Review on Sentiment Analysis on Music

Stuti Shukla[1], Pooja Khanna[2], Krishna Kant Agrawal[3]

[1,2,3]Computer Science and Engineering, ASET, Amity University, Lucknow, India
[1]s.stuti990@gmail.com, [2]inkhanna@gmail.com, [3]kkagrawal@outlook.com

Abstract: In this study, we present a review on how lyrics can be prove its usefulness in mood classification utilizing the features like linguistic lyric feature set and text stylistic feature set. From this study, one may be able to understand the concept and process of sentiment analysis on the basis of lyrics. Some psychological tools like Big five inventory and CAC scale need to be studied to understand which personality type may generate what perception to a particular song. Big five inventory gives the measure of personality type and CAC scale gives the measure of perception. Performance of lyric feature sets was measured individually as well as by combining them. Fusion methods were used to measure performance of combined feature sets. Basically, two fusion methods used were Feature concatenation and Late fusion. Additionally, examination of learning curves was done which proves that the less number of training samples were required for lyrics+audio system. The discoveries have shown the best class in lyrics sentiment analysis and musical mood classification.

Keywords: **Mood classification, sentiment analysis, unigram bigram trigram, fusion method, cac scale, Basic lyric features, Linguistic features, Text stylistic features, feature concatenation, social tag**

I. INTRODUCTION

Human beings have ability to think, understand, make decisions and give their view points on mostly everything. Having this reasoning and decision making ability, the sentiments of humans are effected by almost everything, maybe positively or negatively. The study of these sentiments of human beings is handled under the field of machine learning. This makes utilization of content examination, computational phonetics, normal dialect handling and biometrics to deliberately and effectively distinguish, evaluate, concentrate and study emotional states and subjective data.

As a rule term, the point of sentiment analysis is to distinguish the state of mind of speaker, essayist, or whatever other subject as for some theme or finish logical extremity and passionate response to a record, association or occasion.

Sentiment analysis may be done on various subjects like, movie reviews, twitter (tweets), any text document, comments, public statements or music. In this document we are going to discuss various machine learning concepts used in study of sentiment analysis of a musical lyrics.

Sentiment analysis is a very active field of research. It also includes the study of identification of sentiments expressed towards a product in general or specific product characteristics with the objective of discovering likes and dislikes of the users.

Question may arise that 'Why do people listen to music and why do they enjoy it?'. The answer for this is very simple that 'Music easily induces strong emotions'. Music expresses the feelings of artists and the type of music or genres preferred by the listener expresses the mood of listeners. So, here the reason is clear for the passion of music among people and especially among youth.

Many researchers have done various surveys on young adults (age ranging between 18 to 30 years) and the result had come out that music affects the thought, feelings and action of the person. Psychologists use various tools like Big five inventory and CAC scale to measure the effect of music of different genres on different personality traits [1][2].

The aim is to develop methods for automatically classifying music by mood. Everyone feels the effect of music. It is seen that young adults are more exposed to music by Micheal Jackson, Pink Floyd, Eminem, Rihanna, Taylor Swift, Shakira as well as Bollywood numbers along with Sufi songs and romantic songs. Listening to music sometimes make some people leave their seat and start dancing, on the other hand some music may even make some people cry out their heart. Some music pieces makes a person happy, gushy and think about their romantic partners. Some music pieces provide feeling of calmness and peace.

Not everyone takes the music same way. For some people a song may generate happy feeling whereas for some it may evoke negative feelings. The lyrics of the song have the ability to influence the behavior of the person. By peeping in the history, we can see that at times music was composed to inspire and motivate people to do some specific action like dance, sing, march or fight.

For classification of music various machine learning algorithms are being used. These include supervised learning approaches like, k nearest neighbor, support vector machines(SVMs), logistic regression, decision trees and Gaussian Mixture Models (GMMs) [3].

