Mood Classification of Hindi Songs based on Lyrics

Braja Gopal Patra, Dipankar Das, and Sivaji Bandyopadhyay
Department of Computer Science & Engineering, Jadavpur University, Kolkata, India
{brajagopal.cse, dipankar.dipnil2005}@gmail.com,
sivaji_cse_ju@yahoo.com

Abstract

Digitization of music has led to easier access to different forms music across the globe. Increasing work pressure denies the necessary time to listen and evaluate music for a creation of a personal music library. One solution might be developing a music search engine or recommendation system based on different moods. In fact mood label is considered as an emerging metadata in the digital music libraries and online music repositories. In this paper, we proposed mood taxonomy for Hindi songs and prepared a mood annotated lyrics corpus based on this taxonomy. We also annotated lyrics with positive and negative polarity. Instead of adopting a traditional approach to music mood classification based solely on audio features, the present study describes a mood classification system from lyrics as well by combining a wide range of semantic and stylistic features extracted from textual lyrics. We also developed a supervised system to identify the sentiment of the Hindi song lyrics based on the above features. We achieved the maximum average F-measure of 68.30% and 38.49% for classifying the polarities and moods of the Hindi lyrics, respectively.

1 Introduction

Studies on Music Information Retrieval (MIR) have shown moods as a desirable access point to music repositories and collections (Hu and Downie, 2010a). In the recent decade, much work on western music mood classification has been performed using audio signals and lyrics (Hu and Downie, 2010a; Mihalcea and Strapparava, 2012). Studies indicating contradictory emphasis of lyrics or audio in predicting music moods are prevalent in literature (Hu and Downie, 2010b). Indian music considered as one of the oldest musical traditions in the world. Indian music can be divided into two broad categories, "*classical*" and "*popular*" (Ujlambkar and Attar, 2012). Further, classical music tradition of India has two main variants; namely Hindustani and Carnatic. The prevalence of Hindustani classical music is found largely in north and central parts of India whereas Carnatic classical music dominates largely in the southern parts of India.

Indian popular music, also known as Hindi Bollywood music or Hindi music, is mostly present in Hindi cinemas or Bollywood movies. Hindi is one of the official languages of India and is the fourth most widely spoken language in the World[1]. Hindi or Bollywood songs make up 72% of the total music sales in India (Ujlambkar and Attar, 2012). Unfortunately, not much computational and analytical work has been done in this area.

Therefore, mood taxonomy especially for Hindi songs has been introduced here in order to closely investigate the role played by lyrics in music mood classification. The lyrics corpus is annotated in two steps. In the first step, mood is annotated based on the listener's perspective. In the second step, the same corpus is annotated with polarity based on the reader's perspective. Further, we developed a mood classification system by incorporating different semantic and textual stylistic features extracted from the lyrics. In addition, we also developed a polarity classification system based on the above features.

The paper is organized as follows: Section 2 reviews related work on music mood classification. Section 3 introduces the proposed mood classes. The detailed annotation process and the dataset used in the study have been described in Section 4. Section 5 describes the features of the

[1] www.redlinels.com/2014/01/10/most-widely-

lyrics used in the experiments, which is followed by the results obtained so far and our findings and further prospect are discussed in Section 6. Finally Section 7 concludes and suggests the future work.

2 Related Work

Dataset and Taxonomy: Preparation of an annotated dataset requires the selection of proper mood classes to be used. With respect to Indian music, limited work on mood detection by considering audio features has been reported till today. Koduri and Indurkhya (2010) worked on the mood classification of South Indian Classical music, i.e. Carnatic music. The main goal of their experiment was to verify the *raagas* that really evoke a particular *rasa(s)* (emotion) specific to each user. They considered the taxonomy consisting of ten *rasas* e.g., *Srungaram (Romance, Love), Hasyam (Laughter, Comedy)* etc. Similarly, Velankar and Sahasrabuddhe (2012) prepared data for mood classification of Hindustani classical music consisting of 13 mood taxonomies (*Happy, Exciting, Satisfaction, Peaceful, Graceful, Gentle, Huge, Surrender, Love, Request, Emotional, Pure, Meditative*). Patra et al. (2013a) used the standard MIREX taxonomy for their experiments whereas Ujlambkar and Attar, (2012) experimented based on audio features for five mood classes, namely *Happy, Sad, Silent, Excited* and *Romantic* along with three or more subclasses based on two dimensional "*Energy* and *Stress*" model.

