Computación y Sistemas
ISSN: 1405-5546
computacion-y-sistemas@cic.ipn.mx
Instituto Politécnico Nacional
México

Gopal Patra, Braja; Das, Dipankar; Bandyopadhyay, Sivaji
Multimodal Mood Classification Framework for Hindi Songs
Computación y Sistemas, vol. 20, núm. 3, 2016, pp. 515-526
Instituto Politécnico Nacional
Distrito Federal, México

Multimodal Mood Classification Framework for Hindi Songs

Braja Gopal Patra, Dipankar Das, Sivaji Bandyopadhyay

Department of Computer Science & Engineering,
Jadavpur University, Kolkata,
India

brajagopalcse@gmail.com, dipankar.dipnil2005@gmail.com, sivaji_cse_ju@yahoo.com

Abstract. Music information retrieval is currently an active domain of research. An interesting aspect of music information retrieval involves mood classification. While the Western music captured much attention, research on Indian music was limited and mostly based on audio data. In this work, the authors propose a mood taxonomy and describe the framework for developing a multimodal dataset (audio and lyrics) for Hindi songs. We observed differences in mood for several instances of Hindi songs while annotating the audio of such songs in contrast to their corresponding lyrics. Finally, the mood classification frameworks were developed for Hindi songs and they consist of three different systems based on the features of audio, lyrics and both. The mood classification systems based on audio and lyrics achieved F-measures of 58.2% and 55.1%, respectively whereas the multimodal system (combination of both audio and lyrics) achieved the maximum F-measure of 68.6%.

Keywords. Hindi songs, mood classification, multimodal dataset, mood taxonomy, audio, lyrics.

1 Introduction

The first decade of 21st century witnessed the growth and popularity of music distribution in CDs, DVDs or other portable formats. Another important change was also witnessed recently when the internet connectivity led to the rapid growth in downloading and purchasing of music online. The number of music compositions created worldwide already exceeds a few millions and continues to grow. This fact enhances the importance of developing an automated process for music organization, management, search as well as the generation of playlists and various other music related applications.

Over the centuries, music has shared a very special relationship with human moods and the impact of music on moods has been well documented [16]. We often listen to a song or music which best fits to our mood at that instant of time. Naturally, such phenomenon motivates the music composers and singers and/or performers to express their emotions through piece of songs [20]. It has been observed that people are interested in creating music library that allows them to access songs in accordance with their moods compared to the title, artists and/or genres [6, 26]. Further, people are also interested in creating music libraries based on several other factors e.g., what songs they like/dislike (and in what circumstances), time of the day and their state of mind [6] etc. Thus, organizing music with respect to such metadata is one of the major research areas in the field of playlist generation. Recently, music information retrieval (MIR) based on emotions or moods has attracted the researchers from all over the world because of its implications in human computer interactions.

India is considered to have one of the oldest musical traditions in the World. Hindi is one of the official languages of India and stands fourth with respect to the most widely spoken language in the World[1]. Hindi music or Bollywood music, also known as popular music [35] are mostly present in Hindi cinemas or Bollywood movies [8]. Hindi or Bollywood songs make up 72% of the total music

[1] https://www.cia.gov/library/publications/the-world-factbook/fields/2098.html

sales in India [35]. It is observed that Hindi or Bollywood songs include varieties of Hindustani classical music, folk music, pop and rock music. Indian film music is not only popular in the Indian society, but has also been on the forefront of the Indian's culture around the World [8]. Mood related experiments on Western music based on audio [14, 22], lyrics [39], and multimodal approaches [12, 37, 38], achieved promising milestones in this arena. In contrast, experiments on Indian music moods were limited, for example, mood classifications of Hindi songs were performed using only audio features [25, 26, 35] and lyric features [28]. To the best of the author's knowledge, no multimodal mood classification system was developed for Hindi songs.

In the present article, the authors propose a mood taxonomy suitable for Hindi songs and developed a multimodal mood classification framework based on both audio and lyric features. We collected the lyrics of the audio dataset prepared in Patra et al. [27] and annotated with our proposed mood taxonomy. In case of annotation, the differences in moods were observed between the audio of the songs and their corresponding lyrics. Such differences were analyzed from the perspectives of both listeners and readers. We studied various problems of annotation and developed two mood classification frameworks for Hindi songs based on the audio and lyric features, separately. Further, a multimodal mood classification framework was developed based on both audio and lyric features of Hindi songs. The results demonstrate the superiority of a multimodal approach over a uni-modal approach for mood classification of Hindi songs.

The rest of the paper is organized in the following manner. Section 2 briefly discusses the state-of-the-art mood taxonomies and music mood classification systems developed for Western and Indian songs. Section 3 provides an overview of our proposed mood taxonomy and data annotation process for Hindi songs. Section 4 describes the features collected from audio and lyrics of the Hindi songs, while Section 5 presents the mood classification systems and our findings. Finally, the conclusions and future directions are listed in Section 6.

2 Related Work

The survey work on music mood classification can be divided into two parts, one outlining the mood taxonomies proposed for the Western and Indian songs and second describing the mood classification systems developed for the Western and Indian songs till date.