II. BACKGROUND STUDY

A. *Mood on the basis of Personality traits and Individual Perceptions*

One of the authors have proposed a review to see if songs (with lyrics) of different classifications had any impact on the contemplations (psychological), sentiments (full of feeling) and activity (conative) [1]. The impacts observed on the individuals were classified as cognitive, affective and conative, respectively. This review additionally inspected the identity figure that was more connected with considerations, sentiments and activities inclinations created through songs of different genres. This study was done by conducting a survey over 60 young adults (30 boys and 30 girls) pursuing graduation and post graduation in Amity University Lucknow campus. For this purpose tools used by the researchers were Big five inventory (by John & Srivastava) and CAC scale (developed by authors of this International Journal). The Big five inventory is a 44 item inventory which classifies or measures the personality of an individual into the Big five factors. These five factors or dimensions are named as, Openness to experience, Conscientiousness, Extraversion, Agreeableness and Neuroticism of personality. The measure of reliability of this inventory is 0.75 and its validity is convergent and quite high. Using Big Five Inventory measure of personality is done while using CAC scale measure of perception is done. CAC scale was developed to measure the perception of listeners whether cognitive, affective or conative. The correlation of these sample readings of personality and perception was done and the result was eventually evident that the different personality traits were affected maximum by different genres of music. It was also seen that romantic songs were found most significant genre which evoked maximum thoughts, feelings and action tendencies. 'Extraversion' personalities evoked more thoughts by Item songs as this genre has fast rhythm and pace so the energetic and enthusiastic enjoyed this more. Thoughts and feelings evoked by romantic songs were positively correlated with 'Neuroticism' whereas negatively correlated to 'Conscientiousness'. Along with this 'Conscientiousness' had negatively correlated thoughts and actions for Rap/Outrageous songs whereas 'Openness and Extraversion' had positive correlation with the same. 'Agreeableness' had feelings negatively correlated to the value based songs. Thus, this is evident from this study that music surely has its affect on people but the effect and perception generated on different personality factors is different for different music genres.

Mood on the basis of Music classification

1) *Considering only the lyrics:* A very interesting work was done on music mood classification where the authors had done the classification of music mood considering the single source that is lyrics [4]. The database contains 1000 songs which have all lyrics in English. The categorical approach is used to categorize the mood and 4 categories are considered that are happy, sad, angry and relaxed along with their complementary categories that are not happy, not sad, not angry and not relaxed. For mood classification various supervised learning approaches have been used from which SVM gave the best and most accurate results, thus, SVM is considered as the best classification model for such purposes. The authors have used traditional bag-of-words features and as result the combination of unigram, bigram and trigram is obtained indicating high order bag-of-words which is useful for classification. The authors have also tried to analyze the mood tags. Social tags help to predict the mood of the user. Music tells what the artists want to say but popular music tells what people want to hear [6]. 'Semantic Mood Space' is created using huge number of tags representing the relationships among the mood tags. Experiments were done so that automatically mood is suggested on the basis of given social tags associated with the songs.

2) *Considering lyrics+audio:* The authors (Xiao Hu and J. Stephen Downie) have tried to improve mood classification in music digital libraries (MDL) by combining lyrics and audio [5]. The verses are analyzed on the premise of phonetic assets (linguistic resources) and content elaborate (text stylistic) elements. In this paper, two combination techniques are utilized for joining the best verses highlights with elements of music sound. The experiments were done using hybrid systems (lyric + audio) and it was evident from the learning curves that the hybrid systems needed comparatively less training samples for classification accuracies than the single audio or lyrics systems. The authors have included a large dataset of 5296 songs with 18 mood classification based on social tags. Three types of lyric features were used, first is basic lyrics feature, second is linguistic feature and third one is text stylistic features.

- *Basic Lyric Feature:* It includes bag-of-words elements of four distinct sorts: Content words without stemming (known as Content), Content words with stemming (cont-stem), Part-of-speech (POS) and Function words (FW). Stemming implies consolidating words with same roots, POS are nouns, verbs, pronouns, etc. and FW are the words that are not Content words, for example, in the sentence "She hates me", "She" and "me" are the function words. FW are also known as 'stopwords'. In this research paper, the blend of unigrams and bigrams and unigrams, bigrams and trigrams is done to see the impact of dynamically extended capabilities [4]. Thus, the results assessed accordingly are represented in Table 1.

TABLE I: Basic Lyric Features

Feature Type	n-gram	No. of dimensions
Content words without	Unigram	7227
	Bigram	34133

Feature Type	n-gram	No. of dimensions
stemming (Content)	Trigram	42795
	Unigram+bigram	41360
	Unigram+bigram+trigram	84155
Content word with stemming (Cont-stem)	Unigram	6098
	Bigram	33008
	Trigram	42707
	Unigram+bigram	39106
	Unigram+bigram+trigam	81813
Part of speech (POS)	Unigram	36
	Bigram	1057
	Trigram	8474
	Unigram+bigram	1093
	Unigram+bigram+trigram	9567
Function words (FW)	Unigram	467
	Bigram	6474
	Trigram	8289
	Unigram+bigram	6941
	Unigram+bigram+trigram	15230

