

Buried Markov Model を用いた音声認識モデルの検討*

◎山本隆之, 滝口哲也, 有木康雄 (神戸大)

1 はじめに

現在, 音声認識では, 音響モデルとして隠れマルコフモデル (HMM) が広く用いられているが, HMM を用いた音声認識は, 会話などの話し言葉の認識において, 性能が著しく低下するという問題がある. この問題は現在の状態が 1 フレーム前の過去の状態にのみ依存するという HMM の仮定に起因すると考えられる.

このような問題に対し, HMM の仮定を緩和し, 過去の観測と現在の観測の間の依存関係を表現できる確率モデルである Buried Markov Model (BMM) [1] が研究されている. 従来の BMM の学習方法では, まず HMM を学習し, HMM の学習パラメータを用いて計算した相互情報量を用いて BMM の構造を学習する. しかし, この方法では, BMM の構造が学習した HMM の精度によって, 大きく変化すると考えられる. 本稿では, HMM の初期学習なしでノンパラメトリックに BMM の構造を学習するための手法として, 多重解像度独立性検定 [2] を用いた構造学習法を提案し, 音素認識実験を行い, その有効性を示す.

2 Buried Markov Model

2.1 概要

Buried Markov Model は Fig.1 のようなグラフィカルモデルによって表わされる観測データの特徴次元同士の依存関係を記述した確率モデルである. 依存関係のパターンは各フレームの状態によって決まり, 構造を用いて Gaussian Mixture Model (GMM) のパラメータを学習することで, 別のフレームからの情報が埋め込まれたモデルを構築することができる.

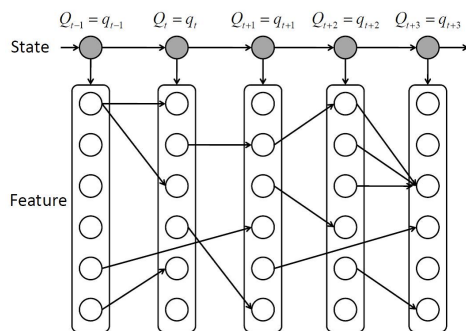


Fig. 1 Buried Markov Model

2.2 構造学習方法

BMM の構造学習には相互情報量を用いた Pairwise アルゴリズムが用いられる. Pairwise アルゴリズムはターゲットノードに対し, 相互情報量が次の条件を満たすノードを親ノードとする方法である.

$$I(X, Z | q) > \delta_1 \quad (1)$$

$$I(Z, Z_i) < \delta_2 I(X, Z) \quad Z_i \in \mathbf{Z} \quad (2)$$

$$I(X, Z) < \delta_3 \quad (3)$$

X はターゲットノード, Z は親の候補, $\mathbf{Z}$ はすでに親であることが決定しているノードの集合, $\delta_1 \dots \delta_3$ はそれぞれの閾値を表す. 式 (1) は相関の高いノードを親とするという条件である. この条件によって, HMM では独立と仮定していた相関の強いノードを特定することが可能となる. 式 (2) は条件付き独立性の検定を行っている. 候補となっているノードがターゲットノードの親ノードとの相関が強いためにターゲットノードとの相関が強くなっているのであれば, 意味のある情報は得られない. この条件は, すでに親であることが決定しているノードとの相関をチェックし, 意味のある情報を持ったノードのみを残すための条件である. これらの条件によって正確に特徴間のモデリングができたとしても, これによって推定誤差を減少させることができるとは限らない. 相関の高いノードを親としても, 他の状態においても相関が高ければ推定誤差は高いままである. そこで式 (3) の条件を導入し, 2つの変数間の相互情報量の状態平均が小さいノードを選択する. ある状態での相互情報量が高く, 他の状態での相互情報量が小さければ, その状態を推定するために有益な情報を与える依存関係であると考えられるためである. この手法で用いる相互情報量は特徴間の依存関係を決定していない状態の BMM(HMM) のパラメータを用いて計算する. この方法では HMM の精度によって得られる依存関係が変わってくる. これを避けるため次章で述べる多重解像度独立性検定を導入する.

3 多重解像度独立性検定

多重解像度独立性検定は, データを分割して作成した 2次元ヒストグラムの情報から, 独立モデルからのデータ発生尤度を計算し, 2つの確率変数が独立であるかどうかをノンパラメトリックに検定する

*Speech Recognition using Buried Markov Model. by Takayuki Yamamoto, Tetsuya Takiguchi and Yasuo Ariki (Kobe Univ.)

