Fast and Efficient Rule Based Recommendation System

Apurva Ashri
Student, Department of Computer Science Engineering, Galaxy Global Group of Institutions, Dinarpur, India

***Abstract*: The main idea of collaborative filtering (CF) methods is to recommend items to a user by taking into account the rates on those items made by other users. CF systems work by collecting user feedback in the form of ratings and exploit similarities and differences among profiles of several users. According to this feedback can be given explicitly, ratings or annotation, or implicitly such as the time spent in examining the content of the recommendations. Although CF has proved to provide good recommendation it presents some troubles. One of the major disadvantages of collaborative filtering is that it cannot recommend new items since the new items don't have any rate. Another disadvantage of CF is that users need to be involved a lot with the system and provide a lot of rates. This is a disadvantage because it is hard to obtain reliable information from many users. In general, current content-based systems have not yet had significant impact on society due to an inability to bridge the semantic gap between computers and humans. It is hard to extract features from the audio signals that have a truly human meaning. Furthermore, the work presented in states that users put a great value in online music reviews, ratings, recommendations, etc. It seems that the information needs of people are quite related with social behaviour and not only based on content or features of music. This thesis will try to solve the main disadvantages of content-based and collaborative filtering recommendation methods. With the purpose of overcoming these disadvantages a hybrid content-based collaborative filtering recommendation system will be built. This hybrid system should possess the best characteristics of both methods and will produce better results than each method individually. To achieve these goal, we will also present research on current content-based and collaborative methods in such a way that we will be able to find out advantages and disadvantages of them.**

***Keywords*: Collaborative filtering, Recommendation**

1. Introduction

In the last years the widespread of several technologies have changed the way people manage, access, and distribute multimedia content. Technologies such as the development and dissemination of P2P networks, and the increase in storage capacity of portable devices had special effect in the worldwide diffusion of multimedia content. Among all the many kinds of multimedia content, music is one of the most popular content now a day. The reason for this is that music is an art and can be shared by many people from different countries, languages, and cultural backgrounds. One point of reference for this affirmation is the number of items sold daily by web-based dealers, or the number of items downloaded and shared via the internet. In its 2007 Digital Music Report, the International Federation of the Phonographic Industry (IFPI), stated that the number of legally downloaded songs in 2006 reach the amount of 795 million. These facts show us that music is commercially and culturally important.

As the amount of audio content available is increasing several questions arise on how to efficiently access, discover, and present it to the final user. In order to answer all these questions there is the need for new techniques for classifying, searching and retrieving, and recommending audio content. In this thesis we will focus mainly on techniques for recommending audio contents to users. Commercial applications such as content-based music recommendation systems may become increasingly important component of e-commerce applications. One of the advantages of these applications is that they do not need a lot of effort from the user, who is simply presented with potentially relevant items.

Recommendation of music is emerging with force nowadays due to the huge amount of music content and because users normally don't have the time to search through these collections looking for new items. The main purpose of a recommendation system is to estimate the user's preferences and present him with some items that he doesn't know yet. Currently, most of the audio recommendation systems can be classified in two major kinds. Recommendation systems based on collaborative filtering techniques and content-based recommendation systems. While both kinds of systems have good characteristics, they fail to provide good recommendations in specific situations. Recently a new kind of recommendation systems is emerging, hybrid content-based collaborative filtering recommendation systems. In the next paragraphs we will explain each of these kinds briefly.

The main idea of collaborative filtering (CF) methods is to recommend items to a user by taking into account the rates on those items made by other users. CF systems work by collecting user feedback in the form of ratings and exploit similarities and differences among profiles of several users. According to this feedback can be given explicitly, ratings or annotation, or implicitly such as the time spent in examining the content of the recommendations. Although CF has proved to provide good recommendation it presents some troubles. One of the major disadvantages of collaborative filtering is that it cannot recommend new items since the new items don't have any rate. Another disadvantage of CF is that users need to be involved a lot with the system and provide a lot of rates. This is a disadvantage because it is hard to obtain reliable information from many users.

On the other hand, content-based methods provide recommendations by comparing representations of content contained in an item to representations of content rated by the user. In order to obtain a representation of the music content, it is necessary to automatically extract features from the audio signals. These features should be as general as possible and try to represent semantically meaningful concepts. To compare the representations of music content it is required to develop efficient similarity metrics. Recently much effort has been put in computational model of music similarity. The area of Music Information Retrieval (MIR) is an emerging, interdisciplinary research field that deals with the way of efficiently representing and finding similarities among music. Applications in this field range from automated music analysis to personalized music recommendation, online music access, query-based retrieval, and automatic playlist generation. A good overview of current MIR tasks can be found by looking at the MIREX competition for MIR algorithms.

