(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7672926号
(P7672926)

(45)発行日　令和7年5月8日(2025.5.8)

(24)登録日　令和7年4月25日(2025.4.25)

(51)国際特許分類

Ｇ０６Ｆ　11/34　(2006.01)	Ｇ０６Ｆ	11/34	１５２
Ｇ０５Ｂ　23/02　(2006.01)	Ｇ０５Ｂ	23/02	Ｚ

請求項の数　14　（全31頁）

(21)出願番号　特願2021-143534(P2021-143534)
(22)出願日　令和3年9月2日(2021.9.2)
(65)公開番号　特開2023-36469(P2023-36469A)
(43)公開日　令和5年3月14日(2023.3.14)
審査請求日　令和6年3月5日(2024.3.5)

(73)特許権者　524132520
日立ヴァンタラ株式会社
神奈川県横浜市戸塚区吉田町２９２番地
(74)代理人　110002365
弁理士法人サンネクスト国際特許事務所
(72)発明者　バックフース ヤナ
東京都千代田区丸の内一丁目６番６号
株式会社日立製作所内
(72)発明者　増田 峰義
東京都千代田区丸の内一丁目６番６号
株式会社日立製作所内
審査官　西間木 祐紀

最終頁に続く

(54)【発明の名称】　外れ値検出装置及び方法

(57)【特許請求の範囲】
【請求項１】
外れ値検出器と、
外れ値判定器と
を有し、
前記外れ値検出器が、ウィンドウ生成器と、一又は複数種類の外れ値サブ検出器とを有し、
前記ウィンドウ生成器が、指定されたウィンドウ長を有する第１の処理ウィンドウ及び第２の処理ウィンドウを生成し、前記第１の処理ウィンドウに対して相対的に前記第２の処理ウィンドウを指定されたスライディング調整長分スライドするスライディング調整を行い、
一又は複数種類の外れ値サブ検出器のうちの一種類以上の外れ値サブ検出器の各々が、実際の値の時系列である実際時系列データのうち前記第１の処理ウィンドウに対応したデータ部分である実際時系列データセットの全体に基づく情報と、予測値の時系列である予測時系列データのうち前記スライディング調整後の第２の処理ウィンドウに対応したデータ部分である予測時系列データセットの全体に基づく情報とを用いて当該実際時系列データセットに外れ値候補があるかを検出することである分布ベースの外れ値サブ検出を行い、
前記外れ値判定器が、前記一種類以上の外れ値サブ検出器の外れ値サブ検出の結果に基づく外れ値候補が外れ値か判定する、
外れ値検出装置。

【請求項２】
前記実際時系列データ及び前記予測時系列データについて複数のパラメータ閾値セットがあり、
前記複数のパラメータ閾値セットの各々が、ウィンドウ長及びスライディング調整長を含むパラメータセットと、外れ値サブ検出において使用される一つ以上の閾値を含む閾値セットとを有し、
前記複数のパラメータ閾値セットの各々について、
前記ウィンドウ生成器が、当該セットにおけるウィンドウ長を有する第１及び第２の処理ウィンドウを生成し、当該生成された第１及び第２の処理ウィンドウに関し、当該パラメータセットにおけるスライディング調整長に従うスライディング調整を行い、
前記一又は複数種類の外れ値サブ検出器のうちの一種類以上の外れ値サブ検出器の各々が、当該セットにおける閾値を用いて分布ベースの外れ値サブ検出を行い、
前記外れ値判定器が、前記複数のパラメータセットの各々について得られた一種類以上の外れ値サブ検出器の外れ値サブ検出結果を基に、外れ値候補が外れ値か判定する、
請求項１に記載の外れ値検出装置。
【請求項３】
前記複数のパラメータ閾値セットにおける複数のパラメータセットが、点ベースと分布ベースのどちらの処理を行うかを表す点／分布ベース分類子を含み、
前記複数のパラメータ閾値セットの各々について、外れ値サブ検出器は、
当該セットにおける点／分布ベース分類子が分布ベースの処理を表す場合、分布ベースの外れ値サブ検出を行い、
当該セットにおける点／分布ベース分類子が点ベースの処理を表す場合、当該セットにおける閾値と、当該セットにおけるウィンドウ長を有する第１の処理ウィンドウに対応した実際時系列データセットにおける各実際値と、当該ウィンドウ長を有する第２の処理ウィンドウに対応した予測時系列データセットにおける各予測値とを基に、当該実際時系列データセットにおける各実際値が外れ値候補かを検出することである点ベースの外れ値サブ検出を行う、
請求項２に記載の外れ値検出装置。
【請求項４】
前記一種類以上の外れ値サブ検出器が、第１種の外れ値サブ検出器を含み、
前記第１種の外れ値サブ検出器が、
前記実際時系列データセットのうち、前記予測時系列データに基づき決定された値閾値より大きい実際値の数である第１の数を特定し、
前記値閾値より大きい予測値の数である第２の数を特定し、
前記分布ベースの比較として、前記第２の数に対する前記第１の数の割合を算出し、
当該算出された割合の大きさに応じて、当該実際時系列データセットに外れ値候補があるかを検出する、
請求項１に記載の外れ値検出装置。
【請求項５】
前記一種類以上の外れ値サブ検出器が、第２種の外れ値サブ検出器を含み、
前記第２種の外れ値サブ検出器が、
前記実際時系列データセットと前記予測時系列データセットとを比較することで予測値よりも大きい実際値の数を特定し、
前記実際時系列データセットにおける実際値の数に対する前記特定された数の割合を算出し、
当該算出された割合の大きさに応じて、当該実際時系列データセットに外れ値候補があるかを検出する、
請求項１に記載の外れ値検出装置。
【請求項６】
前記一種類以上の外れ値サブ検出器が、第３種の外れ値サブ検出器を含み、

前記第３種の外れ値サブ検出器が、
前記実際時系列データセットの分布である第１の分布を特定し、
前記予測時系列データセットの分布である第２の分布を特定し、
前記第１の分布と前記第２の分布との距離を算出し、
当該算出された距離の大きさに応じて、当該実際時系列データセットに外れ値候補があるかを検出する、
請求項１に記載の外れ値検出装置。
【請求項７】
外れ値検出器と、
外れ値判定器と
を有し、
前記外れ値検出器が、ウィンドウ生成器と、一又は複数種類の外れ値サブ検出器とを有し、
前記ウィンドウ生成器が、指定されたウィンドウ長を有する第１の処理ウィンドウ及び第２の処理ウィンドウを生成し、前記第１の処理ウィンドウに対して相対的に前記第２の処理ウィンドウを指定されたスライディング調整長分スライドするスライディング調整を行い、
一又は複数種類の外れ値サブ検出器のうちの一種類以上の外れ値サブ検出器の各々が、実際の値の時系列である実際時系列データのうち前記第１の処理ウィンドウに対応したデータ部分である実際時系列データセットにおける各実際値と、予測値の時系列である予測時系列データのうち前記スライディング調整後の第２の処理ウィンドウに対応したデータ部分である予測時系列データセットにおける各予測値とを基に、当該実際時系列データセットにおける各実際値が外れ値候補かを検出することである点ベースの外れ値サブ検出を行い、
前記外れ値判定器が、前記一種類以上の外れ値サブ検出器の外れ値サブ検出の結果に基づく外れ値候補が外れ値か判定する、
外れ値検出装置。
【請求項８】
前記一種類以上の外れ値サブ検出器が、第１種の外れ値サブ検出器を含み、
前記第１種の外れ値サブ検出器が、
前記予測時系列データセットのうち、前記予測時系列データに基づき決定された値閾値より大きい予測値を特定し、
前記実際時系列データセットのうち、当該特定された予測値に対応した実際値を外れ値候補から除外し、当該実際値以外を実際値候補とする、
請求項７に記載の外れ値検出装置。
【請求項９】
前記一種類以上の外れ値サブ検出器が、第２種の外れ値サブ検出器を含み、
前記第２種の外れ値サブ検出器が、
前記実際時系列データセットのうち、予測値よりも大きい実際値を外れ値候補とし、予測値以下の実際値を外れ値候補から除外する、
請求項７に記載の外れ値検出装置。
【請求項１０】
前記一種類以上の外れ値サブ検出器が、第３種の外れ値サブ検出器を含み、
前記第３種の外れ値サブ検出器が、
日時毎に、前記実際時系列データセットにおける実際値と前記予測時系列データセットにおける予測値との距離を算出し、
日時毎に、前記算出された距離の大きさに応じて、前記実際時系列データセットにおいて当該日時に対応した実際値が外れ値候補かを検出する、
請求項７に記載の外れ値検出装置。
【請求項１１】

前記複数のパラメータ閾値セットの各々について、外れ値サブ検出器は、当該セットを基に、分布ベースの外れ値サブ検出と、実際時系列データセットにおける各実際値が外れ値候補かを検出することである点ベースの外れ値サブ検出とのいずれかを行い、
外れ値サブ検出器は、
点ベースの一つ以上の外れ値サブ検出結果がある場合、当該一つ以上の外れ値サブ検出結果のＡＮＤである一つの外れ値サブ検出結果を算出し、当該一つの外れ値サブ検出結果のうち外れ値候補であることを意味する値の割合である発生率に基づき点ベースの一つの結果値を算出し、
分布ベースの一つ以上の外れ値サブ検出結果がある場合、当該一つ以上の外れ値サブ検出結果のＡＮＤである分布ベースの一つの結果値を算出し、
前記点ベースの一つの結果値と前記分布ベースの一つの結果値とを基に、外れ値候補が外れ値か否かを判定する、
請求項２に記載の外れ値検出装置。
【請求項１２】
前記一種類以上の外れ値サブ検出器の各々が、ログ情報に、外れ値サブ検出の結果を表す情報を出力し、
前記外れ値判定器が、外れ値候補が外れ値と判定されたかの判定結果を表す情報を前記ログ情報に出力し、
前記ログ情報に出力される情報は、検出又は判定の結果に関するログメッセージを含み、
前記ログ情報を基に、外れ値判定結果とログメッセージとを含んだ結果情報を表示する、
請求項１乃至１１のうちのいずれか１項に記載の外れ値検出装置。
【請求項１３】
コンピュータが、指定されたウィンドウ長を有する第１の処理ウィンドウ及び第２の処理ウィンドウを生成し、
コンピュータが、前記第１の処理ウィンドウに対して相対的に前記第２の処理ウィンドウを指定されたスライディング調整長分スライドするスライディング調整を行い、
コンピュータが、一又は複数種類の外れ値サブ検出のうちの一種類以上の外れ値サブ検出を行い、
当該一種類以上の外れ値サブ検出の各々は、実際の値の時系列である実際時系列データのうち前記第１の処理ウィンドウに対応したデータ部分である実際時系列データセットの全体に基づく情報と、予測値の時系列である予測時系列データのうち前記スライディング調整後の第２の処理ウィンドウに対応したデータ部分である予測時系列データセットの全体に基づく情報とを用いて当該実際時系列データセットに外れ値候補があるかを検出することである分布ベースの外れ値サブ検出を行うことを含み、
コンピュータが、前記一種類以上の外れ値サブ検出の結果に基づく外れ値候補が外れ値か判定する、
外れ値検出方法。
【請求項１４】
コンピュータが、指定されたウィンドウ長を有する第１の処理ウィンドウ及び第２の処理ウィンドウを生成し、
コンピュータが、前記第１の処理ウィンドウに対して相対的に前記第２の処理ウィンドウを指定されたスライディング調整長分スライドするスライディング調整を行い、
コンピュータが、一又は複数種類の外れ値サブ検出のうちの一種類以上の外れ値サブ検出を行い、
当該一種類以上の外れ値サブ検出の各々は、実際の値の時系列である実際時系列データのうち前記第１の処理ウィンドウに対応したデータ部分である実際時系列データセットにおける各実際値と、予測値の時系列である予測時系列データのうち前記スライディング調整後の第２の処理ウィンドウに対応したデータ部分である予測時系列データセットにおける各予測値とを基に、当該実際時系列データセットにおける各実際値が外れ値候補かを検出することである点ベースの外れ値サブ検出を行い、

