

Title	Study on Post-processing Method for HMM-based Phonetic Segmentation using Adaptive Neuro Fuzzy Inference system
Author(s)	董, 良
Citation	
Issue Date	2014-09
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/12262
Rights	
Description	Supervisor:Masato Akagi, School of Information Science, Master

Study on Post-processing Method for HMM-based Phonetic Segmentation using Adaptive Neuro Fuzzy Inference system

Liang Dong (1210215)

School of Information Science,
Japan Advanced Institute of Science and Technology

August 7, 2014

Keywords: Automatic speech segmentation, HMM, ANFIS.

Highly accurate and reliable speech segmentation is much more needed for the contemporary speech technology and research. Manual segmentation has been considered the most reliable and precise method to get the segments for speech corpus. In addition, it is used as the standard for the evaluation of automatic speech segmentation. However, manual segmentation is time-consuming and labor-intensive that can be performed only by expert phoneticians. In the age of big data, the shortcoming of manual segmentation is a fatal defect for large speech corpus. There is a strong demand for us to develop precise automatic speech segmentation.

Different kinds of approaches have been proposed for the task of automatic speech segmentation, such as: the detection of variations/similarities in spectral or prosodic parameters of speech, the template matching using dynamic programming or the synthetic speech and the discriminative learning segmentation.

Nowadays, the mainstream technology for automatic speech segmentation is called forced alignment. In this method, Hidden Markov Model (HMM) is commonly used to build a model for each phoneme with different number of states. Speech signal is exacted as a set of feature vectors by frame. The alignment of frames with phonemes is determined by finding

the most likely sequence of hidden states given the observed data and the acoustic model represented by the HMMs. The model can give the rough boundaries for each phoneme, but not accurate enough. The reported performance of traditional HMM-based forced alignment systems range from 80% to 89% agreement within 20ms compared to manual segmentation on the TIMIT corpus

Many methods have been proposed to improve accuracy of automatic speech segmentation within HMM framework, by refining the initial HMM-based segmentation. Some researchers were aware that the difference between HMM-based segmentation and expert phoneticians is the rules and knowledge for segmentation. Fuzzy logic provides a very straightforward way to introduce human knowledge(which is fuzzy by nature) in computer-based systems. So they simulated the process of manual speech segmentation by manually defining the fuzzy rules. The results show that it is a good way to improve the automatic segmentation. But the drawback is that the fuzzy rules need to be carefully designed and adjusted by experts. A suitable post-processing method is needed to solve this problem, which is the purpose of this study.

Adaptive neuro fuzzy inference system (ANFIS) is a Neuro-fuzzy system that uses the learning techniques of neural networks, with the efficiency of fuzzy inference systems. Compared with other learning methods, it has both the advantages of neural networks and fuzzy inference systems, and also have a better performance than others. Based on its advantages: simple implementation and good performance in learning nonlinear and fuzzy rules, it is very suitable to solve our problems.

In this study, we used adaptive neuro fuzzy inference system to compensate the arbitrariness in manual speech segmentation and the systematic segmentation errors produced by HMMs. It is divided into two steps: Firstly, context-independent HMMs was used to obtain the initial time marks. Secondly, a well trained ANFIS was used to refine the time marks. Two experiments were designed on TIMIT database, in order to achieve our purpose.

The results of the experiments show that the ANFIS used in our method significantly improved the accuracy of forced alignment within HMM framework. The proposed system achieved 92.08% agreement within 20 msec

for manual segmentation on the TIMIT corpus, comparing (86.25%) with a traditional HMM-based method. Which also indicates the effectiveness of the ANFIS for the purpose of this research. Moreover, our method is easy to be built and applied to other databases. For the future work, how to make the system effective and easy to be built can be the next topic.