Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence

A Personalized Interest-Forgetting Markov Model for Recommendations *

Jun Chen, Chaokun Wang and **Jianmin Wang**
School of Software, Tsinghua University, Beijing 100084, P.R. China.
chenjun14@mails.thu.edu.cn, {chaokun, jimwang}@tsinghua.edu.cn

Abstract

Intelligent item recommendation is a key issue in AI research which enables recommender systems to be more "human-minded" when generating recommendations. However, one of the major features of human — forgetting, has barely been discussed as regards recommender systems. In this paper, we considered people's forgetting of interest when performing personalized recommendations, and brought forward a personalized framework to integrate interest-forgetting property with Markov model. Multiple implementations of the framework were investigated and compared. The experimental evaluation showed that our methods could significantly improve the accuracy of item recommendation, which verified the importance of considering interest-forgetting in recommendations.

Introduction

Recommender systems have been so common in people's daily lives as it has become increasingly difficult for people to find items which are interesting and useful to them in the era of big data. As a representative of the most successful recommendation methods, the family of collaborative filtering has been seeking ways to reconstruct people's preferences from their feedbacks on previous consumptions (Koren 2009). The modeling of user preferences is essential for systems to better "understand" people. However, understanding is never enough for intelligent recommender systems which are supposed to be more "human-minded". Recommendations could be more accurate and personalized if the system behaves like the person who is using it.

One of the most prominent features of human beings is the memory forgetting (Averell and Heathcote 2011; Ebbinghaus 1885). People's memory on a certain object is not permanent and is losing from time to time unlike that of computers. When returning back from breaks, the quality of people's work negatively correlates with the length of break while positively correlates with the levels of experience that people achieved before the breaks (Anzanello and Fogliatto 2011). Similar to memory, we believe that people's interest on items also shares the feature of forgetting as time elapses. In this paper, we address the item recommendation problem from a novel perspective that incorporates people's forgetting of interest. Specifically, we integrate the interest-forgetting mechanism with Markov model which is widely used in the existing recommendation methods.

Why interest-forgetting is important? Forgetting is like a filter that removes the out-of-date and the irrelevant-to-present information from people's mind. However, this property of human agents has barely been considered in recommender systems in the literature to the best of our knowledge. The forgetting of interest shares similar concept with those of the context-aware recommendations (Wang, Rosenblum, and Wang 2012; Hariri, Mobasher, and Burke 2012) and the temporal recommendations (Koren 2009; Xiang et al. 2010) where people's long-term interest is usually weakened or ignored compared with their short-term interest in recommendations. However, the interest-forgetting in our work puts its focus on simulating people's mind activities for decision making other than only understanding user contexts. Interest-forgetting enables recommender systems to behave more like real people, which will further lead to more accurate recommendations. In addition, it is obvious that different people have different ways in forgetting interest, e.g., the speed of forgetting, the initial amount of interest as well as the relearning rate. To personalize the interest-forgetting for different people is also a vital problem.

Why integrate with Markov model? Many item recommendation applications have been observed with the first-order Markov property or high-order ones (Rendle, Freudenthaler, and Schmidt-Thieme 2010; Cheng et al. 2013). In a Markov chain, discrete stochastic states are represented by random variables, which is naturally in line with the interest-forgetting mechanism that the older the item occurs in the consumption history, the larger amount of interest on it will be lost. Besides, Markov models have successfully been applied in recommender systems due to their solid mathematical foundations (Raftery 1985; Dimitrakakis 2010; Begleiter, El-Yaniv, and Yona 2004). Thus, it is sensible and straightforward to integrate interest-forgetting with Markov model for intelligent recommendations.

Challenges and Contributions To address the issue of performing personalized recommendation considering interest-forgetting with Markov model, we are confronted

*Corresponding author: Chaokun Wang.

with two big challenges: (1) It is nontrivial to determine the appropriate model of interest-forgetting. The existing research on memory forgetting have explored many forms of forgetting mechanisms (Averell and Heathcote 2011; Anzanello and Fogliatto 2011). Unfortunately, there is no prior work to validate which one suits people's forgetting of interest better. (2) The way to integrate interest-forgetting with Markov model is unclear. The Markov model is a probabilistic representation of transitions between states like people's choices. Thus, the interest-forgetting also needs to be interpreted in a probabilistic manner.

In this paper, we attempt to tackle these challenges and address this recommendation problem. The major contributions of this paper are summarized as follows:

- We considered the forgetting of interest in the item recommendation problem, which has barely been mentioned in the literature. We also integrated the interest-forgetting mechanism with Markov models towards a "human-minded" recommender system.
- A personalized framework for interest-forgetting Markov model was brought forward in this paper, together with multiple implementations of the experience and the retention components in the framework.
- The experimental evaluation showed that our methods could significantly improve the accuracy of item recommendation compared with the state-of-the-art.

