Multi-purpose Library of Recommender System Algorithms for the Item Prediction Task

Julius Kolbe

11.6.2013

Contents

1 Abstract

2 Introduction

2.1 Motivation

The library together with this document shall provide a "cookbook" for recommender systems. With the simple syntax and the interactivity of Python it is aimed at beginners to simply experiment with different algorithms. Especially the interactivity is missing in the already existing libraries because none of them is written in Python.

2.2 Task (what a Recommender System does)

A Recommender System works in a scenario with users, items and interactions users can have with items. Such a scenario could be an online shop, where the interactions are purchases of items by users or a video platform, where the users interact with items (videos) by watching them. Based on the past interactions of the users a Recommender System searches for items a user haven't interacted with yet but the probability that he will interact is maximized.

In contrast to this implicit approach you can also use ratings the user gave items explicitly, but in this work the focus lies on implicit feedback. However ratings can be interpreted as the strength of implicit feedback for example how often a user purchased an item. Some algorithms implemented in this library can use this information but none will explicitly predict ratings.

2.3 Objective and Motivation

3 3 Related Work

There is a wide range of projects providing implemantions for recommender system. Some of them are described in this chapter to give a quick overview and comparison.

3.1 MyMediaLite

MyMediaLite[?] is developed at the University of Hildesheim and provides several algorithm for rating prediction and

3.2 PREA

3.3 Mahout

3.4 Duine

3.5 Cofi

3.6 Lenskit

4 Recommendation Algorithms

4.1 Primitive Algorithms

4.2 k-Nearest-Neighbor

4.3 Matrix Factorization

4.3.1 BPRMF

4.3.2 RankMFX

4.3.3 Ranking SVD (Sparse SVD)

4.4 Evaluation Methods

4.4.1 Leave-one-out Protocol

4.5 Evaluation metrics

4.5.1 Hitrate/Recall

4.5.2 Precision

4.5.3 F1

4.5.4 Mean Reciprocal Hitrate

4.5.5 Area under the ROC

5 Datasets for testing

5.1 MovieLens

5.2 Million Song Dataset

5.3 SNAP

6 Experiments

6.1 Execution

6.2 Results

6.3 Comparison

7 Design and Implementation

7.1 General structure

7.2 Interfaces

8 User Manual

8.1 Primitive Algorithms

8.2 k-Nearest Neighbor

8.3 BPRMF

8.4 RankMFX

8.5 Ranking SVD (Sparse SVD)

9 Conclusions

9.1 Future work

9.2 Outlook

10 References

Bibliography

[1] http://www.mymedialite.net/

[2] http://mloss.org/software/view/420/

[3] http://mahout.apache.org/

[4] http://www.duineframework.org/

[5] https://sites.google.com/a/cofirank.org/index/

[6] http://lenskit.grouplens.org/