4-2014

Just-for-Me: An Adaptive Personalization System for Location-Aware Social Music Recommendation

Zhiyong CHENG
Singapore Management University, zy.cheng.2011@smu.edu.sg

Jialie Shen
Singapore Management University, jlshen@smu.edu.sg

Just-for-Me: An Adaptive Personalization System for Location-Aware Social Music Recommendation

Zhiyong Cheng Jialie Shen
School of Information Systems, Singapore Management University, Singapore
{zy.cheng.2011, jlshen}@smu.edu.sg

ABSTRACT

The fast growth of online communities and increasing popularity of internet-accessing smart devices have significantly changed the way people consume and share music. As an emerging technology to facilitate effective music retrieval on the move, intelligent recommendation has been recently received great attentions in recent years. While a large amount of efforts have been invested in the field, the technology is still in its infancy. One of the major reasons for this stagnation is due to inability of the existing approaches to comprehensively take multiple kinds of contextual information into account. In the paper, we present a novel recommender system called Just-for-Me to facilitate effective social music recommendation by considering users' location related contexts as well as global music popularity trends. We also develop an unified recommendation model to integrate the contextual factors as well as music contents simultaneously. Furthermore, pseudo-observations are proposed to overcome the cold-start and sparsity problems. An extensive experimental study based on different test collections demonstrates that Just-for-Me system can significantly improve the recommendation performance at various geo-locations.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Query formulation, Search process; H.5.5 [**Sound and Music Computing**]: Systems

General Terms

Algorithms, Design, Experimentation, Human Factors

Keywords

Music Information Retrieval, Location-Aware, Recommendation, Empirical Study

ICMR'14, April 01-04, 2014, Glasgow, United Kingdom.

1. INTRODUCTION

The growing pervasiveness of online communities and increasing popularity of internet-accessing smart devices have significantly changed the way people consume and share music. As a result, developing intelligent techniques for music information retrieval (MIR) attracts a lot research interests from multimedia and information systems communities [28, 25, 27, 26]. In the recent years, location aware music recommendation system has been gaining in importance due to its wide range applications. With the technique, user's favorite songs can be automatically identified and retrieved based on where the user presents.

The key objective of location-aware music recommender system is to satisfy users' music information needs based on user's location with minimum user efforts on providing feedbacks. Apparently, accurate users' preferences acquisition is essential to the performance of a music recommender system. In general, personal music preferences can be influenced by both the global music popularity and the physical contexts related to where user presents. In the era of Web 2.0, the easy accessibility of the comments from domain experts, peers, and others via various user-generated content (UGC) channels (e.g., microblogging service) is exerting an increasingly impact on user's music preferences. In particular, how to effectively leverage the power of social media (e.g., online music popularity trend detection and analysis) becomes critical to achieve more accurate and reliable performance. On the other hand, smart mobile devices are becoming ubiquitous and people increasingly use the handheld device as a primary platform to access information. When on the move, the preference on music might be dynamically influenced by many factors related to physical environment, such as local social activities/events or geo-location [14]. For example, a user in the gym generally is keen on enjoying energetic music, while she/he prefers peaceful music in library. Also, online music popularity trend has significant influence on user's music taste and preference. The important observations suggest that effective contextual information integration in recommender systems can be very helpful to enhance system performance.

The existing techniques used in music recommender systems can be generally classified into three well-accepted categories: collaborative filtering-based (CF), content-based and hybrid-based [1]. They aim to provide song recommendation by modeling user's long-term music preference. Notwithstanding their great successes, the CF based techniques often suffer from a few problems (e.g., cold-start, sparsity and scalability) when being applied to real system development [22]. Another popular approach when designing music recommender systems is content-based filtering. Content-based filtering methods are based on content signature of music documents. Thus, effective music signature extraction is very important to content-based music recommendation system. While music content analysis has been an active research topic for decades [29, 33], the technologies are still in their infancy and the reported performance is rather poor. Hybrid-based approaches combine both techniques to overcome their own limitations [1, 6]. Recently, with the increasing ubiquity of smart phones, intelligent music recommendation based on where user present attracts more and more attentions from different research communities. Very surprisingly, no existing approach takes both location related context and global popularity trend into account.

Motivated by the above concerns, we develop Just-for-Me system to facilitate accurate location aware social music recommendation. Distinguished from the previous approaches in the domain of context-aware music recommendation, our approach can effectively combine local context and dynamics of global music popularity to facilitate more accurate and robust recommendation. Besides, pseudo-observations are proposed to overcome cold-start and sparsity problems which become much severer in context-aware recommender systems. To the best of our knowledge, no similar approach has been reported in the previous literature. To validate the performance of the proposed system, we conduct a comprehensive set of experimental studies with large music test collections. Experimental results demonstrate that incorporating track popularity can improve the performance significantly. At the same time, a comparative analysis demonstrates that our proposed framework achieves substantial improvement in recommendation accuracy and robustness at various venues (e.g., library and gym).

