

ЗАСЕДАНИЕ ДИССЕРТАЦИОННОГО СОВЕТА Д212.277.04

Повестка дня:

Защита диссертации **Дудариним Павлом Владимировичем**  
на соискание ученой степени *кандидата технических наук*:

**"Исследование и разработка моделей и методов нечеткой  
кластеризации коротких текстов"**

Специальность :

**05.13.01 «Системный анализ, управление и обработка  
информации» (информационные технологии и промышленность) .**

Официальные оппоненты:

**Борисов Вадим Владимирович, доктор технических наук,  
профессор, профессор кафедры  
"Вычислительная техника" ФГБОУ ВО  
«Национальный исследовательский  
университет «МЭИ», г. Смоленск**

**Сухов Сергей Владимирович, кандидат физико-математических  
наук, старший научный сотрудник  
Ульяновского филиала ФГБУН "Институт  
радиотехники и электроники им.  
В.А.Котельникова"**

Ведущая организация – **ФГБУН "Институт автоматизации и процессов  
управления Дальневосточного отделения  
Российской академии наук (ИАПУ ДВО РАН)",  
г. Владивосток**



(Так как председатель диссертационного совета не может председательствовать на заседании совета при рассмотрении диссертации соискателя, у которого он является научным руководителем, то ведение данного заседания возложено на заместителя председателя диссертационного совета Киселева С.К.)

Председатель

**Уважаемые коллеги !**

На заседании диссертационного Совета Д212.277.04 из **23** члена Совета присутствуют **17** человек. Необходимый кворум имеем.

Членам Совета повестка дня известна. Какие будут суждения по повестке дня? Утвердить? (принято единогласно).

По специальности защищаемой диссертации **05.13.01 «Системный анализ, управление и обработка информации» (информационные технологии и промышленность)** (технические науки) на заседании присутствуют **4** докторов наук.

Наше заседание правомочно.

Председатель

Объявляется защита диссертации на соискание ученой степени кандидата технических наук **Дудариним Павлом Владимировичем** по теме: *"Исследование и разработка моделей и методов нечеткой кластеризации коротких текстов"*.

Работа выполнена в Ульяновском государственном техническом университете

Научный руководитель - **д.т.н. профессор Ярушкина Н.Г.**

**Официальные оппоненты:**

**Борисов Вадим Владимирович, доктор технических наук, профессор, профессор кафедры "Вычислительная техника" ФГБОУ ВО «Национальный исследовательский университет «МЭИ», г. Смоленск**

**Сухов Сергей Владимирович, кандидат физико-математических наук, старший научный сотрудник Ульяновского филиала ФГБУН "Институт радиотехники и электроники им. В.А.Котельникова", г. Ульяновск**

Присутствуют оба оппонента, оппонент Борисов В.В. подключен удаленно через средства телеконференцсвязи.

Письменные согласия на оппонирование данной работы от них были своевременно получены.

Ведущая организация - **ФГБУН "Институт автоматизи- ки и процессов управления Дальневосточного отделения Российской академии наук (ИАПУ ДВО РАН)", г. Владивосток.**

Слово предоставляется **Ученому секретарю** диссертационного Совета д.т.н. **А.М. Наместникову Д212.277.04** для оглашения документов из личного дела соискателя.

Ученый секретарь

Соискателем **Дудариним Павлом Владимировичем** представлены в Совет все необходимые документы для защиты кандидатской диссертации (зачитывает):

- заявление соискателя;
- копия диплома о высшем образовании (заверенная);
- справка об обучении в аспирантуре;
- заключение по диссертации от организации, где выполнялась работа;
- диссертация и автореферат в требуемом количестве экземпляров.

Все документы личного дела оформлены в соответствии с требованиями Положений ВАК.

Основные положения диссертации отражены **Дудариним П.В.** в **19** научных работах, в т.ч. в **6** статьях в изданиях из перечня ВАК, **7 публикациях индексируемых Scopus (Web of Science)**. Соискатель представлен к защите **23.06.2021 г.** (протокол №2). Объявление о защите размещено на сайте ВАК РФ **13.07.2021 г.**

Вся необходимая информация по соискателю внесена в ЕГИСМ.

Председатель

Есть ли вопросы по личному делу соискателя к ученому секретарю Совета? (Нет).

Есть ли вопросы к **Дударину П.В.** по личному делу? (Нет).

**Павел Владимирович,** Вам предоставляется слово для изложения основных положений Вашей диссертационной работы.

(Соискатель излагает основные положения).

Уважаемый председатель, уважаемые члены диссертационного совета, уважаемые присутствующие разрешите представить результаты диссертационной работы на тему: Исследование и разработка моделей и методов нечеткой кластеризации коротких текстов.

Стремительный рост массивов информации, состоящих из наборов коротких текстов в электронном виде способствует интенсификации исследований в области развития методов обработки текстов с применением машинного обучения. Примерами коротких текстов могут служить наборы постов в социальных сетях, новостных заголовков, объявлений. Проблеме ежегодно посвящается значительно число исследо-

ваний. Большая часть проводимых исследований относится к текстам на английском языке.

Недостаточная разработанность стандартных средств и низкая эффективность существующих методов для русскоязычных текстов затрудняет их использование в российских автоматизированных системах поддержки принятия решений и управления.

Целью 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.

И для достижения цели поставлены следующие задачи:

- провести исследование моделей и методов машинного обучения для обработки текстов для выявления новых подходов к повышению эффективности четкой и нечеткой кластеризации коротких текстов;
- 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;
- сформулировать перечень этапов программы проведения испытаний метода нечеткой интерактивной кластеризации коротких текстов;
- провести апробацию разработанных модели, методов и алгоритма нечеткой интерактивной кластеризации коротких текстов в качестве элементов функционирующей системы поддержки принятия решений.

Область исследования соответствует паспорту специальности 05.13.01. – «Системный анализ, управление и обработка информации (информационные технологии и промышленность)», а именно:

п. 4 – разработка методов и алгоритмов решения задач системного анализа, оптимизации, управления, принятия решений и обработки информации;

п. 13 – методы получения, анализа и обработки экспертной информации.

Научной новизной обладают:

- предложенная архитектура искусственной нейронной сети, отличающаяся от известных тем, что позволяет решать задачу кластеризации на базе скрытого пространства признаков языковой модели;
- 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;
- 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;
- разработанный алгоритм, автоматизирующий применение предложенных модели и методов для выполнения нечеткой интерактивной кластеризации наборов коротких текстов, интегрированный в систему поддержки принятия решений (СППР).

Основные положения, выносимые на защиту:

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.

Наибольших успехов в области обработки текстов на сегодня удастся добиться при использовании языковых моделей на основе нейронных сетей.

Широкое распространение получили предварительно обученные нейронные сети. Они позволяют осуществить перенос знаний и использовать многозадачное обучение.

Наиболее перспективными предварительно обученными сетями на сегодняшний день являются: ULMFit; ELMo и BERT.

Современные методы кластеризации представляют из себя монолитные нейронные сети, где стадии построения пространства признаков и группировки объектов неразрывно связаны.

При этом Тексты, являясь многомерными объектами, представляют особенную сложность для алгоритмов кластеризации.

А задача кластеризации коротких текстов рассматривается отдельно по причине ряда особенностей коротких текстов к которым относятся: разреженность векторов признаков и частичное или полное отсутствие контекста.

В большинстве случаев решения практических задач кластеризации участие исследователя необходимо либо для построения корректного разбиения на группы, либо принятия решения о структуре иерархии, либо способствует существенному повышению качества кластеризации.

На слайде приведена систематизация методов кластеризации с участием исследователя.

