

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2021-135896

(P2021-135896A)

(43) 公開日 令和3年9月13日(2021.9.13)

(51) Int.Cl.
G06N 20/00 (2019.01)F I
G O 6 N 20/00

テーマコード (参考)

審査請求 未請求 請求項の数 18 O L (全 22 頁)

(21) 出願番号 特願2020-33344 (P2020-33344)
(22) 出願日 令和2年2月28日(2020.2.28)(71) 出願人 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 110000176
一色国際特許業務法人
(72) 発明者 三宅 貫太郎
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 高瀬 誠由
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 野中 康充
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

最終頁に続く

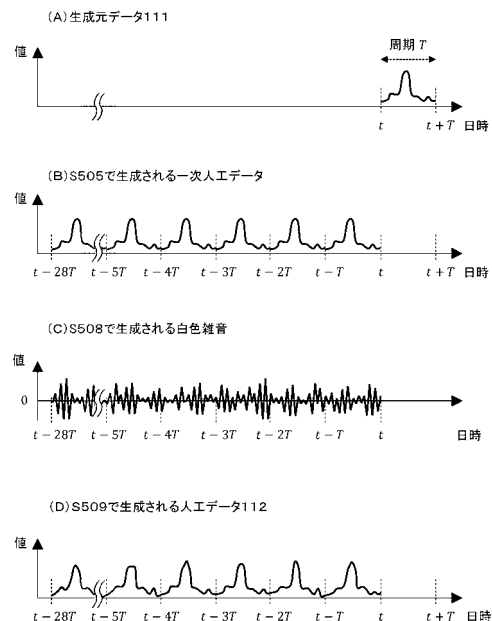
(54) 【発明の名称】 学習データ生成装置、及び学習データ生成方法

(57) 【要約】

【課題】必要とされる期間について適切な内容の学習データを効率よく提供する。

【解決手段】学習データ生成装置は、所定周期分の時系列データである生成元データを複製したデータである複製データを複数連結するとともに複製データの夫々に雑音を与えることにより、要求される期間に応じた期間の時系列データである人工データを生成し、人工データを用いて機械学習モデルの学習に用いる学習データを生成する。学習データ生成装置は、例えば、生成元データよりも過去の期間の人工データを生成し、生成元データに人工データを過去方向に連結して学習データを生成する。

【選択図】 図6



【特許請求の範囲】**【請求項 1】**

情報処理装置を用いて構成され、機械学習モデルの学習に用いる学習データを生成する学習データ生成装置であって、

所定周期分の時系列データである生成元データを複製したデータである複製データを複数連結するとともに前記複製データの夫々に雑音を与えることにより、要求される期間に応じた期間の時系列データである人工データを生成する人工データ生成部と、

前記人工データを用いて学習データを生成する学習データ生成部と、
を備える、学習データ生成装置。

【請求項 2】

10

請求項 1 に記載の学習データ生成装置であって、

前記要求される期間は、ユーザインタフェースを介してユーザから受け付けた期間である、

学習データ生成装置。

【請求項 3】

請求項 1 に記載の学習データ生成装置であって、

前記人工データ生成部は、前記生成元データよりも過去の期間の前記人工データを生成し、

前記学習データ生成部は、前記生成元データに前記人工データを時系列に連結して前記学習データを生成する、

20

学習データ生成装置。

【請求項 4】

請求項 1 に記載の学習データ生成装置であって、

前記人工データ生成部は、前記生成元データよりも未来の期間の前記人工データを生成し、

前記学習データ生成部は、前記生成元データに前記人工データを時系列に連結して前記学習データを生成する、

学習データ生成装置。

【請求項 5】

30

請求項 1 又は 2 に記載の学習データ生成装置であって、

前記学習データ生成部は、前記人工データに、前記機械学習モデルを用いた推論処理を行う機械学習システムの本番運用に際して実際に前記機械学習モデルに入力された時系列データである観測データを連結することにより前記学習データを生成する、

学習データ生成装置。

【請求項 6】

請求項 5 に記載の学習データ生成装置であって、

前記学習データ生成部は、前記観測データを前記人工データよりも優先して採用することにより前記学習データを生成する、

学習データ生成装置。

【請求項 7】

40

請求項 6 に記載の学習データ生成装置であって、

前記学習データの開始時点から終了時点までの期間である学習データ期間を設定する学習データ期間設定部を備え、

前記学習データ期間設定部は、前記観測データの最新の時点进行学习データ期間の終了時点に設定し、

前記終了時点から前記要求される期間だけ遡った時点を前記開始時点に設定する、

学習データ生成装置。

【請求項 8】

請求項 1 に記載の学習データ生成装置であって、

前記人工データ生成部は、前記複製データの夫々に個別に雑音を与える、

50

学習データ生成装置。

【請求項 9】

請求項 1 に記載の学習データ生成装置であって、
前記雑音は白色雑音である、
学習データ生成装置。

【請求項 10】

請求項 1 に記載の学習データ生成装置であって、
前記人工データ生成部は、前記生成元データを、トレンド、周期変動、及び残差の各構成要素に分解し、前記構成要素のうち、前記トレンドと前記周期変動とに基づき前記人工データを生成する、
学習データ生成装置。

10

【請求項 11】

請求項 10 に記載の学習データ生成装置であって、
前記人工データ生成部は、前記残差の分散に基づき前記雑音を生成し、生成した前記人工データに生成した前記雑音を付与する、
学習データ生成装置。

【請求項 12】

請求項 10 に記載の学習データ生成装置であって、
前記人工データ生成部は、前記複製データを連結する際の境界となる時点における、前記境界の前後の前記複製データの前記周期変動の差分を求め、前記差分を反映しつつ複数の前記複製データを連結することにより前記人工データを生成する、
学習データ生成装置。

20

【請求項 13】

情報処理装置が、
所定周期分の時系列データである生成元データを複製したデータである複製データを複数連結するとともに前記複製データの夫々に雑音を与えることにより、要求される期間に応じた期間の時系列データである人工データを生成するステップと、
前記人工データを用いて機械学習モデルの学習に用いる学習データを生成するステップと、
を実行する、学習データ生成方法。

30

【請求項 14】

請求項 13 に記載の学習データ生成方法であって、
前記情報処理装置が、前記人工データに、前記機械学習モデルを用いた推論処理を行う機械学習システムの本番運用に際して実際に前記機械学習モデルに入力された時系列データである観測データを連結することにより前記学習データを生成するステップ、
を更に実行する、学習データ生成方法。

【請求項 15】

請求項 14 に記載の学習データ生成方法であって、
前記情報処理装置が、前記観測データを前記人工データよりも優先して採用することにより前記学習データを生成するステップ、
を更に実行する、学習データ生成方法。

40

【請求項 16】

請求項 13 に記載の学習データ生成方法であって、
前記情報処理装置が、前記生成元データを、トレンド、周期変動、及び残差の各構成要素に分解し、前記構成要素のうち、前記トレンドと前記周期変動とに基づき前記人工データを生成するステップ、
を更に実行する、学習データ生成方法。

【請求項 17】

請求項 16 に記載の学習データ生成方法であって、
前記情報処理装置が、前記残差の分散に基づき前記雑音を生成し、生成した前記人工デ

50

ータに生成した前記雑音を付与するステップ、
を更に実行する、学習データ生成方法。

【請求項 18】

請求項 16 に記載の学習データ生成方法であって、

前記情報処理装置が、前記複製データを連結する際の境界となる時点における、前記境界の前後の前記複製データの前記周期変動の差分を求め、前記差分を順次加算しつつ前記複製データを連結していくことにより前記人工データを生成するステップ、

を更に実行する、学習データ生成方法。

【発明の詳細な説明】

【技術分野】

10

【0001】

本発明は、学習データ生成装置、及び学習データ生成方法に関する。

【背景技術】

【0002】

機械学習システムの実現に際しては、機械学習モデルの精度を確保するため、有効な学習データを準備する必要がある。

【0003】

学習データを生成する技術として、例えば、特許文献 1 には、ニューラルネットワークを利用した判定装置を学習するための学習用データの生成装置について記載されている。学習用データの生成装置は、収集した時系列データのデータ値を変更し、時系列データの各データの時間間隔を変更し、時系列データに歪を付加し、時系列データに雑音を付加する。

20

【0004】

また特許文献 2 には、学習データが少数の場合に、学習データを加工することで、学習の改善に寄与するデータを生成する技術に関して記載されている。具体的には、ニューラルネットワーク学習装置が、学習中のニューラルネットワークを用いて学習データから特徴を抽出し、抽出した特徴から学習中のニューラルネットワークを用いて敵対的特徴を生成し、学習データと敵対的特徴とを用いてニューラルネットワークの認識結果を算出し、認識結果が望ましい出力に近づくようにニューラルネットワークを学習する。

【0005】

30

また特許文献 3 には、監視対象の状態異常を速やかに検知することを目的として構成された異常検知システムに関して記載されている。異常検知システムは、監視対象に対する観測データを収集して時系列観測データとして保存し、観測データを訓練用データおよび検証用データのいずれかに分類し、訓練用データに基づき監視対象の線形状態空間モデルのモデルパラメータを同定し、モデルパラメータと検証用データを入力として監視対象の状態変数の確率分布の推定値を計算し、推定値に基づき監視対象の異常度を計算し、観測データを収集すると新しく収集した観測データを時系列観測データに追加するとともに時系列観測データのデータ数がしきい値よりも大きい場合は最も前に収集した観測データを破棄する。

