

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2019-212159

(P2019-212159A)

(43) 公開日 令和1年12月12日(2019.12.12)

(51) Int.Cl.	F I	テーマコード (参考)
G06N 7/08 (2006.01)	G06N 7/08	5 L049
G06Q 10/04 (2012.01)	G06Q 10/04	

審査請求 未請求 請求項の数 10 O L (全 15 頁)

<p>(21) 出願番号 特願2018-109495 (P2018-109495)</p> <p>(22) 出願日 平成30年6月7日(2018.6.7)</p> <p>(出願人による申告)平成27~30年度国立研究開発法人科学技術振興機構「戦略的創造研究推進事業(多様な環境に自律順応できる水分ストレス高精度予測基盤技術の確立)」委託研究、産業技術力強化法第19条の適用を受ける特許出願</p>	<p>(71) 出願人 304023318 国立大学法人静岡大学 静岡県静岡市駿河区大谷836</p> <p>(74) 代理人 100088155 弁理士 長谷川 芳樹</p> <p>(74) 代理人 100124800 弁理士 諏澤 勇司</p> <p>(72) 発明者 峰野 博史 静岡県浜松市中区城北3丁目5-1 国立大学法人静岡大学内</p> <p>(72) 発明者 柴田 瞬 静岡県浜松市中区城北3丁目5-1 国立大学法人静岡大学内</p>
--	---

最終頁に続く

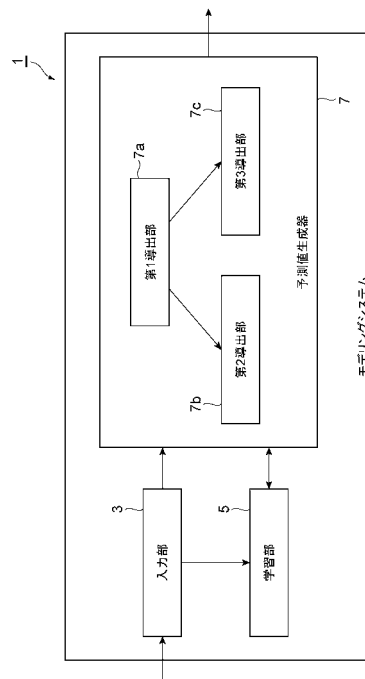
(54) 【発明の名称】 モデリングシステム

(57) 【要約】

【課題】モデリング対象の現象に関する実測値に欠損値が生じやすい場合であっても、モデリングの精度を維持すること。

【解決手段】一形態にかかるモデリングシステム1は、複数の連続した時刻における直接観測できない変数である潜在変数 z_t を、外部から与えられた外部変数 u_t と、直前の時刻の潜在変数 z_{t-1} とから、非線形関数を用いて順次導出する第1導出部7aと、複数の連続した時刻における直接観測できる変数である観測変数 x_t を、同一の時刻の潜在変数 z_t から、非線形関数を用いて順次導出する第2導出部7bと、複数の連続した時刻における直接観測できる変数である観測変数 y_t を、同一の時刻の潜在変数 z_t から、非線形関数を用いて順次導出する第3導出部7cと、を備える。

【選択図】図1



【特許請求の範囲】

【請求項 1】

複数の連続した時刻における直接観測できない変数である潜在変数を、直前の時刻の前記潜在変数から、非線形関数を用いて順次導出する第 1 の導出部と、

前記複数の連続した時刻における直接観測できる変数である第 1 の観測変数を、同一の時刻の前記潜在変数から、非線形関数を用いて順次導出する第 2 の導出部と、

前記複数の連続した時刻における直接観測できる変数である第 2 の観測変数を、同一の時刻の前記潜在変数から、非線形関数を用いて順次導出する第 3 の導出部と、
を備えるモデリングシステム。

【請求項 2】

前記第 1 の導出部は、外部から与えられた外部変数をさらに用いて前記潜在変数を順次導出する、

請求項 1 記載のモデリングシステム。

【請求項 3】

前記第 2 の導出部は、外部から与えられた外部変数をさらに用いて前記第 1 の観測変数を順次導出し、

前記第 3 の導出部は、外部から与えられた外部変数をさらに用いて前記第 2 の観測変数を順次導出する、

請求項 1 又は 2 に記載のモデリングシステム。

【請求項 4】

前記潜在変数、前記第 1 の観測変数、及び前記第 2 の観測変数を、確率分布として導出する、

請求項 1 ~ 3 のいずれか 1 項に記載のモデリングシステム。

【請求項 5】

前記潜在変数、前記第 1 の観測変数、及び前記第 2 の観測変数を、正規分布として導出する、

請求項 4 記載のモデリングシステム。

【請求項 6】

前記潜在変数、前記第 1 の観測変数、及び前記第 2 の観測変数を、離散分布として導出する、

請求項 4 記載のモデリングシステム。

【請求項 7】

前記非線形関数は、ニューラルネットワークを含んでいる、

請求項 1 ~ 6 のいずれか 1 項に記載のモデリングシステム。

【請求項 8】

前記非線形関数のパラメータを、前記第 1 の観測変数の実測値あるいは前記第 2 の観測変数の実測値を用いた機械学習により、予め最適化する学習部をさらに備える、

請求項 1 ~ 7 のいずれか 1 項に記載のモデリングシステム。

【請求項 9】

前記学習部は、前記第 1 の観測変数の実測値及び前記第 2 の観測変数の実測値を基にして、前記非線形関数に関する E L B O 値を計算し、前記 E L B O 値を目的関数とした機械学習により前記パラメータを最適化する、

請求項 8 に記載のモデリングシステム。

【請求項 10】

前記学習部は、前記第 2 の観測変数の実測値が存在する時刻と前記第 2 の観測変数の実測値が存在しない時刻とで異なる演算式を用いて前記 E L B O 値を計算し、複数の時刻で合計した前記 E L B O 値を前記目的関数として用いる、

