Implementation of New Pearson Correlation Coefficient for Recommendation

Htwe Htwe Pyone Faculty of Computer Science, University of Computer Studies (Myitkyina) htwehtwepyone233@gmail.com

Abstract

Recommender system helps the users to choose something actually want or need. So, this system uses the user-based collaborative filtering method for building movies recommendation system. This system knows the user's interest and recommends items which are particular interest of the user. For this target, choosing appropriate similarity measure is a key to success the recommender system. New Pearson correlation coefficient (NewPCC) is one of the most popular similarity measures for collaborative filtering recommender system, to evaluate how much two users are correlated. By using NewPCC similarity method, this system finds out the recommended movie from all movies.

Keywords: Movie, Recommendation, NewPCC.

1. Introduction

Recommendation systems are useful for e-commerce web sites. This system generates the recommended items list that can be customer's interests. For recommendation, recommender system uses the customers purchased items and the customer interested rates. Furthermore, some system can use demographic data, subject interests and favorite artists. Recommender system is applied to the social parts by sharing user's interest information or item (e.g movies, images, music, news, web pages, scientific literatures, books). Some recommender system uses the user's profile that includes the relevant personal information and user's search history. This information is compared with the search history of different users. By using this different information and user's rating (preference), recommender system predicts the user's interested recommendation lists.

So, this system proposes the recommender system for movies by using new Pearson correlation coefficient (NewPCC) method. This system is also implemented as the user-based collaborative filtering system. The proposed system helps the users to deal with the information overload in the Internet in which numbers Thin Thin Swe Faculty of IT Support and Maintenance, University of Computer Studies (Myitkyina) thinthinswettsmdy6 @gmail.com

of contents are growth every time. Moreover, this system provides the personalized recommendations.

2. Related Work

V. Subramaniyaswamy [1] used collaborative based filtering with the help of correlation techniques to get the expected outcome. Moreover, they compared Pearson correlation, cosine similarity, Jaccard similarity, Spearman rank correlation and Mean-squared distance. This system told about which similarity metrics which is good in the context of book recommendation system.

X. Zhou and H. Song [2] used bi-clustering algorithm to generate bi-clusters that contains only one missing data. Missing data are smoothed by the information of data in this bi-cluster. After finishing the bi-clustering process, they got a denser rating matrix. But, the smoothing data was not as reliable as the original data. So, they described a weighted matrix to distinguish between the smoothing data and the original data. Then, they used the user-based CF in the denser matrix to generate recommendation for the active user.

S. D. Punyasamudram [3] described movie recommender system by using item based collaborative filtering technique. For item based on their neighbor's preferences, collaborative filtering techniques create better suggestions than others. In this technique, they first examined the user item rating matrix and they identified the relationships among various items, and then they used these relationships in order to compute the recommendations for the user.

3. Recommendation

In electronic commerce, recommendation provides the suggestions that prune large information spaces. So, the users are directed toward those items that best meet their needs and preferences [4]. There are three steps in the recommender system. First step obtains preference from customer's input data. Second step computes the recommendation using proper techniques. Finally, the recommender system presents the recommendation results to customers [5]. Recommender systems are usually classified into the following categories [6]:

- Content-based recommendations: It tries to recommend items that are more similar to items appreciated by the target user. Similarity between items is defined based on predefined features that are keywords and textual descriptions for each item [7].
- Collaborative filtering recommendations: It is popular recommendation method that bases its predictions and recommendations on the ratings or behavior of other users in the system. The fundamental assumption is that other users' opinions can be selected and aggregated in such a way that provide a reasonable prediction of the active user's preference [8].

3.1. Collaborative Filtering

Collaborative filtering (CF) is popular in the developing recommender systems. CF recommendation includes item-based approach, model-based approach and user-based approach. These are as follows [9]:

- Item-based collaborative filtering: This approach is to calculate a similarity matrix whose entries correspond to pairwise item similarities.
- Model-based collaborative filtering: This method uses the historic rating data for making recommendations. To derive the hidden model, a variety of statistical machine learning algorithms such as Bayesian networks and neural networks can be employed on the training database.
- User-based collaborative filtering: This approach is based on similarity of user interests. This technique recommends the unseen items already rated by similar users.

