A Time Effect based Collaborative Filtering Approach for User Preference Statistics and Recommendation

Yuxi Chen¹, Xiaotong Zhang², Qing Zhao³,⁴* and Faheem Akhtar ⁵,⁶

¹School of Statistics and Mathematics, Country Central University of Finance and Economics, Beijing, 100081, China
²Harpur College of Arts and Sciences, Binghamton University-SUNY, New York, 13902, America
³Faculty of Information Technology, Beijing University of Technology, Beijing, 100124, China.
⁴Department of Computer Science, Vrije Universiteit Amsterdam, 1081 HV, The Netherlands
⁵School of Software Engineering, Beijing University of Technology, Beijing Engineering Research Center for IoT Software and Systems, Beijing, 100124, China
⁶Department of Computer Science, Sukkur IBA University, Sukkur 65200, Pakistan

*Corresponding author’s e-mail: qzo210@vu.nl

Abstract. With the rapid development of the information technologies in the financial field, extracting meaningful information from a massive amount of data is hugely significant for efficient business decision making. The recommendation system is an intelligent system that applies historical knowledge of users to infer their preferences and make a personalized recommendation. However, it suffers from the problem of time effect of user's behaviour, which means a user’s interests may change over time. To overcome this problem, we propose a time effect based collaborative filtering approach to adaptively statistics the change of user preferences. Firstly, Item-based collaborative filtering is used to calculate rating similarity between items. Since an Item-based collaborative filtering algorithm doesn’t consider the time effect; next, the time decay function is proposed to statistics the change of user interests. Experimental results show that the proposed scheme retained higher accuracy compare to traditional collaborative filtering method.

1. Introduction

A recommendation system is a model that recommends relevant products to users through statistics and investigates the potential preferences and behaviours of diverse users in widely used online application domains, such as, e-commerce, news information, products recommendation, and social-networking [1] [2].

Generally, a recommendation system can be divided into three main categories, i.e., rule-based association recommendation [3], content-based filtering system (CB) [4], and collaborative filtering system (CF) [5].

In the rule-based association recommendation, we mine the correlation between different goods during the sales process, which is often used in the retail industry [6]. The association rule is
calculated considering the individual transaction in the database that what proportion of people who buy product A also buy product B. Such as, to highlight percentages of people who bought milk also bought bread.

Content-based (CB) filtering is another widely practiced recommendation approach, which recommends items similar to those that were previously liked by a user [7] [8]. CB usually involves three steps: (1) Item representation; it means extracting features and attributes (content) to represent a specific item. (2) Profile learning; it utilizes the content of an item that a user liked previously to learn the user preference. (3) Recommendation Generation, this step produces a group of most relevant items to users by comparing the content of the user's profile obtained in previously defined steps with candidate items. Furthermore, Collaborative filtering (CF) is the most famous method that plays a pivotal role in the development of an efficient recommendation system, which relies on the item's rating provided by customers [9]. CF can further be divided into two main categories: i.e., user-based collaborative filtering [10] and item-based collaborative filtering [11]. The basic principle of user-based collaborative filtering recommendation is to use a substantial amount of information of all users to find a user group with the similar preference to the current user. For example, if a user A likes an item N, a user B likes items M and N, we can easily assume that the preferences of two users are similar. Whereas, the item-based collaborative filtering utilizes historical data of users to calculate similarity between items, and then recommends similar items for users with similar preference. For example, if a user A likes an item N, and a user B likes an item M and N, we can assume that items M and N are similar, so we can also recommend item M to a user A.

Although the above recommendation approaches have successfully been applied in various user preference statistics tasks and recommendation tasks; but they did not consider the change in user interests over time. In this paper, we proposed a time effect based method to adaptively statistics the change of user preferences from time to time.

2. Model description

The proposed a time effect based collaborative filtering model is comprised of three main steps. Firstly, the rating similarity is calculated based on the item-based CF approach. Secondly, in order to statistics the change of user preferences, inspired by Newton’s law of cooling, we proposed a time decay factor to calculate the change of user preference. Lastly, we combined the rating similarity calculation and time decay factor to produce the final similar item set and top-ranked N items list.

