LDec: One Pass Time Synchronous Decoder

Tomáš Pavelka[1]

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

e-mail: tpavelka@kiv.zcu.cz

Most today's systems for automatic speech recognition are based on the framework of hidden Markov models (HMM) where the search for the most likely word sequence is achieved by searching for the single most likely path through a HMM representing all possible utterances. The usual technique for this is the Viterbi algorithm (see e.g. [2]) and the process (in terminology of speech recognition) is referred to as *decoding*. Since the total number of possible states can be quite large most of the work in construction of the decoder relates to storage and computational efficiency.

Our initial experiments were carried out with hand written grammars describing all possible recognized utterances. The advantage of this approach is that the search space is relatively small and it is not necessary to do any pruning during the search.

When long span language models (such as trigrams) together with larger dictionaries are used it is no longer possible to do an exhaustive search. In order to lower the computational complexity the number of active states (i.e. those that will take part in further computation) must be pruned. A simple but efficient pruning method is the *beam search* which prunes all states having lower score than a given percentage of the highest score. Such search is no longer *admissive*, i.e. does not guarantee to find the most likely word sequence. Despite this it has been shown to work well in practice.

Another way to speed up the search is to reorganize the dictionary. For example if two words start with the same phoneme (phonetic unit) than the computation for their initial states needs to be done only once. This leeds to a tree like structure of the dictionary and its respective HMM. While the tree

reduces the size of the HMM to about 40% of its original size, it can reduce the number of active nodes during beam search by an order of magnitude.

According to a classification scheme proposed in [1] our decoder (named LDec – LASER Decoder) can be described as a single pass time synchronous decoder with dynamic network expansion. Single pass refers to the fact that the final result is known after one run of the decoding algorithm. In the case of a multi pass decoder the decoding process is run several times with different precision of acoustic and language models. Time synchronous means that all HMM states belonging to a time frame are processed before moving to the next frame. A decoder with dynamic network expansion creates the states for the next time frame "on the fly" as opposed to static network where all possible HMM states are known beforehand.

References

[1] X.L. Aubert. An overview of decoding techniques for large vocabulary continuous speech recognition. In *Computer Speech and Language*, volume 16, pages 89–114, 2002.

[2] L.R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, volume 77, 1989.