

Jordan University of Science and Technology

Learning to Recognize Speech from a Small Number of Labeled Examples

Authors: Mark Hasegawa-Johnson, Jui-Ting Huang, Rehab Duwairi, Eiman Mustafawi, Roxana Girju, and Elabbas Benmamoun

Abstract: Machine learning methods can be used to train automatic speech recognizers (ASR). When porting ASR to a new language or dialect, however, we often have too few labeled training data to allow learning of a high-precision ASR. It seems reasonable to think that unlabeled data, e.g., untranscribed television broadcasts, should be useful to train the ASR; human infants, for example, are able to learn the distinction between phonologically similar words after just one labeled training utterance. Unlabeled data tell us the marginal distribution of speech sounds, $p(x)$, but do not tell us the association between labels and sounds, $p(y|x)$. We propose that knowing the marginal is sufficient to rank-order all possible phoneme classification functions, before the learner has heard any labeled training examples at all. Knowing the marginal, the learner is able to compute the expected complexity (e.g., derivative of the expected log covering number) of every possible classifier function, and based on measures of complexity, it is possible to compute the expected mean-squared probable difference between training-corpus error and test-corpus error. Upon presentation of the first labeled training example, then, the learner simply chooses, from the rank-ordered list of possible phoneme classifiers, the first one that is compatible with the single labeled training example. This talk will present formal proofs, and experimental tests using stripped-down toy problems; future work will test larger-scale ASR implementation.