User-based Collaborative Filtering Algorithm Design and Implementation

Hulong Wang\textsuperscript{1}, Zesheng Shen\textsuperscript{1}, Shuzhen Jiang\textsuperscript{1}, Guang Sun\textsuperscript{1}, Ren-jie Zhang\textsuperscript{2}\textsuperscript{*}

\textsuperscript{1}Hunan University of Finance and Economics, Changsha, 410205, China
\textsuperscript{2}Hunan Post and Telecommunication College Network Center, Changsha, 410015, China

\textsuperscript{*}simon5115@163.com

Abstract: With the rapid development of social humanities and technology, especially the development of the Internet, the current Internet era, the amount of information is increasing rapidly with explosive growth rate, the information overload problem is particularly obvious, on the accurate acquisition of information from the massive amount of information users want, from which the personalized recommendation technology was born. In order to solve the problem of acquiring the information users want, this paper researches and discusses a kind of personalized recommendation algorithm - a user-based collaborative filtering algorithm, analyzing the user behavior, comparing the advantages and disadvantages of other related algorithms, using the UserCF algorithm, and optimizing the sparse matrix to reduce the time and complexity of the operation. The algorithm is implemented by software to generate recommendation results. The results of the experimental data in this paper show that the algorithm is effective in recommending projects to users.

Keywords: personalized recommendation; recommendation algorithm; collaborative filtering; user behavior

1. Introduction

There are two ways to solve the problem of information overload in order to get the data that users want accurately\cite{1}: one is a search engine; the other is a recommendation system. The most important part of a good recommendation system is the recommendation algorithm, which determines the performance of the recommendation system and the degree of accuracy of the recommendations, which is a crucial point of the recommendation system.
The work in this paper focuses on user-based collaborative filtering algorithms, introduces the strengths and weaknesses of relevant recommendation algorithms and user-based collaborative filtering recommendation algorithms, and focuses on the algorithmic ideas and specific processes of user-based collaborative filtering algorithms, which are specifically implemented by Python programming software.

2. Related Works

The earliest recommendation system appeared in the 1990s, which to some extent alleviated the problem\cite{2} of information overload caused by the huge amount of information and better solved the problem of information retrieval in the massive information data.

The most important link in the recommendation system is the recommendation algorithm, among which the most well-known and widely used algorithm is the collaborative filtering algorithm, whose ideas are scientific and reasonable, operable, and its recommendation results are also more accurate\cite{3}, so it is favored by many domestic and foreign researchers.

3. User-based Collaborative Filtering Recommendation Algorithm Implementation

3.1. Recommended Models

The user-based collaborative filtering algorithm is to recommend other items that are of interest to other users with similar hobby interests to the target user. The basic idea of the algorithm: the first step is to construct a user-item matrix, which is based on whether the user is interested in a certain item, for example, user A has bought apples, bananas, and pears, and user B has bought bananas and watermelons; the second step is to calculate the set of similar users of the target user by cosine similarity according to the item matrix\cite{4}; finally, the user-item matrix is calculated by K The Nearest Neighbor and User CF algorithm\cite{5} predicts the target user's interest scores for unknown items to generate recommendation results. The flowchart of the user-based collaborative filtering algorithm recommendation model\cite{6} is shown in the following figure.

![Flow chart of user based collaborative filtering algorithm recommendation model](image)

**Fig.1** Flow chart of user based collaborative filtering algorithm recommendation model
3.2. Main processes

The main principle of the user-based collaborative filtering recommendation algorithm[7] is to calculate the similarity of the user's interest through behavioral similarity. The classical algorithm of similarity calculation adopted in this paper: the cosine similarity calculation method. The flowchart of the process of converting textual profile information into numerical calculations is as follows(Fig.2).

![Flow chart of text information conversion](Fig.2)

![Flow chart of user item list](Fig.3)

Cosine similarity calculation using Python programming software with the following key code.

From the above code, it can be concluded that two traversals are performed on the users, and the second traversal is nested inside the first, i.e., each calculation applies cosine similarity to both users to calculate the similarity. Time complexity can be derived. However, when the dataset is very large, the computation is very time-consuming. In reality, most of the users have not behaved towards the same item between two people, i.e., there are very many cases where the above algorithm takes a lot of time to compute this user intersection set of zero. In the spirit of reducing the computation time, the cases
are first computed, i.e., the behavior of the users towards using the same item, and then such cases are divided by the denominator. Make the sparse matrix, with both user u and user v in the inverted list corresponding to K items, and you have =K. Then, traverse the user list for each item, adding 1 to the corresponding between the different two users in the user list, and you end up with a non-zero between all users.

