Time Decay Friend Recommender System for Social Network

Peng Duan[1], Yanhong Li[1], Hua Zhou[2], Li Li[1]

[1] Information Techlology Center,
Yunnan University, China
[2] School of Big data and Intelligent Engineering,
Southwest Forestry University, China
{duanpeng,yhli73,hzhou,lily}@ynu.edu.cn

Abstract. With the rapid development of the Internet, the recommendation system is to solve the problem of information overload and the emergence of a personalized information filtering technology. Collaborative Filtering is the most successful and widely technologies to date. But user's interests are not static, it will vary over time, and in which the social network as well. This paper selects a single forgetting function algorithm to improve research, but different users forgetting rules are not the same, it is the inadequacy of this algorithm. At the same time, the selection of experimental data object is not targeted, it cannot fully considering the characteristic of social platform for all users. The experimental results show that the proposed algorithm can efficiently improve recommendation quality.

Key words: Recommender system, social network, time decay.

1 Introductions

With the continuous development of Internet and mobile applications, information to the explosive speed of the growing information overload problem is increasingly serious. Although the search engine to a certain extent, to meet the needs of users to access to effective information, but still can not effectively solve the information overload problem. Therefore, the recommendation system came into being, it is to solve the problem of information overload on the Internet and proposed a smart agent system, from the Internet a lot of information to the user recommended to meet their preferences or needs of resources, research It is suggested that the effective recommendation can not only satisfy the users' acquisition of the required information, but also have certain effect on the enterprise to improve the customer satisfaction [1].

Nowadays, the recommendation system in e-commerce is increasingly playing its own significant role [2]. Currently, the recommended system is mainly applied in the field of electronic commerce, such as domestic Taobao and Dangdang online store, as well as abroad Amazon. In addition, the recommended system also play a role in other areas, such as Google news recommendation system, the

movie site Netflix on the film recommended, personalized music recommended Ringo, as well as domestic and foreign social networking sites Sina microblog and FACEBOOK friends recommend.

At present, people communicate more and more rely on social software applications, the current mainstream social software, mainly WeChat, Sina microblogging, QQ and so on. Among them, Sina microblog not only can make two strangers or friends on the network to communicate and interact, but also can publish their own feelings or ideas and ideas on a topic, which will attract other people interested in the same topic , And publish their own ideas, so that the two sides can have more in-depth exchanges. In this process, you can see that the impact of time factors can not be ignored [3].

Effective personalized recommendation not only saves users time to get the information they need, but also increases the cost of improving customer satisfaction for many network operators to increase revenue. However, the traditional recommendation system has limitations: it is basically the default user's interest is immutable. However, different users at different times, the views of things and attitude may be different. For example, a user who first likes to watch comedy movies, and after reading a reasoning novel, the preferences may become more inclined to suspense movies, this time can be considered the user to treat the movie preferences have changed. Therefore, it is unreasonable to statically analyze the user's interest and ignore the dynamic changes that may occur, which will directly affect the accuracy of the results. Especially in social networks, the user's comments on other users may have different ideas over time, and the topic of interest will change accordingly, that is, people's interests are not static, for example, a day in the face of user comments or ideas of another user published, endorsed and supported, but contact with a new topic within a certain period of time in the future or affected by other users of speech, then the user may The other user's microblogging content related to the topic and the concept of approval, can be seen as a user of new users concerned about their own preferences have changed [4]. Therefore, we believe that if forgetting to add a function of time, considering the time of impact shift to user interest, dynamic changes to better capture user preferences, in order to make higher quality friends recommend in terms of the user's attention friend The

Based on the above problems, this paper proposes a friend recommendation algorithm based on time forgotten. Compared with the traditional recommendation algorithm, this algorithm adds the time forgetting mechanism to the fusion time forgetting function, and then calculates the user similarity. In this way, you can from the memory impact on the user to start, to a certain extent, solve the problem of user interest drift.

