



SURVEY ON CLASSIFICATION BASED ON AUDIO & LYRICS FOR TAMIL SONGS

Harthi Vasudevan¹ & Sathya M²

Abstract- Songs are present in everyday life and is used for the wide range of objectives. The amount of musical database has to be considerably increased in day to day. Collecting those files are meant to be an important study namely “Music Information Retrieval”. Ever growing study on “The Machine learning” deals with different algorithms for obtaining with the abundance of audio recording that are available in digital audio formats. The availability of tools and Tool-Boxes for the extraction of the musical properties have large quantity on the studies of machines learning and MIR. Relevant problems in MIR include the classification of the songs into genres, making it possible the summarization of common features shared by different songs. Some of the problems that are faced in MIR are the classification of the songs into genres and making it possible for the data sets through the algorithms. In the paper the aim is to present study of various works that are related to audio classifications and the features extraction based on the data sets of statistical method and using the vector file of the lyrics that were manually labeled shows a lot of promise. Especially with the experiment is on two different classification models were achieved. Through this, more information was learned about the relationship of the content of lyrics with the determination of the mood. Calculating normalized vectors for all his songs and then applying a dimensionally reduction algorithm to reduce the vector dimensions, making it easy to visualize. Second, we allocate those songs classify under features in Logistic regression(LR).Furthermore, it can be observed from our experiment that information gain based feature selection gives better and consistent accuracies than other feature selection algorithms Finally, the issues that are related to the songs that are presented and those are addressed in the future work in order to obtain better classification accuracy.

Keywords—Vector format; Songs ,Music retrival information, classification, Analysis songs

1. INTRODUCTION

In the songs, both the lyrics and the audio plays an important roles. In recent years,due to enormous growth of social media, the interest in the field of songs classification and analysis has been growing rapidly. A company can monitor poeple’s opinion on their products and can update themselves by analyzing and understanding the customer’s opinion or perception about their products. Most of the data that is present today is in the form of multimedia that includes songs, movies videos etc. The emotion, genre and theme of the the songs can be obtained only by the lyrics and audio. The Online musical databases and user interactive applications have been considerably increased and makes the fact and effective tools to classify and to retrieve musical content are essential for extracting and organizing the data. The inherent way to organize the song is to classify the songs that this dividing the songs into classification of pieces according to the category described by their music conent.Classificationof the text negative, positive and neutral was the first encouragement to detect emotions in a sentence.the most common ways to label the categories/atist.Moreover, the fuzzy definition of music genre can yield the multi-categorization of individual songs. The songs should be classifiable intot eh stage of automatic classification. The automatic classification of songs are categorized into 5 different stages (i)Audio content format based. (ii)Symbolic features content based. (iii)Lyrical based approach. (iv)Web community data based. (v)Hybrid approaches. Other than this data set model also lays an important roles of classification. So, based on the machine learning Cluster based classification I.e Unsupervised learning will be used in it. In that cluster based algorithm some of the regressions are studies. A song is composed of both verse and chorus segments. Users mood can be drive him in selecting some song, and that song can be used to invoke sentiments in the listeners, which can make the listeners being attached to the emotion of the song or it can make listeners change their present mood also. Thus, mood and song-selection are the features that are interdependent. Thus, the aim of this work is to study about the various methods,techniques and tools used for song emotion recognition. In this paper all the classifications approaches including the regression models will also be presented and explained

2. LITERATURE SURVEY

2.1 Introduction

In the song Classification, most of the works have been carried by the Lyric classification only less number of the songs are classified in the both the basis of the Lyric and the audio format. The Lyrical approach and the audio based approach are the main approaches this field but from the automatic song classification the songs are classified into the 5 different categories. The former is the acoustic representation of music, obtained by sampling the sound waveform. Audio descriptors are mainly

¹ Student, Computer Science Department, Pondicherry University, Pondicherry, India

² Assistant Professor, Computer Science Department, Pondicherry University, Pondicherry, India

extracted with the help of Fourier analysis and signal processing techniques. This category represents the majority of studies in song genre classification and also include music descriptors as the MFCCs (Mel-Frequency cepstral coefficients), spectral shape features and the temporal and energy features. Symbolic data usually take little space and facilitates storage and communication. In the paper there are different papers are studied and analyzed based on the algorithms are other features of classifications and also the different model of data sets for the classified data is also explained. The following section discusses about the various categories of the song classification based on the lyric and audio is explained with their methods, algorithm, tools and also data sets used on their approach is categorized. Sameh Souli The support vector machines method based on Gaussian Kernel is used to classify the datasets due to its capability to deal with high-dimensional data. Our SVM based multiclass classification approach seems well suited for real world recognition tasks. Experimental results have revealed the good performance of the proposed system and the classification accuracy. Paul Ruvolo Audio classification typically involves feeding a fixed set of low-level features to a machine learning method, then performing feature aggregation before or learning. Instead, we jointly learn a selection and hierarchical temporal aggregation of feature, achieving significant performance gains. Mengu Qiau[39] Experimental results show that our approach is successful in discriminating MP3 covers and the steganograms generated by the stenographic tool, MP3Stego, in each category of signal complexity, especially for the audio streams with high signal complexities that are generally more difficult to steganalyze. Antonio M. Rinaldi manage and share ontology in the WEB. The system provides a graphical interface to add multimedia features are automatically extracted using algorithms based on MPEG-7 descriptors.

