Assignment 5

Team name: M2 Robo

Lee Chun Yin	Chiu Yu Ying	Chan Kwan Yin
3035469140	3035477630	3035466978
		Team leader

1. Introduction

1.1. collaborative filtering (CF)

CF is a technique for recommendation system, in which historical feedback data are used to infer connections between users and products [2]. While additional features can be introduced to offset certain bias effects [1], two inputs (user and product) and one output (user rating on the product) are generally sufficient to train a CF model without involving domain-specific data.

Two major approaches for CF include neighbourhood models and latent factor models. Neighbourhood models compare the similarity between users and recommend products positively rated by similar users, while latent factor models perform dimensional reduction on both users and movies to a common, smaller set of feature attributes such that users are recommended with movies of more coherent features.

1.2. The Netflix Prize dataset

The Netflix Prize is a competition for the prediction of users' favour of movies. The dataset provides existing ratings of users on given movies, and models are trained to predict new ratings.

1.2.1 Dataset format

The dataset contains 100480507 rows of data structured in the following format:

User ID	490189 discrete values
Movie ID	17770 discrete values
Rating	1, 2, 3, 4, 5
Date	Dates from 1999 to 2005

1.2.2 Distribution of ratings

Ratings are mostly distributed around 3 and 4, as shown in Figure 1.

Except for some extreme cases, the number of ratings per user over the 7 years mostly follow an exponential relation for users from 10 to 1000 ratings, as shown in Figure 2. The top 10 users with the highest number of ratings range from 17651 to 8877.

For movies with at least 100 ratings (which is the case for the majority), their numbers of ratings demonstrate a similar but more concave relationship, as shown in Figure 3.

1.2.3 Evaluation process

To evaluate performance, 1425333 rows (about 1.42% of all data) are speified as the standard "probe". We perform analysis in the following procedure:

