель нейронный машинный перевод используют глубинное обучение и обучение признаков. Для их работы требуется лишь малая часть памяти по сравнению с традиционными системами статистического машинного перевода [5]. Кроме того, в отличие от традиционных систем перевода, все части модели нейронного перевода обучаются совместно (от начала до конца), чтобы максимизировать эффективность перевода. Двухнаправленная рекуррентная сеть или долгая краткосрочная память, также известная как кодировщик, используется нейронной сетью для кодирования исходного предложения для второй рекуррентной сети, также известной как декодировщик, которая используется для предсказания слов в конечном языке.

Преимущества данной модели машинного перевода заключаются в следующем: рассматриваются целые предложения, происходит понимание сходства между словами. Из недостатков можно выделить: возможность переводиться только на основе предложения, требуется большое количество переведенных текстов.

И заключительной моделью машинного перевода в нашем рассмотрении является Yandex переводчик. Данная модель использует следующие технологии: нейронный машинный перевод, статистический машинный перевод. Исходя из вышесказанного, можно сделать положительные выводы: запоминание редких и сложных слов и фраз, гладкость перевода. К отрицательным сторонам относятся: многочисленные грамматические ошибки, нестабильность перевода.

Проанализировав все преимущества и недостатки рассмотренных моделей можно сделать вывод: для задачи верификации фразы наилучшим подходом будет являться технология нейронного машинного перевода.

Первичной задачей при работе со звуковыми сигналами – является представление в произвольном пространстве embedding. В рамках данной работы был рассмотрен метод Word2Vec.

Word2Vec – это программный инструмент для анализа семантики естественных языков. На рисунке представлена простейшая архитектура данной технологии.

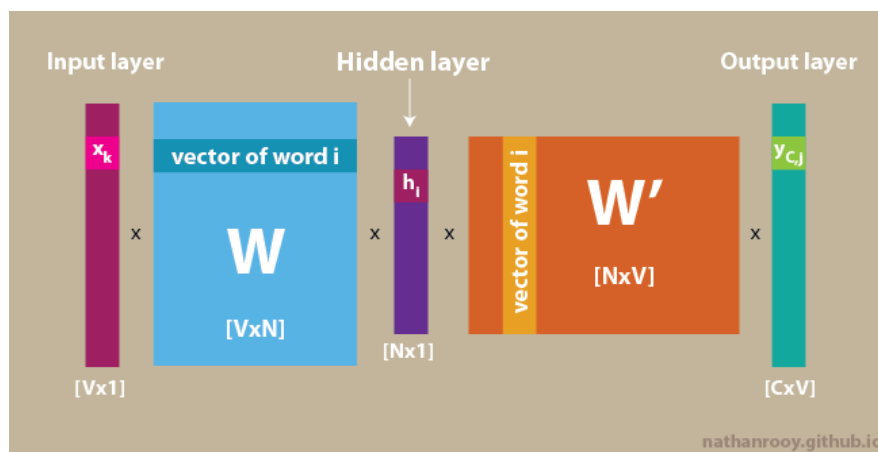


Рисунок. Простейшая архитектура Word2Vec

На рисунке входной слой представляет собой one-hot encoding словаря всего набора данных. Средний уровень является векторным представлением произвольной размерности, в зависимости от матрицы. Выходным слоем является семантическая оценка к каждому слову в алфавите.

Исходя из этого, Word2Vec является преимущественной технологией при работе и анализе языка.

### Литература

1. Wu L., Yen I.E.H., Xu K., Xu F., Balakrishnan A., Chen P.-Y., Ravikumar P., Witbrock M.J. Word Mover's Embedding: From Word2Vec to Document Embedding [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1811.01713.pdf> (дата обращения: 06.03.2019).
2. Mikolov T., Sutskever I., Chen K., Corrado G., Dean J. Distributed Representations of Words and Phrases and their Compositionality [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1310.4546.pdf> (дата обращения: 06.03.2019).
3. Yan-Bo Lin, Yu-Jhe Li and Yu-Chiang Frank Wang. Dual-modality seq2seq network for audio-visual event localization [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1902.07473.pdf> (дата обращения: 06.03.2019).
4. Chang P., Galley M., Manning C. Optimizing Chinese Word Segmentation for Machine Translation Performance // Proceedings of the Third Workshop on Statistical Machine Translation. – 2008. – P. 224–232.
5. Wu Y., Schuster M. et al. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1609.08144.pdf> (дата обращения: 06.03.2019).

**Бебяков Александр Михайлович**

Год рождения: 1995

Университет ИТМО, факультет информационных технологий и программирования, аспирант

Направление подготовки: 09.06.01 – Информатика и вычислительная техника

e-mail: ambebiakov@itmo.ru

**Довидченко Никита Владимирович**

Год рождения: 1984

ООО «Кномикс»

dovidchenko@atlas.ru

**Одинцова Вероника Евгеньевна**

ООО «Кномикс»

odintsova@atlas.ru

**Клименко Наталья Сергеевна**

Год рождения: 1989

ООО «Кномикс»

natasha.klmnk@gmail.com

**Тяht Александр Викторович**

Год рождения: 1983

ООО «Кномикс»

a.tyakht@gmail.com

**Ульянцев Владимир Игоревич**

Год рождения: 1990

Университет ИТМО, факультет информационных технологий и программирования, к.т.н., доцент

e-mail: ulyantsev@itmo.ru

**УДК 004.021****СРАВНИТЕЛЬНЫЙ АНАЛИЗ СОСТАВА МИКРОБИОМА КИШЕЧНИКА  
ЧЕЛОВЕКА ПРИ ЯЗВЕННОМ КОЛИТЕ И БОЛЕЗНИ КРОНА ПО ДАННЫМ  
WGS-МЕТАГЕНОМИКИ****Бебяков А.М.** (Университет ИТМО), **Довидченко Н.В.** (ООО «Кномикс»),**Одинцова В.Е.** (ООО «Кномикс»), **Клименко Н.С.** (ООО «Кномикс»),**Тяht А.В.** (ООО «Кномикс»)**Научный руководитель – к.т.н., доцент Ульянцев В.И.** (Университет ИТМО)

В работе представлены результаты обработки данных полногеномного секвенирования образцов кишечной микробиоты и сопоставления особенностей таксономического и функционального состава среды между когортами пациентов с воспалительными заболеваниями кишечника и группы контроля. Полученные результаты использовались для отладки работы конвейера анализа WGS-данных в составе платформы «Кномикс-Биота». Сравнительный анализ метагеномных данных образцов микробиома являются одним из методов исследования потенциальных микробиотных маркеров заболеваний.

**Ключевые слова:** биоинформатика, метагеномика, WGS, воспалительные заболевания кишечника, SaaS.

Воспалительные заболевания кишечника (ВЗК) являются хроническими рецидивирующими заболеваниями, сопровождающимися поражениями слизистой оболочки желудочно-кишечного тракта (ЖКТ), кишечными кровотечениями, болями в животе. Заболевания в основном поражают лиц трудоспособного возраста, и занимают



одну из ведущих позиций по тяжести течения и летальности среди болезней пищеварительной системы. При диагностике часто возможно выявить общие воспалительные процессы, но дифференциальная диагностика между подклассами заболевания – язвенным колитом (ЯК) и болезнью Крона (БК) является сложной задачей. Возможность применения маркеров [1], основанных на анализе состава микробиоты и его изменения, в качестве средства ранней диагностики считается перспективной ввиду низкой инвазивности относительно применяющихся средств диагностики, таких как колоноскопия и биопсия.

Этиология ВЗК не установлена, генетика является лишь одним из причинных факторов [2]. На развитие заболевания и эффективность средств лечения могут оказывать влияние такие факторы, как дефекты иммунитета, привычки питания, индекс массы тела, возраст, пол, не исключается и роль состава микробиоты. В ходе исследования течения заболеваний было отмечено изменение состава микробиомного сообщества кишечника даже на ранних этапах [1]. На основании разведочного и статистического анализа данных метагеномных исследований микробиома человека при воспалительных заболеваниях кишечника возможно обнаружение ассоциаций изменения состава микробиоты с клиническими проявлениями, выявление потенциальных микробиотных маркеров ВЗК, а также развитие клинической диагностики данных заболеваний.

Технология метагеномного анализа образцов среды широко применяется в исследованиях микробиоты человека и ассоциации изменения состава микробиоты с проявлениями заболеваний, а также влиянием лекарственных средств. Внедрению такой методики, как анализ на основе секвенирования методом «дробовика» (WGS), следует уделить значительное внимание. В отличие от секвенирования 16S рРНК, подобные результаты позволяют более точно оценить генетический состав бактерий, населяющих среду, и так оценить функциональный потенциал микробиома, но приводят к формированию более значительных объемов данных, предполагающих применение более сложных и высокопроизводительных методов обработки [3].

Важным аспектом развития исследований микробиоты человека является обеспечение доступности и воспроизводимости шагов и результатов анализа данных с помощью различных инструментов биоинформатического анализа, представленными в основном общедоступными пакетами Linux-систем [4]. Аналитическая платформа «Кномикс-Биота» (<https://biota.knomics.ru/>) предоставляет облачный сервис проведения высокопроизводительного анализа метагеномных данных для исследователей, не обладающих доступом к необходимым вычислительным ресурсам или не владеющих навыками работы в командной строке и с программным кодом. В рамках данной платформы реализованы вычислительные конвейеры как статистического анализа 16S и WGS метагеномных данных образцов, так и сравнительного анализа с исследованными базами метагеномов.

В работе рассматривался конвейер обработки WGS-данных.

Данные метагеномного WGS представляют собой результат полногеномного секвенирования тотального генетического материала микробиома, при котором в течение нескольких часов генерируются короткие геномные прочтения организмов, населяющих среду.

Одним из основных этапов обработки метагеномных данных является анализ относительной представленности микробных таксонов в изучаемой среде. Для определения таксономического и функционального состава среды применяется HUMAnN2 – система анализа метаболического потенциала, включающая вычисление:

- относительной представленности организмов, используя базу MetaPhlAn2;
- представленности семейств генов, применяя базы ChocoPhlAn и UniRef;
- представленности метаболических путей с помощью базы MetaCyc.

Дальнейший статистический анализ ассоциации общего таксономического и функционального состава среды, характеризуемого относительной представленностью метаболических путей, с предоставленными факторами выполняется с помощью adonis – функции R-пакета «vegan» для анализа ассоциации на основе матриц расстояний с тестированием перестановками. Расчет матрицы расстояний для относительной представленности таксонов использует метрику, учитывающую структурную организацию таксонов – weighted UniFrac. Для относительной представленности метаболических путей применяется метрика Bray-Curtis.

Для выявления отдельных таксонов и отдельных метаболических путей, обладающих ассоциацией с рассматриваемыми факторами, используется masslin – функция R-пакета «MaAsLin» составления обобщенных линейных моделей с последующим отбором переменных и выявлением значимых ассоциаций.

Для отладки работы системы анализа WGS-данных, применяющей отмеченные методы, производилась обработка данных секвенирования образцов микробиоты кишечника:

- 10 образцов пациентов с болезнью Крона;
- 10 образцов пациентов с язвенным колитом;
- 10 контрольных образцов.

Данные для анализа предоставлены Казанским (приволжским) федеральным университетом.

В результате обработки данных получены таблицы относительной представленности таксонов и относительной представленности метаболических путей.

На рисунке изображены в порядке убывания среднего значения относительные представленности двадцати наиболее представленных таксонов ранга род в образцах разных групп: БК – болезнь Крона, ЯК – язвенный колит, К – контроль

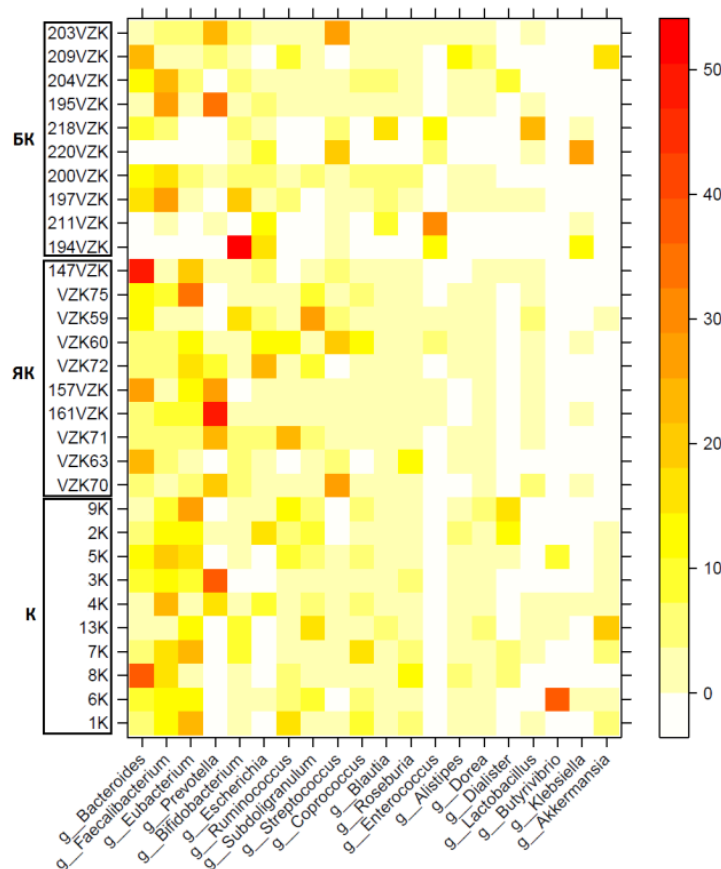


Рисунок. Относительная представленность таксонов ранга род

В качестве фактора для статистического анализа использовалась принадлежность образцов к указанным группам (group): болезнь Крона – CD; язвенный колит – UC; контроль – control.

По результатам анализа выявлено наличие ассоциации с фактором group как общего таксономического состава среды, так и общего функционального состава среды (с p-value равными 0,0003 и 0,0084 соответственно). Метод Adonis, число перестановок – 20000, пороговое p-значение – 0,05.

В таблице отображены значимые результаты анализа ассоциации представленности отдельных таксонов разных рангов с фактором group, где группы сравниваются попарно. Метод MaAsLin, пороговое p-значение – 0,05, применяется поправка на множественные сравнения – метод Бенджамини–Хохберга. Для метаболических путей выявлена значимая ассоциация (p-value=0,03) представленности реакции преобразования  $\alpha$ -D-глюкозы в  $\beta$ -D-глюкозу с принадлежностью образца к группе UC относительно группы CD.

Таблица. MaAsLin, таксоны

taxon	taxa level	covariate	group, compared with	coefficient	p-value	adjusted p-value
f__Eubacteriaceae	family	group[control]	CD	0,257242	0,0002	0,002
f__Eubacteriaceae	family	group[UC]	CD	0,217639	0,0010	0,005
f__Ruminococcaceae	family	group[control]	CD	0,202698	0,0119	0,040
g__Eubacterium	genus	group[control]	CD	0,257521	0,0002	0,004
g__Eubacterium	genus	group[UC]	CD	0,217918	0,0010	0,012
g__Subdoligranulum	genus	group[control]	CD	0,163849	0,0060	0,029
g__Parabacteroides	genus	group[UC]	CD	0,060376	0,0091	0,036
g__Barnesiella	genus	group[UC]	CD	0,055608	0,0037	0,024
g__Barnesiella	genus	group[control]	CD	0,055049	0,0040	0,024
s__Bacteroides_uniformis	species	group[UC]	CD	0,089004	0,0073	0,044
s__Eubacterium_ramulus	species	group[control]	CD	0,062683	0,0001	0,003
s__Barnesiella_intestinihominis	species	group[UC]	CD	0,056408	0,0034	0,033
s__Barnesiella_intestinihominis	species	group[control]	CD	0,055849	0,0036	0,033
s__Coprococcus_catus	species	group[control]	CD	0,047042	0,0034	0,033
s__Coprococcus_catus	species	group[UC]	CD	0,042880	0,0069	0,044
s__Bacteroides_eggerthii	species	group[control]	CD	0,026301	0,0001	0,003

В работе были представлены результаты обработки WGS-данных секвенирования образцов микробиоты пациентов с воспалительными заболеваниями кишечника и

контрольных образцов. Была показана работоспособность системы анализа WGS-данных в составе платформы «Кномикс-Биота».

### Литература

1. Sommer F. et al. Microbiomarkers in inflammatory bowel diseases: caveats come with caviar // *Gu.* – 2017. – V. 66. – P. 1734–1738.
2. Weersma R.K. et al. Multiomics Analyses to Deliver the Most Effective Treatment to Every Patient With Inflammatory Bowel Disease // *Gastroenterology.* – 2018. – V. 155. – № 5. – P. 1–4.
3. Tyakht A. et al. Genetic diversity of *Escherichia coli* in gut microbiota of patients with Crohn's disease discovered using metagenomic and genomic analyses [Электронный ресурс]. – Режим доступа: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6307143/> (дата обращения: 06.03.2019).
4. Efimova D., Tyakht A. et al. Knomics-Biota - a system for exploratory analysis of human gut microbiota data [Электронный ресурс]. – Режим доступа: <https://biodatamining.biomedcentral.com/articles/10.1186/s13040-018-0187-3> (дата обращения: 06.03.2019).

**Большаков Никита Андреевич**

Год рождения: 1996

Университет ИТМО, факультет информационных технологий и программирования, студент группы № М41212

Направление подготовки: 09.04.02 – Информационные системы и технологии

e-mail: snowmanikita@gmail.com

**Шуранов Евгений Витальевич**

Год рождения: 1980

ООО «ЦРТ», лидер продуктовой разработки, к.т.н.

e-mail: shuranov@speechpro.com

**УДК 004.93****ДЕТЕКТИРОВАНИЕ АКУСТИЧЕСКИХ СОБЫТИЙ С ИСПОЛЬЗОВАНИЕМ АУГМЕНТАЦИЙ ДАННЫХ И НЕЙРОННЫХ СЕТЕЙ****Большаков Н.А.** (Университет ИТМО)**Научный руководитель – к.т.н. Шуранов Е.В.**

(ООО «ЦРТ»)

Работа выполнена в рамках темы НИР № 618278 «Синтез моциональной речи на основе генеративных состязательных сетей».

Целью работы являлся обзор способов построения сверточных нейронных сетей с обучением на аугментированных данных, а также методов аугментации данных.

**Ключевые слова:** акустические события, детектирование, искусственные нейронные сети, аугментация данных.

Акустическое событие – это звуковое явление, определенной длительности, времени возникновения, исчезновения и типа. Задача детектирования такого события может быть решена различными способами [1], и один из наиболее эффективных способов, ставший возможным с ростом вычислительных мощностей, использование различных нейронных сетей.

Сама идея создания аналога нейронной сети живого организма для решения задач логического исчисления была высказана еще в середине XX века [2], но наибольший рост популярности и привлекательности применения данной технологии обуславливается появлением совокупности семейств различных методов машинного обучения, впоследствии названных глубоким обучением.

Однако не всегда исходные данные являются достаточными для обучения точной итоговой модели. В определенных случаях возможен вариант, когда имеющийся набор обучающих событий, в силу своей несбалансированности, плохой разметки или иных причин, становится источником таких ошибок как переобучение, потеря точности и т.д.

В работе рассматривались способы решения задачи детектирования акустических событий, реализованные командой Констанцкого университета, в рамках DCASE Community Challenge 2018 [3]. Также в работе описываются методы борьбы с проблемой переобучения и балансировки исходных наборов данных посредством различных аугментаций над ними.

Для начала было рассмотрено визуальное представление нейронных сетей, использованных для решения поставленной задачи детектирования. В качестве основного метода взята искусственная нейронная сеть (ИНС), использующая предобученные веса из двух более простых нейронных сетей.

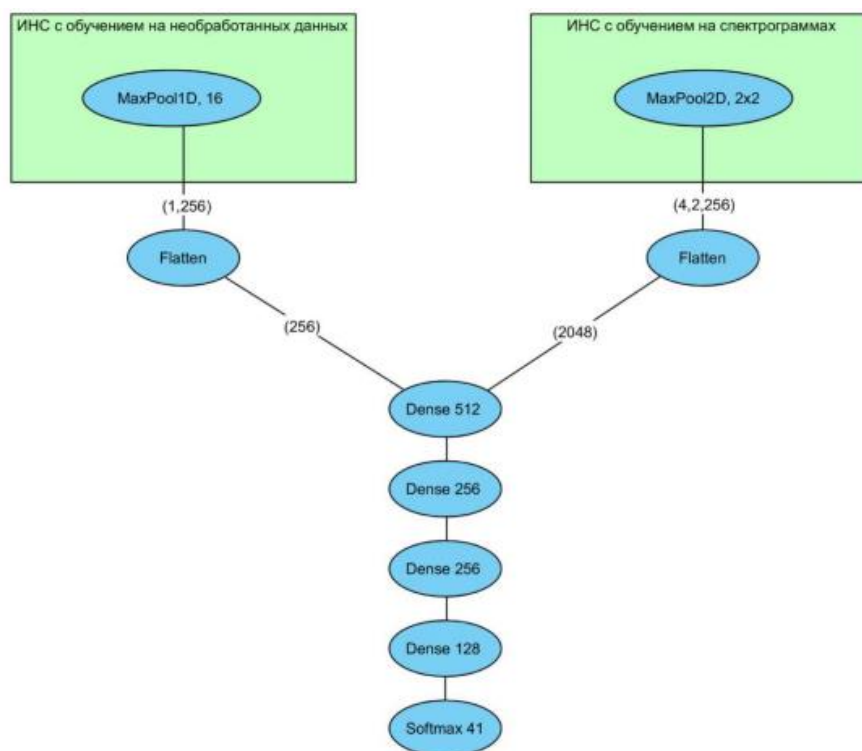


Рис. 1. Визуальное представление ИНС с обучением на комбинированных данных

Как видно из представленного изображения (рис. 1), в архитектуре участвуют ИНС с обучением на необработанных данных, т.е. на сериализованных звуковых файлах, и с обучением на мел-спектрограммах.

Подобный подход реализует собой простой и действенный способ повышения качества решения задачи путем построения нескольких простых ИНС, вместо переусложнения уже существующей.

Теперь перейдем к визуальным представлениям этих ИНС, с которыми можно ознакомиться на рис. 2.

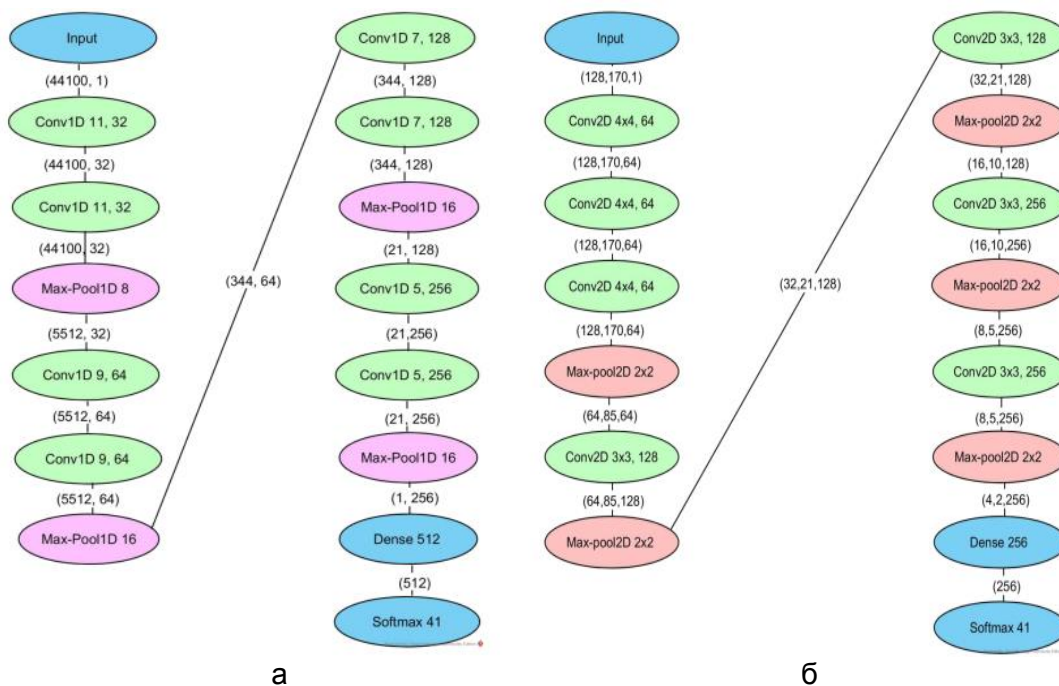


Рис. 2. Визуальное представление используемых ИНС: с обучением на необработанных данных (а); с обучением на спектрограммах (б)

Остановимся подробнее на некоторых из основных элементов:

- Conv – это основные слои сверточной нейронной сети. Они содержат фильтры для разных каналов, и обрабатывают предыдущие слои по частям с помощью так называемых ядер свертки. Построение нейронной сети с использованием Conv2D, дает возможность обучения по спектрограмме, т.е. по двумерному изображению, где будут учитываться не только соседние значения в одномерном массиве как при обычной интерпретации акустического события [4];
- Max-pooling – слой подвыборки, в котором происходит нелинейное уплотнение карты признаков, где несколько выбранных значений с помощью функции максимума уплотняются до одного [4].

Вышеописанные нейронные сети используют хоть и многослойные, но единственные для метода модели.

Следует отметить, что комбинированная модель замещает Dense и Softmax слои, которые расширяют их в своей реализации.

Однако предоставленный организаторами DCASE Community Challenge 2018, набор данных для обучения имел в себе интересующий нас изъян, он не был сбалансирован. И здесь будет уместным рассмотреть некоторые подходы к аугментации акустических событий.

Под аугментацией в области машинного обучения обычно подразумевается искусственное расширение исходного обучающего набора событий, путем проведения различных преобразований над ним, например, линейных.

В рамках данной работы были применены следующие способы аугментации:

- случайная обрезка данных – исходное событие дублируется, затем полученный дублер сокращается путем случайного извлечения части события и присваивания этой части метки исходного класса;
- случайное заполнение данными – исходное событие дублируется, затем полученный дублер увеличивается путем самокопирования до заданного размера и присваивания метки исходного класса;
- временной сдвиг – из одного исходного события берется несколько частей, путем применения случайных обрезки и заполнения данными в разных временных промежутках;
- аугментации одного класса – два события одного класса проходят линейные преобразования и аккумулируются в одно новое событие с присваиванием метки этого класса;
- аугментация смешиванием – практически то же, что и при аугментации одного класса, однако события принадлежат разным классам, и их метки для детектирования также проходят смешивание.

Результаты проведенного тестирования изложенных методов представлены в таблице. Тестирование проводилось при обучении на 9 000 объектов с 41 меткой. Поскольку задачей исследования являлось именно изучение предложенных методов, результаты несколько ниже представленных на DCASE Community Challenge 2018, но они не только подтверждают эффективность изложенных методов, но и позволяют оценить повышение точности решений при использовании различных ИНС и использованием аугментационных данных. Все методы были реализованы с использованием открытой нейросетевой библиотеки Keras, на языке программирования Python, с использованием вышеизложенных визуальных представлений в качестве основы.

Таблица. Результаты тестирования

Тестирование	Необработанные данные		Спектрограммы		Комбинированные данные	
	Best Score	Worst Score	Best Score	Worst Score	Best Score	Worst Score
Без аугментаций	0,701	0,689	0,736	0,712	0,763	0,724
С аугментациями	0,892	0,801	0,924	0,902	0,935	0,908

### Литература

1. Учим компьютер различать звуки: знакомство с конкурсом DCASE и сборка своего аудиоклассификатора за 30 минут [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/company/speechpro/blog/437818/> (дата обращения: 18.03.2019).
2. Мак-Каллок У.С., Питтс В. Логическое исчисление идей, относящихся к нервной активности Архивная копия от 27 ноября 2007 на Wayback Machine // Автоматы. – 1956. – С. 363–384. (Перевод английской статьи 1943 г.)
3. Lederle M., Wilhelm B. Combining high-level features of raw audio and spectrograms for audio tagging [Электронный ресурс]. – Режим доступа: [http://dcase.community/documents/challenge2018/technical\\_reports/DCASE2018\\_Wilhelm\\_109.pdf](http://dcase.community/documents/challenge2018/technical_reports/DCASE2018_Wilhelm_109.pdf) (дата обращения: 18.03.2019).
4. Keras: The Python Deep Learning library [Электронный ресурс]. – Режим доступа: <https://keras.io/> (дата обращения: 18.03.2019).



**Виноградова Таисия Борисовна**

Год рождения: 1995

Университет ИТМО, факультет информационных технологий и программирования, студент группы № М41212

Направление подготовки: 09.04.02 – Информационные системы и технологии

e-mail: taisiyavinogr@gmail.com

**Рыбин Сергей Витальевич**

Год рождения: 1959

Университет ИТМО, факультет информационных технологий и программирования, к.ф.-м.н.

e-mail: svrybin@corp.ifmo.ru

УДК 004.93

**СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ ОПРЕДЕЛЕНИЯ ТОНАЛЬНОСТИ  
КОРОТКИХ ТЕКСТОВ****Виноградова Т.Б.****Научный руководитель – к.ф.-м.н. Рыбин С.В.**

Работа выполнена в рамках темы НИР № 618278 «Синтез моциональной речи на основе генеративных связательных сетей».

Рассмотрена задача автоматического определения тональности текста методами машинного обучения. Проведен обзор различных методов решения задачи (классификация, ранговая регрессия), использованы различные векторные представления слов. Реализован наивный байесовский классификатор для оценки влияния различной предобработки данных на качество предсказания. Модели, определяющие тональность текста, в дальнейшем могут использоваться в системах синтеза эмоциональной речи.

**Ключевые слова:** анализ тональности, машинное обучение, обработка естественного языка, векторные представления слов, синтез речи.

**Введение.** Автоматический анализ эмоциональной составляющей (тональности) текстов играет важную роль при разработке систем синтеза эмоциональной речи. Тональная составляющая оказывает влияние на восприятие человеком синтезированной речи. Речь, синтезированная с учетом особенностей эмоциональной окраски текста, воспринимается более схожей с человеческой.

Начальный этап научно-исследовательской работы был направлен на сбор коллекции текстовых данных и анализ методов, подходящих для решения задачи. Дальнейшей целью являлась разработка и реализация алгоритма автоматического анализа тональности текста, учитывающего особенности рассматриваемой задачи (русский язык, небольшая длина текстов, опечатки).

**Постановка задачи.** Рассматривалась задача автоматического определения эмоциональной окраски (тональности) коротких неформальных текстов на русском языке. Ограничение на длину рассматриваемых текстов в данном случае носит весьма условный характер. Так как наиболее популярными коллекциями данных, предназначенными для разработки моделей анализа неформальных текстов, являются публичные сообщения из социальных сетей (например, Twitter) и отзывы пользователей интернет-сайтов о различных товарах, услугах, рецензии фильмов, которые обычно содержат до 25 слов, удобно рассматривать именно эту длину сообщения в качестве ограничения.

Чаще всего задача определения эмоциональной составляющей текста решается с помощью привлечения дополнительных данных, не являющихся текстовыми: аудиосигнала, соответствующего произнесенной фразе, видеозаписи говорящего, приложенного к тексту изображения, эмодиконов, знаков пунктуации и т.д. [1]. Тем не менее, наиболее интересной кажется задача обработки исключительно текстовой информации, очищенной от знаков пунктуации, эмодиконов и других дополнительных данных, которые часто используются при решении подобных задач (например, для отзывов о кинофильмах такими данными могут быть оценки фильмов по пятибалльной шкале).

Существует несколько различных вариантов постановки задачи анализа эмоциональной окраски. Среди постановок задачи, в которых пространство меток одномерно, можно выделить анализ полярности (классификация текстов на положительно и отрицательно эмоционально окрашенные) и анализ градации эмотивной составляющей в одномерном пространстве меток (определение не только полярности эмоциональной окраски, но и ее силы: от крайне негативной окраски до крайне позитивной). Более сложной задачей является определение эмоциональной окраски как принадлежности текста к одному из нескольких различных классов с метками «радость», «грусть», «отвращение», «страх» и другими типами меток-эмоций.

В данном исследовании рассмотрена задача сопоставления тексту одной из пяти меток: «крайне положительный», «положительный», «нейтральный», «отрицательный», «крайне отрицательный».

### **Различные подходы к решению задачи**

1. Алгоритмы на основе правил и методы машинного обучения. Существует два основных типа методов обработки текстов на естественном языке: методы, основанные на правилах (rule-based), и методы, использующие машинное обучение.

Методы определения тональности текста, основанные на правилах, используют словари эмотивной лексики, которые включают в себя слова и фразы, наличие которых в тексте может указывать на определенную эмоциональную окраску. Такие алгоритмы не требуют большого количества данных, достаточно выделить присущий рассматриваемой предметной области эмотивный лексикон. Такие подходы полностью детерминированы, результаты работы этих алгоритмов интерпретируемы в терминах рассматриваемой предметной области. Недостатком подобных методов является невозможность масштабировать полученные модели: в разных контекстах и при рассмотрении различных предметных областей одни и те же слова могут нести различную эмоциональную окраску.

Алгоритмы машинного обучения основаны на использовании статистических данных. Для их использования требуется большое количество размеченных данных, которые часто бывает трудно получить. Тем не менее, преимущество этих алгоритмов заключается в возможности построения более общих моделей с помощью использования техник обучения с переносом (transfer learning).

Задачу сопоставления тексту одной из пяти возможных меток с использованием алгоритмов машинного обучения можно решать двумя различными способами:

1. рассмотреть задачу классификации без учета порядка на множестве меток (от крайне негативной тональности до крайне позитивной);
2. решать задачу ранговой регрессии, т.е. рассматривать пространство меток как дискретное множество с заданным на нем отношением порядка.

Изучение влияния способа рассмотрения пространства меток на результат работы алгоритма автоматического анализа тональности являлось целью дальнейших исследований.

2. Искусственные нейронные сети. Среди алгоритмов машинного обучения особое место занимают искусственные нейронные сети. Они показывают передовые результаты в задачах анализа текстов большого объема и в некоторых случаях дают результаты распознавания, сравнимые по качеству с анализом, проведенным человеком [2]. В частности, высокие результаты при решении задач автоматического анализа текстов показывают сверточные (CNN) и рекуррентные (RNN) нейронные сети. Различные модификации рекуррентных нейронных сетей, такие как LSTM и GRU, помогают решить проблему затухания градиента при использовании алгоритма обратного распространения ошибки для обучения модели. Подобные методы позволяют контролировать количество информации, которая будет использоваться на каждом шаге обработки входной последовательности слов, регулируя с помощью весовых коэффициентов вклад разных частей входной последовательности в значение целевой функции.

Механизм внимания, используемый при анализе текстов, помогает выделить фрагменты входного текста, которые должны оказать наибольшее влияние на значение оптимизируемой функции. Хотя нейросетевые архитектуры с использованием этого механизма дают значительный прирост качества распознавания при анализе длинных текстов, для текстов небольшой длины разница между моделями, использующими и не использующими механизм внимания, может не быть существенной [3].

