

論文 / 著書情報  
Article / Book Information

論題(和文)	MDL基準を用いたHMMサイズの削減
Title(English)	
著者(和文)	篠田浩一, 磯健一
Authors(English)	Koichi Shinoda
出典(和文)	日本音響学会2002年春季研究発表会講演論文集, 2-5-3, No. , pp. 79-80
Citation(English)	, 2-5-3, No. , pp. 79-80
発行日 / Pub. date	2002, 3

## MDL 基準を用いた HMM サイズの削減\*

○ 篠田 浩一† 磯 健一 (NEC マルチメディア研究所)

## 1. はじめに

音声認識における認識方式として、混合ガウス分布を出力確率分布とした隠れマルコフモデル (HMM) がしばしば用いられている。そこでは、各状態におけるガウス分布数は状態によらず一定であることが多い。しかしながら、一般に各状態に対応する音響特徴量の揺らぎの大きさは状態により異なる。そのため、ある状態では分布数が不足しており、他の状態では分布数が多すぎて認識に寄与しない分布がある、という現象が起きていると考えられる。そこで、各状態の混合分布から、認識に対する寄与が少ない分布を削ることで、認識性能を劣化させずに、計算量・メモリ量を大きく削減できる可能性がある。従来、この観点から、状態クラスタリングの結果を利用した分布数削減手法 (e.g. [1])、半連続 HMM, PTM における削減手法 [2]、などが提案されているが、前者は状態クラスタリングの方法に依存し、後者は各状態の分布数が比較的大きい場合 ( $\approx 1000$ ) に有効な手法である。

ここでは、混合ガウス分布 HMM において各状態の分布数を効果的に削減する手法を提案する [3]。この手法では、まず、十分な学習データ量で学習された大きな分布数をもつモデルを用意し、その状態毎のガウス分布の木構造を作成する。そして、各状態毎に、記述長最小 (MDL) 基準を最小にする分布集合を選ぶ。MDL 基準に付加したペナルティ係数を変更することにより、任意の大きさの総分布数をもつモデルを容易に構築できる。本手法は、状態クラスタリング手法に依存しておらず、また、各状態の分布数が 100 以下の、一般的な混合ガウス分布 HMM に対し効果的な手法である。

## 2. ガウス分布木構造

各状態のガウス分布木構造を作成する手法を説明する [4]。まず、分布  $g_m(\cdot)$  と  $g_n(\cdot)$  との距離  $d_{mn}$  は、一方から他方への Kullback-Leibler 情報量の和として定義される。対角共分散行列を用いる場合、

$$d_{mn} = \sum_i \left[ \frac{\sigma_m^2(i) - \sigma_n^2(i) + (\mu_m(i) - \mu_n(i))^2}{\sigma_n^2(i)} + \frac{\sigma_n^2(i) - \sigma_m^2(i) + (\mu_n(i) - \mu_m(i))^2}{\sigma_m^2(i)} \right] \quad (1)$$

である。ここで、 $\mu_m(i)$ 、 $\sigma_m^2(i)$  はそれぞれ、 $g_m(\cdot)$  の平均ベクトル  $\mu_m$ 、分散  $\Sigma_m$  の第  $i$  成分である。次に、木構造のノード  $k$  に対応する分布パラメー

タは、対応するリーフのガウス分布  $\{g_m^{(k)}(\mathbf{X}) \sim \mathcal{N}(\mathbf{X}|\mu_m^{(k)}, \Sigma_m^{(k)}); m = 1, \dots, M_k\}$  のパラメータを用いて、以下のように定義される。

$$\mu_k(i) = \frac{1}{M_k} \sum_{m=1}^{M_k} \mu_m^{(k)}(i) \quad (2)$$

$$\sigma_k^2(i) = \frac{1}{M_k} \sum_{m=1}^{M_k} \left[ \sigma_m^{(k)2}(i) + \mu_m^{(k)2}(i) \right] - \mu_k^2(i) \quad (3)$$

