Quantitative Study of Music Listening Behavior in a Social and Affective Context

Yi-Hsuan Yang, *Member, IEEE*, and Jen-Yu Liu

***Abstract*—A scientific understanding of emotion experience requires information on the contexts in which the emotion is induced. Moreover, as one of the primary functions of music is to regulate the listener's mood, the individual's short-term music preference may reveal the emotional state of the individual. In light of these observations, this paper presents the first scientific study that exploits the online repository of social data to investigate the connections between a blogger's emotional state, user context manifested in the blog articles, and the content of the music titles the blogger attached to the post. A number of computational models are developed to evaluate the accuracy of different content or context cues in predicting emotional state, using 40,000 pieces of music listening records collected from the social blogging website LiveJournal. Our study shows that it is feasible to computationally model the latent structure underlying music listening and mood regulation. The average area under the receiver operating characteristic curve (AUC) for the content-based and context-based models attains 0.5462 and 0.6851, respectively. The association among user mood, music emotion, and individual's personality is also identified.**

***Index Terms*—Affective computing, music emotion recognition, music listening behavior, social media, user mood recognition.**

I. Introduction

MUSIC is strong in its communicative, expressive, and social functions [1]. Since ancient times, music has been used to modulate the emotional state of an individual, to shape the collective mind of a group of people, and to manage one's self-identity and interpersonal relationships [2]. Although nowadays people have more chances to listen to music with their personal electronic devices alone, the social functions of music still manifest themselves in several ways—from the social tagging of music in websites such as Last.fm, douban.fm and FreeSound,[1] the sharing of custom playlists in Grooveshark,[2] to the instant sharing of music tastes on social network with the service provided by Facebook and Spotify together.[3] It can be argued that every multimedia system that is concerned with the processing or retrieval of music should place the social dimension of music at its core.

Fueled by the tremendous growth of digital music libraries, recent years have witnessed an increasing body of research work on automatic recognition of the affective content of music signals [3]–[5]. Music emotion recognition (MER) is often formulated as a standard classification problem, in which machine learning algorithms are applied to learn the association between emotion labels and features that characterize, for instance, the timbre, tonal and lyrics aspects of music [6]–[12]. However, researchers have begun to address issues specific to MER, such as the subjective nature of emotion perception [13] and the temporal emotion variation as a musical piece unfolds [14]. The ground truth labels for music emotion can be obtained by recruiting human annotators or by harvesting from an online repository of music tags [15]. One may refer to [16], [17] for recent reviews on MER.

While MER focuses on the emotion a musical piece intends to *express* or elicit, user mood recognition (UMR) is concerned with the emotion an individual actually *feels* in response to a stimulus (not necessarily music) [18]–[25].[4] For example, Koelstra *et al.* [19] studied the association between the self-report emotions participants experienced while watching excerpts of music videos and features extracted from the electroencephalogram (EEG) and peripheral physiological signals of the participants. By using functional magnetic resonance imaging (fMRI), Salimpoor *et al.* [22] found that intense pleasure in response to musical stimuli leads to dopamine release in the striatal system. A great deal of research has also been done for recognizing individual emotional state from facial expressions, paralinguistic cues [23], [24], or the articles the individual writes [25], among others.

Although significant progress have been made in MER and UMR to recognize the affective content of music and the emotional state of the listener, little effort has been made to study the two problems under a unified framework. The focus of MER has been on tagging a song with emotion labels that a listener perceives when listening to the song, assuming that the tagging is not biased by the emotional state of the listener [26]. However, in a real-life context, people listen to music with underlying emotional states. People sometimes use music to maintain or intensify their mood, but sometimes to change or alleviate mood [27]–[31]. To create an effective emotion-based music recommendation system, we need to understand the interplay between music emotion and user mood.

Manuscript received September 08, 2012; revised January 31, 2013 and April 11, 2013; accepted April 25, 2013. Date of publication May 29, 2013; date of current version September 13, 2013. This work was supported by a grant from the National Science Council of Taiwan under the contract NSC 101-2221-E-001-017 and the Academia Sinica Career Development Program. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ioannis Kompatsiaris.

The authors are with the Research Center for Information Technology Innovation, Academia Sinica, Taipei 11564, Taiwan (e-mail: yang@citi.sinica.edu.tw; ciaua@citi.sinica.edu.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2013.2265078

[1]http://www.last.fm/, http://www.douban.com/, http://www.freesound.org/

[2]http://grooveshark.com/

[3]http://www.spotify.com/int/about/features/connect-with-facebook/

[4]We use *user mood*, instead of user emotion, as a short term for *individual emotional state* for better differentiation from the term *music emotion*. However, we note that emotion is usually understood as a short experience in response to an object, whereas mood is a longer experience without specific object connection [1]. Therefore, the use of user mood might not be accurate.

The aforementioned study requires a user data set with music listening records, as well as features that characterize the music content and the user context. In addition, as the real-life listening experiences are far from those in a laboratory setting [32], conventional ways of studying self-reports of emotional state or music use are not favorable. In contrast, the user data have to be recorded in a spontaneous and natural fashion, which may have represented a challenge for large-scale study of music listening behavior.

Thanks to the boom in Web 2.0 applications, this challenge can now be circumvented by making use of the social media. It has been a common practice for people to write about their experiences, thoughts, opinions, and feelings through social blogging services such as Twitter.[5] According to Nardi *et al.* [33], "Bloggers are driven to document their lives, provide commentary and opinions, express deeply felt emotions, articulate ideas through writing, and form and maintain community forums." Although bloggers might calibrate what they should and should not reveal because of the awareness of their readers, the vast majority of blogs can be considered as online personal diaries [33]–[36]. In consequence, blogs provide natural and rich information of the everyday context of people. In particular, the social networking website LiveJournal[6] allows a blogger to tag his or her post with a *mood tag* that indicates the blogger's emotional state at posting, and a *music title* that shows which song matches the post well. Although a few research projects have made use of data collected from LiveJournal for MER [8] and UMR [34], [36], little efforts if any have been invested to study the music listening behavior in a variety of real-life contexts kept by LiveJournal.

In this paper, we introduce a novel data set 'LJ40k' that contains 40,000 posts collected from LiveJournal.[7] Each post is labeled with one of the 40 user mood tags adopted by LiveJournal, and is accompanied with a music title whose audio preview can be fetched from the digital media delivery company 7digital [37].[8] To compute the emotion of the LiveJournal tracks, an independent emotion-annotated data set that contains 31,427 tracks is used to train MER models [15]. The ground truth emotion annotation of this 'MER31k' data set is obtained by crawling mood-related tags from Last.fm, with an emotion vocabulary consisting of 190 music emotion tags adopted by the professional music review website AllMusic.[9]

In addition, we collect another set of 751,121 blog posts from LiveJournal (without audio information) for inferring the personal traits of the bloggers from their writing styles [38], [39]. Based on these data sets, we set forth a quantitative study that intends to answer the following novel research questions:

- How well can we predict the emotional state (user mood) of a user from the blog the user posts (user-generated context) or from the music title the user listens to (music listening behavior) in a real-life context?
- How is the association between user mood and the affective content of the selected music title (music emotion)?
- How do the personal traits of a user influence music listening behavior under different user moods?

It should be noted that this work focuses on the emotional state of an individual *prior* to listening to music, instead of the emotion evoked as a result of music listening.

Several content- and context-based models for UMR and MER are built in the course of answering the above questions. A rich set of audio and text features (from both blogs and song lyrics) are extracted to establish the tripartite relationship between music emotion, user mood, and music listening behavior. Moreover, to access the personal traits of the bloggers, the text-based Personality Recognizer [38] is employed. It is hoped that the present study will shed some light on people's music listening behavior and help us develop systems that recommend multimedia content to an individual in a non-intrusive, "mood-optimized" way [31].

The remainder of the paper is organized as follows. First, we briefly review related work in Section II. Next, we provide an overview of this study in Section III. The data sets we created for the study are described in Section IV. Section V presents the computational methods, experimental setup, and the evaluation results. Section VI discusses the results and possible future work. Section VII concludes the paper.

II. Related Work

Music has been widely used for mood and emotion regulation in our daily life, either for negative mood management, positive mood maintenance, or diversion from boredom [27]. According to a psychological study [28], music listening is the second-most used tool for mood regulation, just behind "talking to friends."

The association between personal traits and *long-term* music preference, or taste, has been much studied by psychologists. For example, it has been shown that people prefer styles of music that are consistent with their personalities [40]–[42]. Moreover, it has been found that people who scored high in Neuroticism often use music for affect-regulation and that people who are extravert more often use music as a background for other activities [43]. In another study, Ladinig and Schellenberg [44] found that the personality trait Agreeableness related positively to having more intense emotional responses to music in general, whereas both the personality traits Agreeableness and Neuroticism related to having stronger sad feelings when listening to sad-sounding music excerpts.

In contrast, little has been done to model the *short-term* music preference under different emotional states in a real-life context, mostly because of the difficulty of inducing authentic emotional states in a laboratory and the difficulty of collecting large-scale in-situ data set for such an analysis. Most existing studies in psychology relied on self-reports of music use, which may not translate into actual music use in real life [43]. Moreover, the music data set is usually small and taken only from the classical repertoire [30].

