

Правительство Российской Федерации
Федеральное государственное автономное образовательное
учреждение высшего профессионального образования
Национальный исследовательский университет
«Высшая школа экономики»

Факультет компьютерных наук
Образовательная программа «Прикладная математика и информатика»
Департамент анализа данных и искусственного интеллекта

КУРСОВАЯ РАБОТА

на тему

**Исследование возможности встраивания контекстной
информации в алгоритмы коллаборативной фильтрации на
основе матричных разложений**

Выполнил студент группы 203
Стеценко Макар Александрович

Научный руководитель:
Доцент, кандидат технических наук,
Игнатов Дмитрий Игоревич

Москва 2015

Оглавление

Введение	5
Литература	7

Аннотация

"Исследование возможности встраивания контекстной информации в алгоритмы коллаборативной фильтрации на основе матричных разложений"

Стеценко Макар Александрович

Факультет компьютерных наук. Второй курс. 203 группа. 2015 год.

Игнатов Дмитрий Игоревич

Факультет компьютерных наук. Департамент анализа данных и искусственного интеллекта.

Научный руководитель. dignatov@hse.ru

Наличие контекстной информации является одним из важнейших факторов для построения личных рекомендаций. Однако, классические алгоритмы коллаборативной фильтрации, основанные на матричных разложениях, таких как SVD разложение, используют только информацию о пользователях и предметах и не предоставляют явных методов включения дополнительных факторов. В данной работе будет показан один из методов встраивания контекстной информации в алгоритм, использующий SVD разложение. Для тестирования рассматриваемого метода будет использоваться открытый банк данных [MovieLens](#). База данных содержит пользователей портала MovieLens, каждый из которых оценил не менее 20 фильмов, а так же информацию о каждом фильме.

Abstract

"Integrating Contextual Information into Collaborative Filtering Algorithms based on Matrix Decomposition"

Stetsenko Makar

Faculty of Computer Science. 2nd course. 203 group. 2015 year.

Ignatov Dmitry

Faculty of Computer Science. School of Data Analysis and Artificial Intelligence.

Scientific adviser. dignatov@hse.ru

Context has always been an important factor in personalized Recommender systems. However, standard collaborative filtering algorithms based on matrix factorization rely mainly on user and subject information and don't provide any methods for encapsulating extra data. This work demonstrates such a method based on SVD decomposition. To test results an open data base taken from [MovieLens](#) is used. The database provides information about users and movies.

Введение

Рекомендательные системы стали неотъемлемой частью жизни людей и бизнеса, особенно после популяризации BigData и открытых данных. Задача таких систем угадывать предпочтения пользователей основываясь на их предыдущих действиях. Более формально, задача о рекомендациях состоит в угадывании значения пары (пользователь, предмет). В простейшем варианте, ответом на данный вопрос будет либо 1 - предмет интересен пользователю, либо 0 - предмет не стоит рекомендовать.

Коллаборативная фильтрация один из способов построения рекомендаций. Данный метод использует информацию о других пользователях системы, чтобы понять, какой предмет скорее всего понравится рассматриваемому пользователю. Несмотря на то, что упрощенная модель (пользователь, предмет) подходит для моделирования многих ситуаций, очень часто приходится сталкиваться с дополнительными параметрами, которые играют важную роль при решении задачи о рекомендациях. Таким параметром может быть время, тогда уже имеется 3-х мерный вектор (пользователь, предмет, время). Множество переменных, которые влияют на отношение пользователя к предмету и следовательно на рекомендации для этого пользователя, называется *контекстом*.

Целью данной работы будет решение задачи о рекомендациях с использованием контекстной информации.

План работы:

Таблица 1.1: План работы

Дата	Цель
31.01.2015	Ознакомиться с существующими работами по рекомендательным системам [1–4]
15.02.2015	Ознакомиться с языком программирования Java [5]
22.03.2015	Написать первый вариант программы
05.04.2015	Посчитать быстродействие и точность на различных наборах данных
12.04.2015	Исправление ошибок в работе программы, улучшение
19.04.2015	Закончить работу над отчетом

Литература

1. Ignatov Dmitry I, Nenova Elena, Konstantinova Natalia [и др.]. Boolean Matrix Factorisation for Collaborative Filtering : An FCA-Based Approach.
2. Ignatov Dmitry I, Mikhailova Maria, Zakirova Alexandra Yu [и др.]. Recommendation of Ideas and Antagonists for Crowdsourcing Platform Witology.
3. Xavier Amatriain, Alexandros Karatzoglou, Nuria Oliver Linas Baltrunas. Multiverse Recommendation: N-dimensional Tensor Factorization for Context-aware Collaborative Filtering // ACM Recommender Systems. 2010. URL: <http://xavier.amatriain.net/pubs/karatzoglu-recsys-2010.pdf>.
4. Biclustering neighborhood-based collaborative filtering method for top-n recommender systems / Faris Alqadah, Chandan K. Reddy, Junling Hu [и др.] // Knowledge and Information Systems. 2014. URL: <http://link.springer.com/10.1007/s10115-014-0771-x>.
5. Eckel Bruce. Thinking in Java. 4 изд. Pearson Education, 2006. с. 1079.