Acoustic feature selection for automatic emotion recognition from speech

A sufficient number of training examples is the premise for most machine learning and data mining algorithms to work well. When there are only a few training examples, it is possible to have the problem of overfitting – model can have perfect performance on the training set but cannot generalize well on new examples. If a data set cannot fully cover the whole variable space then it is referred to as a small data set. In this sense the data sets for emotion recognition are small because the typical size is less than 1000 and the number of features is close to 100.

A machine learning framework for emotion recognition

Most existing work can be summarized into the following general procedure:

* Feature extraction stage: extract whole acoustic feature set from the original speech corpus and transform these features into an appropriate format for further processing
* Data preprocessing stage: select the most relevant subset of the whole candidate feature set or reduce the size of the speech data set into fewer dimensions
* Emotion recognition stage: apply machine learning methods on the processed speech data set from the previous stage to recognize emotional states in speech.