

原著論文

SOHMMを用いた隠れマルコフモデルの 状態可視化†

山口 暢彦*

近年、混合分布モデルに対して制約付きEMアルゴリズムを適用することにより、観測データの本質的な構造を探るSOMM（Self-Organizing Mixture Model）が提案され、データの可視化やクラスタリング等への応用が期待されている。SOMMは、同様にデータの可視化等を目的とする自己組織化マップと比べ確率論的にモデル化されており、観測データの分布に適したアルゴリズムを構築することが可能である。本論文では、時系列データを扱えるように、隠れマルコフモデルの考え方を取り入れてSOMMを拡張したSOHMM（Self-Organizing Hidden Markov Model）の提案を行い、SOHMMを用いた時系列データの可視化手法について提案を行う。

キーワード：SOMM、可視化、時系列データ、隠れマルコフモデル

1. はじめに

多変量時系列解析[1, 2]を行うにあたり、その振る舞いが時間的に大きく変化する時系列データを扱う場合がある。例えば、平均気温や降水量などの気象データは季節に応じて循環変動する時系列データとなり、株価や為替相場などは金融危機や政府による政策変更によりその振る舞いが大きく変化する。本論文では、このような時間的に状態が変化する時系列データに対し、その状態推移を可視化する手法について提案を行う。本手法を用いることにより、時間経過にともなう状態の推移が低次元の可視化空間上の軌跡として描かれることとなり、時系列データが生成される環境を視覚的に捉えることが可能となる。

データの可視化を行う一つの方法として、近年、混合分布モデル[3]に対して制約付きEMアルゴリズム[4, 5]を適用することによりデータの可視化を行うSOMM（Self-Organizing Mixture Model）[6, 7]が提案されている。SOMMは、同様にデータの可視化を目的とする自己組織化マップ[8]と比べ確率論的にモデル化されており、観測データの分布に適した可視化アルゴリズムを構築することが可能である。例えば、文献[6, 7]では、データを混合正規分布にあてはめる標準的なSOMMの他に離散データや欠損データの可視化に適したSOMM等が提案されている。また、

SOMMのパラメータ推定は対数尤度関数に自己組織化を行う罰則項を付加した罰則付き最尤推定の一つと解釈することが可能であり、推定アルゴリズムの解釈が容易であるという利点を持つ。本論文では、時系列データを扱えるように、隠れマルコフモデル[9]の考え方を取り入れてSOMMを拡張したSOHMM（Self-Organizing Hidden Markov Model）の提案を行い、SOHMMを用いた時系列データの可視化手法について提案を行う。また、本論文では、SOMMと同様にSOHMMのパラメータ推定法が対数尤度関数に自己組織化を行う罰則項を付加した罰則付き最尤推定の一つと解釈できることを示す。

時系列データの可視化における関連研究として、時系列データを制約付き隠れマルコフモデルにあてはめるGTM-TT（GTM Through Time）[10, 11]が提案されている。GTM-TTは、隠れマルコフモデルのパラメータを基底関数の線形結合を用いて表現し、EMアルゴリズムを用いて最尤推定することにより時系列データの可視化を行う。そのため、GTM-TTでは、基底関数の個数・形状・位置、過学習を抑制する正則化パラメータの適切な設定を行う必要がある。文献[12]では、交差検証法やベイズ法を用いたこれらパラメータの設定手法が提案されているものの、両手法ともに最尤推定を繰り返す必要があり、多くの計算コストが必要となる。一方、提案手法では、近傍関数の広がりを表すパラメータを適切に設定する必要があるものの、GTM-TTと比べより設定する必要があるパラメータの個数が少なく、計算コストを低く抑えることが期待できる。

† Visualizing States in Hidden Markov Models using SOHMM
Nobuhiko YAMAGUCHI

* 佐賀大学大学院工学系研究科
Graduate School of Science and Engineering, Saga University

本論文の構成は以下のとおりである。2では、提案手法の基となるパラメータ推定手法としてEMアルゴリズムについて概説する。3では、文献[6]にて提案されるSOMMについて述べる。4では、提案手法が仮定する時系列データの生成モデルとして隠れマルコフモデルについて述べる。5では、時系列データを隠れマルコフモデルにあてはめるSOHMMの提案を行い、SOHMMを用いた時系列データの可視化手法について提案を行う。6は計算機実験、7はまとめである。

2. EMアルゴリズム

本章では、提案手法の基となるパラメータ推定手法としてEMアルゴリズムの概説を行う。ここでは、観測データを $X=(x'_1, \dots, x'_T)'$ 、潜在変数を $Z=(z_1, \dots, z_T)'$ と書き、観測データ X と潜在変数 Z の対を完全データと呼ぶ。また、確率モデルのパラメータを θ と書くこととする。EMアルゴリズムの目的は、式(1)にて表わされる対数尤度関数 $\mathcal{L}(\theta)$ を極大化するパラメータ θ の推定である。

$$\mathcal{L}(\theta) = \ln p(X; \theta) = \ln \sum_Z p(X, Z; \theta) \quad (1)$$

EMアルゴリズムでは、潜在変数 Z について分布 $q(Z)$ を導入し、対数尤度関数 $\mathcal{L}(\theta)$ の代わりにその下界である $\mathcal{F}(q, \theta)$ を極大化することによりパラメータ推定を行う。下界 $\mathcal{F}(q, \theta)$ は、KLダイバージェンスが非負であることを考慮し、式(2)と定義する。EMアルゴリズムでは、下界 $\mathcal{F}(q, \theta)$ を極大化することにより対数尤度関数 $\mathcal{L}(\theta)$ を極大化するパラメータ θ の推定を行う。

$$\mathcal{L}(\theta) \geq \mathcal{F}(q, \theta) \quad (2)$$

$$= \mathcal{L}(\theta) - \text{KL}(q \| p) \quad (2)$$

$$= \mathbb{E}_q [\ln p(X, Z; \theta)] + \mathcal{H}(q) \quad (3)$$

ここで、 $\text{KL}(q \| p)$ は分布 $q(Z)$ と分布 $p(Z|X; \theta)$ とのKLダイバージェンスであり、 $\mathcal{H}(q)$ は分布 $q(Z)$ のエントロピーである。また、式(3)への変形は $p(X; \theta) = p(X, Z; \theta) / p(Z|X; \theta)$ による。

$$\text{KL}(q \| p) = \mathbb{E}_q \left[\ln \frac{q(Z)}{p(Z|X; \theta)} \right]$$

$$\mathcal{H}(q) = -\mathbb{E}_q [\ln q(Z)]$$

EMアルゴリズムでは、パラメータ θ を適当な初期値に設定し、E-step、M-stepと呼ばれる2個のステップを繰り返すことにより下界 $\mathcal{F}(q, \theta)$ を極大化するパラメータ θ の推定を行う。E-stepでは、直前のM-stepにより求められたパラメータを $\theta^{\tau-1}$ として、パラメータ $\theta^{\tau-1}$ を固定しながら下界 $\mathcal{F}(q, \theta^{\tau-1})$ を最大化

することにより分布 q を決定する。この様な分布 q は、式(2)と対数尤度関数 $\mathcal{L}(\theta)$ が分布 q に依存しないことから $\text{KL}(q \| p)$ を最小化する分布 q と求めればよい。

E-step :

$$\begin{aligned} q^\tau &= \arg \max_q \mathcal{F}(q, \theta^{\tau-1}) \\ &= \arg \min_q \text{KL}(q \| p(Z|X; \theta^{\tau-1})) \end{aligned} \quad (4)$$

M-stepでは、直前のE-stepで求めた分布を q^τ として、分布 q^τ を固定しながら下界 $\mathcal{F}(q^\tau, \theta)$ を最大化することによりパラメータ θ の推定を行う。この様なパラメータ θ は、式(3)とエントロピー $\mathcal{H}(q)$ がパラメータ θ に依存しないことから、完全データの対数尤度の期待値を最大化するパラメータ θ と求めればよい。

M-step :

$$\begin{aligned} \theta^\tau &= \arg \max_\theta \mathcal{F}(q^\tau, \theta) \\ &= \arg \max_\theta \mathbb{E}_{q^\tau} [\ln p(X, Z; \theta)] \end{aligned} \quad (5)$$

