

明 細 書

発明の名称： 隠れ変数モデル推定装置および方法

技術分野

[0001] 本発明は、系列依存性を持つ多変量データの隠れ変数モデル推定装置、隠れ変数モデル推定方法、および隠れ変数モデル推定プログラムに関し、特に、モデル事後確率を近似し、その下界を最大化する事によって系列依存性を持つ多変量データの隠れ変数モデルを推定する隠れ変数モデル推定装置、隠れ変数モデル推定方法、および隠れ変数モデル推定プログラムに関する。

背景技術

[0002] 系列依存性を持つデータが種々存在する。系列依存性を持つデータの例として、例えば、時間依存性を有するデータや、文字系列に依存する文章、塩基系列に依存する遺伝子データ等が挙げられる。

[0003] 自動車から取得されるセンサデータ、健康診断の検査値履歴、電力需要履歴などに代表されるデータは、すべて「系列依存性（本例では時間的な依存性）」を持つ多変量データである。このようなデータの分析は、産業上重要な数多くの分野に適用される。例えば、自動車から取得されるセンサデータの分析によって、自動車の故障原因を解析して素早い修理を実現することが考えられる。また、健康診断の検査値履歴の分析によって、疾患のリスクの推定および疾患の予防を実現することが考えられる。また、電力需要履歴の分析によって、電力の需要を予測して過不足に備えられるようにすることが考えられる。

[0004] このようなデータに対しては、系列依存性を持つ隠れ変数モデル（例えば、隠れマルコフモデル）を用いてモデル化する事が一般的である。例えば、隠れマルコフモデルを利用するためには、隠れ状態数、観測確率分布の種類、及び分布パラメータを決定する必要がある。隠れ状態数と観測確率分布の種類がわかっている場合には、Expectation Maximization法（例えば非特許文献1 参照）を利用してパラメータを推定する事が可能である。

- [0005] 隠れ状態数や観測確率の種類を決定する問題は、一般的に「モデル選択問題」や「システム同定問題」と呼ばれ、信頼性のあるモデルを構築するために極めて重要な問題である。そのための技術が種々提案されている。
- [0006] 例えば、非特許文献2では、隠れ状態数を決定する方法として、変分ベイズ法によって変分自由エネルギーを最大化する方法が提案されている。また、例えば、非特許文献3では、隠れ状態数を決定する方法として、階層Dirichlet過程事前分布を用いたノンパラメトリックベイズ法が提案されている。
- [0007] また、非特許文献4では、時間依存性のない隠れ変数モデルの代表例である混合モデルに対して、完全周辺尤度関数を近似し、その下界を最大化することが述べられている。

先行技術文献

非特許文献

- [0008] 非特許文献1: C. Bishop, Pattern Recognition and Machine Learning, Springer, 2007, pp.610-629
- 非特許文献2: Beal, M. J. Variational Algorithms for Approximate Bayesian Inference. Chapter3, PhD thesis, University College London, 2003
- 非特許文献3: van Gael, J., Saatci, Y, Teh, Y.-W., and Ghahramani, Z. Beam sampling for the infinite hidden Markov model. In ICML, 2008.
- 非特許文献4: Ryohei Fujimaki, Satoshi Morinaga: Factorized Asymptotic Bayesian Inference for Mixture Modeling. Proceedings of the the fifth international conference on Artificial Intelligence and Statistics (AISTATS), 2012

発明の概要

発明が解決しようとする課題

- [0009] 非特許文献2に記載の方法では、周辺化尤度関数の下界を最大化する際に、変分分布上における隠れ状態と分布パラメータの独立性を仮定するため、周辺化尤度の近似精度が悪くなるという問題がある。

- [0010] 非特許文献3に記載の方法では、モンテカルロ法に基づく最適化アルゴリズムが知られているが、計算量が非常に多くなるという問題がある。
- [0011] 非特許文献2に記載の方法と非特許文献3に記載の方法で観測確率の種類を決定することは、計算量が極めて多くなるという問題のため、事実上困難である。
- [0012] この計算量の問題を、観測確率分布が混合多項式曲線となる場合を例に説明する。なお、隠れ状態については、以下の議論には影響しないため省略する。ある隠れ状態に対する観測が多項式曲線となる場合には、1次曲線（直線）、2次曲線、3次曲線など曲線の次数を正しく選ぶ必要がある。上記の方法では、例えば、隠れ状態数が3で直線と2次曲線が2つの場合、隠れ状態数が5で3次曲線が3つと4次曲線が2つの場合等の全てのモデルの候補に対して情報量基準を計算する必要があった。このモデルの候補の数は、例えば隠れ状態数が10で、曲線の最大次数を10とすると約十万通りとなり、隠れ状態数が20で曲線の最大次数を20とすると数百億通りとなる。このようにモデルの候補の数は、探索すべきモデルの複雑さと共に指数的に増加する。従って、非特許文献2および非特許文献3に記載の方法による計算は、事実上困難であった。
- [0013] また、非特許文献4に記載の技術では、隠れ変数間の独立性が必要となり、系列依存性のある隠れ変数モデルに適用できなかった。非特許文献4に記載の技術では、隠れ変数間の系列依存性を考慮していないため、隠れ変数の変分分布が、非特許文献4の式（15）として算出される。しかし、隠れ変数間に系列依存性がある場合には、この式は適切でなく、適切なモデルが得られる保証はない。さらに、隠れ変数間の遷移確率が算出できないという問題がある。
- [0014] そこで、本発明は、多変量データに対する系列依存性を持つ隠れ変数モデルの学習問題において、隠れ状態数および観測確率の種類と共にモデルの候補数が指数的に増加しても高速にモデル選択を実現できるようにすることを目的とする。

