Multimodal Mood Classification - A Case Study of Differences in Hindi and Western Songs

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

Abstract

Music information retrieval has emerged as a mainstream research area in the past two decades. Experiments on music mood classification have been performed mainly on Western music based on audio, lyrics and a combination of both. Unfortunately, due to the scarcity of digitalized resources, Indian music fares poorly in music mood retrieval research. In this paper, we identified the mood taxonomy and prepared multimodal mood annotated datasets for Hindi and Western songs. We identified important audio and lyric features using correlation based feature selection technique. Finally, we developed mood classification systems using Support Vector Machines and Feed Forward Neural Networks based on the features collected from audio, lyrics, and a combination of both. The best performing multimodal systems achieved F-measures of 75.1 and 83.5 for classifying the moods of the Hindi and Western songs respectively using Feed Forward Neural Networks. A comparative analysis indicates that the selected features work well for mood classification of the Western songs and produces better results as compared to the mood classification systems for Hindi songs.

1 Introduction

Global digitization has led to music being available in the form of CDs, DVDs or other portable formats. With the rapid growth in Internet connectivity over the last decade, the audio or video music files are easily available and accessible over the World Wide Web. The number of music compositions created worldwide already exceeds a few millions and continues to grow. Similarly, the popularity of downloading and purchasing of music from online music shops has also been increased at the same pace. Thus, the organization and management of the music files are the important issues to be tackled carefully. Recently, studies on music information retrieval (MIR) have shown that moods are desirable access keys to music repositories and collections (Hu and Downie, 2010a).

In order to find out such access keys, most of the experiments on music mood classification of Western music have been performed based on the audio (Lu et al., 2006; Hu et al., 2008), lyrics (Zaanen and Kanters, 2010) and combination of both (Laurier et al., 2008; Hu and Downie, 2010b; Hu and Downie, 2010a). In case of the Indian music, few tasks have been performed on the Hindi music mood classification based on the audio (Ujlambkar and Attar, 2012; Patra et al., 2013a; Patra et al., 2013b; Patra et al., 2016b), lyrics (Patra et al., 2015c) and combination of both (Patra et al., 2016a). The maximum F-measure achieved for the multimodal system for Hindi songs was 68.6% in Patra et al. (2016a).

Indian music can be divided into two broad categories namely, "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. Hindi or Bollywood music, also known as popular music, is mostly present in Hindi cinemas or Bollywood movies. Hindi is one of the official languages of India and is fourth most widely spoken language in the World[1]. Hindi songs make approximately 72% of the total music sales in India (Ujlambkar and Attar, 2012).

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

Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pages 1980–1989, Osaka, Japan, December 11-17 2016.

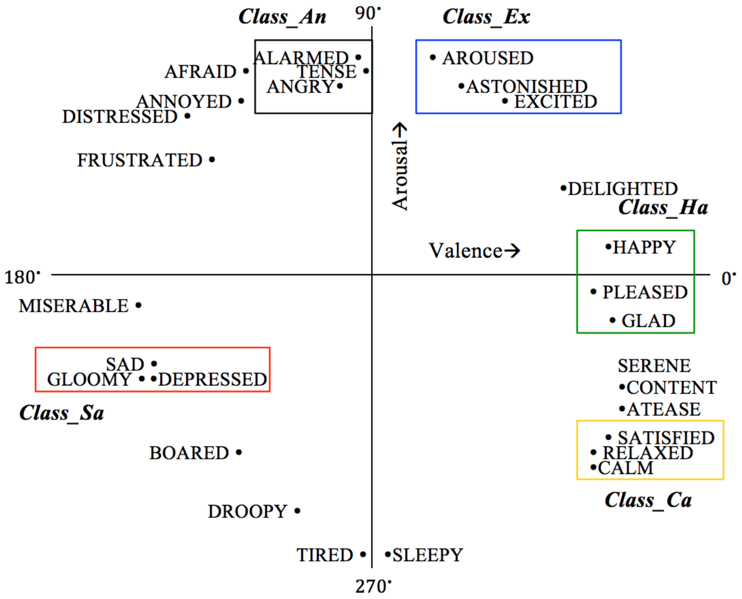

Figure 1: Russell's circumplex model of 28 affect words (Russell, 1980)

In order to deal with the above mentioned issues, the contributions of the authors are given below.

