Collaborative filtering recommendation algorithm based on trust relationship in large data

Piao Lisha

Dianchi College of Yunnan University, Department of Science and Engineering, Yunnan province, Kunming, 650228, China

Keywords: Big data; trust relationship; collaborative filtering recommendation algorithm

Abstract: In recent years, the arrival of big data era has brought new opportunities and challenges to collaborative filtering recommendation system. Introducing trust into traditional collaborative filtering algorithm, a collaborative filtering recommendation algorithm based on improved trust is proposed. However, there is still room for research and improvement on how to expand the limited social relations and how to reveal the impact of user interaction on user characteristics. Hadoop, as an open source metaphysical computing platform, realizes the function of cloud computing, which is widely used by researchers. Therefore, the author studies the collaborative filtering recommendation algorithm based on trust relationship in big data. In order to solve the data sparseness problem commonly associated with the collaborative filtering recommendation algorithm, the trust relationship is combined with the traditional collaborative filtering algorithm. Through the transferability of trust relationship, the relationship between trust degree and similarity is used to improve the data sparsity problem, and a collaborative filtering recommendation algorithm based on trust model is formed. However, the algorithm still needs to be improved in terms of time performance. The next step is to combine the clustering algorithm to make the recommendation algorithm further improve the time performance.

1. Introduction

In recent years, with the rapid development of second-generation Internet technologies and various terminals, hundreds of millions of users have been interacting with other users and servers all the time, resulting in explosive data growth [1]. For targeted information needs, information retrieval solves information overload through tools such as search engines, but the premise is that users must clearly understand what information they need. These trust models have better performance in certain given scenarios, but do not dig deeper into other potential factors that may affect trust relationships [2]. For the Internet field where user experience is highly valued, the problem of information overload is a serious violation of this concept, so how to solve this problem has become an important research work [3]. Mainly for the government, enterprises, universities, National University Science and technology park, scientific research institutions and experts and scholars in the process of government, industry, University and research technology docking and collaborative innovation needs [4]. Using collaborative perception technology, artificial intelligence technology, Internet of Things technology, to solve the technology transformation, negotiation, docking, trading, cooperative R&D and other related issues in different regions and industries [5]. With its open concept, innovative model and good user experience, the number of users on the network has increased continuously [6]. However, due to the fact that the basis of its hypothesis is not very consistent with real life, the recommendation results obtained through comparative analysis have not been well improved.

Recommendation algorithms provide services by simulating real life, while traditional algorithms only consider the similarity between user ratings to generate recommendations, and the results are not optimal recommendations [7]. Nowadays, on the one hand, people enjoy rich and colorful information resources, on the other hand, the huge amount of information makes it difficult for us to accurately obtain the information we really need, resulting in the problem of “information overload” [8]. However, in many cases, the demand is not clear or browsed freely. At this time,

  DOI: 10.25236/mmmce.2019.025

search engines can not help users find interesting goods, sound and books in real time and effectively. Information filtering technology represented by recommendation system emerges as the times require [9]. This recommendation often ignores the impact of user trust on recommendations. In the Internet, the dissemination of information is diverse. The user's demand for information is more personalized and diversified. Although search engines have an irreplaceable role in information retrieval, they face various personalities of users. The request appears to be weak [10]. Enterprises need advanced technology to improve their market competitiveness, but technical research requires a lot of financial, material and human support, and SMEs simply cannot afford it. While the resources on the Internet are becoming more and more abundant, and more and more choices are made, the problem of resource fragmentation, lack of integration, and difficulty for users to choose due to excessive growth is the problem of “information overload” of Internet resources.

