(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7626657号
(P7626657)

(45)発行日　令和7年2月4日(2025.2.4)

(24)登録日　令和7年1月27日(2025.1.27)

(51)国際特許分類

Ｇ０６Ｆ　11/34　(2006.01)		Ｇ０６Ｆ	11/34	１５２
Ｇ０６Ｎ　20/00　(2019.01)		Ｇ０６Ｎ	20/00	

請求項の数　11　（全21頁）

(21)出願番号　特願2021-71101(P2021-71101)
(22)出願日　令和3年4月20日(2021.4.20)
(65)公開番号　特開2022-165669(P2022-165669A)
(43)公開日　令和4年11月1日(2022.11.1)
審査請求日　令和6年2月13日(2024.2.13)

(73)特許権者　000005108
株式会社日立製作所
東京都千代田区丸の内一丁目６番６号
(74)代理人　110001678
藤央弁理士法人
(72)発明者　ジャスワル サティシュ クマル
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　児玉 崇晶

最終頁に続く

(54)【発明の名称】　異常検出装置、異常検出方法、および異常検出プログラム

(57)【特許請求の範囲】

【請求項１】

プログラムを実行するプロセッサと、前記プログラムを記憶する記憶デバイスと、を有する異常検出装置であって、

前記プロセッサは、

監視対象の時系列な第１予測データのうち特定イベントの発生時点後における第２予測データを、スケール変換により補正する補正処理と、

前記補正処理による補正後の第２予測データと、前記監視対象の時系列な第１実測データのうち前記特定イベントの発生時点後における第２実測データと、に基づいて、前記監視対象の異常を検出する検出処理と、

を実行することを特徴とする異常検出装置。

【請求項２】

請求項１に記載の異常検出装置であって、

前記補正処理では、前記プロセッサは、前記第２予測データを、前記第２実測データの変化率を用いた前記スケール変換により補正する、

ことを特徴とする異常検出装置。

【請求項３】

請求項１に記載の異常検出装置であって、

前記補正処理では、前記プロセッサは、前記第２予測データを、前記スケール変換またはシフト変換のうちいずれかの一方の線形変換により補正する、

ことを特徴とする異常検出装置。
【請求項４】
請求項３に記載の異常検出装置であって、
前記補正処理では、前記プロセッサは、前記スケール変換が選択された場合、前記第２予測データを、前記第２実測データの変化率を用いた前記スケール変換により補正し、前記シフト変換が選択された場合、前記第２予測データを、前記第２実測データの変化の差を用いた前記シフト変換により補正する、
ことを特徴とする異常検出装置。
【請求項５】
請求項４に記載の異常検出装置であって、
前記補正処理では、前記プロセッサは、前記第２実測データと前記変化率で拡縮した前記第２予測データとのスケール誤差を算出し、前記第２実測データと前記変化の差でシフトした前記第２予測データとのシフト誤差を算出し、前記スケール誤差と前記シフト誤差とに基づいて、前記スケール変換またはシフト変換のうちいずれかの一方の線形変換を選択する、
ことを特徴とする異常検出装置。
【請求項６】
請求項５に記載の異常検出装置であって、
前記補正処理では、前記プロセッサは、前記スケール誤差と前記シフト誤差とのうち誤差が小さい方の線形変換を選択する、
ことを特徴とする異常検出装置。
【請求項７】
請求項１に記載の異常検出装置であって、
前記プロセッサは、
前記第１実測データの観測時点における実測値の変化率の各々について、変化率しきい値よりも大きい特定の観測時点を前記特定イベントの発生時点候補に決定する決定処理を実行し、
前記補正処理では、前記プロセッサは、前記決定処理によって決定された前記特定イベントの発生時点候補のうちいずれかの発生時点候補の後における前記第２予測データを、前記スケール変換により補正する、
ことを特徴とする異常検出装置。
【請求項８】
請求項７に記載の異常検出装置であって、
前記決定処理では、前記プロセッサは、前記第１予測データと前記第１実測データとの間の誤差が誤差しきい値より大きい場合、前記特定の観測時点を前記特定イベントの発生時点候補に決定する、
ことを特徴とする異常検出装置。
【請求項９】
請求項７に記載の異常検出装置であって、
前記決定処理では、前記プロセッサは、前記特定の観測時点での前記実測値の変化率について所定期間内における出現回数を計数し、前記出現回数が出現回数しきい値よりも小さい前記実測値の変化率に対応する前記特定の観測時点を、前記特定イベントの発生時点候補に決定する、
ことを特徴とする異常検出装置。
【請求項１０】
プログラムを実行するプロセッサと、前記プログラムを記憶する記憶デバイスと、を有する異常検出装置が実行する異常検出方法であって、
前記プロセッサは、
監視対象の時系列な第１予測データのうち特定イベントの発生時点後における第２予測データを、スケール変換により補正する補正処理と、

前記補正処理による補正後の予測データと、前記監視対象の時系列な第１実測データのうち前記特定イベントの発生時点後における第２実測データと、に基づいて、前記監視対象の異常を検出する検出処理と、
を実行することを特徴とする異常検出方法。
【請求項１１】
プロセッサに、
監視対象の時系列な第１予測データのうち特定イベントの発生時点後における第２予測データを、スケール変換により補正する補正処理と、
前記補正処理による補正後の予測データと、前記監視対象の時系列な第１実測データのうち前記特定イベントの発生時点後における第２実測データと、に基づいて、前記監視対象の異常を検出する検出処理と、
を実行させることを特徴とする異常検出プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、異常を検出する異常検出装置、異常検出方法、および異常検出プログラムに関する。
【背景技術】
【０００２】
ＩＴ（Ｉｎｆｏｒｍａｔｉｏｎ Ｔｅｃｈｎｏｌｏｇｙ）システムにおける異常検出は、適切な行動計画が設計され、かつ、ＩＴシステムの円滑かつ効率的な操作を実行されるように、ＩＴシステムの通常とは異なる振る舞いを特定するために利用される。
【０００３】
しかし、ディスクやＲＡＭの容量の追加や削除による構成変更、大量のログファイルの予定外の削除、または、大量のログファイルのバックアップストレージへの転送のようなイベントの直後では、異常検出ができなくなる。これらのイベントが、急激な概念ドリフトを引き起こすからである。これらのイベントは、コンテナベースのマイクロサービスアーキテクチャの採用により、より頻繁に発生すると予想される。そのため、これらのイベントの直後に異常検出を可能にすることが重要である。
【０００４】
特許文献１は、予測分析のためのドリフト検出および補正のための装置、システム、方法、およびコンピュータプログラム製品を開示する。この装置では、予測モジュールは、モデルをワークロードデータに適用して、１つまたは複数の予測結果を生成する。ワークロードデータには、１つ以上のレコードが含まれる。モデルは、トレーニングデータに基づいて１つまたは複数の学習された関数を含む。ドリフト検出モジュールは、１つまたは複数の予測結果に関連するドリフト現象を検出する。予測時間修正モジュールは、ドリフト現象に応答して、少なくとも１つの予測結果を修正する。
【先行技術文献】
【特許文献】
【０００５】
【文献】米国特許出願公開第２００４／０１４８０４７号
【発明の概要】
【発明が解決しようとする課題】
【０００６】
しかしながら、上述した特許文献１では、ドリフト現象が検出された場合、予測時間修正モジュールは、ドリフト現象に応答して、少なくとも１つの予測結果を修正するが、検出後の実測データをトレーニングデータとして用いて関数の再学習が必要となる。したがって、上述したイベント発生から再学習の完了までに時間がかかり、その間、異常検出ができなかったり、再学習前の状態で異常検出をしなければならないため誤検出が発生したりする。

