

ОТЗЫВ

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.

Диссертационная работа Боярова Андрея Александровича «Рандомизированный подход к обучению в условиях отсутствия разметки и малого количества данных» посвящена актуальному направлению — оцениванию параметров модели при обучении в условиях неразмеченных и мало размеченных данных. Для успешной работы основных стандартных алгоритмов машинного обучения с учителем необходима известная модель данных, возможность вычисления градиента для оптимизируемой функции потерь и как можно большее число размеченных тренировочных данных. Однако, в реальных условиях эти требования часто бывают невыполнены. Отсутствие заранее известной структуры и разметки данных является источником неопределенностей, для работы с которым необходимо разрабатывать новые подходы кластеризации. Другой вид неопределенностей, связанный с обработкой слабо размеченных данных, возникает в задачах обучения и классификации по малому количеству примеров. В таких задачах алгоритм должен 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. В этой связи актуальность темы диссертации не вызывает сомнений.

Научная новизна диссертационной работы не подлежит сомнению, так как в диссертации предложен общий рандомизированный алгоритм стохастической аппроксимации для кластеризации в модели смеси гауссовских распределений, способный качественно работать при неизвестных, но ограниченных помехах. Впервые представлена модификация этого алгоритма, способная качественно работать в условиях смеси гауссовских распределений с разреженными параметрами. Впервые представлен рандомизированный подход к обучению адаптивного классификатора по малому количеству размеченных примеров. Для всех предложенных в работе алгоритмов доказана сходимость получаемых с их помощью оценок.

Ценность научных работ соискателя заключается в развитии областей стохастической аппроксимации и оценивания параметров. Основной теоретический вклад состоит в описании общего рандомизированного алгоритма кластеризации в модели смеси гауссовых распределений, способного качественно работать при неизвестных, но ограниченных помехах, и формулировании условий сходимости получаемых с его помощью оценок параметров; описании модификации этого алгоритма стохастической

аппроксимации с рандомизацией на входе для кластеризации в условиях смеси гауссовых распределений с разреженными параметрами и формулировании условий сходимости получаемых с его помощью оценок параметров; описании новой многозадачной функции потерь, предложении рандомизированного метода обучения адаптивного классификатора по малому числу размеченных примеров на каждый класс, использующего эту функцию потерь, и формулировании условий сходимости получаемых с помощью этого метода оценок параметров.

Считаю, что диссертация Боярова Андрея Александровича на тему «Рандомизированный подход к обучению в условиях отсутствия разметки и малого количества данных» представляет несомненный научный интерес, является законченной научно-квалификационной работой. Она отвечает всем требованиям, предъявляемым к кандидатским диссертациям, а ее автор Бояров Андрей Александрович заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 01.01.09 – Дискретная математика и математическая кибернетика.



Поляк Борис Теодорович
7 мая 2020 г.

Доктор технических наук,
Главный научный сотрудник Института проблем управления РАН,
Москва, 117997, Профсоюзная 65, тел. 4953348829
e-mail boris@ipu.ru