CSC 466 Lab 5 Report:
Collaborative Filtering and Recommender Systems

Martin Solomon Hsu: mshsu@calpoly.edu
Rachel Roggenkemper: rroggenk@calpoly.edu
Instructor: Dr. Alexander Dekhtyar
CSC 466: Knowledge Discovery from Data, Fall 2023

Abstract

In this lab, we implement a number of memory-based collaborative filtering techniques and test their accuracy on a stable collection of joke ratings from the Jester project. We performed rigorous tuning of our parameters which are the number of predictions we should evaluate and the number of times the test is to be repeated. We conducted experiments comparing the overall prediction capability between three collaborative filtering techniques (mean utility, weighted sum, and adjusted weighted sum) and comparing user-based versus item-based collaborative filtering. Ultimately, we found that adjusted weighted sum and item-based collaborative filtering performed the best in terms of prediction metrics and mean absolute error.

I. Introduction

Collaborative filtering and recommender systems offer personalized solutions in various domains. At the heart of these systems is the challenge of understanding and predicting user preferences. The primary goal is given a set of users and their incomplete preferences over a set of items, for each user, to find new items for which they would have high preferences. Additionally, a goal is to recommend items that other users with similar preferences find to be of high utility. This is achieved by leveraging the concept of similarity in preferences: if users A and B have similar tastes, then the items highly valued by user A are likely to appeal to user B as well. This leads us into memory-based collaborative filtering methods, where we aggregate known utilities or ratings of items to predict how much a user might value an item they haven't encountered yet. We will delve into a number of memory-based collaborative filtering techniques and test their accuracy.

II. Dataset Description

We utilize the Jester dataset, provided by the Jester project at UC Berkeley. This dataset is a collection of joke ratings, featuring anonymous responses from a diverse user base. Specifically, we focus on the first of the three data files available in the dataset, which includes ratings from 73,421 users. Each user in this dataset has provided ratings for a range of jokes on a continuous scale from -10.00 to +10.00, with the value "99" signifying a null or unrated entry. When evaluating whether to recommend or not recommend a joke in order to evaluate categorical accuracy metrics such as overall accuracy, precision, or recall, we used a threshold of a rating greater than or equal to +5.00. The dataset is structured such that each row corresponds to a unique user, with the first column indicating the number of jokes rated by that user and the subsequent 100 columns representing their ratings for jokes numbered 01 through 100. Notably, a subset of jokes (specifically, columns 5, 7, 8, 13, 15, 16, 17, 18, 19, 20) is densely rated, providing a robust basis for our analysis.

III. Methods

We examined the following collaborative filtering methods, all of which were implemented with a 5-Nearest-Neighbors variation and Pearson Correlation similarity:

A. User-Based Mean Utility

User-based mean utility refers to the average rating that an item has received from all the users who have rated it.

B. User-Based Weighted Sum

User-Based Weighted Sum is a technique used to predict how a user might rate an item they haven't seen before. It works by finding users who have similar tastes to the target user, then calculating how similar these users are to the target user, next, using these similarity scores as weights, the system computes a weighted average of the ratings these similar users have given to the item in question.

C. User-Based Adjusted Weighted Sum

User-Based Adjusted Weighted Sum is a technique where the predicted rating for an item is calculated by adjusting for the average rating of each user. This approach involves finding similar users to the target user, calculating the average rating for each of these similar users, adjusting their ratings for an item by subtracting their average rating, then computing a weighted sum of these adjusted ratings, using similarity scores as weights.

D. Item-Based Mean Utility

Item-based mean utility refers to the average rating that a specific user has given to all the items they have rated.

E. Item-Based Weighted Sum

Item-Based Weighted Sum is a method for predicting a user's rating for an item by using the ratings they have given to similar items. It involves identifying items similar to the target item, calculating similarity weights for these items, and then computing a weighted average of the user's ratings for these similar items. This weighted sum represents the predicted rating for the target item, tailored to the user's preferences as demonstrated by their past ratings.

