Information Processing and Management 53 (2017) 1185–1200

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Utilizing context-relevant keywords extracted from a large collection of user-generated documents for music discovery

Ziwon Hyung[a], Joon-Sang Park[b], Kyogu Lee[a,*]

[a] Music and Audio Research Group, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Korea
[b] Department of Computer Engineering, Hongik University, 94 Wausan-ro, Mapo-gu, Seoul 121-791, Korea

ARTICLE INFO

Article history:
Received 16 October 2016
Revised 26 February 2017
Accepted 17 April 2017
Available online 29 May 2017

Keywords:
Context-relevant keywords
Song-document association
Keyword extraction
Music descriptors
Music retrieval

ABSTRACT

The contextual background of a user is one of the important criteria when deciding what music to listen to. In this paper, we propose a novel method to embed the user context for music search and retrieval. The proposed system extracts keywords from a large collection of documents written by users. Each of these documents contains a personal story about the writer's situation and/or mood, followed by a song request. We consider that there is a strong correlation between the story and the song. Therefore, by extracting keywords from these documents, it is possible to develop a list of terms that can generally be used to describe the user context when requesting a song, which may then be employed to represent a music item in a richer manner. Once each song is represented using the proposed context-relevant music descriptors, we perform Latent Dirichlet Allocation to retrieve similar music based on context similarity. By conducting a series of experiments, we identified a correlation between the proposed music descriptors and conventional approaches, such as acoustic features or lyrics. The identified correlation can be used to auto-tag songs with no document association. We also qualitatively evaluated our system by comparing the performance of our proposed music descriptors with other conventional features for music retrieval. The results showed that the performance of the proposed music descriptors was competitive with conventional features, thereby suggesting their potential use for describing music in semantic music search/retrieval.

1. Introduction

Proliferation of music data available for users increased the demand of searching and retrieving music that perfectly suits the individual listener's situation and entertainment needs. In order to meet such demands, Schedl and Knees (2013) emphasized the importance of personalized and user context-aware systems. As a consequence, music exploration systems implemented various functionalities in an attempt to provide more satisfying results. Some of the methods to search and retrieve music from such systems are searching music using metadata, retrieving recommended music, and browsing predefined playlists (Nanopoulos, Rafailidis, Ruxanda, & Manolopoulos, 2009).

Metadata includes music-related information such as the artist, title, and genre information. Users can query the music exploration system using text to retrieve the exact match. However, metadata lack contextual information, so using a metadata query will not satisfy the user if he or she seeks music in a certain context such as mood or situation. Recently,

* Corresponding author.
E-mail addresses: ziotoss@snu.ac.kr (Z. Hyung), jsp@hongik.ac.kr (J.-S. Park), kglee@snu.ac.kr (K. Lee).

http://dx.doi.org/10.1016/j.ipm.2017.04.006
0306-4573/

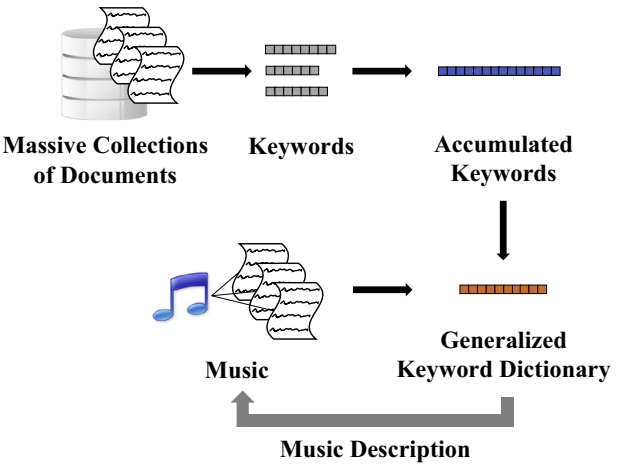

Fig. 1. The concept of the proposed system.

the utilization of social tags to enhance music description has been attempted in the research field of Music Information Retrieval. Symeonidis, Ruxanda, Nanopoulos, and Manolopoulos (2008) gathered social tags obtained from Last.fm to recommend music. However, as Lamere (2008) pointed out, most social tags are related to the artist, title, genre, and instrument, whereas only a small proportion is related to the user context. These social tags enhance text-based music search to some extent but they have difficulties with context-related terms.

Music exploration systems also provide music recommendations to users using various similarity measures. Collaborative filtering algorithms retrieve similar music by discovering similar user preference (Xing, Wang, & Wang, 2014). User preference is inferred by analyzing user ratings and/or user playlist. On the other hand, content-based algorithms utilize music-centric features such as timbre, pitch, and lyrics to discover similar music (Bogdanov et al., 2013; Li, Myaeng, & Kim, 2007; Mayer, Neumayer, & Rauber, 2008). These music-centric features include information obtained from the audio signal itself, but they lack information about the listener. Therefore, these music-centric features have a limited capacity to reflect the needs of the user.

Additionally, playlists tagged with predefined terms are also available in most of the music exploration systems. For instance, Allmusic[1] provides playlists that are tagged with predefined moods and themes. However, the recommendations provided are unbalanced in terms of the song distribution per mood/theme. For instance, the system suggests various songs for the *in love* theme but it only provides one song under the theme of *work*. Another problem is that it is difficult to build a consensus among the users because there is no specific standard for selecting moods and themes.

In this paper, we extract context-relevant keywords from a large collection of user-generated documents to capture the user contextual background when searching for music. Each document comes from a radio program's Internet bulletin board, where it comprises a personal story and a song request. Fig. 1 illustrates the concept of the proposed system. Hyung, Lee, and Lee (2014) showed that there is a strong correlation between personal stories and song requests. There are large number of such documents, so we consider that some general terms will be used to describe the user contextual background when requesting songs. Therefore, by performing keyword extraction based on these documents, it will be possible to create generalized context-relevant music descriptors, which can be applied to music search and retrieval. We collected 186,656 documents sent from the listeners of a radio program aired between 6:00 p.m. and 8:00 p.m.. We chose this program as they only air Western pop songs.

1.1. Contribution

In this paper, we introduce a novel approach that utilizes user-generated documents to extract user context for music search and retrieval. We described music using context-relevant keywords extracted from a large collection of user-generated documents. By utilizing context-relevant music descriptors, our system facilitates natural language text querying when searching for music and thus, it can retrieve music that satisfies the entertainment needs of users. Additionally, the keywords describing the user context are extracted from a large collection of documents, so the coverage of the explainable contextual background when listening to music will be broad. Therefore, users will be able to discover music in various contexts. Finally, we determined a correlation between our proposed context-relevant music descriptors and conventional features, such as acoustic features or lyrics, which could provide some insights into how to overcome the cold start problem where pieces of music with no document associations are never discovered.

[1] www.allmusic.com.

The remainder of this paper is organized as follows. The next section reviews recent papers that embed the user context for music discovery. Studies that exploited text to describe music are also reviewed. Section 3 explains the proposed system in detail. In Sections 4, 5, and 6, various experiments that were conducted in an attempt to validate the proposed approach are described. Finally, Section 7 presents the conclusions by providing a summary and suggesting directions for future research.

2. Background

Recently, novel approaches to the inclusion of contextual information related to users have been attempted actively in order to provide a richer context-relevant representation of music. In this section, we review recent studies that embedded contextual information related to users during music retrieval. We also describe some studies that employed text to enhance music descriptions.

2.1. Utilizing the contextual information of users

Several studies have aimed to infer various types of contextual information related to users when listening to music. Moens, van Noorden, and Leman (2010) investigated the relationship between music tempo and the user's pace while walking or jogging, where they suggested that if the music tempo is sufficiently close to the user's pace, then the user tends to synchronize their steps with the beat. In addition, Baltrunas et al. (2011) showed that a user's musical preferences may depend on differences in their driving status and Helmholz, Vetter, and Robra-Bissantz (2014); Hu et al. (2015) implemented a smartphone-based application that controls music playlist during different driving status. Some works extracted the user location information for music recommendation (Braunhofer, Kaminskas, & Ricci, 2013; Cheng & Shen, 2014; 2016; Schedl, Vall, & Farrahi, 2014).