Win De Mulder structured overview makes it possible to detect the most promising techniques in the fields of recurrent neural networks, applied to language modeling, but it also highlights the techniques for which further research is required. Chu Guan[42] enhanced by this new formulation, we develop the SMO method for optimizing the MKLA dual and present a theoretical analysis to show the lower bound of our method, With the estimated model. We computer the matching degree of users and songs in terms of pitch, volume and rhythm and recommend songs to users. Debora C. Correa[50] A number of approaches usse such musical information to process, retrieve and classify music content. This manuscript provides an overview of the most important approaches that deal with music genre classification and consider the symbolic representation of music data. Current issues inherent to such a music format, as well the main algorithms adopted for the modeling of the music feature space are presented.

Sascha Fruhholz new integrative neural network view unifies the decoding of affective valence in sound, and ascribes differential as well as complementary functional roles to specific nodes within a common neural network. It also highlights the importance of an extended brain network beyond the central limbic and auditory brain systems engaged in the processing of affective sounds. Y.M.G Costa[44] compare the performance of texture features with that of commonly used audio content based features (I.e from the MARSYAS framework), and show that texture features always out performs the audio content based features. We also compare our results with results from the literature. Dongge Li[50] In this paper, we address the problem of classification of continuous general audio data (GAD) for content based retrieval, and describes a scheme that is able to classify audio segments into seven categories, total of 143 classification features for their discrimination capacity. Zhouyo Fu[46] investigate the bag-of-features approach for music classification which can effectively aggregate the local features for song-level feature representation. Moreover, we have extended the standard bag-of-features approach. Prafulla Kalapatapu [47] the feature selection algorithms with the respect to classifiers and percentage of classification accuracies for all the classification algorithms, observed from our experiment that information gain based feature selection gives better and consistent accuracies.

2.2 Analysing the song

Most of the data that is present today is in the form of multimedia that includes songs, movies, videos etc. Among these, Mood of songs has become a hot research topic due to the increasing demand of songs access via mobile phones. Similarly, Mood of songs can be found by following either Lyrical approach by making use of Lyrical features or by following Audio approach by making use of audio features that can be extracted using various toolboxes and can also be found by also by combining both the lyrical and audio based features to enhance classification accuracy. In this survey, a total of twenty seven papers have been studied. Recent papers from various sources have been collected and are analysed in this paper. Eight papers were studied based on lyrical approach, four papers were studied based on audio approach, four papers were studied where both lyrical and audio approaches were considered and a paper was studied which includes text, audio and visual clues. In addition to that, ten papers related to audio analysis, for the extraction of audio features have also been studied. Lyrical approach uses lyrical words uses as features and classifies input song into any atleast one of the emotions. Audio approach uses audio features of each song to classify an input song. The following section discusses about the various approaches used for Analysing the mood of the songs and tools that are used for audio feature extraction to detect the emotion of songs.

2.2.1 Audio content format based

As classifying a song, solely based on lyrics is a challenging task, Ashley M et al. [3] proposed a method to identify the emotional polarity of song through lyrics using Natural Language Processing. A dataset containing 420 songs had been considered, which contained equal number of songs from negative and positive emotions. Last.fm API [32] was used to retrieve tags for songs. Sentiment analysis of songs is a complex task because songs may start up with positive emotion but

can end up in negative emotion and vice-versa, Similarly, some stanzas of song lyrics may constitute to a particular emotion and also songs can talk negative emotions about positive things and vice-versa. To address all those problems, different algorithms were proposed. Word list is simplest algorithm that is based on word counts. The songs are classified by maintaining a counts of the words that appears in the lyrics. Here, the algorithm loops through the words of the songs. In word dictionary, for song classification, the algorithm loops through the lyric words, determining whether each word is present in both positive and the negative dictionaries or in single or in none. Cosine similarity method based on term frequency-inverse document frequency is used to find similarity between an input song and to the songs in training set. The proposed method could not deal with negations. Results shows that song lyrics when considered alone does not show promising results but it improves the classification accuracy when additional features such as audio are combined to lyrical features. Table.3.1 lists out the features and methods that various authors have used to classify songs.