【０００７】
本発明は、イベント発生後における異常検出の即時性の向上を図ることを目的とする。
【課題を解決するための手段】
【０００８】
本願において開示される発明の一側面となる異常検出装置は、プログラムを実行するプロセッサと、前記プログラムを記憶する記憶デバイスと、を有する異常検出装置であって、前記プロセッサは、監視対象の時系列な第１予測データのうち特定イベントの発生時点後における第２予測データを、スケール変換により補正する補正処理と、前記補正処理による補正後の第２予測データと、前記監視対象の時系列な第１実測データのうち前記特定イベントの発生時点後における第２実測データと、に基づいて、前記監視対象の異常を検出する検出処理と、を実行することを特徴とする。
【発明の効果】
【０００９】
本発明の代表的な実施の形態によれば、イベント発生後における異常検出の即時性の向上を図ることができる。前述した以外の課題、構成及び効果は、以下の実施例の説明により明らかにされる。
【図面の簡単な説明】
【００１０】
【図１】図１は、異常検出装置による異常検出例を示す説明図である。
【図２】図２は、異常検出システムのシステム構成例を示す説明図である。
【図３】図３は、異常検出装置のハードウェア構成例を示すブロック図である。
【図４】図４は、ディスク使用率テーブルの一例を示す説明図である。
【図５】図５は、予測結果テーブルの一例を示す説明図である。
【図６】図６は、補正後予測結果テーブルの一例を示す説明図である。
【図７】図７は、異常検出装置による異常検出処理手順例を示すフローチャートである。
【図８】図８は、図７に示した概念ドリフト候補点決定処理（ステップＳ７０５）の詳細な処理手順例を示すフローチャートである。
【図９】図９は、ステップＳ８０２の算出例を示す説明図である。
【図１０】図１０は、ステップＳ８０４の算出例を示す説明図である。
【図１１】図１１は、ステップＳ８０７の算出例を示す説明図である。
【図１２】図１２は、図７に示した予測結果補正処理（ステップＳ７０６）の詳細な処理手順例（前半）を示すフローチャートである。
【図１３】図１３は、図７に示した予測結果補正処理（ステップＳ７０６）の詳細な処理手順例（後半）を示すフローチャートである。
【図１４】図１４は、概念ドリフト候補点ごとのスケールおよびシフト計算例１を示す説明図である。
【図１５】図１５は、概念ドリフト候補点ごとのスケールおよびシフト計算例２を示す説明図である。
【図１６】図１６は、概念ドリフト候補点ごとのスケールおよびシフト計算例３を示す説明図である。
【発明を実施するための形態】
【００１１】
＜異常検出例＞
図１は、異常検出装置による異常検出例を示す説明図である。監視対象は、たとえば、ＩＴインフラストラクチャのディスク使用率である。図１の（Ａ）～（Ｃ）に示すグラフにおいて、横軸は時間軸であり、縦軸はディスク使用率を示す。本実施例では、ディスク使用率を例に挙げて説明するが、ＲＡＭ使用率でもよい。訓練データは、時系列モデルの作成に用いられる時刻ｔ１までの時系列な実測値である。予測データは、時系列モデルから出力される時刻ｔ１～ｔ３までの時系列な予測値である。テストデータは、予測データと比較される時刻ｔ１～ｔ３までの時系列な実測値（正解データ）である。

【００１２】
時刻ｔ１から時刻ｔ２までは予測データとテストデータとの差は許容範囲内であり、時系列モデルによる予測が正しいことを示している。
【００１３】
（１）時刻ｔ２で、イベントが検出されたとする。ここで、イベントとは、ディスクやＲＡＭの容量の追加や削除による構成変更、大量のログファイルの予定外の削除、または、大量のログファイルのバックアップストレージへの転送のように突発的にディスク使用率が変更される挙動である。
【００１４】
このようなイベントの後は、ディスク使用率は通常の挙動を繰り返すが、時刻ｔ２を境にディスク使用率が急激に変化しているため、時刻ｔ２以降、テストデータと予測データとの間に差Ｇが生じる。このようなイベント前後のディスク使用率の挙動を概念ドリフトという。異常検出装置は、ディスク使用率の挙動が正常であるにもかかわらず、異常検出したり、ディスク使用率の挙動が異常であるにもかかわらず、異常検出しなかったりすることになる。
【００１５】
（２）このため、異常検出装置は、時刻ｔ２以降の予測データを線形変換する。（Ｂ）では、異常検出装置は、時刻ｔ２以降の予測データをシフトし、（Ｃ）では、時刻ｔ２以降の予測データをｋ倍にスケール（拡縮）する。異常検出装置は、（Ｂ）のシフト結果後の予測データとテストデータとの差（シフト誤差）と、（Ｃ）のスケール結果後の予測データとテストデータとの差（スケール誤差）と、を比較し、誤差が小さい方を選択する。図１では、シフト誤差が５％、スケール誤差が１０％であるため、シフト変換が採用される。このように、イベントが検出された時刻ｔ２の直後でも線形変換結果を用いることで、時刻ｔ２以降も異常検出が可能になる。
【００１６】
＜システム構成例＞
図２は、異常検出システムのシステム構成例を示す説明図である。異常検出システム２００は、ＩＴインフラストラクチャ２０１と、異常検出装置２０２と、を有する。ＩＴインフラストラクチャ２０１および異常検出装置２０２は、インターネット、ＬＡＮ（Ｌｏｃａｌ Ａｒｅａ Ｎｅｔｗｏｒｋ）、ＷＡＮ（Ｗｉｄｅ Ａｒｅａ Ｎｅｔｗｏｒｋ）などのネットワーク２０３を介して通信可能に接続される。
【００１７】
＜異常検出装置のハードウェア構成例＞
図３は、異常検出装置のハードウェア構成例を示すブロック図である。異常検出装置２０２は、プロセッサ３０１と、記憶デバイス３０２と、入力デバイス３０３と、出力デバイス３０４と、通信インターフェース（通信ＩＦ）３０５と、を有する。プロセッサ３０１、記憶デバイス３０２、入力デバイス３０３、出力デバイス３０４、および通信ＩＦ３０５は、バス３０６により接続される。プロセッサ３０１は、異常検出装置２０２を制御する。記憶デバイス３０２は、プロセッサ３０１の作業エリアとなる。また、記憶デバイス３０２は、各種プログラムやデータを記憶する非一時的なまたは一時的な記録媒体である。記憶デバイス３０２としては、たとえば、ＲＯＭ（Ｒｅａｄ Ｏｎｌｙ Ｍｅｍｏｒｙ）、ＲＡＭ（Ｒａｎｄｏｍ Ａｃｃｅｓｓ Ｍｅｍｏｒｙ）、ＨＤＤ（Ｈａｒｄ Ｄｉｓｋ Ｄｒｉｖｅ）、フラッシュメモリがある。入力デバイス３０３は、データを入力する。入力デバイス３０３としては、たとえば、キーボード、マウス、タッチパネル、テンキー、スキャナ、マイクがある。出力デバイス３０４は、データを出力する。出力デバイス３０４としては、たとえば、ディスプレイ、プリンタ、スピーカがある。通信ＩＦ３０５は、ネットワーク２０３と接続し、データを送受信する。
【００１８】
ここで、記憶デバイス３０２に記憶されているデータを具体的に説明する。記憶デバイス３０２は、ディスク使用率テーブル３２１と、予測結果テーブル３２２と、補正後予測

結果テーブル３２３と、異常検出プログラム３２４と、を有する。ディスク使用率テーブル３２１は、ディスク使用率を時系列に記憶したテーブルであり、図４で後述する。予測結果テーブル３２２は、予測結果を時系列に記憶したテーブルであり、図５で後述する。補正後予測結果テーブル３２３は、補正後の予測結果を時系列に記憶したテーブルであり、図６で後述する。

【００１９】

異常検出プログラム３２４は、プロセッサ３０１に、ＩＴインフラストラクチャ２０１で発生した異常を検出させるプログラムであり、異常検出モジュール３４０、収集モジュール３４１、概念ドリフト候補点決定モジュール３４２および予測結果補正モジュール３４３というプログラムモジュールを含む。異常検出モジュール３４０は、テストデータと予測データとを比較することにより、ディスク使用率の異常を検出するプログラムモジュールである。収集モジュール３４１は、ＩＴインフラストラクチャ２０１からメトリクスデータを収集するプログラムモジュールであり、図７で後述する。概念ドリフト候補点決定モジュール３４２は、概念ドリフト候補点を決定するプログラムモジュールであり、図８～図１１で後述する。概念ドリフト候補点とは、概念ドリフトを引き起こすようなイベントの検出時刻の候補である。予測結果補正モジュール３４３は、予測結果テーブル３２２に記憶された予測結果を補正するモジュールであり、図１２および図１３で後述する。

【００２０】

＜テーブル＞

図４～図６を用いて、記憶デバイス３０２に記憶されたディスク使用率テーブル３２１、予測結果テーブル３２２および補正後予測結果テーブル３２３について説明する。

【００２１】

図４は、ディスク使用率テーブル３２１の一例を示す説明図である。ディスク使用率テーブル３２１は、フィールドとして、タイムスタンプ４０１と、ディスク使用率４０２と、を有する。タイムスタンプ４０１は、ＩＴインフラストラクチャ２０１がディスク使用率４０２を計測した日付時刻である。ディスク使用率４０２は、ＩＴインフラストラクチャ２０１の全ディスク容量のうち使用中のディスクの容量の割合である。

【００２２】

図４において、タイムスタンプ４０１は、たとえば、１５分刻みで記憶される。ディスク使用率４０２は、エントリ３２５Ｘ５まで増加傾向にあり、１５分ごとに５％増加している。エントリ３２１Ｘ６において、ディスク使用率４０２が５０％になると予想されたが、ディスク使用率４０２が突然２５％に下がっている。ディスク使用率４０２が予想（５０％）の半分にまで突発的に下落した理由は、ＩＴインフラストラクチャ２０１のディスク容量が、エントリ３２１Ｘ５のタイムスタンプ４０１である「２０２１－０１－０１：００：００」後に、１０ＧＢから２０ＧＢに２倍にまで増加したからである。

