Exploring User-Specific Information in Music Retrieval

Zhiyong Cheng
National University of Singapore
jason.zy.cheng@gmail.com

Jialie Shen*
Northumbria University, UK
jerry.shen@northumbria.ac.uk

Liqiang Nie
Shandong University, China
nieliqiang@gmail.com

Tat-Seng Chua
National University of Singapore
chuats@comp.nus.edu.sg

Mohan Kankanhalli
National University of Singapore
mohan@comp.nus.edu.sg

ABSTRACT

With the advancement of mobile computing technology and cloud-based streaming music service, user-centered music retrieval has become increasingly important. User-specific information has a fundamental impact on personal music preferences and interests. However, existing research pays little attention to the modeling and integration of user-specific information in music retrieval algorithms/models to facilitate music search. In this paper, we propose a novel model, named User-Information-Aware Music Interest Topic (UIA-MIT) model. The model is able to effectively capture the influence of user-specific information on music preferences, and further associate users' music preferences and search terms under the same latent space. Based on this model, a user information aware retrieval system is developed, which can search and re-rank the results based on age- and/or gender-specific music preferences. A comprehensive experimental study demonstrates that our methods can significantly improve the search accuracy over existing text-based music retrieval methods.

CCS CONCEPTS

•**Information systems** → **Information retrieval; Music retrieval;** Retrieval models and ranking;

KEYWORDS

Semantic music retrieval, user demographic information, re-ranking, topic model

---

*Jialie Shen is corresponding author

1 INTRODUCTION

With the rapid development of mobile computing technology and cloud-based streaming music service, smart devices have become the most popular platforms to consume music daily. Based on Nelsen's Music 360 2015 report[1] , 44% of US music listeners use smartphones to listen to music on a daily basis. Smartphones are typically designed for personal use and it is easy to obtain personal information from smartphones, which can be used in personalized applications to achieve better user experiences when developing personalized applications. With the rapidly growing trend in music consumption with smartphones, there has been an increasing interest in studying user-centered music retrieval [11, 25, 33, 34, 36]. Techniques to support effective user-centered music search are gaining its importance due to its wide range of potential applications [7, 25, 36]. Based on this technology, intelligent music search or recommendation systems can be developed to automatically cater to users' personal music needs.

Text-based music retrieval, as one of the most popular music search paradigms, typically requires users to provide a few text keywords as queries to describe their music information needs [20, 24, 27, 41]. Most previous research efforts in music retrieval have been devoted to the development of retrieval/recommendation algorithms [5, 11, 21, 27], effective music representations [23, 24, 31, 42], similarity measurement [6, 12, 38, 45, 48], and automatic music annotation [1, 3, 13, 26, 28, 37, 40, 41, 46]. However, the effects of user-specific information in music retrieval has not been well studied [25, 33, 34]. User-specific information or user background, such as *age, gender, social status, growing-up environment, culture background, etc.,* has great impact on users' long-term music interests. This hypothesis has received strong support in prior research studies [8, 22, 43]. Due to its great impact, given the same query, users with different backgrounds might expect different search results [11]. For example, given a query "*pop, sad*", the retrieved songs expected by a 40-year-old male could be very different from those favored by a 20-year-old female.

In this study, we develop a text-based music retrieval system, which can effectively leverage user-specific information to improve the search performance. The key challenges on effectively integrating user-specific information in music retrieval include: (1) how to model the influence of user-specific

---

information on music preferences; and (2) how to associate the influence with search queries and songs. To tackle these two challenges, a novel topic model named User Information Aware Music Interest Topic (UIA-MIT) is proposed. UIA-MIT can explicitly model the music preferences of different types of user-specific information. In this model, the music preferences affected by different factors (i.e., age and gender) are represented by probabilistic distributions on a set of latent topics. These latent topics are in turn probabilistic distributions of songs and terms (song's annotations or tags). Therefore, songs, terms, and music preferences (influenced by age and gender) can be associated by latent topics. Based on the UIA-MIT model, we develop a probabilistic text-based music retrieval method, which can effectively exploit user information to improve the search results. In our context, *user-specific information* refers to the *user related information* or *user backgrounds* which have been proven to have great impact on a user's long-term music interests, such as *age*, *gender*, and *country* [8, 22, 43].

In order to evaluate the performance of our proposed method, extensive experiments and comprehensive comparisons over different methods have been conducted on two retrieval tasks: ad-hoc search and re-ranking. The experimental results demonstrate that users' age and/or gender information play an important role on search performance improvement, which indicates the importance of utilizing user-specific information in real music retrieval system development. To the best of our knowledge, our work is the first attempt on designing advanced music retrieval methods to leverage user-specific information in retrieval. In summary, the main contributions of this work include:

- This is the first attempt to explore the incorporation of user-specific information in music retrieval algorithms for search accuracy improvement.
- We propose a UIA-MIT model, which can explicitly model the music preferences affected by different types of user information. Furthermore, based on the model, a text-based music retrieval method is developed to effectively utilize user-specific information for improving the search accuracy.
- We construct two test collections and empirically evaluate the performance of the proposed retrieval method and compare it with a set of baselines. The experimental results demonstrate significant performance improvement.

The remainder of this paper is organized as follows: Section 2 gives a brief overview of related work; Section 3 detailedly describes the proposed UIA-MIT model and retrieval method; Section 4 introduces the experimental datasets and configurations; and Section 5 reports the experimental results and main findings. Finally, Section 6 concludes the paper.

2 RELATED WORKS

In this following, we review the research directions which are closely related to this paper, including text-based music retrieval, the use of user information in music retrieval and recommender systems, and related topic models.

2.1 Text-based Music Retrieval

As one of the most popular music search paradigms, text-based retrieval system is built on the top of mature text retrieval techniques and typically requires users to provide a few text keywords as queries to describe their music information needs [20, 24, 27, 41]. Its search performance heavily relies on the meta-information (e.g., artist and title) and well-defined categorized information (e.g., genre and instrument). In many cases, users would also like to describe their current contexts, such as *emotions* and *occasions*, with the expectation that the music search engine return a playlist with suitable songs [19]. To support the search of such semantic queries, one needs to annotate songs with a rich vocabulary of music terms. However, human annotation is very expensive in terms of both time and labor, and thus it is unlikely to scale with the growth in the amount of recorded songs. To deal with this problem, many automated methods have been proposed to automatically annotate songs with music related concepts by learning the correlation between music acoustic contents and the semantic terms based on a well-annotated music collection. Most automated systems generate a vector of tag weights when annotating a new song for music retrieval [26]. An early work in this direction is described in [41]. They formalized the audio annotation and retrieval as a supervised multi-class labeling task. The dataset CAL500 they created in this study became a standard test collection for subsequent works [1, 13, 26, 28, 42, 44].

