

Recommendation Algorithm of Cloud Computing System Based on Random Walk Algorithm and Collaborative Filtering Model

Feng Zhang¹, Hua Ma, Lei Peng and Lanhua Zhang

Abstract

The traditional collaborative filtering recommendation technology has some shortcomings in the large data environment. To solve this problem, a personalized recommendation method based on cloud computing technology is proposed. The large data set and recommendation computation are decomposed into parallel processing on multiple computers. A parallel recommendation engine based on Hadoop open source framework is established, and the effectiveness of the system is validated by learning recommendation on an English training platform. The experimental results show that the scalability of the recommender system can be greatly improved by using cloud computing technology to handle massive data in the cluster. On the basis of the comparison of traditional recommendation algorithms, combined with the advantages of cloud computing, a personalized recommendation system based on cloud computing is proposed.

Keywords: Random walk algorithm, collaborative filtering model, cloud computing system, recommendation algorithm

Introduction

With the rapid development of network technology and the increasing popularity of Internet applications, more and more people to participate in online shopping, social networking, online movie rental and other network activities, Internet for users to provide information resources also increases exponentially, the formation of a resource. In this context, the information retrieval technology the rapid development of it, the search engine has become a main channel for users to get information, but at the same time, the technology also has some defects. Because in the information retrieval process, the user often by keyword search the information they want, but the search engine of the same keyword feedback search results also, can provide personalized service for the user's preferences and interest, and often through the information retrieval results to a huge number, a considerable part of the results do not meet the needs of users.

Recommendation system is an effective tool to solve the above problems. The main objective is to produce and provide meaningful recommendation information to a group of users who may be interested in the product or project. Compared with traditional search tools and technologies such as search engines, recommender systems have been studied relatively late and began to become a relatively independent research field around the middle of the 1990s that has made great progress in the research and application of a short span of 20 years, especially in the field of electronic commerce. With the rapid growth of the variety and quantity of goods, and vary the quality and price of goods, customers often need to spend a lot of time and effort to

¹School of Medical Information Engineering, Taishan Medical University, Taian, China, 271016

find their own love of the goods. Therefore, the electronic commerce website in order to meet the needs of customers, to design efficient personalized recommendation system, based on customer preferences or needs, providing personalized decision support and information services for the customers to choose goods.

At present, there have been many commercial recommendation systems have been widely used in practice, they provide users such as commodity recommendation, recommend new products, with the same interest and associated user recommendations, the representative ones are Amazon, eBay, GroupLens, Netflix and Moviefinder. At the same time, in distributed computing, grid computing, parallel computing and storage network is generated based on the cloud computing, through the network to a plurality of computing entities with low cost, integrated into a powerful computing ability in distributed system.

The construction of cloud computing has the obvious advantage of personalized recommendation system based on data mining can be carried out with the powerful data processing ability of cloud platform; parallel computing environment using powerful low cost, reduce the cost of data processing and recommendation system on high performance server. Therefore, in the traditional personalized recommendation algorithm based on a comparative study, put forward the cloud based personalized recommendation system and the system can adopt different recommendation strategies according to different recommendation requirements, and can handle large data to provide good support.

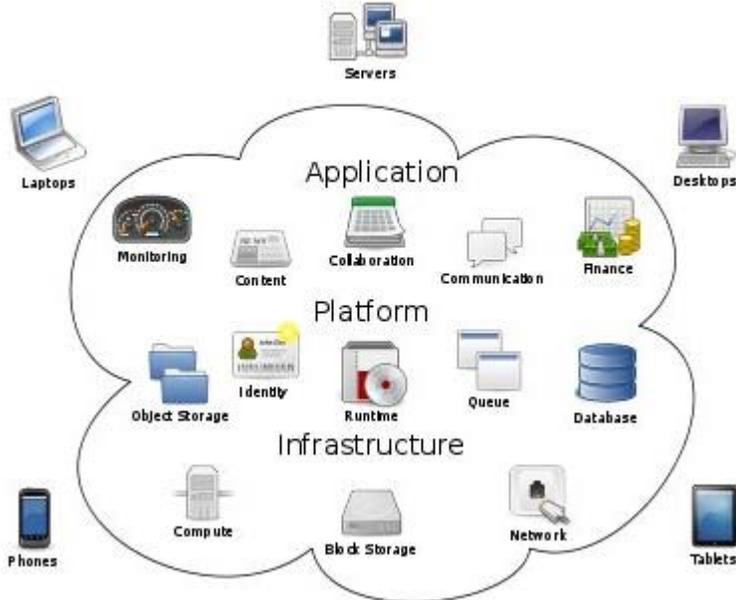


Figure 1. Cloud computing system.