【００２３】

また、エントリ３２１Ｘ７のタイムスタンプ４０１である「２０２１－０１－０１　０１：３０：００」のときに、ディスク使用率４０２が突然上昇した。エントリ３２１Ｘ７では、ディスク使用率４０２が２７．５％になると予想されたが、実測値は５０％である。このエントリ３２１Ｘ７で異常が発生したことがわかる。すなわち、エントリ３２１Ｘ７のディスク使用率４０２は異常値である。異常検出装置２０２は、エントリ３２１Ｘ７のタイムスタンプ４０１で異常が発生したことを検出することになる。

【００２４】

図５は、予測結果テーブル３２２の一例を示す説明図である。予測結果テーブル３２２は、フィールドとして、タイムスタンプ４０１と、予測ディスク使用率５０２と、下側シリーズ５０３と、上側シリーズ５０４と、を有する。予測ディスク使用率５０２は、時系列モデルに目的変数としてタイムスタンプ４０１が入力された場合に出力されるディスク使用率４０２の予測値である。タイムスタンプ４０１のほか、曜日、休日、祝祭日といったパラメータが時系列モデルに入力されてもよい。

【００２５】

下側シリーズ５０３と上側シリーズ５０４との間隔を、予測間隔という。予測間隔は、たとえば、９５％信頼区間である。下側シリーズ５０３の値および上側シリーズ５０４の値は、たとえば、そのエントリの予測ディスク使用率５０２の－３σおよび＋３σの値として算出される（σは標準偏差）。下側シリーズ５０３より小さい値および上側シリーズ５０４より大きい値は棄却域に含まれる。

【００２６】

また、予測間隔はパーセンタイルで規定されてもよい。この場合、下側シリーズ５０３は、たとえば、１０パーセンタイルであり、上側シリーズ５０４は、９０パーセンタイルである。すなわち、１～９番目に小さい予測ディスク使用率５０２およびから９１番目以降の予測ディスク使用率５０２が棄却域に含まれる。

【００２７】

下側シリーズ５０３および上側シリーズ５０４は、異常検出に用いられる。ディスク使用率４０２が、下側シリーズ５０３より低く、または、上側シリーズ５０４より高ければ、そのディスク使用率４０２は異常値として検出される。換言すれば、ディスク使用率４０２は、予測間隔に含まれていれば、そのディスク使用率４０２は異常値として検出されない。

【００２８】

図６は、補正後予測結果テーブル３２３の一例を示す説明図である。補正後予測結果テーブル３２３は、フィールドとして、タイムスタンプ４０１と、補正後予測ディスク使用率６０２と、補正後下側シリーズ６０３と、補正後上側シリーズ６０４と、を有する。補正後予測ディスク使用率６０２は、線形変換により補正された予測ディスク使用率５０２である。補正後下側シリーズ６０３は、補正後予測ディスク使用率６０２に対応する下側シリーズ５０３である。補正後上側シリーズ６０４は、補正後予測ディスク使用率６０２に対応する上側シリーズ５０４である。

【００２９】

＜異常検出処理手順例＞

図７は、異常検出装置２０２による異常検出処理手順例を示すフローチャートである。ステップ７０１～Ｓ７０４は収集モジュール３４１により実行され、ステップＳ７０５は概念ドリフト候補点決定モジュール３４２により実行され、ステップＳ７０６は予測結果補正モジュール３４３により実行される。異常検出装置２０２は、たとえば、ディスク使用率テーブル３２１に未分析のエントリが一定数蓄積されると、図７に示す処理を開始する。

【００３０】

異常検出装置２０２は、ＩＴインフラストラクチャ２０１からメトリクスデータを収集し、ディスク使用率テーブル３２１にエントリを追加する（ステップＳ７０１）。たとえば、ディスク使用率テーブル３２１の１つのエントリは、複数の時刻における各ディスク使用率４０２の統計値である。統計値とは、たとえば、複数の時刻における各ディスク使用率４０２の平均値、中央値、最大値、または最小値である。

【００３１】

ディスク使用率テーブル３２１のタイムスタンプ４０１は、１５分間隔で記録されるため、ＩＴインフラストラクチャ２０１がたとえば１分間隔でディスク使用率４０２を測定する場合、ＩＴインフラストラクチャ２０１はディスク使用率４０２を１５回測定し、１５回分のディスク使用率４０２の統計値を算出する。ＩＴインフラストラクチャ２０１は、１５回の測定うち最後の測定時刻と１５回のディスク使用率４０２の統計値とを異常検出装置２０２に送信する。異常検出装置２０２は、受信した最後の測定時刻をタイムスタンプ４０１に記録し、１５回のディスク使用率４０２の統計値をディスク使用率４０２に記録する。

【００３２】

つぎに、異常検出装置２０２は、分析開始時刻Ｔ$_{start}$を取得する（ステップＳ７０２）。分析開始時刻Ｔ$_{start}$は、異常検出装置２０２がステップＳ７０３の分析を開始

するディスク使用率テーブル３２１のエントリのタイムスタンプ４０１である。すなわち、分析開始時刻Ｔｓｔａｒｔは、ディスク使用率テーブル３２１において、未分析でかつ最古のエントリのタイムスタンプ４０１である。たとえば、ディスク使用率テーブル３２１において、エントリ３２１Ｘ４まで異常検出処理が完了しているものとすると、次のエントリ３２１Ｘ５のタイムスタンプ４０１の値「２０２１－０１－０１　０１：００：００」が分析開始時刻Ｔｓｔａｒｔとなる。

【００３３】

つぎに、異常検出装置２０２は、時系列モデルの再学習に用いられていない分析開始時刻Ｔｓｔａｒｔまでのエントリ群を時系列データとしてディスク使用率テーブル３２１から抽出し、時系列モデルを再学習する（ステップＳ７０３）。時系列モデルは、たとえば、

$$y = f(t) \cdots (1)$$

で表現される関数である。左辺のｙは、ディスク使用率４０２であり、右辺のｔは時刻データであり、たとえば、タイムスタンプ４０１である。時刻データｔとして、タイムスタンプ４０１のほか、曜日の種類（平日、休日、祝祭日）が入力されてもよい。

【００３４】

たとえば、エントリ３２１Ｘ１の１つ前までのエントリが学習済みであり、エントリ３２１Ｘ１～３２１Ｘ４が時系列モデルの学習に用いられていない分析開始時刻Ｔｓｔａｒｔまでのエントリ群であれば、異常検出装置２０２は、エントリ３２１Ｘ１～３２１Ｘ４のタイムスタンプ４０１を時系列モデルの時刻データｔに入力し、それらの出力結果ｙとエントリ３２１Ｘ１～３２１Ｘ４のディスク使用率４０２との差が最小となるように時系列モデルを再学習する。なお、時系列モデルが未生成である場合、異常検出装置２０２は、分析開始時刻Ｔｓｔａｒｔまでのエントリ群を時系列データとしてディスク使用率テーブル３２１から抽出し、時系列モデルを学習する。

【００３５】

なお、時系列モデルは、ランダムフォレストでもよく、ＡＲＩＭＡ（ＡｕｔｏＲｅｇｒｅｓｓｉｖｅ　Ｉｎｔｅｇｒａｔｅｄ　Ｍｏｖｉｎｇ　Ａｖｅｒａｇｅ）モデルでもよく、ＳＡＲＩＭＡ（Ｓｅａｓｏｎａｌ　ＡＲＩＭＡ）モデルでもよい。

【００３６】

つぎに、異常検出装置２０２は、分析開始時刻Ｔｓｔａｒｔから予測の実行を開始し、予測結果テーブル３２２にエントリを追加する（ステップＳ７０４）。具体的には、たとえば、異常検出装置２０２は、ステップＳ７０３で再学習された上記式（１）の時系列モデルの時刻データｔに、分析開始時刻Ｔｓｔａｒｔから最新時刻までのタイムスタンプ４０１を順次入力し、タイムスタンプ４０１ごとのディスク使用率４０２の予測データｐを上記式（１）のｙとして出力し、タイムスタンプ４０１とともに予測ディスク使用率５０２として予測結果テーブル３２２に記録する。また、異常検出装置２０２は、順次入力されたタイムスタンプ４０１ごとに、下側シリーズ５０３の値および上側シリーズ５０４の値を算出して、予測結果テーブル３２２に記録する。

【００３７】

ここで、異常検出モジュール３４０で異常検出されないエントリについて説明する。たとえば、エントリ３２１Ｘ７のディスク使用率４０２（５０％）は異常値である。しかし、エントリ３２１Ｘ７と同一タイムスタンプ４０１の予測結果テーブル３２２のエントリ３２２Ｐ３では、下側シリーズ５０３の値が「４５％」、上側シリーズ５０４の値が「６５％」である。