2.1 Mood Taxonomies

The preparation of an annotated dataset requires the selection of proper mood taxonomies. Identifying an appropriate mood taxonomy is one of the primary and challenging tasks for mood classification. Mood taxonomies are generally categorized into three main classes namely, categorical, dimensional, and social tags [20].

Categorical representation describes a set of emotion tags organized into discrete entities according to their meaning. The earliest categorical music mood taxonomy was proposed by Hevner [10] and is known for its systematic coverage on music psychology [13]. Another traditional categorical approach uses adjectives like *gloomy*, *pathetic* and *hopeful* etc. to describe different moods [21]. On the other hand, *Music Information Retrieval eXchange*[2] (*MIREX*) community proposed a categorical mood taxonomy for audio based mood classification task [14], which is quite popular among the MIR researchers. In case of Indian music mood classification, Koduri and Indurkhya worked on the mood classification of south Indian classical music using categorical mood representation and they considered the mood taxonomy consisting of ten rasas (e.g., Srungaram (Romance), Hasyam (Laughter), Karunam (Compassion) etc.) [17]. Similarly, Velankar and Sahasrabuddhe prepared data for mood classification of Hindustani classical music consisting of 13 different mood classes (e.g., Happy, Exciting, Satisfaction etc.) [36].

Dimensional models of emotion categorization describe emotions with respect to one or more axes. The most well known example of such a space is the "*valence-arousal*" [31] or "*energy-stress*" [33] representation. The

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

ISSN 1405-5546
doi: 10.13053/CyS-20-3-2461

valence indicates positivity and negativity of emotions whereas the *arousal* indicates emotional intensity [16]. One of the earliest researches carried out on the dimensional models was proposed by Russell [31]. The author proposed the *circumplex model of affect* (consisting of 28 affect words) based on the two dimensions, denoted as "*pleasant-unpleasant*" and "*arousal-sleep*". In context of Indian music mood classification, most of the researches adopted the dimensional model. Ujlambkar and Attar used the *Russell's circumplex model of affect* to develop a mood taxonomy of five mood classes and each of the classes consists of three or more sub classes [35]. Patra et al. [27] have used five mood classes with three or more subclasses, which are the subsets of the *Russell's circumplex model of affect*.

Social tags are generally assigned by the non-experts for their own personal use, such as listeners to assist in organization and accessibility of an item [18]. Tags are typically the words or short phrases or unstructured labels that describe resources. In case of Western songs, mood classification was also performed using social tags in [18, 20].

It was observed that the Hevner's adjectives [10] are less consistent in case of the intra-cluster similarity, whereas MIREX mood taxonomy [14] suffers with inter-cluster dissimilarity and confusion between the categories were observed [20]. From the above, Laurier et al. [20] concluded that the psychological models have some similarity with the social tags though it may not be suitable for today's music listening reality [11]. In case of the Indian songs, no such social tags were collected or reported till date.

2.2 Music Mood Classification

The framework of classification systems was divided into three categories based on the type of features and experimental settings.

2.2.1 Audio based Classification

Automatic music mood classification systems were developed based on some popular audio features like *spectral*, *rhythm* and *intensity*. Such features have been used for developing several audio based music mood classification systems in the last decades [7, 12, 20]. Among the various audio based approaches tested at *MIREX*, *spectral* features were widely used and found quite effective for the mood classification of Western songs [12]. The *Emotion in Music task*[3] was started in the year 2014 at MediaEval Benchmark [32]. In the above task, the *arousal* and *valence* scores were estimated continuously for every music clip in a time frame of 0.5 seconds with the help of several regression models [30]. Several experiments were performed specially in mood classification of Western music using only audio features [14, 22].

Few works on music mood classification using audio features are found for several categories of Indian music, such as Carnatic music [17], Hindi music [9, 25, 26, 27, 29, 35], Hindustani classical music [36]. Recently, sentiment analysis of Telugu songs was performed in [1] using several audio features like *prosody*, *temporal*, *spectral*, *chroma* and *harmonic*.

2.2.2 Lyric based Classification

Lyrics based mood classification systems for Western songs were developed by incorporating bag of words (BOW), emotion and sentiment lexicons and other stylistic features in [12, 13, 39]. It was observed that the mood classification systems using lyric features performed better than the mood classification systems using audio features for Western songs [13]. In context to Indian music, Patra et al. [28] performed the mood and sentiment classification using lyric features. But, they have annotated each of the lyrics at the time of listening to its corresponding audio. The above mood classification system obtained very low F-measure of 38.49% using several lyric features of Hindi songs. The sentiment classification system achieved F-measure of

[3]http://www.multimediaeval.org/mediaeval2015/ emotionin-music2015/

68.30% using the same lyric features for Hindi songs. Abburi et al. [1] performed the sentiment analysis on the lyrics of Telugu songs using *word2vec* features.