A great amount of effort has been made by computer scientists and psychologists to study the relationship between music and emotion [3]–[5]. It has been argued that emotion perception of music is subjective and dependent on an interplay between the musical, personal and situational factors of music listening

[5]https://twitter.com/

[6]http://www.livejournal.com/

[7]Data set available at http://mac.citi.sinica.edu.tw/lj/

[8]http://www.7digital.com/; http://developer.7digital.net/

[9]http://www.allmusic.com/moods

[1]. [10] For example, people in a sad mood may find sad-sounding music pleasant. Hunter *et al.* [29] also found that sad mood increases the perception of sadness in music when the music is not clearly happy- or sad-sounding. It has also been pointed out that music emotions are in nature different from the everyday emotions people experience in daily life [4]. For example, emotions such as guilt, contempt, and embarrassment are seldom experienced in music, whereas solemnity is more easily experienced with music [5].

Data from LiveJournal has been used for MER by Dang and Shirai [8], who manually mapped the 50 most popular tags of LiveJournal to 5 clusters and performed five-class emotion classification using 6,000 tracks collected from LiveJournal. LiveJournal data has also been used by Leshed and Kaye to study UMR from text [34], using 812,000 posts for training and 10,000 for testing. The problem is formulated as a tagging problem for the 50 most popular mood tags of LiveJournal. Recently, Tang *et al.* [36] studied UMR by using the social network information and blog text. However, to the best of our knowledge, no existing work has used data from LiveJournal to investigate the tripartite relationship between music emotion, user mood, and music listening behavior.

III. Overview of the Study

The present study can be conceptualized by Fig. 1, where we draw the personal, situational and musical factors and the links interconnecting them. The three factors central to this graph are the user mood, user context, and the music the user listens to. The user mood is influenced by the user context and the personal traits; the user context is determined by factors such as the daily experience, social factors (e.g., listening alone or with friends), time and location; the music listening behavior is influenced by the user mood and user context, but is also conditioned on the individual's music taste and the audio, lyrics, and affective content of music.

In particular, our first research question is concerned with the prediction of user mood from the user-selected music (i.e., the music-based model) and user-generated context (i.e., the context-based model). This question is investigated by making use of the LiveJournal data set LJ40k, with the audio and text features extracted from the music signals, lyrics, and blog posts. The second research question studies the association between user mood and music emotion, which are connected by the music listening records. To facilitate this study, we will train audio-based MER models by using MER31k and apply the resultant models to label the tracks in LJ40k. The last research question explores how the personal traits affect the short-term music preference (i.e., music listening behavior under different user moods). Among the relevant personal factors, we consider the Big-Five personality traits [45], as they can be estimated from the style of writing [38]. Although age and gender are also relevant, they cannot be retrieved from LiveJournal due to privacy concerns.

[10]Personal factors include demographic variables (e.g., gender, age, education), physical state (e.g., feeling well, rested, tired, ill), cognitive factors (e.g., expectations, familiarity, attentiveness, associated memory), emotional state (mood), and personality. Situational factors include physical factors (e.g., location, time, weather, acoustical conditions, live or recorded), social factors (e.g., listening alone or with others), and special occasions, among others [1].

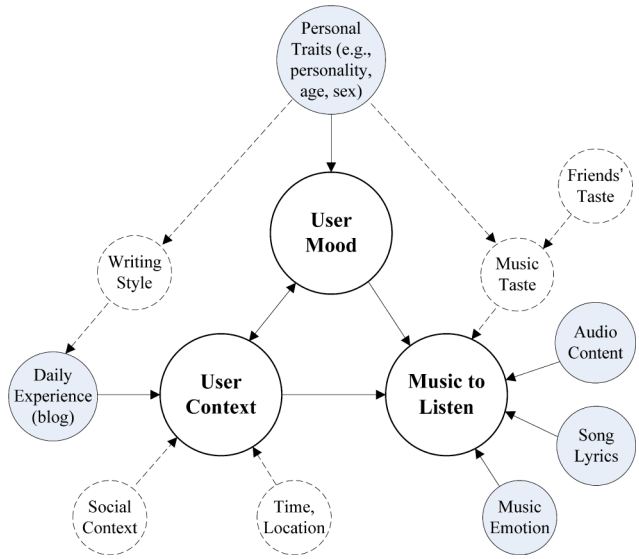

Fig. 1. A graphical model of the musical, personal and situational factors considered in this study. The shaded nodes represent the observed data in our study, whereas the dashed ones are not considered and left as future work.

The factors represented by dashed nodes in Fig. 1 are not considered in this work for the following reasons. First, although it is possible to estimate an individual's music taste from the listening records, such estimation might not be robust as we can obtain at most 25 listening records of a LiveJournal user through its Rich Site Summary (RSS) service. The writing style has been implicitly taken into account with the use of the Personality Recognizer [38], [39]. The time of the post can be retrieved, but we opt for focusing on the daily experience described in the posts in this work. Finally, although the social network of a user can be obtained by the technique described in [36], we leave this as a future work due to the additional complexity in analyzing social network information.

IV. Data Sets

As summarized in Table I, three data sets are used in this work. The first two contain blog posts collected from LiveJournal, which has 42 million registered users and more than 2.1 million active users (54.9% female).[11] One distinguished feature of LJ is that a user can also attach a mood tag and a music title to each blog post, as Fig. 2 displays. The mood tag can either be chosen from a predefined list with 132 tags[12] or be typed freely. The music titles are completely freely-entered song titles and (usually) artist names, so some of them might not correspond to real songs. For posts with valid mood tags and music titles, a $<$ text, mood, music $>$ triple can be obtained.

The first data set, LJ40k, is a subset of the 21 million posts Leshed and Kaye collected in 2005 for their study on text-based UMR [34]. The bloggers in this set are mostly from the United States, and the publication year of these posts fall between 2000 and 2005.

As 132 mood classes are too many for a quantitative study, we considered only the 40 most popular ones. Moreover, we kept

[11]http://www.livejournal.com/stats.bml

[12]http://www.livejournal.com/moodlist.bml

TABLE I
STATISTICS OF THE DATA SETS.

Data set	Content	Metadata	Density	Unique terms	Source
LJ40k	40,000 posts & tracks	User mood labels for 40 classes	1,000 posts & tracks per class	17,849 (for posts)	LiveJournal
LJ750k	751,121 posts	User names for 34,051 bloggers	22.06 posts per user	—	LiveJournal
MER31k	31,427 tracks	Music emotion labels for 190 classes	165.38 tracks per class	—	Last.fm

Fig. 2. Two LiveJournal blog entries posted by the same blogger, with mood tags and music titles attached.

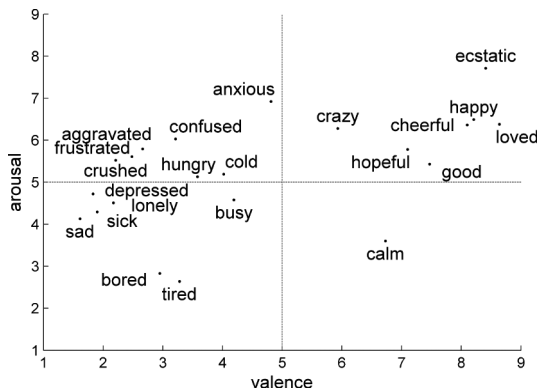

Fig. 3. Valence and arousal values of 22 LiveJournal mood tags according to the Affective Norm for English Words (ANEW) [47].

only the posts whose audio preview can be fetched by the 7digital API. Although the previews are typically 30-second snippets (instead of whole song) sampled from the middle of the songs, state-of-the-art audio feature extraction algorithms normally deal with this [6], [37]. We finally collected 1,000 posts with valid music titles for each mood, giving rise to 40,000 posts in total.

Fig. 3 depicts 22 of the 40 mood tags in an emotion space spanned by *valence* (positive or negative affective states) and *arousal* (energy and stimulation level) — the two fundamental emotion dimensions [46][13]—by referring to the Affective Norm for English Words (ANEW) [47] developed by psychologists. Although only half of the mood classes have an entry in ANEW, it can be found that the mood tags are fairly diverse. In Section V-D, ANEW will be further used to analyze the association between user mood and music emotion.

There are 34,051 unique bloggers in LJ40k. Based on the user-IDs, the second data set, LJ750k, is created by searching for the articles of the bloggers. By using the RSS service at the page of the LJ users,[14] at most 25 entries can be crawled. We obtained a total number of 751,121 posts, with on average 22.06 posts per user. Interestingly, although 39,361 unique user mood tags are observed in LJ750k, 73.38% of the posts use the 132 predefined ones and 49.56% of them use the 40 tags employed in LJ40k. Instead of filtering the posts according to mood tags or the validity of music titles, the whole data set is used as input to the Personality Recognizer [38]. [15]

The third dataset, MER31k, contains 31,427 tracks labeled with music emotion tags defined by AllMusic. We collected our song list from Last.fm, which provides a social platform for online users to tag music. For each of the 190 emotion tags, a song list was obtained by searching for the top songs labeled with that tag in Last.fm, and the audio previews were downloaded from 7digital. On average, we have 165.38 songs for each emotion tag.[16] Among the emotion tags, 43 of them can be visualized in the emotion space, as shown in Fig. 7. MER31k will be used to train MER models to label the music titles contained in LJ40k with music emotion.[17]

V. EXPERIMENTS AND RESULTS

1) Problem 1: How well can we predict the emotional state (user mood) of a user from the blog the user posts (user-generated context) or from the music title the user listens to (music listening behavior) in a real-life context?