2 Related Work

2.1 Recommender System

Personalized recommendation system includes filtering and user-based content-based filtering [5]. Among them, the basic idea of user-based collaborative filtering is to analyze a user's general rating of the project, including the analysis of the user's historical situation, preferences, behavior, and find the best similarity with another user - "neighbors", and in accordance with the recommended "neighbors" of user preferences to the original target, but there is a problem that poor scalability and a user data transferring method based on a filter, in order to solve such a problem, another collaborative filtering method, which is based on Content of the recommendation, this method is mainly using Bayesian probability model, genetic algorithm and other machine learning methods. With the deepening of this application, collaborative filtering also encountered some problems, mainly confined to the following three aspects: sparsity, scalability and synonymous.

In general, the collaborative filtering algorithm is by searching for a large group of people and finding a smaller set, which is similar to the target user's preferences. The basic mechanism of collaborative filtering is [6]: i) through the behavior, preferences and other factors, a group of people's preferences for analysis; ii) the use of similarity measure, select a sub-group, the sub-group of users and target users have the largest; iii) the weight of the user inside the sub-group of the calculation; iv) with the resulting preference function to recommend the target user, this time the recommendations more in line with the user's preferences. where, commonly used methods are metered similarity cosine similarity, Pearson correlation coefficient calculation. Collaborative filtering recommendation system has many advantages, such as from the establishment of the model to the preparation of the program, the whole process is very clear; applicable range is relatively large, movies, music, social, commodity and many other aspects can be used; Higher degree. But the development is still inevitable some shortcomings: the initial system of the lack of basic calculation of the calculation data, can not be related to the calculation of recommended data, that is, cold start problem; in use when the need for the recommended things and user groups secondary filtering, other problems Wait.

In the traditional collaborative filtering recommendation algorithm, both the user-based collaborative filtering algorithm and the project-based collaborative filtering algorithm have the same problem: directly calculate the similarity between users or between projects, and then measure the user Current interest, making recommendations, without taking into account the user's interest in different periods of time offset. In addition, most of the current site and research is based on the project or resources as a recommended object, and for social applications, friends recommended should also be worthy of a study.

2.2 Similarity Measure

In the user-based collaborative recommendation, the core job is to find the most similar "neighbors", and to find the most similar neighbors, one of the most important indicators is the "similarity", "neighbors" more similar, and finally recommended The higher the quality of the results. At present, the most widely used method is to study the similarity of two methods, one is the Pearson correlation coefficient, the other is the cosine similarity [7]. In addition to the relatively extensive use of these two methods, the similarity of other measurement methods, such as Adjusted Cosine Similarity (ACOS) [8], and Proximity-Impact-Popularity (PIP) [9], etc. These different methods of calculating the similarity can be used according to the specific circumstances, in order to achieve different purposes, such as PIP method can effectively reduce the cold start problem on the recommended system.

The Pearson correlation coefficient [10] calculates the similarity by a common score item or a vector of users. When calculating the correlation of the recommendation system, divided into similarity between computing users, the similarity between computing project. This formula can be used to calculate the similarity between an user and other users.

2.3 Time decay

When the user on the Internet aimless browsing, the recommended system can provide users with relevant data or information, allowing users to quickly find their own information. However, if a user's preferences change more frequently, then only the analysis of the history of browsing is not enough, if you can consider the impact of time factors on user preferences, then the final recommendation of the quality of the results will be greatly improved [11, 12]. However, how to find an effective way, the time as a feature added to the recommended system is a problem worthy of study.

Ebbinghaus on memory processes from observations quantitative analysis, the amount of stored memory will be reduced over time, and start a little faster, gradually slow down and stabilize after some time [13, 14]. According to Ebbinghaus's study, people learn to get information or knowledge, only to leave a part of the retained memory, the rest will be forgotten. Ebbinghaus argues that "keeping and forgetting is a function of time". And according to his experimental results, the Ebbinghaus memory forgetting curve is defined as Equation 1.

$$W(u,i) = e^{\frac{-kd}{L}}, k \in [0,1] \tag{1}$$

Wherein, u represents user, i represents item, d represents the time interval of rating, and L represents the time span of the experimental data.