Table.2.1. Features and Methods for Lyrical Analysis

Author Name	Features/ Method	Dataset	No. of Songs	Emotions
Ashley M et al. [3]	TF, IDF	Created manually	420	Positive, Negative
Yanqing Xia et al. [2]	Sentiment related words	5SONGS	2,653 Chinese Pop songs	Positive, Negative
Corona et al. [5]	TF, IDF, BM25	Million Songs	32,302 English songs	-----
Shanmugapriya et al. [6]	Hidden Markov Model	Collection of songs from Website	-----	Happy, angry, sad.
Emil Ian et al. [7]	TF, IDF, Key Graph	Collection of songs from Internet	200 songs	-----
Doran Walsten et al. [28]	K-Means Clustering	Collected from various sources	-----	Classic rock, county, grunge, modern rock, pop, r&b, rap
Govind Sharma et al. [27]	Bag of Words/ Latent Dirichlet Allocation	Collection of songs from Internet	-----	Happy, sad, angry, Tired, Love, Funny
Charulatha et al. [29]	Rakshini Algorithm	-----	-----	Happy, sad, calm

Vector Space Model [VSM] is used for representation of text documents. To address the problems in VSM, Yunqing Xia et al [2] proposed a Sentiment Vector Space model for song sentiment classification based on lyrics. The words that indicate sentiment are used as features in this model. The sentiment units are classified based on the occurrence of sentiment based words, modifiers and negations. In their model, the songs were labelled as heavy-hearted and light-hearted classes. For evaluation, a song corpus named 5SONGS which includes 2,653 Chinese pop songs was created manually, out of which 1,021 were heavy-hearted and 1,632 light-hearted. Songs were divided into testing and training phases of which 2,001 songs were used for training and remaining 652 songs were used for testing purposes. Audio-based, Knowledge based and Machine learning based approaches were used for conducting experiments. In Audio based approach, a twelve- dimension vector was used. Knowledge based approach used Hownet to detect the words that indicate sentiment and to indicate the location of sentiment units in the song lyrics. In Machine Learning approach, an algorithm named SVM-light is implemented which is based on VSM and s-VSM models. Experimental results showed that Vector Space model gives better performance when compared to VSM and Knowledge based approaches. The sentiment vector space model did not consider linguistic rules for sentiment classification.

To explore mood classification in Million songs dataset [34], Corona et al. [5] proposed a comprehensive evaluation of moods based solely on song lyrics. Three granularities for representing moods namely quadrants, moods and group tags were studied. Experiments were conducted on Million songs dataset that had 32,302 English songs. Vector Space Model is used for representation of songs, where each song is represented by a vector. Studies were performed using three different term-weighting schemes. Term Frequency (tf) indicates the number of times a particular term t occurs in a document and it assigns higher weights to those terms that occur very frequently in a document. Term Frequency-Inverse Domain Frequency assigns higher weights to those terms that are rarely used in the collection. BM25 [35] a term weighting scheme, is also used for retrieval and text classification. Delta td-idf has been proposed specifically for sentiment classification and it assigns higher weights to those terms that appear primarily in single class. Experimental results showed that lyrics can alone be useful for music mood classification, working well for some moods and achieving an accuracy of nearly 70%. The dataset constructed did not consider complete lyrics of a song. So, there is a need for new benchmark dataset to achieve better results in this work. By providing such a dataset, it would be easy to compare different approaches that are used for music mood classification.

Shanmugapriya in the method to determine sentiment from lyrics, by using Hidden Markov Model (HMM) which is based on WordNet representation. The present mood of user can be found out by the song they are listening

to at that particular moment. This information obtained can be used in future for song recommendation systems. Here song lyrics are considered to be as a text file, where each song represents a mixture of moods. The input songs are pre-processed using tokenization, stemming, removal of stop words, average calculator, and word sense disambiguation by making use of WordNet representation, and then classification is done by HMM that classifies the songs into one or more classes such as happy, angry, sad. The sentiments of each song were mined using HMM. For performance validation, a dataset was used that consisted of different moods, which was manually annotated by users. Precision and Recall were calculated as a part of experimentation and evaluation. The number of songs considered for evaluation and the output was classified into various classes was not properly explained.

Emil Ian et al in the Lyric based mood recognition for music analysis. The word level features were alone considered for classification and the work was focused on recognizing mood of OPM songs based on lyrics. Song lyrics may contain various sections such as intro, chorus, refrain and other parts. As lyrics are manually submitted by users in websites, the chorus lines may be represented as instructions in short form and these chorus instructions are replaced by the exact portion of lyrics. Word-level features such as Key Graph and Term frequency and Inverse Document Frequency were used for experimental evaluation. The dataset contained 200 song lyrics that were collected from internet websites. Two different approaches were used which includes an automated approach that uses valence and arousal, and a manual approach. It was found that approaches with manual annotation performed better on comparison to automatic approach. On performing experiments, a very high accuracy was achieved when key Graph feature extraction method was applied to song lyrics. The method proposed could only improve the overall accuracy for song lyrics that are sad. This paper considered only a limited number of songs and scope of moods can also be widened by considering more number of moods.

Doran Walsten et al. in the proposed a method for Song Genre classification by the quantitative analysis of song lyrics. A machine learning algorithm, K-means algorithm was applied to cluster the similar songs into same genres. K-means algorithm aims to cluster n observations into k clusters. For input data processing, a java program was written to perform all operations that are associated with K-means algorithm. Experiments were conducted to find out the most valuable features for clustering. The main drawback of the proposed approach is that features have to be recomputed for each song if there is any minor change in the lyrics. Initially experiments were conducted using unsupervised approach, and later conducted on supervised approach such as Support Vector Machine (SVM) as K-Means algorithm had faced some challenges. Even on using supervised learning and thousands of features, their classification accuracy occasionally only reached more than 25%.

