Wuhan University Journal of Natural Sciences

2018, Vol.23 No.1, 009-016

Article ID 1007-1202(2018)01-0009-08

DOI https://doi.org/10.1007/s11859-018-1288-z

# A Framework for Personalized Adaptive User Interest Prediction Based on Topic Model and Forgetting Mechanism

□ **GUI Sisi[1], LU Wei[1], ZHOU Pengcheng[1], ZHENG Zhan[2]**

1. School of Information Management, Wuhan University, Wuhan 430072, Hubei, China;

2. School of Media and Communication, Wuhan Textile University, Wuhan 430073, Hubei, China



**Abstract:** User interest is not static and changes dynamically. In the scenario of a search engine, this paper presents a personalized adaptive user interest prediction framework. It represents user interest as a topic distribution, captures every change of user interest in the history, and uses the changes to predict future individual user interest dynamically. More specifically, it first uses a personalized user interest representation model to infer user interest from queries in the user's history data using a topic model; then it presents a personalized user interest prediction model to capture the dynamic changes of user interest and to predict future user interest by leveraging the query submission time in the history data. Compared with the Interest Degree Multi-Stage Quantization Model, experiment results on an AOL Search Query Log query log show that our framework is more stable and effective in user interest prediction.

**Key words:** user interest; user interest presentation; user interest prediction; topic model; forgetting mechanism

**CLC number:** TP 393

**Received date:** 2017-09-20

**Foundation item:** Supported by the National Natural Science Foundation of China (71473183, 71503188)

**Biography:** GUI Sisi, female, Ph. D. candidate, research direction: user interest, query intent, information retrieval. E-mail: sgui0229@whu.edu.cn

## 0 Introduction

User interest is in the center of personalized information service systems, such as personalized information retrieval system, personalized recommendation system, advertising system, etc. This paper focuses on user interest prediction and considers a specific case where the user interest is not static, as it changes with time. For example, a user may be interested in certain topics for a while and then lose interest in them gradually, unless the interests are triggered again by external factors.

There are two challenges in our user interest prediction. Gauging user interest is the first challenge. With implicit methods, such as questionnaires or interviews, users are reluctant to provide or cannot express their interest accurately [1]. Luckily, explicit rating is a good substitution for implicit methods in gauging user interest [2]. Through explicit rating, user interest can be obtained implicitly from their interactions with the system, without asking users to express their interest explicitly. The interactions (mouse click activities [3], scrollbar activities, keyboard activities, object ratings, spatial information [4,5], social network activities [4,6], etc.) are usually stored in query logs. In the scenario of search engine, a user has his preferred search topic of interest in mind and uses query to express the topic [7,8]. Thus the user interest can be represented as the favored topic distribution over the queries. Topic model, which can discover the "topics" that are behind text, is a good tool for user interest presentation [9].

Capturing the changes of user interest accurately is the second challenge. Most user interest predictions are based on collaborative filtering which hardly considers the evidences from the changes of user interest. Other user interest predictions apply forgetting mechanism or time window mechanism. The two mechanisms can be used to trace the changes of user interest: forgetting mechanism assigns different weights to the interests in different time span, according to a forgetting curve which models the decline of memory; the time window mechanism holds the view that the recent history data (in a certain window size) can reflect user future interest more precise so that it can be used to infer user interest. However, those researches using time window mechanism or forgetting mechanism only consider the data in beginning or in the end of the selected span of historical data and neglect other data in the selected span. We hypothesize that using whether either forgetting mechanism or time window mechanism, each historical data reflects user interest at a single time point and would have potential impact on the accuracy of user interest prediction.

In this paper, in response to the two challenges, we propose a personalized adaptive user interest prediction framework, which represents user interest as a topic distribution and considers every change of user interest in the user historical data for user interest prediction. The framework consists of two models: personalized user interest representation model and personalized user interest prediction model. The former infers user interest using a topic model by analyzing all previous submitted queries in a query log; the latter calculates the weights accumulatively for user interests by considering every data in the selected time span.

We make several contributions to user interest studies:

1) We relieve users of the burden on verbalizing their interests: the proposed personalized user interest representation model infers user interest from user history data and the proposed evaluation method constructs the ground truth from user history data;

2) The proposed personalized user interest prediction model can capture the changes of user interests without neglecting any history records and use all data for user interest prediction;

3) Our framework has been tested on a real query log and the results validate our framework.