The Proposed Methodology

Random walk algorithm. Random walk is a mathematical statistical model consisting of a series of trajectories. The existing recommendation methods can be divided into two categories: content based recommendation and collaborative filtering recommendation. Based on the recommendation method of content to users of previously visited commodity analysis, unknown goods and similar recommended to the user, this method is mainly for goods were analyzed, and then compared with the unknown commodity to find similar goods. The basic

idea of collaborative filtering recommendation method comes from word of mouth marketing, which uses existing users' preferences to predict the preferences of a consumer for a product. Collaborative filtering recommendation method is an effective recommendation method, and has been adopted by many e-commerce websites.

- Simple random walk refers to the random selection of every wandering node, usually random walk, which means simple random walk. Random walk based algorithms have been widely used in network research. Random walk has the advantages of locality, simplicity, low load and robustness to structural change.

Recommender systems recommend information and products that users are interested in according to their interests, features and buying behavior. With the expansion of the scale of e-commerce, the rapid growth of the number and types of goods, customers need to spend a lot of time in order to find the goods they want to buy. This process of browsing large amounts of irrelevant information and products will no doubt continue to drain consumers who are overwhelmed by the problem of information overload. In order to solve these problems, recommender systems came into being. Recommender system is a kind of advanced business intelligence platform based on massive data mining, which can help e-commerce website provide complete personalized decision support and information service for its customers shopping.

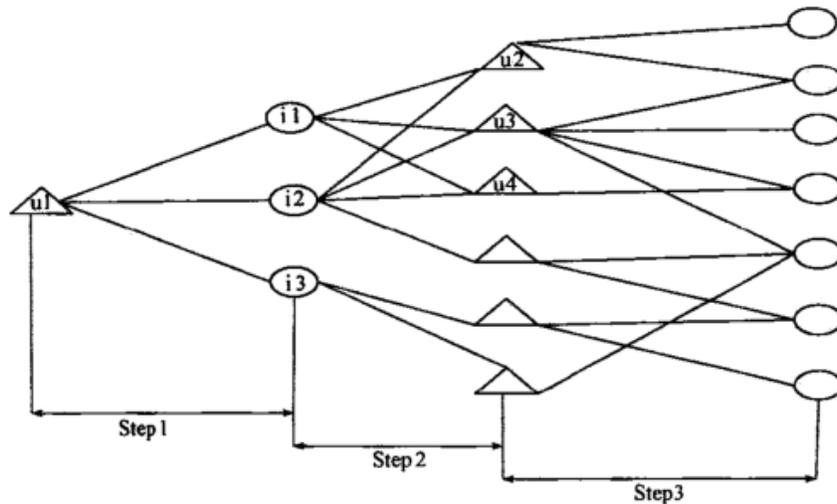


Figure 2. Random walk algorithm.

Collaborative filtering model. Collaborative filtering recommendation technology is currently the most widely used and most successful, its core idea is to find a correlation between users or users of project item score or preference, and then according to the recommendation of correlation. However, with the rapid development and application of network, based on the traditional user or based on existing data sparsity and cold start projects in the collaborative filtering algorithm and other issues, which directly led to greatly reduce the quality of the recommendation. In order to solve these problems, some researchers introduce social information into recommendation, using social network user clustering and social matrix decomposition technology.

The training process based on clustering recommendation method is relatively time-consuming, and the matrix decomposition technique is poor in explaining the recommended results. At present, the typical trust models and inference algorithms of social

networks existing in the Advagato algorithm and Appleseed algorithm for the user entry, advanced and high order is divided into 3 different types, each type of trust relationship between the strength of the former is not the same, the network flow model to calculate the user trust value, the latter used to activate computing the trust value diffusion mechanism. The Tidal Trust algorithm calculates the trust value between the source and the target user in a sequential search based on the breadth first search. The research of Massa and Avesani showed that the introduction of user trust network in recommendation system can effectively alleviate data sparsity and cold start problems. Jamali proposed a random walk model, trust and collaborative filtering method based on project based on the combination, to some extent alleviate the data sparsity and cold start problem in recommendation system, it also has a higher quality of recommendation.

Trust model. In different disciplines, the concept of trust has a different definition, usually most are used to describe an entity to another entity's ability, intent and reliability of recognition. Farag has made the following definition of network trust: trust is the firm belief of an entity to another entity to do the required capacity, and this belief is changing with the entity behavior and the change of time. With the rise of social networks in recent years, more and more scholars have paid more attention to the study of the relationship between trust and accountability in social networks. This paper builds trust model according to the characteristics of trust in social network.

Cloud computing recommendation system. Recommendation system not only pays attention to web queries and web pages and coarse granularity association and sorting, but also study the user model and user preferences, and has the ability to calculate the use of social network. Personalized recommendation system dominated by users of information retrieval and user judgment process of transforming system leading to browse and guide users to access the recommended information oriented process required results. To achieve the personalized recommendation system is based on the recommendation system based recommendation algorithm. At present common recommendation system in the collaborative filtering algorithm based on algorithm and hybrid algorithm recommended recommendation algorithm, the content of association rules. Through reading relevant documents, the recommendation algorithm can be further evidence the specific characteristics of the subdivision algorithm, such as collaborative filtering algorithm can be divided into user based recommendation algorithm, model based recommendation algorithm, such as user based algorithm, also known as storage based algorithm or neighbor based algorithm.