コンピュータが、前記一種類以上の外れ値サブ検出の結果に基づく外れ値候補が外れ値か判定する、
外れ値検出方法。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、概して、外れ値を検出する技術に関する。
【背景技術】
【０００２】
ＩＴ（Information Technology）システムのデータから外れ値を自動検出する方法の一つとして、ＩＴシステムのパフォーマンス負荷をモデル化し、モデルからパフォーマンス負荷を予測し、予測されたパフォーマンス負荷を、実際のパフォーマンス負荷と比較する方法がある。実際のパフォーマンス負荷が予測されたパフォーマンス負荷から大きく乖離している場合、ＩＴシステムの異常に関連する可能性のある外れ値を検出することができる。
【０００３】
検出される外れ値は、いわゆるノイズ外れ値、すなわち、実際のＩＴシステムの異常とは関係のない外れ値が検出されることがある。
【０００４】
特許文献１は、状況依存時系列パターン検出器とユーザからの暗黙的又は明示的なフィードバックデータとから抽出した特徴量に基づいて、外れ値分類子を学習する技術を開示している。学習された外れ値分類子は、最初に識別された異常事象候補からノイズ外れ値を減らすことができる。
【先行技術文献】
【特許文献】
【０００５】
【文献】ＵＳ１０，２６１，８５１
【発明の概要】
【発明が解決しようとする課題】
【０００６】
特許文献１に開示の技術では、教師あり機械学習で外れ値分類子を学習するためにユーザからの暗黙的又は明示的なフィードバックデータが必要である。
【課題を解決するための手段】
【０００７】
外れ値検出装置が、外れ値検出器と外れ値判定器とを有する。外れ値検出器が、ウィンドウ生成器と、一又は複数種類の外れ値サブ検出器とを有する。ウィンドウ生成器が、指定されたウィンドウ長を有する第１の処理ウィンドウ及び第２の処理ウィンドウを生成し、第１の処理ウィンドウに対して相対的に第２の処理ウィンドウを指定されたスライディング調整長分スライドするスライディング調整を行う。一又は複数種類の外れ値サブ検出器のうちの一種類以上の外れ値サブ検出器の各々が、実際の値の時系列である実際時系列データのうち第１の処理ウィンドウに対応したデータ部分である実際時系列データセットと、予測値の時系列である予測時系列データのうちスライディング調整後の第２の処理ウィンドウに対応したデータ部分である予測時系列データセットとを当該外れ値サブ検出器の種類に対応した方法で比較することを含む外れ値サブ検出を行う。外れ値判定器が、一種類以上の外れ値サブ検出器の外れ値サブ検出の結果に基づく外れ値候補が外れ値か判定する。
【発明の効果】
【０００８】
本発明によれば、ユーザからのフィードバックデータを必要とした教師あり機械学習無しに、ノイズが低減された外れ値検出を実現することができる。

【図面の簡単な説明】
【０００９】
【図１】本発明の実施形態に係るノイズ低減外れ値検出装置の機能構成例を示す図である。
【図２Ａ】時系列ＤＢ内の実際時系列データテーブルの構成例を示す図である。
【図２Ｂ】時系列ＤＢ内の予測時系列データテーブルの構成例を示す図である。
【図３Ａ】パラメータ／閾値ＤＢ内のパラメータテーブルの構成例を示す図である。
【図３Ｂ】パラメータ／閾値ＤＢ内の閾値テーブルの構成例を示す図である。
【図４】スパイキング負荷閾値算出処理の流れの一例を示すフローチャートである。
【図５】外れ値検出処理の流れの一例を示すフローチャートである。
【図６】図５のＳ１１００２の流れの一例を示すフローチャートである。
【図７】図５のＳ１１００３の流れの一例を示すフローチャートである。
【図８】図５のＳ１１００４の一例を示すフローチャートである。
【図９】図５のＳ１１００５の流れの一例を示すフローチャートである。
【図１０Ａ】ログＤＢ内のウィンドウ外れ値テーブルの構成例を示す図である。
【図１０Ｂ】ログＤＢ内の外れ値判定テーブルの構成例を示す図である。
【図１０Ｃ】ログＤＢ内の閾値テーブルの構成例を示す図である。
【図１１Ａ】外れ値判定処理の流れの一例を示すフローチャートの一部である。
【図１１Ｂ】外れ値判定処理の流れの一例を示すフローチャートの残りである。
【図１２】外れ値検出結果画面の一例を示す図である。
【図１３】ノイズ低減外れ値検出装置のハードウェア構成例を示す図である。
【図１４】スライディング調整の意義の一例の説明図である。
【図１５】点ベースの予想スパイク検出の意義の一例の説明図である。
【図１６】分布ベースの予想スパイク検出の意義の一例の説明図である。
【発明を実施するための形態】
【００１０】
以下の説明では、「インターフェース装置」は、一つ以上のインターフェースデバイスでよい。当該一つ以上のインターフェースデバイスは、下記のうちの少なくとも一つでよい。
・一つ以上のＩ／Ｏ（Input/Output）インターフェースデバイス。Ｉ／Ｏ（Input/Output）インターフェースデバイスは、Ｉ／Ｏデバイスと遠隔の表示用計算機とのうちの少なくとも一つに対するインターフェースデバイスである。表示用計算機に対するＩ／Ｏインターフェースデバイスは、通信インターフェースデバイスでよい。少なくとも一つのＩ／Ｏデバイスは、ユーザインターフェースデバイス、例えば、キーボード及びポインティングデバイスのような入力デバイスと、表示デバイスのような出力デバイスとのうちのいずれでもよい。
・一つ以上の通信インターフェースデバイス。一つ以上の通信インターフェースデバイスは、一つ以上の同種の通信インターフェースデバイス（例えば一つ以上のＮＩＣ（Network Interface Card））であってもよいし二つ以上の異種の通信インターフェースデバイス（例えばＮＩＣとＨＢＡ（Host Bus Adapter））であってもよい。
【００１１】
また、以下の説明では、「メモリ」は、一つ以上の記憶デバイスの一例である一つ以上のメモリデバイスであり、典型的には主記憶デバイスでよい。メモリにおける少なくとも一つのメモリデバイスは、揮発性メモリデバイスであってもよいし不揮発性メモリデバイスであってもよい。
【００１２】
また、以下の説明では、「補助記憶装置」は、一つ以上の記憶デバイスの一例である一つ以上の補助記憶デバイスでよい。補助記憶デバイスは、典型的には、不揮発性の記憶デバイス（例えば補助記憶デバイス）でよく、具体的には、例えば、ＨＤＤ（Hard Disk Drive）、ＳＳＤ（Solid State Drive）、ＮＶＭｅ（Non-Volatile Memory Express）ドライブ、又は、ＳＣＭ（Storage Class Memory）でよい。

【００１３】
　また、以下の説明では、「記憶装置」は、メモリと補助記憶装置の少なくともメモリでよい。
【００１４】
　また、以下の説明では、「プロセッサ」は、一つ以上のプロセッサデバイスでよい。少なくとも一つのプロセッサデバイスは、典型的には、ＣＰＵ（Central Processing Unit）のようなマイクロプロセッサデバイスでよいが、ＧＰＵ（Graphics Processing Unit）のような他種のプロセッサデバイスでもよい。少なくとも一つのプロセッサデバイスは、シングルコアでもよいしマルチコアでもよい。少なくとも一つのプロセッサデバイスは、プロセッサコアでもよい。少なくとも一つのプロセッサデバイスは、処理の一部又は全部を行うハードウェア記述言語によりゲートアレイの集合体である回路（例えばＦＰＧＡ（Field-Programmable Gate Array）、ＣＰＬＤ（Complex Programmable Logic Device）又はＡＳＩＣ（Application Specific Integrated Circuit））といった広義のプロセッサデバイスでもよい。
【００１５】
　また、以下の説明では、「ｘｘｘＤＢ」又は「ｘｘｘテーブル」といった表現にて（「ＤＢ」はデータベースの略）、入力に対して出力が得られる情報を説明することがあるが、当該情報は、どのような構造のデータでもよいし（例えば、構造化データでもよいし非構造化データでもよいし）、入力に対する出力を発生するニューラルネットワークに代表されるような学習モデルでもよい。従って、「ｘｘｘＤＢ」又は「ｘｘｘテーブル」を「ｘｘｘ情報」と言うことができる。また、以下の説明において、各ＤＢ又は各テーブルの構成は一例であり、一つのＤＢ又は一つのテーブルは、二つ以上のＤＢ又は二つ以上のテーブルに分割されてもよいし、二つ以上のＤＢ又は二つ以上のテーブルの全部又は一部が一つのＤＢ又は一つのテーブルであってもよい。
【００１６】
　また、以下の説明では、「ｙｙｙ器」の表現にて機能を説明することがあるが、機能は、一つ以上のコンピュータプログラムがプロセッサによって実行されることで実現されてもよいし、一つ以上のハードウェア回路（例えばＦＰＧＡ又はＡＳＩＣ）によって実現されてもよいし、それらの組合せによって実現されてもよい。プログラムがプロセッサによって実行されることで機能が実現される場合、定められた処理が、適宜に記憶装置及び／又はインターフェース装置等を用いながら行われるため、機能はプロセッサの少なくとも一部とされてもよい。機能を主語として説明された処理は、プロセッサあるいはそのプロセッサを有する装置が行う処理としてもよい。プログラムは、プログラムソースからインストールされてもよい。プログラムソースは、例えば、プログラム配布計算機又は計算機が読み取り可能な記録媒体（例えば非一時的な記録媒体）であってもよい。各機能の説明は一例であり、複数の機能が一つの機能にまとめられたり、一つの機能が複数の機能に分割されたりしてもよい。
【００１７】
　以下、実施形態を、図面を参照して説明する。なお、以下に説明する実施形態は、特許請求の範囲に記載の発明を限定するものではない。更に、実施形態に記載されている各種の構成要素やその組み合わせは、必ずしも本発明に必須のものではない。
【００１８】
　実施形態の説明において、「外れ値」とは、互いに比較される２種類のデータ間における十分な差のことでよい。当該２種類のデータのうち、一方の種類のデータ（後述の予測時系列データ）は、予想される状態（例えば正常な状態）を表し、他方の種類のデータ（後述の実際時系列データ）は、現在の状態を表してよい。
【００１９】
　「ノイズ外れ値」とは、互いに比較される２種類のデータ間における十分な差のことでよい。但し、ここでは、当該２種類のデータのうち、一方の種類のデータは、予想される正常な状態を表すが、他方の種類のデータは、正常な状態を表すデータにおいて正確に表

現できない正常な状態における予想される変動のために生じる現在の状態を表し、問題視されるべきでないデータでよい。

【００２０】

「実際時系列データ」は、ＩＴシステム（例えば、物理的又は論理的な計算機システム）のような監視対象について得られた現在の状態を表す一種の測定データでよい。本実施形態では、実際時系列データは、パフォーマンス負荷の実測値（実際の値の一例）の時系列であるが、時系列となる実測値は、パフォーマンス負荷以外のデータ項目（例えば、温度や湿度）の実測値でもよい。

【００２１】

「予測時系列データ」は、予測される状態（例えば正常な状態）を表す一種の測定データでよい。本実施形態では、予測時系列データは、パフォーマンス負荷の予測値の時系列である。時系列となる予測値は、実測値と同様、パフォーマンス負荷以外のデータ項目の予測値でもよい。

【００２２】

「予想スパイク」とは、予測時系列データの中で、パフォーマンス負荷の値が特に高くなる期間のことでよい。

【００２３】

「距離」とは、実際時系列データと予測時系列データとの差を定量化できる尺度を指してよい。

【００２４】

「方向」とは、実際時系列データが予測時系列データよりも値が大きいか小さいかを評価するための尺度を指してよい。

【００２５】

「処理ウィンドウ」は、時系列データのうち、実際時系列データと予測時系列データを比較して外れ値結果を出力するための任意の期間を指す。処理ウィンドウの長さは、例えば時間長でよい。

【００２６】

「時系列データセット」は、時系列データのうち処理ウィンドウに対応した範囲のデータでよい。

【００２７】

図１は、実施形態に係るノイズ低減外れ値検出装置の機能構成例を示す。

【００２８】

ノイズ低減外れ値検出装置１００は、ノイズが低減された外れ値検出を行う装置である。ノイズ低減外れ値検出装置１００は、図１３に例示のハードウェア構成を有する物理的な計算機システム（一つ以上の物理計算機）でよいが、物理的な計算機システム（例えばクラウド基盤）に基づく論理的な計算機システム（例えばクラウドコンピューティングサービスシステム）でもよい。

【００２９】

ノイズ低減外れ値検出装置１００は、時系列ＤＢ２００に格納されている実際時系列データ及び予測時系列データと、パラメータ／閾値ＤＢ３００に格納されているパラメータ及び閾値を取得し、実際時系列データと予測時系列データを比較して外れ値を検出し、その外れ値を含む出力結果をディスプレイ４００に可視化する。

【００３０】

時系列ＤＢ２００には、実際時系列データと予測時系列データが格納されている。なお、詳細は、図２Ａ及び図２Ｂを参照して後に説明する。

【００３１】

パラメータ／閾値ＤＢ３００には、ノイズ低減外れ値検出装置１００のユーザによって外部から定義されたパラメータテーブルと閾値テーブルが格納されている。なお、詳細は、図３Ａ及び図３Ｂを参照して後に説明する。

【００３２】

ディスプレイ４００は、ノイズ低減外れ値検出装置１００で得られた結果を視覚化する出力装置である。

【００３３】

ノイズ低減外れ値検出装置１００は、外れ値検出器１１０と、スパイキング負荷閾値算出器１２０と、ログＤＢ１３０と、外れ値判定器１４０とを備える。外れ値検出器１１０は、ウィンドウ生成器１１１と、予想スパイク検出器１１２と、方向算出器１１３と、距離算出器１１４とを含む。

