Journal of Next Generation Technology (ISSN: 2583-021X)
Vol. 2, Issue 1, May 2022

Emotion Based Music Classification

Shashikumar G Totad [1], Spoorthi N Patil[2], Geeta R. Bharamagoudar[3]

[1]Professor, School of Computer Science and Engineering, KLE Technological University Hubli,India

[2]School of Computer Science and Engineering, KLE Technological University Hubballi,India

[3]Professor, Department of Computer Science and Engineering, KLE Institute of Technology, Hubblli, India

Email id: totad@kletech.ac.in[1] , spoorth.sp333@gmail.com[2], geetatotad@yahoo.co.in[3]

Article Received: 02 Feb 2022 *Article Accepted: 19 Feb 2022* *Article Published: 10 May 2022*

Citation
Shashikumar G Totad, Spoorthi N Patil, Geeta R. Bharamagoudar (2022), "Emotion Based Music Classification", Journal of Next Generation Technology, 2(1), 22-29.

Abstract

Music gives strong emotions and feelings to human. Enough research work has been carried out which show that a lot of effort and time is contributed to music-based projection and classifying the emotion-based in music. Also a lot of study on music-based research says that human are overriding of music is being changed and providing platforms for a customized playlist for each user along with it moods are present. Emotions are personally tailored and these personalized natured emotions can be detected. It is very challenging task when it comes to music. Previously, music emotion detection relied on the auditory perception of musical characteristics. But in recent years it's observed that music lyrics are also used to detect emotion. To experiment, we used machine learning approach by using support vector machine (SVM) to identify emotion in 1161 lyrics by using 9 emotions categories with single and multiple labels on emotion, and compared with other classifiers that achieved 0.66 and 0.83 on the F1 scale.

Keywords: Machine learning, support vector machine, F1-score.

I. Introduction

The tie-up between music and emotion has always been a curious research filed in music psychology new technology has been beneficial for the evolution of music, leading to the emergence of world music. Which includes music classification, recommendation, and instrumental classification and Physiology and psychology studies have proven that music can activate the brain areas which are integrated with emotions. Listening to music studies says that humans can generate different emotions based on music it is a well-known part of human lives now. Music is a kind of powerful tool than language to improve strong kind of emotions in humans. Humans have different emotions and feelings which cannot be determined by race, culture, nationality. With the growth of the internet and online media platform, there's a thoroughgoing change in the way how we are consuming music now. There are many online music streaming services such as Apple, Spotify, genesis, Last FM, iTunes, etc. These sites have playlist for users and they provide a generous number of playlist based on mood or emotions, which have limited genres, artist, and albums. Such playlist is mostly created by various automatic playlist generation algorithms. Automatic Music information Retrieval (MIR) helps to detect emotion in music. Emotion of human is not discrete or continuous but detecting is challenging task when it is applied to music and one music holds multiple emotions simultaneously.

On another hand, music is hard to determine with some emotions. Music classification has features such as instrumental, country, track audio signals, and metadata. Features are based on the lyrics along with audio features for songs emotion classification.

II. RELATED WORKS

Many research works are done previously by many researchers in the field of music classification with various classification techniques. This supports for the implementation of music classification-based emotion. Lot of research work is carried on the field of music.

Byeong-jun Han et al.[7] reported that emotions analyzed in different music features like pitch, tempo, loudness, rhythm, etc and done mapping with eleven categorizations of emotion like anger, calm, happy, peaceful, relaxed ,sad and sleepy. The work made use of LIBSVM library and analyzed different music features like the categorization on Juslin's theory with Thayer's emotion model used Support Vector Regression (SVR) as classifiers to coach the regression functions to predict arousal and valence values. The work compares SVR-based method with other classification algorithms viz., Gaussian Mixture Model GMM [1][8] and SVM[7][1], to ascertain the performance. The SMERS (SVR based Music Emotion Recognition System) carryout three steps viz., feature extraction, mapping, and training. In extraction step it holds seven distinct musical features. After extraction these features are mapped with eleven emotion categories with Thayer's 2-D emotion model. Thereafter, the system uses extracted features as input data to train SVR. Result of this paper was based on a Cartesian coordinate system with origin (0,0) to the emotion in Thayer coordinate system and other angles these two classifiers predicts songs emotion which results from the polar Cartesian coordinate and polar coordinate with an accuracy of 91.25% and they increased by changing the coordinate values into polar got the accuracy as 94.55% using SVR and GMM achieved 92.73% accuracy (Byeong-jun Han et al,2009)[7].