以上を用い、ガウス分布の木構造がトップダウンに作成される。なお、ここでは、各ノードからの分岐数は 2 とする (2 分木)。

## 3. MDL 基準

モデル集合  $\{1, \dots, i, \dots, I\}$  とデータ  $x^N = \{x_1, \dots, x_N\}$  が与えられたとき、モデル  $i$  を用いた記述長  $l_i(x^N)$  は以下のように定義される。

$$l_i(x^N) = -\log P_{\theta^{(i)}}(x^N) + \frac{\beta_i}{2} \log N + \log I \quad (4)$$

ここで、 $\beta_i$  はモデル  $i$  の次元 (自由度) であり、 $\hat{\theta}^{(i)}$  はモデル  $i$  のパラメータ  $\theta^{(i)} = (\theta_1^{(i)}, \dots, \theta_{\beta_i}^{(i)})$  の最尤推定量である。MDL 基準 [5] は、この記述長が最小であるモデルが最適なモデルであると主張する。

今、HMM のある状態においてガウス分布木構造が作成され、ルートノード  $S_0$  が  $M$  個のノード  $S_1, \dots, S_M$  に分割されているとき、これらのノードの分布で構成される混合ガウス分布の記述長は、いくつかの近似の後、以下の式で表される [6, 3]。

$$l = \sum_{m=1}^M \frac{1}{2} \Gamma_m (K + K \log(2\pi) + \log |\Sigma_m|) + \alpha K M \log \sum_{m=1}^M \Gamma_m + C \quad (5)$$

ここで、 $\Gamma_m$  は、特徴ベクトルがノード  $m$  の分布から出力される事後確率の全データに渡る和、 $\Sigma_m$  は、ノード  $m$  の分布の分散、 $C$  は定数である。なお、ここでは、(5) に示すように、第 2 項にペナルティ係数  $\alpha$  を乗じ、これを HMM のすべての状態にわたる総分布数の制御に用いる。

## 4. ガウス分布数の削減

上述したガウス分布木構造と MDL 基準を用い、以下の手続きでガウス分布数の削減を行なう。

1. 混合分布 HMM のパラメータを通常の手続きを用いて学習する。この HMM では、各状態の分布数は状態によらず一定の十分大きな値とする。

\*HMM Size Reduction Using MDL Criterion. By Koichi Shinoda and Ken-ichi Iso (NEC Corporation).

†現在東京大学大学院情報理工学系研究科システム情報学専攻

表 1: Comparison of 16g→4g HMM with 16g and 4g HMM. The results of the speed evaluation experiments were obtained by averaging over 52 sentences spoken by one speaker. The machine used was a DEC 21264 (500 MHz, 1024MByte)

HMM	16g	4g	16g→4g
No.Gauss	59056	14792	15153
x realtime	0.96	0.37	0.43

2.HMM の状態毎に以下の処理を行なう。

- 状態のすべてのガウス分布をリーフ分布としたガウス分布木構造を作成する(2章)。
- ガウス分布木構造の中間ノードにおける充足統計量をその下のリーフノードの充足統計量の和として求め、それを用いて式(5)の  $\Gamma_m, \Sigma_m$  を計算する。
- 式(5)に従い、ガウス分布木構造の各ノードセットの記述長  $l$  を計算する。
- 記述長  $l$  を最小にするノードセットを選択する。
- 選択されたノードセットに対応する分布集合を状態の分布とする。

3.全 HMM パラメータの再推定を行なう。

ステップ(2c),(2d)で、すべての可能なノードセットについて記述長を計算するのは計算量的に困難である。そこで以下の準最適な分割アルゴリズムを用いる。まず、ルートノード  $S_0$  を  $S_1, S_2$  の2つのノードに分割することを考える。式(5)より、分割前後の記述長の差は、

$$\Delta = \frac{1}{2}(\Gamma_1 \log |\Sigma_1| + \Gamma_2 \log |\Sigma_2| - \Gamma_0 \log |\Sigma_0|) + \alpha K \log \sum_{m=1}^M \Gamma_m \quad (6)$$

