

論文 / 著書情報
Article / Book Information

論題(和文)	記述長最小原理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年度秋季研究発表会講演論文集, Vol. 3-2-12, No. , pp.
Citation(English)	, Vol. 3-2-12, No. , pp.
発行日 / Pub. date	1995,

◎ 篠田 浩一 渡辺 隆夫
(NEC 情報メディア研究所)

1. はじめに

話者適応化の実用場面では適応化に用いるデータ量はしばしば変化する。適応化の対象として、自由パラメータ数の多い複雑なモデルを採用すると、データ量が少ないときに推定が不安定になる。逆に、自由パラメータの少ない簡単なモデルを用いると、データ量が多いときに性能が低く抑えられる。従って、各々の時点で得られるデータ量に対し、それに適した適当な複雑さをもつモデルを選択し、それに対して適応化を行なうことが望ましい。

筆者らは先にこのような考え方にに基づき、木構造確率分布を用いた自律型話者適応化法を提案した [1]。この手法では、適応化に使用するモデルの選択のために、データ量閾値と呼ぶ、実験的に求める必要のある調節パラメータを用いていた。本稿では、モデル選択において、このような経験に基づくパラメータを用いない方法として、記述長最小原理を用いる手法を提案・評価する。

2. 記述長最小 (MDL) 原理

記述長最小 (Minimum Description Length; MDL) 原理 [2, 3] は、最近の情報理論および計算論的学習理論の研究から、データに対し最適な確率モデルを選択する問題において、有効であることが実証されている [4]。MDL 原理は、確率モデル $i = 1, \dots, M$ のうち、データ $x^N = x_1, \dots, x_N$ の記述長を最小にするモデルが最適なモデルであると主張する。記述長は以下の式で表される。

$$l_R^{(i)} = -\log P_{\hat{\theta}^{(i)}}^n(x^N) + \frac{k^{(i)}}{2} \log N + \log M \quad (1)$$

ここで、 $k^{(i)}$ はモデル i の次数、 $\hat{\theta}$ はデータ x^N に対するモデル i のパラメータ $\theta^{(i)} = (\theta_1^{(i)}, \dots, \theta_{k^{(i)}}^{(i)})$ の最尤推定量である。第 1 項はデータに対する対数尤度を符号反転させたもの、第 2 項はモデルの複雑さを表す量である。モデルが複雑になるほど、第 1 項は小さくなり、第 2 項は大きくなる。このように両者の間にはトレードオフがあり、ある適当な複雑さをもったモデルが最適なモデルとして選択される。

3. MDL 原理を用いた話者適応化

ここでは、混合連続密度分布 HMM における各状態のガウス分布の平均ベクトルを適応化の対象とする。Viterbi アルゴリズムを用いて、入力音

* Speaker Adaptation Using Minimum Description Length Principle, by Koichi SHINODA and Takao WATANABE (NEC Corporation)

声のフレームと適応化の初期 HMM の各ガウス分布との対応付けを行い、各々のフレームの特徴ベクトルと対応するガウス分布の平均ベクトルとの差分 $\delta^N = \delta_1, \dots, \delta_N$ を計算する。ここで N は総データフレーム数である。これら δ^N は対角分散行列をもつ多次元正規分布に従うと仮定し、その平均値を適応化に用いることとする。今、この δ^N を出力する確率モデルを推定する問題に対して MDL 原理を適用する。

データ δ^N に対する様々な複雑さの確率モデルを以下のように用意する。まず、HMM のガウス分布の集合に対し階層的なクラスタリングを行ない、木構造を作成する。リーフノードは HMM の各分布と 1 対 1 に対応する。木構造のそれぞれのノードにおいて δ の確率分布が計算される。その際、リーフノード以外のノードの確率分布は、その下のすべてのリーフノードに付随する δ を用いて計算される。今、木構造を上下に分断するノードの集合を「カット」と呼ぶ。各々のカットがそれぞれ一つの確率モデルに対応する (図 1)。あるカットが選択された場合、カットを構成する各ノードの δ の平均値が、その下のリーフノードに対応する分布の適応化に用いられる。

データ δ^N に対する、ノードの集合 $j^{(i)} = 1, \dots, J^{(i)}$ で構成される、あるカット i の記述長は以下のように計算される。

$$l_R^{(i)} = \sum_{n=1}^N \sum_{l=1}^L \left(\frac{1}{2} \log 2\pi\sigma_{l,j^{(i)}(n)}^2 + \frac{(\delta_n - \Delta_{l,j^{(i)}(n)})^2}{2\sigma_{l,j^{(i)}(n)}^2} \right) + J^{(i)} L \log N + \log M \quad (2)$$

