Lehrstuhl Informatik 6
Human Language Technology and Pattern Recognition

RWTH AACHEN UNIVERSITY

Prof. Dr.-Ing. Hermann Ney

Dr.rer.nat. Ralf Schlüter

Ahornstrasse 55, Building E2
52074 Aachen
GERMANY
Phone: +49 241 80-21612
Fax: +49 241 80-6-21606

schlueter@cs.rwth-aachen.de

| Ahornstrasse 55 | 52074 Aachen | GERMANY   122010

Brno University of Technology
Faculty of Information Technology
Sylva Sadovská
Božetěchova 2
612 66 Brno
CZECH REPUBLIC

02.09.2016

Report on Doctoral Thesis Mirko Hannemann

Dear Prof. Meduna,

enclosed you find my report on the doctoral thesis “Finite-State based Recognition Networks for Forward-Backward Speech Decoding” submitted by Dipl.-Ing. Mirko Hannemann.

Kind regards,

Ralf Schlüter

Enclosed:

- Thesis report (pp. 2-4);

VAT Identification Number
DE 121 689 807

Tax Number
201/5930/5005

RWTH Aachen University
Sparkasse Aachen
Account 18
BIN 390 500 00
SWIFT-BIC: AACSDE 33XXX
IBAN: DE53 3905 0000 0000 0000 18

Report on Doctoral Thesis

"Finite-State based Recognition Networks for Forward-Backward Speech Decoding"

submitted by Dipl.-Ing. Mirko Hannemann

Task Definition

This thesis addresses the search problem within automatic speech recognition (ASR). More specifically, a framework for locally adaptive search space control utilizing a concept for bidirectional search within a finite state transducer approach is developed. Especially, an exact reversal of the search procedure is investigated, including all aspects of finite state transducer based beam search. The aim of the backward search introduced in this thesis is to enable localized pruning. The idea is that forward and backward search directions are independent and complementary. Search beams are locally adapted by either a local comparison of both search directions, or using a first pass in one direction to further guide the search in the opposite direction, which is called *tracked decoding*. An analysis of beam search behaviour was performed to investigate optimal operating points for tracked decoding, followed by a comprehensive performance evaluation of tracked decoding. It is shown that locally adaptive beam search by tracked decoding leads to a considerable reduction in search space and computation time. In addition, iterative forward-backward search utlized for search parallelization.

Structure

In the following, the thesis will be discussed by way of individual sections:

1. *Introduction*
   In the Introduction, the specific motivation, the claims and contributions of this work are presented, and an outline of the thesis is given.
2. *Weighted Finite State Transducers and LVCSR Decoding*
   This chapter discusses the state-of-the-art in automatic speech recognition as relevant for this thesis. The chapter commences with a brief overview of the statistical approach to automatic speech recognition, followed by a thorough discussion of the state-of-the-art in ASR beam search methods, with specific focus on finite-state-transducer based search.
3. *An Alternative Weight Pushing Algorithm*
   In this chapter, a novel method for pushing weights within weighted finite state acceptors is presented. The determinzation and minimization of a search network alone does not necessarily guarantee efficient beam search, which includes discarding unlikely hypotheses. For balanced and effective hypothesis pruning in beam search, it is necessary to obtain locally comparable (partial) scores and apply weights early on in the search process. For this, weight pushing is imperative in weighted finite state transducer based search, but can be inefficient, as shown here. The chapter commences with a discussion of existing weight pushing approaches, which is followed by a well-founded discussion of ergodic Markov chains and their relation to the theory of non-negative matrices. This comprehensive theoretical analysis then leads to the proposal of an alternative iterative weight pushing algorithm aiming at better convergence than existing work on weight pushing. The favorable properties of the alternative weight pushing algorithm developed is validated experimentally on the widely used Switchboard task.