請求項 9 に記載のモデリングシステム。

【発明の詳細な説明】

【技術分野】

10

20

30

40

50

【0001】

本発明は、現象の時間的変化をモデリングするモデリングシステムに関する。

【背景技術】

【0002】

近年、様々な現象の時間的変化をモデル化する技術が浸透し始めている。下記特許文献1には、電力系統の状態量を推定する装置が開示され、この装置は、電力量の出現分布情報を受信する電力系統計測部、電力系統の状態推定を非線形カルマンフィルタを用いて計算するシミュレータ部、電力系統の計測データとシミュレータデータとを用いて確率的に尤もらしい状態値を算出する最尤状態算出部、及び、計測データとシミュレータデータとの差が小さくなるようにシミュレータ部のパラメータを逐次修正するパラメータ補正部からなる。

10

【先行技術文献】

【特許文献】

【0003】

【特許文献1】国際公開2015/079554号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

上述した従来装置では、モデリング対象の現象に関する実測値が時間的に連続して得られず欠損値が生じやすい場合に、シミュレータ部のパラメータを予め適切に設定しておくことが困難な傾向にある。そのため、実測値に欠損値が生じやすい場合にモデリングの精度が低下する傾向にあった。

20

【0005】

本発明の一側面は、上記課題に鑑みて為されたものであり、モデリング対象の現象に関する実測値に欠損値が生じやすい場合であっても、モデリングの精度を維持することが可能なモデリングシステムを提供することを目的とする。

【課題を解決するための手段】

【0006】

上記課題を解決するため、本発明の一形態にかかるモデリングシステムは、複数の連続した時刻における直接観測できない変数である潜在変数を、直前の時刻の潜在変数から、非線形関数を用いて順次導出する第1の導出部と、複数の連続した時刻における直接観測できる変数である第1の観測変数を、同一の時刻の潜在変数から、非線形関数を用いて順次導出する第2の導出部と、複数の連続した時刻における直接観測できる変数である第2の観測変数を、同一の時刻の潜在変数から、非線形関数を用いて順次導出する第3の導出部と、を備える。

30

【0007】

上記形態のモデリングシステムによれば、潜在変数の時間的な変化と、その変化に基づいた第1及び第2の観測変数の時間的な変化をモデリングすることにより、特定の現象における観測値のシミュレーションに利用することができる。特に、2つの観測変数を含んでモデリングされているので、2つの観測変数のうちの片方の観測変数の実測値が欠損しやすい状況下であっても、もう片方の実測値を用いてモデルのパラメータを推定できる。その結果、モデリングの精度を維持することができる。

40

【0008】

ここで、前記第1の導出部は、外部から与えられた外部変数をさらに用いて前記潜在変数を順次導出してもよく、第2の導出部は、外部から与えられた外部変数をさらに用いて前記第1の観測変数を順次導出してもよく、前記第3の導出部は、外部から与えられた外部変数をさらに用いて前記第2の観測変数を順次導出してもよい。こうすれば、外部の環境を考慮に入れてモデリングの精度をより高めることができる。

【0009】

また、潜在変数、第1の観測変数、及び第2の観測変数を、確率分布として導出する、

50

こととしてもよく、特に、正規分布として導出してもよい。かかる構成を備えれば、生物の連続的な状態等の自然現象を適切にモデリングすることができる。

【0010】

また、潜在変数、第1の観測変数、及び第2の観測変数を、確率分布として導出する、こととしてもよく、特に、離散分布として導出してもよい。かかる構成を備えれば、生物の離散的な状態等の自然現象を適切にモデリングすることができる。

【0011】

また、非線形関数は、ニューラルネットワークを含んでいてもよく、任意の非線形関数を選択してもよい。この場合、この場合、生物成長等の非線形な現象を適切にモデリングすることができる。

【0012】

また、非線形関数のパラメータを、第1の観測変数の実測値あるいは第2の観測変数の実測値を用いた機械学習により、予め最適化する学習部をさらに備える、こととしてもよい。かかる構成を採れば、第1及び第2の実測値の実測値を用いてモデルに利用される非線形関数のパラメータを最適化することができる。特に、2つの観測変数のうちの片方の観測変数の実測値が欠損しやすい状況下であっても、もう片方の実測値を用いてパラメータを効率的に最適化することができる。

【0013】

また、学習部は、第1の観測変数の実測値及び第2の観測変数の実測値を基にして、非線形関数に関するELBO値を計算し、ELBO値を目的関数とした機械学習によりパラメータを最適化する、こととしてもよい。

【0014】

また、学習部は、第2の観測変数の実測値が存在する時刻と第2の観測変数の実測値が存在しない時刻とで異なる演算式を用いてELBO値を計算し、複数の時刻で合計したELBO値を目的関数として用いる、こととしてもよい。これにより、2つの観測変数のうちの片方の観測変数の実測値が欠損しやすい状況下であっても、もう片方の実測値を用いてパラメータを効率的に最適化することができる。

【発明の効果】

【0015】

本発明によれば、モデリング対象の現象に関する実測値に欠損値が生じやすい場合であっても、モデリングの精度を維持することができる。

【図面の簡単な説明】

【0016】

【図1】実施形態にかかるモデリングシステムの概略構成を示す図である。

【図2】図1のモデリングシステムを構成するコンピュータのハードウェア構成を示す図である。

【図3】図1の予測値生成器7が用いる学習モデルのデータ系列の生成過程を説明する概念図である。

【図4】図1の予測値生成器7が用いる学習モデルにおける各変数の設定内容の例を示す図表である。

【図5】図1の予測値生成器7が用いる学習モデルのデータ系列の生成過程の他の例を説明する概念図である。

【図6】図1の予測値生成器7が用いる学習モデルのデータ系列の生成過程の他の例を説明する概念図である。

【図7】本実施形態における算出糖度の誤差の評価値MAE、RMSE、RAE、RSEを示すグラフである。

【図8】本実施形態における算出糖度の時間変化を示すグラフである。

【図9】本実施形態における算出糖度の時間変化を示すグラフである。

【図10】本実施形態における算出糖度の時間変化を示すグラフである。

【発明を実施するための形態】

10

20

30

40

50

【 0 0 1 7 】

以下、図面を参照しつつ本発明に係るモデリングシステムの好適な実施形態について詳細に説明する。なお、図面の説明においては、同一又は相当部分には同一符号を付し、重複する説明を省略する。