【００３８】

したがって、異常値であるエントリ３２１Ｘ７のディスク使用率４０２（５０％）は、下側シリーズ５０３の値「４５％」と上側シリーズ５０４の値「６５％」との間の予測間隔［４５，６５］に含まれる。したがって、エントリ３２１Ｘ７のディスク使用率４０２（５０％）は、異常値として検出されない。これは、再学習後の時系列モデルｆが、ＩＴインフラストラクチャ２０１のディスク容量が２倍になった影響を見逃しているからである。

【００３９】
異常検出装置２０２は、概念ドリフト候補点決定処理（ステップＳ７０５）および予測結果補正処理（ステップＳ７０６）により、予測ディスク使用率５０２を補正する。これにより、異常検出装置２０２は、異常検出モジュール３４０により、このように見逃されて正常値として扱われたエントリ３２１Ｘ７のディスク使用率４０２（５０％）を異常値として検出する（ステップＳ７０７）。
【００４０】
＜概念ドリフト候補点決定処理（ステップＳ７０５）＞
図８は、図７に示した概念ドリフト候補点決定処理（ステップＳ７０５）の詳細な処理手順例を示すフローチャートである。ステップＳ７０４のあと、異常検出装置２０２は、操作ログまたは構成管理データベース（ＣＭＤＢ）が利用可能か否かを判断する（ステップＳ８０１）。ＩＴインフラストラクチャ２０１における操作ログまたはＣＭＤＢのいずれかが利用可能である場合（ステップＳ８０１：Ｙｅｓ）、異常検出装置２０２は、概念ドリフト候補点を操作ログまたはＣＭＤＢから取得して（ステップＳ８１０）、予測結果補正処理（ステップＳ７０６）に移行する。
【００４１】
一方、操作ログおよびＣＭＤＢのいずれも利用不可能である場合（ステップＳ８０１：Ｎｏ）、異常検出装置２０２は、予測データとテストデータとの平均誤差を算出する（ステップＳ８０２）。予測データとは、予測ディスク使用率５０２であり、テストデータとは、予測データと同一タイムスタンプ４０１のディスク使用率４０２であり、正解データとも呼ばれる。平均誤差とは、タイムスタンプ４０１ごとの予測データとテストデータとの差分に基づく誤差の平均値である。たとえば、誤差は、図１の（Ａ）の差Ｇであり、下記式（２）で表現される。平均誤差は、下記式（３）で表現される。
【００４２】
【数１】

$$e(t) = \frac{|x_t - p_t|}{x_t} \quad \cdots (2)$$

$$E = \frac{1}{n} \sum_{t=T_{start}}^{t=T_{end}} \frac{|x_t - p_t|}{x_t} \quad \cdots (3)$$

【００４３】
式（２）、（３）のｘｔは、テストデータであり、ｐｔは、予測ディスク使用率５０２であり、ｔは、タイムスタンプ４０１である。ｅ（ｔ）は、タイムスタンプ４０１がｔのときの誤差であり、Ｅは、分析開始時刻ＴｓｔａｒｔからＴｅｎｄまでのｎ個の誤差ｅ（ｔ）の平均値、すなわち、平均誤差である。Ｔｅｎｄは、分析開始時刻Ｔｓｔａｒｔのエントリからｎ番目のエントリのタイムスタンプ４０１が示す分析終了時刻である。
【００４４】
つぎに、異常検出装置２０２は、ステップＳ８０２で算出した平均誤差Ｅが誤差しきい値Ｅｔｈｒｅｓｈｏｌｄよりも大きいか否かを判断する（ステップＳ８０３）。平均誤差Ｅが誤差しきい値Ｅｔｈｒｅｓｈｏｌｄよりも大きくない場合（ステップＳ８０３：Ｎｏ）、予測結果補正処理（ステップＳ７０６）に移行する。一方、平均誤差Ｅが誤差しきい値Ｅｔｈｒｅｓｈｏｌｄよりも大きい場合（ステップＳ８０３：Ｙｅｓ）、ステップＳ８０４に移行する。本例では、誤差しきい値Ｅｔｈｒｅｓｈｏｌｄを、たとえば、Ｅｔｈｒｅｓｈｏｌｄ＝０．２とする。平均誤差Ｅが誤差しきい値Ｅｔｈｒｅｓｈｏｌｄよりも大きければ、急激

な概念ドリフトが発生していると予想される。平均誤差Ｅが誤差しきい値Ｅｔｈｒｅｓｈｏｌｄよりも大きい概念ドリフトを、急激な概念ドリフトと称す。

【００４５】

平均誤差Ｅが誤差しきい値Ｅｔｈｒｅｓｈｏｌｄよりも大きい場合（ステップＳ８０３：Ｙｅｓ）、異常検出装置２０２は、テストデータの変化率ｒｔをタイムスタンプ４０１ごとに算出する（ステップＳ８０４）。変化率ｒｔは、たとえば、下記（４）で算出される。

【００４６】

【数２】

$$(r_t, T_{start-1} \leq t < T_{end}) = \begin{cases} \dfrac{x_{t+1}}{x_t}, & 0 < x_t < x_{t+1} \\ \dfrac{x_{t+1}}{1+x_t}, & 0 = x_t < x_{t+1} \\ \dfrac{x_t}{x_{t+1}}, & x_t \geq x_{t+1} > 0 \\ \dfrac{x_t}{1+x_{t+1}}, & x_t \geq x_{t+1} = 0 \end{cases} \quad \cdots (4)$$

【００４７】

変化率ｒｔは、時刻ｔのテストデータｘｔと時刻ｔ＋１のテストデータｘｔ＋１との比に基づいて算出される。異常検出装置２０２は、タイムスタンプ４０１ごとの変化率のうち、変化率しきい値Ｒｔｈｒｅｓｈｏｌｄよりも大きい特定の変化率ｒｔがあるか否かを判断する（ステップＳ８０５）。特定の変化率ｒｔが１つもない場合（ステップＳ８０５：Ｎｏ）、予測結果補正処理（ステップＳ７０６）に移行する。一方、特定の変化率ｒｔが１つ以上ある場合（ステップＳ８０５：Ｙｅｓ）、ステップＳ８０６に移行する。変化率しきい値Ｒｔｈｒｅｓｈｏｌｄは、ユーザにより任意に設定可能である。

【００４８】

異常検出装置２０２は、特定の変化率ｒｔとそのタイムスタンプ４０１が示す時刻ｔとを選択する（ステップＳ８０６）。そして、異常検出装置２０２は、特定の変化率ｒｔごとに、所定期間（たとえば、１日）においてステップＳ８０６で選択された回数をカウントする（ステップＳ８０７）。この選択された回数を特定の変化率ｒｔの出現回数ｆｒと称す。なお、特定の変化率ｒｔは、完全一致した場合にのみカウントされてもよく、許容範囲内であればカウントされてもよい。たとえば、変化率ｒｔが１．８０４や１．８６３であれば、特定の変化率ｒｔ＝１．８の出現回数ｆｒとしてカウントされる。

【００４９】

このあと、異常検出装置２０２は、出現回数ｆｒが出現回数しきい値Ｆｔｈｒｅｓｈｏｌｄより小さい特定のタイムスタンプ４０１があるか否かを判断する（ステップＳ８０８）。出現回数ｆｒが出現回数しきい値Ｆｔｈｒｅｓｈｏｌｄより小さい特定のタイムスタンプ４０１が１つもない場合（ステップＳ８０８：Ｎｏ）、予測結果補正処理（ステップＳ７０６）に移行する。

【００５０】

一方、出現回数ｆｔが出現回数しきい値Ｆｔｈｒｅｓｈｏｌｄより小さい特定のタイムスタンプ４０１が１つ以上ある場合（ステップＳ８０８：Ｙｅｓ）、異常検出装置２０２は、特定のタイムスタンプ４０１を概念ドリフト候補点に決定し（ステップＳ８０９）、予測結果補正処理（ステップＳ７０６）に移行する。特定の変化率ｒｔが出現回数しきい値Ｆｔｈｒｅｓｈｏｌｄ以上出現しているということは、特定の変化率ｒｔおよびそのタイムスタンプ４０１が概念ドリフトによるディスク使用率４０２の変化を示しているのではな

く、ディスク使用率４０２自体に通常の挙動とは異なる変化が発生していることを示している。したがって、異常検出装置２０２は、出現回数ｆｒが出現回数しきい値$F_{threshold}$より小さい特定のタイムスタンプ４０１が示す時刻で、急激な概念ドリフトが発生したらしいと予測する。

【００５１】

つぎに、概念ドリフト候補点決定処理（ステップＳ７０５）の実行例を図９～図１１を用いて説明する。

【００５２】

図９は、ステップＳ８０２の算出例を示す説明図である。図９では、分析開始時刻となるタイムスタンプ４０１は、「２０２１－０１－０１　０１：００：００」とし、分析終了時刻となるタイムスタンプ４０１は、「２０２１－０１－０１　０１：４５：００」とする。図９の場合、平均誤差Ｅは誤差しきい値$E_{threshold}$よりも大きいため（ステップＳ８０２：Ｙｅｓ）、ステップＳ８０４に移行する。