В неформальных текстах могут встречаться опечатки, ошибки и неточности. Существуют алгоритмы получения векторного представления текстовых данных, которые могут использоваться для обработки зашумленных текстов [4]. Подобные алгоритмы основаны на нейросетевых подходах и позволяют учитывать не только само слово, но и контекст, в котором оно употребляется. Еще одним преимуществом такого подхода является то, что при его использовании размерность полученных векторных представлений имеет порядок мощности алфавита. Представление данных в пространстве невысокой размерности позволяет ускорить обработку больших объемов данных.

Таким образом, для решения задачи автоматического определения эмоциональной окраски текста целесообразно использовать подходы на основе искусственных нейронных сетей. При работе с короткими текстами использование механизма внимания, вероятно, не улучшит результаты распознавания. Если в качестве набора данных для обучения модели выбраны неформальные тексты, можно решить проблему зашумленности данных уже на этапе получения векторных представлений.

### **Подготовка и проведение эксперимента**

1. Подготовка коллекции текстов для дальнейших экспериментов. На данном этапе проведения исследования практическая часть состояла в сборе текстовых данных и разметке его по шкале от крайне негативной окраски до крайне позитивной (всего 5 различных меток). Был собран набор данных, состоящий из 500 коротких фраз, содержащих опечатки, затем он был вручную размечен (по 100 фраз для каждой из 5 меток). Примеры фраз:

- крайне негативная: «как же вы надоели, уже идите делом займитесь»;
- негативная: «что как бы не очень меня устраивает»;
- нейтральная: «директора нет на месте, попозже наберите»;
- позитивная: «как-то получше чем предыдущие, вроде, но что-то не то»;
- крайне позитивная: «всем доволен полностью, отличная работа».

2. Оценка влияния предобработки на качество классификации. Для оценки влияния предобработки на работу методов машинного обучения при анализе коротких текстов на русском языке была реализована простейшая модель классификатора.

Была рассмотрена задача бинарной классификации текстов на негативно и позитивно окрашенные.

Был использован размеченный набор сообщений в Twitter на русском языке [5]. Сообщения из этой социальной сети хорошо отражают специфику рассматриваемой задачи: русский язык, неформальная лексика, короткие тексты, опечатки. Выборка состояла из 10 000 негативных и 10 000 позитивных сообщений.

Был использован наивный байесовский классификатор, использующий упрощенную формулу Байеса:

$$P(c|d) = \frac{P(c)P(d|c)}{P(d)} \approx P(c)P(d|c),$$

где  $P(d)$  – вероятность появления фразы  $d$  – была отброшена, так как принимается гипотеза о том, что она имеет равное значений для всех текстов;  $P(c)$  – вероятность появления фразы определенного класса  $c$ ;  $P(d|c)$  – вероятность принадлежности фразы в  $d$  классу  $c$ .

Было проведено удаление знаков препинания, спецсимволов, имен пользователей. Тексты были переведены в нижний регистр. Были получены копии набора текстов, из которых были удалены все стоп-слова. Под стоп-словами здесь понимаются предлоги, союзы, некоторые наречия, не несущие смысловой нагрузки, но статистически встречающиеся в текстах так часто, что это может существенно повлиять на результат предсказания [6]. К полученным копиям были применены различные варианты предобработки: стемминг и лемматизация.

Были получены различные векторные представления наборов данных с помощью следующих моделей:

- TF-IDF;
- W2V (SkipGram), обученная на имеющихся данных, с размерностью вектора 50;
- W2V (SkipGram), предобученная на национальном корпусе русского языка (НКРЯ), с размерностью вектора 300 и размером окна 5 слов (ruscorpora\_upos\_skipgram\_300\_5\_2018) [7].

Результаты сравнения полученных классификаторов по метрике ассурасу представлены в таблице.

Для предобученной модели W2V (Word2Vec) использовался лишь один метод предобработки, совпадающий с методом обработки данных, на которых была обучена используемая модель (лемматизация).

Таблица. Сравнение полученных классификаторов (ассурасу)

Метод предобработки	Векторные представления слов		
	TF-IDF	Word2Vec	Word2Vec (предобуч.)
Лемматизация, без стоп-слов	0,59	0,55	0,50
Лемматизация	0,62	0,57	0,50
Стемминг, без стоп-слов	0,59	0,56	–
Стемминг	0,60	0,57	–

Заметно, что удаление стоп-слов не приносит улучшения в качество работы алгоритма, а использование модели Word2Vec, предобученной на длинных текстах, показывает результаты хуже по сравнению с другими моделями получения векторных представлений данных.

**Заключение.** Методы предобработки данных влияют на качество работы модели анализа текста. При построении модели важно учитывать особенности анализируемых текстов (длина, уровень зашумленности, тематика). В связи с этими соображениями был собран набор данных для проведения дальнейших исследований. По результатам

сравнения методов предобработки был выбран алгоритм, дающий наилучшие результаты. Собранный набор данных был предобработан исходя из этих соображений (тексты переведены в нижний регистр, проведена лемматизация, удалены знаки препинания, получены векторные представления с использованием статистики TF-IDF). В дальнейшем планируется сравнение алгоритмов восстановления ранговой регрессии с алгоритмами классификации.

### Литература

1. Liu B., Zhang L. A survey of opinion mining and sentiment analysis // Mining Text. Springer U.S. – 2012. – P. 415–463.
2. Arkhipenko K. et al. Comparison of Neural Network Architectures for Sentiment Analysis of Russian Tweets // Proceedings of the International Conference «Dialogue 2016». – 2016. – V. 15. – P. 50–58.
3. Cliche M. BB\_twtr at SemEval-2017 Task 4: Twitter Sentiment Analysis with CNNs and LSTMs // Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017). – 2017. – P. 573–580.
4. Malykh V. Robust word vectors for Russian language // Proceedings of Artificial Intelligence and Natural Language AINL FRUCT 2016 Conference. – 2016. – P. 10–12.
5. Рубцова Ю.В. Построение корпуса текстов для настройки тонового классификатора // Программные продукты и системы. – 2015. – № 1(109). – С. 72–78.
6. Русские стоп слова / Antula: профессиональная студия веб-дизайна [Электронный ресурс]. – Режим доступа: [http://www.antula.ru/noise-word\\_2.htm](http://www.antula.ru/noise-word_2.htm) (дата обращения: 02.03.2019).
7. RusVectořēs: семантические модели для русского языка [Электронный ресурс]. – Режим доступа: <https://rusvectors.org/ru/models/> (дата обращения: 02.03.2019).

**Зено Бассель**

Год рождения: 1986

Университет ИТМО, факультет информационных технологий и программирования, аспирант

Направление подготовки: 09.06.01 – Информатика и вычислительная техника

e-mail: basilzeno@gmail.com

**Матвеев Юрий Николаевич**

Год рождения: 1955

Университет ИТМО, факультет информационных технологий и программирования, д.т.н.

e-mail matveev@speechpro.com

**UDK 004.89****FACE ROTATION USING GENERATIVE ADVERSARIAL NETWORKS**

**Bassel H. Zeno**

**Supervisor – D.Sc., Professor Yuri N. Matveev**

Работа выполнена в рамках темы НИР № 618278 «Синтез моциональной речи на основе генеративных состязательных сетей».

In this work we propose new generative adversarial framework for face rotation based on disentangled representation learning. It consists of five parts: identity encoder, pose encoder, discriminator and classifier. Each input face image is decoded into two embedding vectors, the identity vector and the pose vector. Then by concatenating the identity vector with another target pose vector, the generator synthesis new face image with target view. The model learns in unsupervised fashion on Multi-PIE dataset.

**Keywords:** generative Adversarial Network, convolutional neural networks, disentangled Representation.

The performance of face recognition systems has been significantly improved over the past few years, due to the development of graphics card technology, on the one hand, and more importantly, due to the significant evolution of convolutional neural networks [1, 2]. However, pose variations are still the bottleneck for many real-world face recognition scenarios.

Existing methods that address pose variations can be divided into two categories. One category tries to adopt hand-crafted or learned pose-invariant features [1, 2] while the other resorts to synthesis techniques to synthesize face images of a specific person. For example, TP-GAN [3] and FF-GAN [4] attempt to recover a frontal view image from any face image with large pose. DR-GAN [5] can change the pose of an input face image. However, these methods can only manipulate limited poses of a face image. In additions they also require full annotation of attributes to train the models. Some of these GAN-based methods [4–6] usually have a single-pathway design: an encoder-decoder network is followed by a discriminator network. While the others [3, 7] have two-pathways design. The encoder (E) maps input images into a latent space (Z), and then fed into the decoder (G) with a pose vector to generate novel views.

In practice, in order to learn a meaningful representation of the training dataset without additional constraints, many different GAN frameworks were proposed. They learn interpretable and meaningful latent representations in an unsupervised setting, such as InfoGAN [8], BiGAN [9], or in supervised setting [10]. Despite all the effort in this area, these approaches ignore one of the most fundamental principles of face image generation, which is the disentanglement of the face's identity and pose.

Learning disentangled identity and pose codes is a challenging task since 1) computers need to “imagine” what a given object would look like after a 3D rotation is applied; 2) the multi-view generations should preserve the same “identity”; and 3) the developed algorithms for solving this class of problems train GAN in a supervised fashion, or in its conditional setting.

Despite all the effort in this area, the field still lacks of a coherent framework for unsupervised disentangled pose and identity representations learning. From this consideration, in this article, we propose an end-to-end framework to learn a pair of disentangled identity and pose codes for a given dataset.

Our framework consists of a single generator, in contrast to [3, 7]. The generator has several upsampling, convolutional layers, and a set of residual blocks. The input of the generator is an identity code, while the desired pose code is fed into a multilayer perceptron (MLP) in order to generate parameters of AdaIn layers (Adaptive Instance Normalization) which they were injected in each residual block of generator.

GAN framework was introduced by Goodfellow et al. [11] for generative modeling of data through learning a transformation from points belonging to the simple prior distribution ( $z \sim P_z$ ) to those from the data distribution ( $x \sim P_{data}$ ). A typical GAN model consists of two modules that play an adversarial game or min-max game: a discriminator and a generator. While the generator learns to generate fake samples  $G(z)$  that are indistinguishable from real samples, the discriminator learns to distinguish between these fake samples ( $G(z) \sim P_G$ ) and real ones ( $x \sim P_{data}$ ). The goal of the generator is to fool the discriminator by generating photorealistic samples that resemble those from the real data while that of the discriminator is to accurately distinguish between real and generated samples. The two models, typically designed as neural networks, see Fig. 1, play a min max game with the objective function as shown in Equation:

$$\min_G \max_D V(D, G) = E_{x \sim P_{data}} [\log D(x)] + E_{z \sim P_z} [\log(1 - D(G(z)))]. \quad (1)$$

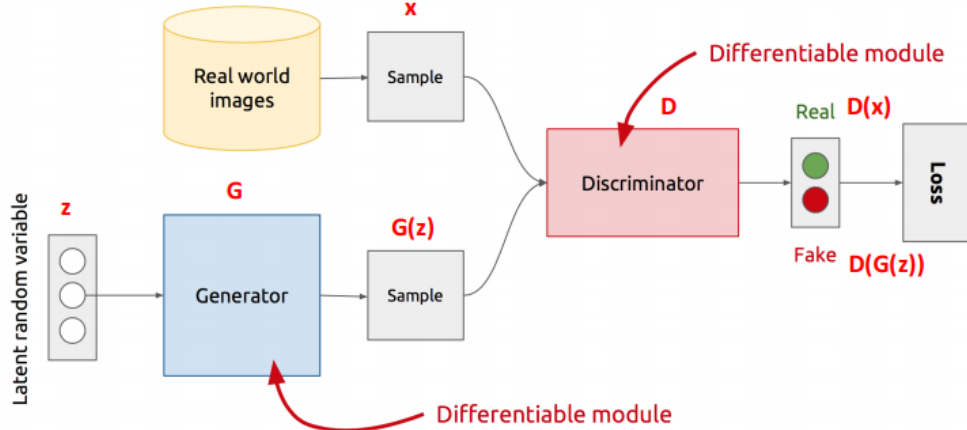


Fig. 1. Generative Adversarial Network (GAN) Architecture

We present our proposed GAN framework for face rotation. To generate a face image of any specific identity with an arbitrary target pose, our framework incorporates the pose information in the synthesis process by using encoder to calculate pose embedding vector.

Let  $x_1^s \in X$  be input image of a certain subject identity, and  $x_2^p \in X$  be another input image to extract the target pose features. Our goal is to synthesize a new face image  $x'$  of the subject of  $x_1^s$  with the extracted pose of  $x_2^p$ . To achieve this goal, we assume that each image  $x \in X$  is generated from an identity latent code  $s \in S$  and a pose latent code  $p \in P$ . In other words,  $x_1^s, x_2^p$  are generated by the pair  $(s_1, p_1)$  and the pair  $(s_2, p_2)$  respectively. As the result, the new face image  $x'$  is synthesized by the pair  $(s_1, p_2)$ .

The overall framework of our proposed GAN model is depicted in Fig. 2, which contains five parts:

1. the identity encoder network  $E_S$ ;
2. the head pose encoder network  $E_P$ ;
3. the generative network  $G$ ;
4. the identity classification network  $C$ ;

5. the discriminative network  $D$ .

The function of the network  $E_S$  is to extract the identity latent code  $s_1$  from the subject image  $x_1^s$ ,  $s_1 = E_S(x_1^s)$ . While the function of the network  $E_P$  is to extract the pose latent code  $p_2$  of the target pose image  $x_2^p$ ,  $p_2 = E_P(x_2^p)$ . We then concatenate these two vectors and feed them into the generator network  $G$  in order to produce the final output image  $x' = G(E_I(x_1^s), E_P(x_2^p))$ . The network  $C$  is used to preserve the identity by measuring the posterior probability  $P(c|x_1^s)$ , where  $c$  is the subject identity of  $x_1^s$ . The Networks  $C$  and  $E_S$  have shared parameters.

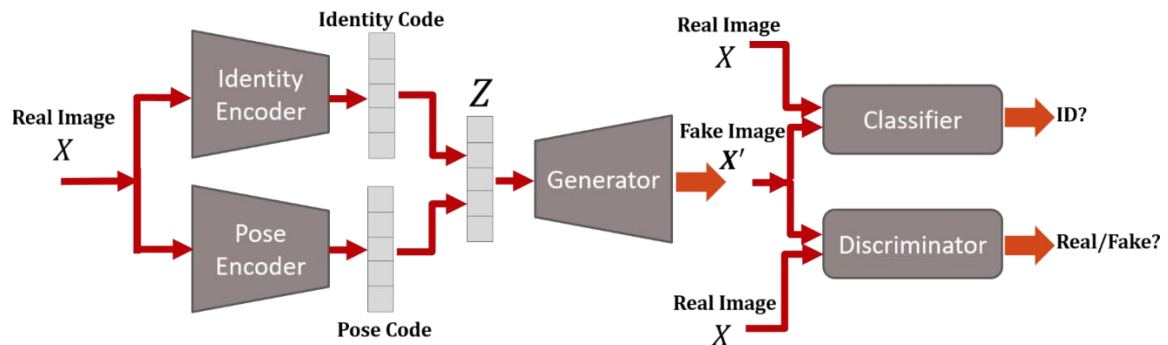


Fig. 2. Our proposed framework

In experiments we use Multi-PIE dataset. Multi-PIE is a labeled database [12] for evaluating face recognition under pose, illumination, and expression variations in controlled settings. It contains 337 identities and provides annotations of the head pose vertical rotation angles. Fig. 3 presents the face synthesis results of the identities that appear in the training dataset. Our method performs well in face synthesis, preserving both identity and pose.

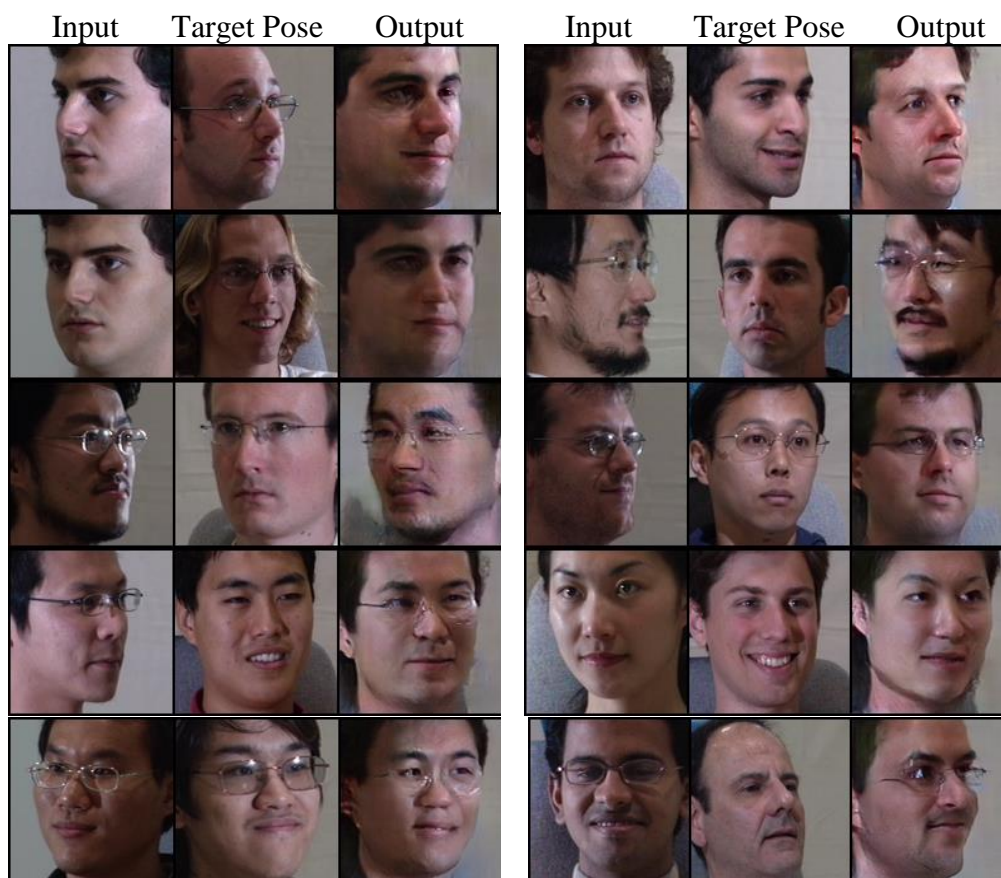


Fig. 3. Results of our GAN model



**References**

1. Chen D., Cao X., Wen F., Sun J. Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification // CVPR. – 2013.
2. Schroff F., Kalenichenko D., Philbin J. Facenet: A unified embedding for face recognition and clustering // CVPR. – 2015.
3. Huang R., Zhang S., Li T., He R. Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1704.04086> (дата обращения: 06.03.2019).
4. Yin X., Yu X., Sohn K., Liu X., Chandraker M. Towards large-pose face frontalization in the wild [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1704.06244> (дата обращения: 06.03.2019).
5. Tran L., Yin X., Liu X. Disentangled representation learning gan for pose-invariant face recognition // CVPR. – 2017. – V. 4. – P. 7.
6. Bao J., Chen D., Wen F., Li H., Hua G. Towards Open-Set Identity Preserving Face Synthesis [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1803.11182> (дата обращения: 06.03.2019).
7. Tian Y., Peng X., Zhao L., Zhang S., Metaxas D.N. CR-GAN: learning complete representations for multi-view generation [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1806.11191> (дата обращения: 06.03.2019).
8. Chen X., Duan Y., Houthoofd R., Schulman J., Sutskever I., Abbeel P. InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets // Advances in neural information processing systems. – 2016. – P. 2172–2180.
9. Donahue J., Krahenbuhl P., Darrell T. Adversarial feature learning [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1605.09782> (дата обращения: 06.03.2019).
10. Antipov G., Baccouche M., Dugelay J.L. Face aging with conditional generative adversarial networks // IEEE International Conference on Image Processing. – 2017. – P. 2089–2093.
11. Goodfellow I.J., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative adversarial nets // Advances in Neural Information Processing Systems. – 2014. – P. 2672–2680.
12. Gross R., Matthews I., Cohn J., Kanade T., Baker S. Multi-pie // Image Vision Computing. – 2010. – V. 28(5). – P. 807–813.

**Казиева Назым**

Год рождения: 1976

Университет ИТМО, факультет информационных технологий и программирования, аспирант

Направление подготовки: 09.06.01 – Информатика и вычислительная техника

e-mail: kaznaz@list.ru

**Кухарев Георгий Александрович**

Год рождения: 1941

Университет ИТМО, факультет информационных технологий и программирования, д.т.н., профессор

e-mail: gakukharev@etu.ru

**УДК 612.087.1****БИОМЕТРИЧЕСКИЕ ШТРИХ-КОДЫ ДЛЯ ВЗАИМОДЕЙСТВИЯ ЛЮДЕЙ  
С ПЛАТФОРМОЙ «ИНТЕРНЕТ ВЕЩЕЙ»****Казиева Н.****Научный руководитель – д.т.н., профессор Кухарев Г.А.**

Работа выполнена в рамках темы НИР № 617040 «Синтез эмоциональной речи на основе глубокого машинного обучения».

В работе обсуждалось использование биометрических штрих-кодов в платформе «Интернет вещей». Предложено использование для этих целей цифрового лицевого портрета, основанного на цветных QR-кодах, которые содержат идентификационную биометрическую и документальную информацию о человеке, участвующем в обменном процессе с платформой «Интернет вещей».

**Ключевые слова:** Интернет вещей, биометрический штриховой код, цветной QR-код, координаты антропометрических точек лица, лицевые биометрические технологии.

Сегодня создаются умные дома (транспорт, предприятия), разрабатываются «дорожные карты», реализуется мониторинг погоды, морей и океанов, состояний сельскохозяйственных полей и лесных угодий, здоровья животных и птиц на фермах, и другие проекты, соединенные в рамках системы Интернета вещей. Интернет вещей представляет вычислительные устройства, цифровые машины, механические машины, людей, животных, другие объекты физического мира, которые взаимодействуют между собой с помощью специального идентификатора [1]. Одним из участников этого взаимодействия и обменного процесса информацией в рамках Интернета вещей является человек.

При этом человек начинает процесс взаимодействия с платформой «Интернет вещей», а все дальнейшее взаимодействие между объектами внутри платформы (по реализации его задания) может происходить уже без его участия, но с его идентификационными параметрами самостоятельно! Поэтому главной проблемой в этом процессе взаимодействия и обмена информацией является безопасность передачи данных, в том числе личностных данных человека (фото, идентификатор, пороли).

В работе обсуждалось использование для целей взаимодействия и обмена – цифрового лицевого портрета. В качестве этого портрета выступал биометрический штриховой код, который содержит: лицевую биометрическую информацию (например, изображение лица – анфас/профиль, координаты антропометрических точек лица и другую биометрическую информацию о лице) и документальную информацию о человеке (ФИО, дата рождения, адрес, телефон и др.), которому принадлежит это лицо. И вся эта информация может быть открыта только в условиях разрешенного доступа со

стороны самого человека и платформы «Интернет вещей». Тогда, находясь на удалении (не выходя из дома, офиса, находясь в пути и др.), человек может безопасно взаимодействовать с платформой «Интернет вещей».

Попытки практического использования штрих-кодов с биометрической информацией ведутся уже более 20 лет [2, 3]. Но как же все-таки сам человек может быть «связан» со штрих-кодом, а не с номером-биркой на его ноге или руке [3]? Оригинальное представление об этой задаче сделано в статье [4], где авторы вводят понятие «биологического кода лица человека». Но, как оказалось, «биологический код лица человека» определяет всю генеральную популяцию людей, но не каждого отдельного в ней человека. Поэтому и практического решения идеи [4] не нашли, однако оставили значительный след.

В работе [5] был показан способ, который возник на развитии идей [4], но с конкретизацией представления для лица каждого отдельного человека – уникально представляя его фенотип в форме линейного штрих-кода в символиках EAN: EAN-8 и EAN-13. Однако решения, предложенные в [5], не получили дальнейшего развития ввиду ограниченности емкости символик кодов EAN и не возможности записи буквенной информации и специальных знаков. Минусом использования кодов EAN-8 и EAN-13 является также необходимость использования первых трех знаков штрих-кода как служебной информации (принадлежности к стране изготовителя товара) или, как свободного штрих-кода с «префиксом» 200–299.

Дальнейшее развитие идей [5] требовало использования штрих-кодов, поддерживающих не только запись цифр и буквенной информации, но и специальных знаков, причем емкость штрих-кодов должна была превышать емкость EAN-8 и EAN-13 не менее, чем на три порядка. Новые решения технологий штрихового кодирования для задач лицевой биометрии, учитывающие эти требования, были предложены в работе [6]. И эти решения были основаны на использовании цветных штриховых кодов: как линейных (Code 128), так и двумерных (QR-кодах). При этом цветные QR-коды фактически создают неограниченные возможности применения их в задачах биометрии и ее приложений.

Например, для двумерных штриховых кодов такое представление заключается в формировании нового цветного графического объекта на основе двух типов компонентов: черно-белых QR-кодов и изображений лиц – путем их размещения на местах слоев «R», «G» и «B» этого графического объекта, что по факту приводит к созданию цветного QR-кода, причем биометрическую и документальную информацию о изображении лица содержат QR-коды, входящие в цветной QR-код. Именно поэтому такие цветные QR-коды были названы как биометрические цветные QR-коды.

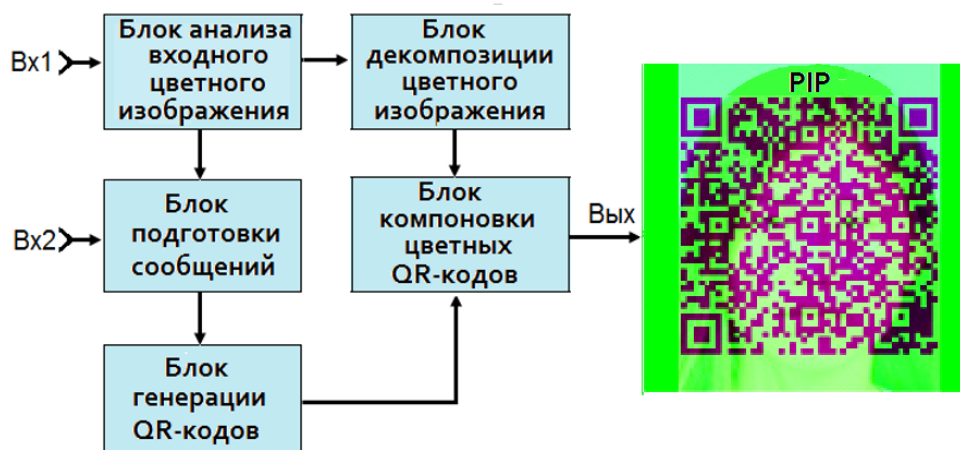


Рис. 1. Блок-схема устройства генерации цветных QR-кодов

Блок-схема устройства и пример цветного QR-кода, полученного по идеям [6], показаны на рис. 1. Содержимое этого QR-кода можно прочитать с помощью программы pdf417/версия 7.0.2. В этом примере слой «QR-код INFO» содержит следующую информацию: CUHK Face Sketch Database (CUFS), photo: f-039-01.jpg ### http://mmlab.ie.cuhk.edu.hk/archive/facesketch.html. При этом база CUFS является широко известной тестовой базой.

Вся остальная информация становится видимой и (или) считывается из отдельных слоев «BIO QR-code» при его декомпозиции.

На рис. 2 показаны два базовых QR-кода и раскрыто их содержание.



Рис. 2. Примеры цветных биометрических QR-кодов

Так, BIO QR-код (PIA) содержит фотоизображение лица человека (Photo), QR-код «INFO» с документальной информацией о человеке, которому принадлежит это лицо, и QR-код «ANTRO», содержащий координаты антропометрических точек лица. Этой информации достаточно, чтобы однозначно идентифицировать человека!

BIO QR-код (PIP) содержит две компоненты цветного изображения лица и QR-код «INFO». Наличие двух компонент изображения лица позволяет полностью реконструировать цветное изображение лица.

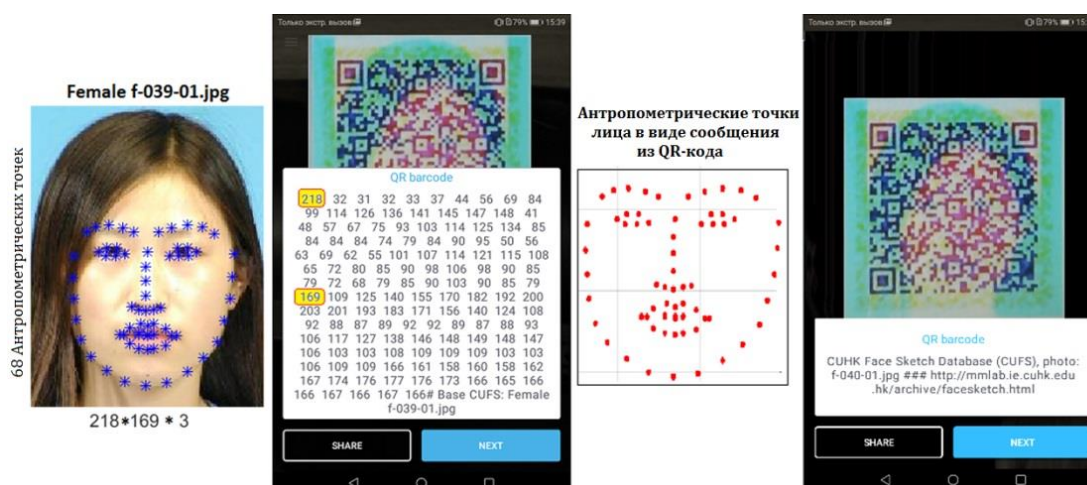


Рис. 3. Примеры чтения данных из цветных QR-кодов

На рис. 3 представлено изображение лица человека, его идентификатор (Female f-039-01), соответствующий номеру в тестовой базе CUFS. На лице показаны 68 антропометрических точек; а ниже показан размер изображения лица. Далее показаны результаты чтения QR-кода «ANTRO» и QR-кода «INFO» с помощью смартфона. Тем самым подчеркивается, что информация, заложенная в QR-коды, читается стандартными способами в случае, если к ней открыт доступ. Очевидно, что в рамках взаимодействия с платформой «Интернет вещей», будут использоваться общие (специальные) инструментальные средства чтения и декодирования этой информации, поскольку она может быть использована многими объектами платформы. В статье [6] также показаны способы шифрования биометрической и документальной информации в цветных QR-кодах, что не допускает ее чтения при несанкционированном доступе к ней.

В заключение можно отметить: на сегодняшний день уже стало ясным, что лицевые биометрические технологии способны обеспечить весь комплекс задач распознавания людей, совершающих доступ к ресурсам и обмен информацией с соответствующими системами, а технологии штрихового кодирования позволяют эффективно реализовать автоматизацию сопутствующих процессов: унификацию и каталогизацию документов и удостоверений личности, быструю проверку их аутентичности и безошибочный обмен данными о клиентах по универсальным стандартам и международным каналам и т.д.

Именно поэтому, интерес к проблеме представления биометрической информации в рамках штрихового кодирования, не ослабевает с момента появления первых проб по использованию штрихового кодирования в биометрии. Сопровождая нас везде, штрих-коды изменяют нашу активность так, что позволяют гораздо эффективнее вести деловой оборот, бизнес, мобильность, нашу повседневную и личную жизнь.

С помощью предлагаемых биометрических штриховых кодов мы можем не только защитить свои данные от стороннего доступа, но и более эффективно реализовать обменный процесс в системе Интернета вещей поскольку обменный процесс реализован в рамках стандартных штриховых кодов.

### Литература

1. Coetzee L., Eksteen J. The Internet of Things-promise for the future? An Introduction // IST-Africa Conference Proceedings. – 2011. – P. 1–9.
2. Soldek J. et al. Image Analysis and Pattern Recognition in Biometric Technologies // Proceedings International Conference on the Biometrics: Fraud Prevention, Enhanced Service. – 1997. – P. 270–286.
3. Heeter T.W. Method for verifying human identity during electronic sale transactions. Patent US № 5878155; опублик. 02.03.99.
4. Dakin S.C., Watt R.J. Biological «bar codes» in human faces // Journal of Vision. – 2009. – V. 9(4). – № 2. – P. 1–10.
5. Кухарев Г.А., Матвеев Ю.Н., Щеголева Н.Л. Формирование штрих-кода по изображениям лиц на основе градиентов яркости // Научно-технический вестник информационных технологий механики и оптики. – 2014. – № 90(3). – С. 89–96.
6. Кухарев Г.А., Казиева Н., Цымбал Д.А. Технологии штрихового кодирования для задач лицевой биометрии: современное состояние и новые решения // Научно-технический вестник информационных технологий, механики и оптики. – 2018. – Т. 18. – № 1. – С. 72–86.

**Калиев Арман**

Год рождения: 1990

Университет ИТМО, факультет информационных технологий и программирования, аспирант

Направление подготовки: 09.06.01 – Информатика и вычислительная техника

e-mail: akaliev@itmo.ru

**Рыбин Сергей Витальевич**

Год рождения: 1959

Университет ИТМО, факультет информационных технологий и программирования, к.ф.-м.н.

e-mail: svrybin@corp.ifmo.ru

**УДК 004.93****ПРОБЛЕМЫ ОБУЧЕНИЯ ГЕНЕРАТИВНЫХ СОСТЯЗАТЕЛЬНЫХ СЕТЕЙ  
В СИНТЕЗЕ РЕЧИ****Калиев А.****Научный руководитель – к.ф.-м.н. Рыбин С.В.**

Работа выполнена в рамках темы НИР № 618278 «Синтез моциональной речи на основе генеративных состязательных сетей».

Последние исследования применения генеративных состязательных сетей в технологии синтеза речи показали улучшение качества синтезируемой речи по сравнению с традиционными подходами. В работе рассмотрены проблемы обучения генеративных состязательных сетей, а также даны описания перспективных направлений улучшения конфигурации генеративных состязательных сетей. Полученные результаты могут быть в дальнейшем использованы в диалоговых системах и в сервисах обслуживания населения.

**Ключевые слова:** синтез речи, генеративных состязательных сети, нейронные сети.

Широкое применение нейронных сетей (НС) для синтеза речи позволило улучшить качество генерируемой речи. Особо сложные взаимосвязи лингвистических и акустических параметров тяжело поддаются человеческому описанию. Можно выделить несколько явных задач, где применение нейронных сетей сделали весомый прорыв в этой области. Первое, и в то же самое время самое несложное, это применение НС для решения проблемы предсказания длины фонемы. Второе, это применение для решения проблемы акустического моделирования речи каждой фонемы предложения или фразы. В обоих этих случаях НС сделали большой прорыв. В конечном счете улучшив качество синтезированной речи. Однако стоит отметить, что применение НС имеет свои серьезные недостатки. Среди них можно выделить следующие, как наиболее проблемные, это требование больших данных для обучения НС и требование к вычислительным мощностям. Решения этих двух проблем являются дорогостоящими в финансовом плане.

