

ЗАКЛЮЧЕНИЕ ДИССЕРТАЦИОННОГО СОВЕТА Д 002.073.05, СОЗДАННОГО
НА БАЗЕ ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО УЧРЕЖДЕНИЯ
«ФЕДЕРАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР «ИНФОРМАТИКА И
УПРАВЛЕНИЕ» РОССИЙСКОЙ АКАДЕМИИ НАУК», ПО ДИССЕРТАЦИИ
НА СОИСКАНИЕ УЧЕНОЙ СТЕПЕНИ КАНДИДАТА НАУК

аттестационное дело № _____
решение диссертационного совета от 28 февраля 2019 г. № 3

О присуждении Потапенко Анне Александровне, гражданке Российской Федерации, ученой степени кандидата физико-математических наук.

Диссертация «Семантические векторные представления текста на основе вероятностного тематического моделирования» по специальности 05.13.17 — «Теоретические основы информатики» принята к защите 13 декабря 2018 года (протокол заседания №24) диссертационным советом, созданным на базе федерального государственного учреждения «Федеральный исследовательский центр «Информатика и управление» Российской академии наук» (ФИЦ ИУ РАН), 119333, г. Москва, ул. Вавилова, д. 40, на основании приказа Министерства образования и науки Российской Федерации №783/нк от 24 июня 2016 г.

Соискатель Потапенко Анна Александровна, 1992 года рождения, в 2014 году окончила федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Московский Государственный Университет имени М.В. Ломоносова» по специальности «Прикладная математика и информатика». В 2018 г. окончила очную аспирантуру федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский университет «Высшая школа экономики», образовательная программа 05.13.17 «Теоретические основы информатики». Работает инженером-исследователем в компании Google UK Limited (Великобритания). Диссертация выполнена на базовой кафедре Яндекс департамента больших данных и информационного поиска факультета компьютерных наук ФГАОУ ВО Национальный исследовательский университет

«Высшая школа экономики» (ведомственная принадлежность – Правительство Российской Федерации).

Научный руководитель – доктор физико-математических наук, профессор РАН Воронцов Константин Вячеславович. Работает в ФГАОУ ВО «Московский Физико-Технический Институт (государственный университет)» в должности руководителя лаборатории машинного интеллекта.

Официальные оппоненты

Бессмертный Игорь Александрович, доктор технических наук, профессор факультета программной инженерии и компьютерной техники ФГАОУ ВО «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики»,

Тутубалина Елена Викторовна, кандидат физико-математических наук, старший научный сотрудник научно-исследовательской лаборатории «Хемоинформатика и молекулярное моделирование» Химического института им. А.М. Бутлерова ФГАОУ ВО «Казанский (Приволжский) федеральный университет»

дали положительные отзывы на диссертацию.

Ведущая организация Федеральное государственное бюджетное учреждение науки «Институт системного программирования им. В.П. Иванникова Российской академии наук» (ИСП РАН) в своем положительном отзыве, подписанном Аветисяном Арутюном Ишхановичем, директором ИСП РАН, указала, что диссертация является законченной научно-квалификационной работой, содержащей новые научные результаты в области семантического анализа текстов. Все результаты, выносимые на защиту, строго обоснованы и подтверждены в ходе вычислительных экспериментов. Автореферат диссертации, опубликованные статьи и тезисы докладов достаточно полно отражают основное содержание работы. В заключении отмечается, что диссертация соответствует всем критериям, установленным Положением о присуждении ученых степеней, предъявляемым к диссертациям на соискание ученой степени кандидата физико-математических наук, а ее автор – Потапенко Анна Александровна заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 05.13.17 – «Теоретические основы информатики»

Соискатель имеет 12 опубликованных работ по теме диссертации, 7 из которых опубликованы в рецензируемых научных изданиях, рекомендованных ВАК. **Наиболее значимые публикации:**

1. Воронцов К. В., Потапенко А. А. Регуляризация, робастность и разреженность вероятностных тематических моделей // Компьютерные исследования и моделирование. — 2012. — Т. 4, № 4. — С. 693–706.
2. Potapenko A. A., Vorontsov K. V. Robust PLSA Performs Better Than LDA // 35th European Conference on Information Retrieval, ECIR-2013, Moscow, Russia, 24-27 March 2013. — Lecture Notes in Computer Science (LNCS), Springer Verlag-Germany, 2013. — Pp. 784–787.
3. Vorontsov K. V., Potapenko A. A. Additive Regularization of Topic Models // Machine Learning Journal. — 2015. — Vol. 101. — Pp. 303–323.
4. Воронцов К. В., Потапенко А. А. Регуляризация вероятностных тематических моделей для повышения интерпретируемости и определения числа тем // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции Диалог. (Бекасово, 4–8 июня 2014 г.). — Вып. 13 (20). — М: Изд-во РГГУ, 2014. — С. 676–687.
5. Potapenko A., Popov A., Vorontsov K. Interpretable Probabilistic Embeddings: Bridging the Gap Between Topic Models and Neural Networks // AINL: Artificial Intelligence and Natural Language Conference / Ed. by Andrey Filchenkov, Lidia Pivovarova, Jan Zizka. — Vol. 789 of Communications in Computer and Information Science. — Springer International Publishing, 2017. — Pp. 167–180.
6. Vorontsov K. V., Potapenko A. A. Tutorial on Probabilistic Topic Modeling: Additive Regularization for Stochastic Matrix Factorization // Analysis of Images, Social networks and Texts (AIST 2014). — Vol. 436 of Communications in Computer and Information Science. — Springer International Publishing Switzerland, 2014. — Pp. 29–46.
7. Vorontsov K. V., Potapenko A. A., Plavin A.V. Additive Regularization of Topic Models for Topic Selection and Sparse Factorization // The Third International Symposium On Learning And Data Sciences (SLDS 2015). — Vol. 9047. — Springer, A. Gammerman et al. (Eds.), LNAI, 2015. — P. 193–202.

