International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 14
October2016

An Approach To Feature Selection Algorithm Based On Ant Colony Optimization For Human Emotion Recognition Using Speech

Swati Pahune
PG student (E & C)
Vidarbha Institute of Technology, Nagpur

Vinay Keswani
Asst.Professor (E & C)
Vidarbha Institute of Technology, Nagpur

Nilesh Bodne
Asst.Professor (E & C)
Vidarbha Institute of Technology, Nagpur

Abstract

Speech is one of the most promising models by which people can express their emotions like anger,sadness, and happiness. These states can be determined using various techniques apart from facial expressions. Acoustic parameters of a speech signal like energy, pitch, Mel Frequency Cepstral Coefficient (MFCC) are important in finding out the state of a person. In this project, the speech signal is taken as the input and by means of MFCC feature extraction method, cepstral coefficients are extracted by using MFCC. The large amount of extracted features may contain noise and other unwanted features. Hence, an evolutionary algorithm called as Ant Colony Optimization (ACO) is used as an efficient feature selection method. By using ACO technique the unwanted features are removed and only best feature subset is obtained. It is found that the total number of features extracted get reduced considerably. The software used is MATLAB

Keywords: ACO Classifier, Emotion recognition, Feature extraction, Feature selection.

I. Introduction

Speech is a complex signal which contains information about the message, speaker, language and emotions. Speech is produced from a time varying vocal tract system excited by a time varying excitation source. Emotion on other side is an individual mental state that arises spontaneously rather than through conscious effort. There are various kinds of emotion which are present in a speech. The basic difficulty is to cover the gap between the information which is captured by a microphone and the corresponding emotion, and to model the specific association. This gap can be bridged by narrowing down various emotions in few, like anger, happiness, sadness, surprise, fear, and neutral. Emotions are produced in the speech from the nervous system consciously, or unconsciously. Emotional speech recognition is a system which basically identifies the emotional as well as physical state of human being from his or her voice [1][2]. Emotion recognition is gaining attention due to the widespread applications into various domains detecting frustration, disappointment, surprise/amusement etc. There are many approaches towards automatic recognition of emotion in speech by using different feature vectors.

A proper choice of feature vectors is one of the most important tasks. The feature vectors can be distinguished into the following four groups: continuous (e.g., energy and pitch), qualitative (e.g., voice quality) spectral (e.g., MFCC), and features based on the Teager energy operator (e.g., TEO autocorrelation envelope area).For classification of speech, methodologies followed are: HMM, GMM, ANN, k-NN, and several others as well as their combination which maintain the advantages of each classification technique. After studying the related literature it can be identified that the feature set which is mostly employed is comprised of pitch, MFCCs, and HNR. Additionally, the HMM technique is widely used by the researchers due to its effectiveness. Feature extraction by temporal structure of the

low level descriptors or large portion of the audio signal is taken could be helpful for both the modeling and classification processes.

II. Speech Emotion Recognition System

Speech emotion recognition aims to automatically identify the emotional state of a human being from his or her voice.It is based on in-depth analysis of the generation mechanism of speech signal, extracting some features which contain emotional information from the speaker's voice, and taking appropriate pattern recognition methods to identify emotional states. Fig.1 indicates the speech emotion system components.

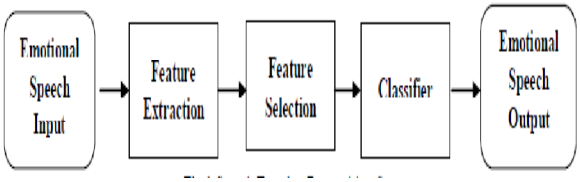

Fig. 1 Speech Emotion Recognition System

Like typical pattern recognition systems, speech emotion recognition system contains four main modules: speech input, feature extraction, feature selection, classification, and emotion output. Since a human cannot classify easily natural emotions, it is difficult to expect that machines can offer a higher correct classification. A typical set of emotions contains 300 emotional states which are decomposed into six primary emotions like anger, happiness, sadness, surprise, fear, neutral. Success of speech emotion recognition depends on naturalness of database. [4][5].

There are six databases accessible: two freely accessible ones, the Danish Emotional Speech corpus (DES) and Berlin Emotional Database (EMO-DB), and four databases from the Interface venture with Spanish, Slovenian, French and English enthusiastic discourse. These databases contain acted enthusiastic discourse.

III. Feature Extraction and Selection

Success of emotion recognition is dependent on appropriate feature extraction as well as proper classifier selection. For classification of various emotions in speech, methodologies followed are: HMM, GMM, SVM, ANN, k-NN, LDC , ACO and several others as well as their combination which maintain the advantages of each classification technique [13][14][15][16].