A. Audio and Lyrics-Based User Mood Recognition

We quantitatively evaluate the association between user mood and music listening behavior by evaluating music-based UMR on LJ40k. The user mood tags are considered as ground truth labels for the corresponding music titles, which are characterized by audio and text features extracted from the music signals and lyrics, respectively. We formulate the problem as a tagging problem and build a binary classifier for each user mood. The output of the mood classifier indicates which songs are adequate for people in that particular emotional state.

1) Feature Extraction: In view of the complexity of music, we consider audio features that represent various perceptual dimensions of music listening, covering the energy, rhythm,

[13]To identify the internal human representations of emotion, psychologists have used factor analysis techniques to analyze the correlation between affective rating scales and found the two most important factors usually correspond to valence and arousal [46]. For example, 'cheerful' is positive and high in arousal, whereas 'sad' is negative and low in arousal.

[14]http://⟨userid⟩.livejournal.com/data/rss

[15]Because LJ750k is directly processed by Personality Recognizer [38], we do not count the number of unique terms of this data set for Table I.

[16]With standard deviation 40.41, maximum 248, and minimum 28.

[17]Although it is possible to query Last.fm with the 40 mood tags to obtain a shared vocabulary for UMR and MER, we found the idea difficult to implement due to the discrepancy between music emotion and user mood [4]. For example, with the mood tag 'annoyed' we retrieved only 5 songs, whereas with the modified version 'amusing' we obtained 245. However, it is unclear whether such modification is appropriate. Moreover, for mood tags such as 'aggravated,' 'exhausted,' and 'drained,' we found no response from Last.fm, and it is not easy to find proper alternative terms. This leads us to use the terms specifically designed for music emotion from AllMusic.

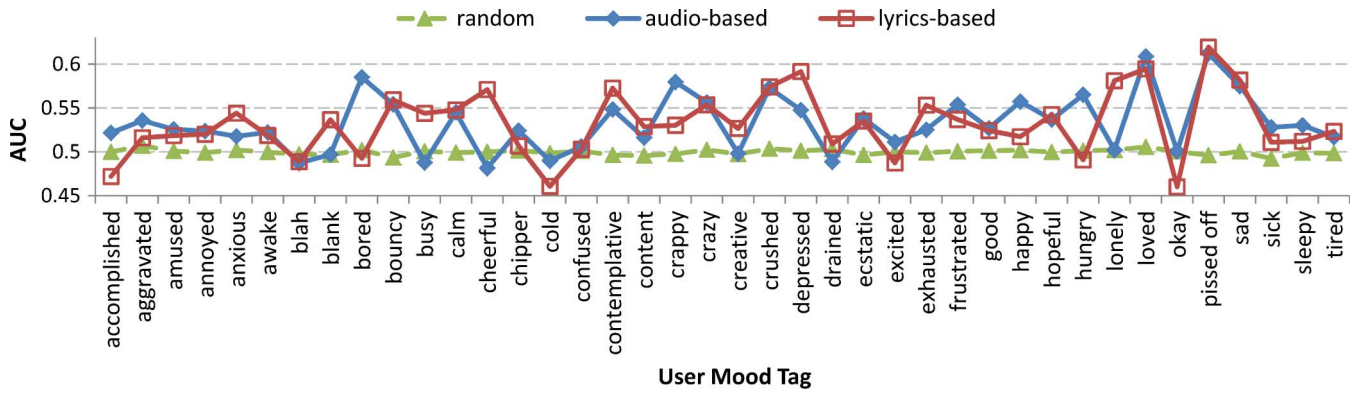

Fig. 4. The per-class AUCs of UMR for the audio-based, lyrics-based models (both after late-fusion), and a random baseline.

TABLE II
EXTRACTED FEATURES.

Aspect	Extracted features
Energy	Perceptual loudness, volume, and sound pressure level [48].
Rhythm	Fluctuation pattern, tempo, and event density [49].
Timbre	Perceptual sharpness, timbre width, dissonance [48] and spectral centroid, brightness, spread, skewness, kurtosis, roughness, flatness, irregularity, flux, rolloff, spectentropy, low energy ratio, time-domain zero-crossing rate, and Mel-frequency cepstral coefficients (MFCCs) [49].
Tonal	Pitch class profile, musical key, key clarity, musical mode, chord change likelihood [49], tonalness, multiplicity [48].
Text (lyrics)	Tf-idf weighted bag-of-words (uni-gram) features [51], and PLSA probability distribution over latent topics [54].

TABLE III
VARIANTS OF TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY.

Abbr.	Formulation	Abbr.	Formulation
TF-1	$1+\log_2 f_{d,t}$	IDF-1	$\ln \frac{\lvert\mathcal{D}\rvert}{F_t}$
TF-2	$\frac{f_{d,t}}{f_{d,t}+\kappa\lvert d\rvert/\lvert\Delta d\rvert}$	IDF-2	$(\max_{t'\in\mathcal{T}} n_{t'}) - n_t$
TF-3	$\frac{(\kappa+1)f_{d,t}}{f_{d,t}+\kappa\left((1-b)+b\cdot\frac{\lvert d\rvert}{\lvert\Delta d\rvert}\right)}$	IDF-3	$1-\frac{n_t}{\ln\lvert\mathcal{D}\rvert}$

$f_{d,t}$ — the number of occurrences of term t in document d
F_t — the number of documents in $\mathcal{D}$ that contain t
$\lvert d\rvert$ — cardinality (length) of d; $\lvert\Delta d\rvert$ — average document length in $\mathcal{D}$
n_t — entropy of term t in $\mathcal{D}$; $n_t = -\sum_d (f_{d,t}/F_t)\ln(f_{d,t}/F_t)$
$\mathcal{T}$ — the universe of terms; $\mathcal{D}$ — the universe of documents

timbre, and tonal aspects of music, as listed in Table II. The features were extracted using the PsySound toolbox [48] and the MIRtoolbox [49].

Most of the features have been used extensively in the music information retrieval field [6], [16]. The details of the features are provided in the appendix. The lyrics of 38,325 tracks can be obtained by making use of the LyricWiki API.[18] We generated the bag-of-words features by counting the occurrence of terms (uni-grams) in the lyrics, with stemming and stopword removal [50].[19] After removing terms that appear less than ten times, we obtain a vocabulary of 11,035 terms. The bag-of-words features are weighted by the term-frequency inverse-document-frequency (tf-idf) measure [51], which helps enhance the significance of terms that have high weight and occur sparsely in the whole corpus. As shown in Table III, we select and compare the performance of three tf and three idf functions, resulting in nine possible combinations. Among them, TF-1, TF-2 are the Okapi formulations widely used in document ranking [52], whereas IDF-2, IDF-3 are information-theoretic measures that reduce the importance of noisy (high entropy) terms [53]. In addition, we apply probabilistic latent topic analysis (PLSA) [54] to compute the probability distribution of each song over 10, 20, 50, and 100 latent topics, using a tempered expectation-maximization (EM) algorithm with at most 100 EM iterations [54]. PLSA reduces the dimension of the feature space and increases the overlapping of semantic terms.

[18]http://api.wikia.com/wiki/LyricWiki_API

[19]We have tested the result without stemming and stopword removal and found the result slightly worse. For example, the AUCs of the lyrics- and context-based models drop to 0.5292 and 0.6845 (after late fusion).

Note that the same set of text features are generated from blog posts for building the context-based UMR (cf. Section V-B). The blog articles contain 17,849 unique terms.

2) Classification Method: We randomly held out 6,000 instances of LJ40k as the validation set, 6,000 for testing, and the remaining 28,000 for training. We ensured that, as a result of the partition, the validation and test sets do not contain blogs from any blogger in the training set. We used the training set to train UMR classifiers, the validation set for parameter tuning, and the test set for performance evaluation. As for classifier training, we adopted the linear-kernel support vector machine (SVM) implemented by the library LIBLINEAR [55], for linear kernel is more efficient. We trained a SVM for each feature type (e.g., four types of audio descriptor) and then fused the decision values from the multiple classifiers with equal weights for final prediction. The decision values are normalized to $[0, 1]$ by a sigmoid function before fusion.[20]

On average, for each mood class, the training set contains 700 positive examples and 27,300 negative examples. As the negative examples greatly outnumber the positive ones, we adopted the EasyEnsemble technique [15], [56] to counteract the data imbalance bias. The technique generates an ensemble of T component classifiers, each of which is trained (using SVM) with a balanced set consisting of all the positive data and a random subset of the negative ones. This way reduces the risk of discarding potentially useful data in the under-sampling process.

[20]Specifically, the L2-regularized L2-loss SVM solver was used, and the value of the cost parameter C was determined from $\{10^{-4}, 10^{-3}, \ldots, 10^{1}\}$ via the validation set. The audio features are standardized (i.e., converted to z-scores), whereas the text (lyrics) features are not normalized.

