

ЗАКЛЮЧЕНИЕ ДИССЕРТАЦИОННОГО СОВЕТА Д 212.277.04,
СОЗДАННОГО НА БАЗЕ
ФГБОУ ВО «УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ» ПО ДИССЕРТАЦИИ
НА СОИСКАНИЕ УЧЕНОЙ СТЕПЕНИ КАНДИДАТА НАУК

аттестационное дело № _____

решение диссертационного совета от 14.09.2022 № 15

О присуждении Андрееву Илье Алексеевичу, гражданину Российской Федерации, ученой степени кандидата технических наук.

Диссертация «Исследование методов и алгоритмов обработки текстовой информации социальных сетей в задачах формирования социального портрета пользователя» по специальности 05.13.01 – Системный анализ, управление и обработка информации (информационные технологии и промышленность) принята к защите 22.06.2022 (протокол заседания № 8) диссертационным советом Д 212.277.04, созданным на базе ФГБОУ ВО «Ульяновский государственный технический университет» (432027, Россия, г. Ульяновск, ул. Северный Венец, д. 32) приказ №678/НК от 18.11.2020 г.

Соискатель Андреев Илья Алексеевич, 24 апреля 1994 года рождения.

В данный момент соискатель обучается на 4 курсе аспирантуры в ФГБОУ ВО «Ульяновский государственный технический университет», работает старшим преподавателем кафедры «Информационные системы» в ФГБОУ ВО «Ульяновский государственный технический университет». Диссертация выполнена в ФГБОУ ВО «Ульяновский государственный технический университет».

Научный руководитель – кандидат технических наук (05.13.12 Системы автоматизации проектирования), Мошкин Вадим Сергеевич, доцент кафедры «Информационные системы» ФГБОУ ВО «Ульяновский государственный технический университет».

Официальные оппоненты:

Куприянов Александр Викторович, доктор технических наук, доцент, заведующий кафедрой технической кибернетики, исполнительный директор

института информатики и кибернетики Федерального государственного автономного образовательного учреждения высшего образования «Самарский национальный исследовательский университет имени академика С.П. Королева»;

Абрамов Максим Викторович, кандидат технических наук, заведующий лабораторией теоретических и междисциплинарных проблем информатики Федерального государственного бюджетного учреждения науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН)

дали положительные отзывы на диссертацию.

Ведущая организация

Федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича» в своем положительном отзыве, подписанном заведующим кафедры инфокоммуникационных систем, к.т.н., Зарубиным Антоном Александровичем, профессором кафедры инфокоммуникационных систем, д.т.н., Гольдштейном Борисом Соломоновичем, доцентом кафедрой инфокоммуникационных систем, к.т.н. Елагиным Василием Сергеевичем и утвержденном Шестаковым Александром Викторовичем, и.о. проректора по научной работе, д.т.н., с.н.с., указала, что диссертация является законченной научно-квалификационной работой, выполненной на актуальную тему. Научные результаты, полученные в диссертации, обладают новизной и направлены на решение научной задачи, которая имеет большое значение для развития методов решения задач системного анализа.

Соискатель имеет 32 опубликованные работы, все они опубликованы по теме диссертации, в том числе статьи в изданиях, рекомендованных ВАК 4 работы, 11 работ в изданиях, индексируемых в Scopus и/или Web Of Science.

Наиболее значимые научные работы по теме диссертации:

1. Андреев, И. А. Алгоритм психолингвистического анализа текстовых данных социальных сетей с применением модели «Большая пятёрка»

/ Андреев И. А., Ярушкина Н. Г., Мошкин В. С. // Онтология проектирования. – 2022. – Т. 12, №1 (43). – С. 82–92. (лично соискателем – 9 страниц).

2. Андреев, И. А. Комбинирование статистического и лингвистического методов для извлечения двухсловных терминов из текста / Андреев И. А, Башаев В. А., Клейн В. В., Ярушкина Н. Г. // Автоматизация процессов управления. – 2013. – № 4 (34). – С. 61–70. (лично соискателем – 6 страниц).

3. Andreev, I.A. Solving the problem of determining the author of text data using a combined assessment // Yarushkina N.G., Moshkin V.S., Andreev I.A. // Proceedings of 8th International Conference «Fuzzy Systems, Soft Computing and Intelligent Technologies 2020 (FSSCIT 2020)», CEUR WS Proceedings, Vol-2782, pp. 112-118. (лично соискателем – 5 страниц).

4. Andreev, I. The Sentiment Analysis of Unstructured Social Network Data Using the Extended Ontology SentiWordNet / Andreev, I., Moshkin V., Yarushkina N. // IEEE, 12th International Conference on Developments in eSystems Engineering (DeSE) – Kazan, Russia – 2019 – pp. 576-580. (лично соискателем – 4 страницы).

