

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6871877号
(P6871877)

(45) 発行日 令和3年5月19日(2021.5.19)

(24) 登録日 令和3年4月20日(2021.4.20)

(51) Int.Cl.

G 0 5 B 23/02 (2006.01)

F I

G 0 5 B 23/02 3 0 2 R

請求項の数 18 (全 23 頁)

(21) 出願番号	特願2018-286 (P2018-286)	(73) 特許権者	000003078
(22) 出願日	平成30年1月4日(2018.1.4)		株式会社東芝
(65) 公開番号	特開2019-121162 (P2019-121162A)		東京都港区芝浦一丁目1番1号
(43) 公開日	令和1年7月22日(2019.7.22)	(74) 代理人	100091982
審査請求日	令和1年8月19日(2019.8.19)		弁理士 永井 浩之
		(74) 代理人	100091487
			弁理士 中村 行孝
		(74) 代理人	100082991
			弁理士 佐藤 泰和
		(74) 代理人	100105153
			弁理士 朝倉 悟
		(74) 代理人	100107582
			弁理士 関根 毅
		(74) 代理人	100118876
			弁理士 鈴木 順生

最終頁に続く

(54) 【発明の名称】 情報処理装置、情報処理方法及びコンピュータプログラム

(57) 【特許請求の範囲】

【請求項1】

状態の検知対象となるコンテンツ変数と、前記コンテンツ変数が得られた条件を示すコンテキスト変数とを含む基準データと、

前記基準データと異なる期間に取得され、前記コンテンツ変数と前記コンテキスト変数とを含む判定対象データとに基づいて、

前記基準データと前記判定対象データ間の前記コンテキスト変数の識別寄与度を算出し、前記識別寄与度に基づき、前記コンテキスト変数から前記判定対象データの状態検知を行うときに使用するコンテキスト変数を選択する、変数選択部

を備え、

前記変数選択部は、前記識別寄与度が第1しきい値以下となった前記コンテキスト変数を、前記状態検知で使用する前記コンテキスト変数に選択する
情報処理装置。

【請求項2】

前記基準データは、前記検知対象が正常状態にある期間において得られたデータである

、

請求項1に記載の情報処理装置。

【請求項3】

前記変数選択部は、前記状態検知に使用される前記コンテキスト変数と、前記基準データに含まれる前記コンテンツ変数と、前記判定対象データに含まれる前記コンテンツ変数

とに基づいて、前記基準データと前記判定対象データ間の前記コンテンツ変数の識別寄与度を算出し、前記コンテンツ変数の識別寄与度に基づき、前記コンテンツ変数から前記状態検知で使用するコンテンツ変数を選択する、

請求項1に記載の情報処理装置。

【請求項4】

前記変数選択部は、前記コンテンツ変数の識別寄与度が第2しきい値以上となった前記コンテンツ変数を、前記状態検知で使用する前記コンテンツ変数に選択する、

請求項3に記載の情報処理装置。

【請求項5】

前記変数選択部は、アンサンブル学習によって前記コンテキスト変数の識別寄与度を算出する、

請求項1に記載の情報処理装置。

【請求項6】

前記アンサンブル学習はランダムフォレストである、

請求項5に記載の情報処理装置。

【請求項7】

前記識別寄与度はPermutation importanceまたはGini Importanceのいずれかである、

請求項1に記載の情報処理装置。

【請求項8】

前記変数選択部は、前記基準データと前記判定対象データにおける同一の前記コンテキスト変数について統計的検定を行い、有意差がない同一の前記コンテキスト変数を、前記状態検知で使用する前記コンテキスト変数に選択する、

請求項1～7のいずれか一項に記載の情報処理装置。

【請求項9】

前記統計的検定はノンパラメトリックな検定手法である、

請求項8に記載の情報処理装置。

【請求項10】

前記ノンパラメトリックな検定手法は、マン・ホイットニーのU検定である、

請求項9に記載の情報処理装置。

【請求項11】

前記基準データと前記判定対象データに含まれる変数について、前記コンテンツ変数と前記コンテキスト変数への分類を実行する、収集部を備えた

請求項1～10のいずれか一項に記載の情報処理装置。

【請求項12】

前記変数選択部が選択した、前記コンテキスト変数および前記コンテンツ変数を含み、前記変数選択部が選択しなかった、前記コンテキスト変数および前記コンテンツ変数が除外された前記基準データに基づき、前記変数選択部が選択した前記コンテキスト変数および前記コンテンツ変数を含み、前記変数選択部が選択しなかった前記コンテキスト変数および前記コンテンツ変数が除外された前記判定対象データの状態検知を実行する、状態検知部を備えた

請求項3又は4に記載の情報処理装置。

【請求項13】

前記状態検知部による状態検知結果を表示する、表示部を備えた

請求項12に記載の情報処理装置。

【請求項14】

前記表示部は、少なくとも前記変数選択部が選択しなかった前記コンテキスト変数または前記変数選択部が選択しなかった前記コンテンツ変数のいずれかを表示する、

請求項13に記載の情報処理装置。

【請求項15】

10

20

30

40

50

前記状態は、異常状態である

請求項 1 ～ 1 4 のいずれか一項に記載の情報処理装置。

【請求項 1 6】

前記選択部は、前記識別寄与度が第 1 しきい値以下の低い識別寄与度となった前記コンテキスト変数を前記コンテキスト変数に選択する

請求項 1 ～ 1 5 のいずれか一項に記載の情報処理装置。

【請求項 1 7】

状態の検知対象となるコンテンツ変数と、前記コンテンツ変数が得られた条件を示すコンテキスト変数とを含む基準データに含まれる前記コンテキスト変数と、

前記基準データと異なる期間に取得され、前記コンテンツ変数と前記コンテキスト変数とを含む判定対象データに含まれる前記コンテキスト変数とに基づいて、

前記基準データ及び前記判定対象データ間の前記コンテキスト変数の識別寄与度を算出し、前記識別寄与度に基づき、前記コンテキスト変数から前記判定対象データの状態検知を行うときに使用するコンテキスト変数を選択するステップ、

を含み、

前記ステップは、前記識別寄与度が第 1 しきい値以下となった前記コンテキスト変数を、前記状態検知で使用する前記コンテキスト変数に選択する

情報処理方法。

【請求項 1 8】

状態の検知対象となるコンテンツ変数と、前記コンテンツ変数が得られた条件を示すコンテキスト変数とを含む基準データに含まれる前記コンテキスト変数と、

前記基準データと異なる期間に取得され、前記コンテンツ変数と前記コンテキスト変数とを含む判定対象データに含まれる前記コンテキスト変数とに基づいて、

前記基準データ及び前記判定対象データ間の前記コンテキスト変数の識別寄与度を算出し、前記識別寄与度に基づき、前記コンテキスト変数から前記判定対象データの状態検知を行うときに使用するコンテキスト変数を選択するステップ、

をコンピュータに実行させ、

前記ステップは、前記識別寄与度が第 1 しきい値以下となった前記コンテキスト変数を、前記状態検知で使用する前記コンテキスト変数に選択する

コンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明の実施形態は、監視装置、監視方法および監視プログラムに関する。

【背景技術】

【0002】

異常の検知が行われる変数と、計測値が得られた条件、状況や背景など（context：コンテキスト）を表現する変数を区別する、文脈的異常検知（contextual anomaly detection）と呼ばれる手法を使うと、高い精度の異常検知を行うことができる。正常／異常属性がラベル付けされ、十分なサンプル数を含む教師データがあれば、異常検知に寄与する変数の選択を行うことができる。

【0003】

しかし、過去の計測値が多く得られていない、または異常発生率が低いなどの理由で正常状態におけるデータしか用意できていない場合がある。このような場合、異常検知への寄与度を基準に変数を行うことは困難である。正常状態におけるデータしかない状況でも、文脈的異常検知の検知精度を高める技術の開発が求められている。

【先行技術文献】

【非特許文献】

【0004】

【非特許文献 1】M. A. Hayes and M. A. M. Capretz

10

20

30

40

50

, “Contextual Anomaly Detection in Big Sensor Data,” 2014 IEEE International Congress on Big Data, Anchorage, AK, 2014, pp. 64–71.