The output of the component classifiers are combined by averaging. We set T to 100, considering that the ratio of the number of negative to positive data is around 39.

We evaluated the accuracy of UMR using the standard information retrieval measures area under the receiver operating characteristic curve (AUC) [57] and the normalized discounted cumulative gain (NDCG) [58]. For each class, we ranked the test instances according to the averaged SVM decision values. AUC measures the ability of a retrieval system to rank positive examples above negative examples, scoring them on a scale from 0.5 (chance level) to 1 (perfect ranking). NDCG summarizes the relevance of a ranking order with the gain of each result discounted at lower ranks. Specifically,

$$\mathrm{NDCG} = \frac{1}{Q}\left(r_1 + \sum_{j=2}^{N} \frac{r_j}{\log_2 j}\right), \tag{1}$$

where r_j is the ground truth relevance of the j-th piece on the ranked list, N =6,000 is the number of test instances, and Q is the normalization term that makes the ideal NDCG equal 1. Note that in the test set there are only 150 relevant (positive) instances on average. We evaluated the accuracy on a per-class basis and then averaged the result across the 40 classes. We will use AUC as the major performance measure.

3) Results: Table IV shows the result of audio-based and lyrics-based UMR using different features, along with a random baseline that randomly assigns tags to tracks. For audio features, the best result is obtained by the timbre descriptors, whereas a late fusion over the four types of descriptor further improves the average AUC and NDCG to 0.5325 $(\mathrm{STD} = 0.0327)$ and 0.4854, respectively. For the bag-of-words lyrics features, we found that the result is more sensitive to the formulation of the idf term (though not shown here); the entropy measure IDF-2 performs the best among others. As Table IV shows, the average AUC 0.5242 obtained by TF-1 $\times$ IDF-2 is close to that attained by the timbre descriptors. In addition, we observed that PLSA greatly reduces the feature dimension without much degradation in either AUC or NDCG. Fusing the six lyrics features improves the average AUC and NDCG to 0.5314 $(\mathrm{STD} = 0.0366)$ and 0.4889, respectively. No significant difference between the audio- and lyrics-based models is observed (p-value $= 0.6727$ for AUC).

To gain more insights, we show the per-class AUC of the two modalities (both after late-fusion) in Fig. 4, from which three observations can be made. First, both models outperform the random baseline by a great margin for about half of the classes. By employing one-tailed t-test, we found that both models are significantly better than the random baseline in both AUC and NDCG (p-value < 0.001, d.f. $= 78$) [59].

Second, the performance of the two models is clearly class-dependent. Comparing the two modalities, the audio-based one performs relatively well for some *valence-neutral mood*, such as 'bored' and 'hungry.' In contrast, the lyrics-based model is better for some *extremely negative or positive mood with low-arousal*, such as 'cheerful,' 'depressed,' and 'lonely.' Interestingly, in MER studies, it has been well-known that audio features perform better for arousal recognition, whereas lyrics features are more informative for valence recognition [11], [16]. We note both models perform well for 'crushed,' 'loved,' and 'sad,' and both perform the best for 'pissed-off,' achieving AUC around 0.6120. If we consider only the top-ten performing mood classes for audio- and lyrics-based models respectively, the average AUCs are 0.5766 and 0.5799.

Third, it can be seen that both models are poor at neutral mood classes such as 'blah' and 'okay,' possibly because people in such mood tend to listen to music randomly or according to their long-term music taste. Similar performance tendency can also be seen from the per-class NCDGs.

TABLE IV
AVERAGE AUC AND NCDG FOR USER MOOD RECOGNITION (UMR).

Modality		Feature	Dim	AUC	NDCG
Random		—	—	0.4998	0.4791
Music	Audio	Energy	24	0.5162	0.4827
		Rhythm	5	0.5161	0.4849
		Timbre	176	0.5270	0.4851
		Tonal	38	0.5238	0.4868
		Late fusion	—	0.5325	0.4854
	Lyrics	TF-1×IDF-2	11,035	0.5242	0.4891
		TF-3×IDF-2	11,035	0.5239	0.4889
		PLSA (10)	10	0.5264	0.4853
		PLSA (20)	20	0.5238	0.4846
		PLSA (50)	50	0.5238	0.4867
		PLSA (100)	100	0.5182	0.4840
		Late fusion	—	0.5314	0.4889
	MER	—	190	0.5381	0.4887
	MER+lyrics		—	0.5462	0.4947
Context	Text	TF-1×IDF-2	17,849	0.6846	0.5554
		TF-3×IDF-2	17,849	0.6923	0.5696
		PLSA (10)	10	0.6397	0.5257
		PLSA (20)	20	0.6514	0.5310
		PLSA (50)	50	0.6598	0.5380
		PLSA (100)	100	0.6608	0.5420
		Late fusion	—	0.6851	0.5585

B. Context-Based User Mood Recognition

To offer a qualitative comparison between music-based UMR and context-based UMR, we extracted text features for the blog posts of LJ40k in the same way as for the song lyrics, and trained binary classifiers for each user mood class. As Table IV shows, the context-based model greatly outperforms the music-based ones. The average AUC after late fusion achieves 0.6851 $(\mathrm{STD} = 0.0718)$, which is significantly better than any music-based model (p-value < 0.001, d.f. $= 78$). This result shows that blog posts contain more direct cues for user mood.[21] Moreover, Fig. 5 shows that the context-based model consistently outperforms the audio-based one in almost all classes. However, we note that the context-based model also performs poorly for the neutral or low-arousal mood such as 'blank,' 'calm,' 'drained,' and 'okay,' for example.

We hypothesize that the worse performance of music-based models can be attributed to the following three factors. First, the audio features we extracted are not accurate enough to represent the content of the music signals, due to the so-called "semantic gap" between low-level signal features and high-level human perception [6]. Second, the accuracy of a tagging system is usually underestimated due to the so-called "weak labeling" problem in the ground-truth annotation [60]; many instances

[21]We note that the performance of PLSA features is much worse than the bag-of-words features for blogs, possibly because blogs contain more direct linguistic cues in describing affect, comparing to lyrics. In addition, late fusion does not improve the result of the context-based model.

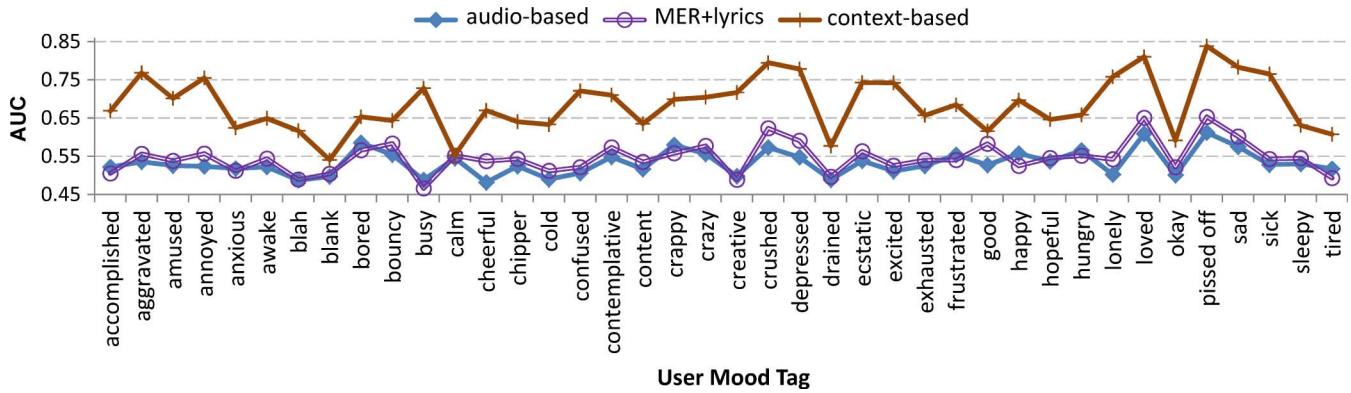

Fig. 5. The per-class AUCs of UMR (after late-fusion) for the audio-based model, the fusion of MER- and lyrics-based models, and the context-based one.