1. We employed our earlier proposed mood taxonomy for music mood classification in Hindi and Western songs (Patra et al., 2015c; Patra et al., 2016a; Patra et al., 2016b).

2. We annotated the audio and lyrics of the Hindi and Western songs using the above mood taxonomy.

3. We observed difference in mood while annotating the mood at the time of listening to the music and reading its corresponding lyric in case of the Hindi songs.

4. We identified important features using correlation based feature selection technique.

5. The Feed Forward Neural Networks (FFNNs) is implemented for mood classification purpose.

6. We have developed a multimodal system based on the audio and lyrics of the songs.

This paper is organized as follows: Section 2 introduces the mood taxonomy and describes the process of dataset preparation. Section 3 describes the audio and lyrics features. The FFNNs and the developed systems with comparison are described in the section 4. Finally, the conclusion is drawn in Section 5.

2 Mood Taxonomy and Dataset

2.1 Mood Taxonomy

We chose the *Russell's circumplex model* (Russell, 1980) to build our own mood taxonomy. The *circumplex model* and a subset of this dimensional model have been used earlier for several mood classification studies (Ujlambkar and Attar, 2012; Patra et al., 2013a; Patra et al., 2013b; Patra et al., 2015c; Patra et al., 2016a; Patra et al., 2016b). The *circumplex model* is based on *valence* and *arousal*, which is widely accepted by the research community. *Valence* indicates the positivity and negativity of emotions whereas *arousal* indicates the emotional intensity. The mood taxonomy is prepared by clustering the similar affect words of the *circumplex model* into a single class and each class contains three affect words of the *circumplex model* as shown in Figure 1. Each of our mood classes has distinct positions in terms of *arousal* and *valence*. We considered the five coarse mood classes, namely "*Class_An*", "*Class_Ca*", "*Class_Ex*", "*Class_Ha*", and "*Class_Sa*" for our experiments.

2.2 Dataset

All audio files of Hindi and Western song were collected from CDs bought from registered stores. The lyrics of the corresponding songs were collected from the web. The lyrics of the Hindi songs were written in *Romanized English* characters while essential resources like Hindi sentiment lexicons and list of stop words are available in *utf-8* character encoding. Thus, we transliterated the *Romanized* Hindi lyrics to *utf-8* characters using the transliteration tool available in the English to Indian Language Machine Translation (EILMT) project[2]. 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.

Related research suggests that the state-of-the-art experiments on music mood classification have been performed on audio clips of 30 seconds (Hu et al., 2008; Ujlambkar and Attar, 2012; Patra et al., 2013b). In case of the Hindi songs, it is difficult to annotate mood of 30 second song clips since the annotators get confused in between adjacent mood classes while annotating short duration audio files. Thus, we sliced each of the audio files into 60 second clips. Each of the audio clips and lyric files of the Hindi and Western songs were annotated by three different annotators. The undergraduate students and research scholars belonging to the age group of 18-35 served as annotators.

From the annotation, we observed that different mood classes were chosen by the annotators during listening to the audio and reading the corresponding lyrics. The difference between listener's and reader'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 lyric of a song in the listeners. For example, a song "Bhaag D.K.Bose Aandhi Aayi"[3] 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 was annotated as "*Class_Sa*" while reading the lyric, whereas it was annotated as "*Class_An*" while listening to the corresponding audio as it contains mostly rock music and arousal is also high. Similarly a song "Dil Duba"[4] was annotated as "*Class_Sa*" and "*Class_Ha*" while reading the lyric and listening to the corresponding audio, respectively. This song portrays negative emotions by using sad or negative words like "tere liye hi mar jaunga (*I would die for you*)", however, this song contains high valence. The above observations emphasize that the combined effect of lyric and audio plays a pivotal role in indicating the final mood inducing characteristics of a music piece. Moreover, the intensity of the emotion felt during listening to a song is much more than the intensity of the emotion while reading a lyric. The main reason behind this may be that the music with the voice induces the emotion.