5. Andreev, I. Approaches to sentiment analysis of the social network text data / Andreev, I. Yarushkina N., Moshkin V. // Proceedings of the Data Science Session at the VI International Conference on Information Technology and Nanotechnology (DS-ITNT 2020) – Vol. 2667. – pp. 198–202. (лично соискателем – 3 страницы).

На диссертацию и автореферат поступили 4 отзыва, все отзывы положительные, в отзывах содержатся следующие замечания:

1. ФГБОУ ВО «Ульяновский институт гражданской авиации имени главного маршала авиации Б.П. Бугаева» (г. Ульяновск). Отзыв подписан доцентом кафедры «Организации аэропортовой деятельности и информационных технологий» Чоракаевым О.Э. (к.т.н., специальность 05.13.12).

Замечания:

• В автореферате можно заметить следующий недостаток: приведены эксперименты по оценке эффективности алгоритмов классификации тональности с использованием только двух языковых моделей – Word2Vec и BERT, при этом не встречается обоснования их применения. В настоящее время существует большой набор модификаций языковых моделей (ELMo, GloVe, TinyBERT), которые могли бы показать высокую эффективность в данной задаче.

2. Филиал ФГБОУ ВО «Национальный исследовательский университет «МЭИ» (г. Смоленск). Отзыв подписан ведущим программистом лаборатории Луферовым В.С. (к.т.н., специальность 05.13.01).

Замечания:

- Одним из пунктов, вынесенных автором в качестве научной новизны, является методика объединения профилей пользователей из различных социальных сетей. В экспериментах показано, что текущего набора алгоритмов сравнения данных профилей, примененных последовательно, достаточно для объединения профилей 100 человек (всего 204 профиля) разных социальных сетей. Однако, данная методика может показать иной уровень эффективности при обработке данных более крупной выборки.

- В списке публикаций встречаются опечатки и ошибки форматирования.

3. ФГБОУ ВО «Финансовый университет при Правительстве Российской Федерации» (г. Москва). Отзыв подписан доцентом департамента анализа данных и машинного обучения Факультета информационных технологий и анализа больших данных Андрияновым Н.А. (к.т.н., специальность 05.13.18).

Замечания:

- В описании алгоритма формирования обучающей выборки Шаг №4 включает в себя предобработку данных. Однако предобработка – это достаточно важный шаг, детальное описание которого в автореферате опущено.

- В таблице 1 и далее приводятся метрики точности разных моделей. Однако отсутствует информация о распределении классов в обучающей и

тестовой выборках, без которой сложно судить о качестве работы модели, даже если точность составляет 87%.

- В таблице 2 в качестве алгоритма классификации указана «Линейная регрессия». Вероятно, автор имел в виду логистическую регрессию?

4. ФГАОУ ВО «Казанский (Приволжский) федеральный университет» (г. Казань). Отзыв подписан доцентом кафедры информационных систем Невзоровой О.А. (к.т.н., специальность 05.13.17).

Замечания:

- Недостатком автореферата является тот факт, что автором не указан объем применяемых им словарей: ключевых фраз на базе тезауруса WordNet-Affect и авторских символов выражения эмоций, которые используются при генерации обучающей выборки

Выбор официальных оппонентов и ведущей организации обосновывается их высокой компетенцией, научными достижениями и наличием публикаций в соответствующей отрасли наук.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований:

проведено сравнение современных интеллектуальных методов анализа текстовых данных, их возможностей и ограничений в рамках психолингвистического и сентимент-анализа ресурсов социальных сетей;

разработан алгоритм формирования обучающей выборки, состоящей из открытых русскоязычных текстовых ресурсов социальных сетей, классифицированных по семи эмоциям, отличающийся совместным использованием словарей авторских символов выражения эмоций и ключевых фраз;

предложен подход к определению психологических характеристик пользователя социальных сетей посредством психолингвистического анализа текстовых сообщений в социальных сетях с использованием методов машинного обучения.

Теоретическая значимость исследования обоснована тем, что:

доказаны положения о необходимости развития алгоритмов классификации

неструктурированных данных социальных сетей по классам тональности, вносящие вклад в развитие моделей и методов обработки текстовой информации, которые основаны на совместном применении искусственных нейронных сетей и семантических подходов.