2. RELATED WORK

Traditional music recommender systems are developed based on three major techniques including collaborative filtering (CF), content-based techniques (CB) and hybrid approaches [1, 14]. The basic idea of CF based systems [22] is to estimate the similarity between users based on their listening history in the past and recommend the songs to a user via inferencing the preference of similar users. CB methods [18] compute the similarity between songs based on their musical contents or associated descriptive information and suggest the songs similar to the ones a targeted user liked in the past. As discussed in Section 1, both methods suffer from their own limitations. Hybrid methods overcome the limitations by combining both techniques [6]. A few hybrid music recommender systems have been proposed recently. One of the good examples is the system developed by Donaldson [9]. The system utilizes an item-based collaborative filtering based on song co-occurrence in playlists and acoustic features of music signals in recommendation. More recently, Yoshii et al. [32] develop an extended pLSA model called three-way aspect model [19] to associate the ratings and audio features of music tracks with a set of latent variables.

With the increasing popularity of smart devices, context-aware music recommender systems (CAMRS) have been received great attentions recently and became a very active research domain [14, 21]. Many existing CAMRS use either CF [15, 16, 20, 30] or CB methods [4, 5, 7, 10, 31], while very few hybrid methods have been developed. The CAMRS aims to satisfy user's local music needs by taking various contextual factors into account. One of the typical examples for recent development in this domain is the In-car music player developed by Baltrunas et al. [3]. The system can recommend music to users based on the landscapes passed. Kaminskas et al. [4, 5, 12, 13] conduct a series of studies on retrieving music tracks suited for place of interests (POI). Rho et al. [20] implement a CAMRS by associating users' contextual preferences with music emotions. Wang et al. [31] develop a mobile music recommender system for daily activities. More comprehensive review on CAMRS can be found in [14, 31]. While a lot of research efforts have been invested in this domain, most of the existing CAMRS only consider either physical environment-related (such as location and time) [3, 5, 15] or user-related contexts (such as activity and emotional state) [7, 20, 31]. To the best of our knowledge, no existing CAMRS take the influence of online music social trends on users' local music preferences into account.

3. JUST-FOR-ME SYSTEM

Just-for-Me system consists of four main components: (1) music popularity detection and analysis module; (2) user contextual listening history collection; (3) music content analysis module; and (4) unified recommendation model. The music popularity detection and analysis module aims to detect and analyze the dynamics of music popularity trends via mining Twitter streaming data. The popularity of music is then integrated into the unified recommendation model, which combines users' music tastes, music content as well as the influence of contextual factors by using a set of latent topics. The latent topics can be treated as the intrinsic factors to explain why users prefer certain piece of music in a particular location and during a particular time period.

3.1 Unified Recommendation Model

In real life, people's music preference and taste can be influenced by many different factors. In particular, Fig. 2 shows the relationship between the number of posts in Twitter[1] of 845 songs in a month and the number of listeners of the songs on Last.fm[2] during the same period. It is not hard to observe strong correlation between popularity trends and users' listening behaviors (the correlation coefficient is 0.52), which demonstrates the effects of music popularity trends (observed in Twitter). In turn, users' listening behaviors can affect the popularity trends of music. On the other hand, for a user, the selection of a song is highly related to his/her preferences on music content under current context. Indeed, the interplay of the music popularity, music content, and users' contextual music preferences is complex and cannot be effectively represented by using a simple combination of various factors, such as separately considering each factor and then combining them together. Thus, we develop a unified probabilistic generative model to model different aspects in a latent space based on a set of latent topics.

Our model is an extension of the three-way aspect model used in [32]. The three-way aspect model in [32] only captures user's long-term music preferences, but it does not consider user's short-term music needs. We extend it to in-

[1]http://www.twitter.com/
[2]http://www.youtube.com/

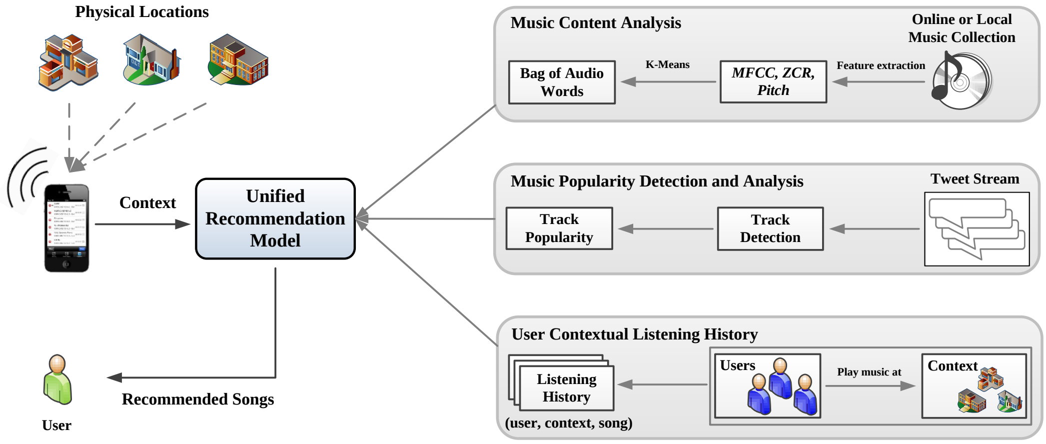


Figure 1: Architecture of the Just-for-Me music recommender system

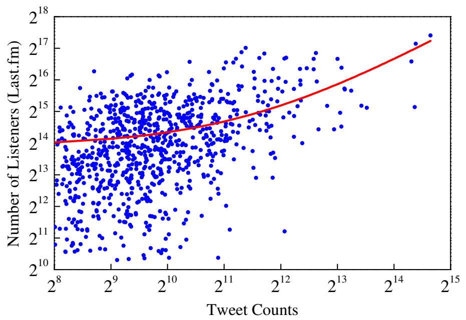


Figure 2: Number of listeners in Last.fm versus the number of tweets in Twitter of 845 tracks from 13 Dec. 2012 to 13 Jan. 2012. The red line is a fitted regression line.