2. Trust in collaborative filtering algorithms

2.1 Define trust in collaborative filtering algorithms

The definition of trust varies according to individual experience and the context, perspective, and problem to be solved. Trust is a concept consisting of many different attributes, including reliability, dependability, honesty, authenticity, security, strength and timeliness. It needs to be considered and defined according to the specific environment in which trust is located. The collaborative filtering recommendation algorithm mainly uses user-item scoring information, analyzes historical score data, mines user's interest preference, and predicts the user's score on items that have not been contacted. From the perspective of the user as an information consumer, the recommendation system can select content that may be of interest from the massive data as a recommendation according to its characteristic information and behavior, which reduces the pain of the user's confused search and improves the user's satisfaction. Recommendation system can actively find useful information for users according to their own preferences and external information characteristics. Fusion of trust relationship: Target users in recommendation system take into account the objective time and user preference factors, and measure the reliability of user recommendation behavior through direct and indirect trust relationship. In this case, the technology of universities and scientific research institutions is rapidly transformed into productivity, and the technological innovation and development of enterprises are facing certain difficulties.

2.2 CF algorithm based on improved trust

The user scoring matrix is established, which converts the explicit, implicit scoring and evaluation behavior of the user to numerical scoring according to certain rules, and then forms the user project scoring matrix. Mixed recommendation, as its name implies, is to mix two or more recommendation technologies to reduce or eliminate the defects of single technology in order to get better recommendation effect. A good recommendation system can automatically find useful information for users by interacting with users' recommendation time and time again, and make users trust and depend on it. Users can access online services such as technology docking, technology services, enterprise exhibition halls, international cooperation and so on by login technology innovation platform through user terminals. These services effectively solve the above problems through the emergence of personalized recommendation system on the Internet, which can filter out many information resources unrelated to users. And can actively recommend information resources for users according to user history records, thereby quickly helping Internet users to quickly find resources and greatly improving the user experience of the Internet. Although this type of model can make up for the above defects to a certain extent, it still does not help in the face of the sparsely scoring data. Then, the reliability threshold is set, and for the user whose score reliability is lower than the threshold, the trust network is reconstructed by considering the positive factor and the negative factor comprehensively.

Calculating the similarity x between users, an initial user similarity matrix is obtained. This

article uses the Person correlation formula as follows:

$$x_n = ax_{n-1} + c\left(mod\ M\right) \tag{1}$$

When the target user has a neighbor, that is, x > 0, the weight F is calculated according to the similarity and trust of the user. The formula is as follows:

$$F\left(\mathrm{x,y}\right) = \begin{cases} 1, & f\left(x,y\right) \geq T \\ 0, & f\left(x,y\right) < T \end{cases} \tag{2}$$

3. Design and Implementation of Recommendation System Based on Hadoop

3.1 System Requirement Analysis

The film recommendation system based on Hadoop platform in this paper needs to be designed to provide expandable storage capacity to store the ever-growing user-item rating records and other data. And can provide excellent computing power, need to maintain the stability of the system and the system to facilitate maintenance. While providing high quality services, Hadoop, a distributed computing framework, has very low hardware requirements. It only needs a large number of low-cost devices to build clusters to achieve specific applications. Hadoop cluster adopts Master/Slave architecture, which still exists in many distributed systems. The preferences of the technology demand side will be similar to those of the technology demand side with similar interests. The technical requirements of the user's registration login information, browsing information, evaluation information, technical docking behavior, technical transactions and other personal behavior models can reflect their preferences. Then, the Hadoop platform traverses each user's project recommendation information to compare with the in-memory data to see which users' scoring mode is most similar to the current user, and record the ID of the most similar TopN user. Therefore, effectively mining the implicit trust relationship between users from the score data is the key to improving the quality of recommendation. Because the algorithm can provide higher recommendation accuracy and low requirements on the data set, that is, it has universality. Unlike the social recommendation algorithm in Chapter 4, it must be based on social relations, and many recommendation systems cannot. The user is required to have a social relationship from the beginning.

Both MapReduce and Spark computing engines are computational engines that are commonly used in today's applications. To verify that multiprocessor distributed processing data can actually reduce computation time and Spark is more suiTable for iterative tasks. In this paper, a set of computational nodes is designed to calculate the time between the MapReduce implementation and the Spark implementation.