Mahesh Bargaje et al., reported that music classification is a fundamental process in MIR where raw data is analyzed according to user needs. Users provided music data is classified into several categories like Genre, Mood, Artist, Instrumental by the classification algorithms [1][2]. They classified music into three categories as a top-level label, mid-level label, and low-level label. To identify eight emotions the suggested approach was arousal-valence method from 2D (Arousal-Valence) plane to 3D (Arousal-Valence-Loudness) plane [7]. Genetic algorithm was used for reducing the feature set and computation time by 9.27%.The proposed system outperformed than older one with accuracy 11.49% with low cost and identified more emotions (Mahesh Bargaje et al, 2005)[1].

Deepti Chaudhary et al., extracted 9 features of emotions from 3150 Hindi songs such as pop, classical, jazz, patriotic etc. In this model, 70% of songs are used for training and 30% for test data. Preprocessing is done with two steps windowing and framing with Fourier transformer function. According to Thayer's two-dimensional approach music is split into distinct classes. The MIR toolbox is used for feature extraction to extract emotional features from audio signals. Which are analyzed using SVM and K-NN classification [8][7]. MATLAB is used to create the model, which achieved 72.5% mean accuracy and 71.8% standard deviation accuracy and for K-NN achieves 71.2% mean accuracy and 71.5% standard deviation accuracy[8][7],indicating that SVM outperformed K-NN ,(Deepti Chaudhary et al,2021).

III. Proposed System

This section discusses about system architecture that defines a structure, views and behavior of the proposed system. System components and sub-systems are developed so that, all work together to implement overall system. The components constitute four stages, as illustrated in Figure 1.