Govind Sharma et al. in the used Latent Dirichlet Allocation (LDA) for mining sentiments of songs. Research is based on the assumption that songs are generally classified based on genres, which won't entirely reflect sentiments. Thus an unsupervised scheme has been proposed to classify songs. Songs have been classified into multiple classes such as happy, sad, angry etc., or into either positive or negative categories, based on the application. To mine the songs LDA has been used. Songs mined by LDA can represent moods. A dataset that considered 6 moods was used for evaluation purpose. LDA uses the bag of words approach. A particular type of song can have a particular kind of lyrical structure that can decide its sentiment which was captured using LDA. Manual annotation was needed for a large number of songs to get better validation results. Negation and ambiguous words were not handled properly.

Charulatha et al. in the proposed a music player that relieves stress by making use of intelligent sentiment analysis. The work was based on the assumption that songs expresses emotion and the mood of a person can be analyzed from the type of song the user is listening to. The proposed system asks the user to create an account, and then a profile is created for each user. The profile consists of musical preferences of user for each emotion. While listening, the user can select any song from the playlist, and correspondingly the lyrics are analyzed and the present mood of user is determined. This particular mood of user is selected and the songs are played as per the preferences of the user based on that emotion or mood. The lyrics of the song are retrieved using API's such as CajunLyrics, LyricsWiki and Musixmatch. An algorithm named Rakshini algorithm was developed to detect the current mood of the user. Three song emotions namely happy, sad and calm were considered. The proposed system could be enhanced to detect the mood of the user by considering user's music preferences. Only a limited range of emotions have been considered and they could be increased.

2.2.2 Analysis using Audio Features

To extract audio features from music files, Kee Moe et al made use of Music Information Retrieval [MIR] toolbox [36]. The features such as pitch, tempo, tonality, dynamics and timbre are extracted from MIR toolbox. A total of 17 audio features are extracted from MIR toolbox. A music database containing 100 music clips was used for analysis. They had considered only four emotions namely sleepy, excited, sad and happy. MIR toolbox is chosen for the process of feature extraction because it can be used an integrated function in MATLAB, it can reduce work complexity and complex computations can also be easily performed using this toolbox. As part of Audio processing, all the input songs are pre-processed and trimmed to two minutes and are converted into .wav format for feature extraction. Then, the features are extracted using MIR toolbox and then the training set is made ready by getting subjective tests from fifteen people. The testing set also undergoes pre-processing and feature extraction stages.

Based on the features generated from the testing set, the features are compared to those of in training set and are comparatively classified. A relatively less number of music clips have been taken into consideration and addition of lyrical

features could have improved their accuracy and also the classification mechanism using extracted features was not mentioned. Table.3.2 lists out the features and methods used for audio analysis used by various authors.

Author	Toolbox/Technique	Features	No. of Songs	Emotions
Kee Moe et al [11]	MIR Toolbox	Pitch, Tempo, imbre, Key, Energy	100	Happy, sad, excited, bored, sleepy, nervous, peace
Souraya et al [24]	Speech Recognition Technology	Textual features	36	Positive, Negative
Braja Gopal et al [26]	jAudio Tool	Rhythm, Timbre, Intensity	230 Hindi songs	Good natured, sweet, fun, aggressive, cheerful etc.
Jayita et al. [31]	MIR Toolbox	Temporal length, low energy, rms energy, pulse clarity, tempo, zero crossing rate, spectral irregularity, rolloff	100	Blues, Classical, Chamber, Orchestral, Jazz, Pop, Hiphop, Techno, Rock, Hard Rock, Soft Rock

Souraya et al. [24] proposed a model that accepts any audio material and then it studies its content by making use of machine learning algorithms by automatically converting the audio files into text files and then the mining the content in the text. Speech recognition technology was used to convert those audio files into text files. Proposed model was applied to analyze telephone calls in a call center, which was focused to distinguish between positive and negative calls. The main task lies in finding the keywords that differentiates between positive and negative calls, and then clustering the calls into similar groups. Dataset considered consisted of only thirty six audios that were extracted in twelve different scenarios. Nineteen of them were manually labelled as positive calls, whereas seventeen of them were labelled as negative calls. On experimentation, proposed method had achieved only 44% accuracy. The conversion of data from audio to text resulted in data loss and thus it lead to less accuracy results.

Braja Gopal Patra et al. [26] proposed a method for automatic mood classification for songs. Dataset consisted of 230 Hindi songs of 30 seconds duration. Three types of audio features were extracted namely rhythm, timbre and intensity. jAudio tool was used for audio analysis. Music Information Retrieval Evaluation eXchange(MIREX) mood taxonomy was used for experimentation and decision tree classifier (J48) was used for classification purpose and an average accuracy of 51.56% was achieved using the 10 fold cross validation. Preparing the large dataset with more number of songs and collecting the corresponding lyrics of those songs could improve classification accuracy. The proposed system achieved quite a low accuracy on comparison to the other existing classification systems for English songs.