corporate the location context and music popularity. Generally, users prefer different music contents in different places. In our model, the location context is integrated by including the location information into observation data. Accordingly, in our model, an observation is represented as a quadruple (u, l, s, w), which represents a user $u \in \mathcal{U}$ listening to a music track $s \in \mathcal{S}$ with the audio content $w \in \mathcal{W}$ at place $l \in \mathcal{L}$. $\mathcal{W}$ is the corpus of *"audio words"*, which are used to represent music content. The generation of audio words is introduced in Section 3.2. The more hours a user u listen to music content w at place l, the more likely the user prefers the music with content w at place l. In the model, a set of latent topics $\mathcal{Z}$ is used to associate music content with a user's music preferences under a certain location context. A user u listens to a music item s in a place l is considered to be related with a latent topic z. The graphical representation of the model is shown in Fig. 3. The generation process of an observation (u, l, s, w) is: a user u randomly selects a topic $z \in \mathcal{Z}$ according to his/her interest distribution, then z in turn "generates" the location l, music track s and audio word w based on their probabilistic distributions over z. The joint probabilistic distribution of user u, place l, music track s and audio word w can be expressed as,

$$Pr(u, l, s, w) = \sum_{z \in \mathcal{Z}} Pr(z)Pr(u|z)Pr(l|z)Pr(s|z)Pr(w|z) \tag{1}$$

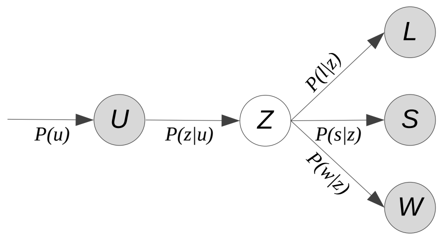


Figure 3: The extended three-way aspect model

For user u, the probability of choosing a music track s at place l is expressed as

$$Pr(s|u, l) = \frac{Pr(u, l, s)}{Pr(u, l)} = \frac{Pr(u, l, s)}{\sum_{s \in \mathcal{S}} Pr(u, l, s)} \tag{2}$$

The joint probability of $Pr(u, l, s)$ can be obtained by marginalizing w in $Pr(u, l, s, w)$, that is

$$Pr(u, l, s) = \sum_{w \in \mathcal{W}_s} Pr(u, l, s, w) \tag{3}$$

where $\mathcal{W}_s$ denotes audio words in the music track s. Combining Eq.1 - Eq.3, $Pr(s|u, l)$ can be computed. Music tracks that best fit the music needs of user u at place l are obtained based on the results.

Next we introduce how to incorporate the music popularity into the model. In general, users have higher chance to consume current popular music (based on the observation in Fig. 2). However, different users may have different degrees of preferences on different contents in current popular music. To integrate the effects of music popularity trends into the modeling of individual user's music preferences in different locations, our model associates the music popularity with the number of observations $n(u, l, s, w)$, which reflects the preferences of user u on music track s with content w in location l. Specifically, let $Pop(s)$ denote the popularity score of music track s,

$$n(u, l, s, w) = n(u, l, s) \times n(s, w) \times Pop(s) \tag{4}$$

where $n(u, l, s)$ is the number of times that the user u has listened to the music track s at location l; $n(s, w)$ is the frequency of audio word w in music track s; $Pop(s)$ is a real value and its computation will be described in Eq. 10. Using this method, the music popularity is associated with user's preferences on music contents in different locations. In the model, a user's personal preference on the music content of a popular music track (encoded in $n(u, l, s) \times n(s, w)$) are strengthened by the popularity of the track ($Pop(s)$). If a track contains more audio words appearing in tracks with higher popularity scores, it has higher chance to be user's preferred track. Thus, the model aims to *recommend tracks with popular content that fits users' music tastes.* In our system, the popularity scores of music tracks are computed once a week. Consequently, the model is updated accordingly so that it can closely track the dynamics of music popularity trends.

A set of model parameters need to be estimated and they include $Pr(z)$, $Pr(l|z)$, $Pr(u|z)$, $Pr(s|z)$ and $Pr(w|z)$. Assuming each observation (u, l, s, w) occurs independently, the log-likelihood of the observation data is

$$\mathcal{L} = \sum_{u,l,s,w} n(u, l, s, w) log(Pr(u, l, s, w)) \tag{5}$$

The model parameters are estimated by maximizing the log-likelihood. Our system applies *tempered EM* (TEM) [11] algorithm, a simulated-annealing-type variant of EM for avoiding overfitting and better generalization. The E-Step and M-Step are iterated alternately until $\mathcal{L}$ converges to a local maximum.

