

Research on Recommendation System Based on Massive Data

Fengqin He

Hohhot Vocational College, Hohhot, Inner Mongolia, 010051

Keywords: recommendation system; massive data; research prospect

Abstract: When dealing with big data, the recommendation performance of traditional recommendation systems, such as conventional collaborative filtering, is limited. The use of easy-to-use K-means clustering algorithm combined with collaborative filtering constitutes a recommendation algorithm. This paper uses genetic algorithm to optimize the combined recommendation algorithm, simplify the combined recommendation algorithm, and reduce the complexity and cost of the combination algorithm. At the same time, the genetic algorithm is improved to enhance the optimization ability of the genetic algorithm and improve the performance of the recommendation system. Finally, the performance of the proposed algorithm is tested based on MovieLens movie scoring dataset. The results show that the optimization ability of the genetic algorithm is improved and the performance of the recommendation system is improved.

1. Introduction

With the rapid development of science and technology and information technology, the society has entered a new era of highly informatized information. The Internet is everywhere, affecting all aspects of human life and completely changing people's way of life. In particular, since the Web 2.0 era, with the sudden emergence of social networking media, Internet users are both consumers of network information and producers of Internet content. The amount of information on the Internet has grown exponentially. Due to the limited ability of the user to distinguish, he often feels that he has no way to deal with large and complex Internet information, making the cost of finding useful information on the Internet huge, resulting in the so-called "information overload" problem. The generation of search engines and recommendation systems provides very important technical means for solving the problem of "information overload". For search engines, when searching for information on the Internet, users need to enter "search keywords" in the search engine. According to the user's input, the search engine performs information matching in the background of the system and displays information related to user queries. user. However, if users can't think of keywords that accurately describe their needs, search engines will be powerless. Unlike search engines, the recommendation system does not require the user to provide explicit requirements, but rather models the user's interests by analyzing the user's historical behavior, thereby actively recommending to the user information that may satisfy their interests and needs. Therefore, the search engine and recommendation system are two complementary tools for the user. The former is active while the latter is passive.

2. The Generation and Development of Recommendation System

After the recommendation system was born, the academic community paid more and more attention to it. Since 199, the American Computer Society holds an annual ACM Conference on Electronic Commerce (ACM EC), and more and more papers related to the recommendation system are published on the ACM EC. The ACM Special Interest Group of Information Retrieval (ACM SIGIR) began to use the recommendation system as an independent research topic for the conference in 2001. The 17th International Joint Conference on Artificial Intelligence also held the recommendation system as a separate topic. In the recent 10 years, the academic community has paid more and more attention to the recommendation system. So far, major international conferences in databases, data mining, artificial intelligence, and machine learning (such as

SIGMOD, VLDB, ICDE, KDD, AAAI, SIGIR, ICDM, WWW, ICML, etc.) have published a large number of research results related to recommendation systems. . At the same time, the first international conference named after the recommendation system ACMRecommender Systems Conference (ACM RecSys) was first held in 2007. In the KDD CUP competition held in recent years by the Data Mining and Knowledge Discovery International Conference (KDD), the contest topics for two consecutive years are recommendation systems. In the KDDCUP 2011 competition, the two competition titles were "music score prediction" and "identify whether the music was scored by the user." In the KDD CUP 2012 competition, the two contest titles were "friend recommendation in Tencent Weibo" and "predicted CTR in computing advertisements."

3. Field Requirements and System Architecture of Recommendation System

The recommendation system has been widely used in many fields, such as news recommendation, microblogging recommendation, book recommendation, movie recommendation, product recommendation, music recommendation, restaurant recommendation, video recommendation and so on. Recommendation systems in different domains have different data sparseness and have different requirements for the scalability of the recommendation system and the relevance, popularity, freshness, diversity, and novelty of the recommendation results. The comparison of the requirements of the recommended systems in different areas is shown in Table 1. Although the requirements are not the same, a complete recommendation system usually includes four parts: data modeling, user modeling, recommendation engine, and user interface.

The data modeling module is responsible for preparing the item data to be recommended, expressing it as a data form favorable for analysis, determining the candidate items to be recommended to the user, and classifying and clustering the items for preprocessing. The user modeling module is responsible for analyzing the user's behavioral information in order to obtain the user's potential preferences. The user's behavior information includes questions and answers, ratings, purchases, downloads, browsing, collections, and time of stay. The recommendation engine module utilizes the back-end recommendation algorithm to filter out the items of interest from the candidate items in real time and recommend them to the user in the form of a list after sorting. The recommendation engine is the core part of the recommendation system, and it is also the part that consumes the most system resources and time. The user interface module bears the functions of displaying recommendation results and collecting user feedback. In addition to basic requirements such as reasonable layout, beautiful interface, and easy use, the user interface should also help users to provide feedback initiatively.