ここで、 L はガウス分布の次元数、 $j^{(i)}(n)$ はカット i において δ_n に対応するノードの番号、 $\Delta_{l,j^{(i)}(n)}$ 、 $\sigma_{l,j^{(i)}(n)}$ は、それぞれ入力データから求めた δ の平均値と分散である。(2) 式において、第 3 項はカットによらず一定なので無視できる。式変形の結果、以下の量を最小にするカットを選択すればよいことがわかる。

$$l_R^{(i)'} = \sum_{j^{(i)}=1}^{J^{(i)}} n_{j^{(i)}} \sum_{l=1}^L \log \sigma_{l,j^{(i)}}^2 + 2J^{(i)} L \log N \quad (3)$$

ここで、 $n_{j^{(i)}}$ はノード $j^{(i)}$ に対応するデータ δ の数である。

なお、実装においては、次に示す方法を用いて探索に要する計算量を削減している。この方法では必ずしも最適解は保証されない。

- 1.初期値として、ルートノードのみのカットを与える。ルートノードに展開フラグを立てる。
- 2.展開フラグの立つノードのなくなるまで以下の3.~5.の処理を行なう。
- 3.展開フラグの立つノードの中から1つのノード j を選択する。 j がリーフノードのときには、 j の展開フラグを下げ使用フラグを立てる。2.に戻る。
4. j を用いたときの記述長 l_p と、 j の子ノードすべてを用いたときの記述長 l_c を計算する。
5. $l_p > l_c$ のとき、 j の展開フラグを下げ、子ノードすべてに展開フラグを立てる。 $l_p < l_c$ のとき、 j の展開フラグを下げ、 j の使用フラグを立てる。
- 6.使用フラグの立つノードの集合を最適なカットとする。

4. 評価実験

4.1. 実験条件

評価実験として、半音節を認識単位とした音声認識システム上で、5000 単語認識をシミュレートした類似 100 単語認識実験 [5] 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、分布間の距離をダイバージェンスで定義し、k-means アルゴリズムを用いてトップダウンに作成した [6]。木構造は、2分木 12 階層とした。ただし、最下層におけるリーフノード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。

4.2. 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、ベイズ事後確率最大基準 [7] とスペクトル内挿話者適応化法 [8, 9] をともに用いた場合の認識率を示す。また、図 2 には、データ量を基準とした自律型話者適応化 [1] の認識率もデータ量閾値が 10、20、50 の場合について示した (D(10), D(20), D(50))。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 87.3% と、誤り率が 3/4 になり、効果が確認できた。データ量を基準とした話者適応化に比べ性能的にはやや劣るが、調節パラメータをもたないという利点を考え合わせれば、実用的には十分な性能である。

5. おわりに

MDL 原理を用いた話者適応化法を提案し、その基本性能を明らかにした。今後は、木構造の作成方法、ノード探索の方法などを工夫し、さらなる性能向上を目指したい。

謝辞

MDL 原理を用いるにあたってご助言を頂いた NEC C&C 研究所の安倍直樹、竹内純一、李航の各氏に感謝致します。

参考文献

- [1] 篠田他: 音学講論, 2-5-10 (1995.3)
- [2] J.Rissanen: *IEEE Trans. IT*, vol.30 (1984)
- [3] 山西、韓: 人工知能学会誌, Vol.7, No.3 (1992.5)
- [4] 李、安倍: 言語処理学会発表論文集, p.53 (1995.3)
- [5] 渡辺他: 信学論 (D-II), J75-D-II,8(1992)
- [6] 渡辺他: 音学講論, 1-8-7 (1993.10)
- [7] C.H.Lee 他: *IEEE Trans. ASSP*, vol.39(1991)
- [8] 篠田他: 音学講論, 1-8-12 (1990.9)
- [9] 篠田他: 信学論 (A), J77-A,2(1994)

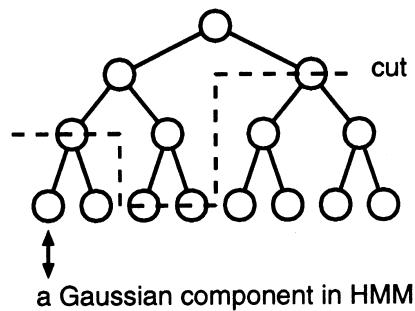


図 1: 木構造

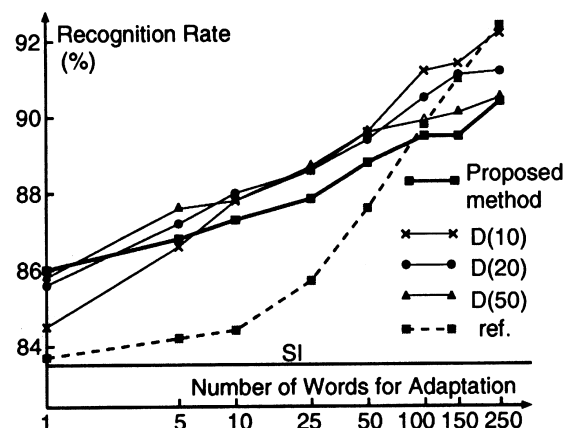


図 2: 5000 単語認識実験結果