The rest of the paper is organized as follows. Section 1 describes the related work and the highlights of our work. Details about the two models (personalized user interest representation model and personalized user interest prediction model) are elaborated in Section 2. Experiments and insights from the experiments are discussed in Section 3. Finally we conclude the whole work in Section 4.

## 1 Literature Review

We summarize related work from two aspects (user interest representation and user interest prediction), corresponding to the two proposed models (personalized user interest representation model and personalized user interest prediction model) in the framework.

### 1.1 User Interest Representation

User interest behind a Web search can be represented by the search topics behind a query. Those search topics can be derived from query text or user interactions with the search engine result pages (SERPs), such as click behaviors recorded in the click-through data.

Using query text only, search topics are derived by mapping queries into pre-defined topic categories. Each category refers to one distinct topic. Jiang *et al*[10] used the sixteen top-level categories in the Open Directory Project (ODP, *http://www.dmoz.org/*) as the pre-defined topic categories. They leveraged a Learning to Rank model (LambdaMART)[11] to rank the candidate queries in a query auto-completion task with 22 distribution-based features derived from the distributions of categories over queries (the target query and the previous issued queries in the same session as the target query).

Using click-through data, search topic are derived as clusters of words [12], Web pages [13] or other query related resources [14]. Each cluster refers to one distinct topic. More specifically: Sadikov *et al*[12] mined search topics by performing a clustering algorithm on a weighted graph where the nodes are the query refinements extracted from the click-thorough data. Based on the cluster results, they investigated the drift between intents. Duan *et al*[13] proposed the concept of click pattern for representing search topics. Each click pattern is a cluster of clicked Web documents and the clusters are identified using an unsupervised divisive clustering algorithm until the average intra-distance of each cluster is below a threshold. The click pattern is applied well to three applications (query ambiguity measurement, query classification and query recommendation). Ren *et al*[14] mined search topics using three types of data sources (queries, Web pages and Wikipedia concepts). Their proposed heterogeneous graph based clustering algorithm outperforms over several baselines that use one or two types of sources or other clustering methods.

Our proposed personalized user interest representation model derives user interest from query text only. However, it does not map the search topics into some pre-defined topic categories. It generates topic categories over query texts using a topic model.

### 1.2 User Interest Prediction

Collaborative filtering is one way for static user interest prediction. Most prior work [15-17] on collaborative filtering cannot reflect the changes of user interest unless the time factor is considered. More specifically, Lee *et al*[15,16] considered user purchase time, comment time, item on sale time and the time span among those time point in an e-commerce system. They proved that time factors can improve the precision of a recommendation system. Pham *et al*[17] record user participation time in the classic user-item model to extend collaborative filtering approach.

Forgetting mechanism and time window mechanism are another two common ways to solve the changes of user interest. Time window mechanism neglects the data out of the window. However, those data may also reflect some user interests and should not be discarded arbitrarily. Maloof *et al*[18] discussed the selection of historical data in time window mechanism. Forgetting mechanism overcomes the shortcoming of time window mechanism. However, previous works on forgetting mechanism only consider the data in the beginning or in the end and does not take data at every time point into consideration. For example, Zhang *et al*[19] used the time span between the initial comment time and the last comment time of the item. Chen *et al*[20] and Wu *et al*[21] used the time span between the initial comment time and the predict time point of the item.

To the best of our knowledge, Interest Degree Multi-stage Quantization (IDMQ) Model [22] is the only prediction model that considers all historical data. The main idea of IDMQ is expressed in Fig.1. It depicts the changes of the weight of an interest for a period of a time. In Fig.1, the $x$-axis refers to time and the $y$-axis refers to the weight of the interest. The curves ($k_0$, $k_1$ and $k_2$) are the forgetting curves and have different start points ($p_0$, $p_1$ and $p_2$). $k_0$ refers to the original or the initial forgetting curve whereas $k_1$ and $k_2$ refer to new forgetting curves when the interests are triggered through interactions at different time point ($t_1$ and $t_2$), respectively. Both starting points ($p_1$ and $p_2$) of $k_1$ and $k_2$ are the sum of two parts: 1) the remaining interest weight from the past ($s_1$ and $s_2$), 2) and the newly added interest weight ($h_1$ and $h_2$) due to the new interactions at $t_1$ and $t_2$.

