

Отзыв научного руководителя

о диссертационной работе Булатова Виктора Геннадьевича
«Методы оценивания качества и многокритериальной оптимизации тематических
моделей в библиотеке TopicNet», представленной на соискание ученой степени
кандидата технических наук по специальности 05.13.18 — «Математическое
моделирование, численные методы и комплексы программ»

Вероятностное тематическое моделирование — одно из направлений автоматической обработки текстов, активно развивающееся последние два десятилетия. Задача тематического моделирования заключается в том, чтобы по заданной коллекции текстовых документов определить, из каких слов состоит каждая тема, и какие темы представлены в каждом документе. Данная задача является некорректно поставленной задачей стохастического матричного разложения и имеет в общем случае бесконечное множество решений. Для получения наиболее подходящего решения в постановку задачи вводятся дополнительные критерии-регуляризаторы. В теории аддитивной регуляризации тематических моделей (ARTM) известны численные методы для решения данной задачи на основе алгоритма EM (expectation-maximization), которые в базовом варианте были реализованы в библиотеке с открытым кодом BigARTM. Практическое применение этой библиотеки наталкивается на ряд проблем: для построения хорошей модели пользователь должен разбираться в теории ARTM и выполнять серии экспериментов по подбору регуляризаторов и последовательности их включения, подбору коэффициентов регуляризации, числа тем и других гиперпараметров. Планирование таких экспериментов требует опыта и часто осложняется отсутствием готовых средств измерения качества тематических моделей, журнализации экспериментов и визуализации промежуточных результатов. Перед Булатовым В. Г. была поставлена задача разработать высокоуровневый программный комплекс поверх BigARTM, который в значительной степени автоматизировал бы проведение серий вычислительных экспериментов в прикладных проектах с применением тематического моделирования.

Булатов В. Г. начал работу в области тематического моделирования под моим научным руководством с 2014 года, будучи студентом магистратуры кафедры анализа данных ФИБТ. Поэтому к выполнению диссертационного исследования он приступил, имея хороший багаж теоретических знаний и практических навыков.

В ходе диссертационного исследования Булатовым В. Г. была разработана архитектура программного комплекса, предложены оригинальные решения для представления структуры эксперимента в виде «кубов гиперпараметров», параметризации оптимизационного критерия через «относительные коэффициенты регуляризации», представления результатов экспериментов и их журнализации с помощью «дерева экспериментов», формирования «рецептов моделирования» для сохранения универсальных сценариев обучения тематических моделей, подходящих для решения различных задач.

Значительная часть диссертационного исследования Булатова В. Г. посвящена проблеме оценивания качества тематических моделей. В работе представлен обширный

обзор литературы по данной теме, предложен и реализован оригинальный численный метод для вычисления новой характеристики согласованности отдельных тем в вероятностных тематических моделях — «внутритекстовой когерентности».

Диссертационное исследование выполнено Булатовым В. Г. в лаборатории машинного интеллекта МФТИ. Он внёс решающий вклад как в разработку проекта TopicNet, так и в выполнение прикладных проектов по заказу двух крупных российских компаний, в которых использовался программный комплекс TopicNet. Все проекты были успешно завершены, по их результатам оформлены РИДы и научные публикации, исходные коды основной части программного комплекса выложены в открытый доступ.

Хотелось бы особо отметить работу Булатова В. Г. по продвижению проекта TopicNet и популяризации тематического моделирования на научных и практических конференциях и семинарах. Также он постоянно консультирует студентов, сотрудников лаборатории машинного интеллекта МФТИ и внешних пользователей библиотеки TopicNet по техническим вопросам и её применению.

Всё это, несомненно, характеризует Булатова В. Г. как сложившегося самостоятельного исследователя и высококвалифицированного специалиста.

Диссертационная работа Булатова В. Г. «Методы оценивания качества и многокритериальной оптимизации тематических моделей в библиотеке TopicNet» является законченной самостоятельной научно-исследовательской работой.

Работа удовлетворяет всем требованиям, предъявляемым ВАК к диссертациям, представляемым на соискание ученой степени кандидата технических наук по специальности 05.13.18 — «Математическое моделирование, численные методы и комплексы программ». Диссертант Булатов В. Г. является квалифицированным специалистом по данной специальности и заслуживает присвоения ему ученой степени кандидата технических наук.

16 октября 2020 г.

Заведующий лаборатории
машинного интеллекта МФТИ,
профессор РАН, д.ф.-м.н.
Моб. +7 916 333 71 69
eMail: k.v.vorontsov@phystech.edu

 Воронцов К. В.



Письмо от руки К.В. Воронцова
КАНЦЕЛЯРИИ АДМИНИСТРАТИВНОГО ОТДЕЛА

