International Journal of Research in Advent Technology, Vol.3, Issue 1, January 2015
E-ISSN: 2321-9637

Human Emotion Recognition from Speech Using SVM Classifier

Jayant Mankar[1], Anjali shelke[2],Roshani Zanzade[3], Chaitalee Choudhari[4], Payal Kirme[5]
Electronics Engg.[1],Asst. Professor, SRMCEW, Nagpur
Electronics & Telecomm.Engg[2],Asst. Professor, SRMCEW, Nagpur
Electronics & Telecomm.Engg[3-8],Schlor, SRMCEW, Nagpur
Email:sushjay@rediffmail.com[1], roshanizanzade@gmail.com[3],cchoudhari9@gmail.com[4]

Abstract- Speech emotion recognition (SER) is a current research topic in the field of human computer interaction (HCI) with a wide range of application. Speech is considered as a powerful means to communicate with intension and emotion. In this paper the emotional state of a human being identified from his or her speech signal. The emotions from speech signal are recognized by considering the features of speech signal. The speech features such as pitch, speaking rate, voice quality, intensity and Mel frequency cepstral coefficients (MFCC) are extracted from speech signal. Support vector machine (SVM) is used as a classifier to classify different emotional state such as anger happiness, sadness, neutral, fear. SVM is simple and efficient algorithm which has a very good classification performance compared to other classifiers.

Index Terms-Emotion recognition, Mel frequency cepstrum coefficient (MFCC), Support Vector Machine.

1. INTRODUCTION

Speech emotion recognition is an important part in emotion recognition. Affective computing is concerned with emotion and machines. Human computer interfaces will become more efficient. Accurate detection of emotion from speech has more benefits for the design more natural. Human machine speech interfaces or for the extraction of useful information from large quantities of speech data. In human interaction there are many ways in which information is exchanged. A speech messages in which people express ideas or communicate has a lot of information that is interpreted implicity. Human emotion recognition from speech of speaker is used to achieve a more natural interaction between humans & machine. Another important aspect to be considered in emotional speech recognition is the data base used for training. In this paper the emotions from the speech signal are recognized by considering the feature of the speech segment by using SVM classifier while classifying different emotional, several features like MFCC (Mel Frequency Cepstral Coefficient)is used.

2.HUMANSPEECH EMOTION RECOGNITION SYSTEM

In the human speech recognition system contains five modules are Emotional speech database input, feature extraction, training, classification and recognized emotion output. The feature selection module is a part of SVM classifier.

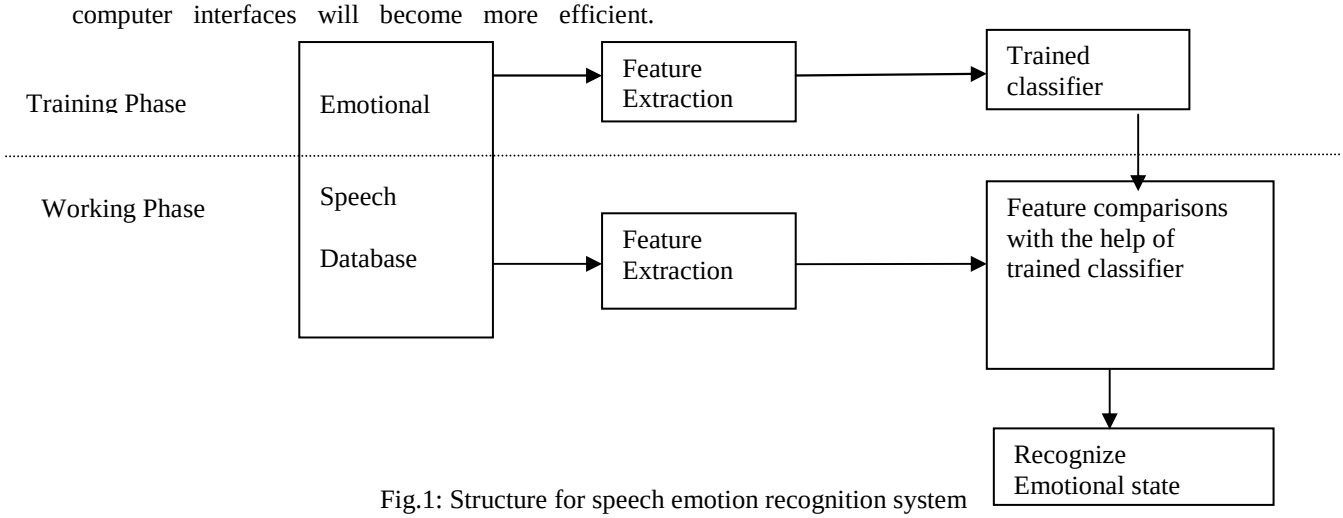

Fig.1: Structure for speech emotion recognition system

2.1. Emotional Speech Database

The basic files are speech signal gives the basic framework of the emotional speech recognition. The emotional speech input already present in the standard database. The input is taken from the database and from the sample of speech after that features are extracted. The emotional speech database used in this study is the Danish Emotional Speech (DES). The speech is express in five emotional states: anger, happy, neutral, sad and fear. In this study, the speech signal was re-sampled to 16KHz, and the silence segments at the beginning and the end of the speech were cut out artificially.

