Journal of Soft Computing Paradigm (ISSN: 2582-2640)
www.irojournals.com/jscp/

Emotion Recognition from Speech using SVM and Random Forest Classifier

A. S. Wincy Pon Annal[1], R. Manonmani[2], C. Booma[3]

[1]Assistant Professor, Electronics and Instrumentation Engineering, Government College of Technology, Coimbatore, Tamil Nadu, India
[2,3]UG student, Electronics and Instrumentation Engineering, Government College of Technology, Coimbatore, Tamil Nadu, India

E-mail: [1]selwincy@gmail.com, [2]manonmanirsm2001@gmail.com, [3]boomasekar421@gmail.com

Abstract

Speech is the most natural way of people to communicate with one another. It is a vital medium for communicating a person's thoughts, feelings, and mental condition to others. The process of identifying the intellectual state is the recognition of basic emotion through speech. In human life, emotions are incredibly significant. In this project, the emotion is recognized from speech using Support Vector Machine (SVM) and Random Forest classifiers. These are supervised machine learning algorithms used for both classification and regression problems. SVM classifies data by creating N-dimensional hyper planes that divide the input into different categories. The classification is accomplished using a linear and non-linear separation surface in the dataset's input feature. Random Forest is a classifier that combines a number of decision trees on different subsets of a dataset and averages the results to increase the dataset's predicted accuracy. These classifiers are used to categorize emotions like happiness, rage, sadness and neutral for a certain incoming voice signal. Here, the system is trained and developed to recognize emotion in real-time speech. The result demonstrates that the Random Forest classifier is significantly better, when compared to the SVM classifier.

Keywords: Emotion Recognition, Support Vector Machine, Random Forest

1. Introduction

Human computer intelligence is a new field of study that tries to teach computers how to learn from human experiences and select how to respond to a given circumstance. As a result, the interface between users and the computer has improved. Computer can be made to recognise numerous qualities contained in the voice sample and derive the emotion

DOI: https://doi.org/10.36548/jscp.2022.1.005
Received: 09.03.2022, received in revised form: 08.04.2022, accepted: 22.04.2022, published: 30.04.2022

underlying them using certain algorithms and techniques. This emotion detection can be done using one of the two information: Speech or Image. Speech is such an important aspect of communication, being able to recognise emotion from it is crucial. There are various methods and classifications created to categorise human emotions based on training datasets, including K-Nearest Neighbor, Support Vector Machines, Decision Tree, Random Forest, Linear Discriminant Analysis and others.

When it comes to recognising emotions, the initial step is to extract features from voice signals. In this project, SVM and Random Forest algorithms are used to train data sets in order to identify emotion or sentiment from extracted features. SVM and Random Forest are supervised machine learning technique that can be used to classify and predict data. SVM tries to categorise data by discovering hyper planes that can separate data with the greatest margin of separation. The fresh values are separated and examined based on the training sets whereas Random Forest is a classifier that uses the average of multiple decision trees on different subsets of the dataset to enhance the dataset's projected accuracy. It's an ensemble method that reduces over-fitting by averaging the results, making it better than a single decision tree. These classifiers classify basic four emotions as happy, angry, sad and neutral. This emotion classification is used to identify human view point or feelings and expressing his/her mental state to others. It is also beneficial to the orators, call centres and human robotic interface.

2. Literature Survey

Emotion recognition using audio signals has been a subject of research in the past. [1] provided an entire overview of voice emotion recognition, including dataset properties and classifier selection. Several acoustic aspects of speech and analyses of CNN was investigated in [2], which is helpful for further research into modern methods of emotion recognition. Mel Frequency Cepstral Coefficients (MFCC) feature was extracted from speech and used SVM classifier [3] to acknowledge the emotion. Perceptral Linear Prediction (PLP) discarded unwanted information from human pitch [4]. Principal Component Analysis (PCA) was proposed in [5] to reduce the feature and employed in SVM and KNN Classification fusion. [6] investigated a way to predict future reactions from emotional speech cues using several sorts of classifiers. [7] used some classification techniques, like Decision Tree and Random Forest, to spot emotion appropriately. Music is used as an input signal [8], and the classification is done using an SVM classifier.

