

適応的遷移確率を導入した隠れマルコフモデルに基づく リカレント確率ニューラルネット

嘉山 敢太[†] 迎田 隆幸[‡] 島 圭介[§]
 横浜国立大学[†] 神奈川県立産業技術総合研究所[‡] 横浜国立大学[§]

1 はじめに

医療・福祉分野において、心電図 (ECG: Electrocardiogram) や筋電位 (EMG: Electromyography) などを利用した疾病の診断支援やマンマシンインターフェース制御など、生体信号の応用が盛んに検討されてきた [1]。これらの実現には生体信号に含まれる特徴や異常をパターン認識技術によって解釈する必要があり、ECG や EMG 信号のような時系列データを扱う場合、データの時間変化を利用した識別器によって高精度な認識が期待できる。

従来、時系列データに対応した様々なモデルが提案されており、ニューラルネットモデルではリカレントニューラルネットや Long-Short Term Memory (LSTM), Gated Recurrent Unit (GRU) など、確率モデルでは隠れマルコフモデル (Hidden Markov Model: HMM) [2] やガウス過程などが存在する。前者は識別精度に優れ、後者は解釈性に優れることが広く知られており、この両モデルを組み合わせた Recurrent Log-Linearized Gaussian Mixture Model (RLLGMN) [3] などの手法も存在する。RLLGMN は HMM と混合正規分布 (Gaussian Mixture Model: GMM) を内包したリカレント確率ニューラルネットであり、両モデルの特徴を有し、事後確率の算出や高速な学習が実現できる。他方、勾配消失問題を解消できないことが課題として挙げられる。

本稿では、HMM の遷移確率をデータの経時的な変化に伴って適応的に変化させることが可能な新しい HMM を提案する。また、それをニューラルネット構造に展開することで確率モデルのパラメータを重み係数として最適化可能にする。これにより、ニューラルネットと確率モデルの両モデルの特徴を有し、RLLGMN の抱える勾配消失問題の解消が可能な識別器の実現が期待される。

2 提案法

2.1 適応的遷移確率を有する隠れマルコフモデル

HMM は時間経過に伴って発生する確率分布の変化を、隠れて存在する状態の確率的な遷移によって表現する確率モデルである。ここで、 C クラス分の時系列データが存在し、クラス c の時系列データを状態数 K_c の HMM でモデル化することを考える。HMM の各状態の出力分布に、平均 $\mu_{c,k}$ 、分散 $\Sigma_{c,k}$ をパラメータとして持つ正規分布 $g(\mathbf{x}(t); \mu_{c,k}, \Sigma_{c,k})$ を導入した場合、任意の時系列データ $\mathbf{S} = \{\mathbf{x}(t) \in \mathbb{R}^D\}_{t=1,2,\dots,T}$ に対する時刻 t の尤度 $l_{c,k}(t)$ は次式で与えられる。

$$l_{c,k}(t) = \sum_{k'=1}^{K_c} l_{c,k'}(t-1) \gamma_{k',k}^c g(\mathbf{x}(t); \mu_{c,k}, \Sigma_{c,k}) \quad (1)$$

$$l_{c,k}(1) = \pi_{c,k} g(\mathbf{x}(1); \mu_{c,k}, \Sigma_{c,k}) \quad (2)$$

T は系列長、 D は観測値 $\mathbf{x}(t)$ の次元数、 $\pi_{c,k}$ は初期遷移確率を表し、状態 k' から状態 k への遷移確率は $\sum_{k=1}^{K_c} \gamma_{k',k}^c = 1$ を満たす。また時系列データ全体に関する事後確率 $p(c|\mathbf{S})$ は次式で定義される。

$$p(c|\mathbf{S}) = \frac{\sum_{k=1}^{K_c} l_{c,k}(T)}{\sum_{c'=1}^C \sum_{k=1}^{K_{c'}} l_{c',k'}(T)} \quad (3)$$

Recurrent probabilistic neural network based on hidden Markov model with adaptive transition probabilities