【００３４】

ノイズ低減外れ値検出装置１００は、まず、外れ値検出器１１０において、取得された実際時系列データと予測時系列データを処理する。具体的には、例えば、外れ値検出器１１０が、実際時系列データと予測時系列データをそれぞれウィンドウ生成器１１１により複数の処理ウィンドウ（複数の時系列データセット）に分割し、３種類の外れ値サブ検出器１１２～１１４により各実際時系列データセットにおける外れ値の可能性を算出する。この処理で得られた結果は、ログＤＢ１３０に格納される。更なる詳細については、外れ値検出器１１０については図５～図９を参照して後に説明し、ログＤＢ１３０については図１０を参照して後に説明する。

【００３５】

外れ値検出器１１０から得られログＤＢ１３０に格納された出力は、外れ値判定器１４０により処理される。つまり、外れ値サブ検出器１１２～１１４の結果に基づいて最終的な外れ値の判定が外れ値判定器１４０によりなされる。必要に応じてログメッセージが外れ値判定器１４０により生成される。最終的な外れ値及びログメッセージは、ログＤＢ１３０に格納され、その後、ディスプレイ４００に可視化される。更なる詳細は、外れ値判定器１４０については図１１を参照して後に説明し、ディスプレイ４００に表示される画面の構成例については図１２を参照して後に説明する。

【００３６】

予測時系列データは、予想スパイクの閾値を算出するスパイキング負荷閾値算出器１２０でさらに処理され、その処理の結果はログＤＢ１３０に格納される。更なる詳細は、図４を参照して後に説明する。

【００３７】

ノイズ低減外れ値検出装置１００により、ユーザからのフィードバックデータを必要とした教師あり機械学習無しに、ノイズが低減された外れ値検出を実現することができる。

【００３８】

時系列ＤＢ２００は、図２Ａに例示の実際時系列データテーブル２０１と、図２Ｂに例示の予測時系列データテーブル２０２とを格納する。

【００３９】

実際時系列データテーブル２０１は、図２Ａに例示するように、実際パフォーマンス負荷（パフォーマンス負荷の実測値）の時系列、つまり、実際時系列データを格納する。実際時系列データテーブル２０１は、日時Ｄ２０１０１及びパフォーマンス負荷Ｄ２０１０２といったカラムを含む。日時Ｄ２０１０１は、パフォーマンス負荷が測定された日時である実際日時（例えば当該日時を表すタイムスタンプ）を格納する。「日時」の単位は、本実施形態では年月日時分秒であるが、それよりも粗い又は細かい単位、或いは別の単位でもよい。パフォーマンス負荷Ｄ２０１０２は、パフォーマンス負荷の実測値（例えば、監視対象のＩＴシステムのパフォーマンスメトリクスを表すデータから取得された数値）を格納する。

【００４０】

予測時系列データテーブル２０２は、図２Ｂに例示するように、予測パフォーマンス負荷（パフォーマンス負荷の予測値）の時系列、つまり、予測時系列データを格納する。予測時系列データテーブル２０２は、日時Ｄ２０２０１及び予測負荷Ｄ２０２０２といったカラムを含む。日時Ｄ２０２０１は、予測パフォーマンス負荷が測定されると予測される日時である予測日時（例えば当該日時を表すタイムスタンプ）を格納する。予測負荷Ｄ２

０２０２は、パフォーマンス負荷として予測される値を格納する。予測時系列データは、任意の方法によって得られてよい。例えば、予測時系列データは、実際時系列データと過去時系列データ（例えば、過去の実際時系列データ、又は、過去に得られた予測時系列データ（予測日時が過去の日時である予測時系列データ））とのうちの少なくとも一部の時系列データを機械学習モデル（例えば、ニューラルネットワーク）に入力することにより当該機械学習モデルから出力されたデータ（又は当該データの加工後のデータ）でよい。或いは、予測時系列データは、過去時系列データ又は他のデータを基に人手により用意されたデータでもよい。

【００４１】

パラメータ／閾値ＤＢ３００は、図３Ａに例示のパラメータテーブル３０１と、図３Ｂに例示の閾値テーブル３０２とを格納する。

【００４２】

パラメータテーブル３０１は、図３Ａに示すように、定義されたパラメータを格納するテーブルである。パラメータテーブル３０１は、例えば、エントリＩＤ　Ｄ３０１０１、実際ウィンドウ長Ｄ３０１０２、予測ウィンドウ長Ｄ３０１０３、スライディング調整長Ｄ３０１０４、及び、点／分布ベース分類子Ｄ３０１０５といったカラムを含む。一つのエントリ（行）において、カラムＤ３０１０２～Ｄ３０１０５に格納される値が、それぞれパラメータである。

【００４３】

エントリＩＤ　Ｄ３０１０１は、エントリのＩＤを格納する。

【００４４】

実際ウィンドウ長Ｄ３０１０２は、実際ウィンドウ（実際時系列データの処理ウィンドウ）の長さである実際ウィンドウ長（を表す数値）を格納する。実際ウィンドウ長は、例えば、時間（例えば、分又は秒の単位）で表現されてよい。

【００４５】

予測ウィンドウ長Ｄ３０１０３は、予測ウィンドウ（予測時系列データの処理ウィンドウ）の長さである予測ウィンドウ長（を表す数値）を格納する。一つエントリにおいて、予測ウィンドウ長は、当該エントリにおける実際ウィンドウ長と同じでもよいし異なっていてもよい。実際ウィンドウ長と予測ウィンドウ長が異なる場合、所定の手法が利用されてよい（例えば、距離計算でDynamic Time Warpingという手法が利用されてよい）。

【００４６】

スライディング調整長Ｄ３０１０４は、実際ウィンドウと予測ウィンドウとの調整時間差（ずれ）の長さであるスライディング調整長（を表す数値）を格納する。スライディング調整長は、例えば、時間（例えば、分又は秒の単位）で表現されてよい。スライディング調整長の詳細は下記の通りである。

・スライディング調整長“０”は、実際ウィンドウと予測ウィンドウと間にずれが無いことを意味する。つまり、実際ウィンドウの開始日時（例えば後述のウィンドウ日時識別子）と予想ウィンドウの開始日時は、同じ日時である。

・スライディング調整長が負の値であることは、実際ウィンドウに対して予測ウィンドウが相対的に過去にスライドすることを意味する。例えば、スライディング調整長“－３０”は、実際ウィンドウの開始日時に比べて、予想ウィンドウの開始日時が３０タイムステップ（例えば３０秒）早いことを意味する。

・スライディング調整長が正の値であることは、実際ウィンドウに対して予想ウィンドウが相対的に将来にスライドすることを意味する。例えば、スライディング調整長“３０”は、実際ウィンドウの開始日時に比べて、予想ウィンドウの開始日時が３０タイムステップ（例えば３０秒）遅いことを意味する。

【００４７】

点／分布ベース分類子Ｄ３０１０５は、外れ値検出に点ベースの処理と分布ベースの処理のどちらを用いるかを表す分類子（例えば“点”又は“分布”といった値）を格納する。

【００４８】

閾値テーブル３０２は、図３Ｂに示すように、定義された閾値を格納するテーブルである。閾値テーブル３０２は、例えば、エントリＩＤ　Ｄ３０２０１、距離閾値Ｄ３０２０２、方向閾値Ｄ３０２０３、スパイク閾値Ｄ３０２０４及び発生率閾値Ｄ３０２０５といったカラムを含む。

【００４９】

エントリＩＤ　Ｄ３０２０１は、エントリのＩＤを格納する。閾値テーブル３０２のエントリ（行）が、パラメータテーブル３０１のエントリと１：１で対応する。従って、例えば、エントリＩＤ“１”をキーに、エントリＩＤ“１”を格納したパラメータテーブルエントリとエントリＩＤ“１”を格納した閾値テーブルエントリとが特定される。エントリＩＤ“１”に対応した種々のパラメータを用いた処理について、エントリＩＤ“１”に対応した種々の閾値が使用される。

【００５０】

距離閾値Ｄ３０２０２は、実際時系列データセットと予測時系列データセットとの距離の閾値である距離閾値を格納する。外れ値候補の評価に距離の算出が必要ない場合には、距離閾値は不要（例えば未定義）でよい。

【００５１】

方向閾値Ｄ３０２０３は、実際時系列データセットと予測時系列データセットとの間の方向の閾値である方向閾値を格納する。「方向」は、例えば、実際時系列データセットと予測時系列データセットとの間において予測パフォーマンス負荷よりも大きい実際パフォーマンス負荷が相対的に多いか否かに依存する。方向閾値は、使用する方向算出方法に合わせて、どのような閾値であってもよい。方向が距離算出において既に得られている場合や、外れ値候補の評価に方向の算出が必要ない場合には、方向閾値は不要（例えば未定義（例えば“０”という値））でよい。

【００５２】

スパイク閾値Ｄ３０２０４は、予想スパイクの閾値であるスパイク閾値を格納する。予想スパイクは、予測時系列データセットから特定され、外れ値候補の評価に用いられる。外れ値候補の評価に予想スパイクが必要ない場合には、スパイク閾値は不要（例えば未定義（例えば“０”という値））でよい。

【００５３】

発生率閾値Ｄ３０２０５は、点ベースの処理において得られる真値の発生率（全てのブール値のうちの真値の割合）の閾値である発生率閾値を格納する。エントリに対応した処理が分布ベースの処理の場合、当該エントリにおいて発生率閾値は不要（例えば未定義（例えば“Ｎｏｎｅ”という値））でよい。

【００５４】

以下、本実施形態において行われる処理の例を説明する。

【００５５】

図４は、スパイキング負荷閾値算出処理の流れの一例を示すフローチャートである。スパイキング負荷閾値算出処理は、スパイキング負荷閾値算出器１２０により行われる処理である。

【００５６】

Ｓ１２００１では、スパイキング負荷閾値算出器１２０は、時系列ＤＢ２００から予測時系列データを取得する。

【００５７】

Ｓ１２００２では、スパイキング負荷閾値算出器１２０は、Ｓ１２００１で取得された予測時系列データ全体について、平均値と標準偏差とを算出する。

【００５８】

Ｓ１２００３では、スパイキング負荷閾値算出器１２０は、ステップＳ１２００２で得られた平均値と標準偏差から、スパイキング負荷閾値を算出する。スパイキング負荷閾値の例は、平均値にｋ倍の標準偏差を加えた値である。

【００５９】

Ｓ１２００４では、スパイキング負荷閾値算出器１２０は、Ｓ１２００３で算出されたスパイキング負荷閾値を予想スパイク検出器１１２に送信するとともに、当該スパイキング負荷閾値をログＤＢ１３０に保存する。

【００６０】

スパイキング負荷閾値は、このように予測時系列データを基に決定されてよい。予測時系列データは、過去の時系列データを基にしたデータであり、期待される実際時系列データ（実際時系列データへの期待値）に相当するため、このような予測時系列データを基にどういうタイミングでスパイクが期待されるかがスパイキング負荷閾値算出器１２０により自動計算される。なお、スパイキング負荷閾値は、手動設定されてもよい。

【００６１】

図５は、外れ値検出処理の流れの一例を示すフローチャートである。外れ値検出処理は、外れ値検出器１１０により行われる処理である。なお、この処理における実際時系列データ及び予測時系列データは、任意のタイミングで例えば時系列ＤＢ２００から外れ値検出器１１０により取得されてよい。また、実際時系列データ及び予測時系列データは、同一期間分のデータを含む。

【００６２】

Ｓ１１００１では、外れ値検出器１１０は、パラメータ／閾値ＤＢ３００に定義された全てのエントリＩＤを取得する。そして、以下のＳ１１００２～Ｓ１１００５が、Ｓ１１００１で取得されたエントリＩＤ毎に実行される。Ｓ１１００２～Ｓ１１００５を、一つのエントリＩＤを例に取り説明する。

【００６３】

Ｓ１１００２では、ウィンドウ生成器１１１が、実際ウィンドウ（第１の処理ウィンドウの一例）と予測ウィンドウ（第２の処理ウィンドウの一例）とを生成する。

【００６４】

Ｓ１１００３では、予想スパイク検出器１１２が、予想される負荷のスパイクを検出する。

【００６５】

Ｓ１１００４では、方向算出器１１３が、方向を算出する。

【００６６】

Ｓ１１００５では、距離算出器１１４が、距離を算出する。

【００６７】

図６は、図５のＳ１１００２の流れの一例を示すフローチャートである。

【００６８】

Ｓ１１１０１では、ウィンドウ生成器１１１が、エントリＩＤに対応したパラメータ（実際ウィンドウ長、予測ウィンドウ長、スライディング調整長）をパラメータ／閾値ＤＢ３００から取得する。

【００６９】

Ｓ１１１０２では、ウィンドウ生成器１１１が、実際ウィンドウ（例えばローリングウィンドウ）を生成する。実際ウィンドウの長さが、Ｓ１１１０１で取得された実際ウィンドウ長である。

【００７０】

Ｓ１１１０３では、ウィンドウ生成器１１１が、予測ウィンドウ（例えばローリングウィンドウ）を生成する。予測ウィンドウの長さが、Ｓ１１１０１で取得された予測ウィンドウ長である。

【００７１】

Ｓ１１１０４では、ウィンドウ生成器１１１が、エントリＩＤが表すスライディング調整長と同じ長さ分、実際ウィンドウに対して予測ウィンドウを相対的にスライドさせる。このように、ウィンドウ生成器１１１が、実際ウィンドウに対して相対的に予測ウィンドウをスライドさせることであるスライディング調整を行う。