更に、観測データ x_t が互いに独立である場合のEMアルゴリズムについて述べる。ここでは、潜在変数 z_t ごとに分布 $q_t(z_t)$ を導入し、分布 q_t を $t=1$ から T まで集めた集合を $q=\{q_1, \dots, q_T\}$ と書くこととする。独立である場合のEMアルゴリズムでは、対数尤度関数の下界 $\mathcal{F}(q, \theta)$ を式(6)、式(7)と定義し、下界 $\mathcal{F}(q, \theta)$ を極大化することにより対数尤度関数 $\mathcal{L}(\theta)$ を極大化するパラメータ θ の推定を行う。

$$\mathcal{L}(\theta) \geq \mathcal{F}(q, \theta) = \sum_{t=1}^T \mathcal{F}_t(q_t, \theta) \quad (6)$$

$$\mathcal{F}_t(q_t, \theta) = \ln p(x_t; \theta) - \text{KL}(q_t \| p(z_t|x_t; \theta)) \quad (7)$$

$$= \mathbb{E}_{q_t} [\ln p(x_t, z_t; \theta)] + \mathcal{H}(q_t) \quad (8)$$

ここで、式(8)への変形は式(9)による。

$$p(x_t; \theta) = p(x_t, z_t; \theta) / p(z_t|x_t; \theta) \quad (9)$$

下界 $\mathcal{F}(q, \theta)$ を極大化するパラメータ θ の推定は、式(10)、式(11)にて表わされるE-step、M-stepを交互に繰り返すことにより行う。

E-step :

$$\begin{aligned} q_t^\tau &= \arg \max_{q_t} \mathcal{F}(q, \theta^{\tau-1}) \\ &= \arg \min_{q_t} \text{KL}(q_t \| p(z_t|x_t; \theta^{\tau-1})) \\ &= p(z_t|x_t; \theta^{\tau-1}) \end{aligned} \quad (10)$$

M-step :

$$\begin{aligned}\theta^\tau &= \arg \max_{\theta} \mathcal{F}(q^\tau, \theta) \\ &= \arg \max_{\theta} \sum_{t=1}^T E_{q_t^\tau} [\ln p(\mathbf{x}_t, z_t; \theta)]\end{aligned}\quad (11)$$

ここで、 q^τ は分布 q_t^τ を $t=1$ から T まで集めた集合であり、 $q^\tau = \{q_1^\tau, \dots, q_T^\tau\}$ である。

3. SOMM**3.1 制約付きEMアルゴリズムを用いた混合分布モデルのパラメータ推定**

本章では、文献[6]にて提案されるSOMMについて述べる。SOMMは、観測データをより低次元の可視化空間上に次元圧縮する手法であり、可視化空間の次元数 L を主に $L=2$ と設定することによりデータの可視化を行う。ここでは、観測データの生成モデルとして混合分布モデルを仮定するSOMMについて、その可視化アルゴリズムの概説を行う。

SOMMでは、観測データ $\mathbf{x}_t = (x_{t,1}, \dots, x_{t,D})'$ は、式(12)にて表わされる混合分布モデルにより独立同一に生成されると仮定する。

$$p(\mathbf{x}_t; \theta) = \sum_{i=1}^M \pi_i p(\mathbf{x}_t | z_t = i; \theta_i) \quad (12)$$

ここで、 π_i は $0 \leq \pi_i \leq 1$ 、 $\sum_{i=1}^M \pi_i = 1$ を満たす混合係数であり、 $p(\mathbf{x}_t | z_t = i; \theta_i)$ は混合要素と呼ばれる確率分布である。また、 z_t は観測データ \mathbf{x}_t が何番目の混合要素により生成されたかを示す潜在変数であり、 θ は混合係数 π_i と混合要素のパラメータ θ_i を全て並べたベクトルである。SOMMでは、観測データ \mathbf{x}_t を可視化するために、可視化空間上に規則的に並べられた M 個の格子点 $\mathbf{g}_1, \dots, \mathbf{g}_M$ を用意し、格子点 \mathbf{g}_i を i 番目の混合要素に対応付ける。そして、観測データ \mathbf{x}_t が i 番目の混合要素により生成される確率 $p(z_t = i | \mathbf{x}_t; \theta)$ を求め、この確率を観測データ \mathbf{x}_t を格子点 \mathbf{g}_i に可視化する確率として用い、データの可視化を行う。

SOMMでは、混合分布モデルに対して、分布 $g_i(z_t)$ を分布 $\bar{q}_t(z_t; w_t)$ と制約する制約付きEMアルゴリズムを適用することによりパラメータ推定を行う。SOMMの推定アルゴリズムとEMアルゴリズムとの唯一の違いは、EMアルゴリズムでは分布 $g_i(z_t)$ を任意の分布としているのに対し、SOMMでは分布 $g_i(z_t)$ を分布 $\bar{q}_t(z_t; w_t)$ と制約している点だけである。この制約により、ある観測データ \mathbf{x}_t を入力した際にそれを可視化する確率 $p(z_t = i | \mathbf{x}_t; \theta)$ が高くなる格子点同士が可視化空間上で互いに近接する様にパラメータ推定が行われる。制約された分布 $\bar{q}_t(z_t; w_t)$ は、自己組織化マッ

プにおける近傍関数を正規化することにより、式(13)と定義する。

$$\bar{q}_t(z_t = i; w_t) = \frac{\exp(-\lambda \|\mathbf{g}_i - \mathbf{g}_{w_t}\|^2)}{\sum_{j=1}^M \exp(-\lambda \|\mathbf{g}_j - \mathbf{g}_{w_t}\|^2)} \quad (13)$$

ここで、 w_t は整数値 $1, 2, \dots, M$ のみを取りうるパラメータであり、 λ は近傍関数の広がりやを決定するパラメータである。この様にして分布 $\bar{q}_t(z_t; w_t)$ を式(13)と定義することにより、分布 $\bar{q}_t(z_t; w_t)$ は潜在変数 z_t が $z_t = w_t$ の時に最大となる分布となり、格子点 \mathbf{g}_{z_t} と格子点 \mathbf{g}_{w_t} との距離に従い滑らかに減少する分布となる。

SOMMでは、分布 $\bar{q}_t(z_t; w_t)$ を $t=1$ から T まで集めた集合を $\bar{q} = \{\bar{q}_1(z_t; w_t), \dots, \bar{q}_T(z_T; w_T)\}$ と書き、集合 q を集合 \bar{q} と制約しながら下界 $\mathcal{F}(q, \theta)$ を極大化することによりパラメータ θ の推定を行う。具体的には、式(14)、式(15)にて表わされるE-step、M-stepを交互に繰り返すことにより、下界 $\mathcal{F}(\bar{q}, \theta)$ を極大化するパラメータ θ の推定を行う。

E-step :

$$\begin{aligned}w_t^\tau &= \arg \max_{w_t \in \{1, \dots, M\}} \mathcal{F}(\bar{q}, \theta^{\tau-1}) \\ &= \arg \min_{w_t \in \{1, \dots, M\}} \text{KL}(\bar{q}_t(z_t; w_t) \| p(z_t | \mathbf{x}_t; \theta^{\tau-1}))\end{aligned}\quad (14)$$

M-step :

$$\begin{aligned}\theta^\tau &= \arg \max_{\theta} \mathcal{F}(\bar{q}^\tau, \theta) \\ &= \arg \max_{\theta} \sum_{t=1}^T E_{\bar{q}_t(z_t; w_t^\tau)} [\ln p(\mathbf{x}_t, z_t; \theta)]\end{aligned}\quad (15)$$

ここで、 \bar{q}^τ は分布 $\bar{q}_t(z_t; w_t^\tau)$ を $t=1$ から T まで集めた集合であり、 $\bar{q}^\tau = \{\bar{q}_1(z_t; w_t^\tau), \dots, \bar{q}_T(z_T; w_T^\tau)\}$ である。

この様にして対数尤度関数の下界 $\mathcal{F}(\bar{q}, \theta)$ を極大化することにより、SOMMでは、混合分布モデルのパラメータ θ が観測データの分布に沿う様に推定される。また、式(6)、式(7)を考慮すると、SOMMのパラメータ推定法は、対数尤度関数に罰則項を付加した罰則付き最尤推定の一つと解釈することができる。罰則項は分布 $\bar{q}_t(z_t; w_t)$ と分布 $p(z_t | \mathbf{x}_t; \theta)$ とのKLダイバージェンスの総和となり、SOMMでは分布 $\bar{q}_t(z_t; w_t)$ と分布 $p(z_t | \mathbf{x}_t; \theta)$ とがKLダイバージェンスの意味で近くなる様にパラメータ推定が行われる。従って、分布 $\bar{q}_t(z_t; w_t)$ が格子点 \mathbf{g}_{z_t} と格子点 \mathbf{g}_{w_t} との距離に従い滑らかに減少する分布であることを考慮すると、対応する格子点 \mathbf{g}_i が互いに近接する場合、観測