Content based recommendation algorithm mainly comes from information retrieval and information filtering technology, according to the correlation between the content of the project and user preferences, to user generated recommendation. By using the method of users in the past on the project's preference (user profile) the establishment of user characteristics, and the contents of the project are analyzed, the correlation between the assessment project and user interest. And then judge and find out the user may be interested in the project is recommended to the user. Collaborative filtering recommendation system is the first generation has been put forward and the recommendation system is widely used, its core idea is the use of user history information to calculate the similarity between users, and then use the high similarity with the target user's neighbors to evaluation of other projects the target user preference prediction to the specific project. The biggest advantage of collaborative filtering system is no special requirements on the recommendation object, Capable of handling non text structured objects such as music, movies, etc.

Customized basis for recommendation strategies. The study found that now most recommender system based on single recommendation algorithm and recommendation strategy,

therefore, the system in personalized and adaptive performance is obviously insufficient, usually according to the specific needs of specific application scenarios establish recommended strategy. Therefore, in the design and development of the recommender system, it is necessary for the current e-commerce recommendation system flexible and changeable business needs, formulate feasible recommendation strategies. Generally speaking, the variable content may face a recommendation system mainly includes 3 aspects: the implementation of specific customer application scenarios; system core recommendation algorithm and strategy; the type and format of the data source.

Based on the analysis of the advantages and disadvantages of different recommendation algorithms and application scenarios, recommendation system based on cloud computing, personalized recommendation system, to recommend specific strategies customized recommendation system is mainly based on the size of data, when the data is small can be used directly to the traditional personalized recommendation algorithm, and when facing the data volume is huge, personalized recommendation algorithm directly uses generally difficult to achieve good performance. Therefore, the data can be sent to each node of the cloud platform is calculated, so as to realize the calculation of multi node distributed data based on cloud computing.

Conclusion

Recommendation system has been developed for nearly twenty years, and it is becoming more and more important in many fields, such as e-commerce, social networking and so on. But in the face of today's big data environment, the original algorithm exposes many insurmountable shortcomings. Relying on cloud computing technology, this paper analyzes the ItemCF algorithm, and solves the scalability and real-time problems of collaborative filtering. At the same time, the integration of cloud recommendation engine into a commercial system and its successful application proved the feasibility of combining cloud computing with recommendation system. In this paper, a random walk method based on trust factor is proposed. This method not only considers the user's scoring information, but also considers the trust information between users.

Acknowledgement

This paper is financially supported by: 1. Shandong Province Science and Technology Project of Higher Education (J15LE12); 2. Shandong Province Science and Technology Development Project (2014GGX101020); 3. Safety Production of Major Accidents Prevention and Control of Key Technology Projects (shandong-0021-2015AQ).

References

- [1] Luo, Huimin, et al. "Drug repositioning based on comprehensive similarity measures and Bi-Random walk algorithm." *Bioinformatics* (2016): btw228.
- [2] Zheng, Jianjun, et al. "Random-walk algorithm for chloride diffusivity of concrete with aggregate shape effect." *Journal of Materials in Civil Engineering* 28.12 (2016): 04016153.
- [3] Lee, Sehyung, et al. "Robust stereo matching using adaptive random walk with restart algorithm." *Image and Vision Computing* 37 (2015): 1-11.

- [4] Wang, H., & Wang, J. (2014, November). An effective image representation method using kernel classification. In *Tools with Artificial Intelligence (ICTAI), 2014 IEEE 26th International Conference on* (pp. 853-858). IEEE.
- [5] Nicosia, Aurélien, et al. "A general hidden state random walk model for animal movement." *Computational Statistics & Data Analysis* 105 (2017): 76-95.
- [6] Peng, Wei, et al. "Applications of Random Walk Model on Biological Networks." *Current Bioinformatics* 11.2 (2016): 211-220.
- [7] Zhang, S., Wang, H., & Huang, W. (2017). Two-stage plant species recognition by local mean clustering and Weighted sparse representation classification. *Cluster Computing*, 1-9.
- [8] Suciu, Nicolae, et al. "A coupled finite element–global random walk approach to advection-dominated transport in porous media with random hydraulic conductivity." *Journal of Computational and Applied Mathematics* 246 (2013): 27-37.
- [9] Bi, C., Wang, H., & Bao, R. (2014, November). SAR image change detection using regularized dictionary learning and fuzzy clustering. In *Cloud Computing and Intelligence Systems (CCIS), 2014 IEEE 3rd International Conference on* (pp. 327-330). IEEE.