Также стоит отметить сложности описания программной реализации кода нейронных сетей. Как правило обучение нейронных сетей – плохо контролируемый процесс. При небольших данных обучения и тестирования, полученные результаты сильно меняются в зависимости от изначальных настроек НС и процесса обучения.

Сложная взаимосвязь лингвистических и акустических параметров крайне трудно раскрыть с помощью простых НС или простых рекуррентных НС. Параметрами акустической модели обычно служат слышимые акустические параметры, но несмотря на это, акустическая модель, построенная на этих параметрах, всегда показывает



результаты ниже, чем записи естественной речи. Возможно существует множество других скрытых параметров речи, которые влияют на человеческое восприятие.

Замена вокодера на НС предполагает улучшение качество синтезированной речи. Однако сложность создания такой модели заключается в оценочных функциях ошибки НС. Современные применяемые оценочные функций ошибки для обучения НС не способны качественно отделить шипящие звуки, как например «ш» или «ж», от простого шума. Как следствие, ученые вынуждены применять вокодеры, такие как WORLD [1] или STRAIGHT [2], для генерации речевого сигнала.

Самая распространенная функция ошибки НС для акустического моделирования и предсказания длины фонем основана на вычислении среднеквадратического отклонения. К сожалению такой подход несет явные сложности в генерации человеческой речи. Очевидным является проблема сглаженности генерируемого речевого сигнала. Функция среднеквадратичного отклонения способствует генерации сглаженного сигнала речи. Такая сглаженная речь легко определяется слушателями как неестественная.

Несмотря на это, самым перспективным направлением исследования применения НС для синтеза речи является поиск функции ошибки обучения НС для акустической модели и для вокодера.

Решить эту проблему призваны генеративные состязательные сети (ГСС). ГСС состоят из двух НС (рис. 1). Одна генерирует последовательность акустических параметров, вторая проверяет их на естественность распределения. Первая НС – это генератор, вторая – дискриминатор. Во время обучения дискриминатору поточно подаются два вектора акустических параметров. Первый получен из базы данных записанных речей, второй сгенерирован с помощью генератора. Дискриминатору ставится задача определить, какой из них естественный и какой из них фальшивый. В то же время генератору ставится задача генерировать распределение акустических параметров так, чтобы дискриминатору было трудно определить их фальшивость.

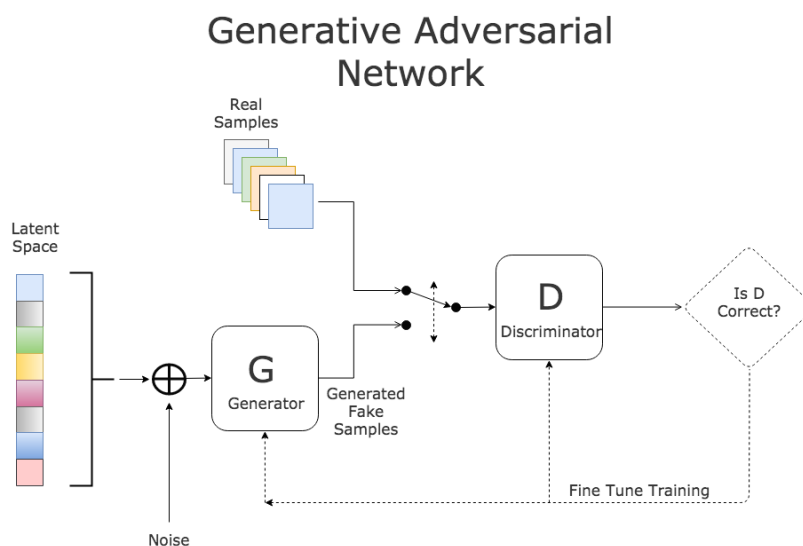


Рис. 1. Архитектура ГСС

Как показали последние исследования в области применения ГСС для синтеза речи, при такой конфигурации НС генератор не стремится генерировать естественную речь, а наоборот, генерировать такую речь, которая обошла бы дискриминатор. Дискриминатор как оказалось не способен точно определить естественную речь от неестественной.

Решением стало использовать рекуррентную НС как генератор, которая заранее была обучена с функций среднеквадратического отклонения [3]. Таким образом, над

полноценной ГСС проводилось только несколько итераций, которая улучшала только распределение акустических параметров. Такой способ обучение ГСС улучшает полноту речевого сигнала и минимизирует основную проблему сглаженности речи (рис. 2).

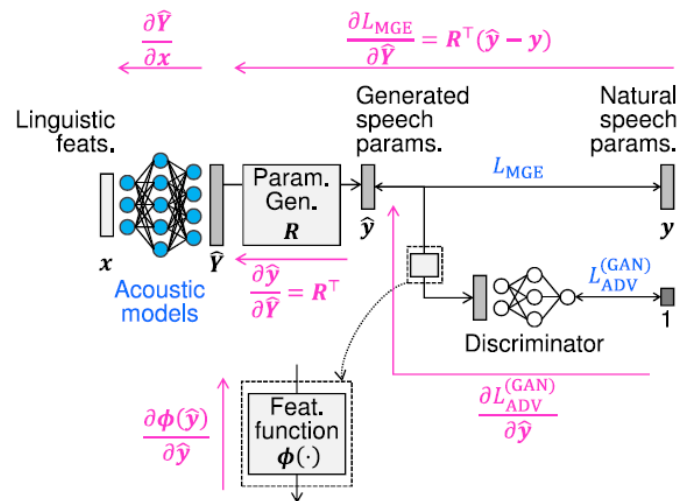


Рис. 2. Архитектура модифицированного ГСС

Но как видно ГСС не в полной мере используется в технологии синтеза речи. Возможности дальнейшей модификации ГСС дают надежды на дальнейшее улучшение качество синтеза речи. Одной из наиболее перспективных направлений – это обучить дискриминатор таким образом, чтобы он мог различать парные звуки, например, «ш» и «щ». В этом случае возможно использовать мел-кепстральные параметры.

Можно использовать аудиокодирование для преобразования речевого сигнала в параметры речи. Для этих целей можно использовать НС, такой кодировщик можно подключить к генератору. Так, после генерирования аудиосигнала генератором, кодировщик преобразует аудио в параметры речи, которые можно будет использовать для подачи дискриминатору. Так можно решить проблему опознавания парных звуков.

Можно отметить, что современные end-to-end методы генерирования речи, как Tacotron [4] и Wavenet [5] имеют похожую философию в своей архитектуре. И весьма предсказуемо, что следующее поколение НС для синтеза речи будут продолжать развивать данный подход.

## Литература

1. Morise M., Yokomori F., Ozawa K. WORLD: a vocoder-based high-quality speech synthesis system for real-time applications // IEICE transactions on information and systems. – 2016. – V. E99-D. – P. 1877–1884.
2. Kawahara H., Masuda-Katsuse I., de Cheveigne A. Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds // Speech Communication. – 1999. – V. 27. – P. 187–207.
3. Saito Y., Takamichi S., Saruwatari H. Statistical parametric speech synthesis incorporating generative adversarial networks // IEEE/ACM Transactions on Audio, Speech, and Language Processing. – 2017. – V. 26. – № 1. – P. 84–96.
4. Wang Y. et al. Tacotron: Towards end-to-end speech synthesis [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1703.10135.pdf> (дата обращения: 06.03.2019).
5. Van Den Oord A. et al. WaveNet: A generative model for raw audio [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1609.03499.pdf> (дата обращения: 06.03.2019).



**Куан Чонг Тхе**

Год рождения: 1984

Университет ИТМО, факультет информационных технологий и программирования, аспирант

Направление подготовки: 09.06.01 – Информатика и вычислительная техника

e-mail: quantrongthe@itmo.ru

**Столбов Михаил Борисович**

Год рождения: 1952

Университет ИТМО, факультет информационных технологий и программирования, к.т.н.

e-mail: stolbov@speechpro.com

**УДК 621.391.8****СРАВНЕНИЕ АЛГОРИТМОВ ВЫДЕЛЕНИЯ РЕЧЕВЫХ СИГНАЛОВ  
ДЛЯ ДВУХЭЛЕМЕНТНОЙ МИКРОФОННОЙ РЕШЕТКИ****Куан Ч.Т.****Научный руководитель – к.т.н. Столбов М.Б.**

Работа выполнена в рамках темы НИР № 618278 «Синтез моциональной речи на основе генеративных состязательных сетей».

Работа посвящена сравнению алгоритмов выделения речевых сигналов на фоне широкополосных когерентных шумов окружения для двухэлементных микрофонных решеток. Рассмотрены адаптивные алгоритмы обработки сигналов двухэлементных микрофонных решеток в частотной области: минимума дисперсии шума и подавления боковых лепестков. Подавление помехи на натуральных сигналах ограничивалось разбросом характеристик микрофонов и достигало значений 20–25 дБ.

**Ключевые слова:** подавление шума, выделение речевого сигнала, микрофонные решетки, алгоритмы MVDR, GSC.

**Введение.** Одноканальные алгоритмы выделения речевых сигналов в нестационарном шуме имеют ограниченную эффективность, вызванную проблемой оценки шума. В таких ситуациях качество речевых сигналов может быть улучшено с помощью микрофонных решеток. Двухэлементные микрофонные решетки (MP2) получили широкое применение благодаря своей простоте и возможности компактного размещения. Вопросы обработки сигналов MP2 рассмотрены во многих работах, например [1–4]. В работе [4] проведено сравнение различных алгоритмов выделения речевых сигналов в присутствии широкополосных когерентных помех и показана эквивалентность для MP2 алгоритма формирования пространственного нуля в направлении источника помехи, алгоритма Minimum Variance Distortionless Response (MVDR) и алгоритма Generalized Sidelobe Canceller (GSC). При этом направление источника помехи предполагалось заданным. Однако данное ограничение является нереалистичным в практических условиях.

В работе рассмотрена задача выделения речевых сигналов в присутствии широкополосного когерентного шума с использованием адаптивных алгоритмов обработки сигналов MP2, в условиях априорной неопределенности направления источника помехи. Работа посвящена сравнению адаптивных алгоритмов MVDR и GSC.

**Модель сигнала.** Рассмотрим сценарий когерентного целевого сигнала и когерентного шума, поступающих на MP2 с направлений, задаваемых углами  $\theta_s$ ,  $\theta_p$ ,

относительно оси, проходящей через микрофоны. В представлении кратковременного дискретного преобразования Фурье (ДПФ) сигнал  $S(f, k)$  акустического целевого источника с направления  $\theta_s$  и когерентный широкополосный шум  $V(f, k)$  с направления  $\theta_v(f)$  формирует на микрофонах вектор сигналов:

$$\mathbf{X}(f, k) = S(f, k)\mathbf{D}_s(f) + V(f, k)\mathbf{D}_v(f) = \mathbf{S}(f, k) + \mathbf{V}(f, k),$$

где  $f, k$  – индексы частоты и номера кадра;  $\mathbf{D}(f, \theta_x) = \mathbf{D}_x(f) = [e^{+j\Phi_x}, e^{-j\Phi_x}]^T$  – вектор фазовых сдвигов сигналов микрофонов относительно центральной точки между ними;  $()^T$  – символ транспонирования;  $\Phi_x(f)$  – фазовые сдвиги:

$$\Phi_x(f) = \pi d \cos(\theta_x) / \lambda = \pi f \tau_0 \cos(\theta_x),$$

где  $d$  – расстояние между микрофонами;  $c$  – скорость звука в воздухе;  $\tau_0 = d/c$  время прохождения звука по оси МР2 между микрофонами;  $\theta_x$  – угол прихода сигнала относительно оси МР2.

Схема обработки сигналов МР2 в частотной области представлена на рис. 1. МР2 состоит из двух ненаправленных микрофонов, разнесенных на расстояние  $d$ .

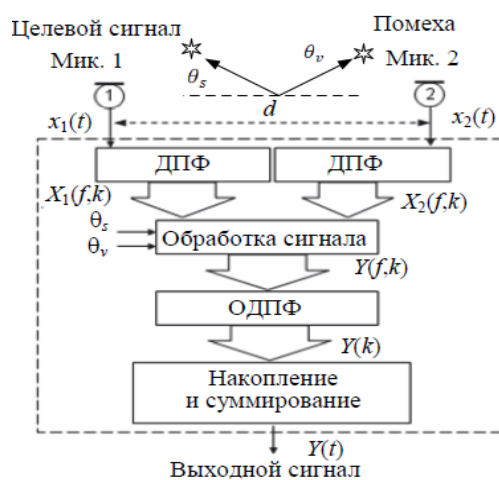


Рис. 1. Схема обработки сигналов МР2 в частотной области

Сигнал на выходе МР2 формируется взвешиванием входных сигналов комплексными весовыми коэффициентами:

$$Y(f, k) = \mathbf{W}^H(f) \mathbf{X}(f, k),$$

где  $\mathbf{W}(f)$  – вектор комплексных коэффициентов;  $()^H$  – символ сопряжения Эрмита;  $Y(f, k)$  – комплексные амплитуды спектра выходного сигнала на  $k$ -м кадре.

Вектор отсчетов сигнала  $\mathbf{Y}(k)$  на  $k$ -м кадре вычисляется с помощью обратного дискретного преобразования Фурье (ОДПФ, IDFT):

$$\mathbf{Y}(k) = IDFT\{Y(f, k)\}.$$

Последовательность векторов отсчетов  $\{\mathbf{Y}(k)\}$  преобразуется в выходную последовательность отсчетов сигнала на основе процедуры пересечения и суммирования (Overlap-and-Add, OLA):  $y(t) = OLA\{\mathbf{Y}(k)\}$ .

**Адаптивный алгоритм MVDR.** Алгоритм MVDR основывается на условии неискаженного приема сигнала с целевого направления  $\theta_s$  и минимизации мощности шума на выходе МР2. Оптимальные веса МР2 являются решением следующей оптимизационной задачи [1]:

$$\min_{\mathbf{w}} E\{|Y(f, k)|^2\} = \min_{\mathbf{w}} [\mathbf{W}^H(f) \mathbf{P}_{vv}(f) \mathbf{W}(f)]$$

при условии неискаженного приема сигнала с целевого направления:

$$\mathbf{W}^H(f) \mathbf{D}_s(f) = 1,$$

где  $\mathbf{P}_{vv}(f)$  – ковариационная матрица спектров шума на микрофонах;  $E\{\cdot\}$  – символ математического ожидания.

Ковариационная матрица спектров шума оценивается для каждой частоты:

$$\mathbf{P}_{vv}(f) = E\{\mathbf{V}(f, k)\mathbf{V}^H(f, k)\}.$$

Решение оптимизационной задачи приводит к следующему соотношению для вектора оптимальных весов [1]:

$$\mathbf{W}_0(f) = \frac{\mathbf{P}_{vv}^{-1}(f)\mathbf{D}_s(f)}{\mathbf{D}_s^H(f)\mathbf{P}_{vv}^{-1}(f)\mathbf{D}_s(f)}.$$

В адаптивном алгоритме MVDR ковариационная матрица и оптимальные веса оцениваются (в паузах целевого сигнала) по поступающим сигналам микрофонов:

$$\mathbf{P}_{vv}(f, k) = (1 - \beta_{MVDR})\mathbf{P}_{vv}(f, k - 1) + \beta_{MVDR}\mathbf{X}(f, k)\mathbf{X}^H(f, k),$$

$$\mathbf{W}_0(f, k) = \frac{\mathbf{P}_{vv}^{-1}(f, k)\mathbf{D}_s(f)}{\mathbf{D}_s^H(f)\mathbf{P}_{vv}^{-1}(f, k)\mathbf{D}_s(f)},$$

где  $\beta_{MVDR}$  – коэффициент сглаживания.

Сигнал на выходе вычисляется с использованием текущих значений коэффициентов:

$$Y(f, k) = \mathbf{W}_0^H(f, k)\mathbf{X}(f, k).$$

**Адаптивный алгоритм GSC.** Схема адаптивного алгоритма GSC для MP2 представлена на рис. 2.

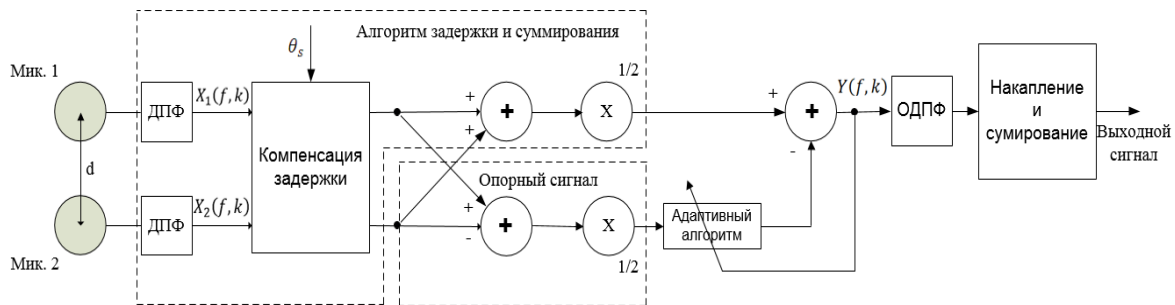


Рис. 2. Схема алгоритма GSC для MP2

Схема состоит из блоков компенсации задержки, блока формирования луча MP2, блока формирования опорного сигнала и адаптивного компенсатора помех.

Компенсация временных задержек (сдвига) сигналов микрофонов (выравнивание фаз сигналов, приходящих с направления  $\mathbf{D}_s$ ) выполняется с помощью соответствующего фазового сигналов микрофонов на каждой частоте:

$$\mathbf{X}_s(f, k) = \mathbf{D}_s^H(f) \cdot \mathbf{X}(f, k),$$

где  $\cdot$  – символ прямого умножения векторов.

Формирование луча MP2 в заданном направлении источника речевого сигнала осуществляется на основе алгоритма задержки и суммирования. Формирование опорного сигнала реализуется с помощью формирования нуля в направлении источника речевого сигнала. При этом формируется оценка опорного сигнала шума, а компоненты целевого речевого сигнала блокируются.

Основной и опорный сигналы  $Y_s(f, k), Y_r(f, k)$  вычисляются по следующим формулам:

$$Y_s(f, k) = \frac{X_{1s}(f, k) + X_{2s}(f, k)}{2}; \quad Y_r(f, k) = \frac{X_{1s}(f, k) - X_{2s}(f, k)}{2}.$$

Взвешенный опорный сигнал вычитается из основного, минимизируя мощность выходного сигнала на каждой частоте:

$$Y(f, k) = Y_s(f, k) - H(f, k) \times Y_r(f, k).$$

Минимизация мощности сигнала в каждом частотном бине  $f$ :  $J(f) = E\{|Y(f, k)|^2\}$ ,  $\partial J(f)/\partial H(f) = 0$  приводит к следующему оптимальному решению для коэффициентов компенсатора:

$$Ho(f, k) = \frac{P_{Y_s Y_r}(f, k)}{P_{Y_r Y_r}(f, k)}$$

В адаптивном алгоритме математическое ожидание аппроксимируется усреднением по времени:

$$P_{Y_s Y_r}(f, k) = (1 - \beta_{GSC}) \times P_{Y_s Y_r}(f, k - 1) + \beta_{GSC} \times Y_s(f, k) Y_r^*(f, k),$$

$$P_{Y_r Y_r}(f, k) = (1 - \beta_{GSC}) \times P_{Y_r Y_r}(f, k - 1) + \beta_{GSC} \times Y_r(f, k) Y_r^*(f, k),$$

где  $\beta_{GSC}$  – коэффициент сглаживания.

Выходной сигнал GSC вычисляется следующим образом:

$$Y(f, k) = Y_s(f, k) - Ho(f, k) \times Y_r(f, k).$$

**Эксперименты.** Целью экспериментов являлось сравнение эффективности подавления шума с помощью адаптивных алгоритмов GSC и MVDR. Кроме того, результаты обработки сравнивались с сигналом на выходе дифференциальной МР2 (Differential Microphone Array, DMA), ноль которой направлен в направлении источника помехи [5]. Экспериментальные сравнения алгоритмов проводились на модельных и натуральных сигналах.

В модельном эксперименте в качестве сигнала был взят гауссов белый шум с частотой дискретизации  $F_s = 16$  кГц, сдвинутый в каналах на два отсчета, что соответствовало его приходу на МР2 ( $d = 5$  см) приблизительно с угла  $\theta_v \approx 30^\circ$ . Речевой сигнал поступал по нормали ( $\theta_s = 90^\circ$ ). Алгоритмы GSC и MVDR автоматически компенсировали временной сдвиг помехи, алгоритм DMA обеспечивал полное подавление шума в случае точного задания временного сдвига сигналов и подавление 3 дБ, если сдвиг не был задан точно. При этом эквалайзер, компенсирующий в DMA искажение речевого сигнала, не применялся [5].

В натурном эксперименте алгоритмы исследовались на сигналах, записанных в безэховой камере на микрофоны, разнесенные на расстояние  $d = 5$  см с частотой  $F_s = 16$  кГц. Схема эксперимента показана на рис. 3.

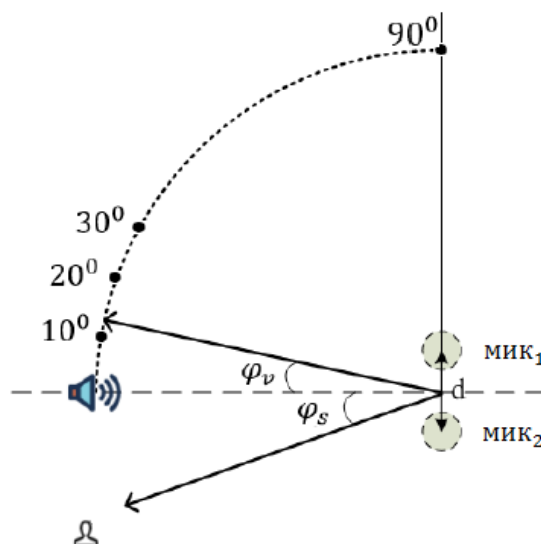


Рис. 3. Схема эксперимента в безэховой камере

Широкополосная помеха с акустической колонки поступала с направлений  $\varphi_v = 0^\circ - 90^\circ$  относительно нормали МР2 с шагом  $5^\circ$ . Каждое новое положение акустической колонки озвучивалось речью диктора, находившегося приблизительно в направлении  $\varphi_s = -30^\circ$  относительно нормали МР2. Обработка проводилась со следующими параметрами алгоритма: размер кадра 512 отсчетов, окно Хэмминга, пересечение 50%, коэффициенты сглаживания  $\beta_{GSC} = 0,5$ ;  $\beta_{MVDR} = 0,5$ . Результаты экспериментов приведены в таблице.

Таблица. Степень подавления шума в MP2

Алгоритмы	DMA	GSC	MVDR
Модельный эксперимент	$+\infty$ дБ при совпадении направлений нуля и источника помехи ... +3 дБ в случае несовпадения	51 дБ	39 дБ
Тест в камере	20 дБ при совпадении направлений нуля и источника помехи	24 дБ	25 дБ

Из эксперимента в безэховой камере можно сделать следующие выводы.

Адаптивные алгоритмы MVDR, GSC автоматически отслеживают направление прихода когерентного шума в отдельных частотных полосах и подавляют этот шум. Степень подавления когерентного шума в алгоритмах MVDR, GSC и DMA (в направлении нуля пространственного отклика) составляла в натуральных экспериментах значений 20–25 дБ и ограничивалась, по всей вероятности, различием характеристик микрофонов.

Формирование нуля пространственного отклика MP2 в направлении источника помехи сопровождается подавлением сигналов (в том числе целевого речевого сигнала), поступающих с других направлений (в наибольшей степени с направлений, близких к направлению помехи). В алгоритме MVDR подавление целевого сигнала компенсируется автоматически. В алгоритме GSC целевой сигнал не будет искажен только в случае, когда он не попадает в опорный канал (ноль опорного канала MP2 сориентирован точно в направлении его источника, и отсутствует реверберация). В алгоритме DMA целевой сигнал может быть сохранен с помощью эквалайзера, компенсирующего искажения при известном направлении прихода целевого сигнала.

Таким образом, алгоритм MVDR можно считать более гибким, а также робастным по отношению к неопределенности положения источников сигнала и помехи по сравнению с алгоритмами GSC и DMA.

**Заключение.** Адаптивные алгоритмы MVDR, GSC используют поступающие сигнала для получения информации о направлении прихода когерентной помехи и автоматического формирования пространственного поля диаграммы направленности в этом направлении. Степень подавления помехи при этом ограничивается статистическим разбросом оценок значений коэффициентов фильтров и разбросом характеристик микрофонов и достигает значений 20–25 дБ.

Полученные результаты могут быть использованы при проектировании систем с большим числом микрофонов.

### Литература

1. Brandstein M. and Ward D. (Eds.). *Microphone Arrays: Signal Processing Techniques and Applications*. – Springer-Verlag, 2001. – 398 p.
2. Benesty J., Chen J., Huang Y. *Microphone Array Signal Processing*. – Berlin, Germany: Springer-Verlag, 2008. – 250 p.
3. Benesty J., Chen J. *Study and Design of Differential Microphone Arrays*. – Springer topics in signal processing, 2013. – 184 p.
4. Столбов М.Б., Перельгин С.В. Алгоритмы двухэлементной микрофонной решетки для выделения речевых сигналов в присутствии когерентных помех // *Цифровая обработка сигналов*. – 2017. – № 4. – С. 34–39.
5. Столбов М.Б., Тхе К.Ч. Прием речевых сигналов в шумовой обстановке с использованием двухэлементных микрофонных решеток // *Научно-технический вестник информационных технологий, механики и оптики*. – 2018. – Т. 18. – № 5. – С. 850–857.

**Лизунова Инна Александровна**

Год рождения: 1996

Университет ИТМО, факультет информационных технологий и программирования, студент группы № М41212

Направление подготовки: 09.04.02 – Информационные системы и технологии

e-mail: inna.lizunova@gmail.com

**Махныткина Олеся Владимировна**

Год рождения: 1982

Университет ИТМО, факультет информационных технологий и программирования, к.т.н.

e-mail: makhnytina@corp.ifmo.ru

**УДК 004.9****ИСПОЛЬЗОВАНИЕ СЛОВАРЯ ЭМОТИКОНОВ ДЛЯ АНАЛИЗА  
ТОНАЛЬНОСТИ ТЕКСТОВ ИЗ СОЦИАЛЬНЫХ МЕДИА****Лизунова И.А.****Научный руководитель – к.т.н. Махныткина О.В.**

Работа выполнена в рамках темы НИР № 618278 «Синтез моциональной речи на основе генеративных связательных сетей».

Целью данной работы являлся сравнительный анализ методов оценки тональности текстов из социальных медиа с применением и без применения словаря эмотиконов. В работе проанализированы русскоязычные сообщения из социальной сети Twitter с использованием различных методов машинного обучения.

**Ключевые слова:** анализ тональности, эмотиконы, машинное обучение, наивный байесовский классификатор, логистическая регрессия, метод опорных векторов.

В настоящее время в социальных медиа получило широкое распространение использование специальных пиктограмм и идеограмм, называемых эмотиконы и эмодзи, для выражения различных эмоций, чувств, настроений. В связи с этим в области анализа тональности текстов большой интерес представляет исследование и разработка алгоритмов, способных обрабатывать и интерпретировать текстовую информацию, содержащую эмотиконы и эмодзи.

Эмотикон – пиктограмма, изображающая эмоцию, составленная из типографских знаков. В отличие от эмотикона, эмодзи – это идеограмма, представляющая собой не набор типографских символов, а готовое изображение с определенным смыслом. Однако некоторые эмодзи, изображающие лица, относят к эмотиконам [1].

В работе осуществлен сравнительный анализ методов, учитывающих и не учитывающих наличие эмотиконов в тексте. Для обработки был выбран корпус коротких текстов Юлии Рубцовой [2], сформированный на основе русскоязычных сообщений из социальной сети Twitter. Данный корпус содержит 114 991 позитивных сообщений и 111 923 негативных. В первую очередь необходимо было произвести анализ корпуса на предмет наличия в нем эмотиконов. Были выбраны наиболее распространенные эмотиконы и произведен их поиск в рассматриваемых текстах из социальной сети Twitter. Результаты анализа количества позитивных и негативных сообщений в данном корпусе представлены в табл. 1 и на рисунке.

В русскоязычном сообществе эмотиконы, выражающие радость и грусть, помимо принятых во всем мире обозначений, могут описываться только закрывающей скобкой (радость) или только открывающей скобкой (грусть). Такое представление распространено намного больше, чем стандартное, что можно увидеть в табл. 1.

Однако, это приводит к дополнительным трудностям при предварительной обработке данных, так как скобки также могут встречаться как выделители фрагментов текста.

Таблица 1. Количество различных эмодзи в позитивных и негативных сообщениях

Эмодзи	Эмоция или состояние	Количество в позитивных сообщениях	Количество в негативных сообщениях
:) :-) ;) ;-) =) x) X)	радость	36 827	0
)	радость	80 715	0
:( :-(: ;-( =( x( X(	грусть	0	47 822
(	грусть	0	67 153
:D :-D ;D ;-D =D xD XD	смех	198 388	149 343
:3 :-3 ;3 ;-3 =3	умиление	2 432	1 131
:* :-* ;* ;-* =* x* X*	поцелуй	4 499	131
:  :-  =	нейтральность	1	1 015
:O :-O =O :o :-o =o :O :-O =O	удивление	21 910	26 076

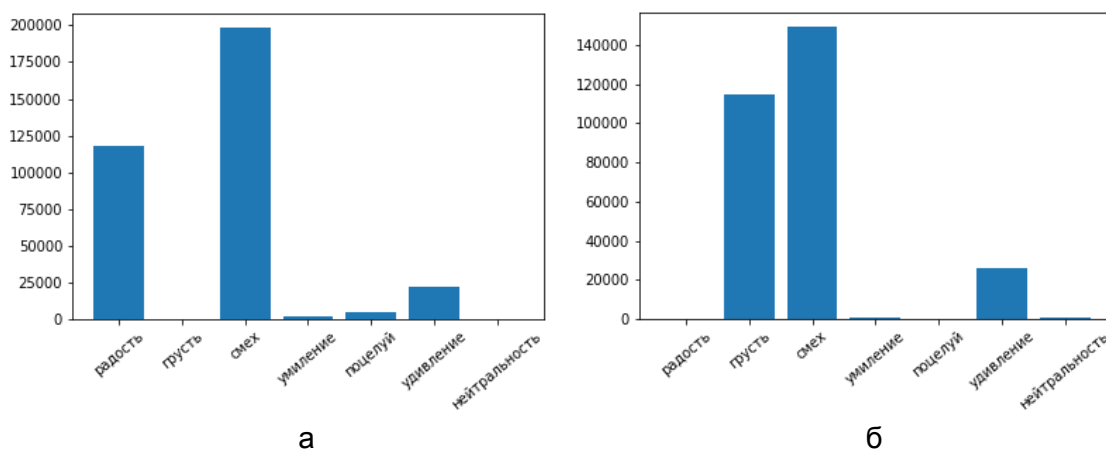


Рисунок. Гистограммы количества различных эмодзи: в позитивных сообщениях (а), в негативных сообщениях (б)

В представленном корпусе ни одно негативное сообщение не содержит эмодзи, выражающих радость, так же, как и ни одно позитивное сообщение не содержит эмодзи, выражающих грусть. Самым часто употребляемым эмодзи в русскоязычном сообществе социальной сети Twitter является эмодзи, обозначающий смех. Эмодзи, по общей принятой классификации считающийся нейтральным, почти всегда встречается в негативных сообщениях, поэтому можно сделать вывод, что он используется как выражение недовольства.

Результаты подсчета показывают, что только 298 из 114 991 позитивных сообщений (0,26%) не содержат эмодзи. Среди негативных 111 923 сообщений не содержат эмодзи 1317 (1,18%). Таким образом, большинство пользователей выражают свои эмоции посредством эмодзи, что является еще одним из поведенческих паттернов, которые целесообразно учитывать при обработке и анализе сообщений из социальных медиа [3].

Исходя из проведенного исследования корпуса можно сделать вывод, что эмодзи являются часто употребляемым способом выражения эмоций. Это подтверждает необходимость их дальнейшего изучения для задач анализа тональности текстов.

На основе проведенного анализа корпуса был составлен словарь наиболее часто используемых эмодзи. На основе словаря текстовые данные были преобразованы

таким образом, что каждый эмодикон был заменен на соответствующее его смыслу слово. Далее была произведена классификация представленного корпуса на основе первоначальных данных, и на основе данных, преобразованных с помощью построенного словаря.

На этапах предварительной обработки текста, построения векторного представления, подбора параметров классификационных моделей в обоих случаях были выполнены одинаковые действия.

В сообщениях из социальной сети Twitter содержится множество упоминаний имен других пользователей вида «@имя пользователя» и url-ссылок. В связи с этим предварительная обработка текста включает в себя следующие операции: замена всех url-ссылок на токен «URL», замена всех упоминаний других пользователей на токен «USER», замена символа «ё» на символ «е», приведение к нижнему регистру, удаление знаков пунктуации, удаление повторяющихся пробелов. В результате проведения операции по удалению знаков препинания вся информация о содержащихся в неизмененных данных эмодиконах удалась. Далее было получено векторное представление слов с помощью технологии Word2Vec [4]. С его помощью каждому слову из входного корпуса был сопоставлен вектор на основе контекстной близости слов. Для решения задачи классификации были рассмотрены следующие методы: наивный байесовский классификатор (NB), логистическая регрессия (LR), метод опорных векторов (SVM) [5].

В качестве оценок эффективности работы алгоритмов используется параметр  $F$ -мера ( $f1$ -score), зависящий от точности (precision) и полноты (recall), которые вычисляются по следующим формулам

$$Precision = \frac{TP}{TP + FP},$$

$$Recall = \frac{TP}{TP + FN},$$

где  $TP$  – истинно-положительное решение;  $TN$  – истинно-отрицательное решение;  $FP$  – ложноположительное решение;  $FN$  – ложноотрицательное решение.  $F$ -мера является гармоническим средним между точностью и полнотой и вычисляется по следующей формуле

$$F_1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}.$$

Оценки, полученные в результате проведения классификации с помощью различных методов на данных, содержащих и не содержащих эмодиконы, представлены в табл. 2.

Таблица 2. Оценки работы различных классификаторов

Метод	Без выделения эмодиконов	С выделением эмодиконов
NB	0,6238	0,8057
LR	0,7278	0,9615
SVM	0,7572	0,9616
CNN	0,7795	0,9716

Таким образом, точность классификации с применением словаря эмодиконов улучшилась в среднем на 20%. Учитывая наличие эмодиконов в 99,3% сообщений рассматриваемого корпуса, можно сделать вывод о том, что анализ тональности без учета данных средств выражения эмоций является неполным. Поэтому исследование и применение методов обработки текстов с использованием эмодиконов является перспективной областью в сфере анализа эмоциональной окраски текстов.



