

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関  
国際事務局  
(43) 国際公開日  
2018年1月18日(18.01.2018)

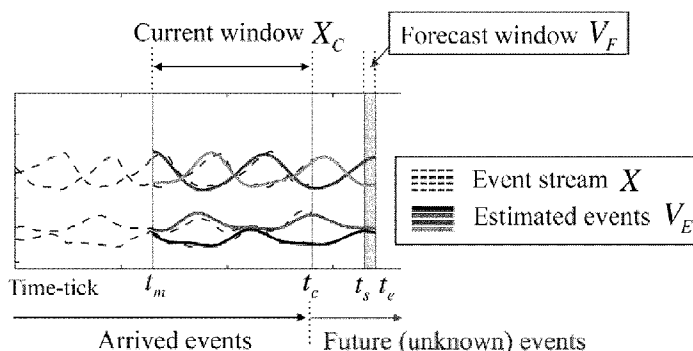


(10) 国際公開番号  
**WO 2018/012487 A1**

- (51) 国際特許分類:  
*G06Q 10/04* (2012.01)
- (21) 国際出願番号: PCT/JP2017/025236
- (22) 国際出願日: 2017年7月11日(11.07.2017)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:  
特願 2016-138075 2016年7月12日(12.07.2016) JP
- (71) 出願人: 国立大学法人熊本大学 (NATIONAL UNIVERSITY CORPORATION KUMAMOTO UNIVERSITY) [JP/JP]; 〒8608555 熊本県熊本市中央区黒髪二丁目39番1号 Kumamoto (JP).
- (72) 発明者: 松原 靖子 (MATSUBARA Yasuko); 〒8608555 熊本県熊本市中央区黒髪二丁目39番1号 国立大学法人熊本大学内 Kumamoto (JP). 櫻井 保志 (SAKURAI Yasushi); 〒8608555
- 熊本県熊本市中央区黒髪二丁目39番1号 国立大学法人熊本大学内 Kumamoto (JP).
- (74) 代理人: 羽立 幸司 (HADATE Koji); 〒8140001 福岡県福岡市早良区百道浜三丁目8番33号 福岡システムLSI総合開発センター 知的財産総合事務所 N E X P A T 内 Fukuoka (JP).
- (81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS,

(54) Title: FORECASTING DEVICE, PARAMETER SET PRODUCTION METHOD AND PROGRAM

(54) 発明の名称: 予測装置、パラメータ集合生産方法及びプログラム



(57) Abstract: The purpose of the present invention is to provide a forecasting device capable of also achieving long-term forecasting with high accuracy by using a large-scale time-series data stream. In the case of the forecasting device 1, a count window calculation unit 11, a regime update unit 13, and a regime addition unit 15 correspond to each hierarchy of a hierarchical structure of a data stream. Furthermore, a numerical model specified by a parameter set stored in a parameter set storage unit 7 includes a non-linear component, and is thus able to express non-linearity of the data stream. The regime update unit 13 updates the parameter set, thereby achieving forecasting by means of a non-linear dynamic system. Furthermore, the regime addition unit 15 adds a new pattern (regime) to the data stream. The regime update unit 13 utilizes regime shift, that is, transition from a certain regime to another regime, in an event stream. Consequently, highly accurate long-term forecasting can also be achieved.

WO 2018/012487 A1

MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類：

- 一 国際調査報告（条約第21条(3)）

---

(57) 要約：大規模な時系列データストリームを用いて、高精度に長期予測をも実現することが可能な予測装置等を提案する。予測装置1において、データストリームの階層的な構造に、カレントウィンドウ計算部11、レジーム更新部13、レジーム追加部15が対応する。さらに、パラメータ集合記憶部7に記憶されたパラメータ集合により特定される数理モデルは、非線形要素を含み、データストリームの非線形性を表現することができる。レジーム更新部13は、パラメータ集合を更新することにより、非線形動的システムによる予測を実現する。さらに、レジーム追加部15は、データストリームにおける新たなパターン（レジーム）を追加する。レジーム更新部13は、あるレジームから別のレジームへと移るイベントストリーム上でのレジームシフトを利用する。これにより、高精度な長期予測をも実現する。

## 明 細 書

**発明の名称**： 予測装置、パラメータ集合生産方法及びプログラム  
**技術分野**

[0001] 本願発明は、予測装置、パラメータ集合生産方法及びプログラムに関し、特に、時刻  $t_c$  までのデータストリーム  $X$  の一部であるカレントウィンドウ  $X_c$  を用いて時刻  $t_c$  から  $l_s$  ステップ以降の一つ又は複数のイベント値  $V_E$  を予測する予測装置等に関する。

### 背景技術

[0002] 発明者らは、これまで、時系列データの解析を研究してきた（非特許文献1参照）。発明者らが提案したAutoPlaitは、時系列データの特徴抽出に関して注目されている。

[0003] 時系列データに基づく予測について、例えば、自己回帰モデル（AR：autoregressive model）、その発展形であるARIMA（Autoregressive integrated moving average model）、線形動的システム（LDS：linear dynamical systems）、カルマンフィルタ（KF：Kalman filters）などが知られている。これらに基づく時系列の解析と予測手法として、AWSOM、TBATS、PLiFなどが提案されている。

### 先行技術文献

#### 非特許文献

[0004] 非特許文献1：Y. Matsubara、外1名、Christos Faloutsos：“AutoPlait: Automatic Mining of Co-evolving Time Sequences”，ACM SIGMOD Conference, pp. 193-204, Snowbird, Utah, June 2014.

### 発明の概要

#### 発明が解決しようとする課題

[0005] しかしながら、従来、ほとんどの手法が線形システムを利用しているため、非線形性を有する時系列データの特徴の表現には不十分であった。また、非線形システムを利用するものは、最近傍探索に基づくものが主流であり、

長期予測のための時系列のモデル化の能力を有さなかった。

[0006] そこで、本願発明は、大規模な時系列データストリームを用いて、高精度に長期予測をも実現することが可能な予測装置等を提案することを目的とする。

### 課題を解決するための手段

[0007] 本願発明の第1の観点は、時刻  $t_c$  までの時系列データ  $X$  の一部であるカレントウィンドウ  $X_c$  を用いて時刻  $t_c$  から  $l_s$  ステップ以降の一つ又は複数のイベント値を予測する予測装置であって、パラメータ集合記憶手段と、レジーム更新手段と、予測手段を備え、前記パラメータ集合記憶手段は、数理モデルを特定するパラメータ集合を記憶し、前記数理モデルは、非線形要素を含み、前記パラメータ集合は、前記非線形要素の係数を特定する非線形パラメータを含み、前記レジーム更新手段は、前記非線形パラメータを変更せず、前記パラメータ集合に含まれる他のパラメータの一部又は全部を更新して、カレントウィンドウ  $X_c$  の各時刻のデータと、更新後の前記パラメータ集合により特定される数理モデルを用いて得られるカレントウィンドウ  $X_c$  の各時刻に対応するイベント値  $V_c$  との違いを小さくし、前記予測手段は、更新後の前記パラメータ集合により特定される数理モデルを用いて時刻  $t_c$  から  $l_s$  ステップ以降の一つ又は複数のイベント値を予測するものである。

[0008] 本願発明の第2の観点は、第1の観点の予測装置であって、前記パラメータ集合記憶手段は、 $c$  個 ( $c$  は自然数) のパラメータ集合  $\theta_i$  ( $i = 1, \dots, c$ ) を記憶し、前記予測手段は、更新後の  $c$  個のパラメータ集合  $\theta_i$  の一部又は全部を用いてイベント値  $V_E$  を予測するものである。

[0009] 本願発明の第3の観点は、レジーム追加手段を備え、前記レジーム追加手段は、カレントウィンドウ  $X_c$  の各時刻でのデータと、更新後の  $c$  個のパラメータ集合  $\theta_i$  を用いて得られる対応する各時刻のイベント値  $V_c$  との違いが追加条件を満たすならば、前記パラメータ集合記憶手段に、新たなパラメータ集合  $\theta_{c+1}$  を追加し、前記レジーム更新手段は、 $c + 1$  個のパラメータ集合  $\theta_i$  ( $i = 1, \dots, c + 1$ ) に含まれる非線形パラメータ以外のパラメータの一部

又は全部を更新し、前記予測手段は、更新後の  $c + 1$  個のパラメータ集合  $\theta_i$  ( $i = 1, \dots, c + 1$ ) により特定される数理モデルの一部又は全部を用いてイベント値を予測する、請求項 2 記載の予測装置。

[0010] 本願発明の第 4 の観点は、第 1 から第 3 のいずれかの観点の予測装置であって、前記数理モデルは、線形要素を含み、前記パラメータ集合は、前記線形要素を特定する線形パラメータを含み、前記レジーム追加手段は、前記非線形パラメータを変更せずに前記線形パラメータを決定し、決定した前記線形パラメータを使用して前記非線形パラメータを決定するものである。

[0011] 本願発明の第 5 の観点は、第 1 から第 4 のいずれかの観点の予測装置であって、カレントウィンドウ  $X_c^{(j)}$  ( $j = 1, \dots, h$ 、 $h$  は自然数) は、 $h$  階層であり、前記パラメータ集合は、 $h$  階層の前記カレントウィンドウ  $X_c^{(j)}$  に対応して  $h$  階層であり、前記レジーム更新手段は、各階層において前記パラメータ集合を更新し、前記予測手段は、各階層において予測されたイベント値から全体のイベント値を予測するものである。

[0012] 本願発明の第 6 の観点は、時刻  $t_c$  までの時系列データの一部であるカレントウィンドウ  $X_c$  を用いて、数理モデルを特定するパラメータ集合の一部のパラメータを変更して新たなパラメータ集合を生産するパラメータ集合生産方法であって、前記数理モデルは、非線形要素を含み、前記パラメータ集合は、前記非線形要素を特定する非線形パラメータを含み、前記情報処理装置が備えるレジーム更新手段が、前記非線形パラメータを変更せず、前記パラメータ集合に含まれる他のパラメータの一部又は全部を更新して、カレントウィンドウ  $X_c$  の各時刻のデータと、更新後の前記パラメータ集合により特定される数理モデルを用いて得られるカレントウィンドウ  $X_c$  の各時刻に対応するイベント値  $V_c$  との違いを小さくする更新ステップを含むものである。

[0013] 本願発明の第 7 の観点は、コンピュータを、第 1 から第 5 のいずれかの観点の予測装置として機能させるためのプログラムである。

[0014] なお、本願発明を、第 7 の観点のプログラムを定常的に記録するコンピュータ読み取り可能な記録媒体として捉えてもよい。

## 発明の効果

- [0015] 発明者らは、レジームという概念に着目した。レジームは、環境生態学において、自然現象内の特徴的な時系列パターンのことを指す。レジームシフトとは、あるレジーム（時系列パターン）から別のレジームに変化する現象を示す。レジームシフトは、近年、様々な分野において活発に研究されており、とりわけ環境生態学分野において活発に取り組まれている課題である。
- [0016] 発明者らは、自然界の動的システムにおけるレジームシフトの概念を拡張し、新たな時系列予測手法を提案する。自然界の動的システムと同様に、実世界におけるデータストリームは、様々な潜在的要素に影響されながら時間発展していく。
- [0017] 本願発明の各観点によれば、大規模な時系列データストリームが与えられたとき、その潜在的なパターンを、非線形要素を含んだ数理モデルにより表現する。そして、非線形パラメータ以外のパラメータ（例えば初期値など）を変更して非線形動的システムを用いて、非線形要素による潜在的なパターンの表現を維持しつつ適応させて、実世界におけるデータストリームを使って精度よく予測することが可能になる。
- [0018] さらに、本願発明の第2の観点によれば、レジームシフトの概念を用いて、複数の非線形モデルを使用することにより時系列データストリーム上でのレジームシフトを発見し、複雑に変化する時系列パターンを表現して、高精度に予測することができる。
- [0019] さらに、本願発明の第3の観点によれば、レジームシフトに伴い、新たなレジームを自動発見して、新たな時系列パターンを追加することができる。
- [0020] さらに、本願発明の第4の観点によれば、新たなレジームを表すモデルを、線形要素を特定した後に非線形要素を特定することにより、計算量を抑えつつ、精度よく推定することが可能になる。
- [0021] さらに、実際の時系列データストリームは、異なる時間発展に基づく多階層の動的システムから構成され、複雑な時系列パターンを有する。つまり、階層的な構造を伴う。本願発明の第5の観点によれば、多階層の構造を利用

