

Assignment 5

Team name: M2 Robo

Lee Chun Yin
3035469140

Chiu Yu Ying
3035477630

Chan Kwan Yin
3035466978
Team leader

1. Introduction

The Netflix dataset contains 100480507 rows of data, each containing the user ID, the movie ID, the rating value (number of stars from 1 to 5) and the date that the rating was given. The ratings are given in the years 1999-2005. from 480189 users on 17770 movies.

The goal of the project is to predict the rating value based on the other fields. This is achieved by identifying user interests and movie characteristics based on their ratings, and comparing these information against the user and movie under prediction. This technique is known as collaborative filtering (CF).

To evaluate performance, 1425333 (about 1.4% of all data) are selected as the "probe" data. We predict the user ratings on the probe data and compare their root mean squared error (RMSE) from the provided data.

In this project, we compare three different models, namely k -nearest neighbours (KNN), singular value decomposition (SVD)++ and neural collaborative filtering (NCF).

2. Conclusion