

Automatic Music Mood Classification of Hindi Songs

Braja Gopal Patra

Dept. of Computer Science & Engineering,
Jadavpur University, Kolkata, India
brajagopal.cse@gmail.com

Dipankar Das

Dept. of Computer Science & Engineering,
NIT Meghalaya, India
dipankar.dipnil2005@gmail.com

Sivaji Bandyopadhyay

Dept. of Computer Science & Engineering,
Jadavpur University, Kolkata, India
sivaji_cse_ju@yahoo.com

Abstract

The popularity of internet, downloading and purchasing music from online music shops are growing dramatically. As an intimate relationship presents between music and human emotions, we often choose to listen a song that suits our mood at that instant. Thus, the automatic methods are needed to classify music by moods even from the uploaded music files in social networks. However, several studies on Music Information Retrieval (MIR) have been carried out in recent decades. In the present task, we have built a system for classifying moods of Hindi songs using different audio related features like rhythm, timber and intensity. Our dataset is composed of 230 Hindi music clips of 30 seconds that consist of five mood clusters. We have achieved an average accuracy of 51.56% for music mood classification on the above data.

1 Introduction

Music, also referred as the “language of emotion” can be categorized in terms of its emotional associations (Kim et al., 2010). Music perception is highly intertwined with both emotion and the context (Bischoff et al., 2009). Due to explosive growth of information and multimedia technologies, digital music has become widely available in different forms of digital format. Thus, the management and retrieval of such music is necessary for accessing music according to their meanings in respective songs. Nowadays, people are more interested in creating music library which allows the accessing of songs in accord-

ance with the music moods rather than their title, artists and or genre. Thus, classifying and retrieving music with respect to emotions has become an emerging research area.

The emotional meaning of the music is subjective and it depends upon many factors including culture (Lu et al., 2006). Moreover, the mood category of a song varies depending upon several psychological conditions of the Human Beings. Representations of music mood with the psychology remain an active topic for research. Apart from such challenges, there are several computational models available for mood classification. On the other hand, the collection of the “ground truth” data is still an open challenge. A variety of efforts have been made towards the collecting labeled data such as listeners’ survey, social tags, and data collection games (Kim et al., 2010).

In our present work, we have developed an automatic mood classifier for Hindi music. Hindi is the national language of India. Hindi songs are one of the popular categories of Indian songs and are present in Bollywood movies. Hindi songs make up 72% of the music sales in India¹. Mainly, we have concentrated on the collection of Hindi music data annotated with five mood classes². Then, a computational model has been developed to identify the moods of songs using several high and low level audio features. We have employed the decision tree classifier (J48) and achieved 51.56% of reasonable accuracy on a data set of 230 songs of five mood clusters.

¹ http://en.wikipedia.org/wiki/Music_of_India

² The term class and cluster are used interchangeably in this paper.

The rest of the paper is organized in the following manner. Section 2 briefly discusses the related work available to date. Section 3 provides an overview of the data and mood taxonomy used in the present experiments while Section 4 describes the feature selection for implementing machine learning algorithm. Section 5 presents the experiments with detailed analysis of results. Finally, conclusions are drawn and future directions are presented in Section 6.

2 Related Works

Music classification has received much attention by the researchers in MIR research in the recent years. In the MIR community, Music Information Retrieval Evaluation eXchange³(MIREX) is an annual competition on several important music information retrieval tasks since 2004. The music mood classification task was included into MIREX in the year of 2007. Many tasks were presented related to music classification such as Genre Classification, Mood classification, Artist Identification, Instrument Identification and Music Annotation etc. We have only surveyed the papers related to music mood classification.

Considerable amount of work has been done on the music mood classification based on audio, lyrics, social tags and all together or in a multi modal approach as described in (Yang et al., 2008; Bischoff et. al., 2009; Kim et al., 2010). Many tasks have been done on the English music mood classification such as lyrics (Hu et. al., 2009a; Hu et. al., 2009b), audio (Lu et al., 2006; Fu et al., 2011) and both (Laurier et al., 2008; Bischoff et al., 2009). Some of the works in Chinese music have been conducted based on audio (Liu et al., 2003) and lyrics (Yang et al., 2008).

