Plagiarism Scan Report

0% Plagiarism

0% Exact Match

0% Partial Match

100% Unique

Words	974
Characters	6604
Sentences	140
Paragraphs	209
Read Time	5 minute(s)
Speak Time	7 minute(s)

Content Checked For Plagiarism

AIE425 Intelligent Recommender Systems, Fall Semester 24/25
Assignment #2: Significance Weighting-based Neighborhood CF Filters
Student ID: [A20000726]
Full Name: [Mohamed Ibrahim Fekry]

1. Introduction
This document focuses on the discussion and conclusion on the implementation of user-based and item-based collaborative filtering models using methodologies such as cosine similarity, Pearson Correlation Coefficient (PCC), bias adjustments, and the introduction of significance weighting.

2. Dataset Preparation
2.1 Synthetically Created Dataset
The dataset was synthetically created using simulated feedback generated as part of Assignment 1. The feedback was adjusted to fit a 1–5 rating scale for normalization purposes. This normalization ensured consistent interpretation of ratings, thereby reducing discrepancies across datasets and enhancing compatibility with collaborative filtering models.

2.2 Key Metrics
* Total Number of Users (tnu): 50 users.
* Total Number of Items (Eni): Extracted from movie datasets from TMDb sources.
* Ratings Distribution:
A bar chart was presented showing how ratings were distributed across items, thereby identifying trends and

patterns of sparsity. This visualization revealed data gaps and the extent of sparsity in the dataset.

2.3 Active Users
Three active users were selected with varying numbers of missing ratings to simulate real-life scenarios where users fail to rate certain items:
* User U1: 2 missing ratings.
* User U2: 3 missing ratings.
* User U3: 5 missing ratings.
These patterns mimic real-world behaviors encountered in collaborative filtering models.

2.4 Target Items
Two items were selected for prediction based on their missing rating percentages:
* Item I1: Approximately 4% missing ratings.
* Item I2: Approximately 10% missing ratings.
These varying levels of sparsity in the selected target items were analyzed to study how prediction models could handle sparse data patterns.

2.5 Co-ratings Metrics
The number of users who co-rated the items with an active user was computed. A 2D matrix was formed with the number of common users (in descending order) and their associated co-rated items. This matrix indicates potential collaborative opportunities within the dataset.

2.6 Threshold (Σ) Computation
For each active user, the maximum number of users who co-rated at least 30% of items was determined. The thresholds varied for U1, U2, and U3 based on their interaction patterns. This approach ensured a more adaptive threshold approach to user behavior and collaboration patterns.

3. Summary of the Comparison of Part 1 and Part 2
3.1 User-Based Collaborative Filtering (Part 1)
Case Study 1.1: Cosine Similarity Without Bias Adjustment
* Results:
* The nearest 20% of users had been defined for each active user based on cosine similarity.
* Predictions relied on unnormalized correspondence of user preferences, which may lead to bias due to rating scale differences.
* Weight effect of discount factor:
* Weaker links were diminished through the discounting of similarities, leading to improved prediction accuracy.
* Incorporating a discount factor prioritized more meaningful user relationships while dampening less relevant ones.

Case Study 1.2: Cosine Similarity with Bias Adjustment
* Results:
* Adjusting for bias revealed latent rating patterns and resulted in more consistent similarity scores.
* Bias adjustments proved essential for datasets with user preferences skewed toward rating items above or below the norm.
* Effect of Discount Factor:
* Predictions derived from discounted similarity showed less noise and focused on more relevant associations.

Case Study 1.3: Pearson Correlation Coefficient (PCC)

* Findings:
* PCC captured user relationships in a linear fashion, offering insights that cosine similarity could not capture.
* Unlike cosine similarity, PCC reflected trends in situations where users rated items differently but consistently.
* Impact of Discount Factor:
* Further discounting refined predictions by assigning less weight to less significant correlations.
* This approach improved prediction accuracy for users with few co-rated items.

Comparison Table for the First 5 Users part 1

User	Item	Cosine	Cosine with Bias	PCC
U1	974453	2.7861	2.8102	2.8106
U1	558449	2.2165	1.9286	1.9189
U1	1100782	2.3382	2.436	2.4294
U1	945961.1	2.7785	1.8356	1.8408
U1	995803	3.2202	3.0707	3.0782
U2	1299652	2.1054	1.9776	2.0743
U2	957119	2.5537	2.4481	2.2716
U2	533535			

1.4485
2.0328
1.6192
U2
1034541
3.3467
3.1992
3.4528
U2
1124641
2.1045
2.3865
2.4718
U3
1147416
2.1036
2.0643
2.0905
U3
1100782
1.6605
1.527
1.5348
U3
791042
2.5493
2.1638
2.1842
U3
1154223
2.6625
2.5191
2.5144
U3
1124641
2.2234
1.95
1.9511
U4
974453
2.677
2.8946
3.2413
U4
845781
2.5581
3.0286
3.0283
U4
558449
3.5573
3.7643
3.552
U4

1118031
2.9932
2.8135
2.7278
U4
945961
1.3323
1.6606
1.7244
U5
912649
1.9962
1.9106
1.901
U5
539972
1.7812
1.9361
1.9285
U5
1034541
2.7941
2.9443
2.9383
U5
1010581
2.3381
2.5561
2.5458
U5
939243
3.123
3.1708
3.1681

3.2 Item-Based Collaborative Filtering (Part 2)

Case Study 2.1: Cosine Similarity Without Bias Adjustment

* Findings:

* For each target item, the top 20% closest items were identified.

* Missing predictions were calculated based on these target items.

* Impact of Discount Factor:

* Improved precision was observed because stronger item relationships were emphasized, while weaker ones were dampened.

Case Study 2.2: Cosine Similarity with Bias Adjustment

* Findings:

* Bias adjustment allowed similarity computation after normalizing rating scales.

* It enabled the discovery of better patterns in the dataset by minimizing skewed similarities.

* Impact of Discount Factor:

* Prioritized item-item interactions, leading to improved prediction reliability.

Case Study 2.3: Pearson Correlation Coefficient (PCC)
* Discoveries:
* PCC proved particularly useful in cases where cosine similarity failed, especially in sparse datasets with few co-rated ratings.
* Effect of Discount Factor:
* Improved prediction accuracy by focusing on correlations with higher significance.

Key Comparisons Across Methods
1. Cosine Similarity vs PCC:
* PCC was better for understanding sensitivity to linear trends.
* Cosine similarity provided better overall alignment for most predictions.
2. With vs Without Bias Adjustment:
* Bias correction consistently improved prediction accuracy by minimizing the effects of outlying behaviors or rating patterns.
3. With vs Without Discount Factor:
* Predictions benefited from discounting, which reduced the influence of weak or less meaningful similarities.

2 | Page

Matched Source

No plagiarism found