Related Works

Markov models have enjoyed a wide popularity in the recommender system community for decades. In the simplest form, the first-order Markov chain has been successfully applied in the next-basket recommendation (Rendle, Freudenthaler, and Schmidt-Thieme 2010) and the successive point-of-interest recommendation (Cheng et al. 2013) where agents' next choices are considered only relevant to the last step. To break the limitations of first-order Markov chain on the simple dependency between states, high-order Markov models (Raftery 1985) and variable-order Markov models (Begleiter, El-Yaniv, and Yona 2004; Dimitrakakis 2010) are also brought forward in the literature where agents' next movements are dependent on multiple states in previous steps. As the number of order increases, the scale of states in high/variable-order Markov models is exponentially expanded, which leads to ineffectiveness in recommendations (Yang et al. 2010). To deal with the scalability problem, one solution is to multiply a stepwise parameter to the state-to-state transition probabilities while overlooking the set-to-state transitions (Raftery 1985). Bonnin's *n*-gram statistical language method (Bonnin, Brun, and Boyer 2010) also attempts to compute recommendations based on Markov models, but it is a domain-dependent approach on web navigation, which may not be applied in general situations.

Recommender systems are supposed to be intelligent and know human agents well. Unlike computers, people are confronted with unintended memory forgetting (Ebbinghaus 1885). The forgetting/learning curves have been studied to explore the relationship between the quality of people's work and the interruption (Nembhard and Osothsilp 2001; Jaber and Bonney 1997; Anzanello and Fogliatto 2011; Averell and Heathcote 2011). Intuitively, it is sensible to incorporate forgetting into the modeling of intelligent agents in systems. (Packer, Gibbins, and Jennings 2011) employs a semantic forgetting algorithm to remove the infrequently used or cheap to relearn concepts to obtain more workable space and therefore diminish response time. Forgetting is considered as a filter to remove the unnecessary, erroneous and out-of-date information to allow for high performance (Freedman and Adams 2011).

Although there are many existing recommendation methods based on Markov models (Rendle, Freudenthaler, and Schmidt-Thieme 2010; Shani, Heckerman, and Brafman 2005; Yang et al. 2010), very few have considered the forgetting of interest to the best of our knowledge. (Zhao, Sheng, and Zhang 2012) models users' drifting interest in temporal recommendations, which factorizes the ratings by employing Ebbinghaus forgetting curve in computing rating deviation. However, it is subject to the user ratings while ours is not and our work doesn't address recommendation as a rating prediction problem, either.

A Personalized IFMM Framework

In this section, we bring forward a novel framework for personalized **I**nterest-**F**orgetting **M**arkov **M**odel, *abbr*. **IFMM**, to broaden the item recommendation research based on Markov models. We incorporate both experience and forgetting effects of people's interest into Markov models to allow for more accurate modeling on agents' behaviors.

Problem Formulation

Let $\mathcal{X} = \{x_1, x_2, ..., x_{|\mathcal{X}|}\}$ denote the set of items (e.g. music, movie, book) in the given data set, and $\mathcal{U} = \{u_1, u_2, ..., u_{|\mathcal{U}|}\}$ be the set of agents (a.k.a. users) in the system. Under the Markov model setting, we formulate the item recommendation as: Given an observing sequence of items $\mathcal{X}^{u,t} \subseteq \mathcal{X}$, which have already been consumed by agent u at time t, recommend Top-N unseen items for u at next time $t+1$ while each item x in the Top-N list should maximize the following recommendation probability:

$$P(x|\mathcal{X}^{u,t}) = P(X_{t+1}^u = x | X_t^u = x_t^u, ..., X_1^u = x_1^u), \quad (1)$$

where $\mathcal{X}^{u,t} = \{x_t^u, ..., x_1^u\}$, x_i^u is the item consumed by u at time i, while X_i^u is a random variable representing an arbitrary item in $\mathcal{X}$.

λ-VOM

Basically, the above is a variable-order Markov model (VOM) problem (Begleiter, El-Yaniv, and Yona 2004) since the length t of the observing sequence is not fixed and varies with input. However, as (Raftery 1985) puts it, a high order of Markov chain (HOM) usually leads to exponential expansion on the number of states, e.g., $|\mathcal{X}|^{k+1}$ parameters for a k-order Markov chain. Moreover, high order Markov chains also suffer from the sparsity of transitions in the given data set, which usually leads to ineffectiveness in modeling the behaviors of agents. VOM is also confronted with these limitations of HOM. Thus, similar to (Raftery 1985), we simplify the expression of $P(x|\mathcal{X}^{u,t})$ by only multiplying a balancing component λ on each one-step transition probability:

$$P(x|\mathcal{X}^{u,t}) \propto \sum_{j=1}^{t} \lambda_j^{u,t} P(X_{t+1}^u = x | X_{t+1-j}^u = x_{t+1-j}^u) \quad (2)$$
$$= \sum_{j=1}^{t} \lambda_j^{u,t} P(x|x_{t+1-j}^u),$$

where $P(x|x_{t+1-j}^u)$ represents the one-step transition probability from item x_{t+1-j}^u to item x. We consider the one-step transition probability $P(x_i|x_j), \forall x_i, x_j \in \mathcal{X}$ to be fixed for all agents so that the whole number of states as well as parameters could be greatly reduced. Therefore, the λ component should be user-specific (u), time-aware (t) and step-aware (j) in this work (see Eq. 2). We call this model λ-VOM, which forms the basis of our IFMM framework.