Jayita et al. [31] worked on feature selection method for classification of audio files. They considered audio files of three types namely speech, music and background sound. They also classified the music into classical and non-classical music. They extracted ten features using MIR toolbox namely temporal length, low energy, rms energy, pulse clarity, tempo, zero crossing rate, spectral irregularity, pitch, inharmonicity, and roll of feature. Their database consisted of 100 .mp3 files with 10 from each emotion namely Blues, Classical, Chamber, Orchestral, Jazz, Pop, Hiphop, Techno, Rock, Hard Rock, Soft Rock. A audio classification was proposed that uses gain ratio for selecting a splitting attribute that classifies the music into various genres. For Gain Ratio's calculation they have used a feature named Pulse clarity that has highest discrimination power when compared to other extracted audio features. The audio files selected by them belongs to a particular genre shares a similar range of values for the features and hence the classification accuracy achieved was about 90%.

2.2.3 Analysis both Lyrics and Audio Features

An approach to detect sentiment from Telugu songs based on multi-modality i.e using song lyrics and audio has been presented by Harika et al [10]. Song lyrics are represented as textual lyrics and features are extracted using bag of words approach. By making use of these features, Doc2Vec generates a vector for each song. Classifiers such as Support Vector Machine, Naïve Bayes and a combination of these classifiers is used. Dataset considered consists of 100 songs, out of which 50 are happy and the remaining 50 are sad. Songs were split up into testing and training sets, of which 40% of the songs for training and the remaining 60% for testing purposes. Three experiments were conducted on each song; Whole song, Beginning 30 seconds of the song, and last 30 seconds of the song. On analysis, it has been observed that beginning of a song gives high classification accuracy as compared to whole song or to the ending of the song. Audio features like prosody, temporal, spectral, Mel-frequency Cepstral Co-efficients, chroma and harmonic tempo were extracted from Open-Source Emotion and Affect Recognition Toolkit [openEAR] toolkit [10]. Features extracted are trained using Classifiers such as Support Vector Machine, Gaussian Mixture Models and a combination of both these approaches. The modalities of both the approaches is combined. The

number of songs used for classification is also very limited. However, it has been found that lyrical features when combined with audio features can be best used to classify the mood of a song. Table.3.3 lists out features and methods used for songs analysis.

Table.2.3. Features and Methods for Songs analysis

Author	Method/Tool	Features	No. of songs	Emotions
Harika et al [10]	openEAR	Bag of Words, prosody,temporal, spectral, Mel-frequency Cepstral Co-efficients, chroma andharmonic tempo	100	Happy, sad
Xiao Hu et al [4]	MARSYAS	Bag of Words, Text Stylistic features	5,296	Calm, sad, romantic, cheerful, aggressive, anxiety, etc.
Adit et al [8]	EchoNest	Arousal, Valence, Tempo, dance ability, energy etc.	1,287	Calm, energetic, dance, sad, happy, romantic, seductive, angry, hopeful.
Xiao Hu et al [25]	Marsyas tool	Bag of words, parts of speech, and function words	4,578	Anxious, hopeful, exciting, gleeful, glad, sad etc.

To improve mood classification, Xiao Hu et al [4] combined both lyrical and audio features. Evaluation was done using lyric text features in music mood classification such as bag-of-words features, text stylistic features and psycholinguistic features. Experiments conducted revealed that the lyric features that are mostly useful in classification were a combination of function words, content words, Affective Norms of English Words [ANEW]

[37] scores, GI psychological features, text stylistic features and affect related words. The experiments were performed on 5,296 songs that includes both audio and lyrics for each song that represents 18 categories of mood that are obtained from social tags. The results showed that when lyrics are combined with audio features, an improved performance was achieved when compared to audio-only features. When lyrics and audio were combined, late fusion outperformed an audio-only system on the same task by 9.6%. When experiments were conducted on learning curves, it had been discovered that when audio and lyrics are combined, it can drastically reduce the samples required to achieve the better performance than using single source systems such as lyrics or audio. Thus, through the experiment findings it has been found that they could improve the effectiveness and efficiency of music mood classification. The drawback is that experiments were conducted only with SVM classification model. Others such as Naive Bayes were not tested which may outperform SVM.

Adit et al [8] proposed a method for detecting the emotion of songs based on both audio and lyrical features. Lyrical features are obtained from segmentation of lyrics into verse and chorus segments. WordNet and ANEW dictionaries were used to compute the valence and arousal features. The audio features such as tempo, dance ability, energy etc. that are derived from EchoNest are used as features to supplement the lyrical features. The testing and training sets are constructed based on the social tags that are extracted from Last.fm [33] website. The classification is done by k-Nearest Neighbour algorithm that retrieves k songs that are similar to a given input song. The similarity measure they have employed is Euclidean Distance. A feature weighting scheme has also been applied to improve the classification accuracy. It has been found that assigning higher weights to chorus segments had provided better results. This observation had led to the idea that the words that are repeated have more influence on the user. This method assigns at least one class to each input song given. The features can also be extended for improving accuracy and this proposed method can also be applied to other languages.

