

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2020年9月17日(17.09.2020)



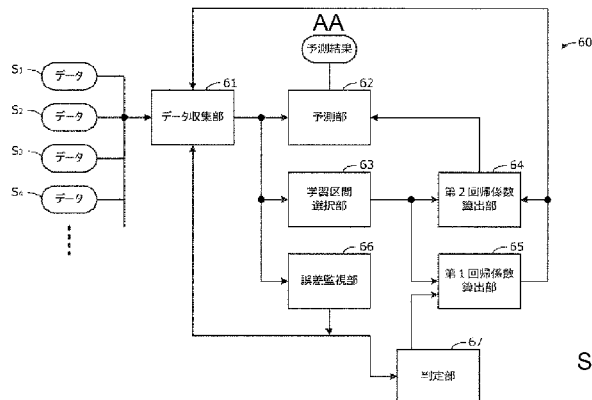
(10) 国際公開番号

WO 2020/184560 A1

- (51) 国際特許分類:
G06N 20/00 (2019.01)
- (21) 国際出願番号: PCT/JP2020/010303
- (22) 国際出願日: 2020年3月10日(10.03.2020)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2019-045555 2019年3月13日(13.03.2019) JP
- (71) 出願人: 日本電信電話株式会社 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目5番1号 Tokyo (JP).
- (72) 発明者: 高嶋 洋一 (TAKASHIMA, Yoichi); 〒1808585 東京都武蔵野市緑町三丁目9番11号 NTT 知的財産センタ内 Tokyo (JP). 湯口 昌宏(YUGUCHI, Masahiro); 〒1808585 東京都武蔵野市緑町三丁目9番11号 NTT 知的財産センタ内 Tokyo (JP). 山田 智広(YAMADA, Tomohiro); 〒1808585 東京都武蔵野市緑町三丁目9番11号 NTT 知的財産センタ内 Tokyo (JP).
- (74) 代理人: 特許業務法人太陽国際特許事務所 (TAIYO, NAKAJIMA & KATO); 〒1600022 東京都新宿区新宿4丁目3番17号 Tokyo (JP).
- (81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH,

(54) Title: DATA PREDICTION DEVICE, DATA PREDICTION METHOD, AND DATA PREDICTION PROGRAM

(54) 発明の名称: データ予測装置、データ予測方法、及びデータ予測プログラム



- 61 Data collection unit
62 Prediction unit
63 Learning segment selection unit
64 Second regression coefficient computation unit
65 First regression coefficient computation unit
66 Error monitoring unit
67 Determination unit
S1, S2, S3, S4 Data
AA Prediction result

(57) Abstract: The purpose of the present invention is to reduce the quantity of data used to calculate regression coefficients, reduce the bandwidth for collecting and transmitting data, and reduce computation throughput in a learning process. A data prediction device according to the present disclosure includes: a first regression coefficient computation unit that computes regression coefficients on the basis of a plurality of data and using a first computation method for calculating regression coefficients by L1 regularization, and transmits selection information for selecting data corresponding to the regression coefficients having an absolute value equal to or greater than a threshold value to a data collection unit; a determination unit that determines which of a first computation method and a second computation method to use; a second regression coefficient computation unit that computes regression coefficients using the determined computation method; and a prediction unit that uses the plurality of data if the computation method is determined to be the first computation method and uses selected data if the computation method is determined to be the second computation method, and outputs a prediction result that was predicted on the basis of the data used and the regression coefficients computed by the second regression coefficient computation unit.

CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IR, IS, JO, KE, KG, KH,
KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY,
MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ,
NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT,
QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, WS, ZA, ZM, ZW.

- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

- 一 国際調査報告(条約第21条(3))

(57) 要約: 回帰係数を求める際に使用するデータの数を低減し、データを収集及び伝送する回線容量を低減し、学習過程における計算処理量を低減する、ことを目的とする。本開示のデータ予測装置は、L1正則化により回帰係数を求める第1算出法を用いて複数のデータに基づいて回帰係数を算出し、絶対値が閾値以上の回帰係数に対応するデータを選択するための選択情報をデータ収集部に送信する第1回帰係数算出部と、第1算出法及び第2算出法のいずれの算出法を用いるかを判定する判定部と、判定された算出法を用いて回帰係数を算出する第2回帰係数算出部と、判定された算出法が第1算出法である場合複数のデータを使用し、判定された算出法が第2算出法である場合選択データを使用し、使用したデータ及び第2回帰係数算出部で算出された回帰係数に基づいて予測した予測結果を出力する予測部と、を含む。

明 細 書

発明の名称：

データ予測装置、データ予測方法、及びデータ予測プログラム

技術分野

[0001] 本開示は、データ予測装置、データ予測方法、及びデータ予測プログラムに関する。

背景技術

[0002] 各所に設置した大量のセンサから取得されるデータから、その中の特定のデータの近い将来の値を予測する問題では、(1)式に例示するように、図1に示す線形回帰による予測が行われることがある。下記(1)式は、図1に例示する1～4の4つのセンサの時刻 $t-3$ から時刻 t までの単位時間毎のデータのうち時刻 $t-2$ から時刻 t までの12個のデータからセンサ1の時刻 $t+1$ におけるデータ $D_{1, t+1}$ を予測する。

$$\begin{aligned} & \beta_0 + \beta_1 D_{1, t-2} + \beta_2 D_{2, t-2} + \beta_3 D_{3, t-2} + \beta_4 D_{4, t-2} + \\ & \beta_5 D_{1, t-1} + \beta_6 D_{2, t-1} + \beta_7 D_{3, t-1} + \beta_8 D_{4, t-1} + \\ & \beta_9 D_{1, t} + \beta_{10} D_{2, t} + \beta_{11} D_{3, t} + \beta_{12} D_{4, t} = D_{1, t+1} \quad \cdots (1) \end{aligned}$$

[0003] β_k ($k=0, \dots, 12$) は回帰係数であり、 $D_{l, t-m}$ は、図1に円で例示するデータであり、 l はセンサの番号、 $t-m$ は時刻を表し、 m は自然数を表す。

[0004] 詳細には、以下の学習過程及び予測過程によりデータの予測が行われる。
まず、予測するデータに関係しそうな各センサの過去のデータを収集して線形回帰式を作成し、予測しようとする現象（予測対象の特異な動き）が生じているサンプル区間のデータを線形回帰式の選択変数に代入して、回帰係数 β_k を連立方程式により求める(学習過程)。次に、求められた回帰係数 β_k により、センサのデータから連続的に予測値を算出する(予測過程)。

[0005] 非常に多量のセンサからの過去の長い期間のデータを用いる場合には、サンプル区間のデータが不足し、連立方程式の解が求まらなくなる場合がある

。例えば、河川の水位の予想を行う場合、センサからのデータの代わり、あるいは、センサからのデータに追加して、例えば、気象庁で作成されるメッシュ状の雨量の予報値を用いることができる。しかしながら、この場合はさらに回帰係数の数が増えることになり、サンプル区間のデータが不足する場合がある。

[0006] サンプル区間のデータが不足した場合、L1正則化という方法（非特許文献1）を用いて、（2）式に例示するように、近似解の回帰係数を求めることができる。このL1正則化には、予測に貢献しにくい回帰係数をゼロに近づける性質がある。

[数1]

$$\hat{\beta} = \arg \min \sum_{i=1}^N \left(y_i - \sum_j \beta_j x_{ij} \right)^2 \quad \dots (2)$$

$$\sum_j |\beta_j| \leq t_c$$

y_i は目的変数であり、 x_{ij} は説明変数であり、 t_c は調整パラメータである。

[0007] L1正則化を用いて構成される、関連技術のデータ予測装置10を図2に示す。データ収集部11は多数のセンサの各々からデータを収集し、予測部12及び学習区間選択部13に同じデータを送信する。学習区間選択部13では、予測しようとする事象が現れていそうなデータ区間内のデータをルールベースで選択し、選択したデータを学習すべきデータとして回帰係数算出部14に送信する。回帰係数算出部14では、L1正則化を用いて回帰係数を求め、当該回帰係数を予測部12に送信する。予測部12は、回帰係数算出部14から受信した回帰係数と、データ収集部11から受信したデータ