方法である．はじめに解像度と境界を固定した場合の独立性検定を考える．まず，解像度を $R \equiv I \times J$ とし，2つの確率変数のサンプルの存在区間を $I \times J$ の領域に分割する．各領域に含まれるサンプルの数をそれぞれ $c_1 \dots c_{I \times J}$ ，サンプルの総数を N ，各領域でサンプルが発生する確率をディリクレ分布で仮定し，各領域の境界の集合を $\mathbf{B}_R$ とするとデータ $\mathbf{D}$ の尤度は次式で表される．

$$Pr(\mathbf{D}) = \frac{\Gamma(\gamma)}{\Gamma(\gamma + N)} \prod_{k=1}^{IJ} \frac{\Gamma(\gamma_k + c_k)}{\Gamma(\gamma_k)} \quad (4)$$

$\gamma = \Sigma \gamma_i$ であり γ_i はディリクレ分布のパラメータを表す．上の式 (4) を $\mathbf{r}(C_{IJ}, \gamma_a)$ とおく．

またデータが依存モデルか独立モデルのどちらか一方から発生すると仮定すると次式が成り立つ．

$$Pr(\mathbf{D}) = Pr(\mathbf{D} | M_I)Pr(M_I) + Pr(\mathbf{D} | M_{-I})Pr(M_{-I}) \quad (5)$$

さらに依存モデルにおいては，1つの多項分布を仮定すれば良いが，独立モデルにおいては，2つの確率変数が独立であるため両軸方向に多項分布を仮定しなければならない．よって C_{IJ} の情報を I 方向の成分と J 方向の成分に分解し，それぞれのデータ尤度を求める．よって依存モデル，独立モデルのデータ尤度は $\mathbf{r}(C_{IJ}, \gamma_a)$ を用いてそれぞれ次のように表せる．

$$Pr(\mathbf{D} | M_{-I}) = \mathbf{r}(C_{IJ}, \gamma_{IJ}) \quad (6)$$

$$Pr(\mathbf{D} | M_I) = \mathbf{r}(C_I, \alpha_I) \mathbf{r}(C_J, \beta_J) \quad (7)$$

この式 (6)(7) を前述の式 (5) に代入し，ベイズの定理で変形することによって次式を得る．

$$Pr(M_I | \mathbf{D}) = \frac{1}{1 + \frac{1-\rho}{\rho} \frac{\mathbf{r}(C_{IJ}, \gamma_{IJ})}{\mathbf{r}(C_I, \alpha_I) \mathbf{r}(C_J, \beta_J)}} \quad (8)$$

ρ は独立と判断するための確率の閾値である．解像度を上げながらそれぞれの解像度において，(8) 式で尤度を計算し，依存度が最大になる解像度における独立性をこのデータの独立性とする．この方法で計算した尤度を，前章で述べた BMM の構築法に利用した相互情報量に代わる基準として導入し，HMM の精度によらず，非線形の関係にも対応した依存関係の学習を可能とする．

4 実験

BMM の学習と評価には日本音響学会の音素バランス文からなる研究用連続音声認識データベース (ASJ-PB)，および，新聞読み上げ音声コーパス (ASJ-JNAS) を用い，男性話者・女性話者・混合話者 (それ

Table 1 認識結果 (%)

	男性	女性	混合
HMM	51.3	46.0	46.0
BMM(従来)	54.6	48.5	51.0
BMM(提案)	55.2	49.1	52.3

ぞれ 100 名 10 発話) の不特定話者モデルを用いた連続音素認識を行った．モデルは GMTK[3] により作成した．テスト音声データのサンプリング周波数は 16 kHz である．音響特徴量は 0 次元を除いた MFCC 12 次元と，その Δ と $\Delta\Delta$ の計 36 次元である．フレームシフトは 10 ms，フレーム幅は 25ms とした．状態には 43 音素各 1 状態を用い，BMM の依存関係は各ノード最大 5 つとし，過去 5 フレームまで探索を行った．結果は Table.1 に示す．この実験において相互情報量を用いて構築した BMM から男性で 0.6 %，女性で 0.6 %，混合で 1.3 % 向上し，多重解像度独立性検定を導入した本手法の有効性が示された．

5 おわりに

BMM の構築法として相互情報量に代わる多重解像度独立性検定を導入した構築法を提案した．GMTK を用いてモデルを実装し，JNAS を用いた連続音素認識実験によってその有効性を確認した．今後の課題として，他の相関基準を用いた構築法との比較，ネットワーク最適化基準の導入などによって更なる精度の向上を目指す．

参考文献

- [1] J.A. Bilmes, Buried Markov models: a graphical-modeling approach to automatic speech recognition, Computer Speech and Language, Volume 17, Issues 2-3, 213-231, 2003.
- [2] D. Margaritis *et al.*, Proceedings of the Twentieth National Conference on Artificial Intelligence (AAAI), 825-830, 2005.
- [3] J.A. Bilmes and G. Zweig, The Graphical Models Toolkit: An open source software system for speech and time-series processing, ICASSP, Orlando, 3916-3919, 2002.