【先行技術文献】

40

【特許文献】

【0006】

【特許文献 1】特開 2019 - 87106 号公報

【特許文献 2】国際公開第 2018 / 167900 号

【特許文献 3】特開 2019 - 191836 号公報

【非特許文献】

【0007】

【非特許文献 1】R. B. Cleveland、外 3 名、“STL: a seasonal-trend decomposition procedure based on loess”、[online]、1990 年、Journal of official statistics、[2020 年 1 月 31 日検索]、インターネット<URL: <https://www.wessa.net/download/stl.pdf>>

50

【発明の概要】**【発明が解決しようとする課題】****【0008】**

時系列データに基づき予兆診断や異常検知等の推論処理を行う機械学習システムの実現に際しては、上記推論処理を行う機械学習モデルの精度を確保する必要がある、そのためには有効な学習データを効率よく準備する必要がある。また機械学習モデルの精度を確保するには、そのために必要とされる期間の時系列データを学習データとして用意する必要がある。

【0009】

しかし特許文献1及び特許文献2は、いずれも必要とされる期間の時系列データを生成する技術については何も開示されていない。また特許文献3に記載の技術では、監視対象に対する観測データを収集する必要がある、例えば、機械学習システムの導入時等のように観測データが得られていない場合には対応することができない。

10

【0010】

本発明の目的は、必要とされる期間について適切な内容の学習データを効率よく提供することが可能な、学習データ生成装置、及び学習データ生成方法を提供することを目的とする。

【課題を解決するための手段】**【0011】**

上記目的を達成するための本発明のうちの一つは、情報処理装置を用いて構成され、機械学習モデルの学習に用いる学習データを生成する学習データ生成装置であって、所定周期分の時系列データである生成元データを複製したデータである複製データを複数連結するとともに前記複製データの夫々に雑音を与えることにより、要求される期間に応じた期間の時系列データである人工データを生成する人工データ生成部と、前記人工データを用いて学習データを生成する学習データ生成部と、を備える。

20

【0012】

その他、本願が開示する課題、及びその解決方法は、発明を実施するための形態の欄、及び図面により明らかにされる。

【発明の効果】**【0013】**

30

本発明によれば、必要とされる期間について適切な内容の学習データを効率よく提供することができる。

【図面の簡単な説明】**【0014】**

【図1】機械学習システムの概略的な構成を示す図である。

【図2】情報処理装置の構成の一例である。

【図3】学習データ生成装置が備える主な機能を示す図である。

【図4】学習データ生成処理を説明するフローチャートである。

【図5】人工データ生成処理の詳細を説明するフローチャートである。

【図6】人工データ生成処理の実行過程で生成されるデータを模式的に示した図である。

40

【図7】生成元データの一例である。

【図8】人工データの一例である。

【図9】学習データ期間設定処理の詳細を説明するフローチャートである。

【図10】観測データの一例である。

【図11】学習データ生成処理の詳細を説明するフローチャートである。

【図12】学習データの一例である。

【図13】第2実施形態における人工データ生成処理を説明するフローチャートである。

【図14】第2実施形態における人工データ生成処理の実行過程で生成されるデータを模式的に示した図である。

【図15】生成元データの一例である。

50

【図 1 6】中間データの一例である。

【図 1 7】複製元データの一例である。

【図 1 8】人工データの一例である。

【発明を実施するための形態】

【0015】

以下、本発明の一実施形態について図面を参照しつつ説明する。尚、以下の説明において、同一の又は類似する機能を有する構成について同一の符号を付して重複する説明を省略することがある。また以下の説明において、符号の前に付した「S」の文字は処理ステップを意味する。また以下の説明において「学習データ」という用語を用いるが、「訓練データ」と同義である。またいわゆる教師あり機械学習に用いる学習データはいわゆるラベルの情報を含むが、説明の簡単のため、本実施形態ではラベルに関する説明や例示を省略する。また以下の説明において、期間は、日時で指定してもよいし、日のみもしくは時間のみで指定してもよい。

10

【0016】

[第1実施形態]

図 1 に、第 1 実施形態として示す学習データ生成装置 100 が適用される情報処理システム（以下、「機械学習システム 1」と称する。）の概略的な構成を示している。同図に示すように、機械学習システム 1 は、推論装置 2 と学習データ生成装置 100 とを含む。

【0017】

推論装置 2 は、時系列データである学習データ 114 を用いて機械学習モデル 23 の学習を行う学習処理部 21、及び機械学習モデル 23 を用いて推論処理を行う推論処理部 22 の各機能を有する。推論処理部 22 は、時系列データである観測データ 113 を機械学習モデル 23 に入力することにより推論処理を行い、結果を推論結果 7 として出力する。機械学習モデル 23 は、例えば、時系列データに基づき予兆診断や異常検知等のための推論処理を行う。

20

【0018】

学習データ生成装置 100 は、時系列データである、生成元データ 111 や観測データ 113 に基づき学習データ 114 を生成する。生成された学習データ 114 は、通信または記録媒体を介して推論装置 2 に入力される。

【0019】

図 2 に、推論装置 2 や学習データ生成装置 100 の構成に用いる情報処理装置 10 の一例を示す。同図に示すように、例示する情報処理装置 10 は、プロセッサ 11、主記憶装置 12、補助記憶装置 13、入力装置 14、出力装置 15、及び通信装置 16 を備える。これらはバス等の通信手段を介して通信可能に接続されている。

30

【0020】

情報処理装置 10 は、例えば、クラウドシステムにより提供される仮想サーバのように仮想化技術やプロセス空間分離技術等を用いて提供される仮想的な情報処理資源を用いて実現されるものであってもよい。また情報処理装置 10 の機能の全部又は一部を、例えば、クラウドシステムが A P I (Application Programming Interface) 等を介して提供するサービスにより実現してもよい。また例えば、通信可能に接続された複数の情報処理装置 10 を用いて学習データ生成装置 100 を構成してもよい。情報処理装置 10 には、例えば、オペレーティングシステム、ファイルシステム、D B M S (DataBase Management System) (リレーショナルデータベース、N o S Q L 等)等のソフトウェアが導入されていてもよい。

40

【0021】

プロセッサ 11 は、例えば、C P U (Central Processing Unit)、M P U (Micro Processing Unit)、G P U (Graphics Processing Unit)、A I (Artificial Intelligence) チップ、F P G A (Field Programmable Gate Array)、A S I C (Application Specific Integrated Circuit) 等を用いて構成されている。

【0022】

50

主記憶装置 12 は、プログラムやデータを記憶する装置であり、例えば、ROM (Read Only Memory)、RAM (Random Access Memory)、不揮発性メモリ (NVRAM (Non Volatile RAM)) 等である。

【0023】

補助記憶装置 13 は、例えば、SSD (Solid State Drive)、ハードディスクドライブ、光学式記憶装置 (CD (Compact Disc)、DVD (Digital Versatile Disc) 等)、ストレージシステム、ICカード、SDカードや光学式記録媒体等の記録媒体の読取/書込装置、仮想サーバの記憶領域等である。補助記憶装置 13 には、記録媒体の読取装置や通信装置 16 を介してプログラムやデータを読み出すことができる。補助記憶装置 13 に格納 (記憶) されているプログラムやデータは主記憶装置 12 に随時読み出される。

10

【0024】

入力装置 14 は、外部からの入力を受け付けるインタフェースであり、例えば、キーボード、マウス、タッチパネル、カードリーダー、音声入力装置等である。出力装置 15 は、処理経過や処理結果等の各種情報を出力するインタフェースである。

【0025】

出力装置 15 は、例えば、上記の各種情報を可視化する表示装置 (液晶モニタ、LCD (Liquid Crystal Display)、プロジェクタ等)、上記の各種情報を音声化する装置 (音声出力装置 (スピーカ等))、上記の各種情報を文字化する装置 (印字装置等) である。

【0026】

入力装置 14 と出力装置 15 はユーザインタフェースを構成する。尚、例えば、情報処理装置 10 が通信装置 16 を介して他の装置 (スマートフォン、タブレット、ノートブック型コンピュータ、各種携帯情報端末等) と情報の入出力を行う構成としてもよい。

20

【0027】

通信装置 16 は、他の装置との間の通信を実現する。通信装置 16 は、通信ネットワークを介した他の装置との間の通信を実現する、無線又は有線方式の通信インタフェースであり、例えば、NIC (Network Interface Card)、無線通信モジュール、USB (Universal Serial Bus) モジュール、シリアル通信モジュール等である。続いて、各装置が備える機能について説明する。

【0028】

図 3 に、学習データ生成装置 100 が備える主な機能を示している。同図に示すように、学習データ生成装置 100 は、記憶部 110、観測データ取得部 120、生成元データ取得部 130、人工データ生成部 140、学習データ期間設定部 150、学習データ生成部 160、及び学習データ出力部 170 の各機能を備える。これらの機能は、学習データ生成装置 100 を構成する情報処理装置 10 のプロセッサ 11 が、情報処理装置 10 の主記憶装置 12 に格納されているプログラムを読み出して実行することにより、もしくは、情報処理装置 10 が備えるハードウェア (FPGA、ASIC、AIチップ等) により実現される。

30

【0029】

上記機能のうち、記憶部 110 は、生成元データ 111、人工データ 112、観測データ 113、及び学習データ 114 を記憶し管理する。記憶部 110 は、例えば、DBMS が提供するデータベースのテーブルや、ファイルシステムが提供するファイルとして、各データを記憶し管理する。

