

ЗАКЛЮЧЕНИЕ ДИССЕРТАЦИОННОГО СОВЕТА Д 212.277.04 , СОЗДАННОГО
НА БАЗЕ ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО БЮДЖЕТНОГО
ОБРАЗОВАТЕЛЬНОГО УЧРЕЖДЕНИЯ ВЫСШЕГО ОБРАЗОВАНИЯ
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»
ПО ДИССЕРТАЦИИ
НА СОИСКАНИЕ УЧЕНОЙ СТЕПЕНИ КАНДИДАТА НАУК

аттестационное дело № _____
решение диссертационного совета от 29.09.2021 № 4

О присуждении Дударину Павлу Владимировичу (Российская Федерация)
ученой степени кандидата ТЕХНИЧЕСКИХ наук.

Диссертация «Исследование и разработка моделей и методов нечеткой кластеризации коротких текстов»
по специальности 05.13.01 - «Системный анализ, управление и обработка информации (информационные технологии и промышленность)»
принята к защите 23.06.2021 (протокол заседания № 2) диссертационным советом
Д 212.277.04 , созданным на базе ФГБОУ ВО «УЛЬЯНОВСКИЙ
ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ» (432027,
Г.УЛЬЯНОВСК, УЛ.СЕВЕРНЫЙ ВЕНЕЦ, 32) №678/НК ОТ 18.11.2020.

Соискатель Дударин Павел Владимирович "26" октября 1983 года рождения.
В 2020 году соискатель окончил аспирантуру по направлению 09.06.01 «Информатика и вычислительная техника» в ФГБОУ ВО «Ульяновский государственный технический университет».

Работает младшим научным сотрудником в Департаменте научных исследований и инноваций ФГБОУ ВО «Ульяновский государственный технический университет»
Диссертация выполнена на кафедре «Информационные системы» ФГБОУ ВО «Ульяновский государственный технический университет»

Научный руководитель – доктор технических наук (05.13.12 Системы автоматизации проектирования) Ярушкина Надежда Глебовна, ФГБОУ ВО «Ульяновский государственный технический университет», ректор, профессор кафедры Информационные системы, профессор.

Официальные оппоненты:

Борисов Вадим Владимирович, доктор технических наук, профессор, профессор кафедры «Вычислительной техники», филиал ФГБОУ ВО «Национальный иссле-

довательский университет «МЭИ» в г. Смоленск;

Сухов Сергей Владимирович, кандидат физико-математических наук, старший научный сотрудник Ульяновского филиала Федерального бюджетного государственного учреждения науки «Институт радиотехники и электроники им. В.А. Котельникова РАН»

дали положительные отзывы на диссертацию.

Ведущая организация «Институт автоматики и процессов управления Дальневосточного отделения РАН» в своем положительном отзыве, подписанном Грибовой Валерией Викторовной, доктором технических наук, заместителем директора по научной работе, научным руководителем лаборатории «Интеллектуальных систем», утвержденным директором Института, членом-корреспондентом РАН Ромашко Р.В., указала, что диссертация является законченной научно-квалификационной работой, выполненной на актуальную тему. Новые научные результаты, полученные в диссертации, направлены на решение научной задачи, имеющей большое значение для развития методов решения задачи системного анализа и обработки экспертной информации.

Соискатель имеет 20 опубликованных работ, в том числе по теме диссертации опубликовано 19 работ, из них в рецензируемых научных изданиях опубликовано 13 работ.

Общий объем опубликованного материала составляет 14.53 п. л., авторский вклад – 5.83 п. л. Недостоверные сведения об опубликованных соискателем работах в диссертационной работе отсутствуют.

Наиболее значимые работы по теме диссертации:

1. Подход к оценке трудоемкости задач в процессе разработки программного обеспечения на основе нейронных сетей. / П.В. Дударин, В.Г. Тронин, К.В. Святков, В.А. Белов, Р.А. Шакуров // Автоматизация процессов управления. 2019. № 3 (57). С. 65-72. (лично соискателем 0,2 п.л.)
2. An approach to fuzzy hierarchical clustering of short text fragments based on fuzzy graph clustering / P.V. Dudarin, N.G. Yarushkina // Advances in Intelligent Systems and Computing. 2018. Т. 679. С. 295-304. (лично соискателем 0,3 п.л.)

3. An approach to customization of pre-trained neural network language model to specific domain / P.V. Dudarin, V.G. Tronin, K.V. Svyatov // Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference «Dialogue 2019». 2019. С. 1-9. (лично соискателем 0,18 п.л.)
4. Подход к обработке обратной связи пользователя для повышения качества результатов работы алгоритма кластеризации / П.В. Дударин, В.Г. Тронин, Н.Г. Ярушкина // Известия Самарского Научного Центра Российской Академии Наук. 2020. № 5 (97). С. 94-105. (лично соискателем 0,5 п.л.)