と、を線形回帰式に代入して連続的に予測値を算出する。

先行技術文献

非特許文献

- [0008] 非特許文献1: Robert Tibshirani, "Regression Shrinkage and Selection via the Lasso", Journal of the Royal Statistical Society, Series B (Methodological), 1996, Vol. 58, No. 1, pp.267-288

発明の概要

発明が解決しようとする課題

- [0009] L_1 正則化により、多くの説明変数を使用する場合にも回帰係数を求めることはできるようになるが、予測に影響を及ぼすセンサの範囲、及びデータ計測時間の範囲は事前に既知でないため、広めに範囲を設定することになる。そのため、入力として使用するデータの数が非常に多くなり、データを収集及び伝送する回線容量を圧迫し、また、学習過程における計算処理量も膨大になる、という問題が発生する。
- [0010] 本開示では、回帰係数を求める際に使用するデータの数やデータを収集及び伝送する回線容量を低減し、学習過程における計算処理量を低減する、ことを目的とする。

課題を解決するための手段

- [0011] 本開示の第1態様のデータ予測装置は、収集した複数のデータ、及び受信した選択情報に基づいて前記複数のデータから選択した選択データを送信するデータ収集部と、 L_1 正則化により回帰係数を求める第1算出法を用いて、前記データ収集部から受信した前記複数のデータに基づいて回帰係数を算出し、算出した回帰係数のうち絶対値が閾値以上の回帰係数に対応するデータを選択するための選択情報を前記データ収集部に送信する第1回帰係数算出部と、回帰係数を算出する際に、前記第1算出法、及び前記第1算出法と異なる第2算出法のいずれの算出法を用いるかを判定する判定部と、前記判定部で判定された算出法を用いて回帰係数を算出する第2回帰係数算出部と

、前記判定部で判定された算出法が前記第 1 算出法である場合、前記データ収集部から受信した前記複数のデータを使用し、前記判定部で判定された算出法が前記第 2 算出法である場合、前記データ収集部から受信した前記選択データを使用し、使用したデータ及び前記第 2 回帰係数算出部で算出された回帰係数に基づいて予測した予測結果を出力する予測部と、を含む。

[0012] 本開示の第 2 態様は、第 1 態様のデータ予測装置であって、前記予測結果と実測値との誤差であって所定値以上の誤差の発生回数を監視する誤差監視部をさらに含み、前記判定部は、前記第 2 算出法を用いると判定した状態で、前記誤差監視部で監視された前記誤差の発生回数が所定回数以上になった場合は、前記第 1 算出法を用いると判定する。

[0013] 本開示の第 3 態様は、第 1 態様のデータ予測装置であって、予め設定された事象が発生したか否かを判定する事象判定部をさらに含み、前記判定部は、前記第 2 算出法を用いると判定した状態で、前記事象判定部で前記事象が発生したことが判定された場合、前記第 1 算出法を用いると判定する。

[0014] 本開示の第 4 態様は、第 1 態様～第 3 態様のいずれかのデータ予測装置であって、前記予測結果と実測値との誤差であって所定値以上の誤差の発生回数を監視する誤差監視部をさらに含み、前記判定部は、前記誤差監視部で監視された前記誤差の発生回数が所定回数以上になった場合は、前記第 1 回帰係数算出部に対して、前記第 1 算出法による回帰係数の再算出の指示、及び前記データ収集部に対する選択情報の再送信の指示を行い、前記第 1 回帰係数算出部は、前記判定部からの指示に応じて回帰係数の再算出及び選択情報の再送信を行い、前記データ収集部は、前記判定部からの指示に応じて再送信された選択情報に応じて、前記第 1 回帰係数算出部に対する選択データの再送信を行う。

[0015] 本開示の第 5 態様はプログラムであって、収集した複数のデータ、及び受信した選択情報に基づいて前記複数のデータから選択した選択データを送信し、 L_1 正則化により回帰係数を求める第 1 算出法を用いて、受信した前記複数のデータに基づいて回帰係数を算出し、算出した回帰係数のうち絶対値

が閾値以上の回帰係数に対応するデータを選択するための選択情報を送信し、回帰係数を算出する際に、前記第1算出法、及び前記第1算出法と異なる第2算出法のいずれの算出法を用いるかを判定し、判定された算出法を用いて回帰係数を算出し、判定された算出法が前記第1算出法である場合、受信した前記複数のデータを使用し、判定された算出法が前記第2算出法である場合、受信した前記選択データを使用し、使用したデータ及び算出された回帰係数に基づいて予測した予測結果を出力する、データ予測処理をコンピュータに実行させる。

[0016] 本開示の第6態様は、データ予測方法であって、コンピュータが、収集した複数のデータ、及び受信した選択情報に基づいて前記複数のデータから選択した選択データを送信し、L1正則化により回帰係数を求める第1算出法を用いて、受信した前記複数のデータに基づいて回帰係数を算出し、算出した回帰係数のうち絶対値が閾値以上の回帰係数に対応するデータを選択するための選択情報を送信し、回帰係数を算出する際に、前記第1算出法、及び前記第1算出法と異なる第2算出法のいずれの算出法を用いるかを判定し、判定された算出法を用いて回帰係数を算出し、判定された算出法が前記第1算出法である場合、受信した前記複数のデータを使用し、判定された算出法が前記第2算出法である場合、受信した前記選択データを使用し、使用したデータ及び算出された回帰係数に基づいて予測した予測結果を出力する。

発明の効果

[0017] 本開示では、データの数を実減し、データを収集及び伝送する回線容量を実減し、学習過程における計算処理量を低減する、ことができる。

図面の簡単な説明

[0018] [図1]線形回帰式によるデータの予測を説明する模式図である。

[図2]関連技術のデータ予測装置を例示するブロック図である。

[図3]本実施形態のデータ予測装置の基本例を例示するブロック図である。

[図4]本実施形態のデータ予測装置の拡張例を例示するブロック図である。

[図5]本実施形態のデータ予測装置の拡張例のハードウェア構成図を例示する

ブロック図である。

[図6]本実施形態の拡張例の学習フェーズの処理の流れを例示する模式図である。

[図7]本実施形態の拡張例の予測フェーズの処理の流れを例示する模式図である。

[図8]本実施形態の拡張例の再算出及び再選択フェーズの処理の流れを例示する模式図である。

[図9]本実施形態のデータ予測装置を例示するブロック図である。

発明を実施するための形態

[0019] 本実施形態のデータ予測装置20を図3に例示する。データ収集部21は多数のセンサ S_1 , S_2 , S_3 , …の各々からデータを収集し、予測部22及び学習区間選択部23に同じデータを送信する。データ収集部21が、予測部22及び学習区間選択部23に送信するデータとしては、多数のセンサ S_1 , S_2 , S_3 , …から収集した複数のデータ、及び後述する第1回帰係数算出部25から受信した選択情報に基づいて収集した複数のデータから選択した選択データがある。

[0020] 学習区間選択部23では、予測しようとする事象が現れていることが予測されるデータ区間をルールベースで選択し、学習すべきデータを第1回帰係数算出部25及び第2回帰係数算出部24に送信する。すなわち、学習区間選択部23は、予測しようとする事象が現れていないことが予測される区間に存在する不要なデータ及び変化の乏しいデータを取り除く。

[0021] 第1回帰係数算出部25は、L1正則化により近似解の回帰係数を求める第1算出法を用いて、データ収集部21から送信され、かつ学習区間選択部23で選択された複数のデータに基づいて回帰係数を算出し、算出した回帰係数のうち絶対値が閾値以上の回帰係数を選択し、選択した回帰係数を示す情報を、データ収集部21及び第2回帰係数算出部24の各々に送信する。

[0022] データ収集部21では、第1回帰係数算出部25から受信した情報に基づいて、収集した複数のデータから絶対値が閾値以上の回帰係数に対応するデ

ータを選択し、選択したデータを選択データとして予測部 2 2、学習区間選択部 2 3、及び誤差監視部 2 6 の各々に送信する。第 1 回帰係数算出部 2 5 からデータ収集部 2 1 に送信される情報は、データ収集部 2 1 においてデータを選択するために使用されるので、以下では選択情報という。

