DYNAMIC LANGUAGE MODEL ADAPTATION USING LATENT TOPICAL INFORMATION AND AUTOMATIC TRANSCRIPTS (WedPmPO1)

Author(s) : Berlin Chen (National Taiwan Normal University, Taiwan)

Abstract : This paper investigates dynamic language model adaptation for Mandarin broadcast news recognition. A topical mixture model was presented to dynamically explore the long-span latent topical information for language model adaptation. The underlying characteristics and different kinds of model structures were extensively investigated, while their performance was verified by comparison with the conventional MAP-based adaptation approaches, which are devoted to extracting the short-span n-gram information. The fusion of global topical and local contextual information was investigated as well. The speech recognition experiments were conducted on the broadcast news collected in Taiwan. Both contemporary newswire texts and in-domain automatic transcripts were exploited in language model adaptation. Very promising results in perplexity as well as word error rate reductions were initially obtained.