2012 IACSIT Hong Kong Conferences
IPCSIT vol. 30 (2012) © (2012) IACSIT Press, Singapore

Collaborative Filtering Algorithm Adapting to Changes Over Dynamic Time

Qian-Qian Zhang [1, 3+] and Ning Ye [1, 2]

[1]School of Information Science and Technology, Nanjing Forestry University, China
[2]School of Computer Science and Technology, Shandong University, China
[3] WaveO Network PteLtd, Singapore

Abstract. Collaborative filtering (CF) as one of the successful technology is widely used in personalized recommendation systems. Considering that existed CF algorithms do not reflect the changes in user interests and item popularity over time, two methods would be introduced in this paper to improve the performance of recommendation systems: first, providing new CF algorithm which is combined of both time-based weighted on user interests and time-based weighted on items, second, reconstructing user-item matrix by logistic function in the first stage on existed CF algorithm. Experimental results show that our improved algorithm can meliorate the recommendation quality of the collaborative filtering recommendation systems.

Keywords: collaborative filtering, time-based weighted on user interests, time-based weighted on items, recommendation system, MAE.

1. Introduction

Collaborative filtering is the most successful technology [1,2] used in recommendation systems, It is based on the assumption that the best way to find interestedinformationof a user is to find users who have similar interests with him firstly, then recommend interested information of these users tohim. The basic idea of CF is to calculate the similarity between the target user andevery basic user, find thenearest neighbours of the target user, and then generate recommendation to him by ratings from his nearest neighbours [3].

User interests change dynamicallyover time, this requires the recommendation system can reflect this dynamic change in time andgenerate more accurate recommendation set to the target user. The advantage of the traditional algorithms isno need to considerthe manifestation of the items, canhelp users find new items they may be interested in, but it did not take into account that the user interests change over time, thusgiveout-dated or currentlyuninterested items to the target user, affect the accuracy of the recommendation. Besides, there are changesin availability and popularity of oneitem over time, such as seasonal changes or a specific holiday.

In this paper, we proposed a CFalgorithm, which is combined of time-weighted on user interests and item popularity, generated item set to the target user. In addition, we reconstructed user-item matrix by adding weight to the ratings of original user-item matrix and generated an amendment user-item matrix, then proceeded similarity calculation and recommendation, improved the efficiency of recommendation.

2. Preparation

Collaborative filteringalgorithm containsthree stages generally: inputting user-item matrix, generating the nearest neighbors of target usersand generatingrecommendationstotargetusers.

Corresponding author. Tel.: +65 97597979; fax: +65 63968132.
E-mail address: *qianqian@waveo.com.*

(1) User-item matrixis a $m \times n$ matrix Rmn, as (1) shown, where, rows mean m users, columns mean n items, r_{ij} in row i and column j means rating of user i on the item j. Itis generallyobtained by real ratings from users. The ratingis usuallyexpressed in figures, for example,usingan integerfrom0 to 5to represent theuserinterestsin specific items.

$$R_{mn} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix} \tag{1}$$

(2) Generatingthe nearest neighbors ofthe target useristhe core of CFalgorithm.There are three main traditionalmethods of calculatingthe similarity: Pearson Correlation Coefficient, Cosine-based Similarity and Adjusted Cosine Similarity [4]. In this paper, we used Pearson CorrelationCoefficient to calculatethe nearest neighbors of the targetuser.

Pearson Correlation Coefficient assumes that $I_{uv} = I_u \cap I_v$ is the item set has ratings from both user u and v, thus the similarity between user u and v is $sim(u,v)$, just as (2), where r_{uk}, r_{vk} represent ratings from user u and v on item k respectively. $\overline{r}_u$, $\overline{r}_v$ represent the average ratings from user u and v on all n items.

$$sim(u,v) = \frac{\sum_{k \in I_{uv}} (r_{uk} - \overline{r}_u)(r_{vk} - \overline{r}_v)}{\sqrt{\sum_{k \in I_{uv}} (r_{uk} - \overline{r}_u)^2} \sqrt{\sum_{k \in I_{uv}} (r_{vk} - \overline{r}_v)^2}} \tag{2}$$