2.1. Rating similarity calculation

Rating similarity calculation is a method in collaborative filtering, to compute the similarity between items, our model chose the item-based method to calculate the rating similarity because the number of users is higher than the number of items in the datasets. Item-based CF recommends the items which are similar to the item’s that a user previously liked, for example, if a user has ever purchased sportswear, the system will automatically recommend other similar sportswear to the user. The recommended system of the item-based CF can mainly be described in the following two steps.

(1) Calculate the similarity between items. The item-based CF assumes that each user is limited to a few of interests, and if two items appear on the positive feedback list of a large number of users, they have a high probability of belonging to the same domain and therefore have a substantial similarity. The similarity between items is calculated by the following equation:

\[ w_{ij} = \frac{|N(i) \cap N(j)|}{|N(i) \cup N(j)|} \]  

Where, \( N(i) \) is the number of users who like item \( i \), \( N(j) \) is the number of users who like item \( j \), \( N(i) \cap N(j) \) is the number of users who like both item \( i \) and item \( j \).

(2) The recommendation list is produced according to the similarity between items and the user historical behaviours. In this stage, we can use equation (2) to compute the preference of user u to an item j.
\[ p_{ui} = \sum_{i \in N(u) \cap S(j,K)} w_{ji} r_{ui} \]  

Where, \( p_{ui} \) represents the preference of the user \( u \) to item \( j \). \( N(u) \) represents a collection of items that the user likes (\( i \) is an item in \( N(u) \)), \( S(j,K) \) represents the \( K \) items that are most similar to item \( i \), \( w_{ji} \) is the similarity between item \( i \) and item \( j \), \( r_{ui} \) represents preference of user \( u \) to item \( i \).

2.2. Time decay factor calculation

Time decay factor exhibits a significant influence on the change of user preference. The closer to the current time, the more user behaviour can reflect the real interest of users. Therefore, in order to make item-based CF objective and accurate, we propose to use a time decay factor that gives different weight to user’s behaviour considering the time it happens.

According to Newton’s law of cooling, the law by which an object whose temperature is higher than its surroundings transfers heat to the surrounding medium as it gradually cools. Which means user preference in the product will decay over time. In the process of voting, the basic principles such as the simulation of vote decay process are based on Newton’s law of cooling. This rule can be defined as the decline of user interest and the length of time is inversely proportional. The formulation can be defined as follows:

\[
\frac{dT(t)}{dt} = -k(T(t) - H) \\
T(t) = T(t_0) \times e^{-k(t_0-t)}
\]

Where, \( T(t) \) represents the user’s current preference for the item that is related to time, \( T(t_0) \) is the initial value, \( H \) is the length of time, \( k \) is the ratio.

Based on the above formulations, we can obtain the time decay function as follows:

\[ N(t) = N_0 e^{-\alpha t} \]  

Where, \( N(t) \) is the value in the time \( T \), \( N_0 = N(0) \) is the initial value in the time 0, it is the highest value when \( t \geq 0 \), \( \alpha > 0 \) is the time decay constant, it will get different values in the different systems, if a user preference is changing quickly in a system, the \( \alpha \) will get a larger value, otherwise, \( \alpha \) will get a smaller value.

2.3. Time effect based collaborative filtering

This module combines the item similarity calculation and the time decay factor to improve the effectiveness of user preference statistics and personalized recommends. The equation can be calculated as follows:

\[ S_{ij} = \sum_{u \in N(i) \cap N(j)} f(|t_{ui} - t_{uj}|) \frac{1}{\sqrt{|N(i)||N(j)|}} \]

Where, \( S_{ij} \) represents the similarity between item \( i \) and item \( j \), \( u \) represents a user, \( N(i) \cap N(j) \) represents the users who like both item \( i \) and item \( j \), \( |N(i)| \) represents the number of users who like item \( i \), \( |N(j)| \) represents the number of users who like item \( j \), \( f(|t_{ui} - t_{uj}|) \) is the time decay function, \( t_{ui} \) represents the time when user \( u \) acts on item \( i \), \( t_{uj} \) represents the time when user \( u \) acts on item \( j \). The value of \( f(|t_{ui} - t_{uj}|) \) can be defined as the longer the user \( u \) acts on item \( i \), the smaller value the \( f(|t_{ui} - t_{uj}|) \) will obtain.

