KV Learning from User-generated Data – SS 2016

Practical Project: Rating Prediction for Movie Recommendation
Description of Task 1

Deadline for submission: Sun, April 17, 2016, 23:59
Date of presentation: Tue, April 19, 2016

For this task you are given a list of about 800,000 anonymized movie ratings of ~3,900 movies made by ~6,000 users (`training.dat`) where each line complies to the following format:

```
user_id <tab> movie_id <tab> rating
```

Your task is to make rating predictions for ~200,000 user-item pairs using collaborative filtering techniques. To this end, you are given another file (`predict.dat`), with each line formatted as

```
test_user_id <tab> test_movie_id
```

and swhich you should extend with your rating prediction and return as:

```
test_user_id <tab> test_movie_id <tab> predicted_rating
```

You are expected to experiment with different approaches (memory- and model-based collaborative filtering) and settings, and to use the parameter combination that yielded best results in your experiments for generating the predictions. For determining which setting yields the best result, you should deploy a reasonable evaluation strategy. (Most important: make sure that you don't use cases for testing that you have used for learning!) To compare your predictions with the ground truth, we will make use of RMSE; so you might want to use that as well.

Your submission must contain:

1. **The movie rating predictions** (quality of approach: max. 15 points)
   The completed `predict.dat` file (in the same line order!) named according to your group letter `task1_group#_predict.dat`
2. **A report on your approach** (quality of report: max. 10 points)
   This report contains a description of the task and of the data that you are working with, descriptions of the approaches that you investigated (detailed enough to show that you actually know what you are doing), your experimental setup for systematic experimentation, a concise presentation of your results (supported, e.g., by tables and/or graphs), and your conclusions. The report must be formatted according to the given LaTeX template (similar to ACM style) and comprise at least three pages. (as PDF)
3. **Your presentation slides** as PDF (quality of presentation: max. 5 points)