(3) Generating recommendation:Two recommendation resultscan be calculated by the nearest neighbors of the target userbased onstep(2):

a. Userinterestsin anyitem: Assumed that N_u is the nearest neighbor set of the target user u, then predicted rating on item i of user u is P_{ui}, as (3)[5].

$$P_{ui} = \overline{r}_u + \frac{\sum_{v \in N_u} sim(u,v)(r_{vi} - \overline{r}_v)}{\sum_{v \in N_u} sim(u,v)} \tag{3}$$

Where, $\overline{r}_u$, $\overline{r}_v$ mean the average ratings of user u and v on items respectively, $sim(u,v)$ means the similarity between user u and v, r_{vi} means rating on item i of user v.

b. $top-N$ recommendation set: Selected N items as $top-N$ recommendation set which had higher ratings and not in the item set thatthe target user hadrated.

3. Collaborative Filtering Algorithm with Changing of Dynamic Time

To makethe userratings on itemsreflect thechanging of time, we amend rating r_{ui} on item i of user u, and get the formula $r_{ui}' = r_{ui} + f_u(t) + f_i(t)$, where r_{ui}' is amended rating on item i by user u, $f_u(t)$ and $f_i(t)$ are functions which represent changes of userinterests and item popularity. Incidentally, the selection of functions $f_u(t)$ and $f_i(t)$ is alsoworthy of study.

3.1. Time-weighted on user interests

User ratingsare notconcurrent, which means that user ratings are changing over time. To make recommendation systemreflectthe changes in user interests timely, we add influence function $f_u(t)$, which reflects user interests changing over time to the actual rating on item i by user u thus amend the actual rating of user u on item i and get formula $r_{ui}' = r_{ui} + f_u(t)$. For different users, the trends over time of their interests are different; we canpersonalize specifictrends for users to obtainmore accurate recommendation by adjusting influence function $f_u(t)$.

3.2. Time-weighted on item popularity

Similarly, we know thatthe item popularity alsochangesover time, which affects user ratings on them to a certain extent. For example,whenan item first appeared, it is always popular, but this will be weakened over time,because peoplealwayshavecuriouspsychological on newitems.So to makerecommendationsystemreflectthe changes in item popularity timely, for further, we add influence function $f_i(t)$, which reflects the changes in item popularity over time to the former formula, $r'_{ui} = r_{ui} + f_u(t)$ and get the final formula $r'_{ui} = r_{ui} + f_u(t) + f_i(t)$, apparently, we can personalize specifictrend for each item by adjusting influence function $f_i(t)$.

After integrating time-weighted on user interests and time-weighted on item popularity, we got amended rating of user u on item i r'_{ui}, the formula is

$$r'_{ui} = r_{ui} + f_u(t) + f_i(t) \tag{4}$$

3.3. Description of collaborative filtering algorithm changing over dynamic time

(1) Input: User-item matrix R_{mn} is rating matrix that m usersratedon n items; we use PearsonSimilarity Algorithm to calculate similarity $sim(u,v)$ between two users; the item set rated by both user u and v is I_{uv}; the number of recommendations is N.

(2) Output: Any target user u and his $top-N$ recommendationset ;

(3) Process:

From formula (2) $sim(u,v) = \frac{\sum_{k\in I_{uv}}(r_{uk}-\overline{r}_u)(r_{vk}-\overline{r}_v)}{\sqrt{\sum_{k\in I_{uv}}(r_{uk}-\overline{r}_u)^2}\sqrt{\sum_{k\in I_{uv}}(r_{vk}-\overline{r}_v)^2}}$, we calculate the similarity between the target user u and any other user, where $N_u = \{N_1, N_2, \cdots, N_k\}, u \notin N_u$ and $sim(u,N_1) > sim(u,N_2) > \cdots > sim(u,N_k)$, the nearest neighbors of the target user is predetermined at k, that is selecting the largest k nearest neighbors based on $sim(u,v)$.

According to ratings from the k nearest neighbors $N_1, N_2, \cdots, N_k$ of the target user rated on item j, we can predict rating P_{uj} from the target user u on item j based on (5).

$$P_{ui} = \overline{r}_u + \frac{\sum_{v\in N_u} sim(u,v)[r_{vi} + f_u(t) + f_i(t) - \overline{r}_v]}{\sum_{v\in N_u} sim(u,v)} \tag{5}$$

Sort P_{ui} obtainedfrom b by descending, recommendthe first N itemsto the target user u.