【００７２】

実際ウィンドウを用いて得られる複数の実際時系列データセットに対応した複数の期間は、互いに非重複の連続した期間でよいが、期間同士で一部が重複してもよい。例えば、実際ウィンドウ長が“３０”の場合、実際時系列データの先頭から３０に相当するデータが先頭の実際時系列データセット（先頭の実際ウィンドウ）であり、次の３０に相当するデータが次の実際時系列データセット（次の実際ウィンドウ）でよい。実際時系列データのうち、実際ウィンドウに対応した範囲のデータが、実際時系列データセットである。実際ウィンドウを用いて複数の実際時系列データセットが得られるため、実際時系列データセット毎に、実際ウィンドウが存在すると言うことができる。各実際ウィンドウの開始日時は、当該実際ウィンドウに対応する実際時系列データセットの開始日時である。

【００７３】

予測ウィンドウを用いて得られる複数の予測時系列データセットに対応した複数の期間は、互いに非重複の連続した期間でよいが、期間同士で一部が重複してもよい。予測時系列データのうち、予測ウィンドウに対応した範囲のデータが、予測時系列データセットである。予測ウィンドウを用いて複数の予測時系列データセットが得られるため、予測時系列データセット毎に、予測ウィンドウが存在すると言うことができる。各予測ウィンドウの開始日時は、当該予測ウィンドウに対応する予測時系列データセットの開始日時である。

【００７４】

Ｓ１１１０２で生成された実際ウィンドウと、Ｓ１１１０３で生成された予測ウィンドウは、ウィンドウセット（ウィンドウのペア）を構成する。従って、当該実際ウィンドウに対応した実際時系列データセットと、当該予測ウィンドウに対応した予測時系列データセットも、ペアを構成し、当該ペアを構成するデータセット間で比較がされることになる。

【００７５】

スライディング調整の意義の一例は、例えば図１４に示す通りである。ＩＴシステムでの所定の処理（例えばバッチ処理）の開始が定刻通りであれば、破線が示す予測時系列データの通りの日時においてスパイクが生じるはずである。しかし、当該所定の処理の開始が定刻より早いといった原因により、実線が示す実際時系列データの通り、スパイクの予想日時よりも早い日時にスパイクが生じる。一比較例では、スパイクの予想日時と異なる日時で生じたスパイクは外れ値として検出され得る。当該日時において、実際パフォーマンス負荷と予測パフォーマンス負荷の差分が大きいためである。しかし、この外れ値はノイズ外れ値である。なぜなら、発生日時が異なるが予想スパイクの発生は異常ではないためである。本実施形態では、上述したスライディング調整がなされることで、スパイクの予想日時とスパイクの実際の日時とを相対的に重ねることができ、以って、このようなスパイク（ノイズ外れ値）を外れ値として検出することを避けること、つまり、ノイズを低減することができる。

【００７６】

図７は、図５のＳ１１００３の流れの一例を示すフローチャートである。

【００７７】

Ｓ１１２０１では、予想スパイク検出器１１２が、パラメータ／閾値ＤＢ３００から、エントリＩＤに対応する点／分布ベース分類子とスパイク閾値とを取得する。

【００７８】

Ｓ１１２０２では、予想スパイク検出器１１２が、Ｓ１１２０１で取得されたスパイク閾値が定義された値であるか否かを判定する。判定結果がＹｅｓの場合、処理はＳ１１２０３に進む。判定結果がＮｏの場合（例えばスパイク閾値の値が未定義値の場合）、処理が終了する。

【００７９】

Ｓ１１２０３～Ｓ１１２１１は、実際ウィンドウと予測ウィンドウとのウィンドウセット（ペア）毎に実行される。Ｓ１１２０３～Ｓ１１２１１の説明では、一つのウィンドウセットを例に取る。なお、当該ウィンドウセットについて、実際ウィンドウと予測ウィンドウとのスライディング調整長はゼロであることもあるしゼロよりも小さい（負の値）又は大きい（正の値）であることもある。従って、一つウィンドウセットについて、実際ウ

ィンドウ（実際時系列データセット）における日時と予測ウィンドウ（予測時系列データセット）における日時とが「対応」するとは、それらの日時が同じ日時（例えば両方とも“2019-12-01 10:00:00”）であることもあれば、スライディング調整長分相対的にずれた日時であること（例えば一方の日時が“2019-12-01 10:00:00”であり他方の日時が“2019-12-01 10:00:30”であること）もある。従って、実際パフォーマンス負荷と予測パフォーマンス負荷との対応関係（言い換えれば、当該実際パフォーマンス負荷の実際日時と予測パフォーマンス負荷の予測日時との差）も、このようなスライディング調整長（時間差）に従う。

【００８０】

Ｓ１１２０３では、予想スパイク検出器１１２が、Ｓ１１２０１で取得された点／分布ベース分類子が“点”か否かを判定する。この判定結果がＹｅｓの場合、Ｓ１１２０４～Ｓ１１２０６が実行される。この判定結果がＮｏの場合（つまり、Ｓ１１２０１で取得された点／分布ベース分類子が“分布”の場合）、Ｓ１１２０７～Ｓ１１２１１が実行される。

【００８１】

Ｓ１１２０４では、予想スパイク検出器１１２が、ブール真値で構成されたブール系列（つまり、全てのブール値が真値“１”であるブール系列）を生成する。ブール系列は、実際ウィンドウ長分の長さを有し、実際ウィンドウ長分の期間を構成する複数の日時に対応した複数のブール値で構成される。

【００８２】

Ｓ１１２０５では、予想スパイク検出器１１２が、Ｓ１１２０４で生成されたブール系列に対応する複数の日時の各々について、当該日時に対応した予想日時の予測パフォーマンス負荷（予測時系列データセットにおける予測パフォーマンス負荷）がスパイキング負荷閾値よりも大きい値である場合、当該日時に対しブール偽値を付与する。つまり、ブール系列のうち、スパイキング負荷閾値よりも大きい予測パフォーマンス負荷に対応したブール値がブール偽値に変わる。

【００８３】

Ｓ１１２０６では、予想スパイク検出器１１２が、Ｓ１１２０５の処理後のブール系列をログＤＢ１３０（ウィンドウ外れ値テーブル１３１の点ベーススパイク結果リスト）に追加する。

【００８４】

図７のＳ１１２０７では、予想スパイク検出器１１２が、実際時系列データセットのうち、スパイキング負荷閾値を超えている実際パフォーマンス負荷の数をカウントする。

【００８５】

Ｓ１１２０８では、予想スパイク検出器１１２が、予測時系列データセットのうち、スパイキング負荷閾値を超えている予測パフォーマンス負荷の数をカウントする。

【００８６】

Ｓ１１２０９では、予想スパイク検出器１１２が、Ｓ１１２０７でカウントされた実際パフォーマンス負荷の数を、Ｓ１１２０８でカウントされた予測パフォーマンス負荷の数で除算することで、パーセンテージを算出する。

【００８７】

Ｓ１１２１０では、予想スパイク検出器１１２が、Ｓ１１２０９で算出されたパーセンテージが、Ｓ１１２０１で取得されたスパイク閾値よりも大きい場合、ブール真値を返す。一方、Ｓ１１２０９で算出されたパーセンテージが、Ｓ１１２０１で取得されたスパイク閾値以下の場合、予想スパイク検出器１１２が、ブール偽値を返す。

【００８８】

Ｓ１１２１１では、予想スパイク検出器１１２が、ブール値（Ｓ１１２１０で返した値）を、ログＤＢ１３０（ウィンドウ外れ値テーブル１３１の分布ベーススパイク結果リスト）に追加する。

【００８９】

以上のようにして、予想スパイク検出器１１２が、予想スパイク検出という観点での外

れ値サブ検出を、点ベース又は分布ベースで行う。分布ベースのアプローチでは、時系列データのうちのデータセット（ウィンドウに対応したデータ部分）が一つの群れ（まとまり）とみなされる。具体的には、実際時系列データセットと予測時系列データセットとの比較に際して、時点毎にパフォーマンス負荷が比較されるのではなく、該当実際パフォーマンス負荷（スパイキング負荷閾値を超えている実際パフォーマンス負荷）の数と、該当予測パフォーマンス負荷（スパイキング負荷閾値を超えている予測パフォーマンス負荷）の数とが比較される。スパイキング負荷閾値は、予測時系列データから算出された閾値であり、且つ、予測時系列データは、実際時系列データと比較される正常な状態を表すデータである。このため、分布ベースの適切な予想スパイク検出が期待される。

【００９０】

点ベースの予想スパイク検出の意義の一例は、図１５に示す通りである。一般に、予測パフォーマンス負荷は、過去の実測パフォーマンス負荷の平均等に基づくため、実際パフォーマンス負荷のスパイクよりも小さい傾向にある。このため、スパイクと検出され得る程に実際パフォーマンス負荷が予測パフォーマンス負荷との差分が大きくても、当該予測パフォーマンス負荷がスパイキング負荷閾値より大きければ、当該スパイクは予定されていたスパイクであるためノイズ外れ値である。Ｓ１１２０４～Ｓ１１２０６に従う点ベースの予想スパイク検出によれば、このようなノイズ外れ値を外れ値として検出する可能性を低減することができる。

【００９１】

分布ベースの予想スパイク検出の意義の一例は、図１６に示す通りである。実測パフォーマンス負荷とそれに対応する予測パフォーマンス負荷との差分がスパイクと判定される程に大きい日時の数が多いことがあり得る。しかし、そのような大きな差分が、予測時系列データセットの精度が低い等のような予め知られている理由により生じた差分の場合、そのような差分に属する実際パフォーマンス負荷がノイズ外れ値である可能性は高い。Ｓ１１２０７～Ｓ１１２１１に従う分布ベースの予想スパイク検出によれば、このような多くの差分に関わる多くのノイズ外れ値を外れ値として検出する可能性を低減することができる。

【００９２】

図８は、図５のＳ１１００４の流れの一例を示すフローチャートである。

【００９３】

Ｓ１１３０１では、方向算出器１１３が、パラメータ／閾値ＤＢ３００から、エントリＩＤに対応するポイント／分布ベース分類子と方向閾値とを取得する。

【００９４】

Ｓ１１３０２では、方向算出器１１３が、方向閾値が定義された値か否かを判定する。この判定結果がＹｅｓの場合、処理がＳ１１３０３に進む。この判定結果がＮｏの場合、処理が終了する。

【００９５】

Ｓ１１３０３～Ｓ１１３０８は、実際ウィンドウと予測ウィンドウとのウィンドウセット毎に実行される。Ｓ１１３０３～Ｓ１１３０８の説明では、一つのウィンドウセットを例に取る。

【００９６】

Ｓ１１３０３では、方向算出器１１３が、点／分布ベース分類子が“点”か否か判定する。この判定結果がＹｅｓの場合、Ｓ１１３０４～Ｓ１１３０５が実行される。この判定結果がＮｏの場合、Ｓ１１３０６～Ｓ１１３０８が実行される。

【００９７】

Ｓ１１３０４では、方向算出器１１３が、ブール値で構成されたブール系列を生成する。ブール系列は、実際ウィンドウ長分の長さを有し、実際ウィンドウ長分の期間を構成する複数の日時に対応した複数のブール値で構成される。当該複数の日時の各々について、実際パフォーマンス負荷がそれに対応する予測パフォーマンス負荷より大きければ、当該日時に対応したブール値は真値であり、実際パフォーマンス負荷がそれに対応する予測パ