We did not notice such differences in mood for the Western music. However, the intensity of the emotion was less while reading a lyric as compared to listening to the corresponding music in case of both Hindi and Western music. The confusion matrix of the Hindi song mood annotation is shown in Table 1. Detailed statistics of the annotated Hindi and Western songs are given in Table 2. We considered only those Hindi songs for our experiments, which were annotated with the same class both after listening to it and reading the corresponding lyric.

We calculated pairwise inter-annotator agreements on the dataset by computing Cohen's κ coefficient (Cohen, 1960). The inter-annotator agreements were calculated separately for audio clip annotation and lyric annotation. The overall inter-annotator agreement scores with five mood classes were found to be 0.94 and 0.84 for Hindi audio and lyrics, respectively. In case of the Western songs, the inter-annotator agreement scores with five mood classes were 0.91 and 0.87 for audio and lyrics, respectively. These correlation coefficients can be interpreted as almost perfect agreements.

[2] http://tdil-dc.in/index.php?option=com vexrtical&parentid=72

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

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

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

		R_{Lyrics}					
		a	**b**	**c**	**d**	**e**	**Total**
L_{Audio}	**Class_An = a**	**48**	0	15	2	10	75
	Class_Ca = b	2	**65**	3	17	13	100
	Class_Ex = c	13	3	**62**	16	6	100
	Class_Ha = d	4	11	15	**66**	4	100
	Class_Sa = e	8	17	7	9	**78**	125

Table 2: Statistics of the Hindi and Western songs

	Hindi Songs		**Western Songs**	
	Clips	**Total Songs**	**Clips**	**Total Songs**
Class_An	203	48	230	60
Class_Ca	252	65	247	72
Class_Ex	258	62	236	63
Class_Ha	232	66	218	58
Class_Sa	285	78	180	45
Total	1230	319	1111	298

3 Feature Extraction

This section describes the process of extracting features from both audio and lyrics. Feature extraction and selection play an important role in machine-learning frameworks. The important features from audio and lyrics were identified using correlation based feature selection technique. We considered different audio and textual features for mood classification in Hindi and Western songs.

3.1 Audio Features

We considered the key audio features like *intensity*, *rhythm* and *timbre* for the mood classification task. These features had been used by researchers for music mood classification in Indian languages (Ujlambkar and Attar, 2012; Patra et al., 2015b). These features were extracted using the jAudio (McKay et al., 2005) toolkit. In addition to these features, *chroma* and *harmonics* features were extracted from the audio files using the openSMILE (Eyben et al., 2010) toolkit.

3.2 Lyric Features

We adopted a wide range of textual features such as sentiment words, stylistic features and N-gram based features which are discussed in the following subsections.

Preprocessing: First we cleaned the lyrics dataset by removing the junk characters and HTML tags. Subsequently we removed the duplicate lines as it was observed that the starting stanza is usually repeated in the song at least a few times. Therefore, we removed these duplicate sentences to remove the biasness in the lyric.

