

УТВЕРЖДАЮ



Проректор
по научной работе, к.ф.-м.н.

Баган Виталий
Анатольевич

« 05 » октября 2020 г.

ЗАКЛЮЧЕНИЕ

федерального государственного автономного образовательного
учреждения высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)» (МФТИ)

Диссертация «Единственность матричного разложения и сходимость регуляризованных алгоритмов в вероятностном тематическом моделировании» выполнена на кафедре анализа данных федерального государственного автономного образовательного учреждения высшего образования «Московский физико-технический институт (национальный исследовательский университет)» Министерства науки и высшего образования Российской Федерации.

В период подготовки диссертации с 2016 г. по настоящее время соискатель **ИРХИН Илья Александрович** трудовую деятельность не осуществлял.

В 2016 г. Ирхин И.А. окончил федеральное государственное автономное образовательное учреждение высшего образования «Московский физико-технический институт (государственный университет)» по направлению подготовки 01.04.02 Прикладная математика и информатика. В 2020 г. окончил очную аспирантуру МФТИ по направлению подготовки 09.06.01 Информатика и вычислительная техника.

Справки о сдаче кандидатских экзаменов выданы в 2020 г. федеральным государственным автономным образовательным учреждением высшего образования «Московский физико-технический институт (национальный исследовательский университет)» и федеральным исследовательским центром «Информатика и управление» Российской Академии Наук (ФИЦ ИУ РАН).

Научный руководитель – доктор физико-математических наук, профессор РАН Воронцов Константин Вячеславович. Основное место работы – федеральное государственное автономное образовательное учреждение высшего образования «Московский физико-технический институт (национальный исследовательский университет)», кафедра интеллектуальных систем, профессор.

По итогам обсуждения диссертации «Единственность матричного разложения и сходимость регуляризованных алгоритмов в вероятностном тематическом моделировании» *принято следующее заключение.*

Тема диссертации является актуальной для развития методологии анализа текстов на основе подхода аддитивной регуляризации тематических моделей поскольку даёт теоретическое обоснование корректности используемых методов, а также предлагает модификации алгоритма, улучшающие основные метрики качества тематических моделей.

Цели и задачи диссертации:

1. Получение достаточных условий сходимости алгоритма аддитивной регуляризации тематических моделей.
2. Получение достаточных условий для единственности матричного разложения в точке алгоритма аддитивной регуляризации тематических моделей.
3. Поиск модификаций исходного алгоритма аддитивной регуляризации тематических моделей, улучшающих сходимость и повышающих метрики качества тематических моделей.

Основные результаты диссертации:

1. Сформулирована и доказана теорема о достаточных условиях сходимости алгоритма аддитивной регуляризации тематических моделей. Возможность практического использования этих условий проверена в экспериментах на текстовых коллекциях.
2. Сформулирована и доказана теорема о достаточных условиях единственности стохастического матричного разложения. Возможность практического использования этих условий проверена в экспериментах на текстовых коллекциях.
3. Предложена и реализована модификация алгоритма аддитивной регуляризации тематических моделей, которая позволяет ускорить сходимость итерационного процесса.
4. Предложен и реализован метод разреживания тематической модели, не ухудшающий метрики качества модели.

Теорема о достаточных условиях единственности стохастического матричного разложения была получена совместно с Дербаносовым Р.Ю.

Остальные результаты диссертации *получены лично соискателем* при научном руководстве доктора физико-математических наук, профессора РАН, Воронцова К.В.

Научная новизна работы. Впервые получены достаточные условия сходимости алгоритма аддитивной регуляризации тематических моделей. Получены новые достаточные условия единственности стохастического матричного разложения. Их применение к задаче тематического моделирования позволило детализировать причины неединственности решения. Предложены

новые модификации постановки оптимизационной задачи и итерационного процесса в рамках аддитивной регуляризации тематических моделей.