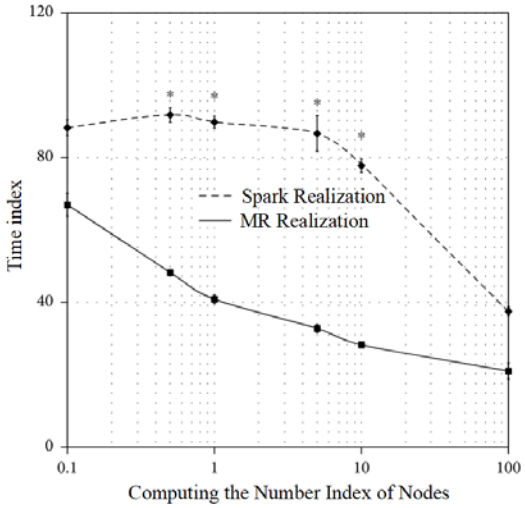

Fig.1. Experimental results of parallelization algorithms on MapReduce and Spark, respectively

3.2 System architecture design

This system uses relational database MySQL to store user and movie information, some recent user history score information, HDFS to store user complete history score information, and the calculated recommendation results are also stored in MySQL database. Hadoop is composed of several sub-projects, including distributed file system HDFS and MapReduce computing model. However, the collaborative filtering algorithm is not perfect, and further improvements are needed to meet the accuracy and speed requirements of the technology recommendation system required by the technology innovation platform. Today's Hadoop platform has been widely used in many software companies, traditional industries and electronic information industries to store and calculate big data. The advantages of HDFS in Hadoop are: strong fault tolerance, the system automatically backs up multiple copies, and when the copy has a problem, the system automatically recovers. Therefore, it is critical to effectively measure the reliability of predictive scores. This paper proposes a corresponding recommendation strategy for trust application. The business layer module is responsible for associating the recommended algorithm layer with the application layer in response to user requests. The application layer provides HTTP network services as modules that interact directly with the user. HDFS provides storage services for large-scale data, while MapReduce provides computational support for large-scale data.

Therefore, it is finally determined to use the recommendation technique based on the item's mode mean to recommend the course information. The experiment is shown in Figure 2 and Figure 3.

