

A Personalized Recommendation Method Based on Collaborative Filtering Algorithm

Liu Hui, Cui Lixin, Yao Ting, Li Rongrong

School of Management & Economics, Beijing Institute of Technology, Beijing, China

Email address:

Liuhui0131@bit.edu.cn (Liu Hui), cuilixin@bit.edu.cn (Cui Lixin)

To cite this article:

Liu Hui, Cui Lixin, Yao Ting, Li Rongrong. A Personalized Recommendation Method Based on Collaborative Filtering Algorithm. *International Journal of Business and Economics Research*. Vol. 8, No. 5, 2019, pp. 297-302. doi: 10.11648/j.ijber.20190805.16

Received: June 28, 2019; Accepted: July 27, 2019; Published: August 23, 2019

Abstract: Collaborative filtering algorithm is a widely used recommendation algorithm. However, when applied to e-commerce personalized recommendation, it faces the following issues: firstly, how to consider the user's interest changes over time when getting similarity between the users more precise; secondly, how to use social networks to more accurately getting the nearest neighbor of users; and thirdly, how to consider the behavior of users who have the same interests and different ratings in making the predicted rating score of item more accurately; fourthly, how to use the inherent relation between product categories, such as internal relations, while recommending. In order to solve these problems, this paper improves the traditional collaborative filtering algorithm by integrating timing updates, trust relationship, optimization of predicted rating score and structured ideas. To distinguish users' past interest characteristics and their recent ones, by introducing the idea of timing update, this paper regards the user's shopping experience as a set of time periods, considering the influences of the users' interest at different time on the similarity of the users, and the influence of trust relationship between target user and similar users on the establishment of nearest neighbor set. On this basis, faced with the difference of evaluation criteria of different users on the same recommendation item, this study optimizes scoring method of similar users and gets a pre-scoring-based predicted rating score method for target user to recommend item. Furtherly, considering the relationship between the recommended item and other items, this paper also proposes an idea of relative recommending based on recommended item as a secondary recommendation. At the end of this paper, the proposed method is verified on the review dataset in MovieLens which is provided by the College of computer science and engineering of University of Minnesota. The experimental results show that the proposed method has obvious recommendation accuracy compared with the traditional collaborative filtering algorithm.

Keywords: Collaborative Filtering, Personalized Recommendation, Timing Update, Trust, Relative Recommending

1. Introduction

With the competition between the e-commerce enterprises becomes intensified, personalized recommendation technology, which brings great benefits to e-commerce enterprises, has caught more and more attention. According to the user's personal characteristics, historical behavior and commodity item's characteristics, using collaborative filtering, content filtering, knowledge discovery, interactive recommendation and other recommendation techniques, personalized recommendation enables e-commerce enterprises to recommend items the users may be interested in [1]. Compared with the traditional search engine, personalized recommendation can meet people's needs with different backgrounds, different purposes and different interests in

different periods [2]. The recommendation servicing has changed the shopping way of e-commerce user, from "people looks for information" to "information looks for people".

Effective personalized recommendation can bring win-win situation for e-commerce enterprises and their users. With providing active service, it finds the latent demands of each user and recommends them to the user by the e-commerce enterprises. It improves the user's shopping experience, meanwhile it creates more value for e-commerce enterprises. Nowadays, how to provide more accurate recommendation to users has become a problem which e-commerce enterprises cannot ignore. Because only when e-commerce enterprises understand users better than users themselves, they will become winners in the competition of e-commerce.

