

## ОТЗЫВ

на автореферат диссертации Дударина Павла Владимировича  
*Исследование и разработка моделей и методов  
нечеткой кластеризации коротких текстов,*  
представленной на соискание ученой степени кандидата технических наук  
по специальности 05.13.01 — «Системный анализ, управление и обработка информации  
(информационные технологии и промышленность)».

1. Задача кластеризации коротких текстов с последующим построением классификатора является трудоемким процессом, требующим вовлечения профессиональных знаний и опыта. Научные исследования в этой области направлены на создание языковых моделей, тезаурусов и онтологий позволяющих автоматизировать извлечение семантической информации. При проведении кластеризации эксперту необходимо понимание контекста и вариантов использования результатов кластеризации, зачастую это понимание приходит к эксперту непосредственно в процессе кластеризации и/или обработки результатов кластеризации. В связи с вышесказанным, работа по созданию методов интерактивной кластеризации представляет собой актуальную **научно-техническую задачу**.

Целью диссертации Дударина П.В. 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.

2. Новизна полученных результатов состоит в следующем.

2.1. Предложена архитектура искусственной нейронной сети, позволяющая решать задачу кластеризации на базе скрытого пространства признаков языковой модели.

2.2. Предложен метод обработки текстов для расширения словаря языковой модели на базе нейронной сети с использованием нечеткого иерархического классификатора, позволяющий учитывать семантическую близость слов.

2.3. Предложен метод обработки обратной связи от эксперта, позволяющий корректировать весовые коэффициенты нейронной сети и проводить интерактивную кластеризацию наборов коротких текстов.

2.4. Разработан алгоритм, автоматизирующий применение предложенных модели и методов для выполнения нечеткой интерактивной кластеризации наборов коротких текстов, интегрированный в систему поддержки принятия решений.

3. Значение результатов диссертации для теории заключается в разработке новых моделей и методов с использованием нейронных сетей и языковых моделей для решения задачи нечеткой кластеризации наборов данных состоящих из коротких текстов.

Значение результатов для практики заключается в повышении эффективности процесса кластеризации коротких текстов, сокращение временных затрат эксперта за счет использования интерактивного подхода в процессе кластеризации и возможности учета обратной связи от эксперта в данном процессе.

4. К недостатку автореферата диссертационной работы следует отнести следующее: *в тексте автореферата диссертации не отражена архитектура автоэнкодера использующегося для инициализации весов в блоке кластеризации основной нейронной сети.*

5. Отмеченный недостаток не снижает научной и практической ценности работы, которая удовлетворяет требованиям, предъявляемым ВАК к кандидатским диссертациям, а ее автор Дударин П.В. заслуживает присвоения ученой степени кандидата технических наук по специальности 05.13.01 — Системный анализ, управление и обработка информации (информационные технологии и промышленность).

Профессор кафедры информатики Санкт-Петербургского государственного университета, профессор, доктор физико-математических наук (специальность 05.13.17 — Теоретические основы информатики),

Александр Львович Тулупьев  
01 сентября 2021 г.

*Документ подготовлен в индивидуальной форме*  
Личную подпись руки А.Л. Тулупьева, профессора кафедры информатики СПбГУ, профессора, д.ф.-м.н., удостоверяю.

и.о. Начальник управления кадров СПбГУ



*С.В. Морозова*  
С.В. Морозова

01 сентября 2021 г.

Федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет», 199034, г. Санкт-Петербург, Университетская набережная, д. 7/9; <http://www.spbu.ru>, [spbu@spbu.ru](mailto:spbu@spbu.ru), +7 812 328-97-88.