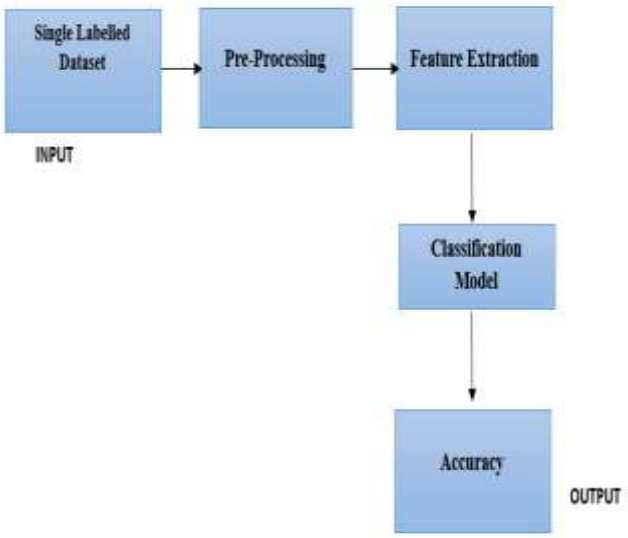

Figure 1: High-Level Design of Emotion-Based Classification

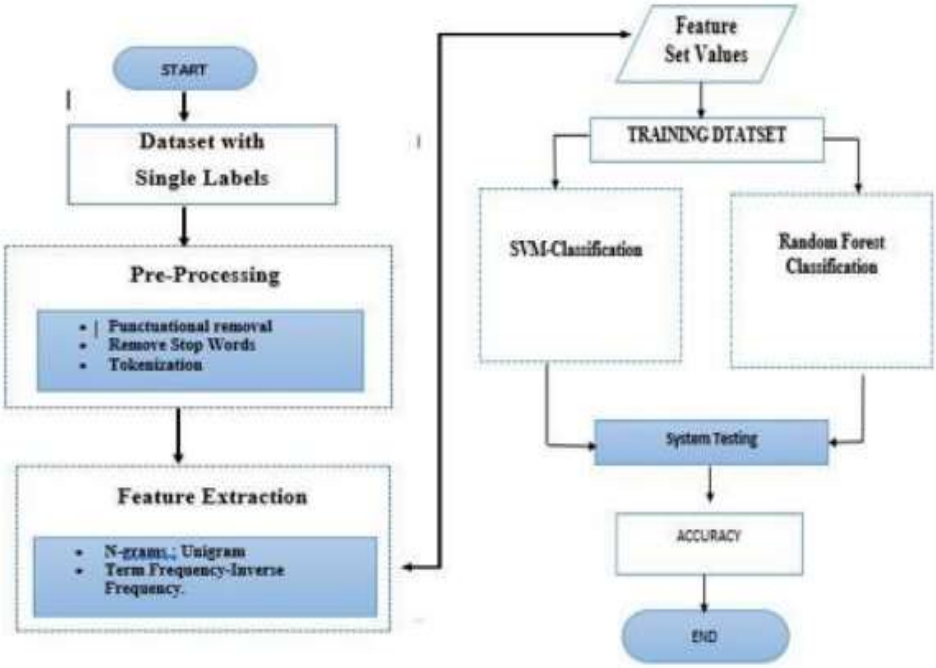

Figure 2: Detailed Level Design of Emotion-Based Classification

Input: The single labeled dataset includes genre, artist, year, title, lyrics, and labels (emotions). Each lyric is assigned an emotion label. Multiple emotion classification dataset contains same attributes of the single label with multiple emotions label such as Sadness Tenderness, Tension, Power and so on.

Pre-Processing : In this stage, text pre-processing is performed to clean the data by removing special characters, punctuation's, and tokenization and stop word's. These carry nonsensical word which cannot be fed as input into the model.

Featuree Extraction: For feature extraction, n-gram is used to predict the next item in the sequence and read the data in contiguous sequence from corpus. Frequency-Inverse Document Frequency (TD-IDF) quantifies the relevant words in corpus. Number of times a term appeared repeatedly and rarely are considered and TD-IDF vector is put up.

Classification Model: In this model linear supervised learning classification is used to fit the data provided by returning the best fit hyper-plane that categorizes our data. Once hyper-plane is obtained, some characteristics can be fed to classifiers to see the predicted class. One Vs. Rest Classifier is utilized for multi-label classification models with SVM[7][1].One Vs. Rest classifier uses a binary classification approach for multi-class classification.

IV. Dataset

The dataset is obtained from Genius.com, eleven hundred sixty-one song lyrics were collected with four genres such as rock, country, rap/hip-hop, and reggae. These lyrics have been hand-annotated that makes data easy and recognizable for machine to train model.These labels are categorized into 9 emotional categories using GEMS using Light Tag tool. It contains two independent datasets: single label and multi-label dataset. By combining both single and multiple labeled emotions yields 7 attributes and 3 emotions in single label dataset and 9 emotions in the multi-label dataset. For single-label dataset each lyric in the dataset is labelled with a single emotion. The key feature of this dataset contains three emotion categories: Sadness, Tension, and Tenderness out of nine emotions. For Multi-labels Dataset each lyric in the dataset is labelled with a multiple-emotion. Interestingly, nine emotions are present, with three to four emotions labelled for each lyric.

Data visualization: Word Cloud is a technique for displaying the keywords that appear most frequently within the given set of data. Size of each word indicates frequency, important textual data or words in lyrics are spotlighted using Word Cloud. These Word clouds are widely used for analysing sadness, tension, tenderness labels in the dataset.

V. Methodology

In this section, modules and the techniques employed are discussed. This section also discusses data-setup and classification performed by the proposed model. As discussed in the previous section two datasets are utilized to evaluate the performance on emotion classification in lyrics. The classification process mainly focuses on binary emotion classification using single label dataset and classifies the dataset involving multi emotions.

Text preprocessing: It is a method to clean the text data and make it ready to feed into the model. Humans have an easy time understanding the language, but machines cannot understand the words. These words need to convert text to binary in an inefficient manner .The dataset contains various

noises in natural language processing like removing stop words, named entity recognition, part of speech tagging, etc.

Punctuation removals: In this module, punctuations are removed using a string library pre-defined list. The string module with a regular expression is replaced with any punctuation in text with an empty string.

Remove Stop Words: Stop words repeatedly occur or commonly utilized words are removed from the corpus because they aren't useful or provide valuable information. These natural language toolkit libraries have a list of words, and a common library is used to remove stop words.

Tokenization: Tokens are the building blocks of natural language. It is a way of separating a piece of text into smaller units called tokens. Tokens can be words, characters or sub worlds.

N-grams: The n-grams are a set of n co-occurring words that read data in a contiguous sequence of n-items for a given sample of text or sentence using the properties of the n-gram. These n-grams collect words corpus to predict the next item in a sequence. It is a probabilistic language model wisely used for sequence analysis, data compression, and natural language processing. An n-gram of size one is called unigram, which takes single words as input to read the word corpus.

Term Frequency -Inverse Document Frequency TF-IDF: It is a statistical measure of how relevant a word is in series or corpus. The number of times the text of a word appears increases the proportionality of the word frequency in the dataset. In a large corpus, some word carries less meaningful information about the actual contents of the document. It is very common to use the TF-IDF transform usage Term Frequency in the dataset. The frequency represents a number of instances in a given word therefore, appears repeatedly in the corpus and rational.