して予測することにより、高精度に予測することができる。

### 図面の簡単な説明

- [0022] [図1] (a) 一般的なレジームシフトの概念と (b) 本願発明のレジームシフトの概念の一例を示す。
- [図2] 本願発明の実施例による現時刻  $t_0$  における分析と予測のスナップショットを示す。
- [図3] 本願発明の実施例であるRegimeCastの3つのアルゴリズムを示す。
- [図4] 本願発明の実施の形態に係る予測装置1の構成の概要を示すブロック図である。
- [図5] 図4の予測装置1の動作の一例を示すフロー図である。
- [図6] 図4の予測装置1において生成されるデータの一例を示す。
- [図7] 身体の運動に関するモーションストリームに対するRegimeCastの解析例を示す。
- [図8] チキンダンスに対するRegimeCastの解析例を示す。
- [図9] Googleにおける3ヶ月先の各キーワードの検索量の予測結果を示す。
- [図10] 本願発明と従来手法の予測精度を示す。
- [図11] 提案手法と従来手法の計算コストを示し、長期的なイベント予測での予測精度と計算コストを示す。

### 発明を実施するための形態

- [0023] 以下では、図面を参照して、本願発明の実施例について説明する。なお、本願発明は、この実施例に限定されるものではない。

#### 実施例

- [0024] まず、図1を参照して、レジームシフトの概念について説明する。レジームは、環境生態学において自然現象内の特徴的な時系列パターンのことを指す。レジームシフトは、あるレジームから別のレジームに変化する現象を示す。
- [0025] レジームシフトは、主に、内的要因（システム内の安定性の変化等）及び／又は外部要因（システムへの外部ショック等）により引き起こされる。例

例えば、図1(a)のように、草原地帯は森林地帯へシフトすることがある。草原地帯は、例えば、草食動物の存在、火事、森林伐採などの要素により木々の成長が抑えられ、安定したシステムとなる。何らかの理由により木々が一定以上の大きさに成長すれば、草食動物や火事の影響を受ける可能性が減少し、森林地帯へ移行する。また、同様に、森林地帯から草原地帯へシフトすることがある。

[0026] 発明者らは、自然界の動的システムにおけるレジームシフトの概念を拡張し、新たな時系列予測手法を提案する。図1(b)のモーションデータストリームのように、時系列データストリームでの時系列パターンをレジームとし、イベントストリーム上でのレジームシフトを利用することにより、予測精度を向上させる。特に、時系列データを適応型非線形動的システムとして表現することにより、複雑な時系列パターンを柔軟に表現する。そして、その適応型非線形動的システムを用いることにより予測精度を向上させる。

[0027] 本願発明によれば、様々な時系列パターン（例えばセンサデータやWebのアクセス履歴等）から構成される大規模データストリームが与えられたとき、それらの中から重要な特徴や潜在的なトレンドを発見し、長期的かつ継続的に将来の時系列予測を行うことができる。

[0028] まず、予測に使用する数理モデルの一例について説明する。

[0029] 生態系における各要素の時間発展は、数1の常微分方程式で表現することができる。ここで、 $s(t)$ は時刻 $t$ における生態系の特性（養分や土壌など）を表す。 $a_0$ は栄養負荷などの $s(t)$ を変化させる環境要因を表す。 $a_1$ はシステム内の $s(t)$ の成長、減衰率を示す（例えば、 $a_1 < 0$ における栄養除去率など）。 $a_2$ は $s(t)$ の関数 $f(s(t))$ による回復率（栄養循環など）を示す。関数 $f$ によりレジームの推移が発生する。

[0030] [数1]

$$\frac{ds(t)}{dt} = a_0 + a_1s(t) + a_2f(s(t))$$

[0031] 発明者らは、数1の動的システムにおけるレジームシフトの概念を拡張した。まず、最もシンプルなパターンとして、単一の動的パターン（レジーム



) の表現方法について述べる。ここでは、イベントシーケンスの中にレジームシフトは存在しない。このモデルは、 $s(t)$  と  $v(t)$  の2種類の時系列活動パターンから構成される。ここで、 $s(t)$  は、 $k$ 次元の潜在値（潜在的な時系列パターン）を示す。 $v(t)$  は、時刻  $t$  における  $d$ 次元の推定イベント（実際のイベント観測値）を示す（例えば、 $d$ 個のセンサから生成される実測値）。これにより、単一のレジームは、数2で表現される。

・  $s(t)$  : 潜在値、時刻  $t$  における  $k$ 次元の潜在的な活動値 ( $s(t) = \{s_i(t)\}_{i=1}^k$ )

・  $v(t)$  : 推定イベント、時刻  $t$  における  $d$ 次元の観測値 ( $v(t) = \{v_i(t)\}_{i=1}^d$ )

[0032] ここで、初期条件を  $s(0) = s_0$  とする。 $ds(t)/dt$  を時刻  $t$  の導関数とする。 $S(t)$  を  $s(t)$  の2次形式とする（すなわち、 $S(t) = s(t)^T s(t)$ ）。また、 $p$ 、 $Q$ 、 $A$  は、潜在値  $s(t)$  を生成するパラメータ集合であり、各成分が、線形、指数、非線形の動的パターンを表現する（なお、この例では、非線形動的パターンの要素  $A$  を2次関数として扱う）。 $u$ 、 $V$  は、時刻  $t$  における潜在値  $s(t)$  から推定イベント  $v(t)$  への射影を示す。さらに、非線形テンソル  $A$  については、動的システムの複雑化を防ぐために、スパースであることが重要である。

[0033] 単一の潜在的な非線形動的システムにおけるパラメータ集合を  $\theta = \{s_0, p, Q, A, u, V\}$  とする。

[0034] [数2]

$$\frac{ds(t)}{dt} = p + Qs(t) + AS(t)$$

$$v(t) = u + Vs(t)$$

[0035] 次に、イベントストリーム上でのレジームシフトについて述べる。例えば、図1(b)における  $c = 2$  種類のレジーム (walking, wiping) が任意のタイミングで交互に繰り返されるものである。このような複雑なイベントを表現するため、順応性の高い時系列モデルが必要になる。そこで、より複雑な時系列パターンを表現するために、 $w(t)$  を導入する。ここで、 $w(t)$

は、時刻  $t$  における  $i$  番目のレジーム ( $i = 1, \dots, c$ ) の推移の強さを示す。

・  $w(t)$  : レジーム活動値、時刻  $t$  における  $c$  個のレジームにおけるレジームシフトの推移値。

[0036] 数2のモデルを拡張し、数3のモデルを提案する。ここで、 $s_i(t)$  は、時刻  $t$  における  $i$  番目のレジームの潜在値である ( $s_i(t) = \{s_{ij}(t)\}_{j=1}^k$ )。  $w(t)$  は、時刻  $t$  における  $i$  番目のレジームの強さである ( $w(t) = \{w_i(t)\}_{i=1}^c$ )。  $v(t)$  は、時刻  $t$  における  $d$  次元の推定イベントである。  $dw(t)/dt$  は、時刻  $t$  の導関数を示す。

[0037] 数3において、新たなパラメータとして  $r(t)$  を導入する。  $r(t)$  は、時刻  $t$  における  $c$  次元のベクトルとして表現される。  $R$  を、レジームシフトのダイナミクスを表現するパラメータ集合とする ( $R = \{r(t)\}_{t=t_0}^{t_c}$ )。ここで、  $t_c$  は、イベントストリームの長さを表す。  $R$  をレジームシフト行列と呼ぶ。  $c = 1$  の場合 (イベントストリームが単一のレジームで構成されている場合) には、数3のモデルは、数2のモデルに一致する。

[0038] レジームのパラメータ集合を、  $\Theta = \{\theta_1, \dots, \theta_c, R\}$  とする。ここで、  $c$  は、イベントストリームに含まれるレジームの個数を示す。

[0039] [数3]

$$\frac{ds_i(t)}{dt} = \mathbb{P}_i + \mathbb{Q}_i s_i(t) + \mathcal{A}_i s_i(t)$$

$$\frac{dw(t)}{dt} = r(t)$$

$$v(t) = \sum_{i=1}^c w_i(t) [\mathbb{U}_i + \mathbb{V}_i s_i(t)]$$

[0040] さらに、多階層の構造への拡張について説明する。数2及び数3のシステムは、単一階層における動的システムであった。しかしながら、実世界における時系列イベントは、異なる時間発展に基づく時系列活動パターンを含んでいる。例えば、Web上のイベントにおける10年周期や1日周期などである。そこで、階層的な構造をとる時系列パターンを表現するため、階層構

造に基づくモデルを使用する。具体的には、多階層のレジーム集合  $M = \{\Theta^{(1)}, \Theta^{(2)}, \dots\}$  を用いる。レジーム集合は、階層構造をとる時系列パターンを表現するモデルの全パラメータ集合である。階層  $i$  においてローカルな推定イベント  $v^{(i)}(t)$  を生成し、重ね合わせることで、実際の推定イベント  $v(t)$  を表現する。

[0041] 続いて、必要な概念を説明する。表 1 は、主な記号と定義を示す。

[0042] [表1]

記号	定義
$d$	時系列の次元数
$t_c$	現在の時刻
$X$	$d$ 次元のイベントストリーム： $X = \{\mathbf{x}(1), \dots, \mathbf{x}(t_c)\}$
$\mathbf{x}(t)$	時刻 $t$ における $d$ 次元のイベント： $\mathbf{x}(t) = \{x_i(t)\}_{i=1}^d$
$\mathbf{s}(t)$	時刻 $t$ における潜在値： $\mathbf{s}(t) = \{s_i(t)\}_{i=1}^k$
$\mathbf{w}(t)$	時刻 $t$ におけるレジーム活動値： $\mathbf{w}(t) = \{w_i(t)\}_{i=1}^c$
$\mathbf{v}(t)$	時刻 $t$ における推定イベント： $\mathbf{v}(t) = \{v_i(t)\}_{i=1}^d$
$X_c$	カレントウィンドウ： $X_c = X[t_m:t_c]$
$V_F$	予測ウィンドウ： $V_F = V[t_s:t_e]$
$c^{(i)}$	$i$ 番目の階層におけるレジームの個数
$\theta_j^{(i)}$	$i$ 番目の階層におけるレジーム $j$ のパラメータ集合
$R^{(i)}$	$i$ 番目の階層におけるレジームシフト行列
$\Theta^{(i)}$	$i$ 番目の階層の全パラメータ集合
$\mathcal{M}$	全パラメータ集合： $\mathcal{M} = \{\Theta(t)\}_{i=1}^h$

[0043]  $X$  を、 $d$ 次元のイベントエントリから構成されるデータストリーム  $X = \{\mathbf{x}(1), \dots, \mathbf{x}(t_c)\}$  とし、 $t_c$  を現在の時刻とする。 $X$  を、イベントストリームと呼ぶ。