2.2.3 Multimodal Classification

Several models on mood classification for the Western music have been developed based on both audio and lyrics [3, 12, 19]. The system developed by Yang et al. [37] is often regarded as one of the earliest studies on combining lyric and audio features in music mood classification [12]. In contrast, Indian music mood classification has been performed based on either audio or lyric features till date. To the best of our knowledge, no research on multimodal mood classification for Indian music has been performed yet. Recently, Abburi et al. [1] performed the multimodal sentiment analysis of Telugu songs using audio and lyric features. Thus, in the present attempt, we emphasized the mood classification of Hindi songs using multimodal features (combination of audio and lyric features).

3 Proposed Mood Taxonomy and Data Preparation

In this section, we described the proposed mood taxonomy and the framework for preparing lyric dataset for Hindi songs.

3.1 Proposed Mood Taxonomy

Most of the taxonomies in the literature were used for evaluating the Western music. Ancient Indian actors, dancers and musicians divided their performance into nine categories based on emotions and called the different emotions together as *Navrasa*, where *rasa* means emotions and *nav* means nine. Unfortunately in the modern context of music making, all the nine types of emotions are not frequently observed. For example, the emotions like *surprise* and *horrific* belonging to the *Navrasa* are rarely observed in current Hindi music. The emotion word *Hasya* (Happiness) need a further subdivision, for instance, *happy* and *excited*. Hence, this model cannot be used for analyzing the mood aspects of Indian popular songs [34].

Another interesting mood taxonomy for classifying Hindi music was proposed by [34] after consulting feedback of 30 users. Observation of many music tracks led us to believe that romantic songs may be associated with largely varying degrees of *arousal* and *valence*, making it difficult to categorize based on Thayer's or Russell's model. The songs from *sad* class need a further subdivision, because there are many sad songs with high *arousal*.

The comparative analysis of different mood taxonomies revealed that the clustering of similar mood adjectives has a positive impact on the classification accuracy. Based on this observation, we opted to use *Russell's circumplex model of affect* [31] by clustering the similar affect words located close to each other on the *arousal-valence* plane into a single class as shown in Figure 1. We considered the mood classes *Angry*, *Calm*, *Excited*, *Happy* and *Sad* for our experiments. Each of the classes contains another two nearby key affect words of the *circumplex model of affect*. Thus, our final mood classes are *Angry* (*Alarmed*, *Tensed*), *Calm* (*Satisfied*, *Relax*), *Excited* (*Aroused*, *Astonished*), *Happy* (*Pleased*, *Glad*) and *Sad* (*Gloomy*, *Depressed*). One of the main reasons for collecting songs and grouping the similar songs into a single mood class is to consider the significant invariability of the audio features at subclass level with respect to their main class. For example, a "Happy" and a "Delighted" song have high *valence*, whereas an "Aroused" and an "Excited" song have high *arousal*.

3.2 Data Preparation

In the present work, we collected the lyrics data from web archives corresponding to the annotated audio dataset available for Hindi songs in [27]. The lyrics are basically written in *Romanized English* characters whereas the prerequisite resources like Hindi sentiment lexicons, emotion lexicons and list of stop words are available in *utf-8* character encoding. Thus, we transliterated the *Romanized English* lyrics to *utf-8* characters using the transliteration tool available in the EILMT

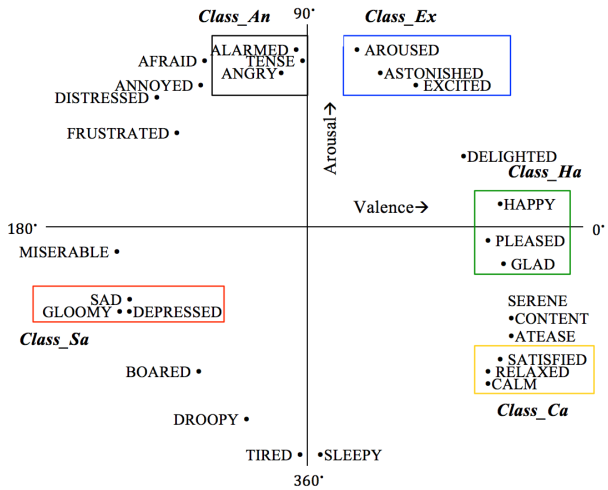

Fig. 1. Russell's circumplex model of affect [31]

project.[4] We observed several errors in the transliteration process. For example, words like 'oooohhhhooo' 'aaahhaa' were not transliterated due to the presence of repeated characters. Again, the words like 'par' and 'paar', 'jan' and 'jaan' were transliterated into different words 'पर' and 'पार', 'जन' and 'जान', but, the above pairs are the same words 'पर' and 'जान'. Hence, these mistakes were corrected manually.

Each of the lyrics was annotated by at least three annotators aged 20±4 years, who were undergraduate students and research scholars and worked as volunteers for annotating the lyrics corpus. The lyrics were asked to annotate after reading it with either of the aforementioned five mood classes. Each of the lyrics was also annotated with positive, negative, and neutral polarities. In several cases, we observed that the mood class that was assigned to an audio is different from the mood class assigned to its corresponding lyric for some of the Hindi songs. The statistics of annotation during listening to audio (L_{Audio}) and reading of the lyrics (R_{Lyrics}) are provided in Table 1.