The function \( f(|t_{ui} - t_{uj}|) \) can be express as follows:

\[ f(|t_{ui} - t_{uj}|) = N_0 e^{-\alpha(t_{ui}-t_{uj})} \]  

Generally speaking, the more recent the user’s behaviour is, the more relevant it is to the current behaviour, therefore, we define the prediction formulation as follows:

\[ p(u,i) = \sum_{j \in N(i) \cap S(i,K)} S_{ij} e^{-\beta(t_0-t_{uj})} \]

Where, \( t_0 \) is the current time, \( \beta \) is the time decay parameter, the value of \( \beta \) will be selected according to the different datasets (the detail is described in part 3.2 Parameter setting).
3. Experience and evaluation

3.1. Datasets and data preprocessing
In this paper, a real-world dataset is used to test the proposed scheme. The dataset includes 11872 financial products and 5164 fund products, 952385 users, and 5031877 user ratings, the corresponding timespan ranges of the user behaviour record from Oct 1, 2017 to Jul 1, 2019.

Primarily data preprocessing technique is used to process the original data into a format on which one can directly apply the recommendation algorithm, which includes two basic steps, in the first step, we filtered the abnormal data, such as, the products that have been removed from shelves, the user who has never produced the user behaviour, and the user who didn’t produce behaviour from a recent year. After filtering abnormal data, the datasets include 3142 financial products, 856 fund products, 27334 users and 43698 user ratings. Moreover, the normalization method is used for all vectors. As shown in the equation (9).

$$\hat{x}_{mn} = \partial_1 + (\partial_1 - \partial_2) \left( \frac{x_{mn} - \min(x)}{\max(x) - \min(x)} \right)$$  \hspace{1cm} (9)

Where, max(x) and min(x) are the max value and min value of corresponding dimension. $x_{mn}$ represents n-th dimension and m-th sample, $\partial_1, \partial_2$ are the upper and lower limits of the mapping interval.

3.2. Parameter setting
According to the characteristics of the financial products, the initial value of time decay is set as the popularity of financial products. The average number of people interacting with financial products in the initial stage is 1,000 days, we set the initial value of $N_0$ to 1, The average life cycle of financial products is 119.65 days, whereas, in the last stage of financial products, the average number of interactions were only 10 per day, thus, we can easily obtain the value of $N(119.65) = 0.01$. Moreover, after applying normalization process with a time decay constant equals to 0.0385. Lastly, we calculated the time decay factor as shown in equation (7). For the time effect based collaborative filtering, the value of $\beta$ is obtained as 0.0385.

$$N(t) = e^{-0.0385t}$$  \hspace{1cm} (10)

3.3. Evaluation criterion
In this paper, we used two widely used evaluation metrics to measure the performance of our method.

1) Precision describes the percentage of items that are recommended correctly in the recommended list.

$$\text{Precision} = \frac{|R(u) \cap T(u)|}{\sum_u |R(u)|}$$  \hspace{1cm} (11)

Where, $R(u)$ is the $N$ items are recommended to user $u$, $T(u)$ represents the collection of items that user $u$ has positive feedback.

