5M-11

混合状態 DBN を用いた環境音のモデル化と分類*

俵 喜美子† 林 朗‡ 末松 伸朗§ 岩田 一貴¶
広島市立大学大学院 † 広島市立大学大学院 ‡ 広島市立大学大学院 § 広島市立大学大学院 ¶

1. はじめに

音声認識の分野では，音声以外の音である環境音を雑音として扱ってきた．しかし，人間は環境音から様々な情報を得て危険予測や状況把握を行っている．このことから，我々は環境音を認識する必要があると考える．

本研究では，環境音の特徴は駆動過程と音源材質の双方により決定される [1] と仮定し，駆動過程と音源材質とを分離してモデル化する．モデル化の手法としては，隠れマルコフモデル（HMM）と線形ダイナミカルシステム（LDS）の結合システムである混合状態 DBN（ダイナミックベイジアンネットワーク）[2] を用いる．駆動過程と音源材質とを分離してモデル化することにより，どのような材質でどのような駆動パターンで環境音が発生したかという駆動過程と音源材質による分類が可能となる．

2. 提案手法

2.1 混合状態 DBN

環境音を駆動過程と音源材質とに分離してモデル化を行う．音源材質の特徴はスペクトル包絡に反映し，駆動過程の特徴は波形の概観に反映する [1]．スペクトル包絡は自己回帰過程（Auto-Regressive）の AR 係数によって決定されることから，音源材質は AR 過程を表現する LDS でモデル化する．駆動過程のモデル化には，波形の時間的変化を表現するために時系列データの解析に一般的に用いられる HMM を適用する．これにより，本研究では環境音のモデル化に LDS と HMM の結合システムである混合状態 DBN を用いることを提案する（図 1）．

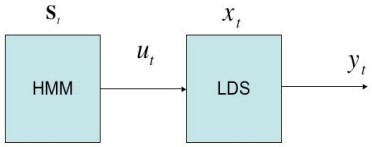


図 1: 混合状態 DBN

混合状態 DBN の LDS サブシステムは以下の式で表される．

$$x_{t+1} = \sum_{n=0}^{N-1} a_n x_{t-n-1} + u_{t+1} + v_{t+1} (N \le t \le T-1) \quad (1)$$

$$y_t = x_t + w_t \quad (2)$$

$$x_t = u_t + v_t (0 \le t \le N-1) \quad (3)$$

ここで，$x_t \in \mathfrak{R}$ は時刻 t における LDS の隠れ状態，$u_t \in \mathfrak{R}$ は時刻 t における LDS システムへの入力，v_t は状態ノイズ過程で $v_t \sim \mathcal{N}(0, Q)$，$y_t \in \mathfrak{R}$ は時刻 t における観測値，w_t は測定ノイズで $w_t \sim \mathcal{N}(0, R)$，$a_0, a_1 \cdots, a_{N-1}$ は N 次の AR 係数である．

HMM サブシステムは以下の式で表される．

$$\mathrm{P}(\boldsymbol{s}_{t+1}|\boldsymbol{s}_t) = \boldsymbol{s}_{t+1}{}' \prod \boldsymbol{s}_t \quad (4)$$

$$u_t = D\boldsymbol{s}_t + r_t \quad (5)$$

$$\mathrm{P}(\boldsymbol{s}_0) = \pi_0 \quad (6)$$

ここで，$\boldsymbol{s}_t$ は時刻 t における HMM の状態変数で $\boldsymbol{s} \in \{\boldsymbol{e}_0, \cdots, \boldsymbol{e}_{S-1}\}$．さらに，$\boldsymbol{e}_i$ は i 番目の要素が 1 となる S 次元の単位ベクトルである．状態遷移行列 $\prod$ は $S \times S$ の行列で (i, j) 要素（$0 \le i \le S-1, 0 \le j \le S-1$）を $\prod(i, j) = \Pr(\boldsymbol{s}_{t+1} = \boldsymbol{e}_i | \boldsymbol{s}_t = \boldsymbol{e}_j)$ とする．各状態の出力の平均を $h_0, h_1, \cdots, h_{S-1}$ とすると，観測行列 D は $1 \times S$ ベクトルで $D = [h_0 h_1 \cdots h_{S-1}]$ で表す．初期状態分布 $\boldsymbol{\pi}_0$ は $S \times 1$ ベクトル，HMM 観測ノイズ過程 r_t は $r_t \sim \mathcal{N}(0, Q)$ とする．

また，LDS の v_t は HMM の r_t により表現されるとし，実際には $v_t = 0$ として扱う．

2.2 隠れ状態変数推定

LDS の隠れ状態変数 $\mathcal{X} = \{x_0, x_1 \cdots, x_{T-1}\}$ と HMM の隠れ状態変数 $\mathcal{S} = \{s_0, s_1 \cdots, s_{T-1}\}$ を観測値 $\mathcal{Y} = \{y_0, y_1, \cdots, y_{T-1}\}$ から推定する．$P(\mathcal{X}, \mathcal{S}|\mathcal{Y})$ を効率的に求めるために，本研究では，近似解法である変分法を適用する．

