Genre and Mood Classification Using Lyric Features

Teh Chao Ying, Shyamala Doraisamy, Lili Nurliyana Abdullah

Faculty of Computer Science and Information Technology

43400, UPM, Serdang, Malaysia

hipohipo87@hotmail.com, {shyamala,liyana}@fsktm.upm.edu.my

***Abstract*— Musical genre and mood classification techniques combining lyric and audio features for Music Information Retrieval (MIR) have been studied widely in recent years. This paper investigates the performances of musical genre and mood classification using only lyric features. In this preliminary study, the Part-of-Speech (POS) feature is utilizes for classification of a collection of 600 songs. Ten musical genre and mood categories were selected respectively based on a summary from the literature. Experiments show that classification accuracies for mood categories outperform genres.**

***Keywords*— Music Information Retrieval; Musical Mood; Musical Genre; Lyrics Analysis**

I. Introduction

Lyrics associated with the music contain information that is vital to the reception and the message of a song. A large number of studies [2, 5, 10, 18] in Music Information Retrieval (MIR) have focused on the application of various principles of text information retrieval to music representation based on lyric. This application of text information retrieval to music representation is useful for lyrics analysis in MIR.

Song lyrics exhibit a certain structure, as they are organized in blocks of choruses and verses. Many songs are organized in rhymes, patterns which are reflected in a song's lyrics and these structures in text are much easier to detect rather than audio.

Many studies have been conducted in the area of musical genre and mood classification based on various combinations of lyric and audio features. These are discussed further in the following section. In this study, we focus on investigating the performance of using the Part-of-Speech feature for musical genre classification in comparison to mood classification.

This paper is organized as follows: Section 2 reviews related work on lyric analysis, musical genre and mood classification in music information retrieval. Section 3 introduces the experimental dataset and data preprocessing for this study. Section 4 describes experiments and lyrics features examined. Section 5 presents the conclusions and future work.

II. Related work

Several research teams have further begun working on adding textual information to the retrieval process, predominantly in the form of song lyrics. Yang and Lee [5] is one of the earliest studies that combined lyrics and audio in MIR. They combined lyric Bag-of-Words (BOW) features and the 182 psychological features proposed in the General Inquirer to disambiguate mood categories, that audio-based classifiers found confusing. An overall improvement in classification accuracy of 2.1% was obtained.

In order to develop a MIR system that uses lyrics information, lyrics recognition systems have been developed [9]. The lyric recognition technique used in these systems is simply a large vocabulary continuous speech recognition (LVCSR) technique, based on HMM (hidden Markov model), acoustic model and a trigram language model.

A semantic and structural analysis of song lyrics is conducted in [13]. It focuses on aspects such as structure detection, e.g. verses and chorus, classification into thematic categories such as 'love', 'violent', 'Christmas', and similarity search. Lyrics are gathered from multiple sources on the web, and are subsequently aligned to each other for matching sequences, to filter out errors like typing errors, or retrieved parts not actually belonging to the lyrics of the songs, such as commercials.

Mood classification in MIR is an emerging metadata type in music digital libraries. The lyrics of a song, which will be heard and understood by a listener, play an important role to help the listener in determining the mood of a song. Hence, the mood of a song has been recognized as one of the important criterion when users organize music objects [4]. The mood of a song is expressed by means of musical features but a relevant part also seems to be conveyed by the lyric texts. In recent years, researchers have started to exploit music lyrics in mood classification.

Cyril et al. [1, 2] research had confirmed the relevance of the lyrics to convey emotions or at least that the mood expressed in music and acoustical data is correlated with information contained in the text. X. Hu and Downie [16] recent research had shown mood classification in music

978-1-4673-1090-1/12/$31.00 ©2012 IEEE

can be significantly improved when combining lyrics and audio.

Musical genres are categorical and typological descriptions that are used to describe musical sounds. Genre classification in MIR has been widely based on audio features, but in recent years, a combination of lyric text and audio [7, 10, 14]. One of the earliest studies on genre classification in lyrics is the study by Neumayer and Rauber [10]. Their study focused on the correlation of the combinational approach using lyric and audio, and genre classification in music retrieval system. In [13, 14], they further study the effect of song lyrics with genre classification. Lyrics analysis that carried out includes rhyme pattern, part of speech characteristic and text statistic features. In their work, lyrics were processed using the Bag-of-Words features and then weighted by the *tf.idf* information. Feature selection was done via document frequency thresholding, i.e. the omittance of terms that occur in a very high or very low number of documents.

