Time Weight Update Model Based on the Memory Principle in Collaborative Filtering

Dan Li
School of Science, Beijing University of Posts and Telecommunication, Beijing, China
Email: lidan9@bupt.edu.cn

Peng Cao *
College of Information Engineering, Beijing Institute of Graphic Communication National, Beijing, China, Email: pc@bigc.edu.cn

Yucui Guo
School of Science, Beijing University of Posts and Telecommunication, Beijing, China

Min Lei
Information Security Center, Beijing University of Posts and Telecommunication, Beijing, China
National Engineering Laboratory for Disaster Backup and Recovery, Beijing University of Posts and Telecommunications, Beijing, China

***Abstract*— Collaborative filtering is the most widely used technology in the recommender systems. Existing collaborative filtering algorithms do not take the time factor into account. However, users' interests always change with time, and traditional collaborative filtering cannot reflect the changes. In this paper, the change of users' interests is considered as the memory process, and a time weight iteration model is designed based on memory principle. For a certain user, the proposed model introduces the time weight for each item, and updates the weight by computing the similarity with the items chosen in a recent period. In the recommend process, the weight will be applied to the prediction algorithm. Experimental results show that the modified algorithm can optimize the result of the recommendation in a certain extent, and performs better than traditional collaborative filtering.**

***Index Terms*—collaborative filtering (CF); recommender system; memory principle; time weight;**

I. Introduction

With the development of the web technology and advent of the big data era, huge amount of information consumes much more time. Directional search technology can help to narrow the search scope. However, people sometimes even have no clear search target and have little idea about what kind of information they want to see. Hence traditional search technology cannot satisfy the information filtering alone. After Resnick [1] first put forward the "personalized recommend research" as an independent concept, more and more recommendation methods are proposed. So recommender system now becomes a hot spot of web technology for its promoting the web sales and improving the adhesion of e-commerce website.

Grudy [2] is the first collaborative filtering recommender system which has been put into application. It can create a model based on users' interest and use this model to recommend books to users. Literature [3] [4] make improvements on the similarity calculation, and prove the recommendation precision. As the item-based CF was proposed, it can be computed offline and also ease the sparsity of rating matrix to improve the efficiency of recommend process. Afterwards literature [5] [6] show that algorithm based on the similarity of products performs better than the algorithm based on user similarity. Naturally, item-based CF has achieved success in practical applications since then. Many large-scale e-commerce systems, e.g. Amazon [7], use this method to process data offline and produce recommendations for their users.

Existing collaborative filtering algorithms focus only on the similarity among users and items [8], and only a few consider the dynamic changes of users' interests. In fact, users' attention is changing with time, and traditional collaborative filtering does not take the time factor into recommend generation. Once user's interests have transferred, the system cannot find that change in time, thus recommend resources may deviate from user's actual demand. Taking the time factor into consideration, Ivan Koychev [9] proposes a method of gradual forgetting to make last observation more important than older ones. Literature [10] builds a time function, and let the importance weight of items decay exponential. But this uniform approach seems unfair to those items that get sustained attention from past until now, and they should

*Corresponding author: pc@bigc.edu.cn

doi:10.4304/jcp.8.11.2763-2767

deserve higher weight in prediction phase. Literature [11] plus the similarity weight of resources in addition to time weight, to avoid ignoring valuable data that produced earlier, thus is suitable for handling users' interests repeated. However, if the importance of item decays by the time window blindly, there would be some useful information overlooked in the meanwhile. Obviously, when a certain item has got the user's continuous attention, it means that the item is attractive to this user in a long term, so the item should have a heavier weight in the subsequent process. Recommender system is responsible for recognizing users' long-term interests and forgetting the interference information, and giving items different weights. Notice that and inspired by the forgetting curve [12], this paper considers the recommend process to be similar to the forgotten phenomenon of the brain.

In view of the analysis of the memory principle and interest drifting law, the time weight of an item is considered as an attribute of memory characteristic in this paper, and a weight iteration model based on the memory principle is proposed. The model puts the user's records into several sets by the time window [13], and determines the evolution route of item weight by computing the similarity in different sets. For this method is to imitate the human memory process, it is possible to learn user's preferences accurately and make recommender systems more personalized. Experimental results show that the new algorithm improves the quality of the recommender system, addresses the impact of the time factor effectively.

The rest of this paper is organized as follows. Section 2 introduces the research background, including the traditional collaborative filtering algorithms, forgetting curve and memory principles. Section 3 proposes the weight iteration model and a modified time-weight CF algorithm. Section 4 gives the experiment results with different parameters. Finally, section 5 makes a conclusion of this paper.