3 Time Decay Friend Recommender System

3.1 System design

As the system is mainly to solve the social network (here Sina microblog, for example) in the Friends of the recommended problem, we will be mentioned in front of the project changed to microblog users posted microblog forwarding number. In order to improve the quality of the recommendation system recommended by the recommendation system, it is proposed to add the time factor to the recommended algorithm. According to the comparison of the time forgetting function, it is decided to use the Ebbinghaus forgetting exponential function to constrain the similarity and then calculate the final recommendation the result. The main procedures are described as follow:

1. Comments, forwarding time and other information are acquired from the raw data.
2. Time indicator is calculated.
3. Similarity is measured through Pearson correlation coefficient with time indicator.
4. Top-N list of recommended friends is generated according to the similarity.

In the introduction of time decay, the realization of friends to achieve the main realization of the idea is: to determine the original Pearson correlation coefficient in the user's assessment of the project resources in the actual representation, and then to determine the degree of similarity between users. The value of the parameter k in the different time period, from the recent score of the score on the final recommendation of the impact of the degree, if the k value is higher, indicating that the most recent period of the score on the recommended results, than the previous paragraph The score within the time has a higher impact on the recommended results. Thus, the improved Pearson correlation coefficient is presented in Equation 2

$$S(u,a) = \frac{\sum_{i \in C_{ua}} \xi_{u,i}(R_{ui} - \overline{R_u})\xi_{u,i}(R_{ai} - \overline{R_a})}{\sqrt{\sum_{i \in C_{ua}} \xi_{u,i}(R_{ui} - \overline{R_u})^2 \xi_{u,i}(R_{ai} - \overline{R_a})^2}}, C_{ua} = i \in C_{ua}|R_{ui} \neq 0, R_{ai} \neq 0$$

(2)

Wherein, C_{ua} represents the forwarded content set from user u and user a. Moreover, $\xi(u,i) = nW(u,i)$, where n represents forwarded number.

3.2 Module Design

Time Decay Processing This part is the highlight of the whole algorithm. The main research thought of this paper is to solve the problem of user interest drift by adding time forgetting function in the original traditional recommendation algorithm. In order to be able to more clearly describe the whole process, define the following symbols:

- $Time_{begin}$: Reference time, the initial time in the experiment with the time span of the collected data as reference time.
- $Time_{rate}$: Absolute time, if the microblogging is the last time to be forwarded for the absolute time, then this time may be the last time to be forwarded with a larger time interval, including the possibility of special data is large, easy to affect the final score, so select The user publishes the microblogging time for the actual scoring time.
- $Length$: Time span, in the experiment to start collecting and stopping the collection of time intervals as a time span.

According to the above definition, if a user's microblogging original forwarding number is N, the decayed forwarding number can be calculated through Equation 3.

$$N_{dec} = N \times e^{\frac{-k(Time_{rate} - Time_{begin})}{Length}} \tag{3}$$

Normalization As in the Sina microblogging, microblogging forwarding number is affected by a variety of factors, there will be a large gap between the sample situation, in order to make the calculation of the results of convergence faster, this paper through the normalization of different sources of data Unified to a reference system.

This article uses Matlab, after the time after the microblogging is the number of forwarded The results were normalized. The normalized method is chosen as the method of linear function transformation, as presented in Equation 4.

$$y = (x - Min)/(Max - Min) \tag{4}$$

Wherein, x, y respectively represent the raw value and normalized value, while Max and Min respectively represent maximum and minimum values of experimental data.

Similar measure processing After the time decay of the score into the Pearson correlation coefficient formula, after the data after the time decay, the time closer to the data similarity weight should be higher. As in the original Pearson correlation coefficient, the calculation of the user as a user of a project score, and in the actual problem of this experiment, the mapping of a microblogging user microblogging is the number of forwarding, this The process from the original initiative into a passive, so the problem of the mean, this paper is simplified to select the sample time after the normalization of the mean.

In other words, time decay results after treatment N_{dec}. After normalization, it can be substituted into Equation 2. Then according to the similarity, the recommendation list is generated.

4 Experiment

4.1 Experimental Dataset

The data used in the experiment is from the data hall in a period of time on the Sina microblogging part of the user's activities recorded. Mainly by recording the user's microblogging forwarding number, release time to calculate the user feedback information. The time span for extracting data is from January 1, 2014 to May 14, 2014, for a total of 134 days of data. There are 63641 users in the database, 84168 articles about microblogging, 27759 microblogging forwarding relationship. The collection of microblogging content, including microblogging ID, release time, microblogging content, microblogging source, microblogging forwarding number, microblogging comments, microblogging was praise number, published user ID, microblogging subject.