Особенно интерактивная кластеризация актуальна при обработке текстовой информации. Без участия эксперта, без выявления его скрытых интенций невозможно заранее определить, какое именно разбиение ожидается в результате работы алгоритма.

И наилучшим образом для решения поставленной задачи подходят методы с обратной связью в виде оценки результата кластеризации.

При этом доля объектов, с которыми взаимодействует эксперт, как правило не превышает 10%.

На слайде представлены наиболее близкие актуальные исследования.

В части интерактивной кластеризации текстов на английском языке наиболее близким является исследование Дэвида Коэна "Semi-supervised Clustering with User Feedback", в этой работе обратная связь давалась по парам объектов.

А в части кластеризации на базе нейронных систем работа Unsupervised Deep Embedding for Clustering Analysis в которой был представлен алгоритм DEC и показаны его преимущества перед другим алгоритмами кластеризации.

Итак, в результате исследований было выявлено, что узким местом применения предварительно обученных языковых моделей является словарь. Словари языковой модели и исследуемого набора данных могут существенно отличаться. При этом существующие методы либо сопоставляют словам не входящим в исходный словарь языковой модели нулевой вектор либо вектор равный среднему всех векторов словаря.

Для решения задачи расширения словаря языковой модели в нейронную сеть добавляется дополнительный слой. В этом слое происходит сопоставление "неизвестным" для языковой модели словам вектора являющегося линейной комбинацией наиболее семантически близких слов из словаря языковой модели.

Итого задача заключается в определении метода расчета весовых коэффициентов в линейной комбинации представленной на слайде.

На первом шаге для множества семантически близких слов к целевому (новому, незнакомому) слову строится нечеткий граф (веса на ребрах равны семантической близости между словами).

Далее при помощи иерархической модификации алгоритма  $\epsilon$ -кластеризации нечеткого графа строится иерархический классификатор, позволяющий выделить и классифицировать лексические значения слова.

На рисунке представлен граф и классификатор для слова "график". Видно, что у этого слова есть два главных лексических значения график как диаграмма или схема, и график как расписание. Расчет весов для линейной комбинации векторов слов производится при помощи алгоритма работа которого представлена на примере на слайде.

Данный алгоритм позволяет распределить веса пропорционально по полученной иерархии, таким образом чтобы более многочисленные группы синонимов не получали преимущества. Таким образом, лексическое значение "Расписание" слова График получит значимый вес в результирующей линейной комбинации.

Таким образом все слова целевого словаря получают в соответствие векторы, которые могут быть использованы в предварительно обученной языковой модели.

Для решения задачи интерактивной кластеризации коротких текстов предлагается следующая архитектура нейронной сети представленная на слайде. В нее встроены предварительно обученная языковая модель как блок, отвечающий за преобразование текста в сжатое векторное представление. И непосредственно блок кластеризации.

В приведенной нейронной сети процессы определения оптимального расположения центров кластеров и построения пространства признаков происходят одновременно за счет определения общей функции потерь на базе метрики Кульбака-Лейблера для двух распределений: фактического распределения объектов по кластерам и целевого, формулы, которых представлены на слайде.

Вышеописанный подход позволяет провести обучение сети и получить первичную кластеризацию в которой еще НЕ были учтены знания и цели эксперта.

Для учета обратной связи от эксперта большинство методов кластеризации предлагают эксперту манипулировать мета-параметрами ал-

горитма, которые в большинстве своем не имеют явной связи с парами "объект - центра кластера".

В отличие от них в данной работе предполагается обратная связь двух видов: "элемент  $X_i$  должен принадлежать кластеру  $C_j$ " и "элементу  $X_i$  не следует находиться в кластере  $C_j$ ".

Такой вид обратной связи легко формулируется экспертами любого уровня подготовки. Первый тип ограничений соответствует ограничениям, используемым в алгоритмах классификации. Второй тип ограничений предложен впервые. Он позволяет формулировать менее жесткие ограничения, что уменьшает трудоемкость работы эксперта. Одновременно может быть получено произвольное количество таких ограничений. В результате работы эксперт формирует матрицу обратной связи.

Формулы 6, 7, и 8 позволяют включить сформулированные ограничения в процесс кластеризации и обучения сети. Добавляемые ограничения не являются жесткими, не приводят к необходимости решать системы уравнений, которые могут потенциально оказаться несовместными.

Для использования предложенных методов в реальных системах требовалось объединить их в единый алгоритм, в котором были выделены следующие шаги: после исправления орфографии и разбиения текста на токены, для токенов отсутствующих в языковой модели производится расширение словаря языковой модели, после чего модель дополнительно обучается под предметную область и исследуемый набор текстов. Затем следуют этапы позволяющие произвести первичную кластеризацию текстов и далее многократно повторяется этап интерактивной кластеризации, до тех пор пока требования эксперта не будут полностью удовлетворены или же число объектов перемещаемых между кластерами с введением очередных ограничений не станет достаточно мало.

Разработанный алгоритм был автоматизирован и применен в рамках Федеральной информационной системы Стратегического планирования.

Одной из составляющих системы стратпланирования является набор ключевых показателей эффективности. Эти показатели присутствуют в каждом из 600 тыс. документов.

Требовалось провести исследование более полумиллиона показателей (представляющих из себя короткие тексты на русском языке, примеры которых приведены на слайде). В самой системе имелся нормативно установленный классификатор состоящий из 11 классов, это крупные верхнеуровневые классы такие как Образование, Здравоохранение и т.п. Очевидно, что сотни тысяч показателей относились к каждой категории. Необходима была более точная классификация имеющегося набора коротких текстов. Ручная обработка при таких объемах не представлялась возможной. В связи с чем была сформулирована задача нечеткой кластеризации набора коротких текстов. При длине текста 10-100 слов.

Для апробации алгоритма с экспертом был выбран раздел "Общее образование". Общий объем набора данных для кластеризации составил 6 975 показателей. Все показатели были кластеризованы предложенной моделью, после чего эксперт точно вносил корректировки в кластеры в виде дополнительных ограничений и запускал очередную итерацию алгоритма.

Подробнее рассмотрим кластер №14, который при первичной кластеризации был назван как "Аттестация". При знакомстве с содержимым



кластера эксперт отметил две принципиально разные группы показателей: аттестация преподавателей и аттестация учащихся. В обратной связи от эксперта первые три показателя из кластера "Аттестация" (№14) были отмечены для вынесения из состава кластера. В результате шага интерактивной кластеризации были сформированы новые кластеры. Среди которых кластер для аттестации персонала №24, а кластер №14 был переименован в "Аттестацию учащихся".

Сравнительный анализ трудоемкости показал эффективность предложенного алгоритма более чем в два раза по сравнению с использованием традиционных методов обработки, таких как тематическая классификация. Суммарная экономия времени на всем наборе составила более 9 человеко-месяцев.

В результате проведенной интерактивной кластеризации на всем объеме ключевых показателей в ходе которой активно использовались экспертные знания, был составлен классификатор из более чем 1000 классов организованных в иерархию с максимальной глубиной в три уровня. Данный классификатор заменил собой первоначальный классификатор ЕМИСС. И был получен инструмент способный автоматически проводить классификацию новых показателей. В результате этого изменилась роль экспертов минэкономразвития и снизилась нагрузка на них.

Для проведения лабораторных испытаний представленных методов, определения точности и границ применимости была сформулирована программа испытаний. Включающая в себя пункты по проверке работоспособности метода на синтетических наборах данных. Важно было убедиться, что интерактивный метод позволяет получить произвольное разбиения множества на заданное число кластеров.

