

ЗАКЛЮЧЕНИЕ ДИССЕРТАЦИОННОГО СОВЕТА Д 002.073.05
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 13.04.2017 № 6.

О присуждении Кузьмину Арсентию Александровичу, гражданину Российской Федерации, ученой степени кандидата физико-математических наук.

Диссертация «Иерархическая классификация коллекций документов» по специальности 05.13.17 – «Теоретические основы информатики» принята к защите 9 февраля 2017 г., протокол №2, диссертационным советом Д 002.073.05 на базе Федерального государственного учреждения «Федеральный исследовательский центр «Информатика и управление» Российской академии наук», 119333, г. Москва, ул. Вавилова, д.44, корп.2, созданным на основании приказа Министерства Образования и Науки Российской Федерации №783/нк от 24.06.2016 г.

Соискатель Кузьмин Арсентий Александрович, 1992 года рождения, в 2015 году окончил ФГАОУ ВО «Московский физико-технический институт (государственный университет)».

С 2015 по настоящее время является аспирантом ФГАОУ ВО «Московский физико-технический институт (государственный университет)».

Диссертация выполнена на кафедре интеллектуальных систем Московского физико-технического института (государственного университета) Министерства образования и науки Российской Федерации.

Научный руководитель – доктор физико-математических наук, Стрижов Вадим Викторович, Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Отдел интеллектуальных систем, научный сотрудник.

Официальные оппоненты:

Миркин Борис Григорьевич, доктор технических наук, доцент, Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет «Высшая школа экономики», Факультет компьютерных наук, Департамент анализа данных и искусственного интеллекта, профессор,

Александров Михаил Аронович, кандидат физико-математических наук, доцент, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации», Кафедра системного анализа и информатики, доцент

дали **положительный** отзыв на диссертацию.

Ведущая организация Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный университет имени М. В. Ломоносова» (Факультет ВМК) в своем **положительном** заключении, подписанном Сергеем Исаевичем Гуровым, кандидатом физико-математических наук, доцентом кафедры Математических методов прогнозирования, указала, что диссертация А.А.КУЗЬМИНА является законченной научно-квалификационной работой, содержащей новое решение актуальной научной задачи – иерархической классификации коллекций документов и имеющей большое значение для развития соответствующей отрасли знаний. Все результаты, вынесенные на защиту, строго научно обоснованы. Автореферат диссертации, опубликованные статьи и тезисы докладов достаточно полно отражают основное содержание работы. В заключении ведущей организации отмечается, что диссертация соответствует всем критериям, установленным Положением о присуждении ученых степеней, предъявляемым к диссертациям на соискание ученой степени кандидата физико-математических наук по специальности 05.13.17 – «Теоретические основы информатики», а ее автор – КУЗЬМИН Арсентий Александрович – заслуживает присуждения ученой степени кандидата физико-математических наук по указанной специальности.

Соискатель имеет 10 работ по теме диссертации, опубликованных в рецензируемых научных изданиях, из которых 4 – в научных изданиях из Перечня, рекомендованного ВАК. Публикации представляют собой статьи в научных изданиях и тезисы выступлений на конференциях. Наиболее значимые публикации по теме диссертации:

1. Кузьмин А. А., Адуенко А. А., Стрижов В. В. Выбор признаков и оптимизация метрики при кластеризации коллекции документов // Известия ТулГУ, 3(2012). С. 119-131.
2. Кузьмин А. А., Стрижов В. В. Проверка адекватности тематических моделей коллекции документов. // Программная инженерия, 4 (2013). С. 16-20.
3. Кузьмин А. А., Адуенко А. А., Стрижов В. В. Тематическая классификация тезисов крупной конференции с использованием экспертной модели // Информационные технологии, 6 (2014). С. 22-26.
4. Златов А. С., Кузьмин А. А. Построение иерархической тематической модели крупной конференции // Искусственный интеллект и принятие решений, 3(2016). С. 77-86.