[0023] なお、データ収集部 2 1 では、絶対値が閾値以上の回帰係数に対応するデータが選択データとして選択され、絶対値が閾値未満の回帰係数に対応するデータが削除される。このため、データ収集部 2 1 に送信する選択情報としては、絶対値が閾値以上の回帰係数を示す情報に代えて、絶対値が閾値未満の回帰係数を示す情報を用い、データ収集部 2 1 において、絶対値が閾値未満の回帰係数に対応するデータを削除することにより、選択データを得るようにしてもよい。

[0024] 第 2 回帰係数算出部 2 4 は、データ収集部 2 1 から送信され、かつ学習区間選択部 2 3 を通過した選択データを使用して、 L_1 正則化ではなく、 L_2 正則化または重回帰を用いて回帰係数を求める。

[0025] 第 2 回帰係数算出部 2 4 で算出された回帰係数は、 L_2 正則化または重回帰を用いた回帰計算による予測が行われる予測部 2 2 に送信される。データ収集部 2 1 では、選択データを予測部 2 2 及び学習区間選択部 2 3 に送信する。即ち、データ収集部 2 1 は、全てのセンサ S_1 , S_2 , S_3 , ... の各々からデータを収集するが、データの送信は、選択的に行う。

[0026] 予測部 2 2 では、データ収集部 2 1 から受信した選択データと、第 2 回帰係数算出部 2 4 から受信した回帰係数とを用いて、線形回帰式で予測した予測結果を出力する。

[0027] すなわち、予測結果は、絶対値が閾値未満の回帰係数に対応するデータが削除され、絶対値が閾値以上の回帰係数と絶対値が閾値以上の回帰係数に対応するデータとから算出されるため、学習過程で使用するデータ量が低減され、計算処理量を低減することができる。

[0028] 誤差監視部 2 6 では、予測部 2 2 による予測結果と、データ収集部 2 1 で収集されたデータである実測値との差である誤差を監視し、所定値以上の誤

差が所定回数(許容回数)以上発生した場合に、データ収集部 21 と第 1 回帰係数算出部 25 に、再選択指示を送信する。データ収集部 21 は、この再選択指示により、選択データの再選択が終わるまでの一定期間、選択データではなく、全てのデータを後段に送信するようになり、第 1 回帰係数算出部 25 では、再度、全てのデータを対象にした L1 正則化によるデータの選択が行われる。この際、L2 正則化または重回帰による予測結果の誤差が、L1 正則化による予測結果の誤差よりも大きい場合にのみ、選択データの再選択を行い、L2 正則化または重回帰による予測結果の誤差が、L1 正則化による予測結果の誤差以下である場合には、再選択を行わない。

[0029] なお、誤差監視部 26 では、予測部 22 による予測結果と、実測値の差である誤差を監視し、所定値以上の誤差が所定回数(許容回数)以上発生した場合に、第 2 回帰係数算出部 24 で、回帰係数の再算出を行うようにしてもよい。回帰係数の再算出後、所定値以上の誤差が所定回数(許容回数)以上発生した場合に、データ収集部 21 と第 1 回帰係数算出部 25 に、再選択指示を送信するようにしてもよい。

[0030] 本実施形態のデータ予測装置 20 の拡張例であるデータ予測装置 60 を図 4 に例示する。データ予測装置 20 と同様の構成及び作用については、説明を適宜省略する。図 4 のデータ予測装置 60 は、データの傾向などに応じて、回帰係数の算出の際に使用する回帰係数算出法を選択する判定部 67 を有する点で、図 3 のデータ予測装置 20 と異なる。

[0031] データ収集部 61 は多数のセンサ S_1 , S_2 , S_3 , ... からデータを収集し、予測部 62 及び学習区間選択部 63 に同じデータを送信する。学習区間選択部 63 では、予測しようとする事象が現れていそうなデータ区間をルールベースで選択し、学習すべきデータを回帰係数算出部 64 に送信する。

[0032] データ予測装置 20 では、第 2 回帰係数算出部 24 で回帰係数を算出する際に、例えば、L2 正則化または重回帰を使用しているが、データ予測装置 60 では、第 2 回帰係数算出部 64 で回帰係数を算出する際に、判定部 67 の判定結果に応じて、使用する回帰係数算出法を選択する。使用する回帰係

数算出法は、例えば、以下の２つから選択される。

- １）選択データを使用する重回帰またはＬ２正則化（第２算出法）
- ２）全てのデータを使用するＬ１正則化（第１算出法）

[0033] 誤差監視部６６において、所定値以上の予測誤差（誤差）が所定回数（許容回数）以上発生したことが検出されるまでは、判定部６７は、１）の回帰係数算出法を選択することを判定し、第２回帰係数算出部６４は、選択データを使用して、１）の回帰係数算出法で回帰係数を算出する。誤差監視部６６において、所定値以上の予測誤差が所定回数（許容回数）以上発生したことが検出された場合、判定部６７は、回帰係数算出法を１）から２）に切り替えることを判定する。

[0034] 第２回帰係数算出部６４は、判定部６７の判定に基づいて、２）の回帰係数算出法で全てのデータを使用して回帰係数を算出する。これにより、データに応じて、より誤差の少ない回帰係数算出法を選択することができる。

[0035] なお、学習区間選択部６３において、事前に設定された事象（例えば、降雨など）が発生したか否かをルールベースで判定する事象判定部をさらに含み、事象判定部が学習すべきデータが更新されたことを判定した場合にも、判定部６７は、２）の回帰係数算出法を選択することを判定してもよい。これにより、学習すべきデータが更新された場合においても、より誤差の少ない回帰係数算出法を選択することができる。

[0036] 図５に、データ予測装置６０のハードウェア構成を例示する。データ予測装置６０は、一例として、図５に示すように、ＣＰＵ（Central Processing Unit）５１、一次記憶部５２、二次記憶部５３、及び、外部インタフェース５４を含む。ＣＰＵ５１は、ハードウェアであるプロセッサの一例である。ＣＰＵ５１、一次記憶部５２、二次記憶部５３、及び、外部インタフェース５４は、バス５９を介して相互に接続されている。

[0037] 一次記憶部５２は、例えば、ＲＡＭ（Random Access Memory）などの揮発性のメモリである。二次記憶部５３は、例えば、ＨＤＤ（Hard Disk Drive）、又はＳＳＤ（Solid State Drive）などの不揮発性のメモリである。

- [0038] 二次記憶部 5 3 は、プログラム格納領域 5 3 A 及びデータ格納領域 5 3 B を含む。プログラム格納領域 5 3 A は、一例として、データ予測プログラムなどのプログラムを記憶している。データ格納領域 5 3 B は、一例として、センサからのデータ及びデータ予測処理中の中間データなどを記憶する。
- [0039] CPU 5 1 は、プログラム格納領域 5 3 A からデータ予測プログラムを読み出して一次記憶部 5 2 に展開する。CPU 5 1 は、データ予測プログラムをロードして実行することで、図 4 のデータ収集部 6 1、予測部 6 2、学習区間選択部 6 3、第 2 回帰係数算出部 6 4、第 1 回帰係数算出部 6 5、誤差監視部 6 6、及び判定部 6 7 として動作する。
- [0040] なお、データ予測プログラムなどのプログラムは、外部サーバに記憶され、ネットワークを介して、一次記憶部 5 2 に展開されてもよい。また、データ予測プログラムなどのプログラムは、Digital Versatile Disc (DVD) などの非一時的記録媒体に記憶され、記録媒体読込装置を介して、一次記憶部 5 2 に展開されてもよい。
- [0041] 外部インタフェース 5 4 には外部装置が接続され、外部インタフェース 5 4 は、外部装置と CPU 5 1 との間の各種情報の送受信を司る。図 5 では、外部インタフェース 5 4 に、センサ 3 1 A、及び危険報知システム 3 1 B が接続されている例を示している。なお、センサ 3 1 A は、多数のセンサを含む。
- [0042] データ予測装置 6 0 で予測されたデータは、例えば、外部インタフェース 5 4 に接続される危険報知システム 3 1 B に送信され、危険報知システム 3 1 B における危険報知処理に使用されてもよい。また、データ予測装置 1 0 で予測されたデータは、例えば、外部インタフェース 5 4 に接続される外部記憶装置に記録されてもよいし、外部インタフェース 5 4 に接続されるディスプレイの画面に文字または画像として表示されてもよい。
- [0043] また、データ予測装置 6 0 は、専用装置であってもよいし、ワークステーション、パーソナルコンピュータ、またはタブレットなどの汎用装置であってもよい。