E-Step

$$Pr(z|u, l, s, w) = \frac{(Pr(z)Pr(u|z)Pr(l|z)Pr(s|z)Pr(w|z))^{\beta}}{\sum_{z'}(Pr(z')Pr(u|z')Pr(l|z')Pr(s|z')Pr(w|z'))^{\beta}}$$

M-Step

$$Pr(u|z) \propto \sum_{l,s,w} n(u, l, s, w)Pr(z|u, l, s, w) \tag{6}$$

$$Pr(l|z) \propto \sum_{u,s,w} n(u, l, s, w)Pr(z|u, l, s, w) \tag{7}$$

$$Pr(s|z) \propto \sum_{u,l,w} n(u, l, s, w)Pr(z|u, l, s, w) \tag{8}$$

$$Pr(w|z) \propto \sum_{u,s,l} n(u, l, s, w)Pr(z|u, l, s, w) \tag{9}$$

In TEM, the parameter β works as an *inverse computational temperature* in physical systems. β is set to 1 initially in TEM. When the performance on held-out data deteriorates, β decreases with ratio η by using $\beta = \beta \times \eta$.

3.2 Music Content Analysis

The audio contents of music tracks are represented by bag of audio words. Three different kinds of acoustic features are extracted to generate audio words, including MFCCs [17], zero-crossing rate (ZCR) and pitch. The MFCCs are the coefficients representing the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale frequency. The first 13 MFCCs is used in this study. Pitch is a major auditory attributes of musical tones. A pitch histogram is an array of 128 integer values indexed by MIDI note numbers, representing the frequency of occurrence of each note in a music item. The note which has the most frequent occurrence is considered here. ZCR is the number of zero-crossings of the waveform within a given frame.

To generate the audio words, each music track is segmented into a sequence of 0.5s frames. For each frame, MFCCs, ZCR and pitch of every 50ms sub-frame with 50% overlapping are extracted; and then the mean values of MFCCs, ZCR and pitch of all the sub-frames in the frame are separately normalized and concatenated to form its feature vector. K-mean clustering method is then used to group frames into clusters based on their feature vectors. The cluster centers are used as *audio words*. Replacing each frame with the nearest audio word, the music track is represented as a sequence of audio words.

3.3 Music Popularity Detection and Analysis

Twitter, as the most popular microblogging service, has millions of users who post what they are doing. By tracking the posts with music related information from the Twitter streaming data, the social popularity of music can be estimated. Schedl et al. [23, 24] have successfully used the Twitter streaming to estimate the spatio-temporal popularity of music artists. In the study, hasttag *#nowplaying* and *#np* are proved to be useful in determining music-related tweets. Both hashtags are used here to collected tweets by using Twitter Streaming API. Each collected tweet is then processed to check whether it contains music track information or not. Each track is represented as a tuple *(artist name, track title)*. By observing the tweets that contain track information, we found that in most cases, a tweet mentions only one track (e.g., *#Nowplaying Friday I'm in Love- The Cure*). Besides, music-related tweets often contain the following patterns:

1. *track title* by *artist name*;
2. *track title* from *artist name*;
3. *track title* - *artist name* or *artist name* - *track title*;
4. *track title* followed by *a hashtag of the artist.*

Because we have not identified all the hashtags of artists in our collection[3], only the first three types of patterns are used in the checking process. Tracks' popularity scores are calculated once a week. With the counted number of tweets for each track in a week, the popularity score of a track is computed as

$$Pop(s) = 1.0 + log(N(s) + 1.0) \tag{10}$$

where $N(s)$ is the number of tweets which mention the music track s in the week. Notice that some songs will have much larger $Pop(s)$ comparing with other songs, which may distort user's music preference by incorporating $Pop(s)$ using Eq. 4. For example, suppose a song s has a large value of $Pop(s)$, and a user only listened to it twice. The large value of $Pop(s)$ will still result in large $n(u, l, s, w)$, which may bias the music interests of the user. To balance the effect of music popularity scores, we scale the popularity scores of tracks into the range of 1 to 5 before using Eq. 4[4].

[3] In Twitter, different users may use different hashtags to

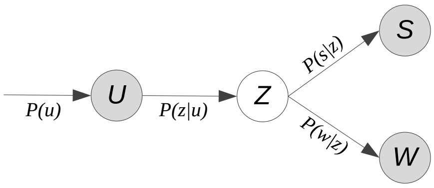


Figure 4: Three-way aspect model

4. EXPERIMENTAL CONFIGURATION

This section provides the details about our experimental configurations. Section 4.1 describes the data collections used in experiments. After that, Section 4.2 introduces the experimental methodology and evaluation metrics used.

4.1 Data Collections

To facilitate the experiments, we construct several datasets: a tweets collection, two music track collections (TC1 and TC2), and a user collection (UC).

4.1.1 Tweets Collections

We collected the tweets including hashtag *#nowplaying* or *#np* by using Twitter Streaming API from 17 Dec. 2012 to 13 Jan. 2013 (four weeks). The tweets are used to detect the popularity of tracks in a *track list*. To construct the track list, we collected artists from the top 100 artists in each week from 2008 to January 2013 in the category of *all place*[5] in Last.fm. Because the data from Last.fm is known to contain misspellings and mistakes, the list is checked by matching each artist name in an expert-based database - AllMusic[6]. After filtering, the list includes 244 artists. The tracks of each artist are collected from the MusicBrainz database[7] to form the track list. The popularity scores of tracks are computed based on the crawled Tweets using Eq. 10. For the top 1000 most popular tracks in each week according to the calculated popularity, the numbers of their listeners in the corresponding week in Last.fm are also collected.