Another issue that closely related with mood classification is to identify the appropriate taxonomy for classification. Ekman (1993) has defined six basic emotion classes such as *happy, sad, fear, surprise, anger and disgust*. However, these classes have been proposed for the image emotion classification as we cannot say a piece of music is disgust. In music psychology, our traditional approach is to describe mood using the adjective like *gloomy, pathetic and hopeful* etc. However, there is no standard taxonomy available which is acceptable to the researchers.

Russel (1980) proposed the circumplex model of affect based on the two dimensional model. These two dimensions are denoted as “pleasant-

unpleasant” and “arousal-sleep”. There are 28 affect words in Russel’s circumplex models and are shown in Figure 1. Later on, Thayer (1989) adapted Russel’s model using the two dimensional energy-stress model. Different researchers used their own taxonomies which are the subsets of Russel’s taxonomy. For example, Katayose et al. (1988) used all the adjectives including *Gloomy, Urbane, Pathetic and Serious*. Yang et al., (2008) used *Contentment, Depression, Exuberance and Anxious/Frantic* as mood taxonomy. MIREX (Hu et al., 2008) has five mood clusters and each cluster has more than four sub classes.

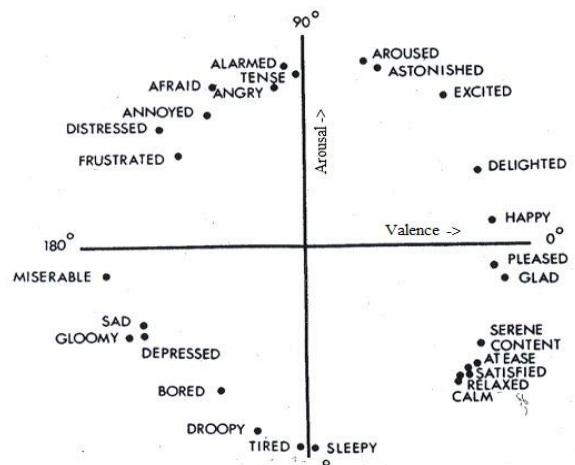


Figure 1. Russell’s circumplex model of 28 affects words

To the best of our knowledge, no work has been carried out on Hindi music mood classification. However, Velankar and Sahasrabudhe (2012) had worked on the preparation of data for Hindustani classical music mood classification. They have performed several sessions for classifying the three Indian Ragas into 13 mood classes.

3 Mood Taxonomy and Data Set

In the present task, a standard data set has been used for the mood classification task. This data has been collected manually and prepared by five human annotators. The songs used in the experiments are collected from Indian Hindi music website⁴. This collected data set includes five moods clusters of MIREX. MIREX mood cluster follows the Theyer’s model (Thayer, 1989).

³ http://www.musicir.org/mirex/wiki/MIREX_HOME

⁴ http://www.songspk.name/bollywood_songs.html

Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
Rowdy	Amiable/ Good natured	Literate	Witty	Volatile
Rousing		Wistful	Humorous	Fiery
Confident	Sweet	Bittersweet	Whimsical	Visceral
Boisterous	Fun	Autumnal	Wry	Aggressive
Passionate	Rollicking	Brooding	Campy	Tense/anxious
	Cheerful	Poignant	Quirky	Intense
			Silly	

Table 1. Five mood cluster of MIREX mood taxonomy

This evaluation forum provides a standard taxonomy for mood classification and many researchers have also used this mood taxonomy (Hu et al., 2008; Cyril et al., 2009). We have also used the MIREX mood taxonomy and are shown in Table 1. Each mood cluster contains five or more moods.

We have faced several problems during the annotation of music. First problem was whether it would be better to ignore the lyrics or not. In Hindi music, we have observed several songs have different music as well as different lyrics. For example, a music having high valence consists of the lyric that belongs to the sad mood class. Hu et al. (2008) prepared the data based on music only and the lyrics of the song were not considered in their work. So, we also tried to avoid the lyrics of the song as much as possible to build a *ground-truth* set.

Second problem is the time frame for a song. We have considered only the first 30 seconds of the song so as to prepare our data. In this frame, some lyrics might present for some of the songs. We have only included the songs that contain lyrics of less than 10 seconds. Finally, we have collected total 230 music tracks out of which 50 tracks were considered from each of the mood clusters except the *cluster5* that contains only 30 tracks.