[0044] 図6～図8に、データ予測装置60におけるデータ予測処理を例示する。

図6は、学習フェーズの流れを例示する。

[0045] データ収集部61は、手順S201で、例えば、多数のセンサを含むセンサ31Aからデータを収集する。データ収集部61は、手順S202で、学習区間選択部63に、収集したデータを送信する。

[0046] 学習区間選択部63は、手順S203で、予測に適切そうなデータ区間を選択して第2回帰係数算出部64と第1回帰係数算出部65に送信する。第1回帰係数算出部65は、L1正則化を用いて回帰係数を算出し、絶対値が閾値以上の回帰係数を選択して、第2回帰係数算出部64とデータ収集部61に送信する。

[0047] 第2回帰係数算出部64は、手順S205で、第1回帰係数算出部65で選択されたデータを使用して、通常的回帰算出法、例えば、L2正則化または重回帰により、回帰係数を求めて予測部62に送信する。

[0048] 図7は、予測フェーズの流れを例示する。データ収集部61は、手順S206で、手順S204で選択されたデータを予測部62及び誤差監視部66に送信する。予測部62は、手順S207で、手順S205で求めた回帰係数とデータとに基づいて、予測結果を算出し、誤差監視部66に送信すると共に、例えば、外部記憶装置に記憶する。

[0049] 図8は、誤差監視及び再算出及び再選択フェーズの流れを例示する。誤差監視部66は、手順S208で、予測結果とデータ（実測値）との予測誤差を算出し、所定値以上の予測誤差が所定回数（許容回数）以上発生した場合、データ収集部61及び判定部67に、回帰係数算出法の切り替えを指示する。

[0050] 回帰係数算出法の切り替えが指示されると、データ収集部61は、全てのデータを学習区間選択部63に送信し、判定部67は、第2回帰係数算出部64で、全てのデータを使用して、L1正則化により回帰係数を算出することを判定し、手順S209で、当該判定結果を第2回帰係数算出部64に送信する。第2回帰係数算出部64は、判定部67の判定を受信すると、回帰

係数の算出法を切り替える。

- [0051] 誤差監視部 66 は、手順 S 210 で、予測結果とデータ（実測値）との予測誤差を算出し、所定値以上の予測誤差が所定回数（許容回数）以上発生した場合、データ収集部 61 及び判定部 67 に、再選択を指示する。
- [0052] 再選択が指示されると、データ収集部 61 は、全てのデータを学習区間選択部 63 に送信し、判定部 67 は、再選択を行うことを判定し、手順 S 211 で、当該判定結果を第 1 回帰係数算出部 65 に送信する。第 1 回帰係数算出部 65 は、判定部 67 の判定を受信すると、回帰係数の再選択を行う。
- [0053] 本実施形態によれば、L 1 正則化で算出した回帰係数だけを予測に用いる場合より、予測誤差が大きくなる傾向を示す場合があるが、予測誤差が大きい場合、あるいは、定期的に、データの再選択を行うことにより、予測誤差をある範囲に収めることが可能である。
- [0054] なお、回帰係数算出法を、第 1 回帰係数算出部 65 で選択済みのデータを使用する L 2 正則化または重回帰から全てのデータを使用する L 1 正則化に切り替える例について説明したが、本実施形態はこれに限定されない。回帰係数算出法を、全てのデータを使用する L 1 正則化から、第 1 回帰係数算出部 65 で選択済みのデータを使用する L 2 正則化または重回帰に切り替えるようにしてもよい。
- [0055] なお、データ予測装置 60 は、図 9 に例示するように、データ収集装置 32A、学習装置 32B 及び予測装置 32C を含んでもよい。データ収集装置 32A は、データ収集部 61 を含み、学習装置 32B は、学習区間選択部 63、第 2 回帰係数算出部 64、第 1 回帰係数算出部 65、誤差監視部 66、及び判定部 67 を含み、予測装置 32C は予測部 62 を含む。
- [0056] センサ 31A とデータ収集部 61 との間、及び、予測部 62 と、例えば、危険報知システム 31B などの予測結果の出力先との間は、伝送回線で接続されている。また、図 9 に示される例では、データ収集部 61 と学習区間選択部 63 との間、第 2 回帰係数算出部 64 と予測部 62 との間も伝送回線で接続されている。

[0057] 本開示のデータ予測装置は、収集した複数のデータ、及び受信した選択情報に基づいて前記複数のデータから選択した選択データを送信するデータ収集部と、 L_1 正則化により回帰係数を求める第1算出法を用いて、前記データ収集部から受信した前記複数のデータに基づいて回帰係数を算出し、算出した回帰係数のうち絶対値が閾値以上の回帰係数に対応するデータを選択するための選択情報を前記データ収集部に送信する第1回帰係数算出部と、回帰係数を算出する際に、前記第1算出法、及び前記第1算出法と異なる第2算出法のいずれの算出法を用いるかを判定する判定部と、前記判定部で判定された算出法を用いて回帰係数を算出する第2回帰係数算出部と、前記判定部で判定された算出法が前記第1算出法である場合、前記データ収集部から受信した前記複数のデータを使用し、前記判定部で判定された算出法が前記第2算出法である場合、前記データ収集部から受信した前記選択データを使用し、使用したデータ及び前記第2回帰係数算出部で算出された回帰係数に基づいて予測した予測結果を出力する予測部と、を含む。

[0058] 本開示では、これにより、データの数を実減し、データを収集及び伝送する回線容量を実減し、学習過程における計算処理量を低減する、ことができる。なお、データ収集部61から後段に送信するデータの実数を低減する代わりに、データ収集部61がデータを受信するセンサの実数を低減してもよい。

[0059] 本開示では、第2回帰係数算出部65での算出に、通常の実帰係数算出法を使用する場合があるため、算出コストを実減することができ、算出リソースの少ない環境における処理を実可能にする。

符号の説明

- [0060] 61 データ収集部
62 予測部
66 誤差監視部
64 第2回帰係数算出部
65 第1回帰係数算出部
31A センサ

5 1 C P U

5 2 一次記憶部

5 3 二次記憶部

請求の範囲

【請求項1】 収集した複数のデータ、及び受信した選択情報に基づいて前記複数の

のデータから選択した選択データを送信するデータ収集部と、

 L 1 正則化により回帰係数を求める第1 算出法を用いて、前記データ収集部から受信した前記複数のデータに基づいて回帰係数を算出し、算出した回帰係数のうち絶対値が閾値以上の回帰係数に対応するデータを選択するための選択情報を前記データ収集部に送信する第1 回帰係数算出部と、

 回帰係数を算出する際に、前記第1 算出法、及び前記第1 算出法と異なる第2 算出法のいずれの算出法を用いるかを判定する判定部と、

 前記判定部で判定された算出法を用いて回帰係数を算出する第2 回帰係数算出部と、

 前記判定部で判定された算出法が前記第1 算出法である場合、前記データ収集部から受信した前記複数のデータを使用し、前記判定部で判定された算出法が前記第2 算出法である場合、前記データ収集部から受信した前記選択データを使用し、使用したデータ及び前記第2 回帰係数算出部で算出された回帰係数に基づいて予測した予測結果を出力する予測部と、

 を含む、データ予測装置。

【請求項2】 前記予測結果と実測値との誤差であって所定値以上の誤差の発生回数を監視する誤差監視部をさらに含み、

 前記判定部は、前記第2 算出法を用いると判定した状態で、前記誤差監視部で監視された前記誤差の発生回数が所定回数以上になった場合は、前記第1 算出法を用いると判定する、

 請求項1 に記載のデータ予測装置。

【請求項3】 予め設定された事象が発生したか否かを判定する事象判定部をさらに含み、

 前記判定部は、前記第2 算出法を用いると判定した状態で、前記事

象判定部で前記事象が発生したことが判定された場合、前記第 1 算出法を用いると判定する、

請求項 1 に記載のデータ予測装置。

[請求項 4]

前記予測結果と実測値との誤差であって所定値以上の誤差の発生回数を監視する誤差監視部をさらに含み、

前記判定部は、前記誤差監視部で監視された前記誤差の発生回数が所定回数以上になった場合は、前記第 1 回帰係数算出部に対して、前記第 1 算出法による回帰係数の再算出の指示、及び前記データ収集部に対する選択情報の再送信の指示を行い、

