To recognizes the SAD emotion with in a speech using SVM and ANN

Puneet Thapar
Student- M-Tech
Department of Computer Science and Engineering
D.A.V Institute of Engineering & Technology.
Jalandhar City, India.

Dinesh Kumar
Associate Professor and Head
Department of Information Technology
D.A.V Institute of Engineering & Technology.
Jalandhar City, India.

Abstract

In this paper we approach to speech emotion recognition using support vector machines (SVM) and artificial neural network (ANN). This is a hot research topic in the field of machine learning approach. Emotion is totally depends on speaker and utterance (phrase). In this paper SVM is used as for training purpose and ANN is used as for classification. In this paper we use SVM and ANN to classify emotion such as sad.

KeyTerms— Speech Emotion Recognition,SVM, ANN, classification.

1. Introduction

Speech emotion recognition is an important part in emotion recognition. Accurate detection of emotion from speech has clear benefits for the design of more natural human – machine speech interfaces or for the extraction of useful information from large quantities of speech data. Affective human computer interaction has been the focus of artificial intelligence research for several years now, and the research has moving ahead from the simple information exchange between human and computer towards the affective communication.Affective human computer interaction technology could be widely applied in virtual reality, especially in the field of entertainment and games. Besides, the virtual human and psychiatric aids are the further application prospects for affective human computer interaction. Making computer recognize the emotion of human being is the foundation of affective human computer interaction. The main carriers of human emotion, including facial expression, posture and speech, are the primary channels for computer to recognize human's emotion. The accustomed way for speech emotion recognition is to distinguish between a defined set of discrete emotions. Manifold classifiers have been employed in this research. Speech is considered as a powerful mode to communicate with intentions and emotions. In the recent years, a great deal of research has been done to recognize human emotion using speech information [1], [2]. Many researcher explored several classification methods including the Neural Network (NN), Gaussian Mixture Model (GMM), Hidden Markov Model (HMM), Maximum Likelihood Bayes classifier

(MLC), Kernel Regression and K-nearest Neighbors (KNN), Support Vector Machine (SVM) [3], [4].

A Berlin Emotional database [5] emotional speech signal files are used for feature extraction and training SVM. The Berlin database of emotional speech was recorded at the Technical University, Berlin. The database German contains speech with acted emotions in language. It contains 493 utterances of 10 professional actors' five males and five females who spoke 10 sentences with emotionally neutral content in 7 different emotions. The emotions were wut (anger), langeweile (boredom), ekel (disgust), angust (fear), freude (happiness), trauer (sadness) and neutral emotional state.

There are various applications of Speech Emotion Recognitionlike Emotion Recognition software for call center it is a full fledgeprototype of an industrial solution for computerized callcenter and can help in detection of the emotional state intelephone call center conversations to provide feedback to anoperator or a supervisor, psychiatric diagnosis, intelligent toys,lie detection, learning environment, educational software, for monitoring purposes.

2. SPEECH EMOTION RECOGNITION SYSTEM

Speech is the primary means of communication between human. Speech refers to the processes associated with the production and perception of sounds used in spoken language. A number of academic disciplines study speech and speech sounds, including acoustics, psychology, speech pathology, linguistics, cognitive science, communication studies, otolaryngology and computer science.

It has been proved that both statistical and temporal features of the acoustic parameters affect the emotion recognition of speech [5]. In this paper, SVM are used to deal with the temporal features, getting likelihood probabilities and state segmentations. Many practical applications proved there is often some physical significance attached to the states of SVM [6]. Finally, the distortions and likelihood probabilities derived from SVMs are combined to be the input of ANN, and ANN is used to classify emotions. Figure 1 illustrates the structure of the speech emotion recognition system developed in this paper.

The basic speech emotion recognition system is shown in Figure 1. This system consists of three main steps: audio segmentation, featureextraction and processing, emotion classification. Here SVM used to audio segmentation, Feature extraction and ANN is used to classify according to the training data generated by SVM.

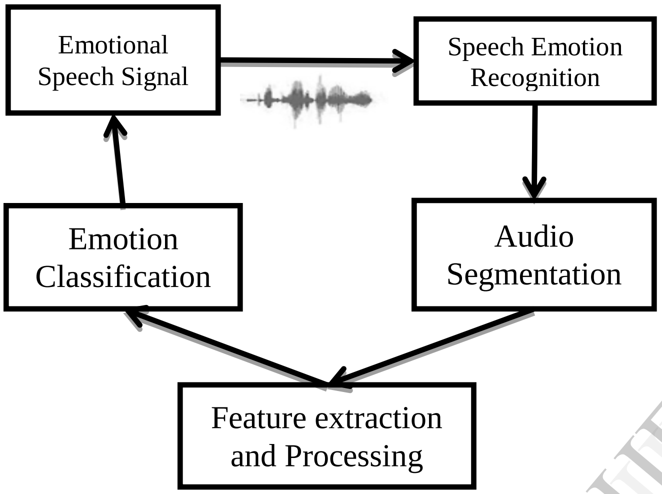

Figure 1: Speech Emotion Recognition System

The first step in this system is to segment the incoming speech signal into the meaningful units that can serve as emotional classification units, such as utterance.

