

ОТЗЫВ

председателя диссертационного совета
на диссертацию Боярова Андрея Александровича
на тему «Рандомизированный подход к обучению в условиях отсутствия разметки и
малого количества данных», представленную на соискание ученой степени
кандидата физико-математических наук по специальности
01.01.09 – Дискретная математика и математическая кибернетика.

Актуальность темы диссертации. Диссертационная работа Боярова Андрея Александровича «Рандомизированный подход к обучению в условиях отсутствия разметки и малого количества данных» посвящена актуальному направлению в области оптимизации: адаптивному построению оценок параметров модели при обучении в условиях отсутствия разметки данных или данных со слабой разметкой. Методы, относящиеся к этому научному направлению, активно исследуются и применяются в таких областях, как теория управления, робототехника, машинное обучение. В диссертационной работе исследуется использование методов адаптивной стохастической аппроксимации в алгоритмах машинного обучения и интеллектуального анализа данных.

Для успешной работы стандартных алгоритмов машинного обучения с учителем необходимы возможность вычисления градиента для оптимизируемого функционала среднего риска, знания о модели описания данных и данные для обучения как можно большего объёма. При этом важно отметить, что эти данные должны быть размечены и очищены от ошибок. Однако, в реальных задачах эти условия часто бывают труднодостижимы, что неизбежно влечёт появление неопределённостей, которые оказывают значительный эффект на классические методы оптимизации и приводят к деградации качества работы полученной в результате модели машинного обучения. Отсутствие заранее известной структуры и разметки данных является одним из источников таких неопределённостей. Для успешной работы с данными такого вида становится крайне актуальной задачей исследования и разработки новых подходов к обучению без учителя и кластеризации. Другой вид неопределённостей возникает в задачах обучения и классификации по малому количеству примеров. При такой постановке задачи для обучения классификатора доступно лишь несколько размеченных элементов для каждого класса. Алгоритм обучения должен иметь возможность адаптироваться для работы с новыми классами, не теряя при этом в качестве работы на старых классах. Необходимость в таком поведении алгоритма обучения часто встречаются во многих прикладных задачах.

Из всего вышеперечисленного следует актуальность темы диссертационного исследования А. А. Боярова.

Основные результаты диссертации приведены во второй главе и включают:

- Формулировку и доказательство общего рандомизированного алгоритма стохастической аппроксимации для кластеризации в модели смеси гауссовских

распределений, анализ качества и скорости работы представленного алгоритма при неизвестных, но ограниченных помехах.

- Формулировку и обоснование модификации общего рандомизированного алгоритма стохастической аппроксимации для кластеризации, применяемой в условиях смеси гауссовских распределений с разреженными параметрами.
- Обоснование подхода к обучению адаптивного классификатора по малому количеству размеченных данных, который использует для оптимизации параметров в многозадачной функции потерь рандомизированный алгоритм стохастической аппроксимации.

В третьей главе эффективность предложенных методов оценивания параметров продемонстрирована на реальных и синтетических данных.

Общий рандомизированный алгоритм стохастической аппроксимации для кластеризации в модели смеси гауссовских распределений, его модификация для случая смеси гауссовских распределений с разреженными параметрами и рандомизированный метод для обучения адаптивного классификатора по малому количеству размеченных примеров представлены в девяти работах автора, одна из которых опубликована в издании из списка ВАК (71 в Списке литературы), три опубликованы в изданиях, индексируемых в международных наукометрических базах Web of Science и Scopus (69, 70, 71 в Списке литературы), а одна (74 в Списке литературы) является свидетельством регистрации программы для ЭВМ.

Научная новизна и теоретическая значимость диссертационной работы обусловлены тем, что в диссертации А. А. Боярова предложен общий рандомизированный алгоритм стохастической аппроксимации для кластеризации в модели смеси гауссовских распределений, способный качественно работать при неизвестных, но ограниченных помехах. Впервые сформулирована модификация этого общего алгоритма, способная успешно работать в условиях смеси гауссовских распределений с разреженными параметрами. Впервые предложен и обоснован рандомизированный подход к обучению адаптивного классификатора по малому количеству размеченных примеров. Для всех предложенных в работе алгоритмов доказана сходимость получаемых с их помощью оценок.

Практическая значимость диссертационной работы не вызывает сомнений, что обосновано её актуальностью и широкой областью применимости полученных в ней результатов. К важным практическим результатам диссертации относится разработанный набор прикладных программ, реализующих предложенные в ходе диссертационного исследования методы. Эти программы показали свою практическую важность в задачах классификации рукописных букв и верификации авторства средневековых арабских манускриптов. Практическая значимость подтверждается рядом прикладных научно-исследовательских работ в рамках грантов РНФ и РФФИ.

Достоверность и обоснованность научных положений и выводов диссертационной работы определяются строгим математическим доказательством основных результатов, апробацией результатов на международных научных конференциях и публикациями в рецензируемых отечественных и зарубежных изданиях. Диссертация хорошо структурирована, аккуратно оформлена, текст диссертации тщательно отредактирован. Все положения диссертации, выносимые на защиту, строго обоснованы автором.

По работе есть замечания и пожелания.

1. Во второй главе при доказательстве сходимости оценок параметров, полученных с помощью рандомизированного алгоритма стохастической аппроксимации для трекинга, используются краткие формулировки из соответствующей статьи. Для улучшения понятности этого раздела в диссертации рекомендуется раскрыть эти формулировки.
2. Некоторые определения и формулировки во Введении страдают от излишней нестрогости и обобщённости.
3. В третьей главе при исследовании влияния внешних возмущений на качество работы общего рандомизированного алгоритма стохастической аппроксимации для кластеризации слабо обоснованы причины выбора рассматриваемых в работе видов возмущений.

Указанные замечания и пожелания не умаляют общей высокой оценки, которую заслуживает диссертационная работа Боярова Андрея Александровича. Все результаты, полученные в диссертации, являются новыми, ясно сформулированными и доказанными. Работа является законченным научным исследованием, выполненным на высоком научном уровне, и обладает несомненной теоретической и практической ценностью.

Считаю, что диссертационная работа Боярова Андрея Александровича на тему «Рандомизированный подход к обучению в условиях отсутствия разметки и малого количества данных» соответствует основным требованиям, установленным Приказом от 01.09.2016 № 6821/1 «О порядке присуждения ученых степеней в Санкт-Петербургском государственном университете», а ее автор Бояров Андрей Александрович заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 01.01.09 – Дискретная математика и математическая кибернетика. Пункт 11 указанного Порядка диссертантом не нарушен.

Председатель диссертационного совета
Доктор технических наук, профессор
Санкт-Петербургского государственного университета



04.09.2020

Фрадков Александр Львович