

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6312630号
(P6312630)

(45) 発行日 平成30年4月18日 (2018. 4. 18)

(24) 登録日 平成30年3月30日 (2018. 3. 30)

(51) Int. Cl.	F I
GO6N 99/00 (2010.01)	GO6N 99/00 1 5 3
GO6F 11/07 (2006.01)	GO6F 11/07 1 5 1
GO6F 11/34 (2006.01)	GO6F 11/34 1 4 7

請求項の数 8 外国語出願 (全 8 頁)

(21) 出願番号	特願2015-98030 (P2015-98030)	(73) 特許権者	000006013
(22) 出願日	平成27年5月13日 (2015. 5. 13)		三菱電機株式会社
(65) 公開番号	特開2015-230727 (P2015-230727A)		東京都千代田区丸の内二丁目7番3号
(43) 公開日	平成27年12月21日 (2015. 12. 21)	(74) 代理人	100110423
審査請求日	平成30年1月15日 (2018. 1. 15)		弁理士 曾我 道治
(31) 優先権主張番号	14/296, 809	(74) 代理人	100111648
(32) 優先日	平成26年6月5日 (2014. 6. 5)		弁理士 梶並 順
(33) 優先権主張国	米国 (US)	(74) 代理人	100122437
早期審査対象出願			弁理士 大宅 一宏
		(74) 代理人	100147566
			弁理士 上田 俊一
		(74) 代理人	100161171
			弁理士 吉田 潤一郎
		(74) 代理人	100161115
			弁理士 飯野 智史

最終頁に続く

(54) 【発明の名称】 時系列データ内の異常を検出する方法

(57) 【特許請求の範囲】

【請求項 1】

機器の動作を表す時系列データ内の異常を検出する方法であって、分割統治手順を用いてトレーニング時系列データを集約することによって、最終的な標本の集合を学習するステップと、テスト時系列データの各ウィンドウについて、前記最終的な標本の集合内で最も近い標本までの距離を求めるステップであって、前記距離は、異常スコアである、ステップと、
ウィンドウの前記異常スコアが閾値よりも大きいとき、ユーザに対して異常を信号で通知するステップと、を含み、前記分割統治手順は、前記トレーニング時系列データ内の初期の標本の集合をチャンクに分割するステップと1つのチャンクのみが残るまで、選択手順および組み合わせ手順のシーケンスを前記チャンクに適用するステップと、前記残りのチャンクの前記標本に最後の選択手順を適用するステップであって、前記最終的な標本の集合を生成するステップとを含み、前記ステップは、プロセッサにおいて実行される、時系列データ内の異常を検出する方法。

【請求項 2】

各標本は、前記時系列データの類似のウィンドウの集合を表す、請求項 1 に記載の方法。

【請求項 3】

前記時系列データの各ウィンドウは、該ウィンドウ内の前記時系列データの形状を特徴付ける軌道成分および確率的成分を特徴付ける統計成分として表され、前記軌道成分および前記統計成分を含む特徴ベクトルは、統計平滑化軌道 (SST) 特徴であり、前記ウィンドウは、時間的にスライドおよび重複している、請求項 1 に記載の方法。

【請求項 4】

前記トレーニング時系列データ内の各ウィンドウについて前記 SST 特徴を求めるステップと、全ての SST 特徴の集合に等しい前記初期の標本の集合を定義するステップとをさらに含む、請求項 3 に記載の方法。

10

【請求項 5】

前記時系列データを平滑化するステップをさらに含む、請求項 1 に記載の方法。

【請求項 6】

前記統計成分は、平均、標準偏差、絶対差の平均、平均交差の数をウィンドウサイズで除算したもの、正の差分のパーセンテージ、0 の差分のパーセンテージ、および一連の正の差分を前記ウィンドウのサイズによって除算した平均サイズを含む、請求項 4 に記載の方法。

20

【請求項 7】

前記選択手順は、最近傍標本を反復的にマージし、前記標本のマージは、前記 SST 特徴の平均を計算することによって行われる、請求項 4 に記載の方法。

【請求項 8】

前記距離は、

【数 1】

$$\text{dist}(v_1, v_2) = \sum_{i=1}^l (v_1 \cdot t(i) - v_2 \cdot t(i))^2 + \frac{l}{7} \sum_{i=1}^7 (v_1 \cdot s(i) - v_2 \cdot s(i))^2$$