3.2. Advantages of CF Recommendation

CF recommendation technique doesn't require contents to be analyzed. It creates user profiles and item profile so it doesn't store huge amounts of term frequency data. User profile are defined by the user's ratings for the items. Item profile consists solely of the item's actual content. So, this technique easier makes the implementation of recommendation system. This technique also provides to improve prediction performance [10].

4. New PCC Similarity Measure

New Pearson correlation coefficient (newpcc) operates by finding users that have rated a number of items in common and by using those ratings to compute a correlation between two users. Equation 1 gives the NewPCC formula of two users u and u'.

newpcc(u,u') =
$$\frac{(|\mu_u \ \mu_{u'}| + \text{max rate}) \text{ meanrate}}{\text{meanrate}}$$
(1)

In the above equation, newpcc (u, u') is the newpcc formula of two users u and u'. The μ_u is the average rating of user u. The $\mu_{\dot{u}}$ is the average rating of user u'. The mean rate is the middle of all rating [11].

5. Proposed System Design

This system is proposed as the movie recommender system by using new Pearson correlation coefficient (NewPCC) based collaborative filtering Method. In this system, the user first must load the movie information for recommendation. According to the collaborative filtering process, this system finds the rated movie items among others. By using rates about movie, this system calculates the similarity between each user by using NewPCC method. The standard rate values are used between 0 and 5. After calculating, this system produces the correlation results between each user. Then, this system predicts the recommended movie list for each user. Finally, this system displays the prediction result for the active user. Figure 1 shows the system flow diagram.



Figure 1 System Flow Diagram

6. Explanation of the System

As a sample, this system is tested by using 10 different film items and 10 users. Film list is shown in Table 1. User list is shown in Table 2. By using rates for each film that voted from each user, this system calculates the similarity between each user. Rate "0" is "no interested" about film. Rate "5" is "strong interested" about film. Rate "2" to "4" is indicated for "intermediate interested". Rate table is shown in Table 3.

No	Film ID	Film Name
1	F01	GoldenEye (1995)
2	F02	American President, The (1995)
3	F03	Nixon (1995)
4	F04	Casino (1995)
5	F05	Get Shorty (1995)

6	F06	Copycat (1995)			
7	F07	Dangerous Minds (1995)			
8	F08	Twelve Monkeys (a.k.a. 12 Monkeys) (1995)			
9	F09	Clueless (1995)			
10	F10	Seven (a.k.a. Se7en) (1995)			

Table 2. User List

No	User ID	User Name
1	U01	Ma Eaint Phoo Nwe
2	U02	Mg Zayar Ko
3	U03	Mg Thiha
4	U04	Mg Thant Sin
5	U05	Ma Phwe Phwe
6	U06	Ma Kyi Phyu
7	U07	Ma May Mi Aung
8	U08	Ma Ei Chawel Ko
9	U09	Ma Shin Lay
10	U10	Ma Kyaw Kyaw

Using NewPCC collaborative filtering method, this system calculates the similarity between each user. Table 4 shows the similarity result about U01 "Ma Eaint Phoo Nwe".

	F01	F02	F03	F04	F05	F06	F07	F08	F09	F10
U01	4	0	0	0	0	0	0	0	0	4
U02	3	0	0	0	3	0	3	0	0	0
U03	3	2	2	3	4	2	0	4	2	5
U04	0	0	0	0	0	0	0	4	0	5
U05	3	3	5	5	3	3	0	3	3	5
U06	3	0	0	0	3	0	0	4	0	4
U07	0	4	4	5	5	0	0	2	0	4
U08	5	0	0	5	3	4	3	0	0	4
U09	3	0	0	4	3	0	3	5	4	5
U10	5	0	0	5	0	0	0	5	5	4

Table 3. Rate Table

No	Users	Similarity
		Result
1	Similarity between U01	-0.21854
	and U02	
2	Similarity between U01	0.12868
	and U03	
3	Similarity between U01	0.23209
	and U04	
4	Similarity between U01	0.21568
	and U05	
5	Similarity between U01	0.0883
	and U06	
6	Similarity between U01	0.30242
	and U07	
7	Similarity between U01	0.30242
	and U08	
8	Similarity between U01	0.30242
	and U09	
9	Similarity between U01	0.09403
	and U10	

Table 4. Similarity Result about U01 "Ma Eaint Phoo Nwe"

For U01 "Ma Eaint Phoo Nwe", this system produces the recommendation film list that includes F04, F05, F06 and F07. Table 5 shows the recommendation list for U01 "Ma Eaint Phoo New".

