BanglaMusicMooD: A Music Mood Classifier from Bangla Music Lyrics

Nazia Parvin Urmi
Dept.of Computer Science and Engineering
United International University
Dhaka, Bangladesh
nurmi161201@bscse.uiu.ac.bd

Nasir Uddin Ahmed
Dept.of Computer Science and Engineering
United International University
Dhaka, Bangladesh
nahmed151086@bscse.uiu.ac.bd

Md.Habibur Rahman Sifat
Dept.of Computer Science and Engineering
United International University
Dhaka, Bangladesh
Habibsifatuiu@gmail.com

***Abstract*—Abstract—This research work aimed at using word-level features for mood classification. The goal of this project is to build a classifier model to detect the mood of the song by analyzing the lyrics of Bangla Song. Naïve Bayes Classifier is used to predict the probability of different classes based on various attributes. Constructing dataset of 500 Bangla songs consisting of Rabindra Sangeet, Nazrul Geeti, Folk and Modern. We have worked on two types of Bangla moods such as happy and sad, corresponding to the Bangla meaning. At the end of the experiment, our proposed method performed well to classify the mood of the song with high accuracy.**

***Index Terms*—Music recommendation system, music classification model, Naive Bayes Classifier, lyrics analysis**

I. Introduction

In our leisure and recreation time, we usually do some activities such as reading books, listening to music, playing. Among these activities, listening to music is one of the popular ones that have an amplifying effect on social consistency, the emotional state of mind and mood of the listeners [6]. Among music lovers, music classification and recommendation model has obtained large popularity.

The power of music to influence mood, create scenes, routines and occasions are widely recognized and this is reflected in a strand of social theory that portrays music as an influence on character, social structure, and action [2]. ***Music*** is an important art that is very soothing and relaxing which can make our bad days better. The more we will be able to examine what we want in music the more music will be appealing to us. Around 68 percent of people who have aged between 18 and 34 years old announce that they listen and enjoy music every day.

In our approach, we created our own Bangla Song dataset for filtering and music mood prediction of listeners that can be associated with happy or sad emotions. Our main focus is to create a system that can be able to predict song mood that will be applied to the song database to select music by its lyrics analysis.

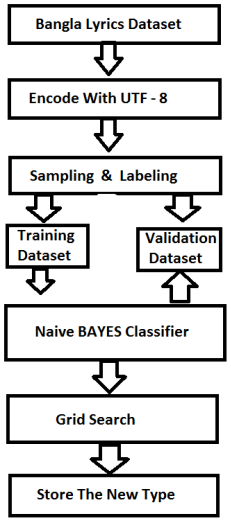

Fig. 1. Flowchart of the proposed method

Based on lyrics analysis, we used a **naive Bayes classification** ml when the conditional independence assumption hold, these classifiers will converge faster than any other models like Logiodel for mood prediction. Naive Bayes classifier performs elastic Regression. Naive Bayes has very high learning efficiency and it can estimate all the probability just need a scan of the training data [3].

Grid Search has used for finding the optical hyper-parameters of a model that outcome the most accurate predictions. Based on the hyper-parameter values, the performance of the entire model is specified. Grid Search is a technique that works iteratively.

In Figure 1, first, we created a dataset. As it contains Bangla character so it needs to encode it in UTF-8 format. Then assigning the class label. Next dividing the dataset into test and

validation. Then using Naıve Bayes classifier, and grid search to find out the mood. Grid search used for hyper-parameter specification and evaluating the model. Then saving the newly classed lyrics.

In machine learning. the value of a **hyper parameter** is set before the learning process begins. It's just a value by which measured how much performance can be gained by tuning it. Depend on the choice of hyper-parameters, a model can require the time to train and test.

K Fold Cross validation is a numerical method used to determine that how skill a machine learning model is. In this method, the data is divided into k subsets. Then the whole method is repeated k times, each time one of the k subsets is used as the test set/ validation set and the other k-1 subsets are put to form a training set.

In II, we provide our project-related work information. III methods that we are using in our project. The process and results we obtained in our project are discussed in IV. there are small discussions about our work in V. In VI, we summarize our whole works and also provided future works of our project.

II. Related Works

In this section, we discuss some related works on music classification, also some recent literature review about mood prediction in text [1], very little has been done so far to address the classification of lyrics according to their mood.

A. Music Classification

Widowati et al. [7] classified the music into moods like happy, peaceful, angry and sad. They performed this classification with 200 track database for training and 50 for testing. Kashyap et al. [4], in their paper, worked on text mining. They used naïve Bayes for strong independent assumptions to divide the model. They also focused on parts of speech, so that word can be easily tokenized. In previous, the majority of work-related with music mood have done by categorical (Happy, sad) models of emotions [5].