前記第 1 回帰係数算出部は、前記判定部からの指示に応じて回帰係数の再算出及び選択情報の再送信を行い、

前記データ収集部は、前記判定部からの指示に応じて再送信された選択情報に応じて、前記第 1 回帰係数算出部に対する選択データの再送信を行う、

請求項 1 ～請求項 3 のいずれか 1 項に記載のデータ予測装置。

[請求項 5]

収集した複数のデータ、及び受信した選択情報に基づいて前記複数のデータから選択した選択データを送信し、

L 1 正則化により回帰係数を求める第 1 算出法を用いて、受信した前記複数のデータに基づいて回帰係数を算出し、算出した回帰係数のうち絶対値が閾値以上の回帰係数に対応するデータを選択するための選択情報を送信し、

回帰係数を算出する際に、前記第 1 算出法、及び前記第 1 算出法と異なる第 2 算出法のいずれの算出法を用いるかを判定し、

判定された算出法を用いて回帰係数を算出し、

判定された算出法が前記第 1 算出法である場合、受信した前記複数のデータを使用し、判定された算出法が前記第 2 算出法である場合、受信した前記選択データを使用し、使用したデータ及び算出された回帰係数に基づいて予測した予測結果を出力する、

[請求項6]

データ予測処理をコンピュータに実行させるためのプログラム。

コンピュータが、

収集した複数のデータ、及び受信した選択情報に基づいて前記複数のデータから選択した選択データを送信し、

 L 1 正則化により回帰係数を求める第 1 算出法を用いて、受信した前記複数のデータに基づいて回帰係数を算出し、算出した回帰係数のうち絶対値が閾値以上の回帰係数に対応するデータを選択するための選択情報を送信し、

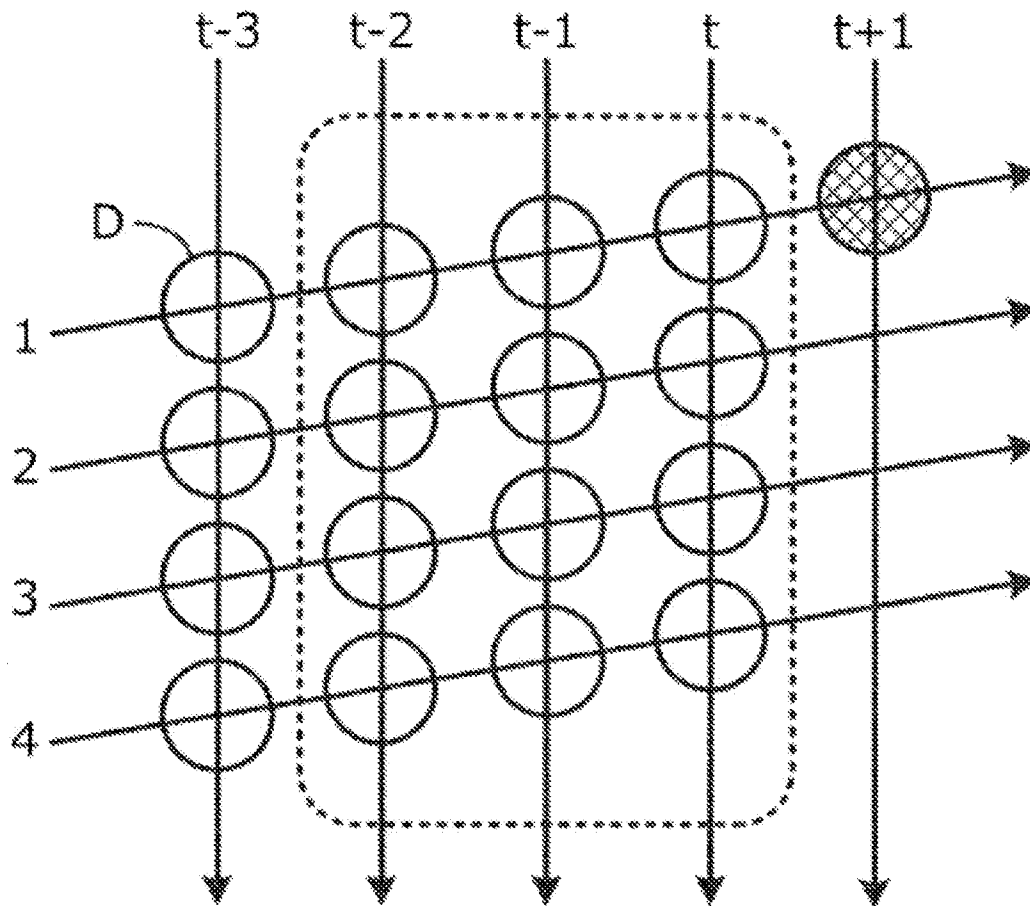
 回帰係数を算出する際に、前記第 1 算出法、及び前記第 1 算出法と異なる第 2 算出法のいずれの算出法を用いるかを判定し、

 判定された算出法を用いて回帰係数を算出し、

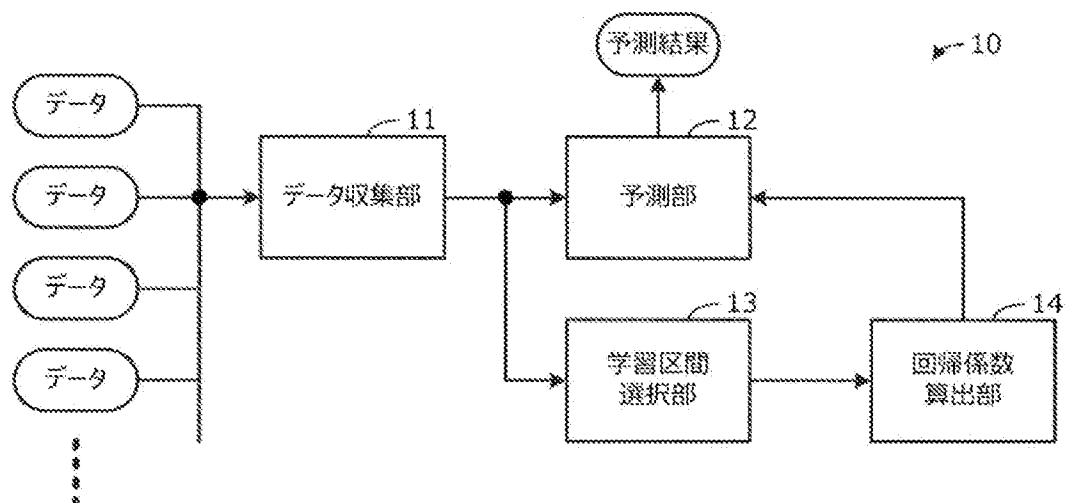
 判定された算出法が前記第 1 算出法である場合、受信した前記複数のデータを使用し、判定された算出法が前記第 2 算出法である場合、受信した前記選択データを使用し、使用したデータ及び算出された回帰係数に基づいて予測した予測結果を出力する、

データ予測方法。

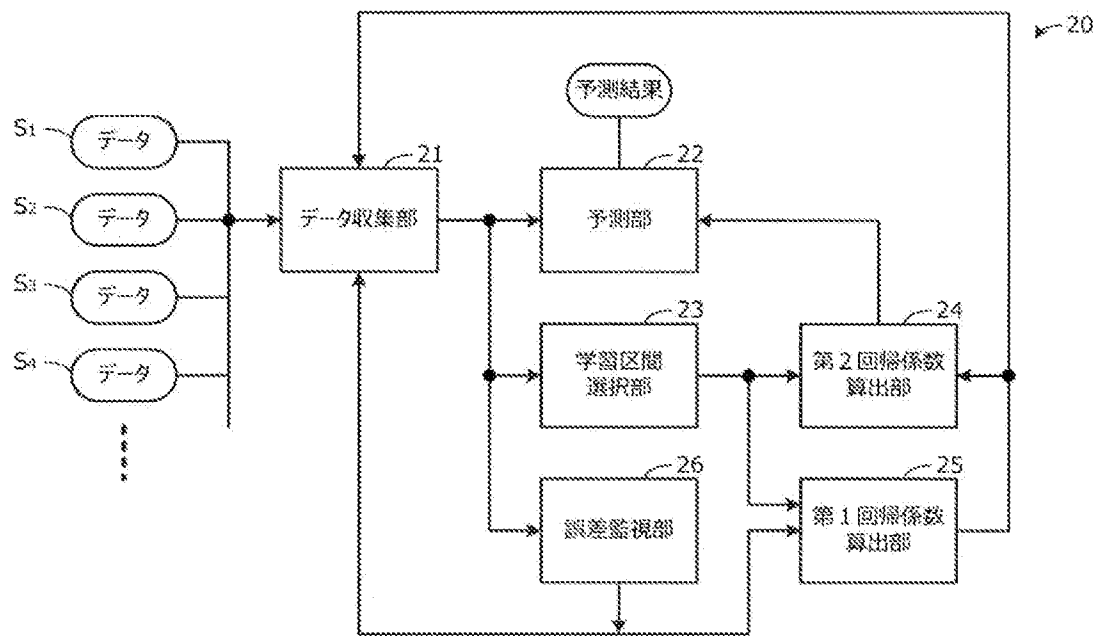
[図1]



