

ОТЗЫВ

официального оппонента Пимонова Александра Григорьевича
на диссертацию Грибкова Егора Игоревича

«Нейросетевые модели на основе системы переходов для извлечения структурированной информации о продуктах из текстов пользователей», представленную на соискание ученой степени кандидата технических наук по специальности 05.13.17 – Теоретические основы информатики

Для анализа представлены диссертация общим объемом 128 страниц и автореферат на 21 странице текста.

Актуальность темы исследования

Мы живем, работаем и отдыхаем в океане информации, в большинстве своем представленной самыми разнообразными интернет-текстами. Эти тексты чаще всего носят неформальный характер и изобилуют расхождениями с письменной нормой а зачастую и со всеми правилами русского языка (употребление сленга, жаргонизмов, просторечных слов, игнорирование правил пунктуации, использование сокращений, неологизмов, заимствований из иностранных языков и т. д.). Все это не позволяет легко анализировать письменные тексты. Поэтому автоматизация обработки и анализа содержимого текстов, подготовленных на естественном языке, для извлечения из них информации об интересующих пользователя объектах (товарах, услугах, свойствах, возможностях и т. п.) является актуальной задачей.

Диссертационная работа посвящена разработке нейросетевых моделей на основе системы переходов для извлечения структурированной информации о продуктах из текстов пользователей.

Анализ содержания диссертационной работы

Диссертационная работа Грибкова Е.И. состоит из введения, трех глав, заключения, списка литературы из 127 наименований, списка сокращений и условных обозначений, списка рисунков и списка таблиц. Материалы диссертационного исследования изложены на 128 страницах текста, содержат 20 таблиц и 37 рисунков.

Во **введении** соискатель обосновывает актуальность исследования, излагает степень разработанности темы, формулирует цель, задачи, объект и предмет исследования, представляет научную новизну, теоретическую и практическую значимость исследования, представляет реализацию результатов и степень их достоверности, апробацию работы, формулирует свой личный вклад и выносимые на защиту основные положения диссертационной работы.

В **первой главе** «Методы машинного обучения в задачах информационной поддержки процессов эксплуатации и сопровождения продуктов» автором проведен обзор методов анализа пользовательских текстов, подготовленных

на естественном языке. Соискатель отмечает, что современные методы, используемые для решения рассмотренных задач извлечения информации, не позволяют учитывать структурные взаимосвязи между извлекаемыми сущностями и подвержены проблемам распространения ошибки, что приводит к низкой точности извлечения объектов. Для решения задачи извлечения структурированной информации из текстов о мнениях, запросах и пожеланиях потребителей соискатель предлагает объединить преимущества моделей, созданных на основе системы переходов, с признаками, получаемыми с помощью рекуррентных и сверточных нейронных сетей.

Во **второй главе** «Нейросетевые модели на основе системы переходов для извлечения структурированной информации о продуктах из текстов пользователей» соискатель рассматривает общую задачу извлечения составных объектов из текста и два варианта ее адаптации под конкретные предметные области. В этой главе автором представлена модель для извлечения составных объектов и их атрибутов из текстов на естественном языке, на основе которой разработаны две нейросетевые модели: 1) нейросетевая модель для извлечения и анализа пользовательских мнений из текстов отзывов о потребительских свойствах товаров; 2) нейросетевая модель для обработки запросов пользователей на этапе эксплуатации программного продукта.

На основе качественного анализа полученных результатов диссертант делает вывод о возможности практического применения предложенных моделей.

В **третьей главе** «Практическая апробация и внедрение моделей, алгоритмов и программного обеспечения» соискателем приведен обзор основных представленных на рынке программных продуктов для обработки текстов, подготовленных на естественном языке.

Здесь же представлены разработанные автором программная система «Quiddi Semantics», предназначенная для сбора и обработки текстов отзывов магазина «AliExpress», и прототип программной системы «Quiddi Support Analyst» для сбора и анализа обращений пользователей программных продуктов.

Из изложенных в этой главе результатов практической апробации предложенных нейросетевых моделей, реализованных в виде компонентов программных систем «Quiddi Semantics» и «Quiddi Support Analyst», диссертант делает вывод об их работоспособности и практической пригодности для решения задач по извлечению мнений и запросов пользователей из текстов.

В **заключении** соискателем изложены основные теоретические и практические результаты диссертационного исследования.

