

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ УЧРЕЖДЕНИЕ «ФЕДЕРАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР «ИНФОРМАТИКА И УПРАВЛЕНИЕ»
РОССИЙСКОЙ АКАДЕМИИ НАУК»

Заключение диссертационного совета Д 002.073.05 на базе Федерального государственного учреждения «Федеральный исследовательский центр «Информатика и управление» Российской академии наук»

Аттестационное дело № _____

Решение диссертационного совета от 15.12.2016 № 11

О присуждении Кудинову Михаилу Сергеевичу, гражданину Российской Федерации ученой степени кандидата технических наук.

Диссертация «Статистическое моделирование русского языка с помощью нейронных сетей» по специальности 05.13.17 – «Теоретические основы информатики» принята к защите 13 октября 2016 г., протокол №8, диссертационным советом Д 002.073.05 на базе государственного учреждения «Федеральный исследовательский центр «Информатика и управление» Российской академии наук», 119333, г. Москва, ул. Вавилова, д.44, к.2, приказ о создании диссертационного совета Министерства Образования и Науки Российской Федерации №783/нк от 24.06.2016.

Соискатель Кудинов Михаил Сергеевич, 1990 года рождения, в 2012 году окончил ФГБОУ ВО «Московский государственный университет имени М.В.Ломоносова». В 2012 – 2015 гг. аспирант ФГБУН «Вычислительный центр им. А. А. Дородницына Российской академии наук». Работает инженером в ООО «Исследовательский Центр Самсунг».

Научный руководитель кандидат физико-математических наук Чучупал В.Я., ведущий научный сотрудник отдела «методы классификации и анализа данных» ФИЦ ИУ РАН.

Официальные оппоненты:

Ромашкин Юрий Николаевич, кандидат технических наук, ведущий научный сотрудник Федерального государственного казенного учреждения "Войсковая часть 35533",

Карпов Алексей Анатольевич, доктор технических наук, доцент, заведующий лабораторией речевых и мультимедийных интерфейсов Федерального государственного бюджетного учреждения науки "Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН)"

дали **положительный** отзыв на диссертацию.

Ведущая организация Федеральное государственное бюджетное учреждение науки Институт проблем передачи информации им. А.А. Харкевича Российской академии наук (ИППИ РАН) в своем **положительном** заключении, подписанном Сорокиным В.Н., д.ф.-м.н., доц., ведущим научным сотрудником лаборатории Информационных технологий передачи, анализа и защиты информации, указала, что диссертация является самостоятельной законченной научно-квалификационной работой, обладает научно-теоретической, практической значимостью и научной новизной, имеет существенное

значение для развития направлений исследований в области теоретических основ информатики. Автореферат достаточно правильно отражает текст диссертации.

В качестве замечания отмечено, что предложенные в диссертации алгоритмы нуждаются в ряде теоретических оценок качества предсказаний модели. Кроме того, отмечается, что эксперименты, поставленные в рамках исследования, проводились на закрытых данных.

Соискатель имеет 7 научных работ, в том числе 5 статей в журналах, рекомендованных ВАК. Работы автора нашли практическое применение в технологиях компании «Самсунг ЭлектрониксКо., Лтд». В частности, автором был получен патент на изобретение «Голосовая связь на естественном языке между человеком и устройством» (RU 2583150).

Наиболее значимые научные работы по теме диссертации:

1. Kudinov M., Romanenko A., Piontkovskaya I. Conditional Random Fields in segmentation and noun phrase inclination tasks for Russian // Computational linguistics and intellectual technologies: Proceedings of the International Conference Dialogue — 2014. — Pp. 297–307.
2. Kudinov M., Piontkovskaya I. Automatic update of the named entities database based on users queries // Computational linguistics and intellectual technologies: Proceedings of the International Conference Dialogue — 2015. — Vol.1 — Pp. 369–376.
3. Kudinov M. Recurrent Neural Networks for Hypotheses Re-Scoring // Speech and Computer — 17th International Conference, SPECOM 2015, Athens, Greece, September 20-24, 2015, — 2015. — Pp. 341–347.
4. Kudinov M., Romanenko A. Hybrid language model based on recurrent neural network and probabilistic topic modeling // Pattern Recognition and Image Analysis. — 2016. — Vol. 26. — Pp.587–592.
5. Кудинов М. С. Использование рекуррентных нейронных сетей для ранжирования списка гипотез в системах распознавания речи // Современная наука: актуальные проблемы теории и практики. Серия «Естественные и технические науки». — 2016. — С.358–364.
6. Кудинов М. Частичный синтаксический разбор текста на русском языке с помощью условных случайных полей // Машинное обучение и анализ данных. — 2013. — Т.1., — С.714–724