課題を解決するための手段

[0015] 本発明による隠れ変数モデル推定装置は、周辺化対数尤度関数を完全変数に対する推定量に関してラプラス近似した近似量の下界として定義される基準値を最大化することによって変分確率を計算する変分確率計算部と、各隠れ状態に対して観測確率の種類とパラメータを推定することによって変分確率を計算する変分確率計算部が変分確率を計算する際に用いた基準値が収束したか否かを判定する収束判定部とを備えることを特徴とする。

[0016] また、本発明による隠れ変数モデル推定方法は、周辺化対数尤度関数を完全変数に対する推定量に関してラプラス近似した近似量の下界として定義される基準値を最大化することによって変分確率を計算し、各隠れ状態に対して観測確率の種類とパラメータを推定することによって変分確率を計算する際に用いた基準値が収束したか否かを判定することとを特徴とする。

[0017] また、本発明による隠れ変数モデル推定プログラムは、コンピュータに、周辺化対数尤度関数を完全変数に対する推定量に関してラプラス近似した近似量の下界として定義される基準値を最大化することによって変分確率を計算する変分確率計算処理、各隠れ状態に対して観測確率の種類とパラメータを推定することによって最適な隠れ変数モデルを推定するモデル推定処理、および、変分確率計算処理で変分確率を計算する際に用いた基準値が収束したか否かを判定する収束判定処理を実行させる。

発明の効果

[0018] 本発明によれば、隠れ状態数および観測確率の種類と共にモデルの候補数が増加しても高速にモデル選択を行える。

図面の簡単な説明

[0019] 図1]本発明の隠れ変数モデル推定装置の第1の実施形態を示すブロック図である。

[図2]隠れ変数変分確率計算処理部104の例を示すブロック図である。

[図3]本発明の第1の実施形態の処理経過の例を示すフローチャートである。
 [図4]隠れ変数変分確率計算処理部104の動作を示すフローチャートである。
 [図5]本発明の隠れ変数モデル推定装置の第2の実施形態を示すブロック図である。
 [図6]本発明の隠れ変数モデル推定装置の概要を示すブロック図である。

発明を実施するための形態

[0020] 以下、本発明の実施形態を図面を参照して説明する。なお、以下の説明では便宜的に、数式内の表記と、文章中の表記とが異なる場合がある。例えば、記号“~”を数式内では、変数の上部に記載するか、文章中では便宜的に右側に記載する等の違いがある。このような数式内の表記と文章中の表記との相違は、当業者が理解し得る範囲内の相違である。

[0021] 本発明の隠れ変数モデル推定装置は、系列依存性を持つ隠れ変数モデルを推定する。以下の説明では、系列依存性を持つデータの例として時間依存性を持つデータを例にして説明するが、本発明におけるデータは系列依存性を持つデータであればよく、時間依存性を持つデータに限定されない。例えば、文字系列に依存するデータ、塩基系列に依存するデータ、あるいはその他

[0022] また、以下の説明では、系列依存性を持つ隠れ変数モデルの最も代表的なモデルである隠れマルコフモデル(式(1)参照)を例にして具体的な説明を行う。

[0023]

[数1]

$$p(\mathbf{x}^N, \mathbf{z}^N | M) = \int \prod_{n=1}^N \left\{ p(\mathbf{z}_n^1 | \boldsymbol{\alpha}) \prod_{k=1}^K \prod_{t=2}^{T_n} p_k(\mathbf{z}_n^t | \boldsymbol{\beta}_k)^{z_{nk}^{t-1}} \right. \\ \left. \times \prod_{k=1}^K \prod_{t=1}^{T_n} p(\mathbf{x}_n^t | \boldsymbol{\phi}_k)^{z_{nk}^t} \right\} p(\boldsymbol{\theta} | M) d\boldsymbol{\theta},$$

式 (1)

[0024] なお、以下の説明では時間依存するデータ列 $\mathbf{x}^n(n=1, \dots, N)$ が入力されると仮定する。ここで、各 $\mathbf{x}^n$ は、長さ T_n の多変量データ列 ($\mathbf{x}^n = \mathbf{x}^{n1}, \dots, \mathbf{x}^{nT}, t=1, \dots, N$) であるとする。次に、観測変数 $\mathbf{x}^{nt}$ に対する隠れ変数 $\mathbf{z}^{nt} = (z^{nt1}, \dots, z^{ntK})$ を定義する。 $z^{ntk}=1$ は、 $\mathbf{x}^{nt}$ が k 番目の隠れ状態から生成されたデータであることを意味し、 $z^{ntk}=0$ は、そうでないことを意味する。また、 $\sum_{k=1}^K z^{ntk} = 1$ である。 $\mathbf{x}$ と $\mathbf{z}$ の組は「完全変数」と呼ばれる。なお、その対比として $\mathbf{x}$ を不完全変数と呼ぶ。完全変数に関する隠れマルコフモデルの同時分布は、式 (1) 中の $P(\mathbf{x}, \mathbf{z})$ として定義される。なお、隠れマルコフモデルの隠れ変数に対する変分分布は、時刻 t における k 番目の隠れ状態 z^{ntk} の分布 $q(z^{ntk})$ 、及び、時刻 $t-1$ から時刻 t において k 番目の状態から j 番目の状態へ遷移する分布 $q(z^{nt-1k}, z^{ntj})$ と表される。

[0025] 式 (1) において、 K は隠れ状態数を表す。また、 $\boldsymbol{\theta} = (\alpha_1, \dots, \alpha_K, \beta_1, \dots, \beta_K, \phi_1, \dots, \phi_K)$ は、モデルのパラメータを表す。ここで、 α_k は k 番目の隠れ状態の初期確率を表し、 β_k は k 番目の隠れ状態からの遷移確率を表し、 ϕ_k は k 番目の隠れ状態に対する観測パラメータを表す。また、 $S_1, \dots, S_K$ は、 ϕ_k に対応する観測確率の種類を表す。なお、 S_1 から S_K となりうる候補は、例えば、多変量データの生成確率の場合には {正規分布、対数正規分布、指数分布} であったり、多項曲線出力の場合では、{0次曲線、1次曲線、2次曲線、3次曲線} であったりする。

[0026] なお、本明細書では具体的な例はすべて隠れマルコフを用いて説明をする

が、その拡張モデル（例えば隠れセミマルコフモデル、因子化隠れマルコフモデルなど）の類似のモデルにも本発明を適用可能である。同様に、本明細書では、ターゲット変数を X とした分布について説明しているが、回帰や判別のように、観測分布が条件付モデル $P(Y|X)$ （ Y はターゲットとなる確率変数）である場合に関しても本発明を適用可能である。

[0027] 実施形態 1.