The main idea of our framework is similar to that of IDMQ: the new weight of the interest composes of remaining weight and newly added weight. However, the proposed framework is different in calculating these two weights. In the experiment, we set IDMQ as baseline.

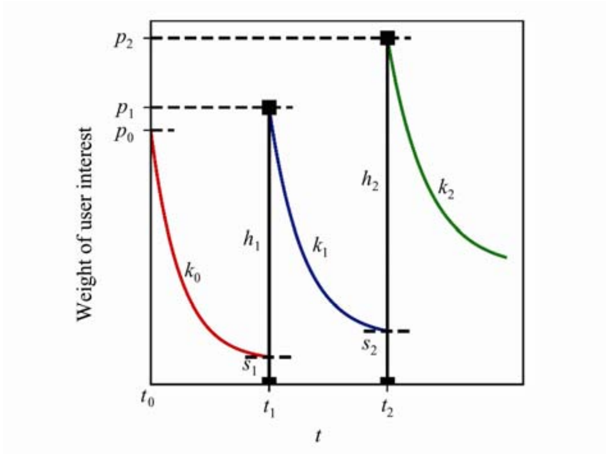


**Fig. 1 Interest degree multi-stage quantization model (IDMQ)**

## 2 Methodology

### 2.1 Basic Idea

We assume that an interest will decrease gradually and can be triggered due to user's interactions later. Thus, the weight of interest consists of the remaining part ($f$) due to forgetting mechanism, and the newly triggered part ($g$) due to user interactions.

In our proposed personalized adaptive user interest prediction framework, the degree of interest decreasing ($f$) is captured by a forgetting curve and the degree of interest increasing ($g$) is estimated as the average of his previous interest weights.

Here is a case: a user is interested in a new topic at $t_0$ with weight $w_0$ and he conducts some interactions related to the same interest at $t_1$ and $t_2$, respectively. We model the changes of his interest from $t_2$ to $t_2$ using personalized user interest representation model and predict his future interest at $t_3$ using personalized user interest prediction model.

Figure 2 illustrates the basic idea of our framework. The three curves in red, blue and green are the forgetting curves. Figure 2(b) is similar to Fig. 2(a), but with shorter time interval between $t_1$ and $t_2$. Due to the space limit, we just elaborate Fig. 2(a) more specifically.

