

論文 / 著書情報
Article / Book Information

論題(和文)	木構造化された確率分布を用いた話者適応化
Title(English)	
著者(和文)	篠田浩一, 渡辺隆夫
Authors(English)	Koichi Shinoda
出典(和文)	日本音響学会平成7年度春季研究発表会講演論文集, Vol. 2-5-10, No. , pp.
Citation(English)	, Vol. 2-5-10, No. , pp.
発行日 / Pub. date	1995,

スペシャル・セッション [音声認識における適応化技術] 2—5—10 木構造化された確率分布を用いた話者適応化 *

◎ 篠田 浩一 渡辺 隆夫
(NEC 情報メディア研究所)

1. はじめに

話者適応化では、話者の発声数が極めて少ないとき、認識性能が向上せず、しばしばかえって劣化する。これは、適応化に用いるデータ量が少ないため、モデルパラメータの推定精度が低いことが原因である。従来、何らかの手段で実質的なパラメータ数を減少させることでデータ量が少量のときの認識性能を高めることが行なわれてきた [1, 2]。しかし、パラメータ数を絞ると、少ないデータ量で認識性能が飽和してしまい、それ以上データを増やしても性能が向上しないという問題が起きる。もし、適応化自体にパラメータ数をデータ量に応じて自律的に調節する仕組みがあれば、データ量の増加に応じてパラメータ数を徐々に増やすことにより、その時点のデータ量に対し最適な適応化が行なわれると期待できる。本稿では、そのような仕組みを実現する一手法を述べる。

2. 木構造を用いた自律型話者適応化

2.1. 原理

適応化用のデータ量によってパラメータ数を自己調節する仕組みとして、パラメータの木構造を導入する (図 1)。パラメータとしては、例えば、連続分布 HMM の出力確率分布の平均ベクトルの適応前後の差分 (適応化ベクトル) を想定している。木構造の作成方法には自由度がある。最も下層のノード (リーフノード) にパラメータを対応づけ、各ノードごとにその下に対応付けられたパラメータ群を代表するパラメータ (代表パラメータ) を定義する。

適応化においては、入力データサンプルとリーフノードとの対応付けを行ない、その対応付けられたデータを用いて木構造の各ノードの代表パラメータを推定する。中間ノードの代表パラメータの推定には、その下のすべてのリーフノードのデータサンプルを用いる。

パラメータの更新には、そのパラメータに対応するリーフノードとその親ノードの代表パラメータを用いる。上層のノードの代表パラメータは、多くのデータサンプルを用いて推定されているため精度が高いが、反面、多くのパラメータをまとめたものであるため、個々のパラメータの推定値としては正確な値ではない。下層に行くに従い、ノードの代表パラメータの値は、その下の個々のパラメータの値に近づくが、データサンプル数が少ないため精度は下がっていく。そこで、図 1 に示すように、データ量が少なくなるときには上層のノードの代表パラメータを用い、多くなるとに従って木構造を下に辿り、より下層の代表パラメータを用いるようにする。用いるノードの階層は各パラメータごとに異なってもよい。データ量の増加にともない、適応化に用いるパラメータ数は増加する。

この手法は、パラメータの補間・平滑化を用いる話者適応化法 [3, 2] において、パラメータの近傍関係

を階層的に定義して拘束の強弱を階層間の上下関係で表し、データ量に従って拘束の強弱を調節する方法と見なすことができる。また、スペクトル空間の階層的クラスタ化による話者適応化 [4] において、適応化に使用する階層をデータ量に従って選択する手法と捉えることも可能である。

2.2. 確率分布の木構造を用いた実現

提案法の一つの実現として、認識方式に混合連続分布 HMM を用い、パラメータに各ガウス分布の適応化ベクトルを用いる場合について述べる。提案法の性能は、木構造の作成方法に大きく左右される。過去の研究 [3, 5] から、確率分布空間において距離の近い出力確率分布はその適応化ベクトルの値も近いことが予想される。そこで、木構造は確率分布間の距離を用いて作成する。