と計算できる。ここで、 $\Delta > 0$  の場合には分割を停止し、 $\Delta < 0$  の場合には分割を実行する。この手続きを、分割すべきノードがなくなるまで繰り返す。

## 5. 評価実験

提案手法を日本語大語彙認識実験により評価した。学習に用いた音声データは、1656名の男性話者の発声した合計188,000文である。評価用には学習データに含まれない10名の男性話者が各々52~67文を発声したデータを用いた。音響特徴量は、MFCC12次元とそれらの $\Delta$ 量および $\Delta\Delta$ 量、 $\Delta$ パワー、 $\Delta\Delta$ パワーの計38次元を用いた。分布数削減の初期モデルとしては、各状態の分布数を一律に16としたHMM(16g HMM)を用いた。また参照のため、8g HMM、4g HMMも用意した。

まず、表1に、提案手法で分布数を削減したときの、総分布数と計算量を示す。ここで、16g→4g HMMは

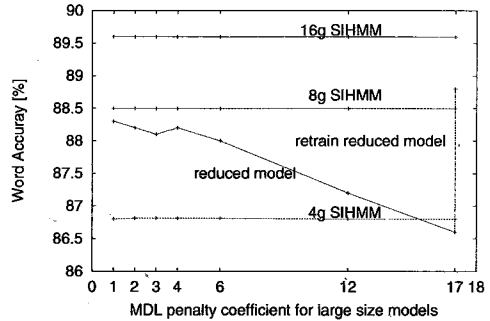


図 1: Comparison of recognition results obtained by averaging over ten test speakers.

提案手法により総分布数を74%削減し、4g HMMとほぼ同じにしたHMMである。このとき、ペナルティ係数  $\alpha = 17$  である。16g→4g HMMの計算量は4g HMMよりわずかに多い。これは、前者における分布数の多い状態に対応する音韻が、分布数の少ない状態に対応する音韻に比べ、評価データ中に多く出現するためと考えられる。次に、図1にペナルティ係数  $\alpha$  を変化させたときの認識結果を示す。16g→4g HMMの再学習後のモデル(図1においてretrainで向上した点)の認識性能は、約2倍のガウス数をもつ8g HMMを上回り、かつ、初期の16g HMMからの劣化も0.8%に押えられていることがわかる。

## 6. おわりに

ガウス分布木構造とMDL基準を用いて、混合ガウス分布HMMのガウス分布数を効果的に削減する方法を提案し、大語彙認識を用いた評価実験で、たかだか0.8%の認識性能の劣化で総ガウス分布数を74%削減することができることを示した。

## 参考文献

- [1] M.-Y. Hwang and X. Huang, "Dynamically configurable acoustic models for speech recognition," *Proc. of ICASSP98*, pp. 669-672, 1988.
- [2] A. Sankar and V. R. R. Gadde, "Parameter tying and Gaussian clustering for faster, better, and smaller speech recognition," *Proc. of EuroSpeech99*, pp. 1711-1714, 1999.
- [3] K. Shinoda and K. Iso, "Efficient reduction of Gaussian components using MDL criterion for speech recognition," *TR of IEICE*, SP2001-83, 2001.
- [4] T. Watanabe, K. Shinoda, K. Takagi, and E. Yamada, "Speech recognition using tree-structured probability density function," *Proc. ICSLP-94*, pp. 223-226, 1994.
- [5] J. Rissanen, "Universal coding, information, prediction, and estimation," *IEEE Trans. IT*, vol. 30, no. 4, pp. 629-636, 1984.
- [6] K. Shinoda and T. Watanabe, "Acoustic modeling based on the MDL criterion for speech recognition," in *Proc. EuroSpeech-97*, vol. 1, pp. 99-102, 1997.