フォーマンス負荷以下であれば、当該日時に対応したブール値は偽値である。
【００９８】
Ｓ１１３０５では、方向算出器１１３が、Ｓ１１３０４で生成したブール系列を、ログＤＢ１３０（ウィンドウ外れ値テーブル１３１の点ベース方向結果リスト）に追加する。
【００９９】
Ｓ１１３０６では、方向算出器１１３が、処理ウィンドウ長分の期間を構成する日時の数に対する、実際パフォーマンス負荷が予測パフォーマンス負荷よりも大きい日時の数のパーセンテージを算出する。
【０１００】
Ｓ１１３０７では、方向算出器１１３が、Ｓ１１３０６で算出されたパーセンテージが、Ｓ１１３０１で取得された方向閾値よりも大きい場合には、ブール真値を返す。一方、Ｓ１１３０６で算出されたパーセンテージが、Ｓ１１３０１で取得された方向閾値以下の場合には、方向算出器１１３が、ブール偽値を返す。
【０１０１】
Ｓ１１３０８では、方向算出器１１３が、ログＤＢ１３０（ウィンドウ外れ値テーブル１３１の分布ベース方向結果リスト）に、Ｓ１１３０７で返したブール値を追加する。
【０１０２】
以上のようにして、方向算出器１１３が、実際時系列データセットと予測時系列データセットとの差の方向（実際時系列データセットが予測時系列データセットよりも大きいという一般的な傾向があるかどうか）という観点での外れ値サブ検出を、点ベース又は分布ベースで行う。
【０１０３】
図９は、図５のＳ１１００５の流れの一例を示すフローチャートである。
【０１０４】
Ｓ１１４０１では、距離算出器１１４が、パラメータ／閾値ＤＢ３００から、エントリＩＤに対応する点／分布ベース分類子と距離閾値とを取得する。
【０１０５】
Ｓ１１４０２では、距離算出器１１４が、Ｓ１１４０１で取得された距離閾値が定義された値か否かを判定される。この判定結果がＹｅｓの場合、処理がＳ１１４０３に進む。この判定結果がＮｏの場合、処理が終了する。
【０１０６】
Ｓ１１４０３～Ｓ１１４１０は、実際ウィンドウと予測ウィンドウとのウィンドウセット毎に実行される。Ｓ１１４０３～Ｓ１１４１０の説明では、一つのウィンドウセットを例に取る。
【０１０７】
Ｓ１１４０３では、距離算出器１１４が、点／分布ベース分類子が“点”か否か判定する。この判定結果がＹｅｓの場合、Ｓ１１４０４～Ｓ１１４０６が実行される。この判定結果がＮｏの場合、Ｓ１１４０７～Ｓ１１４１０が実行される。
【０１０８】
Ｓ１１４０４では、距離算出器１１４が、日時毎に、実際パフォーマンス負荷と予測パフォーマンス負荷との距離（例えば、特徴量の差）を算出する。
【０１０９】
Ｓ１１４０５では、距離算出器１１４が、日時毎に、Ｓ１１４０４で算出された距離がＳ１１４０１で取得された距離閾値を超えている場合、当該日時についてブール真値を決定する。一方、Ｓ１１４０４で算出された距離がＳ１１４０１で取得された距離閾値以下の場合、距離算出器１１４が、当該日時についてブール偽値を決定する。このようにして、複数の日時に対応した複数のブール値で構成されたブール系列が生成される。
【０１１０】
Ｓ１１４０６では、距離算出器１１４が、当該生成されたブール系列を、ログＤＢ１３０（ウィンドウ外れ値テーブル１３１の点ベース距離結果リスト）に追加する。

【０１１１】
Ｓ１１４０７では、距離算出器１１４が、実際ウィンドウ（実際時系列データセット）と予測ウィンドウ（予測時系列データセット）を、それぞれ、同じ処理関数を用いて要約された分布に変換する。実際ウィンドウに対応した分布を「実際分布」と言い、予測ウィンドウに対応した分布を「予測分布」と言う。これらの分布の各々は、例えば、同じビンサイズのヒストグラムでよい。ビンサイズ（ビンの幅）は、パフォーマンス負荷の範囲でよく、ビンの長さは、当該範囲に属するパフォーマンス負荷の数でよい。具体的には、例えば、ビンサイズは固定の幅（例えば１０）、パフォーマンス負荷の範囲が対応するように複数のビンが用意される（例えば、ＣＰＵ使用率は０～１００％の間であり、故に、１０個のビンが必要）。
【０１１２】
Ｓ１１４０８では、距離算出器１１４が、実際分布と予測分布の間の距離を算出する。
【０１１３】
Ｓ１１４０９では、距離算出器１１４が、Ｓ１１４０８で算出された距離がＳ１１４０１で取得された距離閾値を超えている場合、ブール真値を返す。一方、Ｓ１１４０８で算出された距離がＳ１１４０１で取得された距離閾値以下の場合、距離算出器１１４が、ブール偽値を返す。
【０１１４】
Ｓ１１４１０では、距離算出器１１４が、ログＤＢ１３０（ウィンドウ外れ値テーブル１３１の分布ベース距離結果リスト）に、Ｓ１１４０９で返したブール値を追加する。
【０１１５】
以上のようにして、距離算出器１１４が、実際時系列データセットと予測時系列データセットとの距離といった観点での外れ値サブ検出を、点ベース又は分布ベースで行う。
【０１１６】
上述した各種外れ値サブ検出器は、点ベースの外れ値検出も分布ベースの外れ値検出も行うことができるが、それらのうちの一方の外れ値検出を行うようになっていなくてもよい。
【０１１７】
点／分布ベース分類子“点”を含むパラメータセットについて、点ベースの外れ値サブ検出は、実際時系列データセットにおける各実測値と、予測時系列データセットにおける各予測値とを基に、当該実際時系列データセットにおける各実測値が外れ値候補かを検出することである。外れ値候補の場合、当該外れ値候補としての実測値について、ブール真値が出力される。
【０１１８】
点ベースの外れ値サブ検出によれば、個々の実際パフォーマンス負荷について外れ値候補か否かがわかる。点ベースの予測スパイク検出（図７のＳ１１２０４～Ｓ１１２０６）については、図１５を参照して説明した通りである。点ベースの方向算出（図８のＳ１１３０４～Ｓ１１３０５）によれば、予測パフォーマンス負荷以下である実際パフォーマンス負荷を外れ値候補から除外することができる。点ベースの距離算出（図９のＳ１１４０４～Ｓ１１４０６）によれば、予測パフォーマンス負荷との距離が距離閾値以下である実測パフォーマンス負荷を外れ値候補から除外することができる。
【０１１９】
分布ベースの外れ値サブ検出によれば、実際時系列データセット全体について外れ値候補があるか否かがわかる。分布ベースの予測スパイク検出（図７のＳ１１２０７～Ｓ１１２１１）については、図１６を参照して説明した通りである。分布ベースの方向算出（図８のＳ１１３０６～Ｓ１１３０８）によれば、予測パフォーマンス負荷を超えている実際パフォーマンス負荷の割合が方向閾値以下であれば外れ値候補が無いとすることができる。分布ベースの距離算出（図９のＳ１１４０７～Ｓ１１４１０）によれば、予測分布との距離が距離閾値以下である実際分布に対応した実際時系列データセットについては外れ値候補が無いとすることができる。

【０１２０】
ログＤＢ１３０が、図１０Ａに例示のウィンドウ外れ値テーブル１３１と、図１０Ｂに例示の外れ値判定テーブル１３２と、図１０Ｃに例示の閾値テーブル１３３とを格納する。
【０１２１】
ウィンドウ外れ値テーブル１３１は、図１０Ａに示すように、例えば、ウィンドウ日時識別子Ｄ１３１０１、点ベース距離結果リストＤ１３１０２、点ベース方向結果リストＤ１３１０３、点ベーススパイク結果リストＤ１３１０４、分布ベース距離結果リストＤ１３１０５、分布ベース方向結果リストＤ１３１０６、及び分布ベーススパイク結果リストＤ１３１０７といったカラムを有する。
【０１２２】
ウィンドウ日時識別子Ｄ１３１０１は、実際ウィンドウに割り当てられたウィンドウ日時識別子（例えば、実際ウィンドウ長分の期間の開始日時を表す値）を格納する。
【０１２３】
点ベース距離結果リストＤ１３１０２は、点ベースの距離算出において出力されたブール系列のリストを格納する。点ベース方向結果リストＤ１３１０３は、点ベースの方向算出において出力されたブール系列のリストを格納する。点ベーススパイク結果リストＤ１３１０４は、点ベースの予想スパイク検出において出力されたブール系列のリストを格納する。これらのリストＤ１３１０２～Ｄ１３１０４の各々について、ウィンドウ日時識別子毎に（当該ウィンドウ日時識別子から同定される実際ウィンドウを含んだウィンドウセット毎に）、ブール系列がある。ウィンドウ日時識別子毎に、点ベースでのブール系列は、当該ウィンドウ日時識別子に対応した処理ウィンドウの長さ分の期間を構成する複数の日時に対応した複数のブール値で構成される。
【０１２４】
分布ベース距離結果リストＤ１３１０５は、分布ベースの距離算出において出力されたブール値を格納する。分布ベース方向結果リストＤ１３１０６は、分布ベースの方向算出において出力されたブール値を格納する。分布ベーススパイク結果リストＤ１３１０７は、分布ベースの予想スパイク検出において出力されたブール値を格納する。これらのリストＤ１３１０５～Ｄ１３１０７の各々について、ウィンドウ日時識別子毎に（当該ウィンドウ日時識別子から同定される実際ウィンドウを含んだウィンドウセット毎に）、ブール系列がある。ウィンドウ日時識別子毎に、分布ベースでのブール系列は、当該ウィンドウ日時識別子に対応した処理ウィンドウについて出力された一つのブール値で構成される。
【０１２５】
外れ値判定テーブル１３２は、図１０Ｂに示すように、例えば、ウィンドウ日時識別子Ｄ１３２０１、外れ値ブール値Ｄ１３２０２、ノイズブール値Ｄ１３２０３、予想スパイクブール値Ｄ１３２０４、調整ブール値Ｄ１３２０５及びログメッセージＤ１３２０６といったカラムを含む。
【０１２６】
ウィンドウ日時識別子Ｄ１３２０１は、実際ウィンドウに割り当てられた日時識別子を格納する。
【０１２７】
外れ値ブール値Ｄ１３２０２は、実際ウィンドウについて外れ値として識別された場合の結果値としてのブール真値（そうでない場合にブール偽値）を格納する。
【０１２８】
ノイズブール値Ｄ１３２０３は、実際ウィンドウについてノイズ外れ値として識別された場合の結果値としてのブール真値（そうでない場合にブール偽値）を格納する。
【０１２９】
予想スパイクブール値Ｄ１３２０４は、予測時系列データが表す予想スパイクを基に実際ウィンドウについてノイズ外れ値として識別された場合の結果値としてのブール真値（そうでない場合にブール偽値）を格納する。
【０１３０】

調整ブール値Ｄ１３２０５は、ゼロ以外のスライディング調整長を含むパラメータを基に実際ウィンドウが評価された場合に結果値としてのブール真値（そうでない場合にブール偽値）を格納する。調整ブール値Ｄ１３２０５は、ブール値に加えて又は代えて、使用されたスライディング調整長と調整の方向を表す情報（すなわち実際ウィンドウが予測ウィンドウに対して相対的に早いか遅いかについての情報とそれらのウィンドウの時間差を表す情報とを含んだ情報）を格納することもできる。

【０１３１】

ログメッセージＤ１３２０６は、ＩＴシステムの状態に関するデータから外れ値検出処理中に発見されたいくつかの情報、例えば、値が、外れ値であるか、ノイズ外れ値であるか、外れ値ではないか、さらに必要に応じて追加の詳細情報を記述したテキストメッセージを格納する。

【０１３２】

閾値テーブル１３３は、例えば、図１０Ｃに示すように、閾値情報Ｄ１３３０１及び値Ｄ１３３０２といったカラムを含む。

【０１３３】

閾値情報Ｄ１３３０１は、ノイズ低減外れ値検出装置１００において算出された付加的な閾値情報の種類ごとの説明（例えば、利便性や後の参照のための情報）を格納する。閾値情報として、例えば、スパイキング負荷閾値、点ベース調整リスト、及び、分布ベース調整リストがある。

【０１３４】

値Ｄ１３３０２は、閾値情報Ｄ１３３０１における記述に対応して割り当てられたデータ値を格納する。

【０１３５】

図１１Ａ及び図１１Ｂは、外れ値判定処理の流れの一例を示すフローチャートである。外れ値判定処理は、外れ値判定器１４０により行われる。外れ値判定処理は、外れ値検出器１１０の全ての外れ値サブ検出器１１２～１１４での処理結果を使用して最終的に外れ値を判定することを含む。外れ値判定処理は、ディスプレイ４００に出力可能な必要なログメッセージを生成することを含んでよい。

【０１３６】

Ｓ１４００１では、外れ値判定器１４０が、パラメータ／閾値ＤＢ３００を参照し、全ての点ベースエントリ（点／分布ベース分類子“点”を含む全てのエントリ）を評価する。“０”以外のスライディング調整長を含む点ベースエントリがある場合、外れ値判定器１４０が、ログＤＢ１３０の閾値テーブル１３３の点ベース調整リストにブール真値（そうでない場合はブール偽値）を追加する。一つの例として、スライディング調整長“０”を含んだ点ベースエントリについては、図１０Ｃに例示の通り、点ベース調整リストにブール偽値（［０］）が記録される。更に、点ベースエントリとして、スライディング調整長“０”を含んだ点ベースエントリの他に、“０”以外のスライディング調整長を含む点ベースエントリがある場合、閾値テーブル１３３の点ベース調整リストにブール真値が追記される（結果として、当該リストが［０，１］となる）。

【０１３７】

Ｓ１４００２では、外れ値判定器１４０が、パラメータ／閾値ＤＢ３００を参照し、全ての分布ベースエントリ（点／分布ベース分類子“分布”を含む全てのエントリ）を評価する。“０”以外のスライディング調整長を含む分布ベースエントリがある場合、外れ値判定器１４０が、ログＤＢ１３０の閾値テーブル１３３の分布ベース調整リストにブール真値（そうでない場合はブール偽値）を追加する。一つの例として、“０”以外のスライディング調整長を含んだ分布ベースエントリについては、故に、図１０Ｃに例示の通り、分布ベース調整リストにブール真値（［１］）が記録される。更に、分布ベースエントリとして、“０”以外のスライディング調整長を含んだ分布ベースエントリの他に、スライディング調整長“０”を含む分布ベースエントリがある場合、閾値テーブル１３３の分布ベース調整リストにブール偽値が追記される（結果として、当該リストが［１，０］となる）。