Personalized IFMM Framework

So as to improve Markov models to better understand and behave more like the agents in the system, we attempt to incorporate the agents' *memory of interest* on items into Markov models, especially the λ-VOM. The major factor of memory is *forgetting* (or the counterpart *learning*), which plays a great role in agents' behaviors (Averell and Heathcote 2011). The strength of memory is the interplay of both agents' prior experience and amount of forgetting on the items (Wright 1936). Similarly, we believe that agents' interest on items is also influenced by prior experience and forgetting as their memory do. Let $\Upsilon_x^{u,t}$ ($1 \leq \Upsilon_x^{u,t} \leq 2$) be the prior experience of agent u on item x at time t, and $\Phi_x^{u,t}$ ($0 \leq \Phi_x^{u,t} \leq 1$) be the retention of u's interest on x at time t. We fairly set the base values of $\Upsilon_x^{u,t}$ and $\Phi_x^{u,t}$ as 1 where no interest on x is lost while u has only consumed x once before t. However, the experience $\Upsilon_x^{u,t}$ monotonically increases as the agent repeats the consumptions on that item, while the retention of interest $\Phi_x^{u,t}$ is monotonically decreasing as the interval between the time of the consumption on x and t gets larger. Due to the different monotonic properties of $\Upsilon_x^{u,t}$ and $\Phi_x^{u,t}$ as with elapsed time, they naturally have different value ranges, but are still of the same range length.

In our personalized IFMM framework, we combine the λ-VOM and the effect of interest-forgetting ($\Upsilon_x^{u,t}$ and $\Phi_x^{u,t}$) together by defining the λ component in λ-VOM as:

$$\lambda_j^{u,t} = \Upsilon_{x_{t+1-j}^u}^{u,t} \Phi_{x_{t+1-j}^u}^{u,t}, \quad (3)$$

where x_{t+1-j}^u is the item consumed by u at time $t+1-j$ ($1 \leq j \leq t$). Next, we attempt to find proper mathematical expressions of $\Upsilon_x^{u,t}$ and $\Phi_x^{u,t}$ as well as the one-step transition probability between items, so that the item with the largest probability (see Eq. 4) will be the one that agent u finds the most interesting among all the items.

$$P(x|\mathcal{X}^{u,t}) \propto \sum_{j=1}^{t} \Upsilon_{x_{t+1-j}^u}^{u,t} \Phi_{x_{t+1-j}^u}^{u,t} P(x|x_{t+1-j}^u). \quad (4)$$

Suppose each agent u has a set of observing sequences $\mathcal{X}^u = \{\mathcal{X}^{u,t_1}, ..., \mathcal{X}^{u,t_u}\}$, e.g., a set of music playlists, or a set of check-in sequences separated by months. Let Θ denote the set of parameters in the personalized IFMM framework. Then, we define our optimization problem to obtain the optimal Θ^* as follows:

$$\Theta^* = \underset{\Theta}{\operatorname{argmax}} \prod_{u \in \mathcal{U}} \prod_{\mathcal{X}^{u,t} \in \mathcal{X}^u} P(x|\mathcal{X}^{u,t}). \quad (5)$$

The goal of this problem is to maximize the probability of predicting the last item given the rest of an observing sequence $\mathcal{X}^{u,t}$. To solve this optimization problem, we employ the maximum a posteriori (MAP) estimation on the following log-likelihood function:

$$\Theta^* = \underset{\Theta}{\operatorname{argmax}} \mathcal{L} = \sum_{u \in \mathcal{U}} \sum_{\mathcal{X}^{u,t} \in \mathcal{X}^u} \ln(P(x|\mathcal{X}^{u,t})) \quad (6)$$
$$= \sum_{u \in \mathcal{U}} \sum_{\mathcal{X}^{u,t} \in \mathcal{X}^u} \ln(\sum_{j=1}^{t} \Upsilon_{x_{t+1-j}^u}^{u,t} \Phi_{x_{t+1-j}^u}^{u,t} P(x|x_{t+1-j}^u)),$$
$$s.t. \quad 1 \leq \Upsilon_x^{u,t} \leq 2, 0 \leq \Phi_x^{u,t} \leq 1, 0 \leq P(x|x_{t+1-j}^u) \leq 1.$$

Furthermore, the gradient of the log-likelihood $\mathcal{L}$ with respect to the model parameters is given by:

$$\frac{\partial \mathcal{L}}{\partial \Theta} = \sum_{u \in \mathcal{U}} \sum_{\mathcal{X}^{u,t} \in \mathcal{X}^u} \quad (7)$$
$$\frac{\sum_{j=1}^{t} (\frac{\partial \Upsilon_{x_{t+1-j}^u}^{u,t}}{\partial \Theta} \Phi_{x_{t+1-j}^u}^{u,t} + \frac{\partial \Phi_{x_{t+1-j}^u}^{u,t}}{\partial \Theta} \Upsilon_{x_{t+1-j}^u}^{u,t}) P(x|x_{t+1-j}^u)}{\sum_{j=1}^{t} \Upsilon_{x_{t+1-j}^u}^{u,t} \Phi_{x_{t+1-j}^u}^{u,t} P(x|x_{t+1-j}^u)}.$$

Given a training set containing observing sequences of agents, we can iteratively update the parameters Θ using the gradient ascent method. Once the (near) optimal Θ is obtained, our IFMM framework allows for personalized recommendations based on Eq. 4. In the next section, we will introduce the specifications of our IFMM framework, especially the inferences of $\Upsilon_{x_{t+1-j}^u}^{u,t}$ and $\Phi_{x_{t+1-j}^u}^{u,t}$.