2.2.1 変分法

図 1 の混合状態 DBN を図 2 に示すように HMM と LDS に分離する．ここで，追加された変分パラメータを $\eta = \{\boldsymbol{q}_0 \cdots \boldsymbol{q}_{T-1}, u_0 \cdots u_{T-1}\}$，与えられた観測系列 $\mathcal{Y}$ に対して，分離システムの事後確率分布を $Q(\mathcal{X}, \mathcal{S}|\eta, \mathcal{Y})$ とする．

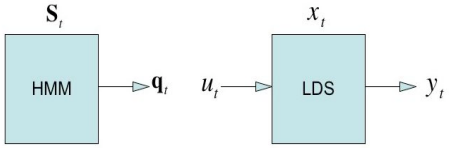


図 2: 分離された混合状態 DBN

変分法では，$Q(\mathcal{X}, \mathcal{S}|\eta, \mathcal{Y})$ ともともとの分布である $P(\mathcal{X}, \mathcal{S}|\mathcal{Y})$ 間のカルバック擬距離が最小になるように η を定義する．

$$\eta^* = \arg\min_{\eta} \Sigma_{\mathcal{S}} \int_{\mathcal{X}} Q(\mathcal{X}, \mathcal{S}|\eta, \mathcal{Y}) \log \frac{Q(\mathcal{X}, \mathcal{S}|\eta, \mathcal{Y})}{P(\mathcal{X}, \mathcal{S}|\mathcal{Y})} d\mathcal{X} \quad (7)$$

*Modeling and the classification of environmental sound using Mixed-State DBN

†Kimiko Tawara · Graduate School, Hiroshima City University
‡Akira Hayashi · Graduate School, Hiroshima City University
§Nobuo Suematu · Graduate School, Hiroshima City University
¶Kazunori Iwata · Graduate School, Hiroshima City University

このようにして定義された変分パラメータ $\boldsymbol{q}_t$ と u_t は以下の式で表される．

$$q_t(i) = e^{(d_i' Q^{-1}(\langle x_t\rangle - \sum_{n=0}^{N-1} a_n \langle x_{t-n-1}\rangle - \frac{1}{2} d_i)} \tag{8}$$

$$u_t = D\langle \boldsymbol{s}_t \rangle \tag{9}$$

ここで，$\boldsymbol{q}_t$ は HMM の出力確率密度とみなし，u_t は LDS の入力とする．d_i は D の i 番目の要素を表す．

$\langle \boldsymbol{s}_t \rangle = E[\boldsymbol{s}_t | \boldsymbol{q}_0, \cdots, \boldsymbol{q}_{T-1}]$ は出力確率密度 $\boldsymbol{q}_t$ をもつ HMM で推論し，$\langle x_t \rangle = E[x_t | u_0, \cdots, u_{T-1}, y_0, \cdots, y_{T-1}]$ は入力 u_t をもつ LDS で推論する．また，式 (8)(9) より，$\boldsymbol{q}_t$ は $\langle x_t \rangle$ に依存し，u_t は $\langle \boldsymbol{s}_t \rangle$ に依存することから，式 (8)(9) に $\langle \boldsymbol{s}_t \rangle$，$\langle x_t \rangle$ の推論を組み合わせたものは不動点方程式を構成し，以下の不動点アルゴリズムで解かれる．

1. $\{\langle x_t \rangle\}$ の初期値を設定する．
2. 式 (8) より $\boldsymbol{q}_t$ を求める．
3. $\boldsymbol{q}_t$ を用いて HMM 推論より $\langle \boldsymbol{s}_t \rangle$ を推定する．
4. 推定した $\langle \boldsymbol{s}_t \rangle$ を用いて式 (9) より u_t を求める．
5. y_t と u_t を用いて、カルマンスムージングより $\langle x_t \rangle$ を推定する．
6. 式 (10) に示すコスト，すなわち負の対数尤度（の上限）が収束するまで 2 から 5 までの推論を繰り返す．

$$\begin{aligned}
\mathrm{Cost} = &\sum_{t=N}^{T-1} \Big(\langle x_t \rangle - \sum_{n=0}^{N-1} a_n \langle x_{t-n-1} \rangle \Big) Q^{-1} \Big(u_t - D\langle \boldsymbol{s}_t \rangle \Big) \\
&+ \frac{1}{2} \sum_{t=N}^{T-1} tr\{D' Q^{-1} D \langle \boldsymbol{s}_t \boldsymbol{s}_t' \rangle\} - \frac{1}{2} \sum_{t=N}^{T-1} Q^{-1} u_t^2 \\
&+ \sum_{t=0}^{N-1} \langle x_t \rangle Q^{-1} (u_t - D\langle \boldsymbol{s}_t \rangle) \\
&+ \frac{1}{2} \sum_{t=0}^{N-1} tr\{D' Q^{-1} D \langle \boldsymbol{s}_t \boldsymbol{s}_t' \rangle\} - \frac{1}{2} \sum_{t=0}^{N-1} Q^{-1} u_t^2 \\
&+ \sum_{t=0}^{T-1} \langle \boldsymbol{s}_t \rangle' \log \boldsymbol{q}_t - \log P_{Q_s} - \log P_{Q_x}
\end{aligned} \tag{10}$$