Mood Classification using Audio Features: Automatic music mood classification systems based on the audio features where spectral, rhythm and intensity are the most popular features, have been developed in the last few decades. The Music Information Retrieval eXchange[2] (MIREX) is an annual evaluation campaign of different Music Information Retrieval (MIR) related systems and algorithms. The "Audio Mood Classification (AMC)" task has been running each year since 2007 (Hu et al., 2008). Among the various audio-based approaches tested at MIREX, spectral features and Support Vector Machine (SVM) classifiers were widely used and found quite effective (Hu and Downie, 2010a). The "Emotion in Music" task was started in the year 2014 at MediaEval Benchmark Workshop. In the above task, the arousals and valence scores were estimated continuously in time for every music piece using several regression models[3].

Notable work on music mood classification using audio features can be found on several music categories, such as Hindi music (Ujlambkar and Attar, 2012; Patra et al., 2014a, 2014b), Hindustani classical music (Velankar and Sahasrabuddhe, 2012) and Carnatic classical music (Koduri and Indurkhya, 2010).

Mood Classification from Lyric Features: Multiple experiments have been carried out on western music mood classification based on bag of words (BOW), emotion lexicons and other stylistic features (Zaanen and Kanters, 2010; Hu and Downie, 2010a, 2010b).

Multi-modal Music Mood Classification: Much literature on mood classification on western music has been published based on both audio and lyrics (Hu and Downie, 2010). The system developed by Yang and Lee, (2004) is often regarded as one of the earliest studies on combining lyrics and audio features to develop a multi-modal music mood classification.

To the best of our knowledge, Indian music mood classification based on lyrics has not been attempted yet. Moreover, in context to Indian music, multi-modal music mood classification also has not been explored either.

3 Taxonomy

In context to the western music, the Adjective list (Hevner, 1936), Russell's circumplex model (Russell, 1980) and MIREX taxonomy (Hu et al., 2008) are the most popular mood taxonomies used by several worldwide researchers in this arena. Though, several mood taxonomies have been proposed by different researchers, all such psychological models were proposed in laboratory settings and thus were criticized for the lack of social context of music listening (Hu, 2010; Laurier et al., 2009).

Russell (1980) proposed the *circumplex model* of affect (consisting of 28 affect words) based on the two dimensions denoted as "*pleasant-unpleasant*" and "*arousal-sleep*" (as shown in Figure 1). The most well-known example of such taxonomy is the *Valence-Arousal* (V-A) representation which has been used in several previous experiments (Soleymani et al., 2013). *Valence* indicates *positive* versus *negative* polarity whereas *arousal* indicates the *intensity* of moods.

[2] www.music-ir.org/mirex/wiki/MIREX_HOME

[3] http://www.multimediaeval.org/mediaeval2015/emotion inmusic2015/

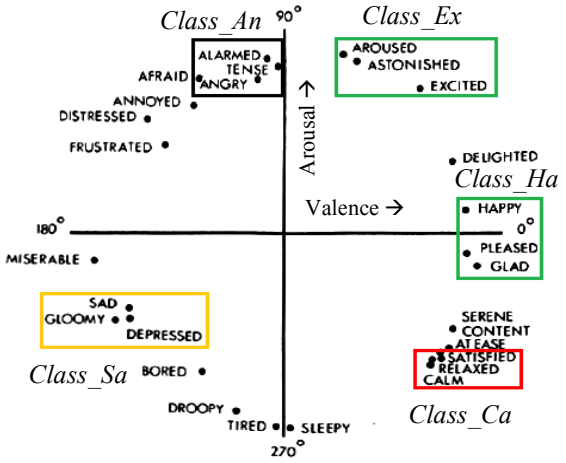

Figure 1. Russell's circumplex model of 28 affect words.

We opted to use Russel's circumplex model by clustering the similar affect words (as shown in Figure 1). For example, we have considered the affect words *calm, relaxed,* and *satisfied* together to form one mood class i.e., *Calm,* denoted as *Class_Ca*. The present mood taxonomy contains five mood classes with three subclasses in each.

