An Approach to Speech Emotion Classification Using k-NN and SVMs

Disne SIVALINGAM

(*Department of Computer Science*, *Trincomalee Campus*, *EUSL*, *Sri Lanka*)

Abstract: The interaction between humans and machines has become an issue of concern in recent years. Besides facial expressions or gestures, speech has been evidenced as one of the foremost promising modalities for automatic emotion recognition. Effective computing means to support HCI (Human-Computer Interaction) at a psychological level, allowing PCs to adjust their reactions as per human requirements. Therefore, the recognition of emotion is pivotal in High-level interactions. Each Emotion has distinctive properties that form us to recognize them. The acoustic signal produced for identical expression or sentence changes is essentially a direct result of biophysical changes, (for example, the stress instigated narrowing of the larynx) set off by emotions. This connection between acoustic cues and emotions made Speech Emotion Recognition one of the moving subjects of the emotive computing area. The most motivation behind a Speech Emotion Recognition algorithm is to observe the emotional condition of a speaker from recorded Speech signals. The results from the application of k-NN and OVA-SVM for MFCC features without and with a feature selection approach are presented in this research. The MFCC features from the audio signal were initially extracted to characterize the properties of emotional speech. Secondly, nine basic statistical measures were calculated from MFCC and 117-dimensional features were consequently obtained to train the classifiers for seven different classes (Anger, Happiness, Disgust, Fear, Sadness, Disgust, Boredom and Neutral) of emotions. Next, Classification was done in four steps. First, all the 117-features are classified using both classifiers. Second, the best classifier was found and then features were scaled to [-1, 1] and classified. In the third step, the with or without feature scaling which gives better performance was derived from the results of the second step and the classification was done for each of the basic statistical measures separately. Finally, in the fourth step, the combination of statistical measures which gives better performance was derived using the forward feature selection method Experiments were carried out using k-NN with different k values and a linear OVA-based SVM classifier with different optimal values. Berlin emotional speech database for the German language was utilized for testing the planned methodology and recognition rates as high as 60% accomplished for the recognition of emotion from voice signal for the set of statistical measures (median, maximum, mean, Inter-quartile range, skewness). OVA-SVM performs better than k-NN and the use of the feature selection technique gives a high rate.

Keywords: Mel Frequency Cepstral Coefficients (MFCC), Fast Fourier Transformation (FFT), Discrete Cosine Transformation (DCT), k Nearest Neighbors (k-NN), Support Vector Machine (SVM), One-Vs-All (OVA)

1 Introduction

The speech signal is the most natural method of communication between humans [1]. This fact has motivated to think of speech as a fast and efficient method of interaction between human and machine. Speech signal contains emotional contents because of that and the body languages human can understand the emotional states of another[8]. However, for a natural human-machine interaction the machine should understand the emotional state of the speaker for satisfactory response to human emotions [5]. This leads to a research field namely Speech Emotion Recognition, which defined as extracting the emotional state of a speaker from his or her speech[6].

Speech emotion recognition is particularly useful for applications, which require natural human-machine interactions where the response of those systems to the user depends on the detected emotion. Such applications are E-learning, Call-center System, Music Player, and Sony Artificial Intelligent Robot (AIBO).

The task of speech emotion classification is incredibly difficult for subsequent reasons. Initially, it's not clear that speech features are most powerful in identifying between emotions, the acoustic variability introduced by the existence of various sentences, speaking rates, speaking styles of the speakers adds another obstacle. As a result, these properties have an effect directly on most of the common extracted speech features like energy contours and pitch. [1]. Sentence may contain several kinds of emotions at the same time, and emotions may be associated with just parts of the sentence. It is very difficult to determine the boundaries between these parts. An example of this situation is a live telecast of a program. One sentence may have more than one emotion. One sentence can speak in more than one language. Such as Tanglish (Tamil and English). No clear boundary for each complex emotion such emotions are shame, pride, satisfaction. Speaker's cultural backgrounds and the environments.

2 Literature Survey

Lalitha et al. (2015) [11] suggested a technique in view of the impact of MFCC and Cepstrum features in emotion detection. In addition, they carried out a relative analysis of Cepstrum, Mel-Frequency Cepstral Coefficients, and frequency scaled MFCC on Emotion Recognition. Creators utilized the Berlin Emo-set [12], which includes 7 fundamental emotions, the furthermore utilized Neural Network to do the classification. A classification rate of 57% with MFCC features and 85.70% with the combination of MFCC features, Cepstrum, and Frequency, scaled MFCC mentioned as the most extreme outcomes.