[†] Kanta Kayama, Yokohama national university

[‡] Takayuki Mukaeda, Kanagawa Institute of Industrial Science and Technology

[§] Keisuke Shima, Yokohama national university

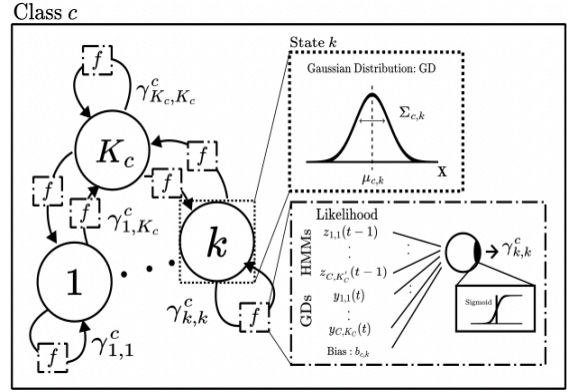


図1 適応的遷移確率を有する HMM

本稿では、従来の HMM の遷移確率 $\gamma_{k',k}^c$ に対し、新たにゲート関数 $G_{c,k}(t)$ を導入することで、観測値の時間変化に伴って遷移確率が適応的に変化する関数として再定義する。ここで、 $y_{c,k}(t) = \log \{g(\mathbf{x}(t); \mu_{c,k}, \Sigma_{c,k})\}$ とし、適応的遷移確率を次式のように定義する。

$$\gamma_{k',k}^c = \exp \left[-G_{c,k}(t) \log \{l_{c,k'}(t-1)\} - (1 - G_{c,k}(t)) y_{c,k}(t) \right] \quad (4)$$

これにより HMM の尤度は次式のように表現される。

$$l_{c,k}(t) = \sum_{k'=1}^{K_c} \exp \left[(1 - G_{c,k}(t)) \log \{l_{c,k'}(t-1)\} + G_{c,k}(t) y_{c,k}(t) \right] \quad (5)$$

また $z_{c,k}(t) = \log \{l_{c,k}(t)\}$ とおくと、 $G_{c,k}(t)$ は次式のように定義する。

$$G_{c,k}(t) = \text{sigmoid} \{ \mathbf{w}_Z^{(c,k)\top} \mathbf{Z}_{c,k}(t-1) + \mathbf{w}_Y^{(c,k)\top} \mathbf{Y}_{c,k}(t) + b_{c,k} \} \quad (6)$$

$\sum_{c=1}^C K_c = N$, $\mathbf{Z}_{c,k}(t) = [z_{1,1}(t), \dots, z_{c,K_c}(t)]^\top \in \mathbb{R}^N$, 及び $\mathbf{Y}_{c,k}(t) = [y_{1,1}(t), \dots, y_{c,K_c}(t)]^\top \in \mathbb{R}^N$ とする。 $\mathbf{w}_Z^{(c,k)} \in \mathbb{R}^N$, $\mathbf{w}_Y^{(c,k)} \in \mathbb{R}^N$ が遷移確率算出に寄与する重み係数、 $b_{c,k}$ がバイアス項である。

2.2 ニューラルネットへの展開

HMM の学習では一般的に EM アルゴリズムや変分ベイズ法、サンプリングによる最適化手法が利用される。しかしながら、提案した適応的遷移確率を有する HMM は複雑なモデル構造を有し、統計的制約を考慮した学習が困難となる。さらに、学習時にパターン識別精度を考慮できないことから上記の学習法の適用は適切ではない。

そこで本稿では、これらの課題を解決するために、提案した HMM の構造を内包するリカレント確率ニューラルネットを提案する。提案法は、提案した HMM に対し対数線形化 [5] を適用することで、複雑な HMM のパラメータを無制約な重み行列として展開する。これにより、誤差逆伝播法を利用した最適化が可能になる。具体的には、正規分布の尤度計算が非線形変換した入力ベクトルと重み行列の積に変換され、 $y_{c,k}(t)$ は次式で表現できる [5]。

$$y_{c,k}(t) = \mathbf{w}_X^{(c,k)\top} \mathbf{X}(t) \quad (7)$$

ただし、入力 $\mathbf{x} = [x_1, \dots, x_D]^\top$ は次式のように非線形変換され、

$$\mathbf{X} = [1, \mathbf{x}^\top, x_1^2, x_1 x_2, \dots, x_1 x_D, x_2^2, x_2 x_3, \dots, x_2 x_D, \dots, x_D^2]^\top \in \mathbb{R}^H \quad (8)$$