МИНОБРНАУКИ РОССИИ

**Федеральное государственное бюджетное учреждение науки  
«Санкт-Петербургский Федеральный исследовательский центр  
Российской академии наук» (СПб ФИЦ РАН)**

14-я линия В.О., д. 39, г. Санкт-Петербург, 199178

Тел.: (812) 328-33-11, факс: (812) 328-44-50,

e-mail: info@spcras.ru, web: http://www.spcras.ru

ОКПО 04683303, ОГРН 1027800514411, ИНН/КПП 7801003920/780101001

02.09.2021 № 60-01-01-757

На № \_\_\_\_\_

### **ОТЗЫВ**

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

Сети глубокого обучения показывают высокие результаты в различных задачах машинного обучения, включая задачи кластеризации. В последние годы интерес российских и зарубежных исследователей в области обработки естественного языка привлекают языковые модели. Языковое моделирование позволяет обучить искусственную нейронную сеть семантике естественного языка, и далее использовать полученную модель для решения различных задач обработки текстов. В связи с этим диссертация соискателя Дударина П.В., целью которой является исследование и разработка моделей и методов нечеткой кластеризации коротких текстов с использованием языковых моделей на базе нейронных сетей, является, бесспорно, актуальной.

Полученные автором диссертационной работы результаты, основанные на использовании искусственных нейронных сетей и языкового моделирования, обладают научной новизной. Достоверность выдвигаемых положений подтверждена проведенными вычислительными экспериментами.

Перечисленные ниже научные результаты являются новыми:

1. Предложенная архитектура искусственной нейронной сети, позволяющая 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, позволяющий учитывать семантическую близость слов.
3. Предложенный метод обработки обратной связи от эксперта, позволяющий корректировать весовые коэффициенты нейронной сети и проводить интерактивную кластеризацию наборов коротких текстов.
4. Разработанный алгоритм, автоматизирующий применение предложенных

модели и методов для выполнения нечеткой интерактивной кластеризации наборов коротких текстов, интегрированный в систему поддержки принятия решений.

Практическая значимость представленной в автореферате диссертационной работы состоит в разработанном в рамках системы поддержки принятия решений программном модуле, который позволяет осуществлять интерактивную нечеткую кластеризацию наборов коротких текстов. Данный модуль показал высокую эффективность в решении задачи анализа набора коротких текстов в рамках НИР в интересах Министерства экономического развития РФ для Системы стратегического планирования РФ.

В качестве недостатка автореферата диссертационной работы следует отметить отсутствие обоснования выбора корпуса «Тайга» и обученной на нем модели word2vec в качестве базовой для работы метода расширения словаря языковой модели.

Отмеченный недостаток не снижает ценность диссертационной работы Дударина Павла Владимировича. Представленные результаты исследования позволяют сделать вывод о том, что диссертационная работа представляет бесспорный научный интерес и практическую ценность, полностью удовлетворяет требованиям ВАК РФ, предъявляемым к кандидатским диссертациям, а соискатель Дударин П.В. заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.01. – «Системный анализ, управление и обработка информации (информационные технологии и промышленность)».

Руководитель лаборатории теоретических и междисциплинарных проблем информатики Санкт-Петербургского Федерального исследовательского центра Российской академии наук (СПб ФИЦ РАН)  
Кандидат технических наук (специальность 05.13.19 — Методы и системы защиты информации, информационная безопасность)

Абрамов Максим Викторович  
+7 (812) 328-33-37  
mva@dscs.pro  
02 сентября 2021 г.

*Личную подпись руки М.В. Абрамова, руководителя лаборатории теоретических и междисциплинарных проблем информатики СПб ФИЦ РАН, к.т.н., удостоверяю*

Начальник отдела кадров СПб ФИЦ РАН

Д.В. Токарев  
02 сентября 2021 г.



Сведения об организации.

Название организации: Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук»

Адрес организации: 199178, г. Санкт-Петербург, 14-линия В.О., д. 39

E-mail: [info@spcras.ru](mailto:info@spcras.ru)

Телефон: +7 (812) 328 33 11

## ОТЗЫВ

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

В настоящее время трудно представить процесс обработки текстов без использования нейронных сетей и языковых моделей. Современные архитектуры нейронных сетей позволяют эффективно обрабатывать тексты с учетом последовательности слов в них, а языковые модели позволяют автоматизировать извлечение смысла из текста. При этом машинная обработка коротких текстов осложняется отсутствием необходимого контекста для извлечения семантики, а проблема кластеризации коротких текстов сложна в первую очередь большим количеством возможных результатов разбиения на кластеры.

