

ЗАКЛЮЧЕНИЕ ДИССЕРТАЦИОННОГО СОВЕТА Д002.073.05  
НА БАЗЕ ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО УЧРЕЖДЕНИЯ  
«ФЕДЕРАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР «ИНФОРМАТИКА И  
УПРАВЛЕНИЕ» РОССИЙСКОЙ АКАДЕМИИ НАУК» ПО ДИССЕРТАЦИИ НА  
СОИСКАНИЕ УЧЕНОЙ СТЕПЕНИ КАНДИДАТА ТЕХНИЧЕСКИХ НАУК

аттестационное дело № \_\_\_\_\_

решение диссертационного совета от 17 декабря 2020 г. № 6

О присуждении Апишеву Мурату Азаматовичу, гражданину Российской Федерации, ученой степени кандидата технических наук.

**Диссертация** «Эффективная реализация алгоритмов тематического моделирования с аддитивной регуляризацией» по специальности 05.13.17 — «Теоретические основы информатики» принята к защите 15 октября 2020 г., протокол №4, диссертационным советом Д002.073.05 на базе Федерального государственного учреждения «Федеральный исследовательский центр «Информатика и управление» Российской академии наук», 119333, г. Москва, ул. Вавилова, д. 44/2, созданным на основании приказа Министерства Образования и Науки Российской Федерации №783/нк от 24.06.2016 г.

Заседание диссертационного совета проводилось в удаленном интерактивном режиме. Основание — приказ ФИЦ ИУ РАН № 3-78 от 11.12.2020 г.

**Соискатель** Апишев Мурат Азаматович, 1993 года рождения, в 2017 году окончил Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Московский государственный университет имени М.В. Ломоносова» по специальности «Прикладная математика и информатика».

В 2020 году закончил очную аспирантуру Федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Московский государственный университет имени М.В. Ломоносова». С 21.09.2020 был прикреплен к аспирантуре Федерального

исследовательского центра «Информатика и управление» Российской Академии Наук для сдачи кандидатского экзамена.

Диссертация выполнена на кафедре математических методов прогнозирования факультета вычислительной математики и кибернетики Федерального государственного бюджетного учреждения высшего профессионального образования «Московский государственный университет имени М.В. Ломоносова».

Соискатель в настоящий момент не работает.

**Научный руководитель** — доктор физико-математических наук, профессор РАН Воронцов Константин Вячеславович. Работает в ФГАОУ ВО «Московский физико-технический институт (национальный исследовательский университет)» в должности руководителя лаборатории машинного интеллекта.

#### **Официальные оппоненты**

Оселедец Иван Валерьевич, доктор физико-математических наук, профессор АНОО ВО «Сколковский институт науки и технологий»,

Краснов Федор Владимирович, кандидат технических наук, директор департамента информационных систем управления АО «Наумен»

**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, врио директора ИСП РАН, указала, что диссертация является законченной научно-квалификационной работой, содержащей новые результаты в областях алгоритмов обучения тематических моделей и методов использования аддитивной регуляризации для анализа данных. 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 технических наук, а ее автор — Апишев Мурат Азаматович заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.17 — «Теоретические основы информатики».

Соискатель имеет 11 опубликованных работ по теме диссертации, 8 из которых опубликованы в рецензируемых научных изданиях, рекомендованных ВАК. **Наиболее значимые публикации:**

1. K. Vorontsov, O. Frei, M. Apishev., P. Romov, M. Suvorova, A. Yanina. Non-bayesian additive regularization for multimodal topic modeling of large collections // Proceedings of the CIKM 2015 Workshop on Topic Models: Post-Processing and Applications. — New York, USA: ACM, 2015. — Pp. 29–37.

2. Vorontsov K. V., Frei O. I., Apishev M. A., Romov P. A., Dudarenko M. A. BigARTM: Open Source Library for Regularized Multimodal Topic Modeling of Large Collections // AIST'2015, Analysis of Images, Social networks and Texts. Communications in Computer and Information Science (CCIS). — Switzerland: Springer International Publishing, 2015. — Pp. 370–384.

3. O. Frei, M. Apishev. Parallel non-blocking deterministic algorithm for online topic modeling // AIST'2016, Analysis of Images, Social networks and Texts. Communications in Computer and Information Science (CCIS). — Vol. 661. — Switzerland: Springer International Publishing, 2016. — Pp. 132–144.

4. Apishev M., Koltcov S., Koltsova O., Nikolenko S., Vorontsov K. Mining Ethnic Content Online with Additively Regularized Topic Models // Computación y Sistemas. — 2016. — Vol. 20, no. 3. — Pp. 387–403.

5. Apishev M., Koltcov S., Koltsova O., Nikolenko S., Vorontsov K., Additive Regularization for Topic Modeling in Sociological Studies of User-Generated Texts // Advances in Computational Intelligence, 15th Mexican International Conference on Artificial Intelligence, MICAI 2016. Proceedings, Part I. Lecture Notes in Artificial Intelligence.