2.2. *Feature Extraction*

In feature extraction process, the features are extract from the speech such as energy, pitch, formant's frequencies etc. all these are prosodic features. In general prosodic features are primary indicator of speaker's emotional state. Here MFCC i.e. Mel frequency Cepstral coefficient is usedas a feature extraction process as shown in fig.

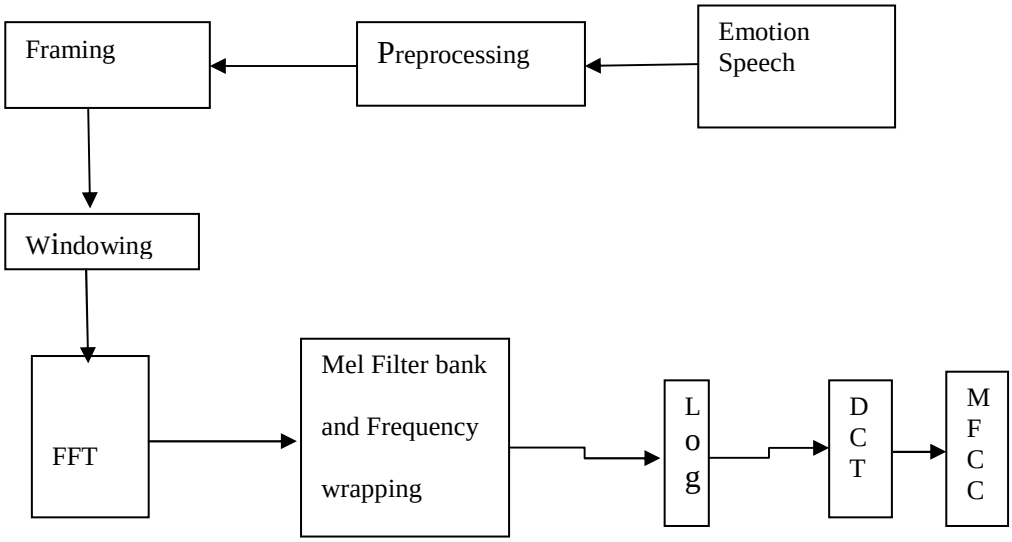

Fig.2: Block Diagram of Speech Emotion Recognition System Feature Extraction

Pre-processing: In this process noise will be removed speech by soft thresholding and the silent part of the signals do not carry any useful information those part including leading & trailing edges are eliminated by thresholding the energy of the signal.

Framing: It is the process of segmenting the speech signal obtained from analog to digital conversion into the small frame with the time length within the range of 20-40 ms.

Windowing: It is used to window each individual frame, in order to minimize the signal discontinuities at the beginning and the end of each frame.

FFT(Fast Fourier Transform): algorithm is used for evaluating the frequency spectrum of speech. It used to convert each frame of N-sample from the time domain into the frequency domain.

MEL Filterbank & Frequency Wrapping: The melfilterbank consist of overlapping triangular filter with the cut-off frequencies determine by the center frequency of the two adjacent filters and it fixed the bandwidth on the mel scale.

Log: In this logarithm simply the multiples of frames is converted into addition of the frame.

Take Discrete Cosine Transform: It is used to orthogonalise the filter energy vectors. Because of this orthogonalisation step, the information of the filter energy vector is compacted into the first number of components and shortens the vector to number of components.

Emotion characteristics	Joy	Anger	Sadness	Fear
Pitch range	High	High	Low	High
Pitch variance	High	Very High	Low	Very High
Intensity mean	High	High	Low	High
Speaking rate	High	Low-Male High – Female	High – Male Low-Female	High

Table.1: Feature Extracted from Speech

2.3. *SVM Training and Classification*

A support vector machine is a supervised machine learning algorithm used for data classification and estimating the relationship between variables. It's"supervised" algorithm because there an initial training phase involved where you feed the algorithm data that has already been classified .After this initial phase is completed, future data sets given to the algorithm can be classified with n minimal human intervention already been classified. After this initial phase is completed, future data sets given to the algorithm can be classified with no or minimal human intervention training process of SVM. The result can be improved if all the above features used properly. In this project the multiclass SVM classifier use with some inputs of different emotional states.

Now once we have different features available with us we can take use of the classifier todistinguish between the different emotional states. First we train the classifier with some input of different emotional states. After training the classifier, we can use it for recognizing the new given input.

The process of SVM training contains labelling of extracted features and training SVM. In the training process, every extracted feature has to assign an associated class label. The SVM is trained according to this labeled feature. The SVM kernel functions are used in the training process of SVM. The result can be improved if we use all the above features properly.

