PLAGIARISM SCAN REPORT

Report Generation Date: 17-12-24

Words: 987

Characters: 6770

Excluded URL : N/A

0% **Plagiarism**	**100%** **Unique**
0 **Plagiarized Sentences**	**46** **Unique Sentences**

Content Checked for Plagiarism

1. Outcomes of Section 3.1

1.1 Adjust Ratings on a 1-to-5 Scale

The ratings in the dataset were already normalized from 1 to 5.

1.2 Total Number of Users in the Dataset (tnu)

tnu. (43)

1.3 Total Number of Items in the Dataset (tni)

tni. (12)

1.4 Count of Ratings for Every Product

1.5 Selection of Active Users

Three active users were selected based on their missing ratings:

* User U1: Missing 2 ratings.

User [5]

* User U2: Missing 3 ratings.

User [2]

* User U3: Missing 5 ratings.

1.6 Selection of Target Items

Two target items were selected based on their missing ratings:

Item: wuthering

1.7 Count of Users Co-Rating Items with Active Users

For each active user, the number of users who co-rated items with them (No_common_users) was calculated. Additionally, the number of co-rated items for each active user (No_coRated_items) was also determined.

Sample of Count of Users Co-Rating Items with Active Users

1.8 2-D Array for Co-Rating Statistics

1.9 Curve of Ratings Quantity for Each Item

.

10. Maximum Number of Users Meeting 30% Co-Rating Threshold

10. 50% Co-Rating Threshold

Case 1.1

This case includes calculating top closest users for active users using Cosine similarity without bias adjustment and also calculating using discounted similarity

As shown in the output above the discounted similarity method and original cosine similarity the top closest do share 2 commons in each active users and that is because discounted similarity follow another criteria in calculating nearest neighbors. The output of discounted similarity is slightly higher as it gives weight to more relevant users. While both similarity measures offer valid predictions, the Discounted Cosine Similarity method seems to provide slightly higher ratings, which may be more desirable in a recommendation system.

Case 1.2

As shown in output below, the top closest is same in both methods there is a change in prediction values , discounted similarity have a more moderate results

Case 1.3

As shown in output below, the Pearson correlation provides a higher rating giving a more diverse rating that recommender systems prefer but it still includes users that have less common users by which they will still effect the calculation with low common number of items. So as dataset is sparser using discounted similarity will be more accurate

Case 2.1

As shown in the output The Original Cosine Similarity produces higher ratings on average compared to Discounted Similarity .The items considered closest to function differ between the two methods. The Original Cosine Similarity and Discounted Similarity but discounted similarity is more relevant, but with lower similarity scores.

Case 2.2

As shown in output below, The top 20% closest items for each item were identified using both methods were relatively very low but in discounted similarity lower suggesting low influence from the data While mean-centered cosine similarity provides stronger relationships between items, discounted similarity is useful for reducing the impact of outliers or noisy data. The choice between the two depends on whether stronger relationships or more balanced predictions are wanted

Case 2.3

As shown in output below ther is small difference between discounted similarity and Pearson correlation is relatively low, while predictions in pearson withount discounted either very high or very low that toconclude that discounted similarity gives more moderate values

Compare 2.1, 2.2, 2.3
1- Cosine similarity: values are extreme
2- Cosine mean-centered : values has slightly small differences with discounted similarity and has nan –values
3- Pearson correlation: outputs are varied , meaning has more variety and has more lmoderate outputs than Cosine similarity
Compare Part 1 and Part 2
Part 1 and part 2 predict ratings:
User-Based Methods–Part 1: Tend to be a little high for users; using the mean-centered cosine similarity

adjusted the ratings by subtracting the average user rating to make them more bias-adjusted predictions. Significance weighting had a strong impact in user based in all methods as it reduced the extremes

Item-Based Methods-Part 2: Predictions are more moderate, with slight differences in ratings due to different similarity measures: Pearson, Cosine, and Discounted Similarity. For example, the predicted ratings using discounted similarity are lower than the ones based on mean-centered cosine similarity, reflecting the discounting effect applied to high similarities to give better effect .

Closest Items:

User-Based methods show closer relationships for items as items is from the same genre.

Item-Based methods tend to get lower similarity scores, especially in Pearson Correlation and Discounted Similarity, which show weaker relationships, such as 0.17 for "Great Expectations" with Pearson.

Discounting Effect:

Discounted Similarity lowers the similarity scores for the highly related items, thus making the predictions to be more balanced than in the non-discounted methods.

Pearson Correlation produces consistently moderate predictions. The discounted similarity provides a more focused user approach as it focuses on user.

Conclusion

On User-Based and Item-Based collaborative filtering, both methods having their advantages and disadvantages, User-Based, which sort of compares people like mean-centered cosine similarity, determines the associates by which users could guess what you would like. Asking your friends about a movie will help you relate it to them. Item-Based is all about the product itself. Think of it as checking out what other similar items are attracted to. This is where Pearson Correlation and Discounted Similarity come in. Common ground between items is established for say, Pearson Correlation, while Discounted Similarity giving higher weights to users that have more common items to highlight that this user is more relevant than other and giving less weight to less important values

Choosing what's better depends upon the purpose. If maximum association among users/items is needed, then mean-centered cosine similarity or Pearson is preferred. But in case extremes or outliers doesn't affect then cosine similarity can be used ,. It is all about what you want the recommendation system to create

One of the advancements that can be made is dynamic discount factor or using grid algorthim to try multiple discount factors and getting the best

Congrats! Your Content is 100% Unique.