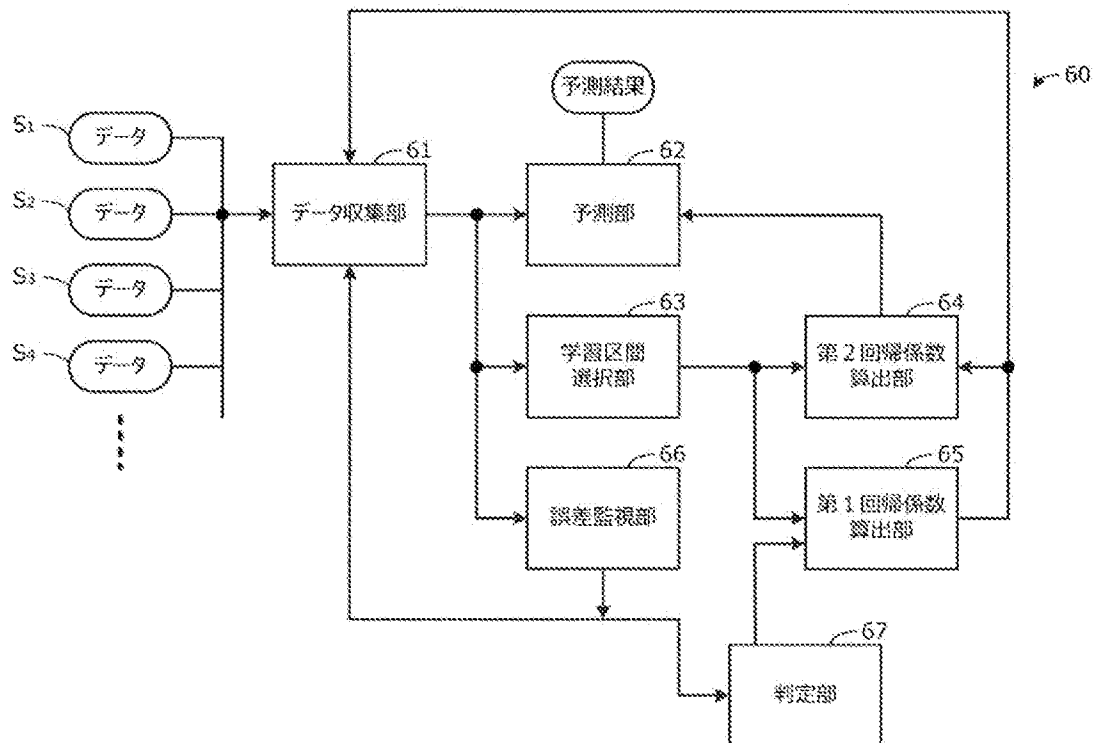
[図2]



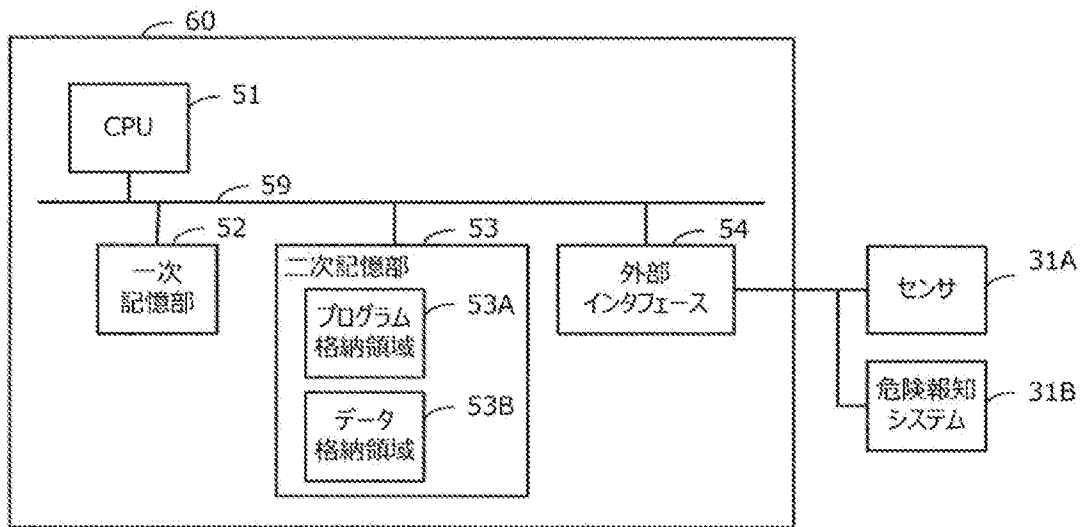
[図3]



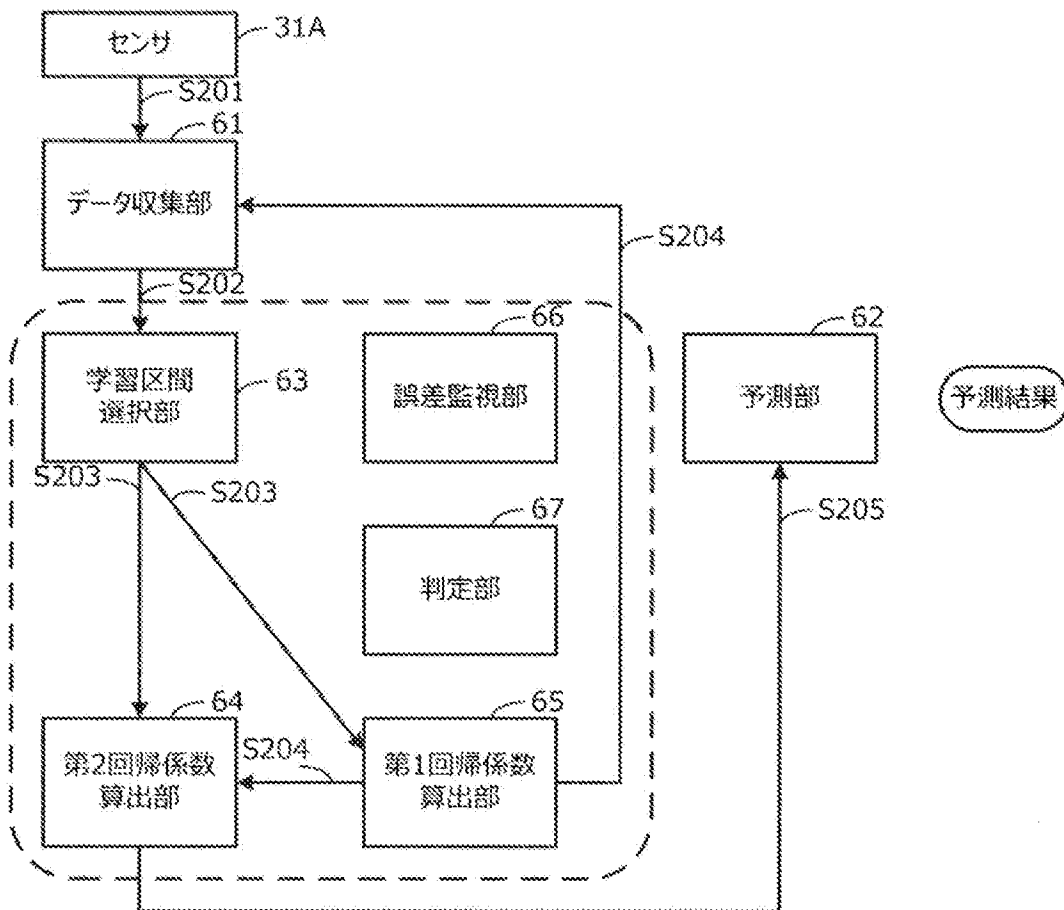
[図4]