To validate the proposed algorithm, you first need to have a lot of data as a basis. The source of the data can be divided into the following sections:

– A microblogging users of a microblogging released a number of forwarding;
– Record microblogging users microblogging release time;
– The time interval compared with the reference time.

The reason why the user selected microblogging is the number of transponders as the experimental object, but did not select the number of comments as the experimental object, is taken into account in the relevant work before the experiment, randomly selected some users to publish microblogging content to investigate, Found that not all users of microblogging comments which are published for their own point of view, compared to microblogging in the forward after the comments on the more targeted, post-forwarding additional comments than the direct comments and microblogging theme Fit higher, so here is selected to be forwarded.

4.2 Experimental Results

First, 10 users randomly selected as a test set, wherein the initial time from the most recent user is the target user, the user similarity computing the remaining nine. The similarity of these 10 users can be measured through Equation 2. Table 1 illustrates the raw data of these 9 users, except the target user.

Second, the target user other than the user 9 microblog number is forwarded to Equation 1 is multiplied. The accuracy of the time interval d in equation 1 is expressed in days, and the closer the value of k is to 1, the greater the proportion of the result closest to the current time, where $k = 0.6$.

Third, after normalization, the final similarity ordering ban be calculated by Equation 2, which is 6,1,4,3,2,5,8,9.

Table 1. User Information

User	Fans	Post Time	Forwarding
1	12496405	2014/5/11	1225
2	4183439	2014/5/9	3535
3	609489	2014/5/9	1525
4	18249317	2014/5/7	6208
5	11286029	2014/5/5	1067
6	223756	2014/5/4	130
7	196430	2014/4/11	3851
8	424497	2014/2/3	2340
9	1477	2014/1/2	4667

4.3 Analysis

From the results of the similarity, the time forgotten factor has a high effect on the sorting of the last similarity degree, followed by the similarity degree of the forwarding number. For example, the user of the sixth order and the seventh user are relatively large in time when the number of times of forwarding is large, and the time is relatively small. But this can be adjusted by adjusting the size of the k value to adjust the results.

For the sake of comparison, we refer to the method of calculating only the user similarity (using the Pearson correlation coefficient), which has not been added to the time forgetting mechanism, as the traditional algorithm. The improved algorithm with the forgetting mechanism is called WS algorithm. The results of these two algorithms are plotted as a curve. As shown in Fig. 1, it illustrates the temporal affect on similarity.
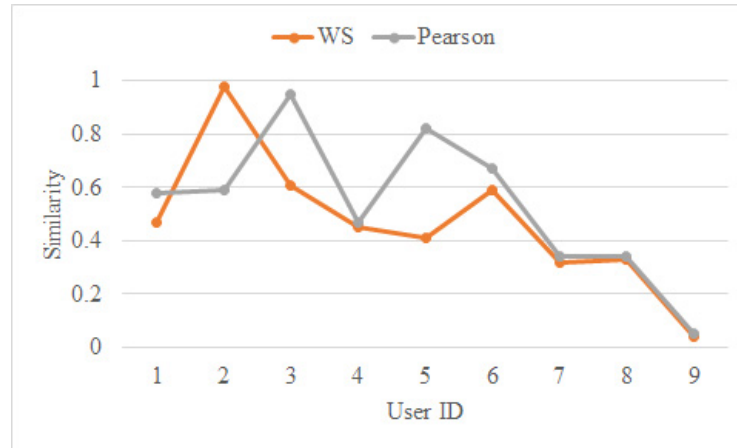
Fig. 1. Temporal affect on similarity.

In order to highlight the characteristics of the improved algorithm in comparison, here is the relative similarity. According to the polyline comparison of Fig.1, compared with the fluctuation of the traditional algorithm, after the time forgotten mechanism is added, it can be seen that the similarity change at this time is declining in general, because the microblogging time is more Far from the user, due to the impact of time attenuation, the lower the similarity. This can be seen that the time forgotten for the user to recommend a certain degree of impact.

