論文 / 著書情報
Article / Book Information

論題(和文)	話者クラスタに基づく初期モデルを用いた話者適応
Title(English)	
著者(和文)	張 志鵬, 古井貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	日本音響学会 2000年春季講演論文集, Vol. , No. 3-9-8, pp. 109-110
Citation(English)	, Vol. , No. 3-9-8, pp. 109-110
発行日 / Pub. date	2000, 3

話者クラスタに基づく初期モデルを用いた話者適応 *

◎張　志鵬　古井 貞熙 (東工大)

1. はじめに

我々はこれまでに、話者交代を混合ガウス分布モデル (GMM) の尤度比較によって自動的に検出しながら、オンライン逐次話者適応を行う手法 [1] を提案し、ニュース音声認識におけるその有効性を確認した。この手法では、男女別の不特定話者モデル (SIHMM) を初期モデルとして用いていたが、話者クラスタに基づく複数の初期モデルを用いた話者適応がより有効と考えられる。これまでに提案された話者クラスタによる話者適応法（例えば [2]）では、各クラスタの HMM に対する尤度を用いてクラスタを選択する必要があるため、認識のための計算量が膨大になる問題点があった。本稿ではまず、話者クラスタによって作成した初期モデルを用いる話者適応法の有効性について検討し、次に計算量を削減するために GMM の尤度比較によるクラスタ選択を用いる方法を提案する。

2. 話者クラスタ

2.1 話者クラスタによる話者適応

入力音声に最も適合した話者クラスタの HMM を選択して認識に用いる手法は、よく利用される話者適応法の一つである。この手法は、入力音声に対して尤度最大となるモデルを選んで認識に用いるだけで、モデルのパラメータを変えないので、クラスタ選択が能率的にできれば、計算量が少なくて済む。幾つかのクラスタのモデル間の内挿によって新しいモデルを構築し、認識を行う方法も提案されている。

2.2 HMM 距離

多数話者の音声データを直接クラスタ化することは困難なので、各話者の音声から作成された特定話者 HMM(SDHMM) のクラスタリングを行う。このために、状態数の等しい二つの HMM 間の距離を以下のように、式 (1) で定義する。但し、b_{js} はモデル j の状態 s 混合 m の出力確率分布、S は状態数、M は混合数をそれぞれ表す。ここでは、話者間の HMM の共分散、混合重みおよび状態遷移確率の違いは無視し、各混合の平均値の違いだけを考慮している。

$$d(i,j) = -\frac{1}{S \cdot M} \sum_{s=1}^{S} \sum_{m=1}^{M} \{\log[b_{js}(\mu_{ism})] + \log[b_{is}(\mu_{jsm})]\} \quad (1)$$

2.3 クラスタリングアルゴリズム

各 SDHMM 間の距離に基づいて、クラスタリングを行う。ここでは、SPLIT 法 [3] で用いられたクラスタリングアルゴリズムを用いた。この手法は LBG 法とは異なり、歪みが最大となるクラスタを順序分割するため、任意の数のクラスタを作成できる。クラスタリングする前に各話者間の距離行列を作成し、これを用いるのでクラスタ中心の初期値を与えなくてもよいというメリットもある。クラスタ数が増加するにつれて、尤度が徐々に増加する。あらかじめ尤度の閾値或いはクラスタ数を指定すれば、自動的にクラスタリングの結果が得られる。

2.4 クラスタリング実験

SIHMM の学習に使われたものと同じデータ (ATR の B セット、ASJ バランス文と ASJ 案内タスク) を用いてクラスタリングを行った。男性 53 名、女性 56 名の話者からなっている。各 SDHMM は、SIHMM を初期値として、Baum-Welch アルゴリズムによる連結学習を行って構築した。各クラスタの HMM は、クラスタに属するあらゆる話者のデータを用いて構築した。

3. 認識実験

3.1 音響モデルと言語モデル

今回の実験で用いた HMM は、tree-based clustering によって状態共有化を行なった文脈依存音素 HMM である。音響特徴量としては 16 次の LPC ケプストラムと正規化対数パワー、及びそれらの一次微分の計 34 次元を用いた。モデルの総状態数は男性が 2106、女性が 2083 である。各状態のガウス分布の混合数はすべて 4 である。

言語モデルの学習に用いたデータは放送ニュース原稿テキスト 5 年分、約 50 万文である。単語出現頻度上位 2 万語を認識語彙とし、間投詞と読みを考慮した言語モデル [4] を用いた。

3.2 評価用データ

実際に放送されたニュース音声から、スタジオで収録されたクリーンな発話をそれぞれ男女 50 文ずつ抽出した。各評価セットには 5〜6 名の話者の音声が含まれている。

3.3 認識実験結果

まず女性テストセットを用いて、HMM を直接用いてクラスタ選択を行う方法により認識実験を行った。認識する際は、まず SIHMM を用いて尤度を計算すると同時に、結果としてのラベルファイル（音素セグメンテーションリスト）を作成する。次に各クラスタの HMM モデルと、このラベルファイルを用いて、入力音声に対しリスコアリングし、そのクラスタの尤度を計算する。SIHMM と各クラスタ HMM の尤度から一番高い尤度を示す HMM を選び、再認識を行う。この手法によれば、尤度最大のクラスタ選択が保証される。bigram を用いる場合と trigram を用いる場合の実験結果を図 1 に示す。図には、各ク