Также были проведены эксперименты на реальных наборах данных. Основная часть экспериментов проводилась с набором коротких текстов от компании Avito. Это объявления разбитые на 4 категории, всего в наборе 489 тысяч объявлений. Примеры объявлений представлены на слайде. Для эксперимента были случайным образом отобраны 2000 объявлений. Метод расширения словаря позволил не только избежать снижения точности языковой модели, но и даже увеличить ее.

При этом первичная кластеризация самих текстов, как и ожидалось, не дала высоких результатов и позволила достичь уровня точности в 52,5%. Далее следовали шаги интерактивной кластеризации.

В итоге были получены следующие результаты:

С ростом количества ограничений точность предсказуемо повышается. При добавлении количества ограничений более 5% от общего числа примеров график точности замедляет свой рост.

В результате 10-ти шагов интерактивной кластеризации по 10 ограничений в каждом была достигнута точность 80.1%

Также важно отметить влияние дополнительного обучения слоев языковой модели, которое позволило увеличить точность на 10% и понизить количество необходимых ограничений.

Так как в литературе нет однозначного определения понятия короткого текста, то был произведен эксперимент для определения границ применимости предлагаемого метода. В результате которого были установлены оптимальные границы в диапазоне от 5 до 100 слов в тексте.

Основные результаты и выводы представлены на слайде:

Проведено исследование моделей и методов машинного обучения для обработки текстов и разработана архитектура искусственной

нейронной сети, реализующей кластеризацию на базе пространства признаков языковой модели русского языка.

Разработан метод обработки текстов для расширения словаря языковой модели на базе нейронной сети с использованием нечеткого иерархического классификатора, который позволяет повысить точность кластеризации в среднем на 10%.

Разработан метод для обработки обратной связи от эксперта, используемый для корректировки весовых коэффициентов нейронной сети, что позволяет проводить интерактивную кластеризацию наборов коротких текстов.

Сформулирован перечень этапов программы проведения испытаний метода нечеткой интерактивной кластеризации коротких текстов. Проведенные исследования по данной программе позволили установить границы применения предлагаемого метода. Метод наиболее эффективен для текстов с количеством слов от 5 до 100. В ходе проведенных исследований была достигнута средняя точность кластеризации 80% при более низком числе дополнительных ограничений по сравнению с аналогичными методами. На разработанный программный модуль, использованный для проведения численных экспериментов, получено Свидетельство о государственной регистрации программы для ЭВМ № 2021615642.

Составлен алгоритм автоматизации работ по нечеткой интерактивной кластеризации коротких текстов в СППР "Федеральная система стратегического планирования РФ (ФИС СП)".

Проведена апробация разработанных модели, методов и алгоритма нечеткой интерактивной кластеризации коротких текстов в качестве элементов СППР "ФИС СП". Внедрение алгоритма позволило решить задачу составления автоматического классификатора ключевых показателей эффективности документов стратегического планирования, при этом экономия трудозатрат эксперта оценивается более чем в 9 человеко-месяцев. Наличие автоматического классификатора позволяет снизить временные затраты экспертов при проведении процедуры проверки корректности классификации входящих (новых) документов.

Разработанный метод интерактивной кластеризации является универсальным и может быть применен для различных наборов коротких текстов. Предложенный метод может быть доработан для использования с различными языковыми моделями и обобщен на случай совместной работы ряда экспертов.

Основные результаты диссертационного исследования опубликованы в 19 печатных работах, в том числе 6 статей в российских рецензируемых научных журналах из Перечня, рекомендованного ВАК РФ, 7 публикаций в изданиях индексируемых в Scopus и Web of Science, 6 публикаций в материалах всероссийских и международных научных конференций.

Результаты диссертационной работы использованы в рамках НИР в интересах Министерства экономического развития РФ. Результаты НИР внедрены в системе ФИС СП.

Архитектура искусственной нейронной сети и алгоритм нечеткой кластеризации коротких текстов, методы расширения словаря языковой модели и корректировки весов нейронной сети для учета обратной связи эксперта в интерактивной кластеризации, а также программная реализация метода нечеткой интерактивной кластеризации внедрены в системе Планета.Аналитика, входящей в реестр отечественного ПО.

На этом доклад закончен, благодарю за внимание.

Председатель

У кого есть вопросы к соискателю?

(Соискатель отвечает на вопросы)

д.т.н, профессор Пиганов М.Н.

Скажите, пожалуйста, у Вас в первом пункте научной новизны указывается предложенная архитектура нейронной сети, при этом в задачах разработку архитектуры нейронной сети Вы не ставите. Как можно это объяснить?

Соискатель

В результате решения первой задачи исследования современных методов кластеризации и обработки текстов, было установлено, что наиболее эффективными являются методы, базирующиеся на искусственных нейронных сетях. В результате при разработке метода интерактивной кластеризации была использована нейронная сеть, архитектура которой и заявлена в положении, выносимом на защиту.

д.т.н, профессор Ташлинский А.Г.

В чем заключается новизна предложенной Вами архитектуры нейронной сети.

Соискатель

Новизна предложенной архитектуры заключается в том, что первый раз языковая модель включается в состав сети кластеризации. В результате в процессе интерактивной кластеризации происходит дополнительное обучение сети.

д.т.н, профессор Ташлинский А.Г.

Дополнительное обучение языковых моделей используется сейчас достаточно активно, чем отличается Ваш подход?

Соискатель

В предложенной нейронной сети дополнительное обучение происходит как часть процесса кластеризации и взаимодействия с экспертом. В этом заключается особенность.

д.т.н, профессор Ташлинский А.Г.

Хорошо. Получается, что сеть обучается за счет использования экспертных знаний. Как Вы оцениваете трудоемкость обучения для достижения точности 80% и выше?

Соискатель

В проведенных экспериментах эксперт взаимодействовал не более чем с 10% объектов от общего объема исследуемых наборов данных. При этом даже при работе с 5% объектов достигаются удовлетворительные результаты кластеризации.

д.т.н, профессор Ташлинский А.Г.

Хотелось бы услышать оценку именно трудоемкости эксперта в часах или днях.

Соискатель

Для 2000 объектов эксперимент длился от 2 до 3-х часов.

д.т.н, профессор Ташлинский А.Г.

Хорошо. У Вас в 4-м выводе указано, что достигается точность 80%. Как связан с этими выводами рисунок 22 в Вашем докладе, где приведены меньшие значения.

Соискатель

Данный рисунок представляет результаты эксперимента по определению границ применимости метода, тут представлены оценки точности по группам текстов определенной длины. Для каждой группы проводились замеры точности первого этапа кластеризации, автоматического, без участия эксперта, точность на них, как и было сказано в докладе, не высокая. Для первой и последней групп текстов проводился также эксперимент с интерактивной кластеризацией и результаты приведены на графике.

д.т.н, профессор Ташлинский А.Г.

Да, это понятно. Но почему для первой группы точность получена меньше 80%, как это можно объяснить? Уточню вопрос, у Вас написано, что в пределах от 10 до 100 слов метод наиболее эффективен. При этом график 22 не достигает 80%.

Соискатель

Дело в том, что для текстов из диапазонов 20-50 и 50-100 отдельно не проводились эксперименты по интерактивной кластеризации, а именно эти тексты и давали вклад в оценку точности в экспериментах, где был получен результат 80%.

д.т.н, профессор Ташлинский А.Г.

Как тогда следует трактовать рисунок 22 в части результатов интерактивной кластеризации?

Соискатель

В данном случае следует опираться на интерполяцию по результа-

там для двух групп текстов первой и последней, и на динамику результатов первичной кластеризации.