【非特許文献2】X. Song, M. Wu, C. Jermaine and S. Ranka, “Conditional Anomaly Detection,” in IEEE Transactions on Knowledge and Data Engineering, vol. 19, no. 5, pp. 631–645, May 2007.

【発明の概要】

10

【発明が解決しようとする課題】

【0005】

本発明の実施形態は、高い精度で文脈的異常検知を行う監視装置、監視方法および監視プログラムを提供する。

【課題を解決するための手段】

【0006】

本発明の実施形態としての監視装置は、異常の検知対象となるコンテンツ変数と、前記コンテンツ変数が得られた条件を示すコンテキスト変数に分類された入力データについて、第1期間に得られた基準データと、第2期間に得られた判定対象データに含まれる前記コンテキスト変数の値に基づき、前記基準データを正常データとして前記判定対象データの異常検知を行うときに使用する変数から除外される、前記コンテキスト変数を決定する、変数選択部を備える。

20

【図面の簡単な説明】

【0007】

【図1】第1の実施形態に係るシステム全体の構成例を示す図。

【図2】コンテキストデータの一例を示した図。

【図3】コンテンツデータの一例を示した図。

【図4】コンテキストデータの一例を示した図。

【図5】コンテンツデータの一例を示した図。

【図6】データの事前処理結果および文脈的異常検知結果の例を示した図。

30

【図7】監視装置に係るハードウェア構成の例を示す図。

【図8】監視装置による処理の例を示したフローチャート。

【発明を実施するための形態】

【0008】

以下、図面を参照しながら、本発明の実施形態について説明する。また、図面において同一の構成要素は、同じ番号を付し、説明は、適宜省略する。

【0009】

(第1の実施形態)

図1は、第1の実施形態に係る監視装置の構成例を示す図である。

【0010】

40

最初に本実施形態に係る監視装置の概要について説明する。図1には、監視装置1が示されている。監視装置1は、データに基づき文脈的異常検知を行うことができる。以下では、データとして物理的なセンサから得られた計測値を使った場合を例に説明するが、適用対象とされるデータはマーケティングデータ、経済指標、情報システムのログデータ、臨床データなどであってもよく、種類については特に問わない。

【0011】

監視装置1は、データに含まれる変数をコンテンツ変数と、コンテキスト変数に分類し、文脈的異常検知を行う。コンテンツ変数（内容変数）とは、異常の検知対象となる変数のことをいう。コンテキスト変数（文脈変数）とは、コンテンツ変数が得られた条件、状況、背景など（context：コンテキスト）を示す変数のことをいう。

50

【0012】

監視装置1は、文脈的異常検知の精度を高めるために、データ中のすべての変数を異常検知に使わず、一部の変数のみを選択することができる。変数選択は、異常検知をする前の前処理において実行される。コンテキスト変数のみについて変数選択を行うこともできるし、コンテキスト変数とコンテンツ変数の両方について変数選択を行ってもよい。

【0013】

監視装置1は、変数選択を行う前に、監視対象が正常状態にあるときの期間に得られたデータであると推定される基準データと、異常検知の対象期間に得られた判定対象データを用意する。基準データと判定対象データは、いずれもコンテンツ変数とコンテキスト変数を含む。

10

【0014】

基準データと判定対象データの間の識別に寄与する程度に基づき、異常検知に使用する変数を選ぶ。以降では、変数の識別に寄与する程度を、識別寄与度とよぶ。識別寄与度は、ランダムフォレストなどの識別器を使って計算する。また、変数の選択は統計的検定などを使って行ってもよい。コンテキスト変数の変数選択を行う場合、識別寄与度が小さい変数を選択する。コンテンツ変数の変数選択を行う場合、識別寄与度が大きい変数を選択する。

【0015】

最後に、監視装置1は選択された（除外されなかった）変数に基づき、文脈的異常検知を行い、結果をディスプレイなどに表示する。これにより、異常検知結果の検証や装置などの保守点検作業などを行うことができる。

20

【0016】

次に、監視装置1の各構成要素について説明する。

【0017】

図1の監視装置1は、収集部2と、コンテキストデータベース（コンテキストDB）3と、コンテンツデータベース（コンテンツDB）4と、変数選択部5と、異常検知部6と、表示部7とを備えている。コンテキストデータベース3は、基準データ3aと、判定対象データ3bとを含む。コンテンツデータベース4は、基準データ4aと、判定対象データ4bとを含む。

【0018】

30

収集部2は、外部の装置またはシステムなどからデータを収集し、コンテキストデータベース3またはコンテンツデータベース4に保存する。収集部2は、電気通信回線10を介して外部の装置やシステムに接続されている。電気通信回線10は、有線の通信媒体であってもよいし、無線であってもよい。有線の通信媒体の例としては、光ファイバ、LANケーブル、電話回線、同軸ケーブルなどがあるが、種類については特に問わない。収集部2が用いる通信規格の例としては、Ethernet、無線LAN、PCI Express、USB、UART、SPI、SDIO、シリアルポート、Bluetoothなどがあるが、その他の規格であってもよい。

【0019】

外部の装置の例としては、空調機器、製造装置、発電所、各種の機械、移動体、電子機器、観測機器、情報通信端末などが挙げられるが、どのような装置であってもよい。外部の装置から収集されるデータの例としては、センサの計測値があるが、その他の種類のデータであってもよい。センサの計測値は物理量であってもよいし、設定値やステータスなどであってもよい。外部のシステムとしては、センサネットワーク、データベースサーバ、ウェブサーバ、ウェブサービスなどがあるが、どのようなシステムであってもよい。

40

【0020】

収集部2は、収集したデータに含まれるコンテキスト変数をコンテキストデータベース3に保存する。また、収集部2は、収集したデータに含まれるコンテンツ変数をコンテンツデータベース4に保存する。あらかじめ外部の装置で変数に対しコンテキスト変数またはコンテンツ変数のラベル付けを行ってもよい。

50

【0021】

また、収集部2で変数をコンテキスト変数またはコンテンツ変数に分類してもよい。変数の分類は、センサに付与された名称やデータ系列名などをパターンマッチングや自然言語処理などによって解析して行ってもよい。例えば、名称に“config”、“status”などの文字列を含む変数は装置の設定や状態に関係している可能性が高いため、コンテキスト変数に分類することができる。各センサに関するメタ情報を解析し、変数の分類を行ってもよい。メタ情報の形式は特に問わない。

【0022】

また、特定の取得先のデータがいずれかの種類の変数のみを含む場合には、データの取得先のアドレスや識別子に基づき分類を行ってもよい。例えば、空調機器の異常検知において、外気温や天気は空調機器の動作条件を示す変数に該当する。条件を示す変数はコンテキスト変数であるため、気象情報を提供する、観測機関のサーバのドメイン名やIPアドレスから得られた変数をすべてコンテキスト変数に分類することができる。変数の分類に係る設定の一部またはすべてを手動で行ってもよい。センサやデータ系列にセンサIDなどの一意的な識別子が付与されていれば、当該識別子に基づき分類設定を行うことができる。