Применительно к проблематике диссертации результативно (эффективно, то есть с получением обладающих новизной результатов)

использован комплекс существующих методов машинного обучения с учителем на основе рекуррентных, свёрточных, двунаправленных рекуррентных нейронных сетей, семантических методов и онтологических моделей;

изложены основные научные положения, гипотезы и рекомендации, позволяющие повысить эффективность построения социального портрета пользователя социальной сети на основе психолингвистического анализа текстовых данных пользователя;

раскрыты принципиальные ограничения методов машинной обработки, заключающиеся в необходимости обучения с подкреплением классификатора, в задачах определения психологических характеристик человека путем обработки коротких текстов социальных сетей;

изучены алгоритмы формирования обучающей выборки, алгоритмы сентимент-анализа, базирующиеся на использовании лингвистических словарей и методах машинного обучения, подходы к объединению профилей пользователей и алгоритмы построения психологического портрета;

проведена модернизация существующих алгоритмов формирования обучающей выборки текстов социальных сетей, подходов к сопоставлению профилей пользователей, алгоритма анализа эмоциональной окраски русскоязычных текстовых данных социальных сетей в системах поддержки принятия решений для обеспечения качественной и эффективной оценки профиля пользователя.

Значение полученных соискателем результатов исследования для практики подтверждается тем, что:

разработан и внедрен в эксплуатацию новый программный комплекс

психолингвистического и сентимент-анализа открытых текстовых русскоязычных данных профилей пользователей социальных сетей. Применение разработанного программного комплекса сократило временные затраты в задачах подбора персонала на основании результатов психолингвистического анализа и анализа эмоциональной окраски текстовых ресурсов профилей соответствующих пользователей. В абсолютных значениях экономия составила более 14 часов для каждой итерации подбора персонала; в рамках проекта «Система интеллектуального поиска и анализа Интернет-СМИ в социальных сетях» в задачах поиска пользователей с учетом параметров достигнуто среднее сокращение времени поиска необходимых профилей на 40%;

определены границы применимости разработанных методов и алгоритмов психолингвистического и сентимент-анализа социальных сетей;

создан и применен программный комплекс классификации профилей пользователей по психологическим характеристикам и анализу высказываний об объектах реального мира в форме текстовых сообщений;

представлены предложения по дальнейшему совершенствованию моделей и алгоритмов психолингвистического и сентимент-анализа текстовых данных социальных сетей.

Оценка достоверности результатов исследования выявила:

для экспериментальных работ результаты получены на основе корректного использования методов анализа неструктурированных ресурсов и современного программного обеспечения, подтверждены вычислительными экспериментами и результатами практического применения в лабораторных и производственных условиях ФГБОУ ВО Ульяновский государственный технический университет, УОСОО «Федерация бадминтона», ООО «ФаззиЛаб» и ФНПЦ АО "НПО"Марс";

теория построена на известных научных данных, которые в полной мере согласуются с ранее опубликованными данными по теме диссертационного исследования;

идея базируется на анализе и обобщении передового опыта и практических

исследованиях ряда российских и зарубежных исследователей по теме диссертации;

использованы сравнения авторских данных и данных, полученных ранее по рассматриваемой тематике для схожего набора данных на английском языке;

установлено качественное совпадение результатов, полученных автором, с опубликованными ранее результатами аналогичных исследований в области психолингвистического и сентимент-анализа текстовых данных социальных сетей;

использованы современных методики сбора и обработки текстовой информации, обеспечивающие воспроизводимость и достоверность результатов.

Личный вклад соискателя состоит в: анализе научных и патентных источников по теме диссертационной работы, обработке и интерпретации аналитической информации, разработке моделей, методов и алгоритмов, планировании и проведении экспериментов, формулировке выводов, внедрении полученных результатов. Все основные исследования проведены лично автором, либо при его непосредственном участии.

В ходе защиты диссертации было высказано следующее критическое замечания – при формировании обучающей выборки не учитываются стилистические особенности русскоязычных текстовых ресурсов, влияющих на процесс классификации текстов по классам эмоциональной окраски.

Соискатель Андреев И.А. ответил на задаваемые ему в ходе заседания вопросы и привел собственную аргументацию: для формирования обучающей выборки использовались короткие текстовые сообщения, которые отличались использованием одного авторского символа выражения эмоций, что снижало вероятность влияния стилистических особенностей на результат классификации текстов с использованием полученной обучающей выборки.

На заседании 14.09.2022 диссертационный совет принял решение за решение научной задачи формирования алгоритмов и методов психолингвистического и сентимент-анализа русскоязычных текстовых данных

социальных сетей, имеющей значение для развития технической отрасли знаний, присудить Андрееву И.А. ученую степень кандидата технических наук.

При проведении тайного голосования диссертационный совет в количестве 18 человек, из них 7 докторов наук по специальности рассматриваемой диссертации, участвовавших в заседании, из 23 человек, входящих в состав совета, дополнительно введены на разовую защиту 0 человек, проголосовали: за 18, против нет, непроголосовавших членов нет.

Председатель

диссертационного совета

Ученый секретарь

диссертационного совета



Ярушкина Надежда Глебовна

Наместников Алексей Михайлович

14.09.2022