

発表概要

純粋関数型言語を用いた 超コンパクト音声認識デコーダの開発

篠崎 隆宏^{†1,†2} 関嶋 政和^{†1,†3}
萩原 茂樹^{†1} 古井 貞熙^{†1}

音声認識研究は小語彙の孤立単語認識の研究からはじまり、今日では大語彙連続音声認識システムが実現されている。音声認識システムを実現する中核的なソフトウェアとして、認識デコーダがあげられる。認識デコーダの役割は、音声特徴量の時系列を受け取り、それに対して最も確からしい単語系列を探索し、出力することである。可能な単語列の種類数は音声の長さに対して指数関数的に増加するため、探索には効率的なアルゴリズムが必要である。大語彙連続認識に対応した現在のデコーダソフトウェアは非常に複雑で、1 万行以上のコードサイズがある。技術の進歩とともにこの複雑性は増加しつつあり、音声認識の研究において新しいアイデアを試すことが困難になりつつある。そこで、プロトタイピングを目的として変更が容易でコンパクトな音声認識デコーダを実現するために、純粋関数型プログラミングの応用を試みた。具体的には、Haskell を用いた大語彙連続音声認識デコーダ「Husky」の開発を行った。純粋関数型プログラミング言語の高度な抽象化能力を用いることで、Husky はわずか 400 行と非常にコンパクトなコードで実装することができた。本発表ではまず音声認識の基本原理について簡単に紹介した後、Husky の設計と実装について述べる。さらに、実際に Husky を用いた大語彙連続音声認識実験を行い、最新の認識システムと同等の認識精度が得られることを示す。

A Compact Speech Decoder Based on Pure Functional Programming

TAKAHIRO SHINOZAKI,^{†1,†2} MASAKAZU SEKIJIMA,^{†1,†3}
SHIGEKI HAGIHARA^{†1} and SADAOKI FURUI^{†1}

The history of automatic speech recognition started with isolated small vocabulary recognition tasks. It has now been improved to work for continuous

speech recognition tasks with large vocabulary. One of the core software in a speech recognition system is called a decoder, which takes speech feature sequence as an input, searches for a word sequence that best matches to the input, and outputs that sequence. The possible number of word sequences exponentially increases for the length of input speech. Therefore, an efficient search strategy is required. The source code of existing decoders is complex and amounts to more than 10,000 lines. Because of the complexity of the software, it is a bottleneck for speech researchers to modify an existing decoder to try a new idea. Seeking for a new framework to compactly describe a decoder for prototyping, we have applied the pure functional programming. Our developed pure functional decoder is written by Haskell and named “Husky”. Thanks to the high abstraction ability of the pure functional language, Husky has only 400 lines, which is significantly smaller than existing decoders. In this presentation, we first briefly introduce the basics of the speech recognition algorithm. Then, the design and implementation of Husky are described. Finally, the performance of Husky is demonstrated by large vocabulary continuous speech recognition experiments.

(平成 23 年 4 月 25 日発表)

^{†1} 東京工業大学大学院情報理工学研究科計算工学専攻

Department of Computer Science, Graduate School of Information Science and Engineering, Tokyo Institute of Technology

^{†2} 千葉大学大学院融合科学研究科

Graduate School of Advanced Integration Science, Chiba University

^{†3} 東京工業大学学術国際情報センター

Global Scientific Information and Computing Center, Tokyo Institute of Technology