Framework Specifications

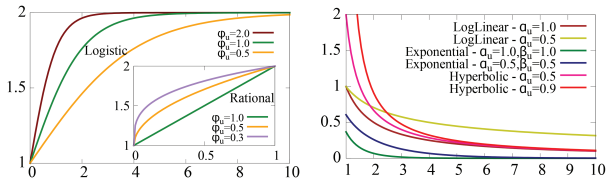
In our personalized IFMM framework, the recommendation probability of an arbitrary item depends on three components — experience, interest retention and one-step transition probability as shown in Eq. 4.

One-Step Transition Probability

The one-step transition probabilities between two items under the personalized IFMM framework are supposed to be fixed. Since λ-VOM is used in the framework, we define the one-step transition probability from item x_j to item x_i as:

$$P(x_i|x_j) = \frac{\sum_{u \in \mathcal{U}} \sum_{\mathcal{X}^{u,t} \in \mathcal{X}^u} \mathbb{1}_{\{x_j, x_i\} \subseteq \mathcal{X}^{u,t}}}{\sum_{u \in \mathcal{U}} \sum_{\mathcal{X}^{u,t} \in \mathcal{X}^u} \mathbb{1}_{x_j \in \mathcal{X}^{u,t}}}, \quad (8)$$

where $\mathbb{1}_{cond}$ is the indicator function, and it returns 1 if $cond$ is satisfied, or otherwise returns 0. $\{x_j, x_i\} \subseteq \mathcal{X}^{u,t}$ represents the condition that $\{x_j, x_i\}$ is a subsequence of $\mathcal{X}^{u,t}$ (x_i appears after x_j in $\mathcal{X}^{u,t}$). This expression of one-step transition probability defines how often that item x_i will be observed after the occurrence of item x_j in a same sequence.



(a) Experience curves (b) Retention curves

Figure 1: Examples of experience curves and interest retention curves.

Experience

The value of experience $\Upsilon_x^{u,t}$ represents the familiarity of agent u on item x at time t. The larger the value of $\Upsilon_x^{u,t}$ is, the more likely that u's choices will be affected by x at time t. Intuitively, agent u's experience on item x should be monotonically increasing with the number of consumptions on x by u. Thus, $\Upsilon_x^{u,t}$ could be considered as a function of the frequency $f_{u,t}(x)$ of u's consumption on item x by time t. Since the value of experience is bounded by 1 (lower) and 2 (upper) in the framework, we employ the logistic function to define experience:

$$\Upsilon_x^{u,t} = \frac{2}{1 + e^{-\phi_u f_{u,t}(x)}} \qquad (f_{u,t}(x) \geq 0, \phi_u \geq 0), \quad (9)$$

where ϕ_u is u's personalized parameter for experience. To explore the proper selection of experience function in another way, we also bring forward a rational function of experience definition:

$$\Upsilon_x^{u,t} = 1 + (f_{u,t}(x))^{\phi_u} \quad (0 \leq f_{u,t}(x) \leq 1, 0 \leq \phi_u \leq 1), \quad (10)$$

where the frequencies of item consumptions are normalized by u's total consumptions before time t. Fig. 1(a) shows some examples of experience curves with both logistic function (Eq. 9) and rational function (Eq. 10). The larger the value of ϕ_u is, the steeper the curve will be for logistic function, but in contrast, the more gradual the curve will be for rational function.

In the personalized IFMM framework, the gradient of experience concerning the parameter ϕ_u is given by:

$$\frac{\partial \Upsilon_x^{u,t}}{\partial \phi_u} = \begin{cases} \frac{2 f_{u,t}(x) e^{-\phi_u f_{u,t}(x)}}{(1+e^{-\phi_u f_{u,t}(x)})^2}, & \text{Logistic function} \\ (f_{u,t}(x))^{\phi_u} \ln(f_{u,t}(x)), & \text{Rational function} \end{cases}. \quad (11)$$

In both cases, the value of experience $\Upsilon_x^{u,t}$ is always bounded by 1 and 2. The implementations of experience can also be replaced by other definitions. In this work, we evaluate our framework with these two experience expressions.

Interest Retention

Interest retention is also the core of our personalized IFMM framework. The memory retention problem has been well studied in the form of forgetting/learning curves in the literature (Anzanello and Fogliatto 2011; Averell and Heathcote 2011; Jaber and Bonney 1997; Ebbinghaus 1885). According to the review in (Anzanello and Fogliatto 2011), the major mathematical expressions of retention can be classified into three categories: log-linear model, exponential model and hyperbolic model. Similarly, as for agents' interest retention on items, we also bring forward three definitions with respect to the memory retention expressions.

