MAXIMUM ENTROPY MODELING FOR SPEECH RECOGNITION

Hong-Kwang Jeff Kuo
IBM T.J. Watson Research Center, USA

ABSTRACT

Maximum entropy (maxent) models have become very popular in natural language processing. This tutorial will begin with a basic introduction of the maximum entropy principle, cover the popular algorithms for training maxent models, and describe how maxent models have been used in language modeling and (more recently) acoustic modeling for speech recognition. Some comparisons with other discriminative modeling methods will be made. A substantial amount of time will be devoted to the details of a new framework for acoustic modeling using maximum entropy direct models, including practical issues of implementation and usage. Traditional statistical models for speech recognition have all been based on a Bayesian framework using generative models such as Hidden Markov Models (HMMs). The new framework is based on maximum entropy direct modeling, where the probability of a state or word sequence given an observation sequence is computed directly from the model. In contrast to HMMs, features can be asynchronous and overlapping, and need not be statistically independent. This model therefore allows for the potential combination of many different types of features. Results from a specific kind of direct model, the maximum entropy Markov model (MEMM), will be presented. Even with conventional acoustic features, the approach already shows promising results for phone level decoding. The MEMM significantly outperforms traditional HMMs in word error rate when used as stand-alone acoustic models. Combining the MEMM scores with HMM and language model scores show modest improvements over the best HMM speech recognizer. This tutorial will give a sense of some exciting possibilities for future research in using maximum entropy models for acoustic modeling.

ABOUT THE SPEAKER

Hong-Kwang Jeff Kuo received the S.B. degree in Computer Science and the S.M. degree in Electrical Engineering and Computer Science in 1992, and the Ph.D. degree in Electrical and Medical Engineering in 1998, all from the Massachusetts Institute of Technology. In 1998, he joined Bell Laboratories as a Member of Technical Staff in Murray Hill, New Jersey, where he worked on research in speech recognition and spoken dialogue systems. In 2002, he joined the IBM T.J. Watson Research Center as a Research Staff Member. Dr. Kuo has published papers in journals and international conferences and workshops on many topics, including discriminative training for natural language call routing, language and pronunciation modeling, natural spoken language parsing and understanding, spoken dialogue systems, and models of normal and pathological speech production.