**Литература**

1. Wegrzyn-Wolska K., Bougueroua L., Yu H., Zhong J. Explore the Effects of Emoticons on Twitter Sentiment Analysis // CS & IT-CSCP 2016. – 2016. – P. 65–77.
2. Рубцова Ю.В. Построение корпуса текстов для настройки тонового классификатора // Программные продукты и системы. – 2015. – № 1(109). – С. 72–78.
3. Мальцева А.В., Махныткина О.В., Шилкина Н.Е. Изучение поведенческих паттернов пользователей социальных сетей: возможности Big Data // Жизнь исследования после исследования: как сделать результаты понятными и полезными. – 2016. – С. 988–991.
4. Mikolov T., Chen K., Corrado G., Dean J. Efficient Estimation of Word Representations in Vector Space [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1301.3781.pdf> (дата обращения: 06.03.2019).
5. Воронцов К.В. Лекции по линейным алгоритмам классификации [Электронный ресурс]. – Режим доступа: <http://www.machinelearning.ru/wiki/images/6/68/voron-ML-Lin.pdf> (дата обращения: 06.03.2019).

**Мамаев Никита Константинович**

Год рождения: 1996

Университет ИТМО, факультет информационных технологий и программирования, студент группы № М41211

Направление подготовки: 04.09.02 – Информационные системы и технологии

e-mail: nikita.mamaev1@gmail.com

**Черных Ирина Александровна**

Год рождения: 1981

ООО «ЦРТ-инновации», научно-исследовательский департамент, научный сотрудник

e-mail: chernykh-i@speechpro.com

**УДК 004.89****ДИАЛОГОВАЯ СИСТЕМА НА ОСНОВЕ ПОИСКА  
С ЯЗЫКОНЕЗАВИСИМЫМИ ВЕКТОРНЫМИ ПРЕДСТАВЛЕНИЯМИ СЛОВ****Мамаев Н.К.** (Университет ИТМО)**Научный руководитель – Черных И.А.** (ООО «ЦРТ-инновации»)

Работа выполнена в рамках темы НИР № 618278 «Синтез моциональной речи на основе генеративных связательных сетей».

В работе приведен анализ эффективности использования языконезависимых векторных представлений слов для инициализации диалоговой системы на основе поиска (retrieval-based).

**Ключевые слова:** диалоговая система, обработка естественного языка, языконезависимые векторные представления слов.

На сегодняшний день подраздел науки о данных, изучающий обработку естественного языка, является одним из самых популярных. В частности, особое внимание уделяется диалоговым системам. Такие лидеры в сфере обработки данных, как Apple, Google и Facebook активно инвестируют средства и открывают команды для таких исследований, чтобы предоставлять максимально эффективные сервисы по автоматической обработке естественного языка как для пользователей, так и для корпоративных партнеров.

Из подходов к разработке диалоговых систем наиболее часто применяется так называемый end-to-end подход, часто подразумевающий использование нейросетевых архитектур. Наиболее подходящими для выделения признаков из текста архитектурами на данный момент считается рекуррентная нейронная сеть с LSTM-ячейками, а также сверточная нейронная сеть.

Также ученые, специализирующиеся в обработке естественного языка, сегодня пристально следят за достижениями по направлению cross-language. Данный подход позволяет переносить знания с одного языка на другой: например, компенсировать недостаток данных на языке решения задачи использованием иноязычного корпуса. Это позволяет применять глубокое обучение моделей на низкоресурсных языках (французский, финский и т.д.).

Одним из примеров применения рекуррентной нейронной сети к построению диалоговой системы является решение Dual Encoder [1] – модель диалоговой системы, основанная на извлечении наиболее подходящего ответа или топ-N наиболее подходящих ответов из базы – иначе говоря, основанная на поиске (retrieval-based). Эта модель обучается на тройках «контекст-ответ-двоичный флаг соответствия», где под контекстом понимается предшествующая часть диалога. Во время прямого прохождения архитектуры контекст и ответ «кодируются», т.е. пересчитываются их

векторные представления. Затем вектор контекста умножается на матрицу  $M$ , и результирующий вектор представляет собой некоторый «сырой» вариант ответа (рис. 1). Этот сгенерированный ответ затем умножается на поданный в качестве обучающего образца ответ, после чего применяется сигмоида для получения вероятности того, насколько данный ответ «подходит» к контексту. Архитектура показывает наилучшие результаты при использовании LSTM-ячейки в качестве «энкодера», извлекающего признаки из текста. Варьируемыми параметрами для этой архитектуры являются методы инициализации матриц и векторных представлений слов, размер скрытой ячейки, а также оптимизационные параметры.

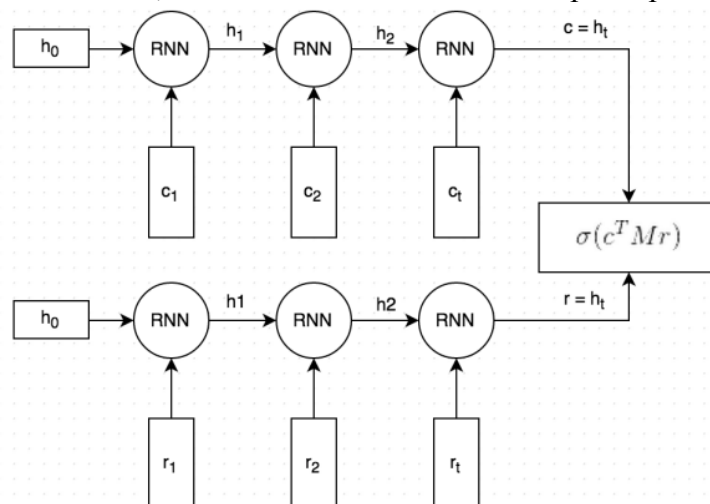


Рис. 1. Архитектура Dual Encoder

В качестве одного из промышленных применений этой модели можно назвать использование в качестве «суфлера» или «помощника оператора». При таком использовании модель предлагает оператору контактного центра или службы поддержки пользователей несколько релевантных вариантов ответа (подсказок) на вопрос, заданный клиентом.

Векторные представления слов в модели Dual Encoder могут быть инициализированы разным способом – либо случайно, используя некоторое вероятностное распределение, либо используя заранее подготовленные векторные представления.

Для построения векторных представлений слов на данный момент существует несколько наиболее часто применяющихся методов: контекстнезависимый TF-IDF, а также контекстозависимые GloVe, Word2Vec, fastText и BERT. В связи с развитием подхода cross-language для повышения обобщающей способности векторных представлений стали применяться методы переноса знаний с одного языка на другой. Эта парадигма нашла реализацию в методе, описанном в работе [2], который позволяет получить языконезависимые векторные представления для любой комбинации языков. В частности, для обучения подобных представлений применяется состязательное (adversarial) обучение с языковым дискриминатором, затем производится поиск прокрустово преобразования (Procrustes transform) (рис. 2), позволяющего сопоставить в векторном пространстве представления пар аналогичных слов на разных языках.

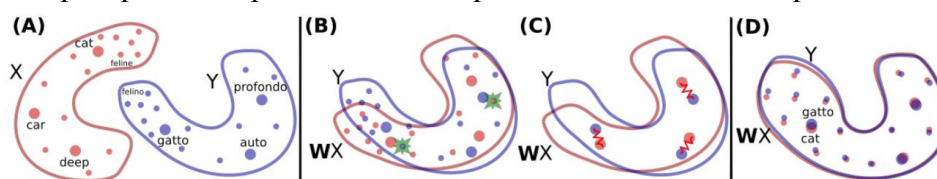


Рис. 2. Иллюстрация метода, описанного в [2]

Опираясь на вышеупомянутые тренды и результаты, авторы провели эксперименты на русскоязычном корпусе, содержащем диалоги службы поддержки одного из крупных мобильных операторов. Корпус содержит 83 тыс. пар «контекст-ответ» и был разбит на train/test/valid подкорпусы в соотношении 3:1:1 соответственно. Для инициализации векторных представлений слов использовали равномерное распределение на отрезке  $[-0,25, 0,25]$  в первом случае, и векторные представления, построенные методом, описанном в работе [2], на паре языков русский-английский во втором случае. В частности, исходные векторные представления во втором случае были построены с помощью метода fastText на веб-данных Common Crawl и Wikipedia на соответствующих языках.

Для оценки точности выбора ответов используется метрика 1 in N Recall@k – отношение количества случаев нахождения правильного ответа среди выбранных  $k$  лучших к общему количеству попыток поиска, при наличии  $N$  вариантов и одного правильного среди них.

Полученные результаты (таблица) демонстрируют, что точность при использовании cross-language векторных представлений слов, полученных по методу [2], в целом возрастает. Таким образом, этот метод представляется оптимальным для применения при инициализации векторных представлений слов в случае, когда размер корпуса недостаточно велик, чтобы обучить векторные представления слов только на нем [3].

Таблица. Результаты эксперимента

	Способ инициализации векторных представлений	
	случайные	cross-language (по [2])
1 in 10, Recall@1	0,44	0,49
1 in 10, Recall@2	0,66	0,7
1 in 10, Recall@5	0,94	0,93

### Литература

1. Lowe R. et al. The Ubuntu Dialogue Corpus: A Large Dataset for Research in Unstructured Multi-Turn Dialogue Systems [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1506.08909.pdf> (дата обращения: 06.03.2019).
2. Conneau A. et al. Word Translation Without Parallel Data [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1710.04087.pdf> (дата обращения: 06.03.2019).
3. Wojanowski et al. Enriching Word Vectors with Subword Information [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1607.04606.pdf> (дата обращения: 06.03.2019).

**Мироненко Александр Алексеевич**

Год рождения: 1995

Университет ИТМО, факультет информационных технологий и программирования, студент группы № М4221

Направление подготовки: 09.04.02 – Речевые информационные системы

e-mail: hexkritor@yandex.com

**Матвеев Юрий Николаевич**

Год рождения: 1955

Университет ИТМО, факультет информационных технологий и программирования, д.т.н.

e-mail: matveev@mail.ifmo.ru

**УДК 004.93**

**МЕТОД ПОВЫШЕНИЯ ТОЧНОСТИ ТРЕКИНГА ЛИЦ ЗА СЧЕТ  
КОМПЕНСАЦИИ ОШИБОК НАЗНАЧЕНИЯ ТРЕКА**

**Мироненко А.А.**

**Научный руководитель – д.т.н. Матвеев Ю.Н.**

В работе описан метод повышения точности трекинга за счет компенсации ошибок назначения трека. Технологии трекинга лиц могут быть использованы во многих областях современной человеческой деятельности. Несмотря на большое количество готовых решений в этой области, их исследование по-прежнему представляет интерес, так как есть простор для их усовершенствования или появления новых.

**Ключевые слова:** трекинг лиц, метод Виолы–Джонса, OpenCV, трекинг.

Цель работы – разработать метод повышения точности трекинга за счет компенсации ошибок назначения трека, в частности исправление ошибок назначения трека при пересечении лиц или долгого отсутствия лица в видеопотоке [1].

Целью данной работы являлась реализация метода повышения точности трекинга лиц за счет уменьшения вероятности ошибок назначения трека при пересечении лиц или долгого отсутствия лица в видеопотоке.

Из [2] можно выделить основные виды ошибок трекинга:

- ошибка неудачного детектирования объекта;
- ошибка долгого отсутствия объекта;
- ошибка при пересечении нескольких треков объектов.

Все эти ошибки влияют на точность работы алгоритма, которую можно оценить при помощи различных метрик. Основными метриками измерения точности работы являются МОТА [3], оценивающая частоту появления ошибок при трекинге, и MOTP [3], оценивающая отклонение позиций трекера с истинным значением. Для точной оценки эффективности трекинга для данной задачи используется метрика IDS (ID Switches) [2], показывающая количество изменений идентификатора трека относительно истинного значения.

Для решения описанной проблемы предложен подход на основе компенсации ошибок назначения трека при пересечении лиц или долгого отсутствия лица в видеопотоке. Он заключается в добавлении дополнительного блока в систему трекинга, в котором находятся ошибки трекинга и компенсируются путем проверки наличия трека, привязанному к определенному лицу.

**Используемые признаки.** В качестве используемых признаков были выбраны гистограммы яркости из-за простоты реализации, высокой производительности и достаточной эффективностью для работы метода.

Гистограммы строятся по монохромному изображению, полученному в результате работы детектора лиц. Количество интервалов в гистограмме – 128. Данное количество интервалов было выбрано по двум причинам.

Во-первых, в монохромном изображении значения яркости имеют 256 различных значений, что кратно 128, позволяя уменьшить количество признаков.

Во-вторых, на изображении, по которому высчитывается гистограмма, возможно появление помех в виде частей объектов, не относящихся к лицу (окружение, другое лицо и пр.), что приводит к ухудшению качества извлекаемых признаков, но уменьшив количество интервалов гистограммы до 128, уменьшается влияние помех, при этом извлеченные признаки остаются различимыми.

Однако полученные признаки на данном этапе являются непригодными для сравнения, так как они могут рассчитываться из изображений различного размера, а значениями в рассчитанных гистограммах является количество пикселей, имеющих яркость в определенном диапазоне, размер которого зависит от количества интервалов гистограммы. Для унификации этих признаков вместо количества пикселей используется вероятность появления пикселя с яркостью в определенном диапазоне.

Сами признаки являются адаптируемыми, так как при обновлении они высчитываются по формуле:

$$\mathbf{f} = f_o m + f_n(1 - m),$$

где  $\mathbf{f}$  – вектор значений обновленного признака;  $f_o$  – вектор значений старого признака;  $f_n$  – вектор значений нового признака;  $m$  – коэффициент памяти.

Сравнение признаков происходит следующим образом. Пусть  $\mathbf{a}$  и  $\mathbf{b}$  – вектора признаков, которые требуется сравнить между собой. Тогда разность признаков  $d$  будет равна:

$$d = \sum_{i=1}^{128} (a_i - b_i)^2.$$

Если  $d > \tau$ , где  $\tau$  – пороговый коэффициент схожести признаков, то сравниваемые лица различны. В остальных случаях лица с признаками  $a$  и  $b$  являются идентичными. Применение данной формулы обусловлено тем, что при большей разнице между  $a_i$  и  $b_i$ , разность  $d$  быстрее будет стремиться к  $\tau$ , тем самым уменьшая риск сделать два различных лица эквивалентными.

**Описание работы метода.** Работа метода (рис. 1) начинается при получении от детектора информации о появлении потенциально нового лица на кадре.

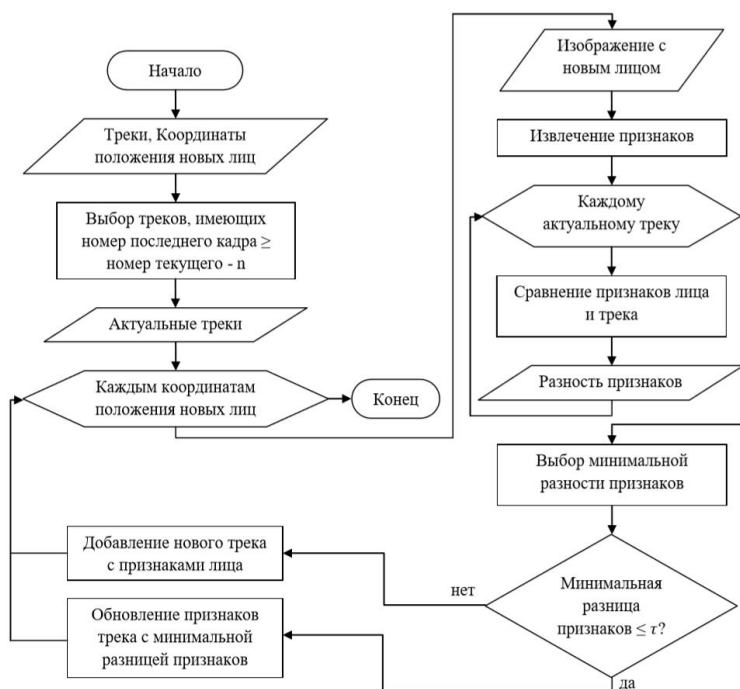


Рис. 1. Алгоритм работы метода компенсации ошибок назначения трека

По новому лицу рассчитываются признаки, и сравниваются с треками, имеющими информацию о лицах в промежутке  $n$  кадров от текущего момента. Это позволяет ограничить использование старых данных, которые могут быть на тот момент неактуальными. При сравнении признаков трека и нового лица выбирается такая пара признаков, что имеет минимальную разность признаков, а также не превышает пороговое значение. Если такая пара существует, то признаки трека обновляются, используя признаки нового лица. В противном случае, создается новый трек для найденного детектором лица, и информация о создании трека передается трекеру, который будет его в дальнейшем отслеживать.

**Тестирование метода.** Для тестирования работы метода компенсации ошибок реализована система трекинга (рис. 2), использующая готовые решения, реализованные в библиотеке OpenCV [4]. В качестве детектора используется метод Виолы–Джонса [5], а в качестве трекеров используются реализации на основе алгоритмов AdaBoost, MIL, MOSSE, TLD, Median Flow, GOTURN, а также фильтров KCF, CSRT.

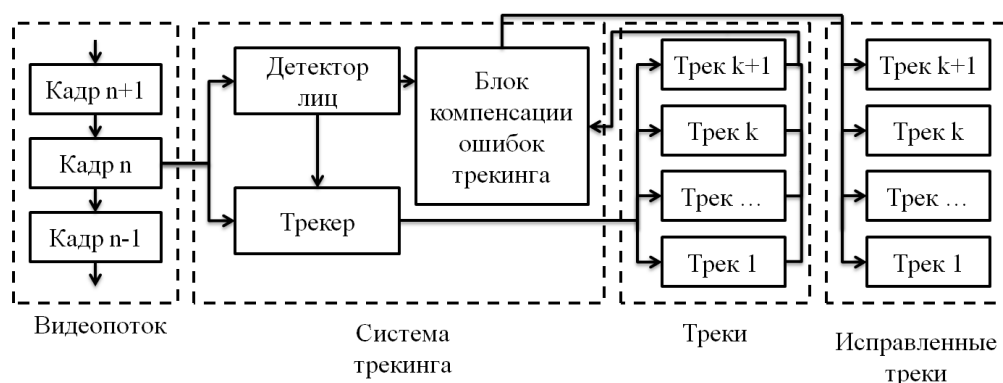


Рис. 2. Структура системы трекинга с блоком компенсации ошибок трекинга

Для оценки работы алгоритмов использована видеозапись, являющаяся частью базы Elliptical Head Tracking [6], разрешением  $128 \times 96$  пикселей и длительностью 101 кадр. На данной видеозаписи присутствуют фрагменты, при которых возникает отсутствие лица в кадре, путем поворота головы человека, а также пересечением лица другим человеком. Камера, использованная для видеозаписи, находится в движении, что усложняет условия работы трекеров. Для подсчета точности работы использованы метрики: MOTP [3], оценивающая частоту появления ошибок при трекинге, и MOTR [3], оценивающая отклонение позиций трекера с истинным значением. Для точной оценки эффективности трекинга для данной задачи используется метрика IDS (ID Switches) [2], база была размечена вручную, что потребовалось в силу отсутствия данных для трекинга лица. Для тестирования работы алгоритмов реализована система трекинга (таблица), использующая готовые решения, реализованные в библиотеке OpenCV [4] (рис. 3). В качестве детектора используется метод Виолы–Джонса [5], а в качестве трекеров используются реализации на основе алгоритмов AdaBoost, MIL, MOSSE, TLD, Median Flow, а также фильтров KCF, CSRT.

Таблица. Результаты работы системы трекинга

Трекер	MOTP, %		MOTA, %		ID Switches		Количество треков		Произошло слияние треков
	(a)	(б)	(a)	(б)	(a)	(б)	(a)	(б)	
AdaBoosting	31,67	31,67	-36,63	-19,8	3	2	6	3	Да
MIL	45,83	43,33	-12,87	-3,96	2	2	5	3	Да

Трекер	МОТР, %		МОТА, %		ID Switches		Количество треков		Произошло слияние треков
	(а)	(б)	(а)	(б)	(а)	(б)	(а)	(б)	
KCF	25,83	25,83	-11,88	-11,88	4	3	10	6	Нет
TLD	23,33	22,5	-50,5	-39,6	3	2	6	4	Нет
Median Flow	40,83	40,83	-21,78	-18,81	2	2	6	5	Нет
MOSSE	18,33	18,33	-3,96	-1,98	8	3	14	6	Нет
CSRT	42,5	40,83	-21,78	-23,76	3	2	6	5	Нет

(а) – результат работы системы трекинга без использования метода;

(б) – результат работы системы трекинга с использованием метода (пороговый коэффициент схожести признаков – 0,025).

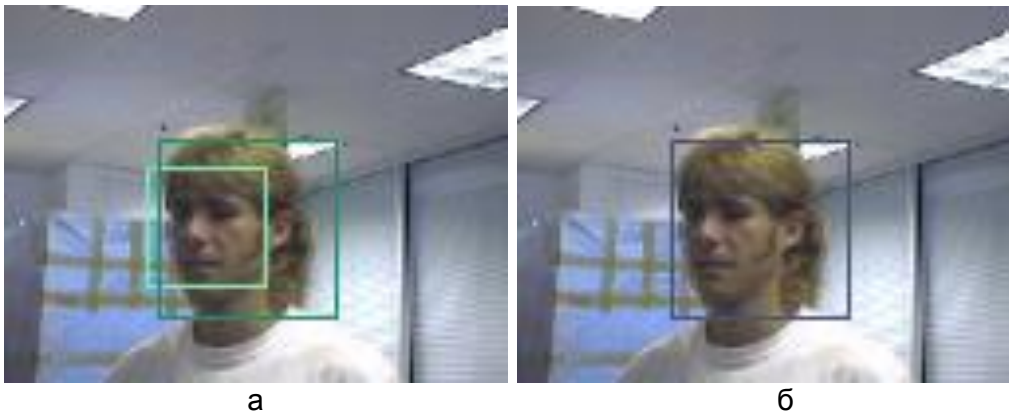


Рис. 3. Пример работы системы трекинга без использования метода (а) и с методом (б)

**Выводы.** В результате исследования работы метода, количество полученных треков и метрика ID Switches уменьшились для всех видов трекеров. Метрики МОТА и МОТР для большинства трекеров не снизилась, что показывает на работоспособность метода, но при применении трекеров MIL и CSRT эти показатели незначительно уменьшились. Это объясняется тем, что при работе метода более точные треки не могли появиться, так как уже существует трек, относящийся к отслеживаемому лицу. Также при работе с трекерами AdaBoosting и MIL произошло слияние треков, принадлежащим двум разным объектам, что было описано в качестве недостатков метода.

Полученные результаты показали, что метод повышения точности трекинга за счет компенсации ошибок назначения трека в текущей реализации способен справиться с поставленной задачей. Перспективами дальнейшего исследования проблемы компенсации ошибок трекинга являются использование других признаков для определения лица, а также добавление алгоритмов для компенсации других ошибок трекинга.

### Литература

1. Олейник А.Л. Применение бинарных дескрипторов для трекинга множества лиц в системах видеонаблюдения // Научно-технический вестник информационных технологий, механики и оптики. – 2016. – Т. 16. – № 4. – С. 670–677.
2. Hofmann M., Haag M., Rigoll G. Unified hierarchical multi-object tracking using global data association // IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS). – 2013. – P. 22–28.



3. Bernardin K., Stiefelwagen R. Evaluating multiple object tracking performance: the CLEAR MOT metrics // Journal on Image and Video Processing. – 2008. – V. 2008. – P. 1.
4. OpenCV (Open source Computer Vision library) [Электронный ресурс]. – Режим доступа: <http://opencv.org/> (дата обращения: 11.02.2019).
5. Viola P., Jones M. Rapid object detection using a boosted cascade of simple features [Электронный ресурс]. – Режим доступа: <https://www.cs.cmu.edu/~srini/15-829/readings/ViJo01.pdf> (дата обращения: 11.02.2019).
6. Birchfield S. Elliptical head tracking using intensity gradients and color histograms // IEEE Computer Society conference on computer vision and pattern recognition. – 1998. – P. 232–237.

**Петров Олег Евгеньевич**

Год рождения: 1989

Университет ИТМО, факультет информационных технологий и программирования, аспирант

Направление подготовки: 09.06.01 – Информатика и вычислительная техника

e-mail: petrov-o@speechpro.com

**Кабаров Владимир Иосифович**

Год рождения: 1959

Университет ИТМО, факультет информационных технологий и программирования, ст. преподаватель

e-mail: kabarov@speechpro.com

**Матвеев Юрий Николаевич**

Год рождения: 1955

Университет ИТМО, факультет информационных технологий и программирования, д.т.н.

e-mail: matveev@mail.ifmo.ru

**УДК 004.934.5****ИСПОЛЬЗОВАНИЕ ОБРАТНЫХ ИНДЕКСОВ ДЛЯ ЭФФЕКТИВНОГО  
ПОЛНОТЕКСТОВОГО ПОИСКА ПО РЕЧЕВЫМ ДАННЫМ****Петров О.Е., Кабаров В.И.****Научный руководитель – д.т.н. Матвеев Ю.Н.**

В рамках работы была исследована возможность использовать обратный индекс для решения задачи полнотекстового поиска по речевым данным. Предложенный в работе подход позволяет использовать готовые решения, такие как Lucene, для поиска по фонограммам, в том числе используя сложные поисковые запросы.

**Ключевые слова:** распознавание речи, речевые информационные системы, полнотекстовый поиск, обратный индекс.

Полнотекстовый поиск является одной из ключевых технологий, значительно повлиявший на развитие сети Интернет. Популярные поисковые системы дают возможность быстро получить нужную информацию, что стимулирует рост объемов информации в сети, прежде всего текстовой. Однако прогресс в технологиях дал толчок для роста объемов не только текстовой информации, но и разного рода медиа-, видео- и аудиоданных [1]. Поиск по медиаданным традиционно решается с помощью сопровождающей метаинформации: ключевым словам, названию, источнику записи, списка участников и прочему. С развитием технологии автоматического распознавания речи стало возможным осуществлять поиск в том числе и по фразам, произнесенным на записи. Целый ряд алгоритмов решает задачу поиска ключевых слов в фонограммах, содержащих речь, но одна из ключевых для поиска по большим объемам данных особенностей опускается – полученные результаты необходимо ранжировать по релевантности перед выдачей пользователю. Задачей настоящего исследования являлась разработка метода полнотекстового поиска как по речевым данным, так и по текстовым, используя единый индекс данных.

Как правило, системы поиска по ключевым словам работают в два этапа. На первом шаге выполняется преобразование речи в текстовый формат с помощью системы автоматического распознавания речи. После этого на основе полученного текста строится поисковой индекс [2]. Самый простой способ организовать полнотекстовый поиск по речевым данным с возможностью ранжировать результат – использовать полнотекстовый поиск по лучшим гипотезам, полученным из системы

распознавания. Это позволяет осуществлять поиск с высокой скоростью. Однако достаточное качество будет достигнуто, только если система распознавания речи имеет низкую частоту ошибок слов. Современные системы распознавания речи все еще не позволяют получить высокую точность, используя только лучший результат. Это особенно критично для языков, для которых качество распознавание еще не достаточно высокое. Чтобы решить эту проблему при построении поискового индекса предполагается использовать всю словную сетку распознавания, учитывая в том числе и гипотезы с низкой достоверностью. Сетки, получаемые в процессе распознавания, имеют произвольную структуру, которая сложно поддается индексации. Вместо нее предполагается использовать более компактное представление словной сети распознавания – сети спутывания [3]. На основе такой структуры можно строить обратные индексы поиска, что, в свою очередь, дает возможность использовать классические подходы организации поисковых индексов. Нужно отметить, что помимо предложенного способа есть и другие варианты использования сетей спутывания, в том числе использованием преобразователей конечных автоматов [4].

Сети спутывания представляют собой графы, которые обладают целым рядом свойств, делающих их удобными для индексирования. В общем виде такие сети представляет собой последовательность бинов. Узлы содержат временные метки, а переходы – распознанные слова с вероятностями. Среди основных свойств такой структуры можно выделить:

- все пути в графе от начального узла к конечному проходят через все вершины графа. Обход такого графа может быть сделан за время  $O(N)$ , где  $N$  – общее количество переходов;
- каждый бин может содержать произвольное количество переходов;
- каждый бин может включать в себя не более чем один эpsilon-переход, означающий, что на данном бине слово может быть пропущено;
- для каждой вершины с временной меткой выполняется правило  $t_i > t_{i-1}$ , т.е. вершины упорядочены по времени, начиная от стартовой вершины и следуя до конечной. Если отсортировать переходы в каждом из бинов по убыванию вероятности, то можно получить лучшую гипотезу за время  $O(N)$ , где  $N$  – количество бинов в сети спутывания.

Благодаря своим свойствам сети спутывания можно считать обобщением обычного текста. Таким образом, обычный текст может быть представлен в виде сети следующим образом:

- каждое слово является бином с одним переходом;
- все слова имеют вероятность 1;
- эpsilon-переходов нет;
- временные метки могут быть расставлены по желанию. Это может быть ссылка на документ, из которого получен текст, или временные метки, полученные с помощью какого-то выравнивания.

Обратный индекс для текста – это структура, которая выглядит следующим образом:

w1: p1, p2, p3  
w2: p3, p4, p6, p123  
...  
wn: ...

Иначе говоря, для каждого слова мы храним те позиции, на которых слово находится в документе. Эта структура довольно простая и позволяет делать сложный поиск по ней, включая пересечения, исключения и другие логические операции.

Для сети спутывания эта структура немного усложняется: сама по себе позиция становится не просто порядковым номером слова в тексте, а номером бина. Другими

словами она становится некой ссылкой на объект бина. Его интерпретация может различаться: временная метка, позиция слова в тексте или что-то еще.

- Каждой позиции ставится в соответствие вероятность слова.
- Несколько слов могут оказаться на одной позиции, но с вероятностями меньше единицы.
- Появляется отдельное эpsilon-слово, которое само по себе смысла не имеет, но требует учета.

Использование таких подходов позволяет строить индексы для полнотекстового поиска по речевым данным, решая среди прочего задачу ранжирования. Для таких индексов также применимы сложные запросы, использующие логические операторы.

### **Литература**

1. Medennikov I., Sorokin I., Romanenko A., Popov D., Khokhlov Yu., Prisyach T., Malkovskii M., Bataev V., Astapov S., Korenevsky M., Zatornitskiy A. The STC System for the CHiME 2018 Challenge // Proc. CHiME 2018 Workshop on Speech Processing in Everyday Environments. – 2018. – P. 1–5.
2. Zobel J., Moffat A. Inverted files for text search engines [Электронный ресурс]. – Режим доступа: <https://docplayer.net/19846805-Inverted-files-for-text-search-engines.html> (дата обращения: 06.03.2019).
3. Mangu L., Brill E., Stolcke A. Finding consensus in speech recognition: word error minimization and other applications of confusion networks // Computer Speech & Language. – 2000. – V. 14(4). – P. 373–400.
4. Mangu L., Kingsbury B., Soltau H., Kuo H.-K., Picheny M. Efficient spoken term detection using confusion networks // ICASSP. – 2014. – P. 7844–7848.

**Рюмин Дмитрий**

Год рождения: 1991

Санкт-Петербургский институт информатики и автоматизации РАН, мл.н.с.;  
Университет ИТМО, факультет информационных технологий и программирования,  
аспирант

Направление подготовки: 09.06.01 – Информатика и вычислительная техника

e-mail: dl\_03.03.1991@mail.ru

**Аксенов Александр**

Год рождения: 1995

Санкт-Петербургский институт информатики и автоматизации РАН, мл.н.с.;  
Университет ИТМО, факультет программной инженерии и компьютерной техники,  
студент

Направление подготовки: 09.04.04 – Информационно-вычислительные системы

e-mail: a.aksenov95@mail.ru

**Карпов Алексей Анатольевич**

Год рождения: 1978

Санкт-Петербургский институт информатики и автоматизации РАН, мл.н.с.;  
Университет ИТМО, факультет информационных технологий и программирования,  
д.т.н., профессор

e-mail: karpov@iias.spb.su

**УДК 004.852****МЕТОД АВТОМАТИЧЕСКОГО ОБНАРУЖЕНИЯ И РАСПОЗНАВАНИЯ  
3D ОДНОРУЧНЫХ ЖЕСТОВ РУК**

**Рюмин Д.** (Санкт-Петербургский институт информатики и автоматизации РАН,  
Университет ИТМО), **Аксенов А.** (Санкт-Петербургский институт информатики  
и автоматизации РАН, Университет ИТМО)

**Научный руководитель – д.т.н., профессор А.А. Карпов**  
(Санкт-Петербургский институт информатики и автоматизации РАН,  
Университет ИТМО)

Работа выполнена в рамках темы НИР № 618278 «Синтез моциональной речи на основе генеративных состязательных сетей».

В работе представлен новый подход обнаружения и распознавания 3D одnorучных жестов для человеко-машинного взаимодействия. Описана логическая структура модулей системы MulGesRecDB для записи жестовой базы данных. Представлена логическая структура базы данных трехмерных жестов. Проиллюстрированы примеры кадров показа жестов в формате FullHD, в режиме карты глубины и в инфракрасном диапазоне. Результаты автоматического определения областей интереса иллюстрируются. Данный метод за счет своей универсальности может использоваться в задачах биометрии, компьютерного зрения, машинного обучения, автоматических системах распознавания лиц, жестовых языков.

**Ключевые слова:** жесты, жестовый язык, распознавание лиц, машинное обучение, компьютерное зрение, нейронные сети.

**Введение.** Задача повышения уровня автоматизации и роботизации всех сфер деятельности человека является одной из ключевых в современном информационном обществе [1]. В связи с этим ученые и руководства развитых, а также развивающихся стран в сотрудничестве с мировыми научными центрами и компаниями уделяют внимание технологиям для эффективного, естественного и универсального взаимодействия человека с компьютерами и роботами [2].

В настоящее время интерактивные информационные системы получают применение в сферах социального обслуживания, медицине, образования, робототехнике, военной индустрии, центрах обслуживания населения, а также для взаимодействия с людьми в различных чрезвычайных ситуациях [3]. Кроме того, все более широкое распространение находят роботы-ассистенты, которые направлены на взаимодействие с людьми для выполнения определенных задач. В этом случае многих классических интерфейсов недостаточно. Вместо этого необходимы более интуитивные и естественные для человека интерфейсы (речевой, жестовый, многомодальный [4] и т.п.). Так, например, жесты могут передавать простые команды роботу, которые будут нести однозначный смысл и эффективны на некотором расстоянии от робота и в шумных условиях, когда речь малоэффективна.