40

【0030】

生成元データ 111 は、人工データ 112 の生成に用いるデータである。人工データ 112 は、学習データ 114 の生成に用いるデータである。観測データ 113 は、機械学習システム 1 が、例えば、本番運用を開始した後に機械学習モデル 23 に入力されたデータである。学習データ 114 は、人工データ 112 や観測データ 113 に基づき生成されるデータであり、機械学習モデル 23 の学習 (訓練) に用いられるデータである。

【0031】

図 3 に示す機能のうち、観測データ取得部 120 は、推論装置 2 から通信や記録媒体を

50

介して観測データ 113 を取得する。記憶部 110 は、観測データ取得部 120 が取得した観測データ 113 を記憶する。

【0032】

生成元データ取得部 130 は、生成元データ 111 を取得もしくは生成する。生成元データ取得部 130 は、例えば、ユーザインタフェースを介してユーザから生成元データ 111 を取得する。生成元データ取得部 130 は、例えば、観測データ 113 に基づき生成元データ 111 を生成する。ユーザがユーザインタフェースを介して観測データ 113 を編集することにより生成元データ 111 を生成してもよい。記憶部 110 は、生成元データ取得部 130 が取得もしくは生成した生成元データ 111 を記憶する。

【0033】

人工データ生成部 140 は、生成元データ 111 に基づき人工データ 112 を生成する。記憶部 110 は、人工データ生成部 140 が生成した人工データ 112 を記憶する。

【0034】

学習データ期間設定部 150 は、学習データ 114 の期間（学習データの開始時点から終了時点まで。以下、「学習データ期間」と称する。）の設定に関する処理を行う。学習データ期間設定部 150 は、例えば、ユーザインタフェースを介してユーザから学習データ期間の設定に関する情報を受け付ける。

【0035】

学習データ生成部 160 は、人工データ 112 や観測データ 113 に基づき学習データ 114 を生成する。

【0036】

学習データ出力部 170 は、学習データ生成部 160 が生成した学習データ 114 を出力する。出力された学習データ 114 は、通信や記録媒体を介して推論装置 2 に入力される。

【0037】

図 4 は、学習データ生成装置 100 が学習データ 114 の生成に際して行う処理（以下、「学習データ生成処理 S400」と称する。）を説明するシーケンス図である。以下、同図とともに学習データ生成処理 S400 について説明する。尚、同図に示す処理の開始時点において、記憶部 110 は、観測データ取得部 120 が取得した観測データ 113、及び生成元データ取得部 130 が取得もしくは生成した生成元データ 111 を既に記憶しているものとする。

【0038】

同図に示すように、まず人工データ生成部 140 が、記憶部 110 が記憶している生成元データ 111 を読み出す（S411）。

【0039】

続いて、人工データ生成部 140 は、ユーザインタフェースを介して、ユーザが生成しようとする学習データ 114 の期間の長さ（以下、「要求期間」と称する。）の指定、生成元データ 111 に含まれている周期の数（以下、「周期数」と称する。）、及び人工データ 112 に与える雑音の生成に用いる分散 σ^2 の入力を受け付ける（S412）。

【0040】

続いて、人工データ生成部 140 は、読み出した生成元データ 111 と、受け付けた要求期間、周期数、及び分散 σ^2 に基づき、人工データ 112 を生成する処理（以下、「人工データ生成処理 S413」と称する。）を行う（S413）。

【0041】

続いて、学習データ期間設定部 150 が、記憶部 110 が記憶している観測データを読み出す（S421）。

【0042】

続いて、学習データ期間設定部 150 は、S412 で人工データ生成部 140 が読み出した生成元データ 111 の期間、S413 で人工データ生成部 140 が生成した人工データ 112 の期間、及び S421 で読み出した観測データ 113 の期間を取得する（S42

10

20

30

40

50

2)。

【0043】

続いて、学習データ期間設定部150は、S412で人工データ生成部140が受け付けた要求期間を取得する(S423)。

【0044】

続いて、学習データ期間設定部150は、S422で取得した各期間とS423で取得した要求期間とに基づき、学習データ期間を設定する処理(以下、「学習データ期間設定処理S424」と称する。)を行う(S424)。

【0045】

続いて、学習データ生成部160が、S411で人工データ生成部140が読み出した生成元データ111、S413で人工データ生成部140が生成した人工データ112、及びS421で学習データ期間設定部150が読み出した観測データ113に基づき、学習データ期間設定処理S424により設定された学習データ期間について学習データ114を生成する(S431)。

【0046】

その後、学習データ出力部170は、生成された学習データ114を出力する。出力された学習データ114は、通信や記録媒体を介して推論装置2の学習処理部21に送信(提供)される。

【0047】

図5は、図4に示した人工データ生成処理S413の詳細を説明するフローチャートである。また図6は、人工データ生成処理S413の実行過程で生成されるデータを模式的に示した図である。人工データ生成部140は、生成元データ111を、S412で受け付けた要求期間に応じた周期数だけ複製し、生成元データ111の日時を適切な日時に置換し、更に観測値に雑音を付与することにより、人工データ112を生成する。以下、図5及び図6を参照しつつ、人工データ生成処理S413について説明する。

【0048】

図7に生成元データ111の一例を示す。以下では同図に示す生成元データ111を例として人工データ生成処理S413を説明する。同図に示すように、例示する生成元データ111は、日時701及び観測値702の各項目を有する複数のエントリ(レコード)を含む。

【0049】

上記項目のうち、日時701には、観測値702の値を取得した日時が設定される。尚、日時701の値は、各エントリを一意に識別するための識別子としても用いられる。観測値702には、観測値が設定される。尚、時系列データにはカテゴリ変数情報が含まれる場合があるが、とくに説明がない限り、観測値は量的変数情報であるものとする。観測値702は、例えば、センサ装置等から取得した値そのもの(生データ)や、複数の観測対象から得られた値を処理(加減乗除、集計処理、統計処理等)することにより得られる値である。上記値は、例えば、観測対象が情報通信システムである場合における通信量や稼働率である。また上記値は、例えば、「上り通信量」と「下り通信量」という2つの観測対象の値を合計した値「合計通信量」である。また例えば、上記値は、ある時点における観測値と別の時点の観測値に基づき計算により求められる。また例えば、上記値は、前回の通信量と今回の通信量との差分(通信量の時間変化量)である。

【0050】

例示する生成元データ111は、2019年11月15日0時0分0秒から2019年11月22日0時0分0秒までの情報を10分間隔で記録した内容からなり、図6(A)に示す1周期分のデータである。尚、以下の説明において、S412で受け付けた生成元データ111の周期は1週間であり、1周期あたり1008個のエントリが含まれるものとする。またS412で要求期間として28週を受け付けているものとする。

【0051】

図5に示すように、まず人工データ生成部140は、S412で受け付けた要求期間以

10

20

30

40

50

上の期間となる、生成元データ 1 1 1 の 1 周期の期間の倍数の最小値（以下、「最小周期数」と称する。）を求める（S 5 0 1）。

【0 0 5 2】

続いて、人工データ生成部 1 4 0 は、求めた最小周期数から生成元データ 1 1 1 に含まれている周期数を減じた値を求め、求めた値を生成元データ 1 1 1 に含まれる周期数で割った値を小数点以下切り上げることにより得られる値を複製回数とする（S 5 0 2）。尚、S 5 0 1 で求めた最小周期数から生成元データ 1 1 1 に含まれている周期数を減じているのは、複製元の生成元データ 1 1 1 の分を複製数から除くためである。例示する生成元データ 1 1 1 の周期数は 1 であり、要求期間は 2 8 週であるため、本例では複製回数として 2 7 が得られる。尚、複製回数は以上の方法以外の方法で取得してもよい。例えば、ユーザインタフェースを介してユーザから複製回数の指定を受け付けるようにしてもよい。

10

【0 0 5 3】

続いて、人工データ生成部 1 4 0 は、S 5 0 2 で取得した複製回数だけ生成元データ 1 1 1 を複製したデータ（以下、「複製データ」と称する。）を生成する（S 5 0 3）。

【0 0 5 4】

続いて、人工データ生成部 1 4 0 は、1 から始まる自然数を各複製データに順に割り当てる。記憶部 1 1 0 は、各複製データに割り当てられた番号（以下、「複製番号」と称する。）を複製データの夫々に対応づけて記憶する（S 5 0 4）。

【0 0 5 5】

続いて、人工データ生成部 1 4 0 は、割り当てた複製番号の逆順に複製データを時系列に連結したデータ（以下、「一次人工データ」と称する。）を生成する（S 5 0 5）。

20

【0 0 5 6】

続いて、人工データ生成部 1 4 0 は、S 5 0 7 で生成した一次人工データの各エントリに、生成元データ 1 1 1 の各エントリの日時 7 0 1 の値を複製したデータ（以下、「参照元日時」と称する。）を付与する（S 5 0 6）。

【0 0 5 7】

続いて、人工データ生成部 1 4 0 は、付与した一次人工データの各エントリの参照元日時を、基準とする日時（同図では例えば日時 t ）から遡った値に更新する（日時 t から生成元データ 1 1 1 の周期と各エントリの複製番号とを乗算することにより得られる日時分遡る）ことにより、各エントリの日時を生成する（S 5 0 7）。例えば、複製番号 2 7 の複製データにおける 2 0 1 9 年 1 1 月 1 5 日 0 時 0 分 0 秒の変更後の日時は、2 7 週分遡った 2 0 1 9 年 5 月 1 0 日 0 時 0 分 0 秒になる。S 5 0 7 を実行することにより生成される一次人工データは、図 6（B）のようになる。