Speech signal composed of large number of parameters which indicates emotion contents of it. Changes in these parameters indicate changes in the emotions. Therefore proper choice of feature vectors is one of the most important tasks. There are many approaches towards automatic recognition of emotion in speech by using different feature vectors. By employing feature extraction technique number of features can be extracted from the emotional speech. To achieve accurate identification of emotion classifier should provided with single best feature. Therefore there is need of systematic feature selection to reduce unuseful features from the base features. To select best features Forward Selection method can be used. The remaining features can be used by classifier to increase classification accuracy.

Feature vectors can be classified as long-time and short-time feature vectors. The long-time ones are estimated over the entire length of the utterance, while the short-time ones are determined over window of usually less than 100ms. The long-time approach identifies emotions more efficiently. Short time features uses interrogative phrases which has wider pitch contour and a larger pitch standard deviation.

IV. Mel-Frequency Cepstrum Coefficients (MFCC):

The most prevalent and dominant method used to extract spectral features is calculating Mel-Frequency Cepstral Coefficients (MFCC). MFCCs are one of the most popular feature extraction techniques used in speech recognition based on frequency domain using the Mel scale which is based on the human ear

scale. Psychophysical studies have shown that human perception of the sound frequency contents for speech signals does not follow a linear scale. Thus for each tone with an actual frequency f measured in Hz, a subjective pitch is measured on a scale called the 'Mel' scale

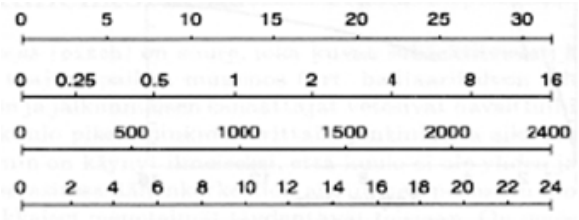

Fig. 2 Mel scale

Mel frequency scale represents subjective (perceived) pitch. It is one of the perceptually movtiated frequency scales (see figure above). Mel scale is constructed using pair wise comparisons of sinusoidal tones: a reference frequency is fixed and then a test subject (human listener) is asked to adjust the frequency of the other tone to be twice higher or lower. Mel scale models the non linear perception of frequencies in the human auditory system. Note that all the scales are related and: $f_{Mel} \approx 100 f_{Bark}$ (very roughly). MFCC is based on the characteristics of the human ear's hearing, which uses a nonlinear frequency unit to simulate the human auditory system. Mel frequency scale is the most widely used feature of the speech, with a simple calculation, good ability of the distinction, anti-noise and other advantages [7][8][10].

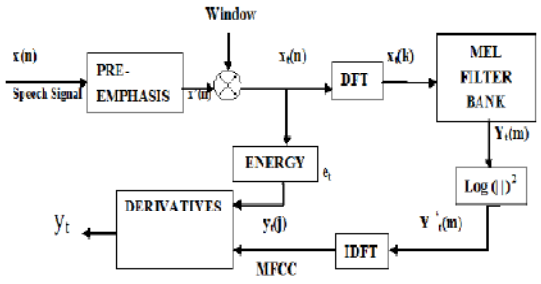

Fig.3 Block Diagram showing MFCC analysis

Mel-Frequency Cepstral Coefficients (MFCC) is a representation of the real cepstral of a windowed short-time signal derived from the Fast Fourier Transform (FFT) of that signal. The difference from the real cepstral is that a nonlinear frequency scale is used, which approximates the behaviour of the auditory system. Additionally, these coefficients are robust and reliable to variations according to speakers and recording conditions. MFCC is an audio feature extraction technique which extracts parameters from the speech similar to ones that are used by humans for hearing speech, while at the same time, deemphasizes all other information. The speech signal is first divided into time frames consisting of an arbitrary number of samples. In most systems overlapping of the frames is used to smooth transition from frame to frame. Each time frame is then windowed with Hamming window to eliminate discontinuities at the edges. The filter coefficients w(n) of a Hamming window of length n are computed according to the formula:

$$w(n) = 0.54 - 0.46cos\left(\frac{2\pi n}{N-1}\right), 0 \leq n \leq N-1$$
$$= 0, \text{Otherwise}$$

