Neural Network Acoustic Model for Recognition of Czech Speech

Tomáš Pavelka and Kamil Ekštein

Laboratory of Intelligent Communication Systems,
Dept. of Computer Science and Engineering,
University of West Bohemia in Plzeň, Czech Republic

E-mail: {tpavelka, kekstein}@kiv.zcu.cz

Keywords: automatic speech recognition, neural networks, hidden Markov models

Abstract: This paper describes the use of artificial neural networks, namely multi layer perceptrons, in the task of automatic recognition of Czech utterances. A speaker independent continuous speech recognition system called LASER (LICS Automatic Speech Extraction/Recognition) is being developed at the laboratory of intelligent communication systems (LICS), University of West Bohemia which takes advantage of the application of neural networks.

Today's most successful speech recognition systems are based on hidden Markov models with Gaussian mixtures as acoustic models. The task of an acoustic model is to provide a "soft" decision (e.g. assign probability) about which class of speech sounds does a small segment of speech belong to. While Gaussian Mixture are the most widely and successfully used solution (e.g. in the Hidden Markov Toolkit (HTK), see [1]) they have some well known limitations. It has been shown for example by [2] that neural networks when employed as acoustic models have some advantages over the more traditional Gaussian mixtures.

We have run a series of tests with both HTK and LASER software to determine the practical advantages and disadvantages of both approaches with the corpora of Czech speech utterances that we have available. Both Neural and Gaussian mixture acoustic models led to almost the same recognition accuracy, but the neural network approach required about four times less trainable parameters than the one with Gaussian Mixtures. This means that the response times of our recognizer are significantly lower than those of the HTK based system applied to the same task while maintaining similar recognition accuracy.

There are several drawbacks of the neural network approach, the most severe one is that the training is about ten times slower than training of a similar system with Gaussian mixtures. But while the training is computationally expensive, the trained recognizer runs well even on slower machines such as PDAs.

References

[1] S. Young, *The HTK Book (for HTK v3.2)*, Cambridge University Engineering Dept, 2002.

[2] H. Bourlard and N. Morgan, "Hybrid HMM/ANN systems for speech recognition: Overview and new research directions," in *Summer School on Neural Networks*, 1997.

[3] T. Pavelka, "Hybrid speech recognizer implementation," M.S. thesis, University of West Bohemia, Pilsen, 2003.