4 Feature Selection

It is observed that the feature selection always plays an important role in building a good pattern classifier. Thus, we have considered the intensity, timbre and rhythm as the key features for our mood classification task. Kim et al., (2010) also found that tempo, sound level, spectrum and articulation are highly related to various emotional expressions. Different patterns of the acoustic cues are also associated with different emotional expressions. For example, exuberance is associated with fast tempo, high sound and bright tim-

bre whereas sadness is with slow tempo, low sound and soft timbre. In our approach, we have concentrated on the features like rhythm, intensity and timbre.

Rhythm Feature: Rhythm strength, regularity and tempo are closely related with people’s moods or responses (Liu et al., 2003). For example, generally, it is observed that, in *Exuberance* cluster, the rhythm is usually strong and steady; tempo is fast, whereas in *Depression* cluster is usually slow and without any distinct rhythm pattern.

Intensity Feature: Intensity is an essential feature in mood detection and is used in several research works (Lu et al., 2006; Liu et al., 2003). Intensity of the *Exuberance* cluster is high, and little in *Depression* cluster. Intensity is approximated by the signal’s root mean square (RMS).

Timbre Feature: Many existing researchers have shown that mel-frequency cepstral coefficients (MFCCs), so called spectral shapes and spectral contrast are the best features for identifying the mood of music (Lu et al., 2006; Liu et al., 2003; Fu et al., 2011). In this paper, we have used both spectral shape and spectral contrast. Spectral shape includes centroid, band width, roll off and spectral flux. Spectral contrast features includes sub-band peak, sub-band velly, sub-band contrast.

Class	Features
Rhythm	Rhythm strength, regularity and tempo
Timbre	MFCCs, Spectral shape, Spectral contrast
Intensity	RMS energy

Table 2. Feature used in mood classification

All the features used in our experiments are listed in Table 2. These features are extracted

using jAudio⁵ toolkit. It is a music feature extraction toolkit developed in JAVA platform. The jAudio toolkit is publicly available for research purpose.

5 Experiments and Evaluation

It is obvious that in order to achieve good results, we require a huge amount of mood annotated music corpus for applying the statistical models. But, to the best of our knowledge, no mood annotated Hindi songs are available to date. Thus, we have developed the dataset by ourselves and it contains 230 songs consisting of five clusters.

The mood classification has been performed using several machine learning algorithms based on the features we discussed in Section 4. We have used the API of Weka 3.7.7.5⁶ to accomplish our classification experiments. Weka is an open source data mining tool. It presents a collection of machine learning algorithms for data mining tasks. We employed several classifiers for the mood detection problem, but the Decision tree (J48) gives the best result as compared to the other classifiers.

The features are extracted using the jAudio Feature Extractor. To get the reliable accuracy, we have performed 10 fold cross validation where the data set are randomly partitioned into 80% training and 20% for testing data. The accuracies have been calculated by the Weka toolkit and are reported in Table 3. The confusion matrix of the classification accuracy is given in Table 4. We have achieved the maximum accuracy of 55.1% in *cluster 1*. It has been observed that the *cluster 5* has lowest accuracy and is about 46.7%. This cluster contains less music as compared to other clusters. The accuracies of *cluster 2*, *cluster 3* and *cluster 4* are 52%, 50% and 54%, respectively.

We have observed that some of the instances from each of the clusters go to its neighboring cluster. For example, some songs from *cluster 2* fall under the *cluster 1* as they have similar RMS energy and tempo. It is observed that the present system achieved quite low accuracy as compared to the other existing mood classification systems for English songs (Liu et al., 2003; Lu et al., 2006) and Chinese (Yang et al., 2008) songs. But, the inclusion of additional features and the feature engineering may remove such kind of biasness and improve the results.

Class	Accuracy
Cluster 1	55.1
Cluster 2	52.0
Cluster 3	50.0
Cluster 4	54.0
Cluster 5	46.7
Average	51.56

Table 3. Accuracies of each class

Clusters	1	2	3	4	5
1	29	8	1	1	11
2	10	27	2	4	7
3	2	12	25	10	1
4	2	3	12	28	5
5	12	1	1	2	14

Table 4. Confusion matrix for the accuracy

6 Conclusion and Future Works

In this paper, we have described a preliminary approach to Hindi music mood classification that exploits simple features extracted from the audio. Three types of features are extracted from the audio, namely rhythm, intensity and timbre. MIREX mood taxonomy has been used for our experiment. We have employed the decision tree classifier (J48) for classification purpose and achieved an average accuracy of 51.56% using the 10 fold cross validation.

