Modeling the Interest-Forgetting Curve for Music Recommendation

Jun Chen  Chaokun Wang  Jianmin Wang
School of Software, Tsinghua University, Beijing 100084, China
chenjun12@mails.thu.edu.cn, chaokun@tsinghua.edu.cn, jimwang@tsinghua.edu.cn

ABSTRACT
Music recommendation plays a key role in our daily lives as well as in the multimedia industry. This paper adapts the memory forgetting curve to model the human interest-forgetting curve for music recommendations based on the observation of recency effects in people’s listening to music. Two music recommendation methods are proposed using this model with respect to the sequence-based and the IFC-based transition probabilities, respectively. We also bring forward a learning method to approximate the global optimal or personalized interest-forgetting speed(s). The experimental results show that our methods can significantly improve the accuracy in music recommendations. Meanwhile, the IFC-based method outperforms the sequence-based method when recommendation list is short at each time.

Categories and Subject Descriptors
H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—Information filtering.

Keywords
Music Recommendation; Interest-Forgetting Curve; Recency Effect.

1. INTRODUCTION
In time domain, people’s consumption of music can be grouped by playlists (equivalently sessions, in this work) which consist of user preferred songs in an order. The analysis on these music playlists gives us insights on how listening behaviors could reveal people’s emotional states, user contexts and short-term preferences [13]. Music recommendations [4, 9], especially the context-aware branches [5, 8, 10, 12], directly benefit from the analysis of user’s playlists. The work in [8] first infers latent topics from the frequently used tags for each song, and mines sequential patterns of latent topics for recommendations. The research in [10] clusters users into groups and separately computes song-to-song transition probabilities within each group, based on which the recommendations are conducted using random walks. The performance of these methods heavily depends on user contexts which can be inferred from mobile sensor data [12], user articles [5], or user interactions with systems [10].

Besides, recency effects have also been observed in people’s listening to music [1, 7], where people have a tendency to recall the recently heard songs other than those heard long ago. A straightforward intuition is that songs consumed more recently contribute more to the understanding of current user context, which leads to non-increasing weights as the consumption timestamps get farther from now. Thus, we believe it is reasonable and appropriate to use the forgetting curve [2, 6] to model this feature.

Given a user’s current music playlist, we would like to explore the impacts of the heard songs upon the user’s next listening choice. However, since the short-term music preference of users usually changes rapidly, it is nontrivial to predict what (s)he likes next, especially when the playlist contains dissimilar songs of different genres, from different artists and periods. Many existing works uniformly or randomly initialize the impacts of the heard songs in recommendations [3, 8, 10]. But it could lead to information loss compared with weighted initializations considering the recency effects in people’s listening to music. This paper attempts to model how recently heard music pieces affect people’s next listening choices by incorporating the concept of interest-forgetting curve which is inspired by human’s memory forgetting curve. Our method models the interest losing rate of users, and assigns descending initial weights to the heard songs in playlists in determining the next listening choice.

The main contributions of this paper include:

- We brought forward a novel interest-forgetting curve model with two music recommendation methods, which could more accurately characterize how recently heard songs affect people’s next listening choices.
- We also proposed a learning method to approximate the global optimal or personalized interest-forgetting speed(s).
- The experimental results showed that our methods outperformed two widely used baselines in music recommendations.

2. INTEREST-FORGETTING CURVE
The memory forgetting curve was first introduced by Ebbinghaus [6] as expressed by an exponential formula (Eq. 1). The
for each user \( u \); the more likely it will affect people’s next listening choice to model the recency effects in people’s listening to music and be strengthened through people’s over-learning on events, which leads to a shallower forgetting curve.

\[
R = e^{-\frac{t}{\alpha}}, \quad \alpha \in \mathbb{R}_{>0},
\]

(2)

where \( t \) is the time interval from the happening of an event till now, and \( \alpha \) is a non-negative real parameter to control the interest-forgetting speed. The larger the value of \( \alpha \) is, the faster the interest loses as time elapses. Particularly, if \( \alpha = 0 \), then IFC is degenerated to the uniform weighted model, i.e., no interest on the past events is lost. We’d like to adapt these aforementioned forgetting curves to model the faster interest loses as time elapses. Particularly, \( R_{s} \) and \( R_{t} \) are the interest-forgetting speed. The larger the value of \( \alpha \) is, the more likely it will affect people’s next listening choice according to the recency effects.