【００５３】

図１０は、ステップＳ８０４の算出例を示す説明図である。エントリ１０００は、分析開始時刻T_{start}の１つ前のタイムスタンプ４０１（すなわち、前回の分析での分析終了時刻T_{end}）の算出結果を示す。エントリ１００４では、変化率が算出されないため、ステップＳ８０４の判定は実行されない。図１０では、エントリ１００１～１００３の各変化率r_tが変化率しきい値$R_{threshold}$よりも大きい特定の変化率であることがわかる。

【００５４】

図１１は、ステップＳ８０７の算出例を示す説明図である。エントリ１１０１～１１０３では、特定の変化率ｒｔの各々の出現回数ｆｒが「１」であるため、出現回数しきい値$F_{threshold}$よりも小さい。したがって、ステップＳ８０６で選択されたエントリ１００１～１００３のタイムスタンプ４０１は、特定のタイムスタンプ４０１に決定される（ステップＳ８０９）。

【００５５】

＜予測結果補正処理（ステップＳ７０６）＞

図１２は、図７に示した予測結果補正処理（ステップＳ７０６）の詳細な処理手順例（前半）を示すフローチャートである。異常検出装置２０２は、テストデータ、予測データ、および概念ドリフト候補点を取得する（ステップＳ１２０１）。つぎに、異常検出装置２０２は、未選択の概念ドリフト候補点があるか否かを判断する（ステップＳ１２０２）。未選択の概念ドリフト候補点がある場合（ステップＳ１２０２：Ｙｅｓ）、異常検出装置２０２は、未選択の概念ドリフト候補点を１つ選択してＴとし、選択した概念ドリフト候補点Ｔの変化率r_Tを、下記式（５）を用いて算出する（ステップＳ１２０３）。

【００５６】

【数３】

$$r_T = \begin{cases} \dfrac{x_{T+1}}{x_T}, x_T \neq 0 \\ \dfrac{x_{T+1}}{1+x_T}, x_T = 0 \end{cases} \quad \cdots (5)$$

【００５７】

つぎに、異常検出装置２０２は、下記式（６）を用いて、概念ドリフト候補点Ｔのスケール誤差$e^T{}_{scale}$を算出する（ステップＳ１２０３）。

【００５８】

【数４】

$$e_{scale}^{T} = \frac{1}{\mathrm{m}} \Sigma_{t=T+1}^{t=T_{end}} \frac{|x_t - r_T * p_t|}{x_t} \cdots (6)$$

【００５９】
上記式（６）において、ｍは、タイムスタンプ４０１が示す時刻Ｔ＋１から分析終了時刻Ｔｅｎｄまでのタイムスタンプ４０１の個数である。スケール誤差ｅ^{T}ｓｃａｌｅは、予測データｐｔを概念ドリフト候補点Ｔの変化率ｒＴでスケール（拡縮）した場合のテストデータとの誤差の平均値を示す。

【００６０】
つぎに、異常検出装置２０２は、下記式（７）を用いて、変化差分ｄＴを算出する（ステップＳ１２０５）。

【００６１】
【数５】

$$d_T = x_{T+1} - x_T \cdots (7)$$

【００６２】
そして、異常検出装置２０２は、下記式（８）を用いて、シフト誤差ｅ^{T}ｓｈｉｆｔを算出して（ステップＳ１２０６）、ステップＳ１２０２に戻る。

【００６３】
【数６】

$$e_{shift}^{T} = \frac{1}{m} \Sigma_{t=T+1}^{t=T_{end}} \frac{|x_t - (p_t + d_T)|}{x_t} \cdots (8)$$

【００６４】
シフト誤差ｅ^{T}ｓｈｉｆｔは、予測データｐｔを概念ドリフト候補点Ｔの変化差分ｄＴでシフト（加減算）した場合のテストデータとの誤差の平均値を示す。そして、ステップＳ１２０２において、未選択の概念ドリフト候補点がない場合（ステップＳ１２０２：Ｎｏ）、異常検出装置２０２は、最小誤差およびそのタイムスタンプ４０１を選択して（ステップＳ１２０７）、図１３のステップＳ１３０１に移行する。具体的には、たとえば、異常検出装置２０２は、概念ドリフト候補点Ｔごとのスケール誤差ｅ^{T}ｓｃａｌｅおよびシフト誤差ｅ^{T}ｓｈｉｆｔのうち、最小誤差を選択する。最小誤差のタイムスタンプ４０１が示す時刻をτとする。

【００６５】
たとえば、概念ドリフト候補点Ｔが３点（Ｔ１、Ｔ２、Ｔ３とする）存在すると仮定すると、スケール誤差ｅ^{T1}ｓｃａｌｅ，ｅ^{T2}ｓｃａｌｅ，ｅ^{T3}ｓｃａｌｅとシフト誤差ｅ^{T1}ｓｈｉｆｔ，ｅ^{T2}ｓｈｉｆｔ，ｅ^{T3}ｓｈｉｆｔとが算出される。ステップＳ１２０７では、異常検出装置２０２は、スケール誤差ｅ^{T1}ｓｃａｌｅ，ｅ^{T2}ｓｃａｌｅ，ｅ^{T3}ｓｃａｌｅとシフト誤差ｅ^{T1}ｓｈｉｆｔ，ｅ^{T2}ｓｈｉｆｔ，ｅ^{T3}ｓｈｉｆｔとの中から最小誤差を選択する。この最小誤差が、たとえば、スケール誤差ｅ^{T3}ｓｃａｌｅであったとすると、異常検出装置２０２は、スケール誤差ｅ^{T3}ｓｃａｌｅが発生した概念ドリフト候補点Ｔ３を選択してτとする。そして、図１３のステップＳ１３０１に移行する。

【００６６】
なお、概念ドリフト候補点決定処理（ステップＳ７０５）において、概念ドリフト候補

点Ｔが１つも決定または取得されなかった場合、ステップＳ１２０７および図１３の処理は実行されない。

【００６７】

図１３は、図７に示した予測結果補正処理（ステップＳ７０６）の詳細な処理手順例（後半）を示すフローチャートである。異常検出装置２０２は、ステップＳ１２０７の選択最小誤差がスケール誤差e^{τ}_{scale}であるかシフト誤差e^{τ}_{shift}であるかを判断する（ステップＳ１３０１）。選択最小誤差がスケール誤差e^{τ}_{scale}である場合（ステップＳ１３０１：スケール誤差）、異常検出装置２０２は、選択最小誤差であるスケール誤差e^{τ}_{scale}が誤差許容値$E_{tolerance}$以下であるか否かを判断する（ステップＳ１３０２）。スケール誤差e^{τ}_{scale}が誤差許容値$E_{tolerance}$以下でない場合（ステップＳ１３０２：Ｎｏ）、予測結果補正処理（ステップＳ７０６）が終了する。

【００６８】

一方、スケール誤差e^{τ}_{scale}が誤差許容値$E_{tolerance}$以下である場合（ステップＳ１３０２：Ｙｅｓ）、異常検出装置２０２は、タイムスタンプ４０１が示す時刻τ後の予測データｐを変化率r_{τ}でスケールする（ステップＳ１３０３）。このとき、異常検出装置２０２は、スケール後の予測データｐに基づいて下側シリーズ５０３および上側シリーズ５０４も補正する。そして、異常検出装置２０２は、スケール後の予測データｐについて、補正後予測結果テーブル３２３のエントリを追加し（ステップＳ１３０４）、予測結果補正処理（ステップＳ７０６）が終了する。

【００６９】

また、ステップＳ１３０１において、選択最小誤差がシフト誤差e^{τ}_{shift}である場合（ステップＳ１３０１：シフト誤差）、異常検出装置２０２は、選択最小誤差であるシフト誤差e^{τ}_{shift}が誤差許容値$E_{tolerance}$以下であるか否かを判断する（ステップＳ１３０５）。シフト誤差e^{τ}_{shift}が誤差許容値$E_{tolerance}$以下でない場合（ステップＳ１３０５：Ｎｏ）、予測結果補正処理（ステップＳ７０６）が終了する。

【００７０】

一方、シフト誤差e^{τ}_{shift}が誤差許容値$E_{tolerance}$以下である場合（ステップＳ１３０５：Ｙｅｓ）、異常検出装置２０２は、タイムスタンプ４０１が示す時刻τ後の予測データｐを変化差分d_{τ}でシフトする（ステップＳ１３０６）。このとき、異常検出装置２０２は、シフト後の予測データｐに基づいて下側シリーズ５０３および上側シリーズ５０４も補正する。そして、異常検出装置２０２は、シフト後の予測データｐについて、補正後予測結果テーブル３２３のエントリを追加し（ステップＳ１３０７）、予測結果補正処理（ステップＳ７０６）が終了する。