Xiao Hu et al [25] used Lyric text mining for classification of music. Audio and Lyrics features were taken into consideration. A total of 4,578 songs were considered, of which 2,829 belongs to positive set of emotions. 18 mood categories were considered for classification. Lyric processing includes feature extraction such as bag of words, parts of speech, and function words. On experimentation it has been found that bags of words is best compared to parts of speech and function words. The best lyric feature was compared to a leading audio feature extraction system. Marsyas tool was used for the process of audio feature extraction. Lyrical features considered alone can classify with high accuracy than the audio features in the categories where the samples are more sparse. It was observed that there is no significant difference between stemming or not stemming representations as part of pre-processing and also combining audio and lyrics features does not improves the classification performance for all mood categories. An accuracy of 70% was achieved with the proposed method.

2.2.4 Analysis Text, Audio Features and Visual clues

Soujanya et al [9] proposed a method to fuse audio, textual and visual clues to detect sentiment from multimedia content. A Youtube dataset had been considered that had 47 videos, of which 20 videos had female speakers and the other 27 had male speakers. All the speakers communicated only through English language. The videos were pre-processed to trim to first 30 seconds. Using MATLAB, all the videos had been converted into image frames, from

which they extracted all features. An openEAR [10] to extract audio features such as MFCC, spectral centroid, beat histogram, spectral flux, pitch etc. On experimentation it has been found that the proposed technique outperformed the existing works. The time complexity taken by the proposed method should be addressed. Gaze and smile features could also be incorporated for improving classification accuracy. Table.3.4 lists out features and methods for song analysis using textual, audio and visual clues.

Table.2.4. Features and Methods for Songs Analysis using Textual, Audio and Visual Clues

Author	Method/Tool	Features	No. of Songs
Soujanya et al[10]	Open EAR	MFCC, spectral centroid, beat histogram, spectral flux, pitch	47

2.3 Tools used for Audio Extraction

There are a large number of audio feature extraction toolboxes available, delivered to the community in differing formats, but usually as at least one of the following formats:

- Stand alone applications
- Plug-ins for a host application
- Software function library

There are so many toolboxes available, but the tools that are most popularly used and those are compatible with programming languages are summarized in this paper. Table.4 briefly describes audio feature extraction toolboxes.

Table.2.5. Toolboxes and their Description.

Tool Name	Brief Description of Toolboxes
Aubio [11]	A high level feature extraction library that extracts features such as onset detection, beat tracking, tempo and melody.
Essentia [12]	Full function workflow environment for high and low level features. Facilitating audio input, pre-processing and statistical analysis of output. Written in C++, with python binding and export data in YAML or JSON format.
jAudio [20]	Java based stand alone application with Graphic User Interface (GUI) and CLI. Designed for batch processing to output in XML format or ARFF for loading into WEKA.
Librosa [13]	API used for feature extraction, for the processing data in Python.
LibXtract [14]	Low level feature extraction tool written with the aim of efficient real time feature extraction, originally in C but now ported to Max-MSP, Pure Data, Super Collider and Vamp formats.
Marsyas [15]	Full real time audio processing standalone framework for dataflow audio processing with GUI and CLI. This programme includes a low level feature extraction tool built in C++, with ability to perform machine learning and synthesis within the framework. The feature extraction aspects have also been translated to Vamp plugin format.
Meyda [16]	Web Audio API based low level feature extraction tool, written in Javascript, Designed for web browser based efficient real time processing.
MIR Toolbox [17]	Audio processing API for offline extraction of high and low level audio features in Matlab. Includes pre-processing, classification and clustering functionality along with audio similarity and distance metrics as part of the toolbox functionality.
Timbre Toolbox [18]	A Matlab toolbox for offline high and low level feature extraction. A toolbox that provides different set of features to the MIR toolbox, specifically made efficient for identifying timbre and to fulfil the Cuidado standards.
YAAFE [19]	Low level feature extraction library designed for computational efficiency and batch processing by utilising data flow graphs, written in C++ with a CLI and bindings for Python and Matlab.

3. CONCLUSION

Music plays a great role in the human life. Music can analysing the mood and changing the mood of the person . The Interest in the field of songs classification and analysis has been growing rapidly. Based on the studies performed, it has been found that either lyrical features or audio features when considered alone doesn't classify with high accuracy. But, when lyrical features are used as a supplement to audio features, performance could be enhanced. So, a framework was developed and was implemented by considering both lyrics and audio features. In this paper the procedure for obtaining the result will be calculated using the models and it will be analyzed into several formats. The proposed work will be under the implementation work and it will be playing the better role than the existing features and model. This model will build the architecture for the basic system and it will be providing the better accuracy ad their result. Lyrics of the song are preprocessed and analysed. Those Lyrics are now in Text

format Then these lyrics are trained using Word2vec then those trained text files are converted into the Binary file. Those Binary file are now analysed using the Features of Wordcount, Distance, wordphase. Audio of the song will be preprocessed using Weka the those files are collected. Then those files taken in the dataset along with the Lyrics (vector file) . Then the Logistic regression is used where those Audio and Lyrics files are taken and analysed with the provided features. By doing the accuracy on the classification of the song is obtained and thus the better classification on the mood of the song will be maintained.