Some studies have focused on modeling the user's emotional state when searching for music. Shan, Kuo, Chiang, and Lee (2009) utilized emotion by building a music emotion model using an affinity discovery from film music. The authors developed a Mixed Media Graph and an affinity graph algorithm to discover the affinities between music features and emotion. Lu and Tseng (2009) combined the content-based, collaboration-based, and emotion-based methods to build a hybrid music recommendation system. The authors altered the weights assigned to each method based on the user's listening behavior. Han, Rho, Jun, and Hwang (2010) modeled human emotional states and their transitions in music by using a novel emotion state transition model. They showed that it is possible to recommend the most appropriate music to users according to their desired emotional state. Yang and Liu (2013) identified the relation between user mood and music emotion, using social data. The authors suggested that using social data to extract user mood is applicable. Deng, Wang, Li, and Xu (2015) used microblogs to infer the user's emotional state when listening to music. Using a predefined list of emotions, the authors developed an association between the user emotion and music based on their analysis on the microblogs.

Other studies have aimed to use multiple types of contextual information related to users. Cunningham, Caulder, and Grout (2008) investigated how the user's emotion, activity status, and surrounding environment are related to music, where they implicitly collected contextual information for users based on their heart rate, motion data, temperature, and other parameters. According to their findings, they proposed a fuzzy logic model to create music playlists that reflect the user context. Su, Yeh, Yu, and Tseng (2010) utilized heartbeat, body temperature, air temperature, noise volume, humidity, light, motion, time, season, and location data, where they combined this information with content analysis based on music data to build a pattern database, thereby linking music with the users. Their results demonstrated that the system provided more effective recommendation lists.

However, most of these previous studies implicitly estimated the contextual information for users with various methods. Thus, the context used to represent music is enriched so the music retrieved matches the user's needs more closely, but the implicit extraction of the user context has limitations in terms of the scope of its coverage. However, if the user context can be obtained explicitly from a source that contains most of the user's possible contextual scenarios when listening to music, this would provide an ideal context-relevant representation of their preferred music. Using this representation could fully satisfy the user's needs when searching for music. The use of text is one method for explicitly expressing the contextual information related to users with an unlimited scope of coverage.

2.2. Exploiting text to describe music

In the past decade, several attempts have been made to exploit text when describing music. Textual representations of music can be obtained from metadata embedded in audio files and from the tags generated by humans. However, due to the size limitations of metadata, they mostly comprise information such as the album details, artist, title, genre, and the year of publication. Bertin-Mahieux, Eck, and Mandel (2010) described tags as keywords that are related to the given content, where these tags provide much more information about the content of the music as well as other information such as the users preferences. According to Turnbull, Barrington, Torres, and Lanckriet (2008), four methods can be used for gathering these tags: surveys, social tags, games, and web documents.

Surveying is the most expensive method because participants must be hired to listen to each song and tag them. Turnbull, Barrington, Torres, and Lanckriet (2007) collected the Computer Audition Lab 500 song (CAL500) dataset by allowing par-

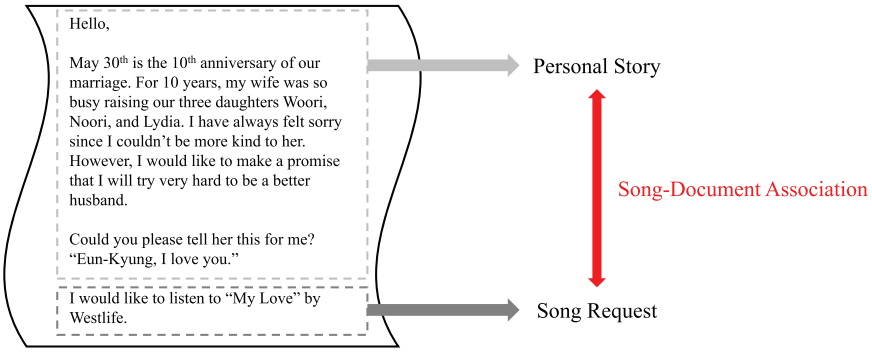

Fig. 2. An example of a document with a song request.

ticipants to listen and annotate songs in order to capture "semantic associations" between music and words, where they showed that it is possible to represent music using "semantic words." Kaminskas, Ricci, and Schedl (2013) built a system that matches tags describing a place of interest (POI) with tags describing music related to the POI. Both tags were gathered by performing a survey. The authors implemented text-based similarity measures to build a location-based music recommendation system.

As Bertin-Mahieux, Eck, and Mandel (2010) mentioned, social tags are tags gathered from a collaborative platform such as Last.fm. Hariri, Mobasher, and Burke (2012) introduced a hybrid music recommendation system that employs social tags obtained from Last.fm. The authors implement topic modeling module and assign a probability distribution of topics for each song. Finally, they represent music with a set of dominant topics used in the hybrid recommendation system. Kamalzadeh, Kralj, Möller, and Sedlmair (2016) introduced a novel mobile platform that allowed users to interact to discover music, where social tags gathered by Last.fm enabled the music to be categorized according to the *genre, mood*, and *other*. Finally, they represented each song using these tags.

Some studies implemented games to obtain the tags. Kim, Schmidt, and Emelle (2008) introduce MoodSwings, a collaborative game, to discover the association between emotion and music. Two players listen to a snippet of a song and they interactively select a point in a valance-arousal space. The authors were able to collect a vast amount of valence-arousal point labels which could be used to describe music. Aljanaki, Wiering, and Veltkamp (2016) performed an experiment by using a crowdsourcing game to study the emotions induced by music. The game provided participants with a snippet of a song categorized into one of four genres, i.e., *classical, rock, pop*, and *electronic*. After listening to the song, the participants selected three out of nine emotion-related terms provided by the game interface. After analyzing the data collected, the authors suggested that it may be possible to represent music using the nine different emotion-related terms.

Web documents are another useful resource and Schedl, Widmer, Knees, and Pohle (2011) developed an *Automatically Generated Music Information System* by crawling web documents related to musical artist or bands. In addition, Sordo, Oramas, and Espinosa-Anke (2015) utilized unstructured text sources to extract music-related entities such as songs, bands, persons, albums, and music genres. The authors used the extracted information to recommend music.

Various methods can be used to obtain text-based tags to describe music, but they have several limitations. Surveys require manpower, which is expensive. In addition, music libraries are enormous. Therefore, using surveys to obtain text representations of music is not practical. In the case of social tags, as mentioned by Lamere (2008), the majority describe the audio content. This biased distribution of social tags limits the capacity to capture contextual information related to users when searching for music. Web documents could overcome the limitation of social tags but previous studies mostly focused on collecting music-related information. Finally, using games limits the scope of coverage in the available contextual information, where they mostly focus on user emotions.

To overcome these limitations, Hyung, Lee, and Lee (2014) utilized documents generated by users that convey associations between personal stories and song requests. Fig. 2 illustrates an example of this type of document. Thus, given a query document, the system retrieves similar documents by performing document analysis. The authors showed that people who share similar contexts also share similar musical preferences, and thus it is possible to recommend songs associated with similar documents.

However, this approach has some limitations. First, the algorithm requires a document with a certain length as an input query, which is insufficient when searching for music. Second, popular songs will have more document associations because they will be requested more frequently by various users. As a result, the system will recommend popular music more frequently, thereby leading to the well-known popularity bias. Third, the system is affected by the cold start problem due to the song-document association requirement. The cold start problem arises when the system cannot draw any inferences about songs due to an insufficient amount of documents. Therefore, the system will never discover songs that are not associated with any documents. To overcome these limitations, we propose a novel method that utilizes generalized context-relevant keywords to describe and discover music.

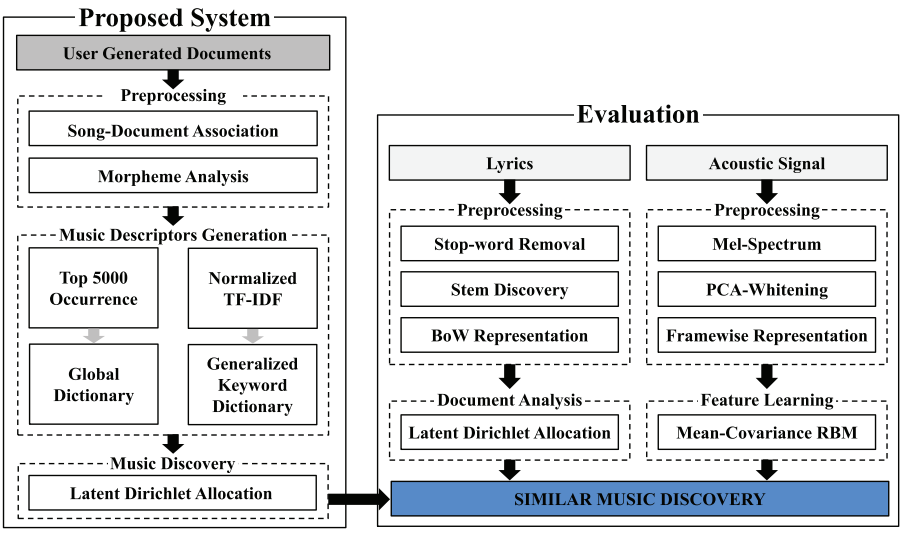

Fig. 3. Block diagram of the proposed system and the evaluation process.