The differences between reader's and listener's perspectives for the same song motivated us to investigate the root cause of such discrepancy. The authors believe that the subjective influence of music modulates the perception of lyrics of a song in the listeners. The poetic and metaphoric usage of language can be observed in the lyrics. For example, a song "Bhaag D.K.Bose Aandhi Aayi"[5] has mostly sad words like "*dekha to katora jaka to kuaa (the problem was much bigger than it seemed at first)*" in the lyric. This song is annotated as "*Sad*" while reading and annotated as "*Anger*" as it contains mostly the rock music and the *arousal* is also high. Similarly, a song "Dil Duba"[6] is annotated as "*Sad*" and "*Happy*" while reading the lyrics and listening to the audio, separately. This song portrays negative emotions by using sad or negative words like "*tere liye hi mar jaunga (I would die for you)*", but, the song contains high *valence*. The above observations emphasize that the combined effect of lyrics and audio is an important factor in indicating the final mood inducing characteristics of a music piece.

It was observed that the annotators were influenced by the moods perceived by the audio of the songs. It was also difficult to feel the mood of a song using only lyrics because the identification of metaphoric and poetic usage is hard without listening to the audio. Hence, the songs were considered which were annotated with the same mood class after listening to the audio as well as reading of its corresponding lyrics for further experiments. We have considered 27, 37, 45, 48 and 53 songs for *Angry*, *Calm*, *Excited*, *Happy* and *Sad* mood classes, respectively and each of these audio files was sliced into 60 seconds of clips. These clips were annotated previously done by Patra et al. [27].

The pairwise inter-annotator agreements were calculated on the dataset by computing Cohen's κ coefficient [4]. The overall inter-annotator agreement scores with five mood classes were found to be 0.80 for Hindi lyrics. However, the inter-annotator agreement was around 0.96 for the lyrics data while annotating with positive, negative, and neutral polarity.

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

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

[6] http://www.hindilyrics.net/lyrics/of-Dil%20Duba %20Dil%20Duba.html

ISSN 1405-5546
doi: 10.13053/CyS-20-3-2461

Table 1. Confusion matrix of annotated songs with respect to five mood classes [after listening to the audio (L_{Audio}) and reading of the lyrics (R_{Lyrics})]

		R_{Lyrics}					
		Angry	**Calm**	**Excited**	**Happy**	**Sad**	**Total**
L_{Audio}	**Angry**	**27**	3	15	7	13	65
	Calm	2	**37**	6	30	25	100
	Excited	13	6	**45**	25	11	100
	Happy	15	11	22	**48**	4	100
	Sad	15	32	15	10	**53**	125

4 Feature Extraction

Feature extraction plays an important role in any classification framework and depends upon the data set used for the experiments. We have considered different audio related features and textual features of lyrics for mood classification.

4.1 Audio Features

We have considered the key features like *intensity*, *rhythm*, and *timbre* for mood classification task. These features have been used for music mood classification for Indian languages in state-of-the-art systems [25, 26, 35]. We have listed the audio features used in our experiments in Table 2 and these features were extracted using the jAudio toolkit[7] [23].

4.2 Lyric Features

A wide range of textual features such as sentiment lexicons, stylistic and n-gram features were adopted in order to develop the music mood classification system.

4.2.1 Sentiment Lexicons (SL)

We used three lexicons to classify the moods present in the lyrics and these lexicons are Hindi Subjective Lexicon (HSL) [2], Hindi SentiWordnet (HSW) [15] and Hindi WordnetAffect (HWA) [5]. 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 ids extracted from the Hindi WordNet.[8] HWA contains 2986, 357, 500, 3185, 801 and 431 words with their parts-of-speech (POS) tags for angry, disgust, fear, happy, sad and surprise classes, respectively.

To the best of our knowledge, the performances of the available POS taggers and lemmatizers for Hindi language are not up to the mark. The CRF based *Shallow Parser*[9] is available for POS tagging and lemmatization, but it also did not perform well on the lyrics data because of the free word order nature of Hindi lyrics. Thus, the number words matched with these sentiment or emotion lexicons are considerably less. The statistics of the sentiment words found in the whole corpus using three sentiment lexicons are shown in Table 3. We also extracted the positive and negative words that were annotated by our annotators. We found 641 and 523 positive and negative unique words from the total corpus.

4.2.2 Text Stylistic (TS)

Text stylistic features have been used effectively in text stylometric analysis such as authorship identification, author identification [24]. These features have also been used for mood classification from lyrics of Western music [12]. The TS features such as the number of unique words, number of repeated words, and number of lines etc. were considered in our experiments. The detailed list of TS features along with their descriptions is given in Table 4.