Inverse Document Frequency, IDF decreases the weight of terms that appear very frequently in the entire document set and increase's the weight of terms that appears rarely. TF-IDF has a higher value the more the significance of the word in collection or corpus.

Support Vector Machine (SVM) It is a supervised machine learning method used mainly to solve classification and regression analysis. This approach creates a linear model that assigns labels to one category after training. We plot each data item as a point in n-dimensional space with the value of each feature and perform classification by locating the hyper-plane that separates the two classes. The SVM classifier is a hyper-plane border [7][8] that best separates the two classes. Determines the correct hyper-plane as follows:

- By considering the thumb rule find the proper hyper plane by choosing the hyper-plane that best separates the two classes.
- By increasing the distance between the closest data points of either class and these hyper-planes, choose the best hyper-plane.
- By raising the margin, SVM [7][1] finds the hyper-plane that accurately classifies the classes.

Random Forest Classifier: It is a supervised machine learning technique that functions similarly to a decision tree. It is a strategy that combines multiple classifiers to solve difficult problems. The

random forest algorithm is made up of multiple decision trees that create a forest, and it is extensively bagged [5][4][3]. The random forest method predicts average or mean of the output from different trees and raises the precision outcome based on the decision tree's predictions. As a result, it decreases dataset over-fitting and boosts precision without requiring a large number of packages.

VI. Implementing Emotion-based Music Classification

For the implementation of the emotion-based music classification, the steps are as follows:

1. Initially, import the appropriate libraries to read and alter the data before performing anything with gathered relevant packages.
2. The collected dataset contains several headers that give information on the collected songs, such as the artist, genre, title, etc. The final column, "label", along with the "lyrics" column is used to classify the problem.
3. Data visualization is performed to display word clouds image and saved the word clouds for 'Sadness', 'Tension', 'Tenderness' classes.
4. In this step we perform data exploration, feature extraction, and data processing.
   - By defining the n-gram models using unigram.
   - Set up a TF/IDF Vectorizer to transform our lyrics into features that indicate their belonging to one of the three classes, using various n-gram models.
   - For feature extraction, import packages of n-gram and TF-IDF.
   - Raw data will be converted to a matrix of TD-IDF features with parameters.

5. The model is being built by importing different classification process by dividing into training and testing to form the model. To initiate, model is trained with lyrics and tested with labels to determine the accuracy of the model.
   - The linear support vector classifier (SVC) similarly known as SVM classification has created a pipeline and compared with another classifier like Random Forest classifier [5][6].
   - The values were being fit and transformed by data the accuracy of different classifiers were noted. Then 5-fold cross validation's mean accuracy is taken into consideration to plot the confusion matrix.
6. The multi-label dataset had been annotated more than one label to transform the information into a machine-readable format. For these problem support-vector classifiers is being used for each class and one vs. rest classifiers. The data is been fed into to the classifier via a pipeline to transform tf/idf features vectors.
   - The model will be split into training and testing sets to determine the accuracy of the model.
   - Before fitting classifiers import the f1-score metric to evaluate the efficiency of learning model.

At this final stage average setting of f1_score is set to micro in order to aggregate the contributions of all classes and compute the average metric.

VII. Result

In single-label classification, lyric and label columns are being utilized and separated the dataset into train and test sets of 70 % and 30%, respectively. The lyric is transformed into feature vectors, which

are being fed into a classifier using SVM and another classifier. Random Forest classifiers, Decision Tree, K-Nearest Neighbor classifier are trained models, and the mean and test accuracy of the model are reported [5]. The accuracy of classifiers on Tests set and 5-cross-validation accuracy is recorded. The accuracy figures obtained are shown in Table 1. Support vector machines utilizing linear SVC have outperformed with other classifiers, with the greatest accuracy recorded of 0.66.

The multi-class classification dataset is split similarly as a single class dataset. It performed TF-IDF vectorization with unigram features to transform the words to vectors. The multi-label dataset performs classification on SVM and one vs. rest classifier.