【 0 0 1 8 】

まず、図 1 ~ 3 を用いて、本発明の一実施形態にかかるモデリングシステム 1 の機能および構成を説明する。図 1 に示すモデリングシステム 1 は、事前に様々な現象を取り巻く環境をモデル化し、それによって生成された学習モデルを用いて現象の時間変化を探索するコンピュータシステムである。より具体的には、学習モデルを基に現象の時間変化を高精度に推定するモデリングシステムである。モデリングシステム 1 の探索対象の現象としては、例えば、トマト等の植物の生育状態とし、栽培環境および生育状態に関する観測値を訓練データとして用いて事前に学習モデルを最適化し、その学習モデルを用いて現在の生育状態を推定し定量化できる。

10

【 0 0 1 9 】

なお、モデリングシステム 1 の探索対象の現象としては、トマト等の植物の生育状態が挙げられるが、これには限定されない。例えば、モデリングシステム 1 は、動物または食品等のその他の有体物の状態、天候、病気、または漁獲等の自然現象の状態、通信ネットワーク、交通、ライフライン、株価、または需給等の社会現象の状態を探索対象としてもよい。また、本明細書でいう「系列」とは、複数の時点における現象の様々な観測値あるいは予測値を示す数値の系列であり、ある規則に基づいて複数の時点において観測することにより得られる数値の系列、学習モデルに従ったある計算式によって複数の時点に関して計算された数値の系列である。

20

【 0 0 2 0 】

図 1 に示すように、モデリングシステム 1 は、入力部 3、学習部 5、及び予測値生成器 7 を含んで構成されている。本実施形態では、モデリングシステム 1 は 1 つの装置によって構成されているが、データ送受信機能を有する複数の装置によって構成されていてもよい。

【 0 0 2 1 】

モデリングシステム 1 を構成するコンピュータ 100 の一般的なハードウェア構成を図 2 に示す。コンピュータ 100 は、オペレーティングシステムやアプリケーション・プログラムなどを実行する演算装置である CPU (プロセッサ) 101 と、ROM 及び RAM で構成される主記憶部 102 と、ハードディスクやフラッシュメモリなどで構成される補助記憶部 103 と、ネットワークカードあるいは無線通信モジュールで構成される通信制御部 104 と、キーボードやマウスなどの入力装置 105 と、ディスプレイやプリンタなどの出力装置 106 とを備える。当然ながら、搭載されるハードウェアモジュールはコンピュータ 100 の種類により異なる。例えば、据置型の PC およびワークステーションは入力装置および出力装置としてキーボード、マウス、およびディスプレイを備えることが多いが、スマートフォンではタッチパネルが入力装置および出力装置として機能することが多い。また、CPU 101 に GPU が付加されていてもよい。

30

【 0 0 2 2 】

後述するモデリングシステム 1 の各機能要素は、CPU 101 または主記憶部 102 の上に所定のソフトウェアを読み込ませ、CPU 101 の制御の下で通信制御部 104 や入力装置 105、出力装置 106 などを動作させ、主記憶部 102 または補助記憶部 103 におけるデータの読み出し及び書き込みを行うことで実現される。処理に必要なデータやデータベースは主記憶部 102 または補助記憶部 103 内に格納される。

40

【 0 0 2 3 】

図 1 に戻って、入力部 3 は、学習モデルのモデル化 (以下、単に「モデル化」ともいう。) のために用いられる訓練データ系列、及び、モデル化した学習モデルに従って現象の時間変化を探索する処理 (以下、単に「探索処理」ともいう。) において用いられる外部変数データ系列を外部から受け付ける。例えば、モデリングシステム 1 が対象とする現象

50

が植物の生育状態である場合には、訓練データ系列に含まれる観測値（実測値）として、糖度、茎径、温度、光量、飽差、開花後の経過日数、灌水後の経過時間、積算温度、二酸化炭素濃度等が挙げられる。また、外部変数データ系列に含まれる数値として、温度、光量、飽差、開花後の経過日数、灌水後の経過時間、積算温度、二酸化炭素濃度、対象花房等が想定される。入力部3は、訓練データ系列を、温度センサ、光センサ、カメラ等の外部センサ装置からの信号を基に生成したデータ系列として受け付けてもよいし、他のコンピュータシステム等の外部装置を介して受け付けてもよい。入力部3は、外部変数データ系列を、ユーザ入力に基づいて受け付けてもよいし、内部のプログラム等によって生成されたものとして受け付けてもよいし、他のコンピュータシステム等の外部装置から受け付けてもよい。

10

【0024】

学習部5は、入力部3によって受け付けられた訓練データ系列を用いて予測値生成器7に対応する学習モデルのパラメータを予め最適化する。学習部5による学習モデルのパラメータの最適化の処理の詳細は後述する。

【0025】

予測値生成器7は、学習部5によってパラメータが最適化された学習モデルを用いて探索処理を実行する機能部分である。予測値生成器7は、入力部3によって受け付けられた外部変数データ系列を用いて探索処理を実行し、適切な生育状態（例えば、経時糖度等）を推定し定量化して出力する。