4.1.2 Track Collection 1 (TC1)

TC1 consists of 1000 music tracks, which are used for music recommendation in experiments. To ensure the collection contains enough popular tracks as well as tracks with low popular scores, TC1 is constructed in the following way: firstly, based on the computed popularity scores using Eq. 10, the most popular 500 tracks in the week from 7 Jan. 2014 to 13 Jan. 2013 are selected; then 500 tracks are sampled from all the tracks of randomly selected 20 artists. The audio contents of these tracks are downloaded from YouTube.

4.1.3 Track Collection 2 (TC2)

TC2 is mainly used as a training dataset for music classification (Section 4.2.2). It contains 1028 tracks for 5 venues used in user study (Section 4.2.2): *office (252)*, *gym (196)*, *library (222)*, *canteen (177)*, *transportation (181)*[8]. Tracks in each category are extracted from the corresponding playlists in Grooveshark[9]. Grooveshark includes many playlists, which are created by users and titled with various contexts. The playlists have been successfully used for activity classification [31]. For each venue, we select the top playlists in the results returned by searching the venue name in Grooveshark[10], and then collect the tracks in the selected playlists. The corresponding audio tracks are downloaded from YouTube.

4.1.4 User Collection (UC)

To construct the user collection (UC), 10,188 recently active users in Last.fm[11] are randomly selected. Music listening records of the users are collected. Last.fm API provides methods to collect the tracks that users played in the past periods (e.g., *'last week'*, *'last month'*, *'last 3 months'*, *'last 6 months'* and *'last year'*) and how many times of these tracks were played in each period. The played tracks of users are ranked in descending order of played times. To collect enough records to capture users' recent music preference, the period of the *'last 3 months'* is a good choice. We collected the listened tracks of users in *'last 3 months'* and *'last week'* on each Monday during 17 Dec. 2012 to 14 Jan. 2013. For example, the collected tracks in *'last week'* from 17 Dec. 2012 to 14 Jan. 2013 are the tracks the users played in each week from 10 Dec. 2012 to 7 Jan. 2013. In experiments, the records of a user played a track for only once in a period are discarded, as playing a track only once cannot reflect the user's preference on the track.

4.2 Methodology & Evaluation Metrics

To analyze the system comprehensively, we first study the effects of music popularity trends on recommendation accuracy (Experiment 1). A user study is then conducted to evaluate the overall performance (Experiment 2). The following sections describe the experimental methodology and evaluation metrics of the two experiments. In both experiments, 20 topics are used in the aspect models, and a vocabulary including 500 audio words is generated[12].

4.2.1 Experiment 1

The effect of music popularity on recommendation is studied by comparing the performance of the three-way aspect model [32] (shown in Fig. 4) without and with the consideration of popularity, represented as USW and USW_P, respectively. In this experiment, TC1 and UC are used. The played tracks of users in *'last 3 months'* are used to predict their music choices in next week. For example, the played tracks in last 3 months before 17 Dec. 2012 are used to recommend music to users in the week from 17 Dec. 2012 to 24 Dec. 2012. We evaluate the recommendation performance in two disconnected weeks (as in successive weeks,

represent an artist.

[4]Several ranges, from 1 to {3,4,5,6,7}, were tested in the experiments described in Section 4.2.1. The range of 1 to 5 obtains the best performance among the tested ranges.

[5]http://www.last.fm/charts/artists/top/place/all?limit=100.

[6]http://www.allmusic.com/ (access: December, 2012).

[7]http://musicbrainz.org/ (access: December, 2012).

[8]Here, *canteen* refers to a type of food service location, and *transportation* refers to public transport, such as bus and train

[9]http://grooveshark.com/

[10]Access in December 2012, when there is only 79 songs for *canteen*, we selected other tracks 98 tracks in playlists of *cafeteria*.

[11]http://www.last.fm/community/users/active (access: December 2012).

[12]The number of topics and the number of audio words are empirically selected.

the played tracks of users could be very similar). Specifically, in the four weeks from 17 Dec. 2012 to 13 Jan. 2013, we use

1. the tracks in *'last 3 months'* of users before *week 1* as training data to recommend tracks for them in *week 1*;

2. the tracks in *'last 3 months'* of users before *week 3* as training data to recommend tracks for them in *week 3*.

The played tracks (ranked in the descending order by played times) of users in *week 1* and *week 3* are seen as the ground truth. To perform the evaluation (P@50 as shown in Section 4.2.3), only the users (in UC) who played at least 50 tracks (in TC1) in *week 1* and *week 3* are selected. Based on the criterion, 138 users are qualified in UC.

4.2.2 Experiment 2 (User Study)

To validate the effectiveness of Just-for-Me system, a user study is conducted to compare it with other two competitors on recommending music tracks at different venues.

1. **R1:** this system recommends tracks randomly. It simulates the cold start stage of recommender systems.

2. **R2:** this system uses a contextual post-filtering approach [2]. Music tracks are first ranked in descending order of relevance based on the results of the three-way aspect model with track popularity (USW_P). Then the most relevant tracks in each venue are recommended to users for the corresponding venue. Notice that tracks are classified into different venues before recommendation (described below).

For simplicity of presentation later on, we use **R3** to represent Just-for-Me system. Both **R2** and **R3** exploit collaborative filtering techniques, users with observations of $n(u,s)$ and $n(u,l,s)$ are needed to train them, respectively. Next, we describe the data used for the training of **R2** and **R3**.

