A Collaborative Filtering Recommendation System with Dynamic Time Decay

Yi-Cheng Chen[1], Lin Hui[2], Tipajin Thaipisutikul[3], Hsuan-Li Chen[4]
[1]*Department of Information Management, National Central University, Taiwan*
[2]*Department of Innovative Information and Technology, Tamkang University, Taiwan*
[3]*Department of Computer Science and Information Engineering, National Central University, Taiwan*
[4]*Department of Computer Science and Information Engineering, Tamkang University, Taiwan*

ycchen@mgt.ncu.edu.tw 121678@mail.tku.edu.tw 106582601@cc.ncu.edu.tw 606780053@s06.tku.edu.tw

***Abstract*—Collaborative Filtering (CF) technique has been widely utilized in recommendation system due to the precise prediction of users' interests. Most prior CF methods adapt the overall rating to make the prediction by collecting preferences information from other users. However, in real applications, people's preferences usually vary with time; the traditional collaborative filtering could not properly reveals the change of users' interests. In this paper, we propose a novel CF-based recommendation, *DDCF*, which captures the preference variations of users including the concept of dynamic time decay. We extend the idea of human brain memory to specify the level of a user's preferences (i.e., instantaneous, short-term, or long-term). According to this concept, DDCF dynamically tunes the decay function based on users' behaviors. The experimental results indicate that DDCF performs better than traditional collaborative filtering with dynamic decay function consideration. The experiments conducted on real dataset also show the practicability of proposed DDCF.**

Keywords- collaborative filtering; decay function; human brain memory; recommendation system

I. INTRODUCTION

Without any doubt, the recommendation system plays an essential role in this information explosion era, since it extracts and predicts what users may want or need from the extremely huge amount of published information or data. The recommendation system has been applied in variety of areas, such as e-commerce, travel recommendations, online video platform, and social tagging prediction, to name a few. Collaborative filtering (CF) is one of the most successful recommendation techniques and has been widely utilized due to the precise prediction of users' interests. Currently, CF can mainly be categorized into user-based and item-based according to the similarity calculation and preference prediction. The user-based and item-based CF considers and derives the similarity between users and items, respectively, and then predicts the rating of the target item based on the exploited similarities.

Most of the prior CF-based recommendation methods (including user- and item-based) usually adapt the overall rating to make the prediction by collecting preferences information from other users. However, in real applications, people's interests usually vary with time; the traditional collaborative filtering could not properly reveals the change of users' preferences. For example, almost all little girls love Barbie doll, but most of them are no longer interesting with it when they grow up.

In this paper, we propose a novel CF-based recommendation, Dynamic Decay Collaborative Filtering (abbreviated as *DDCF*), including the concept of dynamic decay function. DDCF can capture the preference variations of users and depict the evolution of interests. Actually, the interest and memory retention are very similar. People's preferences usually decay and vary with time. As shown in Fig. 1, we could use the Ebbinghaus forgetting curve [24] of human memory to directly describe the change of user's interest. Obviously, the number of reviews effects the preference decay status. Hence, we extend the idea of human brain memory to specify the level of a user's preferences (i.e., instantaneous, short-term, or long-term). Different level of preference, DDCF determines the appropriate decay function based on users' behaviors.

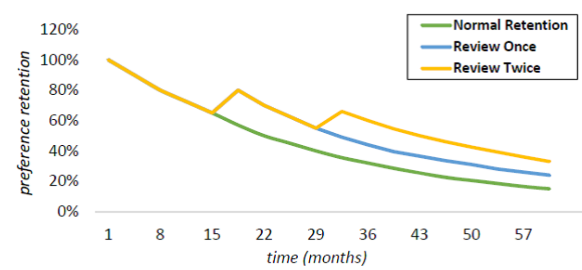

Fig. 1: Evaluation of preference retention by using brain memory curve

The contributions of our work are as follows,

- We point out the significance of time factor in recommendation system, i.e., the interest of user may vary with time. An elegant recommendation algorithm should gradually attenuate the impact of old data and accurately predict the users' future preferences. Tin this study, the preference decay concept has been discussed and included into the CF-based recommendation. We also extend the idea of human brain memory model to describe the preference evolution and decay.
- With the decay function consideration, a novel algorithm, DDCF, has been proposed to effectively recommend items based on users' preferences. To tackle the cold stat and sparsity issues of recommendation system, DDCF utilized the item

clustering to group the similar items together without any predefined parameter.