図 1 は、本発明の隠れ変数モデル推定装置の第 1 の実施形態を示すブロック図である。隠れ変数モデル推定装置 100 は、データ入力装置 101 と、隠れ状態数設定部 102 と、初期化处理部 103 と、隠れ変数変分確率計算処理部 104 と、モデル最適化処理部 105 と、最適性判定処理部 106 と、最適モデル選択処理部 107 と、モデル推定結果出力装置 108 とを備える。隠れ変数モデル推定装置 100 には、入力データ 111 が入力され、入力データ 111 に対して隠れ状態数および観測確率の種類を最適化し、モデル推定結果 112 として出力する。

[0028] また、図 2 は、隠れ変数変分確率計算処理部 104 の例を示すブロック図である。隠れ変数変分確率計算処理部 104 は、前向き確率計算処理部 1041 と、正規化定数記憶部 1042 と、後ろ向き確率計算処理部 1043 と、前向き後ろ向き確率合算処理部 1044 を備える。隠れ変数変分確率計算処理部 104 には、入力データ 111 と、モデル最適化処理部 105 で推定された推定モデル 1045 とが入力され、隠れ変数変分確率 1046 と、前向き確率正規化定数 1047 とを出力する。

[0029] 入力装置 101 は、入力データ 111 を入力するための入力インタフェース装置である。入力装置 101 には、入力データ 111 が入力される際に、観測確率の種類や、隠れ状態数の候補値など、モデルの推定に必要なパラメータも同時に入力される。

[0030] 隠れ状態数設定部 102 は、入力された隠れ状態数の候補値からモデルの隠れ状態数を選択して設定する。以下、設定された隠れ状態数を K と表記する。

[0031] 初期化処理部 103 は、推定のための初期化処理を実施する。なお、初期化は任意の方法によって実施することが可能である。例としては、観測確率の種類を隠れ状態ごとにランダムに設定し、設定された種類に従って、各観測確率のパラメータをランダムに設定する方法や、隠れ変数の変分確率をランダムに設定する方法が挙げられる。

[0032] 隠れ変数変分確率計算処理部 104 は、隠れ変数の変分確率を計算する。ここで、パラメータ θ は初期化処理部 103 あるいはモデル最適化処理部 106 で計算されているため、隠れ変数変分確率計算処理部 104 は、その値を利用する。隠れ変数変分確率計算処理部 104 は、次に定義する最適化基準 A を最大化することによって変分確率を計算する。最適化基準 A とは、周辺化対数尤度関数を完全変数に対する推定量（例えば最尤推定量や最大事後確率推定量）に関してラプラス近似した近似量の下界として定義される。なお、この下界は、完全変数に対する推定量の最適性と対数関数の凹性を用いることで導出することが可能である。

[0033] この手順を、隠れマルコフモデルを例に説明する。まず、周辺化対数尤度関数の下界を考える。この下界は、以下の式（2）で示される

[0034] [数2]

$$\log p(\mathbf{x}^N | M) \geq \sum_{\mathbf{z}^N} q(\mathbf{z}^N) \log \left(\frac{p(\mathbf{x}^N, \mathbf{z}^N | M)}{q(\mathbf{z}^N)} \right)$$

式（2）

[0035] なお、式（2）において、変分確率 $q(\mathbf{z}^N)$ を最大化する事で等号が成立する。ここで、完全変数に対する最尤推定量を用いて分子の完全変数の周辺化尤度をラプラス近似する事で、周辺化対数尤度関数の近似式として、以下に示す式（3）を得る。

[0036]

[数3]

$$\begin{aligned}
\mathcal{J}(q, \bar{\boldsymbol{\theta}}, \boldsymbol{x}^N) &= \sum_{\boldsymbol{z}^N} q(\boldsymbol{z}^N) \left(\log p(\boldsymbol{x}^N, \boldsymbol{z}^N | \bar{\boldsymbol{\theta}}) \right. \\
&\quad - \frac{\mathcal{D}_{\alpha}}{2} \log N - \sum_{k=1}^K \frac{\mathcal{D}_{\beta_k}}{2} \log \left(\sum_{n,t=1}^{N, T_n-1} z_{nk}^t \right) \\
&\quad \left. - \sum_{k=1}^K \frac{\mathcal{D}_{\phi_k}}{2} \log \left(\sum_{n,t=1}^{N, T_n} z_{nk}^t \right) - \log q(\boldsymbol{z}^N) \right).
\end{aligned}$$

式 (3)

[0037] ただし、上付きのバーは完全変数に対する最尤推定量を表す。また、 $\mathcal{D}_*$ は下付きのパラメータ*の次元を表す。

[0038] 次に、式 (3) に対して最尤推定量が対数尤度関数を最大化する性質と、対数関数が凹関数である事を利用して、式 (3) の下界を以下の式 (4) のように算出する。

[0039]