3. Context-relevant music descriptors

In this section, we explain our system, which uses context-relevant keywords to describe music. Fig. 3 shows a block diagram that illustrates our system and the evaluation process.

Our system comprises three steps: preprocessing, music descriptor generation, and music discovery. In the preprocessing step, we extract the song information from each document to create song-document associations. We also remove stop-words and stemmed words during this step. In the subsequent music descriptor generation step, we develop two different music descriptors: 1) a Global Dictionary (GD) comprising terms used in the complete document collection, and 2) a generalized Keyword Dictionary (gKD) comprising terms that users generally employ to describe their context when listening to music. Finally, each song is represented by a bag-of-words (BoW) model using these dictionaries, and our system can discover music with a similar user context background by implementing a text analysis algorithm.

3.1. Preprocessing

As mentioned earlier, our system utilizes documents that comprise personal stories and song requests. First, we needed to extract the song information from each document to create a song-document association. The target radio program airs Western pop songs, so the users typically wrote their personal stories in Korean whereas they wrote their song requests in English. Therefore, we used the Million Song Dataset (MSD) (Bertin-Mahieux, Ellis, Whitman, & Lamere, 2011) to discover the song information within each document. The documents were created by users, so the strings that formed the song requests contained misspelled words, which made it difficult to find exact matches in MSD. Therefore, we found the closest candidate match in MSD by implementing a fuzzy string matching algorithm.

Fuzzy string matching is a process for finding strings that match a pattern approximately rather than exactly. It is widely applied in commercial applications because real-world data contain many errors. We used the well-known Levenshtein distance in the fuzzy string matching process. The Levenshtein distance is the minimum number of single-character edits, such as insertions, deletions, or substitutions, required to change one word into another. After calculating the Levenshtein distance between the song request and each song in MSD, we associate the song that had the lowest Levenshtein distance with the given document.

Our approach utilizes documents that are generated by users, so there are two significant problems when we use the words as they appear in documents: words stemmed from a single word and stop-words. Stemmed words are derived from the same word but they have different spelling formats, e.g., *go, going*, and *gone* are stemmed from the word *go*. If we do not discover the stem of the word, all of the stemmed words are regarded independently, which degrades the performance of our system. Similarly, stop-words such as *and, the*, and *at* have no meaning, and thus they should be removed to improve performance. We used the Korean morpheme analysis tool developed by Kookmin University to eliminate stop-words and to discover the stem of each word.[2]

[2] http://nlp.kookmin.ac.kr/HAM/kor/index.html.

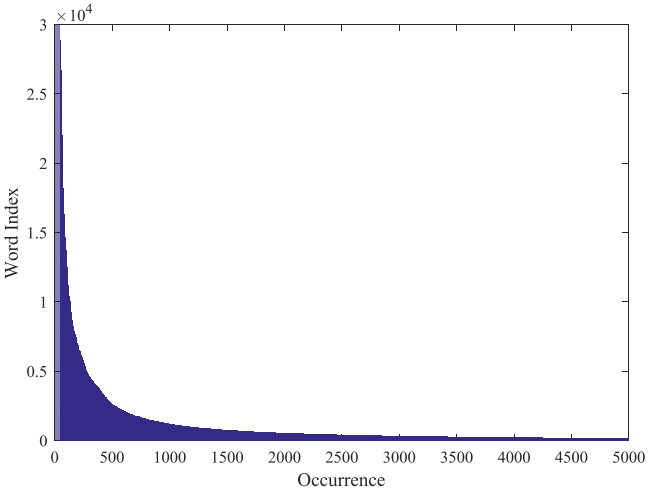

Fig. 4. A snippet of the word distribution.

The documents comprise a personal story and a song request, but excessively short documents contain little or no contextual information. Therefore, we removed documents with less than 20 words after applying the Korean morpheme analysis.

3.2. Generating music descriptors

In this section, we describe two different features used by our system: Global Dictionary (GD) and generalized Keyword Dictionary (gKD). To generate the GD, we first listed the unique terms used in the large collection of documents. We then counted the occurrence of each term to determine the distribution. Part of the distribution is shown in Fig. 4, which indicates that the distribution is extremely long-tailed. The top 5000 terms account for approximately 83% of the total occurrences. We are interested in general terms that can be used to describe the context of the user when requesting a song, so it is not necessary to use terms that are located in the tail. We assume that the top 5000 high occurrence terms will include all the necessary context-relevant terms, and thus they can be used as music descriptors.

During the generation of the GD, we removed occasionally used terms, but the GD still contained frequently used terms. However, frequently used words such as *do, have*, and *song request* are unnecessary because they do not contain meanings that distinguish the contextual characteristics of songs. In addition, we considered that the GD included many noisy terms that were not suitable for use as general terms for describing the user context when listening to music. Therefore, we extracted the keywords by performing term frequency-inverse document frequency (TF-IDF) to remove frequently used terms and noisy terms.

The TF-IDF metric evolved from the IDF metric, which was introduced by Spärck Jones (1972); 2004), and it has been used widely to extract keywords from documents. The TF is the weight of a term occurring in a document and the IDF is a measure that determines whether the term is common or rare in all of the documents. The simplest method for computing TF is counting a term's occurrence. However, if the lengths of the documents differ significantly, TF will be biased toward terms that occur in longer documents. In longer documents, terms are used repeatedly, which leads to a bias over shorter documents if the occurrence of the terms is not normalized. We gathered the documents from an Internet bulletin board, so there were no restrictions on their length, thereby resulting in significantly different document lengths. Therefore, we normalized TF by using the Smart system's augmented TF factor to overcome this problem (Salton & Buckley, 1988). The equations for TF and IDF are shown in (1) and (2), respectively.

$$TF(t,d) = 0.5 + 0.5 \times \frac{f_{t,d}}{max\{f_{t',d} : t' \in d\}} \tag{1}$$

$$IDF(t,D) = \log \frac{N_D}{|\{d \in D : t \in d\}|} \tag{2}$$

In the given equations, $f_{t,d}$ indicates the occurrence of term t within document d, D is the complete document corpus, N_D is the number of documents in the corpus, and $|\{d \in D: t \in d\}|$ is the number of documents that contain the term. TF-IDF is the product of TF and IDF.

After computing TF-IDF for each term, we conducted an experiment to determine the most appropriate number of terms extracted from each document. After varying the number of terms extracted from each document, we manually inspected the extracted keywords. We found that selecting the top 10% of terms provided the most reasonable keywords for use when

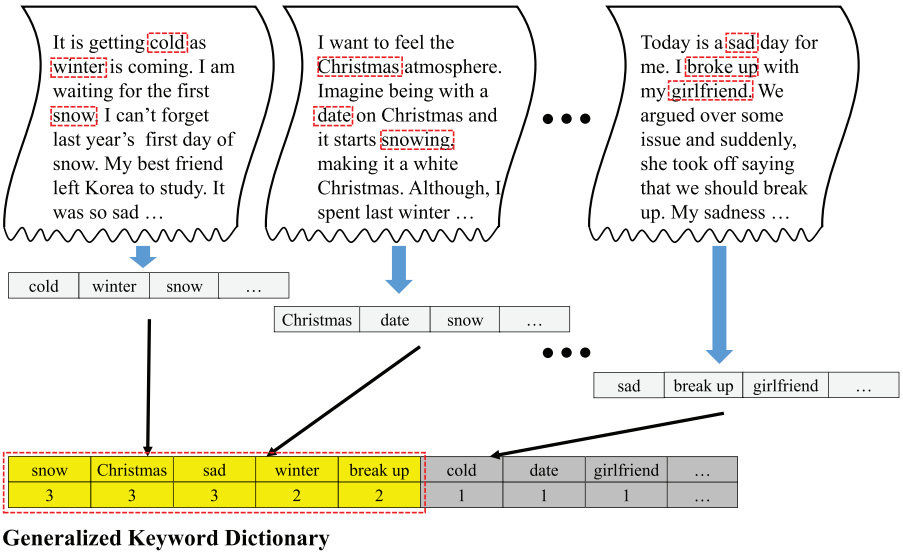

Fig. 5. An example of creating the generalized Keyword Dictionary.

describing the user context. Therefore, we decided to retrieve the top 10% of terms with the highest TF-IDF values from each document to create a keyword list. Using this keyword list, we counted the occurrence of each term in the large collection of documents and we defined 500 keywords with the highest occurrence as the gKD for describing music. This process is illustrated in Fig. 5.