3.2.1 Sentiment Lexicons (SL)

The emotion or sentiment words are one of the most important features for mood classification from lyrics. These words in the Hindi lyrics were identified using three lexicons - Hindi Subjective Lexicon (HSL) (Bakliwal et al., 2012), Hindi SentiWordnet (HSW) (Joshi et al., 2010) and Hindi Wordnet Affect (HWA) (Das et al., 2012). Similarly, we used two lexicons - SentiWordNet (Baccianellaet al., 2010) (SWN) and WordNetAffect (Strapparava and Valitutti, 2004) (WA) for identifying the sentiment words from the lyrics of the Western songs. It was observed that the number of sentiment words found in the lyrics of Hindi songs was less than the number of sentiment words found in the lyrics of Western songs. The main reason was that the the performances of the POS tagger and stemmer/lemmatizer for Hindi language were not up to the mark. The CRF based *Shallow Parser*[5] 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 language. Most of the inflected sentiment words in Hindi lyrics were not matched with the sentiment words available in the Hindi sentiment or emotion lexicons. Thus, the number of words matched with sentiment or emotion lexicons are considerably less. In case of the Western songs, we used the RitaWordNet [6] to get the stemmed words and the parts-of-speech (POS) tags for the lyric words. The statistics of the sentiment or emotion words identified by these lexicons are given in Table 3.

Table 3: Statistics of unique sentiment and emotion words present in the lyrics of Hindi and Western songs

Classes	HWA	WA	Classes	HSL	HSW	SWN
Angry	248	312	**Positive**	1185	872	7853
Disgust	17	32				
Fear	20	52				
Happy	352	412	**Negative**	963	735	5271
Sad	110	231				
Surprise	39	81				

3.2.2 Text Stylistic Features (TSF)

Text stylistic features are widely used in text stylometric analysis such as authorship identification, author identification, etc. These features were also been used for mood classification from Hindi lyrics (Patra et al., 2015c) and Western music lyrics (Hu and Downie, 2010b). It was observed that these features reduce the performance of the system. The TSF such as the number of unique words, number of repeated words, number of lines, etc. were considered for our experiments.

3.2.3 N-Grams (NG)

It was noted by researchers that N-gram based features work well for mood classification using lyrics as compared to the stylistic or sentiment features (Zaanen and Kanters, 2010; Hu and Downie, 2010b; Patra et al., 2015c). The term frequency and document frequency (TF-IDF) scores of unigram, bigram and trigram were considered for the present study. The higher order N-grams tend to reduce the performance of the system. The N-grams having document frequencies more than one were considered to reduce the sparsity of the document vectors. We also removed the stopwords while considering the N-grams as it was observed that the stopwords do not contain any information related to the classification.

3.3 Feature Selection

Feature level correlation (Hall, 1999) was used to identify the most important features as well as to reduce the feature dimension in (Patra et al., 2015a). Thus, we used the correlation based supervised feature selection technique implemented in Weka toolkit [7] to find out the important contributory feature set for audio and lyrics.

A total of 445 audio features were extracted from Hindi and Western music audio files using jAudio and openSMILE. We also collected 12 sentiment features, 12 textual stylistic features and 6832 N-gram features from Hindi lyrics, whereas 8 sentiment features, 12 textual features and 8461 N-gram features were collected from the Western music lyrics. The feature selection technique implemented using Weka yields 154 important audio features for both Hindi and Western songs. 12 sentiment, 8 stylistic, and 1601 N-gram features from Hindi lyrics and 8 sentiment, 8 stylistic and 3175 N-gram features from Western music lyrics were extracted using feature selection technique. We subsequently used these features for the classification purpose.

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

[6] http://www.rednoise.org/rita/

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

4 Classification Framework

We used the FFNNs for the mood classification purpose. It was observed that the FFNNs give better accuracy as compared to other machine learning algorithms like Support Vector Machines (SVMs) and Decision Trees (Patra et al., 2015b). Patra et al., (2015a) achieved low root mean square value for *arousal* and *valence* calculation using FFNNs. Moreover, they (Patra et al., 2015b) also reported higher F-measure for Hindi music mood classification based on audio.