Where N is total number of sample and n is current sample. After the windowing, Fast Fourier Transformation (FFT) is calculated for each frame to extract frequency components of a signal in the time-domain. FFT is used to speed up the processing. The logarithmic Mel-Scaled filter bank is applied to the Fourier transformed frame. This scale is approximately linear up to 1 kHz, and logarithmic at greater frequencies. The relation between frequency of speech and Mel scale can be established as:

$$mel(f) = 2595 * \log_{10}\left(1 + \frac{f}{700}\right)$$

For each speech frame, a set of MFCC is computed. This set of coefficients is called an acoustic vector which represents the phonetically important characteristics of speech and is very useful for further analysis and processing in Speech Recognition. We can take audio of 2 second which gives approximate 128 frames each contain 128 samples (window size = 16 ms). We can use first 20 to 40 frames that give good estimation of speech. Total of forty Two MFCC parameters include twelve original, twelve delta (First order derivative), twelve delta-delta (Second order derivative), three log energy and three 0th parameter.

This all processes are implemented in Maltab program

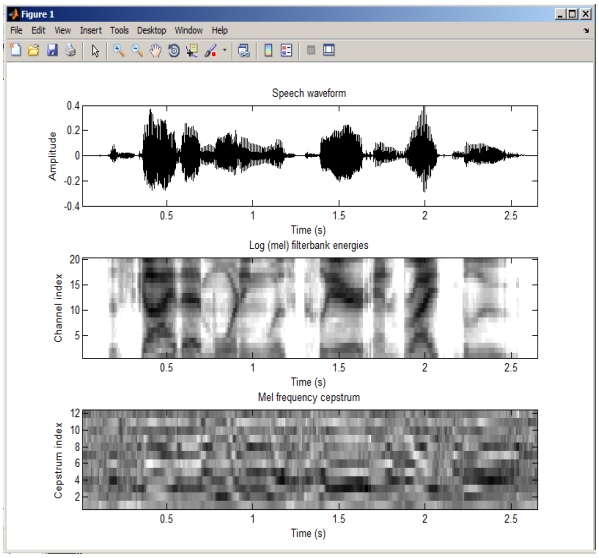

Fig.4 Implementation of MFCCFeatureExtraction

Fig. 4 shows speech signal taken from database whose amplitude is divided into no. of frames of duration 25 ms and each frame shift of 10 ms. We have taken 20 number of filter bank channels whose log filter bank energies are shown by the middle waveform. The filter bank is a set of overlapping triangular band pass filter, that according to mel-frequency scale, the centre frequencies of these filters are linear equally-spaced below 1 kHz and logarithmic equally-spaced above. The speech signal consists of tones with different frequencies. For each tone with an actual Frequency, f, measured in Hz, a subjective pitch is measured on the 'Mel' scale that is shown by third waveform. It also represents cepstral coefficients of the speech signal measured on the mel scale that are phonetically important characteristics of speech. Thus it extracts best features that is useful for processing in speech recognition.

V. Ant Colony Optimization

The main focus of this algorithm is to generate subsets of salient features of reduced size. ACO Feature Selection utilizes a hybrid search technique that combines the wrapper and filter approaches. In this regard, ACO Feature Selection modifies the standard pheromone update and heuristic information measurement rules based on the above two approaches. The reason for the novelty and distinctness of ACO feature selection algorithm versus previous algorithms like PSO, GA, lies in the following two aspects. First, ACO Feature Selection emphasizes not only the selection of a number of salient features, but also the attainment of a reduced number of them. ACO Feature Selection selects salient features of a reduced number using a subset size determination scheme. Such a scheme works upon a bounded region and provides sizes of constructed subsets that are smaller in number[12][18][19]. Thus, following this scheme, an ant attempts to traverse the node (or, feature) space to construct a path (or, subset). However, a problem is that, feature selection requires an appropriate stopping criterion to stop the subset construction. Otherwise, a number of irrelevant features may be included in the constructed subsets, and the solutions may not be effective. To solve this problem, some algorithms, define the size of a constructed subset by a fixed number of iteration for all ants, which is incremented at a fixed rate for following iterations. This technique could be inefficient if the fixed number becomes too large or too small. Therefore, deciding the subset size within a reduced area may be a good step for constructing the subset while the ants traverse through the feature space.