10

【0023】

図2～図5は、収集部2により収集されたデータの例を示している。図2、図3はある電子機器から取得されたデータである。図2、図3のテーブルには、計測時刻ごとに各変数がある値が格納されている。図2、図3の計測時刻は「時間：分」形式で示されているが、年、月、秒、ミリ秒などの情報を含んでもよい。図2、図3にはそれぞれn個の時刻における変数の値が格納されている。図2、図3のテーブルにおけるそれぞれの行は、計測時刻が同一である変数をまとめたものになっている。以降では、同一時刻に得られた変数をまとめたものをレコードと呼ぶものとする。図2、図3のテーブルはそれぞれn個のレコードを含んでいる。

20

【0024】

図2は電子機器のステータス情報に係るセンサの値である。図3は、電子機器の物理センサの計測値である。図2のテーブルは、スイッチAの操作状態や回路ブロックBの電源状態などを含んでおり、コンテキストデータに相当する。図2のコンテキストデータは、m個のコンテキスト変数を含んでいる。一方、図3のテーブルは電子機器の異常動作を判定するときに使われる電圧Cや電圧Dの計測値を含んでいるため、コンテンツデータに相当する。図3のコンテンツデータは、1（小文字のL）個のコンテンツ変数を含んでいる。

30

【0025】

図3のコンテンツデータでは、時刻0：01に電圧Cが2.0Vと他の時刻に比べ高くなっている。しかし、図2のコンテキストデータを参照すると、同じ時刻0：01にスイッチAが押下されたことがわかる。したがって、時刻0：01で検出された電圧変動は電子機器の異常動作ではなく、スイッチの操作に起因していると推定される。

【0026】

また、図3のコンテンツデータでは、時刻0：04に電流Dが25mA、時刻0：05に電流Dが23mAとなっており、電流値が7～8mAである他の時刻に比べて高くなっている。しかし、図2のコンテキストデータを参照すると、時刻0：04、0：05に他の時刻ではOFFになっていた回路ブロックBがONになっていることがわかる。したがって、電流の増加は電子機器の異常動作ではなく、回路ブロックBがONにされ、回路のインピーダンスなどが変化したために起こったと推定される。

40

【0027】

図2、図3のようにコンテキスト変数とコンテンツ変数に分類されたデータを使うと電子機器の文脈的異常検知を行うことができる。

【0028】

図4、図5はあるビルの空調機器を監視するために収集されたデータである。図4、図

50

5のテーブルには、計測時刻ごとに変数の値が格納されている。図2、図3の計測時刻は「時間：分」の形式で示されているが、年、月、秒、ミリ秒などの情報を含んでいてもよい。図4、図5のテーブルは、それぞれn個のレコードを含んでいる。

【0029】

図4のテーブルは、空調機器の設定温度と、ビルの近隣にある観測所における外気温と天気などを格納している。これらの変数は、空調機器の設定や動作条件を示しているため、図4はコンテキストデータに相当する。図4には示されていないが、自動運転、冷房／暖房、除湿など、動作モードの設定情報もコンテキスト変数に分類される。

【0030】

図5のテーブルは、空調機器が設置されている部屋の室内温度と、空調機器の消費電力などが格納されている。室内温度から空調機器の冷暖房効果を確認することができる。空調機器の消費電力から、空調機器の稼働状態を確認することができる。また、室内温度と消費電力の比較を行うことも可能である。例えば、冷房運転時において空調機器の消費電力が大きいものにも関わらず、室内温度が下がらない場合には、空調機器の故障を疑うことができる。図5は、値に基づき異常を検知可能な変数を格納しており、図5のテーブルはコンテンツデータであることがわかる。

10

【0031】

図5のコンテンツデータでは、時刻10：00に310Wであった消費電力が、時刻11：00において350Wに増加している。一方、図4のコンテキストデータを参照すると、時刻10：00に摂氏27度であった設定温度が時刻11：00に摂氏25度に下げられていることがわかる。したがって、時刻11：00における消費電力の増加は機器の異常ではなく、冷房の設定温度が下げられたことに起因していると推定される。

20

【0032】

また、時刻15：00では330Wであった消費電力が、時刻16：00に290Wに減少している。大幅な消費電力の減少がある一方、時刻15：00に摂氏28.2度であった室内温度が時刻16：00で摂氏27.4度に低下している。時刻17：00では消費電力285Wに対し、室内温度が摂氏26.9度であり、消費電力が減っているのにも関わらず、さらに室内温度が低下している。冷房運転中の空調機器の消費電力は室内温度と負の相関関係にあるため、図5のデータでは一般に期待される変数のふるまいとは異なる傾向がみられる。

30

【0033】

図4のコンテキストデータを参照すると、時刻14：00に天気が晴れから曇りに変わっている。さらに時刻15：00に天気が曇りから雨になっている。外気温は、時刻14：00の摂氏35.4度から摂氏27.9度、摂氏26.1度、摂氏24.7度と、大幅な気温低下が認められる。ビルの周辺の気象は、昼過ぎまで猛暑であったが、午後に寒冷前線を伴う低気圧が接近し、冷たい空気の中に激しい雨が降り出したことがわかる。ビルの外の気象が変化した一方、冷房の設定温度は25度で一定である。このような状況であれば、空調機器の消費電力低下と室内温度の低下が同時に発生しうるため、空調機器の異常は発生していないと推定することができる。

【0034】

図2～図5の例で示したように、データに含まれる変数をコンテキスト変数（文脈変数）とコンテンツ変数（内容変数）に分類すると、異常検知の精度を高めることができる。

40

【0035】

コンテキストデータベース3は、各計測時刻におけるコンテキスト変数である、コンテキストデータを保存する。上述の図3、図5は、コンテキストデータの一例である。コンテキストデータは、装置やシステムごとに設けられたテーブルに記憶されていてもよいし、同じテーブルにまとめて記録されていてもよい。複数の装置やシステムのコンテキストデータを同じテーブルに記録する場合、各変数に対し、装置またはシステムに対応する識別子を付与する。

【0036】

50

コンテキストデータは、文脈的異常検知における用途から、基準データ 3 a と判定対象データ 3 b に分けられる。例えば、基準データとして用いられるコンテキストデータと、判定対象データとして用いられるデータを異なるテーブルに格納する。また、用途ごとにテーブルを分けず、基準データとして用いられるデータと判定対象データとして用いられるデータを同一のテーブルに保存してもよい。この場合、基準データに係る開始時刻および終了時刻ならびに判定対象データに係る開始時刻および終了時刻を管理し、両者を識別できるようにする。

【0037】

基準データは、監視対象の装置やシステムが正常状態にあると推定される期間に取得されるデータである。例えば装置の保守点検やセンサの校正作業の直後や、装置がバスタブ曲線における初期故障期間を過ぎて偶発故障期間に入ったときに、装置は正常稼働している可能性が高い。したがって、このような期間に得られたデータを正常状態におけるデータと推定し、基準データとして用いることができる。

【0038】

基準データに含まれるサンプル数は異常検知部 6 に適用される文脈的異常検知の手法に応じて決められる。

【0039】

判定対象データとは、異常検知の対象とされるデータである。判定対象データは、基準データとは異なる期間に得られたデータであり、装置またはシステムの正常／異常状態の判定を行うもののことをいう。判定対象データに含まれるサンプル数は異常検知部 6 に適用される文脈的異常検知の手法に応じて決められる。

【0040】

コンテンツデータベース 4 は、各計測時刻におけるコンテンツ変数である、コンテキストデータを保存する。上述の図 2、図 4 は、コンテンツデータの一例である。コンテンツデータは、装置やシステムごとに設けられたテーブルに記憶されていてもよいし、同じテーブルにまとめて記録されていてもよい。複数の装置やシステムのコンテンツデータを同じテーブルに記録する場合、各変数に対し、装置またはシステムに対応する識別子を付与する。

【0041】

コンテンツデータも、コンテキストデータと同様に基準データ 4 a と判定対象データ 4 b に分けられる。基準データとして用いられるコンテンツデータと、判定対象データとして用いられるコンテンツデータは、同一のテーブルに格納されてもよいし、別々のテーブルに格納されてもよい。