На диссертацию и автореферат поступили отзывы: все отзывы положительные, из них – 0 без замечаний, в 9 отзывах имеются замечания.

В отзыве профессора кафедры информатики Санкт-Петербургского государственного университета, д.ф.-м.н., профессора Тулупьева А.Л. имеется замечание о том, что в тексте автореферата диссертации не отражена архитектура искусственной нейронной сети автоэнкодера, который был использован в методе интерактивной кластеризации коротких текстов для инициализации весов в блоке кластеризации основной нейронной сети.

В отзыве руководителя лаборатории теоретических и междисциплинарных проблем информатики Санкт-Петербургского федерального исследовательского центра РАН, к.т.н. Абрамова М.В. имеется замечание об отсутствии в тексте автореферата обоснования выбора корпуса «Тайга» и обученной на нем модели word2vec в качестве базовой для работы метода расширения словаря языковой модели.

В отзыве ведущего научного сотрудника Федерального исследовательского центра «Информатика и управление» РАН, к.ф.-м.н., доцента Аверкина А.Н., имеется замечание о том, что в тексте автореферата диссертационной работы не представлены результаты проведенного эксперимента, который бы показывал качественное сравнение предложенного метода кластеризации с качеством кластеризации, достигнутым с помощью альтернативного метода, примененного к англоязычному набору текстов.

В отзыве заведующего кафедрой «Системы автоматизированного проектирования» Южного федерального университета, д.т.н., профессора Курейчика В.В.,

имеется замечание об ограниченности проведенных в рамках диссертационной работы экспериментов только русским языком.

В отзыве профессора Института физико-математических наук и информационных технологий Балтийского федерального университета имени И.Канта, д.т.н., профессора Колесникова А.В. имеется замечание о том, что в описании результатов проведенных экспериментов не указаны некоторые характеристики процесса обучения искусственной нейронной сети, в частности, количество эпох обучения, шаг обучения и значения целевой функции на обучающей и валидационной выборках.

В отзыве директора Института системной и программной инженерии и информационных технологий ФГАОУ ВО «Национальный исследовательский университет «Московский институт электронной техники», д.т.н., профессора Гагариной Л.Г. имеются замечания о том, что отсутствует описание обозначения μ_i в формуле расстояния между элементом и центром кластера и не представлено описание компетенций эксперта, необходимых для выполнения интерактивной кластеризации коротких текстов в представленном эксперименте по определению эффективности разработанного метода.

В отзыве доцента кафедры информационных технологий ФГБОУ ВО «Кубанский Государственный Университет», к.т.н. Полупанова А.А. имеется замечание о том, что в тексте автореферата диссертации не указаны размеры входного и выходного слоев в предлагаемой нейронной сети для кластеризации в серии проведенных экспериментов, а также о том, что из текста автореферата диссертации не ясно, каким образом происходит исправление орфографических ошибок на этапе предобработки текста для разработанного алгоритма нечеткой интерактивной кластеризации коротких текстов.

В отзыве доцента кафедры информационных и управляющих систем ФГБОУ ВО «Амурский Государственный Университет», к.т.н. Никифоровой Л.В. имеется замечание о том, что на рисунке 1 (стр. 11), на котором показана разработанная архитектура нейронной сети, не указана используемая функция активации.

В отзыве доцента кафедры радиоэлектроники и информационных технологий ФГБОУ ВО «Государственный морской университет имени адмирала Ф.Ф. Уша-

кова», к.т.н. Лютиковой М.Н. имеется замечание о недостаточности информации в тексте автореферата диссертации о размерах входного и выходного слоя используемой для кластеризации нейронной сети.

Выбор официальных оппонентов и ведущей организации обосновывается их высокой компетенцией, научными достижениями и наличием публикаций в соответствующей отрасли науки.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований:

разработаны методы и алгоритм, позволяющие повысить эффективность интеллектуальной машинной обработки коротких текстов в автоматизированных системах поддержки принятия решений, в частности, разработанный метод интеллектуальной обработки текстов с расширением словаря языковой модели на базе нейронной сети и использованием нечеткого иерархического классификатора позволяет повысить точность кластеризации в среднем на 10%.

предложен оригинальный научный подход к нечеткой интерактивной кластеризации с использованием аппарата искусственных нейронных сетей, языкового моделирования и нечетких графов.

Доказана перспективность использования предлагаемых технических решений для автоматизации интерактивной кластеризации наборов данных, состоящих из коротких текстов.

Введены и обоснованы новая архитектура нейронной сети с объединенными блоками языковой модели и кластеризации, метод расширения словаря языковой модели и алгоритм нечеткой интерактивной кластеризации с учетом обратной связи от специалиста-эксперта.

Теоретическая значимость исследования обоснована тем, что:

Доказаны положения, вносящие вклад в развитие изучаемых процессов нечеткой интерактивной кластеризации коротких текстов, основанных на применении искусственных нейронных сетей и машинного обучения, с привлечением специалистов-экспертов в данной предметной области.