Collaborative filtering, as one of the most widely used

recommendation technologies in e-commerce recommendation system, has played a significant role in practical applications. Domestic and foreign scholars have made a deep research on collaborative filtering recommendation technology theoretically, aiming at improving the accuracy of personalized recommendation. Clustering algorithm is applied to collaborative filtering recommendation, and aimed at the problem of the instability of the initial clustering points in the clustering algorithm, the initial clustering point is determined by Kruskal algorithm, in order to improve the accuracy of recommendation [3]. The concept of fuzzy set is introduced into collaborative filtering algorithm, and a collaborative filtering algorithm based on fuzzy clustering is proposed [4]. A collaborative filtering algorithm based on singular value decomposition is proposed [5]. An improved collaborative filtering algorithm based on implicit and explicit attributes by analyzing the attributes of the learners and the order of access to learning resources is proposed, which improves the accuracy of the recommended e-learning resources for users [6]. A collaborative filtering method named content-boosted is proposed, which introduced additional text information to provide recommendations for "new users" and "new item", effectively alleviating the cold start problem [7]. A method is proposed to solve the problem of data sparsity by fusing the classification information of bulk products into the collaborative filtering algorithm [8]. A collaborative filtering method based on matrix decomposition is proposed. This method reduces the sparsity of data and improves the speed of matrix decomposition convergence [9]. Project-based and user-based models are used to predict missing values in the matrix, which greatly increases the density of the evaluation matrix. Then, the filled evaluation matrix is used for collaborative filtering, which significantly improves the recommendation accuracy [10]. A collaborative filtering algorithm based on Bayesian belief network is proposed, which optimizes Bayesian belief network by extending logistic regression model to provide more accurate recommendation in the case of sparse data [11]. An incremental collaborative filtering algorithm based on SVD is proposed, which reduces the cost of repeated matrix decomposition and improves the scalability of the recommendation system [12]. An algorithm based on weight adjustment is proposed, which reduces the negative impact of inaccurate similarity estimation by reducing the similarity weight of users with too little common labeling [13]. The missing values in the evaluation matrix are filled with the average values of similar user annotations, which improves the recommendation accuracy [14]. It can be found that many scholars are studying from different aspects, but the idea of time series updating is less, the purpose of this paper is to improve the traditional recommendation algorithm from the perspective of timing period updating and enhancing trust.

2. Challenges with Collaborative Filtering Recommendations

The collaborative filtering algorithm includes user-based

collaborative filtering algorithm and project-based collaborative filtering algorithm, according to the differences in consideration of similarity.

User-based collaborative filtering algorithm

User-based collaborative filtering is based on this idea: in real life, people are likely to like something that their friends are interested in. The algorithm finds the similar users through the history of the user's shopping, get the nearest neighbor (or the most similar) of the target user, and then recommend what the nearest neighbor likes to the target user.

Project-based collaborative filtering algorithm

The basic idea of collaborative filtering based on project is that if a user likes a project, something like the project may be popular with the user. It should be noted that project-based collaborative filtering is based on the user behavior that favors two items in the system to find similarities between items.

When the data in the recommendation system is complete and abundant, the traditional collaborative filtering algorithm has the characteristics of accurate recommendation and good recommendation results. However, for new users, due to the lack of interest data on the project, a cold start problem is happened. In addition, the problem of sparsity exists when the data of historical interest between users is relatively small or the number of projects is huge.

In the face of the complexity of business activities and the shortage of collaborative filtering algorithms, collaborative filtering technology will face such problems when realizing personalized recommendation of e-commerce: first, the recommendation of traditional collaborative filtering is to find the similarity between users or projects based on the user-project rating matrix, which does not take into account how the users' interests change over time, that is, the timeliness problem of user's interest does not considered. Second, the recommendation of traditional collaborative filtering ignores the trust relations between user in social network. Third, the traditional collaborative filtering recommendation ignores the inherent relation between commodity categories, such as printers and cartridges, a brand of cosmetics, and other factors which can influence the recommendation efficiency.

3. A Personalized Recommendation Method Based on Collaborative Filtering

Based on the above researches in this paper, considering users' interests characteristics changing over time, the analysis of users' interest characteristics should be based on different time periods, and the interest characteristics of neighbor users will be so. Although the users' interests will change over time, the long-term trust relationship between users still has a positive impact on the recommendation, which is an important aspect in this study. Because there are subjective differences between different users' evaluation criteria, how to eliminate the recommendation interference caused by subjective differences is also a part of this paper. In order to improve the

recommendation experience, this study also considers some commodity items which have inherent relation with the recommended item, and gives an in-depth recommendation by recommending the relevant items to target user.