- Different to previous related studies, we propose a dynamic decay method in this study. DDCF specifies the preference level of items, i.e., instantaneous, short-term, or long-term level, and dynamically determines decay function based on users' rating behaviors.
- To show the practicability of proposed algorithms, we apply DDCF on real datasets. The experimental studies indicate that proposed methods have are both effective and scalable and outperforms state-of-the-art CF-based algorithms.

The remainder of this paper is organized as follows. Section 2 and 3 provide the related work and some preliminaries, respectively. Section 4 describes the DDCF algorithm. Section 5 gives the experiments and performance study, and we conclude in Section 6.

II. Related Works

2.1 Collaberative filtering-based recommendation

SCF [13] combines item- and user-based collaborative filtering techniques together for recommendation. Authors also mention that user-based CF is only suitable for popular item recommendation; by observation, for unpopular items, we should use the item-based CF instead. Zhou et al. [14] utilize a bi-clustering method grouping the items with the order-preserving matrix and then integrate the similarity calculation into the user-based CF recommendation system. Cai et al. [17] borrow the idea of object typicality from cognitive psychology and propose a typicality-based collaborative filtering recommendation system, TyCo. Instead of deriving the similarity by neighbor users, TyCo has more accuracy for prediction based on object typicality calculation.

Niemann et al. [27] propose a collaborative filtering approach based on the items' usage contexts. This approach increases the rating predictions for niche items with fewer usage data available and improves the aggregate diversity of the recommendations. Ma et al. [21] propose a CF-based method combined kmeans clustering, and improve the result with SOM. SOM could do a rough cluster preprocessing as an input, since kmeans clustering need proper *k* setting to get better result. Zhang et al. [8] use a two-layer selection scheme to improve the quality of selected neighbor for CF recommendation. Two-layer neighbor selection consists of two parts, availability evaluation module and trust evaluation module. The modules are used to calculate user influence and improve recommendations.

Some prior studies utilize matrix factorization for improving CF-based recommendation. Nie et al. [15] develop a 3rd-order tensor factorization integrating CF-based techniques for recommendation. Authors also use some latent characteristics to improve the accuracy. Chen et al. [8] propose a tri-factorization method based on orthogonal nonnegative matrix decomposition. After combining with CF method, proposed methods could handle the data sparsity issue effectively. Yehuda et al. [16] propose a multifaceted CF Model, which combines baseline estimates, neighborhood model, and latent factor model [7], to significantly improve the accuracy of similarity calculation and output prediction.

By using global preference and interest-specific latent factors, Kabbur et al. [18] proposed a nonlinear matrix factorization method to recommend top-*n* items that users may be interesting. Pirasteh et al. [26] enhance the recommendation system by exploiting matrix factorization with asymmetric user similarities. Intuitively, two users should be similar when they have common neighbors, even though they do not have any co-rated item.

2.2 Decay collaberative filtering-based recommendation

As already mentioned, the user preferences usually change with time. Some previous works on recommender have investigated how to incorporate temporal information into CF-based approaches. Ding [6] mentions the importance of time weight in CF-based methods for recommendation. The accuracy of prediction of collaborative filtering may be gradually not influence by the old data. Actually, this concept is intuitive, since the users' preference usually vary with time. Wu et al. [4] use power decay function combining user- and item-based collaborative filtering for social tagging label prediction in digital library.

Lee et al. [5] construct a pseudo-rating CF method by using the implicit feedback data. Authors consider the user's purchase time and the item's rating time for weight decay to improve the recommendation accuracy. Gong et al. [11] propose a method to evaluate user's interest change and combine with CF model. Authors use a fixed weight to decay all users' ratings based on item rating time. Santiago et al. [9] discuss the advantages and disadvantages of each decay function using for CF. The experimental result also indicate the post-processing time of each decay combined to CF-based recommendation.

To the best of our knowledge, most prior studies utilize one decay function to evaluate and describe the user preference change. Obviously, only one decay function may not properly describe the complex preference variation of user. In this paper, due to the similarity of preference and memory, we utilize the memory principle of human brain to build a model which with multiple decay function consideration based on the number and time of item rating.