Log-Linear Model Let x_{t+1-j}^u denote the item consumed by agent u at time $t+1-j$ $(1 \leq j \leq t)$ in an observing sequence of length t, and $\Phi_{x_{t+1-j}^u}^{u,t}$ be u's interest retention on x_{t+1-j}^u at time $t+1$ (the recommendation step). The log-linear definition of interest retention based on (Wright 1936) is as follows:

$$\Phi_{x_{t+1-j}^u}^{u,t} = C_u j^{-\alpha_u}, \quad (12)$$
$$1 \leq j \leq t; 0 \leq \alpha_u \leq 1; 0 < C_u \leq 1,$$

where α_u and C_u are u's personalized parameters which are also bounded. As the number of time steps gets further from now (i.e., j increases), the interest retention decreases accordingly. The gradients of $\Phi_{x_{t+1-j}^u}^{u,t}$ w.r.t. parameters are:

$$\frac{\partial \Phi_{x_{t+1-j}^u}^{u,t}}{\partial \alpha_u} = -C_u j^{-\alpha_u} \ln(j), \quad (13)$$

$$\frac{\partial \Phi_{x_{t+1-j}^u}^{u,t}}{\partial C_u} = j^{-\alpha_u}. \quad (14)$$

Exponential Model The exponential definition of interest retention is based on (Knecht 1974), and the mathematical form is:

$$\Phi_{x_{t+1-j}^u}^{u,t} = C_u j^{-\alpha_u} e^{-\beta_u j}, \quad (15)$$
$$1 \leq j \leq t; 0 \leq \alpha_u, \beta_u \leq 1; 0 < C_u \leq 1.$$

Similarly, α_u, β_u and C_u are u's bounded personalized parameters. The gradients w.r.t. these parameters are given by:

$$\frac{\partial \Phi_{x_{t+1-j}^u}^{u,t}}{\partial \alpha_u} = -C_u j^{-\alpha_u} e^{-\beta_u j} \ln(j). \quad (16)$$

$$\frac{\partial \Phi_{x_{t+1-j}^u}^{u,t}}{\partial \beta_u} = -C_u j^{1-\alpha_u} e^{-\beta_u j}. \quad (17)$$

$$\frac{\partial \Phi_{x_{t+1-j}^u}^{u,t}}{\partial C_u} = j^{-\alpha_u} e^{-\beta_u j}. \quad (18)$$

Hyperbolic Model The hyperbolic definition of interest retention is based on (Mazur and Hastie 1978) with personalized parameters α_u and C_u,

$$\Phi_{x_{t+1-j}^u}^{u,t} = \frac{C_u}{j - \alpha_u}, 0 \leq \alpha_u < 1, 0 < C_u \leq 1. \quad (19)$$

Similar to other models, the gradients of interest retention with regard to α_u and C_u are:

$$\frac{\partial \Phi_{x_{t+1-j}^u}^{u,t}}{\partial \alpha_u} = \frac{C_u}{(j - \alpha_u)^2}. \quad (20)$$

$$\frac{\partial \Phi_{x_{t+1-j}^u}^{u,t}}{\partial C_u} = \frac{1}{j - \alpha_u}. \quad (21)$$

In the above models, C_u is the personalized maximum interest retention value of user u. α_u in Log-Linear model and Exponential model, and β_u in Exponential model, are the personalized decaying parameters representing the interest forgetting speeds of users. The larger the values of these parameters are, the faster the interest is lost. Besides, α_u in Hyperbolic model compensates the number of consumption steps, and it personalizes the start step of interest forgetting.

Fig. 1(b) illustrates some examples of interest retention models. For simplicity, the scale parameter C_u here is set to 1.0. We can see that the curves of hyperbolic model are the steepest ones, and those of exponential model are the closest to the horizontal axis. The values on the log-linear curves are relatively higher but also not so differentiable between each other by contrast. In addition, the set of parameters Θ in our framework consists of all the personalized parameters in the definitions of experience and interest retention, such as ϕ_u, α_u, β_u and C_u in this paper. These personalized parameters are randomly initialized in the range $[0.0, 1.0]$ except for the ϕ_u of non-normalized experience which is randomly drawn from absolute $N(0, 0.1)$. By using the learning method introduced in the framework, we can further perform personalized recommendations based on Eq. 4.

Experiments

Experimental Settings

We conducted experiments on the large scale music listening data set collected from Last.fm (Celma 2010). Music recommendation is commonly used in people's daily lives (Wang, Rosenblum, and Wang 2012), and people's listening behaviors are a mixture of new songs and the old songs in the playlists (Anderson et al. 2014). Thus, music recommendation is a typical representative of application scenarios of the personalized IFMM framework. This data set contains 16,986,614 listening records from 992 users on 964,463 songs. We partitioned the listening history of each user into sessions (*a.k.a.* observing sequences) based on a timeout threshold like 1 hour which is the default in our experiments. The threshold measures the maximum acceptable interval between two adjacent records in time order to be considered in the same session. Meanwhile, the listening records whose duration is less than 30 seconds are considered as dislikes of users, and are removed before further experiments.