II. Background

A. Item-based Collaborative Filtering

Item based collaborative filtering algorithm is most widely used in the current recommender systems, which is based on a hypothesis that people would like to choose the similar items that they have chosen before [15]. Its basic principle is generating the recommendation through user's history records and the item similarity metric. The algorithm can be conducted offline due to the stability of the item similarity. Here we defined some symbols in collaborative filtering algorithm and put out the algorithm procedure [16].

Symbols:

$<U_i, I_j, R_{ij}>$: A 3-tuple records user's selection and rating.

U_i : A set of users in the database

I_j : A set of items rated by user U_j

R_{ij} : Represents the rating of I_j by U_i

I_U : Includes all the items that have been purchased or rated by user U .

$sim(I_i, I_j)$: Similarity between item I_i and I_j , is given by $\cos(\vec{I}_i, \vec{I}_j) = \frac{\vec{I}_i \bullet \vec{I}_j}{|\vec{I}_i| \bullet |\vec{I}_j|}$

Procedure:

Input: user U , I_U

Output: top N recommendations

Step1. For each item i belongs to I_U ,get its neighbors by the similarity matrix.

Step2. Get candidate C by delete the repeat items from the neighborhood.

Step3. For each item j belongs to C ,calculate the recommendation level $\Pr e$ of every j to U , $\Pr e = \sum_{i \in I_U} sim(i, j)$

Step4. Sort the items in candidate C by the $\Pr e$,take the top N items[14] as the recommendations for U .

B. Memory Principle

Memory has the ability to reproduce information stored in the brain. Memory can be divided into three types such as instantaneous memory, short-term memory and long-term memory [17]. The brain can distinguish which type is the memory belongs to according to the storage time and repeat intensity. Forgetfulness is the opposite of memory. It refers to the blurred information due to memory capacity limit and the passage of time.

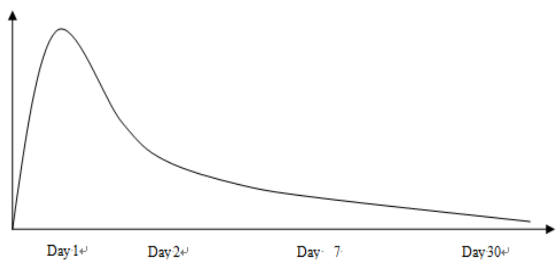

Figure1. Memory phenomenon

As is shown in Fig.1, the amount of information in the brain is to a highest degree in the first day, forgotten phenomenon occurs from the second day and the rate of forgotten becomes lower gradually. Memory can be accumulated [18], so the importance of information can be strengthen through the review. When the same content is repeated, the brain can recognize and keep it better. That is to say, if the brain is exposed to the same information repeatedly, it will take the information seriously and remember them. Specifically, Hermann Ebbinghaus contributed to explain this phenomenon by large number of experiments, and he put up the hypothesis of the exponential nature of forgetting [12].

In this paper we will use this principle for reference, supposing the importance of item reduced along with modified exponential function curve.

Recommendation generating is a process refers to user's behavior, this paper holds that the drift of users' interests is a similar procedure with the information storage in the brain. In view of the memory storage in the brain, this paper put up with a similar conception, the time weight, as an attribute of each item, representing their importance for recommendation procedure. When the user chose a certain kind of item, the weight will be bigger than before, like the process of memory. If the use chose that item no more, that means the item could not get user's attention any longer, like the forgotten process, its weight will decline by a specified function.

III. The Time-Weight Collaborative Filtering Algorithm

In the following parts, we first depict the weight iteration model, and then introduce the model into the item-based collaborative filtering algorithm.

A. The Weight Iteration Model

Some symbols are defined for the collaborative filtering algorithm.

$< U_i, I_j, R_{ij}, T_{ij} >$: A 4-tuple mentioned above, records user's behavior and item's attribute.

I_U: Includes all the items that have been purchased or rated by user U.

w_{ij} : Time weighting coefficient. Denote the importance of j-th item in i-th user's selection set, it's a time-varying figure.

δ: Similarity threshold. Determine the item's weight whether increased or reduced.

λ, μ, k : Characteristic constants. Determined by experiment, different users have different values.

$sim(I_i, I_j)$: Similarity between item I_i and I_j

ε: A threshold value. Whether an item should be removed from database depend on it.