F. Item-Based Adjusted Weighted Sum

Item-Based Adjusted Weighted Sum is a technique where the prediction for a user's rating on an item is refined by adjusting for the average rating of each similar item. This method involves identifying items similar to the target item, calculating the average rating for these similar items, adjusting their ratings by subtracting their average rating, then using these adjusted ratings, a weighted sum is computed, where the weights are based on item similarity.

IV. Research Questions

To investigate these various collaborative filtering algorithms, we asked the following research questions:

A. Comparison of Collaborative Filtering Techniques: Mean Utility, Weighted Sum, and Adjusted Weighted Sum

Among the collaborative filtering techniques of mean utility, weighted sum, and adjusted weighted sum, which method provides the most accurate predictions in the Jester jokes recommender system?

B. User-Based vs. Item-Based Collaborative Filtering

In the context of joke recommendation using the Jester jokes rating dataset, which approach yields more accurate predictions: user-based or item-based collaborative filtering?

V. Experiments

A. Comparison of Collaborative Filtering Techniques: Mean Utility, Weighted Sum, and Adjusted Weighted Sum

We will first conduct tuning of our parameters which are the number of predictions we should evaluate and the number of times the test is to be repeated through rigorous testing. To compare mean utility, weighted sum, and adjusted weighted sum techniques, we will compute the prediction metrics (precision, recall, F1 measure, overall accuracy, and mean absolute error) when using mean utility, weighted sum, and adjusted weighted sum for user-based and item-based collaborative filtering. Note that when the number of times the test is repeated is greater than 1, the prediction metrics reported will be the average of all the runs. For user-based and item-based, we will first see whether mean utility, weighted sum, or adjusted weighted sum performed the best separately. Then, we will evaluate whether the same method works the best for both user-based and item-based or whether another method is more favorable for one versus the other.

B. User-Based vs. Item-Based Collaborative Filtering

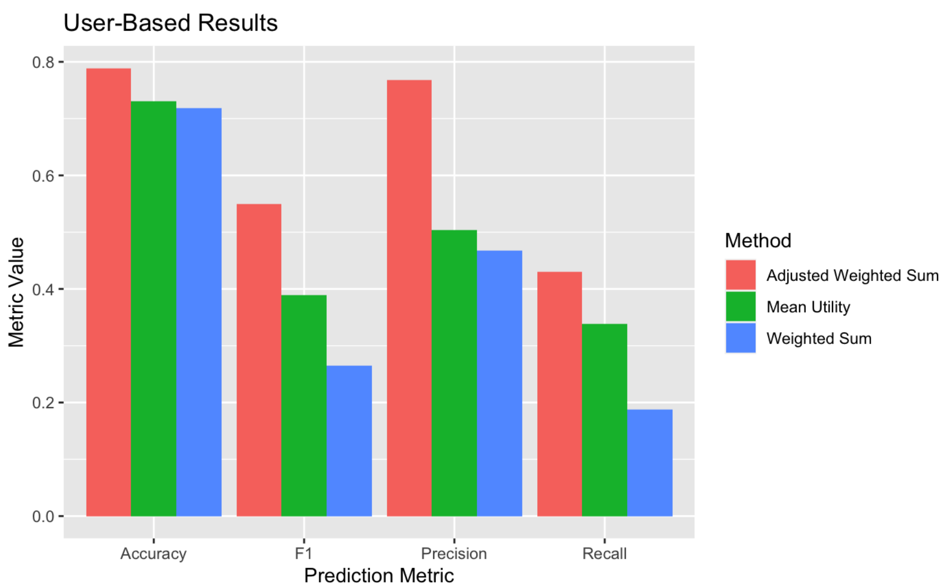
We will first conduct tuning of our parameters which are the number of predictions we should evaluate and the number of times the test is to be repeated through rigorous testing. To compare user-based and item-based collaborative filtering, we will compute the prediction metrics (precision, recall, F1 measure, overall accuracy, and mean absolute error) when using mean utility, weighted sum, and adjusted weighted sum for user-based and item-based collaborative filtering. Note that when the number of times the test is repeated is greater than 1, the prediction metrics reported will be the average of all the runs. Then, we will compute the average difference for each metric for user-based versus item-based for each of the 3 methods (mean utility, weighted sum, and average weighted sum).