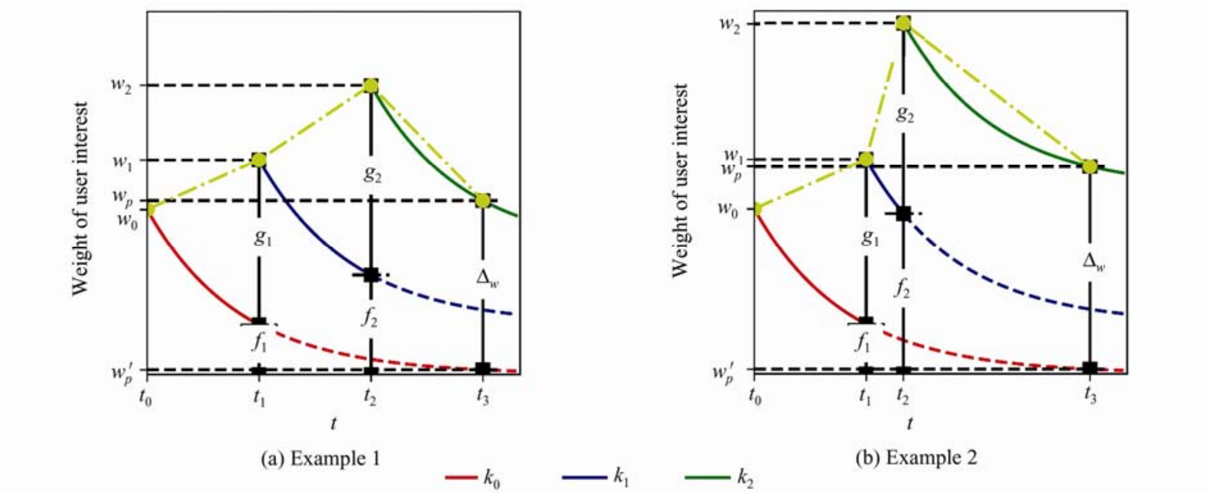


**Fig. 2 Two examples of personalized adaptive user interest prediction framework**

Step 1: We model the changes of interest using the personalized user interest representation model. From $t_0$ to $t_1$, he loses interest gradually from the average of previcus interest weights $w_0$ as in red curve. At $t_1$, the interest is triggered and the weight is raised to $w_1$. $w_1$ consists of two parts: (1) the remaining interest $f_1$ calculated from the red forgetting curve and (2) the newly added interest $g_1$ estimated as $w_0$. From $t_1$ to $t_2$, he loses interest as in the blue curve. At $t_2$, the interest is triggered again and the weight is changed to $w_2$. $w_2$ consists of two parts: 1) the remaining interest $f_2$ calculated from the blue forgetting curve and 2) the newly added interest $g_2$ estimated as the average of $w_0$ and $w_1$. As a result, we capture the changes of his interest from $t_0$ to $t_2$ in yellow dash-dotted line.

Step 2: We predict the weight of future interest at $t_3$ using the personalized user interest prediction model with the green forgetting curve.

The interval between $t_1$ to $t_2$ in Fig. 2(a) is larger than that in Fig. 2(b) and our model predicts a lower weight ($w_p$) of the user interest at $t_3$ in Fig. 2(a) than that in Fig. 2(b). It is reasonable that an interest will be less important if it takes a longer time to recall. Thus Fig.2 shows that our model can capture the impact of different interval time of data (e.g. between $t_1$ and $t_2$) in predicting user interest.

## 2.2 Problem Definition

**Definition 1** The dynamic user interest prediction framework is defined as: $F=\langle U,Z,T,Q\rangle$, where $U$ denotes a set of users, $Z$ denotes a set of interest vectors $z$, $T$ denotes a set of time and $Q$ denotes a set of query vectors. In the framework, an interest is represented by an interest vector $\boldsymbol{z}$ and a query is represented by a query vector $\boldsymbol{q}$. The framework means user $u\in U$ has interest $\boldsymbol{z}\in Z$ which is derived from query $\boldsymbol{q}\in Q$ at time $t\in T$. For each $u\in U$, the framework contains a personalized user interest representation model and a personalized user interest prediction model.

**Definition 2** The personalized user interest representation model (1) generates the user interest vectors to capture the relationship between $Z$ and $Q$ using topic model and (2) generates the query vector $\boldsymbol{q}\in Q$ for each query to capture the relationship between $Q$ and $T$ using forgetting curve.

**Definition 3** The personalized user interest prediction predicts future user interests at $t\in T$ based on personalized user interest representation model of user $u\in U$.

**Definition 4** An interest vector is an $m$-dimensional vector: $\boldsymbol{z}=(z_1,w_{z_1};\cdots;z_m,w_{z_m})$,where $m$ is the total number of topics in the collection, $z_i$ represents an $i$-th topic and $w_{z_i}$ represents the weight of interest for topic $z_i$. The vector is then normalized to be summed to 1.

**Definition 5** A user's query vector is defined as: $\boldsymbol{q}=(word_1,w_{word_1};\cdots;word_c,w_{word_c})$, where $c$ is the number of unique word in the query set, $word_i$ represents $i$-th query term in the collection of query terms and $w_{word_i}$ denotes the weight for word $word_i$.

## 2.3 Generate User Interest Vector Using Topic Model

Topic model outperforms word co-occurrence based clustering or citation-based clustering in topic detection and topic tracking [23] and it is a good way for user profiling [6].

The relationship among topics, query words and query can be presented as $P(word \mid query) = P(word \mid z)P(z \mid query)$ where $query$ denotes the query text. $P(word \mid query)$ is the distribution of query word $word$ over query $query$. $P(word \mid z)$ is the distribution of topic $z$ over query word $word$. $P(z \mid query)$ is the distribution of query $query$ over topic $z$. Topic model is used to obtain $P(word \mid z)$. Latent Dirichlet Allocation (LDA) is the most widely used model because of its capability and low computational complexity. As a result, we get $P(word \mid z)$ for every user using LDA.

