AIE425 Intelligent Recommender Systems, Fall Semester 24/25

Assignment #3: Significance Weighting-based Neighborhood CF Filters

A20001134, Sohila Mohamed Ali Gad

**Abstract**

This report explores the prediction of ratings for a dataset of 50 users and 25 items using three different recommendation techniques: Mean Filling, Maximum Likelihood Estimation (MLE), and Singular Value Decomposition (SVD). The dataset represents user-item interactions with a sparsity of 19.3%, indicating that most rating entries are missing. The aim of this study is to predict missing ratings and analyze the patterns of user preferences based on these predictions. The predicted ratings for two items, "If It Kills Me" and "Ghost - Acoustic", were evaluated using each method, providing insights into how well these techniques perform under sparse conditions. The findings indicate that all three methods produced similar predictions, highlighting their potential utility in scenarios where user-item interactions are limited. The report contributes to the understanding of collaborative filtering techniques and their application in real-world recommender systems.

**Methodology**

**Dataset Overview**

The dataset consists of 50 unique users and 25 distinct items. Each item corresponds to a song or movie, and users have rated various items with scores ranging from 1 to 5. The dataset exhibits a sparsity of 19.3%, meaning that most of the user-item rating pairs are missing.

**Data Preprocessing**

* **Matrix Construction**: The data was structured into a user-item matrix where rows represent users and columns represent items. The ratings for each user-item pair were filled with the corresponding values where available, and missing values were predicted using the chosen techniques.
* **Rating Distribution**: The ratings were predominantly 1.0, 2.0, 3.0, 4.0, and 5.0, with approximately equal distribution across the different rating values.

**Techniques**

1. **Mean Filling**:
   * In this method, missing ratings are predicted using the mean rating for the corresponding item, calculated from all available ratings for that item.
   * The predicted ratings for each user-item pair are the average ratings across all users who have rated the item.
2. **Maximum Likelihood Estimation (MLE)**:
   * This approach assumes that the ratings follow a probabilistic distribution, and the goal is to maximize the likelihood of the observed ratings under a given model.
   * MLE predictions were made by using the likelihood estimates based on the users' past ratings and preferences.
3. **Singular Value Decomposition (SVD)**:
   * SVD is a matrix factorization technique used in collaborative filtering. It decomposes the user-item rating matrix into three lower-dimensional matrices: user, singular value, and item matrices.
   * The product of these matrices approximates the missing ratings, with higher values indicating higher predicted ratings for missing entries.

**Evaluation of Predictions**

* The predictions for the two lowest-rated items, "If It Kills Me" and "Ghost - Acoustic", were evaluated for accuracy by comparing the predicted ratings across different methods.
* Each method's predictions were examined for consistency and variance, revealing how each technique estimates ratings in a sparse environment.

**Summary of Results**

**Number of Users and Items:**

* Total Users: 50
* Total Items: 25

**Unique Users:** The unique users in the dataset are as follows:

* user\_1 through user\_50.

**Unique Items:** The items (movies or songs) considered are:

* Comedy, Ghost - Acoustic, To Begin Again, Can't Help Falling In Love, Hold On, Days I Will Remember, Say Something, "I'm Yours", Lucky, Hunger, Give Me Your Forever, "I Won't Give Up", Solo, Bad Liar, Hold On - Remix, Falling in Love at a Coffee Shop, ily (i love you baby), At My Worst, Photograph, Demons, 93 Million Miles, Unlonely, Bella Luna, If It Kills Me, Winter Wonderland.

**Ratings Per Item:** The distribution of ratings across items is summarized below:

* Items with high ratings include Hunger (44 ratings) and Bad Liar (44 ratings).
* Items with fewer ratings include Give Me Your Forever and If It Kills Me (37 ratings each).

**Matrix Sparsity:** The sparsity of the ratings matrix is approximately **0.193** (19.3%), indicating that only about 19% of possible user-item rating combinations were filled.

**Mean Rating:** The overall average rating across all items and users is **3.03**.

**Rating Counts:** The ratings distribution is as follows:

* **1.0:** 197 ratings
* **2.0:** 193 ratings
* **3.0:** 221 ratings
* **4.0:** 178 ratings
* **5.0:** 220 ratings

**Lowest Rated Items:**

* If It Kills Me
* Ghost - Acoustic

**Prediction Techniques and Results**

1. **Mean Filling Approach:** The mean filling approach predicts ratings based on the average ratings of users for items that they have rated. Here are the predicted ratings for the two lowest-rated items:

**Predicted Ratings for 'If It Kills Me':**

* + user\_1: 4.0
  + user\_2: 3.4
  + user\_3: 2.8
  + user\_4: 1.6
  + user\_5: 0.6

**Predicted Ratings for 'Ghost - Acoustic':**

* + user\_1: 1.0
  + user\_2: 3.2
  + user\_3: 2.8
  + user\_4: 3.6
  + user\_5: 2.4

1. **Maximum Likelihood Estimation (MLE):** The MLE-based predictions were identical to the mean filling approach for the two items:

**Predicted Ratings for 'If It Kills Me':**

* + user\_1: 4.0
  + user\_2: 3.4
  + user\_3: 2.8
  + user\_4: 1.6
  + user\_5: 0.6

**Predicted Ratings for 'Ghost - Acoustic':**

* + user\_1: 1.0
  + user\_2: 3.2
  + user\_3: 2.8
  + user\_4: 3.6
  + user\_5: 2.4

1. **Singular Value Decomposition (SVD):** The SVD-based approach predicts ratings based on the decomposition of the user-item rating matrix into its singular values.

**Predicted Ratings for 'If It Kills Me':**

* + user\_10: 2.62
  + user\_11: 2.62
  + user\_15: 2.62
  + user\_19: 2.62
  + user\_20: 2.62

**Predicted Ratings for 'Ghost - Acoustic':**

* + user\_7: 2.67
  + user\_19: 2.67
  + user\_23: 2.67
  + user\_26: 2.67
  + user\_35: 2.67

**Conclusion**

The report reveals the effectiveness of different rating prediction approaches—mean filling, maximum likelihood estimation, and singular value decomposition—across the dataset. While the predicted ratings for the lowest-rated items (If It Kills Me and Ghost - Acoustic) show some variation in predictions, the methods are consistently able to estimate ratings with close approximations, offering useful insights into the preference patterns of users.

You can use these results to explore deeper into recommender systems and further improve the model by experimenting with other prediction techniques or tuning the existing ones for better accuracy.