4. REFERENCES

- [1] Baixi Xing,Kejun Zhang, Emotion-driven chinese folk Music image retrival based on DE-SVM,2014.
- [2] XIA, Y., Wang, L., WONG, K.-F., & Xu, M. (2008). Lyric-based Song Sentiment Classification with Sentiment Vector Space Model. *International Journal of Computer Processing of Languages*, 21(4),309330.<https://doi.org/10.1142/S1793840608001950>
- [3] Oudenne, A. M., & Chasins, S. E. (2015). Identifying the Emotional Polarity of Song Lyrics through Natural Language Processing. *Cpsc65*, 1–14.
- [4] Hu, X. (2010). Improving mood classification in music digital libraries by combining lyrics and audio. *The 10th Annual Joint Conference on Digital Libraries*,159–168. <https://doi.org/10.1145/1816123.1816146>
- [5] Corona, H. (2015). An Exploration of Mood Classification in the Million Songs Dataset, (August).
- [6] Shanmugapriya, K., & Srinivasan, B. (2015). An Efficient Method for Determining Sentiment from Song Lyrics Based On WordNet Representation Using HMM. *Ijirce.Com*, 1139–1145. Retrieved from http://www.ijirce.com/upload/2015/february/96_AN.pdf
- [7] Ascalon, E. I. V., & Cabredo, R. (2015). Lyric-Based Music Mood Recognition, 3, 1–8.
- [8] Adit J., Jessica A., Karishma K. and Rahul D., (2015), Emotion Analysis of International Journal of AI & Applications(IJAIA), Vol.6, No.3. Songs based on Lyrical and AudioFeatures,
- [9] Poria, S., Cambria, E., Howard, N., Huang, G. Bin, & Hussain, A. (2016). Fusing audio, visual and textual clues for sentiment analysis from multimodal content. *Neurocomputing*, 174, 50–59. <https://doi.org/10.1016/j.neucom.2015.01.095>
- [10] Abburi, H., Sai, E., Akkireddy, A., Gangashetty, S. V, & Mamidi, R. (2016).Multimodal Sentiment Analysis of Telugu Songs, (Saaip), 48–52.
- [11] Han, K. M., Zin, T., & Tun, H. M. (2016). Extraction Of Audio Features For Emotion Recognition System Based On Music, 5(6), 53– 56.
- [12] P. M. Brossier, “The aubio library at MIREX 2006,” MIREX 2006, p. 1, 2006.
- [14] D. Bogdanov, N. Wack, E. Gómez, S. Gulati, P. Herrera, O. Mayor, G. Roma, J. Salamon, J. R. Zapata, and X. Serra, “Essentia: An audio analysis library for music information retrieval,” in *ISMIR*, 2013, pp. 493–498.
- [15] B. McFee, M. McVicar, C. Raffel, D. Liang, and D. Repetto, “librosa: v0.3.1,” Nov. 2014. J. Bullock and U. Conservatoire, “Libxtract: A lightweight library for audio feature extraction,” in *Proceedings of the International Computer Music Conference, 2007*, vol. 43.
- [16] G. Tzanetakis and P. Cook, “Marsyas: A framework for audio analysis,” *Organised sound*, vol. 4, no. 03, pp. 169–175, 2000.
- [17] H. Rawlinson, N. Segal, and J. Fiala, “Meyda: an audio feature extraction library for the web audio api,” in *Web Audio Conference. Web Audio Conference*, 2015.
- [18] O. Lartillot and P. Toivianen, “A matlab toolbox for musical feature extraction from audio,” in *International Conference on Digital Audio Effects*, 2007, pp. 237–244.
- [19] G. Peeters, B. L. Giordano, P. Susini, N. Misdariis, and S. McAdams, “The timbre toolbox: Extracting audio descriptors from musical signals,” *The Journal of the Acoustical Society of America*, vol. 130, no. 5, pp. 2902–2916, 2011.
- [20] G. Peeters, B. L. Giordano, P. Susini, N. Misdariis, and S. McAdams, “The timbre toolbox: Extracting audio descriptors from musical signals,” *The Journal of the Acoustical Society of America*, vol. 130, no. 5, pp. 2902–2916, 2011.
- [21] C.McKay,I.Fujinaga, and P.Depalle, “jAudio:A feature extraction library,” in *Proceedings of the International Conference on Music Information Retrieval*, 2005, pp. 600–3.
- [22] [21] T. Nasukawa, “Sentiment Analysis” : Capturing Favourability Using Expressions,”pp.70-77,2003. *Natural Language Processing Definition of Sentiment*
- [23] K. Dave, I. Way, S. Lawrence, and D.M. Pennock, “Mining the peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews,” 2003
- [24] Souraya E., Neamat E.G, and Moustafa M. , “Sentiment Analysis of Call Centre Audio Conversations using Text Classification”, *International Journal of Computer Information Systems and Industrial Management Applications* ,ISSN 2150-7988 Volume 4 (2012) pp. 619 -627.
- [25] Downie, J. S., & Ehmann, A. F. (2009). Lyric text mining in music mood classification, (*Ismir*), 411–416.
- [26] Patra, B. G. (2013). Automatic Music Mood Classification of Hindi Songs, (Saaip), 24–28.
- [27] Sharma, G., & Murty, M. N. (2011). Mining Sentiments from Songs Using, 328–339.
- [28] Walsten, D., & Orth, D. (n.d.). Song Genre Classification through Quantitative Analysis of Lyrics.
- [29] Vijayakumar, C., & Adaickalam, V. (2016). Smart Stress Relieving Music Player using Intelligent Sentiment Analysis 20–23.
- [30] Harden, J., Marketing, O. S. U., Certification, A., & Candidate, M. B. A. (2016). Lyric Complexity and Song Popularity : Analysis of Lyric Composition and Relation among Billboard Top 100 Songs Yang Gao , *Management Information Systems Graduate Student* , Oklahoma State University Using SAS Enterprise Miner , SAS Certified Base Programmer , SAS Certified Business Analyst , 1–15.
- [31] Iles, C. L. O. F. A. U. F. (2014). A N E FFICIENT F EATURE SELECTION I N, 29–38.
- [32] Cory McKay, John Ashley Burgoyne, Jason Hockman, Jordan B. L. Smith, Gabriel Vigliensoni, and Ichiro Fujinaga. 2010. Evaluating the genre classification performance of lyrical features relative to audio, symbolic, and cultural features. In *Proceedings of the 11th International Society for Music Information Retrieval Conference*, pages 213–218.
- [33] Z. Dong and Q. Dong. *HowNet and the Computation of Meaning*. World Scientific Publishing. 2006.
- [34] Y. Song, S. Dixon, and M. Pearce, “A Survey of Music Recommendation Systems and Future Perspectives,” *9th International Symposium on Computer Music Modelling and Retrieval (CMMR 2012)*, pp. 19–22, 2012.
- [35] K. Sparck Jones, S. Walker, and S. Robertson, “A Probabilistic Model of Onformation Retrieval: development and comparative experiments,”*Information Processing & Management*, vol. 36, pp. 809–840, 2000.(2011) [Online]Available:<http://www.mathworks.com/matlabcentral/fileexchange/24583mirttoolbox/>