【0042】

本実施形態では、コンテキストデータとコンテンツデータを異なるテーブルに分けて記憶したが、コンテキストデータとコンテンツデータを同じテーブルに記憶してもよい。この場合、それぞれの変数にデータの分類を示す識別子を付与する。ここまで、複数のデータの保存形式について説明したが、コンテキストデータと、コンテンツデータが行列として表現可能な形式で保存されているのであれば、どのような形式を用いてもよい。

【0043】

変数選択部 5 は、基準データと判定対象データの間の識別に寄与する程度（識別寄与度）を計算し、計算結果に基づき、異常検知部 6 における文脈的異常検知で使用する変数を選択する。異常検知部 6 は、変数選択部 5 で選択された変数を使って、文脈的異常検知を行う。コンテキスト変数については、識別寄与度が大きいものが異常検知における使用対象から除外される。なお、本実施形態に係る変数選択部はすべてのコンテンツ変数を文脈的異常検知で使用する変数として選択するものとする。

【0044】

あるコンテキスト変数の値、ふるまいなどが基準データと判定対象データで大きくことなっている場合、当該コンテキスト変数の識別寄与度は大きくなる。コンテキスト変数の識別寄与度が大きい場合、基準データと判定対象データは異なる条件、状況や背景など（

10

20

30

40

50

context : コンテキスト) で得られたデータであるとみなすことができる。

【0045】

基準データと判定対象データで異なる挙動を示しているコンテキスト変数を含めて文脈的異常検知を行おうとすると、当該コンテキスト変数の値の違いが過分に反映され、判定が行われてしまう。その結果、判定において真に異常兆候を示唆するその他の変数のふるまいが十分に反映されなくなる。したがって、異常検知において異常の誤検知や見逃しが発生する確率が高まってしまう。文脈的異常検知の精度を高めるためには、識別寄与度の大きいコンテキスト変数があれば、異常検知を行う前に当該コンテキスト変数を除去し、識別寄与度が小さいコンテキスト変数を選択するのが効果的である。

【0046】

識別寄与度の大きいコンテキスト変数を除去することにより、基準データと判定対象データで計測条件に違いがある中でも、基準データと判定対象データで共通している条件に基づいて文脈的異常検知を行うことができる。

【0047】

次に、変数選択処理の詳細について説明する。コンテンツデータと、コンテキストデータを使って行列計算をすることにより、変数選択処理が実行される。以降では「データ」といった場合、コンテンツデータとコンテキストデータの両方を示すものとする。また、単に「変数」といった場合、コンテンツ変数とコンテキスト変数の両方を示すものとする。最初に、変数の表記について説明する。

【0048】

各時刻におけるコンテキストデータを下記の式(1)のようなベクトルで表現することができる。

【数1】

(1)

$$\vec{x} = (b_1, b_2, b_3 \dots, b_m)$$

それぞれのbはコンテキスト変数(文脈変数)であり、条件、状況や背景などを示す設定値や計測値に相当する。式(1)のベクトルは、 $b_1 \sim b_m$ のm個のコンテキスト変数を含んでいる。式(1)のベクトルは、図2、図4におけるレコードに相当する。

【0049】

変数選択処理では、データを用途に基づき基準データと判定対象データに分けている。基準データと判定対象データを識別するため、ベクトルに添え字を付ける。基準データに係るベクトルにはaの添え字を付与し、判定対象データに係るベクトルにはhの添え字を付与するものとする。したがって、ベクトルxを下記の式(2)のように区別する。

【数2】

(2)

$$\vec{x}_a = (b_{a1}, b_{a2}, b_{a3} \dots, b_{am})$$

$$\vec{x}_h = (b_{h1}, b_{h2}, b_{h3} \dots, b_{hm})$$

ここで、 b_{a1} と b_{h1} 、 b_{a1} と b_{h1} はそれぞれ同一のコンテキスト変数であるが、異なる期間に得られた変数であることを明確にするため、aまたはhの添え字で区別している。

【0050】

データは、複数の時刻における変数を含んでいる。異なる時刻に得られたコンテキスト変数を区別するため、さらに添え字を追加する。下記の式(3)では複数の時刻に係るコ

10

20

30

40

50

ンテキスト変数のベクトルが示されている。

【数 3】

$$\begin{aligned} & \vec{x}_{\{1,a\}}, \vec{x}_{\{2,a\}}, \vec{x}_{\{3,a\}}, \dots, \vec{x}_{\{n_a,a\}} \\ & \vec{x}_{\{1,h\}}, \vec{x}_{\{2,h\}}, \vec{x}_{\{3,h\}}, \dots, \vec{x}_{\{n_h,h\}} \end{aligned} \quad (3)$$

10

【0051】

式(3)の下付き添え字{1、a}、{2、a}、{3、a}、{1、h}、{2、h}、{3、h}にある数字は、ベクトル内の変数が得られた時刻を区別している。式(3)では、{n_a、a}の添え字から、基準データはn_a個の時刻における変数の値を含んでいることがわかる。また、{n_h、h}の添え字から、判定対象データについてはn_h個の時刻における変数の値を含んでいることがわかる。

【0052】

次に、上述の表記を用いて第1の実施形態に係る変数選択処理を説明する。

【0053】

基準データとして用いられるコンテキスト変数の複数のレコードを、下記の式(4)のようなn_a×m次元の行列X_aの形式に集約して表記することができる。

20

【数 4】

$$X_a = \begin{bmatrix} \vec{x}_{\{1,a\}} \\ \vec{x}_{\{2,a\}} \\ \vdots \\ \vec{x}_{\{n_a,a\}} \end{bmatrix} \in R^{n_a \times m} \quad (4)$$

30

【0054】

また、下記の式(5)のような全要素の値が1のn_a次元の列ベクトルY_aを定義する。

【数 5】

$$Y_a = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \in R^{n_a} \quad (5) \quad 40$$

【0055】

式(4)と同様に、判定対象データとして用いられるコンテキスト変数の複数のレコー

50

ドを、下記の式（６）のような $n_h \times m$ 次元の行列 X_h の形式に集約して表記することができる。

【数 6】

$$X_h = \begin{bmatrix} \vec{x}_{\{1,h\}} \\ \vec{x}_{\{2,h\}} \\ \vdots \\ \vec{x}_{\{n_h,h\}} \end{bmatrix} \in R^{n_h \times m} \quad (6)$$

10

【 0 0 5 6 】

また、下記の式（７）のような、全要素の値が 0 の n_h 次元の列ベクトル Y_h を定義する。

【数 7】

$$Y_h = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in R^{n_h} \quad (7)$$

20

【 0 0 5 7 】

次に、行列 X_a と行列 X_h を行方向に連結し、下記の式（８）の行列 X_c を生成する。

【数 8】

$$X_c = \begin{bmatrix} \vec{x}_{\{1,a\}} \\ \vec{x}_{\{2,a\}} \\ \vdots \\ \vec{x}_{\{n_a,a\}} \\ \vec{x}_{\{1,h\}} \\ \vec{x}_{\{2,h\}} \\ \vdots \\ \vec{x}_{\{n_h,h\}} \end{bmatrix} \in R^{(n_a+n_h) \times m} \quad (8)$$

40

【 0 0 5 8 】

また、列ベクトル Y_a と Y_h を連結し、下記の式（９）のベクトル Y_c を生成する。

50

【数 9】

$$Y_c = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \in R^{(n_a+n_b)} \quad (9)$$

10

【0059】

行列 X_c と、ベクトル Y_c の準備ができたなら、行列 X_c を説明変数とし、ベクトル Y_c を目的変数とし、それぞれのコンテキスト変数の識別寄与度を計算する。識別寄与度は例えば、ランダムフォレストなどの識別器を使って求めることができる。ランダムフォレストを使った場合、識別寄与度はランダムフォレストにおける変数の重要度に相当する。

