

Title	話し言葉音声認識のためのトリガーペアに基づく言語モデルの適応
Author(s)	Troncoso Alarcon, Carlos
Citation	
Issue Date	2006-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/969
Rights	
Description	Supervisor:党 建武, 情報科学研究科, 博士

Trigger-Based Language Model Adaptation for Conversational Speech Transcription

(話し言葉音声認識のためのトリガーペアに基づく言語モデルの適応)

Carlos Troncoso Alarcón
北陸先端科学技術大学院大学

2006年3月

論文の内容の要旨

近年、音声認識研究の主要な対象は、書き言葉のディクテーションからより自然な話し言葉の書き起こしに移行しつつあるが、話し言葉の音声認識の精度はまだかなり低い。音声認識システムにおいて、言語モデルは可能な単語系列における制約と選好性を与えるもので、非常に重要な役割を担っている。現在最も広く使われている言語モデルは n -gram モデルで、単語間の近距離の依存関係のモデル化には強力であるが、 $n-1$ 単語の履歴で制限されていて、遠距離の依存関係をモデル化できない。本論文は、関連するキーワード (トリガーペア) により、話題に関する遠距離の単語間の関係をモデル化できるトリガー言語モデルについて扱っている。本研究の主な対象である討論や会話では話題が均質である場合が多く、トリガー言語モデルは話題に関する遠距離の制約を捕捉するのに有効であると考えられる。また、話し言葉音声において特徴的な言い淀みに対しても、キーワードの共起に基づくトリガー言語モデルは頑健であることが期待できる。

ただし、この種の遠距離の言語モデルにおいては、信頼できる統計量の推定が鍵であり、特に話し言葉においては、書き言葉に比べて学習データが一般に少ないので、大きな問題となる。本論文では、当該ドメインのデータを十分に活用して、トリガー言語モデルを話し言葉音声に適用するための手法を2つ提案する。ここでは、対象タスクに忠実にマッチしたトリガーペアを抽出する。さらに、統計的推定の信頼性を補完するために、大規模コーパスの統計量も利用するバックオフ手法を導入する。

1章は音声認識における言語モデルの二つの主なアプローチ、特に統計的言語モデルについて紹介する。

2章は主な言語モデルを概観し、提案手法の基本的概念を述べる。そして、言語モデルの評価尺度 (パープレキシティなど) と遠距離の言語モデルの適用法について説明する。

3章では旅行対話と特定の話題に関する模擬講演の音声認識を対象として、トリガー言語モデルの適用を検討する。一般に言語モデルの構築においては、タスクにマッチした学習コーパスのサイズは小さいため、統計量の学習が十分に行えず、逆に、大規模なコーパスでは一般的過ぎて、タスク依存性がなくなるという問題がある。本研究では、タスクにマッチしたコーパスからトリガーペアを抽出し、大規模なテキストコーパスからトリガーペアの生起確率を推定する手法を提案する。評価実験の結果、提案法によるパープレキシティは単一のコーパスから作成した通常のトリガー言語モデルによるパープレキシティより小さくなり、ベースラインより 12.8% の削減が得られた。

4章ではパネル討論の音声認識を対象として、トリガーモデルを用いた言語モデルの適応法を提案する。パネル討論では、与えられた話題について終始話されるので、遠距離でもキーワードの相関が期待できる。トリガー言語モデルはそのような遠距離の依存関係をとらえるためのものであるが、従来は新聞記事などの一般的すぎる大規模コーパスから構築されており、タスクに依存したトリガーペアが十分に得られない。提案法では、ベースラインモデルによる初期認識結果を使用して、当該討論に特化したトリガーペアを抽出し、またそれらの確率を推定する。確率値については、大規模コーパスから推定される統計量も利用するバックオフ手法も提案する。評価実験の結果、大規模コーパスから作成した通常のトリガー言語モデルと比較して、パープレキシティで有意な改善が得られた。また、 n -gram 言語モデルの適応と組み合わせることにより、ベースラインよりパープレキシティを 44% 削減できた。さらに、音声認識精度においても改善が得られた。

5章は論文の結論で、本研究の成果のまとめと今後の課題について述べる。