Также известно, что инвалиды по слуху ограничены в возможностях при общении со слышащими, а при обращении в различные учреждения к ним прикрепляются сурдопереводчики, которых оказывается недостаточно. Поэтому необходимы технологии распознавания жестовых языков глухих людей, при помощи которых будет осуществляться управление и взаимодействие с ассистивными мобильными информационными роботами [5]. Так, например, ученые из Американского института робототехники в Университете Карнеги–Меллона работают над системой, которая может анализировать язык тела и жесты вплоть до положения пальцев. Это исследование представлено на конференции «Computer Vision and Pattern Recognition» [6].

В работе представлен подход к автоматическому обнаружению и распознаванию как статических, так и динамических трехмерных (3D) одноручных жестов русского жестового языка в режиме реального времени с помощью оптической камеры и датчика глубины.

**Набор данных элементов русского жестового языка.** Для процесса обучения и тестирования подхода использовалась собственная база данных 3D-жестов русского жестового языка. Запись данной базы данных осуществлялась при помощи разработанной автоматической системы записи 3D-видеопотока с использованием сенсора-дальномера Microsoft Kinect 2.0. Необходимые слова (фразы) выводились на экран компьютера и демонстрировались дикторам-носителям русского жестового языка по 5 раз минимум в целях дальнейшего обучения системы автоматического распознавания элементов жестового языка на основе вероятностных нейросетевых моделей. Камера Kinect 2.0 была установлена на оптимальном расстоянии 1,5–2 м от диктора, при этом дикторы записывались на однородном светлом фоне.

Всего в записях приняло участие 13 демонстраторов. Все демонстраторы владеют русским жестовым языком. Кроме того, перед записью 3D базы данных весь разработанный авторами словарь жестов был предварительно стандартизирован специалистами центра реабилитации лиц с проблемами слуха. Все жесты сопровождалась устной артикуляцией, однако запись аудиоданных не велась. Общее количество жестовых фраз составило 2 132. Основная тематика записанной базы данных посвящена продуктам питания в супермаркете. Всего в базе данных насчитывается 48 различных одноручных жестов и 116 двуручных. Представленный в данной работе метод направлен на автоматическое обнаружение и распознавание исключительно 3D одноручных жестов. Статическая или динамическая направленность жеста в данном случае не играет никакой роли.

**Описание метода.** На вход разработанного метода подаются видеоданные, на которых происходит воспроизведение одноручных жестов русского жестового языка в двумерном (RGB-камера) и 3D-формате (карта глубины) или непосредственно от сенсора Microsoft Kinect v2. Разрешение цветных видеок кадров составляет 1920×1080

пикселей, а карты глубины 512×424 пикселей с частотой 30 кадров в секунду. Качество цветопередачи для двумерных (2D) данных составляет 8 бит, а для 3D – 16 бит. Осуществляется синхронная потоковая обработка 2D и 3D данных видеоданных. На каждом кадре с помощью карты глубины и набора средств разработки (SDK), поставляемых вместе с сенсором Kinect v2 происходит поиск людей на расстоянии от 1,2 до 3,5 м и вычисление 3D 25-ти точечной модели скелета для каждого человека. После 3D-координаты преобразуются в 2D и формируются прямоугольные области с людьми на 2D-изображении и их 2D 25-ти точечные модели скелета, как показано на рис. 1.

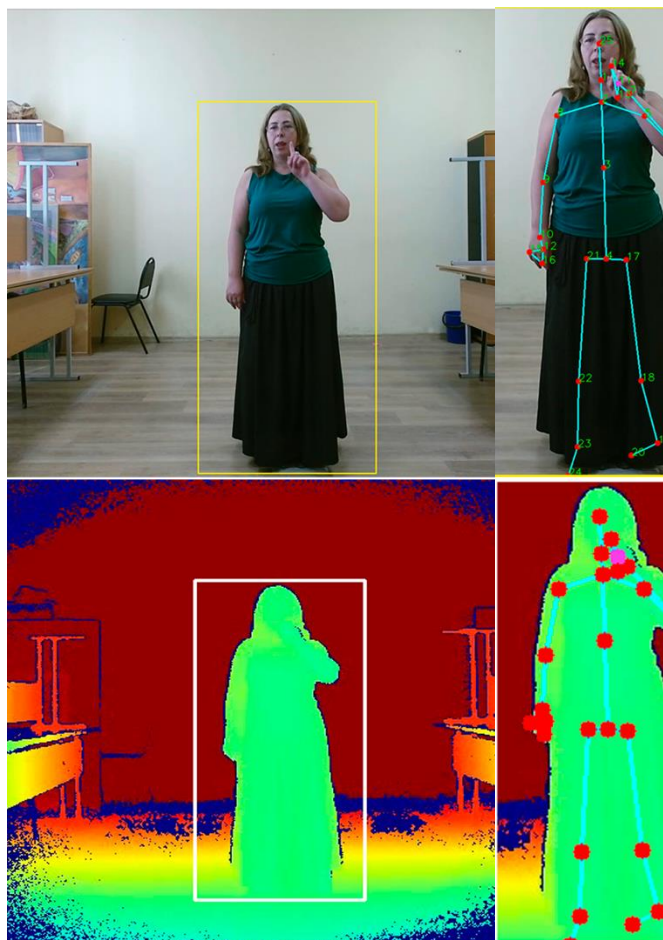


Рис. 1. Пример вычисления 3D и 2D 25-ти точечной модели скелета для каждого человека

Далее используются две модели глубокой сверточной сети для обнаружения лица и формы руки, как показано на рис. 2.

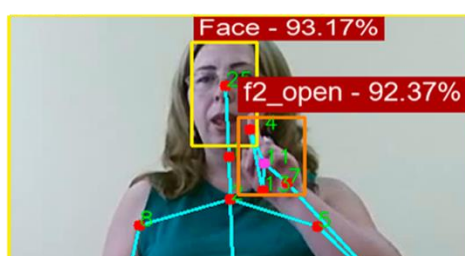


Рис. 2. Пример обнаружения области с лицом и формы руки

Детектор лица основан на структуре Single Shot MultiBox Detector (SSD [7]) с уменьшенной моделью сети ResNet-10 [8]. Данный детектор включен в модуль Deep

Neural Networks (DNN) библиотеки компьютерного зрения Open Source Computer Vision Library [9]. Данный детектор имеет следующие особенности:

1. работает в режиме реального времени, как на Central Processing Unit (CPU), так и на Graphics Processing Unit (GPU);
2. работает для различных ориентаций лица – вверх, вниз, влево, вправо и т.д.;
3. работает даже при существенной окклюзии;
4. обнаруживает лица в различных масштабах.

Детектор форм рук основан на структуре SSD с моделью сети MobileNetV2 [10]. Данный детектор обучался на предварительно собранной и размеченной мультимедийной базе данных 3D-жестов русского жестового языка.

Для выделения области с жестом использовался инструмент для графического аннотирования изображений LabelImg [11]. Аннотации сохраняются в виде файлов eXtensible Markup Language (XML) в формате PASCAL VOC [12], который используется ImageNet [13]. Кроме того, он также поддерживает формат YOLO [14]. Для обучения детектора использовались первые 4 повторения жеста с привязанными к ним координатами необходимых скелетных точек. Данные повторения считаются эталонами показа жеста, а остальные использовались как тестовые данные.

Найденные координаты лиц и форм рук подвергаются нормализации. На основе нормализованных координат происходит вычисление 3D-расстояния от лица до руки, а также подсчет площади пересечения координат руки и лица, как показано на рис. 3. В итоге отличительными характеристиками жеста в определенный момент времени являются:

- нормализованное 3D-расстояние от лица до руки (зона артикуляции жеста);
- нормализованная площадь пересечения лица и руки;
- форма руки.

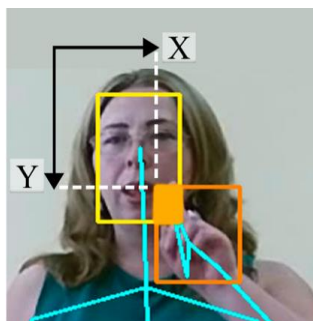


Рис. 3. Пример вычисления отличительных характеристик жеста в определенный момент времени

Тестирование показало, что средняя скорость обработки кадров составляет 12 кадров в секунду.

Заключительным этапом метода является распознавание одноручных жестов, используя описанные характеристики жеста и современные алгоритмы машинного обучения, которые включены в такие инструменты, как scikit-learn [15], Tensorflow Object Detection API [16], Keras [17], которые упрощают создание, обучение и развертывание моделей обнаружения как статических, так и динамических объектов. В частности, используется LSTM [18].

Для обучения используется по 4 повторения жеста одного и того же демонстратора. Общее количество обучающих видео составило 2 496. LSTM тренировалась на последовательностях из 15 кадров до 10 шагов и скоростью обучения 0,003. Обучение нейронной сети производилось с помощью инструмента Keras [17] с использованием RMSprop [19] и асинхронного градиентного спуска.



Результаты экспериментов показали, что средняя точность распознавания составляет 0,63 (63%). Данные результаты получены на 48 различных одноручных жестов из собранной мультимедийной базы данных.

**Заключение.** Таким образом, предложен метод, который объединяет в себе современные алгоритмы компьютерного зрения и глубокие нейронные сети, что позволяет обнаруживать и распознавать 3D одноручные жесты в режиме реального времени. Данный метод за счет своей универсальности может использоваться в задачах биометрии, компьютерного зрения, машинного обучения, автоматических системах распознавания лиц, жестовых языков.

### Литература

1. Ryumin D., Karpov A. Towards Automatic Recognition of Sign Language Gestures using Kinect 2.0 // In Proc. 19th International Conference on Human-Computer Interaction, HCI, Vancouver, Canada, Springer LNCS. – 2017. – V. 10278. – P. 89–104.
2. Ivanko D., Karpov A., Fedotov D., Kipyatkova I., Ryumin D., Ivanko Dm., Minker W., Zelezny M. Multimodal Speech Recognition: Increasing Accuracy using High Speed Video Data // Journal on Multimodal User Interfaces, Springer. – V. 12. – I. 4. – P. 319–328.
3. Toyota Global Site [Электронный ресурс]. – Режим доступа: [http://www.toyota-global.com/innovation/partner\\_robot](http://www.toyota-global.com/innovation/partner_robot) (дата обращения: 06.03.2019).
4. Ivanko D., Ryumin D., Axyonov A., Železný M. Designing Advanced Geometric Features for Automatic Russian Visual Speech Recognition // Proc. 20th International Conference on Speech and Computer SPECOM-2018. – V. 11096. – P. 245–254.
5. Guo X., Yang T. Gesture recognition based on HMM-FNN model using a Kinect // J. Multimodal User Interfaces, Springer. – 2016. – V. 11. – № 1. – P. 1–7.
6. Kinect SDK 2.0 [Электронный ресурс]. – Режим доступа: <https://www.microsoft.com/en-us/download/details.aspx?id=44561> (дата обращения: 06.03.2019).
7. Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C-Y., Berg A. SSD: Single Shot MultiBox Detector // European conference on computer vision, ECCV. – 2016. – P. 21–37.
8. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition // IEEE Conference on Computer Vision and Pattern Recognition. – 2016. – P. 770–778.
9. OpenCV [Электронный ресурс]. – Режим доступа: <https://opencv.org> (дата обращения: 06.03.2019).
10. Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1801.04381.pdf> (дата обращения: 06.03.2019).
11. LabelImg / LabelImg is a graphical image annotation tool and label object bounding boxes in images [Электронный ресурс]. – Режим доступа: <https://github.com/tzutalin/labelImg> (дата обращения: 06.03.2019).
12. Everingham M., Van Gool L., Williams C. K., Winn J., Zisserman A. The pascal visual object classes (voc) challenge // International journal of computer vision. – 2010. – V. 88. – № 2. – P. 303–338.
13. Krizhevsky A., Sutskever I., Hinton G.E. Imagenet classification with deep convolutional neural networks // Advances in Neural Information Processing Systems. – 2012. – P. 1097–1105.
14. Redmon J., Divvala S., Girshick R., Farhadi A. You only look once: Unified, real-time object detection // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2016. – P. 779–788.

15. Scikit-learn / Scikit-learn Machine Learning [Электронный ресурс]. – Режим доступа: <https://scikit-learn.org> (дата обращения: 06.03.2019).
16. Huang J., Rathod V., Sun Ch., Zhu M., Korattikara A., Fathi A., Fischer I., Wojna Z., Song Y., Guadarrama S., Murphy K. Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors // Proceedings of 30th IEEE Conference on Computer Vision and Pattern Recognition. – 2017. – P. 3296–3297.
17. Keras / Keras Documentation [Электронный ресурс]. – Режим доступа: <https://keras.io> (дата обращения: 06.03.2019).
18. Hochreiter S., Schmidhuber J. Long short-term memory // Neural computation. – 1997. – V. 9. – № 8. – P. 1735–1780.
19. Graves A. Generating Sequences with Recurrent Neural Networks [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1308.0850.pdf> (дата обращения: 06.03.2019).

**Сухов Владимир Дмитриевич**

Год рождения: 1994

Университет ИТМО, факультет информационных технологий и программирования, аспирант

Направление подготовки: 09.06.01 – Информатика и вычислительная техника

e-mail: vdsukhov@ya.ru

**Короткевич Геннадий Владимирович**

Год рождения: 1994

Университет ИТМО, факультет информационных технологий и программирования, аспирант

Направление подготовки: 09.06.01 – Информатика и вычислительная техника

e-mail: gkorotkevitch@yandex.ru

**Сергушичев Алексей Александрович**

Год рождения: 1991

Университет ИТМО, факультет информационных технологий и программирования, к.т.н.

e-mail: alserg@itmo.ru

**УДК 004.021****ВЫЧИСЛЕНИЕ БЛИЗКИХ К НУЛЮ P-ЗНАЧЕНИЙ ПРИ АНАЛИЗЕ  
ПРЕДСТАВЛЕННОСТИ ФУНКЦИОНАЛЬНЫХ НАБОРОВ ГЕНОВ****Сухов В.Д., Короткевич Г.В.****Научный руководитель – к.т.н. Сергушичев А.А.**

Работа выполнена в рамках темы НИР № 718574 «Методы, модели и технологии искусственного интеллекта в биоинформатике, социальных медиа, киберфизических, биометрических и речевых системах».

Экспрессия генов является важным процессом, который протекает в живых клетках. Одной из основных задач при проведении биологических экспериментов является интерпретация данных дифференциальной экспрессии генов. Широкое распространение для решения этой задачи получил тест представленности функциональных наборов генов. Данный тест позволяет определять наборы генов, демонстрирующие неслучайное поведение в эксперименте. Так, в основе теста находятся вычисления P-значений, которые, в свою очередь, позволяют делать вывод о статистической значимости исследуемых функциональных наборов генов. В работе предложен корректный переход к логарифму P-значений. Определены теоретические характеристики для распределения P-значений.

**Ключевые слова:** экспрессия генов, марковский процесс, P-значение, набор генов, сэмплирование.

**Введение.** Определение статистически значимых функциональных наборов генов является важной задачей при интерпретации данных дифференциальной экспрессии генов. Актуальным и распространенным решением этой задачи является тест представленности функциональных наборов генов. Данный подход позволяет из заранее известной коллекции функциональных наборов генов определить те, что имеют неслучайное поведение в эксперименте. Классическими методами по определению статистически значимых наборов генов являются GSEA (Gene Set Enrichment Analysis) [1] и FGSEA [2, 3] (Fast Gene Set Enrichment Analysis). Основой методов является применение функции представленности (статистика представленности). Применение функции представленности позволяет проводить сравнение между наборами генов. Для вычисления P-значений в методах GSEA и FGSEA применяется «наивное»

сэмплирование, что является недостатком. Применение «наивного» сэмплирования требует задействования больших вычислительных мощностей в случае, если набор генов имеет высокое значение статистики представленности и, соответственно, близкое к нулю точное Р-значение. В работе [4] был предложен приближенный алгоритм, основанный на применении многоуровневого подхода Монте-Карло. Он позволяет вычислять сколь угодно малые Р-значения за разумное время работы алгоритма с хорошей относительной точностью. При вычислении близких к нулю Р-значений важным аспектом является уровень получаемой ошибки. Для нахождения допустимого интервала, в котором должны находиться приближенные Р-значения, в данной работе осуществлен корректный переход к логарифму вероятности. Определены некоторые теоретические характеристики для распределения приближенных Р-значений, также исследовано отклонение получаемых Р-значений от точных в зависимости от числа шагов в алгоритме.

**Постановка задачи.** Входными параметрами задачи являются вектор весов  $\mathbf{S} = (S_1, S_2, \dots, S_N)$  и коллекция функциональных наборов генов  $G$ . Для определения функциональных наборов генов с неслучайным поведением применяется специальная функция – статистика представленности. Для некоторого набора генов  $gset \in G$  размера  $k$  она задается следующим образом:

$$f_{ES}(gset) = \max_{i \in gset} |ES_i|,$$

где  $ES_i$  вычисляются так:

$$ES_i = \begin{cases} 0, & \text{если } i = 0, \\ ES_{i-1} + \frac{|S_i|}{\sum_{i \in gset} |S_i|}, & \text{если } 1 \leq i \leq N \text{ и } i \in gset. \\ ES_{i-1} - \frac{1}{N-k}, & \text{если } 1 \leq i \leq N \text{ и } i \notin gset \end{cases}.$$

Задача заключается в нахождении следующей вероятности (Р-значения):

$$P(f_{ES}(X) > f_{ES}(gset)),$$

где  $gset$  – фиксированный набор генов размера  $k$ ;  $X$  – случайный набор размера  $k$ . Таким образом, необходимо определить вероятность того, что случайный набор генов размера  $k$  имеет значение статистики представленности больше, чем у исследуемого набора генов. Так, например, для двух наборов генов одинакового размера  $gset_1$  и  $gset_2$  с вероятностями  $p_1$  и  $p_2$  соответственно значения  $p_1$  и  $p_2$  позволяют определить наиболее «неслучайный» набор генов.

**Описание алгоритма.** Предлагается рассмотреть дробление по значениям функции представленности:

$$-1 = f_0 < f_1 < \dots < f_t = f_{ES}(gset).$$

Затем, вместо непосредственного вычисления вероятности  $P(f_{ES}(X) > f_{ES}(gset))$ , предлагается вычислить следующие условные вероятности:

$$P(f_{ES}(X) > f_1 | f_{ES}(X) > f_0) = p_1,$$

$$P(f_{ES}(X) > f_2 | f_{ES}(X) > f_1) = p_2,$$

....

$$P(f_{ES}(X) > f_t | f_{ES}(X) > f_{t-1}) = p_t.$$

Это позволяет ввести оценку для искомой вероятности в следующем виде:

$$P(f_{ES}(X) > f_{ES}(gset)) \approx \prod_{i=1}^t p_i.$$

Таким образом, исходная задача сводится к вычислению условных вероятностей. Для их оценки применяется условное сэмплирование, основанное на применении алгоритма Метрополиса. Так, для вычисления каждой условной вероятности генерируются выборки нечетного размера  $Z$  из необходимого распределения. В

качестве уровней  $f_i, i \in \{1, 2, \dots, t-1\}$  берется медиана выборки из распределения  $P(\cdot | f_{ES}(X) > f_{i-1})$ , поэтому  $p_i = 2^{-1}, i \in \{1, 2, \dots, t-1\}$ . Пусть  $X^t = (X_1^t, X_2^t, \dots, X_Z^t)$  – выборка из распределения  $P(\cdot | f_{ES}(X) > f_{t-1})$ , тогда оценка для искомой вероятности выглядит так:

$$P(f_{ES}(X) > f_{ES}(gset)) \approx 2^{-(t-1)} \cdot \frac{\#\{X_j^t | f_{ES}(X_j^t) > f_t\}}{Z}.$$

**Переход к логарифму вероятности.** Простое применение функции логарифма к Р-значению способно привести к тому, что получаемая оценка не будет являться несмещенной. Для корректного перехода к логарифму Р-значений важно учесть следующее замечание. Известно, что упорядоченные по возрастанию элементы выборки из равномерного распределения являются случайными величинами из бета-распределения. Это дает возможность воспользоваться известными свойствами для математического ожидания и дисперсии логарифма случайной величины из бета-распределения. Так для бета-распределения с параметрами  $(\alpha, \beta)$  выполнено следующее:

$$E[\log \xi] = \psi(\alpha) - \psi(\alpha + \beta), \psi - \text{дигамма-функция};$$

$$D[\log \xi] = \psi_1(\alpha) - \psi_1(\alpha + \beta), \psi_1 - \text{тригамма-функция}.$$

Таким образом, корректный переход к логарифму вероятности имеет следующий вид:

$$\log P(f_{ES}(X) > f_{ES}(gset)) \approx (t-1) \left( \psi\left(\frac{Z+1}{2}\right) - \psi(Z+1) \right) + \log p_t.$$

Также можно ввести теоретическую оценку сверху для среднеквадратического отклонения результатов работы модифицированного алгоритма:

$$\text{expected sd} = \sqrt{t \left( \psi_1\left(\frac{Z+1}{2}\right) - \psi_1(Z+1) \right)}.$$

**Сравнение с точным алгоритмом.** Сравнение результатов работы модифицированного алгоритма проводилось с точным алгоритмом [4]. Точный алгоритм работает для целочисленного вектора весов, поэтому на вход обоим алгоритмам подавался идентичный вектор с целочисленными значениями весов. Результаты работы алгоритмов представлены на рис. 1. Видно, что Р-значения, получаемые модифицированным алгоритмом, находятся в хорошем соответствии с точными значениями.

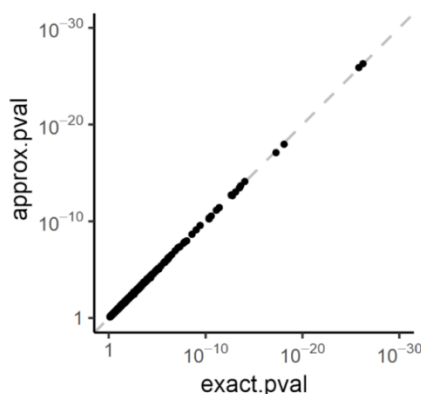


Рис. 1. Сравнение с точным алгоритмом, ось ОХ соответствует точным Р-значениям, ось ОУ соответствует Р-значениям, полученным модифицированным алгоритмом

Точный алгоритм также позволяет определить уровень отклонения наблюдаемого среднеквадратического отклонения от теоретически полученной верхней оценки.

Соответствующие результаты приведены в таблице. Каждое значение «real sd» было получено путем усреднения результатов 1 000 запусков модифицированного алгоритма.

Таблица. Сравнение среднеквадратических отклонений

#gset = k	$f_{ES}(gset)$	Z	real sd	expected sd
50	0,855	1001	0,2995	0,2845
50	0,700	1001	0,1826	0,1816
50	0,550	1001	0,1200	0,1183
100	0,855	1001	0,4158	0,3999
100	0,740	1001	0,2926	0,2828
100	0,580	1001	0,1928	0,1788
100	0,450	1001	0,1211	0,1183
250	0,825	1001	0,6038	0,5752
250	0,710	1001	0,4166	0,4098
250	0,590	1001	0,2906	0,2880
250	0,450	1001	0,1806	0,1788

Видно, что реально наблюдаемые значения среднеквадратических отклонений превышают те, что рассчитаны теоретически. Это объясняется тем, что результаты работы модифицированного алгоритма сходятся к точным результатам, при стремлении к бесконечности числа шагов. На практике число шагов необходимо чем-то ограничивать.

**Число шагов алгоритма.** Для получения выборки из распределения  $P(\cdot | f_{ES}(X) > f_{i-1})$  необходимо совершить следующие действия:

1. рассмотрим наборы генов  $X_1, X_2, \dots, X_Z$  такие, что  $f_{ES}(X_j) > f_{i-1}, j \in \{1, 2, \dots, Z\}$ ;
2. для каждого набора  $X_j$  случайным образом выбираем ген из набора и вне его. Получаем новый набор  $\tilde{X}_j = X_j \setminus \{g\} \cup \{g'\}$ , где  $g \in X_j$ , и  $g' \notin X_j$ . Заменяем  $X_j$  на  $\tilde{X}_j$ , если  $f_{ES}(\tilde{X}_j) > f_{i-1}$ ;
3. для каждого набора  $X_j$  необходимо в количестве  $a \cdot k$  раз совершить замены, удовлетворяющие условию  $f_{ES}(\tilde{X}_j) > f_{i-1}$ .

Для исследования результатов работы алгоритма в зависимости от параметра  $a$  проводилось сравнение с точными P-значениями для целочисленного вектора весов. Результаты запусков приближенного алгоритма для различных наборов генов в зависимости от параметра  $a$  приведены на рис. 2. Каждой точке на графике соответствует усреднение по 100 запускам. На практике  $a = 1$  позволяет получить разумный уровень «случайности» генерируемых выборок без существенного увеличения времени работы алгоритма.

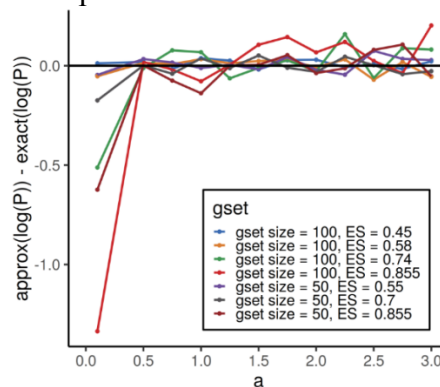


Рис. 2. Отклонение приближенного P-значения от точного в зависимости от числа шагов

**Заключение**

1. Предложен переход к логарифму вероятности с учетом распределения P-значений.
2. Введены характеристики, позволяющие отслеживать отличия наблюдаемого распределения P-значений от теоретического.
3. Полученные результаты внедрены в пакет FGSEA, который реализован на языке программирования R.

**Литература**

1. Subramanian A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles // *Proceedings of the National Academy of Sciences of the United States of America*. – 2005. – V. 102. – № 43. – P. 15545–15550.
2. Сергушичев А.А. Алгоритм для быстрого анализа перепредставленности генов // *Список*. – 2016. – С. 238–244.
3. Sergushichev A.A. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation [Электронный ресурс]. – Режим доступа: <https://www.biorxiv.org/content/10.1101/060012v1> (дата обращения: 06.03.2019).
4. Короткевич Г.В. Алгоритмы для эффективного анализа представленности функциональных наборов генов: выпускная квалификационная работа. – СПб.: Университет ИТМО, 2018.

**Чирковский Артем Дмитриевич**

Год рождения: 1995

Университет ИТМО, факультет информационных технологий и программирования, студент группы № М41212

Направление подготовки: 09.04.02 – Информационные системы и технологии

e-mail: artemioarbaletos@gmail.com

**Бессмертный Игорь Александрович**

Год рождения: 1954

Университет ИТМО, факультет программной инженерии и компьютерной техники, д.т.н., доцент

e-mail: bia@cs.ifmo.ru

УДК 004.912

**КОМПЬЮТЕРНАЯ ФОРМАЛИЗАЦИЯ ТЕКСТОВ НА ЯЗЫКЕ ЭСПЕРАНТО****Чирковский А.Д.****Научный руководитель – д.т.н., доцент Бессмертный И.А.**

В работе рассмотрены способы машинной обработки текстов на языке эсперанто на морфемном, морфологическом и синтаксическом уровнях с учетом специфики строения и использования данного языка. Предложена программная система для автоматизированной формализации и поверхностного извлечения фактов из текстов на языке эсперанто, затрагивающая уровни морфологического и синтаксического анализа с учетом недостатков существующих систем обработки текстов на языке эсперанто.

**Ключевые слова:** автоматическая обработка текстов, плановые языки, эсперанто, синтаксический разбор, извлечение фактов.

Бурное развитие последних лет в сфере обработки естественных языков обошло стороной язык эсперанто, несмотря на большие перспективы использования подобного универсального языка в области разработки систем с человеко-машинным интерфейсом: отсутствие исключений в правилах языка упрощает обработку текстовой информации; однозначность фонетики упрощает процессы преобразования текста в фонемы и обратно, применяемые в задачах синтеза и распознавания речи. Эсперанто – это плановый (искусственный) язык, созданный в 1887 году как универсальное средство коммуникации между людьми разных национальностей. По строению язык является агглютинативным с элементами аналитического, и основными преимуществами данного языка как международного средства общения являются интернациональная лексика и грамматическая простота. Однако видимая простота грамматики языка часто приводит к недооцениванию сложности его синтаксической обработки машиной. Существующие исследования показывают, что, несмотря на высокую точность автоматического морфологического и морфемного анализа, синтаксический и семантический анализ языка эсперанто традиционными средствами не выделяется на фоне естественных языков. Поэтому перспективным представляется использование в отношении к эсперанто современных методов корпусной лингвистики, однако в данном подходе ограничивающим фактором является отсутствие размеченных корпусов языка эсперанто в открытом доступе. Эксперименты по использованию эсперанто для передачи семантической информации не увенчались успехом ввиду отсутствия на тот момент инструментов для захвата контекста повествования, достаточного количества данных, и, возможно, финансирования [1]. В задаче морфемного анализа, т.е. разделения каждого слова на составляющие части, хорошо себя показал словарный метод с разрешением неоднозначностей на основе скрытых марковских моделей [2]. Для морфологического анализа с достаточно высоким качеством (98–99% точность разметки) применяются словари, анализаторы окончаний



и те же марковские модели [3, 4]. Как упоминалось, синтаксический анализ производится не столь успешно: в существующих решениях используются методы на основе грамматик, и точность синтаксической разметки составляет 90% согласно имеющимся источникам, что примерно соответствует точности разметки английского языка современными синтаксическими анализаторами на основе нейронных сетей [3, 4]. Системы, основанные на правилах, действительно способны выдавать достойный результат, но вместе с тем они имеют ряд недостатков: во-первых, высокие требования к квалификации разрабатывающего систему правил эксперта, а также к продолжительности его работы. Кроме того, вольный порядок слов в предложениях на языке требует написания крайне гибких правил грамматики в большом объеме, что приводит к высокой комбинаторной сложности разбора длинных предложений. В настоящей работе предложена система для лингвистической обработки языка эсперанто, состоящая из модуля морфологического и синтаксического анализов, причем используется метод, нивелирующий перечисленные выше недостатки существующих синтаксических анализаторов. Качество работы предлагаемой системы уступает перечисленным аналогам на некоторых целевых текстах, но сие находится допустимым ввиду закрытости существующих аналогов и невозможности более тщательного сравнения на предмет точности, скорости, универсальности. Кроме того, не самая высокая точность синтаксического разбора существующих решений показывает, что системы на правилах сами по себе недостаточно хороши. Но для использования более технологичных статистических методов потребуется корпус, и поэтому основная цель разработанной системы – это не наиболее точная автоматическая разметка, сколько наибольшее автоматизирование процесса ручной разметки текстов на языке эсперанто.

Несмотря на то, что этап морфологического анализа проводится с высокой точностью, ошибки все-таки существуют, и поэтому может появиться вопрос, какого рода ошибки возникают в языке, грамматические правила которого выглядят как «все существительные оканчиваются на *-o*, все прилагательные на *-a*, все глаголы в инфинитиве на *-i*» [5]. На самом деле, многие проблемы, встающие перед разработчиком системы морфологического анализа текста на эсперанто, типичны для прочих европейских языков: это первичная проблема разделения текстов на предложения, это токенизация, расшифровка сокращений и аббревиатур, а также обнаружение именованных сущностей. Некоторую опасность представляет факт, что вышеперечисленные правила окончаний слов не распространяются на имена собственные, которые могут принимать практически любое окончание. Частеречная омонимия представлена в языке эсперанто не очень ярко, но вместе с тем она все же имеет место. По большей части это касается служебных слов, которые могут выступать в роли предлогов или союзов, что может оказаться весьма критичным на этапе синтаксического анализа. Частичное разрешение указанной омонимии возможно посредством анализа частей речи соседних слов или на основе языковой модели, однако вышеупомянутый свободный порядок слов позволяет формировать предложения таким образом, что анализ соседей или языковая модель не будут реагировать правильно, и таким образом работа статистических моделей не безупречна. Однако, несмотря на указанные проблемы и благодаря вышеописанным контрмерам и простой системе определения части речи и грамматических характеристик слов открытого класса (т.е. существительных, прилагательных, наречий, глаголов) по их окончанию, общая точность морфологического анализа весьма велика – 98%. Между тем проблема распознавания именованных сущностей и кореференции существует. Однако никаких подвижек по ее решению в настоящее время не наблюдается, возможно, в связи с упомянутой выше проблемой отсутствия размеченных корпусов, в том числе корпусов именованных сущностей.

Синтаксический разбор языка эсперанто представляется куда более сложной задачей, чем морфологический. Это объясняется очень высокой вариативностью предложений по причине вольного порядка слов. Причем в языке эсперанто сложно выделить преобладающую модель порядка слов, так как он позиционируется как интернациональный, и люди, используя его, могут ненамеренно переносить на гибкий синтаксис эсперанто модель порядка слов из своего родного языка (однако есть предположение, что, скорее всего, в «классическом» эсперанто преобладает модель SVO: субъект-глагол-объект, так как при создании синтаксиса эсперанто его создатель, Людвиг Маркович Заменгоф, ориентировался на синтаксис русского языка). Бороться с такими проблемами можно с помощью хорошо проработанных грамматик, которые весьма трудоемки в разработке и сопровождении. Кроме того, обилие разных правил, уместных в одних и тех же ситуациях приводит к комбинаторному взрыву сложности при обработке длинных предложений, что не очень удобно для практического использования. Задачу синтаксического анализа можно упростить, если ориентироваться не на полный разбор с начала предложения, а на грамматические основы, которые легко выделить после морфологического анализа. Так, субъект предложения обычно представляется существительным или именем собственным без предлогов, объект (прямое дополнение) отображается наличием дополнительного окончания *-n*, глаголы имеют уникальные окончания, а остальные дополнения отображаются с помощью помещения перед ними предлогов. К сожалению, на практике есть некоторые тонкости, заметно усложняющие процесс. Во-первых, для имен собственных может быть не ясно однозначно, являются ли они субъектом или объектом, так как иногда происходит «эсперантификация» имен и применение к ним правил грамматики, а иногда – нет. Во-вторых, все дополнения и подлежащие могут быть распространены через прилагательные и артикли, поэтому перед их определением лучше согласовать все дополнения и сократить их до одного элемента (что дополнительно уменьшает сложность). В-третьих, фиксированной пунктуационной нормы для языка эсперанто предложено не было, поэтому ориентироваться на знаки препинания при разграничении синтаксических конструкций можно, но полагаться на них нельзя. В-четвертых, в языке предусмотрены неочевидные способы применения окончания прямого дополнения *-n* для других целей, например, указания дат, времени и направления движения. В-пятых, для используемых предлогов неочевидна их глубина, т.е. количество последующих дополнений, попадающих под действие этих предлогов. В-шестых, в предложении допустима ситуация, когда невозможно однозначно установить к какому члену предложения относятся дополнения. Частично эту проблему можно решить, задействовав семантический уровень языка и отбросив самые неправдоподобные варианты, однако, это не решает проблему в общем случае. Указанные проблемы могут сильно огорошить исследователя и объясняют, почему синтаксический разбор эсперанто является не такой уж простой задачей.