Недостовверных сведений о работах, опубликованных соискателем ученой степени, не выявлено. В коллективных публикациях автору принадлежат те их части, которые легли в основу положений диссертации, выносимых на защиту.

На автореферат диссертации поступил отзыв от старшего научного сотрудника Лаборатории комбинаторной алгебры Института естественных наук и математики Уральского федерального университета имени первого Президента России Б.Н. Ельцина, к.т.н. Браславского П. И. Отзыв **в целом положительный**, в качестве замечаний отмечается, что из реферата неясно, на каких данных проводились эксперименты, описанные в разделах 4.2 и 4.3 диссертации; в других случаях упоминаются источники данных, но отсутствуют характеристики коллекций.

Выбор официальных оппонентов и ведущей организации обосновывается тем, что они являются широко известными в области машинного обучения и анализа текста на естественном языке, имеют публикации в авторитетных журналах в соответствующих сферах исследований и могут определить научную и практическую значимость диссертации.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований:

разработаны вероятностные тематические модели коллекций текстовых документов, повышающие интерпретируемость и разреженность предметных тем, **предложена** формализация дистрибутивной гипотезы в рамках подхода аддитивной регуляризации тематических моделей,

доказана применимость вероятностных тематических представлений слов к задачам определения семантической близости слов, сегментов текста, документов и нетекстовых мета-данных.

Теоретическая значимость исследования обоснована тем, что:

доказаны теоремы, позволяющие оценивать параметры моделей векторного представления текстов с помощью итерационного ЕМ-алгоритма (Expectation-Maximization), получены формулы Е-шага и М-шага,

применительно к проблематике диссертации результативно использованы методы регуляризации некорректно поставленных задач, численные методы оптимизации, методы дистрибутивной семантики,
раскрыты проблемы многокритериальной оптимизации вероятностных тематических моделей, использования мультимодальных данных и данных о совместной встречаемости слов, выбора числа тем в вероятностных тематических моделях, комбинирования требований разреженности, робастности, различности и интерпретируемости тем, предложены пути решения данных проблем,
изучены связи методов вероятностного тематического моделирования, дистрибутивной семантики и нейросетевых методов.

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

разработаны вычислительно эффективные алгоритмы построения тематических моделей векторного представления текстов, позволяющие совмещать свойство интерпретируемости компонент векторов и применимость таких векторов для определения семантической близости слов или документов,
определены свойства получаемых векторных представлений и качество решения стандартных задач семантического анализа текстов,
представлены рекомендации по выбору гиперпараметров предложенных моделей (коэффициентов регуляризации, весов модальностей, числа тем).

Оценка достоверности результатов исследования выявила:

для экспериментальных работ использовались стандартные наборы данных, в тексте диссертации предоставлены все необходимые детали для воспроизведения полученных результатов,
теория обоснована строгими математическими доказательствами теорем,
установлено совпадение авторских результатов с результатами из других статей для известных методов, обобщенных в рамках данной работы.

Личный вклад соискателя. Результаты по комбинированию тематических моделей с моделями дистрибутивной семантики получены автором лично, за исключением некоторых экспериментов, проведенных совместно с Поповым А.С.

Подход аддитивной регуляризации тематических моделей разрабатывался в соавторстве с Воронцовым К.В. Из работ, опубликованных совместно, в качестве

результатов диссертации приводятся только результаты, полученные лично автором.

На заседании 28 февраля 2019 года диссертационный совет принял решение присудить Потапенко Анне Александровне ученую степень кандидата физико-математических наук.

При проведении тайного голосования диссертационный совет в количестве 28 человек, из них 9 докторов наук по профилю рассматриваемой диссертации, участвовавших в заседании, из 35 человек, входящих в состав совета, проголосовали: за 28, против: 0, недействительных бюллетеней: 0.

Председатель диссертационного
совета Д 002.073.05, академик РАН



Handwritten signature of Yu.I. Zhuravlev

Журавлев Ю.И.

Ученый секретарь диссертационного
совета Д 002.073.05, д.ф.-м.н., профессор

Handwritten signature of V.V. Ryzanov

Рязанов В. В.

«4» марта 2019г.