In general, current content-based systems have not yet had significant impact on society due to an inability to bridge the semantic gap between computers and humans. It is hard to extract features from the audio signals that have a truly human meaning. Furthermore, the work presented in states that users put a great value in online music reviews, ratings, recommendations, etc. It seems that the information needs of people are quite related with social behaviour and not only based on content or features of music.

From the above paragraphs we can see that both collaborative filtering and content-based recommendation systems present problems that are hard to solve. One possible way to overcome these problems is the use of hybrid methods that connect collaborative filtering and content-based methods. The main idea of hybrid recommendations is to reflect both ratings and content data in model the user's preferences. Nevertheless, one problem is that representations of user preferences are different in both methods. Content-based methods represent the preferences as a set of features while collaborative filtering represents the preferences as a set of ratings. In order to mix both methods we have to deal with ad-hoc rules to joint these two kinds of representations.

This thesis will try to solve the main disadvantages of content-based and collaborative filtering recommendation methods. With the purpose of overcoming these disadvantages a hybrid content-based collaborative filtering recommendation system will be built. This hybrid system should possess the best characteristics of both methods and will produce better results than each method individually. To achieve these goal we will also present research on current content-based and collaborative methods in such a way that we will be able to find out advantages and disadvantages of them.

A. *Recommendation systems*

Nowadays, there are a lot of recommendation arrangements, adjacent via internet, that endeavour to counsel to user's countless produce such as music, movies, books, etc. In order to comprehend them early it is vital to have a description. In a finished method, recommendation arrangements are arrangements that target to buy opinions or preferences concerning items from an area of users, and use those opinions to present supplementary users alongside items that are interesting to them. From this finished description we can discern that recommendation arrangements demand two frank things to work properly: Data concerning the preferences of the users, and a method to ascertain if an item is interesting for a user. Normally, the users' data includes external data, such as user profiles, buys reports, and product ratings. The method to ascertain whether an item is interesting to a user or not, depends on the kind of recommendation arrangement, and in the methods utilized to find similarities amid items or users .

The above description is quite finished and might be requested even to persons that counsel items to supplementary persons (the salesman in a records' store). An extra specific meaning of recommendation arrangements is given in

"System that produce individualized recommendations as output or have the result of accompanying the user in personalized method to interesting or functional objects in a colossal space of probable options"

The main keywords in this extra proper meaning are individualized and personalized. These words indicate that every single user will be gave alongside disparate data origins or items depending on the data the arrangement has concerning every single user. In order to tolerate or discussion concerning recommendation arrangements, how do they work, and that kinds continue presently, we will have to define countless words that we will be employing across the rest of this document.

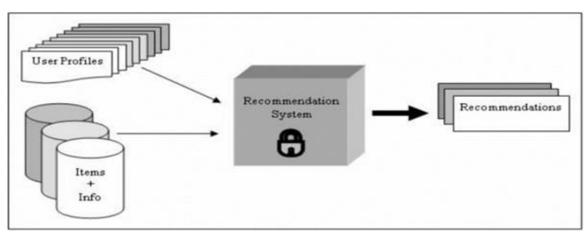

Fig. 1. Recommendation Process

B. *Collaborative filtering*

The development of the Internet has made it far extra tough to efficiently remove functional data from all the obtainable online information. The overwhelming number of data necessitates mechanisms for effectual data filtering. One of the methods utilized for dealing alongside this setback is shouted cooperative filtering [10].

The motivation for cooperative filtering comes from the believed that people frequently become the best recommendations from someone alongside comparable tastes to themselves. Cooperative filtering discovers methods for matching people alongside comparable hobbies and making recommendations on this basis.