30

であり、ここで、 v_1 および v_2 は、2 つの特徴ベクトルであり、 $v_j \cdot t$ は、 v_j の前記軌道成分であり、 $v_j \cdot s$ は、 v_j の前記統計成分である、請求項 6 に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、包括的には、信号処理に関し、より詳細には、実数値時系列データ内の異常を検出することに関する。

【背景技術】

【0002】

時系列データは、多くの用途について生成および解析することができる。例えば、自動機器監視は、産業機器の多くのコストを要する修理を回避することができる。これは、機器上にあるセンサーまたは機器の近くにあるセンサーによって取得された時系列データを解析して、機器の保守または修理が必要とされることを示す場合がある異常を検出することによって行うことができる。

40

【0003】

したがって、1次元時系列データのモデルを効率的に学習することが所望されている。この場合、このモデルは、同じソースからのその後のテスト時系列データ内の異常を検出するのに用いることができる。通常、このモデルは、異常を有しないトレーニング時系列から学習される。

【0004】

50

時系列データのモデルを学習するための多くの方法が知られている。単純かつ効果的な方法は、トレーニング時系列データ全体をモデルとして用いる。換言すれば、トレーニング時系列全体が、モデルとして記憶される。したがって、トレーニングする時間は、無視することができる。異常を検出するために、テスト時系列の各ウィンドウが、トレーニング時系列のあらゆるウィンドウと比較され、最も近くの一一致するウィンドウまでの距離が、異常スコアとして用いられる。この異常スコアが閾値よりも大きい場合、異常が信号で通知される。これについては、非特許文献1を参照されたい。その手法の主な欠点は、非常に大きくなるおそれのある、トレーニング時系列全体を記憶することを要し、また、異常スコアの計算が遅く、これは、実時間の用途に対して、広範囲の異なるタイプの時系列データとともにその方法を用いることを妨げる場合があるということである。

10

【0005】

時系列データをモデル化するための別の部類の方法は、予測技法を用いるものである。そのような方法は、時系列データの多くの以前の値を用いて、現在の値を予測する。これについては、非特許文献2および非特許文献3を参照されたい。それらの予測モデルは、コンパクトにすることができるが、いくつかの時系列データを正確に予測しない場合がある。

【0006】

時系列データをモデル化するための別の方法は、 d 次元特徴空間を通る軌道としてのものである。 d 次元空間における区分的に線形なパスまたはボックスが、トレーニング時系列データ内の有効なパスを効率的に表すのに用いられてきた。これについては、非特許文献4を参照されたい。その学習する方法は、 $O(n \log n)$ の複雑度を有する。

20

【0007】

僅かに異なる手法が、時系列データの短いウィンドウを表すための部分空間を求め、次いで、自己回帰モデルまたは密度推定を用いて、その部分空間において軌道をモデル化する。これについては、非特許文献5を参照されたい。

【0008】

2013年7月1日にJones他によって出願された「Method for Detecting Anomalies in a Time Series Data with Trajectory and Stochastic Components」という発明の名称の特許文献1は、テスト時系列データから抽出された普遍的な特徴を、トレーニング時系列データから取得された普遍的な特徴と比較してスコアを求めることによって、時系列データ内の異常を検出する方法を記載している。これらの普遍的な特徴は、時系列データの軌道成分および時系列データの確率的成分を特徴付ける。次に、異常スコアが閾値よりも大きい場合に、異常が検出される。トレーニング時系列から普遍的な特徴(標本のタイプである)の集合を効率的に学習する方法は、その特許文献1には開示されていない。

30

【先行技術文献】

【特許文献】

【0009】

【特許文献1】米国特許出願第13/932,238号

【非特許文献】

【0010】

40

【非特許文献1】Keogh他「HOT SAX: Finding the Model Unusual Time Series Subsequence: Algorithms and Applications」ICDM 2005

【非特許文献2】Ma他「Online Novelty Detection on Temporal Sequences」SIGKDD 2003

【非特許文献3】Koskivaara「Artificial Neural Networks for Predicting Patterns in Auditing Monthly Balances」J. of the Operational Research Society, 1996

【非特許文献4】Mahoney他「Trajectory Boundary Modeling of Time Series for Anomaly Detection」Workshop on Data Mining Methods for Anomaly Detection at KDD, 2005

【非特許文献5】Liu他「Modeling Heterogeneous Time Series Dynamics to Profile Big Sensor Data in Complex Physical Systems」IEEE Conf. on Big Data, 2013