We evaluated our methods with two experience implementations and three interest retention implementations, i.e., NM+LL, NM+EX, NM+HY, NO+LL, NO+EX and NO+HY where NM, NO, LL, EX and HY represent normalized experience (logistic function), non-normalized experience (rational function), log-linear retention, exponential retention and hyperbolic retention, respectively. We compared our method with the state-of-the-art item recommendation methods, i.e., the Markov-based ones like FPMC (Rendle, Freudenthaler, and Schmidt-Thieme 2010; Cheng et al. 2013) and TSPR (Haveliwala 2002), the graph-based preference fusion STG (Xiang et al. 2010), as well as the sequential pattern based method SEQ (Hariri, Mobasher, and Burke 2012). The sequential pattern based method was evaluated with different supports and window sizes, e.g., SEQ-s5w4 means the method with support as 5 and window size as 4. Since IFMM framework doesn't solve recommendations as a rating prediction problem and there are also no ratings in the Lastfm dataset, we don't consider those rating-oriented approaches, e.g. timeSVD++ (Koren 2009) and NMF, as appropriate comparative ones to our method.

Table 1: Hit ratios of the proposed methods.

Method	Top10	Top30	Top50	Top70	Top90	Top100
NM+LL	0.3288	0.4163	0.4536	0.4756	0.4919	0.4983
NM+EX	0.3335	0.4189	0.4553	0.4766	0.4924	0.4983
NM+HY	**0.3954**	**0.4573**	0.4816	0.4962	0.5068	0.5113
NO+LL	0.3289	0.4164	0.4537	0.4757	0.4921	0.4984
NO+EX	0.3345	0.4198	0.4557	0.4769	0.4927	0.4987
NO+HY	0.3935	**0.4573**	**0.4834**	**0.4991**	**0.5109**	**0.5154**

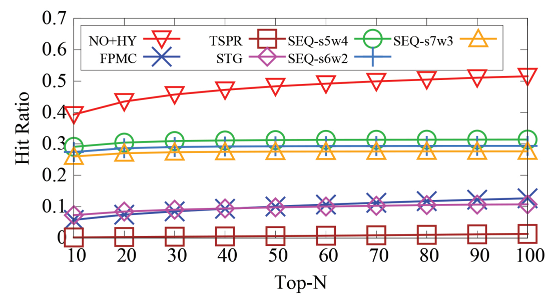


Figure 2: Comparisons on the accuracy of recommendation.

Overall Accuracy Performance

We first evaluated the overall accuracy performance of our proposed methods and the baselines. In the evaluation, we used 80% of each user's observing sequences to learn the personalized parameters in the IFMM framework as well as the parameters of the comparative models. Then, the other 20% observing sequences were used to test the performance by predicting the last song in each test observing sequence while using its rest as input.

Table 1 shows the hit ratios of the proposed methods in the accuracy evaluation under different length of recommendation list. Generally, all the proposed methods exhibit promising accuracy performance in the experiment. As for the choice of retention expressions, we observed that HY methods outperformed the other methods in this evaluation. Meanwhile, the ones with log-linear retention and exponential retention showed similar performance in accuracy. As for the choice of experience expressions, there is not much difference between the hit ratios of methods with normalized experience and non-normalized experience compared with the methods with different retention expressions. Although experience and interest retention are both essential factors in people's forgetting process, this comparison indicates that interest retention is more important than experience on people's listening choices in the Lastfm data set.

Next, we compared the accuracy performance between our methods and the baselines. NO+HY is selected as the representative of our methods since it shows the best perfor-

Table 2: Value distributions of the learned parameters with respect to each method in the experiments. Ranges of parameter values are represented with parentheses.

(a) Value distribution of Normalized ϕ_u.

Method	[0.0,0.2)	[0.2,0.4)	[0.4,0.6)	[0.6,0.8)	[0.8,1.0]
NM+LL	0.945	0.011	0.012	0.014	0.018
NM+EX	0.937	0.015	0.014	0.014	0.020
NM+HY	0.945	0.013	0.009	0.022	0.011

(b) Value distribution of Non-Normalized ϕ_u.

Method	[0.0,2.0)	[2.0,4.0)	[4.0,6.0)	[6.0,8.0)	[8.0,∞)
NO+LL	0.111	0.484	0.403	0.002	0.000
NO+EX	0.128	0.657	0.213	0.002	0.000
NO+HY	0.174	0.492	0.333	0.001	0.000

(c) Value distribution of α_u.

Method	[0.0,0.2)	[0.2,0.4)	[0.4,0.6)	[0.6,0.8)	[0.8,1.0]
NM+LL	0.810	0.066	0.056	0.049	0.019
NM+EX	0.755	0.077	0.080	0.060	0.028
NM+HY	0.877	0.061	0.034	0.020	0.008
NO+LL	0.831	0.052	0.059	0.037	0.021
NO+EX	0.760	0.079	0.068	0.053	0.040
NO+HY	0.889	0.064	0.026	0.018	0.003

(d) Value distribution of β_u.