[数4]

$$\mathcal{G}(q, \tilde{q}, x_N, \theta)$$

$$= \sum_N \sum_{z_n=1}^n q(z_n) \left[\log p(x_n, z_n | \theta) + \sum_{K, T_n} z_{nK}^t \log \delta_K^t \right] + \sum_{N, T_n} \log q(z_N) \left[+ \sum_{n,t=1}^n \log \Delta_t - \frac{D_\alpha}{2} \log N - \sum_{K=1}^K \left(\frac{D_{\beta_K}}{2} \times \right. \right.$$

$$\left. \left. \sum_{N, T_n-1}^{n,t} \log(\tilde{q}(z_{nK}^t) - 1) + \frac{D_{\phi_K}}{2} (\log(\tilde{q}(z_{nK}^t) - 1) \right), \right]$$

$$\delta_K^t = \begin{cases} \frac{1}{\Delta_t} \exp \left(- \frac{D_{\beta_K}}{2} \frac{1}{\sum_{n,t=1}^n q(z_{nK}^t)} \right) & \text{if } t = T_n \\ \frac{1}{\Delta_t} \exp \left(- \frac{D_{\beta_K}}{2} \frac{1}{\sum_{n,t=1}^n q(z_{nK}^t)} \right) & \text{if } t > T_n \end{cases}$$

式(4)

[0040] 隠れ変数の変分分布 $q(z_{ntk})$ 及び $q(z_{ntj})$ は、式(4)を q について最
大化する事によって算出される。ただし、上付き(i)を、隠れ変数変分確率計
算処理部104、モデル最適化処理部105、最適判定処理部106の繰り
返し計算における、(i)回目の繰り返しを表すとして、 $q^{(i)}$ は $\tilde{q} = q^{(i-1)}$ 、 $\theta =$
 $\theta^{(i-1)}$ と固定する。

[0041] なお、式(4)において下線を付した部分をBとする。Bは、後述の式(8)で参照する。

[0042] 図2を参照して、隠れ変数変分確率計算処理部104が備える要素について説明する。前向き確率計算処理部1041には、入力データ111と推定モデルが入力される。そして、前向き確率計算処理部1041は、時刻1から

時刻 t までの観測 ($x^{n1}, \dots, x^{nt}$) が得られた場合の z^{nt} の確率を前向き確率として算出する。ただし、前向き確率は、最適化基準 A で算出されるモデル複雑度（例えば式（４）では δtk に関する項）を考慮して算出される。また、前向き確率計算処理部 1041 は、 z^{nt} の確率の隠れ状態に関する和を 1 とするための正規化定数を、正規化定数記憶部 1042 に記憶させる。

[0043] 同様に、後ろ向き確率計算処理部 1043 は、時刻 $t+1$ から T までの観測 ($x^{nt+1}, \dots, x^{nT}$) が得られた場合の x^{nt} の確率を後ろ向き確率として算出する。なお、後ろ向き確率の計算の際に、前向き確率算出と同時に得られる正規化定数を正規化定数記憶部 1042 から読み込む。ただし、後ろ向き確率は最適化基準 A で算出されるモデル複雑度（例えば式（４）では δtk に関する項）を考慮して算出される。

[0044] 最後に、前向き後ろ向き確率合算処理部 1044 は、前向き確率と後ろ向き確率から、変分分布を算出する。例えば、前向き後ろ向き確率合算処理部 1044 は、 $q(z^{ntk})$ を $x^{n1}, \dots, x^{nT}$ が得られたときの z^{ntk} の確率として計算する。前向き後ろ向き確率合算処理部 1044 は、前向き確率と後ろ向き確率の積として、以下の式（５）の計算によって、 $q(z^{ntk})$ を計算する。

[0045] [数5]

$$q^{(i)}(z_{nk}^t) = f_{nk}^{t(i)} b_{nk}^{t(i)} \quad \text{式（５）}$$

[0046] また、前向き後ろ向き確率合算処理部 1044 は、 $x^{n1}, \dots, x^{nt-1}$ が得られたときの z^{nt-1j} の確率と、隠れ状態 j から隠れ状態 k へ遷移する確率と、隠れ状態 k において x^{nt} が観測される確率と、($x^{nt+1}, \dots, x^{nT}$) が得られた場合の x^{nt} の確率の積として、 $q(z^{nt-1j}, z^{ntk})$ を算出する。具体的には、前向き後ろ向き確率合算処理部 1044 は、以下の式（６）の計算によって $q(z^{nt-1j}, z^{ntk})$ を算出する（式（６）左辺の $q^{\sim}$ の定義は式（７）を参照）。

[0047]

[数6]

$$q^{(i)}(z_{nj}^{t-1}, z_{nk}^t) = \frac{1}{\zeta_n^{t(i)}} f_{nj}^{t-1(i)} \tilde{p}(\mathbf{x}_n^t | \phi_k^{(i-1)}) \beta_{jk}^{(i-1)} b_{nk}^{t(i)}$$

式 (6)

[0048] この手順を隠れマルコフモデルを例に説明すると、前向き確率及び後ろ向き確率は、以下の式 (7) の計算によって算出される。

[0049] [数7]

$$f_{nk}^{t(i)} = \begin{cases} \frac{1}{\zeta_n^{1(i)}} \alpha_k^{(i-1)} \tilde{p}(\mathbf{x}_n^1 | \phi_k^{(i-1)}) & \text{if } t = 1 \\ \frac{1}{\zeta_n^{t(i)}} \tilde{p}(\mathbf{x}_n^t | \phi_k^{(i-1)}) \sum_{j=1}^K f_{nj}^{t-1(i)} \beta_{jk}^{(i-1)} & \end{cases}$$

$$b_{nk}^{t(i)} = \begin{cases} \frac{1}{\zeta_n^{t+1(i)}} \sum_{j=1}^K b_{nj}^{t+1(i)} \tilde{p}(\mathbf{x}_n^{t+1} | \phi_j^{(i-1)}) \beta_{kj}^{(i-1)} & \\ 1 & \text{if } t = T_n \end{cases}$$

$$\tilde{p}(\mathbf{x}_n^t | \phi_k^{(i-1)}) = p(\mathbf{x}_n^t | \phi_k^{(i-1)}) \delta_k^{t(i-1)}.$$

式 (7)