Linguistic lyric Feature: These are described on the basis of General Inquirer (GI) and ANEW (Affective Norms for English Words) and WordNet. General Inquirer is basically a vocabulary that contains some interesting words. These words are in English and 8, 315 in number that are being categorized in 182 psychological classes [5][7]. The feeling generated by each word is manually named with at least one psychological class. As for example 'happiness' is the word for 'emotion', 'pleasure', 'positive', etc. Another specific English vocabulary is ANEW which comprises 1, 034 words with scores in 3 measurements: *valence, arousal* and *dominance* [8]. Valence ranges from lovely to repulsive, arousal or excitement is a scale from quiet to energetic and dominance ranges from compliant to overwhelming (or we can say submissive to dominant). The scoring of these dimensions is done between the range of 1 to 9. Along with these scores in varied three dimensions ANEW also calculates standard deviation of these scores. Since the number of words included by ANEW are few to cover all songs, so wordlist was expanded using WordNet [9]. It's a large lexical database that contains *synset* which means word senses are the synonyms in one *synset* from linguistic view point. WordNet-affect is extension of WordNet which contains 1, 586 one of a kind words in its most recent adaptation [9][10]. Presently the extended ANEW with affect vocabulary contains 7, 756 one of a kind words. This arrangement of words is utilized to construct new bag-of-words elements. The component sort depicted is known as Affect_lex.

Text Stylistic Feature: It consists of addition words (like "oohh", "ahh", etc.), some punctuations (like "!", "?", etc.) and content details (no. of words, average wordLength, etc.). Some other features are, RepeatWordRatio (Number of words – Number of unique words/Number of words), blankLineRatio (Number of blankLines/ Number of lines), avgLineLength (Number of words/ Number of lines), stdLineLength (standard deviation of number of words per line), uniqueWordsPerLine (Number of unique words/ Number of lines), Number of words per minute (Number of words/ song length in minutes) and Number of lines per minute (Number of lines/ song's length in minutes).

The dataset defined by Xiao Hu and J. Stephen Downie [5] in their paper consists of various mood categories like, calm, anxious, exciting, dreamy, hopeful, sad, romantic, aggressive, confident, earnest, glad, gleeful, mournful, gloomy, angry, cheerful, brooding and cynical. The genre distribution of songs in dataset consist of rock, electronic, R&B, jazz, reggae, oldies, metal, new age, blues, country, and hip-hop with 5296 as total number of songs. The lyric features were evaluated and the result was obtained in favor of basic lyric feature types.

TABLE II: Performances of Lyric Feature Types

Feature Types	Accuracy
Content	0.617
Content_stem	0.613
Affect_lex	0.594
FW	0.594
GI	0.586
POS	0.579
ANEW	0.545
TextStyle	0.529

Further, for these feature types, best performance was of Content which has bag_of_words features with multi-order n-grams. These best individual lyric feature types were concatenated with each other and hence the large list of combined feature types were obtained, which contained about 255 features. Since this value range was huge so the features were normalized and represented between [0, 1] before combining.

TABLE III: Performances of Combined Lyric Feature Types

Feature Types	Accuracy
Content+GI+ANEW+Affect_lex+FW+TextStyle	0.638
ANEW+TextStyle	0.637
Content+GI+FW+Affect_lex+TextStyle	0.636
Content+GI+FW+TextStyle	0.635

Best performer is Content+GI+Affect-lex+FW+TextStyle which gained 2.1% higher precision even than the best individual one (0.638 vs 0.617). The most interesting part is to see that worst performers in individual types performed good and were second most suitable types at the point when joined with each other. Also ANEW and TextStyle are the fundamental two component sorts that don't comply with the bag_of_words structure.

Since Content+GI+Affect-lex+FW+TextStyle was found to be the best lyric feature, it was combined with the audio based system to get the combined effect of lyrics as well as music on mood of an individual. Fusion or Combination strategies are utilized to join the heterogeneous information sources to enhance order execution and they are at their best when the elements are adequately assorted. Basically, there are two most efficient fusion methods: first is "feature concatenation" and the other is "late fusion". In feature concatenation, two capabilities are connected and grouping calculation deals with consolidated component vectors. In late fusion, the outputs of the individual classifiers are combined either by averaging or multiplying. Here, weighted averaging estimation technique is utilized and for each testing case, last estimation likelihood is ascertained as:

$$P_{hybrid} = \alpha p_{lyrics} + (1-\alpha) p_{audio} \qquad (1)$$

Here, α is the weight provided to lyrics posterior probability. At the point where the hybrid posterior likelihood is greater than or equivalent to 0.5, then the song is classified as "positive". α ranges from 0.1 to 0.9 with the increment step of 0.1.

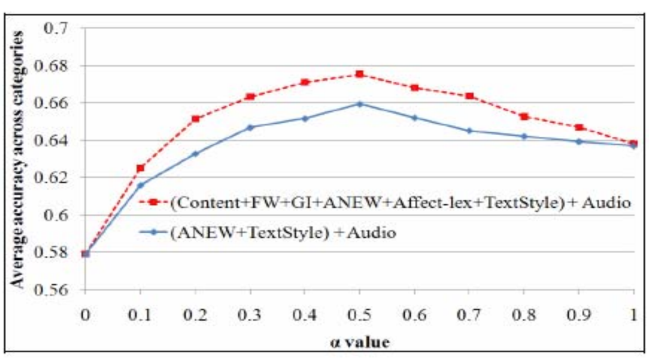

Fig. 1. Impact of α incentive in late fusion on averaged precision

The above figure shows that the most astounding exactness was accomplished when "α=0.5" for both cases. Further, the results of single source and hybrid systems were evaluated and conclusion was made that the feature concatenation was not a decent alternative for joining ANEW+TextStyle features set and sound. Late fusion is useful for both capabilities yet lyric feature combination that performed best is ANEW+TextStyle in consolidating with sound (0.675 versus 0.659) with a measurably irrelevant contrast ($p < 0.05$).