3. MUSIC RECOMMENDATION WITH IFC MODEL

3.1 Problem Formulation

Let \( S = \{s_1, s_2, \ldots, s_{|S|}\} \) denote the set of songs, and \( U = \{u_1, u_2, \ldots, u_{|U|}\} \) be the set of users in the data set. Songs are consumed one after another in a playlist by a user. The directed transition probabilities between songs are represented by an \( |S| \times |S| \) row-normalized asymmetric matrix \( M \) where entry \( M(s_i, s_j) \geq 0 \) is the probability of listening song \( s_j \) right after listening \( s_i \). Each user has a set of music playlists, \( \mathcal{P} = \{p_1, p_2, \ldots, p_n\} \), while each playlist contains a sequence of ordered songs, \( p_i \in \mathcal{S}^1 \).

Our music recommendation problem can be formulated as: Given \( u \)'s current playlist \( p^u \) (current session has not expired yet), past playlists set \( \mathcal{P}^u \) of \( u \), and the transition probability matrix \( M \), we recommend the Top-K most probable songs for \( u \) to listen in the next considering \( u \)'s oblivion of interest on past songs. The recommended songs are selected by their observing probabilities (Eq. 3). To avoid the trivial repetition of a same song, we only recommend songs which have not yet been observed in the current playlist.

\[
s_{rec} = \max_{s_i \in \mathcal{S}} \Pr(s_i | p^u, \mathcal{P}^u, M).
\]

(3)

3.2 Recommendation Framework

We define the probability of recommending song \( s_i \) as the sum of probabilities to visit \( s_i \) via any song \( s_j \) appeared in the current playlist (Eq. 4).

\[
\Pr(s_i | p^u, \mathcal{P}^u, M) = \sum_{s_j \in \mathcal{P}^u} \Pr(s_j | p^u) M(s_j, s_i).
\]

(4)

Since we do not recommend the observed songs in the current playlist as described in Section 3.1, the above computation should be limited on a small set of candidates as,

\[
\mathcal{C}_{p^u} = \{s_x | \exists s_j \in p^u \land M(s_j, s_x) > 0 \land s_x \notin p^u\}.
\]

(5)

\( \mathcal{C}_{p^u} \) contains the unobserved songs in the neighborhood of \( p^u \). Thus, \( M(s_j, s_x) \) in Eq. 4 should be replaced by its neighborhood-normalized value \( M_{\mathcal{C}_{p^u}}(s_j, s_x) \) (Eq. 6).

\[
M_{\mathcal{C}_{p^u}}(s_j, s_x) = \frac{M(s_j, s_x)}{\sum_{s_x' \in \mathcal{C}_{p^u}} M(s_j, s_x')}.
\]

(6)

We can make our recommendations based on the probabilities of songs to be observed using the aforementioned equations. The factor of interest-forgetting is implanted into the definitions of \( M(s_j, s_x) \) and \( \Pr(s_j | p^u) \), which will be discussed in detail by modeling IFC next.

3.3 Modeling IFC in Recommendation

3.3.1 Computing Transition Probabilities

The transition probability \( M(s_j, s_x) \) is very important to represent directed weighted pairwise relations between songs. We proposed two definitions of \( M(s_j, s_x) \) with respect to our two recommendation methods.