データ x_i を格子点 g_i に可視化する確率 $p(z_t = i | x_i; \theta)$ も互いに類似する様にパラメータ推定が行われる。また、SOMMでは、ある観測データ x_i が入力された際にそれを可視化する確率 $p(z_t = i | x_i; \theta)$ が高くなる格子点同士が可視化空間上で互いに近接する様にパラメータ推定が行われる。

この様にしてパラメータ推定を行うことにより、自己組織化マップと類似した混合分布モデルのパラメータ推定法を構築することが可能となる。例えば、文献[6, 7]では、混合係数を全て $1/M$ 、混合要素を平均が μ_i 、分散共分散行列が $\beta^{-1}I$ の等方性正規分布としたシンプルな混合正規分布に対してSOMMを適用した場合、正規分布の平均パラメータのM-stepはバッチ型SOMのパラメータ更新則と等しくなる等、SOMMとバッチ型SOMとの類似性について考察が行われている。

3.2 データの可視化

SOMMでは、観測データ x_i が i 番目の混合要素により生成される確率 $p(z_t = i | x_i; \theta)$ を求め、この確率を観測データ x_i を格子点 g_i に可視化する確率として用い、データの可視化を行う。確率 $p(z_t = i | x_i; \theta)$ は、制約付きEMアルゴリズムにより求められる確率 $p(x_i | z_t = i; \theta_i)$ をベイズの定理を用いて反転することにより式(16)と求める。

$$p(z_t = i | x_i; \theta) = \frac{\pi_i p(x_i | z_t = i; \theta_i)}{\sum_{j=1}^M \pi_j p(x_i | z_t = j; \theta_j)} \quad (16)$$

SOMMでは、確率 $p(z_t = i | x_i; \theta)$ を用いて格子点 g_i の平均を求め、これにより観測データ x_i の可視化を行う。

$$\bar{g}_t = \sum_{i=1}^M p(z_t = i | x_i; \theta) g_i$$

4. 隠れマルコフモデル

本章では、提案手法が仮定する時系列データの生成モデルとして、隠れマルコフモデルについて述べる。隠れマルコフモデルでは、時点 t における時系列データ $x_t = (x_{t,1}, \dots, x_{t,D})'$ は、状態 i 毎に定義される出力確率 $p(x_i | z_t = i)$ に従って生成されたとする。ここで、 z_t は時点 t の状態を表す潜在変数であり、整数値 $1, 2, \dots, M$ のみをとる潜在変数である。また、状態 z_t は一次マルコフ過程に従って生成されたと仮定する。つまり、状態 z_t が状態 j となる確率は直前の状態 z_{t-1} にのみ依存すると仮定する。

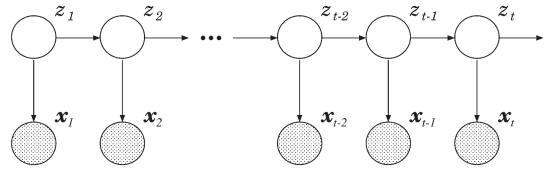


図1 隠れマルコフモデルのグラフィカルモデル

$$\begin{aligned} p(z_t = j | z_{t-1} = i, z_{t-2} = k, \dots) \\ = p(z_t = j | z_{t-1} = i) \end{aligned}$$

図1に隠れマルコフモデルのグラフィカルモデルを示す。本論文では、状態 i から状態 j に推移する確率を推移確率 p_{ij} と書き、最初の状態 z_1 が状態 i となる確率を初期状態確率 ρ_i と書くこととする。

$$p_{ij} = p(z_t = j | z_{t-1} = i)$$

$$\rho_i = p(z_1 = i)$$

もちろん、確率 p_{ij} 、確率 ρ_i は、式(17)、式(18)を満たす。

$$\sum_{j=1}^M p_{ij} = 1, \quad i = 1, \dots, M \quad (17)$$

$$\sum_{i=1}^M \rho_i = 1 \quad (18)$$

5. SOHMM

5.1 制約付きEMアルゴリズムを用いた隠れマルコフモデルのパラメータ推定

本章では、時系列データの状態可視化手法としてSOHMMの提案を行う。SOHMMでは、時系列データは、隠れマルコフモデルにより生成されると仮定する。すなわち、時系列データ x_t は状態 i 毎に定義される出力確率 $p(x_i | z_t = i)$ に従って生成されたとし、状態 z_t は1次マルコフ過程に従って生成されたとする。各状態毎の出力確率 $p(x_i | z_t = i)$ は、平均が μ_i 、分散共分散行列が $\beta^{-1}I$ の正規分布と仮定する。

$$p(x_i | z_t = i) = \left(\frac{\beta}{2\pi} \right)^{D/2} \exp \left(-\frac{\beta}{2} \|x_i - \mu_i\|^2 \right)$$

本章では、隠れマルコフモデルを決定するパラメータ $\mu_i, \beta, p_{ij}, \rho_i$ を全て並べたベクトルを θ と書く。SOHMMでは、状態 z_t を可視化するために、可視化空間上に規則的に並べられた M 個の格子点 g_1, \dots, g_M を用意し、格子点 g_i を状態 i に対応付ける。そして、状態 z_t が状態 i となる確率 $p(z_t = i | X; \theta)$ をforward-

backwardアルゴリズム[9]により求め、この確率を状態 z_t を格子点 g_i に可視化する確率として用い、時系列データの可視化を行う。

SOHMMでは、隠れマルコフモデルに対して、分布 $q(\mathbf{Z})$ を分布 $\bar{q}(\mathbf{Z}; \mathbf{W})$ と制約する制約付きEMアルゴリズムを適用することによりパラメータ推定を行う。SOHMMの推定アルゴリズムとEMアルゴリズムとの唯一の違いは、EMアルゴリズムでは分布 $q(\mathbf{Z})$ を任意の分布としているのに対し、SOHMMでは分布 $q(\mathbf{Z})$ を分布 $\bar{q}(\mathbf{Z}; \mathbf{W})$ と制約している点だけである。この制約により、SOHMMでは、一時刻前の状態が与えられた際にその状態から推移する確率が高くなる格子点同士が可視化空間上で互いに近接する様にパラメータ推定が行われる。制約された分布 $\bar{q}(\mathbf{Z}; \mathbf{W})$ は、自己組織化マップにおける近傍関数を正規化した分布 $\bar{q}_1(z_1; w_1)$ と分布 $\bar{q}_{t,t-1}(z_t | z_{t-1}; w_t)$ とを用いることにより、式(19)と定義する。

$$\bar{q}(\mathbf{Z}; \mathbf{W}) = \bar{q}_1(z_1; w_1) \prod_{t=2}^T \bar{q}_{t,t-1}(z_t | z_{t-1}; w_t) \quad (19)$$

$$\bar{q}_1(z_1 = i_1; w_1) = \frac{\exp(-\lambda \|g_{i_1} - g_{w_1}\|^2)}{\sum_{j_1=1}^M \exp(-\lambda \|g_{j_1} - g_{w_1}\|^2)} \quad (20)$$

$$\begin{aligned} \bar{q}_{t,t-1}(z_t | z_{t-1}; w_t) \\ = \prod_{i_{t-1}=1}^M \bar{q}_{t,t-1}(z_t | z_{t-1} = i_{t-1}; w_{t,i_{t-1}})^{z_{t-1}, i_{t-1}} \end{aligned} \quad (21)$$

$$\begin{aligned} \bar{q}_{t,t-1}(z_t = i_t | z_{t-1} = i_{t-1}; w_{t,i_{t-1}}) \\ = \frac{\exp(-\lambda \|g_{i_t} - g_{w_{t,i_{t-1}}}\|^2)}{\sum_{j_t=1}^M \exp(-\lambda \|g_{j_t} - g_{w_{t,i_{t-1}}}\|^2)} \end{aligned}$$

ここで、 w_1 と $w_{t,i_{t-1}}$ は整数値 1, 2, ..., M のみをとるパラメータであり、 $w_t = (w_{t,1}, \dots, w_{t,M})'$ 、 $\mathbf{W} = (w_1, w'_2, \dots, w'_T)'$ である。また、 λ は近傍関数の広がり方を決定するパラメータである。 z_{t,i_t} は $z_t = i_t$ の時 1 となり、それ以外の時 0 となる潜在変数である。この様にして分布 $\bar{q}_1(z_1; w_1)$ を式(20)と定義することにより、分布 $\bar{q}_1(z_1; w_1)$ は格子点 g_{z_1} と格子点 g_{w_1} との距離に従い滑らかに減少する分布となる。同様に、分布 $\bar{q}_{t,t-1}(z_t | z_{t-1} = i_{t-1}; w_{t,i_{t-1}})$ は格子点 g_{z_t} と格子点 $g_{w_{t,i_{t-1}}}$ との距離に従い滑らかに減少する分布となる。