Разработанная система формализации языка эсперанто включает модуль морфологического анализа и синтаксического анализа. Модуль морфемного анализа решено было не использовать в связи с тем, что задача морфемного анализа и так решена довольно качественно. А также тем, что морфемная информация никак не используется на дальнейших этапах обработки, хотя при построении семантического анализатора использование информации о составе слов представляется весьма перспективным в связи с агглютинативным строением языка. Структуру морфологического анализатора можно назвать классической: Она состоит из блока деления на предложения, нормализатора, токенизатора (задействующего словарь популярных сокращений), словаря для определения частей речи слов закрытого класса, грамматики для определения частей речи по окончаниям слов открытого класса, словаря имен собственных, уже появившихся в тексте, и языковой модели для

разрешения возникших неоднозначностей. Как говорилось, анализатор подобного типа достигает точности 98% на классическом примере эсперанто-литературы «Fundamenta Krestomatio». Для синтаксического анализа используется система на основе комбинаторного метода синтаксического анализа. Метод заключается в разделении каждого входного предложения на значимые сегменты по знакам препинания и союзам, затем находится грамматическая основа в каждом сегменте, и затем из грамматических основ сегментов «собирается» исходное предложение таким образом, чтобы получившаяся конфигурация удовлетворяла правилам синтаксиса языка и транзитивности используемых глаголов. По результатам экспериментов было выяснено, что точность и скорость подобной системы сильно зависит от входных текстов. Таким образом, на целевом тексте «Fundamenta Krestomatio», являющимся художественным, было успешно распознано 88% предложений со скоростью около 300 предложений в секунду на тестовом оборудовании, в то время как проверка той же системы на корпусе новостных текстов из журнала «Monato» дала более удручающие результаты: 81% распознанных предложений при скорости в 60 предложений в секунду.

В результате полученная система не работает, безусловно, качественно, однако она применима для использования по назначению. Выигрыш в скорости разработки, простоты системы и скорости работы на целевых текстах считаются достаточными преимуществами данного решения. Однако доступны способы улучшения системы в ее текущем состоянии. В настоящее время целевой текст проходит через модули системы последовательно, хотя использование итеративного подхода может улучшить качество работы, так как на этапе морфологической обработки имеет смысл использовать некоторую информацию о синтаксисе обрабатываемого текста, а на этапе синтаксического полезным может оказаться информация об семантических отношениях рассматриваемых слов. Перспективным направлением является задействование семантической информации при анализе языка. Его можно осуществить или с помощью баз знаний и размеченных корпусов, которых пока нет в применении к данному языку, но которые можно в теории получить на основе, например, достаточно большой википедии на эсперанто (более 150 000 статей), или с помощью методов дистрибутивной семантики на основе языковых моделей, в том числе с применением таких современных, основанных на нейронных сетях технологий, как Word2Vec, GloVe, Bert, FastText.

### Литература

1. Witkam T. History and Heritage of the DLT (Distributed Language Translation) project. – Utrecht: Press, 2006. – 11 p.
2. Guinard T. An Algorithm for Morphological Segmentation of Esperanto Words // The Prague Bulletin of Mathematical Linguistics. – 2016. – V. 105. – P. 63–76.
3. Bick E. Tagging and Parsing an Artificial Language. An Annotated Web-Corpus of Esperanto. – University of Southern Denmark: Press, 2007. – 12 p.
4. Aasgaard B.C. Parsing of Esperanto. – University of Oslo: Press, 2007. – 144 p.
5. Д-р Эсперанто. Международный язык. Предисловие и полный учебник. Por Rusoj. – Варшава: Типо-Литографія Х. Кельтера, 1887. – 42 с.

**НАПРАВЛЕНИЕ  
ИНФОКОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ**

**Найков Алексей Вячеславович**

Год рождения: 1995

Университет ИТМО, факультет инфокоммуникационных технологий,  
студент группы № К4215Направление подготовки: 11.04.02 – Инфокоммуникационные технологии и системы  
связи

e-mail: anyqov@gmail.com

**Титов Владимир Борисович**

Год рождения: 1949

Университет ИТМО, факультет инфокоммуникационных технологий,  
к.ф.-м.н., доцент

e-mail: v.titov@spbu.ru

**УДК 004.75****РАЗРАБОТКА ПРОЦЕДУРЫ РЕАГИРОВАНИЯ НА АВАРИЙНЫЕ СИТУАЦИИ  
В ГЕОГРАФИЧЕСКИ РАСПРЕДЕЛЕННОМ ЦЕНТРЕ ОБРАБОТКИ ДАННЫХ****Найков А.В.****Научный руководитель – к.ф.-м.н., доцент Титов В.Б.**

Исследования выполнены при финансовой поддержке Министерства науки и высшего образования Российской Федерации (Договор № 03.G25.31.0229).

Университет ИТМО занимается работой по созданию системы управления масштабируемыми географически распределенными центрами обработки данных. В данной работе объяснена значимость процесса реагирования на аварийные ситуации в географически распределенном центре обработки данных, рассмотрен способ реализации данной процедуры. Для реализации в проекте была выбрана программа GLPI, которая помогает легко планировать и управлять изменениями в IT-инфраструктуре и эффективно решать проблемы, когда они появляются.

**Ключевые слова:** облако, распределенный, центр обработки данных, реагирование на аварии, HelpDesk, GLPI.

Университет ИТМО разрабатывает проект по созданию системы управления (СИУ) масштабируемыми географически распределенными центрами обработки данных (МГРЦ) [1]. Термин «географически распределенный» означает, что центры обработки данных (ЦОД) расположены в разных местах на расстоянии нескольких десятков километров. Проект разрабатывается в виде законченного продукта, который может эксплуатироваться без детальных знаний о происходящих процессах внутри ЦОД. При незначительных проблемах (отказ одной машины или нескольких дисков) ЦОД будет функционировать в нормальном режиме. При таком подходе не требуется постоянного присутствия персонала, контролирующего работу ЦОД. Необходимо будет периодически посещать ЦОД только для решения проблем, о которых нужно узнавать, поэтому требуется система реагирования на отказы. Под отказом в дальнейшем понимается незапланированное ухудшение качества какой-либо IT-услуги или полное ее отключение, вызванное программными или аппаратными ошибками. Организация данной системы позволяет добиться следующих преимуществ [2, 3]:

- быстрее восстановить работоспособность IT-услуги;
- получить подробный отчет о случившемся отказе;
- сформулировать полезные рекомендации по недопущению подобных отказов в будущем;
- представить полезные рекомендации по быстрому восстановлению после отказов в будущем;

- создать базу данных с информацией об отказах и способах устранения последствий;
- уменьшить негативные последствия после отказа.

Процедура реагирования на отказ, как и другие компоненты системы управления МГРЦ, реализуется в виде автономного программного агента, который выполняется в изолированной операционной среде (например, в отдельной виртуальной машине или контейнере) [4]. Достоинствами такого решения являются:

- возможность разработчика каждого компонента самому выбирать необходимые ему программные библиотеки, так как некоторые библиотеки конфликтуют между собой при попытках их использования в общей операционной среде;
- относительно легкое перемещение агентов с одного физического сервера на другой;
- возможность размещения и работы агентов на физически разных серверах.

Для взаимодействия друг с другом в асинхронном режиме программные агенты СИУ отправляют сообщения через очередь протокола AMQP, реализованной брокером RabbitMQ.

В качестве системы управления базами данных (СУБД) выбрана свободно распространяемая кроссплатформенная объектно-реляционная СУБД PostgreSQL. Она базируется на языке SQL, поддерживает многие возможности этого языка, большой набор типов данных, предлагает множество современных функций: сложные запросы, внешние ключи, триггеры, изменяемые представления, транзакционная целостность.

Рассмотрим подробнее схему взаимодействия программных компонентов с процедурой реагирования на отказ на примере поломки сервера.

В базе данных вычислительных ресурсов и услуг (центральная база данных, БДУ) имеются две значимые таблицы – список устройств в ЦОД и список выделенных виртуальных ресурсов (например, VM или ВХД). Схема представлена на рисунке.

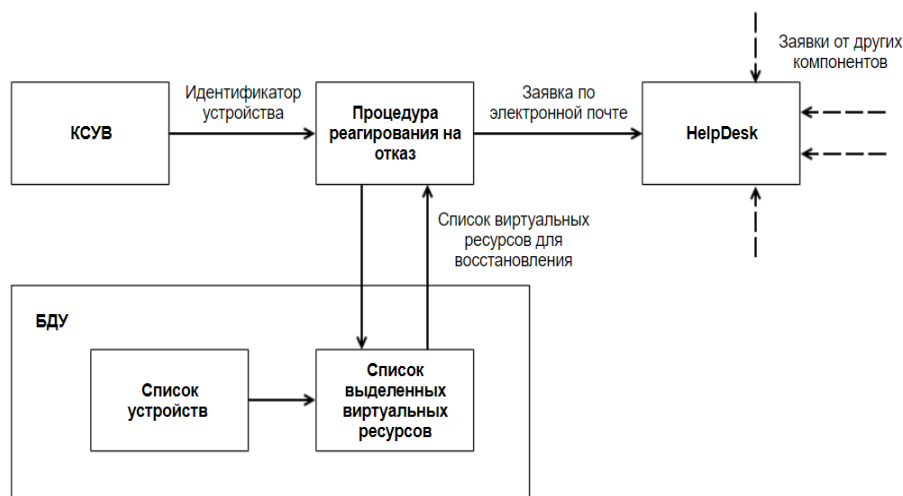


Рисунок. Схема процедуры реагирования на отказ

В записях каждого выделенного виртуального ресурса есть сведения о физическом устройстве, на котором оно работает.

КСУВ – это ядро для управления OpenStack и остальной виртуализацией. Когда КСУВ будет оповещен, что что-то вышло из строя, он передает агенту процедуры оценки ущерба идентификатор устройства (сервера), которое не работает.

Затем происходит обращение в таблицу выделенных виртуальных ресурсов и поиск там тех, которые связаны с этим устройством, т.е. у которых в соответствующем столбце содержится тот же идентификатор. Программа выдает список виртуальных ресурсов, работа которых нарушена, и формирует отчет в формате PDF. Этот отчет содержит информацию о дате и времени отказа IT-услуги, кем отправлено, подробное

описание проблемы, сведения о пострадавшей инфраструктуре, возможное решение. Все эти сведения важны для выяснения причин произошедшего отказа. Далее отправляется письмо по электронной почте в систему поддержки IT-инфраструктуры. По этим мейлам создаются заявки, и автоматически назначаются на них заранее определенные администраторы, которые получают уведомление об этом на свою почту. Для решения проблем администратор пользуется рекомендациями из отчета и базы знаний, дополняя ее по мере нахождения решений новых сбоев. Также в систему поддержки могут посылаться мейлы от других компонентов СИУ, таким образом, будет иметься централизованная система работы с заявками и контроля за ошибками во всем программном комплексе.

Для выбора системы поддержки IT-инфраструктуры необходимо определить требования к ней:

- бесплатная система;
- оповещение по электронной почте;
- учет оборудования;
- FAQ и база знаний;
- простой и понятный интерфейс.

Система GLPI (Gestionnaire Libre de Parc Informatique), являющаяся системой работы с заявками и инцидентами, удовлетворяет этим требованиям, имеет множество полезных функций для администрирования: учет инвентаризации, управление статусом оборудования, назначение задач и планирование. Также имеются активные англоязычные и русскоязычные форумы, где обсуждается много идей по реализации данной системы, и имеется много дополнительных модулей, дополняющих функциональность системы. Возможность оставлять заявки не только через электронную почту, но и посредством веб-интерфейса позволяет расширить функционал данной системы [5].

Таким образом, принимать электронную почту будет система GLPI, являющаяся комплексной системой управления IT-инфраструктурой. Программное обеспечение написано на языке PHP. Система GLPI может быть развернута как на открытых операционных системах типа Linux и BSD, так и на коммерческих системах типа Windows. Была выбрана именно эта система, так как в сравнении с такими системами, как OSTicket и OTRS, GLPI имеет встроенные возможности учета оборудования и активное сообщество, готовое помочь по любым вопросам, связанным как с настройкой, так и с эксплуатацией системы.

В данной работе содержится краткое описание метода реализации процедуры реагирования на аварийные ситуации в географически распределенном ЦОД. Был разработан программный агент, который получает информацию об отказавших устройствах, виртуальных ресурсах, которые на нем были, и через систему поддержки пользователей уведомляет администратора об аварийной ситуации. В результате исследований была выбрана система поддержки пользователей GLPI, но возможно использовать и другие системы. GLPI позволяет выполнять управление заявками и инцидентами, поддерживает базу знаний и FAQ, имеет систему оповещения о событиях. В дальнейшем планируется выполнение некоторых автоматических действий для восстановления виртуальных ресурсов после сбоя, чтобы реагирование происходило быстрее.

## Литература

1. Shevel A.Y., Khoruzhnikov S.E., Grudinin V.A., Sadov O.L., Kairkanov A.B. Globally distributed software defined storage (proposal) // Journal of Physics: Conference Series. – 2017. – V. 898. – № 6. – Art. 062014.

2. Дегтярева А.В. Выявление необходимости и особенности внедрения Service desk в организации // Международный студенческий научный вестник. – 2016. – № 6. – С. 81.
3. Горина Е.В. Service desk как важный элемент функционирования отдела информационных технологий современного предприятия // Актуальные научные исследования в современном мире. – 2016. – № 7-1(15). – С. 77–79.
4. Fedchenkov P.V. et al. The designing of cloud infrastructure consisting of geographically distributed data centers // Proceedings of the VII International Conference «Distributed Computing and Grid-technologies in Science and Education». – 2018. – P. 32–36.
5. Documentation GLPI [Электронный ресурс]. – Режим доступа: <https://glpi-project.org/resources/> (дата обращения: 06.03.2019).



**НАПРАВЛЕНИЕ  
ИНСТИТУТА ДИЗАЙНА И УРБАНИСТИКИ**

**Голубев Кирилл Владимирович**

Университет ИТМО, институт дизайна и урбанистики, аспирант

Направление подготовки: 09.06.01 – Информатика и вычислительная техника

e-mail: golubev1251@gmail.com

**Карсаков Андрей Сергеевич**

Год рождения: 1985

Университет ИТМО, институт дизайна и урбанистики, к.т.н., ст. преподаватель

e-mail: karsakov@itmo.ru

**УДК 004.92****ВИЗУАЛЬНЫЙ АНАЛИЗ И ВАЛИДАЦИЯ РЕЗУЛЬТАТОВ ОБРАБОТКИ  
ГЕОПРИВЯЗАННЫХ ДАННЫХ****Голубев К.В.****Научный руководитель – к.т.н. Карсаков А.С.**

При машинном анализе больших объемов данных не всегда возможно быстро определить корректность работы примененного алгоритма, а также достоверность входных или полученных в результате расчета данных. Решением в этой ситуации может оказаться визуальное представление данных и последующий их анализ при помощи когнитивных способностей человека. В данной работе представлен сценарий использования приложения для визуализации и визуального анализа результатов машинной обработки данных компании-партнера с целью исследования и валидации полученных данных.

**Ключевые слова:** научная визуализация, геопространственная визуализация, визуальный анализ, компьютерная графика, интерактивная визуализация.

Важную часть современной науки занимают направления, связанные с обработкой, анализом и исследованием больших объемов данных. Часто такие исследования посвящены сбору информации о пользователях социальных сетей и сервисов, клиентах организаций, обезличенному анализу действий в сети Интернет или финансовых транзакций. К сожалению, при проведении таких исследований, а также при разработке алгоритмов и методов анализа таких данных, не всегда возможно провести автоматическую валидацию корректности работы алгоритмов и результатов проведенного анализа. При работе с такой задачей важно находить закономерности в отхождении от корректных значений и одиночные выбросы, с чем человеческий мозг в общем случае справляется лучше. Однако с другой стороны, большой объем данных и их однородность создают высокую когнитивную нагрузку на исследователя, что затрудняет такой анализ. Для решения этой проблемы часто используются технику интерактивной визуализации, позволяющей упорядочить и организовать информацию, представив ее в более естественном для человеческого восприятия виде [1]. Кроме того, интерактивность визуализации позволяет еще сильнее уменьшить поток воспринимаемых в единый момент данных, ограничивая детализацию информации и выдавая подробности по запросу.

В ходе исследования, проводимого нашими коллегами, был проведен анализ сети отделений компании-партнера, занимающихся работой с клиентами. Целью исследования было определить отделения, закрытие которых повысило бы загруженность оставшихся офисов, при этом не приводя к неудобству клиентов. Для этой цели были собраны данные о местах жительства, работы и точках интереса клиентов, после чего был проведен анализ этих данных с целью кластеризации точек интереса и определений возможных направлений перехода клиентов между офисами в случае закрытия некоторых из них [2]. Задачей нашей лаборатории стало разработать приложение-визуализатор, позволяющее отображать исходные и расчетные данные,

давая пользователю возможность проверить соответствие исходных данных исследования и результатов анализа. Общая совокупность исходных и аналитических данных, отображаемых визуализатором для конкретного отделения представлена на рис. 1, а.



Рис. 1. Совокупность и охват данных, используемых при проведении анализа (а); общий вид приложения-визуализатора (б)

Для визуализации был использован графический движок Fusion Engine [3], разработанный нашей научной группой и неоднократно использовавшийся во многих других проектах. Этот движок позволяет отображение геопривязанных данных с двойной точностью, что дает возможность точного позиционирования отображаемых объектов даже в случае охвата визуализацией больших территорий. Визуализация включала в себя четыре основных слоя: картографическая подложка, визуализация точечных геопривязанных данных, визуализация потоков (перехода клиентов между отделениями), а также информационные и управляющие панели пользовательского интерфейса.

Рассмотрим подробнее устройство визуализатора. На рис. 1, б, представлен основной экран приложения. Боковая панель интерфейса позволяет пользователю переключать основные слои отображения: отделения компании-партнера, отделения конкурентов, данные о местах жительства, работы и точках интереса клиентов. Это позволяет отфильтровать отображение, оставив только необходимые для текущей задачи данные. Ниже можно выбрать сценарий, определяющий количество выбранных для закрытия отделений, для которых отображается информация по переходу клиентов в другие офисы. При этом точки, отображающие на карте кандидатов на закрытие, выделяются другим цветом.

При выборе конкретного отделения при помощи мыши, пользователь получает возможность отобразить дополнительные данные конкретно по этому отделению. На боковой панели интерфейса появляется обезличенный список привязанных к отделению клиентов, в котором можно выбрать одного человека или группу, просмотрев для выбранных людей оригинальные точки интереса и центры полученных в результате анализа кластеров. Кроме того, в случае если это отделение входит в список кандидатов на закрытие, отдельными цветами выделяются отделения, в которые переходят клиенты. Также есть возможность отображения перехода клиентов в другие офисы организации. Пример отображения такой детализации представлен на рис. 2, а.

Использование визуализатора позволило еще во время работы над проектом обнаружить различные ошибки и недостатки исходного исследования и своевременно их исправить.

Во-первых, была обнаружена неполнота охвата исходных данных. Так, в выборку не вошли данные клиентов, проживающих за пределами Санкт-Петербурга, в частности, в поселках и новых микрорайонах, расположенных сразу за

административной границей города, при том, что организация имеет офисы и в области (рис. 2, б). Своевременное обнаружение этой ошибки позволило собрать недостающие данные и провести повторный расчет, включивший в анализ также клиентов из Ленинградской области.

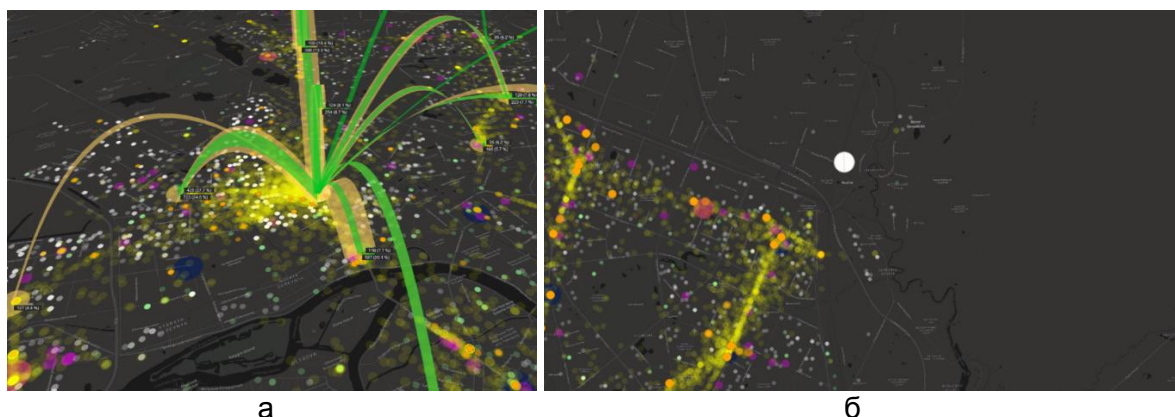


Рис. 2. Отображение данных выбранного отделения (а); недостаток исходных данных по клиентам за пределами Санкт-Петербурга (б)

Другой важной проблемой стали ошибки кластеризации точек интересов. В некоторых случаях алгоритм объединял в один кластер точки с разных концов города, возвращая в качестве результата центры интересов, не связанные с исходными точками. Эту проблему было сложно обнаружить, работая с данными, в то время как в ходе визуального анализа были сразу замечены ошибочные центры кластеров, расположенные в Финском заливе или в реках. Это позволило вовремя выявить ошибку в параметрах алгоритма и исправить их, получив более точное разбиение.

В качестве другого примера – ситуации, еще менее очевидной в отрыве от визуализации, стали ошибки в анализе шаговой доступности, возможно, связанные с ошибками кластеризации. Из-за таких ошибок алгоритм предполагал переход клиентов между отделениями, сильно не соответствующий действительности, переводя массы клиентов в отделения на другом конце города или в области. Отображение потоков перехода клиентов между отделениями позволило легко обнаружить такие проблемы. Примеры некорректного анализа геодоступности представлены на рис. 3.

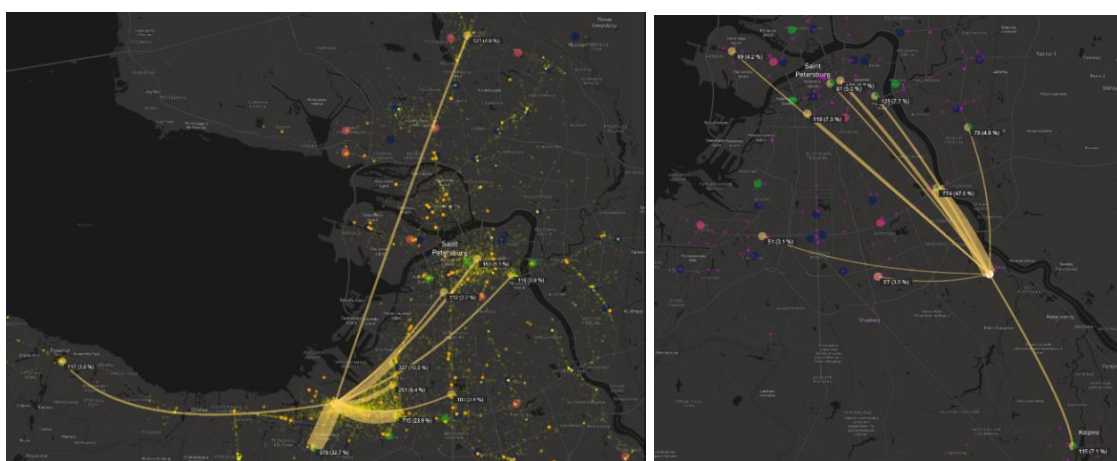


Рис. 3. Ошибки в определении геодоступности

Во всех перечисленных случаях для выявления ошибки в автоматическом режиме пришлось бы проводить множество вычислений и собирать дополнительные данные.

В ходе данного исследования было разработано приложение для визуального анализа и валидации геопривязанных данных. Использование данного приложения

позволило своевременно выявить и предотвратить ошибки в алгоритмах и параметрах анализа таких данных, что поспособствовало повышению качества исследования наших коллег. Представленное исследование показывает важность использования человеческих когнитивных способностей при анализе данных и полезность применения инструментов научной визуализации.

### **Литература**

1. Brodbeck D., Mazza R., Lalanne D. Interactive Visualization-A Survey // Hum. Mach. Interact. – 2009. – P. 1–21.
2. Zaikin O. et al. Optimizing spatial accessibility of company branches network with constraints // Proc. Int. Conf. Comput. Sci. – 2019. – V. 11537. – P. 332–345.
3. Bezgodov A. et al. The Framework for Rapid Graphics Application Development: The Multi-scale Problem Visualization // Procedia Comput. Sci. Elsevier. – 2015. – V. 51. – P. 2729–2733.

**Деева Ирина Юрьевна**

Год рождения: 1994

Университет ИТМО, институт дизайна и урбанистики, аспирант

Направление подготовки: 09.06.01 – Информатика и вычислительная техника

e-mail: iriny.deeva@gmail.com

**Калюжная Анна Владимировна**

Год рождения: 1990

Университет ИТМО, институт дизайна и урбанистики, к.т.н., доцент

e-mail: kalyuzhnaya.ann@gmail.com

**УДК 004.942****ПРОГНОЗИРОВАНИЕ ЛИЧНОСТНЫХ МЕТРИК ДЛЯ КОНТЕКСТНО-НЕЗАВИСИМОГО ЦИФРОВОГО ОБРАЗА ПОЛЬЗОВАТЕЛЯ СОЦИАЛЬНОЙ СЕТИ****Деева И.Ю.****Научный руководитель – к.т.н., доцент Калюжная А.В.**

Процесс оцифровки объектов и операций реального мира довольно активен, что предполагает создание их цифровых образов или модели. Люди регулярно оставляют достаточно своих данных в социальных сетях и на разных сайтах, формируя таким образом свой уникальный цифровой след. На основе полученного цифрового следа можно создать полную цифровую сущность человека в гиперпространстве социальных медиа. Однако человек – это сложная система, следовательно, модель цифрового объекта должна быть многомасштабной. В работе исследована связь такой слабо формализуемой стороны пользователя, как психометрические показатели и его цифрового следа в социальной сети. Для этого были исследованы следующие вопросы. Должен ли психометрический прогноз рассматриваться как задача прогнозирования с несколькими откликами? Какие психометрические черты предсказываются одинаково хорошо в разных онлайн-средах? Каково снижение точности при расширении модели с примерами обучения из других социальных сетей?

**Ключевые слова:** социальные сети, психометрики, многомерная линейная регрессия, одномерная линейная регрессия, кросс-медиа обучение.

Сайты социальных сетей предоставляют уникальную возможность для персонализированных сервисов исследовать поведение пользователей. В дополнение к структурированной информации пользователей, содержащейся в их профилях, такой как биографические данные, пользователи создают различные объемы личных данных различными способами, включая текстовое содержимое (например, обновления статуса, сообщения в блоге, комментарии) или данные предпочтений (лайки и дизлайки). Многие скрытые характеристики, такие как психотип, эмоции и настроение, которые, как правило, явно не установлены пользователями, могут быть извлечены из пользовательского контента.

В работе авторами выполнено исследование в области автоматического прогнозирования личности с использованием данных из социальных сетей. Это новая область, которая привлекает внимание исследователей из-за ее потенциала во многих компьютерных приложениях. Например, это может быть полезно для выявления отклоняющегося поведения пользователей в социальных сетях или для обнаружения депрессивных наклонностей личности.

В работе была использована модель «Big Five» как модель личности, включающую пять психометрик: открытость к опыту, добросовестность (сознательность), экстраверсия, доброжелательность и эмоциональная стабильность (часто наоборот, называется невротизмом). Основная цель исследования состояла в том, чтобы изучить возможность разработки моделей, способных прогнозировать

каждый психометрический показатель на основе данных профиля пользователя в социальных сетях.

В этом исследовании анализ и оценка текущего состояния существующих методов и технологий идентификации личности выполнялись на разнообразном наборе данных социальных сетей, собранных в Facebook (американская социальная сеть) и ВКонтакте (русская социальная сеть). Цель состояла в том, чтобы ответить на следующие вопросы.

Должен ли психометрический прогноз рассматриваться как задача прогнозирования с несколькими откликами? Основываясь на информации профиля пользователя авторам необходимо было получить набор из пяти оценок (действительных чисел по пятибалльной шкале), представляющих измерения метрик Big Five. Эта задача может рассматриваться как проблема регрессии, которая включает в себя различные методы одномерной и многомерной регрессии. Формально одномерная и многомерная регрессии решают следующую проблему: пусть  $F$  будет входным пространством, состоящим из векторов со значениями для  $m$  атрибутов  $f_1, f_2, \dots, f_m$ , и пусть  $T$  будет выходным пространством, состоящим из векторов со значениями для  $n$  целевых переменных  $t_1, t_2, \dots, t_n$ . Цель алгоритма многомерной регрессии – изучить  $M: F \rightarrow T$  модель, минимизирующую ошибку прогноза в обучающей выборке. В этом исследовании  $n=5$ , где  $t_1$  – экстраверсия;  $t_2$  – доброжелательность;  $t_3$  – добросовестность;  $t_4$  – эмоциональная стабильность/невротизм;  $t_5$  – открытость. И  $m=4$ , где  $f_1$  – количество друзей;  $f_2$  – общее количество лайков в профиле;  $f_3$  – общее количество сообщений в профиле;  $f_4$  – общее количество комментариев в профиле. Проведено сравнение методов многомерной регрессии, таких как объединение нескольких моделей регрессии, ансамбль цепочки регрессоров, с одномерным подходом, таким как случайный лес. В качестве базового решения для каждой точки данных рассмотрено среднее значение на обучающих данных. В качестве показателя качества прогноза RMSE-метрика была использована.

Какие психометрические черты предсказываются одинаково хорошо в разных онлайн-средах? Авторы оценивают силу взаимосвязи между различными прогностическими признаками и личностными чертами, выявляя их взаимосвязь. Сравняются результаты корреляции по двум наборам данных. Цель состояла в том, чтобы определить, какие психометрические характеристики предсказаны одинаково хорошо для двух наборов данных из двух разных сайтов социальных сетей. Это в первую очередь необходимо для анализа характера проявления определенных психометрических признаков, которые могут быть полезны для кластеризации пользователей в пространстве прогнозируемых значений.

Каково снижение точности при расширении модели с примерами обучения из других социальных сетей? Прогнозирование личных характеристик – довольно сложная задача. В отличие от демографических данных, результаты социальных опросов являются относительно скудной информацией и измеряются со значительной ошибкой. Кроме того, статистику по профилям пользователей социальных сетей также сложно собирать, поскольку в большинстве случаев эта информация является конфиденциальной [1]. Предложили кросс-медиа обучение или разработку моделей прогнозирования личности с использованием различных цифровых медиа. Преимущество кросс-медиа обучения в том, что примеры обучения из разных социальных сетей могут быть объединены для повышения точности. В этом исследовании была изучена возможность перекрестного обучения прогнозированию личности с использованием наборов данных из двух разных сред (Facebook и ВКонтакте).

Недавно было показано, что цифровой след пользователей может использоваться для автоматической идентификации их личности. Например, [2, 3] показали, что автоматические суждения о человеке, основанные на лайках в Facebook, более точны, чем оценки, сделанные их друзьями. Многие другие подходы были предложены с использованием различных механизмов прогнозирования, пространств признаков и ориентации на различные онлайн-среды [4, 5]. Знание личности человека позволяет нам делать прогнозы относительно предпочтений и интересов в разных контекстах и средах и совершенствовать различные системы рекомендаций [5]. Личность может влиять на процесс принятия решений, и было показано, что это влияет на предпочтения в выборе веб-сайтов [5], продуктов, брендов и услуг, а также контента, такого как фильмы, телешоу и книги [3].

В исследованиях, представленных в данной работе, использованы два набора данных, собранных с самых популярных сайтов социальных сетей. Facebook – это американская социальная сеть, а ВКонтакте – это российская социальная сеть. Набор данных Facebook общедоступен [1]. Для социальной сети ВКонтакте не было свободного доступа к каким-либо наборам данных.

- Набор данных Facebook. Что касается проблемы прогнозирования личности, существует не так много наборов данных с сайтов социальных сетей. Основная причина заключается в том, что сбор помеченных данных занимает много времени и стоит дорого. Кроме того, такого рода информация может быть конфиденциальной. Для сбора личной оценки в этом исследовании использовался подход, при котором пользователи сайтов социальных сетей добровольно сами отвечали на вопросники. MyPersonality [1] – это популярное приложение Facebook, появившееся в 2007 году, которое позволяет пользователям проходить серию психометрических тестов, в том числе стандартную анкету «Big Five» [2].
- Набор данных ВКонтакте. Для социальной сети ВКонтакте было гораздо сложнее собирать данные, потому что не существует открытого набора пользовательских данных. Поэтому для этого исследования нам пришлось собирать данные самостоятельно. Прежде всего, чтобы собрать психометрическую информацию, авторами была распространена ссылка [1] на прохождение вопросника Big Five, и также попросили каждого участника предоставить ссылку на их профиль в ВКонтакте. После этого прогенерировали файлы JSON со стандартными полями VK API, включая информацию о друзьях, лайках, комментариях и постах. Следует отметить, что участие в исследовании было добровольным, и личная информация участников не разглашалась. Было опрошено 107 участников. Размер выборки довольно мал, но количество людей, которые предоставили свою информацию, постоянно увеличивается с момента сбора информации, а прохождение опросов – постепенный процесс.

Авторы использовали основные методы машинного обучения для выбора модели регрессора, включая метод наименьших квадратов, метод построения случайных лесов, метод  $K$ -ближайших соседей, логистическую регрессию и метод опорных векторов. Для оценки качества работы модели использовалась такая метрика, как коэффициент детерминации ( $R$ -квадрат). Как видно из рисунка для обоих наборов данных, лучший результат показал метод случайных лесов. Вот почему эта модель была выбрана в качестве модели регрессора. Каждый набор данных был разделен в соотношении 30% тестовых данных и 70% обучающих данных, чтобы избежать переобучения моделей. Модель использовалась со следующими гиперпараметрами: `nestimators = 100`; `max features = 'sqrt'`. Где `nestimators` – это количество деревьев в лесу, а `max features` – это количество признаков, которые следует учитывать при поиске наилучшего разделения.