[0044] 毎時刻において、新たなイベントエントリ  $\mathbf{x}(t_c)$  が発生し、時刻が進むごとに  $t_c$  が増加するものとする。そこで、最新の時刻において発生したイベント集合をカレントウィンドウとして次のように定義する。 $X_c = X[t_m:t_c]$  を長さ  $l_c$  のカレントウィンドウの部分シーケンスとする。ここで、 $X_c$  は、イベントストリーム  $X$  の時刻  $t_m$  から時刻  $t_c$  まで ( $1 \leq t_m \leq t_c$ ) の部分シーケンスを示す。以下では、 $l_c = 3 \cdot l_s$  とする。

[0045] カレントウィンドウ  $X_c$  が与えられたとき、次の目標は、パラメータ集合  $M$  の中から最適なレジームを発見し、数  $3$  のモデルに基づき  $l_s$  ステップ先の未来のイベント  $V_F = \{v(t_s), \dots, v(t_e)\}$  を推定することである。  $V_F = V[t_s : t_e]$  を  $l_s$  ステップ先の将来イベントシーケンス ( $t_c \leq t_s \leq t_e$ ) とし、  $t_s = t_c + l_s$ 、  $t_e = t_s + l_p$  とする。ここで、  $l_p$  は、出力単位時間の長さである。

[0046] 図 2 は、本願発明の実施例の一つである RegimeCast の現時刻  $t_c$  におけるスナップショットを示す。ここで、黒点線は、オリジナルのイベントストリーム  $X$  を示す。図 2 では、  $d = 4$  次元のイベントシーケンスで構成される。太線は、時刻  $t_m$  から時刻  $t_e$  における RegimeCast によるイベントの推定値  $V_E$  を示す。ここで、時刻  $t_c$  から時刻  $t_e$  までの部分シーケンスは、未来の（つまり未知の）イベント集合である。本願発明は、これらの時系列パターンを高速かつ継続的に推定し続けなくてはならない。本願発明は、オリジナルのストリーム  $X$ （黒点線）が与えられたとき、カレントウィンドウ  $X_c$  に含まれる最新の時系列パターンを発見し、適応型非線形動的システムとして表現することで、現時刻の時系列パターン  $V_E$ （色太線）を推定し、  $l_s$  ステップ先の将来イベント  $V_F$ （矩形内）を高速かつ継続的に出力する。

[0047] まとめると、以下のように表現することができる。イベントストリーム  $X = \{x(1), \dots, x(t_c), \dots\}$  が与えられたとき、  $l_s$  ステップ先の将来イベント  $V_F$  を出力し続ける。具体的には、各時刻  $t_c$  において、カレントウィンドウ  $X_c$  に含まれる最適なレジームのパターンを検出し、  $X_c$  のレジームパターンに基づきモデルのパラメータ集合  $M$  を更新し、  $l_s$  ステップ先の将来イベント  $V_F$  を出力する。

[0048] 図 3 を参照して、本願発明の実施例の一つである RegimeCast の概要を説明する。RegimeCast は、以下の 3 つのアルゴリズムで構成される。

RegimeReader : カレントウィンドウ  $X_c$  とレジームパラメータ集合  $\Theta$  が与えられたとき、レジームのダイナミクスを推定し、イベント  $V_E = V[t_m : t_e]$  を生成する（図 3（a）参照）。

RegimeEstimator : カレントウィンドウ  $X_c$  の中に新たなレジームパターンが含まれていた場合に、 $X_c$  を表現する新たなパラメータ集合  $\theta$  を推定する (図 3 (b) 参照)。

RegimeCast : 各階層  $i$  ( $i = 1, \dots, h$ ) における最適なイベント集合  $V_E^{(i)}$  を推定し、推定イベント  $V_E = V_E^{(1)} + V_E^{(2)} + \dots$  を計算する。その後、 $1_s$  ステップ先のイベント (つまり  $V_F$ ) を報告する。さらに、レジームパラメータ集合  $M$  を更新する (図 3 (c) 参照)。

[0049] 図 3 のアルゴリズムを詳細に説明する。議論の単純化のため、まずは、単一の階層 (つまり、 $h = 1$ ) のみを考慮し、単一のカレントウィンドウ  $X_c$  とレジームパラメータ集合  $\Theta$  が与えられた場合について説明する。

[0050] RegimeReader で、時刻  $t_c$  におけるカレントウィンドウ  $X_c$  とレジームパラメータ集合  $\Theta = \{\theta_1, \dots, \theta_c, R\}$  が与えられた場合を考える。RegimeReader の目的は、現在のレジームパラメータ集合  $\Theta$  に基づき、イベントシーケンス  $V_E = V [t_m : t_c]$  を推定することである。

[0051] 適切なイベントの推定値を得るための最も単純な解決法は、 $\Theta$  内のパラメータ集合を固定し、数 3 のモデルに基づき  $v(t_m)$ 、 $v(t_m + 1)$ 、 $\dots$  を計算することである。しかしながら、実際のイベントストリームでは、カレントウィンドウ  $X_c$  に含まれる潜在的なトレンドは、時間の経過とともに動的かつ連続的に変化していく。そこで、 $\Theta$  内に含まれるレジームのパラメータを最新のカレントウィンドウ  $X_c$  のパターンに基づき最適化する。具体的には、図 3 のアルゴリズムは、 $X_c$  内に含まれる現時刻の活動パターンに応じて  $\Theta$  内のパラメータ集合を柔軟に更新していく必要がある。

[0052] 図 3 (a) は、RegimeReader の処理の流れを示す。RegimeReader は、(1) 個々のレジームの最適化、(2) レジームシフトの同定、の 2 つのパートから構成される。

[0053] (1) 個々のレジームの最適化

個々のレジームのレジームパラメータ  $\theta_i \in \Theta$  ( $i = 1, \dots, c$ ) について、潜在値の初期値  $s_0 \in \theta_i$  を最適化する。具体的には、オリジナルイベントと

推定イベントの2乗誤差を最小化する（つまり、 $\min ||X_c - X_{c_i}||$ ）ような  $s_0$  を求める。ここで、関数  $f_c(s_0 | \theta)$  は、数3のモデルにおける、レジームパラメータ  $s_0$ 、 $\theta$  が与えられたうえでの推定イベント  $V_c = \{v(t_m), \dots, v(t_c)\}$  を示す。

[0054] (2)レジームシフトの同定

(1)で得られた  $c$  個の推定イベントの集合  $\{V_{c_i}\}_{i=1}^c$  に基づき、時刻  $t_c$  におけるレジームシフトの潜在的な動的パターンを推定する。具体的には、 $\Theta$  に含まれるレジームの集合を最適化するためにレジーム活動値  $w(t_c)$  を推定し、数3に基づき  $\Theta$  内のレジームシフト行列  $R$  を更新する（つまり、 $\min ||X_c - f_c(\Theta)||$ ）。その後、推定イベント  $V_E = f_E(\Theta)$  をカレントウィンドウ  $X_c$  に対する最適値として計算する。ここで、平均二乗誤差  $||\cdot||$  を最小化する方法として、非線形性を有する学習に適した LM (Levenberg-Marquardt) アルゴリズムを用いる。

[0055] 次に、図3(b)を参照して、新たなレジームを推定するためのアルゴリズムである RegimeEstimator について説明する。ここでの課題は、カレントウィンドウ  $X_c$  に未知のレジームが含まれていた場合の処理についてである。提案アルゴリズムは、 $X_c$  に含まれる未知の時系列パターンを表現するため、新たなレジーム  $\theta$  を推定し、パラメータ集合  $\Theta$  に挿入する。

[0056] ここで重要な問題として、レジームを表現する  $\theta$  は、非常に多くのパラメータ数から構成されている。一般に、非線形モデルにおける多数のパラメータの同時推定は、最適解の学習が非常に難しく、計算コストも高い。さらに、非線形活動テンソル  $A$  は、単一のレジーム内の時系列パターンの複雑性を抑えるために、スパースであることが重要である。

[0057] そこで、線形、非線形の双方のパラメータ集合を高速かつ効果的に推定するためのアルゴリズムとして RegimeEstimator を提案する。具体的には、パラメータ集合  $\theta$  を、線形、非線形の2種の部分集合： $\theta_L = \{p, Q, u, V\}$ 、 $\theta_N = \{A\}$  に分割し、それぞれのパラメータ集合を個別に推定する。図3(b)を参照して、カレントウィンドウ  $X_c$  が与えられたとき、提案アルゴリ

ズムは、まず非線形活動テンソルを  $A = 0$  とし、 $X_c$  の線形的なパターンを表現するための初期状態  $s_0$  及び線形パラメータ集合  $\theta_L$  を推定する。パラメータの推定には EM (expectation-maximization) アルゴリズムを用いた。続いて、非線形要素  $A$  に関し、推定した  $s_0$  と  $\theta_L$  を用いて、 $X_c$  と潜在値  $V_c$  のエラー値を最小化するように LM アルゴリズムによって最小化する。本研究では、非線形テンソル  $A$  について、モデルの複雑性を抑えるため、対角成分  $a_{ijk} \in A$  ( $i = j = k$ ) のみを推定した。

[0058] これまでは、単一階層のレジーム集合  $\Theta$  に対する推定イベント  $V_E$  の生成方法について述べた。本研究の目的は、多階層における時系列パターン  $M = \{ \Theta^{(1)}, \dots, \Theta^{(h)} \}$  を表現し、 $1_s$  ステップ先の予測ウィンドウ  $V_F$  を推定することである。イベントストリーム  $X$  が与えられたとき、年、週、日単位などの様々な階層の動的パターンを表現したい。そこで、本願発明は、階層モデルに基づく予測手法を提案する。具体的には、カレントウィンドウ  $X_c$  を  $h$  個の階層的なイベント集合  $X_c = X_c^{(1)} + \dots + X_c^{(h)}$  に分割することで、より効果的な予測を実現する。ここで、 $X_c^{(i)}$  は、 $i$  番目の階層におけるイベントを示し、数 4 で計算される。関数  $g(\cdot | t)$  は、長さ  $t$  の移動平均を示す。本論文では、 $H = \{2 \cdot 1_s, 1\}$  とした。

[0059] [数4]

$$X_c^{(i)} = g \left( X_c - \sum_{j=1}^i X_c^{(j)} | H(i) \right)$$

[0060] 図 3 (c) は、RegimeCast の詳細を示す。時刻  $t_c$  における新たなイベント  $x(t_c)$  が与えられたとき、本願発明は、各階層  $i$  におけるカレントウィンドウ  $X_c^{(i)}$  を計算し、(1) イベントシーケンス  $V_E^{(i)}$  を推定する。もし、 $\Theta^{(i)}$  内に適切なレジームが存在しない場合 (つまり、カレントウィンドウと推定イベントの誤差が  $\varepsilon$  以上、例えば、 $\varepsilon = 1 / 2 ||X_c^{(i)}||$  の場合)、(2) 新たなレジーム  $\theta$  を生成し、レジームパラメータ集合  $\Theta^{(i)}$  を更新する。最後に (3)  $1_s$  ステップ先の予測ウィンドウ  $V_F$  を出力する。

[0061] 続いて、ダイナミックポイントセット (DSP) に基づく高速化を説明する。RegimeReaderは、数3のモデルに示したとおり、複雑な動的システムに基づくため、各時刻  $t_c$  において、潜在値  $S_E = \{s(t_m), \dots, s(t_e)\}$  の推定に  $O(l_e)$  の計算量を要する。ここで、 $l_e$  は、 $S_E$  の長さを示す。しかし、この計算時間は、リアルタイム性を要する処理にはボトルネックとなりうる。そこで、本願では、動的なイベント生成を高速化するための手法を提案する。具体的には、すべてのイベント集合  $S_E = \{s(t_m), s(t_m+1), s(t_m+2), \dots, s(t_e)\}$  を生成する代わりに、 $S_E$  の部分集合である  $\hat{S}_E = \{s(t_m), s(t_m+\delta), s(t_m+2\delta), \dots, s(t_e)\}$  のみを生成する。ここで、部分集合  $\hat{S}_E$  を DSP と呼ぶ。  $\delta$  は、潜在値の時間の生成間隔 (例えば、 $\delta = 0.1 \cdot l_s$ ) を示す。  $\hat{S}_E$  は、数5の4次のルンゲ・クッタ法に基づき生成を行う。これにより、モデル推定の計算時間が  $\hat{S}_E$  の長さである  $O(l_e/\delta)$  となり、飛躍的な高速化を実現することができる。