Here we give some related studies about human memory principle. Memory is the ability to reproduce information stored in the brain. Usually, researchers divide memory into three phases, instantaneous memory, short-term memory and long-term memory [30, 31]. Instantaneous memory storage time is very short, and information could be forgotten very fast. On the contract, the information storage in short-term memory could be stayed longer in human brain than instantaneous memory, but still will be forgotten after a while. The information stored in long-term memory phase is able to stay for a long time and not easily forgotten by people.

III. PRELIMINARIES

Suppose that there are a set of users $U = \{u_1, \ldots, u_n\}$, and a set of items $O = \{o_1, \ldots, o_m\}$ in a recommendation system. A rating record is a pair (r_{ij}, t_{ij}) where r_{ij} and t_{ij} are the rate and time of user u_i rating item o_j, respectively. The rating set e_i is the collection of all rating records of user u_i. A user rating vector is defined as $\overrightarrow{u_i} = \langle (r_{i1}, t_{i1}), (r_{i2}, t_{i2}), \ldots, (r_{im}, t_{im}) \rangle$, i.e., rating records in e_i with respect to all items in O. Note that if user u_i does not rate item o_j, the value of r_{ij}, t_{ij} in $\overrightarrow{u_i}$ are both zero. A rating matrix in a recommendation system is defined as,

$$M = \begin{bmatrix} \overrightarrow{u_1} \\ \overrightarrow{u_2} \\ \vdots \\ \overrightarrow{u_n} \end{bmatrix} = \begin{bmatrix} (r_{11}, t_{11}) & \cdots & (r_{1m}, t_{1m}) \\ \vdots & \ddots & \vdots \\ (r_{n1}, t_{n1}) & \cdots & (r_{nm}, t_{nm}) \end{bmatrix},$$ where n and m are the number of users and items, respectively.

Definition 1 (Decayed Rate)
Assume that the current time is t. The decayed rate of a rating record is

$$D(r_{ij}, t_{ij}) = decay_{\mathcal{L}}(\Delta t) \times r_{ij}, \qquad (1)$$

where $\Delta t = t - t_{ij}$. The decay function $decay_{\mathcal{L}}(.)$ could be linear, logistic, power or exponential decay, to name few. The concept and example of decay functions is shown in Fig. 2.

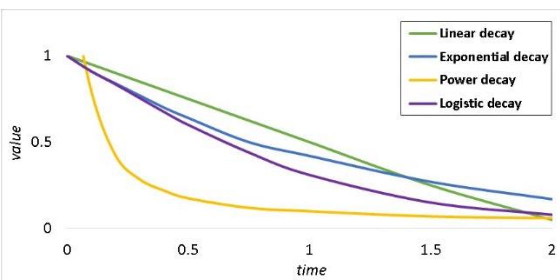

Fig. 2: The example of linear, exponential, power and logistic decay

IV. PROPOSED RECOMMENDATION STSTEM: *DDCF*

As aforementioned discussion, single decay function may not properly describe the complex preference variation of user. In this study, we propose a novel approach, Dynamic Decay Collaborative Filtering (abbreviated as *DDCF*), to effectively predict users' preference. DDCF has four steps: 1) item clustering, 2) interesting level identification, 3) decay function specification, and 4) preference prediction, as shown in Fig. 3.

To tackle the cold stat and sparsity issues of recommendation system, DDCF utilized the item clustering [3] to group the similar items together without any predefined parameter. Then, for each user, we identify each cluster's interesting level according to the time and number of rating record in the cluster. For each level, DDCF utilize different decay functions to describe the preference evolution. Finally, we calculate the similarities among users based on the derived decayed rates and predict the future preferences.

4.1 Item clustering

The cold start and sparsity are two fatal issues in CF-based recommendation. Cold start is related to recommendations for new users or items. Since the system does not have information about new users or items, it is really difficult to make precise recommendations. Sparsity problem is caused by the insufficient number of the transactions and feedback data. Recommendation system is difficult to distinguish the similar interests among users which will downgrade the usability of the collaborative filtering. DDCF uses a parameter-free clustering algorithm to solve the cold start and sparsity issues in CF-based recommendation. As in Definition 2, we derive the relation strength by Jaccard coefficient between two users with filtering out the insignificant relation (i.e., relation value lower than user-specified threshold α).

