

論文内容の要旨

博士論文題目 Pronunciation Estimation with Unconstrained Many-to-Many Alignment and Online Discriminative Learning using Second Order Statistics
(無制約多対多アライメントと2次統計量を用いたオンライン識別学習による発音推定)

氏名 久保 慶伍

(論文内容の要旨)

Open-domain automatic speech recognition (ASR) or text-to-speech synthesis (TTS) used in voice application services is required to recognize or speak arbitrary words. Estimating a pronunciation which connects a voice and word plays an important role to achieve it. In this thesis, we propose a novel grapheme-to-phoneme (g2p) alignment and a novel g2p conversion method to enhance a pronunciation estimation such as g2p conversion.

The joint multigram approach which is a famous g2p alignment used as preprocess in g2p method etc. produces an alignment with fine-grained substrings which improve the expressive ability over out-of-vocabulary (OOV) words, by imposing maximum lengths of both graphemes and phonemes. Because an alignment degrading a performance in g2p method etc. is generated by the forced segment by imposing the maximum lengths, we propose the unconstrained many-to-many alignment introducing city block distance to achieve an alignment with fine-grained substrings without imposing the maximum lengths. We also propose the extensions for our proposed alignment, which are parameter estimation by n-best Viterbi training, and the merge method in an incorrect alignment. Our proposed alignment with extensions improved the expressive ability over OOV words with a 20.3% error rate reduction compared with the joint multigram approach.

Using our proposed alignment as a preprocessing, we propose g2p methods based on an online discriminative learning method called structured Adaptive Regularization of Weights (structured AROW) and structured Narrow AROW (structured NAROW), to address the overfitting problem in a current state-of-the-art g2p method based on the Margin Infused Relaxed Algorithm (MIRA). Structured AROW improves MIRA's overfitting problem not by extremely moving parameters, and by individually controlling the move for each parameter by the confidence represented by second order statistics. Structured NAROW improves the hyperparameter in structured AROW. Structured NAROW significantly improves performances over MIRA on various g2p tasks.

(論文審査結果の要旨)

音声検索といった新語を扱う音声認識を利用したシステムにおいて、新語に正しい発音を付与する自動発音推定の重要性が増している。その理由は音声認識が単語の発音の情報に基づいて音声とその単語の照合を行うことから、辞書に登録されていない新語などを認識可能にするために、その発音の情報が必要となるからである。発音推定手法として現在、データ駆動型の書記素-音素変換(g2p 変換)と、単語とその発音のペアを膨大な Web テキスト内から自動抽出する Web テキストマイニングによる発音獲得がある。本論文では、この課題を解決するため、下記の2つの研究課題に取り組んだ。

1. 発音推定手法を強化するための書記素と音素間の新しいアライメント手法の提案

書記素と音素間のアライメントは発音推定のための事前処理であり、発音推定の性能向上のために正確できめ細かいアライメントが求められる。それを実現するために、市街地距離を導入した無制約多対多アライメントとその拡張手法(n-best Viterbi トレーニングによる学習とコンテキストに基づく正確なアライメント法)を提案する。その提案手法は従来のアライメント手法より正確できめ細かいアライメントを生成する。

2. g2p 変換のための新しい学習手法の提案

2 次統計量を用いたオンライン識別学習である構造化 AROW と構造化 NAROW を提案した。構造化 AROW は過学習に頑健な二値分類手法である重みベクトルの適応的正則化 (AROW) を g2p 変換用に拡張した手法である。これは現在、最も精度の高い MIRA と呼ばれるオンライン識別学習よりも、誤った発音も含む学習データに対して過学習を起こさずに精度良く学習することができる。また、構造化 NAROW は構造化 AROW が持つハイパーパラメータの設定を改善する。

各提案手法に関して評価実験を行ない、拡張された無制約多対多アライメントは、先行研究である joint multigram 手法と比べて、より正確な g2p 変換を実現し、Web テキストマイニングを用いた発音獲得の改善にも寄与すること、構造化 NAROW は様々な g2p タスクにおいて現在、最も精度の高い g2p 変換である MIRA を有意な差で改善することを明らかにした。本研究成果は、2編の学術論文、4編の国際会議論文、3編の国際ワークショップ論文として発表されている。以上より、平成 25 年 12 月 17 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。