Collaborative filtering algorithms frequently need users' alert participation, a facile method to embody users' hobbies to the arrangement, and algorithms that are able to match people alongside comparable interests.

i. Typically, the workflow of a collaborative filtering system is:
ii. A user expresses his or her preferences by rating items (e.g. books, movies or CDs) of the system. These ratings can be viewed as an approximate representation of the user's interest in the corresponding domain.
iii. The system matches this user's ratings against other users' and finds the people with most "similar" tastes.
iv. With similar users, the system recommends items that the similar users have rated highly but not yet being rated by this user (presumably the absence of rating is often considered as the unfamiliarity of an item)
v. A key problem of collaborative filtering is how to combine and weight the preferences of user neighbors. Sometimes, users can immediately rate the recommended items. As a result, the system gains an increasingly accurate representation of user preferences over time.

C. Terms and concepts

The following terms are often used in a recommendation system and the definitions introduced here are based on the work presented in.

- *Item*: in the context of recommendation systems, an item represent the information the system possess about any object. An object can be an electronic document, a product, a person, a service or anything that can be represented by information
- *Recommender:* a recommender is any entity that gives personalized recommendations as output to users' preferences. It may be possible that a recommender does not produce a specific output, such as a list, but they might guide somehow the users in an individual way to useful or interesting items. A recommender could become person or a software system.
- *Recommendation:* this is the output of a recommender; it can be compound by an item or a list of items. The items presented to the users have to be interesting to them, according to the recommender. The criteria used to determine if an item is interesting or not for a user depends exclusively on the technique used by the recommender.
- *User's Interest:* this is an abstract representation of how much a user appreciates an item. This is a subjective concept and it is hard to represent it in an objective way.
- *Prediction:* the expected interest of a user in one item. This concept is different to the concept of recommendation. While some systems might present predictions with the actual recommendations, others can produce recommendations only.
- *Rating:* an objective measure representing a user's interest. The possible values of this measure are given according to a scale established by the designer of the recommendation system.
- *Predicted Rating:* an objective measure representing the expected interest of a user in certain item. This measure is estimated by the system and its possible values are elements of a specific scale.
- *Actual Rating:* objective measure representing the real interest of the user in a specific item. This value is given by the user himself according to the scale of rates of the system.
- *Prediction Accuracy:* a measure that indicates at which extent the predicted rating agrees with the user's actual rating. The more accurate the predictions the better the performance of the recommendation system.
- *Prediction Technique:* the specific algorithm that the recommendation system will use in order to calculate the predicted rating of an item.

2. Objectives

Within this context we make the following contributions:

- To Study and analyze various collaborative filtering approaches.
- To select a dataset that describes a social graph among users, tracks and tags, effectively including bonds of friendship and collaborative annotation.
- We evaluate a Clustering model on this dataset and show that the incorporation of friendship and social tagging can improve the performance of an item recommendation system.
- To evaluate the proposed method outperforms the standard Collaborative Filtering (CF) method, which we also evaluate against the same dataset.
- To measure the accuracy of the filtering method using various metrics such as accuracy, MSE, RMSE, Response time etc.

3. Research Methodology

A. Collaborative Filtering Technology

Description of traditional collaborative filtering algorithm: The detailed description of traditional collaborative filtering algorithm is as follows:
Input: given user set U={u1,u2,……,um}
Resource set ={m1,m2,……,mn}

Rating matrix Rm*n={Rui,Mj},
Rui, Mj are the score given by user ui to resource Mj.
Output: Predict value Pui of resource Mx given by target user Ua
Calculate each user's (ui) similarity (sim ua,ui) of ua and U according to formula (1)(Adjusted cosine similarity).

$$sim(ua,ui) = \frac{\sum_{j\in M}(R_{ua,j}-\bar{R}_j)(R_{ui,j}-\bar{R}_j)}{\sqrt{\sum_{j\in M}(R_{ua,j}-\bar{R}_j)^2}\sqrt{\sum_{j\in M}(R_{ui,j}-\bar{R}_j)^2}}$$

$$P_{ua}, mx = \overline{Rua} + \frac{\sum_{n\in U'} sim(ua,n)(Rn,j-\overline{Rn})}{\sum_{n\in U'} sim(ua,n)}$$

In the formula, sim(ua,ui) shows the similarity between user ua and user ui; M is the resource number ;Rua,j are the values given by user ua to resource j; (Rj)¯shows the average values given by all users to resource j; j is the common resource evaluated by user ua and user ui.

Calculate user ua prediction value (Pua,mx)for resource j, the first n nearest neighbors that has the top similarity should be chosen to calculate. U' shows the nearest neighbor user set of user ua according to formula.