1. 初期 HMM の全状態にわたる出力確率分布の集合の木構造を作成する。リーフノードには各分布を 1 対 1 に対応させる。木構造は、分布間の距離を Kullback-divergence で定義し、k-means アルゴリズムを用いてトップダウンで作成する [6]。図 1 に示すように、階層数を S とし、上から $1, \dots, S$ とする。
2. Viterbi アルゴリズムを用いて、入力音声の特徴ベクトル x_t と正解単語 HMM の各分布 k との対応付けを行ない、各分布、すなわちリーフノードにおける適応化ベクトル Δ を求める。

$$\Delta(S, k) = (\sum_{x_t \in k} x_t) / n(S, k) - \mu_k \quad (1)$$

ここで、 μ_k は分布 k の平均ベクトル、 n はノードに付随するデータサンプルの個数を示す。

3. リーフノード以外のノードにおける適応化ベクトルを計算する。

$$n(s, k) = \sum_{k' \in k} n(s+1, k')$$
$$\Delta(s, k) = \frac{\sum_{k' \in k} n(s+1, k') \Delta(s+1, k')}{n(s, k)}$$

ここで $k' \in k$ は、 k' が k の子ノードであることを示す。

4. 予め定められたデータサンプル数の閾値 N を用いて、各リーフノード k から上層ノードにたどる過程でデータサンプル数 n が初めて N を越えるノード ($s'; k'$) を求める。そのノードの適応化ベクトルを用いて平均ベクトルの更新を行なう。

$$\hat{\mu}_k = \mu_k + \Delta(s', k')$$

上層のノードの適応化ベクトルは、下層のノードのものに比べ、ベクトル空間においてより大局的な変化をあらわすと考えられる。平均ベクトルは、データ量の少ないときには上層のノードの大局的な適応化ベクトルを用いて適応化される。話者による特徴量の大きな相違、あるいは、急激な周囲環境の変化に対し、速やかに追従することができる。そして、データ量の

* Speaker Adaptation Using Tree-Structured Probability Density Function,
by Koichi SHINODA and Takao WATANABE
(NEC Corporation)

増加に従い、より下層のノードの適応化ベクトルが用いられ、局所的なきめ細かい適応が行なわれる。十分なデータ量があるときには、リーフノードの適応化ベクトルがそのまま用いられ、木構造を用いない平均ベクトル推定と等価になる。また、ここで説明した以外にも、例えば、全データ量に従って階層を選択し全分布に対しその階層の適応化ベクトルを用いる、あるいは、複数階層にわたるノードの適応化ベクトルを重み付けて用いる、などの方法も可能である。

3. 評価実験

3.1. 実験条件

提案法の評価として5000単語認識をシミュレートした類似100単語認識実験[7]を行なった。初期HMMとして、男性23名女性20名の音素バランスを考慮した250単語1回発声を用いて学習された不特定話者HMMを用いた。ガウス分布数は各状態について2であり、全分布数は約1500である。木構造は、5階層で、1~3階層の各ノードからの分岐数は4、4階層目からリーフノードへの分岐数はノードにより異なる。評価話者として上の43名には含まれない男性5名女性5名を用意し、適応化用に250単語1回発声、認識評価用に語彙の異なる250単語1回発声を用いた。入力音声は、標本周波数16kHz、分析周期10ms、分析窓長32ms、周波数帯域0.1~7.2kHzの条件で分析され、特徴量として、メルケプストラム10次元、メルケプストラム差分10次元、およびパワー差分を用いた。