3.4. Some ideas on selection of functions $f_u(t)$ and $f_i(t)$

The core of our improved collaborativefilteringalgorithm in this paper is the selections of functions $f_u(t)$ and $f_i(t)$. Appropriate selections of $f_u(t)$ and $f_i(t)$ canimprove theaccuracy ofthe algorithm to a large extent, reflect the influence of time changing on user interests and item popularity timely, improvethe quality of recommendation, make therecommendation systemhas a gooduserexperience, give full play tothe value ofits application. The followingwe put forward some thoughts on selectionsof these two functions:

(1) The introduction offunction $f_u(t)$ isto make our improved collaborative filtering algorithm reflect the changes in user interests over time, generally, recent ratings from a user are more valuable than earlyratings, therefore $f_u(t)$ mustreflect this feature, based on this consideration, we assume that it is a non-decreasing function. But at the same we cannot ignore the value of some early ratings, so we add an adjustmentparameter α to $f_u(t)$, that is $\alpha f_u(t)$.

(2) The function $f_i(t)$ reflects the influence on user rating from the popularity degree ofthe item, because some certaintypes of items (such as electrical items, skin care items and so on) are time-sensitive, in differentseasons, thepopularity ofthese itemsmay change more or less because of the differences among the needs of users; Besides, peoplealwayshavecuriosity on newitems, therefore, thepopularity of a new item often higher when it first appeared, but as timegoes by, the item popularity will showthe trend ofdecreasing.

Based on aboveanalysis, weassume that $f_i(t)$ is a time-segmented function, and is a descending function foremerging time of different items.

(3) Interestsbetween differentusers show differentvariations over time, different items have different features and the popularity changes are different over time, in the algorithm how to select different methods and appropriate functions $f_u(t)$, $f_i(t)$ for different users is important, we may getbetterrecommendations by appropriate selections.

4. Collaborative Filtering Algorithm Based on Reconstructed User-Item Matrix

4.1. Introduction of algorithm

As time affects user interests, we can also improve the CF algorithm by adding weightedtime parameter tospecificratings, giving larger weight to user recent ratings and smaller weight to early ratings. We use logistic to reconstruct user-item matrix, below is the detail:

(1) Gave unique weight to every rating by $\log istic(t_{ui})$ when reconstructing user-item matrix, and replaced the actual rating r_{ui} with $r_{ui} \times \log istic(t_{ui})$, scaledup the original rating. The format of $\log istic(t_{ui})$ is $\log istic(t_{ui}) = \frac{1}{1+e^{-t_{ui}}}$ [6], where $-1 \le t \le 1$, $0 < \log istic(t_{ui}) < 1$ and t_{ui} means the rating value on item i of any user u any time t.

As a monotonicallyincreasing function, the output of $\log istic(t_{ui})$ keeps increasing by time t and remainsintherange (0, 1). Add logistic to original rating means largercontribution from user recent ratings but early ratings reflectpastinterests ofuser, so accounts forless weight.

In thelogisticfunction, the author used standard variables in advance to limit the time t in the range -1 to 1before the experiment, because in this range, they behavealmostlinearandsmall changes in time variable t will lead to small changes in weight, therefore, the recommendation algorithmcan trackchanges inuser interests moreaccurately.

We assume that $r'_{ui} = r_{ui} \times \log istic(t_{ui})$ is amended rating on item of the target user; we can change (1) to (6) by reconstructing user-item matrix Rmn based on original ratings.

$$R'_{mn} = \begin{bmatrix} r'_{11} & r'_{12} & \cdots & r'_{1n} \\ r'_{21} & r'_{22} & \cdots & r'_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r'_{m1} & r'_{m2} & \cdots & r'_{mn} \end{bmatrix} \tag{6}$$

(2) In thestage ofsearching thenearest neighbors ofthe target user, we calculated the similarity between the target user and every other user in database; selected the highestsimilarityas thenearest neighbors. Here, we used thePearsonCorrelationFormula (2) to predict ratingon item i of the target user u P_{ui} by (7), where $\overline{r'_v}$ is amended average rating on items of user v.

$$P_{ui} = \overline{r}_u + \frac{\sum_{v \in N_u} sim(u,v)(r'_{vi} - \overline{r'_v})}{\sum_{v \in N_u} sim(u,v)} \tag{7}$$

(3) In thestage of generating recommendation, sorted P_{ui} by descending and recommended the first N items to the target user u.

