論文 / 著書情報
Article / Book Information

論題(和文)	教師なしクロスバリデーション適応によるタスク適応
Title(English)	
著者(和文)	久保田 雄, 篠崎 隆宏, 古井貞熙
Authors(English)	Takahiro Shinozaki, Yu Kubota, SADAOKI FURUI
出典(和文)	日本音響学会2009年春季講演論文集, , No. 1-5-11, pp. 29-30
Citation(English)	, , No. 1-5-11, pp. 29-30
発行日 / Pub. date	2009, 3

教師なしクロスバリデーション適応によるタスク適応*

◎久保田雄, 篠崎隆宏, 古井貞熙 (東工大)

1 はじめに

音声認識システムにおいて高い認識精度を得るためには認識タスクに適合した大量のデータを用いて音響モデルを学習する必要がある。しかし全てのタスクにおいて大量の学習データを用意することはコストがかかり過ぎ非現実的であり、既存のデータベースから学習したモデルを少量のタスク依存データを用いて適応化させることが実用上非常に重要となる。

音響モデルのタスク適応には、教師あり適応技術または教師なし適応技術を用いることができる。教師あり適応では音声とともにその書き起こしを用いてモデル適応を行うため、性能の高いモデルを得ることができる反面、少量ながら依然としてタスクごとに書き起こしを行わなければならない欠点がある。

これに対し教師なし適応では音声のみを用いてモデルの適応化を行う。教師なし適応には様々な手法が存在するが、適応対象音声の音声認識をまず行い、得られた認識仮説を近似的な書き起こしとしてモデル更新を行う方法が一般的である。人手による書き起こしを全く用いないため様々なタスクに対し簡便に音声認識技術を応用する上で非常に有利であが、得られるモデルの性能が教師あり適応に比べて低くなる問題がある。これは認識仮説には認識誤りが避けられず、それにより適応化性能が低下するためである。

本研究では教師なし適応における認識誤りの影響を低減する手法として我々が提案し話者適応において有効性を示した教師なしクロスバリデーション (CV) 適応手法 [1] をタスク間適応に応用し、タスク間適応においても教師なし CV 適応手法が有効であり、従来のバッチ型適応よりも高い認識性能を持ったモデルが得られることを示す。

2 教師なしクロスバリデーション (CV) 適応手法

従来のバッチ型教師なし適応手法では、適応対象音声の認識処理と認識仮説を用いたモデルの適応を同じデータを用いて交互に繰り返す。認識処理により得られた認識仮説には誤りが存在するが、適応の繰り返しによりこの誤りは減少する。しかし、適応時に認識誤りも正しいものとして処理が行われ、さらにそのようにして適応されたモデルがまた同じデータの認識に用いられるため、逆に強化される誤りも存在する。このため、認識率の向上はしばらくすると収束する。すなわち、この誤りの強化を抑えることが教師なしタスク適応を効果的に行うために重要となる。認識誤りが強化される問題に対処するための手段として、CV 適応手法 [1] では認識処理とモデル更新処理で使用するデータを CV 的な方法により効果的に分離する。認識処理とモデル更新処理で誤りを含めデータ自体が分離されるため、認識誤りが繰り返し強化されることを防ぎ、適応性能を向上させることができる。

3 教師なしタスク適応

教師なしタスク適応は、対象とするタスクの書き起こしのない適応用音声データのみを用いて、そのタスクの音声データの認識率を向上させるための手法である。本実験では読み上げ音声コーパスから学習した不特定話者音響モデルを初期モデル、数時間程度の話し言葉音声をタスク適応用データ、話し言葉音声を目的タスクの評価音声として、実験を行う。さらに補足実験として、教師なしタスク適応によって得られた不特定話者モデルに対して教師なし話者適応を適用し、タスク適応と話者適応を組み合わせた場合の効果についても合わせて検討する。

4 実験条件

認識には東京工業大学で開発を行っている WFST 音声認識デコーダー (T^3 decoder) [2] を用い、適応には HTK [3] を用いた。初期モデルとして用いた音響モデルは、新聞記事読み上げ音声コーパス (JNAS) から EM により学習した状態共有トライフォン HMM であり、状態数 2000、状態ごとの混合数 32 である。教師なし CV 適応の際の学習セット分割数は、予備実験から 20 とした。特徴量は MFCC12 次元とパワー、および

* Application of the unsupervised cross-validation adaptation algorithm to task adaptation. by Yu Kubota, Takahiro Shinozaki, and Sadaoki Furui (Tokyo Institute of Technology)