д.т.н, профессор Ташлинский А.Г.

Хорошо. Вот Вы предлагаете новый метод обработки текстов, и утверждаете, что его точность увеличивается на 10%, но в выводах в докладе не указано по сравнению с чем происходит это увеличение.

Соискатель

Увеличение точности на 10% происходит за счет включения языковой модели непосредственно внутри сети кластеризации, а именно по сравнению с экспериментами в которых языковая модель использовалась лишь для получения сжатых векторных представлений текстов, и не происходило дополнительное обучение в процессе кластеризации.

д.т.н, профессор Ташлинский А.Г.

Таким образом, Вы сравниваете между собой показатели точности из двух экспериментов, проведенных Вами? Происходило ли сравнение с какими-либо известными результатами экспериментов?

Соискатель

Для русскоязычных текстов найти известные результаты экспериментов по интерактивной кластеризации коротких текстов не удалось. При этом наиболее близком аналогом, встретившимся в научной литературе, является эксперимент Дэвида Коэна, результаты которого представлены на слайде. Так как эксперименты проводились на разных наборах коротких текстов, то производить численное сравнение не корректно.

д.т.н, профессор Ташлинский А.Г.

Хорошо, если не количественное, то возможно ли качественное сравнение?

Соискатель

Качественное сравнение показывает схожую динамику, при этом видно, что предложенный метод позволяет использовать меньшее число ограничений при схожей точности, а возможно даже и лучшей.

д.т.н, профессор Ташлинский А.Г.

Понятно. И последний вопрос, правильно ли я понимаю, что у Вас не только сформулирован перечень этапов программы испытаний, как указано в результатах, но и проведены эти испытания?

Соискатель

Да, конечно, были проведены все запланированные эксперимент и

далее в результатах указано, что была проведена апробация и внедрение.

д.т.н, профессор Крашенинников В.Р.

Вы используете понятие точности кластеризации. А как Вы ее вычисляете?

Соискатель

Когда мы говорим о точности кластеризации, предполагается наличие эталонного разбиения на известные классы, далее считается доля корректно соотнесенных объектов.

д.т.н, профессор Крашенинников В.Р.

Откуда появляется этот эталон?

Соискатель

В проводимых экспериментах использовались наборы данных из задач классификации из открытых источников, в которых классы известны для всех объектов.

д.т.н, профессор Крашенинников В.Р.

Насколько мы можем доверять этому известному разбиению на классы?

Соискатель

Соотнесение по классам проводилось большим количеством экспертов и используется в течении длительного времени.

д.т.н, профессор Крашенинников В.Р.

То есть имеется вероятность определенной доли ошибки в этом эталонном разбиении. В связи с этим следующий вопрос. Проводилась ли оценка статистической значимости результатов или же мы говорим только о разнице средних значений?

Соискатель

В данной работе даются усредненные оценки точности по 10 экспериментам в той части, где происходит автоматический расчет и по 3-м экспериментам, где требовалось участие эксперта.

д.т.н, профессор Крашенинников В.Р.

А при сравнении с результатами полученными другими методами использовались ли оценки доверительных интервалов для полученных численных значений?

Соискатель

Нет, для точности такие методы оценивания не применялись.

д.т.н, профессор Васильев К.К.

Павел Владимирович, по Вашему докладу сложилось впечатление, что для английского языка проблема кластеризации коротких текстов решена, а для русского нет, так как Вы работаете только с ним. В связи с чем возникает вопрос, почему нельзя использовать имеющиеся методы для английского языка применительно к русскому? В чем принципиальное отличие проблем кластеризации русскоязычных и англоязычных коротких текстов?

Соискатель

В данной работе ставилась задача кластеризации именно русскоязычных текстов, поэтому не проводились эксперименты для других языков. При этом задача кластеризации коротких текстов для других языков также не решена.

д.т.н, профессор Васильев К.К.

Таким образом, следующим шагом Вам нужно будет развить ваши модели и применить предложенный алгоритм к текстам на других языках?

Соискатель

Да, совершенно верно, планирую такую работу.

д.т.н, профессор Сергеев В.А.

Зависит ли точность от количества текстов в выборке? Допустим при использовании 1000 текстов в выборке вместо 2000, как изменятся результаты?

Соискатель

Эксперименты проводились неоднократно, каждый раз выбиралось случайный набор текстов и проводилось усреднение показателей точности. Разброс значений был не большим, поэтому можно сказать, что результаты на других размерах выборок будут аналогичными.

д.т.н, профессор Сергеев В.А.

Можете ли Вы дать численную оценку разброса?

Соискатель

В пределах 1-2%.

д.т.н, профессор Сергеев В.А.

При обучении Вашей модели на 2000 текстов, и если требуется

добавить к выборке еще какое-то количество текстов, придется ли переобучать вашу модель?

Соискатель

Нет, модель переобучать в таком случае не придется, в данном случае модель будет использована в качестве автоматического классификатора и новым текстам будут присвоены ранее определенные кластеры.

д.т.н, доцент Епифанов В.В.

В положениях выносимых на защиту указан ряд разработанных методов, при этом в раздаточном материале они не выделены явным образом, можете показать их на слайдах?

Соискатель

Да, конечно. Разработано два метода, первый это метод расширения словаря языковой модели, он представлен на слайдах 10 и 11. А также разработан метод учета обратной связи от эксперта в ходе интерактивной нечеткой кластеризации, представленный на слайде 14.

д.т.н, доцент Епифанов В.В.

Можете открыть плакат 11. На нем указаны расстояния и веса в графе, каким образом они получены?

Соискатель

Расстояния получены из модели

Для расчета весов используется следующий алгоритм: после построения иерархического классификатора, задается допустимый уровень, начиная с которого происходит распределения весов по иерархии вниз, от кластеров до листовых элементов. Первоначально распределение происходит пропорционально семантическому расстоянию от слова "график" до центров кластеров. Далее для каждого под-дерева процедура повторяется. В итоге слова из более многочисленных групп синонимов не получают преимущества.

д.т.н, доцент Епифанов В.В.

Далее 13 плакат, пожалуйста. Формулы, указанные на этом плакате, получены Вами или заимствованные?

Соискатель

Формулы, представленные на 13-м слайде, описывают метод DEC, который был взят для реализации шага первичной кластеризации, и далее на его основе мной был предложен метод интерактивной кластеризации.

д.т.н, доцент Епифанов В.В.



Можете отметить формулы предложенные лично Вами?

Соискатель

Да, это слайд 14, формулы 6,7 и 8.

д.т.н, доцент Епифанов В.В.

Далее плакат 15. Описанные шаги алгоритма были использованы только для набора данных системы стратегического планирования, или же они одинаковы для всех проведенных экспериментов?

Соискатель

Все эксперименты выполнялись по одному и тому же алгоритму, представленному на слайде.

д.т.н, доцент Епифанов В.В.

А кем выполнялись шаги алгоритма?

Соискатель

Шаги с 1 по 6-ой выполняются автоматически при помощи разработанного программного модуля, а на шаге 7-м происходит привлечение эксперта для формирования ограничений в рамках интерактивного взаимодействия с модулем кластеризации.

д.т.н, доцент Епифанов В.В.

Можете описать область применения Вашего алгоритма? Например, можно ли применить данный метод кластеризации в приборостроении? Как определяете применимость вашего метода?

Соискатель

Могу рекомендовать применение данного метода при наличии набора коротких текстов длины от 10 до 100-150 слов в количестве от 2000 и более и необходимости разбиения их на ранее неизвестные кластеры. Также важно наличие эксперта по предметной области готового проводить интерактивную кластеризацию. При этом языковой домен исследуемых текстов не важен.