4.2. Metrics and methods of algorithm

In methods for evaluating the quality ofrecommendationsystems, MAE(Mean Absolute Error) is an easy and intuitivemeasureto evaluate the quality ofrecommendation, so it is widely used in recommendation algorithms. In this paper,we used MAEto measure the predictionaccuracy by deviation between predicted ratings and actual ratings of user; the smaller MAEmeans the higher recommendation quality, thus themoreaccuratealgorithm. We assumed that $\{p_1, p_2, \cdots, p_n\}$ is predicted rating set on items from the target user and $\{q_1, q_2, \cdots, q_n\}$ is actual rating set from the target user, and then MAE can be defined as (8), and we can get overall MAE formula (9) by taking average on $MAE_i (i=1,2,\cdots,m)$.

$$MAE_i = \frac{\sum_{i=1}^{n} | p_i - q_i |}{n} \tag{8}$$

$$MAE = \frac{\sum_{i=1}^{m} MAE_i}{m} \tag{9}$$

The result of CFalgorithmis good or notdepended on the number ofthe target user's nearest neighbors, so in the experiment, we divided the experimentaldata intotwo parts: training setandtesting set. We compared our algorithm to traditional CF algorithm as the nearest neighborset sizegradually increasing.

4.3. Experimental analysis

(1) Selection of experimental dataset

In this paper, we selected Movie Lensdataset as our experimentaldata. Movie Lens Site is a research-orientedrecommendation system based on Web, used tocollect user's ratingsonmovies, and provideda list ofrecommendedmovies. It achieved greatsuccess sinceits inception in1996. At currentusersof the systemhave more than43,000;items whichusersrated havemore than 1600. Movie Lens dataset consists of943users'100,000rating data from 1 to 5 and 1682 movies, each userrated at least20movies, andmore importantly, thedatasetcontains thetime attributenecessary tothis paper. We divided rating data to training set and testing set bythe ratio of0.8. Higher rating data means that higher userinterestsonthemovie.

(2) Experimental process

In the experiment, we chosethe first 20 – 30 users as the nearest neighbors of the target user, becausethe numberof neighborswithinthis rangeis abetter choice to predictcurrent interests of the target user[9]. Then wecompared MAE of our algorithm with traditional CF algorithm asthe number changedfrom 20to 30(interval2), the experimental result is as shown in fig.1.