В связи с вышесказанным, работа по созданию методов нечеткой интерактивной кластеризации коротких текстов с применением современных подходов в области построения и обучения нейронных сетей представляет собой актуальную научно-техническую задачу.

Целью диссертации Дударина П.В. 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.

Новыми являются перечисленные ниже результаты исследования.

1. Предложена архитектура искусственной нейронной сети, позволяющая решать задачу кластеризации на базе скрытого пространства признаков языковой модели.

2. Предложен метод обработки текстов для расширения словаря языковой модели на базе нейронной сети с использованием нечеткого иерархического классификатора, позволяющий учитывать семантическую близость слов.

3. Предложен метод обработки обратной связи от эксперта, позволяющий корректировать весовые коэффициенты нейронной сети и проводить интерактивную кластеризацию наборов коротких текстов.

4. Разработан алгоритм, автоматизирующий применение предложенных модели и методов для выполнения нечеткой интерактивной кластеризации наборов коротких текстов, интегрированный в систему поддержки принятия решений.

Значение результатов для практики заключается в экономии трудозатрат эксперта при решении задач кластеризации коротких текстов, а также повышении качества обработки документов стратегического

планирования за счет использования полученного автоматического классификатора.

К недостаткам автореферата диссертационной работы следует отнести следующее: в тексте автореферата диссертации не представлены результаты проведенного эксперимента по сравнению с качеством кластеризации достигаемом с помощью альтернативного метода примененного к англоязычному набору текстов.

Отмеченные недостатки не снижают научной и практической ценности работы, которая удовлетворяет требованиям, предъявляемым ВАК РФ к кандидатским диссертациям, а ее автор Дударин П.В. заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.01 «Системный анализ, управление и обработка информации (информационные технологии и промышленность)».

Ведущий научный сотрудник  
ФИЦ «Информатика и управление» РАН  
к.ф.-м.н., доцент,  
специальность 01.01.09 «Математическая кибернетика»

/Аверкин Алексей Николаевич/



Адрес ФИЦ ИУ РАН 119333, Москва, Вавилова, д.44, кор.2  
E-mail [frccsc@frccsc.ru](mailto:frccsc@frccsc.ru)  
Телефон+7 (499) 135-62-60

## ОТЗЫВ

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

В настоящее время имеет место тенденция применения глубоких нейронных сетей в задачах кластеризации. Повышенный интерес как российских, так и зарубежных исследователей, которые занимаются автоматизацией обработки текстов, связан с языковым моделированием. В связи с этим диссертационная работа соискателя Дударина П.В., назначением которой является исследование и разработка моделей и методов нечеткой кластеризации коротких текстов с использованием языковых моделей на базе нейронных сетей, является, несомненно, актуальной.

Полученные автором диссертации результаты исследования, основанные на применении методов машинного обучения и языкового моделирования, обладают научной новизной. Проведенные вычислительные эксперименты подтверждают достоверность выдвигаемых положений. Несомненной научной новизной обладает разработанный автором метод нечеткой интерактивной кластеризации текстов позволяющий учитывать мнение эксперта в ходе проведения процедуры кластеризации.

Перечисленные ниже научные результаты также являются новыми.

1. Предложенная архитектура искусственной нейронной сети, позволяющая 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, позволяющий учитывать семантическую близость слов.
3. Предложенный метод обработки обратной связи от эксперта, позволяющий корректировать весовые коэффициенты нейронной сети и проводить интерактивную кластеризацию наборов коротких текстов.
4. Разработанный алгоритм, автоматизирующий применение предложенных модели и методов для выполнения нечеткой интерактивной кластеризации наборов коротких текстов, интегрированный в систему поддержки принятия решений.

