

ОТЗЫВ

члена 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. В настоящее время большую актуальность приобретают задачи машинного обучения с использованием данных без разметки или с малым количеством разметки. Такие задачи порождают необходимость в разработке новых методов оптимизации, способных функционировать при существенных неопределённостях и демонстрировать высокое качество распознавания обученных с их помощью моделей. Актуальность этих задач обуславливает актуальность темы диссертационной работы Боярова Андрея Александровича «Рандомизированный подход к обучению в условиях отсутствия разметки и малого количества данных».

Научная новизна и теоретическая значимость. В диссертационной работе получен ряд новых и теоретически значимых результатов. Эти результаты связаны с развитием метода стохастической аппроксимации в задачах кластеризации и обучения по малому количеству данных. Основными теоретическими результатами диссертационной работы являются:

1. Рандомизированный алгоритм стохастической аппроксимации для кластеризации в модели данных, описываемой смесью гауссовских распределений.
2. Модификация указанного алгоритма для случая смеси гауссовских распределений с разреженными параметрами.
3. Алгоритм обучения адаптивного классификатора по малому количеству размеченных примеров, использующий рандомизацию на входе для оптимизации параметров в многозадачной функции потерь.

Полученные результаты расширяют область успешного применения подходов, основанных на стохастической аппроксимации, для оптимизации параметров в машинном обучении.

Практическая значимость. К практическим результатам диссертации относится разработанный в ходе её написания набор прикладных программ, реализующих предложенные в диссертационном исследовании алгоритмы, а так же их применение в задаче верификации авторства средневековых рукописей.

Замечания. К диссертационной работе есть несколько замечаний:

1. В тексте диссертации (стр. 29) указано, что индекс ARI вычисляется для известного, то есть фиксированного, результата кластеризации C . Однако в формуле для ARI используется математическое ожидание. Видимо, речь идёт о рандомизированных возможных результатах кластеризации C ?
2. В разделе 3.1 следует добавить в Таблицы 1 и 2 время выполнения алгоритмов кластеризации для иллюстрирования утверждения о более высоком быстродействии предложенного автором решения.
3. В разделе 3 не исследована чувствительность полученных показателей качества по отношению к настраиваемым параметрам алгоритмов. На сколько трудоёмким является процесс выбора параметров алгоритмов, и на сколько сильным может оказаться снижение качества кластеризации при варьировании этих параметров?
4. В тексте диссертации имеются отсылки к допущениям и определениям, содержащимся в работах из списка литературы, например, в разделе 2.2. Для полноты изложения было бы целесообразно включить эти элементы в текст диссертации.

Приведённые замечания не умаляют общего положительного впечатления от диссертационной работы Боярова Андрея Александровича. Диссертационная работа является законченным научным исследованием и обладает теоретической и практической ценностью, а основные результаты диссертационного исследования изложены в 9 работах, из которых 1 опубликована в издании из списка ВАК, 3 опубликованы в изданиях, индексируемых в базах данных Web of Science и Scopus.



CentraleSupélec

На основании вышеизложенного 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.

Член диссертационного совета

Доктор технических наук,
ЦетральСупелек
(CentraleSupélec)

Арановский Станислав Владимирович

28/08/2020

CAMPUS DE CHÂTENAY-MALABRY

Grande Voie des Vignes
F-92295 CHÂTENAY-MALABRY Cedex
Tél : +33 (0)1 41 13 10 00
Fax : +33 (0)1 41 13 10 10
SIRET : 130 020 761 00024

CAMPUS DE GIF-SUR-YVETTE (SIÈGE)

Plateau de Moulon
3 rue Joliot-Curie
F-91192 GIF-SUR-YVETTE Cedex
Tél : +33 (0)1 69 85 12 12
Fax : +33 (0)1 69 85 12 34
SIRET : 130 020 761 00016

CAMPUS DE METZ

Metz Technopôle
2 rue Edouard Belin
F-57070 METZ
Tél : +33 (0)3 87 76 47 47
Fax : +33 (0)3 87 76 47 00
SIRET : 130 020 761 00040

CAMPUS DE RENNES

Avenue de la Bouffaye
C.S. 47601
F-35576 CESSON-SÉVIGNÉ Cedex
Tél : +33 (0)2 99 84 45 00
Fax : +33 (0)2 99 84 45 99
SIRET : 130 020 761 00032