

Mood Classification based on the Lyrics of Punjabi Songs

Harpreet Kaur

Dr. Gurpreet Singh Josan

Abstract - Music has been an integral part of human beings for emotions, recreation and entertainment as it expresses emotions in a concise and effective manner. People choose to listen to a song or music which best fits to their mood at an instant point of time, making the need to classify the songs in congruence to their moods more of a demand. The study of mood recognition in the field of music has gained a lot of momentum with machine learning and data mining techniques. We have made an effort to build a system for automatic identification of mood according to lyrics by considering tf-idf, POS tagging and n-gram features. Mood classification refers to the use of natural language processing and text analysis to identify what the author meant. We have analyzed various data algorithms in order to learn, train and test the model representing the various moods of songs. We have been successful to achieve a satisfactory accuracy of above 60% of individual features and upto 85% in combination of features in identifying the mood by using KNN approach experimented over a list of 264 songs. Our results show that the use of tf-idf values provides a valuable source of information for automatic mood classifications for Punjabi songs based on lyrics only.

Index Terms - NLP, tf-idf, Part of Speech Tagging, N-gram and K-Nearest Neighbor.

1. Introduction

NLP is a process of extracting meaningful information from natural language input and/or producing natural language output. It enables system to process sentences in a natural language such as English. It is the ability of computer program to understand what a human is saying to it.

1.1 Music and Mood- Mood is an emerging metadata in music libraries and online music repositories. According to research, a mood is an emotional state of mind such as being happy, sad, excited, calm, angry or neutral. Indeed, psychological studies have shown that listening to music has pleasing effect on the emotional state and mood and social consistency of listeners. With so much variety of music easily available, humans do not always listen to the same type of music all the time. They have their interests in favorite artists, albums, music or genre type. This personal choice is governed by emotional state of human being at that instant point of time. Traditionally, classification is done according to artists, albums, music or genre type. Music is based on audio and lyrics (content based) features which associate it with a particular mood or emotion. We here focused on lyrical features.

• Harpreet Kaur has completed her master's degree program (M.tech) in computer science engineering (CSE) from Dept of Computer Science (DCS) Punjabi University, Patiala, Punjab, India. E-mail: preetpanaich@yahoo.com

• Dr. Gurpreet Singh Josan is currently Assistant Professor in Dept. of Computer Science (DCS), Punjabi University, Patiala, Punjab, India. E-mail: josangurpreet@pbi.ac.in

1.2 Issues In Mood Classification-

- ✦ Since individuals have different perceptions about classifying songs according to mood, it becomes a much more difficult task. For e.g. - Some mood categories overlap each other, people perceive folk song as happy and sad which might create difficulty in classification.
- ✦ There are various models for mood classification such as Thayer's, Russell's, Plutchik's, Ekman Model etc. Selecting an appropriate number of classes for mood should be carefully done.
- ✦ If the resources are not available for tasks like POS-tagging then it would be done manually.
- ✦ The words describing mood can be ambiguous and creates difficulty in POS tagging. For e.g. - baag word (in punjabi):- baag (version of phulkari)- Proper Noun, baag (garden)- Noun. How tagging of such word would be done? baago-baag hona (idiom), how idiom would be tagged? When special character is removed, how to tag word "baago"?

1.3 Purpose- To study the lyrics of Punjabi Songs and develop a system to identify the moods.

1.4 Objectives-

- ✦ To collect the lyrics of Punjabi songs and resources that are required in developing the system.
- ✦ To extract the features from the lyrics of a song to create a feature set.
- ✦ To develop the algorithm and flowchart for classification of mood of the song according to the mood divisions given by Thayer's mood model.
- ✦ To classify the songs using KNN classifier.

- To evaluate the performance and various metrics of developed system. In order to achieve these ultimate goals, we have laid down a list of sub-goals that we together help achieve the main objective:

- In Thayer's model of mood, the emotion plane consists of 2 axes – arousal and valence.

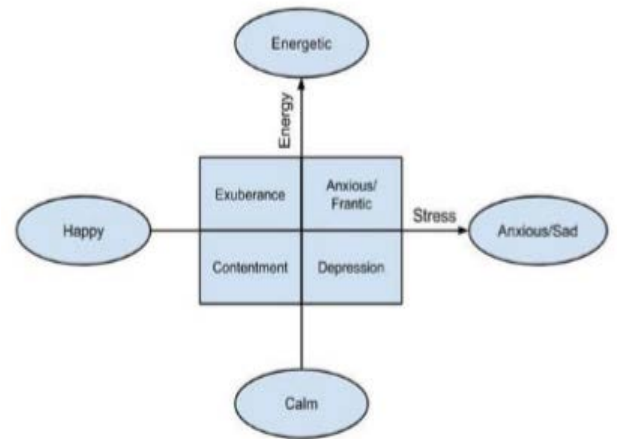
1.5 Scope and Applications-

The scope of the work is limited to Punjabi songs.