### 2.4 Generate Query Vector Using Forgetting Curve

#### 2.4.1 Set forgetting curve

From the aspect of function shape [24], we choose the exponential function to model the forgetting curve for $word_i$.

$$f_{word_i}(\tau, z_j) = \exp\{-\lambda_{z_j}\ \tau\},\ \tau = t_n - t_{n-1} \tag{1}$$

According to formula (1), the weight $f_{word_i}(\tau, z_j)$ of $word_i$ will be calculated through forgetting mechanism each time when $word_i$ is submitted to search engine. $\tau$ is the time interval between $t_n$ and $t_{n-1}$ when $word_i$ is submitted to the search engine. $\lambda_{z_j}$ is the forgetting factor for topic $z_j$.

To get $\lambda_{z_j}$, we get the distribution of all the time intervals over topic $z_j$ and then calculate $\lambda_{z_j}$ from 0 to 1 with a step of 0.005 to make sure over 80% $f_{word_i}(\tau, z_j)$ is over 0.2 and around 20% $f_{word_i}(\tau, z_j)$ is around 0.8. This standard is set to ensure a large difference among $f_{word_i}(\tau, z_j)$ for a good experiment result.

#### 2.4.2 Generate query vector

Recall the premise that user interest will increase if $word_i$ is submitted to the search engine at $t_n$, the weight $word_i$ of $word_i$ will change. The new weight $w^{(n)}_{word_i}$ of $word_i$ at $t_n$ composes of two parts: $f^{(n)}_{word_i}(\tau, z_j)$ which is the remaining weight of word $word_i$ from forgetting curve before $t_n$ (in formula (3)), and $g^{(n-1)}_{word_i, z_j}$ which is the weight of $word_i$ gained by user's interaction at $t_n$ (in formula (4)).

$$w^{(n)}_{word_i} = f^{(n)}_{word_i}(\tau, z_j) + g^{(n-1)}_{word_i, z_j} \tag{2}$$

$$f^{(n)}_{word_i}(\tau, z_j) = \exp\{-\lambda_{z_j} \cdot \tau\},\ \tau = t_n - t_{n-1} \tag{3}$$

$$g^{(n-1)}_{word_i, z_j} = \begin{cases} \dfrac{1}{n-1}\displaystyle\sum_{k=1}^{n-1} w^{(k)}_{word_i},\ n \geqslant 2 \\ 0,\ n = 1 \end{cases} \tag{4}$$

Among Eqs. (2)-(4), Eq. (2) is merely the sum of Eq. (3) and Eq. (4). Eq. (3) is based on forgetting mechanism and Eq. (4) is the average value of $w_{word_i}$ before $t_n$. As a result, we can get $w_{word_i}$ at any time point in user history query logs. The user probability $p(z_{j,t_n})$ for $z_j$ at $t_n$ will be the sum of all $w^{(n)}_{word_i}$ from all $v$ words from $word_1$ to $word_v$ which are related to topic $z_j$ (Eq. (5)).

$$p(z_{j,t_n}) = \sum_{i=1}^{v} w^{(n)}_{word_i} \tag{5}$$

### 2.5 Predict User Interest Using Personalized User Interest Prediction Model

After obtaining the probability of user interest in the past, we predict user interest $p(z_{j,t_{\text{future}}})$ at future time $t_{\text{future}}$ by changing the parameters in formula (2) to formula (5): If $t_n$ is the last time point in user history query logs, we set $\tau$ as the time interval between $t_n$ and $t_{\text{future}}$.

The final distribution of interest for a single user at future time $t_{\text{future}}$ is the normalized interest vector. The normalization can be simply carried by diving the vector length.

## 3 Experiments

### 3.1 Dataset

A query log of a long period is required to validate our framework. However, most query logs are proprietary to commercial companies, apart from four query logs (*http://jeffhuang.com/search_query_logs.html*) and one query log from Yandex (*https://www.kaggle.com/c/yandex-personalized-web-search-challenge/data*). Among all the open-access query logs, an AOL Search Query Log (*http://www.cim.mcgill.ca/~dudek/206/Logs/AOL-user-ct-collection/*) has the most query records, covering the longest log period of three months.