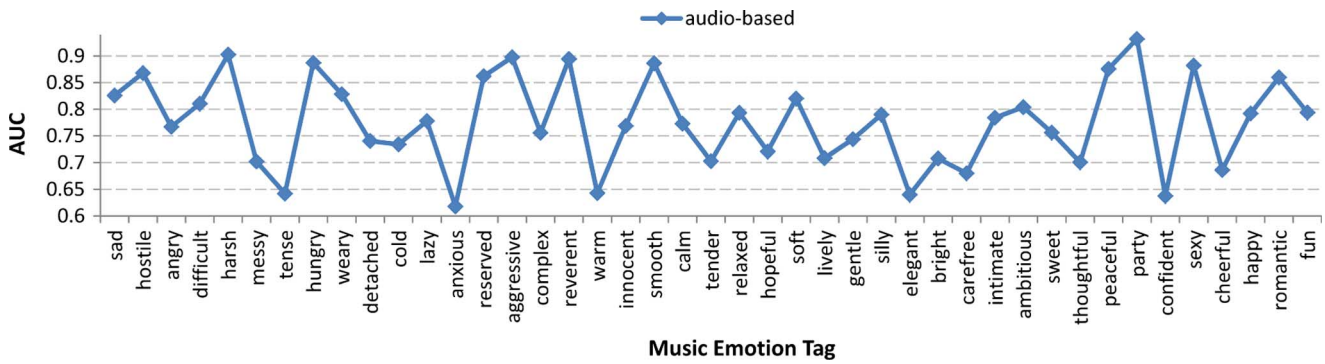

Fig. 6. The per-class AUCs of MER (after late-fusion) for the audio-based model for the 43 music emotion tags from AMG.

have been incompletely tagged, and the absence of a tag does not necessarily mean that the tag does not apply to the instance. This is particularly true for LJ40k, as a blogger can only select one music title to match a post, while there could be dozens of adequate alternatives. Finally, and maybe most importantly, in addition to user mood, the selection of music titles might depend on many other factors such as long-term music taste, time of music listening, and cognitive needs, among others. Even in the same mood, people can listen to virtually any type of songs. There is not always a causal relationship between user mood and music choice. In contrast, the user mood is mostly affected by the context.

In summary, our evaluations show that the user mood is more easily estimated from the blog articles than from the music listening records. The audio- and lyrics-based models can be used to estimate the user mood for about ten mood classes with AUCs close to 0.6, such as 'loved' and 'pissed-off.' In contrast, the context-based model is accurate for almost all classes except for some neutral user mood classes.

1) Problem 2: How is the association between user mood and the affective content of the selected music title (music emotion)?

C. Mer-Based User Mood Recognition

As LJ40k comes with no music emotion tags, we used MER31k to train audio-based MER models. Specifically, we considered MER as a tagging problem and trained audio-based classifier for each of the 190 music emotion classes. To evaluate the accuracy of MER, we randomly held out 6,000 tracks of MER31k for validation, 6,000 for testing, and the remaining 19,427 tracks for training. We trained an ensemble of $T = 50$ component classifiers for each of the four types of audio descriptor, and then performed late fusion by taking the average. For MER, it has been reported that nonlinear kernels such as radial-basis function (RBF) generally performs better [12]. Therefore, the RBF-kernel SVM implemented by the library LIBSVM [61] is adopted.[22]

Table V indicates that the average AUC across the 190 emotion classes attains 0.7614 (STD $= 0.0932$) after late-fusion, which is higher than the average AUC of the previous context-based UMR model. It seems that the association between music emotion and audio features is stronger than that between user mood and text features extracted from blog articles. Fig. 6 shows the AUCs, after late fusion, for the 43 music emotion tags whose valence and arousal values can be found in ANEW. We note that the top-performing classes are mostly high in arousal, such as 'party,' 'harsh,' 'hostile,' and 'sexy.' We also see decent result for low-arousal ones such as 'peaceful,' 'reverent,' 'sad,' and 'smooth.'

We applied the resultant MER models to predict the emotion of the music titles in LJ40k. Instead of assigning binary labels, we used the Platt scaling [61] to compute probability estimates (scores) of class membership, thereby representing a track by a 190-D vector consisting of probability estimates. Based on this feature, we trained an MER-based UMR model with the same setting as the previous UMR models. As shown in Table IV, this MER-based model obtained 0.5381 STD $= 0.0368$) in average AUC, which is better than the audio-based model.

[22]The RBF kernel can be expressed as $\exp(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|_2^2)$, where $\mathbf{x} \in \mathbb{R}^m$ is an input feature vector and γ is a kernel parameter. We optimized C from $\{1, 10, 10^2, 10^3\}$ and γ from $\{1/m, 0.1/m, 0.01/m\}$ via the validation set.

TABLE V
AVERAGE AUC AND NCDG FOR MUSIC EMOTION RECOGNITION (MER).

Modality		Feature	Dim	AUC	NDCG
Random		—	—	0.4999	0.2806
Music	Audio	Energy	24	0.6969	0.3292
		Rhythm	5	0.6229	0.3049
		Timbre	176	0.7432	0.3789
		Tonal	38	0.7060	0.3374
		Late fusion	—	0.7614	0.3913

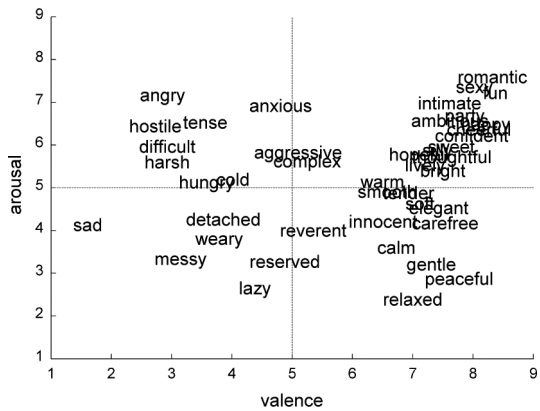

Fig. 7. Valence and arousal values of 43 AMG emotion tags according to the Affective Norm for English Words (ANEW) [47].

However, the performance of MER- and audio-based models does not differ a lot in most classes, possibly because the MER-based model is also derived from audio information. As the audio and lyrics information can be complementary, we further fused the MER- and lyrics-based UMR models by averaging the prediction result and saw the average AUC improved to $0.5462\ (\mathrm{STD} = 0.0410)$, as shown in Table IV and Fig. 5 (denoted 'MER+lyrics').[23] The performance difference between the fused and the audio-only model is, however, not significant ($p - \mathrm{value} = 0.1027, \mathrm{d.f.} = 78$).

Although the MER-based model contributes positively to the accuracy of UMR, the accuracy of music-based models is still inferior to its context-based rival. However, as the same set of audio (text) features performs well for audio-based MER (resp. context-based UMR), we cannot ascribe the low accuracy to the audio (text) features.

D. Association Between User Mood and Music Emotion

In addition to evaluating MER-based UMR, we also studied the correlation between the MER probability scores and the ground truth user mood labels to gain more insights. Specifically, we considered the 43 music emotion tags and 22 user mood tags that are used in ANEW [47] and computed the Pearson's correlation coefficient between every pair of music emotion and user mood. This way, we can not only identify the emotion of music people prefer in a given mood, but also study whether the relevant emotion classes are mood-congruent to the given mood in terms of valence and arousal values. According to the correlation coefficients, we applied Gaussian process regression (GPR) [62] to compute the likelihood for people to listen to every emotion (represented as a point in the emotion space) for a given mood [3].[24]

The result is shown in Fig. 8, in which we sorted the user mood tags according to valence values from left (negative) to right (positive) and attached the top-ten music emotion classes for each mood.[25] Lighter area or larger font size indicates more positive correlation. The following patterns can be found (please see Figs. 3 and 7 for the valence and arousal values):

- when being in a *negative mood* ('sick,' 'depressed,' 'sad,' 'crushed,' 'confused,' and 'lonely'), some people listen to mood-congruent songs such as sad songs, while others listen to incongruent ones such as sweet and tender songs so as to change mood.
- when being in a *low morale* ('frustrated,' 'bored,' 'tired,' and 'aggravated'), people tend to listen to angry music to change their mood.
- when feeling *crazy*, some people listen to party music, while others listen to angry music.
- people prefer smooth, peaceful, and calm music when the mood is *less intense* ('good,' 'calm,' and 'anxious').
- people listen to romantic or sweet music when the mood is *about love* ('loved,' 'hopeful,' and 'cold').
- when being in a *positive and high-arousal mood* ('cheerful,' 'ecstatic,' 'happy,' and 'busy'), most people listen to happy, fun, or party-like music.

Generally, people listen to mood-congruent music when being in a positive mood (i.e., the last three patterns), but tend to listen to mood-incongruent music when being in a negative mood (i.e., the first two patterns). Such tendency is in line with the intuition that people prefer "feeling good" [28]. In addition, because of the "misery loves company" effect [29], people also enjoy listening to sad music when feeling bad.[26]

To sum up, our evaluations identify a number of patterns between user mood and music emotion. Being able to recognize such patterns is encouraging, because they are mined from a large-scale, real-life, and possibly noisy data set LJ40k in a data-driven way, and because the music emotion labels of LJ40k are automatically predicted rather than manually entered. In addition, we found that it is possible to infer user mood from the emotion of the music people listen to, though context-based model still performs better.

1) Problem 3: How do the personal traits of a user influence music listening behavior under different user moods?

[23]Further fusing audio-based UMR model with the MER- and lyrics-based ones does not offer gain; the average AUC attains 0.5449. Moreover, fusing any of the music-based models with the context-based one sees no advantage, possibly because the context information explains most of the variance, in the user mood data, that can be accounted for by the other modalities.

[24]Specifically, we used the GPML (Gaussian process for machine learning) toolkit and selected the isotropic rational quadratic covariance kernel [62]. GPR is attractive mainly because of its flexible non-parametric nature.

[25]We do not show the result for 'lonely' in Fig. 8 due to space limit. Its result is similar to the that of 'sad.'