Using the GD and gKD, each document was represented as a BoW model. In the large collection of documents, multiple documents requested the same song, which indicated that for each song, there were numerous stories by various listeners. Therefore, by aggregating the BoW vectors for each song, we built a generalized, context-relevant representation of the song. The number of documents associated with each song varied, so we normalized each representation by dividing each element of the BoW vector by the total sum of the BoW vector.

3.3. Music discovery

Among the total number of documents in the dataset, only a small proportion was associated with each song. Hence, the BoW vector representation of the song was sparse. According to previous studies of information retrieval, it is not practical to compute the distance between the sparse vectors directly, but instead various algorithms such as latent semantic analysis (LSA), probabilistic latent semantic analysis (pLSA), and latent Dirichlet allocation (LDA) can be used to discover the underlying meanings between terms. The system then transforms the sparse vector into a non-sparse vector, thereby facilitating the appropriate calculation of the distance between vectors. We compared LSA, pLSA, and LDA to examine which algorithm best captures the underlying meanings of the given documents and discovered that LDA performed the best. Hence, we implemented LDA to compute the distance between the BoW vectors that described music.

Blei, Ng, and Jordan (2003) first introduced LDA, which is a generative statistical model where unobserved or latent groups are used to explain the observations. For example, given the task of discovering similar documents, LDA represents each document using a mixture of topics where a distribution over words characterizes each topic. In our approach, we used documents that contained an association between a personal story and a song request. Hyung, Lee, and Lee (2014) showed that people in similar contexts prefer similar music, so people who request the same song will share similar stories. Therefore, the resulting topics will reflect a collaborative shared view of the song.

The modeling process of LDA can be described as finding $P(z|d)$, with each topic described by the probability distribution of words $P(w|z)$. This process can be formalized as shown in (3).

$$P(w_i|d) = \sum_{j=1}^{Z} P(w_i|z_i = j)P(z_i = j|d) \tag{3}$$

$P(w_i|d)$ is the probability of the ith word for a given document d and z_i is the latent topic. $P(w_i|z_i = j)$ is the probability of the ith word in topic j and $Pz_i = j|d)$ is the probability of selecting a word from topic j in the document d. Estimation of $P(w|z)$ and $P(z|d)$ can be accomplished by performing Gibbs sampling (Griffiths & Steyvers, 2004). The estimation of the

Table 1
Topic examples resulted by performing LDA on gKD.

T_6	T_{13}	T_{21}	T_{23}	T_{45}
Request	Birthday	Love	Study	Perform
Song request	Congratulate	Mind	School	Tiresome
Air music	Marriage	Happy	Exam	Exhausted
Ask for	My	Exist	Student	Lose
Hello	Today	Heart	University	Life
Chulsoo Bae	Thank you	Do not	Teacher	Work
Play music	Happy	Boyfriend	Highschool	Drive
Mr.	Ask for	Girlfriend	Limp	Watery
Good	Song request	Forget	English	Care for
Effort	Health	Delighted	Graduate	Difficult

Table 2
Labels used to categorize each term with some examples.

Lamere	Extended	Example words
Genre	Content	heavy metal, punk
Instrumentation		piano, female vocal
Locale	Locale	Korea, office, school
Mood	Mood	exciting, sad, happy
Opinion	Opinion	different, favourite, romantic
Personal	Misc	seen live, I own it
Style		political, humor
Organization		check out
N/A	Situation	love, final exam, birthday
	Date	November, Wednesday, weekend
	Weather	winter, rain, hot day

probabilities is shown in (4) and (5).

$$P(w_i|z_i = j) = \frac{C^{WZ}_{w_i j} + \beta}{\sum_{w} C^{WZ}_{wj} + W\beta} \tag{4}$$

$$P(z_i = j|d_i) = \frac{C^{DZ}_{d_i j} + \alpha}{\sum_{z} C^{DZ}_{d_i z} + Z\alpha} \tag{5}$$

C^{WZ} maintains a count of all word-topic assignments and C^{DZ} counts the document-topic assignments. By discovering the latent topics in the terms that comprised GD and gKD, it was possible to represent each song using a mixture of topics. This representation was no longer sparse so we could compute the distance between songs. We used 500 topics to model GD and used 50 topics to model gKD. Various distance measures can be employed but we used the Cosine similarity. Some of the topics that were modelled are shown in Table 1.

T_6 includes terms that are related to the radio program. For instance, Mr. Bae is the DJ of the program and people request music using the terms listed in the topic. T_{13} included terms related to activities worth congratulating. T_{21} modelled terms that are related to lovers while T_{23} modelled terms related to school. Finally, T_{45} included terms related to being tiresome. While these examples show that the topics were properly modelled conveying the latent meanings, the documents are written in Korean and we had to translate them to English. During this process, some meanings were misinterpreted. For instance, in Korea the term *Effort* is used similarly as *Best regards* in English. Therefore, people usually use them when concluding their personal stories with a song request.

4. Social tags vs. context-relevant music descriptors

In order to show that our music descriptors contained more context-relevant terms than social tags, we compared the terms in gKD with the social tags provided by Last.fm. For this comparison, we first categorized the terms and the tags with several labels. Lamere (2008) used the labels *Genre, Locale, Mood, Opinion, Instrumentation, Style, Misc, Personal*, and *Organization* to categorize Last.fm tags. Our aim was to examine the distribution of music-related words and context-related words, so we combined the music-related labels *Genre* and *Instrumentation* to create a *Content* label. In addition, we labeled artist-related words as *Content*. Finally, we added more specific context-related labels such as *Date, Weather*, and *Situation*. The labels *Style, Personal*, and *Organization* were included in *Misc*, which we used to label words with little meaning or those for which it was difficult to give a specific label. Excluding *Content* and *Misc*, the labels were regarded as context-relevant. The extended labels with some example words are shown in Table 2.

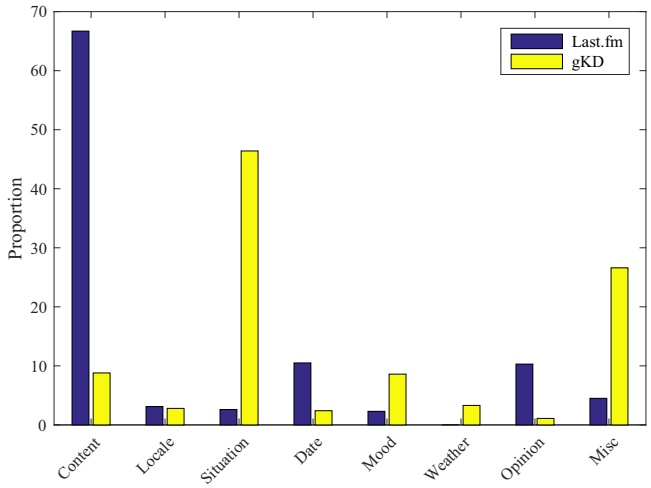

Fig. 6. Label distribution comparison between social tags and terms included in our proposed gKD.

We used songs with 20 or more document associations and collected the 20 most frequently applied social tags for each song from Last.fm. Using the labels described above, we categorized the collected social tags. We also manually categorized the terms in the gKD. Ten different people participated in this process and the label with the majority vote was assigned. Finally, we compare the label distribution of the social tags gathered from Last.fm and the label distribution of the terms in our proposed context-relevant music descriptors. The result is shown in Fig. 6.

The distribution of Last.fm has a similar pattern to that described by Lamere (2008). Over 66% of the social tags are content-related terms, thereby indicating that social tags are more focused on providing information regarding the music content itself. The terms within the gKD only contained 9% of the content-related terms. Over 64% of the terms were context-relevant, which indicates that our method produces far more context-relevant music descriptors than social tags.