B. Clustering

Clustering has been requested to cooperative filtering in two frank ways. First, the items can be clustered to cut the dimension of the item space and aid alleviate locale sparsity. Second, users can be clustered to recognize clusters of users alongside comparable or correlated ratings. Item clustering does not undeviating lead to locale forecast methods. It is a form of pre-processing pace, that needs the consecutive request of a locale forecast method. O'Connor and Her locker have learned item clustering as a pre-processing pace for area established locale prediction. They apply countless clustering methods, but their empirical aftermath display forecast accuracy truly cuts contrasted to the uncluttered center case even though of the clustering method used. A reduction in computational intricacy is attained, however.

Unlike item clustering, user clustering methods can be utilized as the basis of easy locale forecast methods. Locale forecast established on user clustering is the focus of this chapter. We study clustering algorithms from both the average and hierarchical classes. We familiarize a novel K-medians like locale forecast method alongside good forecast accuracy and low forecast complexity. We additionally debate continuing locale forecast methods for hierarchical clustering.

C. Rating Prediction

Seng and Wang present a user clustering algorithm instituted on divisive hierarchical clustering yelled the Recommendation Tree algorithm (Rec Tree). In the Rec Tree algorithm, a cluster node is increased if it is at a depth less than a enumerated maximum, and its size is larger than a enumerated maximum. The precise sequence in that nodes are increased is not critical, and a facile depth-first or breadth-first progress of the nodes can be utilized till one of the termination conditions is met.

D. Collaborative filtering recommendation model based on user's credibility clustering

This paper introduces user's credibility to evaluate user's rating which will obtained by user's counting on the evaluated resource set. Taking movie recommendation system as an example, user's activity, watching rate, rating impartiality are considered mainly.

Definition 1- User's activity: refers to user's activity of resource rating. The more resources user rates, the more contribution they make, the more active. This paper uses user's resource rating numbers as an indicator for evaluating user's activity, which is shown by Act(u), the formula is as follows:

Act(u)=Count(x) (3)

Count(x) is the accumulated number of rating

Definition 2- User's watching rate: refers to the proportion of the movie resources users have watched out of the movie resources users have evaluated. The higher the user's watching rate, the more effective the rating because that means the rating was given rationally by the user after watching movie. User's watching rate is shown by Wat(u), the formula is as follows:

$$\text{Wat (u)} = \frac{Count(y)}{Count(x)}$$

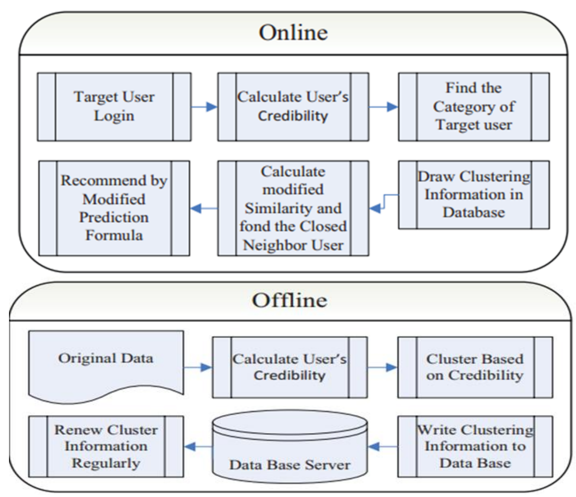

Fig. 2. Basic structure of movie recommendation system

This paper proposes a cooperative filtering recommendation ideal established on user's credibility clustering. This ideal divide recommendation procedure into offline and online phases. Offline, it computes user' credibility early and seizes the insufficient users that has elevated credibility as the clustering center to cluster supplementary users and records the clustered information. Online, the arrangement finds the cluster that target users fit in to, become the clustering data and next gives recommendation. The user clustering numbers are distant less than the user number offline, so merely the similarity amid target user and insufficient clustering centers needs to be computed online. Forecast worth formula of credibility is utilized that the accuracy of recommendation will be increased considerably, period demanded will be decreased.

E. Offline User's Credibility Clustering

The calculation for user's credibility online is so far that encounter the speed of real-time recommendation critically and will stay user's staying time. The satisfaction of user in the direction of the recommendation consequence will be decreased and consequence in the capitulated of client. So, this paper proposes the believed of offline user's clustering, and store the clustering data in the data base.

