

HMM の変分ベイズ学習と環境音認識への応用

坂田尚子 †
広島市立大学林朗 ‡
広島市立大学岩田一貴 ‡
広島市立大学末松伸朗 ‡
広島市立大学

1 はじめに

近年、変分ベイズ学習法が注目を集めている。一般的に、訓練データ数が少ない場合に変分ベイズ学習法 (VB 法) は最尤推定法 (ML 法) に比べ過学習を起こしにくいとされている。

本研究では、連続分布出力隠れマルコフモデル (CHMM) の VB 法を実装し、環境音認識において VB 法の有用性を示すため ML 法と精度を比較する。

なお、環境音とは、周囲に存在する音声以外の音を指す。身の回りでは様々な環境音が存在し、人間は普段の生活の中でその情報をもとに危険予測などの状況把握を行っている。また、普段から聞き慣れている環境音と音声を識別することにより、音声のみに注目することができる。コンピュータが環境音を認識することができれば、音声認識の精度向上、自立移動ロボットや福祉への応用が期待できる。

2 CHMM の変分ベイズ学習法

記号を以下のように定義する。

$S = s_1, s_2, \dots, s_T$: 隠れ状態系列

$L = l_1, l_2, \dots, l_T$: 混合ガウスの出力コンポーネント系列

$\Phi = \{\pi^N, A^{N \times N}, C^{N \times K}, \Theta^{N \times K}\}$: モデルパラメータ

π : 初期状態分布

A : 状態遷移行列

C : コンポーネント出力確率

Θ : 各コンポーネントのガウス分布パラメータ

$X = x_1, x_2, \dots, x_T$: 観測系列

変分ベイズ法ではモデルパラメータを確率変数として扱う。モデルパラメータの事前分布は次の分布とする。

- 初期状態分布 π : ディリクレ分布
- 状態遷移行列 A : ディリクレ分布
- 混合ガウス出力の混合比 C : ディリクレ分布
- ガウス分布の平均と分散 Θ : ガウスウィッシュヤート分布

分布 $q(S, L, \Phi)$ の事後確率への近さを測るため自由エネルギー

ギ $F(q)$ を定義する。

$$F(q) = - \int q(S, L, \Phi) \log \frac{p(X, S, L, \Phi)}{q(S, L, \Phi)} dS dL d\Phi$$

自由エネルギー $F(q)$ を最小化する q を求めるため、確率分布が独立であると仮定する。

$$\begin{aligned} q(S, L, \Phi) &= q(S, L)q(\Phi) \\ &= q(S, L)q(\pi)q(A)q(C)q(\Theta) \end{aligned}$$

$F(q)$ を最小化するために、それぞれの確率分布を残りの 4 つの分布を固定させて順に最適化を行なう。最適化は E ステップ、M ステップの繰り返しとなる。

• E ステップ

$q(\Phi)$ を固定して、 $F(q)$ を最小化するように状態分布 $q(S, L)$ を最適化する。

$$q(S, L) = \frac{1}{Z} \times \pi_{s_1}^* \left[\prod_{t=1}^{T-1} a_{s_t, s_{t+1}}^* \right] \left[\prod_{t=1}^T c_{s_t, l_t}^* f^*(x_t | \theta_{s_t, l_t}) \right]$$

$q(S, L)$ は Forward-backward アルゴリズムで計算できる。つまり、ML 法と同様の計算量で求まる。

• M ステップ

$q(S, L)$ を固定し、 $F(q)$ を最小化するように $q(\Phi)$ を最適化する。 $q(\pi), q(A), q(C), q(\Theta)$ は確率分布が独立であると仮定したので、それぞれを別々に最適化することができる。

3 実験

人工データ、環境音データの分類実験を行い、ML 法と VB 法の分類精度を比較した。

3.1 ハイパーパラメータの設定

初期状態分布: ディリクレ分布

$$p(\pi) = \text{Dir}(\pi_1, \dots, \pi_N | u_1^\pi, \dots, u_N^\pi)$$

のハイパーパラメータ、 $u_i^\pi (1 \leq i \leq N)$ は分布の平均 $E(\pi_i)$ が ML 法で求めた値 $\hat{\pi}_i$ となるように決める。

$$E[\pi_i] = \frac{u_i^\pi}{\bar{u}^\pi} = \hat{\pi}_i \quad (\bar{u}^\pi = \sum_{i=1}^N u_i^\pi)$$

なお、 \bar{u}^π を強さパラメータと呼び、探索して適当な値を求める。

他のパラメータの事前分布についても、分布の平均が ML 法で求めた値となるようにハイパーパラメータを設定する。強さパラメータについては適当に定める。

Variational Bayes HMM Applied to Environmental Sound Recognition

† Sakata Naoko: Faculty of Information Sciences, Hiroshima City University

‡ Akira Hayashi, Kazunori Iwata, Nobuo Sue-matsu: Graduate School of Information Sciences, Hiroshima City University