Angel Urbano Romeu (2016) [4] proposed a technique of Emotion Recognition in view of speech, utilizing a Naïve Bayes Classifier, His research the outcomes from the use of a Naïve Bayer Classifier over different kinds of features. He has used pitch and MFCC related features and used Berlin Dataset focused only on five emotions (Happy, Anger, Fear, Sad and Neutral). OpenSMILE toolkit used to extract features from the audio signal. Initially, he has obtained an accuracy of 70%. For classification improvement, he has used the weighting method and obtained results around 75% accurate the primary contrast between his project and also the others is the programing language used in C language and helpful to develop a final software, which may used in a automaton for various kinds of needs. He has obtained roughly 78% accuracy using proper layers and weights in the classifier. Moreover, they concluded classifying emotions with Naïve bayes provides fast probabilistic results and performs better than classifiers that are more sophisticated.

Shajini Majuran (2017) [13] proposed a Hierarchical Classification technique utilizing MFCC (Mel-Frequency Cepstral Coefficients). Statistical measures of MFCC are investigated in the classification individually. Also, the best fit model for every decision node is developed utilizing a one-versus-all-based SVM classifier. Two benchmark speech sets have been evaluated by the proposed framework: Danish and Berlin provide better performance and results in emotion classification. The statistical measures' subset gives the best recognition. The general results

reveal those essential feelings like anger, sadness, happiness, and neutral can classified in a hierarchical structure. According to the analysis on both DES and Berlin datasets, they discovered the classification values obtained using Berlin datasets are relatively above the values obtained from the DES dataset. Test results of classification acquired on the Berlin dataset demonstrates that emotions can be organized in order: Sad, Disgust, Anger, Fear, Happiness, Bored, and Neutral, respectively. Classification rate using OVA-SVMs for Berlin and DES datasets obtained 70.88% and 55.35% respectively. Classification rate using Hierarchical Scheme (without feature selection) for Berlin and DES datasets obtained 84.17% and 72.21% respectively. Classification rate using Hierarchical Scheme (with feature selection) for Berlin and DES datasets obtained 87.20% and 78.54% respectively.

Many researchers have employed emotion classification from audio signal with different features and classifiers. It can see that a couple of works have focused in on frequency and energy. In such a way, MFCC are considered as the more instructive coefficients. Also, various sorts of features and its statistical characteristics have been taken and assessed all through the research in Emotion Recognition. Plus, successful characteristics from statistical measures can be recognized for a superior classification rate. Hence, this paper proposes a way to deal with Speech Emotion Classification utilizing statistical measures of MFCC features and Its viable subset of features.

3 Background

3.1 State of the Art Approach

Emotion classification from speeches consists of three parts; first, one is Speech processing (noise removal and feature extraction). This part Removes unwanted noise components from signals like wind, rain, etc. The second part is Feature selection. Some features are unhelpful for classification if we use these features, it will affect the classification rate. Feature selection is the process of reducing the no of features. Best feature selection reduces the time and increases the accuracy. Best feature selection allows the development of simpler and faster models. The third part is Classification. In this part classifying, the test set to predict which emotion to assign to the speech.

3.2 Features

Extracting features is the first step in any automatic speech recognition-related system. That is determining the elements of the audio signal that are good for distinguishing the linguistic content and dropping all the other stuff that carries information like background noise, emotion, and so on the main point to know regarding speech is that the shape of the vocal tract including the tongue, teeth that filters the sounds generated by a human. This shape verifies what sound comes out [3]. If we are able to decide the shape correctly, it could provide us with a precise representation of the phoneme delivered.

The shape of the vocal tract appears in the envelope of the short-term power spectrum, and the Mel Frequency Cepstral Coefficients (MFCC) work has precisely portrayed this envelope [9].

The Mel Frequency Cepstral Coefficients (MFCC) are widely used features in automatic speech and speaker recognition. They were introduced by Davis and Mermelstein in the 1980s [9] and have been state-of-the-art ever since. Major steps of computation used to generate MFCC features shown in Fig.2.