Furthermore, in order to evaluate the performance of the proposed scheme, it needs to i) determine whether to meet the same city, with the school and other conditions of the user; ii) determine the performance of the results of the evaluation of the results of the algorithm, based on the original microblogging data, according to Sina microblogging friends recommended procedures - To determine whether to have a common friend; iii) according to the users concerned, the most concerned about the number of users, give the recommended results. Based on the experimental data samples, the recommended table Top-k is obtained. To Top-k as the standard, using precision and recall to test the results, the higher the accuracy and recall rate, said the better performance.

Fig. 2 and Fig. 3 respectively illustrate the comparison between WS and Pearson about precision and recall through different samples.
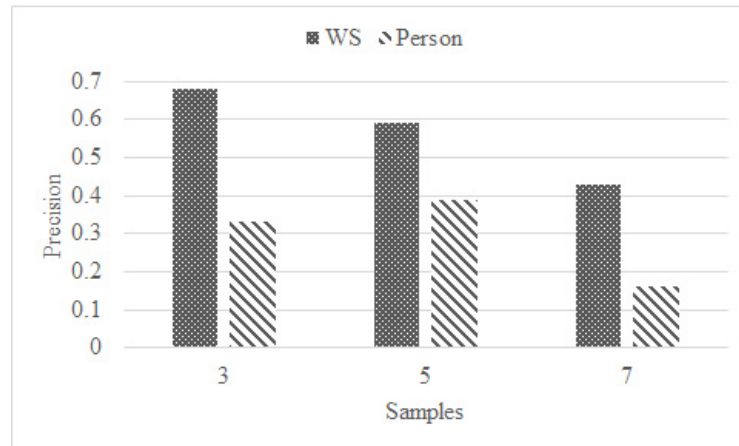

Fig. 2. Precision comparison.

It can be seen from Fig.2 and 3 that when the number of samples of the experiment is 3, 5 and 7 respectively, the improved algorithm has a relatively high value both in accuracy and recall rate. Think that this is due to the improved algorithm more in line with the Ebbinghaus memory forgetting rule. Therefore, the introduction of a time forgetting function can reflect the rules of user memory, thereby reducing the impact of interest drift on the recommended results.
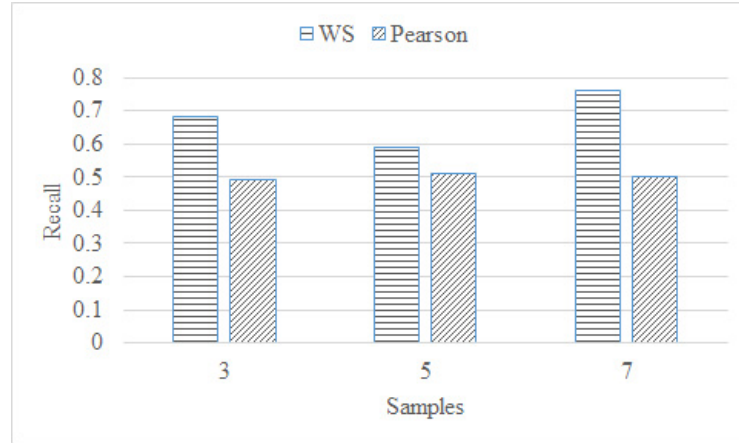
Fig. 3. Recall comparison.

5 Conclusions

According to the user's interest drift phenomenon, we propose a combination of Ebbinghaus forgetting exponential function and Pearson correlation coefficient of an idea, and this idea is mainly implemented in Sina microblogging such a social network of friends recommended process The In this experiment, we will be the user's interest, or in the traditional recommendation algorithm, the score of the project to microblogging users concerned about the degree of microblogging issued by the number of forwarding. Taking into account the memory of the forgotten curve, as the time increases, the curve should be more gentle, so in the time forgotten on the function of choice, we chose the Ebbinghaus forgetting curve as a time forgetting function, in order to gradually reduce the traditional calculation Resource similarity, and then the Pearson correlation coefficient is used to calculate the user recommendation similarity considering the influence of the time factor. From the whole process of this article, such an algorithm to a certain extent, can improve the recommended performance of the recommended system, but also remind us that in the future design of personalized recommendation system in the process, should be more to consider to meet people's natural cognitive rules and Method, to provide users with more effective personalized recommendation service.