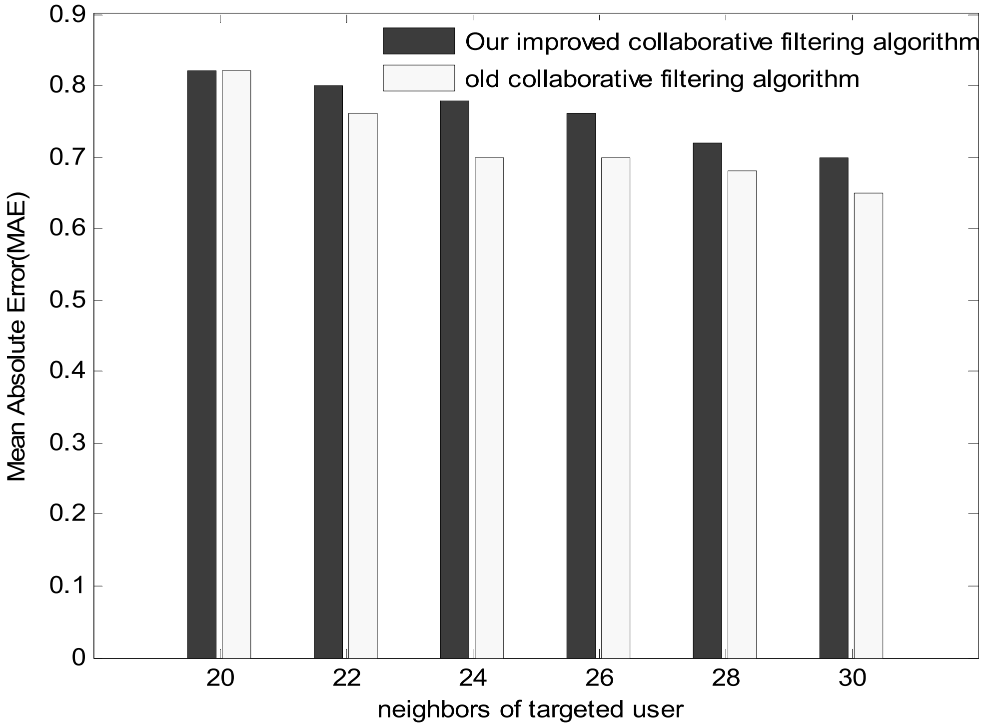

Fig. 1:Comparison of our CF algorithm and traditional CF algorithm

(3) Analysis of experimental result

From figure (1), we know that the average absolute errorMAEof our improved algorithm issmaller thanthe traditionalCFalgorithm, it can be seen,improved algorithmin this paper is moreaccuratecompared tothe traditionalCFalgorithm, therefore, the quality of recommendation system based on thisalgorithm will be improved. From figure (1) we also know that MAE became smaller asthe increasing inthe number ofthe target user'snearest neighbours, so it is also important to find the bestnumber ofthe target user's nearest neighbours.

5. Conclusions and Outlook

In this paper, we proposed twoimproved algorithms: new CF algorithm which is combined of time-weighted on user interests and time-weighted on item popularity and the other is reconstructing user-item matrix. The secondweightedalgorithmis simple and hasreal-time adaptation tochanges inuser interests, experimental results showtheeffectiveness ofour algorithm, it cancaptureuserinterests in real time, and has higher accuracy.

We may also consider usingSVD on our improved user-item matrix to solve the datasparsenessproblem of CF algorithm, improve theaccuracyof predictedratings furthermore, thus improveefficiency andaccuracy ofthe recommendation system.

6. Acknowledgements

This research supported by the National Basic Research Program of China (973 Program)(2012CB114505), China National Funds for Distinguished Young Scientists (31125008), Natural Science Foundation of Jiangsu Province (BK2009393), Jiangsu Qing Lan Project and Scientific and Technological Innovation Foundation of Jiangsu (CXLX11_0525).

7. References

[1] Greg Linden, Brent Smith, and Jeremy York, Amazon.com recommendations: Item-to-Item collaborative filtering [J]. IEEE Internet Computing, 2003, 7(1): 76-80.

[2] G Adomavicius, A Tuzhilin, toward the next generation of recommender systems: A survey of the state-of-the art and possible extensions [J]. IEEE Trans on Knowledge and Data Engineering, 2005, 17(6): 734-729.

[3] Tao Li，JianDong Wang, FeiYue Ye, A collaborative filtering algorithm based on user clustering [J]. System Engineer and Electronic Technology, 2007, 29(7): 1177-1178.

[4] Sarwar B, Karypis G, Konstan J, et al. Item-Based collaborative filtering recommendation algorithms[C],*Proceedings of the 10th International World Wide Web Conference*, 2001: 285-295.

[5] Breeze J, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering [C] ，*Proceedings of the 14th Conference on Uncertainty in artificial Intelligent*, 1998: 43-52.

[6] XiaoQi Cong, HuaiZhen Yu, MeiLian Liu, Weighted time-basedcollaborative filteringalgorithm study [J], ComputerApplication andSoftware, 2009,26(8).

[7] Y Koren, Factorization meets the neighborhood: a multifaceted collaborative filtering model. *Proc, 14th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD' 08)*, 2008, 426-434.

[8] Min Zhu，Bo Su，SVD-based collaborative filtering recommendation algorithms, Computer Security, 2010, (7).

[9] Jin Liu , QianPing Wang , Kun Fang . An Optimized Collaborative Filtering Approach Combined of Item-Based Prediction [C]. *Proceeding of the 2007 11th International Conference on Computer Supported Cooperative Work in Design*, 2007: 158.