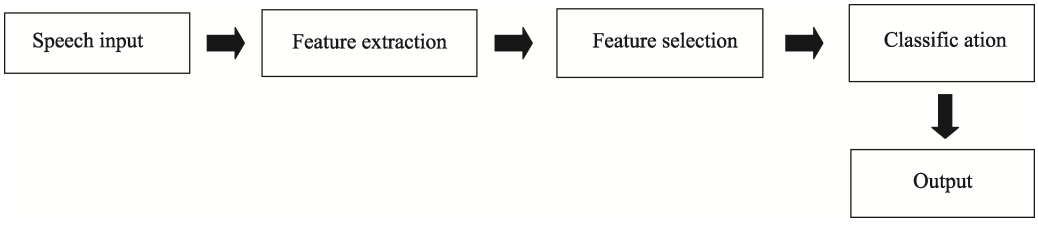

Fig.1 A Generic Framework for State of the Art Approach in Emotion Classification

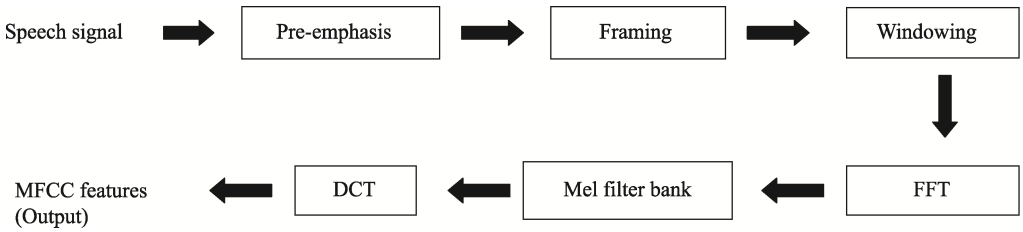

Fig.2 Block Diagram for Generating MFCC Features

a) Pre-emphasis

In the speech process, this signal filter is expected after the sampling to get a smoother spectral shape of the frequency of the speech signal. In other words, this is performed to reduce noise during sound recording.

The filter distortion is based on the input/output relationship in the time domain, expressed in the following equation:

$$y(n) = x(n) - bx(n-1) \quad (1)$$

From equation 1, a- is a filter constant, it is usually $0.9 < b < 1.0$.

b) Framing

In this step, the sound signals have been sectioned into many overlapping so that no single signal is eliminated. This step goes on until all signals have entered at least one frames as displayed (Fig.2). Speech analysis was performed by short-term analysis. A sound signal is continually changing, and to keep things simple we should accept that the sound signal doesn't change much on short time frame scales.

Therefore, we outline the signal in 20-40 ms frames. Assuming that the frame is short, we need more samples to get a reliable estimation of the spectral, assuming it is longer the signal changes a lot all through the frame.

c) Windowing

Windowing is a process for examining long sound signals by taking an adequate delegate area. Windowing is a Finite Impulse Response (FIR) digital filter approach. This process eliminates the associating signal because of the intermittence of the sign pieces.

The discontinuities happen because of the framing process. Assuming that we characterize the window as w (n), $0 \le n \le N - 1$, where N is the number of samples in each frame, the result of windowing is a signal (equation 2):

$$y(n) = x(n)\, w(n),\ 0 \le n \le N-1 \quad (2)$$

From Equation 2, y(n) is the signal resulting from the convolution between the input signal and the window function, and x(n) represents the signal that the window function must convolve. Where w(n) generally uses the Hamming window, which has the form (Equation 3):

$$w(n) = 0.54 - 0.46.\ cos\ (2\pi n/N-1),\ 0 \le n \le N-1 \quad (3)$$

Hamming window is most popular due to its high-resolution performance and it powerfully shapes the separation in amplitude at its ends.

d) Fast Fourier Transformation (FFT)

Fourier transform converts a time series of bounded time-domain signals into a frequency spectrum. After the windowing method, this step will be applied, and the resulting frame is converted into a frequency spectrum.

FFT is a quick algorithm of DFT (Discrete Fourier transform) that is advantageous for changing each frame to N samples from the time domain into the frequency domain.

$$X_n = \sum_{k=0}^{N-1} x^k - e^{-2\pi jkn/N} \quad (4)$$

Fourier Transformation equation 4, shows n = 0, 1, 2…, N-1 and j = sqrt-1. X_n is the n-frequency pattern produced from the FT, x^k is the signal of a frame. The result of this stage is typically called Spectrum.

e) Mel filter bank

The Spectrum still has many data that isn't needed for ASR (Automatic Speech Recognition). Specifically, the cochlea cannot decide the difference between intently spaced frequencies. This impact is more articulated as the frequencies increase. Thus, we take clumps

of period gram bins and sum them to get an idea of ways much energy exists in different frequency areas.