[0050] ただし、 f_{nk}^t (式 (7) の第1式) が前向き確率を表し、 b_{nk}^t (式 (7) の第2式) が後ろ向き確率を表す。より具体的には、式 (7) において、前向き確率と後ろ向き確率の両者とも、漸化式として記述されている。そして、前向き確率は $t=1$ から順に算出することが可能であり、後ろ向き確率は $t=T$ から順に算出することが可能である。なお、正規化定数は ζ_n^t で算出される。後ろ向き確率計算処理部 1043 は、前向き確率計算処理部 1041 が前向き確率を算出する際に計算した正規化定数を利用して後ろ向き確率を算出すればよい。

[0051] また、式 (5) の第3式において、 δ に関する乗算が含まれているが、これは、最適化基準Aで算出されるモデル複雑度を考慮していることを意味する

。

[0052] モデル最適化処理部 105 は、式 (4) に対してモデル (パラメータ θ 及びその種類 S) を最適化する。具体的には、 q 及び $q^{\sim}$ を隠れ変数変分確率計算処理部 104 で算出された隠れ変数の変分分布 ($q^{(i)}$) に固定し、式 (4) における G を最大化するモデルを算出する。この処理において重要な点は、式 (4) によって定義された G は、コンポーネントごとに最適化の関数を分解する事が可能なため、コンポーネント種類の組合せ (S_1 から S_K のどの種類を指定するか) を考慮することなく、 S_1 から S_K 及びパラメータ ϕ_1 から ϕ_K を別々に最適化する事が可能な点である。これによって、コンポーネントの種類を最適化する際に、組み合わせ爆発を回避して最適化を実行する事が可能となる。

[0053] 最適性判定処理部 106 は、式 (4) で計算される最適化基準 A の収束を判定する。最適化基準 A が収束したと判定していない場合には、隠れ変数変分確率計算処理部 104 から最適性判定処理部 106 の処理を繰り返す。なお、式 (4) で計算される最適化基準 A の算出は、隠れ状態が独立ではないため $\sum_n q(z_n) \log q(z_n)$ の計算に指数時間の計算量を必要とするが、正規化定数記憶部 1042 に記憶されている正規化定数を利用して効率的に計算をする事が可能である。例えば隠れマルコフモデルの場合には、以下の式 (8) のように計算される。

[0054] [数8]

$$\sum_{n=1}^N \sum_{t=1}^{T_n} \log \zeta_n^{t(i)} + \mathbf{B}$$

式 (8)

[0055] 式 (8) に示す $\mathbf{B}$ は、式 (4) において下線を付した部分である。

[0056] 隠れ変数変分確率計算処理部 104 から最適性判定処理部 106 の処理を繰り返し、変分分布とモデルを更新する事で、適切なモデルを選択する事が可能となる。なお、この繰り返しによって最適化基準 A が単調に増加する事が

保証される。

- [0057] 最適化基準Aが収束している場合、隠れ状態数設定部102で設定された隠れ状態数Kに対して、隠れ変数変分確率計算処理部104から最適性判定処理部106のループ処理で算出される最適化基準Aと、その1つ前のループ処理で算出された最適化基準Aのうち、大きい方の最適化基準Aに対応するモデルを最適なモデルとして設定する。全ての候補値についてモデルの最適化が完了した場合には、処理がモデル推定結果出力装置108へ移り、まだ最適化の済んでいない候補が存在する場合には、隠れ状態数設定部102へ処理が移る。
- [0058] モデル推定結果出力装置108は、最適な隠れ状態数、観測確率の種類、パラメータ、変分分布などをモデル推定結果出力結果112として出力する。
- [0059] 隠れ状態数設定部102、初期化処理部103、隠れ変数変分確率計算処理部104（前向き確率計算処理部1041、正規化定数記憶部1042、後ろ向き確率計算処理部1043、前向き後ろ向き確率合算処理部1044）、モデル最適化処理部105、最適性判定処理部106、最適モデル選択処理部107およびモデル推定結果出力装置108は、例えば、隠れ変数モデル推定プログラムに従って動作するコンピュータのCPUによって実現される。CPUが、隠れ変数モデル推定プログラムを記録したコンピュータ読み取り可能な記録媒体から隠れ変数モデル推定プログラムを読み込み、上記の各要素として動作すればよい。
- [0060] また、隠れ状態数設定部102、初期化処理部103、隠れ変数変分確率計算処理部104、モデル最適化処理部105、最適性判定処理部106、最適モデル選択処理部107およびモデル推定結果出力装置108が別々のハードウェアで実現されていてもよい。また、隠れ変数変分確率計算処理部104においても、前向き確率計算処理部1041、正規化定数記憶部1042、後ろ向き確率計算処理部1043、前向き後ろ向き確率合算処理部1044が別々のハードウェアで実現されていてもよい。

- [0061] 図3は、本発明の第1の実施形態の処理経過の例を示すフローチャートである。データ入力装置101を介して入力データ111が入力される（ステップS100）。
- [0062] 次に、隠れ状態数設定部102は、入力された隠れ状態数の候補値のうち、まだ最適化の行なわれていない隠れ状態数を選択し設定する（ステップS101）。
- [0063] 次に、初期化処理部103は、設定された隠れ状態数に対して、推定のため、パラメータや隠れ変数変分確率の初期化処理を実施する（ステップS102）。
- [0064] 次に、隠れ変数変分確率計算処理部104は、隠れ変数の変分確率を計算する（ステップS103）。
- [0065] 次に、モデル最適化処理部105は、各隠れ状態に対して観測確率の種類とパラメータの推定を実施する（ステップS104）。この処理は、各隠れ状態のモデルの最適化であると言える。
- [0066] 次に、最適性判定処理部106は、最適化基準Aが収束したかを判定する。（例えば、S105）。最適性判定処理部106は、直近のステップS103～S105のループ処理において得られた最適化基準Aと、その1つ前のステップS103～S105のループ処理で得られた最適化基準Aの差を計算し、その差の絶対値が予め定められた閾値以下になっていれば、最適化基準Aが収束したと判定してよい。また、その差の絶対値が閾値より大きければ、最適性判定処理部106は、最適化基準Aが収束していないと判定してよい。なお、絶対値による最適化基準Aの差の算出は一例であり、例えば相対的な差によって収束を判定する等の方法を採用してもよい。
- [0067] ステップS105において、最適化基準Aが収束していないと判定された場合には、ステップS103～S105の処理を繰り返す。
- [0068] ステップS105において、最適化基準Aが収束したと判定された場合、最適モデル選択処理部107は、直近のステップS103～S105のループ処理において最適化されたモデル（隠れ状態数、観測確率の種類、パラメ