One of the main reasons of developing such taxonomy was to collect similar songs and cluster them into a single mood class. Preliminary observations showed significant invariability in case of audio features of the subclasses over its corresponding main or coarse class. Basically the preliminary observations of annotation are related with the psychological factors that influence the annotation process while annotating a piece of music after listening to the song. For example, a *happy* and a *delighted* song have high valence, whereas an *aroused* and an *excited* songs have high arousal. The final mood taxonomy used in our experiment is shown in Table 1.

Class_Ex	Class_Ha	Class_Ca	Class_Sa	Class_An
Excited	Delighted	Calm	Sad	Angry
Astonished	Happy	Relaxed	Gloomy	Alarmed
Aroused	Pleased	Satisfied	Depressed	Tensed

Table 1. Proposed Mood Taxonomy

4 Dataset Annotation Perspective based on Listeners and Readers

Till date, there is no such mood annotated lyrics corpus available on the web. In the present work, we collected the lyrics data from different web archives corresponding to the audio data that was developed by Patra et al. (2013). Some more lyrics were added as per the increment of the audio data in Patra et al. (2013). The lyrics are basically written in *Romanized English* characters. The pre-requisite resources like Hindi sentiment lexicons and stopwords are available in utf-8 character encoding. Thus, we transliterated the *Romanized English* lyric to utf-8 characters using the transliteration tool available in the EILMT project[4]. As we observed several errors in the transliteration process and hence corrected the mistakes manually.

It has to be mentioned that we have only used the coarse grain classes for all of our experiments. Also to be noted that, we started annotating the lyrics at the same time of annotating their corresponding audio files by listening to them. All of the annotators were undergraduate students worked voluntarily and belong to the age group of 18-24. Each of the songs was annotated by five annotators. We achieved the inter-annotator agreement of 88% for the lyrics data annotated with five coarse grain mood classes (as mentioned in bold face in Table 1). While annotating the songs, we observed that the confusions occur between the pair of mood classes like "*Class_An* and *Class_Ex*", "*Class_Ha* and *Class_Ex*" and "*Class_Sa* and *Class_Ca*" as these classes have similar acoustic features.

To validate the annotation in a consistent way, we tried to assign our proposed coarse mood classes (e.g., *Class_Ha*) to a lyric after reading its lexical contents. But, it was too difficult to annotate a lyric with such coarse mood classes as a lyric of a single song may contain multiple emotions within it. On the other hand, the annotators felt different emotions while listening to audio and reading its corresponding lyrics, separately. For example, *Bhaag D.K.Bose Aandhi Aayi*[5] is annotated as *Calss_An* while listening to it, whereas annotated as *Class_Sa* while reading the corresponding lyric. Therefore, in order to avoid such problem and confusion, we decided to annotate lyrics with one of the coarse grained sentiment classes, viz. *positive* or *negative*.

We calculated the inter-annotator agreement and obtained 95% agreement on the lyrics data annotated with two coarse grained sentiment classes. In order to emphasize the annotation schemes, we could argue that a song is generally considered as *positive* if it belongs to the *happy* mood class. But, in our case, we observed a different scenario. Initially, the annotators annotated

[4] http://tdil-dc.in/index.php?option=com_ vertical&parentid=72

[5] http://www.lyricsmint.com/2011/05/bhaag-dk-bose-aandhi-aayi-delhi-belly.html

a lyric with *Class_Ha* after listening to audio, but, later on, the same annotator annotated the same lyric with *negative* polarity while finished reading of its contents. Therefore, a few cases where the mood class does not always coincide with the conventional moods at lyrics level (e.g., *Class_Ha* and *positive*, *Class_An* and *negative*) are identified and we presented a confusion matrix in Table 2.

	Positive	Negative	No. of Songs
Class_An	1	49	50
Class_Ca	83	12	95
Class_Ex	85	6	91
Class_Ha	96	4	100
Class_Sa	7	117	125
Total Songs			**461**

Table 2. Confusion matrix of two annotation schemes and statistics of total songs.