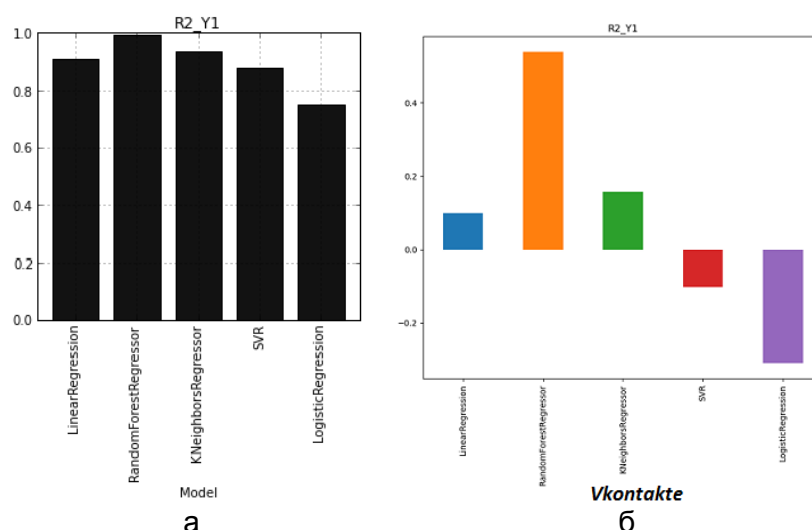


Рисунок. Результаты исследования R2 coef Facebook (а) и ВКонтакте (б)

Интересное наблюдение состоит в том, что многомерные регрессионные подходы не всегда превосходят однофакторный, но в большинстве случаев они дают лучшие результаты. Однако различия между одномерными и многомерными регрессорами незначительны. В качестве одномерной регрессии были использованы ST – Single target regression, MTS – Multi target stacking, MTSC – Multi target stacking corrected, ERC – Ensemble of regressor chains, ERCC – Ensemble of regressor chains corrected. Результаты применения моделей для обоих наборов данных представлены в таблице.

Таблица. Результаты метрики RMSE для разных моделей

Facebook					
Модель	Экстраверсия	Доброжелательность	Сознательность	Невротизм	Открытость
ST	0,78	0,691	0,81	0,625	0,741
MTS	0,79	0,73	0,82	0,67	0,751
MTSC	0,78	0,691	0,83	0,61	0,741
ERC	0,77	0,691	0,81	0,62	0,741
ERCC	0,77	0,691	0,81	0,61	0,741
ВКонтакте					
ST	0,171	0,154	0,168	0,181	0,207
MTS	0,173	0,152	0,168	0,182	0,224
MTSC	0,173	0,154	0,167	0,181	0,207
ERC	0,173	0,154	0,168	0,181	0,207
ERCC	0,171	0,155	0,167	0,181	0,225

Во-первых, были использованы разнообразные методы одномерной и многомерной регрессии, как представлено. Используя алгоритм случайных лесов в качестве базовой модели для двух разных наборов данных, различия между одномерной и многомерной моделями были незначительными.

Затем проанализировали ошибки прогнозирования для различных личных характеристик и обнаружили, что экстраверсия и доброжелательность одинаково хорошо и последовательно прогнозируются в различных социальных средах. В будущем такое извлечение признаков может позволить осуществить кластеризацию в пространстве этих функций для идентификации конкретных групп пользователей.

И расширение модели с примерами обучения из другого источника не улучшило производительность обучаемых моделей. Что позволяет предположить, что контекст данных играет важную роль в успехе кросс-медиа обучения.

**Литература**

1. Farnadi G., Sitaraman G., Rohani M., Kosinski M., Stillwell D., Moens M., Davalos S., De Cock M. How are you doing? Emotions and personality in Facebook // Proceedings of the EMPIRE. – 2014. – P. 45–56.
2. Lambiotte R., Kosinski M. Tracking the digital footprints of personality // Proceedings of the Institute of Electrical and Electronics Engineers. – 2014. – P. 35–39.
3. Kosinski M., Stillwell, D.J., Graepel T. Private traits and attributes are predictable from digital records of human behavior // Proc. Natl. Acad. Sci. – 2013. – V. 110. – P. 5802–5805.
4. Cantador I., Fernández-Tobías I., Bellogín A., Kosinski M., Stillwell D. Relating personality types with user preferences in multiple entertainment domains // Proceedings of the 1st Workshop on Emotion and Personality in Personalized Services. – 2013. – P. 13–28.
5. John O.P., Srivastava S. The Big Five trait taxonomy: history, measurement, and theoretical perspectives // Handb. Pers. Theory Res. – 2008. – P. 132–138.

**Деревицкий Илья Владиславович**

Год рождения: 1992

Университет ИТМО, институт дизайна и урбанистики, аспирант

Направление подготовки: 09.06.01 – Информатика и вычислительная техника

e-mail: ivderevitckii@itmo.ru

**Ковальчук Сергей Валерьевич**

Год рождения: 1984

Университет ИТМО, институт дизайна и урбанистики, к.т.н., доцент

e-mail: sergey.v.kovalchuk@gmail.com

**УДК 004.852****СРАВНИТЕЛЬНЫЙ АНАЛИЗ ПОДХОДОВ К ВЫЯВЛЕНИЮ ПАТТЕРНОВ  
В ТЕЧЕНИЕ САХАРНОГО ДИАБЕТА 2 ТИПА****Деревицкий И.В.****Научный руководитель – к.т.н., доцент Ковальчук С.В.**

Работа выполнена в рамках темы НИР № 618267 «Интеллектуальные технологии гибридного предсказательного моделирования в задачах Р4-медицины».

Для здравоохранения сахарный диабет 2 типа представляет одну из наиболее приоритетных проблем, так как данное заболевание связано с большим числом сопутствующих заболеваний, приводящих к ранней инвалидизации и повышенному сердечно-сосудистому риску. Для целей ценностно-ориентированной медицины необходимы качественные системы поддержки и принятия решений врачам-диагностам, работающим с диабетиками. Такие системы должны быть основаны на знаниях о закономерностях в течении болезни. Все вышесказанное обуславливает необходимость поиска качественных методов выявления паттернов течения болезни из ретроспективной информации. В данном исследовании рассмотрены два подхода, основанные на методах кластеризации машинного обучения, проведен их сравнительных анализ.

**Ключевые слова:** диабет, нарушение толерантности к глюкозе, диабет инсулиннезависимый, машинное обучение, глюкоза плазмы в крови.

Диабет 2 типа одно из самых общих хронических заболеваний в мире [1–3]. Тема исследования диабета 2 типа крайне актуальна, по прогнозам аналитиков в условиях современного мира наблюдается тенденция к росту числа страдающих данной болезнью. Все вышесказанное объясняет актуальность данной темы.

Исследование выполнялось на данных из 6468 медицинских карт, предоставленных центром ФГБУ «НМИЦ им. В.А. Алмазова» Минздрава России.

Для целей ценностно-ориентированной медицины необходимы качественные системы поддержки и принятия решений врачам-диагностам, работающим с диабетиками. Существующие системы, например FINDRISK [4], не позволяют оценить вероятность конкретного диабетического осложнения в конкретный момент течения болезни. Такие системы должны быть основаны на знаниях о закономерностях в течении болезни. Все вышесказанное обуславливает необходимость поиска качественных методов выявления паттернов течения болезни из ретроспективной информации.

Первый подход выявления паттернов диабетических траекторий основан на методах кластеризации.

Разобьем все 6468 последовательности событий на кластеры с помощью Левенштейна и метода  $k$ -средних и попробуем выявить типовые закономерности. Продемонстрируем полученные результаты на рис. 1.

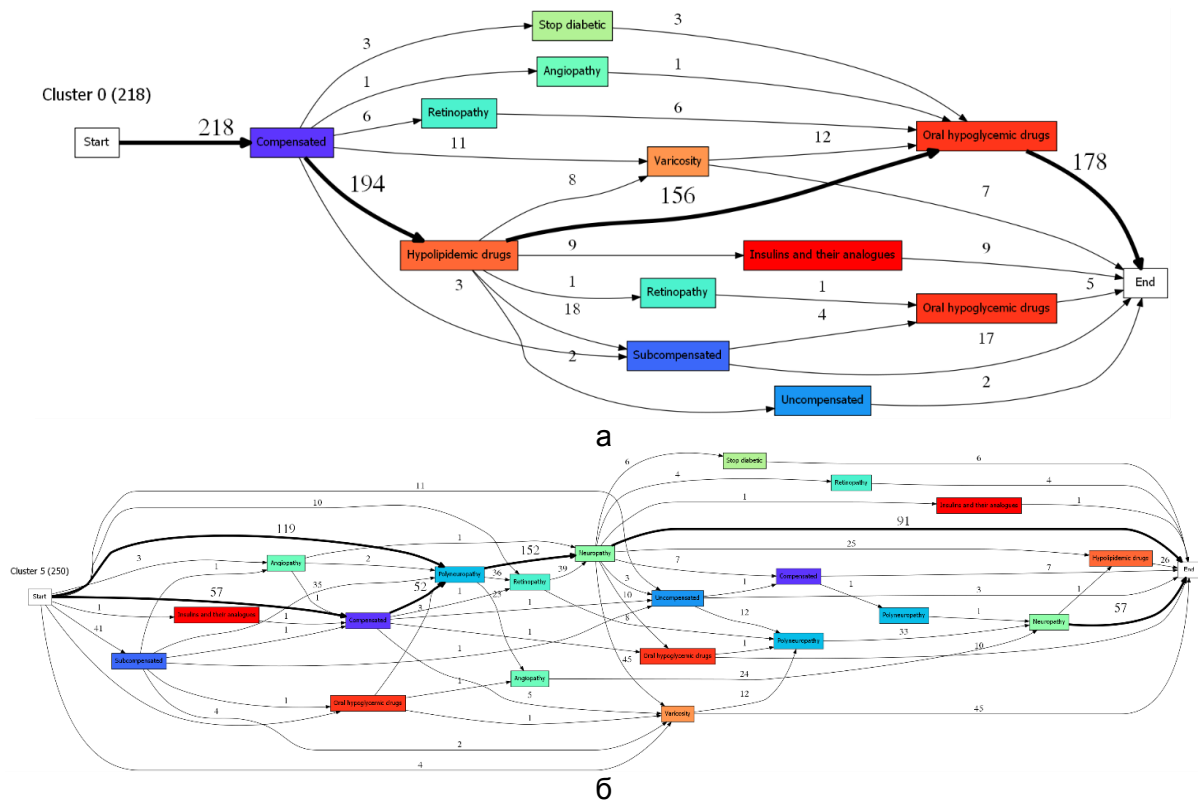


Рис. 1. Короткая схема наблюдения и лечения, кластер 0, 218 пациентов (а); длинная схема наблюдения и лечения, кластер 1, 250 пациентов (б)

У 218 человек наблюдалась короткая схема лечения – поступление – компенсация углеводного обмена. Далее усиленная терапия в виде гиполипидемических препаратов, после сахароснижающие пероральные таблетки, позже в сложных случаях инсулин-терапия и конец наблюдения. Пример выявленного паттерна – для пациентов данного кластера декомпенсация углеводного обмена крайне маловероятна в наблюдаемый период (меньше 5%).

На рис. 1 группа траекторий течения болезни для поступающих с двумя видами состояний диабета (компенсация углеводного обмена и субкомпенсация), и с осложнениями в виде полинефропатия (таких пациентов большинство). Наблюдается осложнение течения болезни в виде появления нефропатии, следует отметить отсутствие медикаментозной терапии в виде гиполипидемических или гипогликемических препаратов, состояние большинства пациентов меняется в лучшую сторону, наблюдение за частью пациентов прекращается, часть пациентов переходит в End конец наблюдения. Закономерности: отсутствует какая-либо терапия на ранних этапах для данных пациентов, необходимо исследовать целесообразность такого подхода.

Второй подход основан на кластеризации пациентов по признакам, влияющим на течение заболевания исходя из априорных медицинских сведений [5], и построении цепей Маркова.

Кластеризация была выполнена с помощью инструментов языка программирования Python, используя пакет sklearn. Выбранные параметры включали пол, возраст, ИБС, вес, индекс массы тела, Курение, максимальный уровень глюкозы в течение наблюдаемого периода.

Метод  $k$ -средних использовался для кластеризации. Количество кластеров было определено на основе максимального показателя соотношения меж кластерного расстояния к внутри кластерам и максимизации силуэта метрики. Оптимальное

количество кластеров было выбрано из интервала от 2 до 25 кластеров. Лучшие метрики показали разделение на 17 групп пациентов.

Качество кластеризации (удаленности кластеров друг от друга) было рассмотрено с помощью метрик из пакета sklearn-метрики. Метрика, которая подсчитывает «Силуэт» переменной для каждого объекта, показала 0,34. Значение силуэта каждого пациента показано на рис. 1, а. Цвета отмечают кластеры, полученные с помощью метода  $k$ -средних. Для визуализации кластеров в двумерном пространстве использовались методы (главная компонента), случайная проекция и  $t$ -SNE.  $t$ -SNE показал наилучшее визуальное разделение кластеров (рис. 1, б). Используя метод нелинейного уменьшения размера и визуализации многомерных  $t$ -распределенных данных ( $t$ -SNE), образцы 6 468 пациентов были переведены в двумерное пространство. Цвета соответствуют группам, полученным с помощью метода  $k$ -средств. Среднее внутрикластерное расстояние в евклидовой метрике составляет 1,01. Среднее межкластерное расстояние в евклидовой метрике составляет 4,8. Соотношение среднего внутрикластерного расстояния и среднего междукластерного составляет 0,22 (данное исследование является продолжением исследования Analysis of clusters of patients suffering from type 2 diabetes with Markov chains, статья подана на конференцию YSC, Greece, 2019.)

Рассмотрим примеры применения данного подхода (рис. 2).

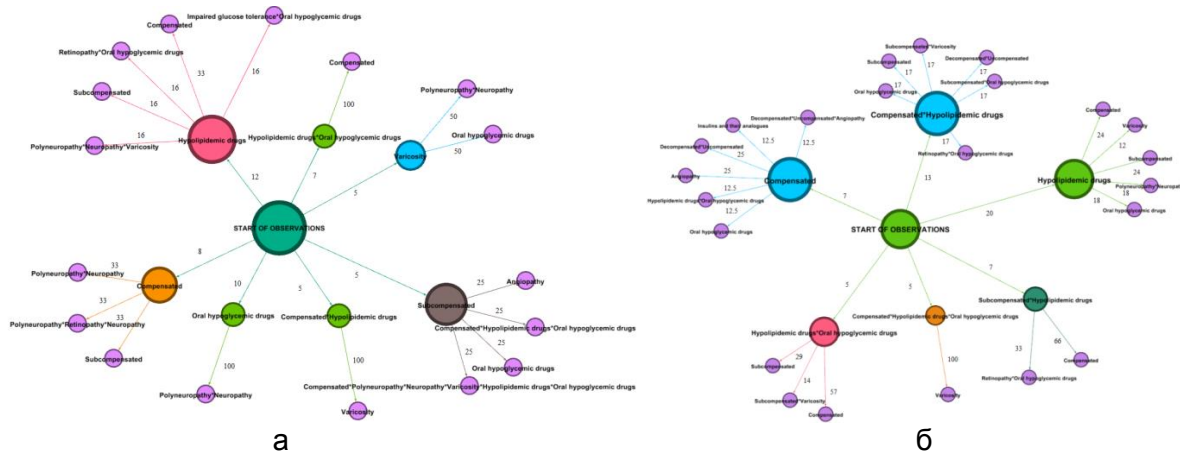


Рис. 2. Короткая схема начала наблюдения и лечения, кластер 7 (а); схема начала наблюдения и лечения, кластер 1 (б)

Для данных пациентов характерна компенсация углеводного обмена на раннем этапе болезни (который схематично изображен на рис. 2, а), после терапии сахароснижающими таблетками наблюдается появление осложнений, стоит исследовать данный вопрос подробнее.

Выше были рассмотрены два подхода к выявлению паттернов в течение сахарного диабета 2 типа. Проведем их сравнительный анализ.

Подход, основанный на кластеризации диабетических траекторий, основанный на расстоянии Левенштайна лучше выявляет закономерности в группах диабетических траекторий, так как в один кластер попадают пациенты со схожими с точки зрения метода  $k$ -средних и расстояния Левенштайна показателями. Однако данный метод слабо применим для целей прогноза дальнейшего течения болезни, в отличие от второго метода, где проводится кластеризация пациентов на основе медицинских данных, которые можно собрать при поступлении больного, метод также учитывает ретроспективную информацию.

**Литература**

1. Panagiotopoulos C. et. al. Type 2 diabetes in children and adolescents // *Canadian Journal of Diabetes*. – 2018. – V. 42. – P. S247–S254.
2. Chiasson J.L., Josse R.G., Gomis R., Hanefeld M., Karasik A., Laakso M. Acarbose for prevention of type 2 diabetes mellitus: The STOP-NIDDM randomised trial // *The Lancet*. – 2002. – V. 359(9323). – P. 2072–2077.
3. Collins G.S., Mallett S., Omar O., Yu L.M. Developing risk prediction models for type 2 diabetes: A systematic review of methodology and reporting [Электронный ресурс]. – Режим доступа: <https://bmcmmedicine.biomedcentral.com/articles/10.1186/1741-7015-9-103> (дата обращения: 06.03.2019).
4. Lindström J., Tuomilehto J. The diabetes risk score: A practical tool to predict type 2 diabetes risk [Электронный ресурс]. – Режим доступа: [https://pdfs.semanticscholar.org/ec58/3bf3b55250b42d803a9945f4fcdced3e6d52.pdf?\\_ga=2.150520886.1443912781.1570707202-1072790387.1570707202](https://pdfs.semanticscholar.org/ec58/3bf3b55250b42d803a9945f4fcdced3e6d52.pdf?_ga=2.150520886.1443912781.1570707202-1072790387.1570707202) (дата обращения: 06.03.2019).
5. Рекомендации по диабету, предиабету и сердечно-сосудистым заболеваниям. EASD/ESC // *Российский кардиологический журнал*. – 2014. – № 3(107). – С. 7–61.

**Калинин Александр Михайлович**

Год рождения: 1994

Университет ИТМО, институт дизайна и урбанистики, аспирант

Направление подготовки: 09.06.01 – Информатика и вычислительная техника

e-mail: amkalinin@niuitmo.ru

**Ваганов Данила Александрович**

Год рождения: 1995

Университет ИТМО, институт дизайна и урбанистики, студент группы № М4217с

Направление подготовки: 01.04.02 – Суперкомпьютерные технологии

в междисциплинарных исследованиях

e-mail: vaganov@itmo.ru

**Боченина Клавдия Олеговна**

Год рождения: 1988

Университет ИТМО, институт дизайна и урбанистики, к.т.н., доцент

e-mail: k.bochenina@gmail.com

**УДК 004.94****ВОССТАНОВЛЕНИЕ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОГО СТАТУСА  
ПОЛЬЗОВАТЕЛЕЙ ОНЛАЙН-СЕТЕЙ ПО ЧАСТИЧНО РАЗМЕЧЕННЫМ  
ДАНЫМ****Калинин А.М., Ваганов Д.А.****Научный руководитель – к.т.н., доцент Боченина К.О.**

Работа выполнена в рамках темы НИР № 618273 «Предсказательное моделирование динамики событий в социо-экономических системах».

В работе поставлена задача прогнозирования среднемесячного объема финансовых трат пользователя социальной сети с использованием графа дружеских связей. В качестве методов применяются статистическое реляционное обучение, графовые эмбединги, а также их комбинация. Предложенные методы применены к реальным деперсонализированным данным, содержащим сведения о взаимодействиях пользователей в социальной сети ВКонтакте и объеме их среднемесячных трат. Для различных методов экспериментально исследовано качество прогнозов «пропущенных» атрибутов при вариации доли известных меток. Комбинированный метод позволил улучшить качество прогнозов (до 4%), показав 64% правильных ответов в случае бинарной классификации и 51% при 3-классовой классификации.

**Ключевые слова:** социальные сети, графовые эмбединги, статистическое реляционное обучение, ассортативность, гомофилия.

**Введение.** Для исследования систем связанных объектов реального мира (транспортные, компьютерные, социальные и т.п.) применяется формализм комплексных сетей. Вершины комплексной сети могут дополнительно характеризоваться набором атрибутов (например, пол, возраст, вектор интересов для пользователей социальной сети). В сетях реального мира часть атрибутов вершин сети может быть недоступна вследствие специфики метода сбора данных или настроек приватности информации. В этом случае возникает задача восстановления значений атрибутов узлов по связям между вершинами.

Одной из движущих сил формирования социальных связей является гомофилия [1] – свойство, при котором люди с похожими интересами образуют связи друг с другом. В теории графов гомофилию можно охарактеризовать значением ассортативности по выделенному атрибуту [2]. Данная метрика характеризует наличие прямой корреляции значений атрибутов с соседними вершинами, при этом один и тот

же атрибут на разных структурах может иметь разную корреляцию. Пользователи социальных сетей имеют социо-демографические признаки, включающие в себя пол, возраст и социо-экономический статус. Последний атрибут представляет наибольшую сложность при классификации, в связи с отсутствием как открытых данных, так и актуальных исследований, в том числе описывающих специфику процесса формирования связей в социальных медиа на основе социо-экономического статуса. Цель данной работы заключалась в восстановлении значений атрибута, обозначающего среднемесячный объем финансовых трат пользователя социальной сети, используя данные об их финансовом состоянии и как информацию о непосредственных связях с другими пользователями, так и схожие глобальные топологические характеристики. Исследование позволит определить метод, наиболее подходящий для задачи восстановления информации о социо-экономическом статусе.

**Обзор предшествующих работ.** На настоящий момент прогнозированию пропущенных значений атрибутов в социальных сетях посвящено множество работ. Методы можно разделить на три типа:

1. Локальные модели. Любые алгоритмы, которые не используют связи между узлами графа [3, 4]. Обычно в качестве локальных моделей применяются традиционные алгоритмы машинного обучения, такие как логистическая регрессия и метод опорных векторов. Признаковое описание строится из атрибутов, не содержащих явных сведений о соседстве вершины.
2. Статистическое реляционное обучение (Statistical Relational Learning, SRL) [5], состоящее из локальной модели, реляционной модели и коллективного вывода [6]. Реляционные модели учитывают непосредственно соседние вершины узла и могут быть преобразованы из локальных моделей путем включения в признаковое описание необходимой информации. Коллективный вывод итеративно классифицирует вершины с помощью реляционной модели, тем самым имитируя влияния вершин друг на друга во всем графе. Локальная модель предоставляет априорные вероятности для инициализации коллективного вывода.
3. Графовые эмбединги (embeddings), проецирующие каждую вершину графа в числовой вектор. Ключевым элементом эмбедингов является функция сходства, определяющая близость двух вершин. Ряд эмбедингов [7, 8] преобразует вершины в вещественные вектора заданной размерности. Другие [9, 10] основаны на методах выделения локального окружения узлов и создают пространство высокой размерности, состоящее из булевых значений. Эмбединги позволяют лишь представить граф в виде признакового описания, для восстановления значений необходимо применять к полученным векторам локальные модели.

Корреляции между социальными связями и денежными тратами пользователей проанализированы в [11]. Восстановление социо-экономических атрибутов было проведено в ряде работ. В [12] прогнозируются профессия и доход пользователей Twitter. Авторы представляют сеть пользователей как двудольный граф и генерируют случайные блуждания от каждой вершины. К полученным моделям цепочек применяется модель Skipgram [13]. В дополнение, к признаковому описанию добавляются  $n$ -граммы, извлеченные из текстов пользователей. В работе [14] пользователи Twitter классифицируются по социальному статусу, который представлен тремя классами (низкий, средний, высокий). В качестве признаков используются информация о поведении пользователя (доли твитов/ретвитов, количество подписчиков/подписок) и  $n$ -граммы. В [4] прогнозируется род деятельности пользователя путем обучения классификатора на предварительно выделенных из текстов тематических кластерах слов.



**Методы.** Для решения поставленной задачи рассматриваются статистическое реляционное обучение и графовые эмбединги, так как данные методы позволяют учитывать связи между узлами графа. Также предложен альтернативный подход, который модифицирует метод, описанный в [6], путем объединения вышеперечисленных методов следующим образом:

1. извлечение эмбедингов из графа. Эмбединги могут быть извлечены как для размеченных, так и неразмеченных вершин. В данном исследовании применяются *node2vec* [7], *VERSE* [8] (с функциями сходства *PPR* [15], *SimRank* [16] и нормализованной матрицей смежности), *ARCTE* [9];
2. обучение классификатора. Обучение любого классификатора из области машинного обучения на извлеченных эмбедингах. В качестве классификатора в данной работе выбрана логистическая регрессия, реализованная в пакете *scikit-learn* [17];
3. реляционная модель. Применение полученных вероятностей принадлежности к классам как априорных на стадии инициализации реляционной модели. На данном этапе применяется модель взвешенного голосования по соседям (*Weighted-vote relational neighbor Classifier, wvRN*), так как она показала лучшее качество в работе [6] среди других моделей;
4. коллективный вывод. Итеративное применение реляционной модели для каждой вершины до определенного числа итераций  $T$ . В качестве коллективного вывода выбран метод ослабления меток (*Relaxation Labeling*), который по сравнению с отбором по Гиббсу (*Gibbs Sampling*) и итеративной классификацией (*Iterative Classification*) требует меньшего числа итераций и имеет возможность распараллеливания [6].

Для тестирования методов применяется кросс-валидация по 5 случайным разбиениям данных на обучающую и тестовую выборки. Также для проверки влияния количества размеченных вершин на качество классификации, кросс-валидация проводится для разбиений с размером обучающей выборки от 10% до 90%. В работе сравниваются как «чистые» эмбединги, на которых обучается логистическая регрессия, так и статистическое реляционное обучение на основе компонент, указанных выше. Последнее использует в качестве априорных вероятностей начальное распределение вершин, тогда как комбинированный подход применяет для этого оценки, полученные из обученного на эмбедингах классификатора. Для всех моделей используются параметры по умолчанию, за исключением количества итераций в комбинированном методе, так как в ходе экспериментов было обнаружено, что его лучшие показатели достигаются примерно на 15-й итерации, в связи с чем количество итераций было уменьшено с 99 до 15.

**Данные.** Для тестирования методов произведен сбор информации из профилей пользователей социальной сети «ВКонтакте», являющихся клиентами российского регионального банка. Все данные деперсонализированы согласно алгоритму, описанному в [18]. Информация включает в себя связи между пользователями, а также записи о транзакциях с указанием их суммы и времени.

Аналогично работам [11], среднемесячный объем трат (*Average Monthly Purchase, AMP*) определен как среднее число трат за каждый уникальный месяц. Так как снятие наличных с дебетовой карты не означает их обязательную трату, в данной работе рассмотрены два типа среднемесячного объема трат: со снятием *AMP* наличных и без (*AMP<sub>nw</sub>*).

Среднемесячный объем трат определен на множестве вещественных чисел. Задача классификации требует дискретных категориальных значений, поэтому признаки разделены следующим образом:

- AMP (binary) – разделение клиентов с малым (5–25 тыс.) и высоким (от 25 тыс.) объемом трат, включая снятие наличных;
- AMP (ternary) – 3 класса: малый (0–5 тыс.), средний (5–25 тыс.) и высокий (от 25 тыс.) объемы трат, включая снятие наличных;
- AMPnw (binary) – 2 класса: малый (5–20 тыс.) и высокий (от 20 тыс.);
- AMPnw (ternary) – клиенты разделены на три равных квантили: с малым (0–5 тыс.), средним (5–20 тыс.) и высоким (от 20 тыс.) объемами трат.

В таблице можно видеть, что значение ассортативности по рассматриваемым признакам относительно низкое для всех рассмотренных наборов данных, что говорит о слабой корреляции между пользователями и их друзьями в социальной сети.

Таблица. Характеристики выделенных графов

Граф	Количество узлов	Количество ребер	Ассортативность (корреляция Спирмена)	Количество классов
AMP (binary)	126291	546975	0,21	2
AMP (ternary)	180297	972591	0,19	3
AMPnw (binary)	106801	392429	0,19	2
AMPnw (ternary)	160945	775926	0,23	3

**Результаты.** Результаты экспериментов представлены на рис. 1–4. Для каждого признака приведен отдельный график с графовыми эмбедингами (рис. 1–4, а) и график с комбинированным методом (рис. 1–4, б) на основе выбранных эмбедингов. Во всех экспериментах VERSE с нормализованной матрицей смежности (см. *verse-neigh* на графиках) в качестве функции схожести показал низкие результаты в сравнении с другими эмбедингами, что может говорить о том, что данный метод построения эмбедингов не учитывает корреляцию объема трат и дружеских отношений. Среди «чистых» эмбедингов лучший результат принадлежит ARCTE, при этом он близок к показателю статистического реляционного обучения, которое, в свою очередь, превосходит (до примерно 5%) VERSE и *node2vec* при высоком проценте известных вершин. Более того, комбинированный с ARCTE метод превосходит остальные методы во всех случаях и позволяет добиться прироста до примерно 4% по сравнению с SRL. Комбинирование с остальными эмбедингами, включая VERSE-*neigh* (кроме случая с AMPnw (binary)), также повышает качество классификации и превосходит SRL, особенно при малом проценте известных вершин (10–50%). Максимальный прирост достигается в случае многоклассовой классификации, для бинарной качество улучшается при малом проценте размеченных вершин (менее 50%). Также в случае тернарной классификации проявляется неожиданный эффект – качество ухудшается, когда в графе размечено более половины вершин. Данный эффект требует дальнейшего изучения. Использование информации о снятии наличных помогает улучшить классификацию на 1% как в бинарном (63%/64%), так и в тернарном (50%/51%) случаях, так как такие данные отражают более полную картину о среднемесячных затратах клиентов.

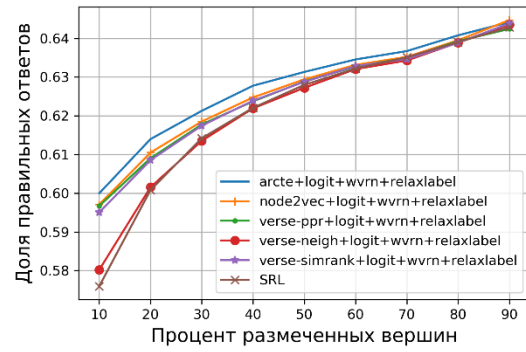
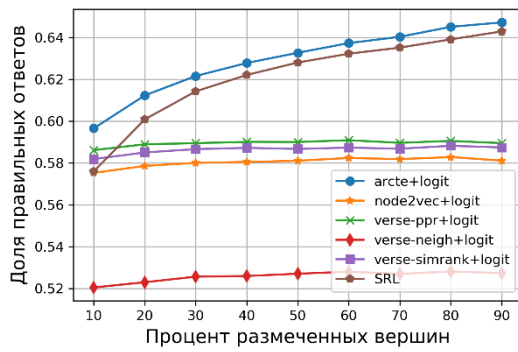


Рис. 1. Результаты классификации по признаку AMP (binary): графовые эмбединги (а); комбинированный метод (б)

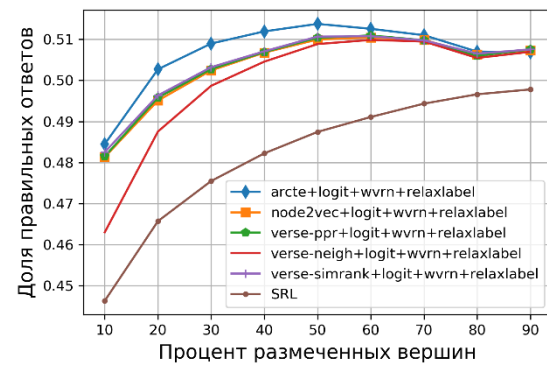
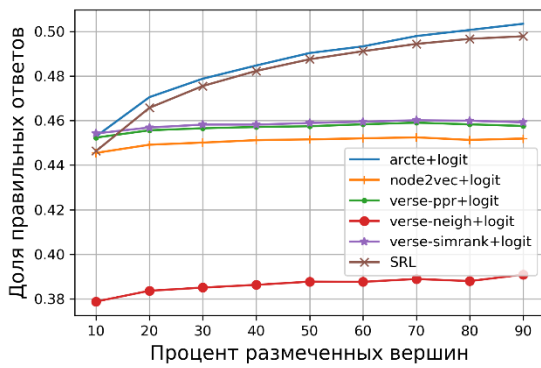


Рис. 2. Результаты классификации по признаку AMP (ternary): графовые эмбединги (а); комбинированный метод (б)

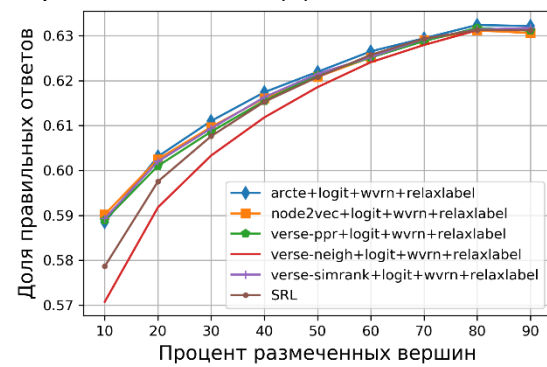
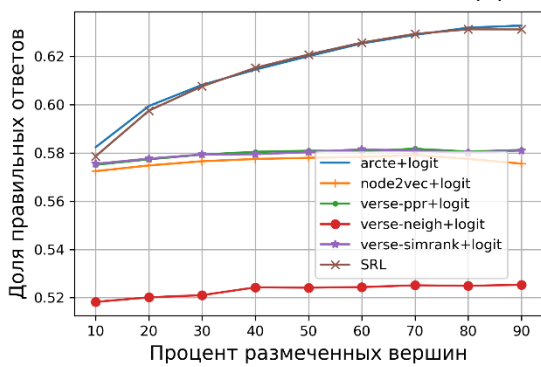


Рис. 3. Результаты классификации по признаку AMPnw (binary): графовые эмбединги (а); комбинированный метод (б)

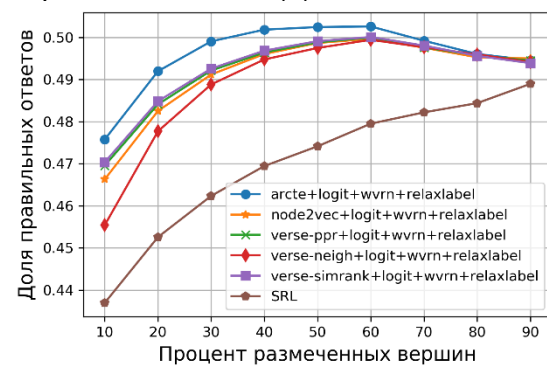
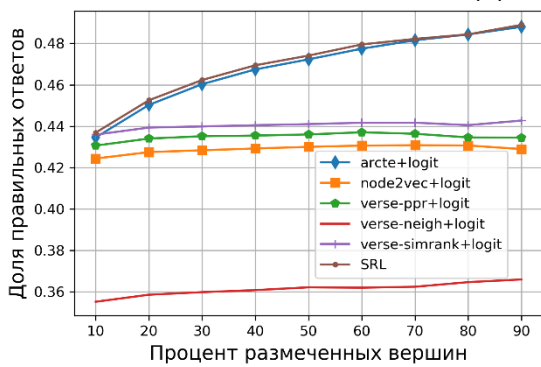


Рис. 4. Результаты классификации по признаку AMPnw (ternary): графовые эмбединги (а); комбинированный метод (б)

**Заключение.** В работе проведено исследование, посвященное классификации пользователей социальных сетей по их среднемесячному объему трат. Лучшее качество прогноза показывает метод ARCTE, основанный на выделении локальных сообществ пользователя, т.е. учитывающий не только прямые связи между соседями, но и другие похожие вершины графа. Следовательно, данный метод лучше учитывает скрытые топологические свойства графа, коррелирующие с процессами формирования дружеских связей на основе социо-экономического статуса пользователей. Обучение классификатора на эмбедингах с низкой размерностью позволяет получить оценки, устойчивые к изменению размера обучающей выборки, однако их качество ниже ARCTE. Возможно, это связано с тем, что классификатор не рассматривает окружение вершин и выявляет корреляции среди построенного ограниченного пространства вещественных признаков. Статистическое реляционное обучение позволяет улучшить оценки, выдаваемые классификатором; при этом также сокращается число итераций, необходимых для вывода. Максимальное преимущество (до 5%) комбинированного подхода достигается при 10% размеченных вершин графа. Таким образом, комбинированный метод восполняет информацию, которую не смогли извлечь эмбединги, статистическим реляционным обучением, учитывающим классы соседних вершин. Полученные результаты позволяют делать вывод о целесообразности применения графового эмбединга ARCTE, а также его объединения со статистическим реляционным обучением для прогнозирования социо-экономического статуса пользователей социальных сетей.