- Shop owners seeking songs to attract certain clients.
- Users can create playlist of songs according to their mood.
- Searching and Sorting of Punjabi songs become easy according to certain mood or occasion.

1.6 Mood Model-

- There are various models for mood classification such as Thayer's model, Russell's model, Plutchik's model, Ekman model etc.



2. Related Work

Author	Objective of Research	Purpose of Study
Xiao Hu, J. Stephen Downie, and Andreas F. Ehman, 2009	Lyric text mining in music mood classification	Lyrics was used in improving audio music mood classification by considering 18 mood categories based on user tags.
Menno V. Zaanen and Pieter Kanters, 2010	Automatic mood classification using TF*IDF based on lyrics	Information retrieval metric: tf*idf values based on lyrics only were used to measure relevance of words to the different mood classes.
R. Mihalcea and C. Strapparava, 2012	Lyrics, Music, and Emotions	Textual and musical features were both be successfully used for emotion recognition in songs with error rate reductions of up to 31%.
B. G. Patra, D. Das, and S. Bandyopadhyay, 2013	Unsupervised Approach to Hindi Music Mood Classification, Mining Intelligence and Knowledge Exploration	Unsupervised classifier for Hindi music mood classification was built using audio features like rhythm, timber & intensity with 48% accuracy.
Oh, S., Hahn, M., & Kim, J., 2013	Music Mood Classification Using Intro and Refrain Parts of Lyrics	Lyrics-based classification was proposed to detect mood of song with only intro and refrain parts of lyrics.
Corona, Humberto, and Michael P. O'Mahony, 2015	An Exploration of Mood Classification in the Million Songs Dataset	The classification of song according to mood from lyrics alone, was based on a vector space model representation with an accuracy of up to 70%.

Emil Ian V. Ascalon, Rafael Cabredo,2015	Lyric-Based Music Mood Recognition	Mood recognition of songs using lyrics was done by using TF-IDF and keyGraph keyword generation.Manual annotation of songs and an automated approach using AV were used.
Patra,B.G.,Das,D.,& Bandyopadhyay,S,2015	Mood Classification of Hindi Songs based on Lyrics	5 mood classes from Russell’s circumplex model were used.SVM, Decision Trees, and FFNN were used with maximum F-measure of 0.725.
S. Chauhan and P. Chauhan,2016	Music mood classification based on lyrical analysis of Hindi songs using LDA	New mood based classification system for hindi songs using Latent Dirichlet Allocation that will be different from classic genre based systems because emotions rule the choice of music.
N. Kashyap, T. Choudhury, D. K. Chaudhary and R. Lal,2016	Mood Based Classification of Music by Analyzing Lyrical Data Using Text Mining	The proposed system obtained 85% accuracy from the trained mood mapper, which lead to a more accurate understanding of the music mood.
Y. An, S. Sun and S. Wang,2017	Naive Bayes classifiers for music emotion classification based on lyrics	Lyrics of music used to analyze and classify music based on emotion for the automatic classification of Chinese music emotions. Music lyrics and their labels have been taken from a website named Baidu music and four different datasets are made.Naive Bayes classifier is used here to classify text and achieved accuracy upto 68%.

3. Research Methodology-

Dataset	No of Songs	Song Category(Mood)
Training Dataset	204	50 lok geet(happy),51 sad songs(sad),53 heroic songs(excited) and 50 religious(calm)
Testing Dataset	60	15 lok geet(happy),15 sad songs(sad),15 heroic songs(excited) and 15 religious(calm)
Total	264	65 lok geet(happy),66 sad songs(sad),68 heroic songs(excited) and 65 religious(calm)

3.1 Lyrics preprocessing- Lyrics of song are given as input in Punjabi language.Lyrics were checked for correctness such as the spelling of the words. Special characters such as ੴ ॥ ੨੩ - ! () ? etc were removed.

3.2 Lyrics normalization- Words that have root words in Punjabi root word list have been replaced with their root word for normalization.