[0062] 理論的な分析を示す。  $l_e$  を、推定イベント集合  $V_E$  の長さとし、  $c$  を、  $M$  に含まれるレジームの個数とする。このとき、各時刻におけるRegimeCastの計算時間は、最小で  $O(c \cdot l_e/\delta)$ 、最大で  $O(c \cdot l_e/\delta + l_c)$  となる。各時刻  $t_c$  において、RegimeReaderは、  $c$  個の最適なレジーム  $V_E$  を推定するために、  $O(c \cdot l_e/\delta)$  の計算時間を要する。もし、カレントウィンドウ  $X_c$  に新たなレジームが含まれていた場合、RegimeEstimatorは、パラメータ集合  $\theta$  の推定に  $O(l_c)$  を要する。したがって、最小で  $O(c \cdot l_e/\delta)$ 、最大で  $O(c \cdot l_e/\delta + l_c)$  の計算時間を要する。

[0063] [数5]

$$s(t+\delta) = s(t) + \frac{1}{6}(K_1 + 2K_2 + 2K_3 + K_4) + O(\delta^5)$$

$$\frac{ds(t)}{dt} = F(s(t)), K_1 = \delta F(s(t)), K_2 = \delta F\left(s(t) + \frac{1}{2}K_1\right),$$

$$K_3 = \delta F\left(s(t) + \frac{1}{2}K_2\right), K_4 = \delta F(s(t) + K_3)$$

[0064] 図4、図5及び図6を参照して、本願発明の実施の形態に係る予測装置の構成及び動作の一例を説明する。

[0065] 図4は、本願発明の実施の形態に係る予測装置1の構成の概要を示すブロック



ック図である。(a)は、予測装置1の構成を示す。(b)及び(c)は、それぞれ、レジーム更新部 $13_j$ 及びレジーム追加部 $15_j$ の構成を示す。

[0066] 図4(a)を参照して、予測装置1は、データストリーム記憶部3と、カレントウィンドウ記憶部5と、パラメータ集合記憶部7(本願請求項の「パラメータ集合記憶手段」の一例)と、 $h$ 個( $h$ は自然数)のカレントウィンドウ計算部 $11_j$ ( $j=1, \dots, h$ )と、 $h$ 個のレジーム更新部 $13_j$ (本願請求項の「レジーム更新手段」の一例)と、レジーム追加部 $15_j$ (本願請求項の「レジーム追加手段」の一例)と、予測部17(本願請求項の「予測手段」の一例)を備える。なお、符号の添え字は、省略する場合がある。データストリーム記憶部3、カレントウィンドウ記憶部5、パラメータ集合記憶部7は、例えばメモリやハードディスクなどの記憶装置によって実現することができる。カレントウィンドウ計算部11、レジーム更新部13、レジーム追加部15、予測部17は、例えばCPUなどの演算装置によって実現することができる。

[0067] 図4(b)を参照して、レジーム更新部 $13_j$ は、 $c$ 個( $c$ は自然数)のアクティビティ演算部 $21_{ji}$ と、重み演算部 $23_j$ と、レジームシフト変数演算部 $25_j$ と、パラメータ更新部 $27_j$ と、推定イベント演算部 $29_j$ を備える。

[0068] 図4(c)を参照して、レジーム追加部 $15_j$ は、線形パラメータ推定部 $31_j$ と、非線形パラメータ推定部 $33_j$ と、パラメータ集合追加部 $35_j$ を備える。

[0069] 図5は、図4の(a)予測装置1と、(b)レジーム更新部 $13_j$ と、(c)レジーム追加部 $15_j$ の動作の一例を示すフロー図である。図5(a)、(b)及び(c)の実質的な処理は、それぞれ、図3(c)、(a)及び(b)と同様である。図6は、図4の予測装置1において生成されるデータの一例を示す。

[0070] 図5(a)を参照して、予測装置1の動作を説明する。予測装置1に、時系列データストリーム $X$ が入力される(ステップST1)。図6(a)は、データストリーム $X$ の一例を示す。データストリーム記憶部3は、入力され

たデータストリーム $X$ を記憶する。

- [0071] 実際の時系列データストリームは、階層的な構造を伴う。カレントウィンドウ計算部 $11_j$ 、レジーム更新部 $13_j$ 及びレジーム追加部 $15_j$ は、階層的な構造を利用して解析するために、時系列データストリームの $h$ 階層のそれぞれに対応する。
- [0072] カレントウィンドウ計算部 $11_j$ は、対応する階層のカレントウィンドウ $X_c^{(i)}$ を生成する（ステップ $ST2$ ）。図6（a）に、カレントウィンドウの設定の一例を示す。カレントウィンドウ記憶部 $5$ は、生成されたカレントウィンドウ $X_c^{(i)}$ を記憶する。
- [0073] パラメータ集合記憶部 $7$ は、全パラメータ集合 $M$ を記憶する。
- [0074] レジーム更新部 $13_j$ は、カレントウィンドウ $X_c^{(i)}$ とレジームパラメータ集合 $\Theta^{(i)}$ から、 $V_E^{(i)}$ と更新後の $\Theta^{(i)}$ を得る（ステップ $ST3$ ）。具体的な処理は、図5（b）を用いて後に説明する。図6（b）及び（c）は、第1階層の2つのパラメータ集合 $\theta_1(1)$ 及び $\theta_2(1)$ により生成されるデータを示す。図6（d）及び（e）は、第2階層の2つのパラメータ集合 $\theta_1(2)$ 及び $\theta_2(2)$ により生成されるデータを示す。なお、実際には、多くのパラメータ集合が使用される。
- [0075] レジーム追加部 $15_j$ は、更新後の $\Theta^{(i)}$ を用いて、カレントウィンドウ $X_c^{(i)}$ の各時刻に対応する $V_c^{(i)}$ を得る（ステップ $ST4$ ）。そして、カレントウィンドウ $X_c^{(i)}$ と $V_c^{(i)}$ との誤差が小さいか否かを判断する（ステップ $ST5$ ）。誤差は、例えば、図3にあるように、二乗誤差が所定の値 $\varepsilon$ 以下か否かによって判断することができる。これらの値が近く、追加条件を満たさないならば、ステップ $ST7$ に進む。これらの値が近くなく、追加条件を満たすならば、レジーム追加部 $15_j$ は、新たなパラメータ集合 $\theta$ を追加し（ステップ $ST6$ ）、ステップ $ST7$ へ進む。ステップ $ST6$ の具体的な処理は、図5（c）を用いて説明する。
- [0076] ステップ $ST7$ で、予測装置は、各階層の予測値 $V_E^{(i)}$ を統合し、全体の予測値 $V_E$ を得る。そして、 $l_s$ ステップ先の未来のイベント値 $V_F$ を出力する。図

6 (f) は、各  $V_E^{(i)}$  を統合して得られる  $V_E$  と、将来イベント  $V_F$  との関係を示す。

[0077] 続いて、図5 (b) を参照して、レジーム更新部 1 3<sub>j</sub> の動作を具体的に説明する。図4 (b) では、アクティビティ演算部 2 1<sub>ji</sub> は、階層のパラメータ集合  $\theta$  の数と同数である。

[0078] レジーム更新部 1 3<sub>j</sub> は、カレントウィンドウ  $X_c^{(i)}$  と、現時点でのレジームパラメータ  $\Theta^{(i)}$  を入力する (ステップ STR 1)。そして、各アクティビティ演算部 2 1<sub>ji</sub> は、対応するパラメータ集合  $\theta_i^{(i)}$  の一部を変更して、カレントウィンドウ  $X_c^{(i)}$  と  $V_c^{(i)}$  の値の誤差を小さくする (ステップ STR 2)。ここで、変更するパラメータは、例えば初期値のように、非線形パラメータ以外のものとする。

[0079] 続いて、レジームのアクティビティを分析する。重み演算部 2 3<sub>j</sub> は、各  $V_{ci}$  の重みを、重み付けした  $V_{ci}$  と  $X_c$  との誤差が小さくするように決定する。レジームシフト変数演算部 2 5<sub>j</sub> は、各  $V_{ci}$  の重みの差分をとり、レジームシフト変数を演算する。パラメータ更新部 2 7<sub>j</sub> は、レジームシフト変数を利用して、対応する階層のレジームパラメータ集合  $\Theta^{(i)}$  を更新する (ステップ STR 3)。推定イベント演算部 2 9<sub>j</sub> は、各階層の推定イベント  $V_E^{(i)}$  を計算し、 $V_E^{(i)}$  と  $\Theta^{(i)}$  を出力する (ステップ STR 4)。

[0080] 続いて、図5 (c) を参照して、レジーム追加部 1 5<sub>j</sub> の動作を具体的に説明する。レジーム追加部 1 5<sub>j</sub> は、カレントウィンドウ  $X_c^{(i)}$  を入力する (ステップ STE 1)。そして、線形パラメータ推定部 3 1<sub>j</sub> は、非線形パラメータを初期化し (ステップ STE 2)、初期値と線形パラメータを求める (ステップ STE 3)。そして、非線形パラメータ推定部 3 3<sub>j</sub> は、求められた線形パラメータを使用して、新たな初期値と非線形パラメータを得る (ステップ STE 4)。パラメータ集合追加部 3 5<sub>j</sub> は、求めたパラメータを用いてパラメータ集合を生成し、新たなパラメータ集合を追加する (ステップ STE 5)。

[0081] 本願発明は、大規模データストリームが与えられたとき、その中から重要

なトレンドを発見し、様々な時系列パターン（レジーム）をモデル化し、長期的な予測を実現する。本願発明の重要な特徴は、以下のように表現することができる。本願では、自然界の生態系モデルにおけるレジームシフトの概念を拡張し、大規模時系列データストリームを適応型非線形動的システムとして表現することで、複雑な時系列パターンを柔軟に表現し、長期的なイベント予測を可能にする。

[0082] (1) 潜在的な非線形動的パターン

自然界の動的システムと同様に、実世界におけるデータストリームは、様々な潜在的要素に影響されながら時間発展していく。例えば、車両センサーデータストリームは、交通状況、天候、運転者等の要素により推移し、Webのアクセス履歴データは、ユーザの嗜好や興味に基づき時間発展する。そこで、本願では、時系列データストリームの潜在的なパターンを非線形の動的システムとして表現する。より具体的には、時系列データシーケンスを潜在的な非線形微分方程式として表現する。

[0083] (2) データストリーム上でのレジームシフト

本願では、さらに、重要な時系列パターンの変化点（データストリーム上でのレジームシフト）を自動発見する。本願では、レジームシフトの概念に基づき、時系列データストリームをモデル化する。複数の非線形モデルを用いることにより、複雑に変化する時系列パターンの全てを表現する。

[0084] (3) 階層的な構造

実際の時系列データストリームは、異なる時間発展に基づく多階層の動的システムから構成され、複雑な時系列パターンを有する。つまり、階層的な構造を伴う。本願は、階層構造に基づくモデルを使用することにより、高精度な予測を実現する。

[0085] 図7～図11を参照して、予測結果の実験例を説明する。

[0086] 図7は、身体の運動に関するモーションストリームに対するRegimeCastの解析例を示す。データセットは左右の腕と足の動きから生成され、walking（歩く）、stretch（腕のストレッチ）等の複数のモーションから構成され

