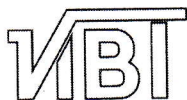


МИНОБРНАУКИ РОССИИ



ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ НАУЧНОЕ УЧРЕЖДЕНИЕ
«ФЕДЕРАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР
ИНФОРМАЦИОННЫХ И ВЫЧИСЛИТЕЛЬНЫХ ТЕХНОЛОГИЙ»
(ФИЦ ИВТ)

Проспект Академика Лаврентьева, д. 6, г. Новосибирск, 630090
Тел.: +7 (383) 330-6150, факс: +7 (383) 333-18-24, e-mail: ict@ict.nsc.ru
ОКПО 05222159, ОГРН 1025403650920, ИНН/КПП 5408105390/540801001

От 30.05.22 № 15312-

на № _____ от _____

ЖДАЮ

директора ФИЦ ИВТ

-мат. наук

Борисович Медведев

май 2022 г.

ОТЗЫВ

ведущей организации на диссертационную работу

Сивак Марии Алексеевны

«Робастное обучение нейронных сетей с простой архитектурой для решения задач классификации данных», представленную на соискание ученой степени кандидата технических наук по специальности 05.13.17 – «Теоретические основы информатики»

На отзыв представлены: диссертация на 111 страницах машинописного текста, включая 18 рисунков, 18 таблиц и список литературы из 105 источников; автореферат диссертации на 20 страницах, включая список из 12 публикаций автора по теме диссертации, в том числе 3 статьи в журналах, рекомендованных ВАК.

Актуальность темы выполненной работы. В настоящее время алгоритмы машинного обучения получают все более широкое развитие и применение на практике. Часто в них используется квадратичная функция потерь, что негативно сказывается на точности результатов при анализе сильно зашумленных данных. Не составляют исключение и искусственные нейронные сети, для обучения которых, как правило, используется алгоритм обратного распро-

странения ошибки. Ввиду того, что квадратичная функция потерь не обеспечивает устойчивость при наличии нетипичных наблюдений в данных, простые нейронные сети не всегда позволяют получить корректные результаты.

В диссертационной работе Сивак Марии Алексеевны предложена оригинальная модификация алгоритма обратного распространения ошибки, заключающаяся в замене квадратичной функции потерь на робастную, что обеспечивает большую устойчивость алгоритма обучения при наличии выбросов в данных. Предложенная модификация позволяет построить ИНС для решения задач классификации с высокой точностью даже для сильно зашумленных данных. Поэтому актуальность темы диссертационной работы Сивак Марии Алексеевны и ее практическая значимость не вызывают сомнений.

Структура и содержание диссертационной работы. Диссертация Сивак Марии Алексеевны состоит из введения, 5 глав основного содержания, заключения, списка литературы и 3 приложений.

Во *введении* описывается степень разработанности темы и обосновывается ее актуальность, формулируются цель, задачи исследования, научная новизна, теоретическая и практическая значимость полученных результатов, результаты, выносимые на защиту, дается краткое содержание работы по главам.

В *первой главе* представлен обзор наиболее распространенных методов и задач машинного обучения, рассмотрены классический и робастный подходы. Приводятся основные понятия теории искусственных нейронных сетей (ИНС), рассматривается алгоритм обучения, а также особенности практического использования ИНС. Кроме того, представлен обзор готовых программных решений для построения нейронных сетей. Приводится обоснование цели и задач исследования.

Во *второй главе* рассматривается подход к построению робастных нейронных сетей. Проводится исследование применимости различных робастных функций потерь при обучении нейронных сетей, предлагается робастная модификация алгоритма обратного распространения ошибки. Кроме того, показано влияние выбора плана эксперимента на точность работы нейронной сети.

