IBM T.J. Watson Research Center
1101 Kitchawan Road
PO Box 218
Yorktown Heights, NY 10598

Department of Computer Graphics and Multimedia
Faculty of Information Technology, Brno University of Technology

Review of doctoral thesis:

FINITE-STATE BASED RECOGNIOTION NETWORKS FOR FORWARD-BACKWARD SPEECH DECODING

Submited by Dipl.-Ing. Mirko Hannemann

The topic of this thesis addresses the issue of speed improvement of a Large Vocabulary Continuous Speech Recognition (LVCSR) decoder. This is a very challenging topic, considered as one of the most difficult areas in the LVCSR field. Any candidate in this area aiming to make an impact has to demonstrate deep theoretical knowledge, expertise and experience in C++, algorithms and computer architecture.

Technical content of the thesis

It order to achieve his goals, the candidate had to address a handful of intermediate problems, which are quite quite interesting on their own. The level of attention given to each such problem is quite commendable, each section is self contained and can serve
as a good reference, with close to tutorial quality. This is one of the reasons to believe this work will contribute to the LVCSR field. I will first briefly describe the content of the thesis and then address how well the candidate fulfilled the expected requirements.

After a brief introduction in Chapter 1, basic concepts of ASR are introduced in Chapter 2. The level of details is sufficient for the thesis to be self contained yet not overwhelming.

Chapter 3 to introduces a novel technique for weight pushing, with excellent theoretical justification and sufficient experimental validation.

Chapter 4 address the theoretical aspects of an exact Language Model reversal, including concisely written steps of the developed technique, with an experimental validation and the end.

Charter 5 combines the techniques introduced in the previous chapter to achieve the main goal of the thesis - efficient method for forward-backward search - with a detailed description of the algorithm and several of its variations, such as parallelization. I particularly appreciate that the candidate discusses alternative methods of tracked search implementation which could possibly be even more efficient.

Chapter 6 nicely summarize the achieved results and shows a roadmap for future work (which I hope will be followed).

Not only this work is relevant to the LVCSR field, but it introduces elegant solutions to known problems not always fully addressed by the available state of the art methods. This is particularly true for the weight pushing and the LM reversal. While the basic idea of the forward-backward search had been explored many times the early works on LVCSR decoders (as the candidate discusses and properly

references), what is certainly a contribution of this work is that the presented forward-backward approach offers the most optimal solution of this problem to my knowledge.

The candidate published his work sufficiently, a clear indication of which is the fact that I was familiar with several of the concepts already, before reading his thesis. Aside from the publication, the candidates academic impact is clear from his contributions to Kaldi, which is becoming a standard toolkit for academic work in the area.

If there is weaker part of this theses (in relative terms), it is its experimental section. The presented results are sufficient to demonstrate the impact of the proposed methods, but it would be very interesting to know the behavior in a more diverse set of tasks. For example, using a test set with much lower word error rate, when the
decoder time tends to be dominated by the cost on acoustic observation evaluation.
Or in a noisy or otherwise mismatched environments.

Summary on the technical content of the thesis :

The thesis clearly demonstrates the qualities of the candidate, his very sound theoretical background, his ability to convey the ideas and his experience in experimental work with attention to details. His work clearly presents a strong contribution to the advancement of the LVCSR.

Comments on the formal aspects:

The thesis is well written in good English (as much as I can judge as a non-native speaker). I did not find any formal errors.

To conclude, I consider this thesis does meet the generally accepted requirements for the doctoral degree and I recommend conferment of this degree by Brno University of Technology.

Specific questions to the candidate:

- The presented weight pushing results are presented when applied to the LG transducer. Was there any attempt to evaluate this method on a fully expanded HCLG transducer, i.e. with a static decoder ?

- The forward-backward methods have an inherent impact on the latency for the LVCSR system. In the thesis, this fact is not discussed in great depth. Can the candidate describe in more detail how does the method apply to various use cases from the offline batch-processing scenario to fully online interactive systems (in terms of latency and real-time factor) ? How would the computation of acoustic observations
be handled more efficiently in the backward path, in particular on platforms without GPUs ?

- The presented speed-up uses a baseline decoder which produces lattices at the state alignment level. This usually comes at additional cost, in comparison to search methods which produce lattices at the word level alignment only. Is this specific cost known for the decoder used to implement the forward-backward search ?

In Yorktown Heights, September 1, 2016

Dr. Miroslav Novak
Research Staff Member
IBM Watson Group

Phone: 1-914-945-2922
E-mail: miroslav@us.ibm.com
http://researcher.ibm.com/researcher/view.php?person=us-miroslav