

ОТЗЫВ

официального оппонента, доктора технических наук, профессора Кравец Аллы Григорьевны на диссертационную работу Мансур Али Махмуд «Модель, метод и алгоритмы Data Mining для интеллектуальной обработки и анализа текстов на естественном языке», представленную на соискание ученой степени кандидата технических наук по специальности 1.2.1. Искусственный интеллект и машинное обучение

1. Актуальность темы диссертационного исследования

В настоящее время обработка текстовых данных с использованием методов искусственного интеллекта (ИИ) является крайне востребованной задачей. Рост объемов неструктурированной текстовой информации требует разработки эффективных алгоритмов для ее анализа и интерпретации. Предложенные автором модель, метод и алгоритмы для решения задачи создания низкоразмерных векторных представлений текста с интерпретируемыми признаками находятся в русле современных и перспективных направлений исследований. Использование низкоразмерных векторных представлений с интерпретируемыми признаками может существенно улучшить эффективность работы моделей машинного обучения, обеспечивая при этом понятность результатов анализа. Кроме того, предложенный метод векторизации текстов, позволяет сохранять ключевые семантические характеристики текста даже при снижении размерности представления. Это открывает новые возможности для анализа больших текстовых массивов, включая задачи классификации, кластеризации и извлечения знаний.

2. Оценка достоверности полученных результатов и новизны диссертационного исследования

Теоретические исследования диссертации основаны на использовании элементов теорий искусственного интеллекта, методов интеллектуального анализа данных, методов обработки и анализа текстов на естественном языке, а также методов объектно-ориентированного программирования.

Достоверность результатов научной работы подтверждается грамотной постановкой задач, полнотой и точностью исходных данных, непротиворечивостью и согласованностью с научными результатами известных ученых, экспериментальной проверкой теоретических результатов, апробацией материалов на научных конференциях и публикациями. Помимо этого, обоснованием достоверности научных результатов, полученных соискателем, является их внедрение, подтвержденное соответствующими актами.

Научная новизна исследования заключается в разработке модели, метода и алгоритмов обработки и анализа текстов на естественном языке, направленных на решение актуальной научной задачи разработки моделей, методов и алгоритмов Data Mining для обработки и анализа текстов на естественном языке в интеллектуальных системах. Предложенные решения позволяют снизить частоту ошибок классификации и кластеризации текстов, что имеет важное значение для развития искусственного интеллекта и машинного обучения. Основные научные результаты включают в себя:

1) Математическую модель векторизации текстов на основе концептов, которая отличается новыми правилами построения эталонных концептов и функциями определения их весов. Это позволяет снизить размерность векторного пространства и улучшить дискриминационную способность векторов признаков.

2) Модифицированный метод генерации векторных представлений документов, использующий интерпретируемые признаки при векторизации. Его применение приводит к снижению ошибок классификации и кластеризации текстовых документов.

3) Алгоритм извлечения и фильтрации ключевых фраз, основанный на частоте их появления и использующий функцию парсера для разметки частей речи. Это обеспечивает извлечение ключевых фраз с правильной грамматической структурой.

4) Алгоритм построения концептов из семантически близких фраз, который решает задачу кластеризации с учетом контекстуальной

семантической близости, что повышает однородность кластеров, представляющих концепты.

Все разработанные решения взаимосвязаны и образуют единый методический комплекс для интеллектуальной обработки текстовых данных.

Теоретическая значимость диссертационного исследования заключается в разработке инновационных модели, метода и алгоритмов обработки естественно-языковых текстов, которые обеспечивают: (1) эффективную обработку и анализ текстов на естественном языке; (2) построение низкоразмерных векторных представлений с сохранением интерпретируемости; (3) точную классификацию и кластеризацию текстовых данных. Эти научные результаты существенно развивают методологический аппарат в области искусственного интеллекта и машинного обучения, предлагая новые решения для задач обработки естественного языка и расширяя возможности работы с семантическими текстовыми представлениями в современных информационных системах.

Практическая ценность работы заключается в создании программного приложения, позволяющего использовать разработанную модель, метод и алгоритмы обработки и анализа текстов на естественном языке в системах искусственного интеллекта для снижения частоты появления ошибок в алгоритмах классификации и кластеризации с учётом условий снижения размерности векторного представления текстов и сохранения его интерпретируемости. Подтверждением практической значимости работы является использование результатов, проведенных соискателем исследований, в ИТ-компании ООО «Ит-Эффект» (г. Москва), а также в учебном процессе ФГАОУ ВО «Южный федеральный университет» (ЮФУ), о чем свидетельствуют представленные акты о внедрении.