[9] has compared the CNN and Decision Tree for emotion recognition and concluded that the Decision Tree is more accurate when it is used for binary class classification. [10] explained the importance of Random Forest and how it is used in various application and how it can be further implemented. [11] used python code to classify emotion using MLP and KNN algorithm. The reason for using MATLAB [12] is detailed. Linear Predictive Coding was implemented in MATLAB [13] to analyse and understand the voice stress level. Image spectrograms using SVM and KNN algorithms, as demonstrated in [14], can be used to recognise speech emotions. Real time human facial expressions are recognized in [15].

3. Emotion Recognition Methodology

The most natural way for humans to communicate with one another is through speech. Emotions colour the language and act as an essential component of natural two-way human communication and interaction. In emotion recognition, basic aspects of speech signals, such as format, energy, and MFCC (Mel Frequency Cepstral Coefficients), are extracted from both offline and real-time speech and sorted into different emotional classes using classifiers like Support Vector Machine (SVM) and Random Forest(RF) Classifier. The block diagram of the system is shown in Figure 1.

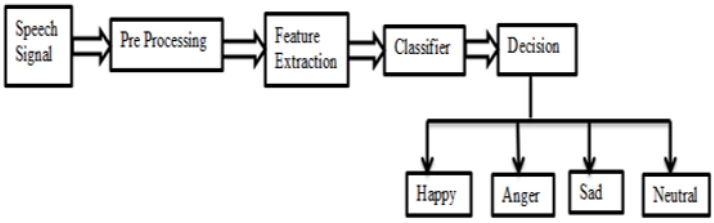

Figure 1. Block diagram of recognizing emotion from speech

The speech signal is initially pre-processed where the input audio is recorded and saved in .wav format audio file. This is the stage where all of the input audio's attributes (sample rate, bits, channel, format, and length) are established in order to get the best possible output from the system. The features of the pre-processed speech signal are extracted. The audio's properties and sample rate are read from the audio file and the Mel-Frequency Cepstral Coefficient(MFCC) is calculated for each sample in the input audio. The classifier receives the extracted feature. It maps the input data to a specific category in the trained data. The classifiers used in this project to recognize emotion are SVM and Random Forest Classifier.

3.1 SVM Classifier

SVM is a supervised machine learning technique that can be used to classify and predict data. The SVM algorithm's goal is to locate a hyper plane in an N-dimensional space that correctly classifies data points. It also contains a marginal plane for removing separation margins. The working of SVM is shown in the Figure 2.

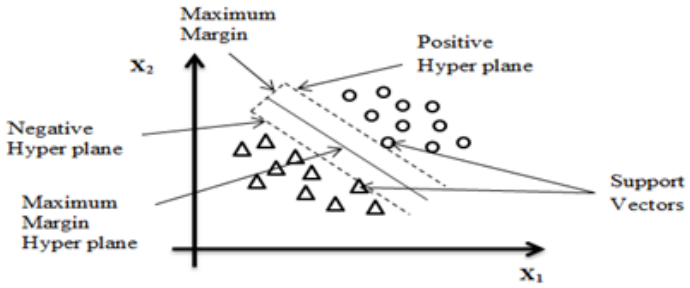

Figure 2. Working of SVM Classifier

In SVM, a hyper plane is a decision border that separates the two classes. Different classes can be assigned to data points that lie on either side of the hyper plane. Then SVM generates a hyper plane and two more hyper planes simultaneously. Support Vectors are the data points or vectors that are closest to the hyper plane and have an effect on the hyper plane's position. The data in the SVM Classifier is of two types. Linear separable and Non-Linear separable data. These classifications are shown in the Figure 3. A single straight line (in 2D) can divide the data points into two classes if they are linearly separable. When data points cannot be split into two classes using a straight line (in 2D), Non-Linear SVM is used to classify them using approaches such as kernel tricks. Kernels converts the data from lower dimension to higher dimensions (or) 2 Dimensions to 3D,4D,5D etc.

