Research on Recommendation Algorithms Based on Collaborative Filtering

Yongxia Zhao¹, Bo Zhang²*
School of Computer Science, Wuhan Donghu University, Wuhan 430212, China
*Corresponding author’s e-mail: bob.cheung@ovspark.com

Abstract. With the rapid development of Internet industry in the era of Web 2.0, creation and sharing of content becomes easier and easier, and people's accessing methods to information have changed. For information consumers, the exponential growth of information makes users use information inefficiently. Users are eager to get more valuable information for themselves. For information producers, they also need a way to make their information be concerned by more potential users. Collaborative filtering algorithm is the most widely used and successful recommendation algorithm. A good recommendation algorithm can help users to find their favorite videos in massive videos, and even relatively unpopular videos can be pushed to users who may be interested in it.

1. Introduction

In the Internet age of data explosion, the rapid development of the Internet industry in the Web 2.0 era makes it easier to create and share content, and the method people access to information has also changed. It is different from the traditional media such as newspapers, radio, television and other methods and now there are so many methods that people can easily access to various kinds of information through the Internet, such as social platforms, portals, information push and so on. With the increase of the amount of information, people have gradually entered the era of information overload from the era of information scarcity. For information consumers, it is inefficient for them to use information because of the exponential growth of information and they are eager to get more valuable information for themselves. For information producers, they also need a way to make their information be concerned by more potential users. In order to solve this contradiction, personalized recommendation system is widely used in various aspects.

Compared with tools and technologies of other classical information systems (such as search engines), personalized recommendation system is an effective tool to solve information overload. In order to realize the core function of recommendation system, which is to identify the useful items for users, it is necessary to predict the items of recommended value. For achieving this function, the recommendation system must be able to predict the utility of the item, or compare the utility of the item, and then determine the recommended item according to the comparison results.

Recommendation algorithm has been developed for more than 20 years since 1992. With the outbreak of the Internet and the emergence of more data, recommendation algorithm plays an important role in using information. On some mainstream video websites such as YouTube, Youku, Potato, IQI and others, video recommendation system can help users find their favorite videos in massive video and even relatively unpopular videos can be pushed to users who may be interested in it. Since recommendation systems have been widely used in the Internet age and related research in academic field has been very active in recent years, and the annual ACM Recommendation System Meeting...
ICSP 2019
IOP Conf. Series: Journal of Physics: Conf. Series 1237 (2019) 022094
doi:10.1088/1742-6596/1237/2/022094

The ACM RecSys held by the Association for Computing Machinery has become an important communication channel and an important window for top researchers in the field of recommendation systems to share their research results.

The concept of collaborative filtering was first proposed by Oki, Coldberg, Nieols and Terry in 1992. Up to now, collaborative filtering algorithm is the most widely used and successful recommendation algorithm.

2. Relative Algorithm of Recommendation System
The purpose of recommendation system is to connect users' interests and objects, which depends on different elements. At present, the popular recommendation system basically contacts users' interests and items through the following three ways.

![Fig.1 Three kinds of contact user interests and items](image)

The first one is through the items that users like, the system would recommend the items that are similar to the ones they like, that is, the collaborative filtering recommendation algorithm based on the items. The second one is through other users who have similar interests to users, the system would recommend to users what other users like similar to their interests, that is, user-based collaborative filtering recommendation algorithm. What’s more, the last one is through some features connecting users and objects, the system would recommend the items that users like. The features here have different manifestations, such as the attribute set of the items or the implicit semantic vector. This kind of approach is content-based recommendation algorithm.

3. Recommendation Algorithm Based on Collaborative Filtering
Generally, collaborative filtering recommendation can be divided into three types. The first is user-based collaborative filtering. The second is item-based collaborative filtering, and the third is model-based collaborative filtering.

The common point of several collaborative filtering algorithms is to calculate the similarity of users or items. There are several common methods to calculate the similarity.

3.1. Calculation methods of similarity
(1) Euclidean Distance Similarity
Euclidean distance is a formula for calculating Euclidean distance in n-dimensional Euclidean space.

\[ d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \]  

(1)
Where x and Y represent the coordinates of two points, D (x, y) represents the distance between points x and y.

(2) Pearson Correlation Similarity

For two data sets X and Y, their correlation can be expressed by Pearson correlation similarity. The Pearson correlation similarity formula:

\[ r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \]  \hspace{1cm} (2)

Among them, n denotes the number of data points in two data sets X and Y. Xi denotes the number of data points in data set X and Yi denotes the number of data points in data set Y.

(3) Log Likelihood Similarity

The log-likelihood ratio (LLR) method calculates the similarity. For event A and event B, we consider the number of occurrences of two events, with the event matrix shown in the following table:

Table 1. The event matrix

| Event A | Everything but A |
|---------|------------------|
| Event B | K11              | K12              |
| Everything but B | K21 | K22 |

Among them: K11: Number of times events A and B occur simultaneously; K12: event B occurred, event A did not occur; K21: event A occurred, event B did not occur; K22: Event A and Event B did not occur.