る。(a)は、オリジナルデータを示す。(b)は、(100:120)-ステップ先の将来イベントを $lp=20$ 時刻ごとに出力する。(c)~(f)は、異なる4つの時刻におけるスナップショットを示す。RegimeCastは自動的かつ効果的に、stretchモーションからwalkingモーションへの移り変わりをはじめとする複数のレジームシフトを検出し、長期的かつ継続的な将来イベントの予測に成功している。

[0087] 図8は、チキンダンスに対するRegimeCastの解析例を示す。具体的には、(a)はオリジナルデータを示し、(b)は(c) beaks、wings、tail feathers、clapsの4つの代表的なダンスステップから構成されている。一般に、ダンスステップのパターンは、多階層のレジームが含まれ、より複雑な時系列パターンから構成される。そのため、予測が非常に難しい。具体的には、図6(c)~(f)に示されるように、各ステップには、いくつかの基本的な動作が含まれており、それぞれの動作が異なるテンポを持っている。RegimeCastは、複数の潜在的なレジームで構成される複雑な時系列パターンを表現し、長期的な動作の予測に成功している。特に、RegimeCastは、事前知識やステップに関する情報を使用しない。RegimeCastは、重要な時系列パターン(レジーム)を高速に発見し、新たなレジームのパラメータを時系列モデルデータベースに格納することで、柔軟なイベント予測を継続的に行うことができる。

[0088] 本願は、モーションキャプチャデータ以外に、Web情報にも応用可能である。図9はGoogleにおける3ヶ月先の各キーワードの検索量の予測結果を示している。図9から分かるように、3ヶ月先のホットピックを知ることが可能である。さらに、本発明は、環境情報(気温と気圧)、経済情報(為替と金プラチナ相場)などでも良好な予測結果が出ている。

[0089] 図10は、本願発明と従来手法の予測結果の比較を示している。従来手法としてARIMA、TBATSと比較している。図10(a)及び(b)は、それぞれ、図8の場合に、オリジナルデータと(100;120)-ステップ先の予測イベントの推定値との二乗平均誤差(RMSE)のエラー値及び平均値を示す。(a

）では、TBATSはエラー値が極めて高いために省略している。図10（c）及び（d）は、それぞれ、ARIMAとTBATSによる予測結果を示す。既存の予測手法であるARIMAとTBATSは、非線形の時系列パターンとその変化点であるレジームシフトを表現できないため、適切に予測することができないことがわかる。

[0090] 図11は、提案手法と従来手法の計算コスト及び長期的なイベント予測での予測精度と計算コストを示す。なお、DPSの効果を検証するため、時間感覚を $\delta = 1$ とした場合についても比較を行った（RegimeCast-F）。（a）及び（b）は、それぞれ、エクササイズでの計算コスト及び平均値を示す。（c）及び（d）は、それぞれ、家の掃除での計算コスト及び平均値を示す。（e）及び（f）は、ステップ数 $l_s$ を50、75、…、200のように変化させたときの予測精度と計算コストを示している。いずれの場合にも、RegimeCastは、既存手法と比較して、長期的なイベント予測に対す大幅な性能向上を達成している。

[0091] 図10及び図11が示すように、提案手法は予測精度と計算コストの両面で優れている。

### 符号の説明

[0092] 1 予測装置、3 データストリーム記憶部、5 カレントウィンドウ記憶部、7 パラメータ集合記憶部、11 カレントウィンドウ計算部、13 レジーム更新部、15 レジーム追加部、17 予測部、21 アクティビティ演算部、23 重み演算部、25 レジームシフト変数演算部、27 パラメータ更新部、29 推定イベント演算部、31 線形パラメータ推定部、33 非線形パラメータ推定部、35 パラメータ集合追加部

## 請求の範囲

- [請求項1] 時刻  $t_c$  までの時系列データ  $X$  の一部であるカレントウィンドウ  $X_c$  を用いて時刻  $t_c$  から  $l_c$  ステップ以降の一つ又は複数のイベント値を予測する予測装置であって、
- パラメータ集合記憶手段と、レジーム更新手段と、予測手段を備え、
  - 前記パラメータ集合記憶手段は、数理モデルを特定するパラメータ集合を記憶し、
  - 前記数理モデルは、非線形要素を含み、
  - 前記パラメータ集合は、前記非線形要素の係数を特定する非線形パラメータを含み、
  - 前記レジーム更新手段は、前記非線形パラメータを変更せず、前記パラメータ集合に含まれる他のパラメータの一部又は全部を更新して、カレントウィンドウ  $X_c$  の各時刻のデータと、更新後の前記パラメータ集合により特定される数理モデルを用いて得られるカレントウィンドウ  $X_c$  の各時刻に対応するイベント値  $V_c$  との違いを小さくし、
  - 前記予測手段は、更新後の前記パラメータ集合により特定される数理モデルを用いて時刻  $t_c$  から  $l_c$  ステップ以降の一つ又は複数のイベント値を予測する、予測装置。
- [請求項2] 前記パラメータ集合記憶手段は、 $c$  個 ( $c$  は自然数) のパラメータ集合  $\theta_i$  ( $i = 1, \dots, c$ ) を記憶し、
- 前記予測手段は、更新後の  $c$  個のパラメータ集合  $\theta_i$  の一部又は全部を用いてイベント値  $V_E$  を予測する、請求項 1 記載の予測装置。
- [請求項3] レジーム追加手段を備え、
- 前記レジーム追加手段は、カレントウィンドウ  $X_c$  の各時刻でのデータと、更新後の  $c$  個のパラメータ集合  $\theta_i$  を用いて得られる対応する各時刻のイベント値  $V_c$  との違いが追加条件を満たすならば、前記パラメータ集合記憶手段に、新たなパラメータ集合  $\theta_{c+1}$  を追加し、

前記レジーム更新手段は、 $c + 1$  個のパラメータ集合  $\theta_i$  ( $i = 1, \dots, c + 1$ ) に含まれる非線形パラメータ以外のパラメータの一部又は全部を更新し、

前記予測手段は、更新後の  $c + 1$  個のパラメータ集合  $\theta_i$  ( $i = 1, \dots, c + 1$ ) により特定される数理モデルの一部又は全部を用いてイベント値を予測する、請求項 2 記載の予測装置。

[請求項 4]

前記数理モデルは、線形要素を含み、

前記パラメータ集合は、前記線形要素を特定する線形パラメータを含み、

前記レジーム追加手段は、

前記非線形パラメータを変更せずに前記線形パラメータを決定し、

決定した前記線形パラメータを使用して前記非線形パラメータを決定する、請求項 1 から 3 のいずれかに記載の予測装置。

[請求項 5]

カレントウィンドウ  $X_c^{(j)}$  ( $j = 1, \dots, h$ 、 $h$  は自然数) は、 $h$  階層であり、

前記パラメータ集合は、 $h$  階層の前記カレントウィンドウ  $X_c^{(j)}$  に対応して  $h$  階層であり、

前記レジーム更新手段は、各階層において前記パラメータ集合を更新し、

前記予測手段は、各階層において予測されたイベント値から全体のイベント値を予測する、請求項 1 から 4 のいずれかに記載の予測装置。

[請求項 6]

時刻  $t_c$  までの時系列データの一部であるカレントウィンドウ  $X_c$  を用いて、数理モデルを特定するパラメータ集合の一部のパラメータを変更して新たなパラメータ集合を生産するパラメータ集合生産方法であって、

前記数理モデルは、非線形要素を含み、



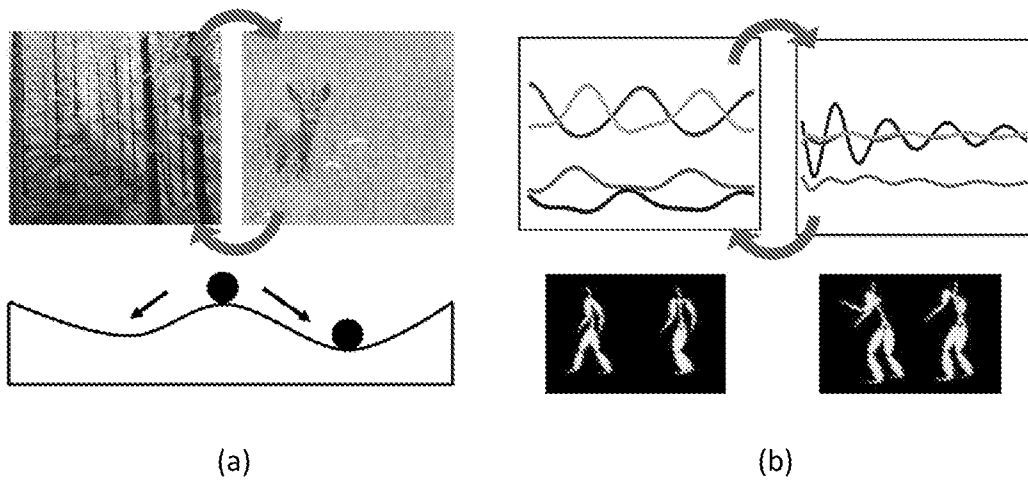
前記パラメータ集合は、前記非線形要素を特定する非線形パラメータを含み、

前記情報処理装置が備えるレジーム更新手段が、前記非線形パラメータを変更せず、前記パラメータ集合に含まれる他のパラメータの一部又は全部を更新して、カレントウィンドウ $X_c$ の各時刻のデータと、更新後の前記パラメータ集合により特定される数理モデルを用いて得られるカレントウィンドウ $X_c$ の各時刻に対応するイベント値 $V_c$ との違いを小さくする更新ステップを含むパラメータ集合生産方法。

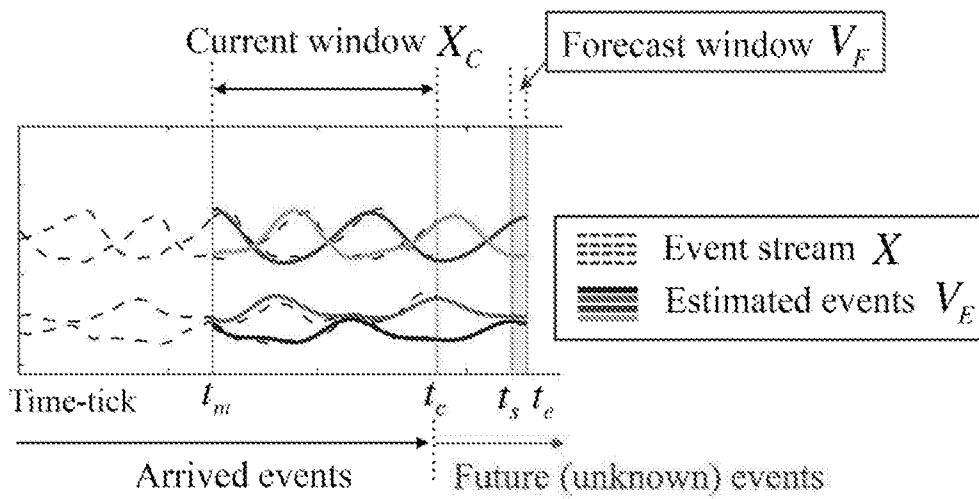
[請求項7]

コンピュータを、請求項1から5のいずれかに記載の予測装置として機能させるためのプログラム。

[図1]



[図2]



[3]