д.т.н, профессор Клячкин В.Н.

Первый пункт в списке результатов говорит о проведенном исследовании моделей и методов машинной обработки текстов. Что Вы включаете в этом понятие, кроме кластеризации, которой Вы непосредственно занимались?

Соискатель

В первую очередь это языковое моделирование, т.е. предсказание

следующего слова по последовательности предшествующих слов. Классификация текстов, определение именованных сущностей, sentiment-анализ. Это задачи обработки текстов. Если говорить о методах решения, то на сегодняшний день решение этих задач базируется на языковом моделировании, а также методах переноса знаний из предварительно обученных языковых моделей и методе многозадачного обучения. В рамках работы исследовались такие языковые модели как BERT, ULMFiT и RuBERT.

д.т.н, профессор Клячкин В.Н.

Вопрос по плакату 13: чем обусловлен выбор метрики использованной для определения фактического расстояния от объекта до центра кластера?

Соискатель

Данная метрика унаследована от алгоритма DEC, положенного в основу. Это обычное распределение Коши, хорошо зарекомендовавшее себя в задачах понижения размерности и кластеризации.

д.т.н, профессор Клячкин В.Н.

Можете пояснить соотнесение представленной метрики с распределением Коши?

Соискатель

В знаменателе находится сумма всех расстояний для нормировки, а если посмотреть на числитель и заменить разность  $z_i - \mu_j$  на  $t$ , то получится формула распределения Коши.

д.т.н, профессор Клячкин В.Н.

Можно подробнее по значениям переменных. В частности, что такое  $k$  над знаком суммирования?

Соискатель

$k$  - это количество кластеров, по которым идет распределение объектов. Далее  $z_i$  - это объекты, а  $\mu_j$  - это центры кластеров.

д.т.н, профессор Дьяков И.Ф.

Скажите, пожалуйста, в научной новизне у Вас указана предложенная новая архитектура нейронной сети. Что Вы вкладываете в понятие новой архитектуры сети?

Соискатель

В данной архитектуре новым является набор слов и их смысловое наполнение, т.е. назначение каждой группы слов, их функция. В частности есть группа, отвечающая за построение векторов признаков

для текстов, и есть группа, отвечающая непосредственно за кластеризацию.

д.т.н, профессор Дьяков И.Ф.

Существует 24 типа нейронных сетей, почему потребовалась разработка вашей архитектуры?

Соискатель

В данной работе не разрабатывает новый тип нейронной сети, используются широко известные искусственные нейронные сети с глубоким обучением. Использовались известные типы слоев: полносвязный и LSTM. Необходима была разработка подходящего набора и группировки слоев для решения поставленной задачи.

д.т.н, профессор Дьяков И.Ф.

Как происходило обучение нейронной сети?

Соискатель

Использовался метод обратного распространения ошибки, и в частности алгоритм Adam, являющийся более эффективной модификацией метода градиентного спуска.

д.т.н, профессор Дьяков И.Ф.

Как Вы определяли точность (качество) обучения вашей сети? Использовался расчет корреляции или какой-то иной программный метод?

Соискатель

Для определения качества обучения нейронной сети использовался метод разделения исследуемого набора данных на обучающую и валидационную выборки. При этом для расчета ошибки использовалось расстояние Кульбака-Лейблера примененное к фактическому и целевому распределениям.

д.т.н, доцент Негода В.Н.

У Вас имеется машинная классификация, не естественная. Возникает вопрос: в какой мере влияет целе-ориентация для принятия решения? Допустим, выходит какой-то документ, который заставляет нас переставлять приоритеты, в результате чего должна измениться классификация. Как это повлияет на ваш процесс кластеризации?

Соискатель

Этот метод и был разработан для того, чтобы дать возможность человеку, эксперту, повлиять на результаты кластеризации и задать направление этой кластеризации. Т.е. в случае полностью автоматической кластеризации эксперт может либо принять результаты кластеризации, либо отклонить их. При этом в случае смены приоритетов

или выхода каких-либо новых нормативных документов, необходима возможность корректировки работы автоматического классификатора, т.е. изменения структуры кластеров. Эту задачу и решает интерактивная кластеризация и учет экспертной информации. При смене приоритетов эксперт должен указать новые ограничения и структура кластеров будет перестроена. Это было показано на примере отделения кластера Аттестации педагогов от кластера Аттестации учащихся.

д.т.н, доцент Наместников А.М.

Можно ли считать, что проблема недостатка контекста, свойственная коротким текстам, как раз и компенсируется интерактивностью метода.

Соискатель

Да, можно сказать, что контекст задается экспертом в ходе интерактивного взаимодействия.

д.т.н, доцент Киселев С.К.

На 15-м плакате, поясните, пожалуйста, что Вы подразумеваете под интерактивностью? Это взаимодействие модуля кластеризации с экспертом? Или же это взаимодействие модулей системы через определенный набор интерфейсов описанных на данном слайде на рисунке в правом нижнем углу?

Соискатель

В диссертационной работе интерактивное взаимодействие подразумевается между модулем кластеризации и специалистом-экспертом, т.е. человеко-машинное взаимодействие.

Председатель

Есть еще вопросы? (Нет).

**Согласны ли члены Совета сделать технический перерыв?** (Нет).  
Тогда продолжаем работу.

Слово предоставляется научному руководителю работы **профессору Ярушкиной Н.Г.**

Уважаемые коллеги! Разноплановая дискуссия уже позволила вам составить впечатление о степени подготовленности и вовлеченности соискателя Дударина Павла Владимировича в тематику, которой он занимается.

Павел Владимирович завершил обучение в аспирантуре не дольше чем год назад, и сейчас представил результаты своей работы. На мой взгляд, Павел Владимирович сложившийся специалист, имеющий замечательную особенность: закончив специальность Прикладная математика и информатика в УлГУ, он хорошо и широко владеет математическими методами, умеет применять их в сфере информационных технологий. В то же время он обладает широким набором компетенций связанных с информационными технологиями, является экспертом в области больших

данных и обработки естественного языка. Именно это позволило использовать данные результаты в важной информационной системе Стратегического планирования Российской Федерации, пользователями которой является Министерство экономического развития РФ.

Уверена, что Павел Владимирович продолжит занятие наукой. Могу его охарактеризовать как ученого только очень хорошо.

Кроме того Павел Владимирович в настоящий момент успешно преподаёт достаточно наукоемкие дисциплины для магистров кафедры Информационные системы.

*(Отзыв прилагается).*

Председатель

**Ученому секретарю Совета** предоставляется слово для оглашения заключения организации, где выполнялась работа и отзыва ведущей организации.

**Ученый секретарь** оглашает заключение организации, где выполнялась работа. Затем зачитывает отзыв ведущей организации.

*(Заключение и отзыв прилагаются).*

Председатель

На автореферат диссертации поступило 9 отзывов, все они положительные. Согласны ли члены Совета заслушать обзор отзывов или зачитать их полный текст? (Согласны)

Слово для обзора отзывов, поступивших на диссертацию, предоставляется **Ученому секретарю Совета**.

### **Ученый секретарь зачитывает обзор отзывов.**

#### **1. Санкт-Петербургский государственный университет.**

Отзыв подписан профессором кафедры информатики, д.ф.-м.н., профессором Тулупьевым А.Л.

Замечания:

• в тексте автореферата диссертации не отражена архитектура автоэнкодера используемого для инициализации весов в блоке кластеризации основной нейронной сети.