20

【0026】

ここで、予測値生成器7が用いる学習モデルにおいてデータ系列を生成する過程を説明する。図3は、予測値生成器7が用いる学習モデルにおけるデータ系列の生成過程を説明する概念図である。このように、この学習モデルは、状態空間モデル（SSM：State Space Model）をベースとしており、直接観測できない状態モデルと状態モデルから生成される観測モデルの2系列のデータ系列を有する。状態モデルは、時刻 t の潜在変数 z_t と、その1つ前の時刻 $t-1$ の潜在変数 z_{t-1} と、その1つ後の時刻 $t+1$ の潜在変数 z_{t+1} とを含み、時刻 t の潜在変数 z_t は前の潜在変数 z_{t-1} と、 z_t と同じ時刻の外部変数 u_t を基に生成され、それ以降の潜在変数 z_{t+1} は同様にして順次生成される。この潜在変数 z_t は、直接観測できない現象に関する数値であり、例えば、植物の状態を示す。観測モデルは、時刻 t の2つの直接観測できる変数である観測変数 x_t, y_t と、その1つ前の時刻 $t-1$ の観測変数 x_{t-1}, y_{t-1} と、その1つ後の時刻 $t+1$ の観測変数 x_{t+1}, y_{t+1} とを含む。時刻 t の観測変数 y_t は潜在変数 z_t と、同じ時刻 t の外部変数 r_t を基に生成され、それ以降の観測変数 y_{t+1} は同様にして順次生成される。時刻 t の観測変数 x_t は潜在変数 z_t と、同じ時刻 t の外部変数 s_t を基に生成され、それ以降の観測変数 x_{t+1} は同様にして順次生成される。すなわち、観測変数 y_t と観測変数 x_t とは同じ潜在変数 z_t を基に生成される。外部変数 u_t, r_t, s_t は、それぞれ、潜在変数 z_t 、観測変数 y_t 、観測変数 x_t に作用する変数である。このように、この学習モデルにおいては、各変数 z_t, y_t, x_t で異なる外部変数を持つことにより、モデル化時に各変数の生成過程に適したデータを設定することで適切に現象の予測を行うことが可能となる。

30

40

【0027】

学習モデルの対象を植物の状態とする場合には、各変数には、図4に示すような数値が設定される。潜在変数 z_t には植物の状態が設定され、観測変数 x_t, y_t にはそれぞれ、植物の糖度、および植物の茎径が設定される。この観測変数 x_t としては、モデル化を行う際に系列のデータを連続して得ることが不可能である、あるいは、系列のデータを連続して得るためのコストが高い数値が設定され、この観測変数 y_t としては、モデル化を行う際に系列のデータを連続して得ることが可能である、あるいは、系列のデータを連続して得るためのコストが低い数値が設定される。また、外部変数 u_t としては、温度、光量、飽差、開花後の経過日数、及び積算温度等が設定され、外部変数 r_t, s_t としては、CO₂濃度、光量、対象花房（第一花房であるか第二花房であるか）等が設定される。

50

【 0 0 2 8 】

図 1 に戻って、予測値生成器 7 は、探索処理を実行する第 1 導出部 7 a、第 2 導出部 7 b、及び第 3 導出部 7 c を有している。第 1 導出部 7 a は、上記の学習モデルを前提にして、連続したそれぞれの時刻 t の潜在変数 z_t を、外部から与えられた時刻 t の外部変数 u_t と、直前の時刻 $t - 1$ の潜在変数 z_{t-1} とから、非線形関数を用いて順次導出する。例えば、非線形関数としては DNN (Deep Neural Network) を用いて最適化された非線形関数を用いる。ここでは、下記式 (1) ;

$$p_{\theta}(z_t|z_{t-1}, u_t) = N(z; \mu_z(z_{t-1}, u_t), \sigma_z(z_{t-1}, u_t)) \quad (1)$$

10

によって表現されるマルコフ性を仮定した生成過程、すなわち、パラメータ θ によって表現された、潜在変数 z_{t-1} 及び外部変数 u_t から平均 μ_z 及び標準偏差 σ_z の正規分布 (確率分布) $N(z)$ を生成する生成過程 P が想定される。より詳細には、第 1 導出部 7 a は、下記式 (2) ;

【 数 2 】

$$\mu_z = NN_z(z_{t-1}, u_t), \log \sigma_z = NN_{\sigma_z}(z_{t-1}, u_t) \quad (2)$$

20

を用いて、潜在変数 z_t を、平均 μ_z 及び標準偏差 σ_z の正規分布として導出する。上記式 (2) 中、 NN_z は、DNN を用いて表現された非線形関数である。

【 0 0 2 9 】

第 2 導出部 7 b は、上記の学習モデルを前提にして、連続したそれぞれの時刻 t の観測変数 x_t を、外部から与えられた時刻 t の外部変数 s_t と、時刻 t の潜在変数 z_t とから、非線形関数を用いて順次導出する。例えば、非線形関数としては DNN を用いて最適化された非線形関数を用いる。ここでは、下記式 (3) ;

【 数 3 】

$$p_{\theta}(x_t|z_t, s_t) = N(x; \mu_x(z_t, s_t), \sigma_x(z_t, s_t)) \quad (3)$$

30

によって表現されるマルコフ性を仮定した生成過程、すなわち、パラメータ θ によって表現された、潜在変数 z_t 及び外部変数 s_t から平均 μ_x 及び標準偏差 σ_x の正規分布 (確率分布) $N(x)$ を生成する生成過程 P が想定される。より詳細には、第 2 導出部 7 b は、下記式 (4) ;

【 数 4 】

$$\mu_x = NN_x(z_t, s_t), \log \sigma_x = NN_{\sigma_x}(z_t, s_t) \quad (4)$$

40

を用いて、観測変数 x_t を、平均 μ_x 及び標準偏差 σ_x の正規分布として導出する。上記式 (4) 中、 NN_x は、DNN を用いて表現された非線形関数である。

【 0 0 3 0 】

第 3 導出部 7 c は、上記の学習モデルを前提にして、連続したそれぞれの時刻 t の観測変数 y_t を、外部から与えられた時刻 t の外部変数 r_t と、時刻 t の潜在変数 z_t とから、非線形関数を用いて順次導出する。例えば、非線形関数としては DNN を用いて最適化された非線形関数を用いる。ここでは、下記式 (5) ;