### Литература

1. McPherson M., Smith-Lovin L., Cook J.M. Birds of a Feather: Homophily in Social Networks // *Annual Review of Sociology*. – 2002. – V. 27. – P. 415–444.
2. Newman M.E.J. Mixing patterns in networks // *Physical Review E – Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*. – 2003. – V. 2(67). – P. 13.
3. Preo D. et al. Studying User Income through Language, Behaviour and Affect in Social Media [Электронный ресурс]. – Режим доступа: [https://www.researchgate.net/publication/282126053\\_Studying\\_User\\_Income\\_through\\_Language\\_Behaviour\\_and\\_Affect\\_in\\_Social\\_Media](https://www.researchgate.net/publication/282126053_Studying_User_Income_through_Language_Behaviour_and_Affect_in_Social_Media) (дата обращения: 06.03.2019).
4. Preoțiu-Pietro D., Lampos V., Aletras N. An analysis of the user occupational class through Twitter content // *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*. – 2015. – P. 1754–1764.
5. Getoor L., Taskar B. Introduction to statistical relational learning // *Journal of the Royal Statistical Society Series A (Statistics in Society)*. – 2010. – V. 173(4). – P. 934–935.
6. Macskassy S.A., Provost F. Classification in Networked Data A Toolkit and a Univariate Case Study // *Journal of Machine Learning Research*. – 2007. – V. 8. – P. 935–983.
7. Grover A., Leskovec J. node2vec: Scalable Feature Learning for Networks // *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. – 2016. – P. 855–864.
8. Tsitsulin A., et al. VERSE: Versatile Graph Embeddings from Similarity Measures [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1803.04742> (дата обращения: 06.03.2019).
9. Rizos G., Papadopoulos S., Kompatsiaris Y. Multilabel user classification using the community structure of online networks. – 2017. – P. 1–34.
10. Wang X., et al. Learning with multi-resolution overlapping communities // *Knowledge and Information Systems*. – 2013. – V. 36. – № 2. – P. 517–535.

11. Leo Y., Karsai M., Sarraute C., Fleury E. Correlations of consumption patterns in social-economic networks // Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. – 2016. – V. 1(8). – P. 493–500.
12. Aletras N., Chamberlain B.P. Predicting Twitter User Socioeconomic Attributes with Network and Language Information [Электронный ресурс]. – Режим доступа: [https://www.researchgate.net/publication/324472319\\_Predicting\\_Twitter\\_User\\_Socioeconomic\\_Attributes\\_with\\_Network\\_and\\_Language\\_Information](https://www.researchgate.net/publication/324472319_Predicting_Twitter_User_Socioeconomic_Attributes_with_Network_and_Language_Information) (дата обращения: 06.03.2019).
13. Mikolov T., Sutskever I., Chen K., Corrado G., Dean J. Distributed Representations of Words and Phrases and their Compositionality [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1310.4546> (дата обращения: 06.03.2019).
14. Lamos V., Aletras N., Geyti J.K., Zou B., Cox I.J. Inferring the Socioeconomic Status of Social Media Users Based on Behaviour and Language // ECIR 2016: Advances in Information Retrieval. – 2016. – P. 689–695.
15. Brin S., Page L. The PageRank Citation Ranking: Bringing Order to the Web [Электронный ресурс]. – Режим доступа: <http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf> (дата обращения: 06.03.2019).
16. Jeh G., Widom J. SimRank: A Measure of Structural-Context Similarity [Электронный ресурс]. – Режим доступа: <https://docviewer.yandex.ru/?lang=en&tm=1570733462&tld=ru&name=2001-41.pdf&text=Jeh+G.%2C+Widom+J.+SimRank%E2%80%AF%3A+A+Measure+of+Structural-Context+Similarity+2002&url=http%3A//ilpubs.stanford.edu%3A8090/508/1/2001-41.pdf&lr=2&mime=pdf&l10n=ru&sign=622ba5dc75588e91b045c54b099b1ffe&keyno=0> (дата обращения: 06.03.2019).
17. Pedregosa F. et al. Scikit-learn: Machine Learning in {P}ython // Journal of Machine Learning Research. – 2011. – V. 12. – P. 2825–2830.
18. Garfinkel S.L. De-identification of personal information [Электронный ресурс]. – Режим доступа: <https://docplayer.net/18976884-De-identification-of-personal-information.html> (дата обращения: 06.03.2019).

**Кокорев Андрей Дмитриевич**

Год рождения: 1995

Университет ИТМО, институт дизайна и урбанистики, аспирант

Направление подготовки: 09.06.01 – Информатика и вычислительная техника

e-mail: a.d.kokorev@yandex.ru

**Лебин Александр Эдуардович**

Год рождения: 1993

Университет ИТМО, факультет информационных технологий и программирования, аспирант

Направление подготовки: 09.06.01 – Информатика и вычислительная техника

e-mail: slebin@mail.ru

**Карсаков Андрей Сергеевич**

Год рождения: 1985

Университет ИТМО, институт дизайна и урбанистики, к.т.н., ст. преподаватель

e-mail: karsakov@itmo.ru

**УДК 004.942****МУЛЬТИАГЕНТНЫЕ ПОВЕДЕНЧЕСКИЕ МОДЕЛИ В ЗАДАЧАХ  
ПЕШЕХОДНОГО МОДЕЛИРОВАНИЯ****Кокорев А.Д., Лебин А.Э.****Научный руководитель – к.т.н. Карсаков А.С.**

Работа выполнена в рамках темы НИР № 618277 «Интеллектуальные технологии разработки и оптимизации поведенческих моделей для создания интерактивного опыта».

Моделирование поведения пешеходов в мультиагентных системах использует множество различных моделей, предназначенных для принятия решений в различных ситуациях. Существующие модели оперируют на смежных, но различных понятийных уровнях. В работе обобщаются уровни принятия решений, возникающие при моделировании пешеходов, обозначаются наиболее острые проблемы и обозначаются подходы, используемые для их решения.

**Ключевые слова:** пешеходное моделирование, мультиагентные системы, планирование пути, математическое моделирование, геометрическое представление агентов.

Задача моделирования поведения большого количества людей возникла в связи с попытками автоматизировать процесс проверки возможности эвакуации из больших сооружений. Существуют различные нормы, описывающие количество и расположение запасных выходов, однако для достижения наименьшего времени эвакуации в различных ситуациях необходимо проводить моделирование на этапе проектирования сооружения. Такое моделирование позволяет устранить недостатки разработанной планировки сразу на стадии проектирования сооружения, что значительно дешевле, чем вносить изменения в уже построенные конструкции.

Другая, не менее важная сфера применения моделей поведения толпы состоит в планировании загруженных транспортных узлов, обслуживающих большие количества людей, например железнодорожные вокзалы, аэропорты и станции метрополитена. При проектировании таких узлов очень важно правильно спроектировать расположение ключевых объектов, чтобы избежать заторов и столкновений потоков людей. Отличительная особенность этой сферы состоит в непрерывности потоков как прибывающих, так и убывающих людей, в отличие от сценария с эвакуацией, когда количество людей в сооружении убывает со временем. Эта особенность важна в ситуациях, когда образуются заторы, которые со временем только усугубляются из-за постоянной «подпитки» людьми.

**Системы моделирования.** На данный момент для моделирования поведения толпы в основном применяются мультиагентные модели. Существуют системы моделирования, предназначенные для изучения и сравнения различных моделей, такие как Menge [1] и MomentUMv2 [2]; также существуют системы, готовые для проведения моделирования, например, PTV Viswalk и Pedestrian Dynamic. Первая группа в основном состоит из систем с открытым исходным кодом, во вторую входят коммерческие продукты с закрытым исходным кодом.

На текущий момент широко распространена трехуровневая архитектура систем моделирования (рис. 1) с внешним анимационным уровнем. Важной особенностью такой архитектуры является независимость уровней друг от друга, что позволяет применять различные комбинации моделей каждого уровня.

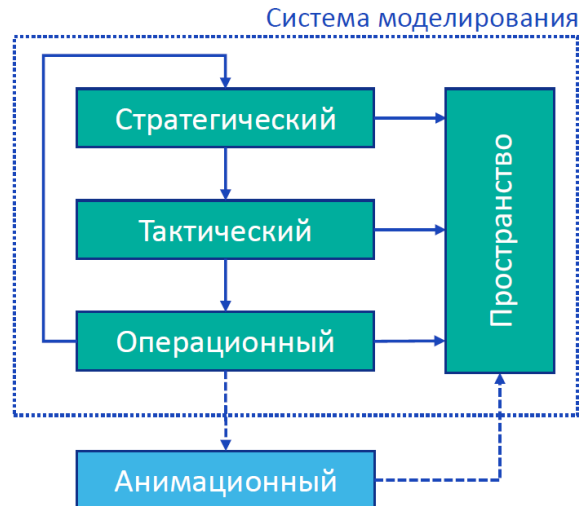


Рис. 1. Принципиальная архитектура системы моделирования

**Стратегический уровень.** На этом уровне происходит управление агентом с точки зрения цели его существования. Например, при моделировании железнодорожного вокзала на стратегическом уровне происходит решение задач, связанных с целью посещения: это может быть покупка билетов, посадка на поезд, встреча прибывающих и т.п. Также на этом уровне может производиться создание новых агентов и удаление агентов, выполнивших свою задачу и вышедших за пределы области моделирования.

**Тактический уровень.** На этом уровне происходит планирование пути до пункта назначения, назначенного на стратегическом уровне. Для поиска пути используются алгоритмы поиска пути в графе, NavMesh и другие подобные алгоритмы.

Основные проблемы, на решение которых направлены современные исследования на этом уровне, следующие:

- высокая вычислительная сложность классических алгоритмов поиска пути. Сложность алгоритма, приемлемая для однократного вычисления может быть слишком высокой в случае большого количества агентов, одновременно перемещающихся по пространству;
- во время поиска пути необходимо учитывать не только статические препятствия, но и динамически образующиеся скопления агентов, которые сами по себе образуют области затрудненного перемещения;
- пространство может динамически изменяться во время моделирования и это необходимо учитывать при построении и динамическом перестроении пути.

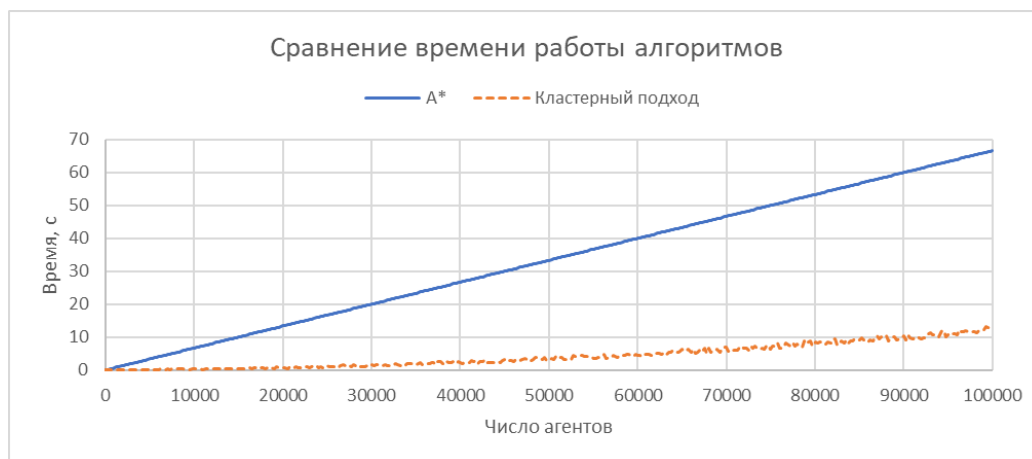


Рис. 2. Кривые зависимости времени построения пути для алгоритма A\* и кластерного подхода

Один из вариантов решения первой проблемы – кластерный подход, т.е. навигация целой группы агентов с совпадающей конечной целью вместо построения пути для каждого агента в отдельности (рис. 2). Такой подход рассматривается в работах [3–5].

Вторая проблема из перечисленных выше сводится к третьей, если представить новообразованное «столпотворение» как динамически появившееся препятствие в моделируемом пространстве (рис. 3). Другой подход состоит в динамической разметке областей пространства трудностью «проходимости». Подобные подходы к решению этих задач исследуют авторы в работах [6, 7].

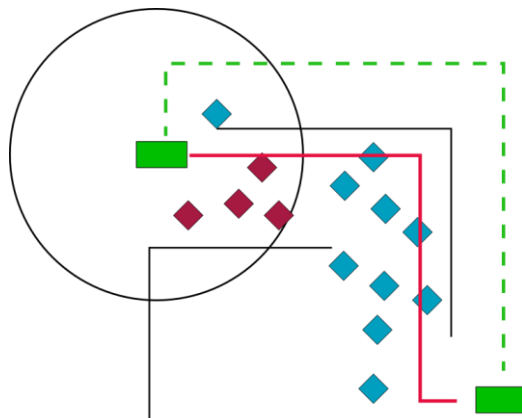


Рис. 3. Пример возможных путей агента, построенных с учетом и без учета скопления других агентов

**Операционный уровень.** Уровень планирования каждого шага и предотвращения столкновений между агентами. Именно на этом уровне сосредоточен интерес исследователей, занимающихся моделированием поведения большого количества людей. Большой интерес представляет реалистичность разработанных моделей в сравнении с поведением людей в реальных ситуациях. В связи с этим внимание исследователей сфокусировано на следующих несоответствиях:

- моделирование столкновений двух плотных потоков агентов часто выявляет нехарактерные для человека отклонения агентов от задуманной траектории;
- может происходить взаимная блокировка агентов в местах их скопления, например, в дверных проемах или узких коридорах;
- рваная или «дребезжащая» траектория движения агента с частыми поворотами и изменением скорости перемещения, нехарактерная для реального человека (рис. 4).



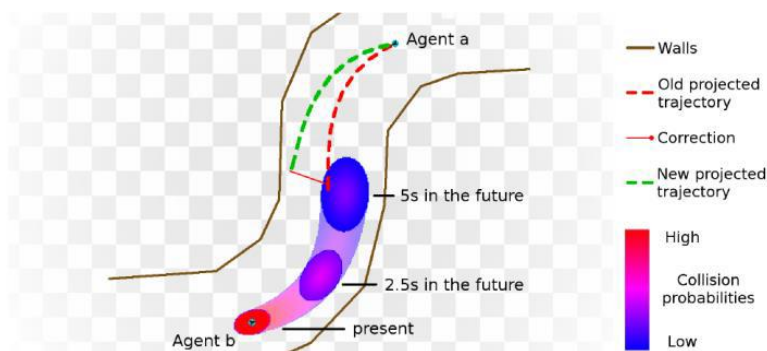


Рис. 4. Визуализация работы алгоритма WarpDriver, изображение из [8]

**Предсказание перемещений.** Для получения реалистичной реакции агента на движение окружающих и построения адекватной траектории движения, проводятся исследования в сфере предсказания позиции других агентов во времени [8, 9].

**Уточнение формы агента.** Для решения задачи взаимоблокировки агентов в ограниченных пространствах ведутся исследования в сторону использования формы агента, более точно описывающей тело реального человека. В качестве такой формы может использоваться, например, эллипс со значительной разницей длин между двумя диаметрами (рис. 5). Это повышает важность расположения агента не только с точки зрения позиции, но и поворота и открывает возможность моделирования доступа в еще более ограниченных пространствах [10, 11].

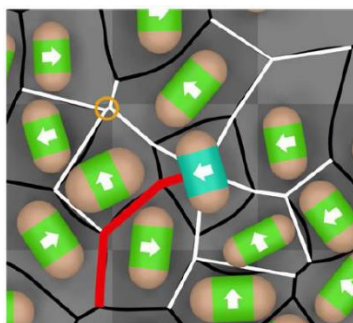


Рис. 5. Визуализация работы алгоритма Torso Crowds, изображение из [10]

**Взаимодействие между агентами.** Еще одно направление исследований касается вопросов взаимодействия агентов, которое не относится напрямую к обработке столкновений. Примером такого взаимодействия может служить распространение паники в толпе, когда агентов постепенно охватывает паника, постепенно распространяясь от источника (рис. 6) [12].

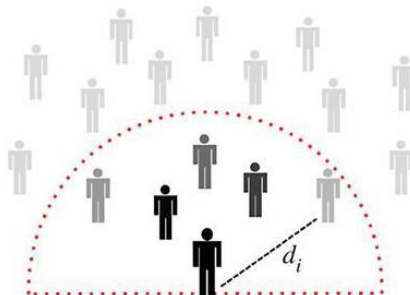


Рис. 6. Схематическое изображение работы алгоритма, взаимодействующего с агентами в определенном радиусе, изображение из [12]

**Литература**

1. Curtis S., Best A., Manocha D. Menge: A Modular Framework for Simulating Crowd Movement // *Collective Dynamics*. – 2016. – V. 1. – P. 1–40.
2. Kielar P.M., Biedermann D.H., Borrmann A. MomenTUMv2: A Modular, Extensible, and Generic Agent-Based Pedestrian Behavior Simulation Framework [Электронный ресурс]. – Режим доступа: [https://www.researchgate.net/publication/308918612\\_MomenTUMv2\\_A\\_Modular\\_Extensible\\_and\\_Generic\\_Agent-Based\\_Pedestrian\\_Behavior\\_Simulation\\_Framework](https://www.researchgate.net/publication/308918612_MomenTUMv2_A_Modular_Extensible_and_Generic_Agent-Based_Pedestrian_Behavior_Simulation_Framework) (дата обращения: 06.03.2019).
3. Vemula A., Muelling K., Oh J. Modeling cooperative navigation in dense human crowds // *IEEE International Conference on Robotics and Automation*. – 2017. – P. 1685–1692.
4. Aroor A., Epstein S.L. Toward Crowd-Sensitive Path Planning [Электронный ресурс]. – Режим доступа: [https://www.researchgate.net/publication/320442617\\_Toward\\_Crowd-Sensitive\\_Path\\_Planning](https://www.researchgate.net/publication/320442617_Toward_Crowd-Sensitive_Path_Planning) (дата обращения: 06.03.2019).
5. Liu H., Xu B., Lu D., Zhang G. A path planning approach for crowd evacuation in buildings based on improved artificial bee colony algorithm // *Applied Soft Computing Journal*. – 2018. – V. 68. – P. 360–376.
6. Aroor A., Epstein S.L., Korpan R. Online Learning for Crowd-sensitive Path Planning // *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. – 2018. – P. 1702–1710.
7. Wang Z., Zlatanova S. Multi-agent based path planning for first responders among moving obstacles // *Computers, Environment and Urban Systems*. – 2016. – V. 56. – P. 48–58.
8. Wolinski D., Lin M.C., Pettré J. WarpDriver // *ACM Transactions on Graphics*. – 2016. – V. 35(6). – P. 1–11.
9. Paris S., Pettré J., Donikian S. Pedestrian reactive navigation for crowd simulation: A predictive approach // *Computer Graphics Forum*. – 2007. – V. 26(3). – P. 665–674.
10. Stüvel S.A., Magnenat-Thalmann N., Thalmann D., Van Der Stappen A.F., Egges A. Torso crowds // *IEEE Transactions on Visualization and Computer Graphics*. – 2017. – V. 23(7). – P. 1823–1837.
11. Song X. et al. Simulation of pedestrian rotation dynamics near crowded exits // *IEEE Transactions on Intelligent Transportation Systems*. – 2018. – V. 20. – № 8. – P. 3142–3155.
12. Rio K.W., Dachner G.C., Warren W.H. Local interactions underlying collective motion in human crowds // *Proceedings of the Royal Society B: Biological Sciences*. – 2018. – V. 285. – № 1878. – P. 20180611.

**Федоров Виктор Викторович**

Год рождения: 1994

Университет ИТМО, институт дизайна и урбанистики, аспирант

Направление подготовки: 09.06.01 – Информатика и вычислительная техника

e-mail: fyodorovvictor@gmail.com

**Кокорев Андрей Дмитриевич**

Год рождения: 1995

Университет ИТМО, институт дизайна и урбанистики, аспирант

Направление подготовки: 09.06.01 – Информатика и вычислительная техника

e-mail: a.d.kokorev@yandex.ru

**Леонов Владислав Романович**

Год рождения: 1996

Университет ИТМО, институт дизайна и урбанистики, студент группы № С41222

Направление подготовки: 01.04.02 – Прикладная математика и информатика

e-mail: vlad\_leonov@bk.ru

**Карсаков Андрей Сергеевич**

Год рождения: 1985

Университет ИТМО, институт дизайна и урбанистики, к.т.н., ст. преподаватель

e-mail: karsakov@itmo.ru

**УДК 004.5****«ПОЛЬЗОВАТЕЛЬСКИЕ ИСТОРИИ» И «АТОМАРНЫЙ ДИЗАЙН»:  
ПРИМЕНИМОСТЬ ДЛЯ ГЕНЕРАЦИИ АДАПТИВНОГО  
ПОЛЬЗОВАТЕЛЬСКОГО ИНТЕРФЕЙСА****Федоров В.В., Кокорев А.Д., Леонов В.Р.****Научный руководитель – к.т.н. Карсаков А.С.**

Работа выполнена в рамках темы НИР № 618273 «Предсказательное моделирование динамики событий в социо-экономических системах».

В работе рассмотрена применимость подходов проектирования интерфейсов «пользовательские истории» и «атомарный дизайн» для технологии генерации адаптивного пользовательского интерфейса сложных систем.

**Ключевые слова:** пользовательский интерфейс, пользовательские истории, атомарный дизайн, адаптивный интерфейс, генерация пользовательского интерфейса.

В настоящее время в связи с быстрым развитием электронных и информационных технологий, все больше операций над информацией производятся с помощью компьютерных программ. Если ранее программы позволяли производить быстрые математические расчеты, то сейчас компьютеры используют для генерации информации, его анализа, поддержке принятия решений и прочих трудоемких задачах. Так как программа является в первую очередь инструментом, именно пользователь задает ей необходимые задачи для решения. Современные программы представляют собой большой набор необходимой функциональности, относящейся к той или иной сфере. В связи с этим программное обеспечение (ПО) обладает сложным графическим интерфейсом.

При работе с программой пользователь может использовать не весь предоставленный функционал, однако, он может достаточно сталкиваться с элементами интерфейса, предназначенными для работы с не используемой функциональностью. «Лишние», по мнению пользователя, графические элементы могут занимать для него полезное виртуальное рабочее пространство или отвлекать внимание от основного процесса работы.

Однако функциональность, бесполезная для одного пользователя, может оказаться значимой для другого пользователя. Поэтому будет опрометчиво убирать ее и программы, или убирать графические элементы, отвечающие за работу с ней. В связи с этим возникает вопрос о персонализации или адаптации интерфейса к отдельно взятому пользователю.

Графический пользовательский интерфейс представляет собой инструмент, при помощи которого пользователь общается с программой [1]. На рисунке представлена модель Сихайма, демонстрирующая логическую структуру пользовательского интерфейса.

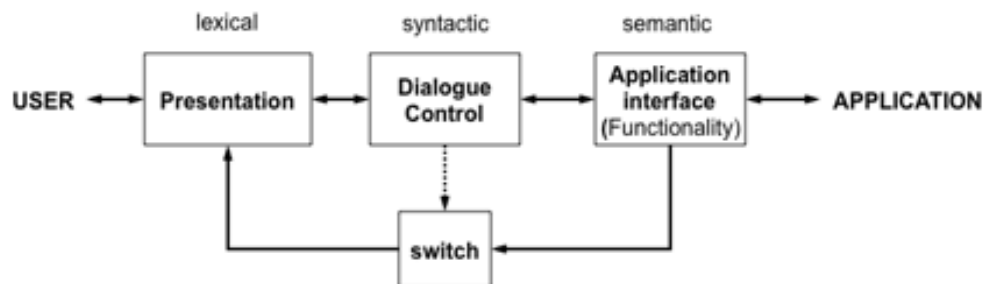


Рисунок. Модель логических компонентов Сихайма

Пользовательский интерфейс состоит из трех компонентов: презентационного, диалогового и функционального. Презентационный компонент принимает ввод команд от пользователя и выводит результаты работы программы в доступном для пользователя виде. Диалоговый компонент интерпретирует команды пользователя в команды программы. Функциональный компонент передает преобразованные команды уже в программу, получает результаты вычисления из программы и передает обратно в диалоговый компонент.

Таким образом, пользовательский интерфейс является важной частью приложения, и его исполнение может повлиять на эффективность работы пользователя с программой.

Пользовательские истории представляют собой способ описания требований к разрабатываемой системе [2]. Они могут иметь различное представление, но в основном создаются при помощи повседневного или бизнес-языка. Клиент записывает требования к программе на удобном ему языке. Пользовательские истории могут содержать описание рабочего процесса в общих чертах. Именно на основе этого описания делается заключение о средствах реализации требований к программе.

Несмотря на то, что процесс создания графического пользовательского интерфейса можно отнести к творческому процессу, он также требует определенных практических подходов. Существует множество подобных подходов. Они могут ориентироваться на функциональность будущего интерфейса, на его простоту в использовании, на его компактность и другие свойства, в зависимости от того, какую цель должен выполнять графический пользовательский интерфейс.

Атомарный дизайн является подходом проектирования, который ориентируется на минимальные, неделимые, самые простые графические элементы интерфейса [3]. Атомарный дизайн подразумевает деление процесса создания интерфейса на 5 этапов: «атом», «молекула», «организм», шаблон, страница.

«Атом» – минимальная частица пользовательского интерфейса. Сюда относят самые базовые графические элементы, доступные при процессе создания интерфейса в определенной среде разработки. Каждый «атом» включает в себе базовую функциональность, которая является достаточной сама по себе. Такие графические элементы могут «взаимодействовать» друг с другом, образуя новые, более сложные структуры для выполнения более сложных задач.

«Молекула» – совокупность «атомов», образующих относительно простые компоненты. Соединение базовых графических элементов на текущем этапе уже формируется для выполнения конкретной задачи. Например, объединение поля для ввода текста с клавиатуры и кнопки образует форму для поиска. Рассмотрение таких совокупностей со стороны состава позволяет обращать большее внимание на наличие элементов и снижает вероятность внедрения малозначащего или лишнего элемента при конструировании интерфейса.

«Организм» – относительно сложный компонент, образующий независимый участок интерфейса. Он может состоять из совокупности «атомов», «молекул» или других «организмов». Такой компонент является логически завершенным и представляет собой модуль, изменения в котором не повлияют на другие подобные модули.

Шаблон – макет, демонстрирующий базовую структуру контента. Расположение всех необходимых компонентов с предыдущих этапов: «атомов», «молекул» и «организмов» – на виртуальной рабочей области позволит увидеть их взаимодействие в определенном контексте. На данном этапе уже формируется структура графического пользовательского интерфейса.

Страница – шаблон вкуче с реальным контентом. После настройки шаблон наполняется содержимым. На данном этапе уже формируется конечный вид пользовательского интерфейса.

Атомарный дизайн широко используется в веб-среде при проектировании страниц сайтов. Однако такой подход можно использовать и при проектировании графического пользовательского интерфейса и десктоп приложения. Такой подход позволяет ясно проследить состав интерфейса и определять необходимые элементы для предоставления требуемой функциональности.

Атомарный дизайн предоставляет модульную систему для создания пользовательского интерфейса, что позволяет четко регулировать состав и структуру интерфейса. Пользовательские истории предоставляют описание функциональности, требуемой от программного обеспечения. Применение принципа модульности к пользовательским историям позволяет определить структуру самой программы и способа ее функционирования. Так как пользовательский интерфейс является инструментом общения пользователя с программой, есть возможность настроить соответствие модулей интерфейса с модулями программы. Таким образом, можно использовать пользовательские истории для создания функциональных модулей, которые, в свою очередь, описывают связанные с ними интерфейсные модули.

Предполагается, что формализация пользовательских историй позволит создать их формальное описание. Подобное описание возможно использовать для анализа и разбора, при котором будет определяться, какая функциональность требуется от программы. При наличии данных о соответствии модулей функциональности и модулей графического интерфейса, проанализированное описание будет предоставлять информацию для автоматической генерации элементов пользовательского интерфейса. Таким образом, по формальному описанию пользовательского интерфейса будет генерироваться функционально полный графический пользовательский интерфейс.

Применение принципа модульности к генерации пользовательского интерфейса позволяет формализовать его описание и предоставить возможность его автоматической генерации. Применение пользовательских историй предоставляет информацию о функциональности программы и, следовательно, предоставляет требования к компонентам, необходимым для реализации подобной функциональности. Объединение данных подходов позволит автоматически генерировать функционально полный интерфейс.

**Литература**

1. Singh G., Green M. Automatic generation of graphical user interfaces // Proc. graphics interface. – 1986. – P. 71–76.
2. Пользовательские истории. Википедия [Электронный ресурс]. – Режим доступа: [https://ru.wikipedia.org/wiki/Пользовательские\\_истории](https://ru.wikipedia.org/wiki/Пользовательские_истории) (дата обращения: 06.03.2019).
3. Atomic Design Methodology [Электронный ресурс]. – Режим доступа: <http://atomicdesign.bradfrost.com/chapter-2/> (дата обращения: 06.03.2019).

## СОДЕРЖАНИЕ

<b>Направление «ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И ПРОГРАММИРОВАНИЯ»</b> .....	<b>4</b>
<b>Аксенов Ю.В., Возисов Н.С., Шорохов С.А.</b> Разработка системы для автоматизации персонализированной поддержки медицинских технологических процессов с применением паттернов проектирования .....	5
<b>Алсуфьев А.А., Старикова Е.Г., Былецкий Д.А.</b> Применение гибридных подходов при управлении содержанием в проектах в рамках ФЦП «Исследования и разработки» .....	11
<b>Аусев Е.В.</b> Исследование технологий машинного перевода для верификации произнесенной фразы .....	16
<b>Бебяков А.М.</b> (Университет ИТМО), <b>Довидченко Н.В.</b> (ООО «Кномикс»), <b>Одинцова В.Е.</b> (ООО «Кномикс»), <b>Клименко Н.С.</b> (ООО «Кномикс»), <b>Тягт А.В.</b> (ООО «Кномикс»). Сравнительный анализ состава микробиома кишечника человека при язвенном колите и болезни Крона по данным WGS-метагеномики.....	19
<b>Большаков Н.А.</b> Детектирование акустических событий с использованием аугментаций данных и нейронных сетей .....	24
<b>Виноградова Т.Б.</b> Сравнительный анализ методов определения тональности коротких текстов .....	28
<b>Bassel H. Zeno</b> Face rotation using generative Adversarial Networks .....	33
<b>Казиева Н.</b> Биометрические штрих-коды для взаимодействия людей с платформой «Интернет вещей» .....	37
<b>Калиев А.</b> Проблемы обучения генеративных состязательных сетей в синтезе речи .....	41
<b>Куан Ч.Т.</b> Сравнение алгоритмов выделения речевых сигналов для двухэлементной микрофонной решетки .....	44
<b>Лизунова И.А.</b> Использование словаря эмодзи для анализа тональности текстов из социальных медиа .....	49
<b>Мамаев Н.К.</b> Диалоговая система на основе поиска с языконезависимыми векторными представлениями слов .....	53
<b>Мироненко А.А.</b> Метод повышения точности трекинга лиц за счет компенсации ошибок назначения трека .....	56
<b>Петров О.Е., Кабаров В.И.</b> Использование обратных индексов для эффективного полнотекстового поиска по речевым данным .....	61
<b>Рюмин Д.</b> (Санкт-Петербургский институт информатики и автоматизации РАН, Университет ИТМО), <b>Аксенов А.</b> (Санкт-Петербургский институт информатики и автоматизации РАН, Университет ИТМО). Метод автоматического обнаружения и распознавания 3D одноручных жестов рук .....	64
<b>Сухов В.Д., Короткевич Г.В.</b> Вычисление близких к нулю P-значений при анализе представленности функциональных наборов генов .....	70
<b>Чирковский А.Д.</b> Компьютерная формализация текстов на языке эсперанто .....	75
<b>Направление «ИНФОКОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ»</b> .....	<b>79</b>
<b>Найков А.В.</b> Разработка процедуры реагирования на аварийные ситуации в географически распределенном центре обработки данных .....	80
<b>Направление «ИНСТИТУТА ДИЗАЙНА И УРБАНИСТИКИ»</b> .....	<b>84</b>
<b>Голубев К.В.</b> Визуальный анализ и валидация результатов обработки геопривязанных данных .....	85
<b>Деева И.Ю.</b> Прогнозирование личностных метрик для контекстно-независимого цифрового образа пользователя социальной сети .....	89

---

<b>Деревицкий И.В.</b> Сравнительный анализ подходов к выявлению паттернов в течение сахарного диабета 2 типа .....	94
<b>Калинин А.М., Ваганов Д.А.</b> Восстановление социально-экономического статуса пользователей онлайн-сетей по частично размеченным данным .....	98
<b>Кокорев А.Д., Лебин А.Э.</b> Мультиагентные поведенческие модели в задачах пешеходного моделирования .....	105
<b>Федоров В.В., Кокорев А.Д., Леонов В.Р.</b> «Пользовательские истории» и «Атомарный дизайн»: применимость для генерации адаптивного пользовательского интерфейса .....	110



**СБОРНИК ТРУДОВ VIII КОНГРЕССА  
МОЛОДЫХ УЧЕНЫХ  
Том 4**

В авторской редакции

Редакционно-издательский отдел Университета ИТМО

Дизайн обложки

Н.А. Потехина

Зав. РИО

Н.Ф. Гусарова

Редактор

Л.Н. Точилина

Подписано к печати 24.10.2019

Заказ № 4244

Тираж 100 экз.