#### **2. Санкт-Петербургский федеральный исследовательский центр РАН.**

Отзыв подписан руководителем лаборатории теоретических и междисциплинарных проблем информатики, к.т.н. Абрамовым М.В.

Замечания:

• в тексте автореферата диссертации отсутствует обоснование выбора корпуса «Тайга» и обученной на нем модели word2vec в качестве базовой для работы метода расширения словаря языковой модели.

#### **3. Федеральный исследовательский центр «Информатика и управление» РАН.**

Отзыв подписан ведущим научным сотрудником, д.ф.-м.н. Аверкиным А.Н.

Замечания:

• в тексте автореферата диссертации не представлены результаты проведенного эксперимента по сравнению с качеством кластери-

зации достигаемом с помощью альтернативного метода примененного к англоязычному набору текстов.

#### **4. Южный федеральный университет.**

Отзыв подписан заведующим кафедрой «Системы автоматизированного проектирования», д.т.н., профессором Курейчиком В.В.

Замечания:

- ограниченность проведенных экспериментов только русским языком.

#### **5. Балтийский федеральный университет имени И.Канта.**

Отзыв подписан профессором Института физико-математических наук и информационных технологий, д.т.н., профессором Колесниковым А.В.

Замечания:

- в описании результатов проведенных экспериментов не указаны характеристики процесса обучения: количество эпох обучения, шаг обучения и значения целевой функции на обучающей и валидационной выборках.

#### **6. Национальный исследовательский университет «Московский институт электронной техники».**

Отзыв подписан директором Института системной и программной инженерии и информационных технологий, д.т.н., профессором Гагариной Л.Г.

Замечания:

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

#### **7. Кубанский государственный университет.**

Отзыв подписан доцентом кафедры информационных технологий, к.т.н. Полупановым А.А.

Замечания:

- в тексте автореферата диссертации не указаны размеры входного и выходного слоев в предлагаемой нейронной сети для кластеризации в серии проведенных экспериментов;
- из текста автореферата диссертации не понятно, каким образом происходит исправление орфографических ошибок на этапе preprocessing текста для разработанного алгоритма нечеткой интерактивной кластеризации коротких текстов.

#### **8. Амурский государственный университет.**

Отзыв подписан доцентом кафедры информационных и управляющих систем, к.т.н. Никифоровой Л.В.

Замечания:

- на рисунке 1 (стр. 11), на котором показана разработанная архитектура нейронной сети не указана используемая функция активации.

#### **9. Государственный морской университет имени адмирала Ф.Ф. Ушакова.**

Отзыв подписан доцентом кафедры радиоэлектроники и информационных технологий, к.т.н. Лютиковой М.Н.

Замечания:

- недостаточность информации в тексте автореферата диссертации о размерах входного и выходного слоя используемой для кластеризации нейронной сети.

*(Отзывы прилагаются).*

Председатель

Слово для ответа на замечания по заключению и отзывам предоставляется соискателю.

*(Соискатель отвечает на замечания по отзывам)*

Спасибо. На замечания ведущей организации у меня есть пояснение к 6 пункту о размере наборов данных для экспериментов: дополнительное обучение языковых моделей проводилось на полном объеме текстов, для интерактивной кластеризации тексты выбирались случайным образом для обеспечения чистоты экспериментов, а размер выборки в 2000 текстов обусловлен располагаемым временем экспертов участвовавших в экспериментах. С остальными замечаниями согласен.

С остальными замечаниями ведущей организации согласен.

С большинством замечаний прозвучавших в отзывах на автореферат диссертации согласен, только на замечание в отзыве доцента кафедры информационных технологий ФГБОУ ВО «Кубанского Государственного Университета», кандидата технических наук Полупанова Алексея Александровича 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, могу пояснить, что исправление орфографических ошибок, происходящее на этапе предварительной обработки текстов, производилось сторонним бесплатным сервисом от компании Яндекс.

Председатель

Слово для отзыва предоставляется официальному оппоненту - **д.т.н. Борису Вадиму Владимировичу.**

Добрый день, уважаемые коллеги! Позвольте мне долго не останавливаться на актуальности работы, о которой уже было достаточно много и убедительно сказано и в докладе и в отзыве ведущей организации. Прежде всего, хочу отметить, что существуют проблемы обработки коротких текстов, характеристики которых достаточно грамотно определил в своей работе соискатель, также определив границы применимости своего метода. Хотя, как мне кажется, предлагаемые подходы могут быть расширены и на более широкие классы текстов и иных объектов, как развитие данного диссертационного исследования.

Естественно, что особенностями коротких текстов являются разреженность векторов, а также полисемия, синонимия, проблемы связанные с расшифровкой аббревиатур. При этом мне бы хотелось отметить очень хороший обширный обзор, представленный в первой главе, показывающий современное состояние дел в области обработки текстов. Этот обзор характеризует анализ не только современного состояния, но и дает некую ретроспективу в области языковых моделей и современных методов машинного обучения, используемых в задачах обработки естественного языка. Тут следует отметить хороший техни-

ческий язык изложения, в том числе, что достаточно редко, отсутствие жаргонизмов, свойственных работам при использовании в них результатов современных исследований из англоязычного сегмента.

Позволю себе остановиться только на основных результатах достигнутых соискателем. Прежде всего, это касается предложенной архитектуры нейронной сети, ценность которой заключается не только в обеспечении решения поставленной задачи расширения языковой модели, но и возможность применения существующих обученных языковых моделей к новым наборам текстов без экстенсивного их расширения.

Второе, что, по моему мнению, является интересным это – метод обработки обратной связи от эксперта встроенной в нейронную сеть языковой модели. В этом случае притязания автора шире, чем просто решение задачи кластеризации текстов, но попутно происходит изменение самой модели, что в результате позволяет получить иерархический классификатор, который в дальнейшем может быть использован в системе поддержки принятия решений. А также в развитии данной работы можно предложить использование онтологии предметной области, которые могут быть погружены в комплексную методику с участием эксперта. Такое человеко-машинное взаимодействие, в ходе которого меняет представление предметной области, видится интересным и перспективным.

Хотелось бы отметить, что соискатель в методе включения обратной связи от эксперта в процесс обучения, говорит о вмешательстве эксперта в результаты кластеризации. Своей обратной связью эксперт косвенно влияет на скрытое признаковое пространство языковой модели. В данной работе существует потенциал для предоставления больших полномочий эксперту для вмешательства в процесс формирования признакового пространства.

Для демонстрации работоспособности и оценки результативности предложенных методов Павел Владимирович был вынужден выйти на уровень надсистемных показателей, численные оценки которых представлены в работе.

На мой взгляд, содержание и качество результатов полученных в представленной диссертационной работе, конечно же, удовлетворяет критериям, предъявляемым к диссертациям на соискание ученой степени кандидата наук.

Практическая значимость подтверждена убедительными итогами внедрения и использования результатов исследования. Все полученные положения и результаты являются обоснованными и достоверными.

Автореферат полно отражает содержание работы.

Диссертация и автореферат диссертации написаны технически грамотным языком.

Апробация результатов диссертационной работы убедительна, опубликовано 19 работ по теме диссертации, из которых 6 работ опубликованы в изданиях из перечня, рекомендованного ВАК РФ, и 7 работ в изданиях индексируемых в Scopus и Web of Science.

Данная работа представляет собой самостоятельно выполненную завершённую квалификационную работу, посвящённую решению актуальной научной задачи.

Замечания:

1. В подразделе 2.1 обоснованию отдельных элементов предложенной архитектуры нейронной сети для обработки коротких текстов не уделено должного внимания.