[7] sourceforge.net/projects/jmir/files/

[8] http://www.cfilt.iitb.ac.in/wordnet/webhwn/

[9] http://ltrc.iiit.ac.in/analyzer/hindi/

ISSN 1405-5546
doi: 10.13053/CyS-20-3-2461

Table 2. List of features extracted from audio

Feature class	Features
Timbre	Spectral Centroid, Spectral Rolloff Point, Spectral Flux, Spectral Variability, Mel-frequency cepstral coefficients (MFCCs), Linear Predictive Coefficient (LPC), Partial Based Spectral Centroid, Partial Based Spectral Flux
Intensity	Root Mean Square, Fraction of Low Energy Windows
Rhythm	Beat Histogram, Strongest Beat, Beat Sum, Strength of Strongest Beat, Compactness, Method of Moments, Zero Crossings, Peak Detection, Peak Based Spectral Smoothness

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

<table>
<tr><th>Classes</th><th>HWA</th><th>Classes</th><th>HSL</th><th>HSW</th></tr>
<tr><td>Angry</td><td>210</td><td rowspan="3">Positive</td><td rowspan="3">927</td><td rowspan="3">785</td></tr>
<tr><td>Disgust</td><td>9</td></tr>
<tr><td>Fear</td><td>15</td></tr>
<tr><td>Happy</td><td>313</td><td rowspan="3">Negative</td><td rowspan="3">891</td><td rowspan="3">613</td></tr>
<tr><td>Sad</td><td>98</td></tr>
<tr><td>Surprise</td><td>35</td></tr>
</table>

4.2.3 N-grams (NG)

N-gram features work well for mood classification using lyrics [28, 39] as compared to the stylistic or sentiment features. We considered the scores of *term frequency and document frequencies* (*TF-IDF*) up to trigram levels, because including the higher order N-grams reduce the accuracy. We considered only those N-grams having document frequency more than one and removed the stopwords while considering the N-grams.

5 Supervised Framework

It was observed that the feature selection improved the performances of the mood classification systems [30]. Thus, the important features were identified from the audio and lyrics using the feature selection technique. The state-of-the-art mood classification systems achieved better results using the Support Vector Machines (SVMs) [13, 14]. Thus, the LibSVM implemented in WEKA tool[10] was used for the classification purpose. We performed 10-fold cross validation in order to get reliable accuracy.

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

5.1 Feature Selection

Feature level correlation is used to identify the most important features as well as to reduce the feature dimension [30]. Thus, the correlation based supervised feature selection technique implemented in WEKA toolkit was used to find out the important contributory features for audio and lyrics. A total of 431 audio features were extracted from the audio files using jAudio. A total of 12 sentiment features, 12 textual stylistic features and 5832 N-gram features were also collected from the lyrics. The feature selection technique implemented using Weka yields 148 important audio features and 12 sentiment, 8 stylistic, and 1601 N-gram features from lyrics. We subsequently use these features for our classification purpose.

5.2 Mood Classification using Audio Features

For music mood classification using audio features, the linear kernel of LibSVM was selected since it provides the higher F-measure in our case as compared to the polynomial kernels. We performed the classification by adding the features one by one. Initially, the timbre features were used to classify the moods, then added intensity features and then rhythm features, incrementally. After adding all the features together, the audio based mood classification system achieved the maximum F-measure of 58.2%. The contribution of each feature in F-measure is given in Table 5.

Table 4. List of TSF and description

Name of the features	Descriptions
Number (No.) of words	Total (Tot.) no. of words in a lyric
No. of unique words	Tot. no. of unique words in a lyric
No. of repeated words	Tot. no. of words with frequency more than one in a lyric
No. of lines	Tot. no. of lines in a lyric
No. of unique lines	Tot. no. of unique lines in a lyric
No. of repeated lines	Tot. no. of repeated lines in a lyric
No. of lines ended with the same words	Tot. no. of lines ended with same words in a lyric
No. of lines ended with the same characters	Tot. no. of lines ended with same characters in a lyric
Average (Avg.) words per line	No. of words/no. of lines
Avg. unique words per line	No. of unique words/no.of lines
Avg. repeated word per line	No. of repeated words/no.of lines
Repeated word ratio	No. of repeated words/no.of words

5.3 Mood Classification using Lyric Features

For mood classification using lyrics, the linear kernel was selected and the classification was performed by adding features one by one. Initially, the experiment was performed using only sentiment features, and then added other features subsequently. The maximum F-measure of 55.1% was achieved using the sentiment and N-gram features for five class mood classification of Hindi songs as shown in Table 5. We also annotated each of the lyrics with positive, negative, and neutral polarity in addition to five mood classes. The maximum F-measure of 69.8% was achieved for polarity classification system using the sentiment and N-gram features of Hindi song lyrics.

5.4 Multimodal Mood Classification

Finally, the experiments were performed using both audio and lyric features. Again, we used the linear kernel of LibSVM for the classification purpose. The TS features reduced the performance of the systems, thus these were excluded while developing the final multimodal system for mood classification of Hindi songs. The multimodal mood classification system achieved the maximum F-measure of 68.6% after adding all features for Hindi songs and the system performance is given in Table 5.

5.5 Observation and Comparison

The confusion matrix for the multimodal mood classification system is also given in Table 6. From the confusion matrix, it is observed that there is biassness in the classification system towards the nearest classes. There were confusions in between the mood class pairs such as "*Angry* & *Excited*", "*Excited* & *Happy*", "*Sad* & *Calm*", "*Sad* & *Angry*".

