International Journal of Electronics and Communication Engineering (IJECE)
Vol.1, Issue 1 Aug 2012 41-45

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

RECOGNIZING OF EMOTIONS FROM SPEECH BY ANN

GYANENDRA PRATAP & RAKESH KUMAR SHARMA

NIT Kurukshetra, India

ABSTRACT

This research is to recognize the state of emotions in speech using Hopfield technique. Speech is uttered by 30 persons who are the speakers selected for this project and being given by five sentences. The Hopfield Neural Network (HFNN) is an algorithm. The primary objective of this project is to develop ANN model to classify the collected voice data into two emotional states, happiness and anger. Several approach and methodology have been introduced in order to achieve the objectives. This project needs more revision and studies to obtain the accuracy in recognizing the emotion through speech.

KEYWORDS*:* Emotions, LPC, Neuralnetwork (HFNN).

INTRODUCTION

With the pace of modern life getting faster, mankind suffers more emotional problems than ever before. The increasingly well-developed technology makes us feel convenient on one hand at the same time it's prone to cause sense of loneliness on the other hand. We use computer every day, and we just treat it as a machine but not a friend because it cannot understand our emotions.

The aim of this research is to develop and evaluate Artificial Neural Network (ANN) model for recognizing emotion in speech expressed by using HopField Neural Network (HFNN). To achieve our objectives we decided to have two stages: research and development. The objectives of the first stage are the following: to learn how well people recognize emotions in speech, to find out which features of speech signal could be useful for emotion recognition, and explore different mathematical models for creating reliable recognizers. If the results of the first stage will be promising then we will start the second stage which objective is to create a real-time recognizer for call center applications.

EMOTION RECOGNIZER

The design of emotion recognizer basically involves different stages such as signal processing, emotion parameter extraction and emotion recognition.

Speech Emotion Recognition System (SER)

SER is tackled as a pattern recognition task. This implies that the following stages are present: feature extraction, feature selection, classifier choice, and testing shown in fig.1. This sequence is called the pattern recognition cycle.

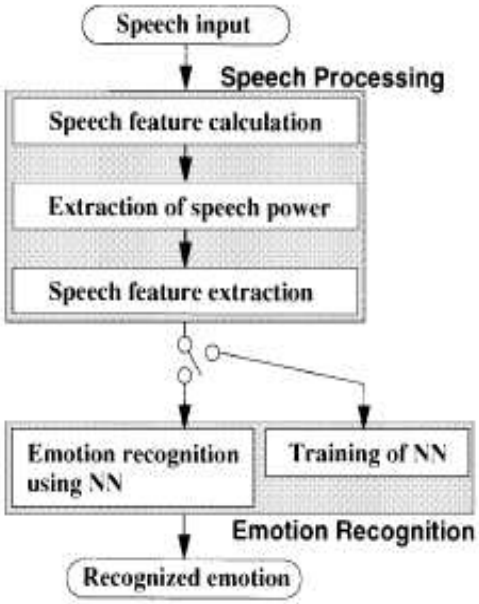


Fig.1. Processing Flow for Recognizing of Emotions

Signal Processing

Signal processing involves digitalization and potentially acoustic preprocessing like filtering, as well as segmenting the input signal into meaningful units.

FeatureExtraction

This stage, involves the speech processor, deals with the emotion features selection and extraction algorithm.

Feature selection (also known as subset selection) is a process commonly used in machine learning, wherein a subset of the features available from the data is selected for application of a learning algorithm. The best subset contains the least number of dimensions that most contribute to accuracy; we discard the remaining, unimportant dimensions. This is an important stage of preprocessing and is one of two ways of avoiding the curse of dimensionality. Linear predictive coding (LPC) is a popular technique for speech compression and speech synthesis.

Speech analysis and synthesis with Linear Predictive Coding (LPC) exploit the predictable nature of speech signals. Cross-correlation, autocorrelation, and auto covariance provide the mathematical tools to determine this predictability.

The features represent the global statistics, which means that values were estimated over the whole utterance. This choice was made, due to the fact that in the literature global statistics is generally thought to be more suitable for SER.

Feature extraction is done with the praat script. On average, the script takes 1.08 to 12.7 sec. to extract 116 features from a speech sample of 1.08 to 12.7 sec of duration (with The modules of the SER

engine are graphically depicted .Given a database with emotional speech, the SER proceeds as follows. The input is the files with speech signal. Depending on the original database format, where necessary, auxiliary database preprocessing scripts convert the files into the required file format and change the speech file format. Then the feature extraction script extracts the features.

The praat script extracts 116 features. The features are statistical functions derived from acoustic parameters. The acoustic parameters are pitch, intensity, formants andharmonicity.The speech features must be extracted from each utterance for the emotion recognition training or testing.