The goal of feature extraction and processing is to the extract relevant features from speech signals with respect to emotions, and to reduce the size of the speech feature set to fewer dimensions. The widely used acoustic features indicating human emotion expression are prosody features and voice quality features [5-9].

Emotion classification, lastly, maps feature vectors onto emotion classes through learning by data examples. The representative emotion classification methods are linear discriminant classifiers (LDC), K-nearest-neighbor (KNN), artificial neural network (ANN) and support vector machines (SVM), etc. In this work, we use an ANN classifier to perform emotion classification.

3. Feature Extraction

To select suited features carrying information aboutemotion is necessary for emotion recognition. Studyon emotion of speech indicates that pitch, energy,formant, Mel prediction cepstrum coefficient (MPCC)and linear prediction cepstrum coefficient (LPCC)are effective features to distinguish certainemotions[5, 7-8]. Feature extraction is based onpartitioning speech into frames. For each frame, sixcommon features, including pitch, amplitude energy,logenergy, 10-order LPCC, 12-order MFCC andformant, are extracted. These features form thecandidate input feature sequences with their first andsecond derivatives.

- **Energy and Related Features**

  The Energy is the basic and most important feature in speech signal. In order to obtain the statistics of energy feature, we use short-term function to extract the value of energy in each speech frame. Then we can obtain the statistics of energy in the whole speech sample by calculating the energy, such as mean value, max value, variance, variation range, contour of energy [2].

- **Pitch and Related Features**

  The pitch signal is another important feature in speech emotion recognition. The vibration rate of vocal is called the fundamental frequency F0 or pitch frequency. The pitch signal is also called the glottal wave-form; it has information about emotion, because it depends on the tension of the vocal folds and the sub glottal air pressure, so the mean value of pitch, variance, variation range and the contour is different in seven basic emotional statuses.

- **MFCC and MEDC features**

  Mel-Frequency Cepstrum coefficients is the most important feature of speech with simple calculation, good ability of distinction, anti-noise. MFCC in the low frequency region has a good frequency resolution, and the robustness to noise is also very good.

  MEDC extraction process is similar with MFCC. The only one difference in extraction process is that the MEDC is taking logarithmic mean of energies after Mel Filter bank and Frequency wrapping, while the MFCC is taking logarithmic after Mel Filter bank and Frequency wrapping. After that, we also compute 1st and 2nd difference about this feature [5].

- **Linear Prediction Cepstrum Coefficients**

  LPCC embodies the characteristics of particular channel of speech, and the same person with different emotional speech will have different channel characteristics, so we can extract these feature coefficients to identify the emotions contained in speech. The computational method of LPCC is usually a recurrence of computing the linear prediction coefficients (LPC) [5].

4. SVM Classification

SVM, a binary classifier is a simple and efficient computation of machine learning algorithms, and is widely used for pattern recognition and classification problems, and under the conditions of limited training data, it can have a very good classification performance compared to other classifiers [9]. The idea behind the SVM is to transform the original input set to a high dimensional feature space by using kernel function. Therefore non-linear problems can be solved by doing this transformation.

The SVM is train according to labeled features. The SVM kernelfunctions are used in the training process of SVM. Binaryclassification can be viewed as the task of separating classes infeature space.

5. ANN Used As Emotion Recognizer

Since ANN possesses excellent discriminate power and learning capabilities, the hybrid classification in this paper takes advantage of a one hidden layer and 9 hidden nodes net to classify emotions. The input of the ANN consists of distortions and likelihood probabilities, while the output is the assumed emotion.

6. Experimental Study

Then consider the various performance measures as precision, recall, F Measure and Accuracy used for machine learning. On the basis of these parameters we calculate the overall accuracy of speech emotion recognition system.

The confusion matrix for SAD emotion is shown as following: -

TABLE 1
Confusion Matrix for Sad Emotion

Emotion	Sad	Other
Sad	85.2	22.54
Other	14.8	77.46

Table 1 shows confusion matrix for a sad emotion of implemented SVM/ANN for Berlin emotion speech utterance using one-to-one multiclass method. Then from this confusion matrix value of Precision is 85.2%, Recall is 79.07% and F Measure is 82.01%. The overall recognition accuracy of sad emotion for Berlin emotions database is 81.33% by using Linear Kernel Function.

Then the Result analysis on the basis of these performance measure is calculated and shown in Fig. 2

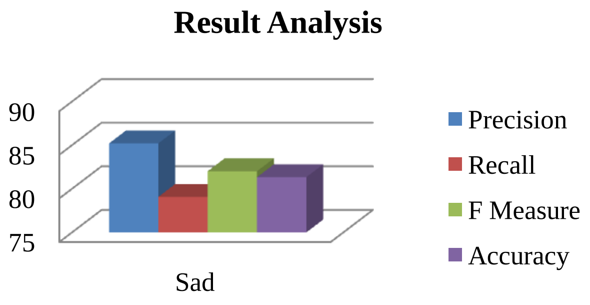

Fig. 2 Result Analysis for Each Emotion

7. Conclusion and Future Scope

In this paper, we have studied on emotion speech recognition by means of SVM / ANN, and we believe that SVM makes significant impact on speech emotion recognition. Furthermore, we use SVM and ANN as a speech emotion recognition classifier to classify only SAD emotion from speech input. The performance of the hybrid classification and isolated SVM were collected by experiment using speech copra (berlin database). From the experiment we analyses that recognition by the Hybrid classification has been proved more effective than isolated SVM.