ータ)の最適化基準Aと、その1つ前のステップS103～S105のループ処理で最適化されたモデルの最適化基準Aとを比較し、値の大きい方の最適化基準Aに対応するモデルを、最適なモデルとして設定する(ステップS106)。

[0069] 次に、隠れ状態数設定部102は、推定されていない隠れ状態数の候補が残っているか否かを判定する(ステップS107)。隠れ状態数の候補が残っている場合には、ステップS102～S107の処理を繰り返す。一方、隠れ状態数の候補が残っていない場合には、モデル推定結果出力装置108がモデル推定結果を出力し(ステップS108)、処理を完了する。

[0070] 図4は、本実施形態における隠れ変数変分確率計算処理部104の動作(換言すれば、ステップS103の処理経過)を示すフローチャートである。

[0071] 前向き確率計算処理部1041は、n番目のデータのt番目の時刻に対する前向き確率 $f^{t(i)}_{nk}$ を算出する(ステップS111)。このとき、前向き確率計算処理部1041は、正規化定数も算出し、正規化定数記憶部1042に記憶させる(ステップS112)。

[0072] 続いて、前向き確率計算処理部1041は、すべての時刻tに対して前向き確率の算出が完了したかを確認し(ステップS113)、未完了の場合にはステップS111、S112の処理を繰り返す。完了した場合には、ステップS114の処理に移る。

[0073] 後ろ向き確率計算処理部1043は、n番目のデータのt番目の時刻に対する後ろ向き確率 $b^{t(i)}_{nk}$ を算出する(ステップS114)。そして、後ろ向き確率計算処理部1043は、すべての時刻tに対して後ろ向き確率の算出が完了したかを確認し(ステップS115)、未完了の場合にはステップS114の処理を繰り返す。完了した場合には、ステップS116の処理に移る。

[0074] 前向き後ろ向き確率合算処理部1044は、n番目のデータのすべての時刻に対して前向き確率、後ろ向き確率の合算処理を行い、変分分布の計算を行う(ステップS116)。

[0075] 続いて、前向き後ろ向き確率合算処理部1044は、nに関してすべてのデ

ータに対して変分分布の算出処理が完了しているかを確認する（ステップS 1 1 7）。未完了の場合には、ステップS 1 1 1以降の処理を繰り返し、完了した場合には、処理を終了する。

[0076] 隠れ状態数および観測確率の種類と共にモデルの候補数が指数的に増加する場合であっても、上記のような本発明の動作（特に、隠れ変数変分確率計算処理部104の動作）によって、高速にモデル選択を実現することができる。

[0077] また、非特許文献4に記載の技術と本発明を比較すると、既に説明したように、非特許文献4に記載の技術では、隠れ変数間の独立性が必要となり、系列依存性のある隠れ変数モデルに適用できなかった。これに対し本発明は、系列依存性を持つ多変量データの隠れ変数モデルを推定することができる。

[0078] 実施形態2.

図5は、本発明の隠れ変数モデル推定装置の第2の実施形態を示すブロック図である。第2の実施形態の隠れ変数モデル推定装置200は、第1の実施形態の隠れ変数モデル推定装置100（図1参照）と比較して、最適モデル選択処理部107を備えず、隠れ状態数選択処理部201を備える。

[0079] データ入力装置101、隠れ状態数設定部102、初期化処理部103、隠れ変数変分確率計算処理部104、モデル最適化処理部105、最適性判定処理部106およびモデル推定結果出力装置108に関しては、第1の実施形態と同様である。

[0080] また、第1の実施形態の隠れ変数モデル推定装置100は、隠れ状態数の候補に対してモデル最適化を行い、最適化基準Aを最大化するモデルを選択する。これに対して、第2の実施形態の隠れ変数モデル推定装置200では、隠れ変数変分確率計算処理部104の処理の後で、隠れ状態数選択処理部201が、小さくなった隠れ状態をモデルから除去する。

[0081] 具体的には、隠れ状態数選択処理部201は、隠れ変数変分確率計算処理部104で算出された $q(z^ntk)$ に対して、以下の式（9）の状態を満たす隠れ

状態を除去する。

[0082] [数9]

$$\sum_{n,t=1}^{N,T_n} q^{(i)}(z_{nk}^t) \leq \varepsilon$$

式 (9)

[0083] 式 (9) の右辺に示す ε は、入力データ 1 1 1 と同時に入力される閾値である。すなわち、隠れ状態数選択処理部 2 0 1 は、閾値 ε 以下である隠れ状態を除去する。

[0084] 式 (9) による隠れ状態の除去は、以下の理由によって正しく動く事が説明される。まず、式 (7) の前向き確率を観察すると、小さい隠れ状態（すなわち小さい δ_{tk} に対応する隠れ状態）に対する前向き確率は小さくなる。また、後ろ向き確率では、小さい隠れ状態は前の状態に対する寄与が小さい。従って、前向き確率と後ろ向き確率から算出される変分分布は、小さい隠れ状態に対する確率が繰り返し最適化を通して徐々に小さくなっていく（前の更新ステップにおいて小さい隠れ状態ほど、次の更新ステップでも小さくなりやすくなるため）。このような構成とする事で、隠れ変数モデル推定装置 1 0 0 のように複数の隠れ状態数の候補に対して最適化をする必要がなく、隠れ状態数、観測確率の種類とパラメータ、変分分布を同時に推定し、計算コストを抑える事が可能という利点がある。