30

【0 0 5 8】

続いて、人工データ生成部 1 4 0 は、図 4 の S 4 1 2 で受け付けた分散 σ^2 を用いて人工データの期間の白色雑音を生成する（S 5 0 8）。S 5 0 8 を実行することにより生成される白色雑音は図 6（C）のようになる。

【0 0 5 9】

続いて、人工データ生成部 1 4 0 は、一次人工データに対して、S 5 0 8 で生成した白色雑音を変動値として付与することにより、人工データ 1 1 2 を生成する（S 5 0 9）。S 5 0 9 を実行することにより生成される人工データ 1 1 2 は、図 6（D）のようになる。

40

【0 0 6 0】

図 8 に、人工データ生成処理 S 4 1 3 により生成される人工データ 1 1 2 の一例を示す。同図に示すように、例示する人工データ 1 1 2 は、日時 8 1 1、観測値 8 1 2、参照元観測値 8 1 3、変動値 8 1 4、参照元日時 8 1 5、及び複製番号 8 1 6 の各項目を有する複数のエントリを含む。

【0 0 6 1】

上記項目のうち、日時 8 1 1 には、S 5 0 7 で生成された日時が設定される。尚、日時 8 1 1 の値は、各エントリを一意に識別する識別子としても用いられる。観測値 8 1 2 に

50

は、S 5 0 9 で生成された人工データ 1 1 2 の当該日時における観測値が設定される。参照元観測値 8 1 3 には、当該日時に対応する、生成元データ 1 1 1 の観測値 7 0 2 が設定される。変動値 8 1 4 には、S 5 0 8 で生成された、当該日時に対応する白色雑音の値が設定される。参照元日時 8 1 5 には、当該日時に対応する、生成元データ 1 1 1 の日時 7 0 1 が設定される。参照元日時 8 1 5 の値は、当該エントリが、当該参照元日時 8 1 5 の値の日時の生成元データ 1 1 1 のエントリに基づくものであることを示す。複製番号 8 1 6 には、S 5 0 4 で割り当てられた複製番号が設定される。

【 0 0 6 2 】

図 9 は、図 4 に示した学習データ期間設定処理 S 4 2 4 を説明するフローチャートである。以下、同図とともに学習データ期間設定処理 S 4 2 4 について説明する。尚、S 4 2 3 で取得した要求期間に重なる期間の観測データ 1 1 3 が既に取得されている場合、人工データ生成部 1 4 0 は、観測データ 1 1 3 を優先して学習データ 1 1 4 として採用されるように学習データ期間を設定する。

10

【 0 0 6 3 】

図 1 0 は、以下の説明で用いる観測データ 1 1 3 の一例である。同図に示すように、例示する観測データ 1 1 3 は、日時 1 0 1 1 と観測値 1 0 1 2 の各項目を有する複数のエントリを含む。上記項目のうち日時 1 0 1 1 には、当該エントリの観測値が取得された日時が設定される。観測値 1 0 1 2 には、観測対象から実際に取得した観測値が設定される。

【 0 0 6 4 】

図 9 に示すように、まず学習データ期間設定部 1 5 0 は、記憶部 1 1 0 が観測データ 1 1 3 を記憶しているか否か（学習データ生成装置 1 0 0 が観測データ 1 1 3 を取得しているか否か）を確認する（S 9 0 1）。記憶部 1 1 0 が観測データ 1 1 3 を記憶している場合（S 9 0 1：YES）、学習データ期間設定部 1 5 0 は、観測データ 1 1 3 の期間の終了時点、学習データ期間の終了時点 t_{end} として設定する（S 9 0 2）。その後、処理は S 9 0 4 に進む。一方、記憶部 1 1 0 が観測データ 1 1 3 を記憶していない場合（S 9 0 1：NO）、学習データ期間設定部 1 5 0 は、生成元データ 1 1 1 の期間の終了時点、学習データ期間の終了時点 t_{end} として設定する（S 9 0 3）。その後、処理は S 9 0 4 に進む。

20

【 0 0 6 5 】

S 9 0 4 では、学習データ期間設定部 1 5 0 は、S 9 0 2 又は S 9 0 3 で設定した学習データ期間の終了時点 t_{end} から、要求期間（人工データ生成部 1 4 0 が S 4 1 2 で取得した要求期間）だけ過去に遡った日時（以下、「仮開始時点 tmp_Tstart 」と称する。）を取得する。

30

【 0 0 6 6 】

続いて、学習データ期間設定部 1 5 0 は、人工データ 1 1 2 の期間、生成元データ 1 1 1 の期間、及び観測データ 1 1 3 の期間と、仮開始時点 tmp_Tstart とを比較する（S 9 0 5）。仮開始時点 tmp_Tstart が人工データ 1 1 2 の期間中である場合（S 9 0 5：人工データの期間中）、学習データ期間の開始時点 t_{start} に仮開始時点 tmp_Tstart を設定する（S 9 0 6）。一方、仮開始時点 tmp_Tstart が、生成元データ 1 1 1 の期間中か観測データ 1 1 3 の期間中である場合（S 9 0 5：生成元データ or 観測データの期間中）、学習データ期間の開始時点 t_{start} に生成元データ 1 1 1 の開始時点を設定する（S 9 0 7）。

40

【 0 0 6 7 】

以上の処理により、学習データ期間の開始時点 t_{start} と終了時点 t_{end} が設定され、学習データ期間の設定が完了する。尚、S 9 0 2、S 9 0 3、及び S 9 0 7 の処理により、学習データ生成処理 S 4 3 1 において、観測データ 1 1 3 又は生成元データ 1 1 1 が人工データ 1 1 2 よりも優先して学習データ 1 1 4 として採用されるようになる。

【 0 0 6 8 】

図 1 1 は、図 4 に示した学習データ生成処理 S 4 3 1 を説明するフローチャートである。以下、同図とともに学習データ生成処理 S 4 3 1 について説明する。

50

【 0 0 6 9 】

まず学習データ生成部 1 6 0 は、記憶部 1 1 0 から、学習データ期間設定処理 S 4 2 4 により設定された学習データ期間に重なる期間の、観測データ 1 1 3、生成元データ 1 1 1、及び人工データ 1 1 2を取得する（S 1 1 0 1～S 1 1 0 3）。

【 0 0 7 0 】

続いて、学習データ生成部 1 6 0 は、取得した生成元データ 1 1 1と取得した人工データ 1 1 2を時系列方向に連結（人工データ 1 1 2、生成元データ 1 1 1の時系列順に連結）した中間連結データを生成する（S 1 1 0 4）。

【 0 0 7 1 】

続いて、学習データ生成部 1 6 0 は、中間連結データと、取得した観測データ 1 1 3を時系列方向に連結して学習データ 1 1 4を生成する（S 1 1 0 5）。尚、学習データ期間の全期間に対応する観測データ 1 1 3が存在する場合、学習データ 1 1 4は全て観測データ 1 1 3によるものとなる。また学習データ期間の一部の期間に観測データ 1 1 3が重なる場合、学習データ 1 1 4の全期間のうち、学習データ期間の開始時点から観測データ 1 1 3の開始時点までは中間連結データによるものとなり、観測データ 1 1 3の開始時点から学習データ期間の終了時点までは観測データ 1 1 3によるものとなる。このように観測データ 1 1 3が存在する場合は観測データ 1 1 3が学習データ 1 1 4として優先的に採用されるので、機械学習システム 1 の本番運用が開始された後、実際に取得されたデータである観測データ 1 1 3のみを学習データ 1 1 4として用いて学習する運用状態に早期に移行することができる。

【 0 0 7 2 】

図 1 2 に学習データ生成処理 S 4 3 1 により生成される学習データ 1 1 4 の一例を示す。例示する学習データ 1 1 4 は、日時 1 2 0 1 及び観測値 1 2 0 2 の各項目を有する複数のエントリを含む。上記項目のうち、日時 1 2 0 1 には、人工データ 1 1 2 の日時 8 1 1、生成元データ 1 1 1 の日時 7 0 1、及び観測データ 1 1 3 の日時 1 0 1 1 のいずれかの値に基づく日時が設定される。観測値 1 2 0 2 には、人工データ 1 1 2 の観測値 8 1 2、生成元データ 1 1 1 の観測値 7 0 2、及び観測データ 1 1 3 の観測値 1 0 1 2 のいずれかに基づく観測値が設定される。

【 0 0 7 3 】

以上に説明したように、第 1 実施形態の学習データ生成装置 1 0 0 によれば、機械学習モデル 2 3 の精度を確保するために必要な期間の学習データを用意することが難しい場合でも、上記期間について有効な学習データをユーザの手を煩わせることなく効率よく生成して提供することができる。

【 0 0 7 4 】

また学習データ生成装置 1 0 0 は、各複製データに個別に雑音を付加した人工データ 1 1 2を用いて学習データ 1 1 4を生成するので、機械学習モデル 2 3 の過学習の抑制効果が期待される多様性を有する学習データ 1 1 4を生成することができ、機械学習モデル 2 3 の推論精度を向上することができる。また人工データ 1 1 2に白色雑音を付加することで、実際の変動に近い変動を再現することができ、例えば、観測データが正規分布に従うことを前提として機能する機械学習モデル 2 3 の推論精度を高めることができる。