This is performed by our Mel filter bank: the basic channel is narrow and offers an indication of what extent of energy exists near zero Hertz. as the frequencies get higher, our filters get wider as we have a tendency to become less involved in variations.

The human ear also propels this: we don't hear tumult on a linear scale. To double the perceived loudness of a sound, we generally have to put eight times more energy into it. This implies that enormous varieties in energy probably won't sound different assuming the tone is loud. This compression process makes our functions more closely match what individuals really hear.

The formula for calculating the Mel frequency for some random frequency f in Hz is given in Equation 5:

$$\text{Mel}(f) = 2595 \log_{10}(1+f/700) \quad (5)$$

f) Discrete Cosine Transformation (DCT)

The final step is to calculate the DCT of the filter bank energies. Still, filter banks are overlapping, the filter bank energies are quite connected with every other. The DCT decorrelates the energies. In any case, notice that only 13 of the 26 DCT coefficients were kept. This is for the reason that the higher DCT coefficients represent quick changes in the filter bank energies, and it appears to be that these quick changes really reduce the performance, thus we get a little improvement by dropping them [17].

4 Feature Scaling

This is a process of normalizing the scope of independent factors or features in the data. In Data Preprocessing, this is called Normalization and is normally performed during the data pre-processing step. Since the range of values of the raw data is much different, in some AI algorithms the objective functions won't work accurately without Normalization. For instance, most classifiers find the distance between two points using Euclidean distance. Assuming one of the features has a wide range of values, that specific feature determines the distance. Therefore, the range of all features should normalize so that each feature contributes approximately proportionately to the final distance. Another reason why feature scaling applied is that gradient descent converges much faster with feature scaling than without it [10]. There are various methods to do feature scaling such as Rescaling (min-max normalization), Mean normalization, Standardization, Scaling to unit length. For this research, rescaling method is used.

Rescaling: Also referred to as minimum-maximum scaling or minimum-maximum normalization, is that the simplest technique and consists in rescaling the range of features to scale the range in [0, 1] or [−1, 1]. Picking the target range relies upon the nature of the data. The overall formula is given as (equation 6):

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (6)$$

Where x - original value, x' - normalized value.

5 Feature Selection

It is important when the number of features is very large there is no need to use every feature. We can use features only those that are important. The top reasons are to use feature selections: It enables the machine-learning algorithm to train faster, it reduces the complexity of a model and makes it easier to interpret, it improves the accuracy of a model if the right subset chosen, and it reduces overfitting. There are various techniques and methodologies that can use to subset feature space and help the models perform better and more efficiently. Such one method is the wrapper method [14].

In wrapper methods, we attempt to use a subset of features and train a model using them. Based on the inferences that draw from the previous model, decide to add or remove features from the subset. The problem essentially reduced to a search problem. These methods are typically computationally very expensive. Some common samples of wrapper methods are forward feature selection, backward feature elimination, recursive feature elimination, and so forth.

Forward Selection: Forward selection is an iterative technique in which we start with having no feature

in the model. In every iteration, we keep adding the feature which best improves our model till an addition of a new variable doesn't improve the performance of the model.

6 Classifiers

a) k-Nearest Neighbors

k-NN classification is one of the most basic and simple classification methods and should be one of the primary selections for a classification study once there's very little or no previous data about the distribution of the data. k-nearest-neighbor classification developed from the necessity to perform discriminant analysis when reliable parametric estimates of probability densities are unknown or difficult to determine. k-NN is a non-parametric method used for classification and regression. In each case, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN used for classification or regression.

In the kNN classification, the result is a class membership. An object is classified by the majority vote of its neighbors, with the object being assigned the most common class among its k nearest neighbors (k is a positive integer, usually small). If k = 1, then the object is simply assigned to that nearest neighbor's class. In kNN regression, the output is the object's property value. This value is the average of the values of its nearest neighbors [18].

b) SVM-based multi-class classifiers

SVM is a supervised machine-learning algorithm, which may be, used for both classification and regression challenges. However, it's mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is that the number of features you have) with the value of each feature being the value of a particular coordinate [1]. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well [7]. There are three standard techniques often used by SVMs to tackle multi-class problems, namely one-versus-one (OVO), one-versus-all (OVA), and directed acyclic graph (DAG). Here we tend to consider the OVA method.