ここで，$\log P_{Q_s}$ は HMM モデルに対する疑似対数尤度で $\log P_{Q_x}$ は LDS モデルの対数尤度である．

2.3 パラメータ学習

観測値 $\mathcal{Y}$ からモデルパラメータ $\theta = \{\boldsymbol{a}, D, Q, R, \Pi, \pi_0\}$ を EM アルゴリズムを用いて学習する．

E ステップでは隠れ変数の期待値を求める．本研究では、前節で説明した不動点アルゴリズムを E ステップとする．

M ステップでは，式 (11) に示すように補助関数が最大となるようなパラメータ $\theta^* = \{\boldsymbol{a}, D, Q, R, \Pi, \pi_0\}$ を求める．

$$\theta^* = \arg\max_{\theta} \Sigma_{\mathcal{S}} \int_{\mathcal{X}} Q(\mathcal{X}, \mathcal{S} | \mathcal{Y}, \eta^*) \log P(\mathcal{X}, \mathcal{S}, \mathcal{Y} | \theta) d\mathcal{X} \tag{11}$$

ここで $*$ は更新後のパラメータを表す．このときの各パラメータの更新式を以下に示す．

$$\sum_{n=0}^{N-1} a_n^* \Big(\sum_{t=N}^{T-1} \langle x_{t-n-1} x_{t-i-1} \rangle \Big) = \Big(\sum_{t=N}^{T-1} \langle x_t x_{t-i-1} \rangle - D^* \langle s_t \rangle \langle x_{t-i-1} \rangle \Big)$$

$$D^* = \Big(\sum_{t=0}^{T-1} \langle x_t \rangle \langle s_t' \rangle - \sum_{t=N}^{T-1} \sum_{n=0}^{N-1} a_n \langle x_{t-n-1} \rangle \langle s_t' \rangle \Big) \Big(\sum_{t=0}^{T-1} \langle s_t s_t' \rangle \Big)^{-1}$$

$$\begin{aligned}
Q^* = \frac{1}{T} \Big\{ &\sum_{t=N}^{T-1} \Big(\langle x_t^2 \rangle - D^* \langle s_t \rangle \langle x_t \rangle - \sum_{n=0}^{N-1} a_n^* \langle x_t x_{t-n-1} \rangle \Big) \\
&+ \Big(\sum_{t=0}^{N-1} \langle x_t^2 \rangle - D^* \langle s_t \rangle \langle x_t \rangle \Big) \Big\}
\end{aligned}$$

$$R^* = \frac{1}{T} \sum_{t=0}^{T-1} (y_t^2 - 2\langle x_t \rangle y_t + \langle x_t^2 \rangle)$$

$$\Pi^* = \Big(\sum_{t=1}^{T-1} \langle \boldsymbol{s}_t \boldsymbol{s}_{t-1}' \rangle \Big) \Big(\sum_{t=1}^{T-1} \langle \boldsymbol{s}_t \boldsymbol{s}_t' \rangle \Big)^{-1}$$

$$\pi_0^* = \langle \boldsymbol{s}_0 \rangle$$

3. 分類

以下の手順により環境音を分類する．

(0) 駆動源のモデル $\{Hmm^{(i)} | 1 \le i \le M\}$ と音源材質のモデル $\{Lds^{(j)} | 1 \le j \le L\}$ が与えられているとする．

(1) 未知の環境音 $\{y_t | 0 \le t \le T-1\}$ が与えられる．

(2) $M \times L$ 通りの組み合わせのモデルに対する環境音のコストをそれぞれ式 (10) より求める．

(3) 最も小さいコストを与える組み合わせである (i, j) に環境音を分類する．

4. まとめ

環境音を駆動過程と音源材質とに分離してモデル化する手法として，混合状態 DBN を用いることを提案した．

参考文献

[1] 佐宗晃，田中和世，AR-HMM を用いた環境音認識の検討, 日本音響学会講演論文集 pp.167-168,2002

[2] Vladimir Pavlovic , Brendan Frey , and Thomas S. Huang ,"Time Series Classification Using Mixed-State Dynamic Bayesian Networks. " IEEE Conf. Computer Vision and Pattern Recognition,pp.609-615,1999