

УТВЕРЖДАЮ

Проректор по научной работе УлГТУ,

т.н. Наместников А.М.



*(Handwritten signature)*

« 23 » *марта* 2021 г.

М.П.

## ЗАКЛЮЧЕНИЕ

федерального государственного бюджетного образовательного учреждения высшего образования «Ульяновский государственный технический университет»

Диссертация Дударина П.В. «Исследование и разработка моделей и методов нечеткой кластеризации коротких текстов» выполнена на кафедре «Информационные системы» УлГТУ. В период подготовки диссертации соискатель Дударин Павел Владимирович обучался в очной аспирантуре и работал по совместительству в Федеральном государственном бюджетном образовательном учреждении высшего образования «Ульяновский государственный технический университет», в Научно исследовательском отделе «Департамента научных исследований и инноваций» младшим научным сотрудником.

В 2005 г. окончил Ульяновский государственный университет по специальности «Прикладная математика и информатика». В 2015 году поступил в очную аспирантуру УлГТУ по специальности 05.13.18 «Математическое моделирование, численные методы и комплексы программ» и закончил ее в 2020 г.

Диплом об окончании аспирантуры выдан в 2020 г. Федеральным государственным бюджетным образовательным учреждением высшего образования «Ульяновский государственный технический университет».

Научный руководитель – Ярушкина Надежда Глебовна, д.т.н., профессор, зав. каф. «Информационные системы», ректор Федерального государственного бюджетного образовательного учреждения высшего образования «Ульяновский государственный технический университет».

По итогам обсуждения принято следующее заключение:

Автор Дударин П.В. выполнил исследование моделей и методов нечеткой кластеризации коротких текстов. Автором точно сформулированы, обозначены и решены задачи исследования, направленные на разработку моделей, методов и алгоритма в системе поддержки принятия решений «Федеральная информационная система стратегического планирования», которая автоматизирует применение предложенных моделей и методов для выполнения нечеткой интерактивной кластеризации наборов коротких текстов.

Автор Дударин П.В. определил следующую цель исследования.

Целью диссертационного исследования является повышение эффективности нечеткой кластеризации коротких текстов путем разработки модели, метода и алгоритма в системе поддержки принятия решений для кластеризации коротких текстов на русском языке с учетом экспертной информации. Эффективность определяется точностью кластеризации и сокращением времени и трудоемкости работы, выполняемой экспертом при использовании предложенного решения.

Цель декомпозирована на ряд задач исследования, которые сформулированы следующим образом:

- исследование моделей и методов машинного обучения для обработки текстов;
- исследование моделей и методов четкой и нечеткой кластеризации коротких текстов;
- разработка метода расширения словаря языковой модели на базе нейронной сети;
- разработка метода для обработки экспертной информации в ходе нечеткой интерактивной кластеризации коротких текстов;

- разработка программы для проведения испытаний по нечеткой интерактивной кластеризации коротких текстов;
- разработка алгоритма в системе поддержки принятия решений для автоматизации работ по нечеткой интерактивной кластеризации коротких текстов;
- оценка эффективности разработанных модели, методов и алгоритма на задаче кластеризации коротких текстов в рамках функционирующей системы.

Объект исследования – кластеризация наборов данных, состоящих из коротких текстов на русском языке и экспертная информация, поступающая в ходе интерактивной обработки текстов.

Предмет исследования – модели и методы нечеткой кластеризации коротких текстов и обработки экспертной информации.

Научную новизну автор формулирует следующим образом:

1. Предложенная архитектура искусственной нейронной сети, отличающаяся от известных тем, что позволяет решать задачу кластеризации на базе скрытого пространства признаков языковой модели;
2. Предложенный метод обработки текстов для расширения словаря языковой модели на базе нейронной сети с использованием нечеткого иерархического классификатора, отличающийся от известных тем, что позволяет учитывать семантическую близость слов;
3. Предложенный метод обработки обратной связи от эксперта, отличающийся от известных тем, что позволяет корректировать весовые коэффициенты нейронной сети и проводить интерактивную кластеризацию наборов коротких текстов;
4. Разработанный алгоритм в системе поддержки принятия решений, автоматизирующий применение предложенных модели и методов для выполнения нечеткой интерактивной кластеризации наборов коротких текстов.

Практическая ценность работы заключается в разработке программного модуля системы поддержки принятия решений на языке Python, позволяющего

осуществлять интерактивную нечеткую кластеризацию наборов данных, состоящих из коротких текстов, и применение его в задаче системного анализа набора коротких текстов в рамках НИР в интересах Министерства экономического развития РФ для Системы стратегического планирования РФ.