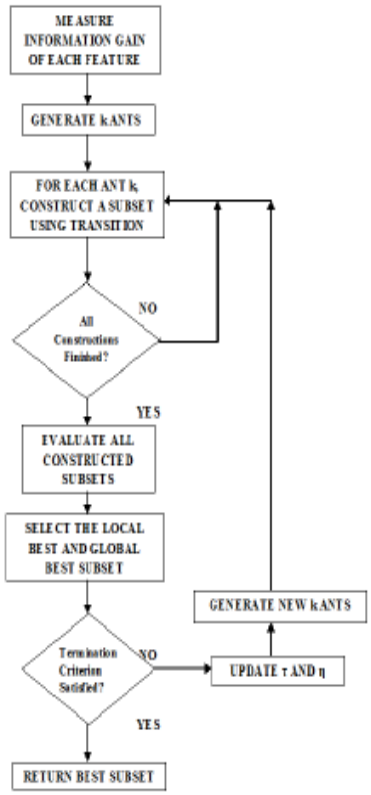

Fig.5 Flow Chart for ACO feature selection

VI. Results

We have combined mfcc feature extraction with ACO so as to get optimized features for our input speech signal.

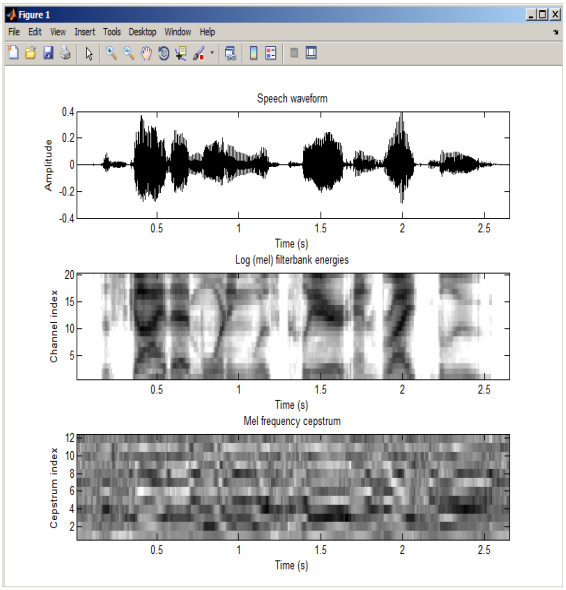

Fig.6 Implementation Of Feature Extraction Using MFCC With ACO

We have taken database of 100 samples having 24 samples of anger & happy emotions each, 31 samples of normal emotions & 21 samples of sad emotions. ACO algorithm along with mfcc tries to extract best features from each emotion. ACO optimizes the values of the extracted features that is **best cost** (from fig.7). After optimization 4 samples of each emotions are sent for training purpose & rest of the samples are used for testing purpose.

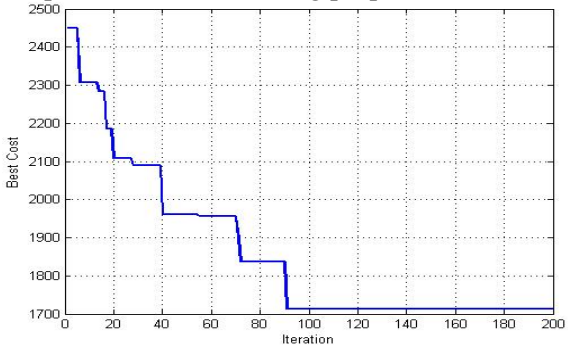

Fig.7 Optimized Features of Emotions

From the above graph, it is clear that 2500 features are observed that get optimized slowly from 2400 till 1700. The features get optimized value that is best cost value between 2100 to 2000, 2000 to mid of 1800 and 1800 to 1700. After 1700, features get optimized between 100^{th} to 200^{th} iteration and no further fluctuations are observed. So we can stop further iterations means we get optimized best cost value that is best features. Fig.8 shows the confusing matrix of accuracy of emotions in percentage which shows the matching with the same emotions diagonally.

	Anger %	Happy %	Normal %	Sad %
Anger	70.83	12.5	8.33	8.33
Happy	16.67	50	12.5	20.83
Normal	6.45	6.45	87.09	0
Sad	14.2	9.52	19.04	57.14

Fig.8 Confusing Matrix Of Accuracy Of Emotions

VI. Conclusion & Future Work

Within this project, we combine basic acoustic features to recognize different emotional states in speech. Some modifications of the algorithm are done and apply it to larger feature vectors containing Mel Frequency Cepstral Coefficients (MFCC) and their delta coefficient, and two energies. ACO algorithm was applied to this work to improve the quality of feature selection and classification performance. From the tabulated results it is observed that the number of features get reduced when number of iterations increased and also number of MFCC coefficients increased. Ant Colony Optimization is able to select the more informative features without losing the performance. Based on our experiment, over **71%** accuracy can be achieved for recognizing **anger**, over **50%** accuracy for identifying **happy**, and over **87%** accuracy can be achieved for recognizing **normal** and about **57%** accuracy can be achieved for recognizing **sad** emotions. Thus speech emotion recognition using the combination of mfcc with ACO gives approximately 66% accuracy.