**Algorithm 1** REGIMEREADER ( $X_C, \Theta$ )

---

```

1: Input: current window  $X_C$  and current regime parameters  $\Theta$ 
2: Output: Estimated events  $V_E = V[t_m : t_e]$  and updated regimes  $\Theta$ 
3: /* (I) Individual regime estimation */
4: for  $i = 1 : c$  do
5:   /* Estimate  $s'_0$  and activity  $V_{C_i}'$  for  $i$ -th regime  $\theta_i$  */
6:    $\{\theta_i | s_0, V_{C_i}\} = \arg \min_{s'_0, V_{C_i}'} \|X_C - V_{C_i}'\|$ ; //  $V_{C_i}' = f_C(s'_0 | \theta_i)$ ;
7: end for
8: /* (II) Estimate regime activity at current time tick  $t_c$  */
9:  $w(t_c) = \arg \min_{w_1, \dots, w_c} \|X_C - \sum_{i=1}^c w_i V_{C_i}\|$ ;
10:  $r(t_c) = w(t_c) - w(t_c - 1)$ ; // Calculate regime shift variable
11:  $\mathbf{R} = \mathbf{R} \cup r(t_c)$ ;  $\Theta = \{\theta_1, \dots, \theta_c, \mathbf{R}\}$ ; // Update full parameters
12:  $V_E = f_E(\Theta)$ ; // Calculate estimated event  $V_E$ 
13: return  $\{V_E, \Theta\}$ ;

```

---

(a)

**Algorithm 2** REGIMEESTIMATOR ( $X_C$ )

---

```

1: Input: current window  $X_C$ 
2: Output: Estimated model parameter set  $\theta = \{s_0, p, \mathbf{Q}, \mathcal{A}, u, \mathbf{V}\}$ 
3: /* Estimate linear dynamical parameters  $\theta_L = \{p, \mathbf{Q}, u, \mathbf{V}\}$  */
4:  $\mathcal{A} = 0$ ; // Initialize tensor  $\mathcal{A}$ 
5:  $\{s_0, \theta_L\} = \arg \min_{s'_0, \theta'_L} \|X_C - V_C\|$ ; //  $V_C = f_C(s'_0, \theta'_L, \theta_N)$ 
6: /* Estimate non-linear dynamical parameters  $\theta_N = \{\mathcal{A}\}$  */
7:  $\{s_0, \theta_N\} = \arg \min_{s'_0, \theta'_N} \|X_C - V_C\|$ ; //  $V_C = f_C(s'_0, \theta'_L, \theta'_N)$ 
8:  $\theta = \{s_0, \theta_L, \theta_N\}$ ; // Full parameter set
9: return  $\theta = \{s_0, p, \mathbf{Q}, \mathcal{A}, u, \mathbf{V}\}$ ;

```

---

(b)

**Algorithm 3** REGIMECAST ( $x(t_c)$ )

---

```

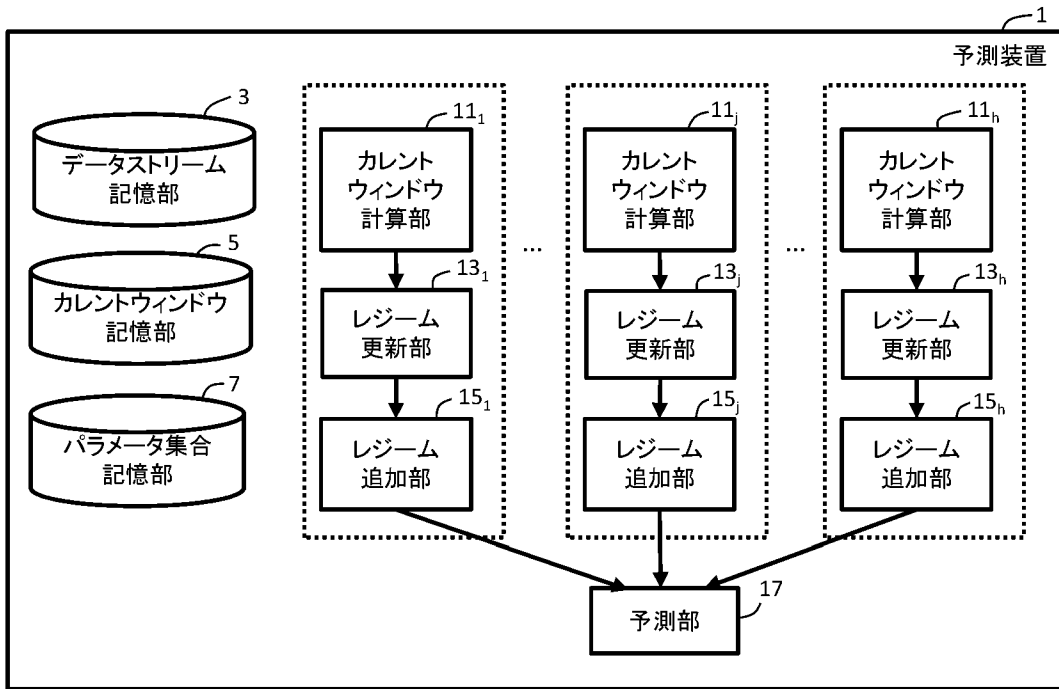
1: Input: a new event  $x(t_c)$  at time tick  $t_c$ 
2: Output:  $l_s$ -steps-ahead future events  $V_F$ 
3: /* Initialize forecast window  $V_F = 0$  */
4: for  $i = 1 : h$  do
5:   Compute  $X_C^{(i)}$ ; // current window at  $i$ -level
6:   /* (I) Parameter fitting for regime activities */
7:    $\{V_E^{(i)}, \Theta^{(i)}\} = \text{REGIMEREADER}(X_C^{(i)}, \Theta^{(i)})$ ;
8:   /* (II) Regime estimation (if required) */
9:    $V_C^{(i)} = V^{(i)}[t_m : t_e]$ ; // Estimated events from  $t_m$  to  $t_e$ 
10:  if  $\|X_C^{(i)} - V_C^{(i)}\| > \epsilon$  then
11:     $\theta = \text{REGIMEESTIMATOR}(X_C^{(i)})$ ;  $\Theta^{(i)} = \{\Theta^{(i)} \cup \theta\}$ ;
12:  end if
13: end for
14: /* (III)  $l_s$ -steps-ahead future event generation */
15:  $V_E = V_E^{(1)} + \dots + V_E^{(h)}$ ;  $V_F = V[t_s : t_e]$ ;
16: return  $V_F$ ;

```

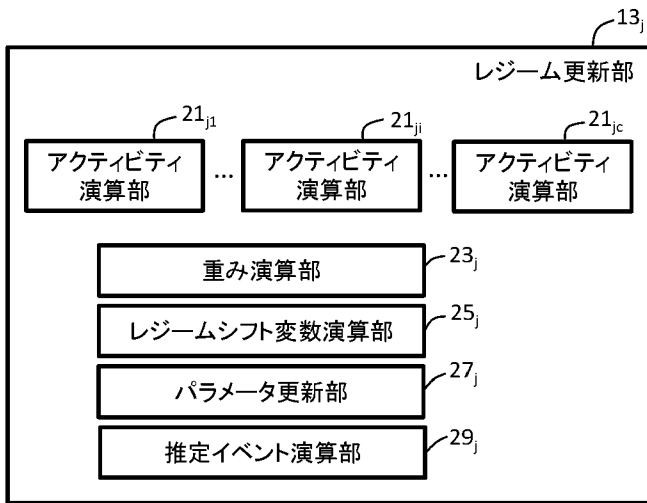
---

(c)

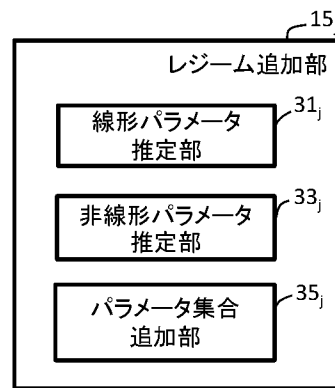
[図4]



(a)

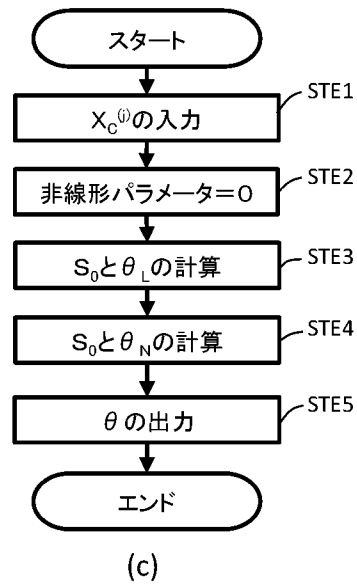
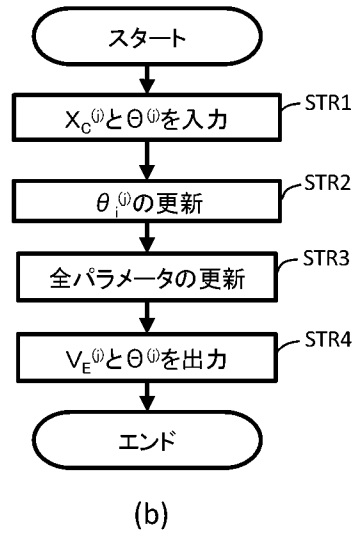
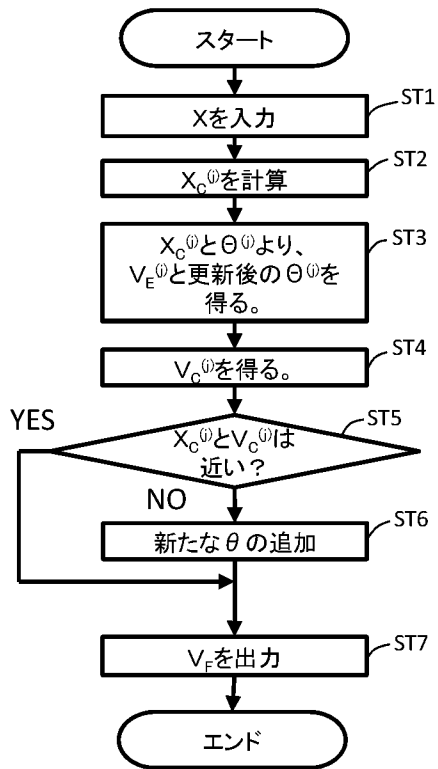


(b)

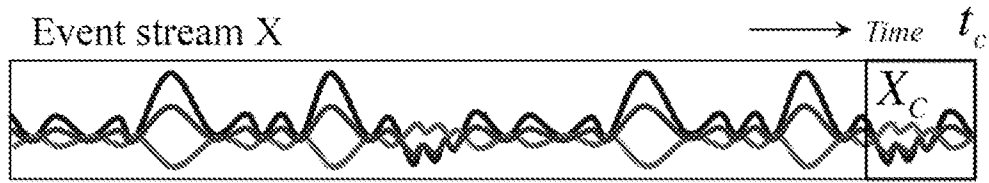


(c)

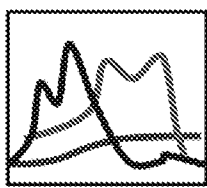
[図5]



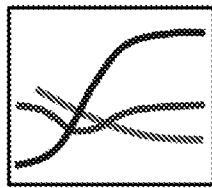
[図6]



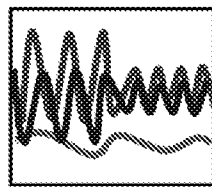
(a)

 $\theta_1^{(1)}$ 

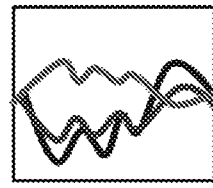
(b)

 $\theta_2^{(1)}$ 

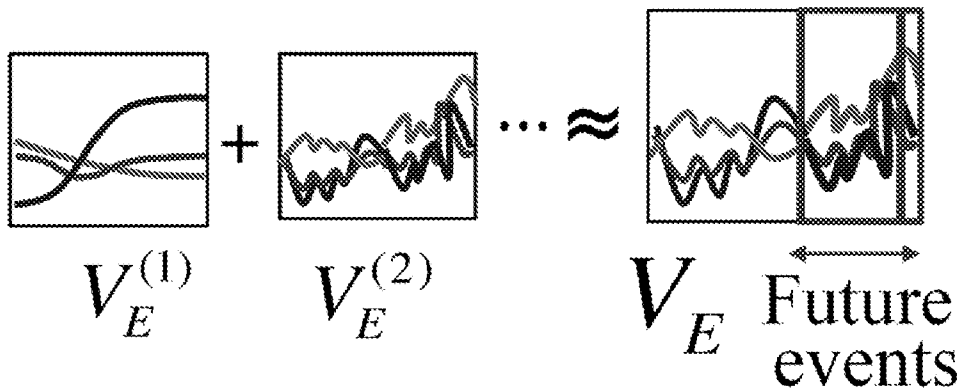
(c)