【 0 0 7 5 】

また図 9 に示したように、学習データ期間設定処理 S 4 2 4 において、観測データ 1 1 3が学習データ 1 1 4に優先して採用されるように学習データ期間が設定されるので、機械学習システム 1 の本番運用が開始された後は、実際に取得されたデータである観測データ 1 1 3のみを学習データ 1 1 4として用いて学習する運用状態に早期に移行することができる。このため、本番運用の開始後、推論装置 2 の推論精度を早期に向上することができる。

【 0 0 7 6 】

また学習データ生成装置 1 0 0 は、生成元データ 1 1 1よりも過去の期間の人工データ 1 1 2を生成することが可能であり、新たに取得される観測データ 1 1 3の期間と重なら

10

20

30

40

50

ないように人工データ 1 1 2 を生成することができ、例えば、人工データ 1 1 2 を観測データ 1 1 3 で置換するといった煩雑な処理を発生させないようにすることができる。

【 0 0 7 7 】

尚、以上では、生成元データ 1 1 1 よりも過去の期間の人工データ 1 1 2 を生成する場合を例示したが、生成元データ 1 1 1 よりも未来の期間の人工データ 1 1 2 を生成してもよい。これにより、例えば、現実の振る舞いを最もよく反映していると考えられる時期における過去の時系列データを生成元データ 1 1 1 として用いて所望の未来の時期の学習データ 1 1 4 を生成することができる。尚、この場合、例えば、図 5 の S 5 0 4 において未来の期間の人工データ 1 1 2 とする各複製データに - 1 から始まる負の整数を複製番号を割り当て、過去の期間の各複製データに割り当てた正の複製番号と負の複製番号の絶対値との合計が図 5 の S 5 0 2 で取得した複製回数と一致するようにする。そのようにすることで、S 5 0 7 で複製データの期間に複製番号を乗算した値を基準とする日時に加算するだけで、日時（期間）情報を容易に算出することができる。

10

【 0 0 7 8 】

[第 2 実施形態]

続いて、第 2 実施形態について説明する。第 2 実施形態の学習データ生成装置 1 0 0 は、生成元データ 1 1 1 を分解することにより得られる構成要素（後述するトレンド、周期変動、及び残差）に基づき人工データ 1 1 2 を生成する。尚、第 2 実施形態の機械学習システム 1 の基本的な構成並びに機械学習システム 1 において実行される処理の流れは、図 1 乃至図 4 とともに説明した第 1 実施形態の機械学習システム 1 と基本的に共通するが、人工データ生成部 1 4 0 の機能の一部が異なる。以下では、第 1 実施形態と異なる部分を中心として説明する。

20

【 0 0 7 9 】

図 1 3 は、第 2 実施形態として示す人工データ生成処理 S 4 1 3 を説明するフローチャートである。また図 1 4 は、人工データ生成処理 S 4 1 3 の実行過程で生成されるデータを模式的に示した図である。また図 1 5 は、以下の説明で用いる生成元データ 1 1 1 の一例である。以下、これらの図を参照しつつ、第 2 実施形態の人工データ生成処理 S 4 1 3 について詳述する。

【 0 0 8 0 】

図 1 4 (A) に示すように、例示する生成元データ 1 1 1 は、小周期 T_p (= 1 日) と大周期 T (= 7 日) を有する、2019 年 1 1 月 1 5 日 0 時 0 分 0 秒から 2019 年 1 1 月 2 2 日 2 3 時 5 0 分 0 秒までの 10 分間隔の 8 日分のデータ (8 回の小周期 T_p (大周期 7 日 $\times$ 1 + 小周期 1 日)) からなる。尚、以下の説明において、生成元データ 1 1 1 の開始時点をととする。また以下の説明において、図 4 の S 4 1 2 で受け付けた要求期間は 28 週とする。また図 4 の S 4 1 2 において、生成元データ 1 1 1 の周期数として 1 周期 (大周期 1 回分) を受け付けているものとする。

30

【 0 0 8 1 】

図 1 3 に示すように、まず人工データ生成部 1 4 0 は、ユーザインタフェースを解して小周期 T_p (1 日) と大周期 T (7 日) の入力を受け付ける (S 1 3 0 1) 。

【 0 0 8 2 】

40

続いて、人工データ生成部 1 4 0 は、S 4 1 2 で受け付けた要求期間 (28 週) 以上の期間となる、生成元データ 1 1 1 の 1 周期の期間の倍数の最小値 (最小周期数) を求める (S 1 3 0 2) 。

【 0 0 8 3 】

続いて、人工データ生成部 1 4 0 は、S 1 3 0 1 で求めた最小周期数から生成元データ 1 1 1 に含まれている周期数を減じた値を求め、求めた値を生成元データ 1 1 1 に含まれている周期数で割った値を小数点以下切り上げ、更に 1 を加算して得られる値を複製回数とする (S 1 3 0 3) 。尚、1 を加算するのは、生成元データ 1 1 1 について後述するトレンドを移動平均により求めることに起因して生じる時間差 (後述する $T_p / 2$) により、生成した人工データ 1 1 2 の期間が要求期間を満たさなくなる可能性があるからである

50

。本例の場合、生成元データ 1 1 1 の周期数が 1 であり、S 4 1 2 で受け付けた要求期間が 2 8 週であるので、複製回数として 2 8 が得られる。

【 0 0 8 4 】

続いて、人工データ生成部 1 4 0 は、小周期 T_p を変動周期として、生成元データ 1 1 1 を構成要素（トレンド、周期変動、残差）に分解する（S 1 3 0 4）。ここでトレンドとは、時系列データにおける長期的な変動を表す要素（Trend component）のことをいう。また周期変動とは、時系列データにおいて一定期間ごとに周期的に現れる要素（Seasonal component）のことをいう。また残差とは、時系列データにおいて、トレンドと周期変動を除くことにより残る細かな変動要素（Redidual component）のことをいう。本実施形態は、上記分解を非特許文献 1 に記載されている S T L（Seasonal-Trend Decomposition Procedure Based on Loess）を用いて行うものとするが、上記分解の方法は必ずしも限定されない。

10

【 0 0 8 5 】

図 1 4（B）は、図 1 4（A）の生成元データ 1 1 1 を分解することにより得られる構成要素である。同図において、（B - 1）はトレンド、（B - 2）は周期変動、（B - 3）は残差である。

【 0 0 8 6 】

図 1 6 に、S 1 3 0 4 で得られるデータ（以下、「中間データ 1 6 0 0」と称する。）を示す。同図に示すように、中間データ 1 6 0 0 は、日時 1 6 0 1、観測値 1 6 0 2、トレンド 1 6 0 3、周期変動 1 6 0 4、及び残差 1 6 0 5 の各項目を有する複数のエントリを含む。同図において、「-」は、データが欠落していることを示す。日時 1 6 0 1 及び観測値 1 6 0 2 は、生成元データ 1 1 1 における日時 1 2 0 1 及び観測値 1 2 0 2 に対応する。トレンド 1 6 0 3、周期変動 1 6 0 4、及び残差 1 6 0 5 には夫々、S 1 3 0 4 で得られた、観測値 1 6 0 2 の構成要素であるトレンド、周期変動、及び残差を示す値が設定される。尚、トレンド 1 6 0 3 と残差 1 6 0 5 は、いずれも期間の両端において、S T L を実行する際に指定した小周期の半分の期間（ $= T_p / 2$ ）の値が欠落する。本例では、2 0 1 9 年 1 1 月 1 5 日 0 時 0 分 0 秒から 2 0 1 9 年 1 1 月 1 5 日 1 1 時 5 0 分 0 秒までの期間と、2 0 1 9 年 1 1 月 2 2 日 1 2 時 0 分 0 秒から、2 0 1 9 年 1 1 月 2 2 日 2 3 時 5 0 分 0 秒までの期間においてトレンド 1 6 0 3 と残差 1 6 0 5 の値が欠落している。

20

【 0 0 8 7 】

図 1 3 に戻り、続いて、人工データ生成部 1 4 0 は、中間データ 1 6 0 0 のトレンド 1 6 0 3 の値が存在する（欠落していない）日時について、同じ日時 1 6 0 1 のトレンド 1 6 0 3 と周期変動 1 6 0 4 の合計値（以下、「複製元観測値」と称する。）を求める（S 1 3 0 5）。S 1 3 0 5 の処理は、図 1 4 では（B - 1）に示すトレンドと（B - 2）に示す周期変動とを合成する処理に相当する。当該処理を実行することにより、図 1 4（C）に示すデータ（以下、「複製元データ 1 7 0 0」と称する。）が得られる。

30

【 0 0 8 8 】

図 1 7 に複製元データ 1 7 0 0 の一例を示す。同図に示すように、複製元データ 1 7 0 0 は、日時 1 7 0 1、観測値 1 7 0 2、トレンド 1 7 0 3、周期変動 1 7 0 4、残差 1 7 0 5、及び複製元観測値 1 7 0 6 の各項目を有する複数のエントリを含む。上記項目のうち、日時 1 7 0 1 には、中間データ 1 6 0 0 のエントリのうち、トレンド 1 7 0 3 の値を有するエントリの日時 1 6 0 1 の値が設定される。観測値 1 7 0 2 には、中間データ 1 6 0 0 のエントリのうち、日時 1 7 0 1 の値に対応する観測値 1 6 0 2 の値が設定される。トレンド 1 7 0 3 には、中間データ 1 6 0 0 のエントリのうち、日時 1 7 0 1 の値に対応するトレンド 1 6 0 3 の値が設定される。周期変動 1 7 0 4 には、中間データ 1 6 0 0 のエントリのうち、日時 1 7 0 1 の値に対応する周期変動 1 6 0 4 の値が設定される。残差 1 7 0 5 には、中間データ 1 6 0 0 のエントリのうち、日時 1 7 0 1 の値に対応する残差 1 6 0 5 の値が設定される。複製元観測値 1 7 0 6 には、中間データ 1 6 0 0 のエントリのうち、日時 1 7 0 1 の値に対応するトレンド 1 6 0 3 の値と日時 1 7 0 1 の値に対応する周期変動 1 6 0 4 の値とを合計した値が設定される。