5 Classification Framework

We adopted a wide range of textual features such as sentiment Lexicons, stylistic features and n-grams in order to develop the music mood classification framework. We have illustrated all the features below.

5.1 Features based on Sentiment Lexicons: We used three Hindi sentiment lexicons to classify the sentiment words present in the lyrics texts, which are Hindi Subjective Lexicon (HSL) (Bakliwal et al., 2012), Hindi SentiWordnet (HSW) (Joshi et al., 2010) and Hindi Wordnet Affect (HWA) (Das et al., 2012). HSL contains two lists, one is for adjectives (3909 *positive*, 2974 *negative* and 1225 *neutral*) and another is for adverbs (193 *positive*, 178 *negative* and 518 *neutral*). HSW consists of 2168 *positive*, 1391 *negative* and 6426 *neutral* words along with their parts-of-speech (POS) and synset id extracted from the Hindi WordNet. HWA contains 2986, 357, 500, 3185, 801 and 431 words with their parts-of-speech from *angry*, *disgust*, *fear*, *happy*, *sad* and *surprise* classes, respectively. The statistics of the sentiment words found in the whole corpus using three sentiment lexicons are shown in Table 3.

5.2 Text Stylistic Features: The text stylistic features such as the number of unique words, number of repeated words, number of lines, number of unique lines and number of lines ended with same words were considered in our experiments.

Classes	HWA	Classes	HSL	HSW
Angry	241	**Positive**	1172	857
Disgust	13			
Fear	13			
Happy	349	**Negative**	951	628
Sad	107			
Surprise	38			

Table 3. Sentiment words identified using HWA, HSL and HSW

5.3 Features based on N-grams: Many researches showed that the N-Gram feature works well for lyrics mood classification (Zaanen and Kanters, 2010) as compared to the stylistic or sentiment features. Thus, we considered Term Frequency-Inverse Document Frequency (TF-IDF) scores of up to trigrams as the results get worsen after including the higher order N-Grams. However, we removed the stopwords while considering the n-grams and considered the N-Grams having document frequency more than one.

We used the correlation based supervised feature selection technique available in the WEKA[6] toolkit. Finally, we performed our experiments with 10 sentiment features, 13 textual stylistic features and 1561 N-Gram features.

6 Results and Discussion

Support Vector Machines (SVM) is widely used for the for the western songs lyrics mood classification (Hu et al., 2009; Hu and Downie, 2010a). Even for the mood classification from audio data at MIREX showed that the LibSVM performed better than the SMO algorithm, K-Nearest Neighbors (KNN) implemented in the WEKA machine learning software (Hu et al., 2008).

To develop the automatic system for mood classification from lyrics, we have used several machine learning algorithms, but the LibSVM implemented in the WEKA tool performs better than the other classifiers available for the classification purpose in our case also. Initially, we tried LibSVM with the polynomial kernel, but the radial basic function kernel gave better results. In order to get reliable accuracy, we have performed 10-fold cross validation for both the systems.

We developed two systems for the data annotated with two different annotation schemes. In the first system, we tried to classify the lyrics into five coarse grained moods classes. In the

[6] http://www.cs.waikato.ac.nz/ml/weka/

second system, we classified the polarities (*positive* or *negative*) of the lyrics that were assigned to a song only after reading its corresponding lyrics. We have shown the system F-measure in Table 4.

In Table 4, we observed that the F-measure of the second system is high compared to the first system. In case of English, the maximum accuracy achieved in Hu and Downie (2010) is 61.72 over the dataset of 5,296 unique lyrics comprising of 18 mood categories. But, in case of Hindi, we achieved F-score of 38.49 only on a dataset of 461 lyrics and with five mood classes. The observations yield the facts that the lyrics patterns for English and Hindi are completely different. We have observed various dissimilarities (w.r.t. singer and instruments) of the Hindi Songs over the English music. There are multiple moods in a Hindi lyric and the mood changes while annotating a song at the time of listening to the audio and reading its corresponding lyric.