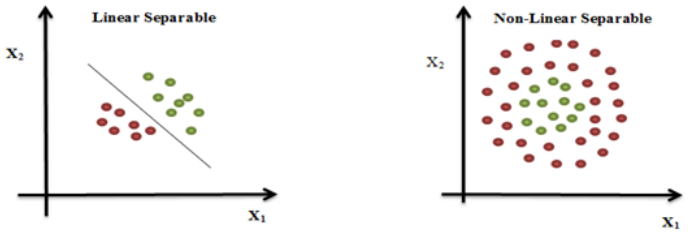

Figure 3. Types of data in SVM

3.2 Random Forest Classifier

It is a supervised machine learning algorithm. It's useful for both classification and regression problems. It is a classifier that uses the average of multiple decision trees on

different subsets of the dataset to enhance the dataset's projected accuracy. Instead of betting on a single decision tree, the random forest collects forecasts from each tree and generates the final output based on the majority of projection votes. The more trees within the forest, the more accurate it becomes, and overfitting is avoided.

It works based on ensemble learning, which is the process of integrating numerous classifiers to solve a complex problem and improve the model's performance. The Ensembling technique is classified as Bagging and Boosting method. Bagging method generates a different training subset with replacement from sample training data, and the final output is determined by majority voting. eg: Random Forest, whereas the Boosting method turns weak learners into strong learners by building successive models with the highest accuracy as the final model. These methods are shown in the Figure 4.

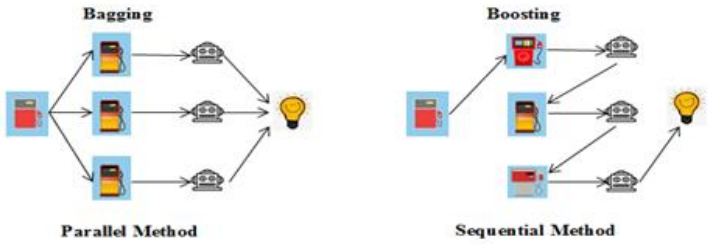

Figure 4. Ensembling Technique

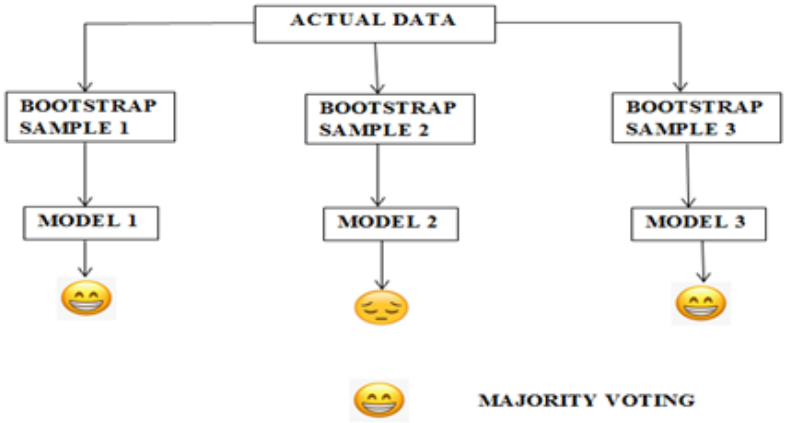

Figure 5. Bagging Ensemble Method

Bagging, commonly called Bootstrap Aggregation, is a random forest ensemble approach. Bagging selects a random sample of information from the whole set. As a result, each model is made using row sampling to switch the samples (Bootstrap Samples) provided by the first Data. Row sampling with replacement is referred known as the bootstrap stage.

The findings are generated after each model is trained independently. After merging the findings of all models, the ultimate outcome is predicated on majority voting. Figure 5 shows the Bagging Ensemble method for Emotion recognition from speech.

4. Results and Discussion

Speech emotion identification performance is influenced by a number of factors, including the quality of noise-free speech, features derived from speech signals, and classification algorithms employed. The process of obtaining a signal from speech is shown in the Figure 6. and the speech signal is represented graphically as shown in the Figure 7.

