

論文内容の要旨

博士論文題目 Rapid Unsupervised Speaker Adaptation Based on
Sufficient Statistics of Hidden Markov Models
(隠れマルコフモデルの十分統計量に基づく高速教師なし話者適応の研究)

氏名 Randy Gomez

(論文内容の要旨)

音声認識の性能を向上させる有力な手段として、話者適応が考えられている。従来の話者適応の手法では、あらかじめ決められた文を間違いなく少なくとも数十文発声することが必要であった。さらに、適応には速くても数分の計算時間がかかった。この論文では、あらかじめ収集されている数百人程度の音声データベースを利用して、ユーザにまったく負担をかけない話者適応アルゴリズムを提案し、かつ、大規模音声データベースを用いて評価を行っている。

この論文では、下記の成果をあげている。

(1) 話者ごとのHMM十分統計量をあらかじめ計算しておき、混合ガウス分布に基づく話者の類似度の計算により、任意の一文発声での話者適応を行う方式の改良を行った。年齢、性別別の音響モデルを利用することによって、話者適応の性能を大幅に向上させた。従来の10文発声の教師あり適応をはるかに上回り、50文発声に近い性能を達成した。かつ、音声データ全体の十分統計量も利用することにより、数秒で話者適応が実行できるアルゴリズムを開発した。

(2) さらに、この十分統計量に基づく話者適応の考えを、最大尤度線形回帰(MLLR)に適用するアルゴリズムを考案して、さらなる性能の向上を達成した。

(3) これらの手法の評価を、数百人の音声データベースと各種雑音環境下で行い、有効性を検証した。

(4) この教師なし話者適応アルゴリズムを実際の音声対話システムに組み込み、良好に動作することを確認した。

この論文の高速教師なし話者適応技術は、音声認識システムの実用化に貢献することが、大いに期待できる。

氏名	Randy Gomez
----	-------------

(論文審査結果の要旨)

音声認識の性能を向上させる有力な手段として、話者適応技術が有用である。この話者適応者技術を、ユーザにまったく負担のかけない教師なし話者適応の枠組みで、数秒の計算時間で実現したことは賞賛に値する。とくに、アルゴリズムの開発、大規模な評価実験、実システムへの組み込みまでを一環として成し遂げたことは評価できる。

これらの成果を、電子情報通信学会などの3編の学術論文、ICASSPなどの重要な国際会議に3編の論文として発表しており、国際的にも高い評価を得ている。

この博士論文の成果をまとめておく。

(1) あらかじめ計算された話者ごとのHMM十分統計量を利用して、任意の一文発声での話者適応を行う方式を改良し、認識性能の向上および処理速度の大幅な高速化を達成した。従来の10文発声の教師あり適応をはるかに上回り、50文発声に近い認識性能を達成した。かつ、音声データ全体の十分統計量も利用することにより、従来法では数分かかる話者適応時間を、数秒で適応できるアルゴリズムに改良した。

(2) さらに、この適応の考えを最大尤度線形回帰(MLLR)に適用して、新しい教師なし話者適応アルゴリズムを考案して、更なる認識性能の向上を達成した。

(3) これらの手法の評価を、数百人の音声データベースを用いて、各種雑音環境下で行い、有効性を報告した。

(4) この教師なし話者適応アルゴリズムを実際の音声対話システムに組み込み、良好に動作することを確かめた。

このように、教師なし話者適応アルゴリズムの改良、評価、さらに、MLLRによる新しい高速話者適応アルゴリズムの考案、大規模音声データによる評価、最後に、実環境システムへの組み込みまでを行い、先端の音声認識研究者として、今後とも十分に活躍できるレベルに到達した。

平成18年8月3日に開催した公聴会の結果も参考にして、本博士論文の審査を行い、本論文は、博士論文(工学)として十分な価値があるものと判断した。