Observations for R2. Because tracks in TC1 are used as test collection, users in UC with enough play records of tracks in TC1 are selected to train **R2**. Specifically, users (in UC) who played at least 10 tracks (in TC1) in *'the last 3 months'* (before 14 Jan. 2012) are used. The requirement of at least 10 played tracks is to better capture user's music preferences and exploit the power of collaborative filtering. Finally, records $n(u,s)$ of 277 users are selected.

Pseudo-observations for R3. Because the records $n(u,s)$ in UC are lack of location contexts, they cannot be directly used to train **R3**, which requires contextual observations $n(u,l,s)$. Indeed, it is a hard problem to obtain sufficient initial observations (of users or items) in most of context-aware recommender systems, due to the additional requirement of contexts to associate with the user-item observations. To solve this problem, we propose to use machine learning methods to infer the context for each user-item observation. Specifically, in our case, a track s is classified into a venue (location context) l by a trained classifier (described below), and then all the records associated with s (e.g., $n(u,s)$) are converted into observations $n(u,l,s)$ by relating $n(u,s)$ to the assigned venue l of s by the classifier. We call the inferred contextual observations as pseudo-observations.

In our implementation, the classifier is a multi-class SVM with RBF kernel. It is trained on TC2 (Section 4.1.3) by

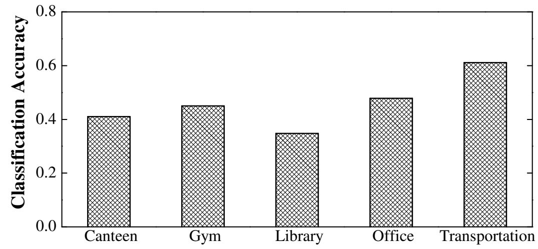


Figure 5: Music classification accuracy in each venue

using LIBSVM [8]. Five-fold cross-validation is adopted, and the parameters are tuned to obtain the best performance. Fig. 5 shows the classification accuracy of each venue. The average classification accuracy over five venues is 45.96%. The trained classifier is used to classify the music tracks in TC1.

Subject Data. 10 subjects with different culture backgrounds participated in the user study. They are 6 males and 4 females from several Asia countries, including China, India, Singapore and Vietnam. Subjects were asked to listen to 250 tracks selected from TC1 and label the tracks they like. The 250 tracks are comprised by randomly selected 50 tracks from each venue based on the classification results of TC1. This is to guarantee that each venue has roughly equal number of tracks. The remaining 750 tracks in TC1 are used to generate the recommended results for the subjects. The average number of labeled tracks of the 10 subjects is 32.5. The played times of their labeled tracks are set to 18, which is the average played times of $n(u,s)$ over all the selected users in UC. Based on the labeled tracks, we can obtained the observations $n(u,s)$ and pseudo-observations $n(u,l,s)$ of the subjects, which are used together with those of the selected users in UC to train **R2** and **R3**.

4.2.3 Evaluation Metrics

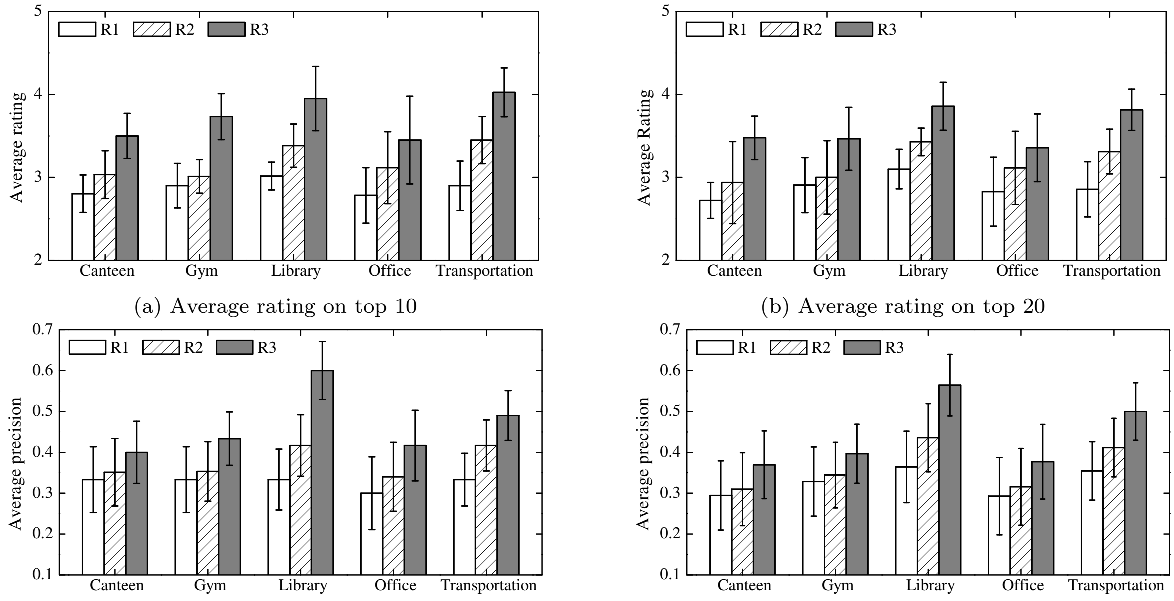
In Experiment 1, the average precisions of top recommended tracks (P@n, $n = \{10, 20, 30, 40, 50\}$) are used as evaluation metrics. Let $R_n(u_i)$ represent the track set with the top n recommended tracks for user u_i; $T_n(u_i)$ represent the most played n tracks of the user in *week 1* or *week 3* (i.e., the ground truth). $|U_n|$ is the number of users used in the experiment. The average precision at n is defined as

$$P@n = \frac{1}{|U_n|} \sum_{u_i \in U_n} \frac{R_n(u_i) \cap T_n(u_i)}{n} \quad (11)$$

