A Survey on Collaborative Recommendation Approach

Vaidehi M. Patel, A/Prof.Upma Vacchani
Computer Department, Silver Oak College of Engineering and Technology
Ahmedabad, Gujarat – India

Abstract- **Collaborative filtering is key technique of recommendation system. But traditional collaborative filtering methods are inefficient especially when the user-rating data is extremely sparse. To solve this problem, we propose an approach to compute the user similarity with the type of users-rating items then we develop a collaborative filtering algorithm based on this approach. Furthermore, we put forward an improved collaborative filtering algorithm based on user similarity combination, which combines the user similarity based on user-rating items and the user similarity based on the types of user-rating items. Last, we carry out an experiment with the classic Movielens data sets to evaluate the algorithm. It shows that the collaborative filtering method based on the user similarity computed with types of user-rating items is more effective than the traditional method based on user similarity computed with user-rating items, and the collaborative filtering approach based on user similarity combination gets the best result.**

Index Terms- **recommendation, collaborative filtering, similarity fusion, user rating type**

I. INTRODUCTION

Most of previous studies only consider user-ratings as users' preferences to specific items, but ignore that user-rating can express their preference on items' types, such as genre, style, color, and brand. And compared to tagging data, the item type information is less parser and more easily to be obtained, so analyzing user-rating information comprehensively is more meaningful. This study regards user-rating data as expressing their preferences on items' types, and then proposes the user-rating item type similarity calculation method. In practice, the number of item types is far less than that of items. So the novel user similarity calculation method can alleviate rating data sparsity problem and enhance scalability of recommendation algorithm. In order to verify the method and put it into practical application, this paper puts forward the user-rating item type collaborative filtering method and improved user-based collaborative filtering approach by similarity fusion which adopts weighted average method to combine the novel user similarity with the traditional Pearson similarity. Experiments on typical data set show that the present methods outperform over typical user based collaborative filtering algorithms.

II. LITERATURE SURVEY

The first recommendation system named Joydeep Das , Shreya Dugar , Harsh Gupta , Subhashis Majumder and Prosenjit Gupta et al. proposed research paper on An Adaptive Approach To Collaborative Filtering Using Attribute Autocorrelation in which they briefly explained scalable CF method by using data clustering techniques. The proposed works partitions the users of the CF system using an adaptive K-means clustering algorithm and then use those partitions (clusters) to select the similar users (neighborhood) of a target user[1]. Kathiravelu Ganeshan et al. proposed research paper on An Intelligent Student Advising System Using Collaborative Filtering in which they briefly explained a web based intelligent student advising system using collaborative filtering, a technique commonly used in recommendation systems assuming that users with simila characteristics and behaviors will have similar preferences[2].Peng-yu LU, Xiao-xiao WU, De-ning Teng et al. proposed research paper on Hybrid Recommendation Algorithm for