Classification

Many classifiers have been tried for SER, and after Weka has appeared it has become easy and straight forward. The most frequently used are Support Vector Machines and Neural Networks. A rapidly evolving area in pattern recognition research is the combination of classifiers to build the so-called classifier ensembles. For a number of reasons (ranging from statistical to computational and representational aspects) ensembles tend to outperform single classifiers.

Training

While training and testing, over fitting should be avoided. Regarding this problem, there are two ways to choose the classifier correctly: just chose the classifier that performs best in the cross-validation test mode (in this work we always use 10-fold cross-validation) or pick the classifier with the smallest MDL. (It can be shown theoretically that classifiers designed with the MDL principles are guaranteed to converge to the ideal or true model in the limit of more and more data.) There are two approaches to training - supervised and unsupervised.

Testing

After training, the network is tested with both open and closed testing. In closed testing, the network is tested using the same set of data on which it was trained. In open testing, the network is tested using the remainder of the data which was not used for training.

To optimize the topology of the sub-neural networks, we carried out the training of the network and the closed testing with a small database (30 subjects). Sub-neural networks were optimized in the same way, but the sub-neural networks were grouped into two sets in order to improve performance: (joy, teasing, fear and neutral) and (sadness, disgust, anger, surprise).

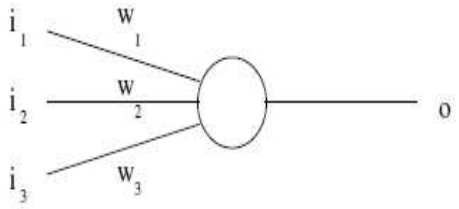
ARTIFICIAL NEURAL NETWORKS

Hopfield networks are constructed from artificial neurons. These artificial neurons have N inputs. With each input i there is a weight w_i associated. They also have an output. The state of the output is maintained, until the neuron is updated. Updating the neurons is done by following operations.

• The value of each input, xi is determined and the weighted sum of all inputs, $\sum_i w_i x_i$ is calculated.

• The output state of the neuron is set to +1 if the weighted input sum is larger or equal to 0. It is set to -1 if the weighted input sum is smaller than 0.

• A neuron retains its output state until it is updated again.



An artificial neuron as used in a Hopfield network.

Operations written as,

$$O = \begin{cases} 1, & \sum i\ wi\,xi \geq 0 \\ -1, & \sum i\ wi\,xi \leq 0 \end{cases}$$

A Hopfield network is a network of N such artificial neurons, which are fully connected. The connection weight from neuron j to neuron i is given by a number w_{ij}. The collection of all such numbers is represented by the weight matrix W, whose components are $w_{ij.}$

Now given the weight matrix and the updating rule for neurons the dynamics of the network is defined if we tell in which order we update the neurons. There are two ways of updating them,

Asynchronous: one picks one neuron, calculates the weighted input sum and updates immediately. This can be done in a fixed order, or neurons can be picked at random, which is called **asynchronous random updating.**

Synchronous: the weighted input sums of all neurons are calculated without updating the neurons. Then all neurons are set to their new value, according to the value of their weighted input sum.

CONCLUSIONS

In this research we will explored how well computers/ machines recognize emotions in speech. Several conclusions can be drawn from the above results. First, decoding of emotions in speech is complex process that is influenced by cultural, social, and intellectual characteristics of subjects. Second, pattern recognition techniques based on neural networks proved to be useful for emotion recognition in speech and for creating customer relationship management systems.

RESULTS

We got the maximum efficiency for anger emotion (60 %) .The overall efficiency for emotion recognition is 46 %. So we can say Results also reveal that LPC is a better choice as feature parameters for emotion classification than the traditional feature parameters.

A Neural network is able to recognize in a satisfying percent a set of emotions pronounced by different speakers, using LPC as input. The results obtained in this study demonstrate that emotion recognition in speech is feasible, and that neural networks are well suited for this task.

REFERENCES

1. Prof. Arun Kulkarni, Prof. Sujata Pathak,"Multimodal approaches for emotional features in speech: A survey", Proc .of EC2IT, KJSCE, Mumbai, pp. 155-160, March 2009.
2. Berlin Emotional Speech Database. http://pascal.kgw.tu-erlin.de/emodb/index-1024.html109
3. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W.Fellenz, and J. Taylor, "Emotion recognition in human-computer interaction," Signal Processing Magazine, IEEE, vol. 18, no. 1, Jan 2001
4. Aishah, A.R., Komiya, R., "A Preliminary Study of Emotion Extraction from Voice," National Conference on Computer Graphics and Multimedia (CoGRAMM'02), Malacca.
5. Rabiner, L.R. and Schafer, .W.(1978). Digital Processing of Speech
6. Signals, Prentice-Hall, Eaglewood Cliffs, NJ.