【０１３８】
Ｓ１４００３では、外れ値判定器１４０が、ログＤＢ１３０からウィンドウ外れ値テーブル１３１を取得する。ウィンドウ外れ値テーブル１３１におけるウィンドウ日時識別子毎に、Ｓ１４００４～Ｓ１４０１６が実行される。Ｓ１４００４～Ｓ１４００６とＳ１４００７が並行して実行されてよい。また、Ｓ１４００４～Ｓ１４００６は、対応する点ベース調整リストのブール値が“０”（偽）とされる点ベースエントリ毎（つまり、スライディング調整長“０”を含んだ点ベースエントリ毎）に行われる。Ｓ１４００４～Ｓ１４００６の説明は、一つのウィンドウ日時識別子且つ一つの点ベースエントリ（スライディング調整長“０”を含んだ点ベースエントリ）を例に取る。Ｓ１４００７の説明は、一つのウィンドウ日時識別子を例に取る。
【０１３９】
Ｓ１４００４では、外れ値判定器１４０が、ウィンドウ外れ値テーブル１３１における全ての点ベースのブール系列（すなわち、点ベースの距離、方向及びスパイクの結果リスト）のＡＮＤ関係を計算することで、単一の点ベースのブール系列を出力する。例えば、一つの日時について、全ての点ベースブール系列におけるブール値が“１”の場合、当該日時について、単一の点ベースブール系列でもブール値が“１”となる。一方、一つの日時について、全ての点ベースブール系列におけるブール値が“０”の場合、又は、それらの点ベースのブール系列にブール値として“１”と“０”が混在する場合、当該日時について、単一の点ベースブール系列ではブール値が“０”となる。
【０１４０】
Ｓ１４００５では、外れ値判定器１４０が、ステップＳ１４００４で得られた単一のブール系列について、ブール真値の発生率（当該単一のブール系列を構成するブール値の数に対する、当該単一のブール系列におけるブール真値の割合）を算出する。例えば、ウィンドウ長が“５”の場合（一つの処理ウィンドウに属する日時（時点）の数が“５”の場合）、Ｓ１４００４で出力されたブール系列は、５つのブール値で構成される。ブール系列が［１，０，１，０，１］場合、Ｓ１４００５において算出されたブール真値の発生率は６０％である。
【０１４１】
Ｓ１４００６では、外れ値判定器１４０が、ステップＳ１４００５で得られた発生率が、パラメータ／閾値ＤＢ３００の発生率閾値（点ベースエントリのエントリＩＤに対応した発生率閾値）よりも大きい場合、ブール真値（そうでない場合はブール偽値）を返す。例えば、Ｓ１４００５において算出されたブール真値の発生率は６０％であり、発生率閾値が７０％の場合、発生率の方が小さいため、ブール偽値が出力される。
【０１４２】
Ｓ１４００７では、外れ値判定器１４０が、対応する分布ベース調整リストにおけるブール値が偽である分布ベースエントリ（スライディング調整長“０”を含んだ分布ベースエントリ）について、ウィンドウ外れ値テーブル１３１における全ての分布ベースのブール系列（すなわち、分布ベースの距離、方向及びスパイクの結果リスト）のＡＮＤ関係を計算することで、単一の分布ベースのブール系列を出力する。分布ベースの処理では、一つの処理ウィンドウにつき、外れ値サブ検出の結果としてのブール値は一つのため、このＳ１４００７で出力されるブール系列は、単一のブール値で構成されている。
【０１４３】
Ｓ１４００８では、外れ値判定器１４０が、Ｓ１４００４～Ｓ１４００６のループの出力である点ベース出力とＳ１４００７の出力である分布ベースの出力とのＡＮＤ関係を計算し、最終的に外れ値のブール値を結果として返す。つまり、Ｓ１４００８では、点ベース出力としての単一のブール値と、分布ベース出力としての単一のブール値のＡＮＤ関係が計算される。
【０１４４】
Ｓ１４００９では、外れ値判定器１４０が、最終的な外れ値のブール値が真であるか否かを判定する。この判定結果がＹｅｓの場合、処理がＳ１４０１０に進む。この判定結果

がＮｏの場合、処理がＳ１４０１４に進む。また、閾値テーブル１３３の調整リストに偽値の点ベース又は分布ベースの対象が無い場合、処理が、Ｓ１４０１０に進んでよい。

【０１４５】

Ｓ１４０１０では、外れ値判定器１４０が、ログＤＢ１３０の閾値テーブル１３３の点ベース又は分布ベースの調整リストのいずれかが真値であるか否かを判定する。この判定結果がＹｅｓの場合、処理がＳ１４０１１に進む。この判定結果がＮｏの場合、処理がＳ１４０１３に進む。

【０１４６】

Ｓ１４０１１では、外れ値判定器１４０が、真値を有する点ベース又は分布ベースの調整リスト（閾値テーブル１３３におけるリスト）に対応する全ての点ベースの発生率評価結果と、分布ベースのブール結果（Ｓ１４００７の出力としてのブール系列）との間でＡＮＤ関係を計算し、その結果を外れ値のブール値の出力として返す。詳細なＡＮＤ関係計算については、ここでは説明しないが、例えば、これまで説明してきたＳ１４００４～Ｓ１４００８と同様の計算でよい。例えば、真値を有する点ベース又は分布ベースの調整リストに対応する全ての点ベースの発生率評価結果としてのブール系列は、Ｓ１４００４～Ｓ１４００６と同様に算出されてよい。なお、Ｓ１４００８とＳ１４０１１との相違点は、次の通りである。すなわち、Ｓ１４００８は、スライディング調整長“０”のエントリについての処理（スライディング調整がされないケースについての処理）であるが、Ｓ１４０１１は、“０”以外のスライディング調整長についての処理（スライディング調整がされるケースについての処理）である。

【０１４７】

Ｓ１４０１２では、外れ値判定器１４０が、Ｓ１４０１１で得られた外れ値のブール値が真であるか否かを判定する。この判定結果がＹｅｓの場合、処理がＳ１４０１３に進む。この判定結果がＮｏの場合、処理がＳ１４０１５に進む。

【０１４８】

Ｓ１４０１３では、外れ値判定器１４０が、既知の時系列情報から外れ値の重大度を算出し、ログメッセージと外れ値ブール値をログＤＢ１３０（外れ値判定テーブル１３２）に格納する。例えば、現在考慮されている処理ウィンドウ（例えばローリングウィンドウ）のウィンドウ日時識別子と対応する処理ウィンドウの実際時系列データセットと予測時系列データセットとを用いて、外れ値判定器１４０が、実際パフォーマンス負荷と予測パフォーマンス負荷との差を定量化することができる。そして、外れ値判定器１４０が、この定量化された情報に基づいて、ログメッセージを生成してよい。更に、外れ値判定器１４０が、処理ウィンドウに対応した期間に存在する予想スパイクを観察し、実際に観察されたスパイキング負荷が予測された予想スパイクよりも十分に長いために異常値として分類された実際時系列データセットを特定してもよい。

【０１４９】

Ｓ１４０１４では、外れ値判定器１４０が、スライディング調整無しに非外れ値と識別された処理ウィンドウ（現在検討されている時間枠）について、ノイズ外れ値があるかどうかがテストされる。例えば、予想スパイクが原因である距離又は方向ベースの外れ値が非外れ値として識別された場合、ノイズ外れ値が観測されてよい。そして、外れ値判定器１４０は、予測時系列に比べて実際時系列がどれだけ大きいか／小さいかという情報や、予測時系列と実際時系列で予想されるスパイクに観察される長さの違いに関する情報を提供し、ノイズ外れ値について警告するログメッセージを作成してよい。ここでは、外れ値判定器１４０が、このテスト結果に応じて、調整ブール値としての偽値と、予想スパイクブール値として真値又は偽値を決定してよい。

【０１５０】

Ｓ１４０１５では、外れ値判定器１４０が、処理ウィンドウ（現在検討されている時間枠）について、ノイズ外れ値（スライディング調整をとった非外れ値）を識別する。この場合、このような処理ウィンドウの日時識別子は、Ｓ１４００９で外れ値として識別され、その後のＳ１４０１２で、スライディング調整を考慮して非外れ値として識別されてい

る。このため、Ｓ１４００９で識別された外れ値がノイズ外れ値であることがわかる。さらに、外れ値判定器１４０が、この処理ウィンドウについての非外れ値が、予想スパイクによってノイズ外れ値になったかどうかをテストしてよい。そして、外れ値判定器１４０が、例えば、予想スパイクよりも早い又は遅い実際のスパイクについて警告するログメッセージを生成してよい。ここでは、外れ値判定器１４０が、このテスト結果に応じて、調整ブール値としての真値と、予想スパイクブール値として真値又は偽値を決定してよい。

【０１５１】

Ｓ１４０１６では、外れ値判定器１４０が、外れ値ブール値、ノイズブール値、予想スパイクブール値、調整ブール値、及び、生成したログメッセージをログＤＢ１３０の外れ値判定テーブル１３２に格納する。外れ値ブール値及びノイズブール値は、Ｓ１４００９及びＳ１４０１２の少なくとも一つの結果に従う値である。予想スパイクブール値、調整ブール値、及び、生成したログメッセージは、Ｓ１４０１４又はＳ１４０１５の結果としての値である。

【０１５２】

Ｓ１４０１７では、外れ値判定器１４０が、実際の外れ値とノイズ外れ値とを分析する（例えばいくつかの連続した処理ウィンドウに対応した期間という大きな文脈に関して分析する）。この分析は、例えば、ログＤＢ１３０（外れ値判定テーブル１３２）における外れ値ブール値、ノイズブール値、予想スパイクブール値及び調整ブール値に基づいて行われる。例えば、実際の外れ値（外れ値ブール値“１”とノイズブール値“０”又は“Ｎｏｎｅ”とに対応した処理ウィンドウでのパフォーマンス負荷）については、外れ値判定器１４０が、実際の外れ値の継続時間等の追加情報を特定してよい。また、例えば、ノイズ外れ値（ノイズブール値“１”に対応した処理ウィンドウでのパフォーマンス負荷）については、外れ値判定器１４０が、例えば、予想スパイクブール値及び調整ブール値等を基に、予想スパイクの発生パターン、及び、実際のスパイクが予想スパイクと比較してどの程度大きいかを識別してよい。実際のスパイクの大きさは、ノイズ外れ値に対応した日時識別子（及び、スライディング調整の大きさ）を基に実際時系列データから特定されてよい。予想スパイクの大きさは、ノイズ外れ値に対応した日時識別子（及び、スライディング調整の大きさ）を基に予測時系列データから特定されてよい。Ｓ１４０１７において、外れ値判定器１４０が、分析結果に基づくログメッセージを生成し、ログメッセージをログＤＢ１３０に格納してもよい。

【０１５３】

図１２は、外れ値検出結果画面の一例を示す。

【０１５４】

外れ値検出結果画面１２００は、ノイズ低減外れ値検出装置１００によりディスプレイ４００に表示されるＧＵＩ（Graphical User Interface）である。外れ値検出結果画面１２００の表示内容は、例えば、ログＤＢ１３０及び時系列ＤＢ２００から全てのログメッセージ、外れ値及び時系列情報を取得して定期的に（例えば頻繁に）更新されてよい。

【０１５５】

外れ値検出結果画面１２００は、グラフィカル可視化エリア４０１、及びログメッセージ出力エリア４０２を有する。

【０１５６】

グラフィカル可視化エリア４０１には、時系列ＤＢ２００の実際時系列データと予測時系列データとに基づき、実際パフォーマンス負荷及び予測パフォーマンス負荷の時系列が、例えばグラフで表示される。また、グラフィカル可視化エリア４０１には、ログＤＢ１３０（例えば外れ値判定テーブル１３２）に基づき特定される外れ値発生時間帯（例えば、外れ値ブール値“１”とノイズブール値“０”又は“Ｎｏｎｅ”とに対応した日時識別子の連続した範囲）が表示されてよい。