E-commerce Website in which briefly explained a hybrid recommendation algorithm based on improved collaborative filtering of user context fuzzy clustering and content-based. For collaborative filtering, firstly, user classification is based on fuzzy clustering according to user context, and then collaborative filtering is used to recommend products for similar users[3].Mehak Maniktala, Shuchita Sachdev, Naveen Bansal, Seba Susan et al.proposed research paper on Finding the Most Informational Friends in a Social Network Based Recommender System in which brieflt explained social Network-based Recommender System (SNRS) [4] is a probabilistic model that takes into account user preferences, item's general acceptance and the friends' influence to provide recommendation.Zhao Xu, Qiao Fuqiang et al. proposed research paper on Collaborative Filtering Recommendation Model Based on User's Credibility Clustering in which they briefly explained a taking movie recommendation system as an example ,proposes a collaborative filtering recommendation model based on user's credibility clustering. This model divides recommendation process into offline and online phases[5] FIDEL CACHEDA, V´ICTOR CARNEIRO, DIEGO FERNA´NDEZ,and VREIXO FORMOSO et al.proposed research paper on Comparison of Collaborative Filtering Algorithms: Limitations ofCurrent Techniques and Proposals for Scalable, High Performance Recommender Systems which explained many algorithms in extracting information from user profilesespecially under sparsity conditions. We have also confirmed the good results of SVD-based techniquesalready reported by other authors. As an alternative, we present a new approach based on the interpretation of the tendencies or differences between users and items. Despite its extraordinary simplicity, in our experiments, it obtained noticeably better results than more complex algorithms. In fact, in the cases analyzed, its results are at least equivalent to those of the best approaches studied[6] Hyung Jun Ahn et al.proposed research paper on new similarity measure for collaborative filtering to alleviate the new user cold-starting problem which briefly explained a new heuristic similarity measure that focuses on improving recommendation performance under cold-start conditions where only a small number of ratings are available for similarity calculation for each user. Experiments using three different datasets show the superiority of the measure in new user cold-start conditions[7]. Jonathan L. Herlocher, Joseph A. Konstan, Al Borchers , and John Riedl et al. research paper on An Algorithm Framework For Performing Collaborative Filtering which briefly explained new algorithm elements that increase the accuracy of collaborative prediction algorithms. We then present a set of recommendations on selection of the right collaborative filtering algorithm components.[8] Queen Esther Booker et al.proposed research paper on Automating “Word Of Mouth” To Recommend ClassesTo Students: An Application Of Social Information Filtering Algorithms which briefly explained the development of a system called OSRS (online student recommender system). We present quantitative and qualitative results obtained from the use of OSRS by current students.[9] Ming-Sheng Shang, Zi-Ke Zhang, Tao Zhou, Yi-Cheng Zhang et al. proposed research paper on Collaborative filtering with diffusion-based similarity on tripartite graphs which briefly explained we study a personalized recommendation model making use of the ternary relations among users, objects and tags. We propose a measure of user similarity based on his preference and tagging information. Two kinds of similarities between users are calculated by using a diffusion-based process, which are then integrated for recommendation. We test the proposed method in a standard collaborative filtering framework with three metrics: ranking score, Recall and Precision, and demonstrate that it performs better than the commonly used cosine similarity [10]

III. TRADITIONAL SIMILARITY CALCULATION METHOD

At present, the well-known similarity calculation methods in collaborative filtering field mainly are cosine similarity calculation method and Pearson correlation similarity calculation method.

Assume user set $U = \{u_1, u_2, \dots, u_m\}$, item set $I = \{i_1, i_2, \dots, i_n\}$.

3.1 Cosine similarity calculation method

The ratings of user *a* and user *u* can be regarded as the two vectors in the n-dimensional vector space, so similarity between the two users can be calculated by the cosine of the angle between the two n-dimensional vectors. The computing equation of cosine similarity is as follow:

(1) $$w(a,i) = \cos(\vec{a}, \vec{i}) = \frac{\vec{a} * \vec{i}}{||\vec{a}||.||\vec{i}||}$$

3.2 Correlation similarity calculation method

That cosine similarity calculation method does not take account of user’s rating scale. For example, the average rating of user A is 4, and the average rating of user B is 3, if both of them rate the same item C with 4, but it cannot tell that user a and user b like item C the same. So the adjusted similarity calculation method computes similarity by subtracting the user's average rating value based on cosine similarity calculation method. The calculation equation is as follow:

(2) $$w(a,i) = \frac{\sum_j (v_{a,j} - \overline{v_a})(v_{i,j} - \overline{v_j})}{\sqrt{\sum_{I_a} (v_{a,j} - \overline{v_a})^2} \sqrt{\sum_{I_i} (v_{i,j} - \overline{v_j})^2}}$$

Where, Va,j is the rating which user a gives to the item j, and $\overline{v_j}$ is the average rating of user j.

3.3 Analysis of traditional similarity calculation methods

Overall, the above two methods calculate the similarity between two users making use of users’ rating information. Formula (1) regards the two different users’ ratings as two n-dimensional vectors, and then computes the cosine of the angle between two vectors, but it ignores the users’ ratings scales affecting similarity calculation. To overcome the shortage, Formula (2) subtracts the user’s average rating to obviate the users’ rating scale affecting similarity calculation. Besides, using users’ ratings over common items makes evaluation of user similarity more reasonable. But in practice common rating items between pairs of users are relatively rare, which makes the correlation similarity calculation method inaccurate. In addition, some literatures suggest filling missing values to relieve the rating sparsity, but filling missing value artificially instead of mining user-rating information comprehensively to solve the data sparsity is also unwise.