One-versus-all (OVA) method is implemented using a "winner-takes-all" strategy. It constructs k separate binary classifiers for k-class classification. The m^{th} binary classifier trained using the data from the m^{th} class as

Positive examples and the remaining k − 1 classes as negative examples. During test, the binary classifier that gives maximum output value determines the class label.

For an example, the winner-takes-all strategy assigns it to the class with the boundary function highest classification. Fig.4 shows the OVA architecture and also

the classification problems at each node for finding the best class out of four classes [19].

7 Methodology

7.1 Feature Extraction

Feature extraction technique is the first step in a speech emotion classification. The feature generating

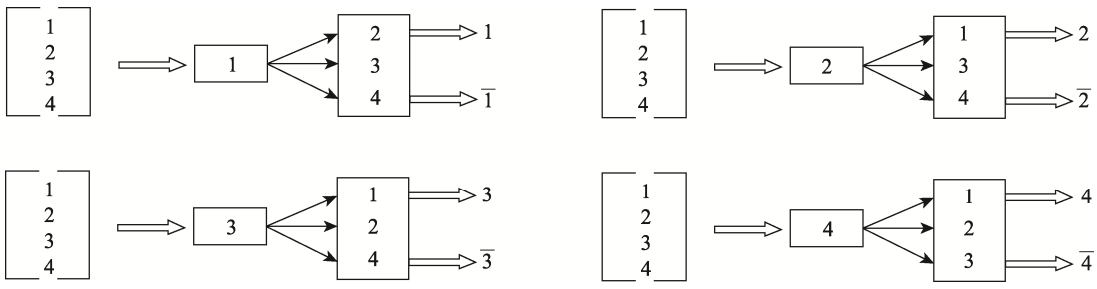

Fig.4 OVA Architecture

process from the dataset called as feature extraction. There are many features available to extract from an audio file. The specialists are as yet examining which ones are best for each particular instance of the speech signal are processing. The most useful and used one is the MFCC features [16].

Therefore, MFCC features selected to use for this research work. MFCC feature extraction shown in Fig.6 and the Fig.5 illustrate the method of emotion recognition system through speech.

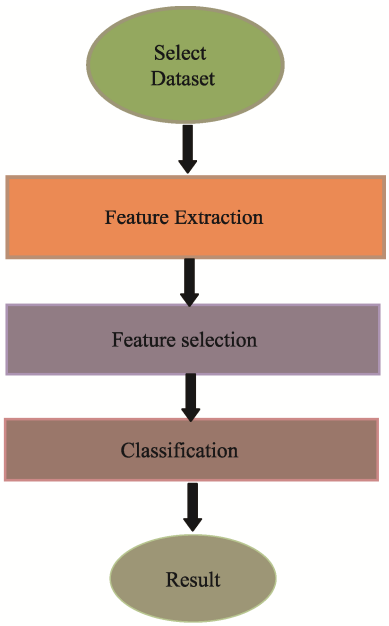

Fig.5 Diagram of Emotion Recognition System through speech

At first, a high-pass filter was applied along with a pre-emphasis coefficient value of 0.97. From there on speech signals changed over into a sequence of feature frames having a window length of 25 msec. Then, at that point, the Hamming Window was applied utilizing 10 ms of hop time. After that, the signals are converted from the time domain to the frequency domain by Fast Fourier Transformation. Then, the filter bank according to the Mel scale is used to calculate the weighted sum of the filter spectral components so the result of the method for a Mel scale approaches. When the channel bank energies are gotten, the logarithm of these energies is calculated.

Finally, the log Mel range spectrum process is changed over into a time-domain utilizing DCT (Discrete Cosine Transformation). This change shows the outcome as MFCC, that depicted as a sequential arrangement of cepstral vectors.13 MFCC was used for this research [16]. Finally, 9 basic statistical measures (median, maximum, minimum, mean, standard deviation, inter-quartile range, skewness, kurtosis, and range) were calculated for each coefficient. And also grouped together resulting in a 117-dimensional feature set. Classification results using k-NN are shown in Table 1 and OVA-SVM showed in Table 2.

7.2 Feature Selection

Each of the basic nine statistical measures selected individually and tested, results shown in Table 3. Then forward feature selection done and obtained results shown in Table 4.