With the rapid development of social music websites, songs are annotated with user-contributed social tags, which provides an alternative way to navigate and search songs (e.g., Last.fm). Social tags have no constraints on the use of text and provide a rich vocabulary, which covers most terms used to describe songs. Extensive research efforts have been devoted to developing tag-based music search systems [24]. However, the user-provided tags are known to be noisy, incomplete and subjective [24], which limits the search performance of tag-based methods. Consequently, many works consider the combination of tags and acoustic similarity to improve the search performance [21, 24, 27].

2.2 User Information in Music Retrieval

In recent years, researchers have realized and advocated the importance of incorporating information about user into music information search and discovery [25, 33, 34, 36]. Previous work has shown that user information can be used to improve music recommendation [8, 43]. In [29], user demographic information was taken into a fuzzy system for context inference using Bayesian network. In [15], user information was used to infer user's mood, which is then used to match with the mood of songs predicted based on music content. [35] studied the influence of user information (e.g., age, gender, and country) on the task of artist recommendation. Although the use of user information in recommender systems is widespread, little effort has been devoted to the exploration of user information in music retrieval. Furthermore, previous studies have not explicitly modeled the music preferences of different ages

and genders in recommendation. In this work, we propose a user-information aware music interests discovery model to capture the music preferences of different ages and genders, and develop a music retrieval framework to use age and/or gender information in music retrieval.

2.3 Topic Models in Music Retrieval

Topic models, such as pLSA [17] and LDA [4], were originally proposed to discover the underlying *themes* or *latent topics* for a large scale collection of text documents. In these models, latent topics are discovered by mining co-occurrence patterns of words in documents that exhibit similar patterns. By treating *users* as *documents* and *items* as *words*, topic models have been applied to discover user's latent interests [18, 30]. In the domain of music information retrieval and recommendation, several previous studies adapted the topic models to capture user's music interests [9, 11, 16, 47]. [47] used the three-way aspect model to discover user music interests in recommendation. [9] extended the three-away aspect model to incorporate both location context and global music trends for location-aware personalized music recommendation. User's local music preference is captured by the co-occurrence of songs in users' music profiles and the co-occurrence of music contents among songs. In [10], a location-aware topic model was proposed to discover the music preferences of different venue types. [16] proposed a variant of LDA to discover user's interests via the combination of co-occurrence of songs in the same user's playlist and co-occurrence of tags in the same song. In [11], a dual-layer music preference topic model was developed to characterize the correlation among users, songs, and search terms for personalized music retrieval.

3 OUR APPROACH

In general, users with similar demographics have more similar music interests than users with different demographics. For example, users in the same age or gender have more similar music interests [2, 22]. To model the influence of such user-specific factors, we propose a User Information Aware Music Interest Topic (UIA-MIT) model. In this model, a set of latent topics (i.e., K topics) are discovered based on the records of users' favorite tracks. Each latent topic represents one type/style of music or a *music interest dimension*. As users' music interests are influenced by different factors. UIA-MIT is designed to capture the influence of different factors on music interests. For example, what are the *general music interests* of users in *a certain age range or gender*; or in other words, the *likelihood* of each type of songs preferred by the users with regard to their ages and genders. In this model, a user's latent music interest is expressed as a mixture of multiple latent topic distributions. Each latent topic distribution represents a certain music interest which is dependent on a user-specific factor (e.g., age). Therefore, the mixture of multiple latent topic distributions in this model represents the music interests of a user, as the result of collective effects of different factors.

Table 1: Notations and their definitions

Notation	Definition
u, s	user, and song, respectively
c, a, g	country, age, and gender category
s_v, s_w	song, audio word, and text word
D	corpus, $(u, a, g, s, \boldsymbol{s_w}, \boldsymbol{s_v}) \in D$
$|D_u|$	number of songs in the u's music profile
$\mathcal{U}, \mathcal{S}$	user set and song set in the corpus
U, S	total number of users and songs in the corpus
$\mathcal{W}, \mathcal{V}$	text and audio word vocabulary, respectively
W, V	text and audio vocabulary size, respecitvely
$\mathcal{A}, \mathcal{G}$	age category set and gender category set
A, G	number of age and gender groups, respectively
K	total number of latent topics
y	indicator variable: decide z generated from θ_u, θ_a, or θ_g
$\boldsymbol{\lambda}$	mixing weight vector
$\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$	Dirichlet priors
$\boldsymbol{\theta_u}, \boldsymbol{\theta_a}, \boldsymbol{\theta_g}$	music preference of user u, age group a, and gender group g, respectively
$\boldsymbol{\phi_s}, \boldsymbol{\phi_v}, \boldsymbol{\phi_w}$	multinomial distributions over songs $\mathcal{S}$, audio words $\mathcal{V}$ and text words $\mathcal{W}$

3.1 Preliminaries

For ease of understanding and presentation, we first introduce some key concepts and notations. Table 1 lists the notations used in this paper.

Dataset The dataset D in our model consists of a set of records, with each record comprising a user, user information (i.e., age and gender), song, and song's content (i.e., tags and audio words), that is, $(u, a, g, s, \boldsymbol{s_w}, \boldsymbol{s_v}) \in D$, where $u \in \mathcal{U}$, $s \in \mathcal{S}$, $a \in \mathcal{A}$, $g \in \mathcal{G}$, $s_w \in \mathcal{W}$, and $s_v \in \mathcal{V}$. [2] One piece of record in the dataset is a user u of age a and gender g who loves a song s with tags $\boldsymbol{s_w}$ and audio content $\boldsymbol{s_v}$: $D_{usc} = \{(u, a, g, s, w, v) : w \in \boldsymbol{s_w}, v \in \boldsymbol{s_v}\}$.

Audio Word An audio word is a representative short frame (e.g., 0.5s in our implementation) of audio stream in a music corpus [11]. Audio words are used to represent the audio content of a song as a "bag-of-audio-words" document.

Latent Topic A latent topic z, or *topic* for short, in a song collection $\mathcal{S}$ is a probabilistic distribution over songs, i.e., $\{P(s|\boldsymbol{\phi_s}) : s \in \mathcal{S}\}$. Similarly, a topic in a text word corpus $\mathcal{W}$ is a probabilistic distribution over text words, i.e., $\{P(w|\boldsymbol{\phi_w}) : w \in \mathcal{W}\}$. A topic in an audio word corpus $\mathcal{V}$ is a probabilistic distribution over audio words, i.e., $\{P(v|\boldsymbol{\phi_v}) : v \in \mathcal{V}\}$.