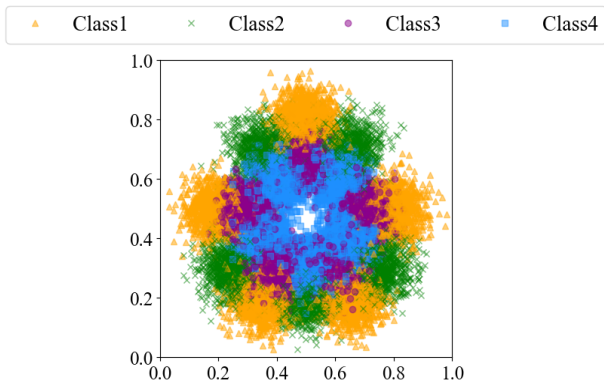


図2 学習データ

非線形変換後の入力ベクトル \mathbf{X} の次元数は $H = 1 + D(D + 3)/2$ となる。これにより、提案した HMM の対数尤度は次式で表現でき、

$$z_{c,k}(t) = \log \left[\sum_{k'=1}^{K_c} \exp \left\{ (1 - G_{c,k}(t)) z_{c,k'}(t-1) + G_{c,k}(t) \{ \mathbf{w}_X^{(c,k)} \top \mathbf{X}(t) \} \right\} \right] \quad (9)$$

$G_{c,k}(t)$ は次式のようになる。

$$G_{c,k}(t) = \text{sigmoid} \{ \mathbf{w}_Z^{(c,k)} \top \mathbf{Z}_{c,k}(t-1) + \mathbf{w}_V^{(c,k)} \top \mathbf{V}_{c,k}(t) + b_{c,k} \} \quad (10)$$

ここで、 $\mathbf{V}_{c,k}(t) = \{ \mathbf{w}_X^{(1,1)} \top \mathbf{X}(t), \dots, \mathbf{w}_X^{(C,K_C)} \top \mathbf{X}(t) \} \top \in \mathbb{R}^N$, $\mathbf{w}_V^{(c,k)} \in \mathbb{R}^N$ である。

さらに、適応的遷移確率を有する HMM をニューラルネット構造に展開した Adaptive Hidden Markov Network (AHMN) を提案する。AHMN の各層の入出力関係について、以下に示す。

まず初めに、入力ベクトルを式 (8) を用いて非線形変換し、これを第 1 層の入力とする。第 1 層の入出力関係は次式で表される。

$${}^{(1)}I_h(t) = X_h(t), \quad {}^{(1)}O_h(t) = {}^{(1)}I_h(t) \quad (11)$$

第 2 層では、隠れ状態が内包する正規分布の対数尤度の計算を行う。第 2 層の入出力関係は次式で表される。

$${}^{(2)}I_{c,k}(t) = \sum_{h=1}^H w_{c,k,h} {}^{(1)}O_h(t), \quad {}^{(2)}O_{c,k}(t) = {}^{(2)}I_{c,k}(t) \quad (12)$$

第 3 層では、提案した HMM の対数尤度の計算を行う。第 3 層の入出力関係は次式で表される。

$${}^{(3)}I_{c,k}(t) = \sum_{k'=1}^{K_c} \exp \left\{ (1 - G_{c,k}(t)) {}^{(3)}O_{c,k'}(t-1) + G_{c,k}(t) {}^{(2)}O_{c,k}(t) \right\}, \quad (13)$$

$${}^{(3)}O_{c,k}(t) = \log \left\{ {}^{(3)}I_{c,k}(t) \right\}$$

最後に、第 4 層の入出力関係は

$${}^{(4)}I_{c,k}(T) = \exp \left\{ {}^{(3)}O_{c,k}(T) \right\}, \quad {}^{(4)}O_c(T) = \frac{\sum_{k=1}^{K_c} {}^{(4)}I_{c,k}(T)}{\sum_{c'=1}^C \sum_{k'=1}^{K_{c'}} {}^{(4)}I_{c',k'}(T)} \quad (14)$$