For musical mood classification, X. Hu and Downie[18] performed their study based on lyrics text. They used only lyric features to classify 18 mood categories derived from user tags. In this study, they examine the role that lyrics text can play in improving audio musical mood classification.

In this study, we only look at lyric analysis using the POS feature. The performance of using the POS feature in musical genre classification is compared with its use in mood classification.

III. DATA PREPROCESING

A total of 600 English songs were selected randomly from a private collection. In order to prevent biased results by too many songs from the same artist and album, we collected the songs from different artists that appeared in different albums as well. At the same time, in order to ensure enough data for our experiments, we only selected those songs with lyrics whose word count was greater than 100. This collection thus comprises songs from different artists, stemming from 200 different albums.

Lyrics are distributed all over the internet and it may contain mistakes since most of it was added by listeners. In order to get the most precise lyrics, we fetch it from three different websites: lyricwiki[1], metrolyrics[2] and amarok[3].

A. *Musical Genres*

Musical genre clasification have been a focus in many MIR studies[10,14]. From the widely used websites which categorised music by genre, allmusic.com[4] and last.fm[5] , and refer to the genres categories in [14], we selected a total number of 10 genre types including pop, blue, country, folk, R & B, reggae, grunge, punk rock, soul, and metal in our dataset.

For each song in the dataset, the genre was determined using the social tag labeled by the listener from the website stated above. Only those songs that have the same genre tag from these two website were selected.

B. *Musical Moods*

Various mood categories were used in previous work in this field. The number of categories in general range from four to ten categories [2, 6, 17, 18]. Among the many emotion models, Russell's model and Tellegen-Watson-Clark model were adopted as the emotion models in this MIR study.

Russell's model is a dimensional model where emotions are positioned in a multidimensional space. There are two dimensions in it: valence and arousal.

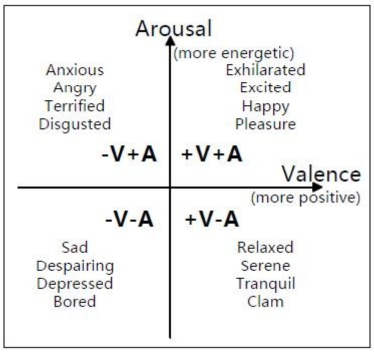

Figure 1. Russell's Model

Russell's model places a total of 28 emotions denoting adjectives in a bipolar space subsuming the two dimensions. In [17], X. Hu and Downie selected a total of 18 moods while Cyril et al [2] selected 4 moods from Russell's model. Besides, in [6], Dan and Yang selected a total of 23 moods from Tellegen-Watson-Clark model. We select a total of 8 moods from Russel's model which were used in [2, 17] and 2 moods from Tellegen-Watson-Clark model which were used in [6]. This builds up the 10 mood categories in our study: happy, sad, angry, relaxed, calm, gloomy, romantic, confident, disgusted, and aggressive.

For the development of the set of words for each mood category we adopted a similar approach in the study by [18]. From the 600 songs, we identified all the unique words and grouped all the synonyms together using WordNet. With WordNet-Affect [4], an extension of WordNet where affective labels are assigned to concepts representing emotions, moods, or emotional responses, we further categories all the unique words and their synonyms into our ten emotion category.

Five under-graduate students were included to check the relevancy of the mood of 600 songs in this study. Given the song lyrics, they were asked to read through it and label the mood of the song. Each of them was asked to check the relevancy of the mood for 120 songs. The mood of the song in our dataset was further evaluated using Conceptnet[6], a freely available commonsense know-

[1] http://lyrics.wikia.com/Lyrics_Wiki
[2] http://www.metrolyrics.com/
[3] http://amarok.kde.org/wiki/Scripts
[4] http://www.allmusic.com/
[5] http://last.fm/
[6] http://conceptnet5.media.mit.edu/

ledge base and natural-language-processing toolkit with mood guessing function. Summarizing the result obtains from the undergraduate student and Conceptnet a comparison of mood relevancy was then perform. If the mood of a song differed widely, another judgment was perform to decide a suitable mood.

C. Lyrics Preprocessing

Lyrics text has unique structures and characteristics. Most lyrics consist of sections such as intro, verse, chorus, and music. Repetitions of words and sections are very common. However, very few available lyric texts were found as verbatim transcripts. Instead, repetitions were annotated as instructions like [repeat chorus 2x, back to intro] [18]. To get the proper lyric texts in our dataset, the steps taken was to check whether the fetched lyrics match the song, followed by comparing three different versions of song in term of spelling, lyric structure (paragraphing) and the words used. In our study we did the lyric preprocessing steps similar to most studies. We cleaned up the lyrics text by removing all the instruction and replaced with the word itself to get the complete lyric text document.