【0060】

ランダムフォレストでは、訓練用のデータから重複を許してランダムにレコードを選択し、ブートストラップ標本を生成する処理と、生成したブートストラップ標本を使い、木を生成する処理を繰り返し行う。このとき訓練用データのうち、約 $1/3$ のレコード (out-of-bag data: 以降では oob データと記す。) が木の生成に使われずに残る。

20

【0061】

Permutation importance による変数の重要度は、上述の oob データを使って算出される。まず、それぞれの oob データを対応する木に適用し、変数が正しく分類されている回数を求める。次に、それぞれの oob データにおける変数の値をランダムに並び替え、対応する木に適用し、変数が正しく分類されている回数を求める。ランダムフォレスト全体について、元の oob データにおける正解数と並び替え後の oob データにおける正解数の差の平均値は、Permutation importance の値となる。

30

【0062】

それぞれの変数について、標準偏差で正規化した Permutation importance を使ってもよい。変数の値を並び替えた後の oob データにおける正解率が元の oob データにおける正解率より大きく低下している場合、当該変数の分類における影響度は大きいと推定される。

【0063】

変数の重要度をジニ係数に基づき計算することもできる。ジニ係数はデータの不純度を示しており、ジニ係数の値が大きいほど、データが未分類であるといえる。したがって、データの分割前後におけるジニ係数の減少量が大きい変数は、重要度が大きいと推定される。それぞれの変数について、ランダムフォレストに含まれるすべての木についてジニ係数の減少量を加算すると、各変数の重要度を求めることができる。この方法により求められた重要度 (識別寄与度) は Gini Importance と呼ばれる。

40

【0064】

上述では、識別器としてランダムフォレストを使う場合を説明したが、これは一例にしかならない。例えば、Adaboost などその他の種類のアンサンブル学習を使って識別を行ってもよく、変数の分類が可能なのであれば、適用する機械学習の手法やアルゴリズムについては特に問わない。また、変数の重要度 (識別寄与度) として、Gini Importance や Permutation importance 以外の指標を使うことを妨げるものではない。

50

【0065】

各コンテキスト変数について、識別寄与度を計算したら、各コンテキスト変数の識別寄与度をしきい値と比較する。コンテキスト変数の識別寄与度がしきい値より大きい場合、当該コンテキスト変数を文脈的異常検知で使用する変数から除外する。一方、コンテキスト変数の識別寄与度がしきい値以下である場合、当該コンテキスト変数は文脈的異常検知で使用する変数として選択される。この判定で用いるしきい値の決め方については特に問わない。例えば、全コンテキスト変数の識別寄与度の平均値をしきい値として使うことが考えられる。

【0066】

以上で変数選択処理は完了する。変数選択部5は、後述する記憶装置105に選択されなかったコンテキスト変数に係る情報または選択されたコンテキスト変数に係る情報を保存してもよい。また、異常検知部6に、選択されなかったコンテキスト変数に係る情報または選択されたコンテキスト変数に係る情報を送信してもよい。異常検知部6は変数選択処理で選択されなかったコンテキスト変数を使用対象から除外して文脈的異常検知を行う。なお、本実施形態では、すべてのコンテンツ変数が選択されるため、異常検知部6はすべてのコンテンツ変数を使って文脈的異常検知を行う。

10

【0067】

異常検知部6は、変数選択部5で選択された変数に基づき、文脈的異常検知を行う。異常検知部6は、コンテキスト変数とコンテンツ変数を含む基準データを正常データとして、コンテキスト変数とコンテンツ変数を含む判定対象データの文脈的異常検知を行う。異常検知部6では、正常データを使って異常検知が可能な方法を使って、文脈的異常検知を実行する。なお、文脈的異常検知の実行前にデータの学習が必要である場合、異常検知部6でデータの学習を行ってもよい。

20

【0068】

正常データを使って異常検知を行う手法の例としては、ニューラルネットワークが挙げられる。ニューラルネットワークによる異常検知を行う場合、正常データであるとみなされる基準データを使ってニューラルネットワークの学習を行う。そして、学習したニューラルネットワークを使い、判定対象データの文脈的異常検知を行う。使用するニューラルネットワークの種類や層の数については特に限定しない。

【0069】

他に、正常データを使って異常検知が可能な手法の例として、正常データの確率密度と判定対象データの確率密度の比を推定する確率密度比推定を使ったものが挙げられる。このような確率密度比に基づく異常検知を行う場合、基準データと判定対象データの確率密度比が1から離れているほど異常の度合いが大きいと推定する。

30

【0070】

上述のニューラルネットワークや確率密度比により異常検知は文脈的異常検知の例であり、その他の手法やモデルを使って文脈的異常検知を行ってもよい。異常検知部6は、文脈的異常検知の結果を記憶装置105に保存する。ここで、文脈的異常検知結果の保存形式の例としては、テキスト、バイナリ、CSV、XMLなどがあるが、形式については特に問わない。文脈的異常検知結果をテキストやグラフィックに変換可能な形式で保存すれば、表示部7は文脈的異常検知の結果をディスプレイに表示することができる。

40

【0071】

表示部7は、異常検知部6が生成した文脈的異常検知結果を、所定の形式を有する画像データやテキストデータに変換し、後述する表示装置103に出力する。表示部7では文脈的異常検知の結果の他に、変数選択部5によるデータ前処理結果を表示してもよい。

【0072】

図6は、表示装置103に表示されたデータの前処理結果および文脈的異常検知結果の例を示している。図6上段には、データ前処理結果が示されている。そして、図6下段には、文脈的異常検知の結果が示されている。図6の例は、空調機器に対して文脈的異常検知を行った場合の表示内容であるため、あるエアコンの点検を促す旨のメッセージが表示

50

されている。本実施形態に係る監視装置はどのような装置やシステムに対して適用してもよく、メッセージの内容を装置やシステムの種類に応じて変更することができる。

【0073】

例えば、情報システムへの標的型攻撃、D o S 攻撃、マルウェア、侵入などを文脈的異常検知により検知する場合、情報システムに対する攻撃が行われた旨のメッセージに加え、セキュリティ対策ソフトウェアやアプライアンスなどによって所定のネットワーク切断、機能の停止、ポートの使用禁止、アクセス遮断などの措置がとられた旨のメッセージを表示してもよい。

【0074】

次に、本実施形態に係る監視装置のハードウェア構成について説明する。本実施形態に係る監視装置は、コンピュータ100により構成される。コンピュータ100には、サーバ、クライアント端末、マイコン、タブレット、スマートフォン、パソコン、汎用コンピュータなど各種の情報処理装置が含まれる。

【0075】

図7は、コンピュータ100の一例を示す図である。図7のコンピュータ100は、プロセッサ101と、入力装置102と、表示装置103と、通信装置104と、記憶装置105とを備える。プロセッサ101、入力装置102、表示装置103、通信装置104、記憶装置105は、バス106により相互に接続されている。

【0076】

プロセッサ101は、コンピュータ100の制御装置と演算装置を含む電子回路である。プロセッサ101として、例えば、汎用目的プロセッサ、中央処理装置（C P U）、マイクロプロセッサ、デジタル信号プロセッサ（D S P）、コントローラ、マイクロコントローラ、状態マシン、特定用途向け集積回路、フィールドプログラマブルゲートアレイ（F P G A）、プログラム可能論理回路（P L D）またはこれらの組合せを用いることができる。

