David McNeil 1876

Speech Recognition Using Linear Prediction Coding and Support Vector Machines

Background

Linear prediction coding (LPC) is a method of low bit rate speech encoding. LPC can be used to find filter coefficients for a time varying N^{th} order FIR filter. An excitation signal can then be applied to this filter to estimate the original signal. The filter coefficients act as a compressed version of the original audio. Individual words or phrases should have relatively unique sets of coefficients allowing these words or phrases to be distinguished by their filter coefficients. These filter coefficients can act as a feature space for training a classification algorithm. For this specific project, I would be using Support Vector Machines (SVMs). SVMs are a non-probabilistic binary linear maximum margin classifier suitable for complex classification problems such as speech or image recognition.

Problem Statement

Humans most notably interact using speech. However, human computer interfaces rarely use speech. The purpose of this project is to bridge that gap by implementing a method of speech recognition using LPC and SVMs.

Approach

The first step in the project is to record or obtain a database of speech clips which can be used as training data. The filter coefficients for these speech segments are then calculated and fed into the training portion of the SVM algorithm. A SVM is developed for each word in the classifiers vocabulary using a one-vs-all training method. New unclassified audio samples can be feed into the SVMs and the corresponding class predicted. If time allowed, I would like to develop an intuitive interface to allow for dynamic training of new speech samples and real time speech prediction. I plan on implementing this software in Matlab. Matlab provides excellent libraries for linear prediction coding and support vector machines.

Anticipated Results

Satisfying results would be the development of a testing and validation framework capable of proving that the LPC filter coefficients are a viable feature vector for training SVMs classifiers for speech recognition. Results would further include statistical data on prediction reliability. A prediction reliability of approximately 90% would be a stretch goal (this prediction rate is on par with commercial speech recognition systems and findings from published academic papers).

Citations

http://www.academia.edu/4874338/Maximizing_the_Speech_Recognition_Accuracy_Using_Linear_predictive_Coding_Guided_by_Maximizing_the_Speech_Recognition_Accuracy_Using_Linear_predictive_Coding

http://www.mathworks.com/help/signal/ref/lpc.html

http://arxiv.org/ftp/arxiv/papers/1305/1305.1145.pdf

http://www.slideshare.net/udaysaikia/lpc-for-speech-recognition

http://www.academia.edu/4874338/Maximizing_the_Speech_Recognition_Accuracy_Using_Linear_predictive_Coding_Guided_by_Maximizing_the_Speech_Recognition_Accuracy_Using_Linear_predictive_Coding

http://www.ipcsit.com/vol6/36-E091.pdf

http://www.seas.ucla.edu/~ingrid/ee213a/speech/vlad_present.pdf