4.1 Feed Forward Neural Networks (FFNNs)

Feed Forward Neural Networks refer to a special topology of neural networks in which each neuron belonging to a layer is connected to all the other neurons in the next layer. Neural networks are widely used for several classification and regression problems for its structural simplicity. The network is divided into multiple layers namely *input layer*, *hidden layer* and *output layer*. The *input layer* consists of inputs to the network. Then, the network follows a *hidden layer* which may consist of any number of *neurons* placed in parallel. Each neuron performs a weighted summation of the inputs which is then passed on to a nonlinear *activation function* (σ), also called the *neuron* function. Mathematically, the functionality of a hidden *neuron* is described as: $\sigma = \sum_{j=1}^{n}(w_j x_j + b_j)$, where the weights $\{w_j, b_j\}$ are symbolized with the arrows feeding into the neurons. The network output is formed by another weighted summation of the outputs of the neurons in the hidden layer (Mathematica Neural Networks- Train and Analyze Neural Networks to Fit Your Data, 2005). This summation on the output is called the output layer. The gradient descent learning principle is used to update the weights as the errors are back-propagated through each layer by the well-known back-propagation algorithm (Rumelhart et al., 1986). The updation rule can be stated as $\theta = \theta - \partial E/\partial\theta$.

4.2 Results Analysis and Discussion

We used FFNNs and LibSVM, a variant of the support vector machines (SVMs) implemented in Weka for the classification purpose. Several systems were developed using the audio features, lyric features and a combination of both. All experiments were conducted on the features selected by the correlation based feature selection technique. To obtain reliable accuracy, a 10-fold cross validation was performed for each of the classifiers.

4.2.1 Mood classification based on Audio

Initially, we performed experiments using only the timbre features and then added the other features incrementally. First we developed LibSVM based mood classification systems for Hindi and Western music. For Hindi music, the audio features based system achieved F-measure of 59.0, whereas for the Western music, the system achieved F-measure of 70.5 using all the audio features. The audio feature based Western music mood classification system achieved better F-measure (11.5 points absolute, 19.5% relative) than the Hindi-one. However, there were less number of instances present in the Western music as compared to the Hindi music. Therefore, we developed another pair of mood classification systems using the same number of instances for both the Hindi and Western music. In this case, the maximum F-measure of 56.8 and 64.5 were achieved for Hindi and Western music mood classification respectively using only audio features. The F-measure of the Western music mood classification system was 7.7 points absolute higher than the one for the Hindi music mood classification system developed on the same number of instances.

In the next phase, we developed audio based mood classification systems using FFNNs on the same set of 154 features. The maximum F-measure of 65.2 and 75.7 were achieved for the Hindi and Western music mood classification systems. The performance of the audio based systems are given in Table 4.

4.2.2 Mood classification based on Lyrics

We performed experiments using sentiment features initially and sequentially added other features incrementally. First, we used LibSVM for the classification purpose for feature ablation study. Subsequently, then we used the FFNNs using all the features together. We observed that the text stylistic features reduced the performance of the system. Thus, we removed text stylistic features from all other systems.

The lyric features based mood classification systems achieved the maximum F-measure of 55.3 and 68.2 for Hindi and Western song lyrics, respectively. The Western song mood classification achieves better F-measure of around 13.0 points absolute than the Hindi song mood classification system based only on lyric features. It was observed that the N-gram features yield good F-measure alone in case of the mood classification systems for Hindi and Western songs. The relative improvement of F-measure in case of Hindi song mood classification was much more than the mood classification system for Western songs. 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.

We also developed lyric based systems for both song categories using the FFNNs. The corresponding mood classification systems achieved the maximum F-measure of 57.1 and 69.2 for Hindi and Western songs respectively. The performance of the lyric based systems are reported in Table 4.

4.2.3 Multimodal Music Mood classification

We developed multimodal music mood classification systems using LibSVM and FFNNs. The multimodal music mood classification system based on both audio and lyric features achieved F-measures of 68.9 and 80.4 for Hindi and Western songs using LibSVM. The multimodal music mood classification system for Western songs performs 11.5 points absolute better than the one for Hindi songs in terms of F-measure.