【０１５７】

ログメッセージ出力エリア４０２には、ログＤＢ１３０に格納されているログテキストメッセージが、グラフィカル可視化エリア４０１における表示の説明的な代替出力として

表示される。
【０１５８】
外れ値検出結果画面１２００は、ＧＵＩ以外のＵＩでもよい。また、外れ値検出結果画面１２００が有する表示エリアは、グラフィカル可視化エリア４０１及びログメッセージ出力エリア４０２に限らないでもよいし、それらの表示エリアは二つ以上のエリアに分離してもよいし一つの表示エリアとされてもよいし、各表示エリアは任意の位置に配置されてよい。
【０１５９】
図１１Ａ及び図１１Ｂが示した処理において、ログメッセージは、外れ値が検出された場合でも非外れ値（例えばノイズ外れ値）が検出された場合でも作成されてよい。これにより、図１２に例示の通り、ログメッセージが表示されれば、オペレータが、例えば、或る日時の正常な実際パフォーマンス負荷について、ノイズ外れ値として検出されたから正常なのか、ノイズ外れ値として検出されていないが元々正常であるのかを、区別することができる。なお、ログメッセージは、どういうステップを経て（上述したフローチャートのどのステップを経て）どのような外れ値検出結果となったのかを表すメッセージを含んでよい。
【０１６０】
図１３は、ノイズ低減外れ値検出装置１００のハードウェア構成例を示す。
【０１６１】
ノイズ低減外れ値検出装置１００は、例えば、一般的な計算機であり、メモリ５０２と、補助記憶デバイス５０３と、通信インターフェース５０４と、メディアインターフェース５０５と、入出力デバイス５０６と、それらに接続されたＣＰＵ５０１とを有する。インターフェース５０４～５０６が、それぞれインターフェースデバイスの一例である。ＣＰＵ５０１が、プロセッサの一例である。
【０１６２】
通信インターフェース５０４は、ネットワーク５０８を介して他の装置（例えば、解析対象のデータを格納する外部データベース）と通信するためのインターフェースデバイスである。
【０１６３】
メモリ５０２は、例えば、ＲＡＭ（Random Access Memory）であり、ＣＰＵ５０１が実行するプログラムやデータ等を記憶する。補助記憶デバイス５０３は、例えば、ＨＤＤ又はＳＳＤであり、ＣＰＵ５０１が実行するプログラムやＣＰＵ５０１が使用するデータ等を記憶する。外部記憶メディア５０７は、メディアインターフェース５０５に着脱可能であり、メディアインターフェース５０５は、外部記憶メディア５０７との間のデータの入出力を仲介する。
【０１６４】
コンソール５００は、入出力デバイス５０６に接続されており、入出力デバイス５０６は、コンソール５００との間で情報の入出力を行う。コンソール５００は、例えばディスプレイ４００を含む。
【０１６５】
ＣＰＵ５０１は、メモリ５０２又は補助記憶デバイス５０３に記憶されたプログラムを実行し、メモリ５０２又は補助記憶デバイス５０３に記憶されたデータを用いて各種処理を実行する。
【０１６６】
ノイズ低減外れ値検出装置１００に実装される各機能は、ＣＰＵ５０１が補助記憶デバイス５０３又はメモリ５０２に格納されたプログラムを実行することにより実現されてよい。上述したＤＢ又はテーブルといった情報は、メモリ５０２、補助記憶デバイス５０３、外部記憶メディア５０７、及び、ネットワーク５０８を介してアクセス可能な外部記憶装置のうちの少なくとも一つに格納される。
【０１６７】

以上、一実施形態を説明したが、これは本発明の説明のための例示であって、本発明の範囲をこの実施形態にのみ限定する趣旨ではない。本発明は、他の種々の形態でも実施することが可能である。

【０１６８】

例えば、ノイズ低減外れ値検出装置１００は、ＩＴシステムの運用管理のユースケースに適用されてよいが、実際時系列データと予測時系列データとの比較による同様のデータ分析が可能な他のユースケースにも適用されてよい。また、例えば、ウィンドウセット毎のループ処理は並列に行われてよい。

【０１６９】

また、例えば、点ベースの処理及び分布ベースの処理の少なくとも一つについて、予想スパイク検出、方向算出及び距離算出のうちの一部の外れ値サブ検出が無くてもよいし、予想スパイク検出、方向算出及び距離算出のうちの少なくとも一部の外れ値サブ検出に代えて又は加えて他種の外れ値サブ検出が採用されてもよい。

【０１７０】

また、例えば、外れ値検出器１１０（予想スパイク検出器１１２）は、予想スパイク検出を点ベースで行うか分布ベースで行うかを、自動で決定してよい。具体的には、例えば、スパイクの発生タイミングの差分が小さい事象を表すデータ（例えば、所定の処理の所定の開始日時と実際の開始日時との差分が許容値以下であることを表すデータ）が外れ値検出器１１０に入力された場合、外れ値検出器１１０（予想スパイク検出器１１２）は、予想スパイク検出を点ベースで行うことを決定してよい。スパイクの発生タイミングの差分が大きい事象を表すデータ（例えば、所定の処理の所定の開始日時と実際の開始日時との差分が許容値を超えていることを表すデータ）が外れ値検出器１１０に入力された場合、外れ値検出器１１０（予想スパイク検出器１１２）は、予想スパイク検出を分布ベースで行うことを決定してよい。

【０１７１】

また、例えば、スライディング調整長は、予定がされた日時と実際の日時との差を表すデータ（例えば、所定の処理の所定の開始日時と実際の開始日時との差分を表すデータ）を基に外れ値検出器１１０（予想スパイク検出器１１２）により自動決定されてよい。

【０１７２】

また、例えば、用意される外れ値サブ検出器は一種類のみでもよい。また、例えば、実際時系列データ及び予測時系列データについて用意されるエントリＩＤ（図３Ａ及び図３Ｂ参照）は一つだけでもよい。言い換えれば、それらの時系列データについて、点ベースの処理と分布ベースの処理のどちらかだけが行われてもよい。例えば、外れ値サブ検出器が一種類だけでエントリＩＤも一つだけの場合、外れ値サブ検出器の出力が外れ値判定器１４０の出力とされてもよい。また、点ベース処理及び分布ベース処理の少なくとも一方について、複数のエントリＩＤがあってもよい。また、各外れ値サブ検出器の出力として、ブール値に代えて又は加えて他種の情報が採用されてもよい。

【符号の説明】

【０１７３】

１００：ノイズ低減外れ値検出装置

【図面】

【図１】

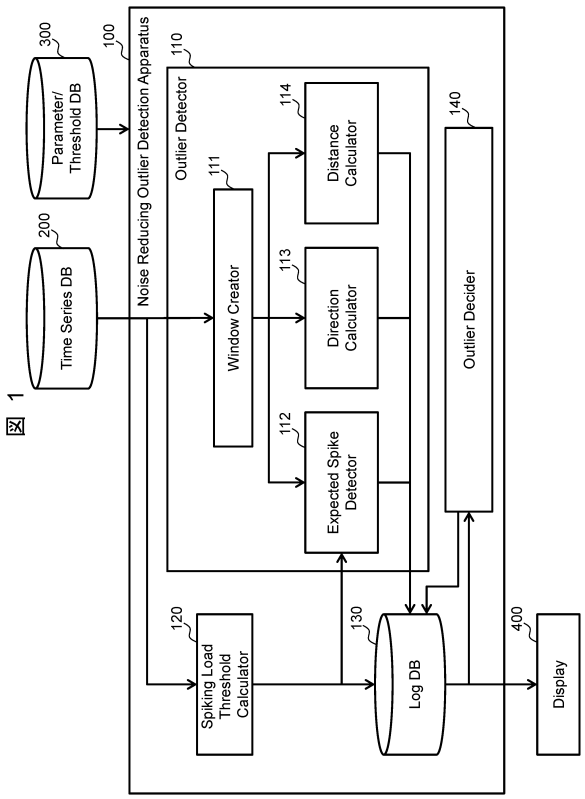


図 1

【図２Ａ】

図 2A

Real Time Series Data Table
201-1~n

D20101 Datetime	D20102 Performance Load
2019-12-01 10:00:00	20
2019-12-01 10:01:00	18
…	…

【図２Ｂ】

図 2B

Forecasted Time Series Data Table
202-1~n

D20201 Datetime	D20202 Forecasted Load
2019-12-01 10:00:00	20
2019-12-01 10:01:00	18
…	…

【図３Ａ】

図 3A

Parameter Table
301

D30101 Entry ID	D30102 Real Window Length	D30103 Forecast Window Length	D30104 Sliding Alignment Length	D30105 Point/Distribution -based Classifier
1	60	60	30	Distribution
2	120	120	0	Point
…	…	…	…	…

【図３Ｂ】

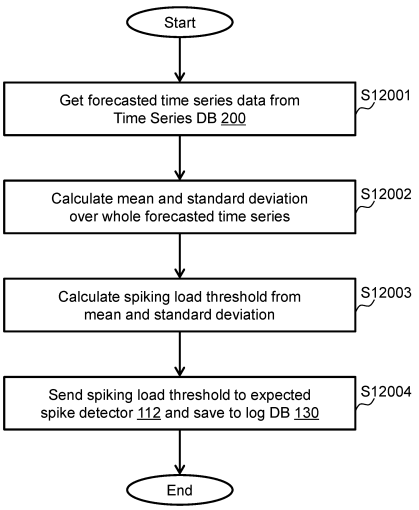
図 3B

Threshold Table
302

D30201 Entry ID	D30202 Distance Threshold	D30203 Direction Threshold	D30204 Spike Threshold	D30205 Occurrence Rate Threshold
1	0.4	70%	130%	None
2	0.7	1	1	80%
…	…	…	…	…