【数5】

$$p_{\theta}(y_t|z_t, r_t) = N(y; \mu_y(z_t, r_t), \sigma_y(z_t, r_t)) \quad (5)$$

によって表現されるマルコフ性を仮定した生成過程、すなわち、パラメータによって表現された、潜在変数 z_t 及び外部変数 r_t から平均 μ_y 及び標準偏差 σ_y の正規分布（確率分布） $N(y)$ を生成する生成過程 P が想定される。より詳細には、第3導出部7cは、下記式(6)；

【数6】

$$\mu_y = NN_y(z_t, r_t), \log \sigma_y = NN_y(z_t, r_t) \quad (6)$$

を用いて、観測変数 y_t を、平均 μ_y 及び標準偏差 σ_y の正規分布として導出する。上記式(6)中、 NN_y は、DNNを用いて表現された非線形関数である。

【0031】

本実施形態のモデリングシステム1では、予測値生成器7によって様々な外部変数データ系列を用いた探索処理が実行され適切な生育状態をモデル化する。そのモデルを用いてモデルベース強化学習システムを構築すれば、目標の生育状態となるように最適な制御内容（例えば、灌水のタイミング等）を獲得することができる。

【0032】

次に、学習部5による機械学習によるモデル化の処理の詳細について説明する。

【0033】

学習部5は、入力部3によって受け付けられた訓練データ系列を用いて、DNNで表現された非線形関数 NN_z , NN_x , NN_y のパラメータを最適化する。モデル化に用いられる訓練データ系列は、各時刻 $t_1 \sim t_m$ (m は自然数) において欠損値が含まれない外部変数 u_t , s_t , r_t 及び観測変数 x_t , y_t のラベルデータ系列 D_l と、各時刻 $t_1 \sim t_n$ (n は自然数) において観測変数 x_t が欠損している外部変数 u_t , s_t , r_t 及び観測変数 y_t のアンラベルデータ系列 D_u とを含んでいる。そして、学習部5は、ラベルデータ系列 D_l とアンラベルデータ系列 D_u とのそれぞれを対象に、変分ベイズ法の枠組みを勾配法で解く最適化手法（「Kingma, D.P., & Welling, M.: Auto-Encoding Variational Bayes, arXiv preprint, arXiv: 1312.6114(2013)」を参照）を用いて非線形関数 NN_z , NN_x , NN_y のパラメータを最適化する。

【0034】

詳細には、学習部5は、ラベルデータ系列 D_l に対し近似事後分布を用い、パラメータ θ を最適化するELBO (Evidence Lower Bound) 値をデータの周辺尤度から、下記式(7)を用いて計算する。

【数7】

$$\begin{aligned} -L^l(x, y; \theta, \varphi) = & \sum_{t=1}^T E_{q_{\varphi}(z_t)} [\log p_{\theta}(y_t|z_t) + \log p_{\theta}(x_t|y_t, z_t)] - \beta D_{KL}(q_{\varphi}(z_1)||p_{\theta}(z_1)) - \\ & \sum_{t=2}^T E_{q_{\varphi}(z_{t-1})} [\beta D_{KL}(q_{\varphi}(z_t)||p_{\theta}(z_t|z_{t-1}))] \end{aligned} \quad (7)$$

上記(7)中、 $\log p(y_t|z_t)$ は、訓練データ系列の観測変数 y_t に関する潜在変数 z_t の対数尤度であり、 $\log p(x_t|y_t, z_t)$ は、訓練データ系列の観測変数 x_t に関する潜在変数 z_t の対数尤度であり、時刻 t の $q(z_t)$ は、1つ前の時刻 $t-1$ の潜在変数 z_{t-1} と、同じ時刻 t の観測変数 x_t , y_t から潜在変数 z_t を導き出すパラメータ φ を有する認識モデル（近似事後確率分布）の式である。また、 D_K

10

20

30

40

50

$D_{\perp}(A||B)$ は、 A と B との間の Kullback-Leibler 情報量を示し、右辺の第 1 項及び第 3 項はそれぞれ、観測変数 x_t に関する潜在変数 z_t の対数尤度の期待値と観測変数 y_t に関する潜在変数 z_t の対数尤度の期待値との和、及び潜在変数 z_t の近似事後確率分布と潜在変数 z_t の事前確率分布との近似度を示す Kullback-Leibler 情報量の期待値を示している。

【0035】

加えて、学習部 5 は、アンラベルデータ系列 D_u に対し近似事後分布を用い、パラメータ θ, φ を最適化する ELBO 値を、上記式 (7) とは異なる下記式 (8) を用いて計算する。

【数 8】

$$-L^u(x, y; \theta, \varphi) = \sum_{t=1}^T E_{q_{\varphi}(x_t|y_t)}[-L^l(x, y; \theta, \varphi) + H[q_{\varphi}(x_t|y_t)]] \quad (8)$$

10

上記 (8) 中、 $q_{\varphi}(x_t|y_t)$ は、時刻 t の観測変数 y_t から同じ時刻 t の観測変数 x_t を導き出すパラメータ φ を有する認識モデル (近似事後確率分布) の式であり、 $H(A)$ は、 A のエントロピーを示す。つまり、ELBO 値 L^u は、観測変数 y_t から導き出される観測変数 x_t の近似事後確率分布における ELBO 値 L^l の期待値と近似事後確率分布 $q_{\varphi}(x_t|y_t)$ の期待値の和を示している。

【0036】

さらに、学習部 5 は、ラベルデータ系列 D_l に対し計算した ELBO 値 L^l と、アンラベルデータ系列 D_u に対し計算した ELBO 値 L^u とを加算することにより、下記式 (9) ;

【数 9】

$$J = \sum_{D_l} L^l(x, y; \theta, \varphi) + \sum_{D_u} L^u(x, y; \theta, \varphi) + \alpha E_{D_u}[-\log q_{\varphi}(x_t|y_t)] \quad (9)$$