The FFNNs based multimodal music mood classification systems achieved the maximum F-measures of 75.1 and 83.5 for Hindi and Western songs, respectively. The performance of the multimodal systems are shown in Table 4 and the confusion matrices for these multimodal music mood classification systems are given in Table 5.

From the confusion matrix, it is observed that multimodal mood classification system for Hindi songs performs better in case of "*Class_Sa*" and performs poorly in case of "*Class_Ha*". This is obvious since the number of instances are more in case of "*Class_Sa*". The maximum number of instances from "*Class_Ha*" are classified as other classes because of the similar audio and lyric features. We also observed that the systems for Hindi songs are quite biased towards the "*Class_Sa*". In case of the Western songs, the "*Class_Ca*" contains the maximum number of instances and thus maximum number of instances are classified correctly for "*Class_Ca*". The system performs better in case of the "*Class_An*" and performs poorly in case of the "*Class_Ex*". It was also observed that some of the instances from each of the classes have tendency to go towards its neighboring classes. The main reason may be the similar features in between the neighbor classes.

4.2.4 Comparison with other systems

The proposed mood classification system for Hindi songs performs poorly as compared to the system of (Ujlambkar and Attar, 2012) which achieved F-measures of 75 to 81 using only audio based features. They used different mood taxonomy and they sliced the songs into 30 second clips. Unfortunately, their dataset is not freely available for research purpose. The features used for the experiments in (Ujlambkar and Attar, 2012) are a subset of our features. The audio based Hindi music mood classification system performs poor as compared to the system developed in (Patra et al., 2015b). Patra et al., (2015b) used more number of instances, but used a subset of our featureset. The audio based mood classification system outperformed other audio based systems reported in (Patra et al., 2013a; Patra et al., 2013b), but they developed their systems using smaller dataset and less number of features.

Our lyrics based mood classification system for Hindi songs outperformed the system reported in (Patra et al., 2015c) by 18.6 points absolute in terms of F-measure. We used similar features, but the number of instances were more in case of the present system. The significant difference in experimental setup is that their dataset was annotated with mood classes after listening to the corresponding audio files, whereas our lyrics dataset was annotated after reading the lexical content of lyrics. To the best of our knowledge, currently there is no other lyrics based mood classification system for Hindi music available in the literature. Patra et al., (2016a) developed multimodal mood classification system for Hindi songs using LibSVM and achieved F-measure of 68.6, which is 0.3 point absolute less than our multimodal mood classification system for Hindi songs using the same LibSVM. The main reason may be that we

Table 4: Performance of the mood classification systems with respect to different features using LibSVM and FFNNs

		Hindi Music			**Western Music**		
Systems	**Features**	P	R	F	P	R	F
Audio Features using LibSVM	**Timbre**	55.2	54.5	54.8	63.7	63.2	63.4
	Timbre+Intensity	55.7	55.3	55.5	66.9	66.6	66.8
	Timbre+Intensity+ Rhythm	58.8	57.8	58.2	70.3	70.0	70.2
	All audio features	58.9	59.1	59.0	70.5	70.5	70.5
Audio Features using FFNNs	**All audio features**	65.3	65.1	**65.2**	75.8	75.7	**75.7**
Lyrics Features using LibSVM	**SL**	41.3	39.4	40.4	60.0	59.6	59.8
	SL+TSF	38.6	38.7	38.6	59.7	59.9	59.8
	NG	46.3	46.7	46.5	60.2	60.3	60.2
	SL+TSF+NG	55.3	52.8	54.1	68.2	68.3	68.2
	SL+NG	55.9	54.7	55.3	68.2	68.3	68.2
Lyrics Features using FFNNs	**SL+NG**	57.2	57.0	**57.1**	69.3	69.1	**69.2**
Multimodal using LibSVM	**Audio+ Lyrics(Excluding TSF)**	69.2	68.6	68.9	80.3	80.5	80.4
Multimodal using FFNNs	**Audio+ Lyrics(Excluding TSF)**	76.8	73.5	**75.1**	84.8	82.2	**83.5**

used more number of instances for the present system. Till date, to the best of the author's knowledge, no other multimodal system has also been developed for Hindi songs based on audio and lyric features.

