

2pSCa9. Semi-supervised learning for speech and audio processing

Session: Tuesday Afternoon, Nov 01

Time: 4:00

Author: Mark A. Hasegawa-Johnson

Location: ECE Dept., Univ. of Illinois. Urbana, IL 61801

Author: Jui-Ting Huang

Location: ECE Dept., Univ. of Illinois. Urbana, IL 61801

Author: Xiaodan Zhuang

Location: ECE Dept., Univ. of Illinois. Urbana, IL 61801

Abstract:

Semi-supervised learning requires one to make assumptions about the data. This talk will discuss two different assumptions, and algorithms that instantiate those assumptions, for the tasks of acoustic modeling and pronunciation modeling in automatic speech recognition. First, the acoustic spectra corresponding to different phonemes overlap, but there is a tendency for the instantiations of each phoneme to cluster within a well-defined region of the feature space—a sort of “soft compactness” assumption. Softly compact distributions can be learned by an algorithm that encourages compactness without strictly requiring it, e.g., by maximizing likelihood of the unlabeled data, or even better, by minimizing its conditional class entropy. Second, the observed phone strings corresponding to coarticulated pronunciations of different words are also, often, indistinguishable, but can be transformed into a representation in which the degree of overlap is substantially reduced. The canonical phonetic pronunciations are transformed into an articulatory domain, possible mispronunciations are predicted based on a compactness criterion in the articulatory domain, and the result is transformed back into the phonetic domain, forming a finite state transducer that is able to effectively use hundreds of alternate pronunciations. 11 1897