3.1. Research on Similar Users of Collaborative Filtering Based on Time Series

In the face of user's interests and time sensitive characteristics, the idea of timing update is introduced into collaborative filtering algorithm. The idea of timing update is to analyze user's buying behavior by stages. The closer the two stage is, the smaller the user interest bias.

According to step size T, the purchase history of user U is divided into different time interval $(t_{i-1}, t_i]$, $1 \leq i \leq k, t_i = t_0 + iT$, where t_0 represents the beginning time of the user U purchase history.

Considering the deviation of user's interests over time, the forgetting curve of German psychologist Ebbinghaus is introduced in this study to express the deviation of users' interest over time. The forgetting curve shows the nonlinear decreasing law of human memory retention. In this paper, a Logistic function is used to express this forgetting curve, which reflects the trend of user interest changing. The expression of forgetting function is (1), which indicates that the longer it goes, the more forgetting is.

$$f(t_i) = \frac{1}{1 + e^{\frac{t_i - t_k}{T}}} \tag{1}$$

t_k indicates user U's current time, and t_i indicates a point in time that there is the i-th recommendation to user U at.

Based on forgetting function, we define the interest offset weight, see (2).

$$g(t_i) = 1 - f(t_i) \tag{2}$$

N_u^{k-1} is defined as the nearest neighbor set of user U at t_{k-1} . ΔN_u^k is the similar neighbor set of user U in the interval $(t_{k-1}, t_k]$, ΔN_u^k is gotten by traditional collaborative filtering algorithm. Then the user U's similar neighbor set $R(N_u^k)$ at t_k is shown as follows (3).

$$R(N_u^k) = N_u^{k-1} \cup \Delta N_u^k \tag{3}$$

In general, the nearest neighbor set of user U at t_k is relatively close to that at t_{k-1} . The nearest neighbor of user U at t_k , that is N_u^k , belongs to the similar neighbors of the user U in the interval $(t_{k-1}, t_k]$, that is ΔN_u^k , and those who have nearest neighbors in N_u^{k-1} . A problem with this processing is that in some cases, users who are less similar to user U in N_u^{k-1} will be eliminated after the K-th recommendation.

The definition of ΔSim_{kv} is the similarity between the target user U and other user V in the interval $(t_{k-1}, t_k]$, and the

similarity between user U and user V in $R(N_u^k)$ at t_k can be Sim_{kv} defined by (4).

$$Sim_{kv} = \frac{1}{C_k} \sum_{i=1}^k g(t_i) * \Delta Sim_{iv}$$

$$C_k = \sum_{i=1}^k g(t_i) \tag{4}$$

According to (4), the nearest neighbor set N_u^k of target user U at t_k is formed from $R(N_u^k)$.

3.2. Research on Similar Users Based on Historical Trust

Facing the situation of data sparsity that the user V has no common shopping activity with the target user U in the interval $(t_{k-1}, t_k]$ (One case is that the user V is the nearest neighbor of the target user U at all time period except current period, but the similarity between user V and user U in the interval $(t_{k-1}, t_k]$ is 0). In this study, the idea of obtaining implicit user trust by John O'Donovan is introduced, and combined with the research in 3.1. Historical trust is used to adjust the user's similarity, in order to make the user V which is wrongly excluded from the nearest neighbor set N_u^k at t_k according to (4) can still recommend items to target user U.

The basic idea of John O'Donovan's implicit obtaining user trust is: according to the recommended history of a user to the target user, the ratio of the numbers recommended correctly for the user to the total recommended numbers can be found in his history. It reflects the potential trust of the target user to the user. The level of the trust can help us to make up for the deviation of the nearest neighbor set caused by removing users incorrectly.