7.3 Classification

First K-Nearest neighbor approach used for classification. That is the simplest classification approach. Would be a nearest neighbor voting strategy that computes all pairwise Euclidean distances between key point representations of a test set to all or any labeled

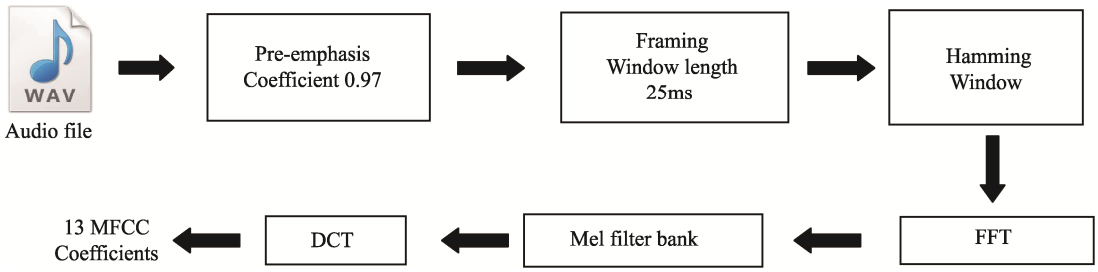

Fig.6 MFCC Extraction

sets within the dataset. Then we used the SVMs approach for classification. SVMs are a supervised learning technique based on a statistical learning theory which will use for pattern classification and regression. Once applied to classification, a linear SVM finds the hyperplane leaving the largest possible fraction of points of the same class on the same side, whereas maximizing the distance of either class from the hyperplane. In this project, we've got used OVA-SVM.

Overview of the experiment shown in Fig.7. The aim of the experiment was to clarify the accuracy of selected groups of features, as well as classification ability of selected classification methods for speech emotion classification. Samples of examination formed from recordings of human speech with various emotional characters. The following settings used in the experiment.

7.4 Datasets

Berlin database of emotional utterances used as Dataset. The database contains of 535 Utterances spoken in German which consisting of seven emotions such as angry (one hundred and twenty-seven), bored (eighty-one), disgust (forty-six), fearful (sixty-nine), happy (seventy-one) emotions, sad (sixty-two) and neutral (seventy-nine) emotions. Numbers in parentheses indicate the number of utterances per emotion, each sentence consists of an average of ten words with an average duration of approximately five seconds, recorded by ten different actors (both genders). The audio files are in WAV file format.

For classification, the k-nearest neighbors' classifier was used for 70% of training data and 30% of testing data. For the classification, the value of k is used as k=1, 3, 5, 7, 11, 13. OVA-SVMs used for 70% of training data and 30% of testing data. The optimal values c was used 2^{-14} to 2^{10}. Then the classifications done in three categories. They are classification without scaling data, classification after Scaling data to [−1, 1] and select appropriate features and classify using SVMs.

Valence of the Berlin Database is shown in Fig.8.

