

## **Assignment 5**

**Team name: M2 Robo**

Lee Chun Yin  
3035469140

Chiu Yu Ying  
3035477630

Chan Kwan Yin  
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.

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

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

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

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

## References

- [1] Robert M Bell, Yehuda Koren, and Chris Volinsky. The belkor 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