3.3 Feature Extraction-

- ✦ Bag of words model was used to create a vocabulary of all unique(non-redundant),unordered words that occur in dataset and each word is associated with a frequency count.
- ✦ Calculate Term Frequency and Inverse Document Frequency.To perform normalization, term frequency is divided by the length of song.

$$TF(w) = (\text{Number of times word 'w' appears in a song}) / (\text{Total number of words in the song})$$

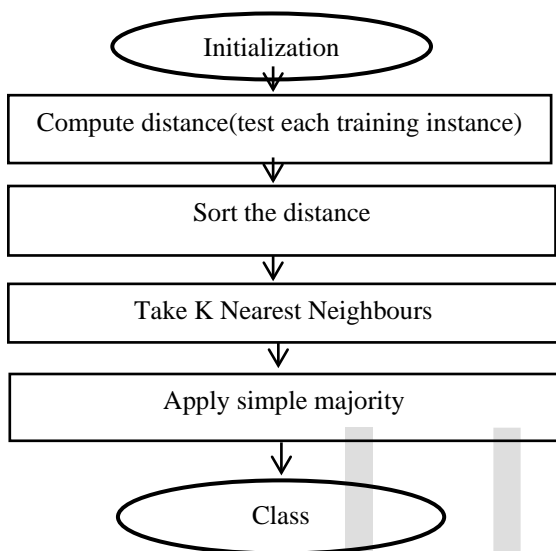
$$IDF(w) = \log_e(\text{Total number of songs} / \text{Number of songs with word 'w' in it})$$

- ✦ Calculate TF*IDF of each tokenized word and clustering of 2 words(bigrams) and 3 words(trigrams) which gives higher weights to words which are rare in a dataset.
- ✦ Perform POS Tagging on each word.
- ✦ Calculate the probability of bigrams.N-gram model measures probability of certain pair of words that would most likely to be found in a Happy, Religious, Heroic or Sad song.Probability of bigram = total no of times two words occur as a

pair/no of times the previous word occurred in document.

- ✦ Now, features are combined to create a feature vector which gives simpler view of the text to a machine learning model.
- ✦ Train the system with KNN classifier according to four mood categories(happy,sad,excited,calm).
- ✦ Calculate the accuracy and other metrics of classification of system.

KNN Classifier



4. Results and Discussion

- ✦ We have studied in literature survey that with the use of basic features solely such as vocabulary usage, use of punctuation marks, top-k word frequencies, sentence count, word count, character count, ratio of words in a sentence etc., classification results are below baseline in mostly cases(if baseline~50%).So, we have used word features.
- ✦ The final results shows the features consisting of tf*idf formula on unigram, bigram and trigram, POS tagging and n-gram feature are of much importance for automatic mood classification. We have calculated bigram probability in n-gram feature.
- ✦ The accuracy values using the tf*idf feature solely and the combination of the tf*idf with other features end up in improved results.
- ✦ The accuracy of classification on average is above 60% individually.TF*IDF of unigram, bigram and trigram yields better results than other features i.e.

	True Positive	True Negative	False Positive	False Negative
--	----------------------	----------------------	-----------------------	-----------------------

up to 85%. The use of the tf*idf feature also increases the performance of the system.

4.1 Metrics Evaluation- The chosen metrics are statistical measures of the performance of KNN classifier which are created by the Confusion matrix of KNN.

- ✦ Sensitivity (TPR or recall)- It measures that how much proportion of positives are there that the classifier can correctly identified.

$$\text{Sensitivity} = \text{TP}/(\text{TP} + \text{FN})$$

- ✦ Specificity (true negative rate)- It measures that how much proportion of negatives are there that the classifier can correctly identified.

$$\text{Specificity} = \text{TN}/(\text{FP} + \text{TN})$$

- ✦ Accuracy- It is the proportion of correctly classified items by classifier.It is used to estimate how well the current dataset can predict an output value for a given input.

$$\text{Accuracy} = \text{TP}+\text{TN}/\text{TP}+\text{TN}+\text{FP}+\text{FN}$$

- ✦ F-Score- It measures the discrimination of two sets of real numbers.It is the harmonic mean of precision and recall(sensitivity). If F-score is larger,more chances are that this feature is much discriminative.

$$\text{F-measure} = 2 * \text{precision} * \text{recall} / (\text{precision} + \text{recall})$$

- ✦ Precision- It implies total number of documents retrieved that are relevant/Total number of documents that are retrieved.

$$\text{Precision} = \text{TP}/\text{TP}+\text{FP}$$

- A TP test result is one that detects the condition when the condition is present.
- A TN test result is one that does not detect the condition when the condition is absent.
- A FP test result is one that detects the condition when the condition is absent.
- A FN test result is one that does not detect the condition when the condition is present.