【００７１】

図１４～図１６は、概念ドリフト候補点Ｔごとのスケールおよびシフト計算例を示す説明図である。図１４の概念ドリフト候補点は、「２０２１－０１－０１　０１：００：００」であり、図１５の概念ドリフト候補点は、「２０２１－０１－０１　０１：１５：００」であり、図１６の概念ドリフト候補点は、「２０２１－０１－０１　０１：３０：００」である。

【００７２】

これら３つの概念ドリフト候補点Ｔにおける３つのスケール誤差e^{T}_{scale}および３つのシフト誤差e^{T}_{shift}のうち最小誤差は、図１４の概念ドリフト候補点Ｔ＝「２０２１－０１－０１　０１：００：００」におけるスケール誤差e^{T}_{scale}＝０．１５９である。したがって、このスケール誤差e^{T}_{scale}が最小誤差として選択され、そのタイムスタンプ４０１である概念ドリフト候補点Ｔ＝「２０２１－０１－０１　０１：００：００」がτとして選択される（ステップＳ１２０７）。そして、エントリ１４０１～１４０３のタイムスタンプ４０１およびスケール後の予測データが、補正後予測テーブルに登録される。

【００７３】

このように、上述した異常検出装置２０２によれば、たとえば、ＩＴインフラストラクチャ２０１において、ディスクの容量がｋ倍に増加すると、ディスク使用率４０２が１／ｋ倍にスケール変換される。また、ディスクの容量がｋ倍に低下すると、ディスク使用率４０２がｋ倍にスケール変換される。また、大量のログファイルの予定外の削除があると、ディスク使用率４０２がその分低下するようシフト変換される。また、大量のファイルがバックアップストレージにコピーされると、ディスク使用率４０２がその分上昇するようシフト変換される。

【００７４】

したがって、概念ドリフトを引き起こすような特定イベント発生後においても、異常検出の即時性が向上し、１日ごとまたは１時間ごとのような定期的なＩＴシステムの異常レポートの即時的な出力が可能になる。したがって、特定イベント発生後に異常検出が停止せず、異常であるのに正常としたり正常であるのに異常としたりするような誤検出も抑制することができる。

【００７５】

また、上述した異常検出装置２０２では、スケール変換とシフト変換のうちいずれか一方の線形変換を選択して、予測データを補正したが、スケール変換とシフト変換のうちいずれか一方の線形変換のみが実装されてもよい。

【００７６】

また、上述した異常検出装置２０２は、下記（１）～（９）のように構成することもできる。

【００７７】

（１）異常検出装置２０２は、プログラムを実行するプロセッサ３０１と、前記プログラムを記憶する記憶デバイス３０２と、を有し、前記プロセッサ３０１は、監視対象（たとえば、ディスク使用率４０２）の時系列な第１予測データ（たとえば、図１に示した時刻ｔ１からｔ３までの予測データ）のうち特定イベントの発生時点Ｔ（たとえば、図１の時刻ｔ２）後における第２予測データを、スケール変換により補正する補正処理（ステップＳ７０６）と、前記補正処理による補正後の第２予測データ（たとえば、図１に示した時刻ｔ２からｔ３までの予測データ）と、前記監視対象の時系列な第１実測データ（たとえば、図１に示した時刻ｔ１からｔ３までのテストデータ）のうち前記特定イベントの発生時点Ｔ後における第２実測データ（たとえば、図１に示した時刻ｔ２からｔ３までのテストデータ）と、に基づいて、前記監視対象の異常を検出する検出処理（ステップＳ７０７）と、を実行する。

【００７８】

（２）上記（１）の異常検出装置２０２において、前記補正処理（ステップＳ７０６）では、前記プロセッサ３０１は、前記第２予測データを、前記第２実測データの変化率r_τを用いた前記スケール変換により補正する。

【００７９】

（３）上記（１）の異常検出装置２０２において、前記補正処理（ステップＳ７０６）では、前記プロセッサ３０１は、前記第２予測データを、前記スケール変換またはシフト変換のうちいずれかの一方の線形変換により補正する。

【００８０】

（４）上記（１）の異常検出装置２０２において、前記補正処理では、前記プロセッサ３０１は、前記スケール変換が選択された場合、前記第２予測データを、前記第２実測データの変化率r_τを用いた前記スケール変換により補正し、前記シフト変換が選択された場合、前記第２予測データを、前記第２実測データの変化の差r_τを用いた前記シフト変換により補正する。

【００８１】

（５）上記（４）の異常検出装置２０２において、前記補正処理（ステップＳ７０６）では、前記プロセッサ３０１は、前記第２実測データと前記変化率r_τで拡縮した前記第２予測データとのスケール誤差e^{τ}_{scale}を算出し、前記第２実測データと前記変化の差

r_τでシフトした前記第２予測データとのシフト誤差$e^\tau{}_{shift}$を算出し、前記スケール誤差$e^\tau{}_{scale}$と前記シフト誤差$e^\tau{}_{shift}$とに基づいて、前記スケール変換またはシフト変換のうちいずれかの一方の線形変換を選択する。

【００８２】

（６）上記（５）の異常検出装置２０２において、前記補正処理（ステップＳ７０６）では、前記プロセッサ３０１は、前記スケール誤差$e^\tau{}_{scale}$と前記シフト誤差$e^\tau{}_{shift}$とのうち誤差が小さい方の線形変換を選択する。

【００８３】

（７）上記（１）の異常検出装置２０２において、前記プロセッサ３０１は、前記第１実測データの観測時点ｔにおける実測値の変化率r_tの各々について、変化率しきい値$R_{threshold}$よりも大きい（ステップＳ８０５：Ｙｅｓ）特定の観測時点ｔを前記特定イベントの発生時点候補Ｔに決定する決定処理（ステップＳ７０５）を実行し、前記補正処理（ステップＳ７０６）では、前記プロセッサ３０１は、前記決定処理（ステップＳ７０５）によって決定された前記特定イベントの発生時点候補Ｔのうちいずれかの発生時点候補τの後における前記第２予測データを、前記スケール変換により補正する。

【００８４】

（８）上記（７）の異常検出装置２０２において、前記決定処理（ステップＳ７０５）では、前記プロセッサ３０１は、前記第１予測データと前記第１実測データとの間の誤差Ｅが誤差しきい値$E_{threshold}$より大きい場合（ステップＳ８０３：Ｙｅｓ）、前記特定の観測時点ｔを前記特定イベントの発生時点候補Ｔに決定する。

【００８５】

（９）上記（７）の異常検出装置２０２において、前記決定処理（ステップＳ７０５）では、前記プロセッサ３０１は、前記特定の観測時点ｔでの前記実測値の変化率r_tについて所定期間内における出現回数f_rを計数し、前記出現回数f_rが出現回数しきい値$F_{threshold}$よりも小さい前記実測値の変化率r_tに対応する前記特定の観測時点ｔを、前記特定イベントの発生時点候補Ｔに決定する。

【００８６】

なお、本発明は前述した実施例に限定されるものではなく、添付した特許請求の範囲の趣旨内における様々な変形例及び同等の構成が含まれる。たとえば、前述した実施例は本発明を分かりやすく説明するために詳細に説明したものであり、必ずしも説明した全ての構成を備えるものに本発明は限定されない。また、ある実施例の構成の一部を他の実施例の構成に置き換えてもよい。また、ある実施例の構成に他の実施例の構成を加えてもよい。また、各実施例の構成の一部について、他の構成の追加、削除、または置換をしてもよい。

【００８７】

また、前述した各構成、機能、処理部、処理手段等は、それらの一部又は全部を、たとえば集積回路で設計する等により、ハードウェアで実現してもよく、プロセッサ３０１がそれぞれの機能を実現するプログラムを解釈し実行することにより、ソフトウェアで実現してもよい。

【００８８】

各機能を実現するプログラム、テーブル、ファイル等の情報は、メモリ、ハードディスク、ＳＳＤ（Ｓｏｌｉｄ　Ｓｔａｔｅ　Ｄｒｉｖｅ）等の記憶装置、又は、ＩＣ（Ｉｎｔｅｇｒａｔｅｄ　Ｃｉｒｃｕｉｔ）カード、ＳＤカード、ＤＶＤ（Ｄｉｇｉｔａｌ　Ｖｅｒｓａｔｉｌｅ　Ｄｉｓｃ）の記録媒体に格納することができる。

【００８９】

また、制御線や情報線は説明上必要と考えられるものを示しており、実装上必要な全ての制御線や情報線を示しているとは限らない。実際には、ほとんど全ての構成が相互に接続されていると考えてよい。