Thus, we choose the AOL Search Query Log for the experiment. More specifically, it consists of 36 389 557 records of over 650 000 users from March to May in 2006. Each record has five fields, namely AnonID, Query, QueryTime, ItemRank and ClickURL: AnonID is the anonymous user unique ID and records with the same AnonID are from the same user; Query is the query text; QueryTime records the time when user submits the query, clicks on a result item, or requests for next "page". No data will be recorded in ItemRank and ClickURL if the user does not click on any result item. With the premise that user expresses their interests through query and will click on the result item if the query is expressed accurately, 19 442 628 records which have data in five fields are extracted, covering 521 682 unique users. Among those records, only data in “AnonID”, “Query” and “QueryTime” are used.

However, the individual data is quite sparse that most individual users do not contribute lots of records: around 49.5% individual users have less than 10 records in the whole period. Though the records of the top 20 individual users only account for 1.11% of all records, the 20th individual user only has 2 698 records of 66 unique queries which are small for generating topic models in the proposed Personalized User Interest Representation Model. Thus, only records of top 20 users are used for the following experiments.

## 3.2 Experiment Setup

We take IDMQ [22] as the baseline.

Both IDMQ and our framework (MY) are tested on individual user with several combinations of different time units (1 s, 60 s, 3 600 s, 8 6400 s) and different time processing methods (Table 1).

**Table 1 Time processing method**

| Notation | Description | An example for output |
|---|---|---|
| 0 | Return the smallest integer value that is greater than or equal to the floating-point value | 2 |
| 1 | Return the largest integer value that is less than or equal to the floating-point value | 1 |
| 3 | Return the floating-point value | 1.5 |

## 3.3 Evaluation Metrics

Most previous researches used recall or mean average precision (MAP) as the evaluation metrics. However, the two measurements require users’ engagement in expressing their actual interests explicitly in the evaluation stage. In terms of privacy, user ID has been coded and no user can be contacted. As a result, recall and MAP are no longer available and a new evaluation method is proposed:

1) We split the whole consecutive AOL Search Query Log into two parts sequentially, with the first part (data in March and April here) being the history data and later part (data in May) being the ground truth. 2) On the history data, we predict the interest vector for each user using our framework and normalize the vector so that the values can be added to 1. 3) On the ground truth data, we normalize the frequencies of words belonging to each topics by dividing the sum of frequencies of all the words and take the final results as the actual user interest vector. 4) We calculate the cosine similarity between the predicted user interest vector and actual user interest vector [25]. The higher the Cosine similarity value is, the better the prediction result will be.

## 3.4 Result and Discussion

Figure 3 is the results of two models (IDMQ and MY) in several combinations of different time units and different time processing methods.

In general, MY statistically significantly outperforms IDMQ, and obtains an average cosine similarity of 0.669 0 whereas IDMQ obtains an average cosine similarity around 0.360 7. A tailed $t$ test for the results of two models on each user proves the improvements are statistically significantly different ($p<0.01$). MY outputs the highest cosine similarity of 0.671 3 when the time unit is 60 s and the time processing method is 0, whereas IDMQ outputs the highest cosine similarity of 0.600 5 when the time unit is 86 400 s and the time processing method is 0.

Also, MY outputs a higher and a more stable prediction than IDMQ. In all combinations of different time units and different time processing methods, the cosine similarities of MY range between 0.665 1 and 0.671 2 with a standard deviation of 0.002 4 whereas the cosine similarities of IDMQ range between 0.194 1 to 0.600 5 with a standard deviation of 0.152 1.

We discuss the results of MY and IDMQ in detail from the following two aspects:

1) Comparison among results within the same time unit (1 s, 60 s, 3 600 s or 86 400 s in Fig. 3).

For three time units (60 s, 3 600 s or 86 400 s), both MY and IDMQ with time processing method 0 outperforms the other two time processing methods (1 and 3). As MY is more stable, the improvements is smaller than that of IDMQ. When time unit equals to 1 s, MY outputs the same cosine similarity of 0.667 0 in all three time processing methods (0, 1 and 3) whereas IDMQ outputs the same cosine similarity of 0.195 9 in time processing method 1 and 3, which is higher than that in time processing method 0 (0.194 1).

2) Comparison among results within the same time processing method (0, 1 or 3 in Fig. 3).