2. Помимо реализованных алгоритмов кластеризации интересным представляется использование алгоритмов, обеспечивающих возможность определения кластеров произвольной формы.

3. Для развития предложенного метода обработки экспертной информации в ходе нечеткой интерактивной кластеризации коротких текстов (см. подраздел 2.3) можно порекомендовать расширение состава правил как с точки зрения возможности учета изменения кластерной структуры, так и с точки зрения учета различной степени принадлежности элементов к соответствующим кластерам.

4. Не представлена оценка алгоритмической сложности реализации разработанных методов обработки коротких текстов и параметрической настройки нейронных сетей для нечеткой интерактивной кластеризации коротких текстов.

Тем не менее, указанные замечания не носят принципиального характера и не снижают высокого уровня диссертационной работы и не влияют на положительный отзыв об этой работе.

Содержание работы соответствует пунктам 4 и 13 паспорта специальности 05.13. Системный анализ, управление и обработка информации.

Диссертация Дударина Павла Владимировича является самостоятельно выполненной, завершенной научно-квалификационной работой. Цель диссертации достигнута. Новые научные и практические результаты имеют существенное значение для науки и практики интеллектуального анализа и обработки текстов.

По содержанию и полученным результатам данная диссертация удовлетворяет критериям Положения о присуждении ученых степеней, предъявляемым к диссертациям на соискание ученой степени кандидата наук, а ее автор Дударин Павел Владимирович заслуживает присуждения искомой ученой степени кандидата технических наук по специальности 05.13.01 – «Системный анализ, управление и обработка информации (информационные технологии и промышленность)».

Спасибо.

(Отзыв прилагается).

Председатель

Соискателю предоставляется слово для ответа на замечания оппонента.

Соискатель

С замечаниями оппонента доктора технических наук, профессора Борисова Вадима Владимировича согласен.

Председатель

Слово для отзыва предоставляется официальному оппоненту – **к.ф.м.н. Сухову Сергею Владимировичу**.

Добрый день, уважаемые члены диссертационного совета! Позвольте мне частично зачитать отзыв.

Актуальность работы. Автоматизированная обработка текстов является одним из приоритетных направлений автоматизации процессов обработки информации для решения широкого круга задач. Важным этапом интеллектуального анализа текстовых данных является кластери-

зация текста. Рост массивов информации, состоящих из наборов коротких текстовых фрагментов, вызывает необходимость развития методов обработки коротких текстов с применением машинного обучения. В отличие от текстов с большим количеством слов, короткие тексты, состоящие из нескольких фраз, представляют большие трудности для кластеризации. Методы кластеризации больших текстов применительно к коротким текстам не дают удовлетворительных результатов. Проблемой кластеризации коротких текстов занимается большое число исследователей, однако, большая часть проводимых исследований относится к текстам на английском языке. Для русского языка методы кластеризации разработаны на недостаточном уровне. Недостаточная эффективность имеющихся методов кластеризации коротких русскоязычных текстов затрудняет их использование в российских автоматизированных системах поддержки принятия решений и управления.

Обычно кластеризация текстов допускает несколько возможных вариантов разбиения на группы. Методы, не позволяющие учесть намерения эксперта при кластеризации, могут отказаться не эффективными для конкретной задачи. Интерактивное участие эксперта позволяет корректировать кластеризацию, произведенную автоматической системой. Разрабатываемые в диссертационной работе Павла Владимировича Дударина методы машинного обучения должны заменить или значительно облегчить кропотливый ручной труд экспертов по кластеризации.

На основе вышеизложенного можно сказать, что исследования в области интерактивной кластеризации коротких русскоязычных текстов, выполненные в диссертационной работе Дударина Павла Владимировича, направлены на решение важной и актуальной задачи.

В качестве научной новизны выделяю следующее:

1. Соискателем предложена оригинальная архитектура искусственной нейронной сети, позволяющая решать задачу кластеризации на базе скрытого пространства признаков языковой модели русского языка.
2. С целью расширения языковой модели разработан оригинальный метод обработки текстов, учитывающий семантическую близость слов и основанный на совместной работе нейронной сети и нечеткого иерархического классификатора.
3. Предложен метод обработки обратной связи от эксперта, позволяющий интерактивно включать и исключать объекты из кластера, что позволило автоматически корректировать весовые коэффициенты нейронной сети.

Научная значимость полученных в работе результатов заключается в развитии методов интеллектуальной обработки текстовой информации, в частности, в разработке новых моделей и методов с использованием нейронных сетей и методов нечеткой логики для решения задачи кластеризации коротких русскоязычных текстов. Предложенные методы являются универсальными и позволяют снизить нагрузку на экспертов при кластеризации больших объемов коротких текстов, что делает целесообразным их дальнейшее применение в прикладных исследованиях для проектирования информационно-аналитических систем.

Практическая значимость диссертационной работы заключается в разработке программного модуля системы поддержки принятия решений, позволяющего осуществить интерактивную нечеткую кластеризацию наборов коротких текстов, и применение этого модуля для Системы стратегического планирования Министерства экономического развития Российской Федерации. Как показано в диссертации, интерактивные

методы кластеризации обеспечили сокращение затрат времени эксперта на обработку результатов и позволили повысить точность кластеризации. Для разработанного автором программного модуля получено свидетельство о государственной регистрации программы для ЭВМ.

Структуру и содержание диссертационной работы позвольте мне опустить, так как она звучала ранее.

К несомненным достоинствам работы следует отнести:

1. Гибкое использование передовых методов машинного обучения, таких как перенос знаний и тонкая настройка нейронной сети. Использование методов из различных областей машинного обучения, позволившее получить эффективные гибридные системы на основе интеграции нечеткой логики и нейронных сетей.
2. Большой объем работ по созданию многомодульной системы, интегрированной в систему поддержки принятия решений, позволившей автоматизировать работу экспертов для решения задачи нечеткой интерактивной кластеризации коротких текстов.

В тоже время у меня имеется ряд замечаний по диссертационной работе, а именно:

1. Выводы главы 3 о том, что разработанный программный модуль позволил эффективно организовать совместную работу экспертов по кластеризации коротких текстов является преждевременным, так как эксперименты по эффективности модели проводятся только в следующей главе.
2. В связи с отсутствием объективных критериев правильности кластеризации в главе 4, было бы целесообразно получить оценку эффективности алгоритма от экспертов. Этого в диссертации сделано не было.
3. В формулах текста диссертации и в автореферате не определены некоторые переменные. Например, не определена переменная  $z_j$  в формуле 2.1 на странице 46 диссертации, не определена переменная  $u$  в уравнении 2.6 на странице 47 диссертации, не определена  $\mu_j$  в уравнениях 5,6 в тексте автореферата.
4. Рисунок 2.17 диссертации неинформативен. Подрисуночная подпись не объясняет рисунок. Рисунок 4 автореферата непонятен без пояснений в тексте.
5. В тексте диссертации отсутствуют ссылки на окончательные результаты кластеризации ключевых показателей эффективности системы стратегического планирования, приведенные в приложении 3.

Указанные замечания не снижают научной значимости результатов исследования.

В заключении, диссертационная работа Дударина Павла Владимировича содержит значимые научные результаты по специальности 05.13.01 Системный анализ, управление и обработка информации. Представленная диссертационная работа является законченным исследованием, в котором содержится решение важной задачи по разработке математических методов и алгоритмов кластеризации коротких текстов с интерактивным участием эксперта.

По объему и научному уровню полученных результатов диссертационная работа удовлетворяет требованиям ВАК РФ, предъявляемым к кандидатским диссертациям. Содержание автореферата соответствует основному содержанию диссертационной работы.