[0085] 第 2 の実施形態において、隠れ状態数設定部 1 0 2、初期化処理部 1 0 3、隠れ変数変分確率計算処理部 1 0 4、隠れ状態数選択処理部 2 0 1、モデル最適化処理部 1 0 5、最適性判定処理部 1 0 6 およびモデル推定結果出力装置 1 0 8 は、例えば、隠れ変数モデル推定プログラムに従って動作するコンピュータの CPU によって実現される。CPU が、隠れ変数モデル推定プログラムを記録したコンピュータ読み取り可能な記録媒体から隠れ変数モデル推定プログラムを読み込み、上記の各要素として動作すればよい。また、

第2の実施形態におけるこれらの各要素が別々のハードウェアで実現されていてもよい。

実施例 1

[0086] 以下、本発明の第1の実施形態の応用例を、自動車のセンサデータに対する走行モード分析を例にして説明する。以下の例では理解を容易にするため1次元の例を説明するが、多次元でも同様に適用可能である。

[0087] 第1の実施形態の隠れ変数モデル推定装置を利用することで、自動車に設置された複数のセンサから取得される多次元時系列データに対して、例えば「走行モード」のように、時系列を、複数の異なる性質へ分解することができる。センサデータからの故障診断や異常挙動の検出を考えた場合に、走行モードによってセンサの振る舞いは大きく異なる。そのため、モードへ分解し分析する事が必要であり、これを自動化する事は非常に重要である。

[0088] 例えば、エンジン回転数を X とし、速度を Y とした、多項回帰出力の隠れマルコフモデルを考える。このとき、推定すべきモデルは、隠れ状態数、各隠れ状態に対する回帰次数 (S_k)、回帰パラメータ (ϕ_k)、初期確率 (α_k)、遷移確率 (β_k)、変分分布 (q) である。

[0089] まず、エンジン回転数と速度の時系列データとともに、隠れ状態数の候補値として $K=1$ から10までを隠れ変数モデル推定装置100に入力する。隠れ状態数設定部102は、 $K=1$ から10まで順に隠れ状態数を設定する。次に、初期化処理部103は、初期化処理として、 K 個の隠れ状態に対して、回帰次数及びその他のパラメータをランダムに設定する。次に、隠れ変数変分確率計算処理部104から最適性判定処理部106によってモデルの推定を行う。この処理を通じて、例えば、エンジン回転数が一定で速度が増加している状態（等加速）に対応する X から Y への0次多項式、エンジン回転数も速度も減少している状態（減速中）に対応する X から Y への1次多項式、エンジン回転数が急激に増加し速度が徐々に増加する状態（急加速）に対応する X から Y への2次多項式など、異なる走行状態が、異なる次数と係数を持つ回帰モデルとして自動的に分離される。さらに、最適モデル選択処理部107が、最もよい

隠れ状態数を自動選択するため、例えばドライバーに応じて異なる運転特性（モード）の数を、自動的に検出し、適切な数の走行モードに分離する事が可能である。

実施例 2

[0090] 以下、本発明の第2の実施形態の応用例を、診療ログ（レセプトデータ）からの疾病パタン分析を例にして説明する。例えば、心筋梗塞などは高血圧や糖尿病といった生活習慣病を事前に併発している事が多く、また生活習慣病は一度治ったとしても、再発する事が多い。そのような疾病のパタンを分析する事で、疾病のリスク低減への方策検討や、生活習慣指導へ活用する事が可能である。

[0091] 本例では、高血圧にかかっているかどうかの論理値（かかっていたら1、かかっていなければ0）を複数の疾病について並べた、多次元の論理値ベクトル時系列を入力データとする。推定するモデルは、多次元のベルヌーイ観測の隠れマルコフモデルとする。

[0092] まず、入力データとともに、隠れ状態数として $K_{\max}$ 、選択のしきい値 ε を入力する。隠れ状態の候補値は $K_{\max}$ に設定され、ベルヌーイ分布のパラメータがランダムに初期化される。次に、隠れ変数変分確率計算処理部104から最適性判定処理部106によってモデルの推定を行う。この処理を通じて、高血圧と糖尿病が合併しているパタン、高脂血症が治ったり再発したりを繰り返すパタン（薬による治療中）、生活習慣病がほとんど発生しないパタンなどに分離するとともに、典型的ではないパタンに対応する隠れ状態は小さくなり、隠れ状態数選択処理部201によって除去され、最終的な推定結果として典型的なパタンのみを抽出する事が可能である。

[0093] 図6は、本発明の隠れ変数モデル推定装置の概要を示すブロック図である。本発明の隠れ変数モデル推定装置は、変分確率計算部71と、モデル推定部72と、収束判定部73とを備える。

[0094] 変分確率計算部71（例えば、隠れ変数変分確率計算処理部104）は、周辺化対数尤度関数を完全変数に対する推定量に関してラプラス近似した近

似量の下界として定義される基準値（例えば、最適化基準 A）を最大化することによって変分確率を計算する。

[0095] モデル推定部 7 2（例えば、モデル最適化処理部 1 0 5）は、各隠れ状態に対して観測確率の種類とパラメータを推定することで最適な隠れ変数モデルを推定する。

[0096] 収束判定部 7 3（例えば、最適化判定処理部 1 0 6）は、変分確率計算部 7 1 が変分確率を計算する際に用いた基準値が収束したか否かを判定する。

[0097] また、変分確率計算部 7 1 が変分確率を計算し、モデル推定部 7 2 が最適な隠れ変数モデルを推定し、収束判定部 7 3 が、基準値が収束したか否かを判定するループ処理を繰り返し、基準値が収束したときに、当該基準値と、その 1 つ前のループ処理における基準値とを比較し、大きい方の基準値に対応する隠れ変数モデルを最適な隠れ変数モデルとして選択する最適モデル選択部（例えば、最適モデル選択処理部 1 0 7）を備える構成であってもよい。