In Experiment 2, human subjects were required to rate the recommended tracks in a 5-point Likert scale from "strong dislike" (1 point) to "strong like" (5 point). The higher rating a subject gives to a song, the subject has stronger interest in the song. The average ratings of recommended tracks at top 10 and 20 for each venue are used as evaluation metrics. The average precision is also computed by regarding the track with rating higher than 3 as a positive recommendation.

5. RESULTS AND ANALYSIS

This section first presents the observed correlation between music popularity trends and music listening behaviors in Section 5.1, and then analyzes the effects of track popularity on music recommendation based on the results



(a) Average rating on top 10

(b) Average rating on top 20

(c) Average precision of top 10

(d) Average precision on top 20

Figure 6: Performance comparison of three recommendation schemes in Experiment 2 (user study)

of Experiment 1 in Section 5.2. Finally, the comparative results of Just-for-Me system with the two competitors in Experiment 2 are presented in Section 5.3.

5.1 Effects of Music Popularity Trends

The music-related posts in Twitter provide a lot of information about the global popularity trends of music, and the numbers of listeners of music tracks disclose general users' listening behaviors. By associating the computed popularity of music tracks on Twitter with the corresponding numbers of listeners in Last.fm, the effect of music popularity trends on users' listening behaviors can be analyzed. With the collected Twitter streams, the popularity scores of tracks are computed in each week. The numbers of listeners of the most popular 1000 tracks in Twitter are crawled from Last.fm. Fig. 7 shows the numbers of listeners versus the popularity scores of the top 1000 popular music tracks in the four weeks from 17 Dec. 2012 to 13 Jan. 2013. Strong correlation can be observed between the popularity scores and the numbers of listeners of tracks in the same week. The average correlation coefficient achieves 0.425. The results demonstrate that the global music popularity trends exert important influence on user listening behaviors.

5.2 Effects of Track Popularity

The performance comparison between the three-way aspect model without (USW) and with popularity (USW_P) is shown in Table 1. From the table, we can see that with the consideration of music popularity in the model, the recommendation accuracy is improved consistently. In USW_P, the co-occurrence of triple (u, s, w) is scaled by the popularity score of the track s to increase the probability of recommending more popular tracks. This is in accord with the fact that users generally tend to listen to current hot music tracks. Moreover, the improvements on P@10 and P@20 are more significant. It is a nice property as users are usually more interested in the top results in recommendation. The

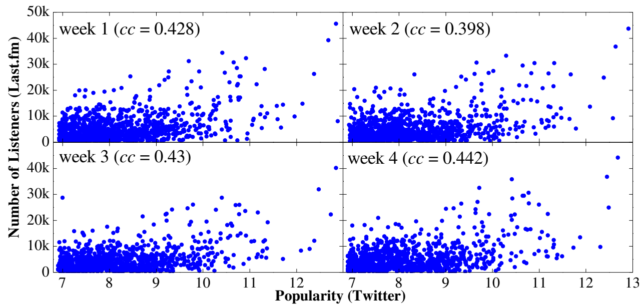


Figure 7: Track popularity using Twitter stream versus the corresponding number of listeners in Last.fm of 1000 music tracks from 17 Dec. 2012 to 13 Jan. 2013 (cc refers to correlation coefficient)

results demonstrate that the incorporation of music popularity can improve the performance of music recommendation.

	Week 1		Week 3	
	USW	USW_P	USW	USW_P
P@10	0.235	**0.301**	0.302	**0.375**
P@20	0.345	**0.413**	0.395	**0.463**
P@30	0.644	**0.684**	0.481	**0.512**
P@40	0.638	**0.675**	0.578	**0.603**
P@50	0.740	**0.780**	0.647	**0.680**

Table 1: The average recommendation precisions over all users by using USW and USW_P in Experiment 1

5.3 Performances of Competitors

Each recommender system (**R1**, **R2** and **R3**) generates a playlist for a subject at a certain venue (u, l). For each (u, l), the top 20 tracks recommended by each system are collected and mixed together to form a single playlist S_{ul}. Thus, subjects do not know that each track is recommended by which recommender. Each subject u was required to listen to each track in S_{ul} at the corresponding venue l. Each track must be played at least one minute before rating (see Section 4.2.3). Fig. 6 shows the average and standard deviation of the ratings and precisions for the top 10 and top 20 recommended tracks in different venues by three recommender systems. We can observe that **R2** performs significantly better than **R1** in most cases, and **R3** greatly outperforms the other two models in all cases. The results verify the effectiveness of integrating the location context and track popularity into recommendation. Notice that the performance of **R3** is obtained based on pseudo-observations. Thus, we can expect that with more real observations, the performance of **R3** will be much better. Besides, the results also demonstrate that pseudo-observations can be used to deal with the cold-start problem in context-aware recommender systems (CARS). Furthermore, comparing with traditional recommender systems, the requirement of context information in observations makes the sparsity problem severer in CARS. It can be expected the proposed pseudo-observations can also be used to alleviate the sparsity problems in CARS.