User's Music Interest UIA-MIT models users' music interest as the mixture of three latent topic distributions (see Eq. 1): (1) $\boldsymbol{\theta_u}$: the music preferences as a collective result based on the influence of all other factors (e.g., personality) besides age and gender , (2) $\boldsymbol{\theta_a}$: age-based music preference, denoting *the music preferences of a certain age range a* or *the general music preferences of users in age range a*, and (3) $\boldsymbol{\theta_g}$:

---

[2] In the paper, unless otherwise specified, notations in bold style denote matrices or vectors, and the ones in normal style denote scalars.
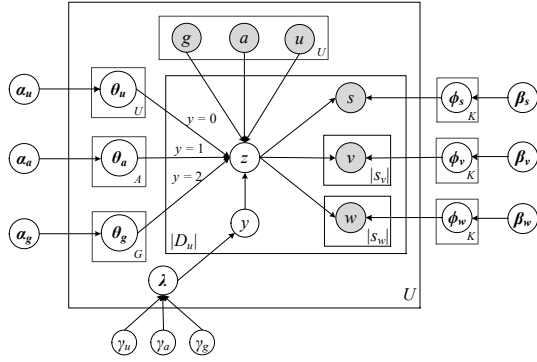
Figure 1: The graphical representation of the UIA-MIT model.

gender-based music preference, denoting *music preferences of male or female.*

3.2 UIA-MIT Model

3.2.1 Model Description. The graphical representation of the model is shown in Fig. 1, in which *age* and *gender* are considered. UIA-MIT explicitly models the music preferences of ages (θ_a) and genders (θ_g). The music preference, as a result of all other factors (excluding age and gender, such as user's personality and country), is modeled as a single probabilistic distribution of latent topics, denoted as user's personal music interest (θ_u). Notice that the UIA-MIT model can be easily extended to model the music preference of other individual factors (e.g., *country*).

From the generation perspective, the model mimics the music selection process by considering the user's music interest, age-based music preference, and gender-based music preference in a unified manner. Given a user with age a and gender g, the likelihood the user u selecting a music track is dependent on the music preferences of user age and gender as well as his/her personal music interest:

$$P(s|u,a,g,\theta_u,\theta_a,\theta_g,\phi_w,\phi_v,\phi_s) = \lambda_u P(s|u,\theta_u,\phi_w,\phi_v,\phi_s)$$
$$+ \lambda_a P(s|a,\theta_a,\phi_w,\phi_v,\phi_s) + \lambda_g P(s|g,\theta_g,\phi_w,\phi_v,\phi_s) \quad (1)$$

where $P(s|u,\theta_u,\phi_w,\phi_v,\phi_s)$ is the probability that song s is generated according to the personal music interest of user u, denoted as θ_u; $P(s|a,\theta_a,\phi_w,\phi_v,\phi_s)$ and $P(s|g,\theta_g,\phi_w,\phi_v,\phi_s)$ denote the probability that song s is generated according to the age music preference of a and gender music preference of g, denoted as θ_a and θ_g respectively. $\lambda = \{\lambda_u,\lambda_a,\lambda_g : \lambda_u + \lambda_a + \lambda_g = 1\}$ is a categorial distribution, which controls the selection motivation of song s. That is, when selecting song s, it is possible that user u selects it according to his/her own music interests θ_u with probability λ_u, or according to the age-based music preference θ_a with probability λ_a, or according to the gender-based music preference θ_g with probability λ_g. Note that λ is a group-dependent parameter, as users in different groups have different tendency to select music from different aspects. For example, from the

training results, female users are more likely to select music tracks according to the general music preferences (namely, mainstreaming music) than male users.

The generation process of UIA-MIT is shown in Algorithm 1. Intuitively, UIA-MIT models user's music interests as the combination of the general music preferences according to certain user-specific information (age and gender here) and user's distinct music interests (affecting by user's personality, etc.). The general music preferences of certain user-specific information can be applied in music-related service.

Collapsed Gibbs sampling [14] is used to estimate the parameters in the topic model. Due to space limitation, we omit the description of parameter estimation in the paper.

Algorithm 1: Generation Process of UIA-MIT

---

1 **for** *each topic $k = 1,...,K$* **do**
2   Draw $\phi_{k,s} \sim Dir(\cdot|\beta_s)$;
3   Draw $\phi_{k,w} \sim Dir(\cdot|\beta_w)$;
4   Draw $\phi_{k,v} \sim Dir(\cdot|\beta_v)$;
5 **for** *each user $u \in \mathcal{U}$* **do**
6   Draw $\theta_u \sim Dir(\cdot|\alpha_u)$;
7 **for** *each age range $a \in \mathcal{A}$* **do**
8   Draw $\theta_a \sim Dir(\cdot|\alpha_a)$;
9 **for** *each gender $g \in \mathcal{G}$* **do**
10   Draw $\theta_g \sim Dir(\cdot|\alpha_g)$;
11 **for** *each user $u \in \mathcal{U}$ with age $a \in \mathcal{A}$ and gender $g \in \mathcal{G}$* **do**
12   **for** *each song $s \in D_u$* **do**
13   Toss a coin according to categorical distribution
        $y_s \sim Dir(\gamma_u,\gamma_a,\gamma_g)$;
14   **if** $y_s == 0$ **then**
15   Draw $z_s \sim Multi(\theta_u)$ according to the music interest of user u;
16   **if** $y_s == 1$ **then**
17   Draw $z_s \sim Multi(\theta_a)$ according to the music preference of age a;
18   **if** $y_s == 2$ **then**
19   Draw $z_s \sim Multi(\theta_g)$ according to the music preference of gender g;
20   After the sampling of the topic $z_s = k$, draw song $s \sim Multi(\theta_{k,s})$;
21   **for** *each word $w \in s_w$* **do**
22   Draw $w \sim Multi(\phi_{k,w})$;
23   **for** *each audio word $v \in s_v$* **do**
24   Draw $v \sim Multi(\phi_{k,v})$;

3.3 Retrieval Method

Given a query q, UIA-MIT can be used to estimate $P(s|q)$ with the consideration of user's age and gender. In music retrieval, songs are then ranked in the descending order of $P(s|q)$ and the top results are returned to the user. Specifically, for a query $q = \{w_1,w_2..,w_n\}$ of user u with age a and gender g, the conditional probability $P(s|q)$ could be computed with estimated parameters $\Theta = \{\theta_u,\theta_a,\theta_g\}$ and $\Phi = \{\phi_s,\phi_w\}$. For the simplicity of presentation, we use $P(s|q,\cdot)$ to denote $P(s|q,u,a,g,\Theta,\Phi)$ in the following.

$$P(s|q,\cdot) = \prod_{i=1}^{n} P(w_i|s,u,a,g,\Theta,\Phi)P(s|u,a,g,\Theta,\phi_s) \quad (2)$$

where query terms in q are assumed to be independent. $P(s|u, a, g, \Theta, \phi_s)$ is computed as:

$$P(s|u, a, g, \Theta, \phi_s) = \sum_{k=1}^{K} P(s|z = k, \phi_s)P(z = k|u, a, g, \Theta)$$

$$= \sum_{k=1}^{K} \phi_{k,s} \cdot (\lambda_u \theta_{u,k} + \lambda_a \theta_{a,k} + \lambda_g \theta_{g,k})$$