3.4 User-rating type similarity calculation method

In the e-commerce websites, users have different preferences, and they purchase commodities with different brand, style, color, etc., so their preferences can reflect from the purchased commodities' types. Similarly, the film is divided into action, science fiction, animation, romance, etc., and user rated movies' types can precisely reflect the user's preferences and tastes on movie, that is, if some users prefer to watch the action movie, they will rate more action movies than other. Taking account of the shortages of traditional similarity calculation method, this study proposes user-rating item type similarity calculation method to overcome those shortages and improve precision of similarity measurement.

The steps of user-rating item type similarity calculation are as follow: firstly, compute the counts of each user-rating each type; secondly, apply cosine similarity calculation method to compute the similarity between pairs of users with regarding the counts of users rating item types as n-dimensional vectors. Thus, the similarity between user *a* and user *i* represents as follow:

(3) $$\mathrm{w}(\mathrm{a},\mathrm{i}) = \cos(\bar{a},\bar{\imath}) = \sum_j \frac{v_{a,j}}{\sqrt{\sum_{k\in I_a} v_{a,k}^2}} \frac{v_{i,j}}{\sqrt{\sum_{k\in I_i} v_{i,k}^2}}$$

Where *j* is the item type that both user *a* and user *i* have rated, $v_{a,j}$ is the count of user *a* rating type *j*, and I_a is the item type set which users *a* have rated.

Above calculation method uses user-rating item type information instead of user-rating information to compute users' similarity, in practice, for the number of item types is far less than that of items, so this method can alleviate rating data sparsity problem and improve scalability of algorithm.

Above calculation method uses user-rating item type information instead of user-rating information to compute users' similarity, in practice, for the number of item types is far less than that of items, so this method can alleviate rating data sparsity problem and improve scalability of algorithm

3.5 User similarity combination

Based on similarity calculation explained in section 3 and 3.4, this study applies weighted average method to combine the user-rating item type similarity and Pearson correlation similarity. After combination, the similarity between user *a* and user *i* is as follow:

$$\mathrm{w}(\mathrm{a},\mathrm{i}) = \mathrm{Pw}(\mathrm{a},\mathrm{i}) * \delta + \mathrm{Vw}(\mathrm{a},\mathrm{i}) * (1-\delta) \quad (4)$$

Where $\mathrm{Pw}(\mathrm{a},\mathrm{i})$ is Pearson correlation similarity between the two users, and $\mathrm{Vw}(\mathrm{a},\mathrm{i})$ is the user-rating

item type similarity, δ is the hybrid ratio, and $\delta \in [0,1]$. The best hybrid ratio can be achieved in training.

3.6 User Similarity by UPS

In the process of clustering, the rating information of clustering centers is with special characteristics, i.e., co prefers to rate high marks, and the determination of a user's preference depends on the similarity between the user and these clustering centers. Therefore, an effective similarity measure
method is helpful for assigning the remaining users into different user groups. In order to highlight the importance of user preference, we propose a new similarity measure method to calculate the similarity between users, as follow:

$$sim(a,b)^{UPS} = \exp\left(-\frac{\sum_{i\in I_{ab}} |r_{ai} - r_{bi}|}{|I_{ab}|} \times |\bar{r}_a - \bar{r}_b|\right) \times \frac{|I_a| \cap |I_b|}{|I_a| \cup |I_b|},$$

We know that two important factors are involved. In the global perspective, user preference is reflected by calculating the average rating on all items, and the higher the difference of average ratings between users, the more different preferences of them are shown. Locally, the factor of common rated items are taken into account to reflect the difference between user preferences. Users who have more common rated items with less difference between their preferences, the higher their similarity is shown. Therefore, users who have consistent preferences are easily assigned to the same user group.

