INTEGRATE TEMPLATE MATCHING AND STATISTICAL MODELING FOR CONTINUOUS SPEECH RECOGNITION

Xie Sun

Dr. Yunxin Zhao, Dissertation Supervisor

ABSTRACT

In this dissertation, a novel approach of integrating template matching with statistical modeling is proposed to improve continuous speech recognition. Commonly used Hidden Markov Models (HMMs) are ineffective in modeling details of speech temporal evolutions, which can be overcome by template-based methods. However, template-based methods are difficult to be extended in large vocabulary continuous speech recognition (LVCSR). Our proposed approach takes advantages of both statistical modeling and template matching to overcome the weaknesses of traditional HMMs and conventional template-based methods.

We use multiple Gaussian Mixture Model indices to represent each frame of speech templates. The local distances of log likelihood ratio and Kullback-Leibler divergence are proposed for dynamic time warping based template matching. In order to reduce computational complexity and storage space, we propose methods of minimum distance template selection and maximum log-likelihood template selection, and investigate a template compression method on top of template selection to further improve recognition performance. Experimental results on the TIMIT phone recognition task and a LVCSR task of telehealth captioning demonstrated that the proposed approach significantly improved the performance of recognition accuracy over the HMM baselines, and on the TIMIT task, the proposed method showed consistent performance improvements over progressively enhanced HMM baselines. Moreover, the template selection methods largely reduced computation and storage complexities. Finally, an investigation was made to combine acoustic scores in triphone template matching with scores of prosodic features, which showed positive effects on vowels in LVCSR.