In the course of this experiment there are some shortcomings, but also need to further improve. The shortcomings are mainly in the following aspects: (1) in the reference time forgotten function, assuming that everyone's forgotten rules are the same, but obviously this is not reasonable, because in actual life, a person's forgetting efforts not only (2) the object of this article is the user to publish microblogging is the number of transponders, this object is more general, will result in experimental results are compared with the results of the experiment, Big deviation. For example, when a user's number of fans is relatively high, even if the published theme and the concept of the target user is similar, but because

of the higher degree of concern, the number of forwarding will be higher than the more. (3) Sina microblogging on such a social platform, the user can focus on the subject of the subject is very broad, if only from a few of the time, if the topic is very controversial, when the number of forwards can not be used as a suitable object of study; Topic range to discuss, will also result in bias in the recommended results. Therefore, in order to improve the recommendation quality of the algorithm, we should use the corresponding forgetting curve for different situations, and should put a more targeted method when we put it on the actual social platform, and study the user's preference recommendation.

References

[1] J. Song, Y. Zhang, K. Duan, M. S. Hossain, and S. M. M. Rahman, "Tola: Topic-oriented learning assistance based on cyber-physical system and big data," *Future Generation Computer Systems*, vol. 75, pp. 200–205, 2017.

[2] Y. Zhang, M. Chen, S. Mao, L. Hu, and V. C. Leung, "Cap: Community activity prediction based on big data analysis," *IEEE Network*, vol. 28, no. 4, pp. 52–57, 2014.

[3] Y. Zhang, Z. Tu, and Q. Wang, "Temporec: Temporal-topic based recommender for social network services," *Mobile Networks and Applications*, pp. 1–10, 2017.

[4] Y. Zhang, "Grorec: a group-centric intelligent recommender system integrating social, mobile and big data technologies," *IEEE Transactions on Services Computing*, vol. 9, no. 5, pp. 786–795, 2016.

[5] Y. Zhang, D. Zhang, M. M. Hassan, A. Alamri, and L. Peng, "Cadre: Cloud-assisted drug recommendation service for online pharmacies," *Mobile Networks and Applications*, vol. 20, no. 3, pp. 348–355, 2015.

[6] B. Yang, Y. Lei, J. Liu, and W. Li, "Social collaborative filtering by trust," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 8, pp. 1633–1647, 2017.

[7] Y. Zhang, M. Chen, D. Huang, D. Wu, and Y. Li, "idoctor: Personalized and professionalized medical recommendations based on hybrid matrix factorization," *Future Generation Computer Systems*, vol. 66, pp. 30–35, 2017.

[8] T. C.-K. Huang, Y.-L. Chen, and M.-C. Chen, "A novel recommendation model with google similarity," *Decision Support Systems*, vol. 89, pp. 17–27, 2016.

[9] Y. Wang, J. Deng, J. Gao, and P. Zhang, "A hybrid user similarity model for collaborative filtering," *Information Sciences*, vol. 418, pp. 102–118, 2017.

[10] F. R. Almeida, A. Brayner, J. J. Rodrigues, and J. E. B. Maia, "Improving multidimensional wireless sensor network lifetime using pearson correlation and fractal clustering," *Sensors*, vol. 17, no. 6, p. 1317, 2017.

[11] C. C. Aggarwal, "Time-and location-sensitive recommender systems," in *Recommender Systems.* Springer, 2016, pp. 283–308.

[12] V. W. Anelli, V. Bellini, T. Di Noia, W. La Bruna, P. Tomeo, and E. Di Sciascio, "An analysis on time-and session-aware diversification in recom-

mender systems," in *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization.* ACM, 2017, pp. 270–274.

[13] J. Chen, F. Tang, J. Xiao, J. Li, J. He, and Y. Tang, "Cogtime_rmf: regularized matrix factorization with drifting cognition degree for collaborative filtering," *Cluster Computing*, vol. 19, no. 2, pp. 821–835, 2016.

[14] Z.-y. Yu, J.-d. Wang, H.-w. Zhang, and K. Niu, "Services recommended trust algorithm based on cloud model attributes weighted clustering," *Automatic Control and Computer Sciences*, vol. 50, no. 4, pp. 260–270, 2016.