-
- [36] Bradley, M. M. and Lang, P. J. 1999. Affective Norms for English Words (ANEW): Stimuli, Instruction Manual and Affective Ratings. Technical report C-1. University of Florida.
- [36] Kharde, V.A. (2016). Sentiment Analysis of Twitter Data : A Survey of Techniques. 139(11), 5-15.
- [37] Sameli souli, Audio sounds classification using scattering features and support vectors machines for mechanical surveillance,2017
- [38] [38] Paul Ruvolo,A learning approach to hierarchical feature selection and aggregation for audio classification,2010.
- [39] Menqyu Qiau, MP3 audio steganalysis,2012.
- [40] Antonio M.Rinaldi, A multimedia ontology model based on linguistic properties and audio-visual features,2014.
- [41] Wim De Mulder.A survey on the application of recurrent neural networks to statistical language modeling , 2014.
- [42] Chu Guan, Efficient karaoke song recommendation via multiple kernel learning approximation,2017.
- [43] Sascha Fruhholz, The sound of emotions-Towards a unifying neural network perspective of affective sound processing, 2016.
- [44] Y.M.G Costa, Music genre classification using LBP textural features,2012.
- [45] Dongge Li,Classification of general audio data for content based retrieval,2001.
- [46] Zhouyu Fu, Music classification via the bag of features approach,2011.
- [47] Prafulla Kalapatatapu, A study on feature selection and classification techniques of Indian music , EUSPN,2016.
- [48] Arianna Mencattini, Speech emotion recognition using amplitude modulation parameters and a combined feature selection procedure, 2014.
- [49] Moataz El,Ayadi,Survey on speech emotion recognition: Features, classification schemes and databases,2010.
- [50] Debora C.Correa, A survey on symbolic data based music genre classification,2016.
- [51] Sibylle Moser, Media modes of poetic reception reading lyrics versus support vectors for medical surveillance,2017.
- [52] Yunqing Xia,Sentiment Vector Space Model for Lyric-based Song Sentiment Classification,2008.
- [53] K.P Shanmugapriya, Dr.B.Srinivasan, An Efficient Method for Determining Sentiment from Song Lyrics Based On WordNet Representation Using HMM, 2015.
- [54] Humberto Corona, An Exploration of Mood Classification in the Million Songs Dataset, 2015.
- [55] Ashley M. Oudenne, Identifying the Emotional Polarity of Song Lyrics through Natural Language Processing, 2012
- [56] Adit Jamdar, Emotion analysis of songs based on lyrical and audio features,2015.
- [57] Kee Moe Han, Extraction Of Audio Features For Emotion Recognition System Based On Music, 2016.
- [58] Xiao Hu, Improving Mood Classification in Music Digital Libraries by Combining Lyrics and Audio, 2010.
- [59] Souraya Ezzat, Sentiment Analysis of Call Centre Audio Conversations using Text Classification, 2012.
- [60] Hanjie Shu, Opinion Mining for Song Lyrics, 2010.
- [61] Soujanya Poria, Fusing audio, visual and textual clues for sentiment analysis from multimodal content, 2015.