

Social recommendation algorithm based on collaborative filter algorithms

Jingwen Tan*, Youxin Hu, Jianqiu Luo

School of Computer Science, Sichuan University of Science and Engineering, Zigong 643000, China

* Corresponding author: Jingwen Tan (Email: wyxshls@163.com)

Abstract: For traditional collaborative filter recommendation algorithm technology, this paper combines the collaborative filtering recommendation algorithm with the community division technology of social networks, use the Louvain community to divide algorithms, divide the recommendation users to a community of similar users, and use the collaborative filter algorithm based on the user similarity formula within the community to recommend. In order to verify the effectiveness and accuracy of the algorithm in this paper, based on the introduction of the Douban dataset and the evaluation criteria used, a variety of comparative experiments are carried out on the Douban dataset with a variety of recommendation algorithms to verify the effectiveness of the proposed algorithm

Keywords: Movie Recommendation; Social Networks; Coordinate Filtering Algorithms and Recommendation Algorithms.

1. Introduction

In today's society, the network has surged massive resources, making people dizzy, which brings great challenges to the group using the network, because it is difficult for users to obtain valuable and available information from a large amount of information, and the recommendation algorithm came into being, the recommendation system is a representative of information filtering technology, which is widely used to solve the problem of information overload, identify the goods of interest from a large amount of data and recommend it to users.

In recent years, many experts have found that the relationship between social network users can be seen as a link [1], these users may follow each other based on hobbies or interests, or real friends, so it means that there may be certain similarities and correlations between users and the users they follow, and we can use this similarity to discover the information and implied content we need. In the recommendation algorithm, there have always been problems in data sparsity, cold start and system performance, and considering multi-dimensional factors combined with the user's information in social relations is an important way to improve the accuracy of the recommendation system [2].

Zhou W et al. [3] propose a new recommendation method SRMP based on social network and preference, which integrates social network information and user preferences into the recommendation matrix decomposition, and in order to improve the accuracy of recommendation, SRMP performs recommendation in each independent subcommunity. FE Walter et al. [4] fused and combined temporal information and knowledge graphs, and some of them used the trust relationship between users as social information.

Zhang Xuefeng et al. [5] based on the existing trust-based recommendation system, using autoencoder to optimize the unsupervised initialization feature of user behavior, and proposed a new trust measurement model that integrates user interaction information, preference degree and trust. In 2021, researchers [6] proposed a service recommendation system that utilizes the social relationship between device owners, in which the recommendation is based on the different

relationships between the service requester and the service provider. Experimental results show that in the Internet of Things environment, incorporating users' social relationships into service recommendation can improve the accuracy and diversity of services. Li Gai et al. [7] propose a new social co-ordering recommendation algorithm that combines explicit/implicit feedback, which simultaneously mines the explicit/implicit information in the user score matrix and the social network matrix.

2. Recommendation algorithm

2.1. Community detection algorithm

This paper uses the Louvain community division algorithm proposed by Blondel et al. [8-9], which is a graph algorithm model based on modularity, different from ordinary modularity-based and modularity-based gain algorithms, this algorithm is suitable for large networks, fast division, and can improve the recommendation efficiency when used on the recommendation system, and has the advantages of simple implementation and high execution effect, so this paper uses the algorithm to similarly divide the social relationships of target users.

Louvain's algorithm is a community discovery algorithm based on modularity optimization, and the modularity formula is as follows:

$$Q = \frac{1}{2m} [A_{uv} - \frac{d_u d_v}{2m}] \delta(c_u c_v) \quad (1)$$

A_{uv} is a user relationship adjacency matrix; d_u represents the degree of user u ; $\delta(c_u c_v)$ is Colorola functions, indicates whether users u and v belong to the same community

From formula, the higher the value of Q , the better the effect of community division, and vice versa.

The goal of Louvain's algorithm is to maximize modularity, and the main algorithm steps are divided into the following steps:

Step1: Think of each node in the graph as a community.

Step2: Try to let a node join the neighbor community, calculate their modularity increment ΔQ , try all the neighbor communities again, and select the neighbor community with the largest modularity increment ΔQ to join. The calculation

formula is as follows (2):

$$\Delta Q = \left[\frac{\sum_{in} + k_{i,in}}{2m} - \left(\frac{\sum_{tot} + k_i}{2m} \right)^2 \right] - \left[\frac{\sum_{in}}{2m} - \left(\frac{\sum_{tot}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right] \quad (2)$$

\sum_{in} is the sum of weights within Community C ; \sum_{tot} is community C Sum of all connection weights (including community to external links); k_i is the sum of all link weights associated with node i ; $k_{i,in}$ is Link from node i to the sum of nodes in community C ; m is the sum of all link weights in the network.

Step3: Merge all communities to become a new Supernode with edge weights that sum the weights of all nodes within the original community.

Step4: If the algorithm has reached its goal, then the algorithm ends, otherwise repeat steps 1-3.