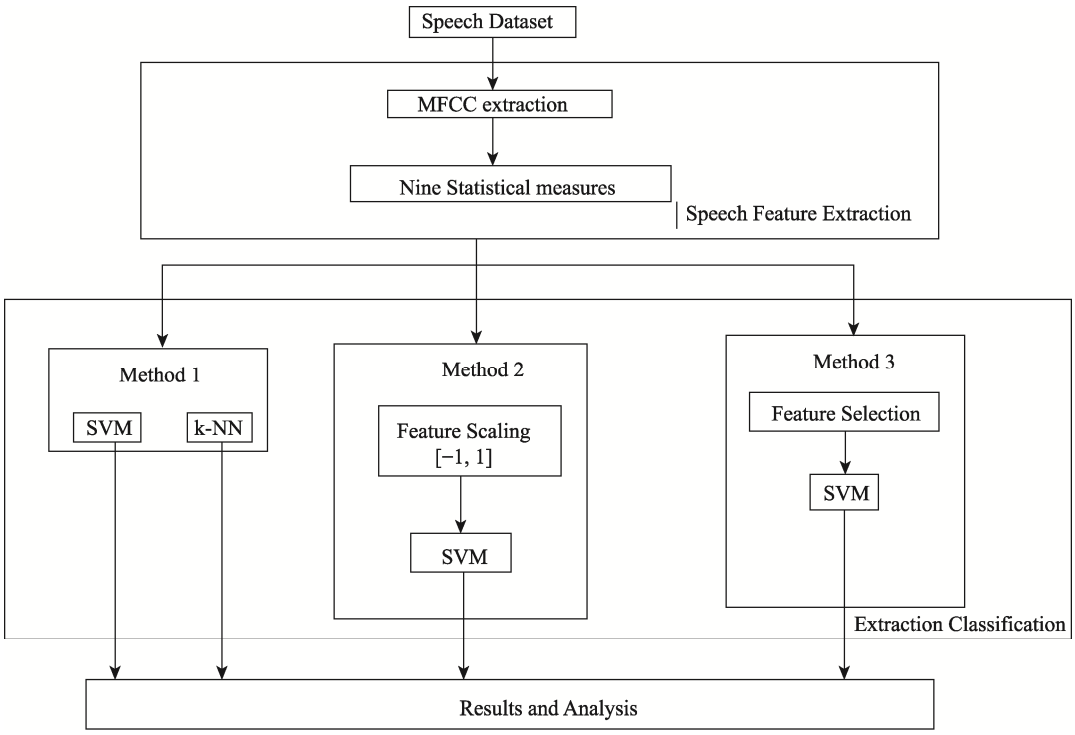

Fig.7 Architecture of Methodology

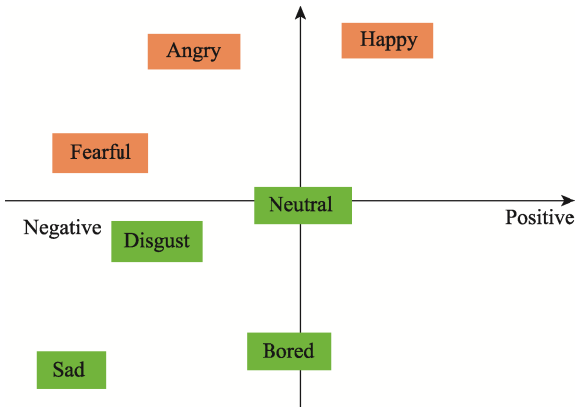

Fig.8 Emotions of Berlin Database

Table 1 No of Samples in Berlin Database

Emotion	Female	Male
Angry	67	60
Happy	44	27
Bore	46	35
Neutral	40	39
Disgust	35	11
Fear	32	37
Sad	37	25
All	301	234

8 Results

Table 2 Classification Rate of MFCC Using k-NN

Rate	k Value
41.88%	1
47.50%	3
44.38%	5
48.13%	7
48.13%	9
51.88%	11
48.75%	13

Table 3 Classification Rate of MFCC with and without Scaling the Data Using SVM

With Scaling		Without Scaling	
Rate	Obtained C Value	Rate	Obtained C Value
56.87	2^{-1}	57.50	2^{-8}

Table 4 Classification Rate of Statistical Measures of MFCC Using SVM

Feature Set	Rate	Obtained C Value
m1, m8	58.75	2^{3}
m4, m8	54.36	2^{5}
m1, m4, m8	57.50	2^{-4}
m1, m3, m4, m6, m9, m8	60.00	2^{-5}

Table 5 Classification Values Using the Subset of Statistical Measurements of MFCC Using SVM

NO	Statistical Measures	Rate	Obtained C value
m1	Median	46.25	2^{-3}
m2	Maximum	43.75	2^{-1}
m3	Minimum	41.88	2^{4}
m4	Mean	44.38	2^{4}
m5	Standard Deviation	42.50	2^{-14}
m6	Inter-quartile Range	43.75	2^{-1}
m7	Kurtosis	41.87	2^{2}
m8	Skewness	50.00	2^{-3}
m9	Range	43.75	2^{5}

9 Conclusion

The objective of this project is to understand and examine the relative performance of nine basic statistical measures obtained from MFCC features for emotion classification. For this purpose, we used MFCC feature extraction and then calculated the nine statistical measurements of MFCC. In this study, 1-fold cross-validation applied for partitioning the data. For classification purposes, we used SVM and k-NN classifiers. SVM performs better than k-NN. When we compare our testing results, without scaling the data gives better performance and less time than scaling data. In this research work, I used the rescaling (min-max normalization) method. Normally scaling increases the performance. In this case, decreasing. Because the audio signal is constantly changing, in short time scales the audio signal does not change much. Therefore, the signal normalized already. Classification rates of the basic nine statistical measures for emotions vary 41-50% and skewness shows better performance. Some features alone have a low classification rate because they have less information and are

useless for classification. Therefore, we appropriate feature selection. Some features are unhelpful for classification; if we use such features, it will affect the classification rate. Best feature selection allows the development of simpler and faster models. For this research, I have used the forward feature selection method. Classification rate using the subset of statistical measures (median, maximum, mean, Inter-quartile range, skewness) set gives better classification rate of 60%. The performance increased with feature selection.

