Identification of Mood based on Lyric Text Mining

V. Kalyani, Dr. Y. Sangeetha

[1]Student of M. Tech, 4th semester Department of information technology
V. R. Siddhartha engineering college,Autonomous Kanuru,Vijayawada, A.P,India

[2]Associate Professor Department of Information Technology
V.R. Siddhartha engineering college,Autonomous Kanuru,Vijayawada,A.P,India

Abstract
The detail study of the subject provides categorizing the songs according to the mood by using lyrics of songs. A new method is proposed to identify the mood of the song by uploading the song files according to the users wish. Before performing data mining tasks we applying data pre-processing for better results. 5 mood categories are maintained according to realistic and user centric perspective. In this research songs can be examine by using different data mining techniques and lex summarization. Finally we canpredict the mood of the song by providing lyrics of the song.

Keywords: Mood, Pre-processing, Categorizing, Examine.

1. Introduction
In the present world data is increased day by day for which music domain is not an exception for thi7s scenario people are focusing to develop and evaluate Music information Retrieval (MIR) that can improve dynamic accessing of the mood in music. Form an idea of the twenty two music systems in the year between 2007 and 2008 in the Audio Mood Classification (AMC) task of the Music information Retrieval Evolution exchange (MIREX)[3]. So these evaluations, severaltechniques haveemerged andresolving these issues in this regard will greatly facilitate further progress on this topic. Machine learning,data mining and other techniques are used in Music sentiment analysis or mood recognition to classify songs in either positive or negative or more emotion categories with highest possible accuracy.AMC dataset is one of the ground truth set that has been used to evaluate in the mood classification systems developed by multiple labs.   Different features of the song that is audio lyrics,metadata or heart of song can be used as the features of the song or combined together is also used.There is a successfully studies that predict music emotions based on the lyrics by using different data mining techniques and other related models. Songs are labelled with emotion or mood categories are crucial requirement to train and using different classification techniques. It is not feasible to prepare large datasets manually.it is very time consuming and highly cost effective. Now a days different songs related websites are in online that are Last.fm and movie rulz.com etc.

1.1Barriers Creating Ground Truth Data Set
Due to characteristicattribute ofsubjectivity of music perception, there is no specified mood category. Music psychologists have introduced various mood models but these have been condemned for missing the true social factors of music listening. Some MIR researchers are taking advantage of professionally assigned mood labels [5]-[6].professionally created labels condemned for not capturing the users' sensitive insight on mood.

1.2 Need for Multimodal Mood Classification
In music psychology and education, the mood of the song can be identified as the most strongly involved with music expressivity. Music information behaviour studies have also recognized music mood as an crucial criterion used by people in music seeking ad organization [1]. Several models have been conducted to classify music by mood [2].there are lot of limitations in music mood classification due to lack of music data and obstacles in identication of mood because mood is a very word net subjective notion,there has yet to unfold a generally agreed mood differentiation that is used within the MIR research and development community[1].

1.3 Selecting the Songs
A song was not selected based on the category if its title or artists contained the similar terms within that category.songs are selected from the different online authentic websites. Last.fm and movierulz.com different online websites provide lyrics of the song. In this paper we provide songs dynamically according to uers wish.word.net is used to identify the important words in the songs.

2. Expermints

2.1 Lyric Pre-processing
Lyric text has unique structures and characteristics that required various pre-processingtechniques. Most of the lyrics contained different sections intro, verse, interlude, pre-chorus, chorus and outro,many with annotation on these segments. Another important thing is repetitions of words in the lyrics and most of the sections are common very few lyric texts were found as verbatim transcripts. Most of the lyrics contain notes about song and instrumentation or the performing artists. When we want to

International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org

Volume 6, Issue 4, July- August 2017   ISSN 2278-6856

perform pre-processing we have to consider these things .in this paper we dynamically provide the songs according to user's wish irrespective of static nature so we ca improve songs data at any time.

2.2 Lyric features
Lyrics are a substantial resource and proper lyrics of the song is crucial to identify the mood of the song .most of the textual features are extracted from the lyrics [4].in this work lyrics of the song play a crucial role so we have to improvise the lyrics of song by putting additional efforts.

2.3 Stop Words
Stop words means the words which are not giving proper meaning for context of the lyrics and we have to remove the articles,pre-portisitions and improper words for better results.Each word is assigned a value that can reprent, among others, the frequency of the word, TF-IDF weight,normalized frequency or a Boolean value indicating presence or absence. In this scenario we use TF-IDF weighting is the most widely used in text analysis and MIR ,but some studies in the text s analysis also reported other representations outperformed TF-IDFweighting.

