

ОТЗЫВ НАУЧНОГО РУКОВОДИТЕЛЯ

на диссертацию Боярова Андрея Александровича
«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

Тема диссертационного исследования А. А. Боярова относится к актуальному направлению в науке — разработке новых алгоритмов машинного обучения, применимых при обучении в условиях неразмеченных и мало размеченных данных. Для успешной работы основных стандартных алгоритмов машинного обучения с учителем необходимы чёткая модель данных, возможность вычисления градиента для оптимизируемой функции потерь (функционала качества) и как можно большее количество тренировочных данных, о вероятностном распределении которых часто предполагается близость к нормальному распределению. Однако, в реальных условиях эти требования часто бывают не выполнены: часто градиент функции потерь нельзя вычислить напрямую и даже гипотеза о центрированности данных относительно некоторых центроидов не подтверждается (например, данные могут по своей природе иметь разреженную структуру). В таких условиях стандартные универсальные методы дают смещённую оценку искомых параметров, настраиваемых при обучении. Приведенные соображения обуславливают необходимость разработки для подобных случаев новых методов, способных работать в таких нестандартных условиях. Стандартные алгоритмы машинного обучения с учителем успешно работают, только когда они обучаются на большом объёме размеченных данных. Задачи самообучения исследуются при обработке неразмеченных данных. Одна из важнейших — задача кластеризации. Отсутствие заранее известной структуры и разметки данных является источником существенных неопределённостей, для работы с которым необходимо разрабатывать новые подходы. Другой вид неопределённостей,

связанный с обработкой слабо размеченных данных, возникает в задачах обучения и классификации по малому количеству примеров. В таких задачах алгоритмы обучения должны иметь возможность адаптироваться для работы с новыми классами, не теряя при этом в качестве работы на старых классах. Методы этого семейства крайне востребованы, так как необходимость в адаптации к новым классам всего по нескольким примерам возникает во многих прикладных задачах.

Андрею Александровичу было поставлено три основные задачи: исследовать, разработать и математически обосновать общий алгоритм кластеризации в модели смеси гауссовых распределений, способный качественно работать при неизвестных, но ограниченных помехах, и реализующий идею «онлайн» обучения; исследовать, разработать и математически обосновать модификацию общего алгоритма кластеризации для условий смеси гауссовых распределений с разреженными параметрами; исследовать, разработать и математически обосновать метод обучения адаптивного классификатора по малому числу размеченных примеров на каждый класс. В рамках первой задачи А. А. Бояровым был предложен и исследован общий рандомизированный алгоритм стохастической аппроксимации с возмущением на входе для кластеризации в модели смеси гауссовских распределений. Была доказана теорема о состоятельности оценок параметров, получаемых с помощью этого метода. В контексте второй задачи А. А. Бояров представил модификацию алгоритма стохастической аппроксимации с рандомизацией на входе для первой задачи, способную эффективно работать в условиях смеси гауссовских распределений с разреженными параметрами. Для третьей задачи А. А. Бояров рассмотрел новую многозадачную функцию потерь и описал новый рандомизированный подход к обучению адаптивного классификатора по малому количеству размеченных примеров. Для разработанных в рамках этого подхода двух новых алгоритмов были доказаны теоремы о состоятельности получаемых с их помощью оценок. На тестовых примерах эти алгоритмы показали своё превосходство в качестве работы над стандартными методами

обучения по малому количеству данных. Все теоретические результаты были апробированы с помощью моделирования на синтетических и реальных данных с помощью разработанного соискателем набора программных средств.

Результаты исследований опубликованы в девяти научных работах и были представлены на международных и российских конференциях.

За время работы над диссертацией Андрей Александрович проявила себя инициативным самостоятельным исследователем, способным преодолевать возникающие трудности.

Считаю, что диссертационная работа А. А. Боярова соответствует всем требованиям Положения «О порядке присуждения учёных степеней в Санкт-Петербургском государственном университете», установленным Приказом № 6821/1. от 01.09.2016 года, а её автор заслуживает присуждения искомой ученой степени кандидата физико-математических наук по специальности 01.01.09 – Дискретная математика и математическая кибернетика.

Профессор кафедры системного
программирования СПбГУ,
доктор физико-математических
наук, профессор

О. Н. Граничин