 $\theta_1^{(2)}$ 

(d)

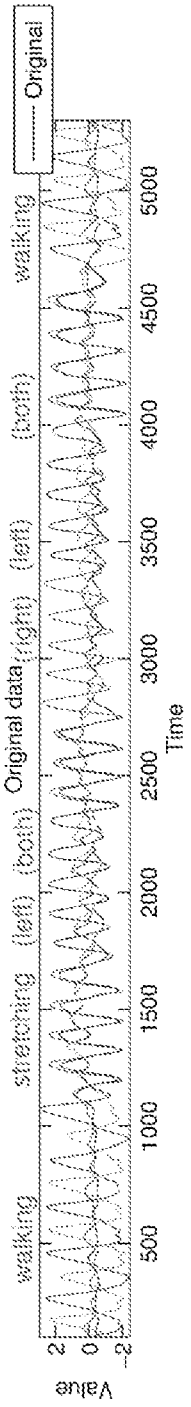
 $\theta_2^{(2)}$ 

(e)

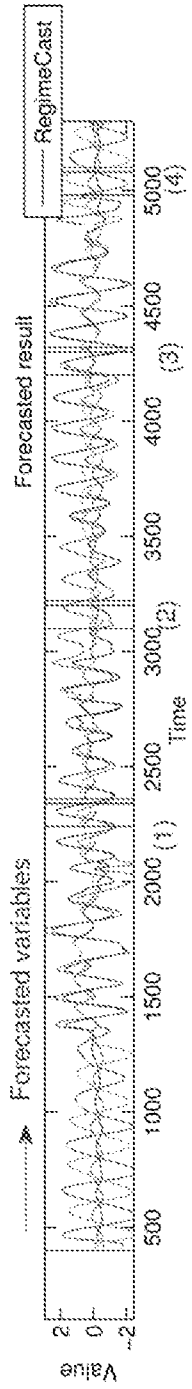


(f)

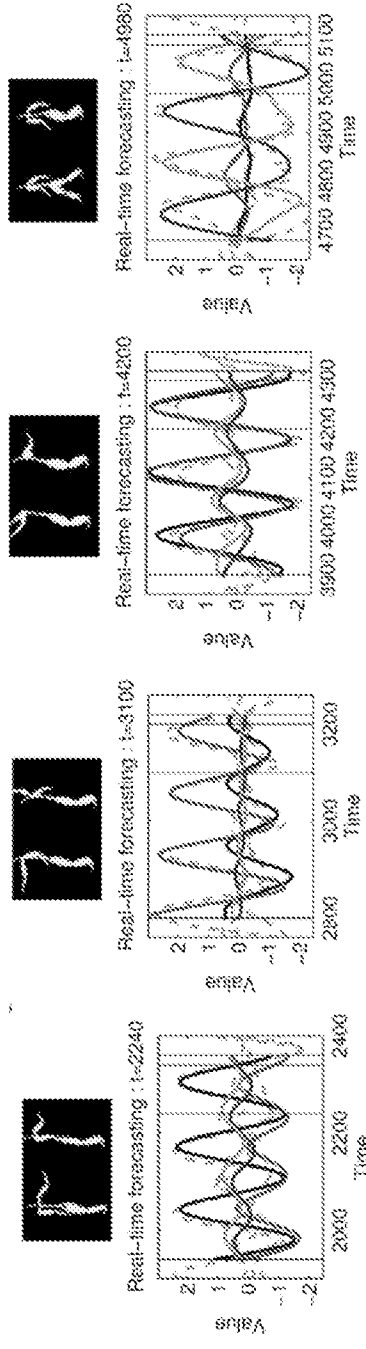
[7]



(a)



(b)



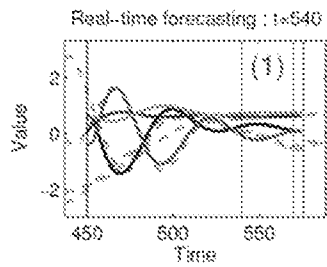
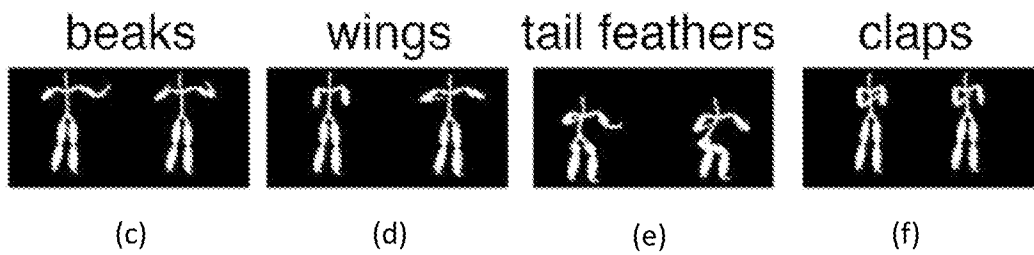
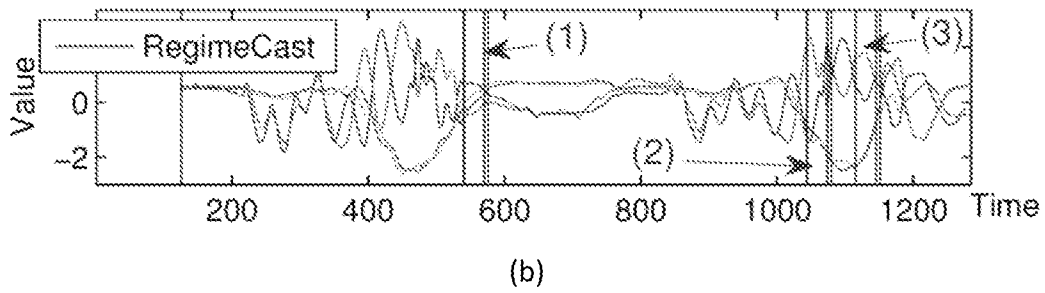
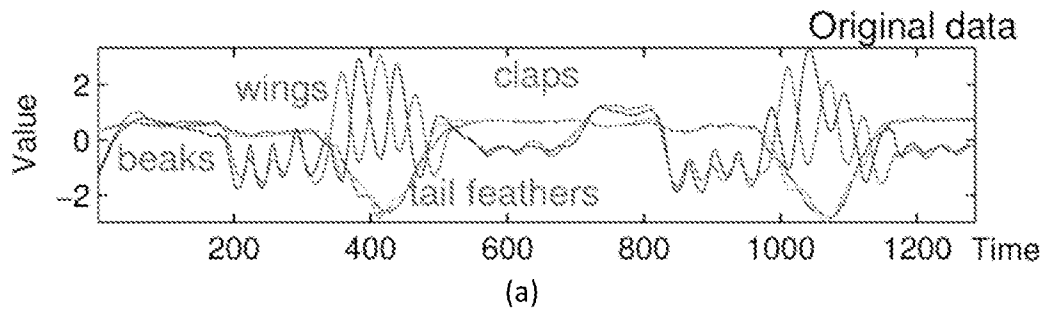
(c)

(d)

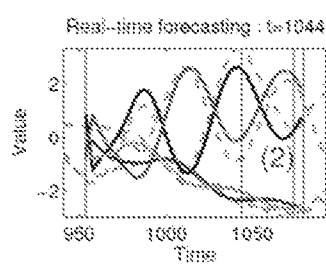
(e)

(f)

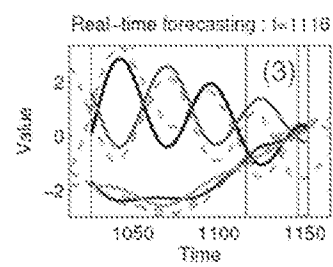
[8]



(g)



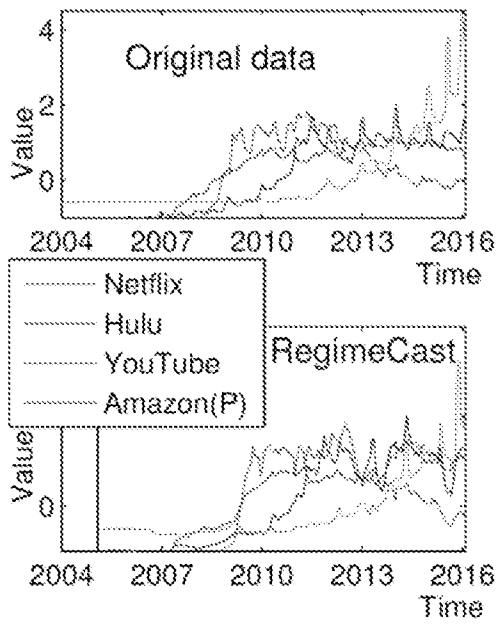
(h)



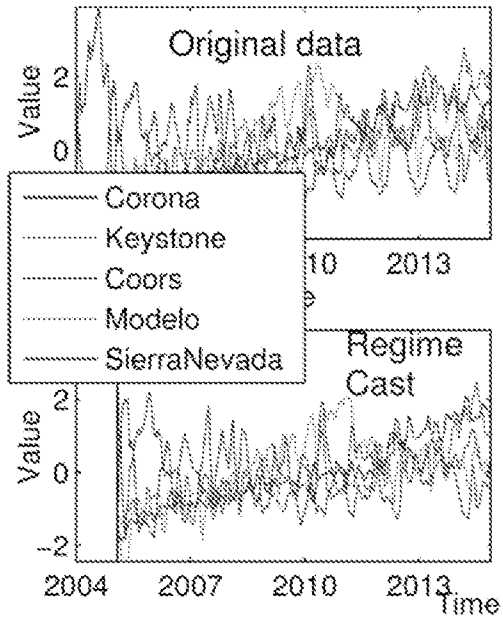
(i)



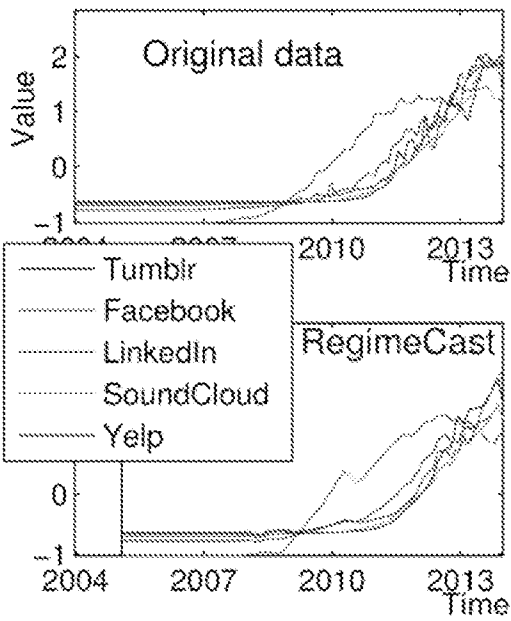
[9]