Sequence-based transition probabilities. Let \( \{x\}_{seq} \) be the representation of a sequence. We say \( \{s_i, s_j\}_{seq} \subseteq p \iff \{s_i, s_j\}_{seq} \) is a subsequence of playlist \( p \). Then, we present the sequence-based \( M(s_i, s_j) \) as,

\[
M_{SEQ}(s_i, s_j) = \frac{\sum_{u \in U} \sum_{p \in \mathcal{P}^u} \mathbb{1}(s_i, s_j)_{seq} \subseteq p}{\sum_{u \in U} \sum_{p \in \mathcal{P}^u} \mathbb{1}(s_i)_{seq} \subseteq p},
\]

(7)

where \( \mathbb{1}_{cond} \) is the indicator function which returns 1 when \( cond \) is satisfied and otherwise returns 0. This definition captures the percentage of sequential co-occurrences of songs.

IFC-based transition probabilities. By considering the oblivion of interest on past songs, we bring forward the IFC-based \( M(s_i, s_j) \) as,

\[
M_{IFC}(s_i, s_j) = \frac{\sum_{u \in U} \sum_{p \in \mathcal{P}^u} \mathbb{1}(s_i, s_j)_{seq} \subseteq p}{\sum_{u \in U} \sum_{p \in \mathcal{P}^u} \mathbb{1}(s_i)_{seq} \subseteq p} e^{-\alpha(p(s_j)-p(s_i))},
\]

(8)

where \( p(s_i) \) represents the position index of \( s_i \) in \( p \), and \( \alpha \) is the interest-forgetting speed. Without loss of generality, the position index starts with 1 from the first heard song in the playlist and increases by 1 when jumping to the next song each time. This IFC-based definition takes the relative listening time steps into accounts when computing the relations between songs in line with Eq. 2. Compared with \( M_{SEQ}(s_i, s_j) \), \( M_{IFC}(s_i, s_j) \) could be more accurate to describe the pairwise relations between songs, but its performance is apt to be affected by the setting of \( \alpha \). The effectiveness of these two methods and the influence of \( \alpha \) will be evaluated in later experiments.
3.3.2 Computing Prior Probabilities

Based on the properties of IFC, the more recently the songs are heard, the more important they are supposed to be to affect people’s next listening choice, and the higher prior probabilities they should have. According to Eq. 2, we proposed our definition of prior probabilities,

\[
Pr(s_j|p^n) = \frac{e^{-\alpha(p^n-s_j)}}{\sum_{s_j \in p^n} e^{-\alpha(p^n-s_j)}},
\]

(9)

where \( |p^n| \) is the number of songs in \( p^n \) while \( p^n(s_j) \) is the position index of \( s_j \) in \( p^n \). This definition gives descending prior probabilities to the heard songs in current session as their listening timestamps get farther from now, which is in line with the properties of IFC in Section 2.

3.3.3 Learning Interest-Forgetting Speed

Throughout the modeling of \( M(s_i, s_j) \) and \( Pr(s_j|p^n) \) based on the exponential form of forgetting curve in Eq. 2, \( \alpha \) is the only parameter remaining unknown. Though Eq. 2 can be replaced by more complex forgetting curve models \cite{2}, we simplify the problem here and leave the extensions in our future work. If \( \alpha \) is properly set, the IFC model will be more accurate to represent the weighted importance of the heard songs in determining people’s next listening choice. Here we proposed a learning method to compute the near-optimal value of \( \alpha \). Given a set of training playlists, our intuitive goal is to maximize the average probability of correctly recommending the last song in each playlist \( p \) while using the rest \( |p| - 1 \) songs as input. The objective function is shown in Eq. 10, where \( P_{train} \) is the set of training playlists, \( C_{sp} \) is the last song in playlist \( p \), \( \hat{s} \) is the set minus operator, and \( M_{sp} \) is the neighborhood-normalized transition probability matrix (Eq. 6).