【0077】

プロセッサ101は、バス106を介して接続された各装置（例えば、入力装置102、通信装置104、記憶装置105）から入力されたデータやプログラムに基づいて演算処理を行い、演算結果や制御信号を、バス106を介して接続された各装置（例えば、表示装置103、通信装置104、記憶装置105）に出力する。具体的には、プロセッサ101は、コンピュータ100のO S（オペレーティングシステム）や、監視プログラムなどを実行し、コンピュータ100を構成する各装置を制御する。

【0078】

監視プログラムとは、コンピュータ100に、監視装置の上述の各機能や構成要素を実現させるプログラムである。監視プログラムは、一時的でない有形のコンピュータ読み取り可能な記憶媒体に記憶される。上記の記憶媒体は、例えば、光ディスク、光磁気ディスク、磁気ディスク、磁気テープ、フラッシュメモリ、半導体メモリであるが、これに限られない。プロセッサ101が監視プログラムを実行することにより、コンピュータ100が監視装置として機能する。

【0079】

入力装置102は、コンピュータ100に情報を入力するための装置である。入力装置102は、例えば、キーボード、マウス、タッチパネルなどであるが、これに限られない。ユーザは、入力装置102を用いることにより、文脈的異常検知（監視）を行うセンサや装置などの指定、コンテキスト変数とコンテンツ変数に係る分類の指定、基準データに係る期間の指定、判定対象期間の指定、変数選択処理で使用する手法の選択、文脈的異常検知処理の開始指令などを入力することができる。

【0080】

表示装置103は、画像や映像を表示するための装置である。表示装置103は、例えば、L C D（液晶ディスプレイ）、C R T（ブラウン管）、有機E L（有機エレクトロルミネッセンス）ディスプレイなどであるが、これに限られない。表示装置103には、上

10

20

30

40

50

述のように判定結果として、データ前処理の結果や異常が検知された箇所などが表示される。

【0081】

通信装置104は、コンピュータ100が外部装置と無線または有線で通信するための装置である。通信装置104は、例えば、NIC (Network Interface Card)、通信モジュール、モデム、ハブ、ルータなどであるが、これに限られない。収集部2は、通信装置104を介して、センサが設置された建物などから計測値データを遠隔収集してもよい。

【0082】

記憶装置105は、コンピュータ100のOSや、監視プログラム、監視プログラムの実行に必要なデータ、監視プログラムの実行により生成されたデータなどを記憶する記憶媒体である。記憶装置105には、主記憶装置と外部記憶装置が含まれる。主記憶装置は、例えば、RAM、DRAM、SRAMであるが、これに限られない。また、外部記憶装置は、例えば、ハードディスク、光ディスク、フラッシュメモリ、磁気テープなどであるが、これに限られない。コンテキストデータベース3、コンテンツデータベース4は、記憶装置105上に構築されてもよいし、外部のサーバやストレージ上に構築されてもよい。

10

【0083】

なお、コンピュータ100は、プロセッサ101、入力装置102、表示装置103、通信装置104、記憶装置105を、それぞれ1つずつまたは複数備えてもよいし、プリンタやスキャナなどの周辺機器を接続されていてもよい。

20

【0084】

また、監視装置は、単一のコンピュータ100により構成されてもよいし、相互に接続された複数のコンピュータ100からなるシステムとして構成されてもよい。

【0085】

さらに、監視プログラムは、コンピュータ100の記憶装置105に予め記憶されていてもよいし、コンピュータ100の外部の記憶媒体に記憶されていてもよいし、インターネット上にアップロードされていてもよい。いずれの場合も、監視プログラムをコンピュータ100にインストールして実行することにより、監視装置の機能が実現される。

【0086】

30

(第2の実施形態)

第1の実施形態に係る監視装置では、アンサンブル学習などの機械学習により、識別寄与度を計算し、それぞれの変数の識別寄与度に基づいて変数選択処理を実行していた。基準データとして用いられるコンテキスト変数と、判定対象データとして用いられるコンテキスト変数に対して統計的検定を行い、検定結果に基づき変数選択を行ってもよい。

【0087】

第2の実施形態に係る変数選択処理では、変数について一切の前提を設けず、特定の分布を仮定しないノンパラメトリックな検定手法を用いることができる。ノンパラメトリックな検定手法の例としてはマン・ホイットニーのU検定 (Mann-Whitney U test) などがあるが、その他の検定手法を用いてもよい。また、使用する検定手法は必ずノンパラメトリックな検定手法でなくてもよい。

40

【0088】

基準データの行列 X_a である式(4)の要素を明示すると、下記の式(10)のようになる。

【数 1 0】

$$X_a = \begin{bmatrix} b_{\{1,a1\}} & b_{\{1,a2\}} & b_{\{1,a3\}} & \cdots & b_{\{1,am\}} \\ b_{\{2,a1\}} & b_{\{2,a2\}} & b_{\{2,a3\}} & \cdots & b_{\{2,am\}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{\{n_a,a1\}} & b_{\{n_a,a2\}} & b_{\{n_a,a3\}} & \cdots & b_{\{n_a,am\}} \end{bmatrix} \in R^{n_a \times m} \quad (10)$$

10

【0 0 8 9】

一方、判定対象データの行列 X_h である式 (6) の要素を明示すると、下記の式 (11) のようになる。

【数 1 1】

$$X_h = \begin{bmatrix} b_{\{1,h1\}} & b_{\{1,h2\}} & b_{\{1,h3\}} & \cdots & b_{\{1,hm\}} \\ b_{\{2,h1\}} & b_{\{2,h2\}} & b_{\{2,h3\}} & \cdots & b_{\{2,hm\}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{\{n_h,h1\}} & b_{\{n_h,h2\}} & b_{\{n_h,h3\}} & \cdots & b_{\{n_h,hm\}} \end{bmatrix} \in R^{n_h \times m} \quad (11)$$

20

【0 0 9 0】

式 (10)、(11) の各行は、各時刻で得られたコンテキスト変数のレコードに相当する。一方、式 (10)、(11) の各列は、複数の時刻における、ひとつのコンテキスト変数の値に相当する。式 (10) の行列 X_a と式 (11) の行列 X_h はいずれも m 個のコンテキスト変数を含んでいる。

30

【0 0 9 1】

本実施形態に係る変数選択部は、行列 X_a と行列 X_h から同じコンテキスト変数に係る列を抽出する。そして、それぞれの列について統計的検定を行い、有意差が存在する場合には、当該コンテキスト変数を、文脈的異常検知の使用対象から除外する。すなわち、有意差が存在しないコンテキスト変数を、文脈的異常検知で使用するコンテキスト変数に選択する。この処理をすべてのコンテキスト変数に係る列について実行することにより、文脈的異常検知に使用するコンテキスト変数を選択することができる。

【0 0 9 2】

変数選択部で実行される処理が異なる点を除けば、第2の実施形態に係る監視装置の機能および構成は、第1の実施形態に係る監視装置と同様である。

40

【0 0 9 3】

(第3の実施形態)

上述の各実施形態に係る監視装置では、コンテキスト変数（文脈変数）についてのみ変数選択を行っており、コンテンツ変数については、すべてのコンテンツ変数を選択して文脈的異常検知を行っていた。しかし、さらにコンテンツ変数（内容変数）の変数選択を行ってもよい。第3の実施形態に係る変数選択部は、コンテキスト変数の変数選択をした後、コンテンツ変数の変数選択を実行する。コンテキスト変数の変数選択処理については、第1の実施形態または第2の実施形態で説明した方法のいずれを用いてもよい。

【0 0 9 4】

本実施形態に係る変数選択部の処理を説明する前に、変数の表記について説明する。下

50

記の式（１２）は、基準データとして用いられるコンテンツ変数の行列 Z_a である。

【数１２】

$$Z_a = \begin{bmatrix} S_{\{1,a1\}} & S_{\{1,a2\}} & S_{\{1,a3\}} & \cdots & S_{\{1,al\}} \\ S_{\{2,a1\}} & S_{\{2,a2\}} & S_{\{2,a3\}} & \cdots & S_{\{2,al\}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ S_{\{n_a,a1\}} & S_{\{n_a,a2\}} & S_{\{n_a,a3\}} & \cdots & S_{\{n_a,al\}} \end{bmatrix} \in R^{n_a \times l} \quad (12)$$

