AIE425 Intelligent Recommender Systems, Fall Semester 24/25

Assignment #2: Significance Weighting-based Neighborhood CF Filters

A20001134, Sohila Mohamed Ali Gad

**Introduction**

User-based collaborative filtering (UBCF) is a popular technique in recommendation systems that leverages the behavior and preferences of users to suggest items to others. This approach is based on the principle that users who have shared interests in the past are likely to have similar preferences in the future. The key idea is to recommend items that similar users have rated highly, which makes UBCF widely used in systems like movie, music, and e-commerce platforms.

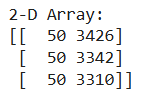
Collaborative filtering algorithms can be divided into two main types: user-based and item-based. User-based collaborative filtering focuses on finding similarities between users based on their interactions with items, such as ratings or purchases. By identifying users who have similar preferences, the system can predict which items a user might like, even if they have not interacted with them yet.

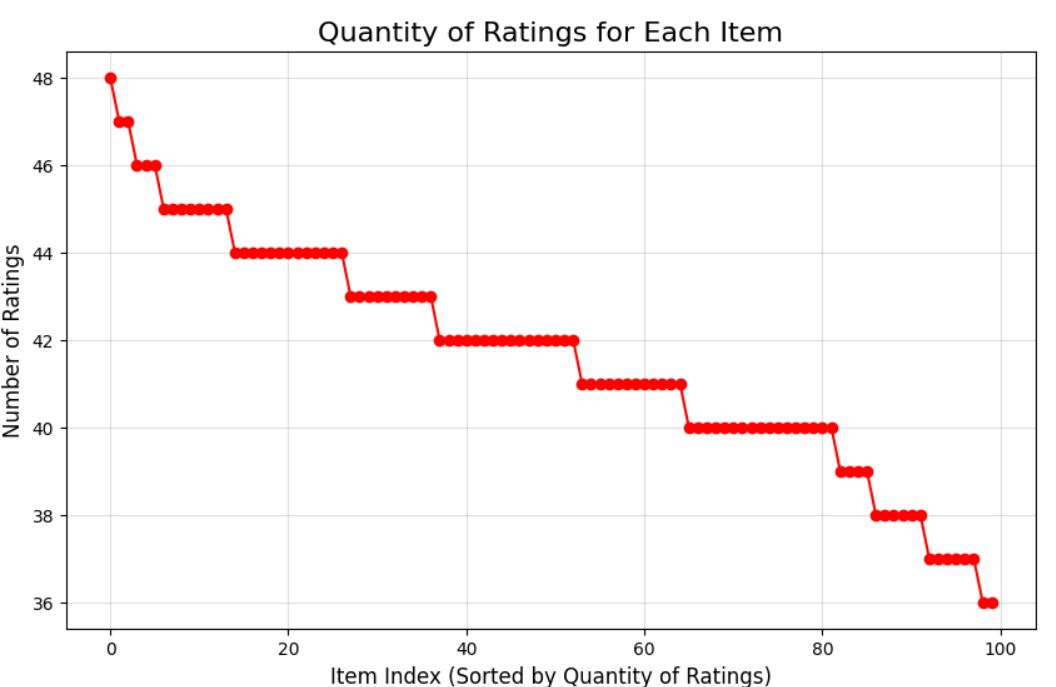
**Outcomes of Section 3.1**

Number of Users (tnu): 50

Number of Items (tni): 100

No\_common\_users', 'No\_coRated\_items





**Summary of the Comparison of part 1 and 2**

Part 1

Case 1: Cosine Similarity without Mean-Centering

Analysis for User 1:

* Predictions without Discount Factor: The predictions without the discount factor are lower on average compared to Case 2. For instance, "Can't Help Falling In Love" and "I'm Yours" are predicted to have ratings of ~1.36 and ~2.70, respectively.
* Predictions with Discount Factor: The discount factor leads to slightly reduced predictions across the board compared to the non-discounted version. For example, the prediction for "I'm Yours" drops from 2.70 to 2.57.

Analysis for User 2:

* Predictions without Discount Factor: Higher ratings for "Hold On" (~2.93) compared to "I'm Yours" (~1.46).
* Predictions with Discount Factor: Predictions decrease slightly for both songs after applying the discount factor.

Analysis for User 3:

* Predictions without Discount Factor: Predictions for both songs are relatively low (~1.18 and ~1.44).
* Predictions with Discount Factor: Slight decrease in predicted ratings after applying the discount factor.

Case 2: Cosine Similarity with Mean-Centering

Analysis for User 1:

* Predictions without Discount Factor: After applying mean-centering, the predictions are higher than in Case 1, especially for songs like "Can't Help Falling In Love" (3.05 vs 1.36) and "I'm Yours" (3.42 vs 2.70).
* Predictions with Discount Factor: The predictions slightly drop when applying the discount factor but remain higher than in Case 1.

