Various Robust Search Methods in a Hungarian Speech Recognition System

Gábor Gosztolya, András Kocsor, László Tóth, and László Felföldi

Research Group on Artificial Intelligence
of the Hungarian Academy of Sciences and University of Szeged
H-6720 Szeged, Aradi vértanúk tere 1., Hungary
{gosztg, kocsor, tothl, lfelfold}@inf.u-szeged.hu
http://www.inf.u-szeged.hu/speech

Extended Abstract

In any speech recognition application we have to identify spoken words, based on the information provided by various features. In this process a large number of word-combinations must be tried out, and the best fitting ones must be chosen. A reduction of this search space (ie. word-sequences) is quite important for both speed and memory reasons, because most of these hypotheses will, for one reason or another, turn out to be quite unsuitable.

To tackle this problem, a number of standard algorithms are available like Viterbi beam search, stack decoding, forward-backward search and A* [1][2]. We have implemented some of them and focused mainly on an extension of the general purpose stack decoding method. Our OASIS Speech Laboratory package incorporates most of these methods, which we then tested on a set of (Hungarian) speech databases.

In order to find the best fitting word-sequences, language information is obviously quite important. Incorporating this kind of knowledge into a speech recognition system usually means some kind of language model has to be used. Although this paper focuses on the search process, we cannot ignore another related point, that of choosing a good representation for the Hungarian language.

References

1. JELINEK, F., *Statistical Methods for Speech Recognition,* The MIT Press, 1997.
2. HUANG, X., ACERO, A., HON, H.-W., *Spoken Language Processing,* Prentice Hall PTR, 2001.