In this study, the idea of getting nearest neighbor based on historical trust is: Facing the removed user V which is in the similar neighborhood set $R(N_u^k)$ of the target user U and the Sim_{kv} is lower, considering the trust level between the user V and the target user U, if the target user U has a high degree of trust to the user V, then the user V can be saved as a nearest neighbor in the N_u^k set, otherwise it will be removed completely from N_u^k .

The standard of measuring the trust level of the target user U to the user V can also be defined as: if user V is the user U's nearest neighbor in the more than R or R interval in history experience of the user V, the trust level of the target user U to the user V is R. R can be set to k-1, 3k/4, or other reasonable functions.

3.3. Research on Prediction Rating Score of Target User

In order to reduce the deviation between different user rating criterion and reduce the interference caused by similar user scoring standards, an optimized prediction rating method is presented in this paper to predict the target user's rating on the target item.

According to the research in 3.1. and 3.2., the nearest neighbor set of target user U is formed. $N_u^k(i)$ is defined as the nearest neighbor set of target user U in that user rate item I at t_k . Setting $V \in N_u^k(i)$, considering the similarity relationship between target user U and nearest neighbor user V, we can get the user U's pre-rating of item I by referring to user V's rating of item I. see (5).

$$P_u(i) | R_{vi} = \bar{R}_u + Sim_{kv} * (R_{vi} - \bar{R}_v) \quad (5)$$

$$P_u(i) = \frac{\sum_{v \in N_u^k(i)} Sim_{kv} * P_u(i) | R_{vi}}{\sum_{v \in N_u^k(i)} |Sim_{kv}|} = \bar{R}_u + \frac{\sum_{v \in N_u^k(i)} Sim_{kv}^2 * (R_{vi} - \bar{R}_v)}{\sum_{v \in N_u^k(i)} Sim_{kv}} \quad (6)$$

3.4. Related Recommendation to Recommended Item

Based on the research in 3.1.-3.3., using the inherent relation between commodity items, items relating with recommended items are be found and further recommendation of related item may be existed. That is secondary recommendation.

For commodities sold in e-commerce, there are usually some relation between them, as follows.

- (1) Category relation: classification of commodity attributes has formed tree hierarchy relation among items of commodity set.
- (2) Demand dependency relation: in economics, there are obvious usage dependence between some commodity items. Users must use another commodity item while using one kind of commodity, thus it forms a clear dependence between the two commodity items. For example, users who choose printer need printing paper.
- (3) Complementation relation: some items with certain attributes are highly expendable and unsatisfiable. Users usually need a lot of similar commodities to meet their needs and desires. Such as film and television, dramas, music songs, etc.
- (4) Substitutional relation: there is an obvious mutual exclusion relationship between certain commodity items. A user's need for one commodity can substitute for his need for another one at a given time period. For example, users who just bought HUAWEI mobile phone don't want to buy OPPO mobile phones soon.

Based on the above relations between items, we can make further recommendation with items related to the recommended items. For example, when an item which is a complementary item of a recommended item exists, it can be recommended to target users. When an item which has a demand dependence relation with recommended item exists, according to the analysis result of shopping history data, it will be decided whether to be recommended to target user. While deciding related recommendation, we make decision of selecting recommendation items based on nearest neighbor user's rating.

R_{vi} represents the user V's rating of item I, $P_u(i) | R_{vi}$ represents the user U's pre-rating of item I under condition of R_{vi} , \bar{R}_u and \bar{R}_v represent the mean rating of target user U and its nearest neighbor user V.

Based on the rating score of all the nearest users of user U in $N_u^k(i)$ for item I, the target user U's predicted rating score for item I is shown in (6).

4. Validation

On MovieLens, a group of film rating dataset collected by the GroupLens project team of the school of computer science and engineering of University of Minnesota, this paper validates the effectiveness and feasibility of the above proposed method.

4.1. Experimental Environment and Data

The experimental environment is as follows:

- (1) hardware environment: CPU is AMD 2.0GHz, memory is 2GB, hard disk is 250G;
- (2) software environment: Windows 7, Eclipse, JDK1.6;
- (3) programming language: Java.