Table 5: Confusion matrix for multimodal systems using FFNNs

	Hindi Songs					**Western Songs**				
Classified as –>	**a**	**b**	**c**	**d**	**e**	**a**	**b**	**c**	**d**	**e**
Class_An = a	**153**	11	18	7	14	**195**	2	23	3	7
Class_Ca = b	1	**185**	3	35	28	0	**208**	2	12	25
Class_Ex = c	37	10	**192**	12	7	25	2	**192**	12	5
Class_Ha = d	5	35	10	**170**	12	3	9	16	**182**	8
Class_Sa = e	14	37	2	8	**224**	9	12	2	6	**151**

For Western music, it is very difficult to compare our mood classification system with other systems available in the literature, as our mood taxonomy is totally different from the mood taxonomy proposed by the existing multimodal mood classification systems (Hu and Downie, 2010a; Hu and Downie, 2010b). The number of mood classes present in their taxonomy is much higher (eighteen) than ours. Taking this into consideration, our present Western music mood classification system performed better than those systems. We used almost similar audio and lyric features as compared to the above mentioned systems. They used sentiment lexicons like General Inquirer, ANEW and WordNet-Affect, whereas we used SentiWordNet and WordNet-Affect for identifying the sentiment words.

4.2.5 Observations

Each mood classification system for Western songs outperforms the corresponding mood classification system for Hindi songs developed with the same classifier. The reasons for such results are listed below.

1. The moods experienced during listening to audio and reading the corresponding lyric are different in case of Hindi songs.

2. The mood is not very clear in the first 60 seconds clip of a Hindi song. Starting 20-30 seconds of the first clip of a song is mostly calm.

3. Western songs are usually much more rhythmic than Hindi songs.

4. The intensity of the mood felt in case of reading a lyric is less than the intensity of the mood felt at the time of listening to the audio in both song types.

5. We need more sophisticated features for audio to identify the mood in case of the Hindi music.

5 Conclusions

We developed mood annotated multimodal (lyrics and audio) datasets for Hindi and Western songs. Based on these multimodal datasets, we developed automatic multimodal music mood classification systems using LibSVM and FFNNs. The best performing systems developed using FFNNs achieved the maximum F-measures of 75.1 and 83.5 for Hindi and Western songs, respectively. It was observed that the different moods were perceived by the annotators while listening to audio and reading the corresponding song lyric in case of the Hindi songs. The main reason for such difference may be that the audio and lyrics were annotated by different annotators. Another reason may be that the mood is not transparent in lyrics as compared to the mood present in the audio of the corresponding song. In future, we intend to perform deeper analysis of the listener's and reader's perspectives of mood aroused from songs. We would also like to collect more instances for mood annotated datasets. We are also planning to use bagging and voting approach for the classification purpose.

Acknowledgments

The work reported in this paper is supported by a grant from the "Visvesvaraya Ph.D. Scheme for Electronics and IT" funded by Media Lab Asia of Ministry of Electronics and Information Technology (MeitY), Government of India.

References

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. *LREC*, 2200–2204.

Akshat Bakliwal, Piyush Arora, and Vasudeva Varma. 2012. Hindi subjective lexicon: A lexical resource for hindi polarity classification. *LREC*.

Jacob Cohen. 1960. A Coefficient of Agreement for Nominal Scales *Educational and Psychological Measurement* 20(1):37–46.

Dipankar Das, Soujanya Poria, and Sivaji Bandyopadhyay. 2012. A classifier based approach to emotion lexicon construction. *Natural language processing and information systems*, 320–326.