(3)

According to Bayes rule and the graphical representation of the UIA-MIT model, $P(w_i|s, u, a, g, \Theta, \Phi)$ is estimated as:

$$P(w_i|s, \cdot) = \sum_{k=1}^{K} P(w_i|z = k, \phi_w) \frac{P(s, z_k|u, a, g, \Theta, \phi_s)}{P(s|u, a, g, \Theta, \phi_s)}$$

$$= \sum_{k=1}^{K} \phi_{k,w_i} \cdot \frac{\phi_{k,s} \cdot (\lambda_u \theta_{u,k} + \lambda_a \theta_{a,k} + \lambda_g \theta_{g,k})}{\sum_{k=1}^{K} \phi_{k,s} \cdot (\lambda_u \theta_{u,k} + \lambda_a \theta_{a,k} + \lambda_g \theta_{g,k})}$$

(4)

Based on Eq. 3 and Eq. 4, Eq. 2 becomes

$$P(s|q, \cdot) = \prod_{i=1}^{n} \sum_{k=1}^{K} \phi_{k,w_i} \cdot \phi_{k,s} \cdot (\lambda_u \theta_{u,k} + \lambda_a \theta_{a,k} + \lambda_g \theta_{g,k}) \quad (5)$$

Typically, when θ_u of a particular user u is known, Eq. 5 can be used for personalized music search. However, for a new user to the system, θ_u is unknown, while his/her age/gender information is relatively easier to be found. In such cases, we normalize λ_a and λ_g to $\lambda_a + \lambda_g = 1$, and the following equation is used for retrieval:

$$P(s|q, a, g, \Theta, \Phi) = \prod_{i=1}^{n} \sum_{k=1}^{K} \phi_{k,w_i} \cdot \phi_{k,s} \cdot (\lambda_a \theta_{a,k} + \lambda_g \theta_{g,k}) \quad (6)$$

If either age or gender information is available, only the corresponding music preferences will be used (namely, set $\lambda_a = 1$ or $\lambda_g = 1$ in the equation). Intuitively, $\phi_{k,w_i} \cdot \phi_{k,s}$ evaluates the similarity of the song s with respect to query q in the music dimension k in the music interest space, and $\lambda_a \theta_{a,k} + \lambda_g \theta_{g,k}$ estimates the music preferences with respect to age range a and gender g in the music dimension k. Thus it can be seen as that the model re-weights the original query in different music dimensions based on user's age and gender information. Given a new users with only age and/or gender information, our system can be used to search songs based on Eq. 6. Accordingly, the exploitation of age and gender (i.e., Eq. 6) can alleviate the cold-start problem in personalized music retrieval, in which user's music preference is unknown.

4 EXPERIMENTAL CONFIGURATION

We conduct a comprehensive experimental study to investigate the performance of our methods on two music retrieval tasks: *ad-hoc* and *re-rank*. The experiments mainly answer the following research questions:

RQ1 What is the performance of UIA-MIT based retrieval methods on ad-hoc search with the use of age and/or gender information as compared to other text-based music retrieval methods? (See Sect. 5.1.1)

RQ2 Whether the use of user-specific information (i.e., age and gender information in our study) for re-ranking can improve the music search performance? If so, how much can be improved by using age and/or gender information? (See Sect. 5.1.2 and Sect. 5.2)

RQ3 Whether the UIA-MIT model can be extended to model the music preference of other user information, such as country? (See Sect. 5.2)

RQ4 What are the impact of different types of user information on user's music preference, such as age, gender and country information? (See Sect. 5.3)

To answer these questions, we constructed two test collections to conduct experiments. In our experiments, test collections are split into a training set and a testing set. The training set is used to train the UIA-MIT model. User-specific information and users' favorite songs are used in the training stage. In the testing stage, only user-specific information is used in retrieval. The information of favorite songs are only used in the result evaluation. It is similar to the cold-start problem in personalized search, where user's personal preferences are unknown.

4.1 Datasets

To evaluate the search accuracy of retrieval systems with respect to query users, a great challenge is how to obtain the ground truth of the test queries with respect to the corresponding query users. In our retrieval task, given a query q of a user with certain information, a relevant song should not only be relevant to the query but also loved by the users with such information. We developed two test collections by crawling user information from Last.fm. Thousands of queries and corresponding ground truth are generated for testing. The dataset will be released for the repeatability of the experiments and other related studies.[3]

User Profile Dataset We construct a dataset with users' demographic information and their favorite music tracks from Last.fm. The dataset is collected in the following procedures. 160 recent active users were randomly selected from Last.fm[4]. The friends of these users and the friends of their friends were also collected with their demographic information, including age, gender, country. In total, 90,036 users were collected. Users who provide both age and gender information were retained. As the number of users with age under 16 or above 54 years old is small, we removed these users and only focused on studying the influence of ages between 16 to 54. Finally, the dataset remains 45,334 users. Users' favorite tracks were collected using Last.fm public API "User.getLovedTracks", and a 30-second audio stream of each song was downloaded from 7digital[5]. By removing users with less than 10 favorite songs and songs preferred by less than 10 users, there are 29,412 users (15,826 males and 13,586 females) and 15,323 songs in the final dataset. The social tags of these songs were

---

[3]Experimental datasets are accessible in:https://www.dropbox.com/sh/eue9it0lqlpzo7q/AAAe-v2MS0KYln5qSpSgPQQna?dl=0.
[4]Accessed http://www.last.fm/community/users/active on Mar 3, 2015.
[5]https://www.7digital.com/

collected from Last.fm using API "Track.getTopTags". The social tags of songs from Last.fm are used in the topic model training and the tag-based music retrieval method (see the TAG method in Section 4.2). We denote the dataset as D for ease of presentation.

Retrieval Test Collection 1 (TC1) To judge the relevance of songs with respect to queries, it is necessary to label the songs in the dataset with query concepts. CAL10K [39] is a labeled song collection. The annotations are used as ground truth in pervious text-based music retrieval research [27]. This dataset contains 10,870 songs from 4,597 different artists. The label vocabulary is composed of 137 "genre" tags and 416 "acoustic" tags. The number of tags of songs varies from 2 to 25 tags. The song tags are mined from the Pandora Web sites. The annotations in Pandora are contributed by music experts and are considered highly objective [39]. 2,839 songs in D are contained in the CAL10K. These songs are used as retrieval dataset in TC1. In experiments, we categorized the users into 7 age groups, as shown in Table 2. Thus, there are 14 user groups (7 age groups × 2 gender groups) in total.

Table 2: Number of users in each age group in Test Collection 1.