To the best of our knowledge, there is no existing system available for lyrics based mood classification in Hindi. As the lyrics data is developed on the audio dataset in Patra et al., (2013a), thus, we compared our lyrics based mood classification system with the audio based mod classification system developed in the Patra et al., (2013a; 2013b). Our lyrics based system performed poorly as compared to the audio based systems (accuracies of 51.56% and 48%), although lyrics dataset contain more instances than the audio based system. They divided a song into multiple audio clips of 60 seconds, whereas we considered the total lyrics of a song for our experiment. This may be one of the reasons for the poor performance of the lyrics based mood classification system as the mood varies over a full length song. But in the present task, we performed classification task on a whole lyric. It is also observed that, in context of Hindi songs, the mood aroused while listening to the audio is different from the mood aroused at the time of reading a lyric. The second system achieves the best F-measure of 68.30. We can observe that the polarity all over the music does not change, i.e. if a lyric is positive, then the positivity is observed through the lyric. We also observed that the N-Gram features yield F-measure of 35.05% and 64.2% alone for the mood and polarity classification systems respectively. The main reason may be that the Hindi is free word order language. The Hindi lyrics are also more free word order than the Hindi language itself as it matches the end of each line.

7 Conclusion and Future Work

In this paper, we proposed mood and polarity classification systems based on the lyrics of the songs. We achieved the best F-measure of 38.49 and 68.3 in case of the mood and polarity classification of Hindi songs, respectively. We also observed that the listener's perspective and reader's perspective of emotion are different in case of audio and its corresponding lyrics. The mood is transparent while adopting the audio only, where the polarity is transparent in case of lyrics.

In future, we plan to perform the same experiment on a wider set of textual features. Later on, we plan to develop a hybrid mood classification system based audio and lyrics features. We also plan to improve accuracy of the lyrics mood classification system using multi-level classification.

Acknowledgments

The first author is supported by Visvesvaraya Ph.D. Fellowship funded by Department of Electronics and Information Technology (DeitY), Government of India. The authors are also thankful to the anonymous reviewers for their helpful comments.

Systems	Features	Precision	Recall	F-Measure
System 1: Mood Classification	Sentiment Lexicon (SL)	29.82	29.8	29.81
	SL+Text Stylistic (TS)	33.60	33.56	33.58
	N-Gram (NG)	34.1	36.0	35.05
	SL+TS+ NG	**40.58**	**36.4**	**38.49**
System 2: Polarity Classification	SL	62.30	62.26	65.28
	SL+TS	65.54	65.54	65.54
	NG	65.4	63.0	64.2
	SL+TS+NG	**70.30**	**66.30**	**68.30**

Table 4. System performance

References

Aditya Joshi, A. R. Balamurali and Pushpak Bhattacharyya. 2010. A fall-back strategy for sentiment analysis in Hindi: a case study. In: *Proc. of the 8ht International Conference on Natural Language Processing (ICON* -2010).

Akshat Bakliwal, Piyush Arora and Vasudeva Varma. 2012. Hindi subjective lexicon: A lexical resource for Hindi polarity classification. In: *Proc. of the 8th International Conference on Language Resources and Evaluation (LREC)*.

Aniruddha M. Ujlambkar and Vahida Z. Attar. 2012. Mood classification of Indian popular music. In: *Proc. of the CUBE International Information Technology Conference*, pp. 278-283. ACM.

Braja G. Patra, Dipankar Das and Sivaji Bandyopadhyay. 2013a. Automatic music mood classification of Hindi songs. In: *Proc. of 3rd Workshop on Sentiment Analysis where AI meets Psychology (IJCNLP 2013),* Nagoya, Japan, pp. 24-28.

Braja G. Patra, Dipankar Das and Sivaji Bandyopadhyay. 2013b. Unsupervised approach to Hindi music mood classification. In: *Mining Intelligence and Knowledge Exploration*, pp. 62-69. Springer International Publishing.

Braja G. Patra, Promita Maitra, Dipankar Das and Sivaji Bandyopadhyay. 2015. MediaEval 2015: Feed-Forward Neural Network based Music Emotion Recognition. In *MediaEval 2015 Workshop*, September 14-15, 2015, Wurzen, Germany.

Braja G. Patra, Dipankar Das and Sivaji Bandyopadhyay. 2015. Music Emotion Recognition. In *International Symposium Frontiers of Research Speech and Music (FRSM - 2015)*.