5. Examining the correlation with conventional features

5.1. Dataset

In order to examine the correlation with conventional features, we built a noise-free dataset. The dataset comprises songs with more than 20 document associations. We manually removed documents associated with incorrect songs. Thus, we had 350 songs with a minimum of 20 and a maximum of 365 documents associated with a song. The lyrics data were manually downloaded from various Internet sites, such as MetroLyrics[3] and LyricsFreak.[4] We selected these two websites due to their abundance of lyrics data. Finally, we downloaded the whole audio file from a Korean online music service provider to extract the acoustic features.

5.2. Examining the correlation with conventional features

There are many works that use acoustic features and lyrics to retrieve similar music. Therefore, we use these two features to examine if it is possible to use our proposed music descriptors for similar music discovery. If the proposed music descriptors contain valuable information which can be used during similar music retrieval, it will have some correlation with the well-used acoustic features and lyrics. Therefore, we examine the correlation between the proposed features with acoustic features and lyrics by adopting recent music retrieval systems that use acoustic features and lyrics when identifying similar music.

Recently, Schluter and Osendorfer (2011) implemented the mean-covariance restricted Boltzmann machine (mcRBM) for music retrieval. We employed this method as the baseline algorithm using acoustic features. Following the previously described procedure, we first extracted the Mel-spectrum of the audio clip and concatenated 39 neighboring frames to generate the Mel-spectral blocks. Due to the high dimension of the Mel-spectral block, PCA whitening was applied where the least significant components were discarded. Finally, the mcRBM was trained using 200,000 blocks. We avoided over-fitting by using different segments of each song for training and evaluation.

Sasaki, Yoshii, Nakano, Goto, and Morishima (2014) introduced a system that uses lyrics for music retrieval. Thus, by using their procedure for lyrics analysis, we implemented LDA to represent the lyrics of the songs, included in the dataset

[3] www.metrolyrics.com.
[4] www.lyricsfreak.com.

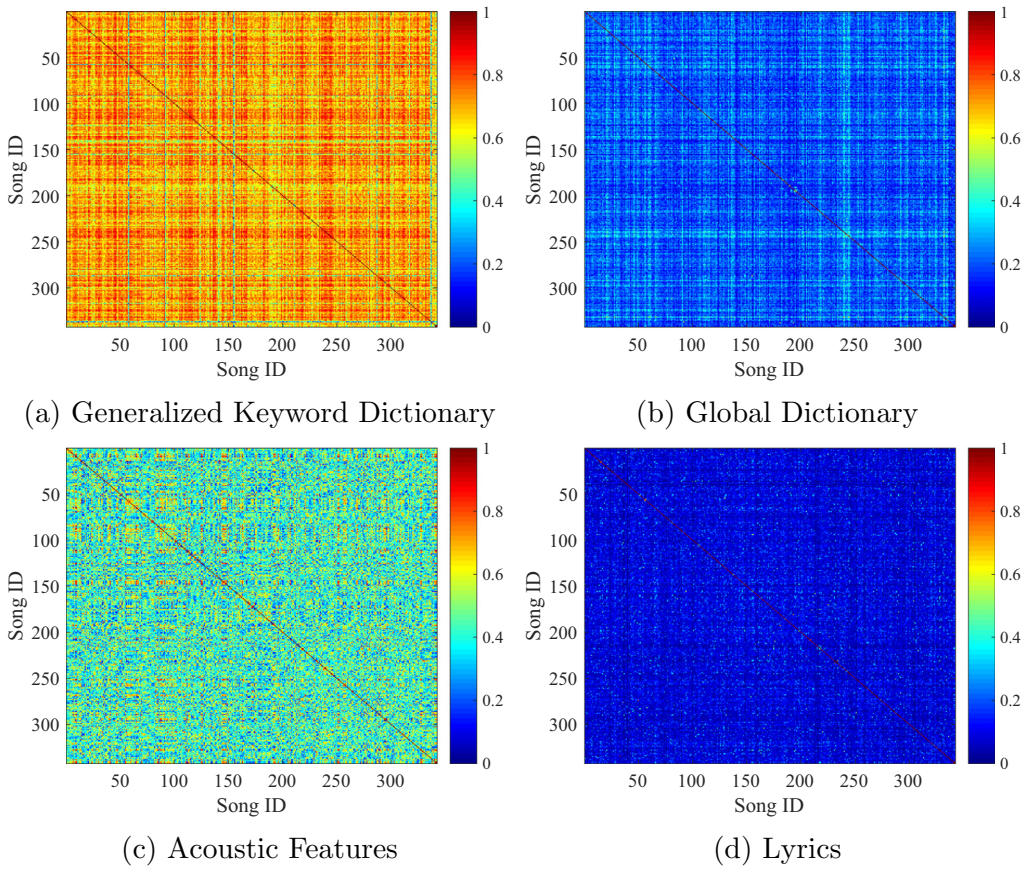

Fig. 7. Similarity matrix using different music descriptors as features.

explained above, with a mixture of latent topics. To avoid over-fitting when training the LDA, we used 50,000 lyrics of songs that were not included in the dataset.

Our proposed algorithm employs LDA to discover music, and thus a latent topic space that captures the characteristics of the context-relevant music descriptors is created. We refer to this feature space as the "context-relevant feature space" (CFS). In addition, by implementing the baseline algorithms explained above, we created two feature spaces: one representing the acoustic characteristics of music and another representing the characteristics of lyrics. We refer to the former feature space as the "acoustic-relevant feature space" (AFS) and the latter as the "lyrics-relevant feature space" (LFS).

We project the songs into these feature spaces and compute the distance between each song using the Cosine distance metric. Using the distance matrix, we then compute the similarity matrix. We normalized each similarity matrix by dividing the maximum value so that all matrices would have values between zero and one. Once we find the similarity matrix for all three feature spaces, we examine the correlation between CFS and other feature spaces by performing the Mantel test. Mantel test is a statistical test used to compute the correlation between two similarity (distance) matrices. We then provide examples that show the highest correlation between CFS and AFS, and between CFS and LFS.

5.3. Results

Using each generalized Keywords Dictionary, acoustic features, and lyrics, we computed the similarity matrix between songs. The similarity matrices are shown in Fig. 7. From the similarity matrices, we can observe that if the proposed music descriptors are used as musical features, most songs are regarded similar while if the lyrics are used as musical features, most songs are regarded different. Such difference can be explained by the different size of the dictionary. The dictionary size of the proposed music descriptors is 500 while the dictionary size of lyrics is 5000. Hence, lyrics are much more descriptive than the proposed music descriptors distinguishing each song more precisely. On the other hand, acoustic features show that they are somewhere in between lyrics and the proposed music descriptors. However, from the similarity matrices, it is difficult to identify rather if our music descriptor is feasible or not.

We additionally performed the Mantel test to observe the correlation between similarity matrices. The results are shown in Table 3. The results show that the proposed music descriptors are more correlated to lyrics than the acoustic features.

Table 3
Correlation coefficients between the similarity matrix computed using the proposed gKD and the similarity matrix computed using the conventional features.

	Correlation coefficient
Acoustic features	0.0009
Lyrics	0.0243

(a) "Dancing queen" by Abba

(b) "Nothing's gonna change my love for you" by Glenn Medeiros

Fig. 8. Word cloud of the context-relevant representation for the song with best performance. The relevant songs were assigned from (a) AFS and (b) LFS.

This can be interpreted as lyrics having more influence on which music to listen to given a music listening context. In many cases, lyrics include information of when it is good to listen to the song. For instance, if a song has lyrics that include a story about love, people who are dating someone will likely listen to the song. On the other hand, acoustic features are less related to the music listening context. The acoustical characteristic is vague information in determining which music to listen to given a music listening context.

While the proposed music descriptors are more correlated to lyrics than acoustic features, there are some songs that showed high correlation between the proposed music descriptors and acoustic features. Therefore, we provide examples that showed highest correlation for each acoustic features and lyrics. The song with the highest correlation between the proposed music descriptors and acoustic features was "Dancing Queen" by Abba. The word cloud for the context-relevant keywords assigned to the song is shown in Fig. 8a. On the other hand, the song with the highest correlation between the proposed music descriptors and lyrics was "Nothing's gonna change my love for you" by Glenn Medeiros. We illustrate the context-relevant keywords assigned to the song in Fig. 8b.