4.2 Experimental Results

Lok geet	12	11	34	3
Sad songs	2	44	1	13
Heroic songs	5	39	6	10
Religious	0	45	0	15

Table 1(i)- Confusion Matrix of POS Tagging feature

	Sensitivity	Specificity	Precision	F-Score	Accuracy
Lok geet	0.80	.24	.26	0.39	0.38
Sad songs	0.13	0.98	0.67	0.22	0.77
Heroic songs	0.33	0.87	0.45	0.38	0.73
Religious	0	1	NaN	0	0.75

Table 1(ii)- Confusion Matrix of POS Tagging feature

	True Positive	True Negative	False Positive	False Negative
Lok geet	0	45	0	15
Sad songs	1	45	0	14
Heroic songs	15	3	42	0
Religious	0	43	2	15

Table 2(i)- Confusion Matrix of Bigram Probability feature

	Sensitivity	Specificity	Precision	F-Score	Accuracy
Lok geet	0.40	1	1	0.57	0.85
Sad songs	0.80	0.78	0.54	0.65	0.78
Heroic songs	0.87	0.58	0.41	0.55	0.65
Religious	0	1	NaN	0	0.75

Lok geet	0	1	NaN	0	0.75
Sad songs	0.07	1	1	0.12	0.77
Heroic songs	1	0.07	0.26	0.42	0.30
Religious	0	0.96	0	0	0.72

Table 2(ii)- Confusion Matrix of Bigram Probability feature

	True Positive	True Negative	False Positive	False Negative
Lok geet	6	45	0	9
Sad songs	12	35	10	3
Heroic songs	13	26	19	2
Religious	0	45	0	15

Table 3(i)- Confusion Matrix of TF*IDF(Unigram)

	Sensitivity	Specificity	Precision	F-Score	Accuracy
Lok geet	0.40	1	1	0.57	0.85
Sad songs	0.80	0.78	0.54	0.65	0.78
Heroic songs	0.87	0.58	0.41	0.55	0.65
Religious	0	1	NaN	0	0.75

Table 3(ii)- Confusion Matrix of TF*IDF(Unigram)

	True Positive	True Negative	False Positive	False Negative
Lok geet	10	45	0	5

Sad songs	12	44	1	3
Heroic songs	8	42	3	7
Religious	13	32	13	2

Table 4(i)- Confusion Matrix of TF*IDF of bigram-trigram feature

	Sensitivity	Specificity	Precision	F-Score	Accuracy
Lok geet	0.67	1	1	0.80	0.92
Sad songs	0.80	0.98	0.92	0.86	0.93
Heroic songs	0.53	0.93	0.73	0.61	0.83
Religious	0.87	0.71	0.50	0.63	0.75

Table 4(ii) - Confusion Matrix of TF*IDF of bigram-trigram feature

	True Positive	True Negative	False Positive	False Negative
Lok geet	1	45	0	14
Sad songs	7	45	0	8
Heroic songs	15	8	37	0
Religious	0	45	0	15

Table 5(i)- Confusion Matrix of TF*IDF(unigram) + bigram probability feature

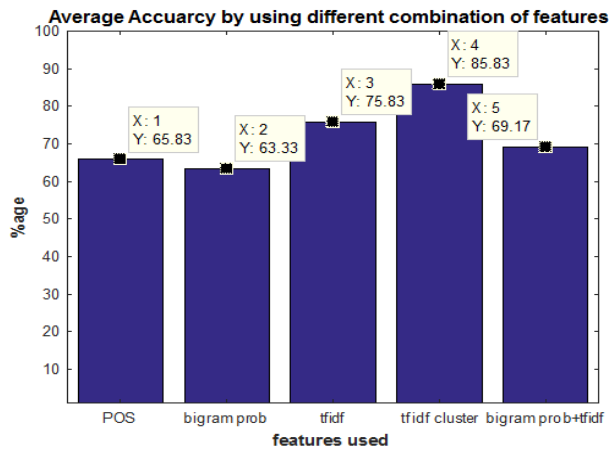
	Sensitivity	Specificity	Precision	F-Score	Accuracy
Lok geet	0.07	1	1	0.12	0.77

Sad songs	0.47	1	1	0.64	0.87
Heroic songs	1	0.18	0.29	0.45	0.38
Religious	0	1	NaN	0	0.75