The audio based mood classification system developed in Patra et al. [29] achieved F-measure of 72%. The F-measure achieved by our audio based mood classification system was less as compared to the above system by around 14%. One of the main reason for such low performance is that the authors used 1540 number of audio clips whereas we used xx number of audio clips only. We had to select less number of audio clips to the cope up with the clash happened during mood level annotations of audio and lyrics. They used Feed-Forward Neural Networks (FFNNs) for identifying the moods of a song whereas we used LibSVM for mood classification. It is found in the literature that the performances of the

Table 5. System Performances

Systems	Features	Precision	Recall	F-measure
Audio Features	Timbre	55.2	54.4	54.8
	Timbre + Intensity	55.7	55.1	56.9
	Timbre + Intensity + Rhythm	58.6	57.8	**58.2**
Lyric Features	SL	41.1	39.3	40.2
	SL + TSF	38.5	38.5	38.5
	NG	46.3	46.7	46.5
	SL + TSF + NG	52.2	52.6	52.4
	SL + NG	55.7	54.5	**55.1**
Lyric Features (Polarity Classification)	SL + NG	69.7	69.9	**69.8**
Multimodal	Audio + Lyrics (Excluding TS features)	68.9	68.3	**68.6**

Table 6. Confusion matrix for the multimodal music mood classification system

		Predicted				
		Angry	**Calm**	**Excited**	**Happy**	**Sad**
Actual	**Angry**	**16**	1	5	2	3
	Calm	2	**26**	1	3	5
	Excited	10	0	**30**	4	1
	Happy	3	2	8	**33**	2
	Sad	7	5	1	1	**39**

state-of-the-art mood classification systems based on audio are better using the FFNNs [29, 30].

Ujlambkar and Attar [35] achieved the maximum F-measure of around 75-81%. Our system performed less as compared to the above state-of-the-art system by around 22%. They developed their system using a different mood taxonomy and with more number of audio files which are not available for research. Again, each of the clips was of 30 seconds for their experiment whereas in case of ours, the size of the music clips was 60 seconds. The performance obtained by our audio based mood classification system shows an improvement of around 8% over the state-of-the-art audio based mood classification systems [25, 26] (achieved accuracy of around 50%), because they used less number of audio files.

It was observed that the N-gram features yield F-measure of 46.5% alone for the mood classification system based on the lyric features. The main reason may be that the Hindi is free word order language and the Hindi lyrics are also more free in word order than the Hindi language itself. Whereas, the text stylistic features do not help much in our experiments as it reduces the F-measure of the system by around 2.7%.

The state-of-the-art mood classification system available for the Hindi songs in [28] achieved F-measure of 38% using lyric features only. Our system out performs the above system by around 17%. The main differences was that they annotated the moods to the lyrics after listening to the corresponding audio whereas in our case, we considered the moods annotated after listening to the audio and reading the corresponding lyrics to avoid the biasness in the mood annotation process. More number of lyrics were used for the current experiment. The polarity classification system outperforms by around 1% as compared to the polarity classification system available in [28]. To the best of the author's knowledge, there is no other mood classification system based on the lyric features available for Hindi songs till date. While multiple experiments were performed on mood classification of Western songs based on lyrics, but the differences in number of mood classes made comparisons among these works and our proposed method difficult.

The multimodal mood classification system achieved the maximum F-measure of 68.6% after adding all the audio and lyrics (excluding the TS features) features using LibSVM. To the best of our knowledge, there is no state-of-the-art multimodal mood classification system available for Hindi songs. Laurier et al. [19] performed multimodal mood classification of Western songs using both audio and lyric features and achieved 98.3%, 86.8%, 92.8%, and 91.7% for *Angry*, *Happy*, *Sad* and *Relaxed* mood classes, respectively. They made the classification much easier by classifying one mood class at a time, i.e. for the first time they classified "*Angry*" or "*not Angry*" and so on. Hu and Downie [12] achieved 67.5% for multimodal mood classification using late fusion on a dataset of 5,296 unique songs comprising of 18 mood classes. Our multimodal mood classification system outperforms the above system by 1%.

6 Conclusion and Future Work

The multimodal mood annotated dataset (lyrics and audio) was developed for research in music mood classification of Hindi songs. The automatic music mood classification system was developed from the above multimodal dataset and achieved the maximum F-measure of 68.6%. The different moods were perceived while listening to a song and reading the corresponding lyric of song. The main reason for this difference may be that the audio and lyrics were annotated by different annotators. Another reason may be that the moods are transparent in audio as compared to lyrics of Hindi songs. Later on, we intend to perform deeper analysis of the listener's and reader's perspectives of mood aroused from the song.

In near future, we wish to collect more mood annotated dataset. We will use the neural networks for the classification purpose as it gives better results in Patra et al. [29]. We are also planning to use bagging and voting approach for the classification purpose. The songs having different moods while listening and reading it were excluded from the present study and we intend to perform deeper analysis on these songs in future.

Acknowledgements

The work reported in this paper is supported by a grant from the "**Visvesvaraya Ph.D. Fellowship**" funded by "Department of Electronics and Information Technology (DeitY)", Ministry of Communications & Information Technology (MCIT), Government of India.