The experimental dataset, MovieLens, consists of 943 users whose basic information are included, 1682 movies and 100000 rating scores. In this dataset, the rating is divided into 5 levels. The higher the rating score, the more the user's favor.

User's basic information includes user demographic information, such as user age, gender, occupation and other information.

Movie information includes movie ID, movie name, movie release time, rating time, movie link and movie subjects.

User's rating of movies includes rating score and rating time.

The dataset is divided into 80:20. 80% of the dataset is used as training set, and 20% of the dataset is used as a test set. In order to better validate the accuracy of the algorithm, the dataset is divided into 5 groups, and there is no intersection between corresponding test sets.

4.2. Evaluation Criterion

The MAE (that is Mean Absolute Error) is a statistical accuracy measurement index. It is used to evaluate the recommendation accuracy of a recommendation method.

$\{p_1, p_2, p_3, \dots, p_n\}$ is defined as the predicting rating score of recommended items in test set, and $\{q_1, q_2, q_3, \dots, q_n\}$ is as the actual score of recommended items in test set. MAE is defined as follow (7).

$$MAE = \frac{\sum_{i=1}^n |p_i - q_i|}{n} \quad (7)$$

The lower the MAE, the higher the accuracy of prediction.

4.3. Experimental Design and Results

Under the same experimental conditions, this paper uses the Java language to verify and compare the traditional collaborative filtering algorithm and the method proposed in this paper on the data set MovieLens. The comparison

verification is carried out on the 30 day, the 60 day and the 90 day as a time period, and the experimental results are shown in Figure 1. The results show that the recommendation accuracy of the proposed method is higher than that of the traditional collaborative filtering algorithm, and as the time series period increases to a certain, the MAE value decreases continuously. From Figure 1, when the time period is 90 days or 60 days, the difference of MAE is very little. Therefore, in order to maintain a certain precision, selecting 60 days as a time period is better.

MAE Comparison

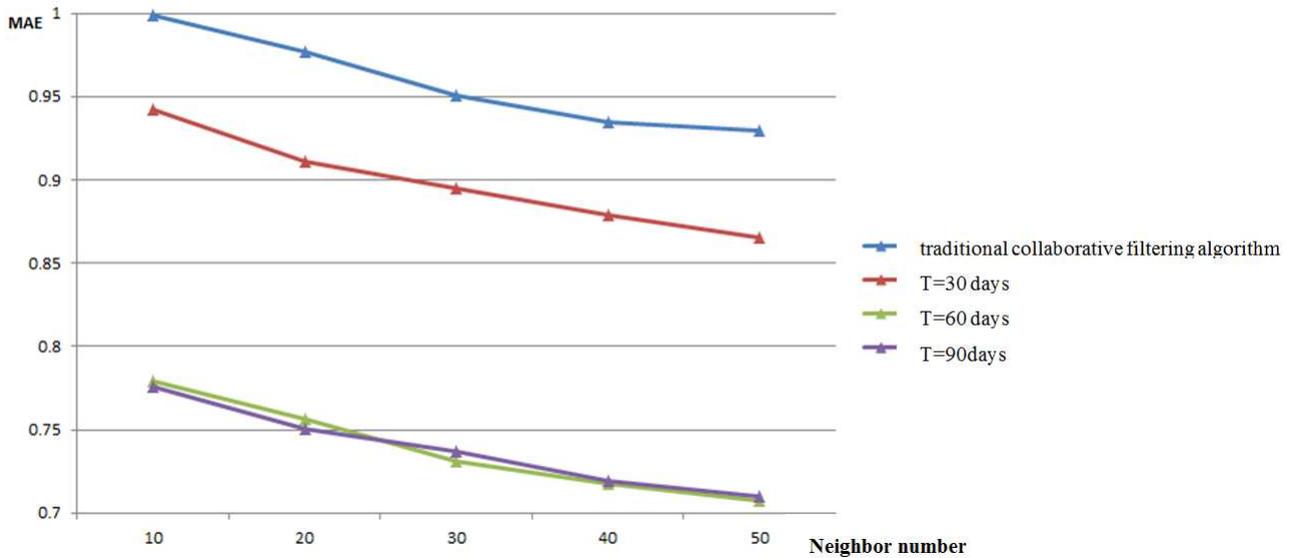


Figure 1. MAE comparison of the traditional collaborative filtering algorithms and the proposed method.