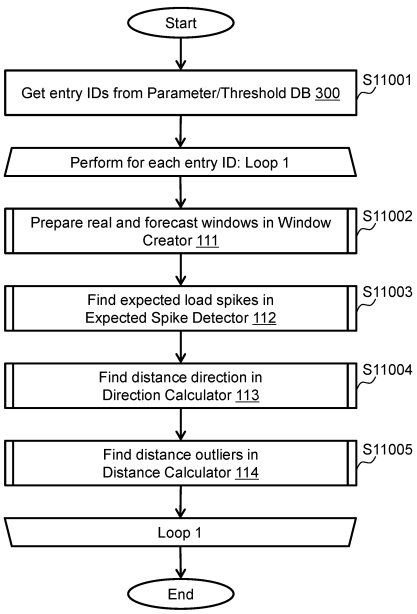
【図４】

図 4



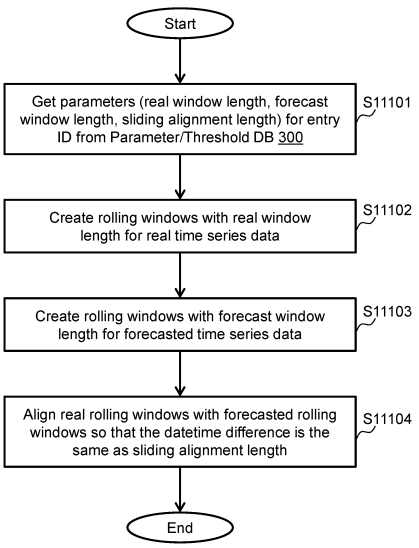
【図５】

図 5



【図６】

図 6



【図７】

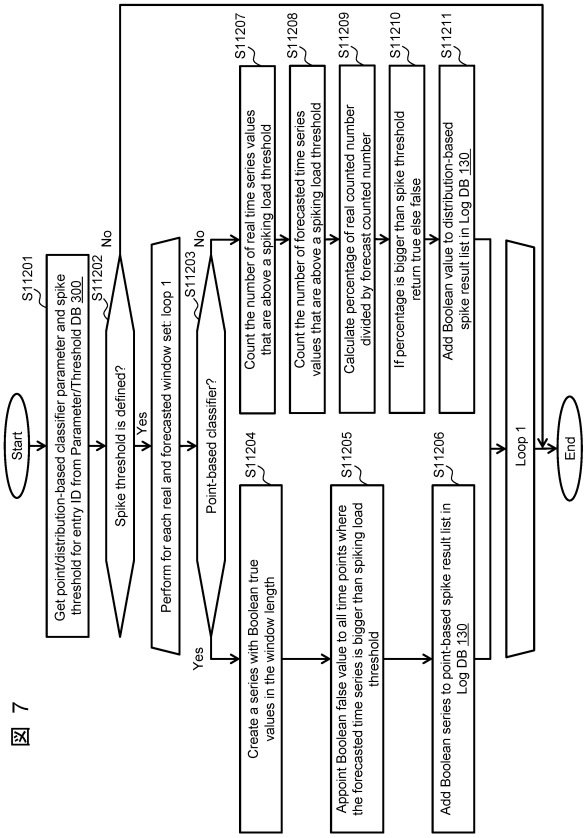


図 7

【図８】

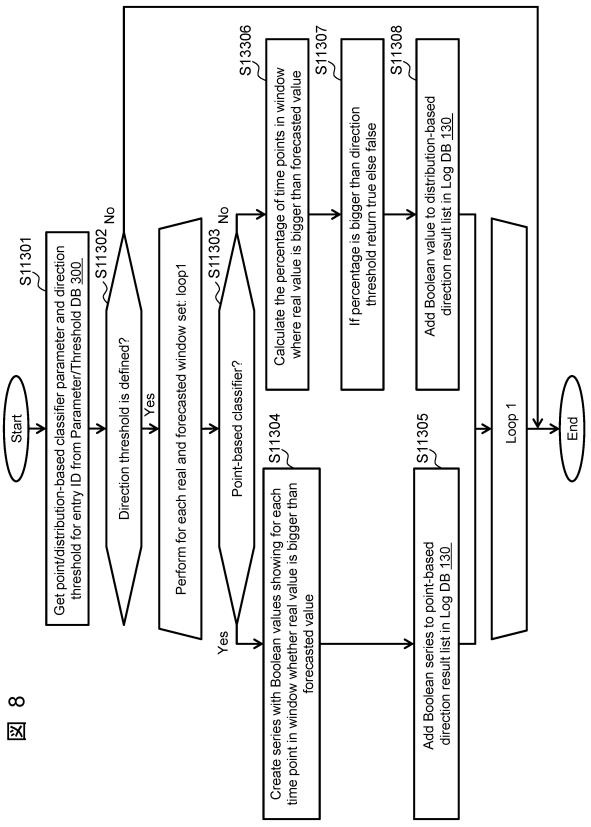


図 8

【図９】

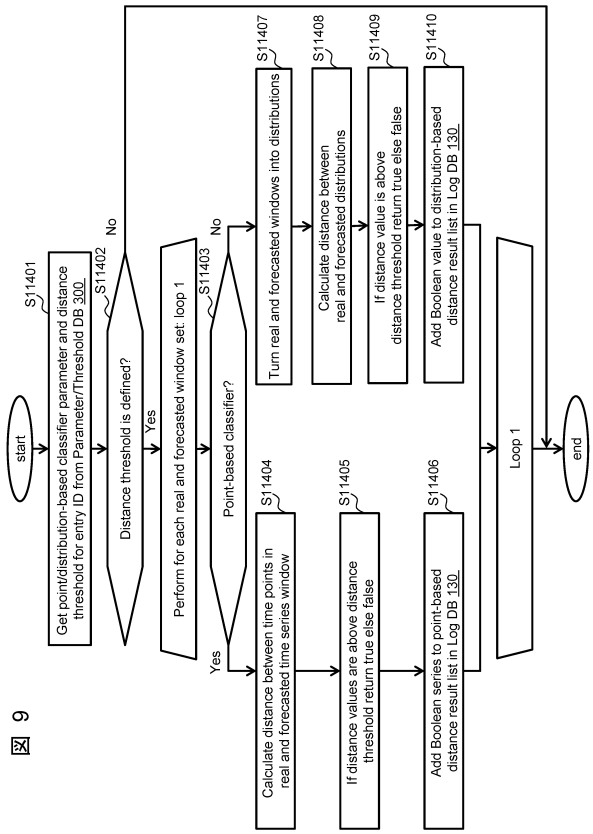


図 9

【図１０Ａ】

図 10A

Window Outlier Table 131

D13101 Window datetime identifier	D13102 Point-based Distance Result List	D13103 Point-based Direction Result List	D13104 Point-based Spike Result List	D13105 Distribution-based Distance Result List	D13106 Distribution-based Direction Result List	D13107 Distribution-based Spike Result List
2019-12-01 10:00:00	[[1,0,0,...,1]]	[[1,0,1,...,0]]	[[0,0,1,...,1]]	[1]	[1]	[1]
2019-12-01 10:10:00	[[0,0,1,...,1]]	[[0,1,1,...,1]]	[[0,0,1,...,1]]	[1]	[1]	[1]
...	...	...	...	...	...	...

【図１０Ｂ】

図 10B

Outlier Decision Table 132

D13201 Window datetime identifier	D13202 Outlier Boolean Value	D13203 Noise Boolean Value	D13204 Expected Spike Boolean Value	D13205 Aligned Boolean Value	D13206 Log Message
2019-12-01 10:00:00	0	1	1	1 (+30min)	Noisy outlier detected where real values are 30% bigger than forecast but 70% of the data being covered by an expected spike. The real spike started 30 min. earlier than expected but there is no problem.
2019-12-01 10:10:00	1	None	None	None	ALERT: Outlier detected with real values being 40% bigger for 70% of the observed time frame
...	...	...	...	...	...

【図１０Ｃ】

図 10C

Threshold Table 133

D13301 Threshold info.	D13302 Values
Spiking Load Threshold	8.7
Point-based Alignment List	[0]
Distribution-based Alignment List	[1]

【図１１Ａ】

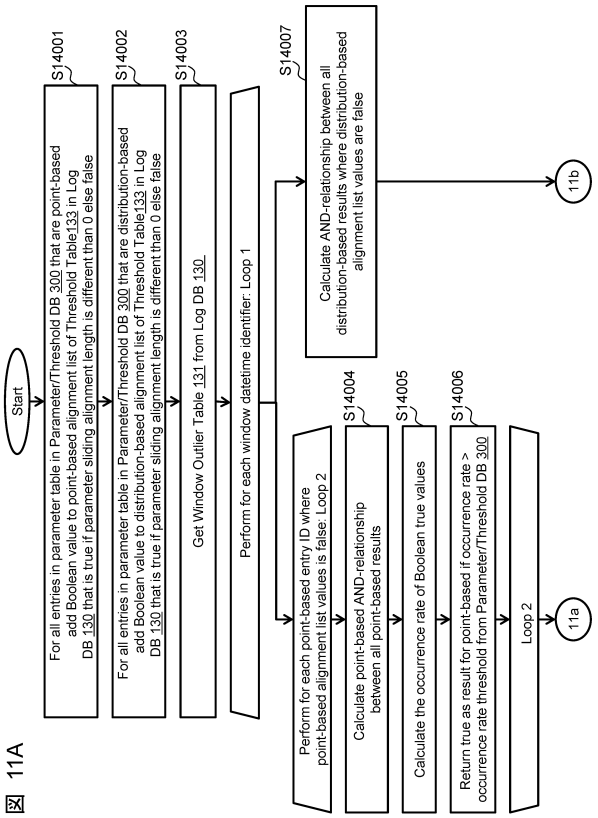


図 11A

【図１１Ｂ】

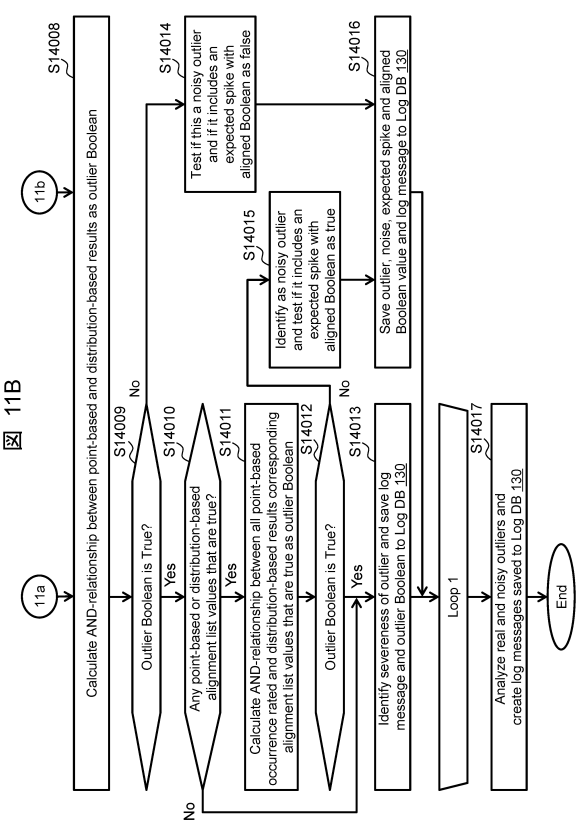


図 11B

【図１２】

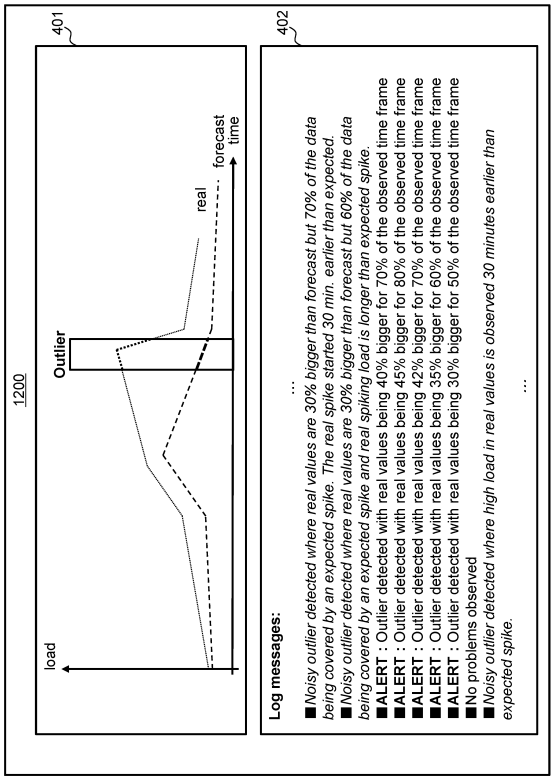


図 12

【図１３】

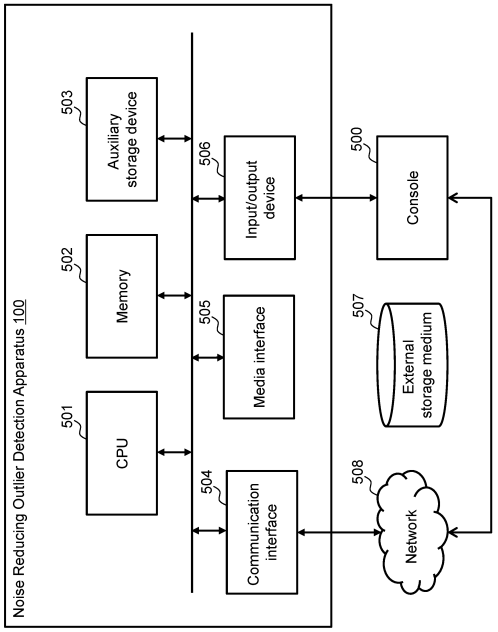


図 13

【図１４】

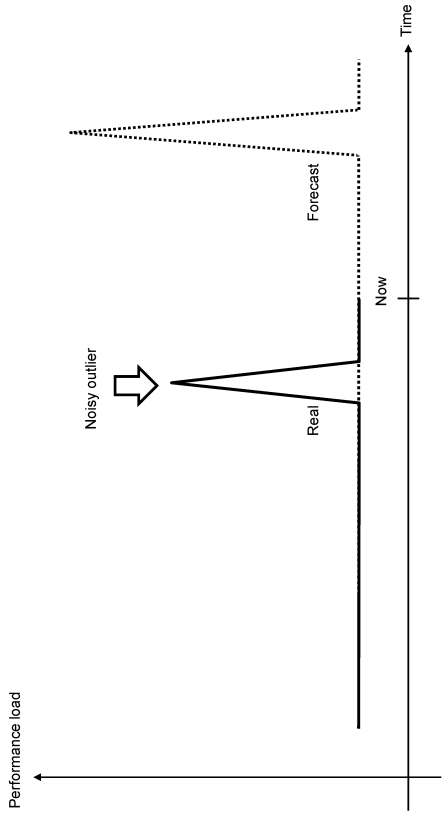


図 14

【図１５】

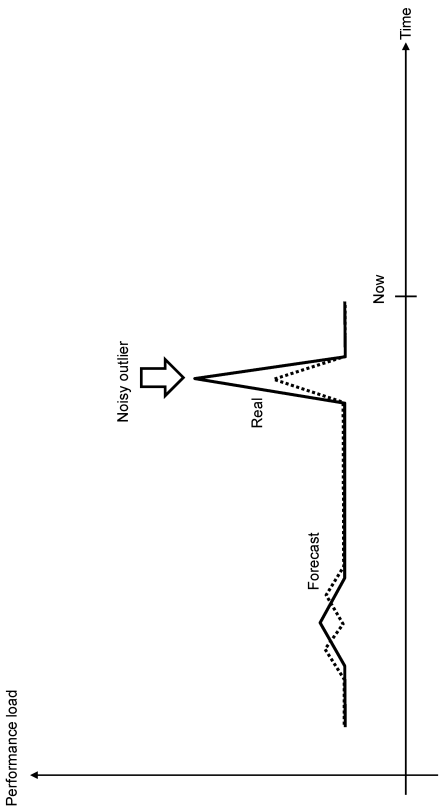
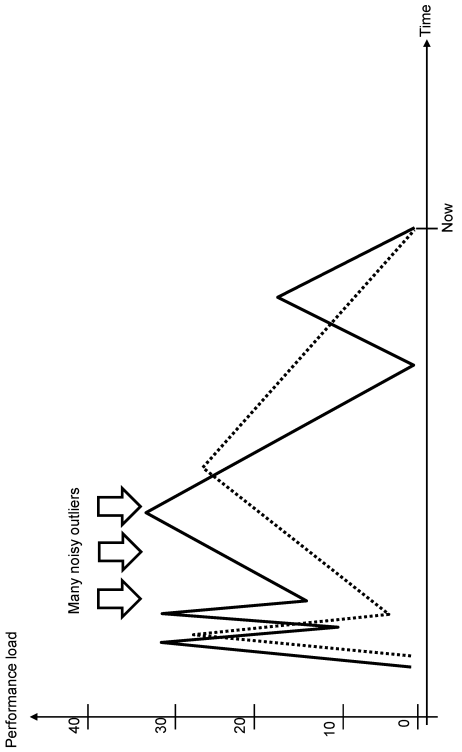


図 15

【図１６】

図 16



フロントページの続き

(56)参考文献　特開２０１１－２４７６９５（ＪＰ，Ａ）
特開２０２１－１８２２８７（ＪＰ，Ａ）
特開２０１８－１９５９２９（ＪＰ，Ａ）
特開２０１９－０８２８０１（ＪＰ，Ａ）
特開２０１５－０２６２５２（ＪＰ，Ａ）

(58)調査した分野　(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　１１／３４
Ｇ０５Ｂ　２３／０２