40

50

【 0 0 8 9 】

図 1 3 に戻り、続いて、人工データ生成部 1 4 0 は、S 1 3 0 3 で求めた複製回数だけ複製元データ 1 7 0 0 を複製する (S 1 3 0 6)。以下、複製された各データのことを「複製データ」と称する。

【 0 0 9 0 】

続いて、人工データ生成部 1 4 0 は、1 から始まる自然数を、生成した各複製データに順に割り当て、記憶部 1 1 0 が、各複製データに割り当てられた番号 (以下、「複製番号」と称する。) を複製データの夫々に対応づけて記憶する (S 1 3 0 7)。尚、第 1 実施形態で述べたのと同様に、当該処理において 1 から始まる自然数とは別に - 1 から始まる負の整数を複製番号として割り当てることにより生成元データ 1 1 1 よりも未来の期間における人工データ 1 1 2 を生成してもよい。この場合、第 1 実施形態の場合と同様に、正の複製番号と負の複製番号の絶対値との合計値が S 1 3 0 3 で取得した複製回数と一致するようにする。

10

【 0 0 9 1 】

続いて、人工データ生成部 1 4 0 は、割り当てた複製番号の逆順に、複製データを時系列方向に連結していくことにより一次人工データを生成する (S 1 3 0 8)。

【 0 0 9 2 】

続いて、人工データ生成部 1 4 0 は、一次人工データのうち、複製番号が 1 で日時が $t + T$ から $t + T + T_p / 2$ の期間に該当するエントリを削除する (S 1 3 0 9)。

【 0 0 9 3 】

続いて、人工データ生成部 1 4 0 は、一次人工データの各エントリに対して、生成元データ 1 1 1 の各エントリの日時 7 0 1 の値を複製したデータ (以下、「参照元日時」と称する。) を付与する (S 1 3 1 0)。

20

【 0 0 9 4 】

続いて、人工データ生成部 1 4 0 は、一次人工データの各エントリの参照元日時を、基準とする日時から遡った値に更新することにより各エントリの日時を生成する (S 1 3 1 1)。この処理により、例えば、複製番号 2 の複製データにおける 2 0 1 9 年 1 1 月 1 5 日 1 2 時 0 分 0 秒の変更後の日時は、2 週分遡った 2 0 1 9 年 1 1 月 1 日 1 2 時 0 分 0 秒となる。

【 0 0 9 5 】

続いて、人工データ生成部 1 4 0 は、複製データを連結する際の境界となる時点における、境界の前後の複製データの周期変動の差分 d を求める。具体的には、人工データ生成部 1 4 0 は、周期変動について、 $t + T_p / 2$ の時点のエントリの値と当該時点から一つ前の時点のエントリの値との差分 d を求める (S 1 3 1 2)。例えば、図 1 6 の中間データ 1 6 0 0 の例では、 $t + T_p / 2$ は 2 0 1 9 年 1 1 月 1 5 日 1 2 時 0 分 0 秒であるため、同日時の周期変動 1 6 0 4 として 1 5 2 が得られる。また同日時の一つ前の時点である 2 0 1 9 年 1 1 月 1 5 日 1 1 時 5 0 分 0 秒の周期変動 1 6 0 4 として 1 5 1 が得られる。このため、本例では差分 d として 1 が得られる。

30

【 0 0 9 6 】

続いて、人工データ生成部 1 4 0 は、一次人工データの各エントリに対して、差分 d と各エントリの複製番号との積として求められる値を、一次人工データの各エントリの観測値に反映 (例えば、加算又は減算) する (S 1 3 1 3)。即ち、短期間のデータから取得されるトレンド (差分 d) が要求期間において継続していたと仮定した場合における一次人工データを生成する。当該処理の実行後、一次人工データは図 1 4 (D) のようになる。

40

【 0 0 9 7 】

続いて、人工データ生成部 1 4 0 は、S 1 3 0 4 で得られた残差の分散 s^2 を求める (S 1 3 1 4)。

【 0 0 9 8 】

続いて、人工データ生成部 1 4 0 は、上記分散 s^2 を有する一次人工データの期間に

50

対応する期間について白色雑音を生成する（S 1 3 1 5）。当該処理を実行することにより生成される白色雑音は図 1 4（E）のようになる。

【0 0 9 9】

続いて、人工データ生成部 1 4 0 は、一次人工データに対して、各エントリの観測値を複製した値（以下「参照元観測値」と称する。）を生成する（S 1 3 1 6）。

【0 1 0 0】

続いて、人工データ生成部 1 4 0 は、一次人工データに対して、S 1 3 1 5 で生成した白色雑音を変動値として付与する（S 1 3 1 7）。

【0 1 0 1】

続いて、人工データ生成部 1 4 0 は、人工データ 1 1 2 の各エントリの観測値に、夫々の参照元観測値に夫々の変動値を加算した値を設定して人工データを生成する（S 1 3 1 8）。当該処理を実行することにより生成される人工データ 1 1 2 は、図 1 4（F）のようになる。

【0 1 0 2】

図 1 8 に人工データ 1 1 2 の一例を示す。同図に示すように、人工データ 1 1 2 は、日時 1 8 0 1、観測値 1 8 0 2、参照元観測値 1 8 0 3、変動値 1 8 0 4、参照元日時 1 8 0 5、複製番号 1 8 0 6 の各項目を有する複数のエントリを含む。上記項目のうち、日時 1 8 0 1 には、S 1 3 1 1 において生成された日時が設定される。日時 1 8 0 1 の値は、各エントリを一意に識別するための識別子としても機能する。観測値 1 8 0 2 には、S 1 3 1 8 で求めた観測値が設定される。参照元観測値 1 8 0 3 には、S 1 3 1 6 で生成された参照元観測値が設定される。変動値 1 8 0 4 には、S 1 3 1 7 で付与された白色雑音の値が設定される。参照元日時 1 8 0 5 には、S 1 3 1 0 で付与された日時が設定される。参照元日時 1 8 0 5 は、生成元データ 1 1 1 の日時 7 0 1 に対応し、当該エントリが生成元データ 1 1 1 の日時 7 0 1 のエントリに基づくものであることを示す。複製番号 1 8 0 6 は、S 1 3 0 7 で割り当てられた複製番号が設定される。

【0 1 0 3】

以上に説明したように、第 2 実施形態の学習データ生成装置 1 0 0 は、2 つの周期を含む時系列データである生成元データ 1 1 1 を、トレンド、周期変動、残差に分解し、トレンドと周期変動とに基づき雑音のない複製元データを生成し、また残差から得た分散 s^2 に基づき白色雑音を生成し、複製元データと白色雑音から人工データを生成する。このため、現実にかかる変動過程に近い変動過程を再現した学習データを生成することができ、これを用いて機械学習モデル 2 3 の学習を行うことで推論装置 2 の推論精度を向上することができる。

【0 1 0 4】

また学習データ生成装置 1 0 0 は、複製データを連結する境界となる時点の前後の複製データの周期変動の差分 d （短期間のトレンド）を取得し、上記境界において複製番号と差分 d との積の値だけ観測値を変化させつつ複数の複製データを連結することにより人工データ 1 1 2 を生成する。このため、長期のトレンドを考慮した学習データ 1 1 4 を生成することができ、機械学習モデル 2 3 を精度よく学習することができる。

【0 1 0 5】

以上、本発明の実施形態につき説明したが、本発明は上記した実施形態に限定されるものではなく、様々な変形例が含まれる。また例えば、上記した実施形態は本発明を分かりやすく説明するために構成を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限定されるものではない。また各実施形態の構成の一部について、他の構成に追加、削除、置換することが可能である。

【0 1 0 6】

また上記の各構成、機能、処理部、処理手段等は、それらの一部又は全部を、例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い。また実施形態で示した各機能を実現するソフトウェアのプログラムコードによっても実現できる。この場合、プログラムコードを記録した記憶媒体を情報処理装置（コンピュータ）に提供し、その情報処理装

10

20

30

40

50

置が備えるプロセッサが記憶媒体に格納されたプログラムコードを読み出す。この場合、記憶媒体から読み出されたプログラムコード自体が以上の実施形態の機能を実現することになり、そのプログラムコード自体、及びそれを記憶した記憶媒体は本発明を構成することになる。このようなプログラムコードを供給するための記憶媒体としては、例えば、ハードディスク、SSD (Solid State Drive)、光ディスク、光磁気ディスク、CD-R、フレキシブルディスク、CD-ROM、DVD-ROM、磁気テープ、不揮発性のメモ리카ード、ROM等が用いられる。

【0107】

以上の実施形態において、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、製品上必ずしも全ての制御線や情報線を示しているとは限らない。全ての構成が相互に接続されていてもよい。また以上では各種の情報を表形式で例示したが、これらの情報は表以外の形式で管理してもよい。