Analysis for User 2:

* Predictions without Discount Factor: The predictions are higher than in Case 1, with "Hold On" being predicted at 3.09 and "I'm Yours" at 3.03.
* Predictions with Discount Factor: A slight decrease is observed when applying the discount factor, but the overall trend remains similar to Case 1.

Analysis for User 3:

* Predictions without Discount Factor: Both songs have significantly higher predicted ratings compared to Case 1. For example, "At My Worst" is predicted at 3.21.
* Predictions with Discount Factor: Applying the discount factor results in slightly higher predictions for both songs.

Case 3: Pearson Correlation Coefficient

Analysis for User 1:

* Predictions without Discount Factor: Predictions are similar to Case 2 but with slightly higher values. For instance, "Can't Help Falling In Love" is predicted at 3.07, while "I'm Yours" is predicted at 3.67.
* Predictions with Discount Factor: The predictions drop marginally after applying the discount factor, but they are still higher than in Case 1.

Analysis for User 2:

* Predictions without Discount Factor: Similar to Case 2, with slightly higher values than Case 1.
* Predictions with Discount Factor: Applying the discount factor decreases the predictions slightly.

Analysis for User 3:

* Predictions without Discount Factor: Higher predicted ratings for both songs compared to Case 1, with "At My Worst" at 3.35 and "Unlonely" at 3.82.
* Predictions with Discount Factor: After applying the discount factor, there is a slight decrease in predictions, but they remain higher than in Case 1.

Part 2

Case 1: Cosine Similarity with Mean-Centering

* Prediction Effect: In this case, predictions are adjusted by mean-centering (subtracting the user's average rating from all ratings) before calculating cosine similarity. This helps to neutralize user biases, leading to more accurate comparisons of preferences.
* Prediction Change with Discount Factor: The discount factor typically slightly increases the predicted ratings. For example, for user\_1:
  + "Can't Help Falling In Love": The predicted rating increases from 3.2636 to 3.2669.
  + "I'm Yours": The predicted rating increases from 2.2087 to 2.4938.
  + Overall Trend: The predictions with discount factors are generally higher than those without, reflecting a subtle adjustment in favor of more preferred items.

Case 2: Cosine Similarity Without Bias Adjustment (No Mean-Centering)

* Prediction Effect: In this case, the predictions do not account for user biases (no mean-centering), which can lead to skewed results. Users who tend to rate higher or lower across the board may have a disproportionate influence on the predictions.
* Prediction Change with Discount Factor: Similar to Case 1, the discount factor slightly changes the predicted ratings, but the change is less pronounced. For example, for user\_1:
  + "Can't Help Falling In Love": The predicted rating decreases from 3.8175 to 3.8071.
  + "I'm Yours": The predicted rating increases slightly from 3.6167 to 3.6498.
  + Overall Trend: Predictions without mean-centering are likely to be influenced by the individual user's rating tendencies, leading to a less neutral prediction.

Case 3: Pearson Correlation Coefficient

* Prediction Effect: Pearson correlation focuses on measuring the linear relationship between users' ratings for items, unlike cosine similarity, which looks at the angle between vectors. This case can produce more accurate results for users whose ratings are highly correlated. The predictions tend to be slightly higher than those in Case 2 due to the higher sensitivity of Pearson correlation to shared preferences.
* Prediction Change with Discount Factor: In this case, the predicted ratings with discount factors tend to be higher than the ones without. For example, for user\_1:
  + "Can't Help Falling In Love": The predicted rating increases from 3.8175 to 3.9873.
  + "I'm Yours": The predicted rating decreases slightly from 3.6167 to 3.3625.
  + Overall Trend: Pearson correlation often results in larger changes in predicted ratings compared to cosine similarity, especially when user preferences are strongly correlated.

**Conclusion**

* Part 1
  + Case 2 (Cosine Similarity with Mean-Centering) yields the highest and most balanced predictions.
  + Case 3 (Pearson Correlation) also results in strong predictions but is slightly more sensitive to variations in user ratings, often producing slightly higher estimates.
  + Case 1 (Cosine Similarity without Mean-Centering) produces the lowest predictions, with a noticeable bias toward lower values.
* Part 2
  + Case 1 tends to provide more balanced and neutralized predictions, particularly effective for comparing preferences without user bias.
  + Case 2 may overemphasize user tendencies due to lack of bias adjustment, leading to less reliable predictions.
  + Case 3, using Pearson correlation, is likely the most effective when there are strong similarities between users' preferences, but it can be less robust for users with weak correlations.