VI. Results

A. Comparison of Collaborative Filtering Techniques: Mean Utility, Weighted Sum, and Adjusted Weighted Sum

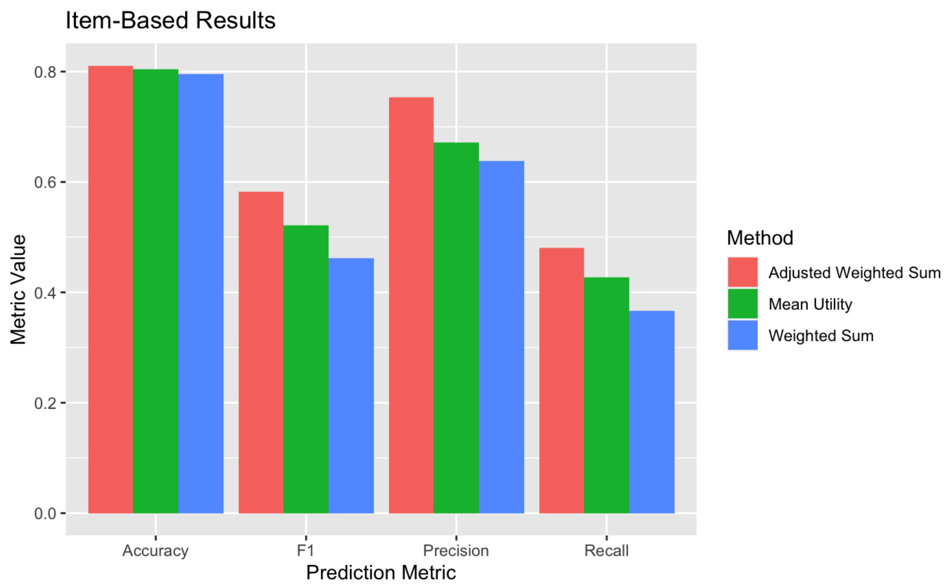
To compare the collaborative filtering techniques, for each user-based and item-based, we will visually compare the prediction metrics for mean utility, weighted sum, and adjusted weighted sum.

Figure 1: User-Based Comparison of Collaborative Filtering Techniques Results



As we can see in Figure 1, for all four prediction metrics (accuracy, F1, precision, and recall), adjusted weighted sum performs the best when using user-based aggregation. Mean utility performs second best across all four metrics, with weighted sum performing the worst across all four metrics. The accuracy for all three methods are fairly comparable, with the accuracy for mean utility and weighted sum being very similar. For F1 and recall scores, there is a clear separation between the three methods. The precision for adjusted weighted sum is substantially higher than mean utility and weighted sum.

Figure 2: Item-Based Comparison of Collaborative Filtering Techniques Results



As we can see in Figure 2, for all four prediction metrics (accuracy, F1, precision, and recall), adjusted weighted sum performs the best when using item-based aggregation, similar to what we saw above using user-based aggregation. Likewise, mean utility also performs second best across all four metrics, with weighted sum performing the worst across all four metrics. The accuracy for all three methods are very close. For F1, precision, and recall scores, there is a clear separation between the three methods.

Figure 3: Mean Absolute Error Comparison Results

Lastly, looking at the Mean Absolute Error (MAE) for item-based and user-based for all three methods in Figure 3, we see the same trends as above. Adjusted weighted sum performs the best, while mean utility also performs second best, with weighted sum performing the worst for both item-based and user-based aggregation.

B. User-Based vs. Item-Based Collaborative Filtering

To compare user-based versus item-based collaborative filtering, we will compute the average difference for each metric for user-based versus item-based for each of the 3 methods (mean utility, weighted sum, and average weighted sum).

Table 1: Mean Utility: Average Difference for Prediction Metrics

Prediction Metric	Difference (Item-Based – User-Based)
Precision	0.168
Recall	0.088
F1	0.132
Accuracy	0.073
MAE	-0.631

After evaluating the results in Table 1, we can see that for mean utility, item-based collaborative filtering outperforms user-based collaborative filtering. For all four prediction metrics, item-based aggregation yields higher metric values. Additionally, item-based collaborative filtering has a smaller MAE compared to user-based collaborative filtering, which further supports that item-based aggregation performs better.

