論文の内容の要旨

題目

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)

氏名

倉田　岳人

要旨

大語彙連続音声認識を利用するアプリケーションは実世界の様々な場面に導入され始めており，その種類は今後も増えていくことが期待される．本論文では，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のために必要な技術を，以下の二つの研究課題を設定し，検討する．

1. 大語彙連続音声認識用モデルを自動構築する．
2. アプリケーション全体の性能を改善する．

最初に，大語彙連続音声認識を利用する様々なアプリケーションを紹介し，アプリケーションごとに話題や語彙は異なり，またアプリケーションによってはそれらが頻繁に変化することを指摘する．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することは現実的ではなく，これを踏まえて，第一の研究課題として，大語彙連続音声認識用モデルの自動構築方法を検討する．次に，多くのアプリケーションは，大語彙連続音声認識と他の技術を組み合わせて実現されていることを指摘する．ユーザから見ると，アプリケーションの目的がどれだけ達成されるか，ということが，中間的な音声認識結果の正確さよりも重要である．例えば，音声検索システムは，大語彙連続音声認識とウェブ検索によって構成されているが，ユーザにとっては，音声認識結果の正確さよりも，意図したページが提示されることの方が大切であろう．これを踏まえて，第二の研究課題として，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を改善する方法を検討する．

本論文は 4 部（7 章）から構成されている．第 1 部は導入部分として，大語彙連続音声認識の歴史・理論についてまとめ，それを利用するアプリケーションを紹介し，上述した研究課題の重要性を述べる．また，第 4 部では本論文の結論を述べる．第 2 部，第 3 部では，以下のように，二つの研究課題に対しての取り組みをまとめる．

第 2 部は，第 3 章と第 4 章から成り，第一の研究課題に取り組む．第 3 章では，音声デ

ータとそれに対応する書き起こしデータを必要としない，言語モデルの識別学習の方法を提案する．従来の言語モデルの識別学習では，音声データを音声認識システムで認識し，誤りを含む認識結果を得た上で，それを対応する書き起こしテキストと比較して，言語モデルのパラメータを調整していたため，人手による音声データの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，誤りを含む認識結果を，テキストデータから直接生成し，識別学習に利用する．これにより，音声データを収集し，それを人手で書き起こす必要がなくなる．大学の講義音声を利用して評価実験を行い，提案手法により認識性能を改善できることを確認した．次に，第 4 章では，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する．大語彙連続音声認識の枠組みでは，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データから獲得することが一般的であるが，自動単語分割器はそのような単語を正確に単語分割することができず，またその読みを自動推定することも困難であった．提案手法では，テキストデータ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する．評価実験を通じて，自動構築された辞書は，音声認識精度の向上に貢献することを確認した．第 3 章，第 4 章で提案する手法は，大語彙連続音声認識を利用するアプリケーションの新しい分野への導入や，音声検索システム・音声インデクシングシステムのようなアプリケーションでのモデルの頻繁な更新に貢献する．

第 3 部は，第 5 章と第 6 章から成り，二つのアプリケーションを例として，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を改善する方法を検討する．第 5 章では，コールセンターにおける電話音声からの固有表現検出システムについて検討する．コールセンターでの会話には，主に固有表現からなる機密情報が含まれており，それらを検出することができれば，ビジネス上の価値がある．帯域幅や自由な発話内容により，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い．そして，誤りを多く含むテキストからの固有表現検出は困難であった．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．本論文では，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を提案する．提案手法では，音声認識の仮説を単語コンフュージョンネットワークで表現し，単語ではなく，単語コンフュージョンネットワークの各節を単位として固有表現検出を行う．実際のコールセンターの音声データを利用して評価実験を行い，固有表現検出の精度が上昇することを確認した．第 5 章で提案する手法は，音声検索語検出システムのような，大語彙連続音声認識の結果に対して情報検索を行うアプリケーションの精度改善にも貢献する．次に，第 6 章では，自動車内音声操作システムについて検討する．ユーザは自動車内で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なかった．結果として，ユーザは意図した機能を音声で操作できず，自動車内での音声操作システムの普及が進んでこなかった．本論文では，自動車内でのユーザの発話例を収集した上で，それらに対して定量的・定性的な分析を加え，従来の文法に基づく音声認識の限界を示す．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する．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る．また，後段の処理としてアクション分類を導入することで，音声認識の結果に誤りが含まれていても，ユーザの意図を頑健に推定することができる．実際の市販車の音声操作システムを利用して評価実験を行い，提案手法により，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た．大語彙連続音声認識の結果に対してアクション分類を適用する，という処理の流れは，コールセンターにおける自動での顧客振り分け（コールルーティング）システムや，携帯電話におけるパーソナルアシスタントでも同様であり，第 6 章で提案する手法は，これらのアプリケーションの性能改善にも貢献する．