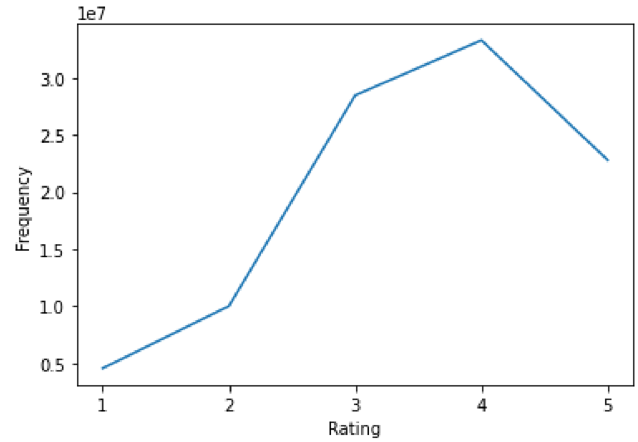

Figure 1. Frequency of ratings

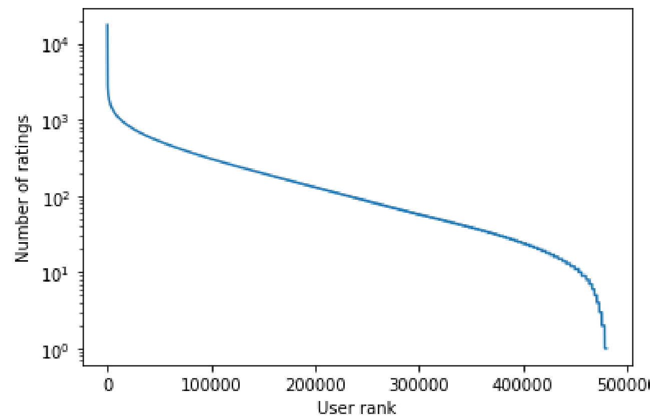

Figure 2. Number of ratings per user

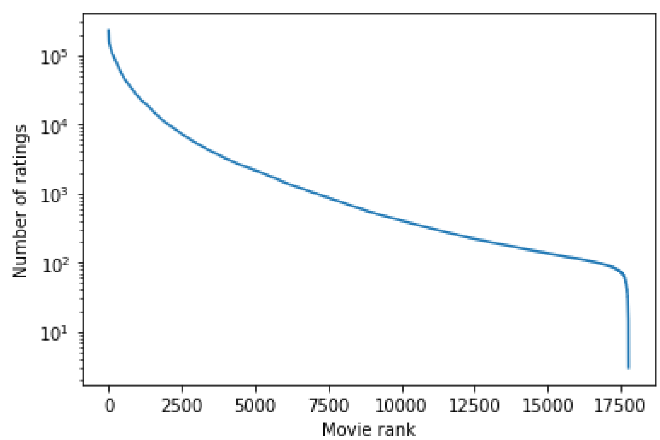

Figure 3. Number of ratings per movie

1. Train the model with the 99055174 non-probe rows.
2. Predict user ratings with the 1425333 probe rows.
3. Compute the root mean squared error (RMSE) between the predicted data and actual data.

In this project, we evaluate three models, namely:

- k-nearest neighbours (KNN), a simple neighbourhood model
- singular value decomposition (SVD), a latent factor model that accounts for user bias
- neural collaborative filtering (NCF), a latent factor model that represents features with neural network weights

2. Conclusion

2.1. KNN model

In the KNN recommendation system, we discover that the result using cosine with mean-corrected ratings has the best performance among all (RMSE = 1.1046) while the result using cosine with 3-corrected ratings does not work (with RMSE = 1.1800). The Netflix rating is a 5-star rating which is like a Likert scale commonly used in research surveys. In fact, such a scale has a known issue called the central tendency bias, which means people having surveys tend to have a neutral or middle rating and avoid the endpoints of scale [3]. The poor performance of 3-corrected rating shows the dataset does not exactly follow the same situation of the bias mentioned.

However, according to the frequency graph of each value of the scale, we can see the ratings accumulate at 3-4. This may be because viewers, who are potentially giving unfavour rating, will simply close the movie and choose the other one after clicking and watching the preview of certain movie. The potential raters only remain those who are still interested in the movie after watching the preview. With the central tendency bias, those remaining viewers have a high tendency to give a rating among the favorable rating range (i.e. between 3 and 5). In this case, 4 is the "middle" to be selected. Apart from that, we also observe that there are less counts for unfavorable ratings (i.e. less than 3). This may be because only those viewers, who watched most of the movie, feel disappointed will give unfavorable ratings. This may be the reason for the less counts for unfavorable ratings.

Therefore, based on these cases, the "middle" of the rating is not just 3 and also 4. The mean-corrected ratings can consider the mentioned situation.

2.2. SVD model

In the SVD model, we observed that train and test RMSE are inversely related over different selections of hyperparameters. Figure 4 visualizes the results of different hyperparameters we tested with.

2.3. NCF model

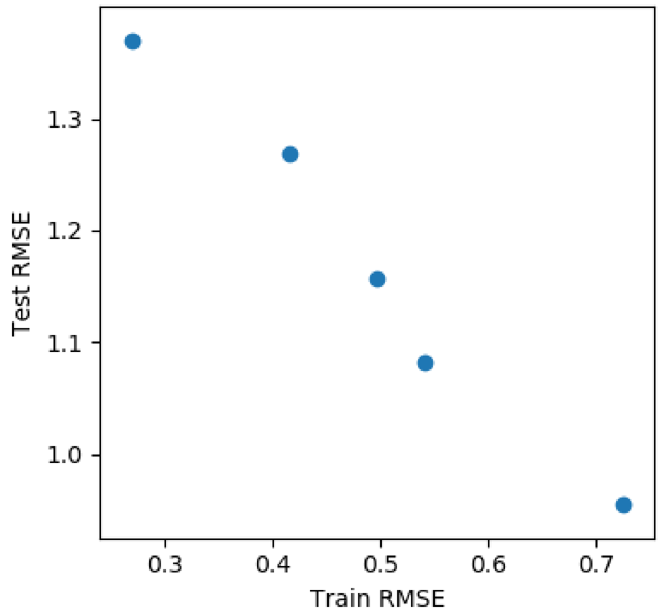

Figure 4. Train RMSE vs Test RMSE

References

[1] Robert M Bell, Yehuda Koren, and Chris Volinsky. The bellkor 2008 solution to the netflix prize. *Statistics Research Department at AT&T Research*, 1(1), 2008. 2

[2] Y. Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2008. 2

[3] Stanley S Stevens. Issues in psychophysical measurement. *Psychological review*, 78(5):426, 1971. 3