Method	[0.0,0.2)	[0.2,0.4)	[0.4,0.6)	[0.6,0.8)	[0.8,1.0]
NM+EX	0.696	0.084	0.095	0.085	0.040
NO+EX	0.700	0.093	0.101	0.067	0.039

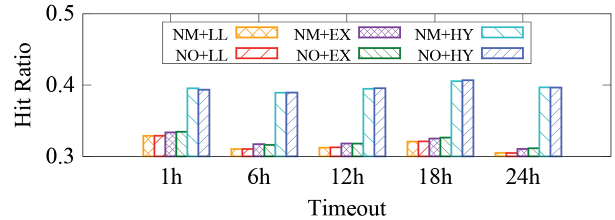
(e) Value distribution of C_u.

Method	[0.0,0.2)	[0.2,0.4)	[0.4,0.6)	[0.6,0.8)	[0.8,1.0]
NM+LL	0.005	0.019	0.030	0.046	0.900
NM+EX	0.002	0.013	0.037	0.056	0.892
NM+HY	0.003	0.023	0.030	0.053	0.891
NO+LL	0.004	0.021	0.032	0.050	0.893
NO+EX	0.006	0.021	0.027	0.057	0.889
NO+HY	0.002	0.013	0.036	0.051	0.898

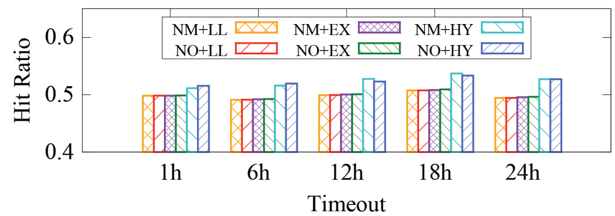
mance among the 6 proposed methods (see Table 1). Fig. 2 illustrates the comparison results where we can see NO+HY dominates all the baselines under different length of recommendation list. The absolute improvement of NO+HY is about 10% to 20% compared with the best of the baselines. The sequential pattern based methods showed competitive results by reaching hit ratios around 30%, while FPMC, STG and TSPR are not observed with high accuracy in the experiments. In all, our method can significantly improve the accuracy of item recommendations in the experiments.

Personalized Parameters

We analyzed the value distributions of the learned personalized parameters in our methods to explore the functionality of parameters as shown in Table. 2. The values of parameters are grouped and represented with parentheses. For instance, in Table 2(a), the value 0.018 in the cell with respect to method NM+LL and parameter range $[0.8,1.0]$ means there are about 1.8% of users whose normalized ϕ_u values fall between $[0.8,1.0]$ when computed with method NM+LL. By comparing the distributions of experience parameter ϕ_u in



(a) Top-10 recommendations.



(b) Top-100 recommendations.

Figure 3: The impact of timeout on the performance.

Table. 2(a) and Table. 2(b), we can see that for the NM methods, most users have the normalized ϕ_u values between $[0.0,0.2)$, which are not very differentiable. While for the NO methods, a large number of users show their ϕ_u values around $2.0 - 4.0$, which leads to wider difference in contrast.

Table. 2(c) shows the distribution of α_u from the retention expressions. Many users tend not to be personalized on this parameter since the α_u values of about $75\% - 90\%$ users are within $[0.0,0.2)$. However, the fractions of α_u-less-personalized users are different where the EX methods have lower fractions while the HY methods have higher ones.

Similarly, Table. 2(d) shows the distribution of β_u value from the exponential retention expression. The personalization on β_u is more significant compared to the other parameters although there are still 70% users having β_u within $[0.0,0.2)$. However, the fractions of personalized users contribute to the improvement in the recommendation accuracy.

Besides, the distribution of C_u parameter from the retention expressions is shown in Table. 2(e). C_u is a scaling parameter which controls the maximum interest retention a user has. Unsurprisingly, most users have $C_u > 0.8$ while a small fraction of users have smaller C_u values.

From the above analysis, we know that not all users tend to be personalized enough in all dimensions. In contrast, different human agents tend to be personalized on different aspects. However, even only one personalized parameter exists among the above for any human agent, the recommendation results will be much different. The diversified personalization ways may be one of the reasons why our methods outperformed the baselines in the accuracy as shown in Fig. 2.

Timeout

We also evaluated the performance of our methods as with the change of timeout which partitions the observing sequences along the history of people's consumptions. Timeout influences the general length of observing sequences, i.e., the order of λ-VOM. A larger timeout will usually lead to longer sequences which incorporate more contextual information into user "contexts" for next-step recommendations. As illustrated in Fig. 3 of the impact of timeout on the recommendation accuracy, we found slight fluctuations on hit ratios of our methods as timeout changes for both Top-10 and Top-100 recommendations. In our experiments, the best choice of timeout seems to be about 18 hours from Fig. 3. However, the overall performance of our methods is not apt to be significantly affected by the setting of timeout.

In sum, the experiments showed that our methods under the IFMM framework can significantly improve the item recommendations with diversified personalization strategies, and the performance of our methods is stable and promising. Besides, the time complexity of online recommendation using IFMM framework is linear to the length of the observing sequence and the number of transferable items, which makes IFMM efficient and scalable. The IFMM framework is applicable as long as the consumption time (or order) information is available regardless of the type of items. However, for scenarios where few reconsumption behaviors on a same item from a same user are observed, e.g. movie, IFMM still works since the interest retention could be computed though the experience would always be 1.

