

Buried Markov Model の構造構築 における独立性検定法の検討*

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. HMM は隠れ状態と観測のみからなる単純な構造の確率モデルであり, 明瞭な発音の音声に対しては高い精度で認識を行うことができるが, 複雑な事象を表現するには適しておらず, ノイズが含まれる音声や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.

J. Bilmes によって考案された Buried Markov Model (BMM) [1] は, HMM の仮定を緩和することでその表現能力を高めたモデルであり, HMM の隠れ状態と観測の構造に加え, 現在の観測と他フレームの観測間や現在の観測の内部での依存関係を表現できるため, より複雑な事象に対応することができる. この BMM の依存関係はデータから学習するが, この学習法については十分に検討がなされていない. 本研究では, BMM の構造の学習に必要な独立性の検定法について複数の手法を導入し, 音素認識精度を比較し, より識別性能の高い認識モデルを構築できる独立性検定法について検討を行った.

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り, 依存関係のパターンは各フレームの状態によって決まる. 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.

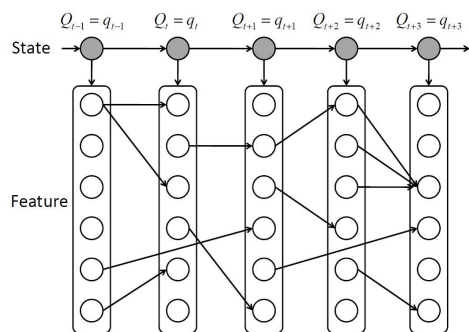


Fig. 1 Buried Markov Model

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 X に対し, 相互情報量が次の条件を満たすノード Z を親ノード集合 $\mathcal{Z}$ に加える方法である. $\delta_{1...3}$ はそれぞれの閾値を表す.

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

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

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

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できる. しかし, ある状態での相互情報量が高くと, 他の状態においても同様に高ければ, その状態を推定するために有益な情報を持ったノードとはいえない. 式 (3) はそういった相互情報量が高いが不要なノードを判定するため, 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条件である. 式 (2) は冗長性の検査を行っている. $\mathcal{Z}$ と $\mathcal{Z}$ のノードとの相関が強ければ, $\mathcal{Z}$ のノードから与えられる情報と同質の情報しか得られないため, こういったノードを除く必要がある.

この手法で用いる相互情報量は依存関係を学習していない状態の 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影響を受ける可能性がある. これを改善するため次章で述べる独立性検定を導入する.

3 独立性検定

3.1 多重解像度独立性検定

多重解像度独立性検定 [2] 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 つの確率変数の独立性をノンパラメトリックに検定する方法である. 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 B_R とするとデータ D の尤度は次式で表される.

*Independence Test for structuring Buried Markov Model. by Takayuki Yamamoto, Tetsuya Takiguchi and Yasuo Ariki (Kobe Univ.)

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

$\gamma = \sum \gamma_i$ であり γ_i はディリクレ分布のパラメータを表す。データが依存モデルか独立モデルのどちらから一方から発生すると仮定すると次式が成り立つ。

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

また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2つの多項分布を仮定しなければならない。よって依存モデル、独立モデルのデータ尤度は式(4)を用いてそれぞれ次のように表せる。

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

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

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。

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

ρ は事前確率を表す。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採用する。この方法で計算した尤度を、Pairwise アルゴリズムに導入し、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。

3.2 順位相関係数

スピアマンの順位相関係数は2つの変数の順位の間の相関の強さを表す指標であり、次式で表わされる。

$$\rho = 1 - \frac{6 \sum_{i=1}^n D_i^2}{(n-1)n(n+1)} \quad (9)$$

長さ n の2つのデータ系列 X, Y のあるインデックス i におけるデータ X_i, Y_i の順位の差を D_i とする。

ケンドールの順位相関係数はスピアマンの順位相関係数と同様の2つの変数の順位の相関の強さを表す指標であり、次式で表わされる。

$$\tau = \frac{P-Q}{\frac{1}{2}n(n-1)} \quad (10)$$

長さ n の対応したデータ系列 X, Y から2つのデータを選ぶ。選んだデータのインデックスを a, b とすると、 P は「 $X_a < X_b$ かつ $Y_a < Y_b$ 」または「 $X_a > X_b$ かつ $Y_a > Y_b$ 」の条件に当てはまる組み合わせの数、 Q は「 $X_a < X_b$ かつ $Y_a > Y_b$ 」または「 $X_a > X_b$ かつ $Y_a < Y_b$ 」の条件に当てはまる組み合わせの数である。すなわち、 P がデータ間の正相関を、 Q が負相関を表しており、 P が大きくなれば正相関が強くなり τ は1に近づき、 Q が大きくなれば負相関が強くなり τ は-1に近づく。 P と Q が同じであれば τ は0となり無相関であると判断できる。

順位相関係数は線形・非線形に関わらず相関関係を判別することができるが、複雑な相関関係を判別することができず、単調関数性を判別する指標となる。

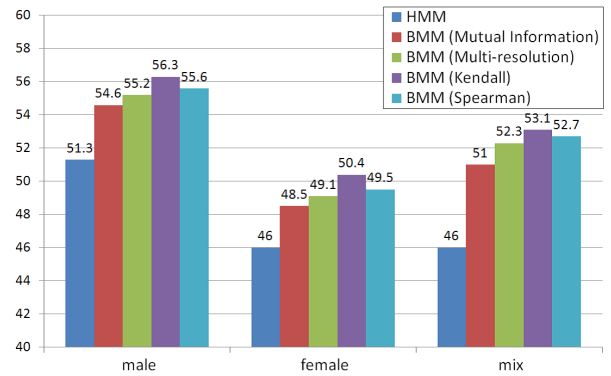


Fig. 2 音素認識精度 [%]

4 実験

HMM と独立性検定に相互情報量、多重解像度独立性検定、順位相関係数 (ケンドール, スピアマン) を用いて学習した BMM の音素認識精度の比較実験を行った。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 (それぞれ 100 名 10 発話) の不特定話者モデルを用いた。特徴量には 0 次元を除いた MFCC 12 次元と、その Δ と $\Delta\Delta$ の計 36 次元を用いた。43 音素各 1 状態とし、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 Fig. 2 に示す。この実験においてケンドール順位相関係数を用いた場合、相互情報量を用いて構築した BMM から男性で 1.7 %, 女性で 1.9 %, 混合で 2.1 % 向上した。

5 おわりに

BMM の構築法として相互情報量に代わるノンパラメトリック独立性検定を導入した構築法を提案した。音素認識実験の結果、相互情報量では分布を限定するため得られる情報が少なく、多重解像度独立性検定では不必要な関係も得てしまうため、単調関数のみを特定する順位相関係数が効果的であることが分かった。今後の課題として、構築アルゴリズムの検討などを行い、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.