Based on this study, we can plan to develop an automatic emotion recognizer, which can help people who have difficulties in understanding and identifying emotions to improve their social and interaction skills.Emotion composition and how to extract more distinctive features for different types of emotions should be studied in the future. Future work would be integrating speech emotion recognition with facial emotion recognition for increased accuracy. Brain-Computer interface systems based on emotion recognition can be a scope of work in the future.

References

[1] Dr. C. Sunil Kumar, C.N Ravi and J. Dinesh, **"**Human Face Recognition and Detection system with Genetic and Ant Colony Optimization Algorithm", IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 16, Issue 4, Ver. VII (Jul – Aug. 2014), PP 11-15.

[2] Anjali Shelke, Sonali Joshi, "Design of Human Emotion Recognition System from Speech Using Particle Swarm Optimization ",IJCAT Feb 2014.

[3] Namrata Dave, "Feature Extraction Methods LPC, PLP and MFCC in Speech Recognition", International Journal For Advance Research In Engineering And Technology, Volume 1, Issue VI, July 2013.

[4] Dipti D. Joshi, Prof. M. B. Zalte**, "**Speech Emotion Recognition: A Review**",** IOSR Journal of Electronics and Communication Engineering (IOSR-JECE) ISSN: 2278-2834, ISBN: 2278-8735. Volume 4, Issue 4 (Jan. - Feb. 2013), PP 34-37.

[5] C. Poonkuzhali, R. Karthiprakash, Dr. S. Valarmathy, M. Kalamani, "An Approach To Feature Selection Algorithm Based On Ant Colony Optimization For Automatic Speech Recognition " IJAREEIE, Vol. 2, Issue 11, November 2013.

[6] Shivanker Dev Dhingra, Geeta Nijhawan, Poonam Pandit, "Isolated Speech Recognition Using MFCC And DTW", IJAREEIE, Vol. 2, Issue 8, August 2013.

[7] S. Ramakrishnan, **"**Recognition of Emotion from Speech: A Review, Speech Enhancement, Modeling and Recognition- Algorithms and Applications**"**, Dr. S Ramakrishnan (Ed.), ISBN:978-953-51-0291-5, InTech, Available from: http://www.intechopen.com/books/speech-enhancement-modeling-and-recognition-algorithms and- applications/recognition-of-emotion-from-speech-a-review 2012.

[8] Vaishali M. Chavan, V.V. Gohokar," Speech Emotion Recognition by using SVM classifier", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249–8958, Volume-1, Issue-5, June 2012.

[9] Dilbag Singh, "Human Emotion Recognition System ", I. J. Image, Graphics and Signal Processing, 2012, pp 50-56. http://www.mecs-press.org

[10] Md.Afzal Hossan, Sheeraz Memon, Mark A Gregory, "A Novel Approach for MFCC Feature Extraction", School Electrical and Computer Engineering, RMIT University, Melbourne, Australia. Source: IEEE Xplore January 2011 https://www.researchgate.net/publication/224217606.

[11] Jasmina Novakovica, Milomir Minica, Alempije Veljovic, "Classification Accuracy of Neural Networks with PCA in Emotion Recognition",Theory and Applications of Mathematics & Computer Science, 2011.

[12] Dr. S. Srinivasa Rao Madane, S. Venkatesan, "Face Detection by Hybrid Genetic and Ant Colony Optimization Algorithm", International Journal of Computer Applications (0975 – 8887) Volume 9– No.4, November 2010.

[13] Firoz Shah.A, Raji Sukumar .A, Babu Anto. P, "Automatic Emotion Recognition from Speech using Artificial Neural Networks with Gender- Dependent Databases", International Conference on Advances in Computing, Control, and Telecommunication Technologies. © 2009 IEEE.

[14] W. N. Widanagamaachchi and A. Dharmaratne, **"**Emotion Recognizer :A neural network approach**"**, Proceedings of 9th International Conference on Intelligent System Design and Applications, 2009.

[15] Keshi Dai, Harriet J. Fell, and Joel MacAuslan, "Recognizing Emotion In Speech Using Neural Networks", College of Computer and Information Science, Northeastern University, Boston, MA, USA, 2008.

[16] Kamran Soltani, Raja Noor Ainon **,** "Speech Emotion Detection Based On Neural

Networks", Faculty of Computer Science and Information Technology, University of Malaya Kuala Lumpur, Malaysia1-4244-0779-6/2007 IEEE.
[17] A book on "Ant Colony Optimization", by Marco Dorigo and Thomas Stutzle.