10

【００９５】

本実施形態に係る変数選択部の処理を説明する前に、変数の表記について説明する。下記の式（１３）は、判定対象データとして用いられるコンテンツ変数の行列 Z_h である。

【数１３】

$$Z_h = \begin{bmatrix} S_{\{1,h1\}} & S_{\{1,h2\}} & S_{\{1,h3\}} & \cdots & S_{\{1,h1\}} \\ S_{\{2,h1\}} & S_{\{2,h2\}} & S_{\{2,h3\}} & \cdots & S_{\{2,h1\}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ S_{\{n_h,h1\}} & S_{\{n_h,h2\}} & S_{\{n_h,h3\}} & \cdots & S_{\{n_h,h1\}} \end{bmatrix} \in R^{n_h \times l} \quad (13)$$

20

【００９６】

式（１２）の行列 Z_a と式（１３）の Z_h の各行は、各時刻で得られたコンテンツ変数のレコードに相当する。一方、行列 Z_a 、 Z_h の各列は、複数の時刻における、ひとつのコンテンツ変数の値に相当する。行列 Z_a 、 Z_h はいずれも l （小文字の l ）個のコンテンツ変数を含んでいる。

30

【００９７】

次に、変数選択後のコンテキストデータの表記について説明する。下記の式（１４）は基準データとして使われるコンテキストデータである。式（１４）は変数選択が行われた後のデータであるため、列の数を示す記号として m^* が使われている。

【数１４】

$$X_a^* = \begin{bmatrix} \vec{x}_{\{1,a\}}^* \\ \vec{x}_{\{2,a\}}^* \\ \vdots \\ \vec{x}_{\{n_a,a\}}^* \end{bmatrix} \in R^{n_a \times m^*} \quad (14)$$

40

【００９８】

一方、下記の式（１５）は判定対象データとして使われるコンテキストデータである。式（１５）も変数選択が行われた後のデータであるため、列の数を示す記号として m^* が

50

使われている。

【数 1 5】

$$X^*_h = \begin{bmatrix} \vec{x}^*_{\{1,h\}} \\ \vec{x}^*_{\{2,h\}} \\ \vdots \\ \vec{x}^*_{\{n_h,h\}} \end{bmatrix} \in R^{n_h \times m^*} \quad (15)$$

10

【0 0 9 9】

式 (1 4)、(1 5) で、それぞれの行は変数選択部で除外されなかったコンテキスト変数である。列の数 m^* については、変数選択前の列の数を m とすると、 $m^* \leq m$ の関係が成り立つ。以降では、上述の表記を用いて、第 3 の実施形態に係る変数選択処理について説明する。

【0 1 0 0】

まず、行列 X^*_a と行列 Z_a を列方向に連結し、下記の式 (1 6) に示した行列 W_a を生成する。

20

【数 1 6】

$$W_a = \begin{bmatrix} b_{\{1,a1\}} & \cdots & b_{\{1,am^*\}} & s_{\{1,a1\}} & \cdots & s_{\{1,al\}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ b_{\{n_a,a1\}} & \cdots & b_{\{n_a,am^*\}} & s_{\{n_a,a1\}} & \cdots & s_{\{n_a,al\}} \end{bmatrix} \in R^{n_a \times (m^*+l)} \quad (16)$$

30

【0 1 0 1】

同様に、行列 X^*_h と行列 Z_h を列方向に連結し、下記の式 (1 7) に示した行列 W_h を生成する。

【数 1 7】

$$W_h = \begin{bmatrix} b_{\{1,h1\}} & \cdots & b_{\{1,hm^*\}} & s_{\{1,h1\}} & \cdots & s_{\{1,h1\}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ b_{\{n_h,h1\}} & \cdots & b_{\{n_h,hm^*\}} & s_{\{n_h,h1\}} & \cdots & s_{\{n_h,h1\}} \end{bmatrix} \in R^{n_h \times (m^*+l)} \quad (17)$$

40

【0 1 0 2】

そして、行列 W_a と行列 W_h を行方向に連結し、下記の式 (1 8) に示した行列 W_c を生成する。

【数 1 8】

$$W_c = \begin{bmatrix} b_{\{1,a1\}} & \cdots & b_{\{1,am^*\}} & s_{\{1,a1\}} & \cdots & s_{\{1,al\}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ b_{\{n_a,a1\}} & \cdots & b_{\{n_a,am^*\}} & s_{\{n_a,a1\}} & \cdots & s_{\{n_a,al\}} \\ b_{\{1,h1\}} & \cdots & b_{\{1,hm^*\}} & s_{\{1,h1\}} & \cdots & s_{\{1,hl\}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ b_{\{n_h,h1\}} & \cdots & b_{\{n_h,hm^*\}} & s_{\{n_h,h1\}} & \cdots & s_{\{n_h,hl\}} \end{bmatrix} \in R^{(n_a+n_h) \times (m^*+l)} \quad (18)$$

10

【0103】

次に、式(18)の行列 W_c を説明変数、式(9)のベクトル Y_c を目的変数とし、それぞれのコンテンツ変数 s の識別寄与度(重要度)を計算する。識別寄与度は第1の実施形態における変数選択処理と同様、ランダムフォレストやAdaboostなどの識別器を使って求めることができる。分類が可能なのであれば、適用する機械学習の手法やモデルについては特に問わない。識別寄与度を示す指標として、例えばPermutation importanceや、Gini Importanceなどを使うことができるが、その他の指標を使ってもよい。

20

【0104】

各コンテンツ変数 s について、識別寄与度を計算したら、各コンテンツ変数の識別寄与度をしきい値と比較する。コンテンツ変数の識別寄与度がしきい値より小さい場合、当該コンテンツ変数を文脈的異常検知で使用する変数から除外する。一方、コンテンツ変数の識別寄与度がしきい値以上である場合、当該コンテンツ変数は文脈的異常検知で使用する変数として選択される。この判定で用いるしきい値の決め方については特に問わない。例えば、全コンテンツ変数の識別寄与度の平均値をしきい値として使うことが考えられる。

【0105】

基準データと判定対象データとの間で挙動が類似したコンテンツ変数の識別寄与度(分類における重要度)は小さくなる。基準データと判定対象データとの間で似たふるまいや、近い値を示しているコンテンツ変数を含めて文脈異常検知を行っても、正常/異常の判定に寄与するところが少なく、却って判定精度を低下させるおそれすらある。したがって、識別寄与度の小さいコンテンツ変数を使用対象から外すことにより、文脈的異常検知の精度を改善することができる。また、異常検知部に入力される変数の数が減れば、文脈的異常検知における処理負荷が軽減される。

30

【0106】

以上で第3の実施形態に係る変数選択処理は完了する。変数選択部は、記憶装置105に選択されなかった変数に係る情報または選択された変数に係る情報を保存してもよい。また、異常検知部に、選択されなかった変数に係る情報または選択された変数に係る情報を送信してもよい。異常検知部は変数選択処理で選択された変数(コンテキスト変数およびコンテンツ変数)を使う一方、変数選択処理で選択されなかった変数(コンテキスト変数およびコンテンツ変数)を除外して、文脈的異常検知を行う。

40

【0107】

変数選択部で実行される処理が異なる点を除けば、第3の実施形態に係る監視装置の機能および構成は、上述の各実施形態に係る監視装置と同様である。なお、図6のデータの前処理結果および文脈的異常検知結果の例では、コンテキスト変数のみならず、一部のコンテンツ変数も選択から除外されている。したがって、図6の画面は第3の実施形態に係る監視装置により生成され、出力されたものであることがわかる。