Практическая ценность рассматриваемой диссертационной работы состоит в разработанном программном модуле, позволяющем осуществлять интерактивную нечеткую кластеризацию коротких текстов. Данный модуль был использован в решении задачи анализа набора коротких текстов в рамках НИР в интересах Министерства экономического развития РФ для Системы



стратегического планирования РФ.

В качестве недостатка автореферата диссертационной работы следует отметить ограниченность проведенных экспериментов только русским языком.

Данный недостаток не снижает ценность диссертационной работы Дударина Павла Владимировича. Содержание результатов исследования позволяет сделать вывод о том, что диссертационная работа представляет несомненный научный интерес и практическую ценность, полностью удовлетворяет требованиям ВАК, предъявляемым к кандидатским диссертациям, а соискатель Дударин П.В. заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.01. – «Системный анализ, управление и обработка информации (информационные технологии и промышленность)».

Заведующий кафедрой  
«Систем автоматизированного проектирования»  
Южного федерального университета  
доктор технических наук, профессор  
Курейчик Владимир Викторович  
347928 г. Таганрог, пер. Некрасовский, д. 44  
Тел. (8634)383451, e-mail: vkur@sfedu.ru  
Персональная страница: <https://sfedu.ru/person/vkur>  
Специальности: 05.13.01 – Системный анализ,  
управление и обработка информации и  
05.13.18 – Математическое моделирование,  
численные методы и комплекты программ

Я, Курейчик Владимир Владимирович, даю согласие на включение моих персональных данных в документы, связанные с работой диссертационного совета и их дальнейшую обработку.

«09» сентября 2021 г.

Подпись Курейчика В.В. заверяю  
Директор ИКТИБ ЮФУ, д.т.н., проф.



В.В. Курейчик

Г. Е. Веселов

## ОТЗЫВ

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

1) Кластеризация коротких текстов с последующим построением классификатора – ресурсоемкая задача, требующая вовлечения эксперта в процесс кластеризации. Современные научные исследования в этой области направлены на создание инструментов, автоматизирующих извлечение семантической информации из наборов текстов. В частности, для этой цели служат: языковые модели, тезаурусы и онтологии. Приступая к кластеризации, эксперт исследует набор данных и формирует представления об этом наборе непосредственно в процессе исследования. При таком подходе, отдельные суждения эксперта могут быть противоречивыми, неточными, и меняться по мере углубления в исследуемый набор текстов. В связи с вышесказанным, работа по созданию методов интерактивной кластеризации, позволяющих учитывать мнение эксперта и/или экспертов, представляет собой актуальную *научно-техническую задачу*.

Цель диссертационной работы Дударина П.В. – повышение эффективности нечеткой кластеризации коротких текстов разработкой модели, метода и алгоритма в системе поддержки принятия решений для кластеризации коротких текстов на русском языке с учетом экспертной информации.

2) Научная новизна полученных результатов заключается в следующем:

а) Предложена архитектура искусственной нейронной сети для решения задачи кластеризации на базе скрытого пространства признаков языковой модели.

б) Предложен метод обработки текстов для расширения словаря языковой модели на базе нейронной сети с использованием нечеткого иерархического классификатора для учета семантической близости слов.

в) Предложен метод обработки обратной связи от эксперта, позволяющий корректировать весовые коэффициенты нейронной сети, и проводить интерактивную кластеризацию наборов коротких текстов.

г) Разработан алгоритм, автоматизирующий применение предложенных модели и методов для выполнения нечеткой интерактивной кластеризации наборов коротких текстов, интегрированный в систему поддержки принятия решений.

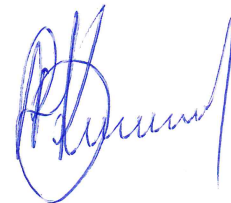
3) Практическое значение результатов диссертационной работы заключается в повышении эффективности процесса кластеризации коротких текстов.

Реализованный алгоритм сокращает временные затраты эксперта за счет использования интерактивного подхода в процессе кластеризации и дает возможность учета обратной связи от эксперта в данном процессе.

