

ОТЗЫВ

официального оппонента на диссертационную работу ДУДАРИНА Павла Владимировича, выполненную на тему «Исследование и разработка моделей и методов нечеткой кластеризации коротких текстов» и представленную на соискание ученой степени кандидата технических наук по специальности 05.13.01 – «Системный анализ, управление и обработка информации (информационные технологии и промышленность)»

Актуальность темы диссертационного исследования

В настоящее время активно развиваются интеллектуальные методы, модели и технологии интеллектуального анализа и обработки текстовых сообщений из разнородных источников в условиях неопределенности.

Эти тексты характеризуются рядом особенностей: относительно небольшой размер затрудняет их статистический анализ; отсутствие структуризации (специальной разметки и полей) существенно усложняет процедуры анализа таких текстов; наличие грамматических и синтаксических неточностей требует дополнительной обработки; нестационарность тезауруса (состава и важности слов) приводит к необходимости использования процедур предварительной кластеризации этих текстов.

Поэтому тематика данной диссертации, посвященная исследованию и разработке интеллектуальных методов, моделей и алгоритмов кластеризации коротких текстовых сообщений в условиях неопределенности, является несомненно актуальной.

Структура и характеристика диссертационной работы

Диссертационная работа включает в себя введение, 4 главы, заключение, список литературы и приложения.

Во введении охарактеризована актуальность темы диссертационной работы, определены объект, предмет и цель работы, поставлены исследовательские задачи, сформулированы теоретическая значимость, научная новизна и практическая ценность полученных результатов, перечислены выносимые на защиту положения.

В первой главе описаны и проанализированы методы и подходы и кластеризации коротких текстовых сообщений с учетом неопределенности. Обоснован выбор используемой языковой модели.

Во второй главе П.В. Дудариним предложен метод интерактивной кластеризации коротких текстов, учитывающий возможности языковых моделей с высокими показателями точности и перплексии.

Третья глава посвящена разработке алгоритмического обеспечения и программных средств нечеткой интерактивной кластеризации коротких текстов, реализующих предложенные методы и модели.

В четвертой главе проведены экспериментальные исследования и выполнена оценка эффективности интерактивной кластеризации на основе использования предложенного научно-методического, алгоритмического и программного обеспечения.

В заключении отражены основные результаты диссертации.

В приложениях представлены результаты внедрения и использования.

Теоретическая значимость, научная новизна и практическая ценность результатов диссертации

Теоретическая значимость диссертационной работы заключается в развитии интеллектуальных методов и моделей кластеризации текстов в условиях неопределенности.

Научная новизна результатов диссертации заключается в следующем.

1. Предложена разновидность искусственных нейронных сетей, ориентированных на решение задач кластеризации коротких текстов с учетом неявных признаков языковой модели.

2. Разработан метод обработки коротких текстов для расширения словаря языковой модели на основе использования нейронной сети и нечеткого иерархического классификатора, учитывающий семантическую близость слов.

3. Предложен метод настройки нейронных сетей для нечеткой интерактивной кластеризации коротких текстов.

4. Разработано алгоритмическое обеспечение, реализующее предложенные методы нечеткой интерактивной кластеризации наборов коротких текстов.

Практическая ценность результатов диссертационной работы состоит в разработанных программных средствах для интерактивной нечеткой кластеризации текстов, используемых в составе интеллектуальной системы поддержки принятия решений в интересах Министерства экономического развития РФ для Системы стратегического планирования РФ. Программно

реализованное научно-методическое обеспечение также внедрено в системе Планета.Аналитика 4.0 ООО «ИБС “Экспертиза».

***Обоснованность научных положений, выводов и рекомендаций,
их достоверность***

Все полученные в ходе диссертационного исследования положения, выводы и рекомендации являются *обоснованными*, базируются на методах системного анализа, теории искусственных нейронных сетей, нечеткой кластеризации, проектирования и разработки программных средств. *Достоверность* результатов диссертации обеспечивается корректным использованием методов исследований, подтверждена результатами экспериментов и сравнительной оценки с существующими методами интеллектуального анализа текстов, итогами внедрения и использования, апробацией положений диссертации на международных и всероссийских конференциях.

***Соответствие содержания автореферата положениям диссертации,
оценка языка и стиля изложения диссертации и автореферата***

Автореферат полно и точно отражает содержание диссертации. Диссертация и автореферат написаны технически грамотным языком.

Апробация работы, публикации по теме диссертации

Основные результаты работы докладывались и получили одобрение на 12 международных и национальных конференциях. Содержание диссертации полно отражено в 19 научных работ, в том числе в 6 статьях в изданиях из перечня ВАК и в 7 статьях в изданиях, индексируемых в базе данных Web of Science и Scopus.

Мнение официального оппонента по диссертации, замечания

Данная диссертация представляет собой самостоятельно выполненную завершённую научно-квалификационную работу, посвящённую решению актуальной научной задачи. Диссертация содержит новые научные и практические результаты.

По работе следует отметить следующие замечания и рекомендации.

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 нечёткой 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 нечёткой интерактивной кластеризации коротких текстов.

Указанные замечания не носят принципиального характера, не снижают высокого уровня диссертационной работы и не влияют на положительное мнение официального оппонента о ней в целом.

Формулировка научной задачи, решенной в диссертации и вывод о ее соответствии Положению о присуждении ученых степеней

Содержание диссертационной работы соответствует формуле и пунктам области исследования Паспорта специальности 05.13.01 – «Системный анализ, управление и обработка информации (информационные технологии и промышленность)», прежде всего, пункту 4 «Разработка методов и алгоритмов решения задач системного анализа, оптимизации, управления, принятия решений и обработки информации» и пункту 13 «Методы получения, анализа и обработки экспертной информации».

Диссертация Дударина Павла Владимировича на тему «Исследование и разработка моделей и методов нечеткой кластеризации коротких текстов» является самостоятельно выполненной, завершенной научно-квалификационной работой. Цель диссертации достигнута. Новые научные и практические результаты, полученные П.В. Дудариним, 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)».

Официальный оппонент

Профессор кафедры вычислительной техники
филиала федерального государственного бюджетного
образовательного учреждения высшего образования
«Национальный исследовательский университет «МЭИ»
в г. Смоленске,
доктор технических наук, профессор
(специальность 05.13.15)

В.В. Борисов

«30» августа 2021 г.

Подпись Борисова Вадима Владимирович заверяю



Ученый секретарь Ученого совета
филиала ФГБОУ ВО «НИУ «МЭИ»
в г. Смоленске, к.э.н.
Е.А. Кириллова

Филиал ФГБОУ ВО
«Национальный исследовательский университет «МЭИ»
в г. Смоленске
Адрес: 214013, г. Смоленск, Энергетический проезд, д. 1.
Тел. 8 (4812) 65-14-61
Web-сайт: <https://sbmpei.ru/>
E-mail: vbtor67@mail.ru