Считаю, что диссертационная работа удовлетворяет требованиям пунктов 9-11,13,14 Положения о присуждении учёных степеней, предъ-

являемых к кандидатским диссертациям, а ее автор, 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)».

Спасибо.

(Отзыв прилагается).

Председатель

Слово для ответа на замечания оппонента предоставляется соискателю.

Соискатель

С замечаниями оппонента кандидата физико-математических наук Сухова Сергея Владимировича согласен.

Председатель

Кто хочет выступить?

д.т.н, профессор Ташлинский А.Г.

Несколько непонятным для меня остался вопрос об архитектуре, видимо речь идет не о новой архитектуре нейронной сети, а об оптимизации известной архитектуры сети для решения определенной научной задачи, т.е. речь идет о составляющих архитектуры, например кодере и декодере. Но сама по себе задача группировки слоев, понимания того что и как должно дообучаться: целая сеть или конкретные слои и как это влияет на решение задачи, все это является не менее весомым результатом, расширяющим наше понимание работы нейронной сети.

Второе по поводу численных значений точности кластеризации, ответы на эту группу вопросов меня не до конца удовлетворили. Потому, что точность обычно рассматривается применительно к одному известному набору данных, с известным объемом и составом. К этому набору данных применяются различные методы и далее происходит сравнение результатов. В этом случае результаты таких экспериментов имеют одну и ту же достоверность. Если же набор данных является новым, то он нуждается в определении каких-либо характеристик достоверности, следующих из объема тестирования. Второй момент, по представленным результатам на рисунке 22 линейная аппроксимация дает максимум 75%, хотелось бы более строго научного доказательства связи этого эксперимента с другим экспериментом, в ходе которого была получена цифра 80%.

В целом работа сильная, считаю, что диссертация вполне соответствует искомой степени. У меня в этом плане сомнений не возникает. Спасибо.

д.т.н, доцент Негода В.Н.

Когда мы занимаемся кластеризацией, хотелось бы, чтобы целеориентация продукта кластеризации была каким-то образом учтена.

Когда я задавал вопрос, то ожидал услышать идеальный ответ, что мы помещаем новое постановление по тех. регламенту в контекстную базу и больше ничего не делаем. Система в этом случае сама нам говорит, что вступает в противоречие и что необходимо добавить в классификатор. Второй вариант решения задачи учета нового тех. регламента это изменение весов в классификаторе непосредственно экспертом, т.е. вариант прямого воздействия на модель нейронной сети. Третий вариант, когда эксперт работает с продуктами классификации. Именно это и приходится ему делать, когда имеется естественная классификация, т.к. он хорошо понимает продукты естественной классификации и нового постановления меняющего приоритеты. При этом эксперт знает, какие пункты вступают в противоречие и какие пункты нуждаются в изменении.

В данном случае, получается, что эксперт совершает ту же самую работу, но уже над автоматически полученной классификацией. Хотелось бы увидеть в диссертации моделирование использование продуктов, получаемых с помощью средств автоматизации.

В тоже время, безусловно, сила всех результатов, которые были представлены полностью соответствует всем требованиям. Диссертация очень хорошая, она мне очень понравилась. Буду голосовать "за".

д.т.н, доцент Епифанов В.В.

Коллеги, мне работа понравилась своей логичностью и цельностью. В работе проведена серьезная аналитика существующих подходов и предшествующих работ. В теоретической части предложена архитектура искусственной нейронной сети, предложены методы которые так или иначе апробированы. Результаты работы внедрены в государственной организации. Единственное, в раздаточном материале приведены справки и акты, в которых плохо видно, отражена ли в них эффективность внедрения, хотя бы качественная или в единицах трудоемкости, или в денежном эквиваленте.

Также остается неясно, каким образом должна проводиться последующая эксплуатация системы, в которой функционирует представленный модуль кластеризации. Необходимо программа обучения эксперта для интерактивной работы с предложенным алгоритмом.

Тем не менее, соискатель отвечал на вопросы хорошо, имеется большое количество качественных публикаций, поэтому я буду голосовать "за".

д.т.н, профессор Дьяков И.Ф.

Работа, которая нам была представлена, имеет большое значение, потому что в последнее время все больше расширяется круг задач, в которых используются нейронные сети, для оборонной промышленности, медицины и т.д.

Хотелось бы отметить, что предложенную архитектуру нейронной сети надо было бы сравнить с другими возможными архитектурами, например с сетями Кохонена или Хебба.

Поставленная задача решена полностью, и работа соответствует требованиям, предъявляемым к кандидатским диссертациям, а автор заслуживает присвоения ученого звания кандидата технических наук.

Председатель

Кто еще хочет выступить? Нет желающих? (Нет).

**Соискателю предоставляется заключительное слово.**

Соискатель

Хочу выразить благодарность членам диссертационного совета за оценку работы, свою признательность оппонентам, которые внимательно изучили работу и выразить благодарность ведущей организации за признание работы. Благодарю всех уважаемых специалистов, которые дали свои замечания и отзывы на автореферат.

Огромная благодарность научному руководителю д.т.н., профессору Ярушкиной Надежде Глебовне за возможность заниматься интересным исследованием, за поддержку и терпение. А также сотрудникам кафедры «Информационные системы» за своевременную помощь в исследовании и обстоятельную критику.

Планирую дальнейшую работу в этом направлении. Спасибо.

Председатель

Переходим к голосованию. Какие будут предложения по составу счетной комиссии? Поступили предложения включить в состав счетной комиссии профессора Браже Р.А., профессора Васильева К.К. и профессора Дьякова И. Ф.

Прошу голосовать. Возражений нет.

Председатель

Прошу счетную комиссию приступить к работе.

(Счетная комиссия организует тайное голосование.)

Председатель

Коллеги! Продолжаем нашу работу. Слово предоставляется председателю счетной комиссии Васильеву К.К.

Оглашается протокол счетной комиссии.  
(Протокол счетной комиссии прилагается).

Кто против? (Нет).

Кто воздержался? (Нет).

Протокол счетной комиссии утверждается.

Таким образом, на основании результатов тайного голосования (за - 16 , против - 0 , недействительных бюллетеней - 1) диссертационный совет Д212.277.04 при Ульяновском государственном техническом университете признает, что диссертация **Дударина П.В.** содержит новые решения по исследованию и разработке моделей и методов нечеткой кластеризации коротких текстов, соответствует требованиям, предъявляемым к кандидатским диссертациям (**п.9** "Положения"

ВАК), и присуждает **Дударину Павлу Владимировичу** ученую степень кандидата технических наук по специальности **05.13.01**.

Председатель

У членов Совета имеется проект заключения по диссертации **Дударина П.В.** Есть предложение принять его за основу. Нет возражений? (Нет). Принимается.

Какие будут замечания, дополнения к проекту заключения?

**(Обсуждение проекта) .**

Председатель

Есть предложение принять заключение в целом с учетом редакционных замечаний. Нет возражений? Принимается единогласно.

**Заключение объявляется соискателю.**

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 \_\_\_\_\_ № \_\_\_\_\_

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 исследовательский 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  $\mu_i$  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. Ушакова», 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  $\varepsilon$ -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.

Защита окончена. Есть ли замечания по процедуре защиты? (Нет).

Поздравляет соискателя с успешной защитой. Благодарит членов совета и всех участников за внимание.

**Заседание объявляется закрытым.**

Зам. Председателя Совета Д214.2.277.04,  
д.т.н., доцент

 С.К. Киселев

Ученый секретарь Совета Д214.2.277.04,  
д.т.н., доцент

 А.М. Наместников



14.10.2021