\[
G(\alpha) = \max_{\alpha \in \mathbb{R}_>0} \frac{1}{|P_{train}|} \sum_{p \in P_{train}} Pr(s_j|p \setminus \{s_j\}, P_{train}, M_{sp})
\]

\[
= \max_{\alpha \in \mathbb{R}_>0} \frac{1}{|P_{train}|} \sum_{p \in P_{train}} \sum_{s_j \in p(1:p)} Pr(s_j|p \setminus \{s_j\}) M_{sp}(s_j, s_j)
\]

(10)

Since the transition probability matrix \( M \) is considered to be fixed for a given data set, we therefore only need to learn a proper \( \alpha \) for \( Pr(s_j|p) \). Thus, when using \( M_{IFC}(s_i, s_j) \) as the representation of transition probabilities, we first compute \( M_{IFC}(s_i, s_j) \) with a pre-defined \( \alpha_{pre} \) (e.g., \( \alpha_{pre} = 1.0 \)) and then we conduct the learning method to get a near-optimal \( \alpha \) for \( Pr(s_j|p) \). Since \( Pr(s_j|p) \) (Eq. 9) is derivable on \( \alpha \), \( G(\alpha) \) is therefore also derivable on \( \alpha \). However, it is difficult to compute the optimal \( \alpha \) given \( \frac{\partial G(\alpha)}{\partial \alpha} = 0 \), because the partial derivative is the sum of many different \( \alpha \)-related exponential components. But we can use a gradient method to get a near-optimal \( \alpha \) value by iteratively increasing or decreasing \( \alpha \) according to the sign of the partial derivative.

Moreover, the proposed learning method can be extended to compute personalized interest-forgetting speed \( \alpha_u \). If we replace \( P_{train} \) with \( P^u \) in Eq. 10, we will get personalized \( \alpha_u \) for each individual user \( u \) with the minimum effort. The personalized IFC model is more suitable than IFC model if the data set is considered to be personalized.

4. EXPERIMENTS

Now that we have developed our music recommendation methods based on the IFC model, we turn to conduct experiments to evaluate the performance of our method compared with the baselines. We begin by introducing the experimental settings, and then analyze the evaluation results.

4.1 Experimental Settings

We conducted the experiments on the Last.fm data set \cite{2} which contains users’ history of binary music consumption with timestamps. This data set consists of 10,244,067 music listening records collected from 992 users on 964,463 songs. We partitioned each user’s listening history into playlists with a timeout bound as 30 minutes which is an empirical setting — neither too long nor too short. If the idle time between two adjacent listening records is beyond 30 minutes, these two listening records are partitioned into the former and the latter playlists, respectively. Inside a playlist, if the idle period between any two adjacent listening records is less than 30 seconds, we assumed the user disliked the former song, and we removed the former listening record from the playlist. If one song appeared more than once in a same playlist, we only preserved its first occurrence and removed the rest repetitions. Furthermore, to minimize the accuracy caused by very short playlists, we only considered those playlists with at least 5 distinct songs. For playlists containing more than 30 distinct songs, we only preserve their first 30 distinct songs as playlists. There are totally 599,353 playlists generated from the data set. We randomly sampled 67,911 playlists as the test set which covers all users. The training set consists of the rest 531,442 playlists. Apparently, the larger the training set is, the more accurate the IFC will be to model people’s listening preferences.

Since we do not predict ratings on songs like \cite{9} and no other information besides binary listening records is available in this data set, we compared our method with two widely used baselines — the Neighborhood method (abbr. NH) \cite{3} and the Sequential Pattern method (abbr. SP) \cite{8}. NH and SP are similar to our work and only use binary listening records as input. NH computes Top-N neighbors for the input playlist using cosine similarity, e.g., \( N = 100 \). It recommends songs with the highest frequencies among the Top-N neighbors. SP firstly mines the sequential patterns using PrefixSpan method \cite{11}, and then recommends songs with the largest confidence by appending the candidate songs to the tail of the input sequence. For the SP method, we conducted experiments with different support s of frequencies and window sizes, e.g., SP-s5w3 represents the SP method with support as 5 and window size as 3.