3.2. 実験結果

提案法の話者10名の平均認識率を図2に示す。図中SIは不特定話者認識の認識率、refは、従来法として、ベイズ事後確率最大基準[8]とスペクトル内挿話者適応化法[3, 5]をともに用いた場合の結果を示す。例えば、データ量サンプル数の閾値 N が20のとき、適応化用単語数10単語で不特定話者認識の認識率83.5%のところ提案法の認識率は88.0%と、誤り率が3/4になり、効果が確認できた。図3は、閾値 N が20のときの提案法の結果を、すべての分布に対し同一の階層の適応化ベクトルを用いて適応化したときの提案法と比較したものである。提案法は、単語数の少ないときには上位の階層の傾向を、単語数の多いときには下位の階層と同様の傾向を示し、各階層ごとの適応化の認識性能の高い部分を、うまく切り出していることがわかる。

4. おわりに

木構造を用いた自律型話者適応化法を提案し、その基本性能を明らかにした。今回は提案法の一実現例を示したが、検討すべき選択肢は多く、今後は順次それらを実現していく予定である。

参考文献

- [1]高橋他: 音学講論, 3-8-3 (1994.10)
- [2]大倉他: 音学講論, 2-Q-17 (1992.3)
- [3]篠田他: 音学講論, 1-8-12 (1990.9)
- [4]S. Furui, Proc. ICASSP-89, pp.286-289(1989)
- [5]篠田他: 信学論 (A), J77-A,2(1994)
- [6]渡辺他: 音学講論, 1-8-7 (1993.10)
- [7]渡辺他: 信学論 (D-II), J75-D-II,8(1992)
- [8]C.H.Lee 他: IEEE Trans. ASSP, vol.39, no.4(1991)

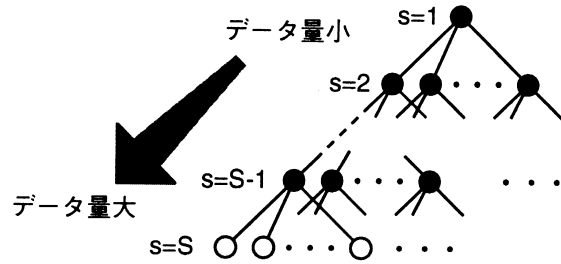


図1: パラメータの木構造

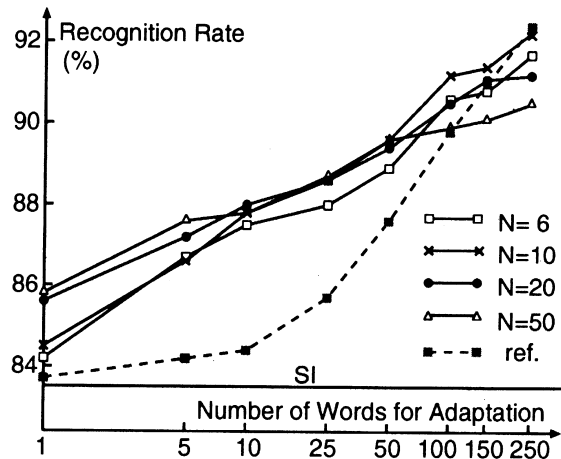


図2: 木構造を用いた自律型話者適応化の効果

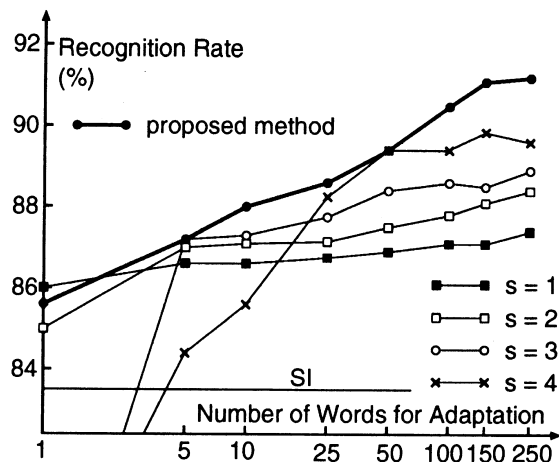


図3: 階層別の適応化との比較