[26]As the correlation coefficients only provide an isolated (or direct) measure of the association between music emotion and user mood, we have also considered the "beta weights," which are essentially the weight for each music emotion in the linear SVM model for each user mood [63]. Beta weights can be used to quantify the contribution of each music emotion to the prediction model when the variance contributions of all other emotion classes have been accounted for [63]. We found similar patterns from the beta weights.

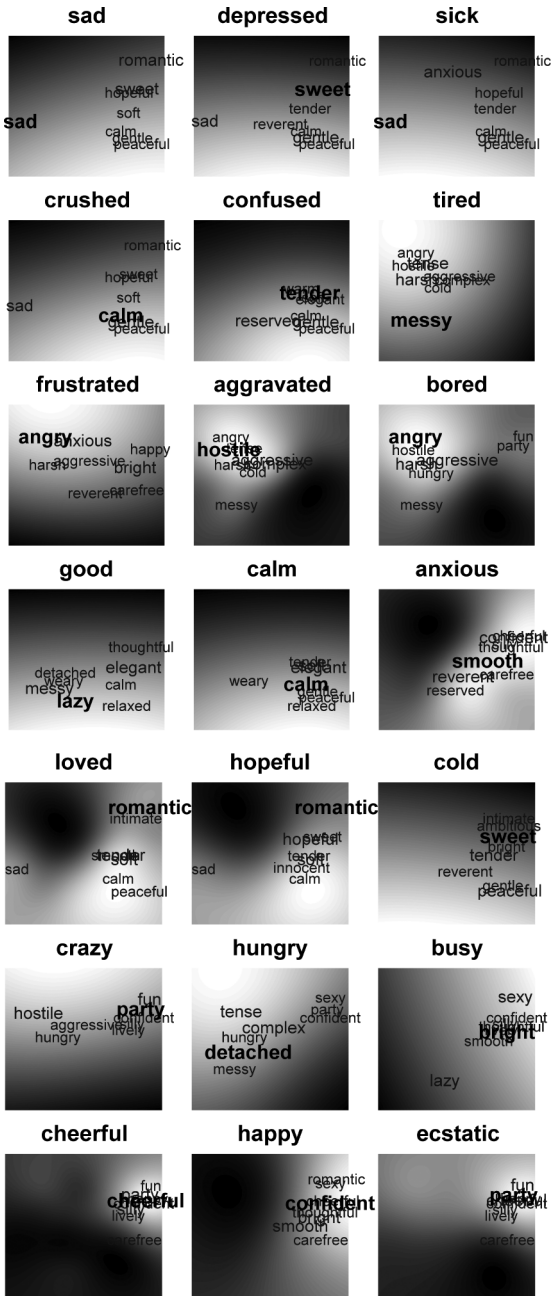

Fig. 8. Visualization (in the valence-arousal emotion space [46]) of the music emotion tags whose correlation with one of the 21 user mood tags (shown in the title of each sub-figure) is among the highest in the listening records of LJ40k. Lighter area or larger font size indicates more positive correlation between the corresponding music emotion and the given user mood.

E. Personality, Mood and Short-Term Music Preference

As described in Section III, we focus on the personality aspect of individuals for it can be automatically estimated with Personality Recognizer [38], a toolkit developed by Mairesse *et al.* based on previously researched correlations between Big-Five personality traits (see below for details) [45] and people's writing styles.[27] Several linguistic features are employed by Personality Recognizer, including the frequency counts of 88 word categories (e.g., Friends, Family, Money, Cognitive processes, negation terms, swear words, etc.) from Linguistic Inquiry & Word Count (LIWC) [64] and 14 features derived from MRC (Medical Research Council) Psycholinguistic database [65]. Therefore, although Personality Recognizer may not be always accurate, it exploits most linguistic cues that have been found correlated with personality [39]. As LJ40k and LJ750k are from the same pool of bloggers, we estimated personality from the larger set LJ750k and then applied the result to LJ40k for better accuracy. For a blogger with multiple posts, the average estimate from all his/her posts is adopted. The output of Personality Recognizer falls within 1 (low) and 7 (high) for each of the personality traits.

The Big-Five traits (or "OCEAN") are Openness (creative vs. caution), Conscientiousness (organized vs. careless), Extraversion (outgoing vs. solitary), Agreeableness (compassionate vs. unkind), and Neuroticism (sensitive vs. confident). They were identified by psychologists in the 1990s by performing factor analysis on self-report trait ratings [45].

We performed the following statistical test independently for each trait. From the listening records associated with a certain user mood, we partitioned the records into two sets according to the personality estimate in that trait: one set contains records of people whose scores in that personality trait are top 10% highest (or lowest), and the other set contains the remaining. We tested whether the music titles in one set are significantly different from those in the other with respect to a music emotion. Specifically, we used two-sample, one-tailed t-test for the hypothesis that, for instance, *when being in a sad mood, people who scored high on the Extraversion trait listen to more happy songs than people who exhibited Extraversion close to the average*, assuming unknown but equal variances for the records in the two sets [59]. Here the two sets contain 100 and 900 records, respectively.

In this study, we focus on the following four music emotion classes, 'sad,' 'angry,' 'peaceful,' and 'party,' for 1) they are prototypical emotions that represent each of the four quadrants of the emotion space (cf. Fig. 7), and 2) they are easier to be automatically predicted; the AUCs are 0.8258, 0.7670, 0.8759, 0.9320, respectively (cf. Fig. 6).

Table VI shows the $<$ user personality, user mood, music emotion $>$ triples for which the null hypotheses (i.e., one is not greater than the other) can be rejected at the 0.1% significance level $(\mathrm{d.f.} = 998)$. A total number of 18 triples are found, most of which are related to the Extraversion and Agreeableness traits.[28] For instance, people who scored high in Extroversion listen to more party-like music when feeling 'loved' or 'sick,' while Fig. 8 shows that people in these two moods can listen to music of a variety of emotions. We also see that people who scored low in Extroversion listen to less angry but more peaceful music when feeling 'bored' and 'drained'; people who scored low in Agreeableness listen to less sad but more angry music when feeling 'chipper' and 'lonely'; and people who scored low in Conscientiousness prefer angry music over sad ones when feeling 'depressed.'

[27]For example, people high in the personality trait Extraversion tend to use first person plural pronouns, second person pronoun, and mention friends, music, physical states, and sexuality, etc., in their articles, whereas people high in Conscientiousness tend to mention things about achievement, but talk less about negative emotions, death, or swear words [39].

[28]There are 3,200 possible triples (5×2) personality traits (high/low), 40 user mood classes, 4×2 music emotion preference (more/less). We got 93 and 295 triples if the significance level was set to 1% and 5%, respectively.

TABLE VI
PREFERENCE OF CERTAIN EMOTION-SOUNDING MUSIC GIVEN A USER MOOD FOR PEOPLE HIGH/LOW IN A SPECIFIC PERSONALITY TRAIT

Personality	User mood	Music emotion
High in Extraversion	loved	more party music
High in Extraversion	sick	more party music
Low in Extraversion	anxious	less angry music
Low in Extraversion	awake	less angry music
Low in Extraversion	blank	less party music
Low in Extraversion	bored	less angry music
Low in Extraversion	bored	more peaceful music
Low in Extraversion	cheerful	less angry music
Low in Extraversion	drained	less angry music
Low in Extraversion	drained	more peaceful music
Low in Agreeableness	chipper	less sad music
Low in Agreeableness	chipper	more angry music
Low in Agreeableness	cold	less sad music
Low in Agreeableness	frustrated	less sad music
Low in Agreeableness	lonely	less sad music
Low in Agreeableness	lonely	more angry music
Low in Conscientiousness	depressed	less sad music
Low in Conscientiousness	depressed	more angry music

Interestingly, in psychology it has been found that people who scored high in Extroversion prefer happy music [41], [42], people who scored high in Neuroticism prefer sad music and tend to use music for mood regulation [43], and that the trait Agreeableness relates positively to having intense emotional responses to music [44]. Our result matches well with existing psychological studies except for less evidence for the Neuroticism trait.

In sum, our result indicates that people with different personalities prefer different music even when being in the same mood. Our empirical findings from the in-situ LJ40k data set appears to be in concert with those reported in psychological studies conducted in laboratory settings.

VI. DISCUSSION

In the process of our study, we have read some posts in order to subjectively evaluate the consistency among articles, mood tags, and music titles. We found it relatively easy to find good-matching posts; for example, posts labeled with 'happy' mood tag usually describe good events like birthday parties, hanging out with friends, etc., while posts labeled with 'sad' usually describe unfavorable events like problems with girl/boyfriends, death, etc. However, it is not trivial to find ill-matching posts because 1) some mood tags are ambiguous in nature, such as 'okay,' and 2) for us who are not in the same context of the blogger, we cannot fully understand the events and the metaphors used in the post. Instead of attempting to remove possibly ill-matching posts, we argue that it is important to investigate the real-life music listening behavior manifested in every record of the LiveJournal repository.

Although the LJ40k data set can be noisy, we have identified interesting patterns among user mood, music emotion, and individual's personality in our study of the second and third research problems (cf. Sections V-D and V-E). Depending on user mood and personality, people have different preferences of mood-congruent or incongruent music. The findings are consistent with those derived from psychological studies.