Cyril Laurier, Mohamed Sordo, Joan Serra and Perfecto Herrera. 2009. Music mood representations from social tags. In: *Proc. of the ISMIR*, pp. 381-386.

Dan Yang and Won-Sook Lee. 2004. Disambiguating music emotion using software agents. In: *proc. of the ISMIR*, pp. 218-223.

Dipankar Das, Soujanya Poria and Sivaji Bandyopadhyay. 2012. A classifier based approach to emotion lexicon construction. In: *proc. of the 17th International conference on Applications of Natural Language Processing to Information Systems (NLDB - 2012)*, G. Bouma et al. (Eds.), Springer, LNCS vol. 7337, pp. 320–326.

Gopala K. Koduri and Bipin Indurkhya. 2010. A behavioral study of emotions in south Indian classical music and its implications in music recommendation systems. In: *Proc. of the ACM workshop on Social, adaptive and personalized multimedia interaction and access*, pp. 55-60. ACM.

James A. Russell. 1980. A Circumplx Model of Affect. *Journal of Personality and Social Psychology*, 39(6):1161-1178.

Kate Hevner. 1936. Experimental studies of the elements of expression in music. *The American Journal of Psychology*: 246-268.

Makarand R. Velankar and Hari V. Sahasrabuddhe. 2012. A pilot study of Hindustani music sentiments. In: *Proc. of 2nd Workshop on Sentiment Analysis where AI meets Psychology (COLING-2012).* IIT Bombay, Mumbai, India, pages 91-98.

Menno V. Zaanen and Pieter Kanters. 2010. Automatic mood classification using TF*IDF based on lyrics. In: *proc. of the ISMIR*, pp. 75-80.

Mohammad Soleymani, Micheal N. Caro, Erik M. Schmidt, Cheng-Ya Sha and Yi-Hsuan Yang. 2013. 1000 songs for emotional analysis of music. In: *Proc. of the 2nd ACM international workshop on Crowdsourcing for multimedia*, pp. 1-6. ACM.

Philip J. Stone, Dexter C. Dunphy and Marshall S. Smith. 1966. *The General Inquirer: A Computer Approach to Content Analysis.* MIT Press, Cambridge, US.

Rada Mihalcea and Carlo Strapparava. 2012. Lyrics, music, and emotions. In: *Proc. of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 590-599. Association for Computational Linguistics.

Xiao Hu, J. Stephen Downie, Cyril Laurier, Mert Bay and Andreas F. Ehmann. 2008. The 2007 MIREX Audio Mood Classification Task: Lessons Learned. In proceedings of the 9th International Society for Music Information Retrieval Conference (ISMIR 2008), pages 462-467.

Xiao Hu, Stephen J. Downie and Andreas F. Ehmann. 2009. Lyric text mining in music mood classification. In proceedings of 10th International Society for Music Information Retrieval Conference (ISMIR 2009), pages 411-416.

Xiao Hu and J. Stephen Downie. 2010a. Improving mood classification in music digital libraries by combining lyrics and audio. In: *Proc. of the 10th annual joint conference on Digital libraries*, pp. 159-168. ACM.

Xiao Hu and J. Stephen Downie. 2010b. When lyrics outperform audio for music mood classification: A feature analysis. In: *proc. of the ISMIR*, pp. 619-624.

Xiao Hu. 2010. Music and mood: Where theory and reality meet. In: *proc. of the iConference.*

Xiao Hu, J. Stephen Downie, Cyril Laurier, Mert Bay and Andreas F. Ehmann. 2008. The 2007 MIREX

audio mood classification task: lessons learned. In: *proc. of the ISMIR,* pp. 462-467.

Yi-Hsuan Yang, Yu-Ching Lin, Ya-Fan Su and Homer H. Chen. 2008. A regression approach to music emotion recognition. *IEEE Transactions on Audio, Speech, and Language Processing,* 16(2): 448-457.

Youngmoo E. Kim, Erik M. Schmidt, Raymond Migneco, Brandon G. Morton, Patrick Richardson, Jeffrey Scott, Jacquelin A. Speck and Douglas Turnbull. 2010. Music emotion recognition: A state of the art review. In: *proc. of the ISMIR*, pp. 255-266.