6. CONCLUSION

In this paper, we present a novel location-aware music recommendation system called Just-for-Me, which can effectively integrate music content, location related context and music popularity trends into a unified probabilistic generative model. To support music popularity detection, we develop a novel method to analyze Twitter streaming data. We also demonstrate a strong correlation between music popularity trends and user listening behaviors, with the computed popularities of tracks and their played information in Last.fm. An experimental study validates the effects of the consideration of music popularity in recommendation. The recommendation performance can be greatly improved by applying the strategies to incorporate music popularity into music recommender systems. Further, a set of user studies have been conducted to verify the performance of the Just-for-Me system. The results show that with effective integration of various contextual factors, Just-for-Me achieves a significant performance improvement at various geo-locations.

7. ACKNOWLEDGEMENTS

Jialie Shen and Zhiyong Cheng are partially supported by Mobile plus Cloud Computing Theme Research Program Award - Microsoft Research.

8. REFERENCES

[1] G. Adomavicius et al. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE TKDE*, 17:743–749, 2005.
[2] G. Adomavicius et al. Context-aware recommender systems. In *Recommender systems handbook*, pages 217–253. Springer, 2011.
[3] L. Baltrunas et al. Incarmusic: Context-aware music recommendations in a car. In *EC-Web*, 2011.
[4] M. Braunhofer et al. Recommending music for places of interest in a mobile travel guide. In *ACM RecSys*, 2011.
[5] M. Braunhofer et al. Location-aware music recommendation. *IJMIR*, 2(1):31–44, 2013.
[6] R. Burke. Hybrid web recommender systems. In *The adaptive web*, pages 377–408. Springer, 2007.
[7] R. Cai et al. Musicsense: Contextual music recommendation using emotional allocation modeling. In *ACM MM*, 2007.
[8] C. Chang et al. LIBSVM: A library for support vector machines. *ACM TIST*, 2, 2011.
[9] J. Donaldson. A hybrid social-acoustic recommendation system for popular music. In *ACM RecSys*, 2007.
[10] S. Dornbush et al. Xpod: A human activity aware learning mobile music player. In *Proc. Workshop on Ambient Intelligence, IJCAI*, 2007.
[11] T. Hofmann. Probabilistic latent semantic analysis. In *UAI*, 1999.
[12] M. Kaminskas et al. Knowledge-based music retrieval for places of interest. In *MIRUM*, pages 19–24. ACM, 2012.
[13] M. Kaminskas et al. Location-aware music recommendation using auto-tagging and hybrid matching. In *ACM RecSys*, 2013.
[14] M. Kaminskas et al. Contextual music information retrieval and recommendation: state of the art and challenges. *Computer Science Review*, 6(2-3):89–119, 2012.
[15] J. Lee et al. Context awareness by case-based reasoning in a music recommendation system. In *Ubiquitous Computing Systems*, pages 45–58. 2007.
[16] A. Lehtiniemi. Evaluating supermusic: streaming context-aware mobile music service. In *ACM ACE*, 2008.
[17] B. Logan et al. Mel frequency cepstral coefficients for music modeling. In *ISMIR*, 2000.
[18] M. Pazzani et al. Content-based recommendation systems. In *The adaptive web*, pages 325–341. Springer, 2007.
[19] A. Popescul et al. Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments. In *UAI*, 2001.
[20] S. Rho et al. Svr-based music mood classification and context-based music recommendation. In *ACM MM*, 2009.
[21] F. Ricci. Context-aware music recommender systems: workshop keynote abstract. In *ACM WWW*, 2012.
[22] J. Schafer et al. Collaborative filtering recommender systems. In *The adaptive web*, pages 291–324. 2007.
[23] M. Schedl. Analyzing the potential of microblogs for spatio-temporal popularity estimation of music artists. In *Proceedings of the IJCAI*. Citeseer, 2011.
[24] M. Schedl et al. What's hot? estimating country specific artist popularity. In *ISMIR*, 2010.
[25] M. Schedl et al. Harvesting microblogs for contextual music similarity estimation: a co-occurrence-based framework. *Multimedia Systems*, pages 1–13, 2013.
[26] M. Schedl et al. Multimedia information retrieval: music and audio. In *ACM MM*, 2013.
[27] M. Schedl et al. Location-aware music artist recommendation. In *MMM*, 2014.
[28] J. Shen et al. Towards effective content-based music retrieval with multiple acoustic feature combination. *IEEE TMM*, 8(6):1179–1189, 2006.
[29] J. Shen et al. Modeling concept dynamics for large scale music search. In *ACM SIGIR*, 2012.
[30] J.-H. Su et al. Music recommendation using content and context information mining. *IEEE Intelligent Systems*, 25:16–26, 2010.
[31] X. Wang et al. Context-aware mobile music recommendation for daily activities. In *ACM MM*, 2012.
[32] K. Yoshii et al. An efficient hybrid music recommender system using an incrementally trainable probabilistic generative model. *IEEE TASLP*, 16(2):435–447, 2008.
[33] B. Zhang et al. Compositemap: a novel music similarity measure for personalized multimodal music search. In *ACM MM*, 2009.