Then, the formula for calculating LLR is as follows:

\[ \text{LLR} = 2 \ast (\text{rowEntropy} + \text{columnEntropy} - \text{matrixEntropy}) = 2 \ast (H(k_{11} + k_{21}, k_{12} + k_{22}) + H(k_{11} + k_{12}, k_{21} + k_{22})) - H(k_{11} + k_{21}, k_{12} + k_{22}) \]  \hspace{1cm} (3)

Among them, H represents information entropy.

(4) Manhattan Distance

Manhattan distance has many names: city block distance, L1 distance, L1 norm, Manhattan length. Formula for calculating Manhattan distance:

\[ d_i(x, y) = \|x - y\|_1 = \sum_{i=1}^{n}|x_i - y_i| \]  \hspace{1cm} (4)

Where x and Y represent two n-dimensional vectors.

(5) Spearman Correlation Similarity

The calculation formula of Spearman correlation coefficient is as follows:

\[ \rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \]  \hspace{1cm} (5)

Among them, \( d_i = x_i - y_i \), n denotes the size of the data set. Xi and Yi are data points in the data set X and Y respectively.

3.2. User-based collaborative filtering recommendation algorithm

User-based collaborative filtering mainly considers the similarity between users and users. As long as the items can be found that the similar users like and predict the score that target users comment on the corresponding items. Then we can find some items with the highest score and recommend them to users.

3.3. Item-based collaborative filtering recommendation algorithm

Item-based collaborative filtering is similar to user-based collaborative filtering, but we just turn to find the similarity between items. Only when we find the target user's score on certain items, we can predict similar items with high similarity and recommend some similar items with the highest score to users.

3.4. Model-based collaborative filtering recommendation algorithm

Model-based collaborative filtering recommendation is to train a recommendation model based on sample users' preference information, and then predict and calculate the recommendation according to real-time user preference information.
The recommendation mechanism based on collaborative filtering is the most widely used recommendation mechanism nowadays. It has the following remarkable advantages: it does not need to model objects or users strictly and does not require the description of objects to be machine-understandable, so this method is domain-independent.

The recommendation calculated by this method is open and can share other people's experience. It can well support users to find their potential interests and preferences. However, it also has the following problems:

The core of the method is based on historical data, so there are "cold start" problems for new items and new users.

The effectiveness of recommendation depends on the amount and accuracy of user historical preference data.

In most implementations, historical preferences of users are stored in sparse matrices and there are some obvious problems in the calculation of sparse matrices, including the possibility that a small number of people's erroneous preferences will have a great impact on the accuracy of recommendation.

For some users with special tastes, they can not get good recommendation.

Because it is based on historical data, it is difficult to modify or evolve according to users' preferences after capturing and modeling, which leads to the lack of flexibility of this method.

4. Conclusion

Collaborative Filtering is to analyze the user's personal preference model through the user's historical browsing, viewing, purchasing and other behavior habits, and find similar users or items similar to the designated items in the user group. By synthesizing these similar users, the system can predict the preference of the designated users for this information.

Acknowledgments

This paper was co-supported by the Foundation for Young Scholars in Wuhan Donghu University under grant 2018dhzk006. Thanks for the support of Wuhan Donghu University. Thank you for providing the experimental platform in Wuhan Donghu University. We have completed the implementation of the algorithm on this platform and the experimental data is presented.

References

[1] Steven Bird Ewan Klein Edward Loper. Python Natural Language Processing [M]. Beijing: People's Posts and Telecommunications Press. 2014.6.
[2] Zhou Zhizhua. Machine Learning [M]. Beijing: Tsinghua University Press. 2016.1.
[3] Mitchell Python Network Data Acquisition [M]. Beijing: People's Posts and Telecommunications Press. 2016.3.
[4] Yin Lutong, Yu Jing, Lu Liang, et al. Video recommendation algorithm integrating comment analysis and implicit semantic model [J]. Computer Applications, 2015, 35 (11): 3247-3251.
[5] Xie Fachuan. Modeling of video recommendation system based on data mining [D]. University of Electronic Science and Technology, 2012.
[6] Yin Lutong. Research on Web Video Recommendation System Based on Hidden Semantic Model [D]. Xinjiang University, 2016.
[7] Mao Zhi. Video Recommendation System Based on Hybrid Semantic Matrix [D]. Xiangtan University, 2016.
[8] Duan Xiaokang. Personalized Book Recommendation Based on User Information [D]. South China University of Technology, 2013.
[9] She Zhengwei. Research and Implementation of Text Tendency Analysis System Based on Neural Network [D]. Fudan University, 2013.
[10] Xie Kege. Research on Academic Recommendation Based on Collaborative Filtering [D]. Shanghai University, 2015.