Group	0	1	2	3	4	5	6
Age	16-20	21-25	26-30	31-35	36-40	41-45	46-54
#Users	9,003	12,820	4,941	1,482	595	324	247

Retrieval Test Collection 2 (TC2) Due to the limited size of the well-labeled songs, TC1 contains only 2,839 songs in the retrieval stage, which is relatively small. TC1 can be used to to test the performance of the proposed retrieval methods by leveraging age and gender information. To examine the performance of our methods in large datasets and demonstrate the extendability of UIA-MIT to other user information ("*country*" here), we constructed TC2. TC2 uses social tags as annotations in relevance judgement. 26,468 users in D have age, gender, and country information. These users are from 179 different countries. With the age, gender, and country information, we categorize users into groups based on {age, gender, country}, e.g., 16-20_male_US. The top 30 user groups with the most number of users are used in experiments to examine the performance of our methods. These groups are shown in Table 3. By removing users with less than 10 favorite songs and songs liked by less than 10 users, TC2 has 14,715 users and 1, 0197 songs.

Query Set In experiments, we use a combination of k distinct terms as queries. Following the methodology in [27, 41], queries composed by $k = \{1, 2, 3\}$ terms are used. The method described in [27] is used to construct the query set. In TC1, all the terms in CAL10K dataset are treated as 1-term query candidates; and for 2-term and 3-term queries, all the term combinations are considered as candidates. In TC2, social tags are used to generate the queries. We first filtered the tags which appear less than 10 times in the dataset, and removed the tags which express personal interests in the song, such as "favorite", "great", "favor", "excellent", etc. After

Table 3: Number of users in different groups in Test Collection 2.

Country	Male			Female		
	16-20	21-25	26-30	16-20	21-25	26-30
Brazil	1605	1320	253	1665	809	156
Poland	179	270	83	472	401	98
Russia	83	269	109	184	314	92
UK	186	416	223	198	364	120
US	539	1360	684	457	1273	533

tokenizing the tags of each song into terms, all the terms are treated as candidates. For 2-term and 3-term queries, all the term combinations which appear in a song are treated as candidates. Next, query candidates with less than 10 relevant songs in the ground truth of all user groups were removed. For the 3-term query in TC2, we retain a random sample of 3,000 queries as [27]. Table 4 summarizes the number of queries in TC1 and TC2. Notice that the queries are the same for each group. Table 5 shows some query examples.

Table 4: Number of queries in TC1 and TC2.

Test Collection	# 1-Term	# 2-Term	# 3-Term
TC1	33	122	542
TC2	79	1691	3000

Table 5: Few examples for each type of queries.

1-Term Query	2-Term Query	3-Term Query
aggressive	aggressive, guitar	aggressive, angry, guitar
angry	aggressive, rock	angry, guitar, rock
breathy	bass, tonality	drums, angry, guitar
country	blues, guitar	guitar, aggressive, angry
danceable	country, guitar	guitar, pop, romantic

Ground Truth As the query is evaluated with respect to different user groups (i.e., age and gender groups), a relevant song with respect to a query should (1) contains all the query terms in the annotations (in the CAL10K dataset for TC1) or social tags (for TC2); and (2) be loved by at least 10 users in this user group. The second criterion is to guarantee that the relevant song is loved by users in a user group (i.e., with a certain age and gender). Based on the criteria, the relevant songs in the retrieval datasets of TC1 and TC2 are labeled. Notice that for each query, the numbers of relevant songs in different groups are different.

4.2 Experimental Setup

In our experiments, users were split into two sets for two-fold cross-validation: one set (users with their favorite tracks) is used for model training, and the other set is used to create the query set and generate the corresponding ground truth. The dataset is split in the way to guarantee each set has

approximately equal number of users in the two sets. In the result presentation, the presented results are the average performance of the two sets. For the training of UIA-MIT, the corpus is formulated into three types of documents:

User-Song Document It represents a user's music profile. For each user, a user-song document is generated based on his/her favorite songs. The document is the concatenation of all the songs preferred by the users.

Song-Text Word Document This document represents the semantic content of a song. Songs' tags from Last.fm are used to form their text documents. In our implementation, tags that appeared less than 10 songs are filtered. The remaining tags of a song are concatenated and tokenized with a standard stop-list to form its text document.

Song-Audio Word Document This document represents the audio content of a song, namely, the audio words used in the UIA-MIT model. The audio contents of one song are represented by "bag-of-audio-words" document. An audio word is a representative short frame of audio stream (e.g., 0.5 second) in a music corpus. For each song, a 30-second audio track downloaded from 7digital is used to generate its "bag-of-audio-word" document. We follow the method described in [11] to generate the song-audio word documents.

4.2.1 Baselines.

- **TAG** In this method, the social tags of each song in Last.fm are used as the text description for retrieval. The standard tf-idf weighting scheme is used to compute the similarity between query and songs with the standard cosine distance in the Vector Space Model [32].
- **WLC** The first result returned by TAG is used as the seed for a content-based music retrieval (CBMR) method. Then the score of the TAG method and CBMR method are linearly combined together to generate the final search results. This method is implemented as described in [11].
- **PAR** This method is proposed in [21]. It incorporates audio similarity into an already existing ranking. In our experiments, the results of tag-based method (TAG) are used as the initial ranking list. In the implementation, we followed the details reported in the referred paper.
- **GBR** This method is proposed in [27]. It is a graph-based ranking method, which combines both tag and acoustic similarity in a probabilistic graph-based representation for music retrieval. In our implementation, we followed the details reported in the referred paper.
- **Music Popularity Based Re-ranking (MPR)** This method re-ranks the top (e.g., 100) songs returned by other retrieval methods according to the popularity of these songs in each user group. The popularity score is computed as:

$$POP(s) = \frac{N(s,a,g)}{N(a,g)} \qquad (7)$$

where $N(s,a,g)$ is the number of users in group (a,g) favoring song s, and $N(a,g)$ is the total number of the users in group (a,g).

- **Group User Music Representation (GUMR)** For each group, we aggregate the social tags of songs loved

by the users in this group to form a music representation document for the group. For a given query q of user u in age a and gender g, the similarity score is computed as:

$$Sim(q,s,a,g) = w \cdot TAG(q,s) + (1-w) \cdot Sim(s,a,g) \qquad (8)$$

in which $TAG(q,s)$ is the cosine similarity between the song s and q using TAG method, and $Sim(s,a,g)$ is the similarity between user group preference and the song s using cosine distance with the standard tf-idf weighting scheme (based on song's text document and the group's music document). The combination weight is tuned in experiments.

The following methods are the variants of the proposed method, which simulate the scenarios when partial user information is available, and study the improvements of using such information individually or together in retrieval.

- **A-MIT**: only considering age information in UIA-MIT;
- **G-MIT**: only considering gender information in UIA-MIT;
- **C-MIT**: only considering country information in UIA-MIT (only tested in TC2);
- **AG-MIT**: considering both age and gender information in UIA-MIT.
- **AGC-MIT**: considering age, gender, and country information in UIA-MIT (only used in TC 2).