The process flow chart is shown above (Fig. 3).

4. Experimental design and analysis

4.1. Experimental data and environment

The experimental data for this test is the Movielens dataset provided by the Statistical Agency, which contains a large number of users, movies, and ratings of movies, this is a dataset of users rating movies with different levels of ratings. This paper focuses on the TopN recommendation problem in the implicit feedback dataset, so for this dataset, the rating records will not be used for the TopN recommendation[8]. In this experiment, the recommendation solves the problem of determining whether the target user will rate the movie, not the size of the rating.

4.2. Experimental results and analysis

Results of user interest calculations and data presentation via Python.

1) Taking K=2, i.e., when there are 2 neighboring users, the results of the recommended items to each user are as follows.

From the experimental results, it can be concluded that: the items recommended to the user a are q and s. The two movies q and s have the highest interest ratings, and user a is more likely to be interested in q and s. The following table shows all the items that users may be interested in (Tab.1).

| target user | Recommendation results and interest ratings |
|-------------|--------------------------------------------|
| A           | 'q':0.8,'s':0.8                            |
| B           | 'q':1.0,'y':1.0                            |
| C           | 'w':1.1,'t':1.1                            |
| D           | 'b':1.0,'c':1.0,'g':1.0,                   |
| E           | 'k':1.2999                                 |
| F           | 'b':0.8,'p':0.8                            |
| G           | 'a':'q'='d'='y'='i'='r'='z'='p':0.6324555320336759 |
| H           | 'f':0.7,'v':0.7,'s':0.7                    |
| J           | 'f'='g'='j':0.6324555320336759             |
| K           | 'a':1.2,'g':1.2                            |
| L           | 'g':1.2,'t':1.2                            |
| P           | 'g':1.2                                   |

Tab.1 Project maps that users may be interested in
From the above data and the table above

From the above graph, it can be easily seen that the curve trends for both accuracy and recall are almost identical, both increasing and then decreasing, while the curve trend for coverage has been decreasing. In accuracy and recall, when the top 50 users are taken from the neighboring users, the corresponding values of both are maximal, where accuracy max=0.2216 and recall max=0.1209; that is, in this data set, when k=50, the recommendation has the best accuracy. The user CF algorithm is feasible and effective when analyzing the recommendation as a whole, but the highest accuracy is 22.16% when comparing with the current mainstream recommendation system.

5. Summary and outlook of the paper

In this paper, the idea of a user-based collaborative filtering recommendation algorithm is implemented by Python programming software by studying the characteristics of its algorithm and introducing cosine similarity algorithm for calculating user similarity and UserCF algorithm for calculating user interest degree. At the same time, in order to reduce the time complexity of the algorithm, the user-item matrix is inverted into the project-user matrix by the pairwise inverted table method, which greatly reduces the time complexity of the algorithm and shortens the computation time.

Based on the relevant algorithm proposed in this paper and combined with the collected user data, a simple personalized recommendation is designed and implemented to recommend items for users and implement a personalized recommendation service.
References

[1] Liwei H, Bitao J, Shouye L and Yanbo L 2018 J. Review of research on deep learning-based recommendation systems (vol 41, Online Publication) 22 pp4-6

[2] Tsang E P K Foundations of constraint satisfaction (DBLP), 1993 semi-fragile watermarking technique for image authentication (2002 6th Aizu-Ai Conference on Signal Processing, vol 2) pp 1592-1595

[3] Cheng H T, Koc L, Harmsen J et al 2016 Wide & Deep Learning for Recommender Systems pp 7-10

[4] P-N Tan, M. Steinbach, V. Kumar 2005 Introduction to Data Mining (Addison-Wesley)

[5] Liang X 2012 Practice of recommendation system (People's Posts and Telecommunications Press) pp 46-47

[6] Yao Z, Weibin C 2013 Review of research on collaborative filtering recommendation J. Machine Intelligence and Applications 6 pp 4-10

[7] Bo F, Jiujun C. 2012 Collaborative filtering and recommendation algorithm based on multiple similarities among users J. Computer Science 39 pp 23-26

[8] McLaughlin M. R, Herlocker J. L, 2004 A collaborative filtering algorithm and evaluation metric that accurately model the user experience Proc. 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '04) pp 329-336