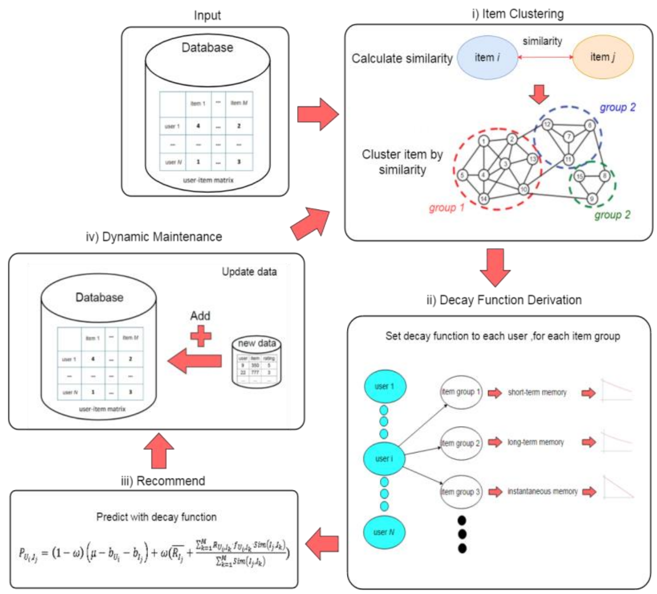

Fig. 3: The concept of DDCF

Definition 2 (Relation strength)
Given an item o, the profile $p = \{p_1, p_2, \ldots, p_k\}$ consist of k features of item o. The relation between two items can be derived by

$$R(o_i, o_j) = \frac{|p_i \cap p_j|}{\sqrt{|p_i| \times |p_j|}}. \qquad (2)$$

With the user-specified threshold α, the relation strength is defined as,

$$RS(o_i, o_j) = \begin{cases} R(o_i, o_j), \text{ if } R(o_i, o_j) \geq \alpha \\ 0 \quad , \text{ if } R(o_i, o_j) < \alpha \end{cases}. \qquad (3)$$

Obviously, α could control how dense of the relations among items when clustering and then effect the efficiency of process.

After deriving the relation strength, we use a parameter-free algorithm to cluster the item in the system. The pseudo code is given in Algorithm 1. DDCF propose a modularity-like evaluation, as shown in Definition 3, to be the terminated criteria of hierarchical clustering. At each iteration, based on the clustering result from the last iteration, we merge all pairs of items with the strongest relation strength among their neighbors to form larger clusters. Suppose the clustering result in the last iteration and in the current iteration are C and C', respectively. If the strength gain from C to C' is negative, DDCF will stop clustering, since the previous clustering result is good enough. Obviously, we can significantly decrease the time consumed in the clustering due to reduce the computation iteration.

Algorithm 1: *Item_clustering* (O)

```
C ← ∅;
C ={c1, ..., cm} ← set each o∈O as a community in C
while true do   // community detection
  for each ci∈ C do
    N(ci) ← collect all neighbors have RS value to ci;
    for each cj∈ N(ci) do
      if max_RS(ci) = max_RS(cj)
        ci ← merge ci and cj;
        C ← C−cj;
  claculate S(C);
  if ΔS < 0   // strength gain in Definition 3
    break;
output C;
```

Definition 3 (Strength gain)
Given an item set $O = \{o_1, \ldots, o_m\}$ in a recommendation system and the clustering result $C = \{c_1, c_2, \ldots, c_p\}$, the strength function is defined as,

$$S(C) = \sum_{k=1}^{p} \left[\frac{IS_k}{TS} - \left(\frac{OS_k}{TS} \right)^2 \right], \quad (4)$$

where $IS_k = \sum_{o_i, o_j \in c_k} RS(o_i, o_j)$ is the summation of total relation strengths among items inside cluster c_k, $DS_k = \sum_{o_i \in c_k, o_j \in O} RS(o_i, o_j)$ is the summation of relation strengths of items in cluster c_k and other items not in c_k, and $TS = \sum_{o_i, o_j \in O} RS(o_i, o_j)$ is the summation of all relation strengths between any two items in the recommendation system. With two different clustering results C and C', the strength gain from C to C' is

$$\Delta S_{C \to C'} = S(C) - S(C'). \quad (5)$$

4.2 Interest level & decay function identification

After clustering items, for each user, DDCF identifies the interest level of each cluster based on his/her rating behavior. We borrow the concept of human brain memory [30, 31] to describe the preference variation. DDCF categorizes users' preferences into instantaneous, short-term, and long-term interest level extending from the idea of Ebbinghaus forgetting curve [24]. The preference in instantaneous level usually consist very short and may be decayed fast. On the contract, the preference in short-term level may stay longer in brain than instantaneous level, but still will decay after a while. The preference in long-term level is able to stay for a long time and not easily forgotten by people.