10

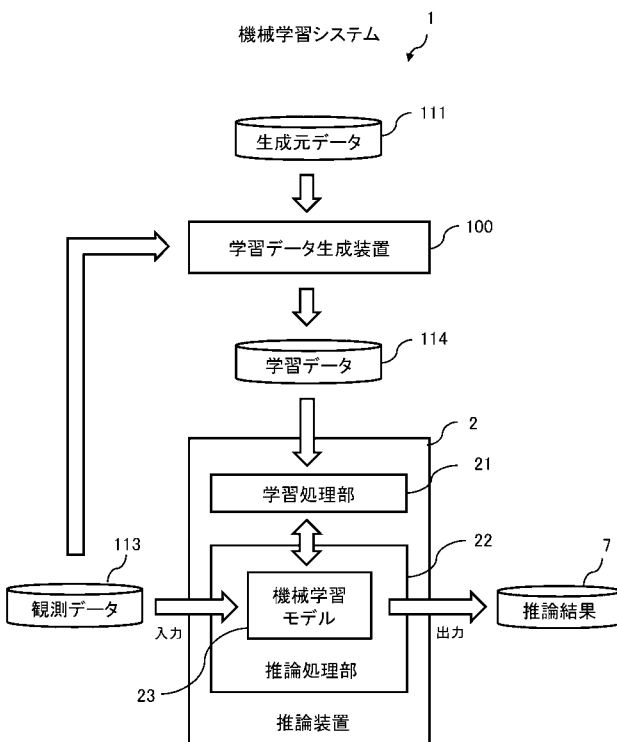
【符号の説明】

【0108】

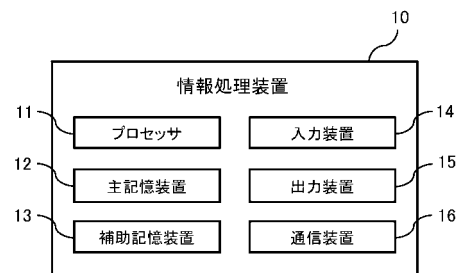
1 機械学習システム、2 推論装置、21 学習処理部、22 推論処理部、23 機械学習モデル、100 学習データ生成装置、110 記憶部、112 人工データ、111 生成元データ、113 観測データ、114 学習データ、120 観測データ取得部、130 生成元データ取得部、140 人工データ生成部、150 学習データ期間設定部、160 学習データ生成部、170 学習データ出力部、S400 学習データ生成処理、S413 人工データ生成処理、S424 学習データ期間設定処理、S431 学習データ生成処理

20

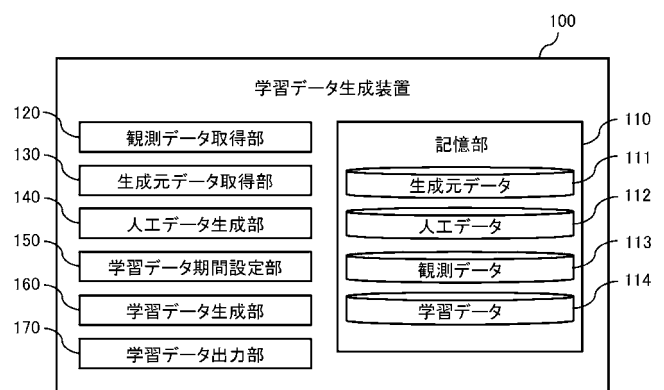
【図1】



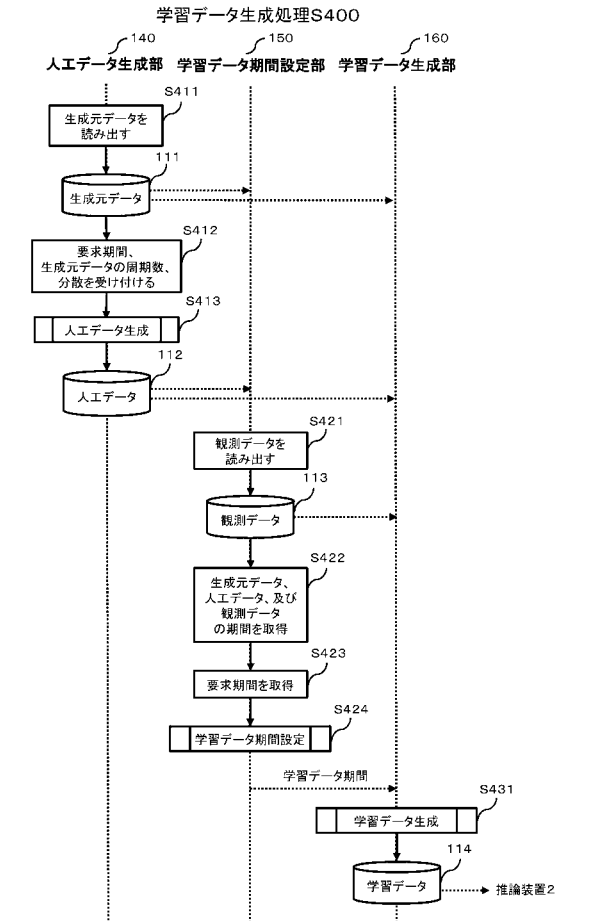
【図2】



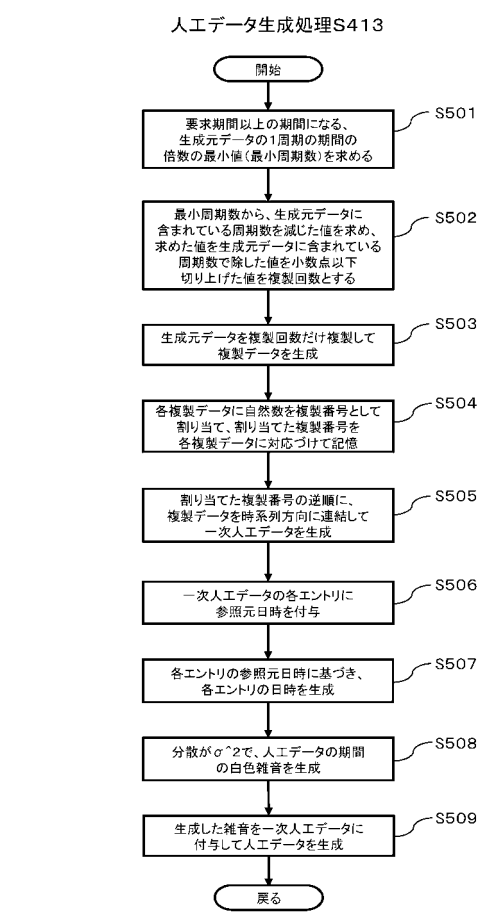
【図3】



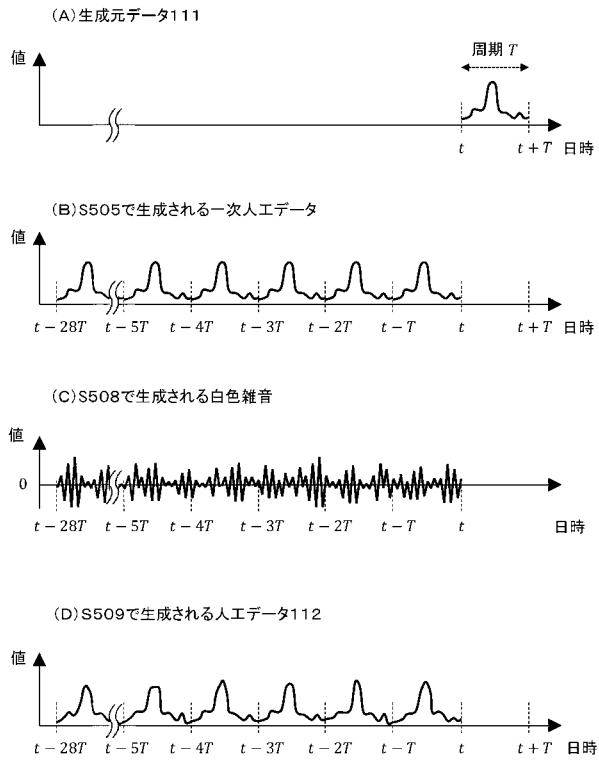
【 図 4 】



【 図 5 】



【 図 6 】



【 図 7 】

生成元データ 111

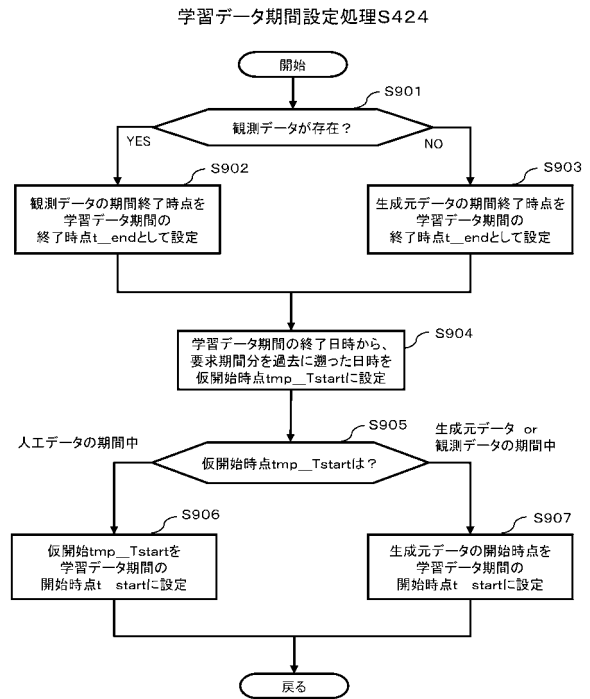
日時 701	観測値 702
2019/11/15 00:00:00	51
2019/11/15 00:10:00	72
2019/11/15 00:20:00	60
:	:
2019/11/21 23:30:00	66
2019/11/21 23:40:00	48
2019/11/21 23:50:00	63

