

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ
Кафедра дифференциальных уравнений и системного анализа

Аннотация к дипломной работе

**МЕТОДЫ ФАКТОРИЗАЦИИ И ЭМБЕДДИНГОВ ДЛЯ АНАЛИЗА
ДАННЫХ ВЫСОКОЙ РАЗМЕРНОСТИ**

Мамуль Даниил Иосифович

Научный руководитель:
ст. преподаватель
Н.А. Азаров

В дипломной работе 48 страниц, 15 рисунков, 8 источников.

РЕКОМЕНДАТЕЛЬНАЯ СИСТЕМА, ФАКТОРИЗАЦИЯ, ЭМБЕДДИНГИ, ГРАФОВАЯ МОДЕЛЬ, SWING, BERT, ВЫСОКОРАЗМЕРНЫЕ ДАННЫЕ, ХОЛОДНЫЙ СТАРТ, ПАЙПЛАЙН

В дипломной работе рассматривается задача построения рекомендательной системы на основе пользовательских взаимодействий с товарами в условиях высокой размерности и разреженности данных.

Целью дипломной работы является разработка и сравнение алгоритмов рекомендаций, использующих факторизационные, эмбединговые и графовые подходы, а также построение гибридной модели, пригодной для применения в электронной коммерции.

Для достижения поставленной цели применялись методы: матричной факторизации (SVD, EASE), генерации эмбедингов товаров на основе трансформеров (BERT), графового анализа взаимодействий (SWING), а также вычисление метрик качества рекомендаций: MAP, MAR, NDCG, BNDCG.

В дипломной работе получены следующие результаты:

1. Проведён обзор и реализовано сравнение четырёх подходов: SVD, EASE, BERT, SWING;
2. Построена гибридная архитектура, сочетающая поведенческие и семантические сигналы;
3. Выявлены сильные и слабые стороны моделей, предложены сценарии их практического применения;
4. Применена схема рекомендаций по типу товара с фильтрацией через эмбединги и уточнением внутри подкаталога на основе SWING.

Новизна работы заключается в интеграции эмбединговой фильтрации по типам с последующей локальной графовой рекомендацией, что позволяет формировать сопутствующие товары даже при отсутствии истории.

Дипломная работа носит прикладной характер. Её результаты могут быть использованы в системах персональных и витринных рекомендаций, в том числе в блоках «С этим товаром покупают». Разработанный пайплайн показал эффективность при внедрении в условиях промышленной среды.

Все результаты дипломной работы обоснованы экспериментально, подтверждены расчётами и сопоставлены с известными подходами. Обоснованность и достоверность выводов подтверждаются сопоставимыми значениями метрик и визуальным анализом выдачи.

Дипломная работа выполнена автором самостоятельно.

Thesis project is presented in the form of an explanatory note of 48 pages, 15 figures, 8 references.

RECOMMENDER SYSTEM, FACTORIZATION, EMBEDDINGS, GRAPH MODEL, SWING, BERT, HIGH DIMENSIONAL DATA, COLD START, PIPELINES

The object of research in this thesis project is the task of building a recommender system based on user-item interaction data under conditions of high dimensionality and data sparsity

The purpose of the thesis is to develop, implement, and compare several recommendation algorithms including factorization-based, embedding-based, and graph-based approaches and to construct a hybrid pipeline suitable for use in real-world e-commerce scenarios.

To achieve this goal, the following methods and tools were applied: matrix factorization (SVD, EASE), transformer-based item embeddings (BERT), graph analysis using co-occurrence models (SWING), and quality evaluation using metrics such as MAP, MAR, NDCG, and BNDCG

The main results of the thesis project are as follows:

A comparative analysis of four recommendation models (SVD, EASE, BERT, SWING) was conducted;

A hybrid architecture combining behavioral and semantic signals was implemented and tested;

Strengths and weaknesses of each model were identified, and their applicability was analyzed;

A type-level recommendation pipeline was proposed and realized, using BERT for semantic filtering and SWING for refined graph-based recommendations.

The novelty of this work lies in combining type-level semantic filtering with local graph-based recommendation logic, which allows generating relevant companion items even in the absence of prior user interactions.

This thesis project is of an applied nature. Its results can be used in recommender systems for personalized or cross-sell scenarios, including "frequently bought together" blocks on e-commerce platforms. The developed pipeline has demonstrated practical effectiveness under production-like conditions.

All results in the thesis are supported by experimental evaluations, validated using standard ranking metrics, and confirmed by visual analyses. The validity and reliability of the conclusions are ensured by consistent outcomes across multiple methods and alignment with known practices.

The thesis project is complete, all tasks have been successfully fulfilled, and the work opens avenues for further development in areas such as generative modeling and adaptive ranking strategies.

The thesis project was completed independently by the author.