For a given user U, all those projects which have been purchased or rated compose a set denoted by I_U, and they will be divided into different clusters like $I_{U_1}, I_{U_2}, \cdots, I_{U_k}$ by tag T_i, as there are k-th time windows. At the beginning, a certain item is given a weight w_n to represent its importance for recommendation, the initial value is w_0. When first Δt period passed, w_n of the item in I_{U_1} will reduce to a lower degree; simultaneously, there comes some new data into set I_{U_2}, and they need to compare with the items in set I_{U_1} on the similarity, the pre-set δ will help w_n choose their new evolution routes, exponential decline or linear decline. The procedure will be repeated continually. The mathematical description for the iteration is:

$$w_{ij,n+1} = \begin{cases} w_{ij,n} \cdot e^{-\lambda \cdot \Delta t} & if \quad \max(sim(I_{i,n}, I_{i,n+1})) < \delta \\ -kt + (w_{ij,n} + \mu) & else \end{cases}$$

After a round of iteration, there will be a selection on those items according to their weight and the pre-set threshold ε. If $w_n < \varepsilon$, the corresponding item becomes unimportant for the recommend, then the model will delete it from database. For those items that represent user's long-term interests will be picked out, composed of new interests clusters.

Since Ebbinghaus has put up the exponential function to fit the memory decay [12], paper [10] also proved the exponential function fits better than the logistic function, which is another decreasing function. So in this paper, we cite the exponential curve as the decline route directly.

B. The Modified Collaborative Filtering Algorithm

The above-mentioned weight was introduced to the traditional CF algorithm, and an improved algorithm was proposed. The algorithm first gets neighbor sets by similarities among the chosen items, further to produce the candidate set. Let I_U be the set of items that user had selected, get the w by weight iteration function, and take w to calculate the weighted recommended values of a certain item. And then sort the recommended value to generate the top-N recommend items.

Input: user U, I_U

Output: top N recommendations

Step1. For each item i belongs to I_U, get its neighbors by the similarity matrix.

Step2. Get candidate C by delete the repeat items from the neighborhood.

Step3. For each item i belongs to I_U, calculate the time weight w by the iteration function

Step4. For each item j belongs to C, calculate the recommendation level $\Pr e$ of every j to U, $\Pr e = \sum_{i \in I_U} w_i \bullet sim(i, j)$

Step5. Sort the items in candidate C by the $\Pr e$, take the top N items as the recommendations for U.

Fig.2 shows the process flow of the new time weight model.

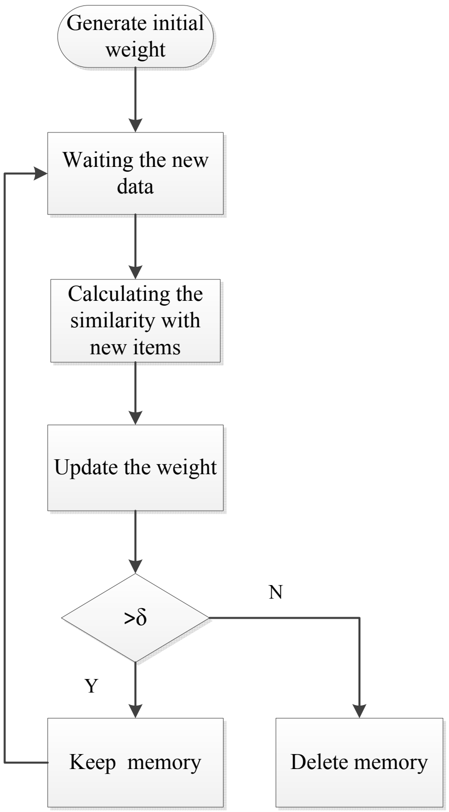

Figure2. Process of time weight model

IV. EXPERIMENTE RESULTS AND ANALYSIS

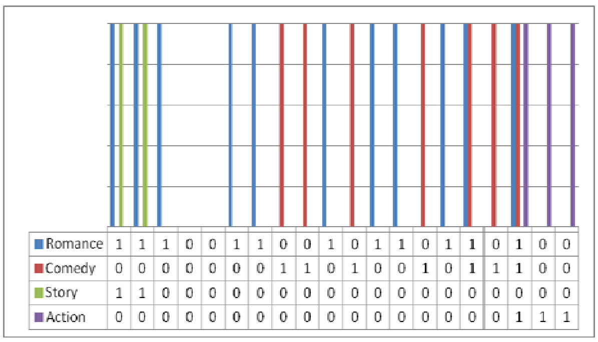

Figure3. Sample data in four months

The experiment made a simulation of the selection records for four months, and it can reflect the drift of users' interests, that including the interference selection and long term attention. There are some sample data shown in Fig.3, these four sample lines respectively represent four different types of movies. And the horizontal axis divides 4 months into 10 segments. If the bar occurs in a certain segment, which means the user has selected this kind of movie in that period.