4) К недостатку автореферата диссертационной работы следует отнести следующее: *в описании результатов проведенных экспериментов не указаны характеристики процесса обучения: количество эпох обучения, шаг обучения и значения целевой функции на обучающей и валидационной выборках.*

5) Отмеченный недостаток не снижает научной и практической ценности работы, которая удовлетворяет требованиям, предъявляемым ВАК к кандидатским диссертациям, а ее автор Дударин П.В. заслуживает присвоения ученой степени кандидата технических наук по специальности 05.13.01. – Системный анализ, управление и обработка информации (информационные технологии и промышленность).

Профессор Института физико-математических наук и информационных технологий Балтийского федерального университета имени И. Канта, д.т.н., профессор,  
05.13.01 – «Системный анализ, управление и обработка информации (информатика)»  
Колесников Александр Васильевич  
2 сентября 2021 года



236016, г. Калининград, ул. А. Невского, 14.  
post@kantiana.ru  
+7 (4012) 59-55-97

Балтийский федеральный университет им. И. Канта  
Кадровая служба  
Подпись  
*Александр В. Колесников*  
"2" сентября 2021 г.



## ОТЗЫВ

на автореферат диссертации Дударина Павла Владимировича «Исследование и разработка моделей и методов нечеткой кластеризации коротких текстов», представленной на соискание ученой степени кандидата технических наук по специальности 05.13.01 «Системный анализ, управление и обработка информации (информационные технологии и промышленность)»

Стремительный рост массивов информации, состоящих из наборов коротких текстовых фрагментов, способствует интенсификации исследований в области развития методов обработки текстов с применением машинного обучения. Недостаточное количество исследований в области русского языка объясняется не только меньшим числом исследователей, например, по сравнению с английским языком, но и объективно большей сложностью русского языка для автоматизированной обработки. Недостаточная разработанность стандартных средств кластеризации для коротких текстов и низкая эффективность существующих методов для русскоязычных текстов затрудняет их использование в российских автоматизированных системах поддержки принятия решений и управления.

Проведенный автором автореферата анализ современных моделей и методов кластеризации коротких текстов определил следующие основные задачи диссертационной работы: разработать метод расширения словаря языковой модели на базе нейронной сети; разработать метод для обработки экспертной информации в ходе нечеткой интерактивной кластеризации коротких текстов; составить алгоритм автоматизации работ по нечеткой интерактивной кластеризации коротких текстов в системе поддержки принятия решений и провести его апробацию.

Результаты исследования, полученные автором диссертации, обладают научной новизной. Вычислительные эксперименты подтверждают достоверность выдвигаемых положений. Практическая ценность работы состоит в том, что разработанный программный модуль, позволил осуществить интерактивную нечеткую кластеризацию наборов данных, состоящих из коротких текстов, в задаче анализа набора коротких текстов в рамках НИР в интересах Министерства экономического развития РФ для Системы стратегического планирования РФ.

В тексте автореферата можно выделить следующие незначительные недостатки:

1. Отсутствует описание обозначения  $\mu_i$  в формуле расстояния между элементом и центром кластера;
2. Не представлено описание компетенций эксперта, необходимых для выполнения интерактивной кластеризации коротких текстов в представленном эксперименте по определению эффективности разработанного метода.

Указанные недостатки не снижают общей ценности работы Дударина П.В. Считаю, что диссертационная работа представляет научный интерес и

практическую ценность и полностью удовлетворяет всем требованиям, предъявляемым к кандидатским диссертациям, а соискатель Дударин П.В. заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.01 «Системный анализ, управление и обработка информации (информационные технологии и промышленность)».

Директор Института системной  
и программной инженерии  
и информационных технологий

Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет «Московский институт электронной техники»

Д.т.н., профессор

Гагарина Лариса Геннадьевна

Дата 06.09.2021 г.