50

【発明の概要】**【発明が解決しようとする課題】****【0011】**

したがって、多くの異なるタイプの時系列データに適用することができる時系列のための正確でコンパクトなモデルを効率的に学習することが必要とされている。

【課題を解決するための手段】**【0012】**

本発明の実施の形態は、トレーニング時系列データから学習されたモデルを用いて、1次元時系列データ内の異常を効率的、かつ正確に検出する方法を提供する。

【0013】

本発明の主なアイデアは、トレーニング時系列データを標本の集合としてモデル化することである。これらの標本は、時系列データ内の様々な異なるウィンドウまたは部分シーケンスを表す。最終的な標本の集合は、トレーニング時系列データ内の重複するウィンドウの全体集合よりも大幅に小さい。

【0014】

この標本学習手順は、分割統治のアイデアを用いて、時系列データをそれよりも小さなチャンクに分割し、チャンクごとに標本の集合を学習し、その後、それらの結果を組み合わせる。結果として得られた標本学習手順は、 $O(nw)$ 時間（平均の場合）の複雑度を有する。ここで、 n は、トレーニング時系列データのサイズであり、 w は、ウィンドウのサイズである。

【発明の効果】**【0015】**

結果として得られた標本の集合は、異常検出を非常に高速にする。なぜならば、トレーニング時系列データは、相対的に少ない数の標本によって効率的に集約されているからである。さらに、この学習されたモデルによって、多種多様な時系列データのための非常に正確な異常検出方法が得られる。

【図面の簡単な説明】**【0016】**

【図1】本発明の実施の形態による時系列データ内の異常を検出する方法の流れ図である。

【図2】本発明の実施の形態による異常の集合を学習するための分割統治手順の概略図である。

【発明を実施するための形態】**【0017】****異常検出方法**

図1は、時系列データ内の異常を検出する方法を示している。この方法のステップは、当該技術分野において知られているように、バスによってメモリおよび入力/出力インターフェースに接続されたプロセッサ100において実行することができる。

【0018】

標本の集合111が、分割統治手順200を用いてトレーニング時系列データ101を集約することによって学習される。標本は、時系列データの類似のウィンドウの集合を表したものである。分割統治手順の詳細は、図2に示されている。

【0019】

テスト時系列データ102のウィンドウごとに、最も近い標本までの距離が求められる(120)。この距離は、異常スコア121として用いられる。

【0020】

次に、ウィンドウの異常スコアが閾値 T よりも大きいとき、異常131が信号で通知される(130)。

【0021】**分割統治手順**

10

20

30

40

50

図 2 に示すように、分割統治手順は、初期の標本の集合 2 0 1 をチャンク 2 0 2 に分割する。次に、選択手順および組み合わせ手順 2 2 0 および 2 3 0、2 4 0 および 2 5 0 のシーケンス 2 0 5 がチャンクに適用される。選択手順は、所与の標本の集合から、より小さな標本の集合を選択する。このより小さな標本の集合は、所与の集合を十分に表すように選ばれる。単純な選択手順の一例は、チャンク内の標本間の最小の最近傍距離が閾値 T よりも大きくなるまで、チャンク内の 2 つの最も近い標本を反復的にマージすることである。

【 0 0 2 2 】

組み合わせ手順は、2 つのチャンクからの標本の全てを単純に 1 つのチャンクにするものである。選択および組み合わせのこのプロセス 2 0 5 は、単一のチャンクが残るまで継続される。その後、最後の選択動作 2 6 0 が、最終的な標本の集合 2 0 3 を生成する。

【 0 0 2 3 】

統計平滑化軌道 (S S T : Statistical and Smoothed Trajectory) 特徴

時系列データのウィンドウは、当該ウィンドウ内の時系列データの形状を特徴付ける軌道成分および確率的成分を特徴付ける統計成分として表される。これらのウィンドウは、時間的にスライドするとともに重複する。軌道成分および統計成分を含む特徴ベクトルは、統計平滑化軌道 (S S T) 特徴と呼ばれる。

【 0 0 2 4 】

時系列データの類似のウィンドウの集合を表す S S T 特徴の集合が与えられると、この集合の標本が、軌道成分および統計成分の各要素の標準偏差とともに平均軌道成分および平均統計成分によって求められる。

【 0 0 2 5 】

トレーニング時系列データを集約する標本の集合を学習するための本発明者らの手順は、次のステップを含む。

【 0 0 2 6 】

トレーニング時系列データ内の重複するウィンドウごとに S S T 特徴ベクトルを求め、全ての S S T 特徴ベクトルの集合に等しい初期の標本の集合 2 0 1 を定義する。