Table-1: TF-IDF scores of the song

	L1	L2	L3	L4	L5	L6
L1	1.003	o.19	0.0	0.27	0.35	0.209
L2	0.197	1.0	0.206	0.153	0.158	0.093
L3	0.0	0.206	1.001	0.168	0.0	0.0
L4	0.27	0.153	0.168	1.0	0.05	0.033
L5	0.35	0.158	0.0	0.05	1.001	0.167
L6	0.209	0.093	0.0	0.033	0.167	0.998

3. Evaluation Measures and Classifiers
In this study we use k-means clustering technique for classification. Clustering means grouping of similar objects together. For better results we use cosine similarity measure .finally we applied lex summarization for improvement of the results. Lexrank summarization is the one of the technique which we use generally in text mining. The LexRank algorithm measure the importance of sentences in the document by considering its relative importance to its neighboring sentences, where a positive contribution will raise the importance of a sentence's neighbor, while a negative contribution will lower the importance value of a sentence's neighbor. This idea is basically the same with PageRank, unless it is used in counting the importance of sentence in a given set of sentences. We have to maintain dictionary for the emotions. Dictionary must be as large as possible for better results.

Table2:Clustering technique for the song

ID	LINE	CLUSTER
1	Everybody hateful clever	Cluster2
2	Knows dice hateful clever	Cluster3
3	Knows war	Cluster3
4	Everybody hateful knows good guys	Cluster3
5	Sitting hateful clever	Cluster3
6	Morning sun hateful clever	Cluster3

4.Similarity Measures
Similarity Measure is the one of technique in order to calculate similarity between the documents .there are different similarity measures. in this study we use cosine similarity measure to evaluate the song based on the lyrics of the song for better results.followed by lexrank summarization which is generally used in the text mining for optimal solution.

Table3: Cosine similarity matrix

S.No	Line	everybody	hateful	clever	knows	dice
1	Everybody hateful clever	0.201	0.114	0.133	0.0	0.0
2	Knows dice hateful clever	0.0	0.068	0.08	0.14	0.223
3	Everybody crossed	0.151	0.0	0.0	0.0	0.0
4	Everybody hateful clever	0.201	0.114	0.133	0.0	0.0
5	Knows clever	0.0	0.0	0.557	0.349	0.0

5. Conclusions and future work
This study investigates the importance of text features in identification of mood on 5 moodcategories which is provide in the dictionary for better results we have remove stop words, and performing the data preprocessing techniques.we identify the useful lyric text after that we applied different datamining techniques in order to get the results.Finally we identify the mood of the song.All this work is done on lyric text only Sometimes lyric features alone can give better results rather than audiofeatures in the area of where the samples are spread randomly and sparse or semantic meaning of the lyrics connect with the mood category but combining audio and lyric features provide better performance in some categories of mood identification.Most of the lyrics having the repeated data and outliers so for that we have to perform lot of preprocessing for optimal results.Combining audio features

along with lyric features can provide best results.Future work includes examination of other features of text and considers the statistics of the text and important words given by lexicons.we have examine single category of text and find out which category will give crucial feature of text .for optimal solution we use svm,naïve bayes and other machine learning technique.

References

[1] P. N. Juslin and P. Laukka: "Expression, Perception, and Induction of Musical Emotions: A Review and A Questionnaire Study of Everyday Listening," Journal of New Music Research, Vol. 33, No. 3, pp. 217-238, 2004.

[2] X. Hu and J. S. Downie: "Exploring Mood Metadata: Relationships with Genre, Artist and Usage Metadata," Proceedings of the International Conference on Music Information Retrieval, 2007.

[3] X. Hu, J. S. Downie, C. Laurier, M. Bay, and A. Ehmann: "The 2007 MIREX Audio Music Classification Task: Lessons Learned," Proceedings of the International Conference on Music Information Retrieval, 2008.

[4] R. Mayer, R. Neumayer, and A. Rauber: "Rhyme and Style Features for Musical Genre Categorization by Song Lyrics," Proceedings of the International Conference on Music Information Retrieval, 2008. [20] Y.-W. Chen and C.-J. Lin: "Combining SVMs with Various Feature Selection Strategies," In Feature Extraction, Foundations and Applications, Springer,2006.

[5] Y.-H. Yang, Y.-C.Lin, H.-T.Cheng, I.-B. Liao, Y.C. Ho, and H. H. Chen: "Toward multi-modal music emotion classification", Proceedings of Pacific Rim Conference on Multimedia (PCM'08), 2008.

[6] X. Hu, M. Bay, and J. S. Downie: "Creating a Simplified Music Mood Classification Groundtruth Set," Proceedings of the 8th International Conference on Music Information Retrieval, 2007.