

審査の結果の要旨

氏名 倉田 岳人

本論文は「Spoken Language Processing for Applying Large Vocabulary Continuous Speech Recognition to the Real World (大語彙連続音声認識の実世界適用に向けた音声言語処理に関する研究)」と題し、英文で全7章から成る。

第1章「Introduction (序論)」では、大語彙連続音声認識を利用するアプリケーションの実世界適用に向けて、(1) 大語彙連続音声認識用モデルの自動構築、(2) 大語彙連続音声認識を利用するアプリケーションの性能改善、という二つの研究課題を設定している。第一の研究課題の背景として、大語彙連続音声認識を利用するアプリケーションでは、アプリケーションごとに話題や語彙が異なり、また同じアプリケーションであってもそれらが頻繁に変化することを指摘し、人手を介して多くのアプリケーションのモデルを用意したり、頻繁にモデルを更新したりすることが現実的ではないことを述べている。第二の研究課題の背景として、多くのアプリケーションが大語彙連続音声認識と他の技術の組み合わせにより実現されていることを指摘し、ユーザの観点からは中間的な音声認識結果の正確さよりも、アプリケーションの目的がどれだけ達成されるか、ということの方が重要であることを述べている。

第2章「Large Vocabulary Continuous Speech Recognition and Its Applications (大語彙連続音声認識とそれを利用するアプリケーション)」では、最初に大語彙連続音声認識の概要および性能評価方法についてまとめている。次に、大語彙連続音声認識を利用するアプリケーションを、実装されるプラットフォームやアプリケーションの目的ごとに分類し、これらの実世界適用に第1章で設定した研究課題が重要であることを述べている。

第3章と第4章では、「大語彙連続音声認識用モデルの自動構築」という第一の研究課題に取り組んでいる。

第3章「Acoustically Discriminative Language Model Training with Pseudo-LVCSR (擬似音声認識を利用した言語モデルの識別学習)」では、音声データとそれに対応する書き起こしテキストを必要としない、言語モデルの識別学習の方法を提案している。従来の言語モデルの識別学習では音声データから、音声認識システムにより誤りを含む認識結果を得て、それを書き起こしテキストと比較していたため、人手による書き起こし作業が必要であった。提案手法では、音声認識システムが出力すると考えられる認識結果を、音素間の類似度を考慮してテキストから直接生成し、識別学習に利用している。これにより、音声データを収集し、人手で書き起こす必要がなくなる。大学講義音声による評価実験を行い、認識精度を改善できることを確認している。

第4章「Unsupervised Lexicon Acquisition from Speech and Text (音声とテキストからの語彙自動獲得)」では、大語彙連続音声認識のための辞書の自動構築について検討している。大語彙連続音声認識では、辞書に含まれない単語は認識されない。しかし、大語彙連続音声認識を新しい分野に導入する場合、辞書に含まれないその分野特有の単語が存在

する。分野特有の単語はその分野のテキストに含まれるが、そのような単語を自動的に獲得し、その読みを自動推定することは困難であった。提案手法では、テキスト中のすべての部分文字列を単語候補とみなし、それらを音声データと照合することで、適切な単語とその読みを獲得している。自動構築された辞書が音声認識精度の向上に貢献することを実験的に確認している。

第5章と第6章では、「大語彙連続音声認識を利用するアプリケーションの性能改善」という第二の研究課題に取り組んでいる。

第5章「Leveraging Word Confusion Networks for Named Entity Modeling and Detection from Conversational Telephone Speech (単語コンフュージョンネットワークを利用した日本語電話音声からの固有表現検出)」では、コールセンターにおける電話音声からの固有表現検出について検討している。電話音声に対する音声認識は困難であり、認識誤り率が高くなりやすく、誤りを多く含むテキストからの固有表現検出は困難で、特に音声認識の一位の認識結果に固有表現が現れなければ、後段の固有表現検出は正しく動作しない。ここでは一位の認識結果だけではなく、競合する他の仮説も考慮に入れて、固有表現検出を行う方法が提案されている。実際の電話音声による評価実験を行い、固有表現検出の精度が上昇することを確認している。

第6章「Voice Control System in Automobiles Reflecting Analysis on How Users Speak (ユーザの発話傾向分析に基づく車載機器操作のための音声入力手法の検討)」では、自動車内音声操作システムについて検討している。ユーザは音声操作を行う場合、幅広い表現の発話を行う。従来用いられていた文法による音声認識では、ユーザの幅広い発話を認識できず、結果として意図した機能を音声で操作することは困難であった。ここでは、自動車内でのユーザの発話例を収集し、それらに定量的・定性的な分析を加え、文法に基づく音声認識の限界を示している。その上で、大語彙連続音声認識と自然言語処理によるアクション分類とを組み合わせることで、ユーザの幅広い発話に対応する方法を提案している。大語彙連続音声認識を導入することで、文法では表現しきれない発話を認識することができ、後段の処理としてアクション分類を導入することで、認識結果に誤りが含まれていても、ユーザの意図を推定することができる。市販車の音声操作システムを利用して評価実験を行い、ユーザが直観的な発話で操作できる音声操作システムを実現できる見通しを得ている。

第7章「Conclusion (結論)」では、本論文の研究成果をまとめている。

以上を要するに、本論文は大語彙連続音声認識を利用するアプリケーションを実世界適用するために、人手を伴わずに音声認識用言語モデル、および辞書を構築・改善する手法を提案している。また、二つのアプリケーション（日本語電話音声からの固有表現検出及び自動車内音声操作システム）を例として、大語彙連続音声認識を利用するアプリケーション全体の性能改善法を提案している。提案手法は実験的に効果が検証され、さらに本論文で紹介されている他の様々なアプリケーションへも適用できる可能性が示されており、音声認識システムに関する電子情報学に貢献するところが少なくない。

よって本論文は博士（情報理工学）の学位論文として合格と認められる。