Применительно к проблематике диссертации результативно (эффективно, то есть с получением обладающих новизной результатов)

использован комплекс существующих методов машинного обучения с участием учителя, искусственных нейронных сетей глубинного обучения, кластерного анализа, нечеткой логики и авторских алгоритма, методов и модели для усовершенствования процедуры нечеткой кластеризации коротких текстов.

Изложены основные научные положения, гипотезы и рекомендации, позволяющие повысить эффективность автоматизированной кластеризации наборов данных, состоящих из коротких текстов.

Раскрыты принципиальные ограничения существующих методов кластеризации коротких текстов без использования аппарата языкового моделирования в рамках единой искусственной нейронной сети и без вовлечения в процесс эксперта, выполняющего функцию частичного обучения.

Изучены процессы дополнительного обучения языковых моделей для предметных областей исследуемых наборов текстов, алгоритмы нечеткой и интерактивной кластеризации, архитектуры нейронных сетей для обработки текстов и методы вовлечения специалистов-экспертов в процесс кластеризации коротких текстов.

Проведена модернизация существующих алгоритма ϵ -кластеризации нечеткого графа и алгоритма кластеризации DEC (deep embedding for clustering) с целью оптимизации их использования в системах поддержки принятия решений для обеспечения качественной и эффективной кластеризации коротких текстов.

Значение полученных соискателем результатов исследования для практики подтверждается тем, что:

разработаны и внедрены в промышленную эксплуатацию новый высокоэффективный программный модуль, обеспечивающий повышение эффективности обработки больших массивов информации в виде наборов коротких текстов; предлагаемый подход сократил временные затраты экспертов на этапе кластеризации и построения классификатора в 3 раза по сравнению с полностью ручной кластеризацией и в 2 раза – в случае использования методов тематической кластеризации.

В абсолютных значениях экономия составила более 180 человеко-дней высококвалифицированного специалиста.

определены границы применимости предлагаемого алгоритма нечеткой интерактивной кластеризации и перспективы его использования;

создана и интегрирована в Федеральную информационную систему стратегического планирования Российской Федерации искусственная нейронная сеть для решения задачи кластеризации и классификации отдельных фрагментов входящих документов стратегического планирования;

представлены предложения по дальнейшему совершенствованию интеллектуальной машинной обработки коротких текстов на основе нечеткой интерактивной кластеризации для использования с различными языковыми моделями и в случаях совместной работы группы экспертов-специалистов по предметной области.

Оценка достоверности результатов исследования выявила:

для экспериментальных работ результаты получены на основе корректного использования математических методов численного анализа и современного высокопроизводительного программного обеспечения, подтверждены вычислительными экспериментами и результатами практического применения в лабораторных и производственных условиях ФГБОУ ВО УлГТУ и ООО “ИБС Ульяновск”;

теория построена на известных проверенных научных данных, которые в полной мере согласуются с ранее опубликованными данными по теме диссертационного исследования;

идея базируется на анализе и обобщении передового опыта и практических исследований ряда российских и зарубежных ученых по теме диссертации;

использованы сравнения авторских данных и результатов, полученных ранее по рассматриваемой тематике для схожего набора данных на английском языке:

установлено качественное совпадение результатов, полученных автором, с опубликованными ранее результатами аналогичных исследований в области интерактивной кластеризации текстов на английском языке;

использованы современные методики сбора и обработки текстовой информации, обеспечивающие воспроизводимость и достоверность результатов.

Личный вклад соискателя состоит в анализе научных и патентных источников по теме выполненного исследования, обработке и интерпретации аналитической информации, разработке модели, методов и алгоритма нечеткой интерактивной кластеризации, планировании экспериментов, формулировке выводов, апробации и внедрении полученных результатов. Все основные исследования проведены лично автором, либо при его непосредственном участии.

На заседании 29.09.2021 диссертационный совет принял решение

за решение научной задачи разработки модели, методов и алгоритма нечеткой кластеризации коротких текстов на русском языке с учетом экспертной информации, имеющей большое значение для развития методов решения задач системного анализа, управления и обработки информации, в частности, для повышения эффективности обработки значительных массивов информации в виде наборов коротких текстов, в том числе документов стратегического планирования развития страны, присудить Дударину П.В. ученую степень кандидата технических наук.

При проведении тайного голосования диссертационный совет в количестве 17 человек, из них 4 докторов наук по специальности рассматриваемой диссертации, участвовавших в заседании, из 23 человек, входящих в состав совета, дополнительно введены на разовую защиту 0 человек, проголосовали: за 16, против **нет**, недействительных бюллетеней 1.

Заместитель председателя
диссертационного совета
Ученый секретарь
диссертационного совета



29.09.2021

Киселев Сергей Константинович

Наместников Алексей Михайлович