---

*Speaker Adaptation Using Initial Models Made by Speaker Clustering
By Zhipeng Zhang and Sadaoki Furui (Tokyo Institute of Technology)

ラスタ数における単語誤り率を示す。結果から、クラスタ数が 4 の場合に一番良い結果が得られることがわかる。

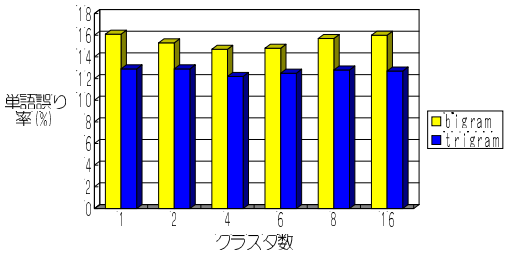

図 1．各クラスタ数に対する認識結果

4. GMM による計算量の削減

4.1 GMM によるクラスタ選択

前述した手法では、クラスタ選択の際、SIHMM による認識を行った後に各クラスタの HMM でリスコアリングすることが必要なので、膨大な計算量を要する。この問題に対処するために、GMM をクラスタ選択に用いることを試みる。GMM は話者の特徴を反映でき、かつ簡単なモデルなので、テキスト独立型話者認識に広く使われている。GMM を尤度比較に用いるために、まず HMM の学習と同じデータで不特定話者と各クラスタの GMM を構築する。GMM の混合数は話者認識実験の結果などをもとに 64 とした。入力音声に対して、不特定話者と各クラスタの GMM に対する尤度を計算し、その中の一番高い尤度を示す GMM モデルに対応する HMM を選んで認識を行う。

4.2 認識実験結果

4 クラスタを用い、bigram を用いる場合と trigram を用いる場合の男女別の実験結果を図 2 に示す。図には、適応化を行う前 (baseline)、HMM の尤度比較方法（”HMM”）、提案する GMM の尤度比較方法（”GMM”）による単語誤り率を示す。いずれの評価セットに対しても、適応化により誤り率が低下していることが分かる。”HMM”法により、適応化前に比べて誤り率は男女平均で 7.0%(相対値) 低下している。”GMM”法では、適応化前に比べて誤り率は男女平均で 4.1%低下している。

4.3 オンライン逐次話者適応

次に、オンライン逐次話者適応を行った。まず GMM によって尤度最大のクラスタを選び、そのクラスタの HMM を初期モデルとして、文献 [1] に報告した方法によってオンライン逐次適応を行った。実験結果を図 3 に示す。図には、適応化を行う前 (baseline)、不特定話者モデルを初期モデルとしての逐次適応法（”1 cluster”）、および提案したモデル選択法（”4 cluster”）による誤り率を示す。モデル選択法により不特定話者モデルを初期値とする適応化法よりも、誤り率が低下することが確認された。適応化前に比べて誤り率は男女平均で 11.6%低下している。

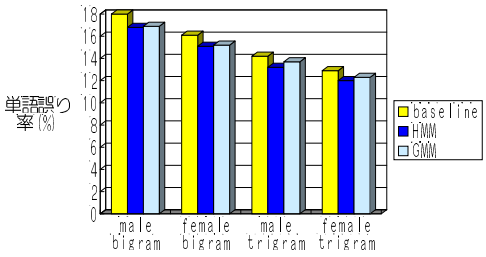

図 2．提案した GMM による尤度比較法と HMM 法の認識結果

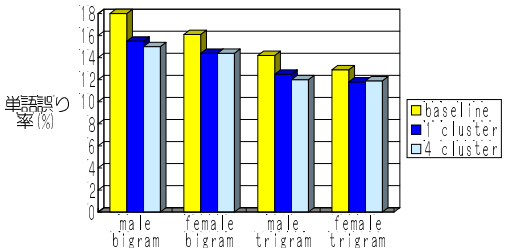

図 3．GMM によって選択したクラスタの HMM を初期モデルとする逐次適応による認識結果

5. まとめ

話者クラスタリングによって作成した初期モデルを用いる話者適応法について検討した。計算量を削減するために GMM を用いるクラスタ選択法を提案した。適応実験によって提案した手法の有効性が確認された。今後は音声区間の自動切り出し、文の区切りの自動決定、雑音への対処法などと組み合わせていく予定である。

謝辞

ニュース原稿及び音声データを提供して頂いた NHK 放送技術研究所に感謝します。ご助言を頂いた NTT サイバースペース研究所の大附克年氏に感謝します。日頃討論頂く東工大の研究室の方々に感謝します。

参考文献

[1] 張, 古井 秋季音学議論,pp.45-46, 1999-9

[2] 小坂, 他 信学論,Vol.J78-D pp.1-9, 1995-1

[3] 管村, 他, 音声研究会資料,S82-64, 1982-12

[4] 桜井 他, 春季音学議論,pp.57-58, 1999-3