F. Online recommendation

Based on the consequence of offline data, counsel online that will cut staying period for users considerably, at the alike period increases the accuracy of recommendation across the user's credibility clustering method. The user's satisfaction will be enhanced all-round. Solve the problem of new user's cold start When the presently list users log in, recommendation will not be given as no locale record exists, that is the chilly onset problem. Across observation, it is facile to find out that the favoured resources will be close for users that have comparable interests. This paper holds that new users can be categorized into users that have the alike qualities like hobbies, sex, period, etc. Resources can be categorized correspondingly too to form the correspondence of user and resource. Cluster and give recommendation in this range. It is extra pertinent to give recommendation to new users in the scope of specific user's cluster and resource cluster.

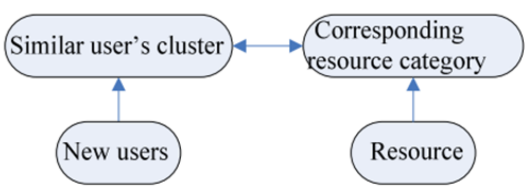

Fig. 3. User-Resource model

4. Conclusion

This paper presented an overview on fast and efficient rule based recommendation system.

References

[1] Ahn, Hyung Jun. "A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem." Information Sciences 178, no. 1 (2008): 37-51.
[2] Bach, Francis, Jacob Abernethy, Jean-Philippe Vert, and TheodorosEvgeniou. "A New Approach to Collaborative Filtering: Operator Estimation with Spectral Regularization." (2008).
[3] Bell, Robert M., and Yehuda Koren. "Scalable collaborative filtering with jointly derived neighborhood interpolation weights." In Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on, pp. 43-52. IEEE, 2007.
[4] Cacheda, Fidel, Víctor Carneiro, Diego Fernández, and VreixoFormoso. "Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems." ACM Transactions on the Web (TWEB) 5, no. 1 (2011): 2.
[5] Del Castillo, H. Siles. "Hybrid Content-Based Collaborative-Filtering music recommendations." (2007).
[6] Duling, Edward Burger. "The development of pedagogical-content knowledge: Two case studies of exemplary general music teachers." PhD diss., The Ohio State University, 1992.
[7] Ekstrand, Michael D., John T. Riedl, and Joseph A. Konstan. "Collaborative filtering recommender systems." Foundations and Trends in Human-Computer Interaction 4, no. 2 (2011): 81-173.
[8] Goggin, Gerard. Cell phone culture: Mobile technology in everyday life. Routledge, 2006.
[9] Groh, Georg, and Christian Ehmig. "Recommendations in taste related domains: collaborative filtering vs. social filtering." In Proceedings of the 2007 international ACM conference on Supporting group work, pp. 127-136. ACM, 2007.
[10] Harris, Ronald J. "The Evaluation of Constraints Within Information Technology Departments Concerning the Statistical and Analytical Problems of Big Data." PhD diss., Northcentral University, 2018.
[11] Hastings, W. Reed, Neil Duncan Hunt, Marc B. Randolph, John Robert Ciancutti, and Stanley Miguel Lanning. "Approach for estimating user ratings of items." U.S. Patent 7,403,910, issued July 22, 2008.
[12] He, Xiangnan, Xiaoyu Du, Xiang Wang, Feng Tian, Jinhui Tang, and Tat-Seng Chua. "Outer product-based neural collaborative filtering." arXiv preprint arXiv:1808.03912 (2018).
[13] Hu, Yifan, Yehuda Koren, and Chris Volinsky. "Collaborative filtering for implicit feedback datasets." In Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on, pp. 263-272. Ieee, 2008.
[14] Huang, Zan, Daniel Zeng, and Hsinchun Chen. "A comparison of collaborative-filtering recommendation algorithms for e-commerce." IEEE Intelligent Systems 5 (2007): 68-78.
[15] Karatzoglou, Alexandros, Xavier Amatriain, LinasBaltrunas, and Nuria Oliver. "Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering." In Proceedings of the fourth ACM conference on Recommender systems, pp. 79-86. ACM, 2010.
[16] Kluver, Daniel, Michael D. Ekstrand, and Joseph A. Konstan. "Rating-based collaborative filtering: algorithms and evaluation." In Social Information Access, pp. 344-390. Springer, Cham, 2018.
[17] Lee, Joonseok, Mingxuan Sun, and Guy Lebanon. "A comparative study of collaborative filtering algorithms." arXiv preprint arXiv:1205.3193 (2012).
[18] Li, Bin, Qiang Yang, and XiangyangXue. "Transfer learning for collaborative filtering via a rating-matrix generative model." In Proceedings of the 26th Annual International Conference on Machine Learning, pp. 617-624. ACM, 2009.
[19] Liang, Dawen, Rahul G. Krishnan, Matthew D. Hoffman, and Tony Jebara. "Variational autoencoders for collaborative filtering." In Proceedings of the 2018 World Wide Web Conference, pp. 689-698. International World Wide Web Conferences Steering Committee, 2018.
[20] Lim, Kian-Tat, SLAC David Maier, Oliver Ratzesberger, and Stan Zdonik. "Requirements for science data bases and scidb." In Conference on Innovative Data Systems Research. 2009.
[21] Liu, Fengkun, and Hong Joo Lee. "Use of social network information to enhance collaborative filtering performance." Expert systems with applications 37, no. 7 (2010): 4772-4778.
[22] Lops, Pasquale, Marco De Gemmis, and Giovanni Semeraro. "Content-based recommender systems: State of the art and trends." In Recommender systems handbook, pp. 73-105. Springer US, 2011.
[23] Marlin, Benjamin. "Collaborative filtering: A machine learning perspective." PhD diss., University of Toronto, 2004.
[24] McEvoy, Arthur F. The fisherman's problem: ecology and law in the California fisheries, 1850-1980. Cambridge University Press, 1986.
[25] Nilashi, Mehrbakhsh, Othman Ibrahim, and Karamollah Bagherifard. "A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques." Expert Systems with Applications 92 (2018): 507-520.
[26] Ortega, Fernando, Bo Zhu, Jesús Bobadilla, and Antonio Hernando. "CF4J: Collaborative filtering for Java." Knowledge-Based Systems 152 (2018): 94-99.
[27] Pasquale, Frank. The black box society. Harvard University Press, 2015.
[28] Poli, Corrado. Environmental Politics: New Geographical and Social Constituencies. Springer, 2015.