4. *Exact Reversal of ARPA Back-Off Language Models*
   The reversal of the overall search procedure especially requires an exact reversal of the language model to ensure that backward search gives the same scores as forward search. In this chapter, the exact reversal of back-off language models is investigated with specific focus on the corresponding weighted finite state acceptor implementation. A comprehensive discussion of the related problems, and a proof of correctness for the proposed solution is provided. In addition, an alternative derivation based on 'textitBayes rule is provided.
5. *Combining Forward and Backward Search in Decoding*
   In this chapter, the complete construction of backward search in a weighted finite state transducer framework is presented, and two methods aiming at local adaptation of beam search to improve the search efficiency are presented. First of all, all steps of the construction of the time-reversed search graph by way of individually reversing its components are discussed. Finally, two methods for efficient locally adaptive beam search are presented: *incremental forward-backward search* to detect and iteratively improve local search errors, and *tracked decoding* to optimally guide search in one direction utilizing information from search in the opposite direction. In both cases, a structured discussion and experimental study of the pruning parameters involved is presented, followed by independent experimental evaluations of the proposed approaches.
6. *Conclusions*
   A summary of the main contributions, and an outlook into possible future work is given.

Assessment and Recommendation

In his thesis, Mr. Hannemann presents pioneer work on locally adaptive beam search by way of an exact time-reversal of the corresponding search networks. As his analyses show, the search space required usually varies strongly during the course of speech to text decoding. Former static pruning methods do not sufficiently acknowledge this dynamic variation in search space and regularly lead to unnecessarily large local search complexity and/or constrain the search space to strongly, resulting in search errors.
Mr. Hannemann proposes the combination of a standard forward search pass with an exactly time reversed, “backward” search pass for better local search space control in ASR. Even though the use of multiple search passes is not new for the purpose of efficient search, the core novelty is Mr. Hannemann's insight about the independence of (partial) forward and backward search passes, and its usefulness for local adaptation of beam search. Overall, the novel locally adaptive beam search framework proposed by Mr. Hannemann provides an important improvement in the efficiency of beam search for ASR and related applications.
In his investigation of the problems underlying the time reversal of beam search in a weighted finite state transducer framework, Mr. Hannemann shows a very good understanding and fundamental knowledge about the modeling and the corresponding algorithms involved. This is reflected in his theoretically sound solutions for weight pushing within weighted finite state acceptor based seaerch networks, as well as the time reversal of such networks and especially the language model component. Mr. Hannemann applied his proposed forward-backward search framework within two approaches for locally adaptive search, tracked decoding and incremental forward-backward search.

Tracked decoding, has been proposed by Mr. Hannemann and is analyzed and evaluated extensively within his thesis. It shows considerable improvements in computational complexity by allowing for locally adaptive pruning.
Incremental decoding was introduced in [Nolden et al. 2013], but was closely building on Mr. Hannemann's forward-backward search framework and was derived from the tracked decoding approach. Mr. Hannemann reimplemented the approach the approach during the course of an Internship at Microsoft Research aiming at chunked decoding and parallelization. Within the internship, the approach was implemented into Microsoft's recognition system, which might also be the reason for the missing direct comparison of tracked decoding and incremental forward-backward search, which might have been of interest.
Mr. Hannemann's research findings were published on the main conferences for the field of ASR, i.e. the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), the ISCA Interspeech conference, the IEEE Spoken Language Technology (SLT) workshop, and the IEEE Automatic Speech Recognition and Understanding (ASRU) Workshop. His results are acknowledged by the ASR community. Specifically, his findings were taken up by one of the Ph.D. students of our institute (as cited in this thesis), who applied his definition of backwards search for incremental forward-backward search and obtained very good results. During the course of his work, Mr. Hannemann also participated twice in the prestigious Johns Hopkins University workshops, and also contributed to the widely acknowledged Kaldi ASR toolkit.
Mr. Hannemann participated in a number of projects during the course of his thesis-related work. I know him personally from our participation in competing teams within the large, highly competetive US-IARPA project BABEL, where he within FIT/BUT provided strongly to their team Babelon headed by BBN, which consistently obtained excellent performance throughout the course of the project.
Overall, Mr. Hannemann shows outstanding and promising work on the search problem in automatic speech recognition, which already resulted in successful follow-up work by another research group during the course of his work [cf. pp. 68], which was again taken up by Mr. Hannemann. With his thesis, which can be expected to induce further investigations to take advantage of forward-backward search, Mr. Hannemann provides a notable contribution to the ASR community and specifically to the search problem in ASR. I am pleased to accept his thesis without reservations. His doctoral thesis meets the requirements of the proceedings leading to Ph.D. title conferment.