References

[1] Moataz El Ayadi, Mohamed S Kamel, and Fakhri Karray. "Survey on speech emotion recognition: *Features, classification schemes, and databases.* Pattern Recognition, 44 (3): 572-587, 2011.

[2] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, Visual Categorization with Bags of Keypoints, *Proceedings of Workshop on Statistical Learning in Computer Vision, ECCV*, pp.1-22, 2004.

[3] Ali Hassan, *On Automatic Emotion Classification Using Acoustic Features*, University of Southampton, Faculty of Physical and Applied Sciences, 2012.

[4] Angel Urbano Romeu, *Emotion recognition based on the speech, using a Naïve Bayes Classifier*, Institute for Computer Technology, 2016.

[5] Pooja Yadav and Gaurav Aggarwal, Speech Emotion Classification using Machine Learning, *International Journal of Computer Applications*, 118(13), 2015.

[6] S. Casale, A. Russo, and G. Scebba, Speech Emotion Classification using Machine Learning Algorithms, *The IEEE International Conference on Semantic Computing*, 2008.

[7] Peipei Shen, Zhou Changjun, Xiong Chen, Automatic Speech Emotion Recognition Using Support Vector Machine, *IEEE International Conference on Electronic and Mechanical Engineering and Information Technology (EMEIT) volume 2*, Page(s): 621-625, 12-14 Aug. 2011.

[8] Akalpita Das, Purnendu Acharjee, Laba Kr. Thakuria, A brief study on speech emotion recognition, *International Journal of Scientific and Engineering Research (IJSER), Volume 5*, Issue 1, pg-339-343, January-2014.

[9] Mel Frequency Cepstral Coefficient (MFCC), *[Online]. Available: [http://practicalcryptography.com/miscallaneous/machine learning/guide-mel-frequency-cepstral- coefficients-mfccs/]*. Accessed June 10, 2018.

[10] Feature Scaling, *[Online] Available: [https://en. wikipedia.org/wiki/Feature_scaling]*, Accessed: June 14, 2018

[11] S. Lalitha, D. Geyasruti, R. Narayanan and M. Shravani, Emotion detection using MFCC and Cepstrum Features, in *Fourth International Conference on Eco-friendly Computing and Communication Systems*, *volume 70*, pp. 29-35, 2015.

[12] Berlin Database of Emotional Speech [Online]. Available *[http://emodb.bilderbar.info/docu/]*. Accessed July 12, 2017.

[13] Shajinii Majuran and Amirthalingam Rananan, A Feature-Driven Hierarchical Classification *Approach to Emotions in Speeches Using SVMs, IEEE*, 2017.

[14] Feature Selection Methods, [Online]. Available*: [https: //www.analyticsvidhya.com/blog/2016/12/introduction-to-feature-selection-methods-with-an-example-or-how-to-select- the-right-variables/]*. Accessed June 10, 2018.

[15] The Implementation of Speech Recognition using Mel-Frequency Cepstrum Coefficients (MFCC) and Support Vector Machine (SVM) method based on Python to Control Robot Arm, [Online]. Available: *http://iopscience.iop.org/article/10.1088/1757899X/288/1/012042*Accessed June 10, 2018.

[16] FEATURE EXTRACTION USING MFCC, [Online], Available *http://aircconline.com/sipij/V4N4/4413sipij08.*pdf, Accessed: May 29, 2018.

[17] Speech Processing for Machine Learning: Filter banks, Mel-Frequency Cepstral Coefficients (MFCCs) and What's In-Between, [Online] *Available: [http:// haythamfayek.com/2016/04/21/speech-processing-for machine-learning.html]*, Accessed: June 12, 2018.

[18] A Quick Introduction to K-Nearest Neighbors Algorithm[Online].Available[*https://medium.com/@adi.bronshtein/a-quick-introduction-to-k-nearest-neighbors-algorithm- 62214cea29c7]*. Accessed: June 21, 2018.

[19] Understanding Support Vector Machine algorithm frome xamples[Online]. Available*: [https://www. analyticsvidhya.com/blog/2017/09/understaing-support-vector achine-example-code/]*, Accessed: June 12, 2018.

Author Biography

Disne SIVALINGAM, received her B.Sc. (Hons) Degree from Eastern University Sri Lanka, in 2018. She is a M.Sc. candidate at Moratuwa University, Sri Lanka. Her main research interests include Natural Language Processing, Machine Learning, and Sentiment Analysis.

Email: disnes@esn.ac.lk