[29] Raghuwanshi, Sandeep K., and R. K. Pateriya. "Collaborative Filtering Techniques in Recommendation Systems." In Data, Engineering and Applications, pp. 11-21. Springer, Singapore, 2019
[30] Robinson, Jack, Peter Muller, Timothy Noke, Teng Lew Lim, Wallace Glausi, Larry Fullerton, and Dusan Hamar. "Dynamic information management system and method for content delivery and sharing in content-, metadata-and viewer-based, live social networking among users concurrently engaged in the same and/or similar content." U.S. Patent 8,707,185, issued April 22, 2014.
[31] Sarwar, Badrul M., Joseph A. Konstan, Al Borchers, Jon Herlocker, Brad Miller, and John Riedl. "Using filtering agents to improve prediction quality in the grouplens research collaborative filtering system." In in the GroupLens Research Collaborative Filtering System. Proceedings of the ACM Conference on Computer Supported Collaborative Work (CSCW. 1998.
[32] Susanto, Johanes Lilik. "A practical theological evaluation of the divine healing ministries of Smith Wigglesworth and John G. Lake: A continuationist reformed perspective." PhD diss., University of South Africa, 2007.
[33] Takács, Gábor, IstvánPilászy, BottyánNémeth, and DomonkosTikk. "Scalable collaborative filtering approaches for large recommender systems."The Journal of Machine Learning Research 10 (2009): 623-656.
[34] Tso-Sutter, Karen HL, Leandro BalbyMarinho, and Lars Schmidt-Thieme. "Tag-aware recommender systems by fusion of collaborative filtering algorithms." In Proceedings of the 2008 ACM symposium on Applied computing, pp. 1995-1999. ACM, 2008.
[35] Xu, Zhao, and QiaoFuqiang. "Collaborative Filtering Recommendation Model Based on User's Credibility Clustering." In Distributed Computing and Applications to Business, Engineering and Science (DCABES), 2014 13th International Symposium on, pp. 234-238. IEEE, 2014.
[36] Yang, Xiwang, Yang Guo, Yong Liu, and Harald Steck. "A survey of collaborative filtering based social recommender systems." Computer Communications 41 (2014): 1-10.
[37] Zhao, Zhi-Dan, and Ming-Sheng Shang. "User-based collaborative-filtering recommendation algorithms on Hadoop." In Knowledge Discovery and Data Mining, 2010. WKDD'10. Third International Conference on, pp. 478-481. IEEE, 2010.