As shown in Fig.3, the romance movie appears from the beginning to the end, that means the user like this movie type that even not have interest shifting. As to the red bar, denoted comedy movie, occurs from the eighth segment, the user didn't pay attention at first, but after contact he or she began choose this type frequently, finally it becomes another stable interest. The story movie presented by green bar appears no more after the second time. For the Action movie, newly got the user's attention, and have no further information lately. We can see that the behavioral characteristics and hobby of this user in the period. Romantic movies in the top line got the attention along the time, the comedy got attentions from the second month to the fourth month, the story movies in the third line hadn't been selected any more after the first month, and the action movie in the last line got attentions until the last month.

TABLE I.
WEIGHT OF 4TYPE MOVIE IN 4 MONTHS

Romance	1	0.9006	0.8662	0.8522	0.8465
Comedy	0	1	0.9006	0.3662	0.5489
Story	1	0.4006	0.1629	0.0662	0.0269
Action	0	0	1	0.4006	0.6629

Table1 shows the exact time weight value in this period, Here λ=0.9 μ=0.5 .The romantic movies got the attention along the time, hence this kind of movie can keep a high weight in Fig.3, its weight keep a stable decrease from 1.000 to 0.8465 in last two months. As for the comedy movies, an initial value generated in the end of the first month and decline to 0.3662 in the next period, lately this kind of movies got attentions again, and the weight is on the rise then. Since the user selected story movies no more after the first months, its weight has no reason to rise up. At last, the action films' weight is generated in the middle time and slows down lightly.

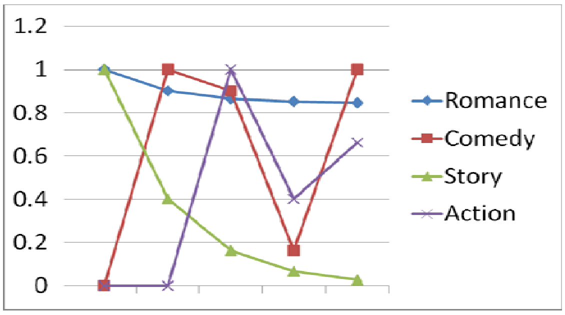

Figure4. Trend of weight

As is shown in Fig.4, this process of weight iteration was depicted by the broken lines. The interference selection means that the user didn't choose this kind of products never again, and its weight should be weakened, the algorithm can recognize the item which have a most low weight value less than ε and remove it from user's records; At the same time, to those of high weight items, since they could get continuous attention of the user, the system will remember them and take as user's long-term interests.

In order to test the feasibility and effectiveness of the model, we take the traditional item-based CF to compare with the new algorithm. Select the records of a previous period of time to speculate the subsequent choice and compare with the real selection with reference to MAE [19].

MAE (Mean Absolute Error) is a statistical accuracy measurement between predictions and real ratings. Each prediction p_i was corresponding to a rating q_i; The MAE is computed by the sum of N pairs of $< p_i, q_i >$ and then gets the average. Formally $MAE = \frac{\sum_{i=1}^{N} |p_i - q_i|}{N}$. Obviously, the more accurate recommendation should get the lower MAE.

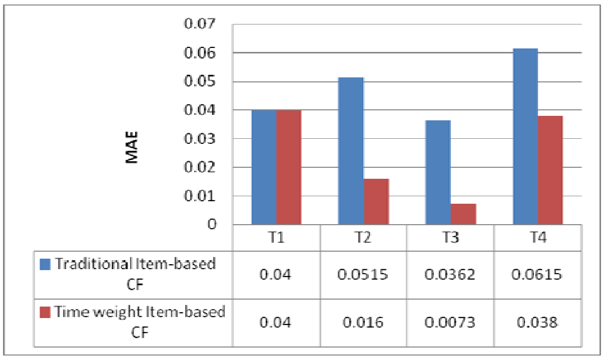

Figure5. MAE using different algorithm

Fig.5 describes the comparison between the traditional CF and new time weight CF. In the first period; the two algorithms have got the same results due to the initialization of the weight at the beginning. With the passage of time, the MAE of the new algorithm is much smaller than the traditional algorithm, because of the importance of interference information been weakened and the long term interesting selection been kept, and all of that make the prediction results more close to the user preferences.