2.2. User-based collaborative filtering algorithm

The user-based collaborative filtering algorithm [10] refers to recommending items to the user that the user likes with similar interests. The general recommendation process is to recommend to user 1 who needs to be recommended, find user 2 similar to user 1, and recommend items that similar user 2 likes but user 1 has not touched before to user 1, so as to complete the recommendation.

The detailed steps of the user-based collaborative filtering algorithm are:

Step1: Establish a user-item scoring matrix based on the user's historical behavior data, and map the user's rating data to the elements in the table, as shown in Table 1 below:

Table 1. User-Item matrix

	User 1	User 2	User 3	User 4
Item1	2	4	1	2
Item2	-	3	5	-
Item3	1	-	-	-

Step2: Calculate user similarity according to Table 1 matrix.

Cosine similarity is a method of evaluating the similarity of two vectors by calculating their angular cosine value. In the user-based collaborative filtering algorithm, cosine similarity represents the size of the vector angle between the user vector i and the user vector j , which can be expressed by the following equation 3:

$$sim(i, j) = \cos(i, j) = \frac{i \cdot j}{\|i\| \cdot \|j\|} \quad (3)$$

If both the user vector i and the user vector j are n -dimensional, then cosine similarity can be expressed as the following equation:

$$\cos(\theta) = \frac{\sum_{k=1}^n x_{1k} x_{2k}}{\sqrt{\sum_{k=1}^n x_{1k}^2} \sqrt{\sum_{k=1}^n x_{2k}^2}} \quad (4)$$

Step3: Sort according to the scores of the items in the above formula, and select the N items with the highest rating to recommend to the target user.

The following figure shows the recommendation algorithm process:

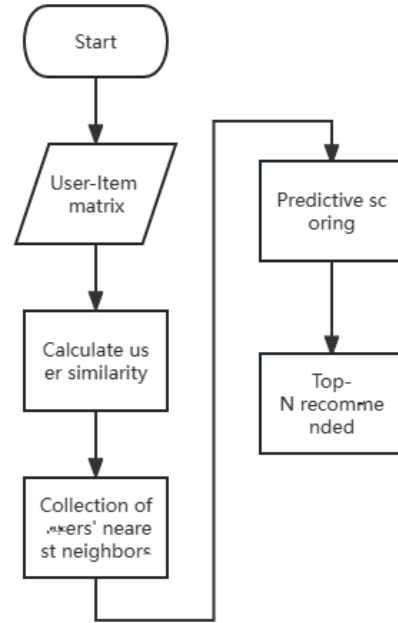


Figure 1. Recommendation Algorithm Process.

3. Results and Analysis

3.1. Introduction to datasets

In order to verify the performance of the algorithm in this paper, the Douban dataset will be used as the experimental dataset.

The Douban dataset (<https://www.librec.net/datasets>) uses a 5-star scoring mechanism. There are 7375 user data and 99746 movie data in the dataset, 278483 ratings, and 111761 social relationships, of which there are a total of 15 movie types in the Douban movie dataset. This is shown in Table 2 below:

Table 2. Douban Movie Dataset

User	Movie	Ratings	Social Relationship	Movie genre
129490	99746	278483	111761	15

This experiment adopts cross-validation technology to avoid the chance of experimental results. The dataset was divided into 5 parts, 80% of the data was randomly selected as the training set, 20% of the data was used as the test set, and the average of the five experiments was taken as the final result of the experiment.

3.2. Evaluation indicators

Since the beginning of the recommendation system research, the evaluation of prediction and recommendation results has been a very important link, and the advantages and disadvantages of a recommendation algorithm are directly reflected in its performance on these evaluation indicators.

Common scoring prediction tasks are generally calculated by RMSE and MAE. See the following formula 5 and 6:

$$MAE = \frac{\sum_{u,i \in T} |r_{ui} - \hat{r}_{ui}|}{|T|} \quad (5)$$

$$RMSE = \sqrt{\frac{\sum_{u,i \in T} (r_{ui} - \hat{r}_{ui})^2}{|T|}} \quad (6)$$

3.3. Model comparison

In order to verify the recommendation efficiency of the

algorithm proposed in this paper, the experiment in this paper will compare with the results of multiple recommendation algorithms similar to this paper, and the recommended evaluation indicators of their respective algorithms are obtained.

(1) UK-Means [11]

The K-means clustering [12] method is used to cluster all users of the recommender system, and then find the similar user collection of the target user in the cluster to which the user belongs, and finally complete the recommendation according to the similar user.

(2) SVD++[13]

The SVD++ is based on the improved algorithm of singular value decomposition, the user-item scoring matrix is used to solve the sparsity problem of data to a certain extent.

(3) Louvain-CF

The algorithm presented in this paper, collaborative filtering recommendation algorithm based on Louvain algorithm

In this experiment, the Douban dataset is divided into five parts, one of which is the test set and the other four copies are used as the training set. Calculate the experimental results on five datasets and take the average to get the final result.