Florian Eyben, Martin Wöllmer, and Björn Schuller. 2010. Opensmile: the munich versatile and fast open-source audio feature extractor. *Proceedings of the 18th ACM international conference on Multimedia*, 1459–1462, ACM.

Mark A. Hall. 1999. Correlation-based feature selection for machine learning. *PhD dissertation*, The University of Waikato.

Xiao Hu, J. Stephen Downie, Cyril Laurier, Mert Bay, and Andreas F. Ehmann. 2008. The 2007 MIREX audio mood classification task: Lessons learned. *Proceedings of the 9th International Society for Music Information Retrieval Conference (ISMIR 2008)*, 462–467.

Xiao Hu and J. Stephen Downie. 2010a. When Lyrics Outperform Audio for Music Mood Classification: A Feature Analysis. *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, 619–624.

Xiao Hu and J. Stephen Downie. 2010b. Improving mood classification in music digital libraries by combining lyrics and audio. *Proceedings of the 10th Annual Joint Conference on Digital Libraries*, 159–168.

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

Cyril Laurier, Jens Grivolla, and Perfecto Herrera. 2008. Multimodal music mood classification using audio and lyrics. *Proceedings of the 7th International Conference on Machine Learning and Applications (ICMLA'08)*, 688–693, IEEE.

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*, 5–18.

Cory McKay, Ichiro Fujinaga, and Philippe Depalle 2005 jAudio: A feature extraction library. *Proceedings of the International Conference on Music Information Retrieval*, 600–603.

Braja G. Patra, Dipankar Das, and Sivaji Bandyopadhyay. 2013. Automatic Music Mood Classification of Hindi Songs. *Proceedings of the 3rd Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2013)*, 24–28.

Braja G. Patra, Dipankar Das, and Sivaji Bandyopadhyay. 2013. Unsupervised approach to Hindi music mood classification. *Mining Intelligence and Knowledge Exploration*, 62–69.

Braja G. Patra, Promita Maitra, Dipankar Das, and Sivaji Bandyopadhyay. 2015. MediaEval 2015: Feed-Forward Neural Network based Music Emotion Recognition. *Proceedings of MediaEval 2015 Workshop.*

Braja G. Patra, Dipankar Das, and Sivaji Bandyopadhyay. 2015. Music Emotion Recognition System. *Proceedings of the International Symposium Frontiers of Research Speech and Music (FRSM-2015)*, 114–119.

Braja G. Patra, Dipankar Das, and Sivaji Bandyopadhyay. 2015. Mood Classification of Hindi Songs based on Lyrics. *Proceedings of the 12th International Conference on Natural Language Processing (ICON-2015).*

Braja G. Patra, Dipankar Das, and Sivaji Bandyopadhyay. 2016. Multimodal Mood Classification Framework for Hindi Songs. *Computación y Sistemas*, 20(3):515-526.

Braja G. Patra, Dipankar Das, and Sivaji Bandyopadhyay. 2016. Labeling Data and Developing Supervised Framework for Hindi Music Mood Analysis. *Journal of Intelligent Information Systems.*

David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. 1986. Learning representations by back-propagating errors. *Nature*, 323:533–536.

James A. Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161-1178.

Carlo Strapparava and Alessandro Valitutti. 2004. WordNet Affect: an Affective Extension of WordNet. *LREC*, 4:1083–1086.

Aniruddha M. Ujlambkar and Vahida Z. Attar. 2012. Mood classification of Indian popular music. *Proceedings of the CUBE International Information Technology Conference*, 278–283.

Menno Van Zaanen and Pieter Kanters. 2010. Automatic Mood Classification Using TF*IDF Based on Lyrics. *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, 75–80.

Mathematica Neural Networks- Train and Analyze Neural Networks to Fit Your Data. 2005. *Wolfram Research Inc., First Edition*, Champaign, Illinois, USA.