【図 8】

人工データ 112

日時	観測値	参照元観測値	変動値	参照元日時	複製番号
2019/05/10 00:00:00	58	51	7	2019/11/15 00:00:00	27
2019/05/10 00:10:00	68	72	-4	2019/11/15 00:10:00	27
2019/05/10 00:20:00	62	60	2	2019/11/15 00:20:00	27
⋮	⋮	⋮	⋮	⋮	⋮
2019/05/16 23:40:00	54	48	6	2019/11/21 23:40:00	27
2019/05/16 23:50:00	69	63	6	2019/11/21 23:50:00	27
2019/05/17 00:00:00	46	51	-5	2019/11/15 00:00:00	26
2019/05/17 00:10:00	69	72	-3	2019/11/15 00:10:00	26
⋮	⋮	⋮	⋮	⋮	⋮
2019/11/14 23:40:00	52	48	4	2019/11/21 23:40:00	1
2019/11/14 23:50:00	55	63	-8	2019/11/21 23:50:00	1

【図 9】

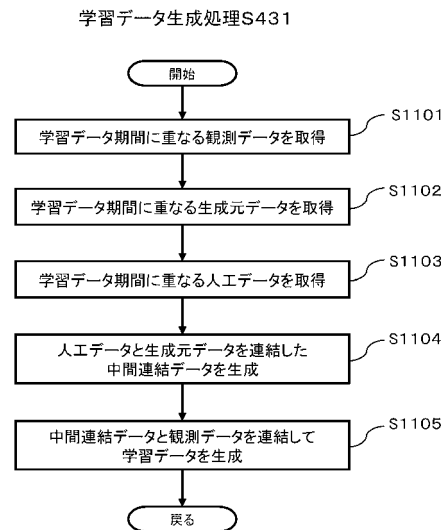


【図 10】

観測データ 113

日時	観測値
2019/11/22 00:00:00	51
2019/11/22 00:10:00	72
2019/11/22 00:20:00	60
⋮	⋮
2019/11/28 23:50:00	63
2019/11/29 00:00:00	52
⋮	⋮
2019/12/19 23:50:00	63

【図 11】



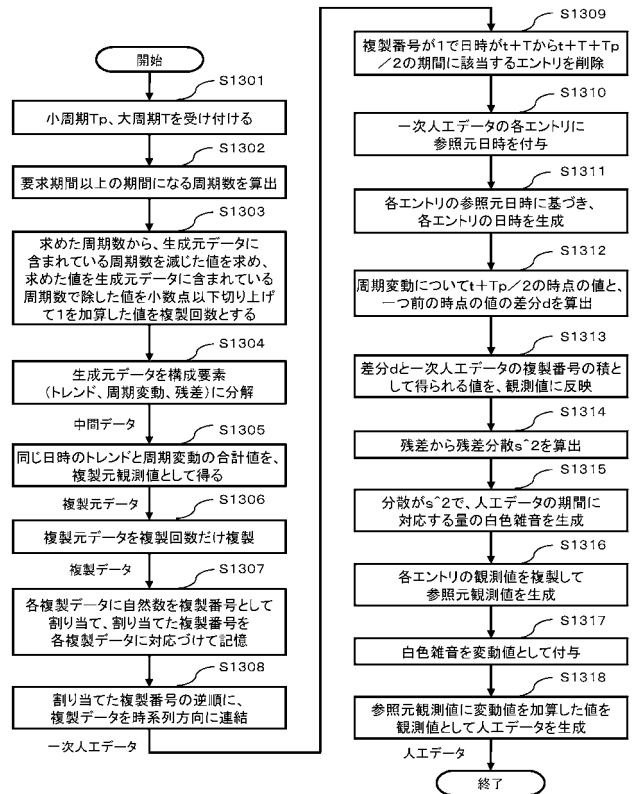
【図 1 2】

学習データ 114

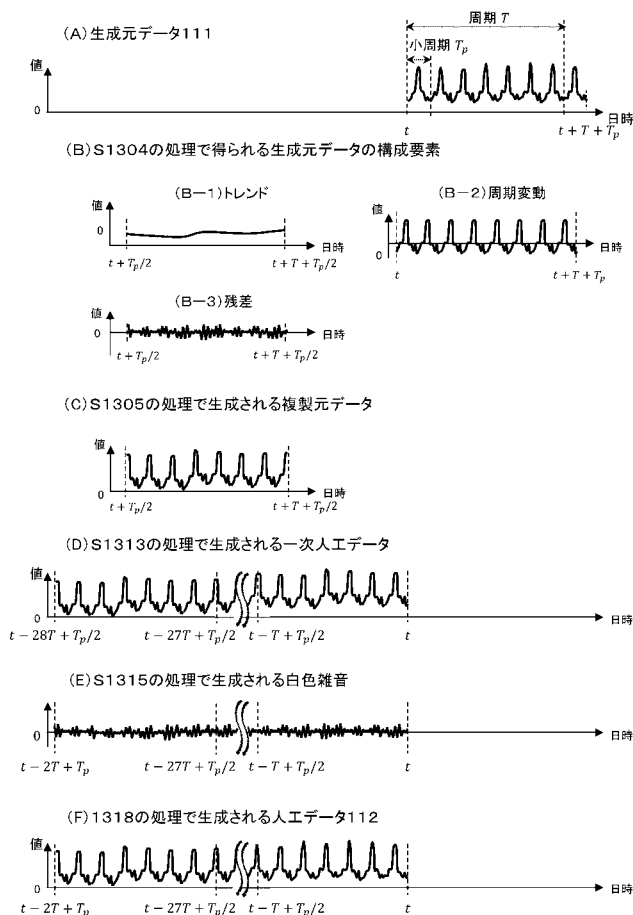
日時	観測値
2019/05/10 00:00:00	58
2019/05/10 00:10:00	68
2019/05/10 00:20:00	62
⋮	⋮
2019/11/14 23:50:00	55
2019/11/15 00:00:00	51
⋮	⋮
2019/11/21 23:50:00	60
2019/11/22 00:00:00	51
⋮	⋮
2019/12/19 23:50:00	63

【図 1 3】

人工データ生成処理S413



【図 1 4】



【図 1 5】

生成元データ 111

日時	観測値
2019/11/15 00:00	51
2019/11/15 0:10:00	72
2019/11/15 0:20:00	60
2019/11/15 0:30:00	97
⋮	⋮
2019/11/16 23:50:00	60
2019/11/16 0:00:00	52
⋮	⋮
2019/11/21 23:50:00	63
2019/11/22 0:00:00	50
⋮	⋮
2019/11/22 23:50:00	63

【図 16】

中間データ1600

1601 日時	1602 観測値	1603 トレンド	1604 周期変動	1605 残差
2019/11/15 0:00:00	51	—	-132	—
2019/11/15 0:10:00	72	—	-143	—
⋮	⋮	⋮	⋮	⋮
2019/11/15 11:40:00	248	—	133	—
2019/11/15 11:50:00	267	—	151	—
2019/11/15 12:00:00	278	126	152	0
2019/11/15 12:10:00	270	122	150	-2
⋮	⋮	⋮	⋮	⋮
2019/11/22 11:40:00	278	148	133	-3
2019/11/22 11:50:00	286	156	151	-21
2019/11/22 12:00:00	301	—	152	—
2019/11/22 12:10:00	290	—	150	—
⋮	⋮	⋮	⋮	⋮
2019/11/22 23:40:00	88	—	-133	—
2019/11/22 23:50:00	63	—	-123	—

【図 17】

複製元データ1700

1701 日時	1702 観測値	1703 トレンド	1704 周期変動	1705 残差	1706 複製元観測値
2019/11/15 12:00:00	278	126	152	0	278
2019/11/15 12:10:00	270	122	150	-2	272
⋮	⋮	⋮	⋮	⋮	⋮
2019/11/22 11:40:00	278	148	133	-3	281
2019/11/22 11:50:00	286	156	151	-21	307

【図 18】

人工データ112

1801 日時	1802 観測値	1803 参照元 観測値	1804 変動値	1805 参照元日時	1806 複製番号
2019/05/03 12:00:00	284	278	7	2019/11/15 12:00:00	28
2019/05/03 12:10:00	276	272	4	2019/11/15 12:10:00	28
2019/05/03 12:20:00	267	269	-2	2019/11/15 12:20:00	28
⋮	⋮	⋮	⋮	⋮	⋮
2019/05/16 23:40:00	26	20	6	2019/11/21 23:40:00	28
2019/05/16 23:50:00	41	35	6	2019/11/21 23:50:00	28
2019/05/17 00:00:00	19	24	-5	2019/11/22 00:00:00	27
2019/05/17 00:10:00	42	45	-3	2019/11/22 00:10:00	27
⋮	⋮	⋮	⋮	⋮	⋮
2019/11/14 23:40:00	52	48	4	2019/11/21 23:40:00	1
2019/11/14 23:50:00	55	63	-8	2019/11/21 23:50:00	1

フロントページの続き

(72)発明者 山崎 伊織

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