4. CONCLUSION

In many learning algorithms, we can only do linear classification only using a straight line by separation of data points. But there are algorithms that support vector machines that can also do non-linear classification using a Kernel method.

REFERENCES

[1] Vaishali M Chavan, V.V. Gohkar , "Speech Emotion Recognition by using SVM Classifier", March 2012.

[2] Kamran soltani, Raja Noor Ainon , "Speech emotion detection based on neural network" 2007.

[3] S. Ramakrishnan , Department of Information Technology "Recognition of Emotion from Speech"

[4] BhoomikaPanda andSanghamitraMohanty, "Use of SVM classifier and MFCC in speech emotion recognition system", March 2012.

[5] D.Ververid is, and C. Kotropoulos, "Automatic speech classification to five emotional states based on gender information, proceedings of the EUSIPCO2004 Conference, Austria, 341-344, Sept. 2004.

[6] Ralf Kompe and JM Pardo, "Emotional space improves emotion recognition", in Proc. ICSLP 2002.USA, 2002.

[7] Kaxuhiko Takahashi, RyohciNakatsu and J.Nicholson, "Emotion Recognition in Speech Using Neural Network", Dec. 2000.

[8] Proceedings of the Fourth International Conference on Machine Learningand Cybernetics, Guangzhou, 18-21 August 2005, "SPEECH EMOTIONRECOGNITION BASED ON HMM AND SVM" by YI-LIN LIN,GANG WEI.

[9] "HIDDEN MARKOV MODEL-BASED SPEECH EMOTIONRECOGNITION" by BjörnSchuller, Gerhard Rigoll, and Manfred Lang,Institute for Human-Computer Communication, 2003 IEEE.

[10] "Timing Levels in Segment-Based Speech Emotion Recognition", byBjörnSchuller and Gerhard Rigoll of Institute for Human-MachineCommunication in 2006.

[11] "A NEURAL NETWORK APPROACH FOR HUMAN EMOTIONRECOGNITION IN SPEECH", by Muhammad WaqasBhatti, YongjinWang and Ling Guan in 2006.
[12] "Comparison Between Fuzzy and NN Method for Speech EmotionRecognition", by Aishah Abdul Razak, Ryoichi Komiya, MohamadIzaniZainalAbidin, in 2005 IEEE.
[13] 2005 IEEE International Workshop on Robots and Human InteractiveCommunication, "Fuzzy Emotion Recognition in Natural SpeechDialogue", by AnjaAustermann, Natascha Esau, Lisa Kleinjohann andBernd Kleinjohann.
[14] "Emotion Recognition on the Basis of Human Speech", by ZygmuntCiotaTechnical University of Lodz, Department of Microelectronics andComputer Science.
[15] SICE-ICASE International Joint Conference 2006, Oct. 18-21, 2006 inBexco, Busan, Korea, "Robust Speech Emotion Recognition Using LogFrequency Power Ratio", by Kyung-Hak Hyun, Eun-Ho Kim andYoon-KeunKwak.
[16] "Speech Emotion Recognition Based on Rough Set and SVM", by JianZhou, Guoyin Wang, Yong Yang, Peijun Chen, 2006 IEEE.
[17] "GMM SUPERVECTOR BASED SVM WITH SPECTRAL FEATURESFOR SPEECH EMOTION RECOGNITION", by Hao Hu,Ming-XingXu, and Wei Wu, Tsinghua University, Beijing, 2007 IEEE.
[18] "SPEECH EMOTION RECOGNITION USING GAUSSIAN MIXTUREVECTOR AUTO REGRESSIVE MODELS" byMoataz M. H. El Ayadi,Mohamed S. Kamel, and FakhriKarray, Pattern Analysis and MachineIntelligence Lab, Electrical and Computer Engineering, University ofWaterloo,2007 IEEE.
[19] IMACS multi-conferenceon "Computational Engineering in SystemApplications"(CESA), October 4-6, 2006, Beijing, China."Emotion-detecting Based Model Selection for Emotional SpeechRecognition", by Y. C. Pan, M. X. Xu, L. Q. Liu, P. F. Jia.
[20] "Speech Emotion Recognition in E-learning System Based on AffectiveComputing" by Wu Li, Yanhui Zhang, Yingzi Fu, Third InternationalConference on Natural Computation (ICNC 2007).
[21] 2007 IEEE International Conference on Control and Automation FrBP-27Guangzhou, CHINA - May 30 to June 1, 2007. Research on "SpeechEmotion Recognition in E- learning By Using Neural Networks Method",by Qian Zhang, Yan Wang, Lan Wang and Guoqiang Wang, 2007 IEEE.
[22] 16th IEEE International Conference on Robot & Human InteractiveCommunication August 26 - 29, 2007 / Jeju, Korea, "Speech EmotionRecognition Using Eigen-FFT in Clean and Noisy Environments", EunHo Kim , Kyung HakHyu, Soo Hyun Kim and Yoon KeunKwak, 2007IEEE.