2) Recall describes the proportion of items recommended correctly to the list of items that the user has positive feedback.

$$\text{Recall} = \frac{|R(u) \cap T(u)|}{\sum_u |T(u)|}$$  \hspace{1cm} (12)

3.4. Experimental results
This section presents the results of the comparisons of the proposed models with baseline methods. In order to evaluate the effectiveness of the proposed time decay factor, this paper only focused on the baseline methods that used traditional item-based collaborative filtering. Because of the limited number of recommendation results displayed by recommendation systems, we first determined the $N$ values with the best recommendation performance through experiments (as shown in Figure 1), and then, the value of similar set $K$ is calculated according to $N$ (as shown in figure 2).
First, we evaluated the effect of top N results and similar item size K on accuracy and recall. From the Figure 1, we can observe that as the number of recommended results increases, the recall rate also increases. The main reason might be the recall rate, which is the ratio of the products that users like in the recommendation list to all the products those users like in the system, thus, as the number of recommendation lists increases, the probability of products user preference being recommended also increases. When N=6, the accuracy rate is the highest i.e., 23.6%, therefore during the experimentation, we set N=6 as the final value.

As per selecting the value of N, we can calculate the similar item set size K. From Figure 2, we can observe that with K=25, both the accuracy and the recall reaches to the highest i.e., 45.4% and 16.6% respectively.

Finally, we compared the proposed scheme with the baseline. Form Table 1; it is observed that our method has better performance than baseline, which conclusively demonstrates the importance of the proposed time decay factor and the effectiveness of the proposed scheme. Furthermore, the
experimental result illustrates that effectively fitting the time decay factor of the data can better statistics the change of user preferences.

4. Conclusion
In this research, we proposed to use a time effect based collaborative filtering approach for solving the problem of change in the user’s preferences and interests over time. The proposed method can fully consider the time effects of financial products and user preferences. In the proposed approach, initially, the item-based CF method is utilized to calculate the rating similarity, later, we proposed to use a time decay factor to calculate the change of user preference, and finally, we integrated the rating similarity and time decay factor. The experimental results demonstrate that the proposed scheme has significantly enhanced the accuracy of the recommendation system, and I believe this research can serve as an essential reference during the development of efficient recommendation systems in various applications domains.

References
[1] S. R. Gandhi, J. Gheewala (2017), A Survey on Recommendation System with Collaborative Filtering Using Big Data, International Conference on Innovative Mechanisms for Industrial Application, pp. 457-460.
[2] SS. Li, E. Karahanna (2015), Online Recommendation Systems in a B2C E-Commerce Context: A Review and Future Directions, Journal of the Association for Information Systems, Volume 16, pp. 72-107.
[3] L. Moonen, S. Di Alesio, D. Binkley, T. Rolfnes (2016), Practical guidelines for change recommendation using association rule mining, IEEE/ACM International Conference on Automated Software Engineering (ASE), pp. 252-260.
[4] HJ Gao, JL Tang, X Hu, H Liu (2015), Content-Aware Point of Interest Recommendation on Location-Based Social Networks, Association for the Advancement of Artificial Intelligence, Austin, pp. 155-164.
[5] Y. Koren, R. Bell (2015), Advances in Collaborative Filtering, Recommender systems Handbook, Springer, pp. 77-118.
[6] Luberg A., Tammet T., Järv P. (2011), Smart City: A Rule-based Tourist Recommendation System, In: Law R., Fuchs M., Ricci F. (eds) Information and Communication Technologies in Tourism, pp.51-62.
[7] P. Lops, M. De Gemmis, G. Semeraro (2011), Content-based Recommendation Systems: State Of The Art and Trends, Recommender Systems Handbook, Springer, pp. 73-75.
[8] Kompan M., Bieliková M. (2010), Content-Based News Recommendation, In: Buccafurri F., Semeraro G. (eds) E-Commerce and Web Technologies, Vienna, pp. 61-72.
[9] Mehdi Elahi, Francesco Ricci, Neil Rubens (2016), A survey of active learning in collaborative filtering recommender systems, Computer Science Review, Volume 20, pp. 29-50.
[10] Hamidreza Koohi, Kourosh Kiani (2016), User based Collaborative Filtering using fuzzy C-means, Measurement, Volume 91, pp. 134-139.
[11] Dubey A., Gupta A., Raturi N., Saxena P. (2018), Item-Based Collaborative Filtering Using Sentiment Analysis of User Reviews, Applications of Computing and Communication Technologies. Communications in Computer and Information Science, vol 899. Springer, Singapore, pp. 77-87.