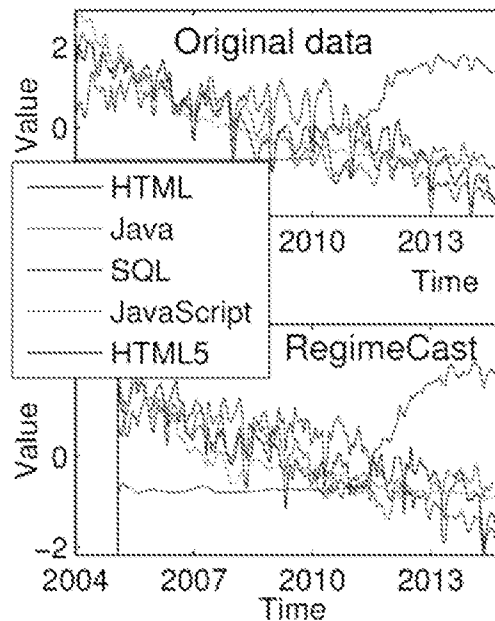
(a)



(b)

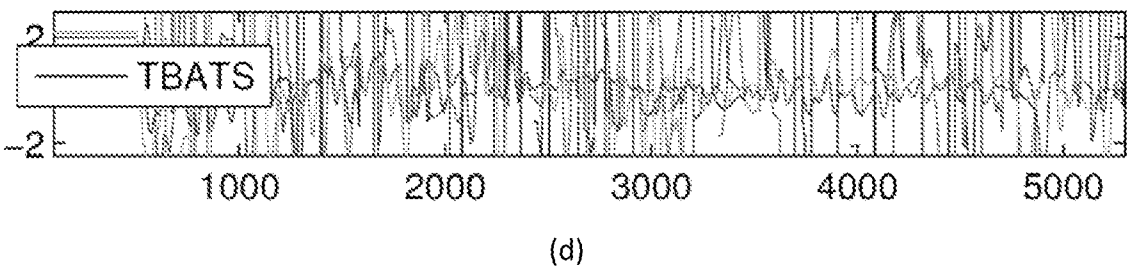
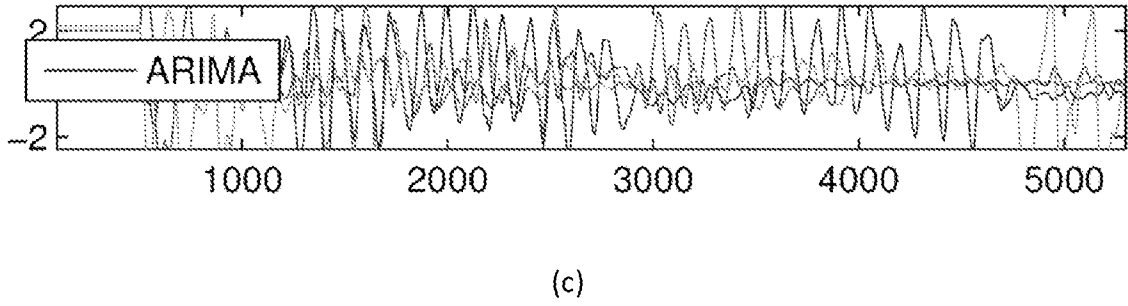
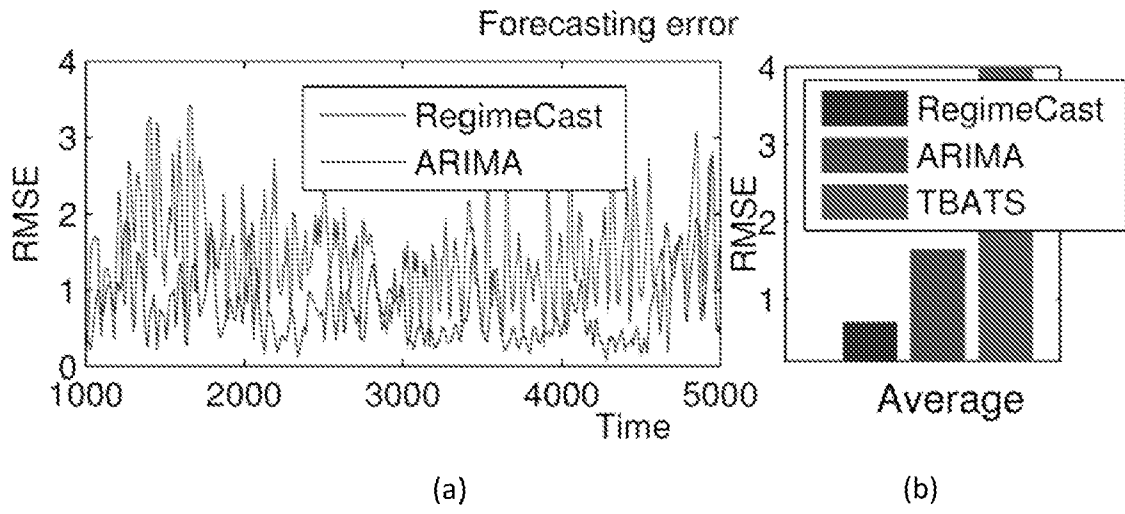


(c)

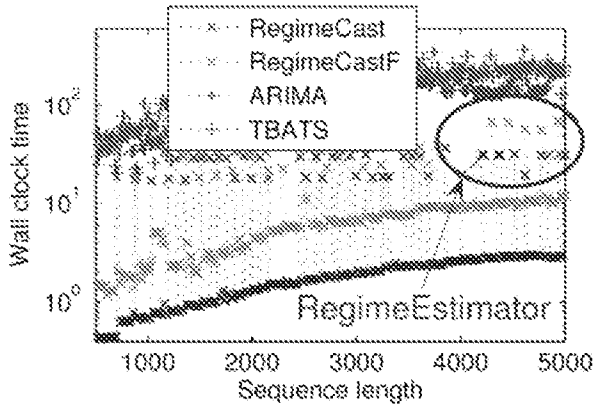


(d)

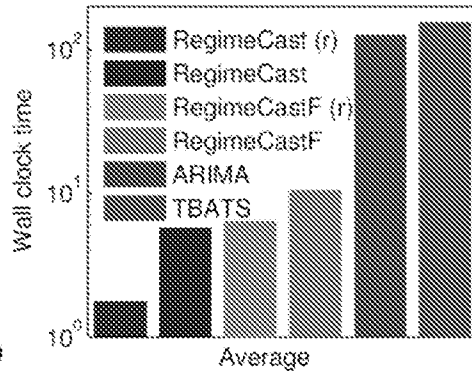
[10]



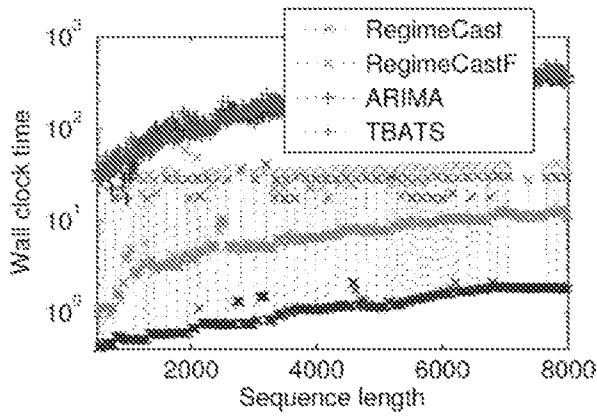
[11]



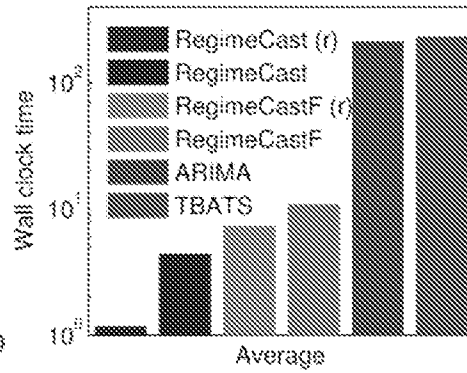
(a)



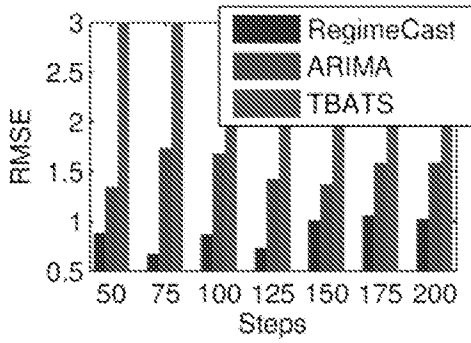
(b)



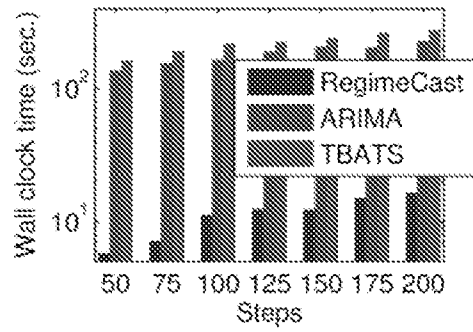
(c)



(d)



(e)



(f)

**INTERNATIONAL SEARCH REPORT**

International application No.

PCT/JP2017/025236

**A. CLASSIFICATION OF SUBJECT MATTER**  
 G06Q10/04(2012.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
 G06Q10/04

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2017
Kokai Jitsuyo Shinan Koho	1971-2017	Toroku Jitsuyo Shinan Koho	1994-2017

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2016-126718 A (Nippon Telegraph and Telephone Corp.), 11 July 2016 (11.07.2016), entire text (Family: none)	1-7
A	JP 11-134320 A (Nippon Telegraph and Telephone Corp.), 21 May 1999 (21.05.1999), entire text (Family: none)	1-7
A	JP 6-214975 A (Hitachi, Ltd.), 05 August 1994 (05.08.1994), entire text (Family: none)	1-7

Further documents are listed in the continuation of Box C.       See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 06 September 2017 (06.09.17)	Date of mailing of the international search report 19 September 2017 (19.09.17)
---	--

Name and mailing address of the ISA/ Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan	Authorized officer  Telephone No.
--	---

A. 発明の属する分野の分類 (国際特許分類 (IPC)) Int.Cl. G06Q10/04(2012.01)i											
B. 調査を行った分野 調査を行った最小限資料 (国際特許分類 (IPC)) Int.Cl. G06Q10/04											
最小限資料以外の資料で調査を行った分野に含まれるもの <table style="width:100%; border-collapse: collapse;"> <tr> <td style="width:30%;">日本国実用新案公報</td> <td>1922-1996年</td> </tr> <tr> <td>日本国公開実用新案公報</td> <td>1971-2017年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996-2017年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994-2017年</td> </tr> </table>				日本国実用新案公報	1922-1996年	日本国公開実用新案公報	1971-2017年	日本国実用新案登録公報	1996-2017年	日本国登録実用新案公報	1994-2017年
日本国実用新案公報	1922-1996年										
日本国公開実用新案公報	1971-2017年										
日本国実用新案登録公報	1996-2017年										
日本国登録実用新案公報	1994-2017年										
国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)											
C. 関連すると認められる文献											
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号									
A	JP 2016-126718 A (日本電信電話株式会社) 2016.07.11, 全文 (ファミリーなし)	1-7									
A	JP 11-134320 A (日本電信電話株式会社) 1999.05.21, 全文 (ファミリーなし)	1-7									
☑ C欄の続きにも文献が列挙されている。		☐ パテントファミリーに関する別紙を参照。									
* 引用文献のカテゴリー 「A」 特に関連のある文献ではなく、一般的技術水準を示すもの 「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの 「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す) 「O」 口頭による開示、使用、展示等に言及する文献 「P」 国際出願日前で、かつ優先権の主張の基礎となる出願		の日の後に公表された文献 「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの 「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの 「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの 「&」 同一パテントファミリー文献									
国際調査を完了した日 06.09.2017		国際調査報告の発送日 19.09.2017									
国際調査機関の名称及びあて先 日本国特許庁 (ISA/J P) 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号		特許庁審査官 (権限のある職員) 石田 信行	5M   9469								
		電話番号 03-3581-1101	内線 3599								

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 6-214975 A (株式会社日立製作所) 1994.08.05, 全文 (ファミリーなし)	1-7