Table 2: Weighted Sum: Average Difference for Prediction Metrics

Prediction Metric	Difference (Item-Based – User-Based)
Precision	0.170
Recall	0.178
F1	0.197
Accuracy	0.078
MAE	-0.555

Looking at Table 2, we see the same conclusion for weighted sum as we saw above with mean utility. Item-based collaborative filtering outperforms user-based collaborative filtering using the four prediction metrics and MAE to measure this statement.

Table 3: Adjusted Weighted Sum: Average Difference for Prediction Metrics

Prediction Metric	**Difference (Item-Based – User-Based)**
Precision	-0.014
Recall	0.050
F1	0.033
Accuracy	0.022
MAE	0.110

As we can see in Table 3, we see slightly different trends for adjusted weighted sum as we saw with mean utility and weighted sum. Looking at accuracy, recall, and F1 score, item-based collaborative filtering performs better than user-based collaborative filtering. However, the precision using user-based aggregation is very slightly better compared to using item-based collaborative filtering. Additionally, the MAE is also slightly better for user-based collaborative filtering compared to item-based collaborative filtering. Nevertheless, these differences are both very small differences. Ultimately, the differences between item-based and user-based collaborative filtering are much smaller for adjusted weighted sum compared to what we saw above in mean utility and weighted sum. Thus, for adjusted weighted sum, the performance for item-based and user-based collaborative filtering are fairly comparable.

VII. Conclusions

Overall, after comparing the collaborative filtering techniques (mean utility vs. weighted sum vs. adjusted weighted sum), for each user-based and item-based, adjusted weighted sum performed the best for both user-based and item-based collaborative filtering. Utilizing adjusted weighted sum produced an accuracy rate of around 80%. Additionally, after comparing user-based versus item-based collaborative filtering, item-based collaborative filtering performed the best overall, also producing an accuracy rate of around 80%. Putting these conclusions together, out of the six pairs of methods we tested, item-based adjusted weighted sum performed the best with an accuracy rate of 81.1%.

VIII. Appendix

A. User-Based Mean Utility Results

```
Random Sample Report
Method: User Mean Utility
Sample Size: 150
N Neighbors: 5
--------------------------------------------------------------------------------
Repeat: 0
Confusion Matrix:
obs    False  True
pred
False     95    17
True      20    18

Precision:         0.474
Recall:            0.514
F1 Measure:        0.493
Overall Accuracy: 0.753
MAE:               3.627
--------------------------------------------------------------------------------
Repeat: 1
Confusion Matrix:
obs    False  True
pred
False    104    30
True       8     8

Precision:         0.500
Recall:            0.211
F1 Measure:        0.296
Overall Accuracy: 0.747
MAE:               3.571
--------------------------------------------------------------------------------
Repeat: 2
Confusion Matrix:
obs    False  True
pred
False     90    34
True      12    14

Precision:         0.538
Recall:            0.292
F1 Measure:        0.378
Overall Accuracy: 0.693
MAE:               3.717
```

B. User-Based Weighted Sum Results

```
Random Sample Report
Method: User Weighted Sum
Sample Size: 150
N Neighbors: 5
--------------------------------------------------------------------------------
Repeat: 0
Confusion Matrix:
obs    False  True
pred
False     97    39
True       8     6
```

```
Precision:         0.429
Recall:            0.133
F1 Measure:        0.203
Overall Accuracy: 0.687
MAE:               3.635
--------------------------------------------------------------------------------
Repeat: 1
Confusion Matrix:
obs    False  True
pred
False    103    33
True       7     7

Precision:         0.500
Recall:            0.175
F1 Measure:        0.259
Overall Accuracy: 0.733
MAE:               3.496
--------------------------------------------------------------------------------
Repeat: 2
Confusion Matrix:
obs    False  True
pred
False    100    29
True      11    10

Precision:         0.476
Recall:            0.256
F1 Measure:        0.333
Overall Accuracy: 0.733
MAE:               3.832
```