SOHMMでは、分布 $q(\mathbf{Z})$ を分布 $\bar{q}(\mathbf{Z}; \mathbf{W})$ と制約しながら下界 $\mathcal{F}(q, \theta)$ を極大化することによりパラメータ θ の推定を行う。具体的には、式(22)、式(23)にて表わされるE-step、M-stepを交互に繰り返すことに

より下界 $\mathcal{F}(\bar{q}, \theta)$ を極大化するパラメータ θ の推定を行う。

E-step :

$$\begin{aligned} \mathbf{W}^\tau &= \arg \max_{\mathbf{W} \in Q} \mathcal{F}(\bar{q}, \theta^{\tau-1}) \\ &= \arg \min_{\mathbf{W} \in Q} \text{KL}(\bar{q}(\mathbf{Z}; \mathbf{W}) \| p(\mathbf{Z} | \mathbf{X}; \theta^{\tau-1})) \end{aligned} \quad (22)$$

M-step :

$$\begin{aligned} \theta^\tau &= \arg \max_{\theta} \mathcal{F}(\bar{q}^\tau, \theta) \\ &= \arg \max_{\theta} \mathbb{E}_{\bar{q}(\mathbf{Z}; \mathbf{W}^\tau)} [\ln p(\mathbf{X}, \mathbf{Z}; \theta)] \end{aligned} \quad (23)$$

ここで、 $Q = \{1, \dots, M\}^{M(T-1)+1}$ であり、 $\bar{q}^\tau = \bar{q}(\mathbf{Z}; \mathbf{W}^\tau)$ である。

この様にして対数尤度関数の下界 $\mathcal{F}(\bar{q}, \theta)$ を極大化することにより、SOHMMでは、隠れマルコフモデルのパラメータ θ が時系列データの分布に沿う様に推定される。また、式(2)を考慮すると、SOHMMのパラメータ推定法は対数尤度関数に罰則項 $\text{KL}(\bar{q}(\mathbf{Z}; \mathbf{W}) \| p(\mathbf{Z} | \mathbf{X}; \theta))$ を付加した罰則付き最尤推定の一つと解釈できる。ここで、罰則項は付録Aの展開により式(24)と展開することが可能であり、分布 $\bar{q}_{t,t-1}(z_t | z_{t-1} = i_{t-1}; w_{t,i_{t-1}})$ と分布 $p(z_t | z_{t-1} = i_{t-1}, \mathbf{X}; \theta)$ とのKLダイバージェンスの線形和と展開することが可能である。それゆえ、SOHMMでは、分布 $\bar{q}_{t,t-1}(z_t | z_{t-1} = i_{t-1}; w_{t,i_{t-1}})$ と分布 $p(z_t | z_{t-1} = i_{t-1}, \mathbf{X}; \theta)$ とがKLダイバージェンスの意味で近くなる様にパラメータ推定が行われる。

$$\begin{aligned} \text{KL}(\bar{q}(\mathbf{Z}; \mathbf{W}) \| p(\mathbf{Z} | \mathbf{X}; \theta)) \\ = \text{KL}(\bar{q}_1(z_1; w_1) \| p(z_1 | \mathbf{X}; \theta)) \\ + \sum_{t=2}^T \sum_{i_{t-1}=1}^M \bar{q}_{t,t-1}(i_{t-1}; w_{1:t-1}) \\ \times \text{KL}(\bar{q}_{t,t-1}(z_t | i_{t-1}; w_{t,i_{t-1}}) \| p(z_t | i_{t-1}, \mathbf{X}; \theta)) \end{aligned} \quad (24)$$

$$\begin{aligned} \bar{q}_t(z_t = i_t; w_{1:t}) &= \sum_{i_1=1}^M \cdots \sum_{i_{t-1}=1}^M \bar{q}_1(i_1; w_1) \\ &\times \prod_{u=2}^t \bar{q}_{u,u-1}(i_u | i_{u-1}; w_{u,i_{u-1}}) \end{aligned} \quad (25)$$

ここで、 $w_{1:t} = (w_1, w'_2, \dots, w'_T)'$ である。従って、分布 $\bar{q}_{t,t-1}(z_t | z_{t-1} = i_{t-1}; w_{t,i_{t-1}})$ が格子点 g_{z_t} と格子点 $g_{w_{t,i_{t-1}}}$ との距離に従い滑らかに減少する分布であることを考慮すると、対応する格子点 g_i が互いに近接する場合、状態 j から状態 i へと推移する確率 $p(z_t = i | z_{t-1} = j, \mathbf{X}; \theta)$ が互いに類似する様にパラメータ推定

が行われる。また、SOHMMでは、一時刻前の状態が与えられた際にその状態から推移する確率が高くなる格子点同士が可視化空間上で互いに近接する様にパラメータ推定が行われる。同様の理由により、対応する格子点 g_i が互いに近接する場合、最初の状態 z_1 が状態 i となる確率 $p(z_1=i|X; \theta)$ も互いに類似する様にパラメータ推定が行われる。

5.2 E-step

E-stepは、式(22)にて表わされるKLダイバージェンスを最小化することによりパラメータ W の更新を行う。しかしながら、パラメータ W の取りうるパターン数は $M^{M(T-1)+1}$ 通りであり、これら全てのパターンからKLダイバージェンスを最小化するパラメータ W^T を探索することは困難である。そこで本節では、より探索回数の少ないパラメータ W^T の探索アルゴリズムについて提案を行う。本アルゴリズムでは、 $t=T, \dots, 1$ の順にパラメータ $w_1, w_{t,i_{t-1}}$ を逐次的に探索することによりパラメータ W^T の探索を行う。それゆえ、各パラメータ $w_1, w_{t,i_{t-1}}$ 毎に M 回の探索を行えば良く、 $M \times (M(T-1)+1)$ 回の探索によりパラメータ W^T の探索が可能となる。

ここでは、式(22)にて表されるKLダイバージェンスを最小化するパラメータを $w_1^T, w_{t,i_{t-1}}^T$ と書き、 $w_i^T = (w_{i,1}^T, \dots, w_{i,M}^T)'$ 、 $W^T = (w_1^T, w_{t,i_{t-1}}^T, \dots, w_T^T)'$ と書く。また、簡単のためパラメータ θ^{T-1} を省略して書くこととする。まず、式(22)にて表わされるKLダイバージェンスを付録Bの展開により式(26)と求める。また、式(22)にて表わされるKLダイバージェンスからパラメータ $w_{t,i_{t-1}}$ に依存しない項を除いた関数を $R_{t,i_{t-1}}$ と定義し、付録Bの展開により式(27)、式(28)と求める。

$$\text{KL}(\tilde{q}(Z; W) \| p(Z|X)) = r_1(w_1, w_{2:T}) \quad (26)$$

$$R_{t,i_{t-1}} = \tilde{q}_{t-1}(i_{t-1}; w_{1:t-1}) \times r_{t,i_{t-1}}(w_{t,i_{t-1}}, w_{t+1:T}) \quad (27)$$

$$R_{T,i_{T-1}} = \tilde{q}_{T-1}(i_{T-1}; w_{1:T-1}) \times r_{T,i_{T-1}}(w_{T,i_{T-1}}) \quad (28)$$

ここで、 $w_{t:T} = (w_t^T, \dots, w_T^T)'$ であり、関数 r_1 と関数 $r_{t,i_{t-1}}$ は、式(29)、式(30)、式(31)と定義される関数である。また、確率 $p(z_1=i_1|X)$ と確率 $p(z_t=i_t|z_{t-1}=i_{t-1}, X)$ はforward-backwardアルゴリズムにより求める。

$$r_1(w_1, w_{2:T}) = \text{KL}(\tilde{q}_1(z_1; w_1) \| p(z_1|X)) + \sum_{i_1=1}^M \tilde{q}_1(i_1; w_1) r_{2,i_1}(w_{2,i_1}, w_{3:T}) \quad (29)$$

$$r_{t,i_{t-1}}(w_{t,i_{t-1}}, w_{t+1:T}) = \text{KL}(\tilde{q}_{t,t-1}(z_t|i_{t-1}; w_{t,i_{t-1}}) \| p(z_t|i_{t-1}, X)) + \sum_{i_t=1}^M \tilde{q}_{t,t-1}(i_t|i_{t-1}; w_{t,i_{t-1}}) \times r_{t+1,i_t}(w_{t+1,i_t}, w_{t+2:T}) \quad (30)$$

$$r_{T,i_{T-1}}(w_{T,i_{T-1}}) = \text{KL}(\tilde{q}_{T,T-1}(z_T|i_{T-1}; w_{T,i_{T-1}}) \| p(z_T|i_{T-1}, X)) \quad (31)$$