Научная новизна проведенных исследований и полученных результатов

Автором в диссертации представлены новые научно обоснованные нейросетевые модели на основе переходов, реализованные в виде программных систем и позволяющие извлекать структурированную информацию из пользо-

вательских текстов на естественном языке, что способствует повышению качества товаров и услуг, предоставляемых компаниями. Научная новизна диссертационной работы Грибкова Е.И. заключается в том, что:

1) была предложена оригинальная нейросетевая модель на основе системы переходов для извлечения составных объектов и их атрибутов из текстов на естественном языке, позволяющая одновременно предсказывать структуру объекта и значения его атрибутов, с возможностью адаптации этой модели для решения конкретных задач через определение множеств, описывающих семантику фрагментов и атрибутов;

2) предложенная модель была положена в основу разработанной оригинальной нейросетевой модели для извлечения и анализа мнений из текстов пользовательских отзывов о продуктах, отличающейся от известных моделей использованием подхода на основе системы переходов и позволяющей получить лучшие показатели качества извлечения частей составных объектов;

3) предложенная модель общего вида была адаптирована и положена в основу разработки оригинальной нейросетевой модели для анализа запросов пользователей на этапе эксплуатации и сопровождения программного продукта, отличающейся от известных моделей использованием подхода на основе системы переходов и позволяющей получить лучшие показатели качества извлечения частей составных объектов.

Обоснованность и достоверность научных положений, выводов и рекомендаций, сформулированных в диссертации

Выводы и основные положения, выносимые соискателем на защиту, логично вытекают из содержания диссертационной работы.

Достоверность и обоснованность исследований и полученных в диссертации результатов не вызывают сомнений и в достаточной мере обусловлены корректным применением аппарата нейронных сетей при разработке модели и результатами сравнительного анализа использования предложенных моделей с аналогичными. Адекватность предложенных в работе моделей и алгоритмов подтверждается результатами их практического использования в коммерческих программных продуктах компании ООО «ТомскСофт».

Основные результаты исследований отражены в 8 публикациях, в числе которых 4 статьи в научных изданиях из перечня ВАК РФ. Соискателем получены 2 свидетельства об официальной регистрации программ для ЭВМ: № 2019612276 от 14.02.2019 г. и № 2020614799 от 24.04.2020 г.

Значимость и использование результатов, полученных в диссертационной работе

Теоретическая значимость диссертационной работы Грибкова Е.И. заключается в развитии методов предсказания составных объектов с использованием нейронных сетей для извлечения структурированной информации из пользовательских текстов на естественном языке с применением моделей

на основе системы переходов в задачах, связанных с обработкой текстов мнений пользователей о потребительских и программных продуктах.

Практическую значимость результатов исследования составляют комплекс нейросетевых моделей и программные средства, позволяющие маркетологам компаний выполнить анализ текстов отзывов пользователей о продуктах для определения сильных и слабых сторон собственных продуктов и продуктов фирм-конкурентов; специалистам службы технической поддержки на этапе эксплуатации и сопровождения программного продукта выполнить анализ предложений пользователей для обеспечения эксплуатации программного продукта в соответствии с его техническими характеристиками и развития продукта в соответствии с требованиями рыночной ситуации.

Замечания по диссертационной работе

По содержанию диссертационной работы можно сделать следующие замечания.

1) В п. 3 научной новизны (с. 7) указано, что «*На основе предложенного метода разработана оригинальная модель...*», хотя ни до, ни после этого в тексте диссертации не описан какой-либо *предложенный метод*.

2) При описании новизны в п. 2, 3 (с. 7) соискателем используется «*показатель качества*» (терминология автора) F_1 . Сам же этот показатель определяется только в главе 2 на с. 48-49. При этом автор называет его то мерой (с. 48), то метрикой (с. 85), то показателем метрики (с. 13 автореферата). В формуле (2.10) (с. 48) определения F_1 отсутствуют пояснения смысла P и R , что очень затрудняет понимание их способа расчета. Не определены границы изменения F_1 , что не позволяет сопоставить численному значению F_1 соответствующий качественный смысл.

3) В пояснениях к формуле (2.3) (с. 42) отсутствует описание E . И только из автореферата становится понятно (с. 11 автореферата), что это «*отображение последовательности слов в последовательность векторных представлений*».

4) Формула (2.5) (с. 43) содержит b . Не поясняется, что это. В автореферате (формула (5), с. 12) b без объяснения причин исчезает.