6. Kochedykov D., Apishev M., Golitsyn L., Vorontsov K. Fast and modular regularized topic modelling // Proceeding Of The 21St Conference Of FRUCT

Association. The seminar on Intelligence, Social Media and Web (ISMW). — 2017. — Pp. 182–193.

7. M. Apishev, K. Vorontsov. Learning Topic Models With Arbitrary Loss // Proceedings of the 26th Conference of FRUCT Association. — 2020. — Pp. 30–37.

8. М. Апишев. Эффективные реализации алгоритмов тематического моделирования // Труды ИСП РАН. — 2020. — Т. 32, №. 1. — С. 137–152.

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.

**Выбор официальных оппонентов и ведущей организации** обосновывается тем, что оппоненты и сотрудники ведущей организации являются известными специалистами в области разработки и исследования моделей и алгоритмов анализа данных и обработки естественного языка, имеют публикации в авторитетных научных журналах в соответствующих сферах исследований и могут определить научную и практическую значимость диссертации.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований:

**разработана** вероятностная аддитивно регуляризованная мультимодальная тематическая модель для выделения специфических тем по заданному словарю,

**предложены** вычислительно эффективные алгоритмы обучения аддитивно регуляризованных тематических моделей,

**доказана** применимость гиперграфовых тематических моделей для анализа данных, имеющих транзакционную природу.

**Теоретическая значимость** исследований обоснована тем, что:

экспериментально **доказаны** возможность использования модификации итерационного EM-алгоритма (Expectation-Maximization) с E-шагом без нормировки для ускорения обучения тематических моделей и улучшения их качества, превосходство разработанной модели для выделения специфических тем над более простыми моделями в задаче информационного поиска,



**применительно к проблематике диссертации результативно использованы** методы регуляризации некорректно поставленных задач и численные методы оптимизации,

**раскрыты** сложности, возникающие при обучении тематических моделей на больших объемах данных, с одновременным использованием нескольких критериев регуляризации, а также при обучении моделей с помощью EM-алгоритма с ускоренным E-шагом,

**проведена модернизация** алгоритмов обучения мультимодальных аддитивно регуляризованных тематических моделей.

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

**разработаны** эффективный параллельный асинхронный алгоритм DetAsync для онлайн-обучения аддитивно регуляризованных тематических моделей и модификация алгоритмов обучения с разреженным хранением параметров,

**определены** перспективы практического использования предложенной модели для выделения специфических тем в задаче информационного поиска,

**создана** программная реализация предложенных алгоритмов и комбинации регуляризаторов в библиотеке с открытым кодом BigARTM,

**представлены** рекомендации по использованию предложенной модели для выделения специфических тем в задаче поиска межнационального дискурса в текстах социальных сетей, а также рекомендации по выбору гиперпараметров в предложенных моделях (коэффициентов регуляризации, весов модальностей).

Разработанные алгоритмы включены в учебный курс «Машинное обучение и большие данные» кафедры АТП ФПМИ МФТИ, их реализации в составе библиотеки BigARTM используются в коммерческих проектах ООО «ДжиВиЭй» и лаборатории машинного интеллекта МФТИ, разработанная модель для выделения специфических тем использовалась в проекте Санкт-Петербургского филиала НИУ ВШЭ по анализу этно-социального дискурса в текстах социальной сети «ВКонтакте». Каждое из указанных внедрений подтверждено актом.

**Оценка достоверности результатов исследования выявила:**

для экспериментальных работ использовались стандартные для данной области наборы данных, метрики качества и открытое программное обеспечение, в тексте диссертации представлены все необходимые детали для воспроизведения полученных результатов,

теория, лежащая в основе работы, многократно излагалась в научных публикациях и является логически и математически обоснованной,

идеи базируются на анализе подходов к повышению эффективности обучения тематических моделей, предложенных в научной литературе.

**Личный вклад соискателя.** Разработка и реализация первой версии алгоритма DetAsync производились совместно с Фреем А.И., вклад соискателя является решающим. Устранение ряда технических недоработок алгоритма, первое сравнение его по производительности с конкурентами и все прочие результаты диссертационной работы, заключающиеся в разработке и реализацией алгоритмов и моделей, а также проведении экспериментов, выполнены соискателем лично. Из работ, опубликованных совместно, в качестве результатов диссертации приводятся только результаты, полученные лично ее автором.

На заседании 17 декабря 2020 года диссертационный совет принял решение присудить Апишеву Мурату Азаматовичу ученую степень кандидата технических наук.

При проведении открытого голосования диссертационный совет в количестве 21 человека, из них 7 докторов наук по профилю рассматриваемой диссертации, участвовавших в заседании, из 30 человек, входящих в состав совета, дополнительно введены на разовую защиту 0 человек, проголосовали «за» 21, «против» 0.

И.о. председателя диссертационного совета Д002.073.05,

доктор технических наук

Ученый секретарь диссертационного совета Д002.073.05,

кандидат технических наук

«17» декабря 2020



*Handwritten signature of I.A. Matveev*

Матвеев И.А.

*Handwritten signature of I.A. Reyer*

Рейер И.А.