By examining the word cloud of "Dancing Queen", it is possible to find terms that are related with the musical characteristics of the song. Since "Dancing Queen" is a dance song with a bright melody, people requesting this song often used positive terms. For example, *happiness, exciting*, and *joyful* was frequently used which harmonize with the acoustic characteristics of the song. On the other hand, the word cloud of "Nothings gonna change my love for you" have direct connection with the lyrics. The lyrics of the song are about passionate love. Therefore, users who requested this song used words such as *love, like*, and *marriage* often. Also, words that indicate the subject of the love, for instance, *mother, husband*, and *wife*, were also frequently used.

From these examples, we were able to examine the correlation of the proposed music descriptors and conventional features. Since the proposed music descriptors can explain the music piece with terms related to music listening context and there are some correlations with conventional features, we assume that it can be feasible to use these music descriptors for similar music retrieval.

6. Qualitative evaluation

By comparing the label distribution of social tags obtained from Last.fm with the label distribution of terms within our proposed context-relevant music descriptors, we showed that our approach contains richer context-relevant terms. Additionally, we examined correlation between the proposed music descriptors and conventional features. However, this does not verify if the proposed music descriptors can be used for music discovery. Therefore, in an attempt to show that our music descriptors are valid, we conducted a user evaluation test.

Table 4
Information of the query songs used for user evaluation.

Title	Artist	Genre
What a Wonderful World	B. B. King	Blues
All Aboard	The Wiyos	Blues
Cry	Faith Hill	Country
Somewhere Over the Rainbow	Impellitteri	Country
Everytime	A1	Electronic
As Long as you Love me	Backstreet Boys	Electronic
White Lies	Rose Cousins	Folk
It's a Long Way to the Top	Lucinda Williams	Folk
Take Five	Dave Brubeck	Jazz
Don't Know Why	Norah Jones	Jazz
Que Lloro	Sin Bandera	Latin
Mas Que Nada	Tamba Trio	Latin
Rebirth	AZ	Rap
The Anthem	K-OS	Rap
Need Your Love	Aswad	Reggae
Danger	Thriller U	Reggae
Hold My Hand	Van Hunt	R&B
C.R.U.S.H.	Ciara	R&B
Last Christmas	Wham	Pop-Rock
November Rain	Guns & Roses	Pop-Rock

6.1. User study setup

We used songs with at least one document association for the user evaluation test. No manual noise filtering was performed. The lyrics were obtained from MusiXMatch, which is the official lyrics collection for the MSD. The audio clips for songs were collected by downloading audio preview clips provided by 7Digital.[5] The lengths of the clips varied between 30 and 60 seconds. There were total 7192 songs used in the user study.

We randomly selected two songs from each genre to obtain a total of 20 query songs. The genre information was retrieved from Allmusic. Each participant was randomly assigned 10 query songs. An audio clip with duration of 30 s was provided as well as song-related information such as the artist, title, and lyrics. While listening to the query song, the participants were asked to recollect their memories of the song or to imagine a situation where the song would be likely to be heard. After listening to the query song, the participants were given a questionnaire containing four songs. The songs were retrieved using lyrics, acoustic features, GD, and gKD. We followed the methods explained in the previous section to discover similar music using lyrics and acoustic features. The method for discovering similar music using GD and gKD is explained in Section 3.3. For each retrieved song, the participants were asked to answer two questions.

- Q1. Is the song similar to the query song?
- Q2. Does the song fit with the memories/situation recalled/imagined from the query song?

The first question was taken from the human evaluation task used by MIREX Audio Music Similarity and Retrieval,[6] and the second question was obtained from a study by Wang et al. (2013) with some modification. We asked the participants to answer both questions on a five-point scale. Finally, the participants were required to provide demographic information such as their gender and age. The songs used for user evaluation is described in Table 4.

6.2. Results

We collected the responses from the user's evaluation for 10 days. In total, 50 participants made valid responses, where 34 participants were male and the other 16 were female. Age of the participants varied from 20 to 60 years. The mean ratings for Q1 and Q2 are shown in Fig. 9.

Excluding lyrics, all the other features had significantly different ratings for Q1 and Q2 ($p < 0.05$), where the ratings for Q2 were higher than those for Q1. These results indicate that there were some discrepancies between music similarity and context similarity. Some participants regarded the recommended music as suitable for the context of the query song regardless of the music's similarity, which suggests that it is important to consider the user context when recommending music. We mainly analyzed the results of Q2 because our key aim is identifying suitable features so the contextual information of users is considered when discovering music. The mean ratings for Q1 and Q2 showed that there was a significant difference between Q1 and Q2, but the mean ratings between different features in Q2 did not differ significantly. Therefore, we performed a further analysis by computing the mean ratings for each genre. The results are shown in Fig. 10.

[5] www.7digital.com.
[6] www.music-ir.org/mirex/wiki.

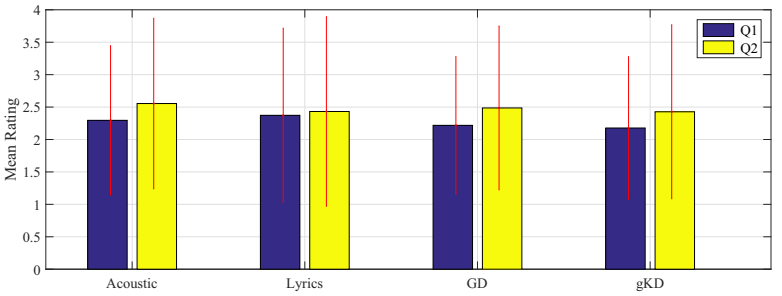

Fig. 9. Mean ratings for Q1 and Q2.

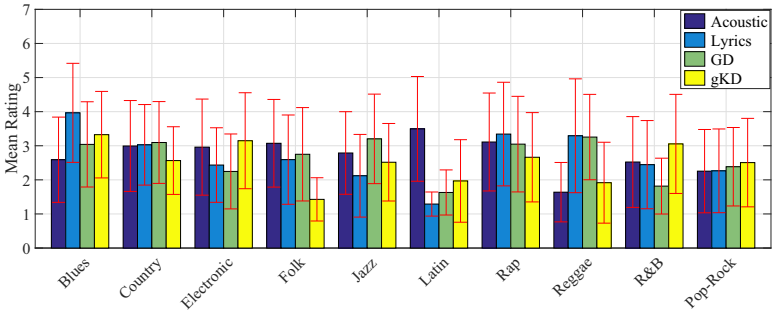

Fig. 10. Mean ratings for Q2 per genre.

The results showed that using lyrics performed best for Blues, whereas our proposed features gKD and GD ranked second and third, respectively. A possible explanation for this result may be the characteristics of the recommended songs. One of the songs in the Blues category was “What a beautiful world.” When lyrics were used to discover similar music, a cover song was retrieved because cover songs share almost identical lyrics. This boosted the ratings for the music retrieved based on lyrics similarity. Our proposed features also proved competitive because people shared a similar context when requesting songs in the Blues category. For example, “Somewhere over the rainbow” by Il Divo was retrieved when the gKD was used. Some common terms used to describe the query song and the retrieved song were *classical, calm*, and *delightful*. Therefore, it is possible to imply that people shared a similar contextual background when listening to songs in the Blues category.

For the songs in the Latin category, the use of acoustic features performed considerably better than the other three features. This result can be explained by the rhythmic characteristics of Latin music. There are variants of the rhythm, but the most fundamental rhythm is called a clave. In most Latin songs, the rhythm structure is based on the clave and this strong rhythmic characteristic is distinguishable from other genres. The AFS explained the rhythmic structure well, thereby resulting in good retrieval performance. By contrast, the remaining three features obtained poor performance, which indicates that for Latin songs, the participants did not agree on lyrics similarity or contextual background similarity. A possible explanation for this result may be found in the language of the lyrics. Latin songs are in Latin languages so the participants could not understand what the lyrics meant. Therefore, only the acoustic characteristics were considered when providing the ratings.