*Лариса Геннадьевна Гагарина*  
*Иванов Иван Иванович*  
*Иванов Иван Иванович*



Докторская диссертация защищена по специальности 05.13.07- Автоматизация технологических процессов и производств

Адрес организации: 124498, г. Москва, г. Зеленоград, площадь Шокина, д 1

E-mail: netadm@miee.ru

Телефон: (499) 731-44-41.

## ОТЗЫВ

на автореферат диссертации Дударина Павла Владимировича «Исследование и разработка моделей и методов нечеткой кластеризации коротких текстов», представленной на соискание ученой степени кандидата технических наук по специальности 05.13.01 «Системный анализ, управление и обработка информации (информационные технологии и промышленность)»

В настоящее время, в связи с бурным развитием информационных технологий, при анализе больших объемов структурированных и неструктурированных данных, активно применяется кластерный анализ текстов на различных языках. Проблемой кластерного анализа активно занимаются российские и зарубежные учёные, о чём свидетельствуют ежегодно проводимые научно-практические конференции и многочисленные публикации в научных изданиях. Автоматизация процесса решения задач кластеризации требует интерактивного участия экспертов, их знаний и опыта, что позволяет качественно повысить эффективность решения поставленной задачи с учётом современных методов и подходов при обработке текстовой информации. Учитывая вышесказанное, исследования в области интерактивной нечёткой кластеризации коротких текстов на русском языке, представляют собой важную и *актуальную* научно-техническую задачу.

Целью диссертационной работы Дударина П.В. являлось повышение эффективности нечёткой кластеризации коротких текстов путём разработки модели, метода и алгоритма в системе поддержки принятия решений для кластеризации коротких текстов на русском языке с учётом экспертной информации, а также разработке программного модуля, позволяющего проводить вычислительные эксперименты метода нечёткой интерактивной кластеризации коротких текстов.

В качестве новых результатов, полученных автором и выносимых на защиту, следует отметить:

– предложена архитектура искусственной нейронной сети, позволяющая решать задачу кластеризации на базе скрытого пространства признаков языковой модели;

– предложен метод обработки текстов для расширения словаря языковой модели на базе нейронной сети с использованием нечёткого иерархического классификатора, позволяющий учитывать семантическую близость слов;

– предложен метод обработки обратной связи от эксперта, позволяющий корректировать весовые коэффициенты нейронной сети и проводить интерактивную кластеризацию наборов коротких текстов;

– разработан алгоритм, автоматизирующий применение предложенных модели и методов для выполнения нечёткой интерактивной кластеризации

наборов коротких текстов, интегрированный в систему поддержки принятия решений.

Теоретическая значимость результатов диссертации заключается в разработке новых моделей и методов с использованием языковых моделей на базе нейронных сетей для решения задачи нечёткой кластеризации коротких текстов.

Практическая значимость заключается в повышении эффективности процесса кластеризации коротких текстов, сокращении временных затрат эксперта за счёт итеративности и интерактивности предлагаемого метода кластеризации с учётом обратной связи от эксперта, задаваемой в виде ограничений для процесса кластеризации на каждой итерации.

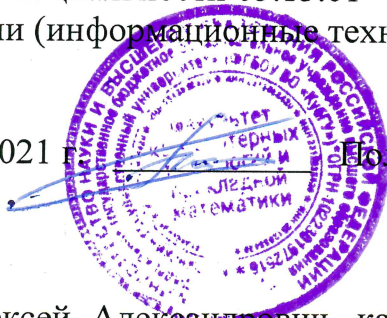
В то же время по автореферату диссертационной работы имеются следующие замечания:

1) в тексте автореферата диссертации не указаны размеры входного и выходного слоев в предлагаемой нейронной сети для кластеризации в серии проведённых экспериментов;

2) из представленного в тексте автореферата диссертации не понятно, каким образом происходит исправление орфографических ошибок на этапе предобработки текста для разработанного алгоритма нечёткой интерактивной кластеризации коротких текстов (НИККТ), приведённого на стр. 15, в п. 1.