Our future work will further explore the possibility to recognize more emotion states as Fear, Angry and Happy, etc. to increase the recognition rate.

8. REFERENCES

[1] Aastha Joshi and RajneetKaur, “A Study of Speech Emotion Recognition Methods”, International Journal of Computer Science and Mobile Computing (IJCSMC), Vol. 2, Issue.4, April 2013, pp.28 – 31.

[2] Christopher J. C. Burges, “A tutorial on support vector machines for pattern recognition”, Springer, Data Mining and Knowledge Discovery, Vol. 2, Issue. 2, 1998, pp.121-167.

[3] C.W Hsu, C.C. Chang and C.J. Lin, “A Practical Guide to Support Vector Classification”, Technical Report, Department of Computer Science & Information Engineering, National Taiwan University, Taiwan, 2003.

[4] L.R. Rabiner and B.H. Juang, “Fundamentals of speech recognition”, Englewood Cliffs, NJ: Prentice-Hall, 1993.

[5] Felix Burkhardt, Astrid Paeschke, Miriam Rolfes, Walter F. Sendlmeier and Benjamin Weiss, “A Database of German Emotional Speech”, Proceedings of Interspeech, Lissabon, Portugal. 2005, pp. 1517-1520.

[6] Corinna Cortes and Vladimir Vapnik, "Support vector machine", Machine learning, Vol. 20, 1995, pp. 273-297.

[7] PranitaN.Kulkarni and Prof.D.L.Gadhe, "Comparison between SVM & Other Classifiers for SER", International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Vol. 2, Issue 1, January- 2013, pp. 1-6.

[8] D. Ververidis, C. Kotropoulos, and I. Pitas, "Automatic emotional speech classification", in Proc. 2004 IEEE Int. Conf. Acoustics, Speech and Signal Processing, Vol. 1, Montreal, May 2004, pp. 593-596.

[9] Yixiong Pan, PeipeiShen and LipingShen, "Speech Emotion Recognition Using Support Vector Machine", International Journal of Smart Home, Vol. 6, April 2012, pp. 101-108.

[10] J.H. Tao and Y.G. Kang, "Features importance analysis for emotional speech classification", In Proceedings of lecture notes in computer science, Springer, 2005, pp.449-457.

[11] R.Cowie and E. Douglas-Cowie, "Automatic statistical analysis of the signal and prosodic signs of emotion in speech", In Proc. 4th Int. Conf. Spoken Language Processing, Philadelphia, IEEE, Vol. 3, 1996, pp.1989-1992.

[12]Moataz M. H. El Ayadi, Mohamed S. Kamel, and FakhriKarray, "Speech Emotion Recognition Using Gaussian Mixture Vector Autoregressive Models", IEEE International Conference on Acoustics, Speech and Signal Processing, Pattern Analysis and Machine Intelligence Lab, Electrical and Computer Engineering, University of Waterloo, Vol.4, 2007, pp. 957-960.

[13]Vaishali M. Chavan and V.V. Gohokar, "Speech Emotion Recognition by using SVM-Classifier", International Journal of Engineering and Advanced Technology (IJEAT), ISSN: 2249 – 8958, Vol. 1, Issue. 5, June 2012, pp.11-15.

[14]Xia Mao, Bing Zhang and Yi Luo, "Speech Emotion Recognition Based On A Hybrid of HMM/ANN", Proceedings of the 7th WSEAS International Conference on Applied Informatics and Communications, Athens, Greece, August 2007, pp. 367-370.

[15]M. D. Skowronski and J. G. Harris, "Increased MFCC Filter Bandwidth for Noise-Robust Phoneme Recognition", IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Florida, Vol. 1, May 2002, pp.1-4.

[16] R. Huber, A. Batliner, J. Buckow, E. Noth, V. Warnke, and H. Niemann, "Recognition of emotion in a realistic dialogue scenario," in the International Conference on Spoken Language Processing (ICSLP 2000), Beijing, China, Vol. 1, 2000, pp. 665–668.

[17]S.McGilloway, R. Cowie, E. Douglas-Cowie, S. Gielen, M. Westerdijk, and S. Stroeve, "Approaching automatic recognition of emotion from voice: A rough benchmark," in the ISCA Workshop on Speech and Emotion, Belfast, Northern Ireland, 2000, pp. 200–205.

[18] T. S. Polzin and A. Waibel, "Emotion-sensitive human-computer interfaces," in the ISCA Workshop on Speech and Emotion, Belfast, Northern Ireland, 2000.