With respect to MY, the results do not change a lot when the time unit changes. With respect to IDMQ, the larger time unit is, the better prediction accuracy achieves.

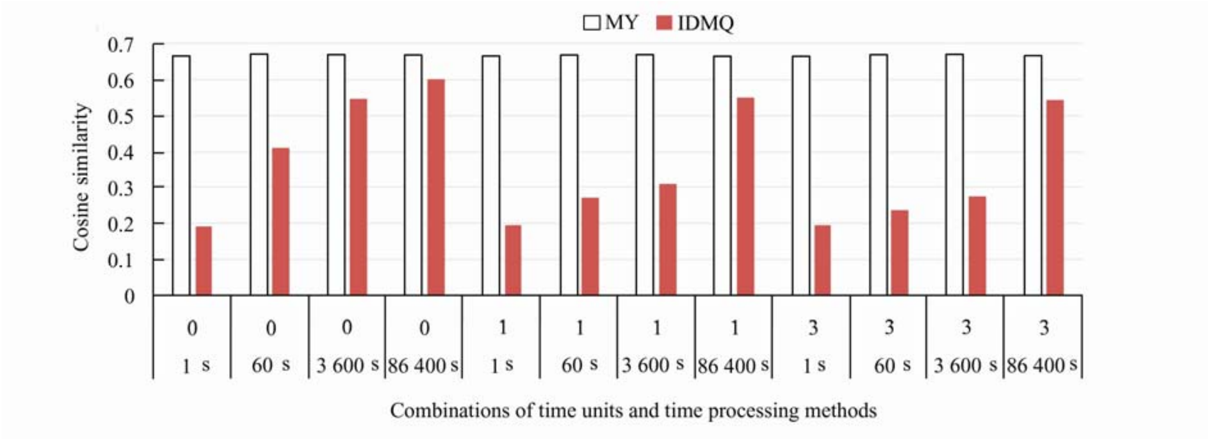


**Fig. 3 Cosine similarity for two methods**

0, 1 and 3 in the first line under the *x*-axis denote different time processing method in Table 1, and 1 s, 60 s, 3 600 s, 86 400 s in the second line denote different time units

## 4 Conclusion

On the premise that user interest will be aroused every time when user submit the same or similar queries to the search engine, this paper presents a personalized adaptive user interest prediction framework: it first uses a personalized user interest representation model to infer user interest from queries in his history data using a topic model; then it uses a personalized user interest prediction model to capture the dynamic changes of user interest and to predict future user interest by leveraging the query submission time in the history data. Experiments on AOL Search Query Log show good results. However, more efforts can be made to improve the model: 1) Testing the model with other prediction models, such Memory-based User Profile [26]; 2) Obtaining other latest query logs of longer period and validateing the framework on those query logs.

## References

[1] Carroll J. *Interfacing Thought: Cognitive Aspects of Human-Computer Interaction* [M]. Cambridge: MIT Press, 1987.

[2] Claypool M, Brown D, Le P, *et al.* Inferring user Interest[J]. *IEEE Internet Computing*, 2001, **5**(6): 32-39.

[3] Qiu F, Cho J. Automatic identification of user interest for personalized search[C] // *Proceedings of the* 15*th International Conference on World Wide Web (WWW* '06*)*. New York: ACM, 2006: 727-736.

[4] Yin H, Cui B, Chen L, *et al.* Dynamic user modeling in social media systems[J]. *ACM Transactions on Information Systems*, 2015, **33**(3): 10-44.

[5] Li M, Sagl G, Mburu L, *et al.* A Contextualized and personalized model to predict user interest using location-based social networks[J]. *Computers*, *Environment and Urban Systems*, 2016, **58**: 97-106.

[6] Wu L, Wang D, Guo C, *et al.* User profiling by combining topic modeling and pointwise mutual information (TM-PMI) [C] // *Proceedings of the* 22*th MultiMedia Modeling* (*MMM* 2016), 2016:152-161.

[7] Ahmed A, Low Y, Aly M, *et al.* Scalable distributed inference of dynamic user interests for behavioral targeting[C] // *Proceedings of the* 17*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (*SIGIR* '11). New York: ACM, 2011: 114-122.