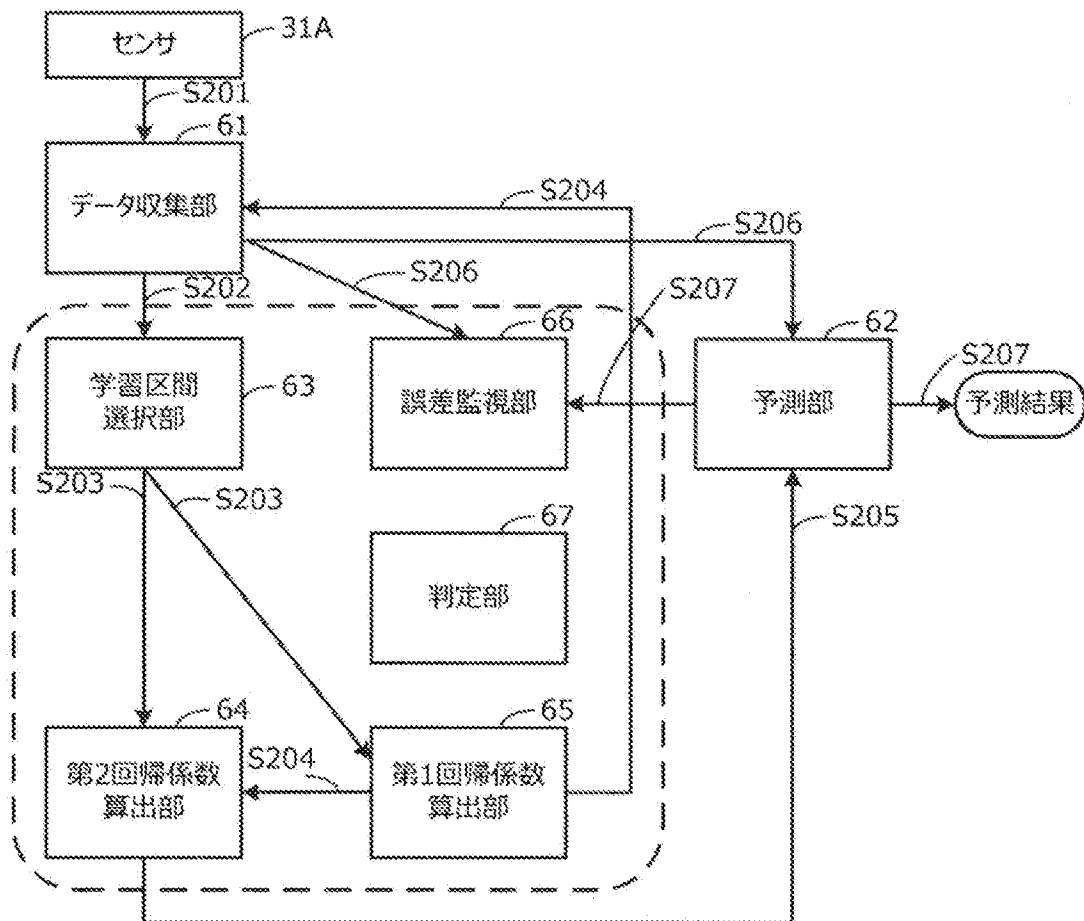
[図5]



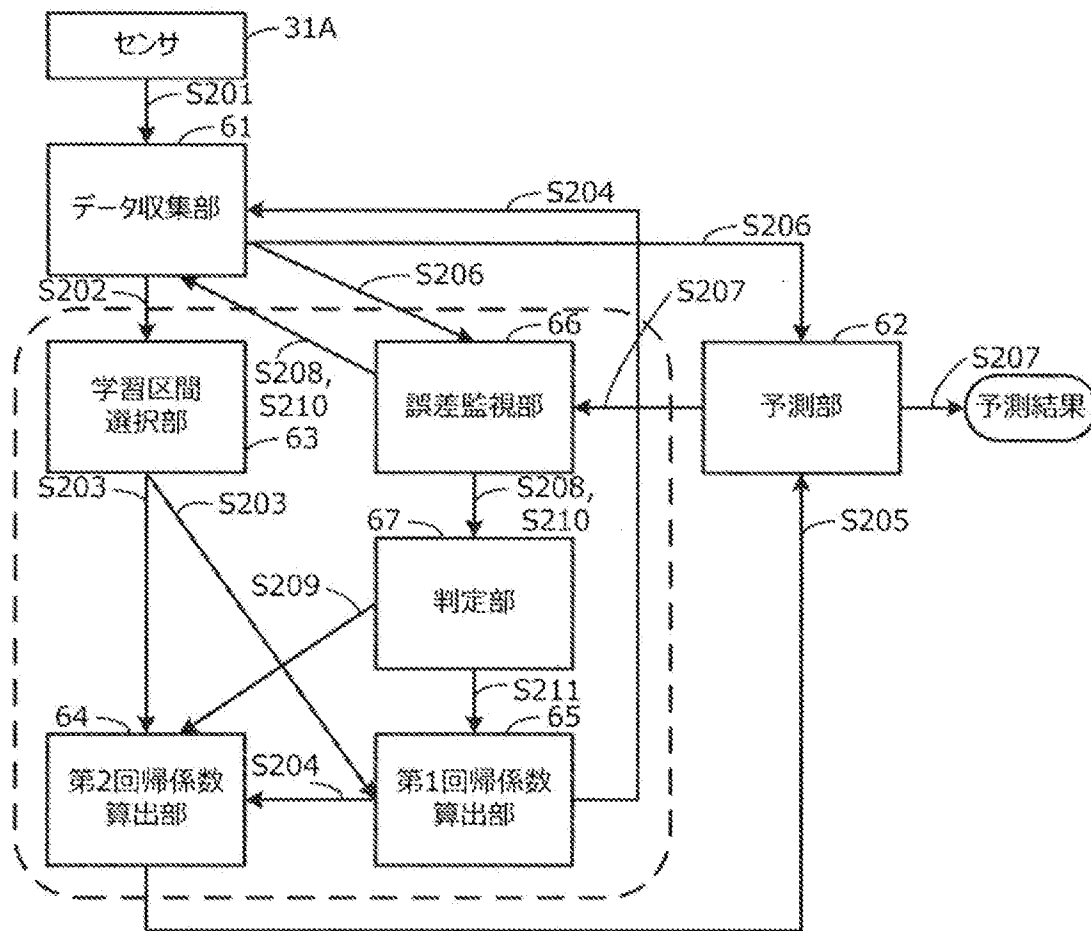
[図6]