20

を用いて目的関数 J を計算する。上記式 (9) 中、右辺の第 3 項は観測変数 x_t の周辺尤度の期待値である。そして、学習部 5 は、目的関数 J を最大化するように学習モデルのパラメータ θ, φ の最適化することによってモデル化の処理を実行する。なお、ELBO 値 L^l, L^u がパラメータ θ, φ に関して微分可能なため、目的関数もパラメータ θ, φ に関して微分可能である。学習部 5 は、逆伝搬を介した確率的勾配降下法 (Stochastic Gradient Descent) によってパラメータ θ, φ を最適化することができる。

30

【0037】

以上説明した実施形態に係るモデリングシステム 1 によれば、外部変数 u_t に基づいた潜在変数 z_t の時間的な変化と、その変化に基づいた観測変数 x_t, y_t の時間的な変化をモデリングすることにより、特定の現象における観測値 x_t, y_t のシミュレーションに利用することができる。特に、2つの観測変数 x_t, y_t を含んでモデリングされているので、2つの観測変数 x_t, y_t のうちの片方の観測変数 x_t の実測値が欠損しやすい状況下であっても、もう片方の実測値 y_t を用いて学習モデルのパラメータを機械学習によって推定できる。その結果、モデリングの精度を維持することができる。

40

【0038】

本実施形態では、潜在変数 z_t 、観測変数 x_t, y_t を正規分布として導出している。このようにすれば、生物の連続的な状態等の自然現象を適切にモデリングすることができる。

【0039】

また、学習モデルで用いられる非線形関数はニューラルネットワークを含んで表現されているので、生物成長等の非線形な現象を適切にモデリングすることができる。

【0040】

50

また、モデリングシステム 1 には、非線形関数のパラメータを、観測変数 x_t, y_t の実測値を用いた機械学習により、予め最適化する学習部 5 を備えている。このような構成により、2 つの観測変数 x_t, y_t のうちの片方の観測変数 x_t の実測値が欠損しやすい状況下であっても、もう片方の実測値 y_t を用いてパラメータを効率的に最適化することができる。

【0041】

さらに、学習部 5 は、観測変数 x_t の実測値が存在する時刻に対応するラベルデータ系列 D_L と観測変数 x_t の実測値が存在しない時刻に対応するアンラベルデータ系列 D_U とで異なる演算式を用いて E L B O 値 L^L, L^U を計算し、それらを合計した値を目的関数として用いている。これにより、2 つの観測変数 x_t, y_t のうちの片方の観測変数 x_t の実測値が欠損しやすい状況下であっても、もう片方の実測値 y_t を用いてパラメータを効率的に最適化することができる。

10

【0042】

なお、本発明は、上述した実施形態の態様に限定されるものではない。

【0043】

例えば、モデリングシステム 1 の予測値生成器 7 が用いる学習モデルは、様々なデータ系列の生成過程を反映したものが採用可能である。

【0044】

上記実施形態のモデリングシステム 1 は、図 5 に示すような生成過程を有する学習モデルに対応したモデル化処理及び探索処理を実行してもよい。図 5 に示す生成過程は、図 3 に示す生成過程と比較して、状態モデルにおいて、潜在変数 d_t を離散分布として導出する点異なる。モデリングシステム 1 は、上述した実施形態と同様にして、学習部 5 がモデル化処理を実行し、予測値生成器 7 が探索処理を実行することができる。このような学習モデルに対応した処理を実行することで、生物の離散的な状態（例えば、生育段階）等の自然現象を適切にモデリングすることができる。

20

【0045】

また、上記実施形態のモデリングシステム 1 は、図 6 に示すような生成過程を有する学習モデルに対応したモデル化処理及び探索処理を実行してもよい。図 5 に示す生成過程は、図 3 に示す生成過程と比較して、さらにもう 1 つの状態モデルが追加され、追加された状態モデルにおいて、潜在変数 d_t を離散分布として導出する点異なる。追加された状態モデルでは、時刻 t における潜在変数 d_t が、同じ時刻 t における外部変数 v_t と、1 つ前の時刻 $t - 1$ の潜在変数 d_{t-1} に基づいて生成される。また、観測モデルにおいては、時刻 t の観測変数 x_t, y_t が、それぞれ、同じ時刻 t の潜在変数 z_t に加えて、同じ時刻 t の潜在変数 d_t に基づいて生成される。モデリングシステム 1 は、上述した実施形態と同様にして、学習部 5 がモデル化処理を実行し、予測値生成器 7 が探索処理を実行することができる。このような学習モデルに対応した処理を実行することで、生物の離散的な状態（例えば、生育段階）と連続的な状態（例えば、水分状態）等を組み合わせた自然現象を適切にモデリングすることができる。

30

【0046】

また、上記実施形態のモデリングシステム 1 においては、第 1 導出部 7 a が、外部変数 u_t を用いなくて、潜在変数 z_{t-1} のみから、潜在変数 z_t を導出してもよいし、第 2 導出部 7 b が、外部変数 s_t を用いなくて、潜在変数 z_t のみから、潜在変数 x_t を導出してもよいし、第 3 導出部 7 c が、外部変数 r_t を用いなくて、潜在変数 z_t のみから、潜在変数 y_t を導出してもよい。同様に、図 5 あるいは図 7 に示す生成過程に対応する学習モデルが採用された場合も、いずれかあるいは全ての外部変数を用いなくて各変数が導出されてもよい。

40

【0047】

以下、本実施形態に係るモデリングシステム 1 における観測変数の推定精度について評価する。ここでは、図 3 に対応する学習モデルを欠損値の無いラベルデータ系列 D_L を用いてモデル化した場合（以下、「Cont-S」と示す。）、図 3 に対応する学習モデル