Our methods is not vulnerable to new users because we can assume several consumptions randomly or based on popularity for new users, and recommend items to them. As long as consumptions are generated, we can remove the assumptions and leave the real consumptions in the observing sequences. As for new items, our method may need to recalculate the transition probabilities in a batch mode periodically. However, only the transition probabilities of items in the new batch of observing sequences will be updated, which means only a local updating is required.

Conclusions and Future Work

In this paper, we attempted to incorporate the forgetting of interest into Markov models to better simulate the decisions of intelligent agents for personalized recommendations. To address this issue, we brought forward a framework of personalized interest-forgetting Markov model as well as the specifications of the framework. The experimental evaluations showed that our methods could significantly improve the item recommendation accuracy. In the future, we prospect this framework to behave in a human way to compute recommendations, i.e. human intelligent features should be incorporated into this framework, especially the mind activities like learning, forgetting and relearning.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (No. 61373023, No. 61170064, No. 61133002) and the National High Technology Research and Development Program of China (No. 2013AA013204).

References

Anderson, A.; Kumar, R.; Tomkins, A.; and Vassilvitskii, S. 2014. The dynamics of repeat consumption. In *WWW*, 419–430.

Anzanello, M. J., and Fogliatto, F. S. 2011. Learning curve models and applications: Literature review and research directions. *Int J Ind Ergonom* 41:573–583.

Averell, L., and Heathcote, A. 2011. The form of the forgetting curve and the fate of memories. *J Math Psychol* 55:25–35.

Begleiter, R.; El-Yaniv, R.; and Yona, G. 2004. On prediction using variable order markov models. *J Artif Intell Res* 22:385–421.

Bonnin, G.; Brun, A.; and Boyer, A. 2010. *Skipping-based collaborative recommendations inspired from statistical language modeling*. InTech.

Celma, O. 2010. *Music Recommendation and Discovery in the Long Tail*. Springer. chapter 3.

Cheng, C.; Yang, H.; Lyu, M. R.; and King, I. 2013. Where you like to go next: successive point-of-interest recommendation. In *IJCAI*, 2605–2611.

Dimitrakakis, C. 2010. Bayesian variable order markov models. In *AISTATS*, 161–168.

Ebbinghaus, H. 1885. *Memory: A Contribution to Experimental Psychology*. New York by Teachers College, Columbia University.

Freedman, S., and Adams, J. 2011. Filtering data based on human-inspired forgetting. *IEEE Trans Syst Man Cybern B Cybern* 41(6):1544–1555.

Hariri, N.; Mobasher, B.; and Burke, R. 2012. Context-aware music recommendation based on latent topic sequential patterns. In *RecSys*, 131–138.

Haveliwala, T. H. 2002. Topic-sensitive pagerank. In *WWW*, 517–526.

Jaber, M. Y., and Bonney, M. 1997. A comparative study of learning curves with forgetting. *Appl Math Model* 21(8):523–531.

Knecht, G. 1974. Costing, technological growth and generalized learning curves. *Oper Res Quart* 25(3):487–491.

Koren, Y. 2009. Collaborative filtering with temporal dynamics. In *KDD*, 447–456.

Mazur, J., and Hastie, R. 1978. Learning as accumulation: a reexamination of the learning curve. *Psychol Bull* 85(6):1256–1274.

Nembhard, D. A., and Osothsilp, N. 2001. An empirical comparison of forgetting models. *IEEE Trans Eng Manag* 48(3):283–291.

Packer, H. S.; Gibbins, N.; and Jennings, N. R. 2011. An on-line algorithm for semantic forgetting. In *IJCAI*, 2704–2709.

Raftery, A. E. 1985. A model for high-order markov chains. *J Roy Statist Soc Ser B* 47(3):528–539.

Rendle, S.; Freudenthaler, C.; and Schmidt-Thieme, L. 2010. Factorizing personalized markov chains for next-basket recommendation. In *WWW*, 811–820.

Shani, G.; Heckerman, D.; and Brafman, R. 2005. An MDP-based recommender system. *J Mach Learn Res* 6:1265–1295.

Wang, X.; Rosenblum, D.; and Wang, Y. 2012. Context-aware mobile music recommendation for daily activities. In *ACM Multimedia*, 99–108.

Wright, T. 1936. Factors affecting the cost of airplanes. *J. Aeronaut. Sci.* 3(2):122–128.

Xiang, L.; Yuan, Q.; Zhao, S.; Chen, L.; Zhang, X.; Yang, Q.; and Sun, J. 2010. Temporal recommendation on graphs via long- and short-term preference fusion. In *KDD*, 723–732.

Yang, Q.; Fan, J.; Wang, J.; and Zhou, L. 2010. Personalizing web page recommendation via collaborative filtering and topic-aware markov model. In *ICDM*, 1145–1150.

Zhao, L.; Sheng, F.; and Zhang, B. 2012. Collaborative filtering based on user's drifting interests. *IJACT* 4(15):336–345.