To the best of our knowledge, we have not found any music search methods using such information in retrieval. MPR and GUMR are two heuristic methods on utilizing age and gender information. Notice that if these methods can also improve the search accuracy, it further demonstrates the importance of considering user-specific information in music retrieval. In the above methods, PAR and MPR are re-ranking methods and thus are compared in the re-ranking task (Sect. 5.2). Other methods are compared in the ad-hoc search task (Sect. 5.1).

4.2.2 Metrics.
Precision at k (**P@k**) and Mean Average Precision (**MAP**) are used as evaluation metrics. As the top search results are more important, we report P@10 and MAP@10 in experimental results.

4.2.3 Parameter Setting.
In our implementation, the hyperparameters in the topic model are turned in a wide range of values. In the UIA-MIT model, without prior knowledge about the topic distributions of users in different ages and genders, we set $\boldsymbol{\alpha_u}$, $\boldsymbol{\alpha_a}$ and $\boldsymbol{\alpha_g}$ to be symmetric. For simplicity, we set them to be the same and tune them in the range of $\alpha = \alpha_u = \alpha_a = \alpha_g \in \{0.01, 0.05, 0.1, 1.0, 5.0\}$. Similarly, β_w, β_v and β_s are also set to be symmetric and tuned in similar manners: $\beta = \beta_w = \beta_v = \beta_s \in \{0.01, 0.05, 0.10, 0.15, 0.20, 0.25\}$. The values of γ_u, γ_a and γ_g bias the tendency of choosing music according to user's personal, age or gender music preferences. We would like the tendency to be learned from the data, thus γ_u, γ_a, γ_g are all set to 1. In Gibbs sampling for the training of topic models, 100 sampling iterations were run as burn-in iterations and then 50 sampling iterations with a gap of 10 were taken to obtain the final results. In

Table 6: Comparison of retrieval performance on TC1

Method	1-Term Query		2-Term Query		3-Term Query	
	P@10	MAP	P@10	MAP	P@10	MAP
TAG	.164	.134	.054	.041	.022	.014
WLC	.178	.142	.058	.051	.022	.023
GBR	.248	.267	.136	.147	.122	.125
GUMR	.133	.115	.030	.026	.015	.012
G-MIT	.276	.283	.139	.155	.116	.134
A-MIT	.250	.259	.135	.151	.111	.128
AG-MIT	**.339***	**.335***	**.177***	**.184***	**.149***	**.166***

the result presentation in Sect. 5, the reported results are based on the parameters with the best results.

5 EXPERIMENTAL RESULTS

In all the reported results, the symbol (*) after a numeric value denotes significant differences ($p < 0.05$, a two-tailed paired t-test) with the corresponding second best measurement. In our experiments, for each user group (i.e., male users between 16-20 years old), all the 1-term, 2-term and 3-term queries are used for retrieval and evaluation. All the results presented are the average values over all user groups in each test collection.

5.1 Performance on TC1

5.1.1 Retrieval Performance. Retrieval results of the proposed methods using age and/or gender information with the baselines are reported in Table 6. From the table, we observe that: First, for the three types of queries, it is obvious that queries with more terms is more difficult for all methods. As can be seen, the proposed method using age and gender information (AG-MIT) outperforms all the other methods over all types of queries. Besides, A-MIT and G-MIT methods obtain comparable or better performance over the GBR method, which shows the best performance besides our proposed methods. The results demonstrate the effectiveness of the proposed retrieval methods.

Second, compared to the TAG method which only uses text information in retrieval, WLC and GBR that use both text and acoustic features obtain better performance. WLC uses a linearly combination of similarities based on TAG and acoustic features. It can only slightly improve the search performance. Notice that the WLC uses the first search results of TAG as acoustic query, the search accuracy of TAG thus affects the improvement of the WLC method. GBR method explore both text and acoustic information by discovering and using the intrinsic correlation between the semantics of terms and the acoustic contents. It obtain much better results than the TAG method.

Third, it can be seen that the G-MIT and A-MIT methods can improve the search performance over the GBR method for 1-term query. The AG-MIT method can further improve

the performance for 1-term, 2-term, and 3-term queries. The effects of using age or gender information in retrieval are comparable, and the gender information seems to be slightly more effective than age information. The performance of the AG-MIT method is obviously better than the G-MIT and A-MIT methods, indicating that the use of both age and gender information together is more effective than using them individually. The results of GUMR are quite poor, because of the simple method on modeling user's music preferences with age and gender information.

5.1.2 Re-ranking Performance. This section presents the re-ranking performance based on the top 100 results of different retrieval methods (i.e., TAG, WLC, and GBR). The results are reported in Table 7. In the table, the row starting with "-" shows the performance obtained by the corresponding baseline methods. Overall, the results are improved greatly and significantly by the re-ranking methods, even for 2- and 3-term queries whose initial results are very poor. An interesting finding is that the initial results of the TAG method is worse than the WLC method, however, the TAG method can obtain much better re-ranking results than the WLC method by all re-ranking methods. The results indicate that the WLC method can obtain better search results in top positions (e.g., top 10 results), while it reduces the number of relevant results in a longer list (i.e. top 100 results).

The effectiveness of the re-ranking methods based on the proposed models can be observed by comparing our methods (A-MIT, G-MIT, or AG-MIT) with the PAR method. The improvements of our methods are much greater than that of the PAR method. Notice that the UIA-MIT model explores the relevance between queries (semantic concepts or tags) and songs in a latent music interest space, which is discovered based on the music preferences of a large number of listeners. In other words, the method leverages the collaborative knowledge of crowds to estimate the relevance between query concepts and songs , or the music preference of general users on songs with respect to query concepts. The external knowledge is complementary to the information used by the TAG, WLG, or GBR methods, which compute the relevance between the query and song only based on the contents. Consequently, using the estimated relevance between query and song based on the UIA-MIT model for re-ranking can significantly improve the search performance.

The benefits of utilizing user information in music retrieval can be well demonstrated by the MPR method, which can improve the search performance greatly using a heuristic method - re-ranking the songs according to their popularity in user groups. For the 1-term query on $P@10$, the relative improvement by MPR achieves more than 133% and 43% over the TAG and GBR methods, respectively. The improvement over 2-term and 3-term queries is even larger. G-MIT and A-MIT methods obtain much better results than the MPR method. The AG-MIT method can further improve the performance. Notice that improvements achieved by A-MIT, G-MIT and AG-MIT are contributed by both the learned associations between query and songs and the