C. User-Based Adjusted Weighted Sum Results

```
Random Sample Report
Method: User Adjusted Weighted Sum
Sample Size: 150
N Neighbors: 5
--------------------------------------------------------------------------------
Repeat: 0
Confusion Matrix:
obs    False  True
pred
False    104    25
True       4    17

Precision:         0.810
Recall:            0.405
F1 Measure:        0.540
Overall Accuracy: 0.807
MAE:               2.788
--------------------------------------------------------------------------------
Repeat: 1
Confusion Matrix:
obs    False  True
pred
False     97    23
True       8    22

Precision:         0.733
Recall:            0.489
F1 Measure:        0.587
```

```
Overall Accuracy: 0.793
MAE:              2.779
--------------------------------------------------------------------------------
Repeat: 2
Confusion Matrix:
obs    False  True
pred
False     96    29
True       6    19

Precision:        0.760
Recall:           0.396
F1 Measure:       0.521
Overall Accuracy: 0.767
MAE:              2.822
```

D. Item-Based Mean Utility Results

```
Random Sample Report
Method: Item Mean Utility
Sample Size: 150
N Neighbors: 5
--------------------------------------------------------------------------------
Repeat: 0
Confusion Matrix:
obs    False  True
pred
False    107    17
True       7    19

Precision:        0.731
Recall:           0.528
F1 Measure:       0.613
Overall Accuracy: 0.840
MAE:              2.631
--------------------------------------------------------------------------------
Repeat: 1
Confusion Matrix:
obs    False  True
pred
False    111    20
True       7    12

Precision:        0.632
Recall:           0.375
F1 Measure:       0.471
Overall Accuracy: 0.820
MAE:              3.185
--------------------------------------------------------------------------------
Repeat: 2
Confusion Matrix:
obs    False  True
pred
False     96    28
True       9    17

Precision:        0.654
Recall:           0.378
F1 Measure:       0.479
Overall Accuracy: 0.753
MAE:              3.206
```

E. Item-Based Weighted Sum Results

```
Random Sample Report
Method: Item Weighted Sum
Sample Size: 150
N Neighbors: 5
--------------------------------------------------------------------------------
Repeat: 0
Confusion Matrix:
obs    False  True
pred
False    105    27
True       7    11

Precision:        0.611
Recall:           0.289
F1 Measure:       0.393
Overall Accuracy: 0.773
MAE:              3.133
--------------------------------------------------------------------------------
Repeat: 1
Confusion Matrix:
obs    False  True
pred
False    101    27
True       8    14

Precision:        0.636
Recall:           0.341
F1 Measure:       0.444
Overall Accuracy: 0.767
MAE:              3.060
--------------------------------------------------------------------------------
Repeat: 2
Confusion Matrix:
obs    False  True
pred
False    113    16
True       7    14

Precision:        0.667
Recall:           0.467
F1 Measure:       0.549
Overall Accuracy: 0.847
MAE:              3.103
```

F. Item-Based Adjusted Weighted Sum Results

```
Random Sample Report
Method: Item Adjusted Weighted Sum
Sample Size: 150
N Neighbors: 5
--------------------------------------------------------------------------------
Repeat: 0
Confusion Matrix:
obs    False  True
pred
False    106    16
True       9    19

Precision:        0.679
```

```
Recall:            0.543
F1 Measure:        0.603
Overall Accuracy: 0.833
MAE:               3.272
--------------------------------------------------------------------------------
Repeat: 1
Confusion Matrix:
obs    False  True
pred
False     99    24
True       9    18

Precision:         0.667
Recall:            0.429
F1 Measure:        0.522
Overall Accuracy: 0.780
MAE:               2.899
--------------------------------------------------------------------------------
Repeat: 2
Confusion Matrix:
obs    False  True
pred
False    101    25
True       2    22

Precision:         0.917
Recall:            0.468
F1 Measure:        0.620
Overall Accuracy: 0.820
MAE:               2.548
```