

Public Abstract

Xiaolong Li

Ph.D.

Computer Science

REAL-TIME SPEAKER-INDEPENDENT LARGE VOCABULARY CONTINUOUS
SPEECH RECOGNITION

Advisor: Dr. Yunxin Zhao

Graduation Term Fall 2005

ABSTRACT

In this dissertation, a real-time decoding engine for speaker-independent large vocabulary continuous speech recognition (LVCSR) is presented. Three indispensable and correlated performance measurements -- accuracy, speed, and memory cost, are carefully considered in the system design, with the main innovations in fast and memory-efficient decoding algorithms.

A novel algorithm, Order-Preserving Language Model Context Pre-computing (OPCP) is proposed for fast Language Model (LM) lookup, resulting in significant improvement in both overall decoding time and memory space without any decrease of recognition accuracy. Experimental results are reported on two LVCSR tasks (Wall Street Journal 20K and Switchboard 33K) with three sizes of trigram LMs (small, medium, large). In comparison with MPH and LMCP methods, OPCP reduced LM lookup time from about 30~80% of total decoding time to about 8%~14%, without any loss of word accuracy. Except for the small LM, the total memory cost of OPCP for LM lookup and storage was about the same or less than the original N-gram LM storage, and was much less than the compared methods. The time and memory savings in LM lookup by using OPCP became more pronounced with the increase of LM size.

By using the OPCP method and other optimizations, our one-pass LVCSR decoding engine, named TigerEngine, reached real-time speed in both tasks of Wall Street Journal 20K and Switchboard 33K, on the platform of a Dell workstation with one 3.2 GHz Xeon CPU. TigerEngine is to be used in automatic captioning for Telehealth.