Table 7: Comparison of re-ranking performance on TC1

| Method | 1-Term Query | | | | | | 2-Term Query | | | | | | 3-Term Query | | | | | |
| | TAG | | WLC | | GBR | | TAG | | WLC | | GBR | | TAG | | WLC | | GBR | |
	P@10	MAP	P@10	MAP	P@10	MAP	P@10	MAP	P@10	MAP	P@10	MAP	P@10	MAP	P@10	MAP	P@10	MAP
-	.164	.134	.178	.142	.248	.267	.054	.041	.058	.051	.136	.147	.022	.014	.022	.023	.112	.125
PAR	.233	.249	.086	.074	.330	.309	.099	.116	.036	.035	.197	.199	.068	.078	.022	.023	.182	.175
MPR	.383	.409	.185	.148	.355	.347	.237	.257	.063	.069	.241	.257	.184	.204	.039	.042	.206	.223
G-MIT	.460	.470	.197	.223	.439	.450	.260	.278	.069	.079	.263	.268	.191	.209	.048	**.053**	.215	.230
A-MIT	.444	.470	.195	.221	.431	.434	.254	.277	.077	.086	.273	.285	.192	.211	.045	.046	.215	.233
AG-MIT	**.479***	**.495**	**.228***	**.242**	**.444***	**.463***	**.267**	**.294***	**.087***	**.091***	**.285***	**.287**	**.200**	**.225***	**.053**	.048	**.226**	**.246***

Table 8: Comparison of re-ranking performance on TC2 based on the results of TAG

| Method | 1-Term Query | | 2-Term Query | | 3-Term Query | |
	P@10	MAP	P@10	MAP	P@10	MAP
TAG	.122	.120	.042	.020	.022	.017
MPR	.216	.277	.135	.138	.103	.106
A-MIT	.240	.318	.146	.182	.103	.106
G-MIT	.228	.304	.135	.158	.104	.108
C-MIT	.244	.318	.144	.181	.104	.107
AG-MIT	.252	.330	.144	.176	.104	.108
AGC-MIT	**.375***	**.505***	**.160**	**.199***	**.105**	**.110**

Table 9: Mean values of λ in the UIA-MIT model

| Test | Male Groups | | | Female Groups | | |
Collection	λ_a	λ_g	λ_c	λ_a	λ_g	λ_c
TC1	.119	.170	-	.149	.217	-
TC2	.089	.137	.183	.143	.179	.185

captured age and gender music preferences in UIA-MIT. The relative improvement of the AG-MIT for the 1-term query on $P@10$ achieves more than 192% and 79% over the TAG and GBR methods, respectively. UIA-MIT captures the age music preference and gender music preference together and achieves consistent and better improvement over A-MIT and G-MIT. It shows that the influence of age and gender are correlated in affecting users' music interests. Thus, it is not optimal to use age-based music preference and gender-based music preference individually.

5.2 Performance on TC2

Similar results can be observed in TC2 for both ad-hoc and re-ranking tasks. Due to space limitation, we only present the performance of re-ranking performance based on TAG, because from the experimental results on TC1: (1) using our method in re-ranking can greatly improve the search results; and (2) re-ranking performance based on TAG are comparable or better than the performance based on WLC and GBR. Table 8 presents the re-ranking results of different methods on TC2. The second row shows the search results of TAG method, and 3 - 9 rows show the re-ranking results of different methods. From the table, it can be observed that user's country information (C-MIT) can also be used to improve the performance. AGC-MIT obtains the best performance, demonstrating that the UIA-MIT model can be easily extended to include other types of user information and the utilization of more types of user information can obtain better performance.

5.3 Impact of Different User Information

In the UIA-MIT model, λ_u, λ_a, and λ_g control the contributions of users' unique music interests and general music interests of age and gender in music selection. Thus, the values of these parameters show the relative importance of different types of user information in users' music preferences. Table 9 shows the mean values of λ_a, λ_g, and λ_c of male and female groups in two test collections. λ_c is a similar parameter as λ_a, controlling the weight of the general music interest of users in different countries. Notice that UIA-MIT in TC1 has not considered country information. The values of those three parameters vary greatly across different user groups, indicating that those three factors are inter-correlated to affect users' music preferences. Overall, the differences of three parameters between male and female groups are more obviously (comparing to different age or country groups). In the table, we show the values of three parameters in male groups and female groups separately. Some interesting observations can be found on the global level: (1) users' personal music interests ($\boldsymbol{\theta_u}$) dominate the music selection, as $\lambda_u > 0.5$[6]; (2) the value of λ_c is greater than those of λ_g and λ_a, indicating that the effects of country on music selection is larger than age and gender; (2) the value of λ_g is slightly larger than that of λ_a, indicating that gender has more impact on the music preferences than age.

6 CONCLUSION

In this paper, we proposed a novel User-Information-Aware Music Interest Topic (UIA-MIT) model to discover the latent music interest space of general users and capture the music preferences of users in different ages and genders. Based on the proposed model, a music retrieval method is developed for text-based music retrieval, which can effectively incorporate user information to improve the search results. Extensive experiments were conducted to demonstrate the effectiveness of exploiting user's age and gender information

---

[6]$\lambda_u = 1 - \lambda_a - \lambda_g$ in TC1 or $\lambda_u = 1 - \lambda_a - \lambda_g - \lambda_c$ in TC2

in music retrieval. The results demonstrate the importance and potential of utilizing user-specific information in music retrieval systems. We hope this work can shed light on the direction of developing user-centric music retrieval systems and motivate more research efforts in this area.

ACKNOWLEDGMENTS

REFERENCES

[1] L. Barrington, M. Yazdani, D. Turnbull, and G. RG Lanckriet. 2008. Combining feature kernels for semantic music retrieval. In *ISMIR*.

[2] P. Berkers. 2012. Gendered scrobbling: Listening behaviour of young adults on Last.fm. *Interactions: Studies in Communication & Culture* 2, 3 (2012), 279–296.

[3] T. Bertin-Mahieux, D. Eck, F. Maillet, and P. Lamere. 2008. Autotagger: A model for predicting social tags from acoustic features on large music databases. *Journal of New Music Research* 37, 2 (2008), 115–135.

[4] D. Blei, A. Ng, and M. Jordan. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.* 3 (2003), 993–1022.

[5] J. Bu, S. Tan, C. Chen, C. Wang, H. Wu, L. Zhang, and X. He. 2010. Music recommendation by unified hypergraph: combining social media information and music content. In *ACM MM*.

[6] M. Casey, C. Rhodes, and M. Slaney. 2008. Analysis of minimum distances in high-dimensional musical spaces. *IEEE Trans. Audio, Speech, Language Process.* 16, 5 (2008), 1015–1028.

[7] M. A Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney. 2008. Content-based music information retrieval: Current directions and future challenges. *Proc. IEEE* 96, 4 (2008), 668–696.

[8] Òscar Celma. 2010. Music Recommendation. In *Music Recommendation and Discovery*, Òscar Celma (Ed.). Springer, Chapter 3, 43–86.

[9] Z. Cheng and J. Shen. 2014. Just-for-me: An adaptive personalization system for location-aware social music recommendation. In *ACM ICMR*.

[10] Z. Cheng and J. Shen. 2016. On effective location-aware music recommendation. *ACM Trans. Inf. Syst.* 34, 2 (2016), 13.