Both of our proposed features performed better than the conventional features for Pop-rock, which indicates that the participants agreed on the similarity of the contextual background for the query song and the retrieved song. The two query songs in the Pop-rock genre were “Last Christmas” by Wham and “November rain” by Guns & Roses. Terms such as *winter, present, Christmas*, and *snow* were frequent for “Last Christmas,” and terms such as *autumn, November*, and *rain* were common for “November Rain.” These terms indicate that the two songs are often heard and requested at specific times or seasons. For example, “Last Christmas” is mostly heard during the winter season and specifically during the Christmas period. The song retrieved for “Last Christmas” was “All I want for Christmas” by Rupaul. This song is also a Christmas song that shares a similar contextual representation. By contrast, “November rain” seems to be heard most often during the autumn and on rainy days. The song retrieved for “November rain” was “Shape of my heart” by Sting. After inspecting the context-relevant representation of “Shape of my heart,” we found that autumn-related terms were frequently used for this song. For these two songs, a consensus was built on when the songs are likely to be heard, and thus the participants gave high ratings to the songs retrieved using our context-relevant representations.

An interesting but trivial finding was that lyrics performed the best for the Rap genre. This result agrees with the characteristic of Rap. Due to its complex rhyme structures, lengthy lyrics, and distinctive vocabulary, Rap was easily distinguished from other genres by using lyrics as a feature when retrieving similar music.

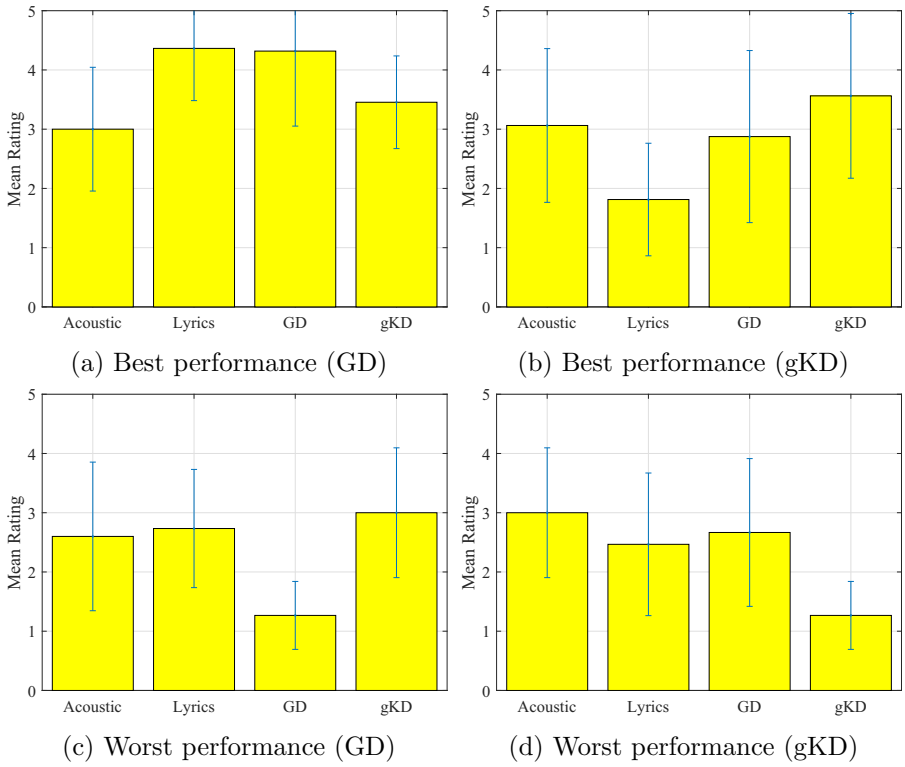

Fig. 11. Best and worst ratings for Q2 per proposed features.

Given the task of discovering music that suits the user context, we found that our proposed features could compete with conventional lyrics and acoustic features in most of the genres. However, analysis based on genre had some limitations. For example, two songs in the Jazz category, i.e., "Take five" by Dave Brubeck and "Don't know why" by Norah Jones, obtained variable results. By computing the mean value, it was possible to say that GD had the best performance, but different conclusions could be derived when the results for each song were inspected separately. Two different reasons may explain this discrepancy: 1) the vagueness of dividing music based on a single genre; and 2) the contextual background when listening to music depends on each specific song rather than the genre. Therefore, we analyzed songs with the best and worst performance using our proposed descriptors, and the results are shown in Fig. 11.

The best performance obtained when we used the GD to retrieve music was for the query song "Somewhere over the rainbow" by Impellitteri. The song retrieved was "Aubrey" by Bread. By manually inspecting the GD, we found that terms such as *tranquil, peaceful*, and soft were often used to describe both songs, possibly because both songs have a soft beat and moderate tempo. Another interesting finding was that both songs shared family-related terms such as *mother, father*, and *uncle*. We also inspected the documents associated with the songs and found that family-related terms were used because both songs were released more than 20 years ago; thus, several people requested these songs because their mother, father, or uncle liked the song. The GD provided a rich contextual representation, so it was possible to capture this pattern.

The query song "As long as you love me" by Backstreet Boys had the best performance when the gKD was used, where the song retrieved was "It's my life" by Bo Bice. In this case, it was hard to find a common theme within the lyrics. The query song is about love whereas the retrieved song is about a way to live one's life. However, this result may be attributed to the acoustical characteristics of the songs. Both songs share a similar tempo and melody structure, and thus they shared several mood-related words in the gKD, thereby receiving high ratings from the participants. This result implies that context-relevant music descriptors could capture the acoustic characteristics of the two songs to satisfy the participants.

The worst performance obtained using the GD was for the query song "Everytime" by A1, where the retrieved song was "My life" by Guce. This low performance was due to the inaccuracy of the fuzzy string matching algorithm. When we manually inspected the documents associated with "My life," the context was similar to that described in documents associated with "Everytime," but the song request was not "My life." The fuzzy string matching algorithm tries to discover the most likely candidate, so "My life" was assigned, thereby resulting in a mismatch in the song-document association.

When we used the gKD to retrieve similar songs, the query "White lies" by Rose Cousins had the worst mean ratings, where the retrieved song was "Metamorphosis" by Evile, and this poor performance was possibly due to the lack of documents associated with the song. Only one document was associated with each song and the context written in each

document was similar. However, it is difficult to generalize the contextual background when listening to music with only one document association, thereby resulting in the failure to convince the participants to give high ratings.

Based on the user evaluation, we discovered some limitations but we found that our proposed music descriptors could compete with the conventional features at music discovery. Therefore, our proposed algorithm based on context-relevant keywords could be used to describe music.

7. Conclusion

In this paper, we proposed a novel method for representing music by conducting keyword extraction on user-generated documents. Utilizing a large collection of documents containing the contextual information from users and a song request, we first built song-document associations by discovering the request song through calculating the Levenstein distance. We then extracted keywords from each document to obtain context-relevant music descriptors, which we used to explain the user's general contextual information when searching for music. Finally, we represented each song using the generalized context-relevant music descriptors.

We evaluated our method by comparing the keywords generated using our system with the social tags gathered from Last.fm, which showed that our approach provides richer contextual information than social tags. We also performed various quantitative evaluations by comparing our proposed music descriptors with conventional features such as lyrics and acoustic features when discovering music. Our statistical analyses showed that there were correlations between the context-relevant representations with lyrics and acoustic features, and thus our proposed music descriptors could be used to enhance music descriptions. We also performed qualitative evaluations and showed that our proposed features can compete with conventional features at discovering music.

The main contribution of our method is that we employed user-generated documents to represent music with context-relevant keywords. The keywords were extracted from the documents created by the users, so the scope of the context-relevant terms is not limited, which facilitates richer context-relevant music descriptions. By including contextual information when describing music, it is possible to search for music using context-relevant terms and this is an improvement compared with current music search systems, which are focused mostly on content-relevant words, such as the artist, title, and genre. Therefore, it could be implemented to enable users to search music in a semantic way.

In addition, we statistically demonstrated that there were correlations between the proposed context-relevant music descriptors with lyrics and acoustic features. These correlations mean that is possible to utilize our music descriptors to auto-tag songs without song-document associations, thereby addressing the cold start problem. This will be investigated in future research.

Based on our different experiments, we discovered some limitations. First, incorrect song-document associations could be produced because of inaccurate context-relevant representations of music. This limitation was due to errors generated when computing the Levenshtein distance using the fuzzy string matching algorithm. However, this limitation may be resolved by providing a formal platform that could allow users to write their personal stories and song requests. Another possible solution could be using a better algorithm for the fuzzy string matching process. The second limitation was the production of insufficient documents associated with a song, which could be resolved by employing the correlations that we determined between our proposed music descriptors and conventional features. For the songs with few document associations, we could implement an algorithm to auto-tag them by discovering similar songs with abundant document associations using lyrics and/or acoustic features. However, as mentioned above, this was beyond the scope of the present study and it will be addressed in future research.