それらのΔとΔΔの計39次元である。評価セットには、日本語話し言葉コーパス (CSJ) の男性話者による学会講演10講演からなる標準の評価セットを用いた。これらの講演話者は全て異なり、1講演は10分から20分程である。タスク適応は、適応データとしてCSJの学会講演から1講演あたり5分を単位として必要な時間分だけランダムに選択したものを用いて、MAP [4] 手法により行った。適応の繰り返し数は5回である。話者適応は、話者ごとにMLLR [5] を10回繰り返すことにより行った。

5 実験結果

5.1 教師なしタスク適応

図1は教師なしタスク適応の結果で、横軸は適応データの量を、縦軸は最終的な単語誤り率を示す。またグラフ中でInitは適応をかける前の不特定話者モデルの場合、Batchは従来法の教師なしバッチ型適応をかけた場合、CVは教師なしクロスバリデーション適応をかけた場合をそれぞれ示している。

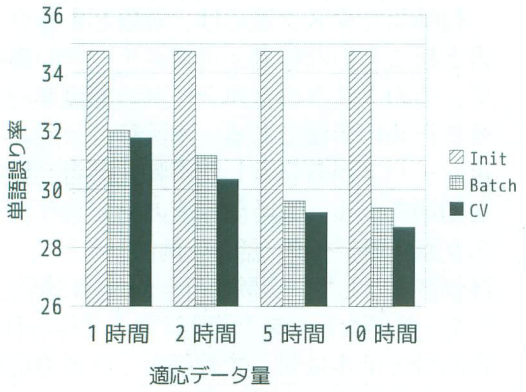

Fig. 1 教師なしタスク適応

全ての適応セット量において常にCV適応手法が従来法よりも低い単語誤り率となり、教師なしタスク適応においてもCV適応手法が効果的であることが示された。特に適応セットが2時間分の場合に提案法と従来法とでの単語誤り率の差が最大になった。その場合Initの単語誤り率34.7%に対し、従来法での単語誤り率は31.2%、提案法を用いた場合は30.4%であり、従来法を基準とした単語誤り率の相対削減率は2.6%であった。

5.2 タスク適応後の教師なし話者適応

図2は2時間分の適応用音声データでタスク適応を行った音響モデルに対し教師なし話者適応を行った結果で、横軸は適応を行った回数、縦軸はそれぞれの単語誤り率を示す。

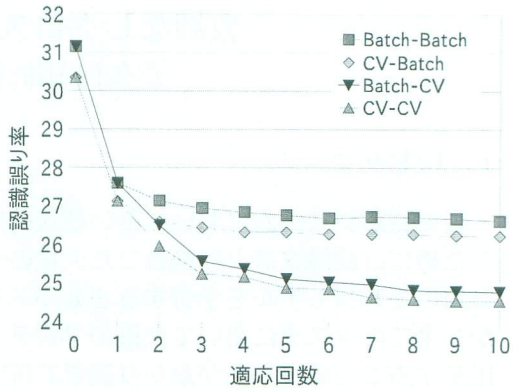

Fig. 2 教師なしタスク適応と話者適応の組み合わせ: (例:CV-Batchは教師なしCVタスク適応を行いその後教師なしバッチ型話者適応を行った場合)

タスク適応にバッチ型適応手法とCV適応手法を用いた場合を比較すると、話者適応にバッチ型適応手法およびCV適応手法のどちらを用いた場合についても、話者適応手法を10回繰り返した後でもタスク適応時のCV適応手法の効果が残り、より低い単語誤り率が得られることが分かる。単語誤り率の最低値は、タスク適応と話者適応の両方にCV適応手法を用いた場合に得られた。話者適応を10回繰り返した後での単語誤り率は、タスク適応と話者適応の両方に従来法を用いた場合が26.6%、CV適応法を用いた場合が24.5%であり、従来法の適応結果を基準としたCV法による単語誤り率の相対削減率は7.9%であった。

6 まとめ

教師なしCV適応法をタスク適応に応用し、話者適応とともにタスク適応においても同手法が有効であることを示した。

参考文献

[1] 篠崎隆宏 他, 日本音響学会 (2009春), 1-5-10.
[2] P.R.Dixon 他, IEEE ASRU, 443-448, 2007.
[3] S.Young 他, The HTK Book, Cambridge University Engineering Department, 2005.
[4] C.H.Lee 他, IEEE Transactions on Signal Processing, 39, 199-206, 1999.
[5] C.J.Leggetter, P.C.Woodland, Eurospeech, 1155-1158, 1995.