[11] Z. Cheng, J. Shen, and S. CH Hoi. 2016. On effective personalized music retrieval by exploring online user behaviors. In *ACM SIGIR*.

[12] J S. Downie. 2008. The music information retrieval evaluation exchange (2005-2007): A window into music information retrieval research. *Acoustical Science and Technology* 29, 4 (2008), 247–255.

[13] K. Ellis, E. Coviello, A. B. Chan, and G. RG Lanckriet. 2013. A bag of systems representation for music auto-tagging. *IEEE Trans. Audio, Speech, Language Process.* 21, 12 (2013), 2554–2569.

[14] T. L Griffiths and M. Steyvers. 2004. Finding scientific topics. *PNAS* 101, Suppl 1 (2004), 5228–5235.

[15] B.-J. Han, S. Rho, S. Jun, and E. Hwang. 2010. Music emotion classification and context-based music recommendation. *Multimed. Tools Appl.* 47, 3 (2010), 433–460.

[16] N. Hariri, B. Mobasher, and R. Burke. 2013. Personalized text-based music retrieval. In *Workshops at AAAI Conference on Artificial Intelligence*.

[17] T. Hofmann. 1999. Probabilistic latent semantic indexing. In *ACM SIGIR*.

[18] T. Hofmann and J. Puzicha. 1999. Latent class models for collaborative filtering. In *IJCAI*.

[19] J.-Y. Kim and N. J Belkin. 2002. Categories of music description and search terms and phrases used by non-music experts. In *ISMIR*.

[20] P. Knees, T. Pohle, M. Schedl, D. Schnitzer, and K. Seyerlehner. 2008. A document-centered approach to a natural language music search engine. In *ECIR*.

[21] P. Knees, T. Pohle, M. Schedl, D. Schnitzer, K. Seyerlehner, and G. Widmer. 2009. Augmenting text-based music retrieval with audio similarity. In *ISMIR*.

[22] A. LeBlanc, Y. Jin, L. Stamou, and J. McCrary. 1999. Effect of age, country, and gender on music listening preferences. *Bull. Counc. Res. Music Educ.* 141 (1999), 72–76.

[23] M. Levy and M. Sandler. 2008. Learning latent semantic models for music from social tags. *J. New Music Res.* 37, 2 (2008), 137–150.

[24] M. Levy and M. Sandler. 2009. Music information retrieval using social tags and audio. *IEEE Trans. Multimed.* 11, 3 (2009), 383–395.

[25] C. Liem, M. Müller, D. Eck, G. Tzanetakis, and A. Hanjalic. 2011. The need for music information retrieval with user-centered and multimodal strategies. In *ACM workshop on Music Information Retrieval with User-centered and Multimodal Strategies*.

[26] R. Miotto and G. Lanckriet. 2012. A generative context model for semantic music annotation and retrieval. *IEEE Trans. Audio, Speech, and Language Process.* 20, 4 (2012), 1096–1108.

[27] R. Miotto and N. Orio. 2012. A probabilistic model to combine tags and acoustic similarity for music retrieval. *ACM Trans. Inf. Syst.* 30, 2 (2012), 8.

[28] J. Nam, J. Herrera, M. Slaney, and J. O. Smith. 2012. Learning sparse feature representations for music annotation and retrieval. In *ISMIR*.

[29] H.-S. Park, J.-O. Yoo, and S.-B. Cho. 2006. A context-aware music recommendation system using fuzzy bayesian networks with utility theory. In *FSKD*.

[30] A. Popescul, D. Pennock, and S. Lawrence. 2001. Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments. In *UAI*.

[31] M. Riley, E. Heinen, and J. Ghosh. 2008. A text retrieval approach to content-based audio retrieval. In *ACM SIGIR*.

[32] G. Salton, A. Wong, and C.-S. Yang. 1975. A vector space model for automatic indexing. *Commun. ACM* 18, 11 (1975), 613–620.

[33] M. Schedl and A. Flexer. 2012. Putting the user in the center of music information retrieval. In *ISMIR*.

[34] M. Schedl, A. Flexer, and J. Urbano. 2013. The neglected user in music information retrieval research. *J. Intell. Inf. Syst.* 41, 3 (2013), 523–539.

[35] M. Schedl, D. Hauger, K. Farrahi, and M. Tkalčič. 2015. On the influence ofuser characteristics on music recommendation algorithms. In *ECIR*.

[36] M. Schedl, S. Stober, E. Gómez, N. Orio, and C. CS Liem. 2012. User-aware music retrieval. *Multimod. Music Process.* 3 (2012), 135–156.

[37] J. Shen, W. Meng, S. Yan, H. Pang, and X. Hua. 2010. Effective music tagging through advanced statistical modeling. In *ACM SIGIR*.

[38] J. Shen, H. Pang, M. Wang, and S. Yan. 2012. Modeling concept dynamics for large scale music search. In *ACM SIGIR*.

[39] D. Tingle, Y. E Kim, and D. Turnbull. 2010. Exploring automatic music annotation with acoustically-objective tags. In *ISMIR*.

[40] D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet. 2008. Semantic annotation and retrieval of music and sound effects. *IEEE Trans. Audio, Speech, Language Process.* 16, 2 (2008), 467–476.

[41] D. Turnbull, L. Barrington, D. Torres, and G. RG Lanckriet. 2007. Towards musical query-by-semantic-description using the cal500 data set. In *ACM SIGIR*.

[42] D. R Turnbull, L. Barrington, G. Lanckriet, and M. Yazdani. 2009. Combining audio content and social context for semantic music discovery. In *ACM SIGIR*.

[43] A. L Uitdenbogerd and R. Schyndel. 2002. A review of factors affecting music recommender success. In *ISMIR*.

[44] J.-C. Wang, Y.-C. Shih, M.-S. Wu, H.-M. Wang, and S.-K. Jeng. 2011. Colorizing tags in tag cloud: a novel query-by-tag music search system. In *ACM MM*.

[45] M. Wang, W. Fu, S. Hao, D. Tao, and X. Wu. 2016. Scalable semi-supervised learning by efficient anchor graph regularization. *IEEE Trans. Knowledge Data Eng.* 28, 7 (2016), 1864–1877.

[46] M. Wang, X. Liu, and X. Wu. 2015. Visual Classification by ell_1-Hypergraph Modeling. *IEEE Trans. Knowledge Data Eng.* 27, 9 (2015), 2564–2574.

[47] K. Yoshii, M. Goto, K. Komatani, T. Ogata, and H. G Okuno. 2008. An efficient hybrid music recommender system using an incrementally trainable probabilistic generative model. *IEEE Trans. Audio, Speech, Language Process.* 16, 2 (2008), 435–447.

[48] B. Zhang, J. Shen, Q. Xiang, and Y. Wang. 2009. CompositeMap: a novel framework for music similarity measure. In *ACM SIGIR*.