Email: naoko@pr1.info.hiroshima-cu.ac.jp

3.2 人工データ分類実験

人工データを 2 つの HMM(状態数 2) から生成し、訓練データおよびテストデータとした。訓練データ(長さ 100)の本数を変え VB 法による分類精度と ML 法による分類精度を比較した。テストデータ(長さ 100)は 100 本とした。学習する HMM は過学習の影響を調べるため 10 状態とした。

図 1 に結果を示す。縦軸を精度、横軸を訓練数とし、訓練数ごとに、左に ML 法、右に VB 法の分類精度を示した。どの訓練データ数の場合でも VB 法が ML 法より分類精度が高くなっている。

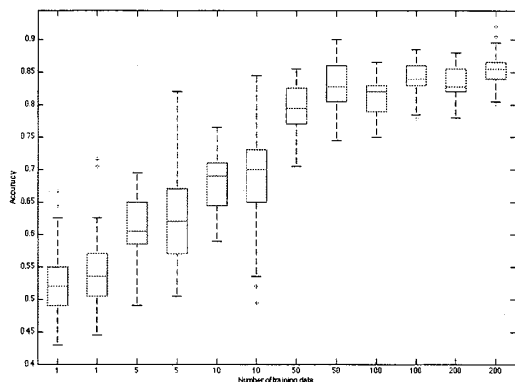


図 1 人工データの分類精度

3.3 環境音分類実験 (環境音 9 種類)

環境音の分類実験を以下のように行った。

- 対象データ：
表 1 の環境音 9 種類 (RWCP 実環境音・音響データベース) [3]

表 1 使用環境音

音	発生要因
bank	貯金箱を手で持ち、金属棒で叩く
cherry1	木板 A を手で持ち、木棒で叩く
magnol	木板 B を手で持ち、木棒で叩く
metall15	1.5mm 厚の金属板を手で持ち、木棒で叩く
sandpp1	紙ヤスリで木片を擦る
sandpp2	木片に巻き付けた紙ヤスリで木片を擦る
saw1	金ノコギリで金属片を掻く
aircap	エアキャップを手で潰す
sticks	割り箸を手で折る

bank から metall15 までの 4 つの音は衝突系、sandpp1 から saw1 までの 3 つの音は摩擦系、aircap と sticks の 2 つの音は破裂破壊系の音であり、材質、駆動過程が偏らないように音を選んでいる。

- 学習する HMM: 3 状態, 混合数 1
- 特徴抽出: メルケプストラム [三木 '99]

- 訓練データ数: 5~50 サンプル
- テストデータ数: 50 サンプル
- 試行回数: 30 回

図 2 に結果を示す。特に訓練データ数が少ないときほど ML 法に比べ VB 法の分類精度が高いことが分かる。

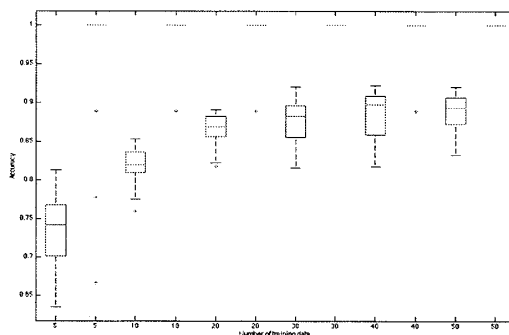


図 2 環境音の分類精度

3.4 環境音分類実験 (環境音 81 種類)

RWCP 実環境音・音響データベースで提供されている 92 種類の環境音のうち、100 サンプル以上提供されている 81 種類の環境音について前の実験と同様の分類実験を行った。ただし、訓練データ数は 5、試行回数は 1 である。

結果の分類精度は ML 法が 63.4%、VB 法が 77.8% だった。

4 まとめ

連続分布出力隠れマルコフモデル (CHMM) の VB 法を実装し、人工データ分類、および環境音分類実験において、VB 法の ML 法に対する優位性を示した。

参考文献

- [1] C.M.Bishop, パターン認識と機械学習 上下, シュプリンガー・ジャパン, 2008
- [2] 三木 一浩, 西浦 敬信, 中村 哲, 鹿野 清宏: “HMM を用いた環境音識別の検討,” 情報処理学会研究報告. SLP, 音声言語情報処理., vol. 99, pp.79-84, 1999.
- [3] 技術研究組合 新情報処理開発機構 実環境音響サブワーキンググループ, RWCP 実環境音声・音響データベース.