

(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11月19日 (19.11.2020)



W I P O I P C T

図 1

関 1111 画 1

(10) 国際公開番号

¥0 2020 /230324 入 1

(51) 国際特許分類:

G06N 20/00 (2019.01)

(21) 国際出願番号:

PCT/JP 19/019547

(22) 国際出願日:

2019年5月16日 (16.05.2019)

(25) 国際出願の言語:

日本語

(26) 国際公開の言語:

日本語

(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: 外間 正浩 (SOTOMA Masahiro); 〒1808585 東京都武蔵野市緑町3丁目9-1 NTT 知的財産センター内 Tokyo (TP).

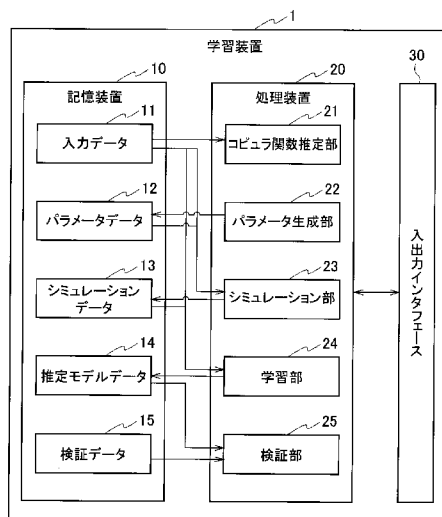
津田 昌幸 (TSUDA Masayuki); 〒1808585 東京都武蔵野市緑町3丁目9-1 NTT 知的財産センター内丁〇砂〇 (見).

(74) 代理人: 三好 秀和, 外 (H0 Y 08111 Hidekazu et al.); 〒1050001 東京都港区虎ノ門1丁目2番8号 虎ノ門琴平タワー丁〇砂〇 (見).

(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, 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, JP, 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,

(54) Title: LEARNING DEVICE, LEARNING METHOD, AND LEARNING PROGRAM

(54) 発明の名称: 学習装置、学習方法および学習プログラム



1 Learning device
10 Storage device
11 Input data
12 Parameter data
13 Simulation data
14 Estimation model data
15 Verification data
20 Processing device
21 Copula function estimation unit
22 Parameter generating unit
23 Simulation unit
24 Learning unit
25 Verification unit
30 Input/output interface

(57) Abstract: A learning device 1 is provided with: a storage device 10 for storing input data 11 including a plurality of data sets relating to a first event and a plurality of data sets relating to a second event, wherein the number of the data sets relating to the second event is smaller than the number of the data sets relating to the first event; a copula function estimation unit 21 which estimates, from the data sets relating to the second event, a copula function and a parameter used by the copula function; a simulation unit 23 which generates data sets relating to the second event by simulation using

NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT,
QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA,
UG, US, UZ, VC, VN, 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, 0 K, £ ¼ £8^1, 戸良 03, 011, 1111, 1111, 1¼ IS, II ; 1.T ,L11, LV, 1^0, MK ,MT, 見 , No , 卩レ 卩丁, 110, 118, 8¾ 81, 8K, 81^ ,見) ,0 处 1 田 ろ 3 I , 〇尸 ,00, 〇 , 〇M , 。ん 。凡 GQ, GW, 疆 , 他 , M民 呢 SN, TD, 丁。)

添付公開書類 :

- 国際調査報告 (条約第21条 (3))

the copula mction and the parameter; and a learning unit 24 which, with reference to the input data 11 and the data sets relating to the second event generated by the simulation unit, learns an estimation model for distinguishing between the first event and the second event.

(57)要約 :学習装置 1 は、第 1 の事象に関する複数のデータセットと、第 2 の事象に関する複数のデータセットを含む入力データ 1 1 であって、第 2 の事象に関するデータセットの数は、第 1 の事象に関するデータセットの数よりも少ない入力データ 1 1 を記憶する記憶装置 1 0 と、第 2 の事象に関するデータセットから、コピュラ関数とコピュラ関数で用いられるパラメータを推定するコピュラ関数推定部 2 1 と、コピュラ関数およびパラメータによるシミュレーションにより、第 2 の事象に関するデータセットを生成するシミュレーション部 2 3 と、入力データ 1 1 と、シミュレーション部によって生成された第 2 の事象に関するデータセットを参照して、第 1 の事象と第 2 の事象を区別する推定モデルを学習する学習部 2 4 を備える。

明 細 書

発 明 の 名 称 : 学 習 装 置 、 学 習 方 法 お よ び 学 習 プ ロ グ ラ ム

技 術 分 野

[0001] 本発明は、複数のデータセットを参照して、機械学習を行う学習装置、学習方法および学習プログラムに関する。

背景技術

[0002] 一般的に、機械、デバイス等の各種設備の保守点検において、各種設備に設けられたセンサの値から、各種設備の故障が推定される。センサの値から推定される故障は、各種設備の劣化等の不具合を含んでも良い。各種設備の故障の推定は、保守点検の効率化、性能またはサービス品質等の維持において、有効である。