Table 5. Recommendation List for U01 "Ma EaintPhoo Nwe"

No	Film ID	Film Name
1	F04	Casino (1995)
2	F05	Get Shorty (1995)
3	F06	Copycat (1995)
4	F07	Dangerous Minds (1995)

7. Experimental Result of the System

For experimental result, this system is tested by using 500 movies and 500 users. These testing data are obtained from the "250 MB" movie dataset. These movies data are extracted from the "movie lens" website.

This website address is <u>https://grouplens.org/datasets/movielens/</u>. This dataset includes the movie name and the user rating on these movies. According to the collaborative filtering process, this system uses the 0, 1, 2, 3, 4, 5 rating values from

the user. At the movie lens website, these rate values are already trained and obtained from the user.

This system uses the mean absolute error (MAE) to measure the accuracy of the system. The more the MAE result is obtained, the more the error rate is occurred. Mean absolute error measures the deviation of predictions generated by the recommender system from the true rating values, as they were specified by the user. The MAE equation is as follows:

$$MAE_{i} = \frac{\sum_{j=1}^{n_{i}} \left| ar_{ij} - r_{ij} \right|}{n_{i}}$$
(2)

The count of these film items is n_i , where $n_i \leq n_i$, since users are not required to express their preferences over all n available items. The predictions generated for those film items are r_{ij} , for $j = 1, 2, ..., n_i$, while the actual ratings provided from the user are denoted by ar_{ij} , for $j = 1, 2, ..., n_i$. Then, the MAE for user u_i is computed by first summing the absolute errors of the n_i corresponding actual ratings-prediction pairs.

After this system has calculated the accuracy according to the MSE method, this system obtains the 80% correct rate and 20% error rate. These results are obtained by testing the 500 movies and 500 users. The accuracy of the system is shown in Figure 2.



Figure 2 Accuracy of the Proposed System

8. Conclusion

Recommender systems are powerful for onlinebusiness. The proposed recommender system give the most effective and efficient information to the user according to the user's interest. By using this system, the user can select exactly what they want to see and it is time saving. The proposed recommender system can personalize and customize the web environment towards the user's preference. Finally, this system provides personalized recommendations, contents and services about Films to the users. Personalized services are useful in most E-commerce domains, benefiting both the customer and merchant.

References

[1] V. Subramaniyaswamy, "Neighborhood-based Approach of Collaborative Filtering Techniques for Book Recommendation System", *International Journal of Pure and Applied Mathematical*, vol. 1, pp. 13241-13250, 2018.

[2] X. Zhou and H. Song, "A Collaborative Filtering Recommendation Algorithm based on Bi-clustering", pp. 803-807, *IEEE*, 2015.

[3] S. D. Punyasamudram, "Movie Recommender System using Item based Collaborative Filtering Technique", *IEEE*, 2016.

[4] V. Vekariya and G. R. Kulkarni, "Hybrid Recommender Systems: survey and experiments", pp. 469-473, *IEEE*, 2012.

[5] W. Kangning, H. Jinghua and F. Shaohong, "A Survey of E-Commerce Recommender System", *IEEE*, 2007.

[6] A. Gediminas and T. Alexander, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions", *IEEE Transactions on Knowledge and Data Engineering*, vol. 6, no. 17, pp.734-749, 2005.

[7] A. Pooyan and T. L. Behrouz, "A Collaborative Filtering Recommender System based on User's Time Pattern Activity", pp. 252-257, IEEE, 2013.

[8] P. Phongsavanh and Y. Lasheng, "Movies Recommendation System using Collaborative Filtering and K-means", *International Journal of Advanced Computer Research*, vol. 7, no. 29, pp. 52-59, 2017.

[9] N. W. Khin, "Integrating Rating Styles of Users in Item-based Collaborative Filtering Techniques", *International Conference on Science and Engineering (ICSE)*, 2013.

[10] M. A. Hameed and O. A. Jadaan, Collaborative Filtering Based Recommendation System: A Survey, *International Journal on Computer Science and Engineering (IJCSE)*, vol. 4, no. 5, pp.859-876, 2012.

[11] S. Leily and H. A. Sasan, "A Note on Pearson Correlation Coefficient as a Metric of Similarity in Recommeder System", *IEEE*, 2015.