[0098] また、変分確率計算部 7 1 が変分確率を計算し、モデル推定部 7 2 が最適な隠れ変数モデルを推定し、収束判定部 7 3 が、基準値が収束したか否かを判定するループ処理を繰り返し、変分確率計算部の計算結果に応じて、所定の条件を満たす隠れ状態を除去する隠れ状態除去部（例えば、隠れ状態数選択処理部 2 0 1）を備える構成であってもよい。

[0099] また、モデル推定部 7 2 が、隠れ変数モデルとして隠れマルコフモデルを推定する構成であってもよい。

[0100] この出願は、2012 年 5 月 31 に出願された米国仮出願 61/653,855 を基礎とする優先権を主張し、その開示の全てをここに取り込む。

[0101] 以上、実施形態を参照して本願発明を説明したが、本願発明は上記の実施形態に限定されるものではない。本願発明の構成や詳細には、本願発明のスコープ内で当業者が理解し得る様々な変更をすることができる。

産業上の利用の可能性

[0102] 本発明は、系列依存性を持つ多変量データの隠れ変数モデル推定装置に好

適に適用される。

符号の説明

- [0103] 1 0 1 データ入力装置
- 1 0 2 隠れ状態数設定部
- 1 0 3 初期化処理部
- 1 0 4 隠れ変数変分確率計算処理部
- 1 0 5 モデル最適化処理部
- 1 0 6 最適性判定処理部
- 1 0 7 最適モデル選択処理部
- 1 0 8 モデル推定結果出力装置
- 2 0 1 隠れ状態数選択処理部

請求の範囲

- [請求項1] 周辺化対数尤度関数を完全変数に対する推定量に関してラプラス近似した近似量の下界として定義される基準値を最大化することによって変分確率を計算する変分確率計算部と、
- 各隠れ状態に対して観測確率の種類とパラメータを推定することで最適な隠れ変数モデルを推定するモデル推定部と、
- 変分確率計算部が変分確率を計算する際に用いた基準値が収束したか否かを判定する収束判定部とを備える
- ことを特徴とする隠れ変数モデル推定装置。
- [請求項2] 変分確率計算部が変分確率を計算し、モデル推定部が最適な隠れ変数モデルを推定し、収束判定部が、基準値が収束したか否かを判定するループ処理を繰り返し、
- 基準値が収束したときに、当該基準値と、その1つ前のループ処理における基準値とを比較し、大きい方の基準値に対応する隠れ変数モデルを最適な隠れ変数モデルとして選択する最適モデル選択部を備える
- 請求項1に記載の隠れ変数モデル推定装置。
- [請求項3] 変分確率計算部が変分確率を計算し、モデル推定部が最適な隠れ変数モデルを推定し、収束判定部が、基準値が収束したか否かを判定するループ処理を繰り返し、
- 変分確率計算部の計算結果に応じて、所定の条件を満たす隠れ状態を除去する隠れ状態除去部を備える
- 請求項1に記載の隠れ変数モデル推定装置。
- [請求項4] モデル推定部は、隠れ変数モデルとして隠れマルコフモデルを推定する請求項1から請求項3のうちのいずれか1項に記載の隠れ変数モデル推定装置。
- [請求項5] 周辺化対数尤度関数を完全変数に対する推定量に関してラプラス近似した近似量の下界として定義される基準値を最大化することによっ

て変分確率を計算し、

各隠れ状態に対して観測確率の種類とパラメータを推定することで最適な隠れ変数モデルを推定し、

変分確率を計算する際に用いた基準値が収束したか否かを判定することを特徴とする隠れ変数モデル推定方法。

[請求項6] 変分確率を計算し、最適な隠れ変数モデルを推定し、基準値が収束したか否かを判定するループ処理を繰り返し、

基準値が収束したときに、当該基準値と、その1つ前のループ処理における基準値とを比較し、大きい方の基準値に対応する隠れ変数モデルを最適な隠れ変数モデルとして選択する

請求項5に記載の隠れ変数モデル推定方法。

[請求項7] 変分確率を計算し、最適な隠れ変数モデルを推定し、基準値が収束したか否かを判定するループ処理を繰り返し、

変分確率の計算結果に応じて、所定の条件を満たす隠れ状態を除去する

請求項5に記載の隠れ変数モデル推定方法。

[請求項8] 隠れ変数モデルとして隠れマルコフモデルを推定する請求項5から請求項7のうちのいずれか1項に記載の隠れ変数モデル推定方法。

[請求項9] コンピュータに、

周辺化対数尤度関数を完全変数に対する推定量に関してラプラス近似した近似量の下界として定義される基準値を最大化することによって変分確率を計算する変分確率計算処理、

各隠れ状態に対して観測確率の種類とパラメータを推定することで最適な隠れ変数モデルを推定するモデル推定処理、および、

変分確率計算処理で変分確率を計算する際に用いた基準値が収束したか否かを判定する収束判定処理

を実行させるための隠れ変数モデル推定プログラム。

[請求項10] コンピュータに、

変分確率計算処理、モデル推定処理、収束判定処理のループ処理を繰り返して実行させ、

基準値が収束したときに、当該基準値と、その1つ前のループ処理における基準値とを比較し、大きい方の基準値に対応する隠れ変数モデルを最適な隠れ変数モデルとして選択する最適モデル選択処理を実行させる請求項9に記載の隠れ変数モデル推定プログラム。

[請求項11]

コンピュータに、

変分確率計算処理、モデル推定処理、収束判定処理のループ処理を繰り返して実行させ、

変分確率計算処理の計算結果に応じて、所定の条件を満たす隠れ状態を除去する隠れ状態除去処理

を実行させる請求項9に記載の隠れ変数モデル推定プログラム。

[請求項12]

コンピュータに、

モデル推定処理で、隠れ変数モデルとして隠れマルコフモデルを推定させる

請求項9から請求項11のうちのいずれか1項に記載の隠れ変数モデル推定プログラム。