Отмеченные недостатки не снижают научной и практической ценности работы, которая удовлетворяет требованиям, предъявляемым ВАК РФ к кандидатским диссертациям и установленным «Положением о присуждении ученых степеней», а автор диссертационной работы Дударин Павел Владимирович, заслуживает присуждения учёной степени кандидата технических наук по специальности 05.13.01 – Системный анализ, управление и обработка информации (информационные технологии и промышленность).

«10» сентября 2021 г.



Полупанов Алексей Александрович

Полупанов Алексей Александрович, канд. техн. наук, доцент кафедры информационных технологий.

Федеральное государственное бюджетное образовательное учреждение высшего образования «Кубанский государственный университет»,

350040, г. Краснодар, ул. Ставропольская, 149, факультет компьютерных технологий и прикладной математики,

телефон: +7 (861) 219-95-01, +7 (928) 273-44-90,

e-mail: polualex@mail.ru

## ОТЗЫВ

на автореферат диссертации Дударина Павла Владимировича «Исследование и разработка моделей и методов нечеткой кластеризации коротких текстов», представленной на соискание ученой степени кандидата технических наук по специальности 05.13.01 «Системный анализ, управление и обработка информации (информационные технологии и промышленность)»

Задачи систематизации и классификации являются трудоемкими и требуют вовлечения экспертных знаний и опыта. Для автоматизации процесса решения таких задач зачастую используют алгоритмы кластеризации. Исследования современных ученых в области кластеризации коротких текстов направлены на создание моделей способных определить сжатые векторные представления для таких текстов и тем самым оцифровать содержащуюся в них семантическую информацию. При этом важно учитывать контекст, цели исследования, варианты использования результатов и прочую информацию, не содержащуюся непосредственно в текстах, но известную исследователю заранее или открывающуюся в процессе проведения исследования. В связи с вышесказанным работа по созданию методов интерактивной кластеризации представляет собой *актуальную* научно-техническую задачу.

Основной целью диссертационной работы Дударина П.В. являлось 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.

В качестве значимых и новых результатов, полученных автором, следует отметить следующие:

– предложена архитектура искусственной нейронной сети, позволяющая решать задачу кластеризации на базе скрытого пространства признаков языковой модели;

– предложен метод обработки текстов для расширения словаря языковой модели на базе нейронной сети с использованием нечеткого иерархического классификатора, позволяющий учитывать семантическую близость слов;

– предложен метод обработки обратной связи от эксперта, позволяющий корректировать весовые коэффициенты нейронной сети и проводить интерактивную кластеризацию наборов коротких текстов;

– разработан алгоритм, автоматизирующий применение предложенных модели и методов для выполнения нечеткой интерактивной кластеризации наборов коротких текстов, интегрированный в систему поддержки принятия решений.

Теоретическая значимость результатов диссертации заключается в



разработке новых моделей и методов с использованием языковых моделей на базе нейронных сетей для решения задачи нечеткой кластеризации коротких текстов.

Практическое значение заключается в повышении эффективности процесса кластеризации коротких текстов, сокращение временных затрат эксперта за счет итеративности и интерактивности предлагаемого метода кластеризации с учетом обратной связи от эксперта задаваемой в виде ограничений для процесса кластеризации на каждой итерации.

К недостатку автореферата диссертационной работы следует отнести: *в тексте автореферата диссертации не указаны размеры входного и выходного слов предложенной нейронной сети для кластеризации, в серии проведенных экспериментов.*

Отмеченный недостаток не снижает научной и практической ценности работы, которая удовлетворяет требованиям, предъявляемым ВАК РФ к кандидатским диссертациям и установленным «Положением о присуждении ученых степеней», а автор диссертационной работы Дударин Павел Владимирович, заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.01 – Системный анализ, управление и обработка информации (информационные технологии и промышленность).

«09» 09 2021 г.  Лютикова Марина Николаевна

Лютикова Марина Николаевна, канд. техн. наук, доцент кафедры «Радиоэлектроники и информационных технологий».