Investigation of learning curves associated with single source system and late fusion system was done to find that if lyrics can help in reducing the sample training data. The figure 2 shows the result. It is evident that with more training data all system performances were increased except the sound based system which expanded gradually. This validates that by combining lyrics and audio training samples can be reduced required to achieve certain classification performance level.

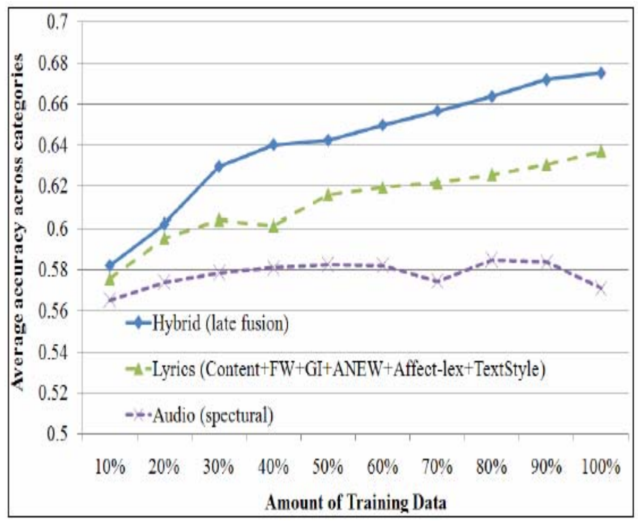

Fig. 2 Learning curves of hybrid and single source system[5]

III. CONCLUSION AND FUTURE SCOPE

As we have seen in the best combined lyric features, ANEW and TextStyle are the only type which are combination of two feature types rest all are the combination of atleast four featuretypes, hence they have high dimensionality. ANEW+TextStyle have just 37 dimensions which is more proficient when contrasted with others. Along with this, having high dimensionality gives a space to feature selection and reduction. So this investigation of feature selection and reduction for consolidated lyric feature with high dimensionality can be done in our future work.

Many researchers have used classification models like SVM. As for upcoming work, the association of features and classifiers is valueable for further examination. Utilizing different classification models, other than SVM (e.g., Naïve Bayes), the top-positioned components may be not quite the same as those chosen by SVM. With legitimate component choice techniques, other grouping models may outflank SVM.

In our future work, we may review different methods for classification of musical mood on the basis of genres, social tags, audio and lyrics and find an efficient fusion method combining all these feature sets, providing wide range hybrid system. These experiments may help to improve the mood classification by providing more precise and accurate results with large range of input dimensions.

REFERENCES

[1] Ms Aheli Chakraborty, Dr. Durgesh K. Upadhyay, Prof. (Dr.) Manju Agrawal: Exploring the Effect of Songs of Various Genres on Young Adults' Thoughts, Feelings and Actions: The International Journal of Indian Psychology.

[2] John, O.P., & Srivastava, S. The big-five trait taxonomy: history, measurement, and theoretical perspectives. In L.A. Pervin & O.P. John (Eds.), Handbook of Personality: Theory and Research (2nd ed.). New York: Guilford Press (1999).

[3] F.O. Isinkaye, Y.O. Folajimi, B.A. Ojokoh: Recommendation systems: Principles, methods and Evaluation: Cairo University, Egyptian Informatics Journal.

[4] He, H., Jin, J., Xiong, Y., Chen, B., Sun, W., and Zhao, L. Language feature mining for music emotion classification via supervised learning from lyrics. In Proceedings of Advances in the 3rd International Symposium on Computation and Intelligence (ISICA 2008).

[5] Xiao Hu, J. Stephen Downie: Improving Mood Classification in Music Digital Libraries by Combining Lyrics and Audio: JCDL 2011.

[6] Kevin Buhrer, Micheal Johns, Scott Stephens: Modern music sentiment analysis: Team Lyrics Lab: http: //michaeljohns.github.io/lyrics-lab

[7] Bradley, M. M. and Lang, P. J. Affective Norms for English Words (ANEW): Stimuli, Instruction Manual and Affective Ratings. Technical report C-1. University of Florida 1999.

[8] Fellbaum, C. WordNet: An Electronic Lexical Database, MIT Press 1998.

[9] http: //sentiwordnet.isti.cnr.it/, last accessed 2017/06/21.