【 0 0 2 7 】

類似の重複するウィンドウの標本の初期マージ 2 2 0 を行う。

【 0 0 2 8 】

分割統治学習手順 2 0 0 を用いて初期の標本の集合をさらに圧縮する。

【 0 0 2 9 】

これらのステップのそれぞれを、以下でより詳細に説明する。

【 0 0 3 0 】

特徴ベクトル

初期の S S T 特徴の集合に等しい初期の標本の集合の計算は、簡単である。スライディングウィンドウ 1 0 5 が用いられ、各ウィンドウ内で、軌道成分および統計成分が求められる。軌道成分は、ウィンドウ内の時系列値の単純な固定ウィンドウ移動平均値を用いて求められ、ウィンドウの平均を減算した後、平滑化時系列データが得られる。

【 0 0 3 1 】

平滑化に起因して、平滑化時系列データ内の値の半分は、重要な情報を失うことなく廃棄することができる。これによって、 $w / 2$ 個の要素を有する軌道成分が得られる。統計成分は、初期の時系列ウィンドウ内の確率的情報を特徴付けるように設計されたウィンドウ内の時系列データ値にわたって求められる統計の小さな集合である。

【 0 0 3 2 】

統計成分は、平均、標準偏差、絶対差の平均、平均交差の数をウィンドウサイズで除算したもの、正の差分のパーセンテージ、0 の差分のパーセンテージ、および一連の正の差分をウィンドウサイズによって除算した平均サイズを含む。

【 0 0 3 3 】

軌道成分および統計成分を単一の特徴ベクトルに組み合わせることによって、 $w / 2 +$

10

20

30

40

50

7個の実数値からなるベクトルが得られる。初期の標本の集合は、標準偏差成分が全て0に設定された初期のSST特徴ベクトルの集合に等しい集合である。

【0034】

重複する部分シーケンスは、多くの場合、シーケンス間の距離が小さいので、類似の重複するシーケンスのSST特徴ベクトルをマージすることによって、初期標本集合のサイズが削減される。2つのSST特徴ベクトルをマージするために、これらの2つのベクトルの加重平均がとられる。各特徴ベクトルの重みは、その特徴ベクトルに既に平均されている特徴ベクトルの数である。

【0035】

初期マージ210の後、分割統治手順200を用いて、最終的な標本の集合203が効率的に選択される。この分割統治手順は、初期の標本の集合201をサイズCのチャンクに分割する(210)。ここで、Cは、この手順のパラメータである。チャンクごとに、選択手順が実行される。1つの実施の形態では、各チャンクに対して実行されるこの選択手順は、反復的なマージ手順である。この反復的なマージ手順は、まず、チャンク内の標本ごとに、チャンク内の最近傍標本を見つける。この手順は、以下の距離関数を用いる。

【0036】

【数1】

$$\text{dist}(v_1, v_2) = \sum_{i=1}^l (v_1 \cdot t(i) - v_2 \cdot t(i))^2 + \frac{l}{7} \sum_{i=1}^7 (v_1 \cdot s(i) - v_2 \cdot s(i))^2 \quad (1)$$

【0037】

ここで、 v_1 および v_2 は、2つの特徴ベクトルであり、 $v_j \cdot t$ は、 v_j の軌道成分であり、 $v_j \cdot s$ は、 v_j の統計成分である。距離関数の計算は、部分和がそれまで最近傍の距離よりも既に大きい場合には早期に計算を終了することを意味する早期放棄を用いる。標本間の距離が最小の2つの標本は、マージされる。2つの標本をマージすることは、それらの2つの標本の加重平均を計算することからなる。ここで、標本の重みは、その標本に既に平均されている標本の数に等しい。チャンク内のマージされた標本の最近傍が決定され、上記2つの標本のうちの一方を最近傍として有する他のいずれの標本も、それらの最近傍が再計算される。

【0038】

2つの最も近い標本をマージするこのプロセスは、チャンク内の残りの全ての標本の最小の最近傍距離が閾値Tよりも大きくなるまで継続される。その時点で、選択手順は、終了し、残りの標本の集合をチャンクにするために返す。