B. Mood Prediction

Widowati et al. [7] used some music features like pitch, pulse clarity, tempo, key, and scale. To classify, Convolutional Neural Network (CNN) is used in this paper. For this, the accuracy they have got is 82 percent. Kashyap et al. [4] took motivation from mood detection challenges. The features they focused on Mood perception, Mood cataloging, Acoustic cues. Acoustic cues made their work significant. They followed the divide and conquer approach to break a model into multiple ones. By gathering the information they trained the model. They divided their dataset based on models in the categories like stop words, sad words, angry words, aggression words, average words, happy words etc. It is like a bunch of trees. So feature mapping becomes easier. In this paper [5], they predict the mood of a song based on song lyrics. Happy or Sad mood can be detected by high precision based on the text lyrics features.

C. Bangla Mood Processing

Bangla's mood processing depends on word embedding. A single Bangla word can occur in a sad song or a happy song. The divide and conquer approach can be followed. At the same time, the probabilistic approach applied to find the best match. Though not much information available for Bangla mood processing.

III. Methods

A. Dataset

For building the machine learning model we need a dataset. But for Bangla music mood dataset in .csv format was not available. So we created our dataset. We categorized the dataset of Bangla song lyrics by using lyrics, artist, mood, year, genre. We made the dataset aiming 500 songs. We divide it 60 percent for training, 40 percent for validation. Another thing is we created the dataset is (.txt format) because UTF-8 encoding was creating a problem to fit (.csv files). We labeled the mood in Bangla in the dataset. We also manually created the stop words list, so that the model can find map them.

B. Feature Extraction

From the lyrics, tokenization has to be performed. For tokenization, the lyrics were converted into vectors. But it's not enough. Because it could have different n-gram sequences. Porter stemming algorithm is used for inflexional endings from words. Binarization is used to transform both the discrete attributes and continuous attributes into binary attributes. For counting the number of times each word occurs in each lyrics term frequency (tf) has been used.

$$tf - idf(t,d) = tf(t,d) * idf(t) \tag{1}$$

The equation was used to calculate term frequency depending on normalized term frequency. Term document frequency used as a form of inverse document frequency.

$$idf(t) = log(\frac{(1+nd)}{1+df(t,d)}) + 1 \tag{2}$$

"nd" is the total number of lyrics. In the lyrics dataset, the columns correspond to terms and rows correspond to documents in the collection. Value of "nd" increases proportionally with the number of times along with the word appears as a document.

C. Tokenization

In the dataset, we have lyrics. Lyrics are like sentences. From the lyrics for it is important to trim desired content. Tokenize texts, select tokens of interests, create an NLTK text. As the work is going on Bangla music mood, so CLTK [Classical Language Toolkit] is used. By using CLTK, normalizing the words making vocabulary. TokenizeSentence named function is used to divide sentences into smaller parts

of words. It works by detecting spaces between two words. Then finding the most common words is the task. This thing is done by stop words. It is a kind of pre-processing. Then the words converted into vectors. Count vectorize is used to build vocabulary. In the vocabulary one word can occur many times, to remove the inflation porter stream algorithm is used. We wrote a tokenization function by using CLTK to token the Bangla strings.

D. Model Selection

Grid search applied to optimize the "Naïve Bayes Model". F1 score was applied to the class label of the dataset. To support F1 score Recall and Precision were used. The confusion matrix was generated to see the number of the true positive rate as well as the negative rate.

$$F1 = 2 * \frac{precision * recall}{precision + recall} \tag{3}$$

where

$$precision = \frac{TP}{TP + FP} \tag{4}$$

and

$$recall = \frac{TP}{TP + FN} \tag{5}$$

TP = number of true positives, FP = number of false negatives, and FN = number of false negatives.

$$P(W_j/X_i) = \frac{P(X_i/W_j) * P(W_j)}{P(X_i)} \tag{6}$$

The Bayesian theory was applied to calculate the Prior and Posterior probability. It is the way to classify test data. Binary feature vector supports the Bayesian Theory to count probabilities.

IV. Experiment and Results

A. Experiment

Manual class labeled assigned. The data set consists of 40 percent sad, and 60 percent happy song lyrics. Cross-validation was performed measuring the F1 score. In figure-2, we generate a happy word cloud of frequent words as well as in figure-3, we generate a sad word cloud. After collecting the Bangla song lyrics dataset, we manually assigned a class label. Then we divided the dataset into a test and train. The training dataset consists of 40 percent happy and 60 percent sad song lyrics in Bangla. So it can be said that the model is a bit biased to sad lyrics, it is because of the lack of Bangla lyrics resources. For model selection, we applied the grid search technique. For hyper-parameter specification, we used a grid search approach. K-fold cross-validation was used. Here we used k=2 because there was some decoding problem. We used the ROC curve for measuring the quality of our model. Grid search used three separate Naïve Bayes models.