As the music emotion labels were predicted by audio-based MER models, it is fair to say that there is also a certain association between user mood and the audio content of the music pieces. In the study of either audio-based UMR or audio-based MER models, fusing energy, rhythm, timbre, and tonal features always improves the prediction accuracy, showing that music perception is indeed multi-faced.

However, in addressing the first research problem, we found that it is much easier to predict user mood from the user context manifested in the posts, rather than from the audio content or lyrics of the music pieces the bloggers attached to the posts (cf. Sections V-A to V-C). Audio- and lyrics-based models only perform well for a limited number of user mood classes. Considering that the same set of audio and text features perform well for audio-based MER and context-based UMR, we cannot attribute the phenomenon to the semantic gap or the immaturity of the feature extraction algorithms. Rather, it is more likely that there is a causal relationship between user mood and blog post, but not in the case between user mood and music choice.

That being said, it is important to note that the music title a blogger attached to a post is not necessarily a song for the user context described in the post, but for the user context while posting the article. Therefore, the music listening records from LiveJournal may still be different from the music listening history recorded by, for example, a music player.

Given that users' short-term music preference is usually influenced by the context of music listening, context-aware music recommendation offers a promising tool by which users can access their favorite music pieces one after another in a query-free and context-dependent manner [66], [67]. Based on the above findings, it seems feasible to infer an individual's mood from the user-generated text (blog posts, short messages, etc.) and then recommend either mood-congruent or mood-incongruent music depending on user mood and personality.

VII. CONCLUSION

In this paper, we have presented a preliminary study that explores the relationship between the musical, personal and situational factors of music listening, using large collections of social media data. The study has been structured around three core research questions, each of which investigates the connection between a pair of factors. Our result shows that it is more accurate to predict the individual's emotional state from the user-generated context, although we can identify several patterns between user mood and music emotion. Our study has also identified tripartite association among user mood, music emotion, and individual's personality traits similar to those reported in the psychology literature. These findings, as a whole, suggest that the social functions of music can be well explored from a real-life, in-situ data set.

Due to the available data and space limitation of the paper, in this study we have not taken into account individual's long-term music preference, social network, time and location of music listening, among others. More insights can be obtained if we have access to an individual's complete listening history [68]. In that case, we can study the relationship between user mood and music preference before and after mood regulation by music, for example.

Although the present study might be at best preliminary, we hope it can call for more attention towards the investigation of personal and situational factors in music listening using the online repository of social data.

APPENDIX
DETAILS OF THE AUDIO FEATURES

Below are the brief introduction of the audio features we employ for both UMR and MER. The features were generated by PsySound [48] and MIRtoolbox [49] with default settings.

Energy. PsySound aims to model parameters of auditory sensation based on psychoacoustic models [48]. We use PsySound to estimate the loudness (low/high), volume (small/large), and sound pressure level (A-weighted or Z-unweighted using slow or fast integration times) of music.

Rhythm. We implemented the following five rhythm features proposed in [7] based on utility functions in MIRtoolbox: average tempo (beats per minute), event density (average onset frequency), rhythm strength (average onset strength), rhythm regularity and clarity. They are all statistics derived from the detection curve of onset, which refers to the starting time of each musical events (notes) [49].

Timbre. We used MIRtoolbox to extract a large number of short-time timbre features based on the short-time Fourier transform of the audio signal [49]. The features were computed on a per-frame basis for every 50 ms, half-overlapping frames. To represent the entire musical piece in the vector space, the running features were summarized by taking the mean and standard deviation. Most features are statistics computed from the spectral domain. For example, spectral flux is defined as the square of difference between the normalized magnitudes of successive frames, spectral irregularity measures the degree of variation of the successive peaks of the spectrum of single frame, and spectral flatness represents the ratio between the geometric mean of the power spectrum and its arithmetic mean [3], [6]. We also use PsySound to get psychoacoustic features including sharpness (dull/sharp), timbre width (flat/rough), spectral and tonal dissonance (dissonant/consonant) [48].

Tonal We used MIRtoolbox to compute the pitch class profile (or wrapped chromagram; the intensity of 12 semitone pitch classes) for each frame and took the centroid and peak as two pitch features [49]. By comparing a chromagram with the 24 major and minor key profiles [6], we can estimate the strength of the frame in association with each key (e.g., C major) and the musical mode. In addition, we used PsySound to compute the tonalness (tone-like or noise-like) and multiplicity (the number of pitches heard) of music [48].

ACKNOWLEDGMENT

The authors would like to thank Gilly Leshed for sharing the 21M LiveJournal posts with us and the anonymous reviewers for valuable comments that improved the quality of this paper.

REFERENCES

[1] M. Clayton, "The social and personal functions of music in cross-cultural perspective," in *The Oxford Handbook of Music Psychology*, S. Hallam, I. Cross, and M. Thaut, Eds. New York, NY, USA: Oxford Univ. Press, 2008.

[2] D. J. Hargreaves and A. C. North, "The functions of music in everyday life: Redefining the social in music psychology," *Psychol. Music*, vol. 27, no. 1, pp. 71–83, 1999.

[3] Y.-H Yang and H. H. Chen, *Music Emotion Recognition*. Boca Raton, FL, USA: CRC, 2011.

[4] M. Zentner, D. Grandjean, and K. R. Scherer, "Emotions evoked by the sound of music: Characterization, classification, and measurement," *Emotion*, vol. 8, no. 4, pp. 494–521, 2008.

[5] P. N. Juslin and P. Laukka, "Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening," *J. New Music Res.*, vol. 33, no. 3, pp. 217–238, 2004.

[6] M. Müller *et al.*, "Signal processing for music analysis," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 6, pp. 1088–1110, 2011.

[7] L. Lu, D. Liu, and H. Zhang, "Automatic mood detection and tracking of music audio signals," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 1, pp. 5–18, 2006.

[8] T. T. Dang and K. Shirai, "Machine learning approaches for mood classification of songs toward music search engine," in *Proc. Int. Conf. Knowledge & Syst. Eng.*, 2009.

[9] X. Hu, "Improving music mood classification using lyrics, audio and social tags," Ph.D. dissertation, Univ. Illinois at Urbana-Champaign, Urbana, IL, USA, 2010.

[10] B. Schuller, F. Weninger, and J. Dorfner, "Multi-modal non-prototypical music mood analysis in continuous space: Reliability and performances continuous space: Reliability and performances," in *Proc. Int. Soc. Music Info. Retrieval Conf.*, 2011, pp. 759–764.

[11] C. Laurier, "Automatic classification of musical mood by content-based analysis," Ph.D. dissertation, Universitat Pompeu Fabra, Barcelona, Spain, 2011.

[12] Y.-H. Yang and H. H. Chen, "Ranking-based emotion recognition for music organization and retrieval," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 4, pp. 762–774, 2011.

[13] J.-C. Wang *et al.*, "The acoustic emotion Gaussians model for emotion-based music annotation and retrieval," in *Proc. ACM Multimedia*, , 2012, pp. 89–98.

[14] Y. E. Kim *et al.*, "Music emotion recognition: A state of the art review," in *Proc. Int. Soc. Music Info. Retrieval Conf.*, 2010.

[15] Y.-C. Lin, Y.-H. Yang, and H. H. Chen, "Exploiting online music tags for music emotion classification," *ACM Trans. Multimedia Comput., Commun., Applicat.*, vol. 7S, no. 1, 2011.

[16] Y.-H. Yang and H.-H. Chen, "Machine recognition of music emotion: A review," *ACM Trans. Intell. Syst. Technol.*, vol. 3, no. 4, 2012.

[17] M. Barthet, G. Fazekas, and M. Sandler, "Multidisciplinary perspectives on music emotion recognition: Implications for content and context-based models," *Proc. Int. Soc. Comput. Music Modelling & Retrieval*, pp. 492–507, 2012.

[18] R. W. Picard, *Affective Computing*. Cambridge, MA, USA: MIT Press, 1997.

[19] S. Koelstra *et al.*, "Deap: A database for emotion analysis; using physiological signals," *IEEE Trans. Affective Comput.*, vol. 3, no. 1, pp. 18–31, 2012.

[20] K. Jonghwa and E. Andé, "Emotion recognition based on physiological changes in music listening," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 12, pp. 2067–2083, 2008.

[21] R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Trans. Affective Comput.*, vol. 1, no. 1, pp. 18–37, 2010.

[22] V. N. Salimpoor *et al.*, "Anatomically distinct dopamine release during anticipation and experience of peak emotion to music," *Nature Neurosci.*, vol. 14, pp. 257–262, 2011.

[23] M. Pantic, G. Roisman, and T. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 39–58, 2009.

[24] M. A. Nicolaou *et al.*, "Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space," *IEEE Trans. Affective Comput.*, vol. 2, no. 2, pp. 92–105, 2011.

[25] P. D. Turney and M. L. Littman, "Measuring praise and criticism: Inference of semantic orientation from association," *ACM Trans. Inf. Syst.*, vol. 21, no. 4, pp. 315–346, 2003.

[26] A. Gabrielsson, "Emotion perceived and emotion felt: Same or different?," *Musicae Scientiae*, pp. 123–147, 2002.