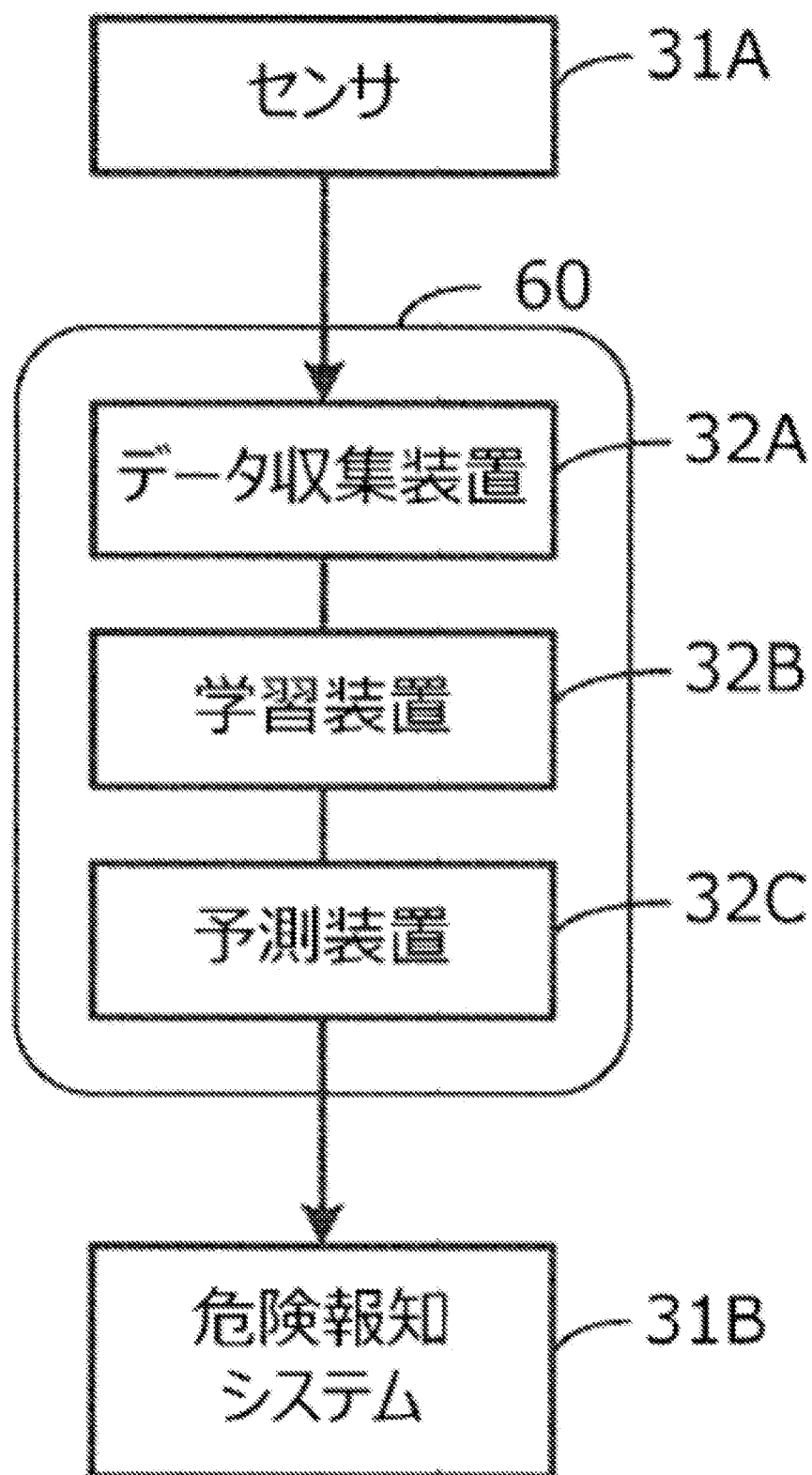
[図7]



[図8]



[図9]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2020/010303

A. CLASSIFICATION OF SUBJECT MATTER

G06N 20/00 (2019.01) i

FI: G06N20/00 130

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06N20/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan	1922-1996
Published unexamined utility model applications of Japan	1971-2020
Registered utility model specifications of Japan	1996-2020
Published registered utility model applications of Japan	1994-2020

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2019-32185 A (TOSHIBA CORP.) 28.02.2019 (2019-02-28) paragraphs [0009]-[0100], fig. 1-11	1-6
A	JP 2018-151883 A (TOSHIBA CORP.) 27.09.2018 (2018-09-27) paragraph [0020]	1-6
A	JP 2018-109876 A (TOSHIBA CORP.) 12.07.2018 (2018-07-12) paragraphs [0068]-[0080]	1-6
A	JP 2018-10477 A (FUJITSU LTD.) 18.01.2018 (2018-01-18) entire text, all drawings	1-6
A	JP 2017-130041 A (FUJITSU LTD.) 27.07.2017 (2017-07-27) entire text, all drawings	1-6



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"I" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

28 April 2020 (28.04.2020)

Date of mailing of the international search report

19 May 2020 (19.05.2020)

Name and mailing address of the ISA/

Japan Patent Office

3-4-3, Kasumigaseki, Chiyoda-ku,

Tokyo 100-8915, Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/JP2020/010303

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
JP 2019-32185 A	28 Feb. 2019	US 2019/0041235 A1 paragraphs [0023]- [0116], fig. 1-11	
JP 2018-151883 A	27 Sep. 2018	US 2018/0260726 A1 paragraph [0036]	
JP 2018-109876 A	12 Jul. 2018	KR 10-2018-0104542 A US 2018/0189242 A1 paragraphs [0090]- [0102]	
JP 2018-10477 A	18 Jan. 2018	EP 3346428 A1 (Family: none)	
JP 2017-130041 A	27 Jul. 2017	(Family: none)	

A. 発明の属する分野の分類（国際特許分類（IPC）） G06N 20/00(2019.01)i FI: G06N20/00 130														
B. 調査を行った分野 調査を行った最小限資料（国際特許分類（IPC）） G06N20/00 最小限資料以外の資料で調査を行った分野に含まれるもの <table border="0"> <tr> <td>日本国実用新案公報</td> <td>1922 - 1996年</td> </tr> <tr> <td>日本国公開実用新案公報</td> <td>1971 - 2020年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996 - 2020年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994 - 2020年</td> </tr> </table> 国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）			日本国実用新案公報	1922 - 1996年	日本国公開実用新案公報	1971 - 2020年	日本国実用新案登録公報	1996 - 2020年	日本国登録実用新案公報	1994 - 2020年				
日本国実用新案公報	1922 - 1996年													
日本国公開実用新案公報	1971 - 2020年													
日本国実用新案登録公報	1996 - 2020年													
日本国登録実用新案公報	1994 - 2020年													
C. 関連すると認められる文献														
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号												
A	JP 2019-32185 A（株式会社東芝）28.02.2019（2019 - 02 - 28） [0009]-[0100], 図1-11	1-6												
A	JP 2018-151883 A（株式会社東芝）27.09.2018（2018 - 09 - 27） [0020]	1-6												
A	JP 2018-109876 A（株式会社東芝）12.07.2018（2018 - 07 - 12） [0068]-[0080]	1-6												
A	JP 2018-10477 A（富士通株式会社）18.01.2018（2018 - 01 - 18） 全文、全図	1-6												
A	JP 2017-130041 A（富士通株式会社）27.07.2017（2017 - 07 - 27） 全文、全図	1-6												
☐ C欄の続きにも文献が列挙されている。 ☒ パテントファミリーに関する別紙を参照。														
<table border="0"> <tr> <td>* 引用文献のカテゴリー</td> <td>“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの</td> </tr> <tr> <td>“A” 特に関連のある文献ではなく、一般的技術水準を示すもの</td> <td>“X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの</td> </tr> <tr> <td>“E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの</td> <td>“Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの</td> </tr> <tr> <td>“L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）</td> <td>“&” 同一パテントファミリー文献</td> </tr> <tr> <td>“O” 口頭による開示、使用、展示等に言及する文献</td> <td></td> </tr> <tr> <td>“P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献</td> <td></td> </tr> </table>			* 引用文献のカテゴリー	“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの	“A” 特に関連のある文献ではなく、一般的技術水準を示すもの	“X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの	“E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの	“Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの	“L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）	“&” 同一パテントファミリー文献	“O” 口頭による開示、使用、展示等に言及する文献		“P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献	
* 引用文献のカテゴリー	“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの													
“A” 特に関連のある文献ではなく、一般的技術水準を示すもの	“X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの													
“E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの	“Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの													
“L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）	“&” 同一パテントファミリー文献													
“O” 口頭による開示、使用、展示等に言及する文献														
“P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献														
国際調査を完了した日 28.04.2020		国際調査報告の発送日 19.05.2020												
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号		権限のある職員（特許庁審査官） 山本 俊介 5B 5087 電話番号 03-3581-1101 内線 3545												

国際調査報告
 パテントファミリーに関する情報

国際出願番号

PCT/JP2020/010303

引用文献			公表日	パテントファミリー文献	公表日
JP	2019-32185	A	28.02.2019	US 2019/0041235 A1 [0023]-[0116], FIGs. 1-11	
JP	2018-151883	A	27.09.2018	US 2018/0260726 A1 [0036] KR 10-2018-0104542 A	
JP	2018-109876	A	12.07.2018	US 2018/0189242 A1 [0090]-[0102] EP 3346428 A1	
JP	2018-10477	A	18.01.2018	(ファミリーなし)	
JP	2017-130041	A	27.07.2017	(ファミリーなし)	