Практическая ценность результатов диссертации заключается в том, что в работе были получены достаточные условия сходимости алгоритма аддитивной регуляризации тематических моделей, выполнение которых может быть легко гарантировано на практике. Также были получены достаточные условия на единственность стохастического матричного разложения, выполнение которых было подтверждено в экспериментах. Помимо этого, на основе теоретической части работы были предложены и реализованы модификации итерационного алгоритма и оптимизационной задачи, которые показали существенное улучшение качества получаемых тематических моделей в экспериментах. Это подтверждает применимость и полезность результатов работы при практическом построения тематических моделей.

Обоснованность и достоверность результатов и выводов

Выводы диссертации обоснованы аналитическими расчетами и подтверждены данными научных экспериментов, полученными с помощью разработанных алгоритмов и комплекса программ. Теоретическую и методологическую основу проведенных разработок и исследований составили труды отечественных и зарубежных авторов в области анализа текстов, а также технические решения, созданные и опубликованные в российских и зарубежных патентах и свидетельствах на изобретения РФ. Для анализа полученных результатов использовались методы статистической обработки данных.

Материалы диссертации опубликованы автором достаточно полно в трёх работах из списка журналов, рекомендованных ВАК:

1. Автоматическое извлечение атрибутов водителя из логов мобильного приложения такси / Селезнев Н.К., Ирхин И.А., Кантор В.В. // Труды МФТИ. — 2018. — Т. 10, No 3. — С. 5—15.
2. Сходимость алгоритма аддитивной регуляризации тематических моделей / Ирхин И.А., Воронцов К.В. // Труды Института математики и механики УрО РАН. — 2020. — Т. 26, No 3. — С. 57—68.
3. Проблемы устойчивости и единственности стохастического матричного разложения / Дербаносов Р. Ю., Ирхин И.А. // Журнал вычислительной математики и математической физики. — 2020. — Т. 60, No 3. — С. 19—28.

Личный вклад соискателя в работах с соавторами заключается в следующем:

В [1] предложена идея применения подхода тематического моделирования и критерии качества решению прикладной задачи; в [2] предложены достаточные условия сходимости, доказаны все утверждения и теоремы, реализованы и проведены все эксперименты; в [3] доказана основная лемма, реализованы и проведены все эксперименты.

Основные результаты работы докладывались на следующих научных конференциях и семинарах:

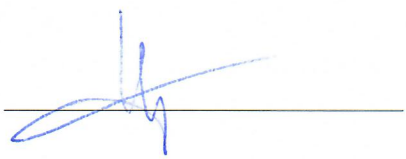
1. 5th International Symposium, Conformal and Probabilistic Prediction with Applications, 2016
2. Научный семинар Школы Анализа Данных, 2016.
3. Научный семинар лаборатории искусственного интеллекта, 2018.
4. Научный семинар Федерального исследовательского центра «Информатика и Управление» Российской Академии Наук, 2020.

Содержание диссертации соответствует паспорту специальности 05.13.17 – Теоретические основы информатики, в частности, пунктам:

- п. 5. Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечения разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений.
- п. 6. Разработка методов, языков и моделей человекомашинного общения; разработка методов и моделей распознавания, понимания и синтеза речи, принципов и методов извлечения данных из текстов на естественном языке.

Диссертация «Единственность матричного разложения и сходимостью регуляризованных алгоритмов в вероятностном тематическом моделировании» **ИРХИНА Ильи Александровича** рекомендуется к защите на соискание ученой степени кандидата физико-математических наук по специальности **05.13.17 – теоретические основы информатики.**

Заключение принято на заседании кафедры анализа данных МФТИ. Присутствовало на заседании 5 человек. Результаты голосования: «за» – 5 чел., «против» – нет, «воздержались» – нет, протокол № 2020-10-1 от 5 октября 2020 года.



Федотов Станислав Николаевич,
к.ф.-м.н., заместитель заведующего
кафедрой «Анализ данных» ФПМИ МФТИ