There are several directions for future work. One of them is to incorporate more audio features for enhancing the current results of mood classification. Later on lyrics of the song may be incorporated for multimodal mood classification. Preparing the large audio data and collecting the lyrics of those songs may be considered as the other future direction of research.

Acknowledgments

The work reported in this paper is supported by a grant from the India-Japan Cooperative Programme (DST-JST) 2009 Research project entitled ‘‘Sentiment Analysis where AI meets Psychology’’ funded by Department of Science and Technology (DST), Government of India.

References

Cyril Laurier, Jens Grivolla and Perfecto Herrera. 2008. Multimodal music mood classification using audio and lyrics. In *proceedings of the Seventh In-*

⁵ <http://sourceforge.net/projects/jmir/files/>

⁶ <http://weka.wikispaces.com/>

- ternational Conference on Machine Learning and Applications (ICMLA'08), pages 688-693. IEEE.
- Cyril Laurier, Mohamed Sordo and Perfecto Herrera. 2009. Mood cloud 2.0: Music mood browsing based on social networks. In *Proceedings of the 10th International Society for Music Information Conference (ISMIR 2009)*, Kobe, Japan.
- Dan Liu, Lie Lu and Hong-Jiang Zhang. 2003. Automatic Mood Detection from Acoustic Music Data. In *proceedings of the International Society for Music Information Retrieval Conference (ISMIR 2003)*.
- Haruhiro Katayose, Hasakazu Imai and Seiji Inokuchi. 1988. Sentiment extraction in music. In *proceedings of the 9th International Conference on Pattern Recognition, 1988*, pages 1083-1087. IEEE.
- James A. Russell. 1980. A Circumplex Model of Affect. *Journal of Personality and Social Psychology*, 39(6):1161-1178.
- Kerstin Bischoff, Claudiu S. Firan, Raluca Paiu, Wolfgang Nejdl, Cyril Laurier and Mohamed Sordo. 2009. Music Mood and Theme Classification-a Hybrid Approach. In *proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, pages 657-662.
- Lie Lu, Dan Liu and Hong-Jiang Zhang. 2006. Automatic mood detection and tracking of music audio signals. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(1):5-18.
- Makarand R. Velankar and Hari V. Sahasrabudde. 2012. A Pilot Study of Hindustani Music Sentiments. In *Proceedings of 2nd Workshop on Sentiment Analysis where AI meets Psychology (COLING-2012)*. IIT Bombay, Mumbai, India, pages 91-98.
- Paul Ekman. 1993. Facial expression and emotion. *American Psychologist*, 48(4):384-392.
- Robert E. Thayer. 1989. *The Biopsychology of Mood and Arousal*. Oxford University Press, NY.
- Soujanya Poria, Alexander Gelbukh, Amir Hussain, Sivaji Bandyopadhyay and Newton Howard. 2013. Music Genre Classification: A Semi-supervised Approach. In *Pattern Recognition*, Springer Berlin Heidelberg, pages 254-263.
- 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. 2009a. Lyric text mining in music mood classification. In *proceedings of 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, pages 411-416.
- Yajie Hu, Xiaoou Chen and Deshun Yang. 2009b. Lyric-based Song Emotion Detection with Affective Lexicon and Fuzzy Clustering Method. In *proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, pages 123-128.
- Yi-Hsuan Yang, Chia-Chu Liu and Homer H. Chen-Yang. 2006. Music emotion classification: a fuzzy approach. In *Proceedings of the 14th annual ACM international conference on Multimedia*, pages 81-84. ACM.
- Yi-Hsuan Yang, Yu-Ching Lin, Heng-Tze Cheng, I-Bin Liao, Yeh-Chin Ho and Homer H. Chen. 2008. Toward multi-modal music emotion classification. In *proceedings of the Advances in Multimedia Information Processing (PCM-2008)*. Springer, Berlin Heidelberg, pages 70-79.
- 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 *proceedings of 11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, pages 255-266.
- Zhouyu Fu, Guojun Lu, Kai Ming Ting and Dengsheng Zhang. 2011. A survey of audio-based music classification and annotation. *IEEE Transactions on Multimedia*, 13(2): 303-319.