式(27)、式(28)と確率 $\tilde{q}_{t-1}(z_{t-1}=i_{t-1}; w_{1:t-1})$ が非負であることから、パラメータ $w_{t,i_{t-1}}^T$ を求めるにはパラメータ $w_{t+1:T}^T = (w_{t+1}^T, \dots, w_T^T)'$ を固定しながら関数 $r_{t,i_{t-1}}(w_{t,i_{t-1}}, w_{t+1:T}^T)$ を最小化すれば良い。従って、E-stepでは、 $t=T, \dots, 1$ の順に式(32)、式(33)、式(34)を逐次的に用いることによりパラメータ W^T の探索を行う。

E-step :

$$w_1^T = \arg \min_{w_1 \in \{1, \dots, M\}} r_1(w_1, w_{2:T}^T) \quad (32)$$

$$w_{t,i_{t-1}}^T = \arg \min_{w_{t,i_{t-1}} \in \{1, \dots, M\}} r_{t,i_{t-1}}(w_{t,i_{t-1}}, w_{t+1:T}^T) \quad (33)$$

$$w_{T,i_{T-1}}^T = \arg \min_{w_{T,i_{T-1}} \in \{1, \dots, M\}} r_{T,i_{T-1}}(w_{T,i_{T-1}}) \quad (34)$$

5.3 M-step

M-stepは、式(23)にて表わされる完全データの対数尤度の期待値を最大化することによりパラメータ θ の更新を行う。完全データの対数尤度の期待値は、式(35)と計算できる。

$$\begin{aligned} E_{\tilde{q}(Z; W^T)} [\ln p(X, Z; \theta)] &= \sum_{i_1=1}^M \dots \sum_{i_T=1}^M \tilde{q}(i_1, \dots, i_T; W^T) \\ &\times \left\{ \ln p(i_1) + \sum_{t=1}^T \ln p(x_t|i_t) + \sum_{t=2}^T \ln p(i_t|i_{t-1}) \right\} \end{aligned} \quad (35)$$

式(17)、式(18)にて表わされる制約条件のもとで対数尤度の期待値を最大化するパラメータ θ^T は、Lagrange の未定乗数法を用いることにより式(36)、式(37)、式(38)、式(39)と求められる。

$$\rho_i^T = \frac{\tilde{q}_1(i; w_1^T)}{\sum_{j=1}^M \tilde{q}_1(j; w_1^T)} \quad (36)$$

$$p_{ij}^{\tau} = \frac{\sum_{t=2}^T \tilde{q}_{t-1}(i; \mathbf{w}_{1:t-1}^{\tau}) \tilde{q}_{t,t-1}(j|i; \mathbf{w}_{t,i}^{\tau})}{\sum_{t=2}^T \tilde{q}_{t-1}(i; \mathbf{w}_{1:t-1}^{\tau})} \quad (37)$$

$$\mu_i^{\tau} = \left\{ \sum_{t=1}^T \tilde{q}_t(i; \mathbf{w}_{1:t}^{\tau}) \right\}^{-1} \sum_{t=1}^T \tilde{q}_t(i; \mathbf{w}_{1:t}^{\tau}) \mathbf{x}_t \quad (38)$$

$$\beta^{\tau} = \frac{DT}{\sum_{t=1}^T \sum_{i=1}^M \tilde{q}_t(i; \mathbf{w}_{1:t}^{\tau}) \|\mathbf{x}_t - \mu_i^{\tau}\|^2} \quad (39)$$

M-stepでは、式(36)、式(37)、式(38)、式(39)によりパラメータ θ の更新を行う。

5.4 SOHMMのアルゴリズム

以上から、SOHMMのアルゴリズムをまとめると以下になる。

step 1

$\tau = 1$ とし、パラメータ $\theta^{\tau-1}$ に初期値を設定する。

step 2

forward-backwardアルゴリズムにより確率 $p(z_1 = i_1 | X; \theta^{\tau-1})$ と確率 $p(z_t = i_t | z_{t-1} = i_{t-1}, X; \theta^{\tau-1})$ とを求める。

step 3 : E-step

$t = T, \dots, 1$ の順に式(32)～式(34)を用いてパラメータ W^{τ} を求める。

step 4 : M-step

式(36)～式(39)を用いてパラメータ θ^{τ} を求める。

step 5

$\tau = \tau + 1$ とし、下界 $\mathcal{F}(q, \theta)$ が収束するまで step 2 に戻って繰り返す。

5.5 状態の可視化

SOHMMでは、確率 $p(z_t = i | X; \theta)$ をforward-backwardアルゴリズムにより求め、この確率を状態を格子点 g_i に可視化する確率として使い、データの可視化を行う。具体的には、確率 $p(z_t = i | X; \theta)$ を用いて格子点の g_i 平均を求め、これにより状態 z_t の可視化を行う。

$$\bar{g}_t = \sum_{i=1}^M p(z_t = i | X; \theta) g_i$$

6. 計算機実験

本論文では、(1)人工時系列データ、(2)気象データの2種類の時系列データを用意し、可視化実験を行

う。ここでは、時点 t の時系列データ \mathbf{x}_t を入力する際に直近 $p-1$ 期間の過去の値 $\mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-p+1}$ も併せて入力するSOMMをTDSOMM(Time-Delay SOMM)と呼び、SOHMM、SOMM、TDSOMMの3種類を用いて可視化実験を行う。実験では、パラメータは $L=2$ 、 $M=6 \times 6$ 、 $\lambda=10$ と設定し、格子点 g_i は $[-1, 1] \times [-1, 1]$ の領域上に均等に配置を行う。また、TDSOMMの次数は $p=2$ と設定する。

6.1 人工時系列データ

6.1.1 実験方法

式(40)にて表わされる隠れマルコフモデルを用いて学習用時系列データ \mathbf{x}_t とテスト用時系列データ $\mathbf{x}_t^{\text{test}}$ をそれぞれ10000個づつ生成し、可視化実験を行う。

$$\mathbf{x}_t = \mu_{z_t^*} + \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d. } N(\mathbf{0}, \mathbf{I}) \quad (40)$$

状態数は $M^* = 5, \dots, 8$ の4種類を用意し、パラメータ μ_{z^*} はそれぞれが3次元立方体の頂点となるように式(41)と設定する。

$$\begin{aligned} \mu_1 &= (1, 1, 1)', & \mu_5 &= (-1, 1, 1)', \\ \mu_2 &= (1, 1, -1)', & \mu_6 &= (-1, 1, -1)', \\ \mu_3 &= (1, -1, 1)', & \mu_7 &= (-1, -1, 1)', \\ \mu_4 &= (1, -1, -1)', & \mu_8 &= (-1, -1, -1)' \end{aligned} \quad (41)$$

また、初期状態確率 ρ_i は $1/M^*$ と設定し、推移確率 p_{ij} は3次元立方体の辺に沿ってのみ状態推移するように式(42)と設定する。

$$\rho_i = 1/M^* \quad p_{ij} = \begin{cases} 0.9 & \text{if } i = j \\ 0.1/|L_i| & \text{if } \|\mu_i - \mu_j\| = 2 \\ 0 & \text{otherwise} \end{cases} \quad (42)$$

ここで、 $L_i = \{j | \|\mu_i - \mu_j\| = 2 \wedge 1 \leq j \leq M^*\}$ であり、 $|L_i|$ は集合 L_i の要素数である。

本実験では、可視化手法の評価尺度として状態の誤分類率 ER の定義を行い、これにより各可視化手法の評価を行う。まず、学習用時系列データを用いてSOMMのパラメータ推定を行い、パラメータ推定が終了した時点でテスト用時系列データを入力し、状態の誤分類率 ER を求める。ここでは、推定されたパラメータを用いて学習用時系列データの状態を可視化したベクトルを \bar{g}_t と書き、同様にテスト用時系列データの状態を可視化したベクトルを \bar{g}_t^{test} と書くこととする。そして、ベクトル $\bar{g}_1, \dots, \bar{g}_T$ の内ベクトル $\bar{g}_{t'}^{\text{test}}$ に最も類似したベクトル \bar{g}_{ω} に対応する状態 z_{ω}^* をテスト用時系列データの時点 t' の状態と分類する。本論文では、上記の分類結果とテストデータの真の状態

$z_{t'}^{*,\text{test}}$ とで状態が一致しない割合を状態の誤分類率 ER と定義し、これにより各可視化手法の分類のしやすさについて評価を行う。

$$ER = \frac{1}{T^{\text{test}}} \sum_{t'=1}^{T^{\text{test}}} \mathbb{I}[z_{t'}^{*,\text{test}} \neq z_{\omega_{t'}}^*],$$

