

## ОТЗЫВ

на автореферат диссертации "Модель, метод и алгоритмы Data Mining для интеллектуальной обработки и анализа текстов на естественном языке", представленной Мансур Али Махмуд на соискание ученой степени кандидата технических наук по специальности 1.2.1. Искусственный интеллект и машинное обучение

Работа посвящена актуальной научной задаче, связанной с разработкой модели, алгоритмов и метода обработки и анализа текстов на естественном языке в условиях экспоненциального роста объемов текстовых данных. Решение данной задачи позволяет существенно повысить эффективность инструментов и систем искусственного интеллекта и машинного обучения. Ключевым элементом алгоритмов классификации и кластеризации текстовых документов является векторное представление, которое преобразует текстовую информацию в числовые форматы, пригодные для машинной обработки.

Важнейшей особенностью данной работы, определяющей ее научную новизну, является разработка модели, метода и алгоритмов, которые в совокупности позволяют строить более информативные векторные представления с интерпретируемыми характеристиками. Построенная математическая модель векторизации текстов на основе концептов, характеризуется использованием новых правил построения эталонных концептов и новых функций определения их весовых коэффициентов. Это позволяет сократить размерность векторного пространства и повысить дискриминационную способность результирующих векторов признаков. Новые весовые функции определяют значимость концептов для документов путем сопоставления ключевых фраз документа с набором извлеченных эталонных концептов с использованием меры семантического сходства. Для построения эталонных концептов были разработаны новые алгоритмы извлечения ключевых фраз и формирования семантически однородных концептов.

Модель и алгоритмы образуют модифицированный метод генерации низкоразмерных и интерпретируемых векторных представлений текстовых документов, где каждая размерность соответствует определенному концепту, а не отдельному слову. Проведенные вычислительные эксперименты подтвердили достоверность и эффективность полученных результатов. Предложенное решение способствовало снижению частоты ошибок в алгоритмах классификации и кластеризации текстовых документов.

По автореферату имеются следующие замечания:

- 1) из текста автореферата неясно, как обрабатываются пересекающиеся концепты;
- 2) не в полной мере описан процесс определения порога сходства  $\Theta$  и количества эталонных концептов, которые использовались для построения векторов.

На основе автореферата в целом можно констатировать, что диссертационная работа соответствует уровню кандидатской диссертации и обладает необходимой теоретической и практической значимостью. Работа соответствует специальности 1.2.1. Искусственный интеллект и машинное обучение, а ее автор Мансур А. М. достоин присуждения ему ученой степени кандидата технических наук по данной специальности.

Еремеев Александр Павлович

профессор кафедры прикладной математики и искусственного интеллекта

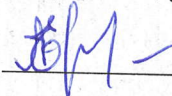
НИУ «Московский энергетический институт» (НИУ «МЭИ»),

доктор технических наук (научная специальность 05.13.11 – Математическое и программное обеспечение вычислительных машин, комплексов, систем и сетей), профессор.

111250, г. Москва, вн. тер. г. муниципальный округ Лефортово, ул. Красноказарменная, 14, стр. 1.

Тел.: +7 (495) 362-79-62, e-mail: [eremeev@appmat.ru](mailto:eremeev@appmat.ru)

20.05.2025 г.



Еремеев Александр Павлович



ЗАМЕСТИТЕЛЬ ДИРЕКТОРА  
УПРАВЛЕНИЯ НА РАБОТЕ С ПЕРЕКОНАЛАМ  
Л.И. ПОЛЕВАЯ

Подпись Еремеева Александра Павловича заверяю