References

1. **Abburi, H., Akkireddy, E. S. A., Gangashetty, S. V., & Mamidi, R. (2016).** Multimodal sentiment analysis of telugu songs. *Proceedings of the 4th Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2016)*, pp. 48–52.
2. **Bakliwal, A., Arora, P., & Varma, V. (2012).** Hindi subjective lexicon: A lexical resource for hindi polarity classification. *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC)*.
3. **Bischoff, K., Firan, C. S., Paiu, R., Nejdl, W., Laurier, C., & Sordo, M. (2009).** Music mood and theme classification-a hybrid approach. *Proceedings of the 10th International Society for Music Information Retrieval (ISMIR)*, pp. 657–662.
4. **Cohen, J. (1960).** A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, Vol. 20, No. 1.
5. **Das, D., Poria, S., & Bandyopadhyay, S. (2012).** A classifier based approach to emotion lexicon construction. *Proceedings of the International Conference on Application of Natural Language to Information Systems*, Springer, pp. 320–326.
6. **Duncan, N. & Fox, M. (2005).** Computer–aided music distribution: The future of selection, retrieval and transmission. *First Monday*, Vol. 10, No. 4.
7. **Fu, Z., Lu, G., Ting, K. M., & Zhang, D. (2011).** A survey of audio-based music classification and annotation. *IEEE Transactions on Multimedia*, Vol. 13, No. 2, pp. 303–319.

8. **Gopal, S. & Moorti, S. (2008).** *Global Bollywood: Travels of Hindi song and dance*. U of Minnesota Press.

9. **Hampiholi, V. (2012).** A method for music classification based on perceived mood detection for Indian bollywood music. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, Vol. 6, No. 12, pp. 1636–1643.

10. **Hevner, K. (1936).** Experimental studies of the elements of expression in music. *The American Journal of Psychology*, Vol. 48, No. 2, pp. 246–268.

11. **Hu, X. (2010).** Music and mood: Where theory and reality meet. *Proceedings of the iConference 2010*.

12. **Hu, X. & Downie, J. S. (2010).** Improving mood classification in music digital libraries by combining lyrics and audio. *Proceedings of the 10th annual joint conference on Digital libraries*, ACM, pp. 159–168.

13. **Hu, X. & Downie, J. S. (2010).** When lyrics outperform audio for music mood classification: A feature analysis. *Proceedings of the 11th International Society for Music Information Retrieval (ISMIR)*, pp. 619–624.

14. **Hu, X., Downie, J. S., Laurier, C., Bay, M., & Ehmann, A. F. (2008).** The 2007 mirex audio mood classification task: Lessons learned. *Proceedings of the 9th International Society for Music Information Retrieval (ISMIR)*, pp. 462–467.

15. **Joshi, A., Balamurali, A., & Bhattacharyya, P. (2010).** A fall-back strategy for sentiment analysis in hindi: a case study.

16. **Kim, Y. E., Schmidt, E. M., Migneco, R., Morton, B. G., Richardson, P., Scott, J., Speck, J. A., & Turnbull, D. (2010).** Music emotion recognition: A state of the art review. *Proceedings of the 11th International Society for Music Information Retrieval (ISMIR)*, pp. 255–266.

17. **Koduri, G. K. & Indurkhya, B. (2010).** A behavioral study of emotions in south indian classical music andits implications in music recommendation systems. *Proceedings of the 2010 ACM workshop on Social, adaptive and personalized multimedia interaction and access*, ACM, pp. 55–60.

18. **Lamere, P. (2008).** Social tagging and music information retrieval. *Journal of New Music Research*, Vol. 37, No. 2, pp. 101–114.

19. **Laurier, C., Grivolla, J., & Herrera, P. (2008).** Multimodal music mood classification using audio and lyrics. *Proceedings of the Seventh International Conference on Machine Learning and Applications, 2008. ICMLA'08*, IEEE, pp. 688–693.

20. **Laurier, C., Sordo, M., Serra, J., & Herrera, P. (2009).** Music mood representations from social tags. *Proceedings of the 10th International Society for Music Information Retrieval (ISMIR)*, pp. 381–386.

21. **Liu, D., Lu, L., & Zhang, H. (2003).** Automatic mood detection from acoustic music data. *Proceedings of the 4th International Society for Music Information Retrieval (ISMIR)*, pp. 81–87.

22. **Lu, L., Liu, D., & Zhang, H.-J. (2006).** Automatic mood detection and tracking of music audio signals. *IEEE Transactions on audio, speech, and language processing*, Vol. 14, No. 1, pp. 5–18.

23. **McKay, C., Fujinaga, I., & Depalle, P. (2005).** jaudio: A feature extraction library. *Proceedings of the 6th International Conference on Music Information Retrieval*, pp. 600–3.

24. **Patra, B. G., Banerjee, S., Das, D., Saikh, T., & Bandyopadhyay, S. (2013).** Automatic author profiling based on linguistic and stylistic features. *Proceedings of the Notebook for PAN at CLEF*.