IV. Experiments

A. Lyrics Feature

Lyrics are a very rich resource and many types of textual features can be extracted from them. In this preliminary work, from the most commonly used feature types in related text classification tasks, we only focused on one lyric feature which was Part-of-Speech.

Part-of-speech (POS) tagging, also called word-category disambiguation is a lexical categorization or grammatical tagging of words according to their definition and the textual context they seem in. Basic POS categories are for example nouns, verbs, conjunction, articles, and adjectives. In [13], Rauber and Mayer presume that different genres will differ in the category of words used. In their study, 9 POS categories which include nouns, verbs, pronouns, relational pronouns, prepositions, adverbs, articles, modals and adjectives were used. We similarly used the same POS categories. and another POS category, interjections was added. Interjections were chosen because it can express emotion [19].

LingPipe[7] suite of libraries was employed in the analysis of POS tagging similar to [13]. We fed the lyrics text into LingPipe and the outputs were obtained as lyric texts with the POS tagger. All the 600 lyric text with POS tagger were then analyzed in terms of the occurrence of each unique word in each different category of POS. This would then act as features for genre and mood classification. A general English analysis tagger, Brown Corpus was used.

[7] http://allias-i.com/lingpipe/

B. Results Discussion

For the experiments, we employed the WEKA machine learning toolkit. We utilized the k-Nearest-Neighbour, Naïve Bayes and Support Vector Machines (SVM) based on a ten-fold cross–validation, which was further averaged over five repeated runs. The classification experiments were perform based on POS features and Table 2 shows the accuracies obtain for musical genre and mood classifications.

TABLE 2
ACCURACIES FOR MUSICAL GENRE AND MOOD

POS	K-NN	Naïve Bayes	SVM
Genre	0.3841	0.3658	0.3994
Mood	0.5662	0.5064	0.5967

From the results we can see that the accuracy for mood classification was better than genre classification using the POS feature. From the unique words identified in Section III, many (but not all) of the words convey emotional information rather than genre information. There is more significant data that we can get in musical mood compared to musical genre. One of the reasons for lower result for genre classification could be lack of genre information from the words. As a result, the accuracy for genre classification is lower than mood classification.

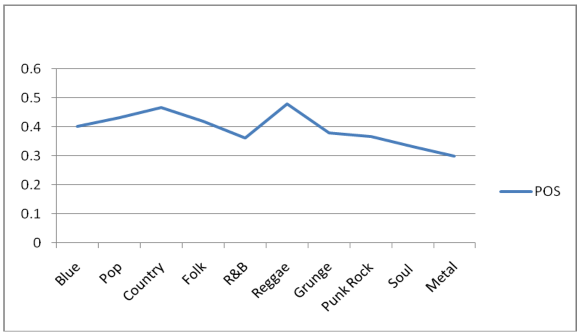

Figure 2. Accuracies for genre classification

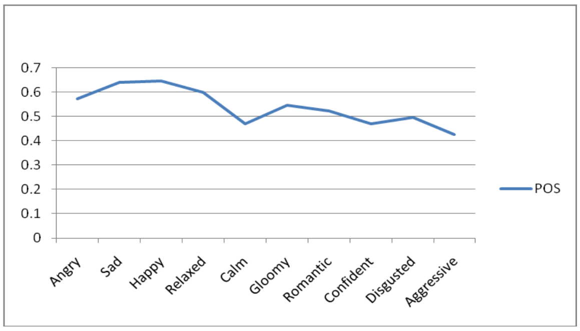

Figure 3. Accuracies for mood classification

As shown in Fig. 2 and Fig. 3, the accuracy for genre and mood on different categories vary greatly. In genre classification, Country and Reggae achieved higher accuracy compared to other genre type, while Metal had the lowest accuracy. In Country songs, we can identify some typical words such as “road”, “cowboy”, “house”, “highway”, “American”, and “Whiskey” while in Reggae songs, words like “dem”, “mi”, “fi”, “inna”, and “jah” happen in a high frequency. The occurrence of these typical words contributes to the higher classification accuracy in both of the genre.

In mood classification, it can be observed that sad and happy had better accuracy while calm and aggressive had lower accuracy. The number of words that represented the emotions happy and sad were greater compared to the number of words that represented the emotion calm and aggressive. This is highly likely the reason for the lower classification accuracy of calm and aggressive.

