D 351 Mini Project 4

Due: 12/03/23

Movie Recommendation System using User-Based Collaborative Filtering

Objective:

Build a movie recommendation system using user-based collaborative filtering using k-Nearest Neighbours(KNN) on the Movielens 100k dataset. Train the model on 80% of the data and evaluate its performance by predicting movie ratings for the remaining 20%.

Dataset:

Movielens 100k dataset (https://grouplens.org/datasets/movielens/100k/)

Instructions for Python Implementation(10 points):

Make sure to include the following steps of the code:

I. **Data Loading and Exploration**: Load and explore the dataset to understand its structure and features. Make sure to handle missing values (if any).

II. **Data Preprocessing**: Convert the data into a suitable format for user-based collaborative filtering.

III. **Data Split**: Hide the rating of 20% of movies from 20% of users and use the remaining data to predict the rating of the hidden movie by the hidden users. In other words, you can know 80% of the ratings by the selected 20% testing users.

IV. **User-Based Collaborative filtering:** Implement the user-based collaborative filtering algorithm using KNN and predict the ratings of the testing set based on the ratings of similar users in the training set. You may use Cosine similarity or Pearson Correlation to calculate user similarities.

V. **Evaluation:** Evaluate the model's performance using appropriate metrics like Mean Squared Error(MSE) or Root Mean Squared Error(RMSE). Compare the predicted ratings with the actual ratings in the testing set.

VI. **Results and Analysis:** Analyze the performance of the user-based collaborative filtering model and discuss the strengths and limitations of the model.

Bonus(1 points):

Implement Item-based collaborative filtering and compare the approach with user-based collaborative filtering. Highlight any difference in their performance.

Submission:

Students should submit a python code and a PDF file of the code detailing the steps taken in each phase of the project, including code and interpretations of results.