50

を欠損値のあるアンラベルデータ系列 D_u を用いてモデル化した場合（以下、「Cont - SS」と示す。）、図5に対応する学習モデルを欠損値の無いラベルデータ系列 D_l を用いてモデル化した場合（以下、「Disc - S」と示す。）、図5に対応する学習モデルを欠損値のあるアンラベルデータ系列 D_u を用いてモデル化した場合（以下、「Disc - SS」と示す。）、図6に対応する学習モデルを欠損値の無いラベルデータ系列 D_l を用いてモデル化した場合（以下、「2L - S」と示す。）、及び、図6に対応する学習モデルを欠損値のあるアンラベルデータ系列 D_u を用いてモデル化した場合（以下、「2L - SS」と示す。）の評価結果を示した。

【0048】

図7には、それぞれの場合での算出糖度の誤差の評価値 MAE (Mean Absolute Error)、RMSE (Root Mean Square Error)、RAE (Relative Absolute Error)、RSE (Relative Square Error) を示している。このように、欠損値のあるデータ系列も用いてモデル化した Cont - SS の場合は、欠損値の無いデータ系列のみを用いてモデル化した Cont - S に比較して、MAE が 41%、RMSE が 46% 低下した。また、その他の生成過程を有する学習モデルについても、同様に、欠損値のあるデータ系列も用いてモデル化した場合のほうが、精度が向上していることが評価された。特に、Cont - SS、Disc - SS、及び 2L - SS の全てにおいて、RAE 及び RSE が 1 未満となっており高精度が実現できている。

10

【0049】

図8～10には、それぞれの場合での算出された糖度平均値の時間変化を糖度の実測値と比較して示している。これらの結果から、欠損値のあるデータ系列も用いてモデル化した Cont - SS、Disc - SS、及び 2L - SS の場合は、欠損値の無いデータ系列のみを用いてモデル化した Cont - S、Disc - S、及び 2L - S に比較して、糖度の時間変化に追従できていることが分かった。このように、観測変数に欠損値が生じても相補的なモデル化処理によって潜在空間を適切に表現できおり、少量の訓練データによっても高い表現力が維持された学習モデルをモデル化できていることが分かった。

20

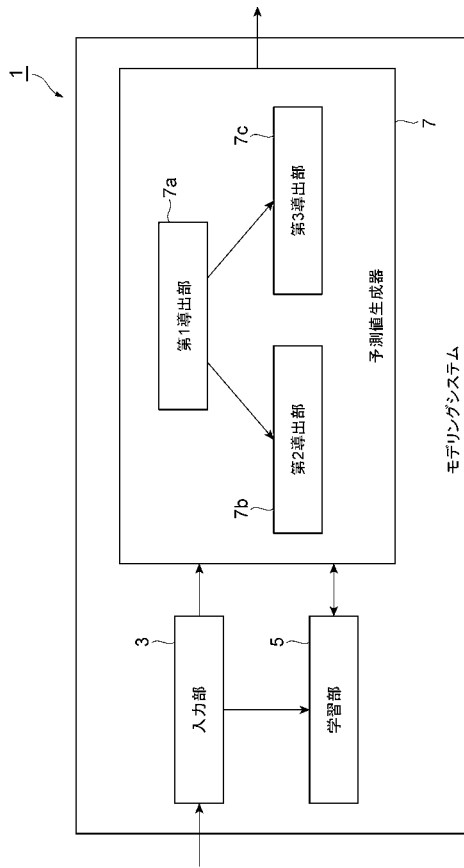
【符号の説明】

【0050】

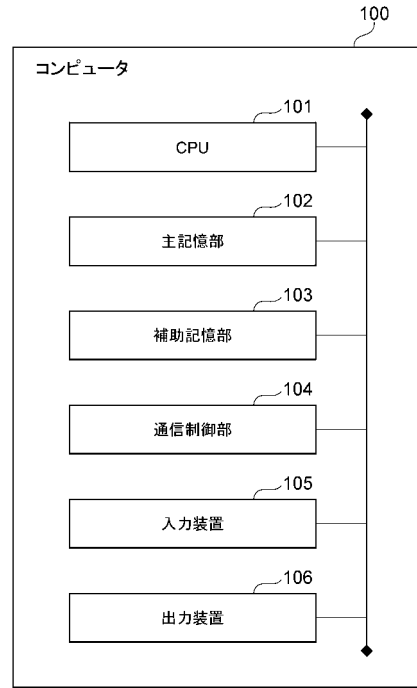
1 ... モデリングシステム、3 ... 入力部、5 ... 学習部、7 ... 予測値生成器、7a ... 第1導出部、7b ... 第2導出部、7c ... 第3導出部、 d_t, z_t ... 潜在変数、 r_t, s_t, u_t, v_t ... 外部変数、 x_t, y_t ... 観測変数。

30

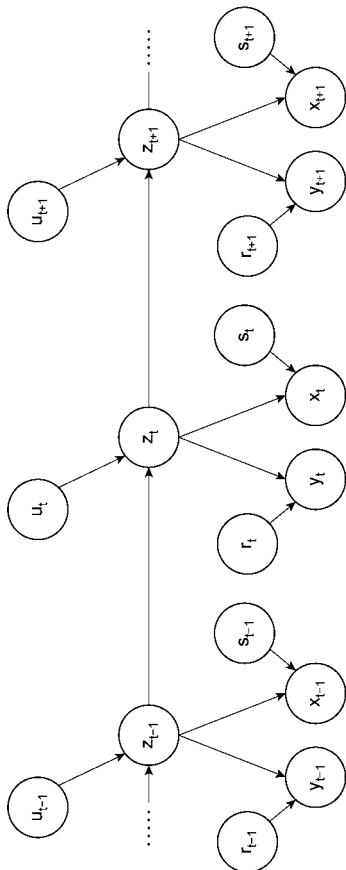
【図1】



【図2】



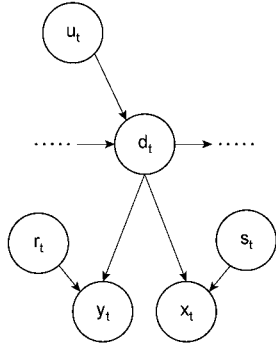
【図3】



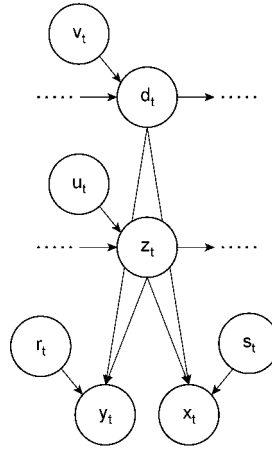
【図4】

変数	設定内容
x_t	糖度、等
y_t	茎径、等
u_t	温度、光量、飽差、開花後の経過日数、積算温度、等
r_t, s_t	CO ₂ 濃度、光量、対象花房、等
z_t	植物の状態