Suppose the clustering result of item set O in a recommendation system is $C = \{c_1, c_2, \ldots, c_p\}$. For a user u_i and his/her rating set e_i, we could collect all rating records of items clustering in c_k and derive a rating sequence $\langle (r_{i1}, t_{i1}), (r_{i2}, t_{i2}), \ldots, (r_{i\ell}, t_{i\ell}) \rangle$ by sorting the rating record with t_{ij} in non-decreasing order. Given a user-specified time size w, the significant set $se_{ik} = \{(r_{ij}, t_{ij}) \mid t_{ij+1} - t_{ij} \leq w, t_{i0} = t_{i1}, 0 \leq j \leq \ell\}$. According to Ebbinghaus memory curve [24], human usually will not forget one thing easily after reviewing or mentioning 7 times. We borrow this idea and extend to describe preference variation. Hence, the interest level $\mathcal{L}_{ik}$ of c_k for u_i is defined as,

$$\mathcal{L}_{ik} = \begin{cases} instantaneous\_level, & \text{if } 0 < |se_{ik}| \leq 3 \\ short-term\_level, & \text{if } 4 \leq |se_{ik}| \leq 6 \\ long-term\_level, & \text{if } 7 \leq |se_{ik}| \end{cases} \quad (6)$$

According to the level characteristic, DDCF assigns different decay function for three interest level, instantaneous, short-term, and long-term levels. As mentioned above, instantaneous level usually consist very short and may be decayed fast. We choose the power decay to simulation the preference change. However, when a user rates the items in one cluster over 4 and 7 times, it means that he/she is quite interesting with this type of items. We could utilize the logistic and exponential decay functions to simulate the preference evolutions of short-term and long-term levels, respectively. The decay function of each level is defined as,

$$decay_{instant}(\Delta t) = \Delta t^{-\lambda} \cdot \alpha,$$
$$decay_{short}(\Delta t) = \frac{2}{1 + e^{\lambda \cdot \Delta t}},$$
$$decay_{long}(\Delta t) = e^{-\lambda \cdot \Delta t}. \quad (7)$$

Notice that the parameters λ and α could tune the decay degree of function, which usually are derived by heuristic evaluation.

4.3 Preference prediction

Different to traditional CF-based recommendation, DDCF uses baseline estimation and similarity calculation with decay consideration to predict the rate of the item. As several discussions in prior studies [XX], item-based CF methods usually have better accuracy than user-based CF. Hence, we extend the idea of item-based CF for recommendation. There are several methods can derive the similarity between two items, such as cosine, adjusted cosine, Pearson, Jaccard coefficients, to name a few. In this study, we use the adjusted cosine to calculate the item similarity.

Definition 5 (Item similarity)
Suppose that there are a set of users $U = \{u_1, \ldots, u_n\}$, and a set of items $O = \{o_1, \ldots, o_m\}$ in a recommendation system. Given two items $o_x, o_y \in O$, U_{o_x, o_y} is the set of users in U have rated o_x and o_y simultaneously. The similarity between two items o_x and o_y is defined as,

$$sim(o_x, o_y) = \frac{\sum_{u_k \in U_{o_x, o_y}} (r_{u_k, o_x} - \overline{r_{o_x}}) \times (r_{u_k, o_y} - \overline{r_{o_y}})}{\sqrt{\sum_{k=1}^{n} (r_{u_k, o_x} - \overline{r_{o_x}})^2} \times \sqrt{\sum_{k=1}^{n} (r_{u_k, o_y} - \overline{r_{o_y}})^2}}, \quad (8)$$

where the $\overline{o_x}$ and $\overline{o_y}$ are the average rates of items o_x and o_y in the recommendation system, respectively.