The experimental results are shown in Figures 2 and 3 below and Table 3:

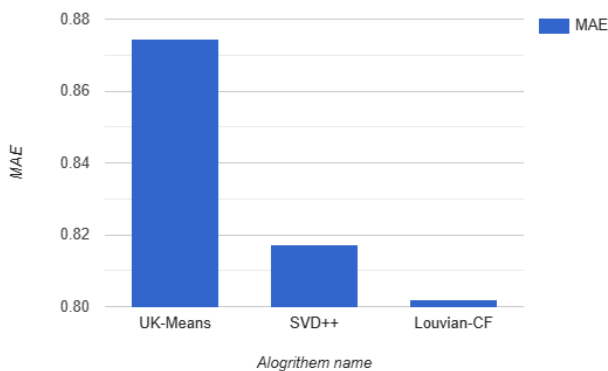


Figure 2. The MAE value compares the results with other algorithms.

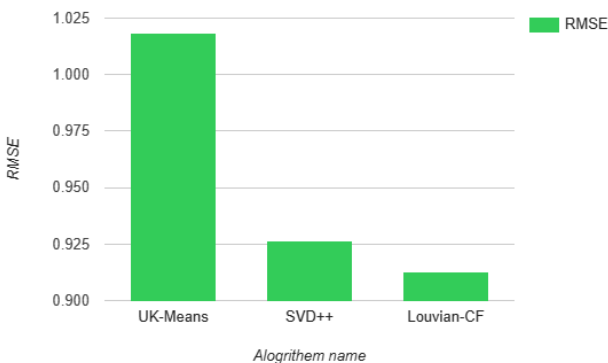


Figure 3. The RMSE value compares the results with other algorithms.

Table 3. The MAE and RMSE value Compare of the experiment

Name	UK-Means	SVD++	Louvain-CF
MAE	0.8746	0.8172	0.8018
RMSE	1.0182	0.9263	0.9126

This experimental result proves that the Louvain-CF

algorithm is superior to the other two algorithms

4. Conclusion

(1) The algorithm in this paper combines the user-based collaborative filtering algorithm with the similarity calculation of the community-based collaborative filtering algorithm, and integrates the two dimensions of user item type preference and score similarity to achieve the purpose of improving user similarity and improving recommendation effect. Finally, using the collection of neighbors in the community, the target unscored items are predicted and the Top-N recommendation list is generated. The experiment was verified on the Douban public dataset, and compared with other algorithms, the experiment verified that the recommendation accuracy and operation effect of the algorithm were better.

(2) This paper combines the community division algorithm with the user-based collaborative filtering algorithm, and uses the Louvain community division algorithm, on this basis, although the sparsity problem and cold start problem are effectively solved, but this paper should consider a variety of community division algorithms, so as to make multiple comparisons, choose the best division method, and need to consider whether the algorithm is suitable for the recommendation of more platforms, and the evaluation index of the prediction score in this paper is better than other recommendation algorithms. However, it can be considered that the algorithm compares with other algorithms in multiple evaluation indicators.

References

- [1] Bedi P, Sharma C. Community detection in social networks [J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2016, 6(3): 115-35.
- [2] Zhang Q,Weng L. A systematic review of social recommendation [J].Computer Engineering and Application, 2020, 56(01):1-10.
- [3] Zhou W , Li J , Zhang M , et al. Incorporating Social Network and User's Preference in Matrix Factorization for Recommendation[J]. Arabian Journal for Science and Engineering, 2018, 43(12):8179-8193.
- [4] FE Walter, Battiston S , Schweitzer F . A model of a trust-based recommendation system on a social network[J]. Autonomous Agents and Multi-Agent Systems, 2008, 16(1):57-74.
- [5] Zhang X,Chen X,et al. Top-N recommendation algorithm integrating user trust and influence[J].Journal of Zhejiang University(Engineering Science),2020,54(02):311-319.)
- [6] Khelloufi A, Ning H, Dhelim S, et al. A Social-Relationships-Based Service Recommendation System for IIoT Devices [J]. IEEE Internet of Things Journal, 8.
- [7] LI Gai, LI Lei, ZHANG Jiaqiang. Social co-ordering recommendation algorithm based on explicit/implicit feedback[J].Computer Applications,2021,41(12):3515-3520.)
- [8] Newman M, Girvan M.Finding and evaluating community structure in networks:, 10.1103/PhysRevE.69.026113[P]. 2003.
- [9] Blondel V D, Guillaume J L, Lambiotte R, et al. Fast unfolding of communities in large networks [J]. 2008.
- [10] Nilashi M, Bagherifard K, Ibrahim O, et al. Collaborative Filtering Recommender Systems[J]. Research Journal of Applied Sciences Engineering & Technology, 2013, 5(16): 4168-4182.

- [11] Kanungo T , Mount D M , Netanyahu N S , et al. An efficient k-means clustering algorithm: analysis and implementation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2002, 24(7):881-892.
- [12] Esfahani M H, Alhan F K. New hybrid recommendation system based On C-Means clustering method; proceedings of the Information & Knowledge Technology, F, 2013 [C].
- [13] Kumar R, Verma B K, Rastogi S S. Social Popularity based SVD++ Recommender System[J]. International Journal of Computer Applications, 2014, 87(14):33-37.