В *третьей главе* представлены результаты исследований, проведенных с использованием технологии статистического моделирования. Результатом первого этапа исследований стал набор интервалов, на которых рекомендуется выбирать значения параметров робастных функций потерь. На втором этапе выполнялось исследование устойчивости робастных нейронных сетей при классификации данных с различной долей засоряющих наблюдений. На третьем этапе проводилось исследование устойчивости построенных нейронных сетей при анализе

наборов данных, содержащих в себе различное число объектов. Исследования показали, что использование робастных нейронных сетей позволяет получить существенный выигрыш в точности при анализе зашумленных данных.

В *четвертой главе* рассматривается разработанный программный модуль для построения робастных нейронных сетей «RobustNN». Приводятся минимальные системные требования, архитектура, особенности реализации модуля. Описываются основные методы, позволяющие пользователю задавать архитектуру нейронной сети и выбирать функцию потерь. Приводится пример структуры файлов, получаемых при сохранении модели.

В *пятой главе* с использованием разработанного инструментария решается задача классификации нефтяных месторождений и задача определения местоположения проводника в кровеносном сосуде.

В *заключении* диссертационной работы приведены выводы и сформированы основные результаты. Список литературы содержит 105 наименований.

В *приложении А* представлены фрагменты исходного кода, демонстрирующие использование разработанного программного модуля в различных режимах. В *приложении Б* представлены свидетельства о государственной регистрации программы для ЭВМ. В *приложении В* представлены акты внедрения результатов научных исследований.

Научная новизна работы. Новизна результатов диссертационной работы Сивак Марии Алексеевны заключается в следующем.

1. Предложена робастная модификация алгоритма обучения искусственных нейронных сетей в общем виде, позволяющая строить и обучать нейронные сети прямого распространения. В основе данной модификации лежат идеи алгоритма обратного распространения ошибки, однако вместо квадратичной функции потерь используются робастные. Такой подход позволяет не исключать выбросы из рассмотрения, а снизить их негативное влияние при обучении нейронной сети.

2. На основе результатов вычислительных экспериментов даны рекомендации по выбору значений внутренних параметров робастных функций потерь, что позволяет ускорить процесс настройки робастных ИНС.

3. Проведены исследования работоспособности робастных нейронных сетей с различными функциями потерь при анализе наборов данных с различной долей засоряющих наблюдений и с различным числом объектов. Показано, что использование робастных ИНС позволяет существенно увеличить

точность результатов. Кроме того, продемонстрировано, каким образом точность работы робастной нейронной сети зависит от выбора плана эксперимента.

Достоверность основных научных положений, выводов и рекомендаций. В работе использованы методы и положения математического анализа, прикладной математической статистики, регрессионного анализа, теории планирования эксперимента и методы оптимизации. Используемые в диссертации методы теоретических исследований научно обоснованы. Достоверность научных положений, рекомендаций и выводов обеспечивается корректным использованием методов исследования, а также подтверждением полученных выводов результатами вычислительных экспериментов, проведенных с использованием технологии статистического моделирования.

Теоретическая и практическая значимость диссертации.

Предложенная модификация алгоритма обратного распространения ошибки и проведенные исследования применимости робастных функций потерь в нейронных сетях позволяют строить нейронные сети для работы с зашумленными данными. При этом влияние нетипичных наблюдений снижается, т.е. не требуется их исключение из набора данных на этапе предобработки. Проведенные исследования показали, что использование робастного подхода позволяет улучшить точность результатов и увеличить скорость обучения нейронных сетей. Разработанный программный модуль позволяет построить робастные нейронные сети с произвольным количеством скрытых слоев и числом нейронов на них. Полученные в ходе исследования результаты могут быть рекомендованы к использованию в различных областях: генетика, биология, медицина, нефтегазовая отрасль, контроль качества продукции, экономика и др. Внедрение робастных нейронных сетей может быть полезным в таких организациях, как, например, Институт нефтегазовой геологии и геофизики им. А.А. Трофимука СО РАН, Федеральный исследовательский центр «Институт цитологии и генетики СО РАН», Институт химической биологии и фундаментальной медицины СО РАН.