Fig. 2. Happy wordcloud visualizations of the most frequent words.

Fig. 3. Sad wordcloud visualizations of the most frequent words.

For features vector mapping Multi-Variate Bernoulli Bayes was used. For term frequency, Multi-Variate Bernoulli Bayes was used and tf-idf was used for checking the occurrence of a word.

B. Results

The word cloud showed the most frequently used Bangla words. Grouping the songs by lyrics showed the subsample. We applied the grid search by using three different Naïve Bayes models. Multinomial Naïve Bayes classifier showed good performance. But for some technical issues, we could not generate enough evaluation curves.

After selecting the model the final classifier was used for training. The model is generating a result. As we did not get enough number of data, the model has some limitations to find out the actual results that we were expecting. If we can train the model with a high amount of data then the result would become more clear. The results we have obtained are shown in Table I.

TABLE I
RESULT GENERATED BY THE BANGLA MUSIC MOOD MODEL

ACC	PRE	REC	F1	ROC	AUC
100.0	100.0	100.0	100.0	100.	100.0
100.0	100.0	100.0	100.0	100.	100.0

V. DISCUSSION

Our model worked on our dataset. As there are limited resources on Bangla music mood prediction so, the comparison is a little bit challenging. Our model is better because we made the dataset and class labeled them manually. We used the optimal grid search technique. And the removal of stop words made that more redundancy free. After a deep analysis of result data, it shows that the number of sad songs increases over the years passed. It is very clear in fig. It strongly gives knowledge that in recent years Bengali people are very much comfortable with the sad song rather than happy songs. Or the artists have very much interest in sad songs over happy songs. However in modern society people are having so much trouble with their emotions, that they are listening to a large number of sad songs than previous generations. It can be used as a mood filter which will describe a music recommendation system. In our naive Bayes model among all three of them that we have used all perform better than before if they stop words are not used for a grid search. The higher ROC AUC derives that in the used model some non-relevant words are both describes for happy sad. As explained before the multinomial naive Bayes model worked better than Bernoulli naive Bayes model which was used to input the binary feature vectors. The mood classifier has high training and validation precision but the result for cross-validation for electing the model did not completely satisfied overfitting data. However, the high precision for the classifier is still successive for reaching the goal.

VI. CONCLUSION AND FUTURE WORK

In this paper, we tried to focus on Music Mood Prediction by using Naïve Bayes Classifier. We worked with the "Bangla Song" dataset. We focused on the strong independence assumption of Naïve Bayes. Besides Naïve Bayes need fewer data to train our model. Naive Bayes classifier plateaus above a certain threshold in our model. Till now the model is working with a little size dataset. As we increase the instances of the dataset, it would work much better. While using Bangla Dataset encoding was a bigger issue. Tokenizing was another issue. By using Classical Language Toolkit we overcome the problem. Grid search was used to find out the combinations of "Hyper-parameters". Due to a lack of Bangla resources, the model still struggles to find the best solution. But in most of the cases, it can identify the mood. By using this model, we can get to know how people categorize music besides, what actually a person feels. Music trends can be plotted and assumed also.

In the near future, we would like to predict the genre of songs and also try to improve the accuracy level of our music classification model.

REFERENCES

[1] Cecilia Ovesdotter Alm, Dan Roth, and Richard Sproat. "Emotions from text: machine learning for text-based emotion prediction". In: *Proceedings of the conference on human language technology and empirical methods in natural language processing*. Association for Computational Linguistics. 2005, pp. 579–586.

[2] Tia DeNora. *Music in everyday life*. Cambridge University Press, 2000.

[3] Zheng Gong and Tian Yu. "Chinese web text classification system model based on Naive Bayes". In: *2010 International Conference on E-Product E-Service and E-Entertainment*. IEEE. 2010, pp. 1–4.

[4] Nirbhay Kashyap et al. "Mood based classification of music by analyzing lyrical data using text mining". In: *2016 International Conference on Micro-Electronics and Telecommunication Engineering (ICMETE)*. IEEE. 2016, pp. 287–292.

[5] Sebastian Raschka. "MusicMood: Predicting the mood of music from song lyrics using machine learning". In: *arXiv preprint arXiv:1611.00138* (2016).

[6] Daniel Västfjäll. "Emotion induction through music: A review of the musical mood induction procedure". In: *Musicae Scientiae* 5.1_suppl (2001), pp. 173–211.

[7] Fika Ulfa Widowati, Fitriyanto Santi Nugroho, and Guruh Fajar Shidik. "Classification of Music Moods Based on CNN". In: *2018 International Seminar on Application for Technology of Information and Communication*. IEEE. 2018, pp. 318–321.