[8] Song T, Zhao H, Zheng S, *et al.* Construction of user interest model based on queries and concept[J]. *Journal of Jilin University* (*Information Science Edition*), 2015, **33**(3): 310-314(Ch).

[9] Veningston K, Shanmugalakshmi R. Combining user interested topic and document topic for personalized inform- ation retrieval [C] // *Proceedings of the* 3*rd International Conference on Big Data Analytics* (*BDA* 2014). 2014: 60-79.

[10] Jiang J, Cheng P. Classifying user search intents for query auto-completion [C] // *Proceedings of the* 2016 *ACM International Conference on the Theory of Information Retrieval* (*ICTIR* '16). New York: ACM, 2016: 49-58.

[11] Burges C. From RankNet to LambdaRank to LambdaMART: An overview[EB/OL]. [2017-10-31]. *http://citeseerx.ist.psu.edu/viewdoc/download?doi=*10.1.1.180.634*&rep=rep1&type=pdf*.

[12] Sadikov E, Madhavan J, Wang L, *et al.* Clustering query

refinements by user intent[C]// *Proceedings of the* 19*th International Conference on World Wide Web* (*WWW'*10). New York: ACM, 2010: 841-850.

[13] Duan H, Kiciman E, Zhai C. Click Patterns: An empirical representation of complex query intents[C] // *Proceedings of the* 21*th ACM International Conference on Information and Knowledge Management* (*CIKM* '12). New York: ACM, 2012: 1035-1044.

[14] Ren X, Wang Y, Yu X, *et al*. Heterogeneous graph-based intent learning with queries, web pages and Wikipedia concepts[C]//*Proceedings of the* 7*th ACM International Conference on Web Search and Data Mining* (*WSDM* '14). New York: ACM, 2014: 23-32.

[15] Lee T, Park Y, Park Y. A time-based approach to effective recommender systems using implicit feedback [J]. *Expert Systems with Applications*, 2008, **34**(4): 3055-3062.

[16] Lee T, Park Y, Park Y. An empirical study on effectiveness of temporal information as implicit ratings [J]. *Expert Systems with Applications*, 2009, **36**(2): 1315-1321.

[17] Pham X, Jung J, Nam B, *et al.* User timeline and interest-based collaborative filtering on social network [J]. *Context-Aware Systems and Applications*. 2015, **165**: 132- 140.

[18] Maloof M, Michalski R. Selecting examples for partial memory learning[J]. *Machine Learning*, 2000, **41**(1): 27-52.

[19] Zhang Y, Liu Y. A collaborative filtering algorithm based on time period partition[C]// *Proceedings of the* 3*rd International Symposium on Intelligent Information Technology and Security Informatics* (*IITSI* 2010). Piscataway: IEEE Computer Society, 2010: 777-780.

[20] Chen Z, Jiang Y, Zhao Y. A collaborative filtering recommendation algorithm based on user interest change and trust evaluation [J]. *International Journal of Digital Content Technology and its Applications*, 2010, **4**(9): 106-113.

[21] Wu Y, Wang Y, Tang Z. A collaborative filtering recommendation algorithm based on interest forgetting curve [J]. *International Journal of Advancements in Computing Technology*, 2012, **4**(10): 148-157.

[22] Yu H, Cui R, Dong Q. Micro-blog user interest model based on forgetting curve [J]. *Computer Engineering and Design*, 2014, **35**(10): 3367-3372(Ch).

[23] Ding W, Chen C. Dynamic topic detection and tracking: A comparison of HDP, C-word, and co-citation methods [J]. *Journal of the Association for Information Science and Technology*, 2014, **65**(10): 2084-2097.

[24] Ding Y, Li X. Time weight collaborative filtering[C]// *Proceedings of the* 14*th ACM International Conference on Information and Knowledge Management* (*CIKM* '05). New York: ACM, 2005: 485-492.

[25] Jeong D, Song M. Time gap analysis by the topic model-based temporal technique [J]. *Journal of Informetrics*, 2014, **8**(3): 776-790.

[26] Wu D, Zhao D, Zhang X. An adaptive user profile based on memory model [C] // *Proceedings of the* 9*th International Conference on Web-Age Information Management* (*WAIM*'08). Piscataway: IEEE Computer Society, 2008: 461-468.

□