Нейронные сети, построенные с помощью разработанного программного модуля, использовались при решении задачи классификации нефтяных месторождений в деятельности Института «ТатНИПИнефть» ОАО «Татнефть», что подтверждается соответствующим актом о внедрении.

Полнота опубликования научных результатов и апробация.

По теме диссертации опубликовано 12 работ, в том числе 3 статьи в журналах, включенных в перечень ВАК, в которых должны быть опубликованы

основные научные результаты диссертации на соискание ученой степени кандидата наук, 2 статьи в журналах, индексируемых в базах данных Scopus и Web of Science, 2 свидетельства о государственной регистрации программы для ЭВМ. Результаты работы докладывались на научных конференциях международного и всероссийского уровня. Все полученные новые результаты диссертационного исследования достаточно полно отражены в опубликованных работах.

Замечания по содержанию диссертации.

1. Неясно, почему при проведении большинства исследований используется набор данных «Ирисы Фишера», который является достаточно простым по своей структуре и содержит малое количество наблюдений.
2. Неясно, почему моделирование выбросов в наборе «Ирисы Фишера» производилось только для 3-х и 4-х признаков.
3. Не проведено сравнение полученных результатов классификации с более сложными нейронными сетями (например, сверточными).
4. Неясно, почему оценка качества работы построенных нейронных сетей с помощью ROC не выполняется при решении задачи классификации нефтяных месторождений.

Заключение.

Отмеченные выше недостатки не носят принципиального характера и не снижают общей положительной оценки работы.

Работа Сивак Марии Алексеевны выполнена на высоком научно-техническом уровне и написана грамотным научным языком. Содержание диссертации достаточно полно представляет научные исследования и практические разработки, проведенные автором. Содержание автореферата соответствует содержанию диссертации и отражает основные результаты исследований.

Достоверность результатов работы не вызывает сомнений, сами результаты достаточно полно опубликованы и известны специалистам в области обработки данных дистанционного зондирования. По своему содержанию работа соответствует паспорту специальности 05.13.17 – Теоретические основы информатики, а именно п. 5 – «Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечение, разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений».

Таким образом, диссертация Сивак Марии Алексеевны имеет внутреннее единство и является завершенной научно-квалификационной работой, в которой содержится решение важной научно-технической задачи: разработки методов робастного обучения искусственных нейронных сетей для решения различных задач классификации.

По своей актуальности, научной новизне, объему выполненных экспериментальных исследований и практической значимости, представленная работа соответствует требованиям, установленным в пп. 9-14 Положения «О присуждении учёных степеней», утвержденного Постановлением Правительства Российской Федерации от 24.09.2013 г. № 842, предъявляемым к диссертациям на соискание учёной степени кандидата технических наук по специальности 05.13.17 – Теоретические основы информатики, и является законченной научно-квалификационной работой, а её автор Сивак Мария Алексеевна достойна присуждения учёной степени кандидата технических наук по указанной специальности.

Отзыв на диссертацию заслушан, обсужден и одобрен на объединенном семинаре ФИЦ ИВТ и НГУ «Информационные технологии» (протокол № 18 от 17 мая 2022 г.).

Пестунов Игорь Алексеевич,
кандидат физико-математических наук, доцент,
ведущий научный сотрудник,
и.о. заведующего лабораторией аэрокосмического мониторинга и
обработки данных (совместно с АлтГУ),
Федеральное государственное бюджетное научное учреждение
«Федеральный исследовательский центр информационных и вычислительных технологий» (ФИЦ ИВТ)
Россия, 630090, г. Новосибирск, пр-т Академика Лаврентьева, 6.
Телефон: +7(383) 334-91-55
Эл. почта: pestunov@ict.sbras.ru

Подпись _____ И.А. Пестунов

25.05.2022 г.

*Сотворил отпечаток
07.05.2022*

25.05.2022

*Отзыв посылу
в свет 06.05.2022*