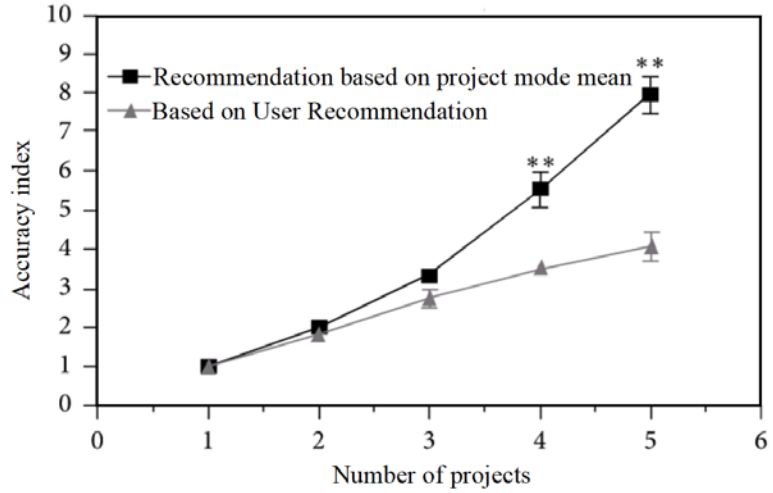

Fig.2. Analysis of the accuracy calculation results of two algorithms

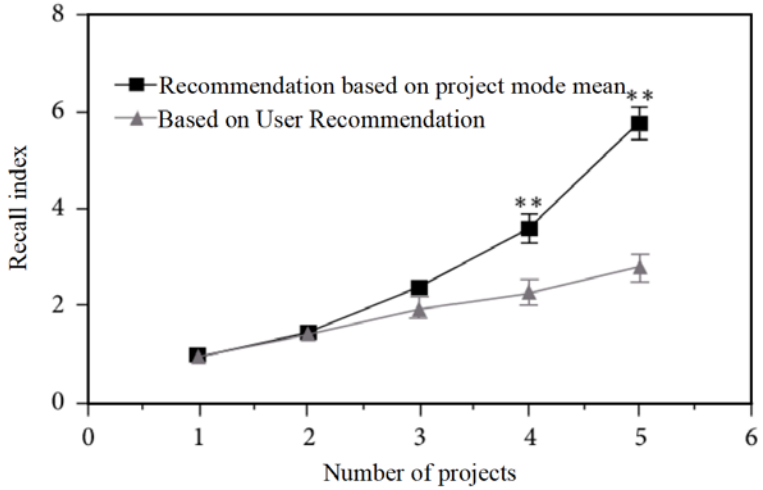

Fig.3. Analysis of recall rate calculation results of two algorithms

4. Conclusion

Collaborative filtering recommendation system has always maintained a high research enthusiasm, and has been widely used in the commercial field, especially in the e-commerce industry, bringing huge commercial wealth. Collaborative filtering technology is not confined to the field. It has strong versatility and principle of life. It is widely used in various fields. However, it is facing many challenges because of the impact of large data. Especially, the large amount of data is sparse, which brings a test to the recommendation effect and computational efficiency. Compared with the traditional recommendation algorithm, the accuracy of recommendation has been greatly improved. The specific steps of the algorithm are described in detail, including the establishment of user trust model, the calculation of mixed weights, the generation of nearest neighbors, score prediction and so on. The effectiveness of the improved algorithm is verified by experiments. For example, the rapid growth of network resources has made it difficult for network users to extract truly effective information resources from massive information. At the same time, the algorithm proposed in this paper still has some shortcomings: the recommended coverage has not been significantly improved. How to better use the scoring information to optimize the algorithm model to achieve the effect of improving recommendation accuracy and coverage at the same time is the focus of the next research work. The experimental results show that it is feasible to introduce the trust factor into the collaborative filtering algorithm, and the recommended result is significantly improved. And the trust degree calculation method in this paper is also flexible and adapTable, which greatly improves the accuracy of the recommendation algorithm.

Acknowledgement

This paper was supported by Yunnan Province Department of Education Science Research Foun dation Research on the Recommended Algorithm of Collaborative Filtering with Large Data Techno logy (2019J0828).

References

[1] Chen M, Mao S, Liu Y. Big Data: A Survey[J]. Mobile Networks and Applications, 2014, 19(2):171-209.

[2] Fan J, Han F, Liu H. Challenges of Big Data analysis[J]. National Science Review, 2014, 1(2):293-314.

[3] Gandomi A, Haider M. Beyond the hype: Big data concepts, methods, and analytics[J]. International Journal of Information Management, 2015, 35(2):137-144.

[4] Assuncao M D, Calheiros R N, Bianchi S, et al. Big Data Computing and Clouds: Trends and Future Directions[J]. Journal of Parallel & Distributed Computing, 2013, 79-80:3-15.

[5] Kitchin, Rob. The real-time city? Big data and smart urbanism[J]. GeoJournal, 2014, 79(1):1-14.

[6] Lv Y, Duan Y, Kang W, et al. Traffic Flow Prediction With Big Data: A Deep Learning Approach[J]. IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, 2015, 16(2):865-873.

[7] Specht, D. Book Review: The Data Revolution: Big Data, Open Data, Data Infrastructures and their Consequences[J]. Media, Culture & Society, 2015, 37(7):1110-1111.

[8] Pagare R, Patil S A. Study of Collaborative Filtering Recommendation Algorithm Scalability Issue[J]. International Journal of Computer Applications, 2014, 67(25):10-15.

[9] Xiaojun, Liu. An improved clustering-based collaborative filtering recommendation algorithm[J]. Cluster Computing, 2017, 20(2):1281-1288.

[10] Wang X Y, Liu H. Collaborative Filtering Recommendation Algorithm Integrated into Co-Rating Impact Factor[J]. Advanced Materials Research, 2014, 926-930:3004-3007.