4. Recommendation Engine

The basic recommendation methods of the recommendation engine may be classified into content-based recommendation and collaborative filtering-based recommendation. The basic principle of the content-based recommendation method is to select other similar items as recommendation results according to the user's favorite items [2]. For example, if you now have a new movie that has the same actor or subject matter as a user's past movie, the user may like the new movie. The user model's vector features are often used to describe the user's interests and hobbies, and for each item, feature extraction is performed as the content model's content features. Then, the matching degree between the vector feature of the user model and the vector feature of the candidate item model is calculated, and the candidate object with a higher matching degree can be pushed as the recommendation result to the target user. Collaborative filtering, proposed by David Goldberg in 1992, is the most successful and widely used technology in personalized recommendation systems. The well-known foreign commercial websites such as Amazon, the well-known domestic watercress network, shrimp network and other sites, have adopted a collaborative filtering method. Its essence is based on the technology of association analysis, which uses the common preferences of the user's group to recommend to users. Collaborative filtering utilizes the user's historical behavior (preferences, habits, etc.) to cluster users into clusters. This

recommendation is calculated by similar users, assuming that the current user is also interested in items that are similar to other similar users. The collaborative filtering recommendation method usually includes two steps: finding a set of users (users' groups or clusters) that are similar to the target user according to the user behavior data; finding items that the user likes in the set and the target user has not purchased. In practical use, collaborative filtering technology faces two major constraints: one is data sparseness, and the other is cold start-up problems. Collaborative filtering requires the use of relevance between the user and the user or the item and the item for recommendation. The most popular memory-based collaborative filtering approach is based on the neighbor relationship approach. The method first finds out the neighbors of the user that are similar to the specified user evaluation history, and predicts the results according to the behavior of these neighbors or finds items similar to the query items. The premise of doing this is that if two users have similar ratings on a group of items, then they will have similar ratings for other items; or if the two items have similar ratings on a group of users, then they will also have similar evaluations for other users. The key to collaborative filtering algorithms is to find the nearest neighbor of the user (item). When the data is sparse, the items purchased by the users are difficult to overlap, and the effect of collaborative recommendation is not good. One of the improvements is that in addition to direct neighbors, the behavior of indirect neighbors can also influence the current user's decision-making behavior. In addition, some methods to solve the sparse problem can add some default values, artificially make the data more dense, or use iterative completion methods, first add some values, and then add other values on the basis. In addition, there is the use of migration learning methods to make up for data sparseness. However, these methods can only partly solve the problem of data sparseness and cannot completely overcome them. In real applications, due to the large scale of data, the problem of sparse data becomes more prominent. Data sparsity limits the effectiveness of collaborative filtering methods. Identifying algorithms that match the sparseness of data in order to be able to make the right choice based on the specific application is a very valuable research topic.

The model-based method predicts results through a parametric model that fits the training set. It includes cluster-based CF, Bayesian classifier, regression-based method. The basic idea of a clustering-based approach is to cluster similar users (or items). This technique helps solve data sparsity and computational complexity issues. The basic idea of Bayesian is to calculate the conditional probability of each possible rating value (such as 1 to 5 points in the movie recommendation) given the user A's other ratings and other user ratings, and then choose a rating with the largest probability value. As a predictor. The basic idea of the regression-based approach is to first use a linear regression model to learn the relationship between the scores of items, and then predict the user's score on the items based on these relationships. The Slope-one algorithm uses a linear model on the evaluation matrix to enable it to quickly calculate results with relatively good accuracy. The most recent type of successful model-based method is based on low-rank matrix decomposition. For example, SVD and SVD++ decompose the evaluation matrix into three low-rank matrices. The product of the three matrices can restore the original matrix to some extent, so that missing values can be evaluated. The other method is non-negative matrix factorization. The difference is that the result of matrix decomposition must not be negative. A method based on low-rank matrix decomposition extracts a set of potential (hidden) factors from the scoring matrix and describes the users and items through these factor vectors. In the movie field, these automatically identified factors may correspond to common labels for a movie, such as style or type (drama or action), or may not be interpretable.

The single database storage technology of the traditional recommendation system is no longer applicable under large data sets. There is an urgent need for a storage architecture that provides a unified interface externally and uses multiple mixed mode storage to satisfy the storage of various data files under a large data set. In addition, the traditional recommendation system adopts the calculation method of the single node in the recommendation algorithm, which cannot meet the calculation requirements of the large data set generated by the massive users under the large data set. The complexity, uncertainty and emergence of big data itself also bring many new challenges to the

recommendation system. The time efficiency, space efficiency and recommendation accuracy of traditional recommendation systems all encounter serious bottlenecks.

5. Conclusion

Although people are highly concerned about the changes in the volume of data, they should meet this challenge through collaboration. The conclusion of this paper is that the recommendation system is still in the stage of big data being immature, and providing real-time recommendations still faces challenges, such as weighing accuracy and coverage, lack of data, unpredictable items, changing user settings, and adaptability. . Not only must the evaluation indicators be listed in the development process, but the 4V dimension (capacity, speed, diversity, and authenticity) of big data should also be kept in mind to find a more comprehensive recommendation system solution.

References

- [1] Li Heping, Chen Yuxin. Based on LabVIEW industrial robot automatic sorting system design [J]. Machine Tools and Hydraulics, 2015, 43 (9): 90 - 93.
- [2] Wang Hongbo, Li Jiandong, Cui Xiaohui, et al. Design of group control communication system for sorting production line based on industrial robot[J]. Manufacturing Technology & Machine Tool, 2016(3): 93-98.
- [3] Xu Hailing, Wu Hao, Li Xiao Dong, et al. Inter-network recommendation system compared research [J]. Journal of Software, 2009(2):350-362
- [4] Liu Jianguo, Zhou Tao, Wang Binghong. Personalized recommendation system research progress [J]. Progress in Natural Science, 2009(1): 1-15
- [5] Wei Ruibin. H-b index application research [J]. Intelligence Theory and Practice, 2011, 34(3): 66-69.