In DDCF, for a user u_i, the rate prediction of a certain item o_j could be derived by follows:

$$P_{u_i,o_j} = (1-\rho)\times\left(\mu - b_{u_i} - b_{o_j}\right) + \rho\times\left(\overline{r_{o_j}} + \frac{\sum_{k=1}^{m} D(r_{u_i,o_k}, t_{u_i,o_k})\times sim(o_j,o_k)}{\sum_{k=1}^{m}\left|sim(o_j,o_k)\right|}\right), \quad (9)$$

where $0 \le \rho \le 1$. Actually, Equation (9) can be decomposed into two parts: baseline estimation and decay CF. The parameter ρ is used to control the portion of baseline estimation and decay CF contributing the final prediction result. We utilize $\mu - b_{u_i} - b_{o_j}$ as the baseline estimation to predict the rating value; μ is the average rate of all items in the recommendation system, and b_{u_i} and b_{o_j} are the deviations of the rates of user u_i and item o_j, respectively. Then, when calculating the decay CF for prediction, we use Equation (1) to derive the decayed rate $D(r_{u_i,o_k}, t_{u_i,o_k})$ based on the time t_{u_i,o_k} and the corresponding decay function in Equation (7).

V. EXPERIMENTAL RESULTS

To evaluate the performance of proposed DDCF, five CF-based methods: 1) traditional item-based CF (IBCF), 2) fixed-exponential decay CF (DCF-exp), 3) fixed-power decay CF (DCF-pow), 4) fixed-logistic decay CF (DCF-log), 5) fixed-linear decay CF (DCF-lin), are implemented for comparison. All algorithms were coded in C++ language and tested on a workstation with Intel i7-3370 3.4 GHz with 8 GB main memory. A comprehensive performance study has been conducted on two real datasets to show the applicability of DDCF, as shown in Table 1. MovieLens-100K dataset contains 100,000 ratings (1-5 scale) from 716 users for 3,952 movies. MovieLens-1M dataset contains 1,000,000 ratings (1-5 scale) from 6,040 users for 3,952 movies.

Table 1: characteristics of MovieLens

	MovieLens-100K	MovieLens-1M
number of users	716	6040
number of movies	3,952	3,952
Average rated Items/User	125	149.7
Rating range	1 - 5	1 - 5

To measure the statistical accuracy, we use the mean absolute error (MAE) and root mean square error (RMSE) as the metrics to evaluate the quality of prediction results.

$$MAE = \frac{\sum_{i=1}^{n}\left|\widetilde{r_i} - r_i\right|}{n}. \quad (10)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\widetilde{r_i} - r_i)^2}{n}}. \quad (11)$$

n is the number of total predicted rating, $\widetilde{r_i}$ the predicted rating for the ith item, r_i is the user's true rating for the ith item. MAE is defined as the average absolute difference between predicted ratings and actual ratings; likewise, RMSE is the average root square difference between predicted values and actual values. Both measures are frequently used to assess the goodness of predicted values by a model or an estimator.

In the first experiment, to show the accuracy of DDCF under different training-testing partition, we vary the ratio of training and testing portion of MovieLens-1M dataset form 50% to 90%. As shown in Figs. 4 and 5, comparing to other CF-based methods, DDCF has better accuracy. Notice that, DDCF still has more precise prediction than CF using the fixed decay functions (i.e., power, logistic, linear and exponential). This is partly because that dynamically tuning the decay function could simulate the variance of preference more properly.

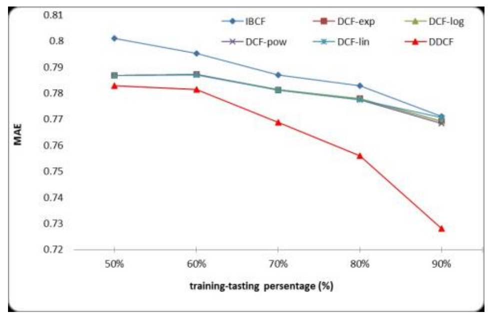

Fig. 4: The MAE of six algorithms on MovieLens-1M dataset

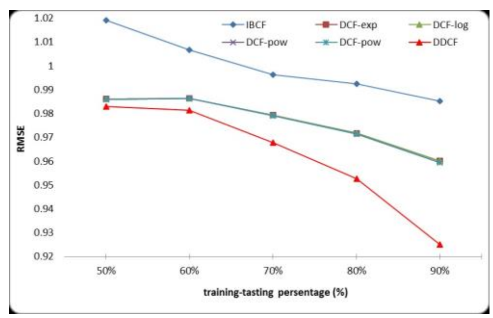

Fig. 5: The RMSE of six algorithms on MovieLens-1M dataset

In second experiment, we compare MAE and RMSE of DDCF with other 5 CF-based methods under two real datasets, as shown in Tables 2 and 3. Obviously, the proposed DDCF has best accuracy of prediction comparing to other algorithms.