ここで、 $\mathbb{I}[\cdot]$ は引数が真の時に 1 を、それ以外の時に 0 を返す指示関数であり、 T^{test} はテスト用時系列データの個数である。

6.1.2 実験結果

図 2 から図 4 に状態 z_t を \bar{g}_t を用いて可視化した結果を示す。図 2 より、SOHMM では $M^* = 5, \dots, 8$ 個のクラスタの可視化に成功している。一方、図 3、図 4 より、SOHMM と比べ SOMM、TDSOMM ではクラスタが大きく広がっており、各クラスタを識別することがより困難である。また、表 1 に人工時系列データにおける状態の誤分類率を示す。状態の誤分類率は各実験ごとに時系列データを作成し直し、10 回実験を行った平均値である。表 1 より、SOMM、TDSOMM と比べ SOHMM の方が状態の誤分類率を低く抑えており、より状態の分類がしやすい可視化であることが分かる。

図 5 に時点 $t = 0, \dots, 100$ における状態数 $M^* = 8$ の人工時系列データ x_t と状態 z_t^* とを示す。図 6 に期間 A ($t = 0, \dots, 23$) と期間 B ($t = 24, \dots, 97$) における \bar{g}_t の推移を示す。図 6 より、SOHMM では状態 z_t^* の推移にともない \bar{g}_t が図 2 にて表わされる 8 個のクラスタ間を移動して行く様が見て取れる。一方、SOMM、TDSOMM では、状態 z_t^* ごとにクラスタが形成されているものの、SOHMM と比べクラスタが大きく広がっており、状態推移を視覚的に把握することが難しい。

表 2 に人工時系列データのパラメータ推定に要した計算時間の平均値を示す。計算時間は状態数 $M^* = 5, \dots, 8$ の 4 種類に対してそれぞれ 10 回ずつ実験を行った平均値である。表 2 より SOHMM の方がより計算時間を必要としており、SOHMM の計算時間短縮は今後の課題である。

表 1 人工時系列データにおける状態の誤分類率

M^*	SOHMM	SOMM	TDSOMM
5	6.2%	32.0%	18.8%
6	6.4%	35.4%	20.5%
7	6.7%	38.3%	22.0%
8	10.6%	46.5%	27.8%

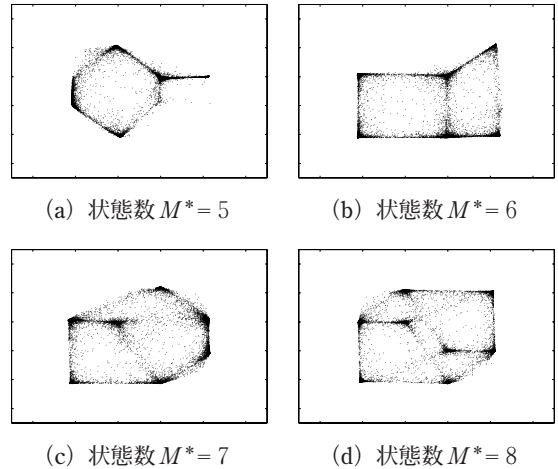


図 2 SOHMMによる人工時系列データの状態可視化

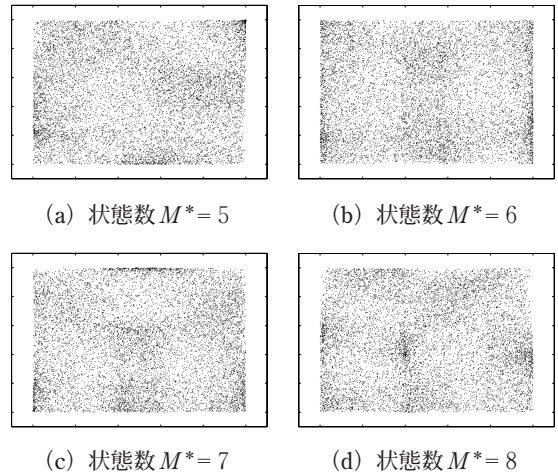


図 3 SOMMによる人工時系列データの状態可視化

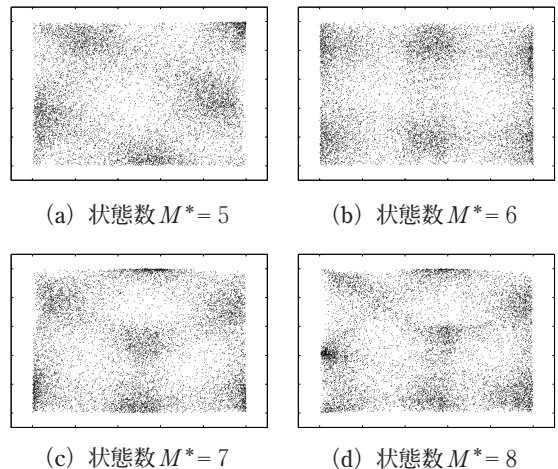


図 4 TDSOMMによる人工時系列データの状態可視化

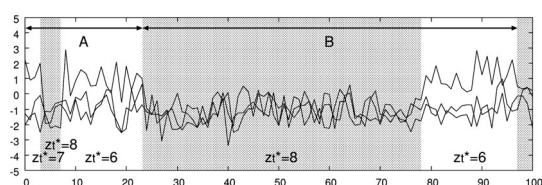
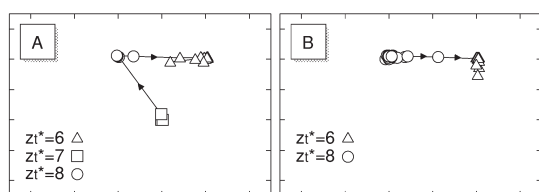
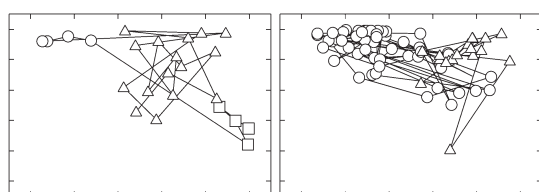


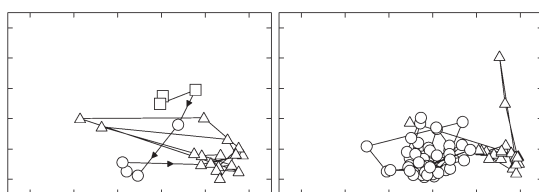
図5 時点 $t = 0, \dots, 100$ における状態数 $M^* = 8$ の人工時系列データとその状態



(a) SOHMM



(b) SOMM



(c) TDSOMM

図6 人工時系列データの期間A(左), 期間B(右) の状態推移の可視化

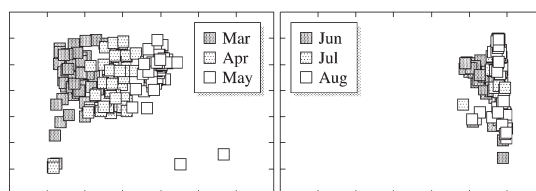
表2 人工時系列データにおける計算時間

SOHMM	SOMM	TDSOMM
295.4 sec	60.7 sec	59.6 sec

6.2 気象データ

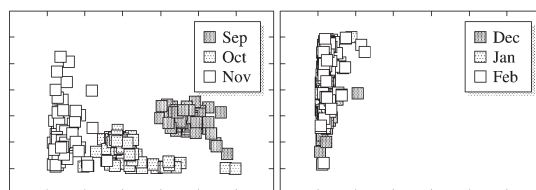
6.2.1 実験方法

気象庁が公表する東京都東京の気象データの内から日平均気温, 海面気圧, 現地気圧, 相対湿度, 蒸気圧, 降水量をまとめて6変量時系列データの生成を行い, 可視化実験を行う. 気象データは, 1951年1月から2010年5月までの713ヵ月分のデータを使用する.