[27] A. J. Lonsdale and A. C. North, "Why do we listen to music? a uses and gratifications analysis," *British J. Psychol.*, vol. 102, pp. 108–134, 2011.

[28] A. V. Goethem and J. A. Sloboda, "The functions of music for affect regulation," *Musicae Scientiae*, vol. 15, no. 2, pp. 208–228, 2011.

[29] P. G. Hunter, E. G. Schellenberg, and A. T. Griffith, "Misery loves company: Mood-congruent emotional responding to music," *Emotion*, vol. 11, no. 5, pp. 1068–1072, 2011.
[30] A. J. M. van den Tol, "A self-regulatory perspective on people's decision to engage in listening to self-selected sad music when feeling sad," Ph.D. dissertation, Univ. Limerick, Limerick, Ireland, 2012.
[31] S. J. Breckler, R. B. Allen, and V. J. Konecni, "Mood-optimizing strategies in aesthetic-choice behavior," *Music Percept.*, vol. 2, no. 4, pp. 459–470, 1985.
[32] D. Watson and R. Mandryk, "An in-situ study of real-life listening context," in *Proc. Sound and Music Computing Conf.*, 2012, pp. 11–16.
[33] M. A. Cohn, M. R. Mehl, and J. W. Pennebaker, "Why we blog?," *Commun. ACM*, vol. 47, no. 12, pp. 41–46, 2004.
[34] G. Leshed and J. Kaye, "Understanding how bloggers feel: Recognizing affect in blog posts," in *Proc. ACM Int. Conf. Comput. Human Interaction*, 2006.
[35] A. Miura and K. Yamashita, "Psychological and social influences on blog writing: An online survey of blog authors in Japan," *J. Comput.-Mediated Commun.*, vol. 12, pp. 1452–1471, 2007.
[36] J. Tang *et al.*, "Quantitative study of individual emotional states in social networks," *IEEE Trans. Affective Comput.*, vol. 3, pp. 132–144, 2012.
[37] A. Schindler, R. Mayer, and A. Rauber, "Facilitating comprehensive benchmarking experiments on the million song dataset," in *Proc. Int. Soc. Music Info. Retrieval Conf.*, 2012, pp. 469–474.
[38] F. Mairesse *et al.*, "Using linguistic cues for the automatic recognition of personality in conversation and text," *J. Artif. Intell. Res.*, vol. 30, pp. 457–501, 2007.
[39] T. Yarkoni, "Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers," *J. Res. Personality*, vol. 44, no. 3, pp. 363–373, 2010.
[40] K. D. Schwartz and G. T. Fouts, "Music preferences, personality style, and developmental issues of adolescents," *J. Youth Adolescence*, vol. 32, pp. 205–213, 2003.
[41] P. J. Rentfrow and S. D. Gosling, "The Do Re Mi's of everyday life: The structure and personality correlates of music preferences," *J. Personality Soc. Psychol.*, vol. 84, no. 6, pp. 1236–1256, 2003.
[42] P. J. Rentfrow, L. R. Goldberg, and D. Levitin, "The structure of musical preferences: A five-factor model," *J. Personality Soc. Psychol.*, vol. 100, no. 6, pp. 1139–1157, 2011.
[43] T. Chamorro-Premuzic *et al.*, "Personality, self-estimated intelligence, and uses of music: A Spanish replication and extension using structural equation modeling," *Psychol. Aesthet., Creativ. Arts*, vol. 3, no. 3, pp. 149–155, 2009.
[44] O. Ladinig and E. G. Schellenberg, "Liking unfamiliar music: Effects of felt emotion and individual differences," *Psychol. Aesthet., Creativ. Arts*, vol. 6, no. 2, pp. 146–154, 2011.
[45] L. R. Goldberg, "An alternative description of personality: the big-five factor structure," *J. Personality Soc. Psychol.*, vol. 59, no. 6, pp. 1216–1229, 1990.
[46] J. A. Russell, "A circumplex model of affect," *J. Personality Soc. Sci.*, vol. 39, no. 6, pp. 1161–1178, 1980.
[47] M. Bradley and P. J. Lang, Affective Norms for English Words ANEW: Instruction Manual and Affective Ratings.The Center for Research in Psychophysiology, Univ. Florida, Tech. Rep., 1999.
[48] D. Cabrera, S. Ferguson, and E. Schubert, "Psysound3: software for acoustic and psychoacoustic analysis of sound recordings," in *Proc. Int. Conf. Auditory Display*, 2007, pp. 356–363. [Online]. Available: http://psysound.wikidot.com/.
[49] O. Lartillot and P. Toiviainen, "MIR in matlab (II): a toolbox for musical feature extraction from audio," in *Proc. Int. Soc. Music Info. Retrieval Conf.*, 2007, pp. 127–130. [Online]. Available: http://users.jyu.fi/lartillo/mirtoolbox/.
[50] F. Sebastiani, "Machine learning in automated text categorization," *ACM Comput. Surveys*, vol. 34, no. 1, pp. 1–47, 2002.
[51] A. Aizawa, "An information-theoretic perspective of tf-idf measures," *Inf. Process. Manage.*, vol. 39, pp. 45–65, 2003.
[52] S. E. Robertson *et al.*, "Okapi at trec-4," in *Proc. Text REtrieval Conf. (TREC-4)*, 1995, pp. 73–96.
[53] M. Lan *et al.*, "Supervised and traditional term weighting methods for automatic text categorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, pp. 721–735, 2009.
[54] T. Hofmann, "Probabilistic latent semantic indexing," in *Proc. ACM SIGIR Conf. Res. and Develop. in Inf. Retrieval*, 1999, pp. 50–57.
[55] R.-E. F. Fan *et al.*, "LIBLINEAR: A library for large linear classification," *J. Mach. Learn. Res.*, vol. 9, pp. 1871–1874, 2008.
[56] X.-Y. Liu, J. Wu, and Z.-H. Zhou, "Exploratory under-sampling for class-imbalance learning," *IEEE Trans. Syst., Man, Cybern.*, vol. 39, no. 2, pp. 539–553, 2009.
[57] C. Cortes and M. Mohri, "Auc optimization vs. error rate minimization," in *Proc. Conf. Advances in Neural Inf. Processing Syst.*, 2004.
[58] K. Jarvelin and J. Kekalainen, "Cumulated gain-based evaluation of IR techniques," *ACM Trans. Inf. Syst.*, vol. 20, no. 4, pp. 422–446, 2002.
[59] D. C. Montgomery, G. C. Runger, and N. F. Hubele, *Engineering Statistics*. New York, NY, USA: Wiley, 1998.
[60] D. Tingle, Y. E. Kim, and D. Turnbull, "Exploring automatic music annotation with acoustically objective tags," in *Proc. ACM Int. Conf. Multimedia Inf. Retrieval*, 2010, pp. 55–62.
[61] C.-C. Chang and C.-J. Lin, LIBSVM: A Library for Support Vector Machines, 2001. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm.
[62] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006. [Online]. Available: http://www.gaussianprocess.org/.
[63] L. L. Nathans, F. L. Oswald, and K. Nimon, "Interpreting multiple linear regression: A guidebook of variable importance," *Practical Assess., Res., Eval.*, vol. 17, no. 9, pp. 1–19, 2012.
[64] J. W. Pennebaker, M. E. Francis, and R. J. Booth, Linguistic inquiry and word count -LIWC2001, 2001. [Online]. Available: http://www.liwc.net/descriptiontable1.php.
[65] M. Coltheart, "The MRC Psycholinguistic database," *Quart. J. Exp. Psychol. Section A*, vol. 33, no. 4, pp. 497–505, 1981.
[66] M. Kaminskas and F. Ricci, "Contextual music information retrieval and recommendation: State of the art and challenges," *Comput. Sci. Rev.*, vol. 6, no. 2–3, pp. 89–119, 2012.
[67] G. T. Elliott and B. Tomlinson, "Personalsoundtrack: contextaware playlists that adapt to user pace," in *Proc. ACM CHI*, 2006, pp. 736–741.
[68] N. Bolger, A. Davis, and E. Rafaeli, "Diary methods: capturing life as it is lived," *Annu Rev. Psychol.*, vol. 54, pp. 579–616, 2003.

Yi-Hsuan Yang (M'11) received the Ph.D. degree in Communication Engineering from National Taiwan University, Taiwan, in 2010. Since September 2011, he has been with the Academia Sinica Research Center for Information Technology Innovation, where he is an Assistant Research Fellow. His research interests include music information retrieval, multimedia content analysis, machine learning, and affective computing. He was awarded the 2011 IEEE Signal Processing Society (SPS) Young Author Best Paper Award. He is an author of the book Music Emotion Recognition, published by CRC Press in 2011. He gave a tutorial on music affect recognition in the International Society for Music Information Retrieval Conference (ISMIR) in 2012. His work on emotion-based music video generation won the first prize in ACM Multimedia Grand Challenge 2012.

Jen-Yu Liu received the M.S. degree in Computer Vision and Artificial Intelligence from Universitat Autonoma de Barcelona, Spain, in 2011. Since October 2011, he has been a Research Assistant at the Academia Sinica Research Center for Information Technology Innovation. His research interests include music information retrieval, user-centered data mining, and machine learning.