Table 2: MAE and RMSE in MovieLens-100K dataset

	MAE	RMSE
IBCF	0.810133	1.088575
DCF-exp	0.802042	1.078469
DCF-log	0.802041	1.078469
DCF-pow	0.802221	1.078666
DCF-lin	0.80954	1.04201
DDCF	0.737513	0.946979

Table 3: MAE and RMSE in MovieLens-1M dataset

	MAE	RMSE
IBCF	0.771146	0.985259
DCF-exp	0.769135	0.960049
DCF-log	0.769136	0.960023
DCF-pow	0.768411	0.95936
DCF-lin	0.770700	0.964040
DDCF	0.728071	0.924965

VI. CONCLUSION

In this paper, we propose a novel CF-based recommendation, *DDCF*, which captures the preference variations to predict users' interests. DDCF extends the idea of human brain memory to dynamically adjust the decay functions based on users' behaviors. The experimental results indicate that DDCF performs better than traditional collaborative filtering and other fixed decay function consideration. We also apply the proposed DDCF on real datasets to show the practicability.

REFERENCES

[1] F. Bodendorf, and C. Kaiser. "Detecting Opinion Leaders and Trends in Online Social Networks," *The 2nd ACM workshop on Social web search and mining*, pp. 65-68, 2009.

[2] Y. Chen, Y. Chen, C. Hsu, H. You, J. Liu, and X. Huang, "Mining Opinion Leader in Big Social Graph," *The 31st IEEE International Conference on Advanced Information Networking and Applications (AINA 2017)*, pp.XXX, 2017.

[3] Y. Chen, W. Zhu, W. Peng, W. Lee, and S. Lee. "CIM: community-based influence maximization in social networks," *ACM Transactions on Intelligent Systems and Technology*, vol. 5, no.2, article 25, 2014.

[4] Y. Cho, J. Hwang, and D. Lee. "Identification of Effective Opinion Leaders in the Diffusion of Technological Innovation: A Social Network Approach," *Technological Forecasting and Social Change*, vol. 79, no.1, pp. 97-106, 2012.

[5] J. Duan, J. Zeng, and B. Luo. "Identification of Opinion Leaders Based on User Clustering and Sentiment Analysis," *The 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, pp. 377-383, 2014.

[6] Z. Feng, X. Xu, N. Yuruk, and T. Schweiger, "A novel similarity-based modularity function for graph partitioning," *The 9th International Conference on Data Warehousing and Knowledge Discovery (DaWaK'07)*, pp. 385–396, 2007.

[7] E. Katz, "The two-step flow of communication: An up-to-date report on hypothesis," *Public opinion quarterly*, vol. 21, no. 1, pp. 61-78, 1957.

[8] F. Li, and T. Du, "Who is Talking? an Ontology-Based Opinion Leader Identification Framework for Word-of-Mouth Marketing in Online Social Blogs," *Decision Support Systems*, vol. 51, no. 1, pp. 190-197, 2011.

[9] Y. Li, S. Ma, Y. Zhang, R. Huang, "An Improved Mix Framework for Opinion Leader Identification in Online Learning Communities," *Knowledge-Based Systems*, vol. 43, pp. 43-51, 2013.

[10] Q. Miao, S. Zhang, Y. Meng, and H. Yu, "Domain-Sensitive Opinion Leader Mining from Online Review Communities," *The 22nd international conference on World Wide Web companion*, pp. 187-188, 2013.

[11] Mobile01. http://www.mobile01.com

[12] K. Song, D. Wang, S. Feng, and G. Yu, "Detecting Opinion Leader Dynamically in Chinese News Comments," *Web-Age Information Management.Springer*, pp. 197-209, 2012.

[13] X. Yu, X. Wei, and X. Lin. "Algorithms of BBS Opinion Leader Mining Based on Sentiment Analysis," *Web Information Systems and Mining.Springer*, pp. 360-369, 2010.

[14] H. Zhou, D. Zeng, and C. Zhang. "Finding Leaders from Opinion Networks," *2009 Intelligence and Security Informatics (ISI'09)*, pp. 266-268, 2009.

[15] Z. Zhai, H. Xu, and P. Jia. "Identifying Opinion Leaders in BBS," *2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT'08)*, pp.398-401, 2008.