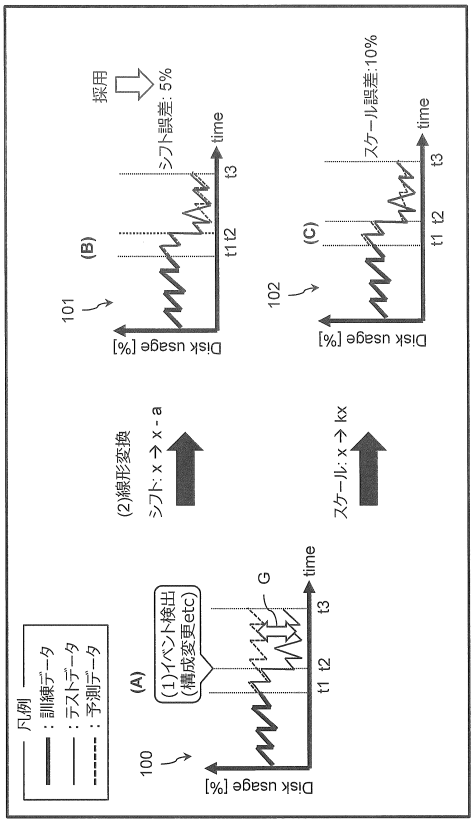
【符号の説明】

【００９０】

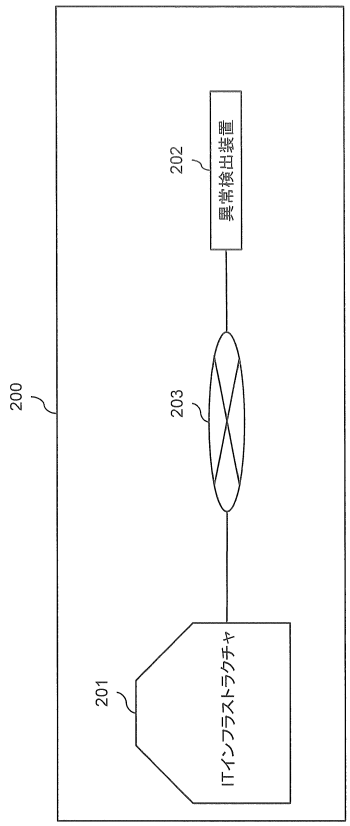
２００　異常検出システム
２０１　ＩＴインフラストラクチャ
２０２　異常検出装置
３０１　プロセッサ
３０２　記憶デバイス
３２１　ディスク使用率テーブル
３２２　予測結果テーブル
３２３　補正後予測結果テーブル
３２４　異常検出プログラム
３４０　異常検出モジュール
３４１　収集モジュール
３４２　概念ドリフト候補点決定モジュール
３４３　予測結果補正モジュール

【図面】

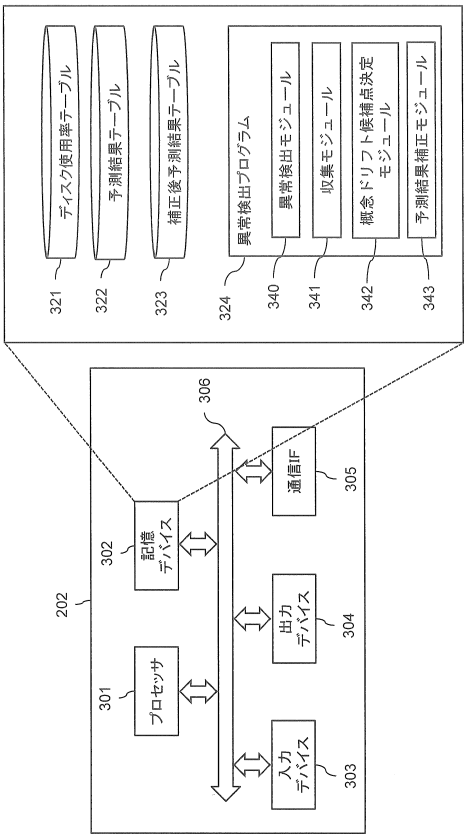
【図１】



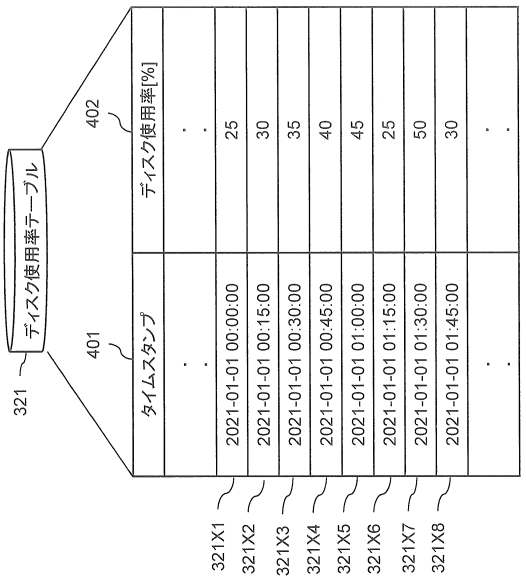
【図２】



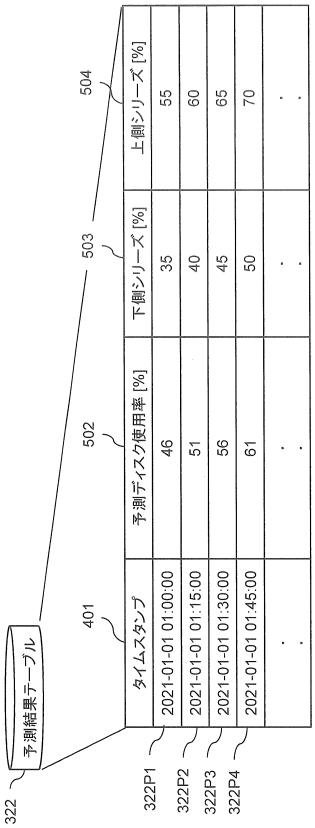
【図３】


ディスク使用率テーブル
予測結果テーブル
補正後予測結果テーブル
異常検出プログラム
異常検出モジュール
収集モジュール
概念ドリフト候補点決定モジュール
予測結果補正モジュール
321
322
323
324
340
341
342
343
202
プロセッサ
記憶デバイス
出力デバイス
入力デバイス
通信IF
301
302
303
304
305
306

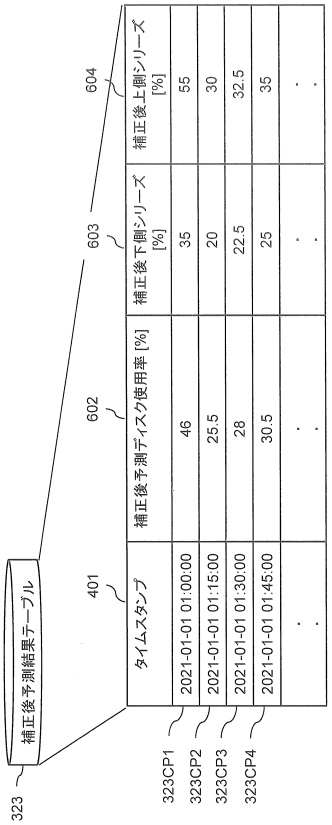
【図４】


ディスク使用率テーブル
321
タイムスタンプ
ディスク使用率[%]
401
402
2021-01-01 00:00:00 25
2021-01-01 00:15:00 30
2021-01-01 00:30:00 35
2021-01-01 00:45:00 40
2021-01-01 01:00:00 45
2021-01-01 01:15:00 25
2021-01-01 01:30:00 50
2021-01-01 01:45:00 30
321X1
321X2
321X3
321X4
321X5
321X6
321X7
321X8

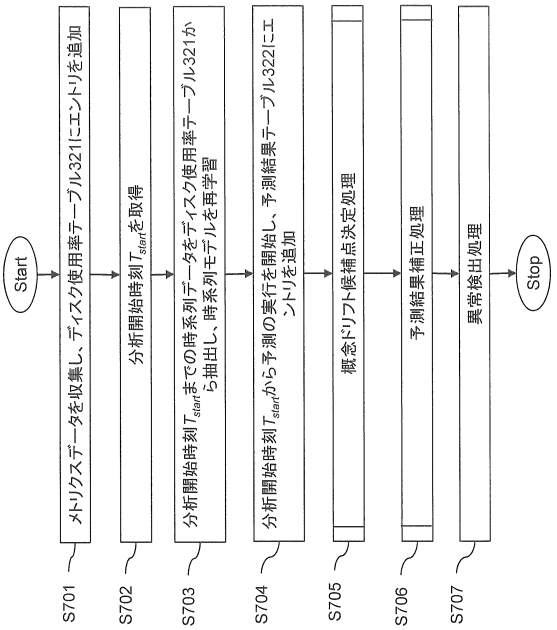
【図５】


予測結果テーブル
322
タイムスタンプ
予測ディスク使用率 [%]
下側シリーズ [%]
上側シリーズ [%]
401
502
503
504
2021-01-01 01:00:00 46 35 55
2021-01-01 01:15:00 51 40 60
2021-01-01 01:30:00 56 45 65
2021-01-01 01:45:00 61 50 70
322P1
322P2
322P3
322P4