Результаты, изложенные в диссертации, получены автором самостоятельно. В коллективных публикациях автору принадлежат те их части, которые использованы в диссертации.

На диссертацию и автореферат поступили отзывы:

Хританкова Антона Сергеевича, к.ф.-м.н., доцента кафедры Алгоритмов и технологий программирования МФТИ (ГУ). Отзыв положительный, в качестве замечаний отмечено, что следовало бы указать время работы и объем использованной алгоритмами памяти в вычислительном эксперименте, а также сравнить их асимптотику с полученными теоретическими оценками. Для сравнения с другими алгоритмами следует также использовать общедоступные наборы данных, используемые другими исследователями, такие как веб-каталоги (например, данные Open Directory Project).

Выбор официальных оппонентов и ведущей организации обосновывается тем, что они являются широко известными специалистами в области машинного обучения, текстового анализа, классификации и кластеризации, 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 методика иерархической классификации частично размеченных коллекций документов с экспертной кластерной структурой тем в виде дерева, вероятностная модель классов документов, комплексы программ для численного исследования методов и проведения вычислительных экспериментов;

предложены метод иерархической классификации коллекций документов, вероятностная модель классов документов коллекции и методы оптимизации ее параметров, энтропийный метод оценки важности слов, способ оценки совместного апостериорного распределения параметров модели и классов неразмеченных документов, метод оценки вероятности принадлежности документа кластеру и метод ранжирования кластеров по релевантности заданному документу;

доказана эффективность и перспективность использования разработанных математических моделей и методов при решении задач иерархической классификации документов частично размеченных коллекций с экспертно заданной структурой тем в виде дерева.

Теоретическая значимость исследования обоснована тем, что:

доказаны теоремы об оценке апостериорного распределения параметров предложенной модели классов документов, об оптимальных значениях вариационных параметров, используемых в оценке апостериорного распределения, об эквивалентности качества предложенного способа оценки вероятности принадлежности документа классу методу, использующему принцип максимума апостериорной вероятности для оценки параметров, об оценке совместного апостериорного распределения параметров модели и классов неразмеченных документов, о сложности EM алгоритма оптимизации параметров и гиперпараметров предложенной модели;

применительно к проблематике диссертации результативно использованы методы вариационного байесовского вывода, методы локальной вариации, методы векторного представления слов, метод энтропийной оценки важности признаков и комплекс программ для предобработки и иерархической классификации текстовых документов;

проведена модернизация существующих подходов иерархической классификации путем обобщения взвешенной функции сходства и энтропийного метода оценки важности признаков на иерархический случай, а также методов оценки параметров взвешенной функции сходства.

Значение полученных соискателем результатов исследования для практики подтверждается тем, что предложенные в работе методы могут быть использованы для иерархической классификации коллекций текстов с учетом уже существующих экспертных моделей. В частности, на базе предложенных методов разработан программный комплекс для классификации тезисов крупной конференций и веб-сайтов компаний индустриального сектора.

Достоверность научных положений, результатов и выводов диссертации обеспечена корректным использованием линейной алгебры, математической статистики и теории вариационного байесовского вывода, математическими

доказательствами и экспериментальной проверкой предложенных методов на реальных задачах иерархической классификации.

Личный вклад соискателя: все теоретические и экспериментальные результаты, представленные в диссертационной работе, являются новыми и получены соискателем самостоятельно. Из совместных публикаций в качестве результатов диссертации приводятся только полученные лично автором.

На заседании 13 апреля 2017 года диссертационный совет принял решение присудить Кузьмину А.А. ученую степень кандидата физико-математических наук. При проведении тайного голосования диссертационный совет в количестве 26 человек, из них 10 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 -0, недействительных бюллетеней- 1.

Председатель диссертационного совета Д 002.073.05
доктор физико-математических наук
академик РАН



Журавлев Ю. И.

Ученый секретарь диссертационного совета Д 002.073.05
доктор физико-математических наук
профессор

Рязанов В. В.

13 апреля 2017 г.