V. CONCLUSION

In this paper, we show the feasibility of POS in musical genre and mood classification. Our comparison of the accuracy for musical genre and mood classification shows that mood classification was better than genre classification. We found that lack of typical words in genres and moods categories respectively will affect the classification accuracy.

In future, we plan to improve our work in the following ways: (1) In order to achieve better result, others lyric features such as Bag-of-Words and Text Statistic features will be investigate. (2) As described in Section III, under-graduates students were included to check the relevancy of the mood of the dataset. Expert in MIR and linguistic field will be invited to train our under-graduates students so that their knowledge in musical mood is strong enough to determine the mood of a song. (3) Work on a larger data collection, and (4) study how will be the accuracy goes when we combine the musical genre and mood classification together.

REFERENCES

[1] Cyril Laurier et al, "Indexing Music by Mood: Design and integration of an automatic content-based annotator," Multimedia Tools Application 2010.

[2] C. Laurier, J. Grivolla and P. Herrera, "Multimodal music mood classification using audio and lyrics," In Proc. of the 7th International Conference on Machine Learning and Applications, San Diego, California, USA, 2008.

[3] C. Strapparava and A. Valitutti, "WordNet-Affect: An Effective Extension of WordNet," In Proc. of the International Conference on Language Resources and Evaluation, pp. 1083-1086, 2004.

[4] D. Bainbridge, S.J. Cunningham and J.S. Downie, "How People Describe Their Music Information Needs: A Grounded Theory Analysis of Music Queries," In Proc. Of the 2nd International Conference on Music Information Retrieval (ISMIR'03).

[5] D. Yang, and W. Lee, "Music Emotion Identification from Lyrics," In the Proc. of 11th IEEE International Symposium on Multimedia 2009.

[6] D. Yang, and W. Lee, "Disambiguating Music Emotion Using Software Agents," In Proc. of the 5th International Conference on Music Information Retrieval (ISMIR'04).

[7] G. Tzanetakis, G. Essl, and P. Cook, "Automatic Musical Genre Classification Of Audio Signals," In Proc. of the International Conference on Music Information Retrieval (ISMIR'01).

[8] K. Trohidis, G. Tsoumakas, G. Kalliris, and I. Vlahanas, "Multi-label Classification of Music into Emotions," In Proc. of the 9th International Conference on Music Information Retrieval (ISMIR'08).

[9] M. Suzuki, T. Hosoya, A. Ito, and S. Makino, "Music Information Retrieval from a Singing Voice Using Lyrics and Melody Information," EURASIP Journal on Advances in Signal Processing, Volume 2007.

[10] Robert Neumayer and Andreas Rauber, "Integration of Text and Audio Features for Genre Classification in Music Information Retrieval," In Proc. of the 29th European Conference on Information Retrieval(ECIR'07), pp. 724-727, Rome, Italy, 2007.

[11] Tao Li and Mitsunori Ogihara, "Toward Intelligent Music Information Retrieval," IEEE Transactions on Multimedia, Volume 8, No.3 (2006), pp. 564-574. Multilabel Music.

[12] T. Wang, D. J. Kim, K. S. Hong, and J. S. Youn, "Music Information Retrieval System using Lyrics and Melody Information," In Proc. of the Asia-Pacific Conference on Information Processing.

[13] Rudolf Mayer and Andreas Rauber, 2010, "Multimodal Aspect of Music Retrieval: Audio, Song Lyrics – and Beyond?" Z.W.Raś and A.A. Wieczorkowska (Eds.): Adv. In Music Inform. Retrieval, SCI 274, pp. 333-363.

[14] R. Mayer, R. Neumayer, and A. Rauber, "Combination of Audio and Lyrics Features for Genre Classification in Digital Audio Collection,"

[15] R.Neumayer and A.Rauber, "Multimodal Music Retrieval: Visualization and Evaluation of Clustering by Both Audio and Lyrics," In Proc. Of the 8th Conference Recherehe d'Information Assistee par Ordinator(RIAO'07).

[16] X. Hu, and J.S. Downie, "Improving mood classification in music digital libraries by combining lyrics and audio," Proceeding of the 2010 Joint International Conference on Digital Libraries, JCDL 2010.

[17] X. Hu, and J.S. Downie, "When Lyrics Outperform Audio For Music Mood Classification: A Feature Analysis," In Proc. Of the 11th International Conference on Music Information Retrieval (ISMIR'10).

[18] X. Hu, J.S. Downie and A.F. Ehmann, "Lyric Text Mining In Music Mood Classification," In Proc. Of the 10th International Conference on Music Information Retrieval (ISMIR'09).

[19] http://en.wikipedia.org/wiki/Interjection