となり、 ${}^{(4)}O_c(T)$ は時系列データ \mathbf{S} が与えられた際のクラス c に対する事後確率 $p(c|\mathbf{S})$ となる。

3 実験

3.1 方法

実験では、5 つの隠れ状態を持つ HMM から生成される時系列データを使用した。各状態の出力分布は 2 次元正規分布に従い、ある時刻 t から時刻 $t + 1$ となる状況において、現在の隠れ状態に留まる確率を 0.8、異なる別の隠れ状態に遷移する確率をそれぞれ 0.05 とした。この条件に従う HMM から 4 クラス分の時系列データを生成し、このとき、各クラスの時系列データの系列長を 100 とした。ここで系列長 100 の 2 次元データを 1 サンプルとした時、各クラス 50 サンプルの学習データを作成した。データの一例として学習時に使用した時系列データを

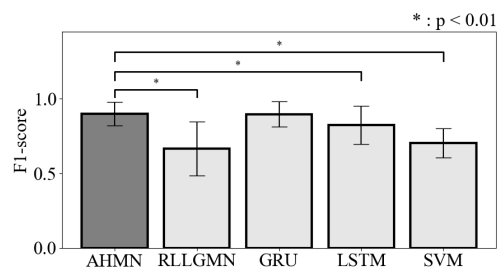


図3 精度評価結果

2 次元平面上にプロットしたものを図 1 に示す。

提案法の比較手法として、RLLGMN, LSTM, GRU, RBF カーネルを用いたサポートベクターマシン (SVM) を使用した。また、学習データと同条件で検証データ、及びテストデータを作成し、検証時の損失を最小にするモデルをテストデータによる精度評価に使用した。モデルの精度評価指標として F1-Score を使用し、提案法の有効性を検証するために、各モデル 30 回ずつ試行した際の F1-score に対し Welch の t 検定を行なった。

3.2 結果

F1-score による精度評価結果を図 2 に示す。実験で使用したデータは HMM から生成した時間変化を有する時系列データであり、SVM のような静的な識別器は識別精度が低下した。他方、時間情報を考慮可能な識別器のうち AHMN, GRU, LSTM は高い識別精度を得ることができた。この結果から、時系列データにおける識別問題において、そのデータの時間情報を考慮することの有効性が確認された。また、AHMN は RLLGMN に対して有意な精度改善が認められた。これは HMM の遷移確率をデータの時間変化に伴って適応的に変化させるようにしたことで、状態遷移をより柔軟に表現可能になったことを示している。また、(9) 式の対数尤度計算は GRU のネットワーク構造と類似していることから、これが勾配消失問題の抑制に寄与し、長期的な時間情報を考慮可能なモデルとなった可能性が示唆される。

4 おわりに

本稿では、遷移確率をデータの時間変化に適応的に変化させられる新たな HMM を提案した。また、提案した新たな HMM をニューラルネット状に展開することで、確率モデルのパラメータをネットワークの重み係数として最適化することを可能にした。実験結果から、提案法によって比較手法と同等かそれ以上の識別精度が得られることが示された一方、生体信号やベンチマークデータに対する提案法の有効性については検証できていないため、今後更なる追加検証に取り組む必要がある。

参考文献

- [1] B. Rim, *et al.*, “Deep Learning in Physiological Signal Data: A Survey”. *Sensors* 2020, 20, 969.
- [2] L. E. Baum, and T. Petrie, “Statistical Inference for Probabilistic Functions of Finite State Markov Chains”. *The Annals of Mathematical Statistics*, vol.37, no.6, 1966, pp.1554-63. JSTOR.
- [3] T. Tsuji, *et al.*, “A recurrent log-linearized Gaussian mixture network”, *IEEE Transactions on Neural Networks*, vol.14, no.2, pp.304-316, March 2003.
- [4] S. Z. Yu, and H. Kobayashi, “Practical implementation of an efficient forward-backward algorithm for an explicit-duration hidden markov model”. *IEEE Transactions on Signal Processing*, vol.54, no.5, pp 1947-1951, May 2006.
- [5] T. Tsuji, *et al.*, “A log-linearized Gaussian mixture network and its application to EEG pattern classification”, *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol.29, no.1, pp.60-72, Feb. 1999.