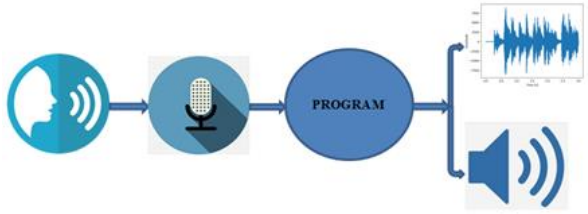

Figure 6. Block diagram to obtain a speech signal

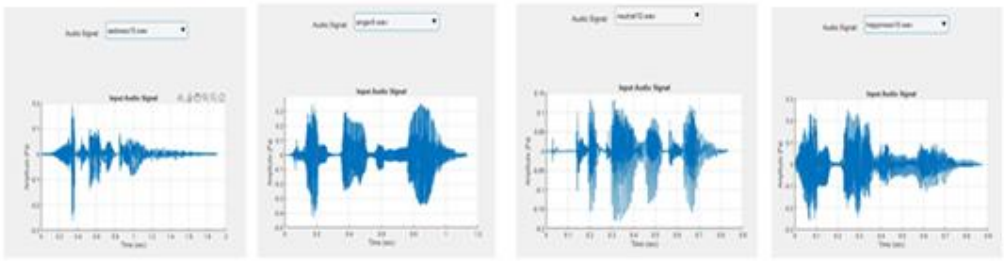

Figure 7. Graphical representation of different speech signal

The SVM and Random Forest Classifier are used for further recognition of emotions such as happy, sad, angry and neutral. The emotions of the speech signal is obtained. This is accomplished with the help of the MATLAB GUI. The emoji type emotion is retrieved in the GUI. Figure 8 shows the output obtained from SVM Classifier. For better prediction Random Forest Classifier is used. The output for Random forest Classifier is shown in the Figure 9.

The comparison of the performances of the classifiers are made from a graph and is shown in Table 1. Results show that, the Random Forest Classifier is better at classifying voice data based on emotion. However, when the trained model is used for binary class classification rather than multi class classification, the SVM Classifier performs better.

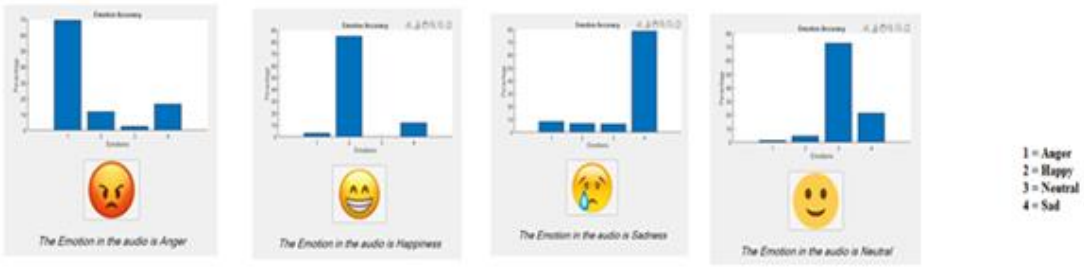

Figure 8. SVM Classifier

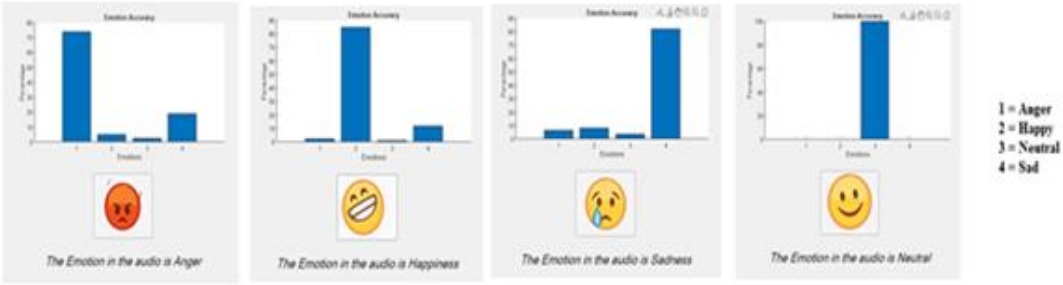

Figure 9. Random Forest Classifier

Table 1. Comparison of a Classifier performance

Classifier / Audio	SVM Classifier	Random Forest Classifier
Happy	83%	85%
Anger	70%	73%
Sad	79%	82%
Neutral	72%	100%

5. Conclusion

The emotion recognition from speech data is done using SVM and Random Forest classifier. Datasets of speech signals with different emotions are used for training and prediction of emotions. Based on the speech signal, the emotions are classified as happy, anger, sad and neutral. Among the two classifiers, Random Forest performs better in recognizing the emotions up to 85% whereas Support Vector Machine (SVM) gives 76% accuracy. Hence it shows that Random Forest is better for classification of speech data and can be used to detect a person's emotion through speech in real-life scenarios.