Acknowledgement

This research was supported by the MSIP (Ministry of Science, ICT, and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2016-H8501-16-1016) supervised by the IITP (Institute for Information & communications Technology Promotion).

References

Aljanaki, A., Wiering, F., & Veltkamp, R. C. (2016). Studying emotion induced by music through a crowdsourcing game. *Information Processing & Management, 52*(1), 115–128.

Baltrunas, L., Kaminskas, M., Ludwig, B., Moling, O., Ricci, F., Aydin, A., et al. (2011). Incarmusic: Context-aware music recommendations in a car.. In *Ec-web: 11* (pp. 89–100). Springer.

Bertin-Mahieux, T., Eck, D., & Mandel, M. (2010). Automatic tagging of audio: The state-of-the-art. *Machine Audition: Principles, Algorithms and Systems*, 334–352.

Bertin-Mahieux, T., Ellis, D. P., Whitman, B., & Lamere, P. (2011). The million song dataset.. In *Ismir: 2 No. 9* (p. 10).

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research, 3*(Jan), 993–1022.

Bogdanov, D., Haro, M., Fuhrmann, F., Xambó, A., Gómez, E., & Herrera, P. (2013). Semantic audio content-based music recommendation and visualization based on user preference examples. *Information Processing & Management, 49*(1), 13–33.

Braunhofer, M., Kaminskas, M., & Ricci, F. (2013). Location-aware music recommendation. *International Journal of Multimedia Information Retrieval, 2*(1), 31–44.

Cheng, Z., & Shen, J. (2014). Just-for-me: An adaptive personalization system for location-aware social music recommendation. In *Proceedings of international conference on multimedia retrieval* (p. 185). ACM.

Cheng, Z., & Shen, J. (2016). On effective location-aware music recommendation. *ACM Transactions on Information Systems (TOIS), 34*(2), 13.

Cunningham, S., Caulder, S., & Grout, V. (2008). Saturday night or fever? context-aware music playlists. *Proceeding audio mostly*.
Deng, S., Wang, D., Li, X., & Xu, G. (2015). Exploring user emotion in microblogs for music recommendation. *Expert Systems with Applications, 42*(23), 9284–9293.
Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National academy of Sciences, 101*(suppl 1), 5228–5235.
Han, B.-j., Rho, S., Jun, S., & Hwang, E. (2010). Music emotion classification and context-based music recommendation. *Multimedia Tools and Applications, 47*(3), 433–460.
Hariri, N., Mobasher, B., & Burke, R. (2012). Using social tags to infer context in hybrid music recommendation. In *Proceedings of the twelfth international workshop on web information and data management* (pp. 41–48). ACM.
Helmholz, P., Vetter, S., & Robra-Bissantz, S. (2014). Ambitune: Bringing context-awareness to music playlists while driving. In *International conference on design science research in information systems* (pp. 393–397). Springer.
Hu, X., Deng, J., Zhao, J., Hu, W., Ngai, E. C.-H., Wang, R., et al. (2015). Safedj: A crowd-cloud codesign approach to situation-aware music delivery for drivers. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 12*(1s), 21.
Hyung, Z., Lee, K., & Lee, K. (2014). Music recommendation using text analysis on song requests to radio stations. *Expert Systems with Applications, 41*(5), 2608–2618.
Kamalzadeh, M., Kralj, C., Möller, T., & Sedlmair, M. (2016). Tagflip: Active mobile music discovery with social tags. In *Proceedings of the 21st international conference on intelligent user interfaces* (pp. 19–30). ACM.
Kaminskas, M., Ricci, F., & Schedl, M. (2013). Location-aware music recommendation using auto-tagging and hybrid matching. In *Proceedings of the 7th acm conference on recommender systems* (pp. 17–24). ACM.
Kim, Y. E., Schmidt, E. M., & Emelle, L. (2008). Moodswings: A collaborative game for music mood label collection.. In *Ismir: 8* (pp. 231–236).
Lamere, P. (2008). Social tagging and music information retrieval. *Journal of New Music Research, 37*(2), 101–114.
Li, Q., Myaeng, S. H., & Kim, B. M. (2007). A probabilistic music recommender considering user opinions and audio features. *Information processing & management, 43*(2), 473–487.
Lu, C.-C., & Tseng, V. S. (2009). A novel method for personalized music recommendation. *Expert Systems with Applications, 36*(6), 10035–10044.
Mayer, R., Neumayer, R., & Rauber, A. (2008). Rhyme and style features for musical genre classification by song lyrics.. In *Ismir* (pp. 337–342).
Moens, B., van Noorden, L., & Leman, M. (2010). D-jogger: Syncing music with walking. *Proceeding SMC*.
Nanopoulos, A., Rafailidis, D., Ruxanda, M. M., & Manolopoulos, Y. (2009). Music search engines: Specifications and challenges. *Information Processing & Management, 45*(3), 392–396.
Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information processing & management, 24*(5), 513–523.
Sasaki, S., Yoshii, K., Nakano, T., Goto, M., & Morishima, S. (2014). Lyricsradar: A lyrics retrieval system based on latent topics of lyrics.. In *Ismir* (pp. 585–590).
Schedl, M., & Knees, P. (2013). Personalization in multimodal music retrieval. In *Adaptive multimedia retrieval. large-scale multimedia retrieval and evaluation* (pp. 58–71). Springer.
Schedl, M., Vall, A., & Farrahi, K. (2014). User geospatial context for music recommendation in microblogs. In *Proceedings of the 37th international acm sigir conference on research & development in information retrieval* (pp. 987–990). ACM.
Schedl, M., Widmer, G., Knees, P., & Pohle, T. (2011). A music information system automatically generated via web content mining techniques. *Information Processing & Management, 47*(3), 426–439.
Schluter, J., & Osendorfer, C. (2011). Music similarity estimation with the mean-covariance restricted boltzmann machine. In *Machine learning and applications and workshops (icmla), 2011 10th international conference on: 2* (pp. 118–123). IEEE.
Shan, M.-K., Kuo, F.-F., Chiang, M.-F., & Lee, S.-Y. (2009). Emotion-based music recommendation by affinity discovery from film music. *Expert Systems with Applications, 36*(4), 7666–7674.
Sordo, M., Oramas, S., & Espinosa-Anke, L. (2015). Extracting relations from unstructured text sources for music recommendation. In *International conference on applications of natural language to information systems* (pp. 369–382). Springer.
Spärck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation, 28*(1), 11–21.
Spärck Jones, K. (2004). Idf term weighting and ir research lessons. *Journal of Documentation, 60*(5), 521–523.
Su, J.-H., Yeh, H.-H., Yu, P. S., & Tseng, V. S. (2010). Music recommendation using content and context information mining. *Intelligent Systems, IEEE, 25*(1), 16–26.
Symeonidis, P., Ruxanda, M. M., Nanopoulos, A., & Manolopoulos, Y. (2008). Ternary semantic analysis of social tags for personalized music recommendation.. In *Ismir: 8* (pp. 219–224).
Turnbull, D., Barrington, L., Torres, D., & Lanckriet, G. (2007). Towards musical query-by-semantic-description using the cal500 data set. In *Proceedings of the 30th annual international acm sigir conference on research and development in information retrieval* (pp. 439–446). ACM.
Turnbull, D., Barrington, L., Torres, D., & Lanckriet, G. (2008). Semantic annotation and retrieval of music and sound effects. *IEEE Transactions on Audio, Speech, and Language Processing, 16*(2), 467–476.
Wang, M., Kawamura, T., Sei, Y., Nakagawa, H., Tahara, Y., & Ohsuga, A. (2013). Context-aware music recommendation with serendipity using semantic relations. In *Joint international semantic technology conference* (pp. 17–32). Springer.
Xing, Z., Wang, X., & Wang, Y. (2014). Enhancing collaborative filtering music recommendation by balancing exploration and exploitation.. In *Ismir* (pp. 445–450).
Yang, Y.-H., & Liu, J.-Y. (2013). Quantitative study of music listening behavior in a social and affective context. *IEEE Transactions on Multimedia, 15*(6), 1304–1315.