【0108】

50

本発明の実施形態に係る監視装置全体の処理を説明する。図8は、監視装置による処理の例を示したフローチャートである。以下では、図8を参照しながら処理を説明する。

【0109】

まず、データをコンテンツ変数とコンテキスト変数に分類する。(ステップS101) ステップS101におけるデータとは、監視装置に入力されるデータのことをいう。コンテンツ変数とコンテキスト変数に分類された後のデータを入力データとよぶものとする。変数の分類処理の詳細は、収集部2に係る説明で述べた通りである。

【0110】

次に基準データと判定対象データを用意する。(ステップS102) 基準データと判定対象データの定義については、コンテキストデータベース3に係る説明で述べた通りである。そして、基準データと判定対象データに基づき、変数選択処理を行う。(ステップS103) 変数選択部による変数選択処理は上述の各実施形態のいずれの処理を用いてもよい。変数選択処理の詳細は変数選択部5および上述の各実施形態の説明で述べた通りである。

10

【0111】

そして、選択された変数を使って、判定対象データの文脈的異常検知を実行する。(ステップS104) 文脈的異常検知に係る処理の詳細は異常検知部6の説明で述べた通りである。最後に、監視装置は変数選択の結果および文脈的異常検知を表示する。(ステップS105)

【0112】

20

なお、本発明は上記各実施形態そのままに限定されるものではなく、実施段階ではその要旨を逸脱しない範囲で構成要素を変形して具体化できる。また、上記各実施形態に開示されている複数の構成要素を適宜組み合わせることによって種々の発明を形成できる。また例えば、各実施形態に示される全構成要素からいくつかの構成要素を削除した構成も考えられる。さらに、異なる実施形態に記載した構成要素を適宜組み合わせてもよい。

【符号の説明】

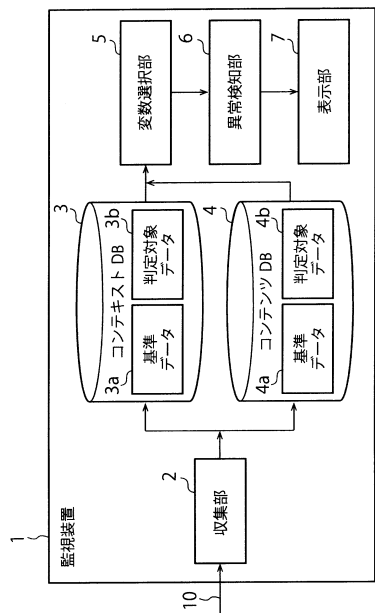
【0113】

- 1 監視装置
- 2 収集部
- 3 コンテキストデータベース
- 3 a、4 a 基準データ
- 3 b、4 b 判定対象データ
- 4 コンテンツデータベース
- 5 変数選択部
- 6 異常検知部
- 7 表示部
- 10 電気通信回線
- 100 コンピュータ
- 101 プロセッサ
- 102 入力装置
- 103 表示装置
- 104 通信装置
- 105 記憶装置
- 106 バス

30

40

【図 1】



【図 2】

計測時刻	m			
	スイッチ A	回路ブロック B		
0:00	0 (操作なし)	0 (OFF)		
0:01	1 (押下)	0 (OFF)		
0:02	0 (操作なし)	0 (OFF)		
0:03	0 (操作なし)	0 (OFF)		
0:04	0 (操作なし)	1 (ON)		
0:05	0 (操作なし)	1 (ON)		
0:06	0 (操作なし)	0 (OFF)		
⋮	⋮	⋮		

コンテキストデータ

【図 3】

計測時刻	l		
	電圧 C	電流 D	
0:00	1.2V	7mA	
0:01	2.0V	8mA	
0:02	1.1V	6mA	
0:03	1.2V	7mA	
0:04	1.2V	25mA	
0:05	1.1V	23mA	
0:06	1.1V	7mA	
⋮	⋮	⋮	

コンテンツデータ

【図 4】

計測時刻	m			
	設定温度	外気温	天気	
10:00	27℃	33.6℃	0 (晴れ)	
11:00	25℃	33.8℃	0 (晴れ)	
12:00	25℃	34.2℃	0 (晴れ)	
13:00	25℃	35.0℃	0 (晴れ)	
14:00	25℃	35.4℃	1 (曇り)	
15:00	25℃	27.9℃	2 (雨)	
16:00	25℃	26.1℃	2 (雨)	
17:00	25℃	24.7℃	2 (雨)	
⋮	⋮	⋮	⋮	

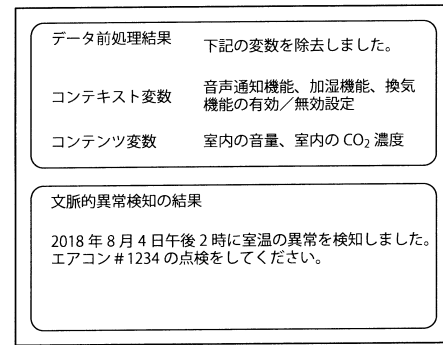
コンテキストデータ

【図 5】

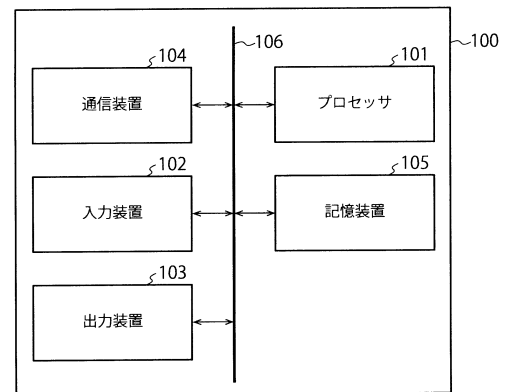
n	l			
	計測時刻	室内温度	消費電力	
	10:00	30.3℃	310W	
	11:00	28.9℃	350W	
	12:00	27.8℃	340W	
	13:00	28.1℃	330W	
	14:00	28.3℃	335W	
	15:00	28.2℃	330W	
	16:00	27.4℃	290W	
	17:00	26.9℃	285W	
	⋮	⋮	⋮	

コンテンツデータ

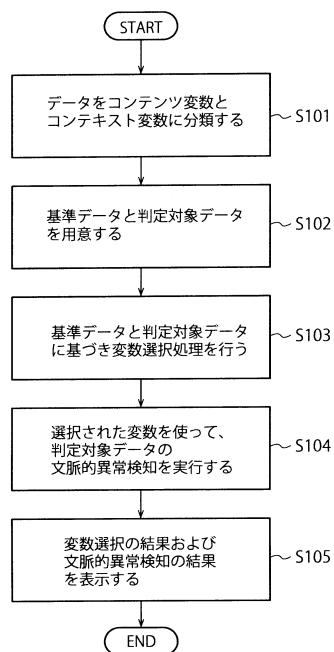
【図 6】



【図 7】



【図 8】



フロントページの続き

- (72)発明者 森山 拓郎
東京都港区芝浦一丁目1番1号 株式会社東芝内
- (72)発明者 石井 岳
東京都港区芝浦一丁目1番1号 株式会社東芝内

審査官 山村 秀政

- (56)参考文献 特開2014-186402 (JP, A)
特開2010-191556 (JP, A)
国際公開第2017/011734 (WO, A1)
特開2006-319220 (JP, A)
特開2004-169989 (JP, A)
特開平07-181097 (JP, A)
特開2017-120504 (JP, A)
特開2008-204282 (JP, A)
特開平08-030323 (JP, A)

- (58)調査した分野(Int.Cl., DB名)
G05B 23/02