На автореферат поступили отзывы:

1. **Кушнира Дмитрия Алексеевича**, к.т.н., руководителя группы речевых технологий ООО «ДСС-Лаб». Отзыв положительный, однако содержит замечание о том, что в автореферате не упомянуты модель LSTM рекуррентной нейронной сети и не рассматривается возможность учета правого контекста предсказываемого слова.
2. **Шайдунова Георгия Яковлевича**, д.т.н., научного руководителя Военно-инженерного института ФГАОУ ВО «Сибирский Федеральный Университет», профессора. Отзыв положительный, замечаний нет.
3. **Жигулевцева Юрия Николаевича**, к.т.н. старшего научного сотрудника кафедры «Системы автоматического управления» МГТУ им. Баумана. Отзыв положительный, замечаний нет.

Выбор официальных оппонентов обосновывается тем, что Ю.Н.Ромашкин и А.А.Карпов являются известными специалистами в области речевых технологий и, в частности, распознавания речи.

Ведущая организация – ФГБУН «Институт проблем передачи информации им. А.А.Харкевича Российской академии наук (ИППИ РАН)» является одним из передовых научных центров в области теоретической информатики, речевых технологий и автоматической обработки текста.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований:

предложена и протестирована статистическая языковая модель на основе нейронных сетей, использующая отдельные классификаторы для предсказания лемм и морфологических признаков,

предложена и протестирована языковая модель, задействующая признаки, полученные путем отображения контекста слова в вектор действительных чисел методами вероятностного тематического моделирования,

предложена и протестирована модель предсказания морфологической формы на основе сверточной нейронной сети.

Теоретическая значимость

Предложена статистическая языковая модель с отдельным предсказанием лемм и морфологических признаков, основанная на применении рекуррентной и сверточной нейронных сетей, позволяющая учитывать более длинный левый контекст для предсказания; разработана и реализована гибридная статистическая языковая модель на рекуррентной нейронной сети и тематическом разложении левого контекста, повышающая качество на более длинных текстах; предложен, обоснован и экспериментально исследован метод расширения словаря языковой модели на рекуррентной нейронной сети за счет отдельной модели предсказания морфологической формы русских слов, основанной на сверточной нейронной сети.

Научная новизна результатов исследования заключается в том, что:

- 1) *экспериментально исследованы* новые алгоритмы и методы статистического моделирования языка, позволяющие повысить производительность и качество предсказаний на русских текстах относительно общепринятых методов;
- 2) *модифицированы* существующие методы статистического моделирования языка с целью улучшения качества предсказаний: для учета более длинного левого контекста *задействованы* методы вероятностного тематического моделирования;
- 3) для уменьшения объема словаря модели производится отдельное предсказание начальной формы с помощью рекуррентной нейронной сети и морфологических показателей с помощью сверточной нейронной сети;
- 4) *экспериментально проверяется* гипотеза о том, что порядок слов в русском языке не оказывает негативного влияния на качество языковой модели на рекуррентной нейронной сети при предсказании начальных форм слов.

Практическая значимость результатов исследования выражается в том, что предложенные в рамках исследования модели могут быть задействованы в программных продуктах в области распознавания речи, распознавания рукописного текста и других.

Достоверность научных положений, результатов и выводов диссертации обеспечена корректным использованием математического аппарата и сходимостью теоретических оценок с экспериментальными результатами. **Новизна** результатов подтверждается сравнением с известными результатами по тематике диссертационного исследования.

Личный вклад соискателя состоит в том, что все теоретические и экспериментальные данные получены автором лично.

На заседании 15.12.2016 г. Диссертационный совет принял решение присудить Кудинову Михаилу Сергеевичу степень кандидата технических наук.

При проведении тайного голосования диссертационный совет в количестве 26 человек, из них 9 докторов наук по специальности 05.13.17 – «Теоретические основы информатики», участвовавших в заседании, из 35 человек, входящих в состав совета, дополнительно введены на разовую защиту: 0 человек, проголосовали за: 25, против: 1, недействительных бюллетеней: 0.

Председатель диссертационного
совета Д 002.073.05, академик РАН

Журавлев Ю.И.



Ученый секретарь
диссертационного совета
Д.002.017.02, д.ф.-м.н., профессор



Рязанов В.В.



19 декабря 2016 г.