(a) 春 (3月～5月)

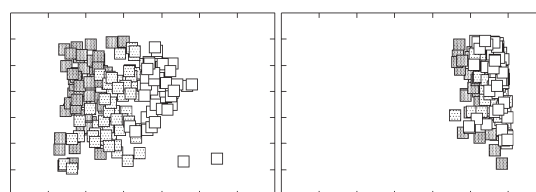
(b) 夏 (6月～8月)



(c) 秋 (9月～11月)

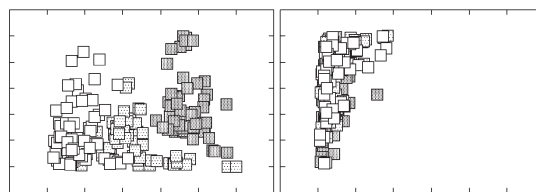
(d) 冬 (12月～2月)

図7 SOHMMによる気象データの状態可視化



(a) 春 (3月～5月)

(b) 夏 (6月～8月)



(c) 秋 (9月～11月)

(d) 冬 (12月～2月)

図8 SOMMによる気象データの状態可視化

6.2.2 実験結果

図7から図9に各月毎の状態 z_t を \bar{g}_t を用いて可視化した結果を示す. 図7から図9により, SOHMM, SOMM, TDSOMMの全てにおいて, \bar{g}_t が隣り合う月同士近接してまとめられていること, 1年を周期として可視化空間上を周回していることが見て取れる. また, 春(3月～5月), 夏(6月～8月), 秋(9月～11月), 冬(12月～2月)と季節ごとに見た場合, 夏と冬に比べ春と秋の方がよりクラスタが近接しており, 従って, 春と秋の方がより類似した季節であることが視覚的に理解できる. ただし, SOHMMと比べSOMM, TDSOMMの方がよりクラスタが大きく広がっており, SOMM, TDSOMMの方がより季節を識別することが困難である. また, 表3に気象データにおける状

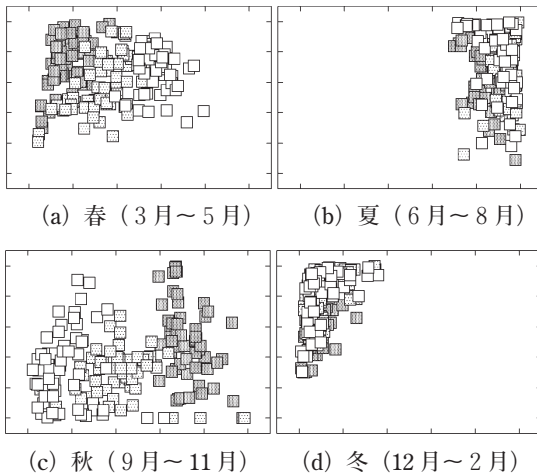


図9 TDSOMMによる気象データの状態可視化

表3 気象データにおける状態の誤分類率

SOHMM	SOMM	TDSOMM
39.1 %	55.5 %	48.9 %

表4 気象データにおける計算時間

SOHMM	SOMM	TDSOMM
178.8 sec	7.6 sec	7.6 sec

態の誤分類率を示す。状態の誤分類率は気象データを5個の時系列データ X_0, \dots, X_4 に分割し、 X_s を学習用時系列データとした際には $X_{(s+1) \bmod 5}$ をテスト用時系列データとし、 $s = 0, \dots, 4$ の5回実験を行った平均値である。表3よりSOMM、TDSOMMと比べSOHMMの方がより状態の誤分類率を低く抑えており、より状態の分類がしやすい可視化であることが分かる。

表4に気象データのパラメータ推定に要した計算時間の平均値を示す。計算時間は上記の時系列データ X_0, \dots, X_4 にて5回実験を行った平均値である。表4よりSOHMMの方がより計算時間を必要としており、SOHMMの計算時間短縮は今後の課題である。

7. むすび

本論文では、時系列データを隠れマルコフモデルにあてはめ、制約付きEMアルゴリズムを用いてパラメータ推定を行うSOHMMについて提案を行った。また、SOHMMのパラメータ推定法が対数尤度関数に自己組織化を行う罰則項を付加した罰則付き最尤推定の一つと解釈できることを示した。SOHMMを用いるこ

とにより、各時点における時系列システムの状態を可視化し、時系列システムの状態推移を視覚的に表現することが可能となる。計算機実験では、気象データ等を用いて可視化実験を行い、従来手法であるSOMMやTDSOMMと比べSOHMMの方がより季節の識別等を視覚的にしやすいことを確認した。

提案手法が持つ問題点についてまとめる。SOHMMでは、近傍関数の広がりを表すパラメータを適切に設定する必要がある。しかしながら、提案手法ではパラメータを経験的に設定する必要がある。今後の課題としてパラメータを自動的に設定する手法が望まれる。

参考文献

- [1] G. Edward, P. Box, and G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1994.
- [2] James D. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.
- [3] Geoffrey J. McLachlan and David Peel. *Finite Mixture Models*. John Wiley & Sons, Inc., New York, 2000.
- [4] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 39, No.1, pp.1-38, 1977.
- [5] Radford M. Neal and Geoffrey E. Hinton. A view of the EM algorithm that justifies incremental, sparse, and other variants. In *Learning in graphical models*, pp.355-368, Cambridge, MA, USA, 1999. MIT Press.
- [6] Jakob Verbeek, Nikos Vlassis, and Ben Kröse. Self-organizing mixture models. *Neurocomputing*, Vol.63, pp.99-123, jan 2005.
- [7] Jakob Verbeek. *Mixture Models for Clustering and Dimension Reduction*. PhD thesis, University of Amsterdam, 2004.
- [8] T. Kohonen, M. R. Schroeder, and T. S. Huang, editors. *Self-Organizing Maps*. Springer-Verlag, 2001.
- [9] L.R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, Vol.77, No.2, pp.257-286, 1989.
- [10] Christopher M. Bishop, Geoffrey E. Hinton, and Iain G. D. Strachan. GTM through time. In *IEE Fifth International Conference on Artificial Neural Networks*, pp.11-116, 1997.
- [11] Iván Olier and Alfredo Vellido. Advances in clustering and visualization of time series using GTM through time. *Neural Networks*, Vol.21, No.7, pp.904-913, 2008.
- [12] Christopher M. Bishop, Markus Svensén, and Christopher K. I. Williams. GTM: the generative topographic mapping. *Neural Computation*, Vol.10, No.1, pp.215-234, 1998.
- [13] J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann Publishers, 1988.

A 式 (24) の導出

本節では、簡単のためパラメータ $\theta, w_1, w_{t,i_{t-1}}, w_t, w_{1:t}, W$ を全て省略して記述する。まず、式 (21) より式 (43) を得る。

$$\sum_{i_t=1}^M \tilde{q}_{t,t-1}(i_t|z_{t-1}) = 1 \quad (43)$$

また、図 1 のグラフィカルモデルにおいて、 z_t は (z_{t-1}, X) により (z_1, \dots, z_{t-2}) から有向分離されており、従って、式 (44) を得る [13]。

$$p(z_t|z_{t-1}, \dots, z_1, X) = p(z_t|z_{t-1}, X) \quad (44)$$

式 (44) より式 (45) を得る。

$$\begin{aligned} p(Z|X) &= p(z_1|X) \prod_{t=2}^T p(z_t|z_{t-1}, \dots, z_1, X) \\ &= p(z_1|X) \prod_{t=2}^T p(z_t|z_{t-1}, X) \end{aligned} \quad (45)$$

式 (19) と式 (45) より $\text{KL}(\tilde{q}(Z)||p(Z|X))$ は、式 (46) と変形できる。

$$\begin{aligned} \text{KL}(\tilde{q}(Z)||p(Z|X)) &= E_{\tilde{q}} \left[\ln \frac{\tilde{q}(Z)}{p(Z|X)} \right] \\ &= E_{\tilde{q}} \left[\ln \frac{\tilde{q}_1(z_1) \prod_{t=2}^T \tilde{q}_{t,t-1}(z_t|z_{t-1})}{p(z_1|X) \prod_{t=2}^T p(z_t|z_{t-1}, X)} \right] \\ &= E_{\tilde{q}} \left[\ln \frac{\tilde{q}_1(z_1)}{p(z_1|X)} \right] + \sum_{t=2}^T E_{\tilde{q}} \left[\ln \frac{\tilde{q}_{t,t-1}(z_t|z_{t-1})}{p(z_t|z_{t-1}, X)} \right] \end{aligned} \quad (46)$$

式 (46) の第 1 項は、式 (19) と式 (43) より式 (47) と変形できる。

$$\begin{aligned} E_{\tilde{q}} \left[\ln \frac{\tilde{q}_1(z_1)}{p(z_1|X)} \right] &= \sum_{i_1=1}^M \dots \sum_{i_T=1}^M \tilde{q}(i_1, \dots, i_T) \ln \frac{\tilde{q}_1(i_1)}{p(i_1|X)} \\ &= \sum_{i_1=1}^M \tilde{q}_1(i_1) \ln \frac{\tilde{q}_1(i_1)}{p(i_1|X)} = \text{KL}(\tilde{q}_1(z_1)||p(z_1|X)) \end{aligned} \quad (47)$$