Table 5(ii)- Confusion Matrix of TF*IDF(unigram)+bigram probability feature

Mood / Features	Category-1 Happy Mood Lok geet(folk songs)	Category-2 Sad Mood Sad songs	Category-3 Excited Mood Heroic songs	Category-4 Calm Mood Religious
POS Tagging Feature	38.33	76.67	73.33	75
Bigram Probability Feature	75	76.67	30	71.67
TF*IDF(Unigram)Feature	85	78.33	65	75
TF*IDF (Bigram+Trigram) feature	91.67	93.33	83.33	75
TF*IDF(Unigram)+Bigram Probability feature	76.6667	86.667	38.3333	75

Table 6- Classification Accuracy by combining various features (in %)



Average accuracy by using different combination of features

5. Conclusion

- ✦ The lingual part of music reveal sufficient information to end up with classifications with an accuracy on average more than 60% of individual feature.
- ✦ Test results show that TF*IDF of unigram, bigram and trigram yields better results i.e. upto 85% as compared to other features. Other features such as POS tagging and n-gram feature also yields satisfactory results but not upto mark yet.
- ✦ The best accuracy of classification of testing data lies in range between 63% to 85% .
- ✦ The current system can recognize the 4 moods such as happy, sad, excited and calm from lyrics of songs.
- ✦ This is just a first step in exploring mood based on lyrics of Punjabi songs.

6. Future Work

- ✦ The implementation in this thesis was only performed with a limited database. Extend the work by enriching the dataset of Punjabi songs.
- ✦ In this thesis classifications were performed according to four moods. Choosing other categories of mood could give new insights.
- ✦ Other musical aspects such as tempo, key ,theme or genre or combination with audio features can be used for expressing or revealing a certain mood or state of emotion to improve system.
- ✦ In Future, the mood model may likely undergo some changes to enhance and to suit best for the Songs scenarios.
- ✦ Moreover, mood classification based on lyrics in Punjabi Songs have not been explored so far.

References

- [1] Xiao Hu, J. Stephen Downie, and Andreas F. Ehman. Lyric text mining in music mood classification. In Proceedings of the tenth International Society for Music Information Retrieval Conference (ISMIR); Kobe, Japan, pages 411–416, October 2009.
- [2] Menno V. Zaanen and Pieter Kanters. Automatic mood classification using TF*IDF based on lyrics. In proc. of the ISMIR, pp. 75-80. 2010.
- [3] R. Mihalcea and C. Strapparava, "Lyrics, Music, and Emotions," In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 590–599, 2012.
- [4] B. G. Patra, D. Das, and S. Bandyopadhyay, Unsupervised Approach to Hindi Music Mood Classification, Mining Intelligence and Knowledge Exploration, R. Prasath and T. Kathirvalavakumar (Eds.): LNAI 8284, pp. 62–69, Springer International Publishing, 2013.
- [5] Oh, S., Hahn, M., & Kim, J. Music Mood Classification Using Intro and Refrain Parts of Lyrics. In 2013 International Conference on Information Science and Applications (ICISA 2013): Suwon, South Korea, 24-26 June 2013. (pp. 1-3). 2013
- [6] Corona, Humberto, and Michael P. O'Mahony. An Exploration of Mood Classification in the Million Songs Dataset, 12th Sound and Music Computing Conference, Maynooth University, Ireland, 26 July-1 August 2015. Music Technology Research Group, Department of Computer Science, Maynooth University, 2015.
- [7] Emil Ian V. Ascalon, Rafael Cabredo .2015. Lyric-Based Music Mood Recognition. Presented at the DLSU Research Congress 2015 De La Salle University, Manila, Philippines March 2-4, 2015.
- [8] Patra, B.G., Das, D., & Bandyopadhyay, S. Mood Classification of Hindi Songs based on Lyrics. In Proceedings 12th International Conference on Natural Language Processing. a. pp. 78–93. 2015.
- [9] S. Chauhan and P. Chauhan, "Music mood classification based on lyrical analysis of Hindi songs using Latent Dirichlet Allocation," 2016 International Conference on Information Technology (InCITe) - The Next Generation IT Summit on the Theme - Internet of Things: Connect your Worlds,

Noida, 2016, pp. 72-76.
doi: 10.1109/INCITE.2016.7857593.2016.

- [10] N. Kashyap, T. Choudhury, D. K. Chaudhary and R. Lal, "Mood Based Classification of Music by Analyzing Lyrical Data Using Text Mining," 2016 International Conference on Micro-Electronics and Telecommunication Engineering (ICMETE), Ghaziabad, pp. 287-292.2016
- [11] Y. An, S. Sun and S. Wang, "Naive Bayes classifiers for music emotion classification based on lyrics," 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan, pp. 635-638.2017.

IJSER