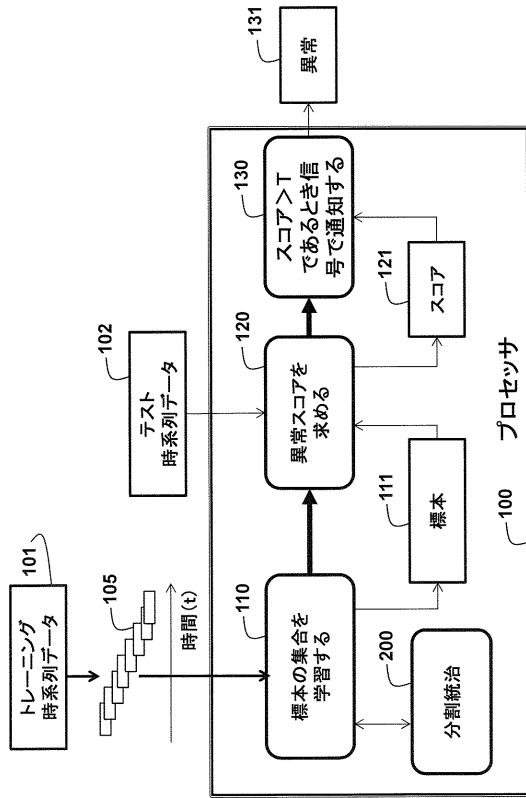
【0039】

別の実施の形態では、各チャンクに対して実行される選択手順は、k平均クラスタリング等の従来のクラスタリングアルゴリズムである。

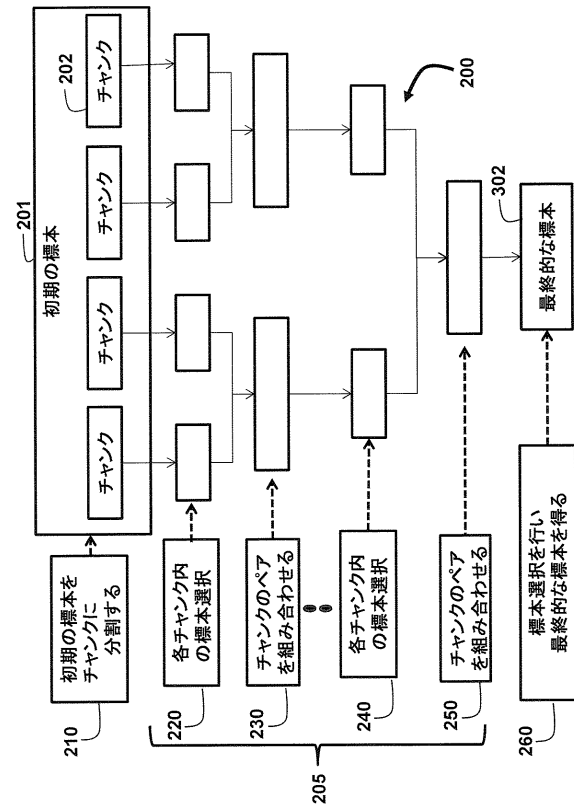
【0040】

選択手順が各チャンクに適用された後、チャンクのペアが、それらの2つのチャンクからの全ての標本を単純に1つの新たなチャンクにすることによって、1つの新たなチャンクに組み合わされる。次に、選択手順は、それらの新たなチャンクのそれぞれに対して実行され、次いで、これらのチャンクは、組み合わされ、残りのチャンクが1つだけになるまで、以下同様に行われる。最終的な選択手順が、残りの1つのチャンクに対して実行され、全体のトレーニング集合のための最終的な標本の集合が得られる。これは、図2に示すような階層的な標本学習手順によって返される最終的な標本の集合である。

【図 1】



【図 2】



フロントページの続き

- (72)発明者 マイケル・ジェイ・ジョーンズ
アメリカ合衆国、マサチューセッツ州、ケンブリッジ、ブロードウェイ 201、ケアオブ・ミツ
ビシ・エレクトリック・リサーチ・ラボラトリーズ・インコーポレイテッド
- (72)発明者 ダニエル・ニコラエフ・ニコヴスキ
アメリカ合衆国、マサチューセッツ州、ケンブリッジ、ブロードウェイ 201、ケアオブ・ミツ
ビシ・エレクトリック・リサーチ・ラボラトリーズ・インコーポレイテッド

審査官 多胡 滋

- (56)参考文献 特開2006-236280(JP,A)

- (58)調査した分野(Int.Cl., DB名)

G06N 99/00
G06F 11/07
G06F 11/34