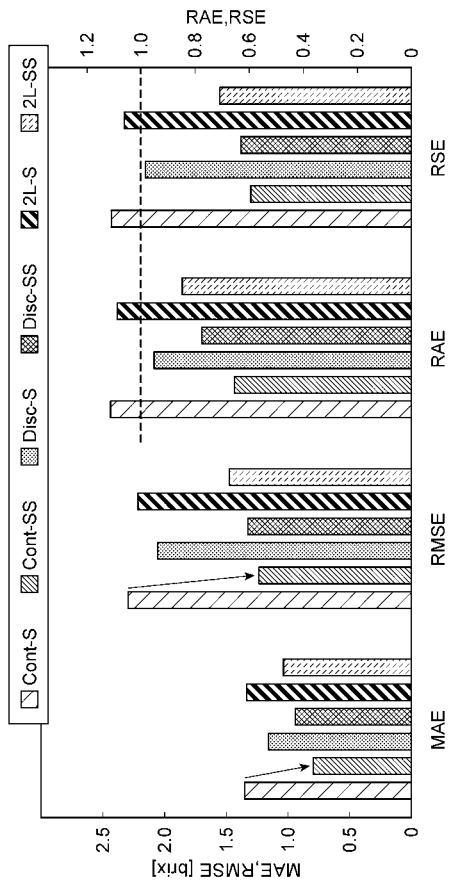
【 図 5 】



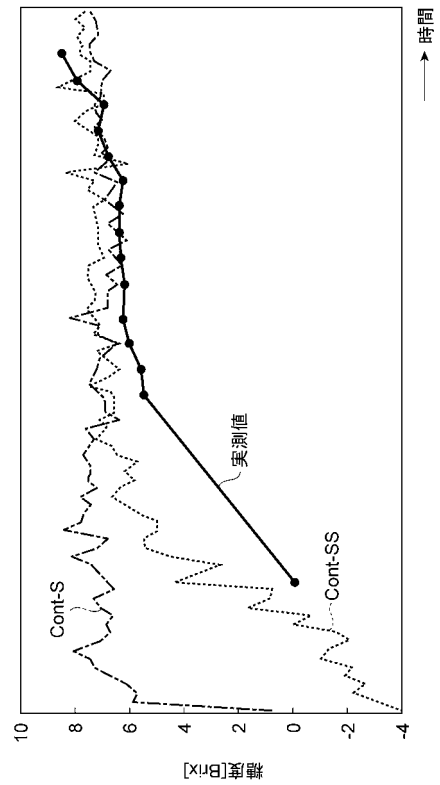
【 図 6 】



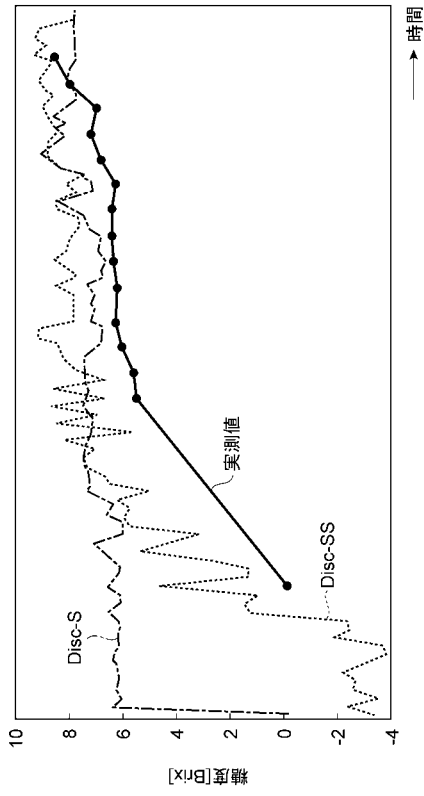
【 図 7 】



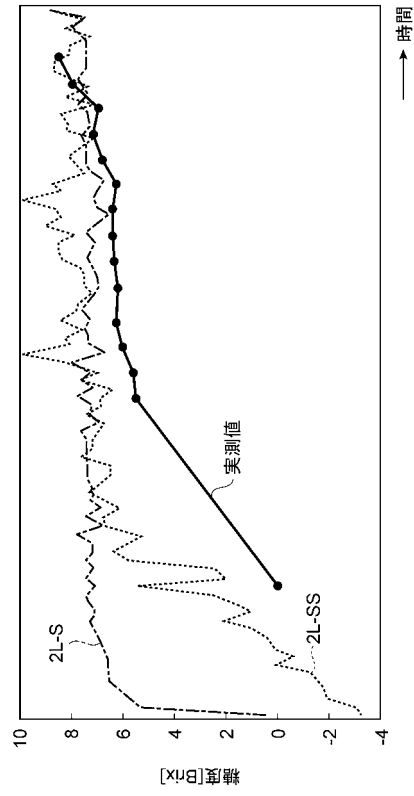
【 図 8 】



【 図 9 】



【 図 10 】



フロントページの続き

(72)発明者 水野 涼介

静岡県浜松市中区城北3丁目5-1 国立大学法人静岡大学内

(72)発明者 中西 豪太

静岡県浜松市中区城北3丁目5-1 国立大学法人静岡大学内

Fターム(参考) 5L049 DD02