また、式 (46) の第 2 項は、式 (49) と変形できる。

$$\begin{aligned} \sum_{t=2}^T E_{\tilde{q}} \left[\ln \frac{\tilde{q}_{t,t-1}(z_t|z_{t-1})}{p(z_t|z_{t-1}, X)} \right] &= \sum_{t=2}^T \sum_{i_1=1}^M \dots \sum_{i_{T-1}=1}^M \tilde{q}(i_1, \dots, i_{T-1}) \ln \frac{\tilde{q}_{t,t-1}(i_t|i_{t-1})}{p(i_t|i_{t-1}, X)} \end{aligned} \quad (48)$$

$$\begin{aligned} &= \sum_{t=2}^T \sum_{i_1=1}^M \dots \sum_{i_{t-1}=1}^M \tilde{q}_1(i_1) \prod_{u=2}^{t-1} \tilde{q}_{u,u-1}(i_u|i_{u-1}) \\ &\quad \times \text{KL}(\tilde{q}_{t,t-1}(z_t|i_{t-1})||p(z_t|i_{t-1}, X)) \\ &= \sum_{t=2}^T \sum_{i_{t-1}=1}^M \tilde{q}_{t-1}(i_{t-1}) \\ &\quad \times \text{KL}(\tilde{q}_{t,t-1}(z_t|i_{t-1})||p(z_t|i_{t-1}, X)) \end{aligned} \quad (49)$$

ここで、式 (48) への変形は、式 (19) と式 (43) による。また、式 (49) への変形は、式 (25) による。式 (46)、式 (47)、式 (49) より式 (24) を得る。

B 式 (26), 式 (27), 式 (28) の導出

本節では、簡単のためパラメータ $\theta, w_1, w_{t,i_{t-1}}, w_t, w_{1:T}, w_{t:T}, W$ を全て省略して記述する。また、KL ダイバージェンスを式 (50) と省略して記述する。

$$\begin{aligned} \text{KL}_{t,t-1}(i_{t-1}) &= \text{KL}(\tilde{q}_{t,t-1}(z_t|i_{t-1})||p(z_t|i_{t-1}, X)) \end{aligned} \quad (50)$$

まず、式 (46)、式 (47)、式 (48) より式 (51) を得る。

$$\begin{aligned} \text{KL}(\tilde{q}(Z)||p(Z|X)) &= \text{KL}(\tilde{q}_1(z_1)||p(z_1|X)) \\ &+ \sum_{t=2}^T \left\{ \sum_{i_1=1}^M \dots \sum_{i_{t-1}=1}^M \tilde{q}_1(i_1) \prod_{u=2}^{t-1} \tilde{q}_{u,u-1}(i_u|i_{u-1}) \right. \\ &\quad \left. \times \text{KL}_{t,t-1}(i_{t-1}) \right\} \end{aligned} \quad (51)$$

更に、 $\sum_{t=2}^T$ を展開して書くことにより、式 (51) は式 (52) と変形できる。

$$\begin{aligned} (51) &= \text{KL}(\tilde{q}_1(z_1)||p(z_1|X)) \\ &+ \sum_{i_1=1}^M \tilde{q}_1(i_1) \times \text{KL}_{2,1}(i_1) \\ &+ \sum_{i_1=1}^M \sum_{i_2=1}^M \tilde{q}_1(i_1) \tilde{q}_{2,1}(i_2|i_1) \times \text{KL}_{3,2}(i_2) \\ &\quad \vdots \\ &+ \sum_{i_1=1}^M \dots \sum_{i_{T-1}=1}^M \tilde{q}_1(i_1) \prod_{u=2}^{T-1} \tilde{q}_{u,u-1}(i_u|i_{u-1}) \\ &\quad \times \text{KL}_{T,T-1}(i_{T-1}) \end{aligned} \quad (52)$$

ここで、 $\tilde{q}_1(i_1), \tilde{q}_{2,1}(i_2|i_1), \dots, \tilde{q}_{T-1,T-2}(i_{T-1}|i_{T-2})$ の順で括ることにより、式 (52) は式 (53) と変形できる。

$$\begin{aligned}
(52) = & \left\{ \text{KL}(\tilde{q}_1(z_1) \| p(z_1 | \mathbf{X})) + \sum_{i_1=1}^M \tilde{q}_1(i_1) \right. \\
& \times \left\{ \text{KL}_{2,1}(i_1) + \sum_{i_2=1}^M \tilde{q}_{2,1}(i_2 | i_1) \right. \\
& \times \left\{ \text{KL}_{3,2}(i_2) + \sum_{i_3=1}^M \tilde{q}_{3,2}(i_3 | i_2) \right. \\
& \vdots \\
& \times \left. \left. \left. \left. \left. \left. \text{KL}_{T,T-1}(i_{T-1}) \right\} \cdots \right\} \right\} \right\} \right\}
\end{aligned} \quad (53)$$

式(53)の波括弧を内側より順に $r_{T,i_{T-1}}, \dots, r_{2,i_1}, r_1$ と定義することにより式(26)を得る。

同様に、式(52)からパラメータ $w_{t,i_{t-1}}$ に依存しない項を除くことにより、関数 $R_{t,i_{t-1}}$ を式(54)と求める。

$$\begin{aligned}
R_{t,i_{t-1}} = & \tilde{q}_{t-1}(i_{t-1}) \times \text{KL}_{t,t-1}(i_{t-1}) \\
& + \tilde{q}_{t-1}(i_{t-1}) \sum_{i_t=1}^M \tilde{q}_{t,t-1}(i_t | i_{t-1}) \times \text{KL}_{t+1,t}(i_t) \\
& \vdots \\
& + \tilde{q}_{t-1}(i_{t-1}) \sum_{i_t=1}^M \cdots \sum_{i_{T-1}=1}^M \prod_{u=t}^{T-1} \tilde{q}_{u,u-1}(i_u | i_{u-1}) \\
& \times \text{KL}_{T,T-1}(i_{T-1})
\end{aligned} \quad (54)$$

ここで、式(54)は、 $\tilde{q}_{t-1}(i_{t-1}), \tilde{q}_{t,t-1}(i_t | i_{t-1}), \dots, \tilde{q}_{T-1,T-2}(i_{T-1} | i_{T-2})$ の順に括弧することにより、式(55)と変形できる。

$$\begin{aligned}
R_{t,i_{t-1}} = & \tilde{q}_{t-1}(i_{t-1}) \\
& \times \left\{ \text{KL}_{t,t-1}(i_{t-1}) + \sum_{i_t=1}^M \tilde{q}_{t,t-1}(i_t | i_{t-1}) \right. \\
& \times \left\{ \text{KL}_{t+1,t}(i_t) + \sum_{i_{t+1}=1}^M \tilde{q}_{t+1,t}(i_{t+1} | i_t) \right. \\
& \vdots \\
& \times \left. \left. \left. \left. \left. \left. \text{KL}_{T,T-1}(i_{T-1}) \right\} \cdots \right\} \right\} \right\}
\end{aligned} \quad (55)$$

式(55)の波括弧を内側より順に $r_{T,i_{T-1}}, \dots, r_{t,i_{t-1}}$ と定義することにより、式(27)、式(28)を得る。

(2012年4月2日 受付)

(2012年6月19日 採録)

[問い合わせ先]

〒840-8502 佐賀県佐賀市本庄町1

佐賀大学大学院 工学研究科 知能情報システム学専攻

山口 暢彦

TEL: 0952-28-8566

FAX: 0952-28-8650

E-mail: yamag@is.saga-u.ac.jp

著者紹介



やまぐち のぶひこ
山口 暢彦 [正会員]

1996年名工大・工・知能情報システム卒。平10同大大学院修士課程了。同年富士通名古屋通信システム(株)入社。平15名工大大学院博士課程了。現在、佐賀大学大学院工学系研究科助教。パターン認識・学習・データの可視化に関する研究に従事。

Visualizing States in Hidden Markov Models using SOHMM

by

Nobuhiko YAMAGUCHI

Abstract :

The self-organizing mixture models (SOMMs) were proposed as an expectation-maximization (EM) algorithm that yields topology preserving maps of data based on probabilistic mixture models. Compared to self-organizing maps, the SOMM algorithm has a clear interpretation: it maximizes the sum of data log likelihood and a penalty term that enforces self-organization. The object of this paper is to extend the SOMM algorithm to deal with multivariate time series. The standard SOMM algorithm assumes that the data are independent and identically distributed samples. However, the i.i.d. assumption is clearly inappropriate for time series. In this paper we propose the extension of the SOMM algorithm for multivariate time series, which we call self-organizing hidden Markov models (SOHMMs), by assuming that the time series is generated by hidden Markov models (HMMs).

Keywords : self-organizing mixture models, visualization, time series, hidden Markov models

Contact Address : **Nobuhiko YAMAGUCHI**

Graduate School of Science and Engineering, Saga University
1, Honjo-machi, Saga, 840-8502, JAPAN
TEL : 0952-28-8566
FAX : 0952-28-8566
E-mail : yamag@is.saga-u.ac.jp