Автор выносит на защиту следующие положения:

1. Предложенная архитектура искусственной нейронной сети, позволяет эффективно решать задачу кластеризации на базе пространства признаков языковой модели русского языка;
2. Предложенный метод обработки текстов для расширения словаря языковой модели на базе нейронной сети с использованием нечеткого иерархического классификатора повышает точность кластеризации;
3. Предложенный метод учета обратной связи от эксперта, используемый для корректировки весовых коэффициентов нейронной сети позволяет проводить интерактивную кластеризацию наборов коротких текстов;
4. Разработанный алгоритм на основе предложенных моделей и методов реализован в системе поддержки принятия решений и автоматизирует применение предложенных моделей и методов для выполнения нечеткой интерактивной кластеризации наборов коротких текстов.

Реализация результатов исследования, а именно архитектуры искусственной нейронной сети, алгоритма нечеткой кластеризации коротких текстов, методов расширения словаря языковой модели и корректировки весов нейронной сети для учета обратной связи эксперта в интерактивной кластеризации, производилась в системе Планета.Аналитика 4.0 (входит в реестр отечественного ПО) компании ООО “ИБС “Экспертиза”.

Практическая значимость работы подтверждается использованием ее результатов и рекомендаций в рамках НИР в интересах Министерства экономического развития РФ по темам: “Разработка рекомендаций по совершенствованию информационного обеспечения участников стратегического планирования в части осуществления мониторинга и контроля реализации документов стратегического планирования с использованием Федеральной

информационной системы стратегического планирования (ФИС СП)” и “Разработка методического обеспечения интеллектуальной системы проверки уведомления об утверждении (одобрении) документа стратегического планирования или внесении в него изменений при ведении федерального государственного реестра документов стратегического планирования Федеральной информационной системы стратегического планирования (ФИС СП)”. Результаты НИР внедрены в системе ГАС “Управление”, что подтверждено соответствующими документами, представленными в приложении к диссертации.

Апробация результатов исследования.

Основные положения и результаты диссертации опубликованы в журналах из Перечня, рекомендованного ВАК РФ: «Автоматизация процессов управления», «Радиотехника», «Нечеткие системы и мягкие вычисления», «Известия Самарского научного центра Российской Академии Наук»; докладывались, обсуждались и получили одобрение на: всероссийской научно-практической конференции “Нечеткие системы и мягкие вычисления” (Санкт-Петербург 2017); международной конференции “Интеллектуальные информационные технологии в технике и на производстве” ИТИ (Варна 2017, Сочи 2018, Острава 2019); всероссийской научной конференции «Нечеткая логика и мягкие вычисления в промышленности» (Ульяновск: 2017, 2018, 2019); национальной конференции по искусственному интеллекту (Москва 2018, Ульяновск 2019); международной конференции “World Conference on Soft Computing” (Баку 2018); международной конференции “Mexican International Conference on Artificial Intelligence” (Гвадалахара 2018); международной конференции “European Society for Fuzzy Logic and Technology” (Прага 2019); международной конференции по компьютерной лингвистике и интеллектуальным технологиям “Диалог” (Москва, 2019); I-ом национальном конгрессе по когнитивным исследованиям, искусственному интеллекту и нейроинформатике (Москва, 2020).

По результатам работы было опубликовано 19 статей, в том числе, 6 – в журналах из Перечня, рекомендованного ВАК РФ, а также 7 статей в изданиях,

индексируемых в Scopus и Web of Science. Подана заявка на получение свидетельства о государственной регистрации программ для ЭВМ.

Структура и объем диссертации. Диссертационная работа состоит из 136 страниц машинописного текста, содержащих введение, четыре главы и заключение, 9 таблиц, 46 рисунков, список литературы из 134 наименований, 3 приложений.

Личный вклад соискателя.


Все результаты, составляющие содержание диссертации, получены диссертантом самостоятельно. Подготовка к публикации некоторых результатов проводилась совместно с соавторами, причем вклад диссертанта был определяющим.

Исходя из вышеизложенного, следует считать, что работа Дударина Павла Владимировича выполнена на важную научно-техническую тему, является актуальной, представляет собой законченное научное исследование, решает поставленные задачи в полном объеме.

Диссертация «Исследование и разработка моделей и методов нечеткой кластеризации коротких текстов» Дударина Павла Владимировича рекомендуется к защите на соискание ученой степени кандидата технических наук по специальности 05.13.01 – «Системный анализ, управление и обработка информации (информационные технологии и промышленность)».

Заключение принято на заседании Научно-технического совета факультета информационных систем и технологий (ФИСТ).

Присутствовали на заседании 13 чел. Результаты голосования: «за» – 13 чел., «против» – 0 чел., «воздержались» – 0 чел., протокол № 1 от «24» февраля 2021 г.



Председатель Научно-технического совета факультета информационных систем и технологий УлГТУ Киселев С.К., д.т.н., доцент, заведующий кафедрой «Измерительно-вычислительные комплексы»