Федеральное государственное бюджетное образовательное учреждение высшего образования «Государственный морской университет имени адмирала Ф.Ф. Ушакова»,  
353924, г. Новороссийск, пр. Ленина, 93, факультет эксплуатации водного транспорта и судовождения

телефон: +7(8617) 71-75-25 , +7 (903) 458-60-02.  
e-mail: mnlyutikova@mail.ru

  
Подпись \_\_\_\_\_  
  
Наличие ОК  
Сергеев Ело  
09.09.2021

## ОТЗЫВ

на автореферат диссертации Дударина Павла Владимировича «Исследование и разработка моделей и методов нечеткой кластеризации коротких текстов», представленной на соискание ученой степени кандидата технических наук по специальности 05.13.01 «Системный анализ, управление и обработка информации (информационные технологии и промышленность)»

Задача кластерного анализа является одним из методов системного анализа и в настоящее время активно применяется при анализе больших объемов структурированных и неструктурированных данных, в частности текстов на русском языке. Озвученной проблемой активно занимаются российские и зарубежные учёные, о чём свидетельствуют многочисленные публикации в научных изданиях. Автоматизация процесса решения задач кластеризации требует интерактивного участия экспертов, их знаний и опыта, что позволяет качественно повысить эффективность решения поставленной задачи с учётом современных методов и подходов при обработке текстовой информации.

Таким образом, исследования в области интерактивной нечёткой кластеризации коротких текстов на русском языке, представляют собой важную и *актуальную* научно-техническую задачу.

Основной целью диссертационной работы соискателя являлось 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.

В качестве новых результатов, полученных автором и выносимых на защиту, следует отметить:

– предложена архитектура искусственной нейронной сети, позволяющая решать задачу кластеризации на базе скрытого пространства признаков языковой модели;

– предложен метод обработки текстов для расширения словаря языковой модели на базе нейронной сети с использованием нечеткого иерархического классификатора, позволяющий учитывать семантическую близость слов;

– предложен метод обработки обратной связи от эксперта, позволяющий корректировать весовые коэффициенты нейронной сети и проводить интерактивную кластеризацию наборов коротких текстов;

– разработан алгоритм, автоматизирующий применение предложенных модели и методов для выполнения нечеткой интерактивной кластеризации наборов коротких текстов, интегрированный в систему поддержки принятия решений.

Теоретическая значимость результатов диссертации заключается в разработке новых моделей и методов с использованием языковых моделей на базе нейронных сетей для решения задачи нечеткой кластеризации коротких текстов.

Практическое значение заключается в повышении эффективности процесса кластеризации коротких текстов, сокращении временных затрат эксперта за счет итеративности и интерактивности предлагаемого метода кластеризации с учетом обратной связи от эксперта задаваемой в виде ограничений для процесса кластеризации на каждой итерации.

К недостатку автореферата диссертационной работы следует отнести:

*в тексте автореферата диссертации на рисунке 1 (стр. 11), показана разработанная соискателем архитектура нейронной сети для кластеризации коротких текстов, однако из рисунка не понятно, какая функция активации используется в приведённой архитектуре.*

Отмеченный недостаток не снижает научной и практической ценности работы, которая удовлетворяет требованиям, предъявляемым ВАК РФ к кандидатским диссертациям и установленным «Положением о присуждении ученых степеней», а автор диссертационной работы Дударин Павел Владимирович, заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.01 – Системный анализ, управление и обработка информации (информационные технологии и промышленность).

«10» 09 2021 г.  Никифорова Лариса Владимировна

Никифорова Лариса Владимировна, канд. техн. наук, доцент кафедры информационных и управляющих систем.

Федеральное государственное бюджетное образовательное учреждение высшего образования «Амурский государственный университет»,  
675027, Амурская обл., г. Благовещенск, Игнатьевское шоссе, 21,  
факультет математики и информатики,  
телефон: +7 (4162) 234-674, +7 (914) 381-81-65,  
e-mail: chepak@inbox.ru