IV. CONCLUSION

Our work proposes the method which outputs the user-ratings item type similarity with accuracy. As Number of item types is far less than that of items, so it can easily avoid rating data sparsity problems of similarity calculation and improve the scalability of similarity calculation method. In order to yield even better recommendation result, this work comes up with improved user-based collaborative filtering algorithm by similarity fusion. The experiment has demonstrated that the method is superior to traditional user-based collaborative filtering algorithm and the collaborative filtering algorithm based on user-rating item type. And the improved collaborative filtering algorithm can overcome the effect on similarity measurement due to the inaccuracy of one similarity method and take advantage of each similarity calculation method. The best hybrid ratio is varying in different data.

REFERENCES

[1] Emam, A. Intelligent Advising System. WorldComp 2011 Proceedings.The 2011 World Congress in Computer Science, Computer Engineering,and Applied Computing. Las Vegas, Nevada, USA July 18-21.Retrieved http://worldcompproceedings.com/proc/p2011/IKE3175.pdfon 28/1/2015.

[2] Collaborative Filtering. WebWhompers. Retrieved 10th April 2014from: http://webwhompers.com/collaborative filtering.html

[3] K means Clustering. OhMyPHD. Retrieved 23rd March 2014 from:http://www.onmyphd.com/?p=k-means.clustering#h3_badexample.

[4] J. Das, S. Majumder, D. Dutta, and P. Gupta, "Iterative use of weightedvoronoi diagrams to improve scalability in recommender systems,"in Proceedings of the 19th pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2015), vol. LNAI 9077, 2015, pp. 605–617.

[5] J. Das, A. K. Aman, P. Gupta, A. Haider, S. Majumder, and S. Mitra, "Scalable hierarchical collaborative filtering using bsp trees," in Proceedings of the International Conference on Computational Advancement in Communication Circuits and Systems (ICCACCS 2014), 2014, pp. 269–278.

[6] Manzhao Bu, Shijian Luo, Ji He. A fast collaborative filtering algorithm for implicit binary data. International Conference on Computer-Aided Industrial Design & Conceptual Design, 2009

[7] Guo Feipeng, Lu Qibei, A novel contextual information recommendation model and its application in e-commerce customer satisfaction management, J. Discrete Dynamics in Nature and Society, (2015)1-11.

[8] I. H. Ting, P. S. Chang, and S. L. Wang, "Understanding Microblog Users for Social Recommendation Based on Social Networks Analysis,"J. UCS 18, vol 4, 2012, pp 554-576.

[9]] Guo yanhong, deng guishi. Research on a personalized recommendation algorithm of collaborative filtering[j]. Computer application research, 2008,25(1):39.

[10] S. Susan, and M. Hanmandlu, "A Non-Extensive entropy feature and its application to texture classification," Neurocomputing, vol 120, November 2013, pp. 214–225.

[11] Ning ZhengYuan, Wang Lijin. Common Algorithm And Its Realization for Statistics And Decision Making [M]. Beijing:Tsinghua University Press, 2009:210-213.

[12] S.Wan, Y. Lan, J. Guo, C. Fan, and X. Cheng, "Informational Friend Recommendation in Social Media," Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval, 2013, pp 1045-1048.

[13] Wang Yonggu, Qiu Feiyue, Zhao Jianlong, Liu Hui, Research on personalized recommendation of learning resources based on collaborative filtering recommendation technology, J. Journal of Distance Education, 2011(3)66-71.

[14] Song Yating, Xu Tianwei, Overview of personalized recommendation technologies based user interest, J. Journal of Yunnan University, 2012, 34(S1)20-23.

[15] Li Ying, Wang Bo, Sui Zhanli, Yu Juan, Research on personalized recommendation technologies for e-commerce, J. Fujian Computer, 2015(1)29-30.

[16]S.Susan, and M. Hanmandlu, "A Non-Extensive entropy feature and its application to texture classification," Neurocomputing, vol 120, November 2013, pp. 214–225.

[17] S. Susan, and M. Hanmandlu, "Unsupervised detection of nonlinearity in motion using weighted average of non-extensive entropies," Signal, Image and Video Processing 9.3, 2013, pp 511-525.

[18] Liu N-H (2013) Comparison of content-based music recommendation using different distance estimation methods. Appl Intell 38(2):160–174