V. Conclusion

This paper researches the weight of items changing with time in the item-based collaborative filtering. To reflect the change of users' interests, this paper proposes a weight iterative model which is inspired by memory principle and the curve of forgetting. The model has a comprehensive consideration both on time window and interests drifting. Experiments show that the new algorithm can improve the accuracy of the recommended.

Acknowledgment

This work is supported by the National Nature Science Foundation of China (No.60973146, No.611 70269, No.61170259, No.61003285, No.61170272), Beijing Natural Science Foundation (No.4122026) and the Fundamental Research Funds for the Central Universities (No. BUPT2013RC0308).

References

[1] Resnick P, Iakovou N, Sushak M, et al. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. Proc 1994 Computer Supported Cooperative Work Conf, Chapel Hill, 1994:175-186
[2] Rich E. User Modeling via Stereotypes. Cognitive Science, 1979, 3(4):329-354
[3] Yang MH, Gu ZM. Personalized Recommendation Based on Partial Similarity of Interests. Advanced Data Mining and Applications Proceedings, 2006, 4093:509-516
[4] Chen YL, Cheng LC. A Time-Based Approach for Effective Recommender Systems Using Implicit Feedback. Expert Systems with Applications, 2008, 34(4):3055-3062
[5] Sarwar B, Karypis G, Konstan J, et al. Item-based Collaborative Filtering Recommendation Algorithms. Proc 10t Int'l WWW Conf, Hong Kong, 2001:1-5
[6] Deshpande M, Karypis G. Item-based top-N recommendation algorithms. ACM Trand Information Systems, 2004, 22(1):143-177
[7] Linden G, Smith B, York J. Amazon.com recommendations: Item-to-item collaborative filtering. IEEE Internet Computing, 2003, 7(1):76-80
[8] Zeng Chun, XingCX,Zhou LZ. A Survey of Personalization Technology. Journal of Software, 2002, 13(10)
[9] Ivan Koychev, Gradual Forgetting for Adaptation to Concept Drift. Proceedings of ECAI 2000 Workshop on Current Issues in Spatio-Temporal Reasoning, Berlin, p. 101-107
[10] Y Ding, X Li. Time Weight Collaborative Filtering. CIKM '05 Proceedings of the 14th ACM international conference on Information and knowledge management. Pages 485-492.
[11] XingCX,GaoFR,Zhan SN, et al. A Collaborative Filtering Recommendation Algorithm in Corporated with User Interest Change. Journal of Computer Research and Development,2007,44(2): 296 301.
[12] Ebbinghaus H. Memory: A Contribution to Experimental Psychology [M]. New York: Dover Publications Inc., 1963.
[13] Shen Jian, Yang Yu-pu. Dynamic Collaborative Filtering Recommender Model Based on Rolling Time Windows and its Algorithm. Computer Science. 2013 40(2).
[14] Zhao Liang, HU Nai-Jing, Zang SZ. Algorithm Design for Personalization Recommendation Systems. Journal of Computer Research and Development. Vol.39, No.8 Aug.2002
[15] Marko Balabanovic, Yav Shoham. Content-Based, Collaborative Recommendation. Communications of the ACM. Vol. 40, No. 3. March 1997
[16] Liu JianGuo, Zhou Tao, Wang BingHong. Progress of Personalized Recommender System. Progress in Natural Science. Vol. 19, No. 1. January 2009.
[17] Zuo Saizhe, Guo Yucui, Gong Shangbao, Wu Xu, Zhang Hu. Trust Value Update Model Based on the Memory Theory. Journal of Southeast University (Natural Science Edition). Vol. 40 Sup (II) Nov.2010
[18] HL Roediger III, F Craik Varieties of Memory and Consciousness: Essays in honour of Endel Tulving [M]. Hillsdale, USA: Lawrence Erlbaum Associates, 1989:5-36
[19] Badrul Sarwar, George Karypis, Joseph Konstan, John Riedl. Item-Based Collaborative Filtering Recommendation Algorithms. ACM 1-58113-348-0/01/0005. May 1-5, 2001

Dan Li, Master. Born in September 1990, in Shandong, China. Study for M.S. degree on Applied Mathematics in Beijing University of Posts and Telecommunications (BUPT). Her research interests include recommendation system and collaborative filtering algorithm.