References

[1] T. Sai Samhith, G. Nishika, M. Prayuktha, "Speech based Emotion Recognition using Machine Learning Algorithms." International Journal of Creative Research Thoughts(IJCRT), 09, 06 (2021) : 815-823.

[2] Mao, Q., Dong, M., Huang, Z., & Zhan, Y, "Learning Salient Features for Speech Emotion Recognition Using Convolutional Neural Networks." IEEE transactions on multimedia, 16, 8 (2014) : 2203-2213.

[3] Ms. Shambhavi.S.Sheerur, Dr. V.N.Nitnaware, "Emotion Speech Recognition using MFCC and SVM." International Journal of Engineering Research & Technology(IJERT), 04, 06(2015).

[4] R. Thiruvengatanadham, "Speech/Music Classification using PLP and SVM." International Journal of Engineering and Computer Science(IJECS), 08, 02(2019) : 24469-24472.

[5] Mohammad Jawad AI Dujaili, Abbas Ebrahimi-Moghadam, "Speech Emotion Recognition using SVM and KNN Classifications fusion." International Journal of Electrical and Computer Engineering(IJECE), 11, 02(2021) : 1259-1264.

[6] Ajay Gupta, Siddhesh Morye, Mukul Sitap, "Speech based Emotion Recognition using Machine Learning." International Research Journal of Engineering and Technology(IRJET), 08, 04(2021) : 3289-3295.

[7] Fatemeh Noroozi, Tomasz Sapinski, "Vocal based Emotion Recognition using Random forest and Decision Tree." International Journal of Speech Technology, 20, 05 (2017).

[8] P.Dhanalakshmi, Suresh P. Kumar, R. Thiruvengatanadham. "Speech/Music Classification using SVM." International Journal of Computer Applications, 65, 06(2013) : 36-41.

[9] Navya Damodar, Vani H Y and Anusuya M A, "Voice Emotion Recognition using CNN and Decision Tree." International Journal of Innovative Technology and Exploring Engineering(IJITEE), 08, 12(2019) : 4245-4249.

[10] Mohammed Zakariah, "Classification of large datasets using Random Forest Algorithm in various applications: Survey." International Journal of Engineering and Innovative Technology(IJEIT), 04, 03(2014) : 189-198.

[11] Anjani Reddy J, Dr. Shiva Murthy G, "Emotion Recognition from Speech using MLP and KNN." International Journal of Engineering Research and Applications(IJERA), 11, 06(2021) : 34-38.

[12] Ritu D. Shah, Dr. Anil C. Suthar, "Speech Emotion Recognition based on SVM using MATLAB." International Journal of Innovative Research in Computer and Communication Engineering(IJIRCCE), 04, 03(2016) : 2916-2921.

[13] Dinesh Kumar, Upasana Sharma, Shewta Solanki, "Voice Stress Analysis using Linear Predictive Coding in MATLAB." International Journal of Engineering Technologies in Engineering Research(IJETER), 04, 07(2016) : 210-213.

[14] Hend Ab.ELLaban, A.A Ewees, Elsaeed E.AbdElrazek, "A Real-Time System for Facial Expression Recognition using Support Vector Machines and K-Nearest Neighbor Classifier." International Journal of Computer Applications(IJCA), 04, 08(2017) : 23-29.

[15] Anusha A. U, Anusha C, Bhumika S, Harshini S, "Development of Real Time Facial Emotional Video Analysis with Speech Data Fusion using Random Forest Algorithm." International Journal of Engineering Research & Technology(IJERT), 09, 12(2021).