5) На с. 75 не определено множество $V(a)$. В автореферате (с. 16) это сделано.

6) На с. 17 автореферата *пример 2* не соответствует *примеру 5* диссертации (с. 83).

Содержание автореферата соответствует тексту диссертационной работы. В тексте автореферата встречаются опечатки, ошибки, несогласованность слов в предложениях, тавтология.

1) Не понятна фраза (с. 2, абз. 1) «*...что может послужить как увеличить, так и обрушить продажи...*».

2) Аналогично (с. 6, абз 7) – «*...оценка качества которого в наибольшей степени подкреплена информацией их отзывов других покупателей.*».

- 3) Аналогично (с. 9, абз. 2) – «...применение методов, позволяющих предсказывать структуру объекта в рамках единой модели с учетом структурных взаимосвязей между сущностями в тексте, что **обеспечивают** более высокую точность...».
- 4) Аналогично (с. 11, абз. 9) – «В качестве *f* будет использовать вероятностный классификатор следующего вида:».
- 5) Тавтология (с. 12, абз. 4) – «...**представлена** в виде диаграммы, **представленной** на рисунке 2.».
- 6) Не согласованы слова (с. 15, абз. 4) – «Улучшение относительно следующей наилучшей **альтернативой** составило:...».
- 7) (С. 17, абз. 1) – «...Google **Paly** Market...».
- 8) (С. 18, абз. 2) – «...для обработки текстов **не** естественном языке.».
- 9) (С. 18, абз. 2) – «...что из всех рассмотренных продуктов **Только** два из шести рассмотренных продуктов содержат...».
- 10) (С. 18, последний абз.) – «...что подтверждается актом **на** внедрение и **использовании** программных систем...».

Сделанные замечания не снижают научной и практической ценности проведенного диссертационного исследования и не влияют на общую положительную оценку.

Заключение о соответствии диссертации критериям, установленным Положением о присуждении ученых степеней

Диссертационная работа Грибкова Е.И. представляет собой научно-квалификационную работу, написана на актуальную тему, отличается научной новизной и практической значимостью, имеет завершенный характер, выполнена на высоком научно-техническом уровне.


Автором в диссертации описаны новые научно обоснованные нейросетевые модели на основе системы переходов и представлены программные средства, позволяющие маркетологам компаний выполнить анализ текстов отзывов пользователей о продуктах для определения сильных и слабых сторон собственных продуктов и продуктов фирм-конкурентов; специалистам службы технической поддержки на этапе эксплуатации и сопровождения программного продукта выполнить анализ предложений пользователей для обеспечения эксплуатации программного продукта в соответствии с его техническими характеристиками и развития продукта в соответствии с требованиями рыночной ситуации, что способствует повышению эффективности управления деятельностью компаний.

Содержание диссертации соответствует пунктам 5 и 6 паспорта специальности 05.13.17 – Теоретические основы информатики: «Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечениях, разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений»; «Разработка методов, языков и моделей человеко-машинного общения; разработка методов и моделей распо-

знания, понимания и синтеза речи, принципов и методов извлечения данных из текстов на естественном языке». Автореферат соответствует содержанию диссертации, а ее основные положения опубликованы в научных работах.

Считаю, что диссертационная работа Грибкова Е.И. «Нейросетевые модели на основе системы переходов для извлечения структурированной информации о продуктах из текстов пользователей» удовлетворяет требованиям п. 9 «Положения о присуждении ученых степеней», предъявляемым к кандидатским диссертациям, а ее автор, Грибков Егор Игоревич, заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.17 – Теоретические основы информатики.

Заведующий кафедрой прикладных информационных технологий ФГБОУ ВО «Кузбасский государственный технический университет им. Т. Ф. Горбачёва», доктор технических наук, профессор


04.12.20

Пимонов Александр Григорьевич

Докторская диссертация защищена по специальности 05.13.16 – Применение вычислительной техники, математического моделирования и математических методов в научных исследованиях (по отраслям наук).

Адрес: 650000, г. Кемерово, ул. Весенняя, д. 28.

Телефон (приемная): (3842) 68-23-14, факс: (3842) 39-69-60.

E-mail: pag_vt@kuzstu.ru.

Веб-сайт: kuzstu.ru.

Подпись Пимонов Е.И.
ЗАВЕРЯЮ
ученый секретарь совета
Э.В. Хейлиник
«04» 12 2020

