

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
Факультет прикладной математики и информатики
Кафедра вычислительной математики

Аннотация к дипломной работе

**Вычислительные алгоритмы
положительной матричной факторизации**

Крачковский Даниил Янович

Научный руководитель - кандидат физико-математических наук, доцент
Фалейчик Б. В.

Минск, 2019

РЕФЕРАТ

Дипломная работа, 52 страниц, 2 таблиц, 4 рисунков, 14 источника.

Ключевые слова: НЕОТРИЦАТЕЛЬНАЯ МАТРИЧНАЯ ФАКТОРИЗАЦИЯ (НМФ), АППРОКСИМАЦИЯ НИЗКОГО РАНГА, ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ, ЧИСЛЕННЫЕ МЕТОДЫ, ЛИНЕЙНАЯ АЛГЕБРА, РАЗРЕЖЕННЫЕ МАТРИЦЫ, ТЕРМ-ДОКУМЕНТНАЯ МАТРИЦА, МЕТОД НАИМЕНЬШИХ КВАДРАТОВ.

Объект исследования - вычислительные алгоритмы неотрицательной матричной факторизации.

Цель работы - исследование эффективности алгоритмов неотрицательной матричной факторизации и их применение к задаче интеллектуального анализа текста.

Основным методом проведения работы является вычислительный эксперимент. В ходе работы было программно реализовано 4 метода решения задачи НМФ (метод мультипликативного обновления и 3 различных модификации метода попеременных наименьших квадратов), проведено их экспериментальное сравнение. Реализованные алгоритмы были применены к задаче выделения ключевых слов и предложений в текстах. Результаты исследования показали что наиболее эффективным алгоритмом оказался метод попеременных наименьших квадратов на основе нормальных уравнений. Метод выделения ключевых слов и предложений на основе НМФ был применён к тексту первой главы данной работы. Полученные результаты хорошо резюмируют данный текст.

Областью применения являются рекомендательные системы, задачи интеллектуального анализа текстов, поисковые системы, компьютерное зрение, биоинформатика, спектральный анализ данных и другие.

ABSTRACT

Diploma work, 52 pages, 2 tables, 4 pictures, 14 references.

Key words: NONNEGATIVE MATRIX FACTORIZATION (NMF), LOW-RANK APPROXIMATION, DATA MINING, NUMERICAL METHODS, LINEAR ALGEBRA, SPARSE MATRIX, TERM-DOCUMENT MATRIX, LEAST SQUARES METHOD.

The object of study is numerical methods of nonnegative matrix factorization.

The purpose of the work is to study the efficiency of nonnegative matrix factorization algorithms and to apply them to a data mining problem.

The main method of study is the computational experiment. During the study 4 methods of NMF problem solution were implemented (multiplicative update rule and 3 different modifications of alternating least squares method). The comparative numerical experiment was performed.

The implemented algorithms were applied to automatic key words and key sentences extraction problem. The results showed that the most efficient algorithm is alternating least squares based on normal equations. Key words and sentences extraction algorithm based on NMF was applied to the text of the first chapter of this work. The results summarise the text well.

The applications areas are recommendation systems, text mining problems, search engines, computer vision, bioinformatics, spectral data analysis and others.