5. Conclusion

In this paper, the traditional collaborative filtering algorithm is optimized by introducing the idea of interest time updating, trust level of target user and evaluation criterion differentiation. Meanwhile, based on item inherent relation, the paper further considers the recommendation of related commodity items, and proposes a personalized recommendation method. The verification on the dataset MovieLens shows that the personalized recommendation method in this paper is more accurate than the traditional collaborative filtering algorithm, and the recommendation accuracy is different with the difference of time period. Due to the lack of inherent relation between commodity items in the experimental dataset, the verification of related recommendation cannot be effectively reflected. This part will be carried out in our next research work.

Acknowledgements

We acknowledge the support of the following research fund and project: (1) the National Key R&D Program of China (2016YFF0204100); (2) Beijing Natural Science Foundation (9192018).

References

- [1] Xiang L, "Recommended system practice," Beijing: People's Posts and Telecommunications Press, 2012.
- [2] Shan M, "Design and implementation of e-commerce recommendation system based on personalized recommendation," Changchun: Jilin University, 2014.
- [3] Yuan L, "Research on Collaborative filtering personalized recommendation algorithm based on Clustering," Wuhan: Huazhong Normal University, 2014.
- [4] Wang M J, "Collaborative filtering algorithm based on fuzzy clustering," Computer Engineering, 2012 38 (24): 50-52.
- [5] Chen Q H, "Collaborative filtering recommendation algorithm based on SVD," Chengdu: Southwest Jiao Tong University, 2012.
- [6] Salehi M, "Application of implicit and explicit attribute based collaborative filtering and BIDE for learning resource recommendation," Data & Knowledge Engineering, 2013, 87 (9): 130-145.
- [7] Melville P, Mooney R J, Nagarajan R, "Content-boosted collaborative filtering for improved recommendations," Proceedings of the National Conference on Artificial Intelligence, Menlo Park, CA; Cambridge; MA; London; AAAI Press; MIT Press; 1999, 2002. 187-192.

- [8] Ziegler C N, Lausen G, Schmidt-Thieme L, "Taxonomy-driven computation of product recommendations," Proceedings of the thirteenth ACM international conference on Information and knowledge management, ACM, 2004, 406-415.
- [9] Nasiri M, Minaei B, "Increasing prediction accuracy in collaborative filtering with initialized factor matrices," Journal of supercomputing, 2016, 72 (6): 2157-2169.
- [10] Ma H, King I, Lyu M R, "Effective missing data prediction for collaborative filtering," Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, ACM, 2007, 39-46.
- [11] Greiner R, Su X, Shen B, et al, "Structural extension to logistic regression: Discriminative parameter learning of belief net classifiers," Machine Learning, 2005, 59 (3): 297-322.
- [12] Sarwar B, Karypis G, Konstan J, et al, "Incremental singular value decomposition algorithms for highly scalable recommender systems," Proceedings of Fifth International Conference on Computer and Information Science, Citeseer, 2002, 27-28.
- [13] Herlocker J L, Konstan J A, Borchers A, et al, "An algorithmic framework for performing collaborative filtering," Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, ACM, 1999, 230-237.
- [14] Chee S H S, Han J, Wang K, Rectree, "An efficient collaborative filtering method," Proceedings of Data Warehousing and Knowledge Discovery, Springer, 2001, 141-151.
- [15] Y. Ding, X. Li, "Time weight collaborative filtering," ACM international conference on Information and knowledge management, Bremen, 2005, 485-492.