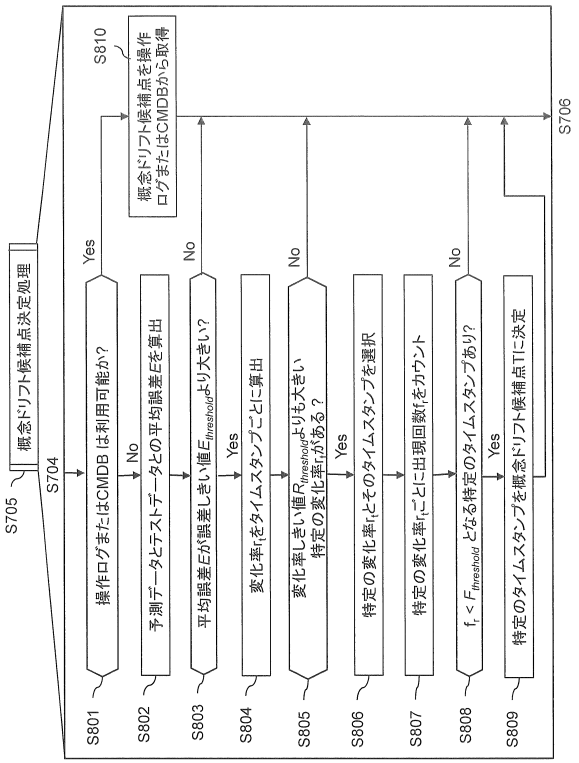
【図６】


補正後予測結果テーブル
323
タイムスタンプ
補正後予測ディスク使用率 [%]
補正後下側シリーズ [%]
補正後上側シリーズ [%]
401
602
603
604
2021-01-01 01:00:00 46 35 55
2021-01-01 01:15:00 25.5 20 30
2021-01-01 01:30:00 28 22.5 32.5
2021-01-01 01:45:00 30.5 25 35
323CP1
323CP2
323CP3
323CP4

【図７】



【図８】



【図９】

予測値と正解データとの平均誤差算出例(S802)

タイムスタンプ401	ディスク使用率 (x_t) (402)	予測ディスク使用率 (p_t) (502)	誤差e(t) $\frac{\lvert x_t - p_t \rvert}{x_t}$
[T_{start}] 2021-01-01 01:00:00	45	46	0.022
2021-01-01 01:15:00	25	51	1.040
2021-01-01 01:30:00	50	56	0.120
[T_{end}] 2021-01-01 01:45:00	30	61	1.033
			平均誤差E = 0.554
誤差しきい値$E_{threshold}$=0.2			

【図１０】

変化率rtの算出例(S804)

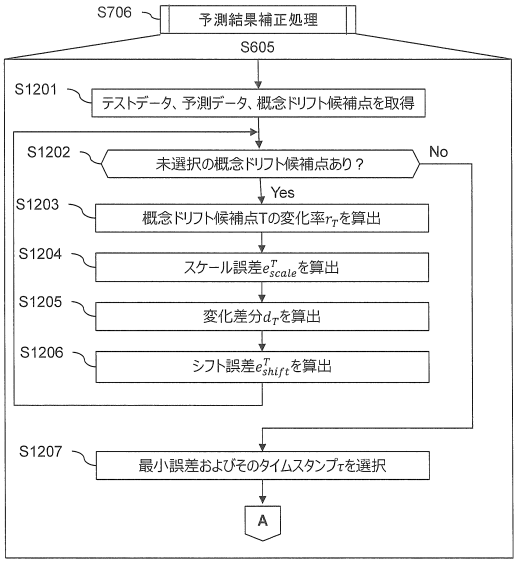
タイムスタンプ401	ディスク使用率 (x_t) (402)	変化率 rt(S804)	rt> $R_{treshold}$? (S805)	
[$T_{start-1}$] 2021-01-01 00:45:00	40	1.125	NO	1000
[T_{start}] 2021-01-01 01:00:00	45	1.800	YES	1001
2021-01-01 01:15:00	25	2.00	YES	1002
2021-01-01 01:30:00	50	1.667	YES	1003
[T_{end}] 2021-01-01 01:45:00	30	NA	NA	1004
変化率しきい値$Ratio_{threshold}$=1.5				

【図１１】

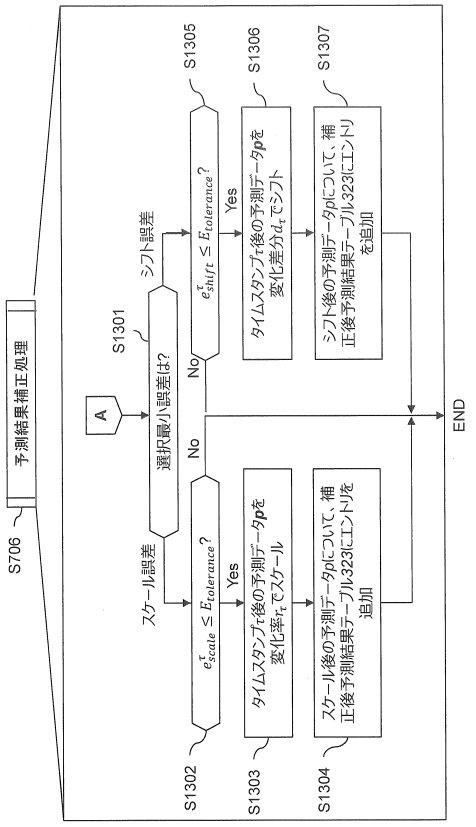
出現回数ftの算出例（S807）

選択タイムスタンプ（S806）	変化率 rt(S804)	出現回数fr（S807）	fr < $F_{threshold}$?（S808）	
2021-01-01 01:00:00	1.800	1	YES	1101
2021-01-01 01:15:00	2.00	1	YES	1102
2021-01-01 01:30:00	1.667	1	YES	1103
$F_{threshold}$=2 per day				

【図１２】



【図１３】



【図１４】

概念ドリフト候補点= 2021-01-01 01:00:00
変化率r_T = 25/45 = 0.5
変化差分d_T = 25-45 = -20

	タイムスタンプ401	テストデータx	予測データp	スケール後予測データ (r_T p_i)	スケール誤差 $\left(\frac{|x_i - r_T p_i|}{x_i}\right)$	シフト後予測データ (p_i+d_T)	シフト誤差 $\left(\frac{|x_i - (p_i + d_T)|}{x_i}\right)$	
	[T] 2021-01-01 01:00:00	45	46					
1401	2021-01-01 01:15:00	25	51	25.5	0.02	31	0.24	
1402	2021-01-01 01:30:00	50	56	28	0.44	36	0.28	
1403	[T_{end}] 2021-01-01 01:45:00	30	61	30.5	0.017	41	0.367	
				スケール誤差 e_{scale}^{T} = 0.169		シフト誤差 e_{shift}^{T} = 0.296		

【図１５】

概念ドリフト候補点= 2021-01-01 01:15:00

変化率r_T = 50/25 = 2.0

変化差分d_T = 50-25 = 25

タイムスタンプ401	テストデータx	予測データp	スケール後予測データ $(r_T{}^*p_t)$	スケール誤差 $(\frac{|x_t-r_T{}^*p_t|}{x_t})$	シフト後予測データ $(p_t{}^*d_T)$	シフト誤差 $(\frac{|x_t-(p_t{}^*d_T)|}{x_t})$
2021-01-01 01:00:00	45	46				
[T] 2021-01-01 01:15:00	25	51				
2021-01-01 01:30:00	50	56	112	1.24	81	0.62
$[T_{end}]$ 2021-01-01 01:45:00	30	61	122	3.067	86	1.867
			スケール誤差e^T_{scale} = 2.154		シフト誤差e^T_{shift} = 1.244	

【図１６】

概念ドリフト候補点= 2021-01-01 01:30:00

変化率r_T = 30/50 = 0.6

変化差分d_T = 30-50 = -20

タイムスタンプ401	テストデータx	予測データp	スケール後予測データ $(r_T{}^*p_t)$	スケール誤差 $(\frac{|x_t-r_T{}^*p_t|}{x_t})$	シフト後予測データ $(p_t{}^*d_T)$	シフト誤差 $(\frac{|x_t-(p_t{}^*d_T)|}{x_t})$
2021-01-01 01:00:00	45	46				
2021-01-01 01:15:00	25	51				
[T] 2021-01-01 01:30:00	50	56				
$[T_{end}]$ 2021-01-01 01:45:00	30	61	36	0.2	41	0.367
			スケール誤差e^T_{scale} = 0.2		シフト誤差e^T_{shift} = 0.367	

フロントページの続き

(56)参考文献　特開２０１５－０９０６９１（ＪＰ，Ａ）
特開平１１－１２５４４９（ＪＰ，Ａ）
特開２００９－１６９９５９（ＪＰ，Ａ）
特開２０１９－１３５４５１（ＪＰ，Ａ）

(58)調査した分野　(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　１１／３４
Ｇ０６Ｎ　２０／００