Generalization Problem in ASR Acoustic Model Training and Adaptation

Sadaoki Furui

Department of Computer Science, Tokyo Institute of Technology

Since speech is highly variable, even if we have a fairly large-scale database, we cannot avoid the data sparseness problem in constructing automatic speech recognition (ASR) systems. How to train and adapt statistical models using limited amounts of data is one of the most important research issues in ASR. This paper summarizes major techniques that have been proposed to solve the generalization problem in acoustic model training and adaptation, that is, how to achieve high recognition accuracy for new utterances. One of the common approaches is controlling the degree of freedom in model training and adaptation. The techniques can be classified by whether a priori knowledge of speech obtained by a speech database such as those spoken by many speakers is used or not. Another approach is maximizing "margins" between training samples and the decision boundaries. Many of these techniques have also been combined and extended to further improve performance. Although many useful techniques have been developed, we still do not have a golden standard that can be applied to any kind of speech variation and any condition of the speech data available for training and adaptation.