3. Оценка содержания диссертации, степени ее завершенности, подтверждение публикаций автора

Диссертационное исследование структурно организовано в соответствии с действующими требованиями ГОСТ 7.0.11-2011 и содержит в себе: введение;

четыре раздела с выводами; общее заключение; список литературы из 146 источников и два приложения. Приложения представлены на 7 страницах.

Диссертация Мансур А.М. является завершенным научным исследованием, изложена последовательно и грамотно, правильно оформлена. Характерной особенностью работы является решение весомой научной задачи, имеющей важное значение для развития технологий искусственного интеллекта и машинного обучения. Разработанные модель, метод и алгоритмы отличаются новизной и доведены до практической реализации.

Проведенный анализ содержания работы свидетельствует о полном решении сформулированных исследовательских задач и достижении основной цели диссертационного проекта. Научная ценность исследования проявляется в принципиальной новизне полученных результатов, их теоретической важности, а также в перспективах практического применения. Полученные результаты прошли апробацию на международных научно-технических конференциях.

Научные положения и выводы исследования нашли отражение в 17 публикациях автора, включая: 3 статьи – в издании, рекомендованном ВАК РФ, в т.ч. 1 – без соавторов; 2 статьи – в изданиях, индексируемых в базах данных Scopus и Web of Science; 2 свидетельства о государственной регистрации программ для ЭВМ.

4. Соответствие содержания автореферата диссертации

Содержание автореферата диссертационной работы Мансур Али Махмуд соответствует всем требованиям, предъявляемым ВАК при Минобрнауки РФ, и полностью отражает содержание диссертации, её основные положения и выводы.

5. Замечания по диссертационной работе

1. Автор провел экспериментальные исследования и сравнил предложенный метод генерации векторных представлений с существующими аналогами, однако не рассмотрел сравнение интерпретируемости и вычислительной эффективности своих результатов с рядом современных

генеративных моделей (например, BERT и GPT).

2. Хотя автором декларируется, что интерпретируемость улучшает производительность алгоритмов, отсутствует количественная оценка этого влияния.

3. В работе не приведены данные о времени обучения и обработки больших текстов.

4. Алгоритм построения концептов использует сферический алгоритм K-средних, но недостаточно исследованы альтернативные варианты (например, иерархическая кластеризация).

Редакционные замечания:

- а. В работе имеются орфографические ошибки и опечатки, например: на стр. 8 «Разработан программное» вместо «разработано»; стр. 110 «Униграммы» вместо «Униграммы».
- б. В первом абзаце на стр.72 перепутаны английские ключевые слова и их переводы.

6. Заключение

Отмеченные замечания не снижают общей положительной оценки работы и направлены на дальнейшее совершенствование исследования. Диссертационное исследование Мансур А.М. является завершённой научно-квалификационной работой, направленной на решение важной задачи в области искусственного интеллекта и машинного обучения.


Проведённое исследование отличается выраженной новизной и научной самостоятельностью, а его результаты обладают значительным теоретико-прикладным потенциалом. Основные её положения представлены в рецензируемых научных изданиях из перечня ВАК и международных базах данных, а также апробированы на различных научных конференциях.

Тема диссертации, цель и задачи работы, научная новизна, методы и алгоритмы, используемые в работе, подтверждают её соответствие пунктам 4 и 5 паспорта специальности 1.2.1. Искусственный интеллект и машинное обучение (технические науки).

Оценивая диссертационную работу в целом, считаю, что она полностью соответствует требованиям пп. 9-14 Положения о присуждении ученых степеней, утверждённого Постановлением Правительства РФ от 24 сентября 2013 г. № 842 «О порядке присуждения ученых степеней» (с изменениями и дополнениями в редакции от 16 октября 2024 года), предъявляемым ВАК РФ к диссертациям на соискание ученой степени кандидата технических наук, а ее автор, Мансур Али Махмуд, заслуживает присуждения ученой степени кандидата технических наук по специальности 1.2.1. Искусственный интеллект и машинное обучение.

Официальный оппонент, доктор технических наук, профессор
(н.с. 05.13.10 Управление в социальных и экономических системах),
профессор кафедры «Системы автоматизированного проектирования
и поискового конструирования» федерального государственного
бюджетного образовательного учреждения высшего образования
«Волгоградский государственный технический университет»
(ВолгГТУ), 400005, г. Волгоград, проспект им. В. И. Ленина, д. 28,
Тел.: +7 (8442) 24-81-00, e-mail: AllaGKravets@yandex.ru,
https://www.vstu.ru/university/personalii/kravets_allya_grigorevna/

«30» мая 2025 г.

 Алла Григорьевна Кравец

