

2019-05-22

Exercise Sheet 7

Exercise 1 (Collaborative Filtering Basics)

[6 Points]

- a) 0 Points Firstly, make yourself familiar with the widely used **MovieLens dataset** provided by the GroupLens research group.
- b) 2 Points To get started with collaborative filtering recommender systems, we aim to predict the rating of a given user (with a given history of interactions with the recommender system and hence, already performed ratings). As one of many recommender libraries, we will get to know the **Surprise library**, which already provides the MovieLens dataset as a built-in dataset. Make use of Surprise's basic k-nearest neighbor (KNN) recommender and find suitable parameters for this algorithm. What is difference between the three KNN variants provided?
- c) 2 Points We aim to evaluate these approaches by utilizing k -fold cross evaluation. In principle, we can do two types of evaluations: The first option are error measures that capture the difference between the original rating and the predicted rating (mean absolute error (MAE), root mean squared error (RMSE)). The second option for evaluation a RecSys is to make use of $\text{recall}@k$, $\text{precision}@k$ and $F_1@k$, which requires a ranked list of recommendations. Please compute both evaluation options and compare the results.
- d) 1 Point Which role does the similarity measure employed play in such a KNN recommendation scenario? How do the similarity measures provided by Surprise work and how do they compare in terms of performance?
- e) 1 Point The MovieLens dataset comprises ratings between 1 and 5 stars (so-called explicit feedback as the user has to explicitly rate the item). The alternative to explicit feedback is implicit feedback that is based on implicit user actions (e.g., the movies watched or clicked). Implicit feedback is mostly binary. We can transform the MovieLens dataset to an implicit dataset if we choose to consider all ratings of > 3 stars as positive feedback and all other ratings as negative feedback. Please repeat the previous experiments now on the implicit feedback dataset and elaborate on the differences.

Exercise 2 (Matrix Factorization)

[4 Points]

- a) 3 Points Please read the following introductory article to Matrix Factorization:

Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, (8):30–37, 2009. URL <http://www.inf.unibz.it/~ricci/ISR/papers/ieeecomputer.pdf>

... answer the following questions:

- 1 Point What is the basic idea of matrix factorization and how does it contribute to recommender system performance?
 - 1 Point How does matrix factorization work?
 - 1 Point Which learning algorithms are proposed in the paper and how do those work in principle?
- b) 1 Point Based on the Surprise library, we can easily also compute SVD recommendations. What is the difference between matrix factorization and SVD? How do the results compare to the results of the rather simple KNN recommender algorithm of the previous exercise?

Important: Submit your solution to OLAT and mark your solved exercises with the provided checkboxes. The deadline ends at 23:59 on the day before the discussion.