TABLE 1: PERFORMNACE OF EMOTION-BASED CLASSIFICATION MODELS

Classifier	Test Accuracy	5-fold cross validation mean accuracy
SVM	0.66	0.59
Random Forest	0.58	0.52
Decision Tree	0.61	0.56
KNN	0.56	0.53

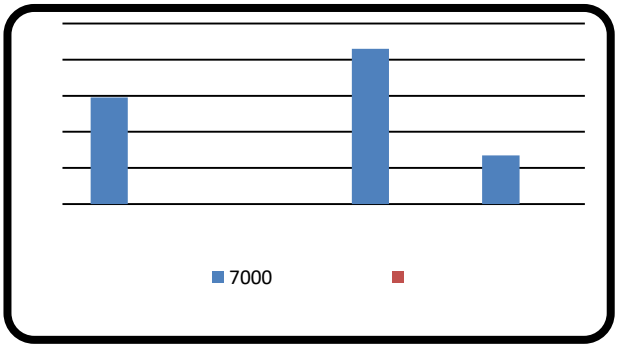

Figure 3: Performance of emotion-based classification Models

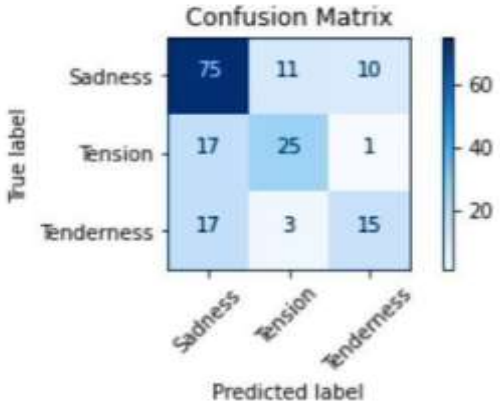

Figure 4: Confusion matrix for SVM classifier.

Through the test set, the classifier has achieved an F1 score of 0.83. This F1 score measures test accuracy and it considers both the precision and the recall to compute the score.

VIII. Conclusion

Emotion-based Music classification is built on the SVM classification model. To classify the emotion expressed in the lyrics and identify the categories of emotions implied in lyrics .It is an interesting field for research. There are many ways to improve the classification to detect the emotion present in the music. Annotation was performed on 1161 song lyrics and categorized into nine emotions. The proposed system performed pre-processing and extracted the features present in lyrics and vectorized using TD-IDF and unigram and fed the data into classifier to detect the emotion of the lyrics. In Single labeled dataset, we used SVM and achieved the highest accuracy as 0.66 and 5-fold cross-validation achieved 0.59 accuracy .The multiple label SVM classifier has achieved a 0.83 F1 accuracy.

References

[1] Mahesh Bargaje, “On Emotion Recognitionand Emotion Based Classification of Audio using Genetic algorithm-AnOptimized approch,”inInternational Conference on Industrial Instrumention and Control(ICIC) College of Engineering Pune,India May 28-30,2015.

[2] Geeta R.B, Gouri C.Khadabadi, Shashikumar G Totad, "A Survey on Identification and Classification of Diseases in crops using Image Processing techniques", Grenze International Journal of Engineering and Technology, Jan Issue, Grenze ID: 01.GIJET.7.1.506, Jan 2021, Vol-7, pp-174-179I.

[3] Amruta Patil, Shashikumar G. Totad "Non-invasive Soya Bean Seed Analysis Using Machine Learning", International Journal of Recent Technology and Engineering,Jan-2019, V0l-7,279-282 ISSN: 2277-3878.

[4] Abhilash Biradar & Shashikumar G Totad, "Detection Depression in Social Media Posts using Machine Learning", International Conference on Recent Trends in Image Processing and Pattern Recognition(RTIP2R-2018) on 21-24 Dec-2018 at Solapur University, Solapur.

[5] Mulimani D., Totad S.G., Patil P., Seeri S.V. (2021) “Adaptive Ensemble Learning with Concept Drift Detection for Intrusion Detection.” 4th International Conference on Intelligent Computing and Communication (ICICC - 2020) 18th,19th & 20th Sept, 2020, Dayanand Sagar University, Bengaluru, In: Data Engineering and Intelligent Computing. vol 1. Pp 331-339 , 2021,Springer, Singapore. https://doi.org/10.1007/978-981-16-0171-2_31

[6] Shashikumar G Totad., Geeta R.B., Satwik Belaldavar. (2019) "Scalability and Speedup Enhancement of SVM Algorithm for Big Data Processing ". AICTE Sponsored International Conference (ICCEEE2020) 24th-26th August 2020 at St. Peter's Engineering College, Hyderabad.

[7] Byeong-jun Han, Seungmin Rho Roger B. Dannenberg Eenjun Hwang “smers: music emotion recognition using support vector regression” 10th International Society for Music Information Retrieval Conference ,ISMIR 2009.

[8] Deepti Chaudhary, Niraj Pratap Singh, and Sachin Singh “Emotion-Based Hindi Music Classification” under exclusive licence to springer nature Singaore,2021.