4.2 Evaluations

We first evaluated the accuracy performance of our method compared with the baselines. The training set was used to construct the transition probability matrix \( M \), and to learn the near-optimal forgetting speed \( \alpha \). The evaluation was performed on the test set. We evaluated the methods using the average hit ratio which is the average ratio of correctly predicting the last heard song in each playlist \( p \) while using its rest \( |p| - 1 \) listening records as input. Two implementations of \( M \) defined in Section 3.3.1 were both evaluated. The method using \( M_{SEQ}(s_i, s_j) \) and \( Pr(s_j|p^n) \) is denoted as IFCM-SEQ, while the one using \( M_{IFC}(s_i, s_j) \) and \( Pr(s_j|p^n) \) is denoted as IFCM-IFC.

Fig. 1 shows the comparison on accuracy performance. It is evident that our methods have significant improvement.
on the average hit ratio compared with the baselines, especially when only a small number of songs are recommended each time. Usually, recommender systems are not expected to give a large number of recommendations for a user each time, which will jeopardize the user friendliness. IFCM-IFC dominates IFCM-SEQ in accuracy when Top-K ≤ 40. This proves that IFCM-IFC stresses more interest loss in computing transition probabilities and therefore promotes the rankings of related songs. As we have pointed out, the higher accuracy the method has when only a small number of songs are recommended each time, the higher practical value the method will be of. Thus, we consider IFCM-IFC as a little optimization based on IFCM-SEQ. In average, our methods (IFCM-SEQ and IFCM-IFC) have about 5% absolute improvement in average hit ratios compared with NH, and even larger compared with SP.

Fig. 2 illustrates the influence of $\alpha$ on the average hit ratio of Top-10 results. The larger the value of $\alpha$ is, the faster the interest is forgetting and the more significant recency effect the model assumes. However, the decline of curves in Fig. 2 after $\alpha > 1$ elucidates that over-assumption on the recency effect could also fail the IFC model. The optimal setting of $\alpha$ is around 0.5 which is also the value obtained using our learning method. We can see the influence of $\alpha$ is smooth when $\alpha \geq 0.2$, especially for IFCM-IFC. Thus, the proposed method based on IFC model is considered to be robust.

We also evaluated the accuracy performance of our method given different lengths of input playlists (Fig. 3). The distribution of the length of playlists in the training set and the test set is illustrated in Fig. 3(a). Generally, there are more short playlists than the long ones since the number of playlists drops as the playlist length increases. This observation fits the real situation that most people usually listen to a small number of distinct songs during one session. Fig. 3(b) shows the average hit ratio of our method within the different-length groups of playlists. Interestingly, we observed higher average hit ratio in the group of playlists whose length is between 10 and 14. Beyond or below this range, it leads to lower average hit ratios, and worst in the $> 30$ group (playlists with no less than 30 distinct songs). We may conclude that the effect of human retention of interest has upper and lower boundaries, and inside the range, IFC model has the best performance, while outside the range, IFC model may suffer from random effects (below the lower bound) or noises (beyond the upper bound).

In sum, the experimental results show that IFC model can significantly improve the accuracy performance in music recommendations.

5. CONCLUSIONS

In this paper, we proposed a novel music recommendation method based on the IFC model to predict a user’s next choice considering the recency effects in people’s listening to music. We presented the recommendation framework with two open ends — the computation of transition probabilities and prior probabilities. We brought forward the sequence-based and the IFC-based solutions to our problem. The experimental results showed a significant improvement in the accuracy of our music recommendation method compared with the baselines. In the future, we will consider to improve the IFC model by considering more other existing forgetting curve models.

6. ACKNOWLEDGEMENT

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

7. REFERENCES