25. **Patra, B. G., Das, D., & Bandyopadhyay, S. (2013).** Automatic music mood classification of hindi songs. *Proceedings of the 3rd Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2013)*, pp. 24–28.

26. **Patra, B. G., Das, D., & Bandyopadhyay, S. (2013).** Unsupervised approach to hindi music mood classification. *Proceedings of the Mining Intelligence and Knowledge Exploration*, Springer, pp. 62–69.

27. **Patra, B. G., Das, D., & Bandyopadhyay, S. (2015).** Labeling data and developing supervised framework for hindi music mood analysis. *Unpublished*.

28. **Patra, B. G., Das, D., & Bandyopadhyay, S. (2015).** Mood classification of hindi songs based on lyrics. *Proceedings of the 12th International Conference on Natural Language Processing (ICON-2015)*.

29. **Patra, B. G., Das, D., & Bandyopadhyay, S. (2015).** Music emotion recognition system. *Proceedings of the International Symposium Frontiers of Research Speech and Music (FRSM-2015)*, pp. 114–119.

30. **Patra, B. G., Maitra, P., Das, D., & Bandyopadhyay, S. (2015).** Mediaeval 2015: Music emotion recognition based on feed-forward neural network.

Proceedings of the MediaEval 2015 Workshop (2015).

31. **Russell, J. A. (1980).** A circumplex model of affect. *Journal of Personality and Social Psychology*, Vol. 39, No. 06, pp. 1161–1178.
32. **Soleymani, M., Caro, M. N., Schmidt, E. M., Sha, C.-Y., & Yang, Y.-H. (2013).** 1000 songs for emotional analysis of music. *Proceedings of the 2nd ACM international workshop on Crowdsourcing for multimedia*, ACM, pp. 1–6.
33. **Thayer, R. E. (1990).** *The biopsychology of mood and arousal*. Oxford University Press.
34. **Ujlambkar, A. M. (2012).** Automatic mood classification of indian popular music. *Master's thesis*, College of Engineering, Pune.
35. **Ujlambkar, A. M. & Attar, V. Z. (2012).** Mood classification of indian popular music. *Proceedings of the CUBE International Information Technology Conference*, ACM, pp. 278–283.
36. **Velankar, M. R. & Sahasrabuddhe, H. V. (2012).** A pilot study of hindustani music sentiments. *Proceedings of the 2nd Workshop on Sentiment Analysis where AI meets Psychology (COLING 2012)*, pp. 91–98.
37. **Yang, D. & Lee, W.-S. (2004).** Disambiguating music emotion using software agents. *Proceedings of the 5th International Society for Music Information Retrieval (ISMIR)*, pp. 218–223.
38. **Yang, Y.-H., Lin, Y.-C., Cheng, H.-T., Liao, I.-B., Ho, Y.-C., & Chen, H. H. (2008).** Toward multi-modal music emotion classification. *Proceedings of the Pacific-Rim Conference on Multimedia*, Springer, pp. 70–79.
39. **Zaanen, M. V. & Kanters, P. (2010).** Automatic mood classification using tf* idf based on lyrics. *Proceedings of the 11th International Society for Music Information Retrieval (ISMIR)*, pp. 75–80.

Braja Gopal Patra is a Ph.D. Scholar in the Department of Computer Science and Engineering, Jadavpur University, India. He received Master's degree from the Department of Computer Science and Engineering, National Institute of Technology (NIT), Agartala, India in 2012 and Bachelor's degree in Computer Science and Engineering from West Bengal University of Technology, India in 2010. He is a recipient of the Visvesvaraya Ph.D. fellowship of "Department of Electronics and Information Technology", Government of India. His research interests include Music Information Retrieval, Sentiment Analysis, and Natural Language Processing. He is a member of the ACL and IEEE.

Dipankar Das is an Assistant Professor in the Department of Computer Science and Engineering, Jadavpur University, India. He received Ph.D. and Master's degrees from the Department of Computer Science and Engineering, Jadavpur University in 2013 and 2009 respectively. He received Bachelor's degree in Computer Science and Engineering from West Bengal University of Technology in 2005. His research interests are in the area of Natural Language Processing, Emotion and Sentiment Analysis, Affect Computing, Information Extraction and Language Generation. He has more than 50 publications in top conferences and journals and has served as an author over 15 Book Chapters. He is a member of the IEEE, ACL, HUMAINE groups.

Sivaji Bandyopadhyay is a Professor in the Department of Computer Science and Engineering, Jadavpur University, India. He received the Ph.D., Master's and Bachelor's degrees from the Department of Computer Science and Engineering, Jadavpur University in 1998, 1987, and 1985, respectively. He is engaged with several national and international projects. His research interests are in the area of Natural Language Processing, Machine Learning, Machine Translation, Sentiment Analysis, Question Answering Systems and Information Extraction. He has more than 300 publications in top conferences and journals. He has served as program chair, workshop chair and PC member of COLING, IJCNLP, NAACL, NLPKE, ICON and others. He is a member of the ACL, AAMT.

Article received on 11/01/2016; accepted on 03/03/2016.
Corresponding author is Braja Gopal Patra.