論文の内容の要旨

Advancing Finite State-based Spoken Language Processing Systems: Theoretical Principles and Practical Techniques

（有限状態制御の音声言語処理系の先鋭化：理論と実装）

氏　　名　　ノバック　ジョセフ　ロバート

This thesis describes a variety of advanced research conducted in the area of Spoken Language Processing using the Weighted Finite-State Transducer (WFST) paradigm. In particular this thesis looks at WFST-based approaches to Automatic Speech Recognition (ASR) and Grapheme-to-Phoneme (G2P) conversion.

In the area of WFST-based ASR, it describes experimental results and theoretical approaches to the construction of integrated WFST-based transducer cascades for ASR applications, with a focus on three-component cascades and the use of advanced dynamic composition algorithms. It also provides empirical results illustrating the superior performance characteristics of WFST-based methods in Speech Recognition with respect to classical approaches.

It proposes a novel theoretical algorithm for constructing integrated WFST-based cascades for dynamic grammars and dialog applications.

In the area of G2P conversion, this work investigates a variety of novel model construction and decoding techniques. These include Lattice Minimum-Bayes Risk decoding for G2P pronunciation lattices, and N-best rescoring with a Recurrent Neural Network Language Model. In addition to these theoretical applications, novel engineering solutions for several problems in WFST-based G2P conversion are presented. These include a novel method for the use of failure transitions in FST composition and simple perplexity analysis for joint G-P pronunciation lattices.

In addition to the theoretical descriptions, this work provides several complete open source toolkits. These ASR-related toolkits include a modified WFST-based ASR decoder capable of leveraging state-of-the-art dynamic composition algorithms, a set of tools for building and integrating WFST-based ASR cascade components, and a simple education-oriented library suitable for training statistical language models. The G2P work has also been released in the form of an open source C++ toolkit. This toolkit has been evaluated by several independent third party research groups and shown to achieve state-of-the-art accuracy on large-scale pronunciation databases.

Beyond specific contributions, the primary goal of this work has been to establish a strong connection between the often competing interests of education, theoretical advancement and the necessity of practical, open implementations.