[0003] 昨今において、センサから得られたデータおよび周辺状況を示す様々なデータを用いた機械学習により、各種設備の故障を判断する場合がある。機械学習において、各種設備の故障を検出するためのモデルが生成される。機械学習において、故障を表す故障データと、故障していないことを表す未故障データが、教師データとして参照される。

[0004] しかしながら一般的に、故障データ数に対して未故障データ数が多い傾向がある。教師データのうち、一方の事象を表すデータセット数が多いデータをメジャーデータと称し、データセット数が少ないデータをマイナーデータと称する。またメジャーデータとマイナーデータで構成される教師データを、不均衡データと称する。

[0005] 機械学習は、不正解率を最小にするモデルを構築する。しかしながら教師データにおいて、メジャーデータのデータセット数とマイナーデータのデータセット数の不均衡の程度が大きい場合、機械学習で得られたモデルは、メジャーデータの状態や現象を正解する傾向を有する場合がある。すなわち機械学習により得られるモデルは、メジャーデータの不正解率を最小にする傾向を有する。したがって、故障データセット数に対して未故障データセット

数が多い教師データにより得られたモデルは、本来知りたいはずの故障の正解率を下げる結果につながる可能性がある。

[0006] 不均衡データによる機械学習結果の偏りに対処する方法として、大きく2つのアプローチが知られている。ひとつのアプローチは、機械学習のモデル構築プロセスにおいて、学習手法に内包されている各種パラメータの調整等を行う方法である。この方法は、学習器において、実際の結果と推定結果を比較してパラメータの調整やその結果を推定モデルにフィードバックする機能を工夫することで、推定精度を上げる。この場合、マイナーデータのデータセット数を変更しないので、学習器がマイナーデータから直接得ることのできる特徴量は変化しないため、原理的に、母集団に対するデータの代表性の影響を受ける。

[0007] もう一つのアプローチは、リサンプリング手法である。リサンプリング手法では、マイナーデータをなんらかの手段で増やし、あるいはメジャーデータを何らかの手段で減らして、データ数のバランスをとる。一般的に、前者はアップサンプリング、後者はダウンサンプリングと呼ばれる（非特許文献1）。機械学習において、アップサンプリングとダウンサンプリングの両者が同時に用いられる場合もある。

[0008] また変量間の相互依存性を表現でき、かつ関数のパラメータで相互依存性の強弱や様相を変化させ得る数学的手法として、コピュラがある。相互依存性とは、ピアソンの相関係数が表すような、正規分布に従う分布全体の直線的な関係のみではなく、多様な分布形状や、分布の位置による関係の違いを含んだ関係性を意味する。

[0009] またUCI Machine Learning Repositoryにて中性子星の観測データが、公開されている（非特許文献2および非特許文献3）。

先行技術文献

非特許文献

[001 0] 非特許文献1：Foster Provost, Machine Learning from Imbalanced Data Sets 101, AAAI Technical Report WS-00-05, 2000

非特許文献2：R. L. Lyon, "HTRU2" data, UCI Machine Learning Repository, DOI: 10.6084/m9.figshare.3080389.v1., <https://archive.ics.uci.edu/ml/datasets/HTRU2>

非特許文献3：R. L. Lyon, B. W. Stappers, S. Cooper, J. M. Brooke, J. D. Knowles, Fifty Years of Pulsar Candidate Selection: From simple filters to a new principled real-time classification approach, Monthly Notices of the Royal Astronomical Society 459 (1), 1104-1123, DOI: 10.1093/mnras/stw656

発明の概要

発明が解決しようとする課題

- [001 1] 多くの場合、機械学習で参照されるデータは多次元であり、データの様々な分布や、多変量間の様々な関係性を反映できるリサンプリング手法が求められることから、コンピュータを用いたリサンプリング手法が有効であると考えられる。しかしながら、非特許文献1に記載のリサンプリング方法では、コンピュータは用いられていない。
- [001 2] 従って本発明の目的は、コンピュータを用いてリサンプリングする学習装置、学習方法および学習プログラムを提供することである。

課題を解決するための手段

- [001 3] 上記課題を解決するために、本発明の第1の特徴は、複数のデータセットを参照して、機械学習を行う学習装置に関する。本発明の第1の特徴に係る学習装置は、第1の事象に関する複数のデータセットと、第2の事象に関する複数のデータセットを含む入力データであって、第2の事象に関するデータセットの数は、第1の事象に関するデータセットの数よりも少ない入力データを記憶する記憶装置と、第2の事象に関するデータセットから、コンピュータ関数とコンピュータ関数で用いられるパラメータを推定するコンピュータ関数推定部と、コンピュータ関数およびパラメータによるシミュレーションにより、第2の事象に関するデータセットを生成するシミュレーション部と、入力データと、シミュレーション部によって生成された第2の事象に関するデータセッ

トを参照して、第1の事象と第2の事象を区別する推定モデルを学習する学習部を備える。

[0014] コピュラ関数推定部によって推定されたパラメータ以外の新たなパラメータを生成するパラメータ生成部をさらに備え、シミュレーション部は、コピュラ関数および新たなパラメータによるシミュレーションにより、新たなパラメータについて第2の事象に関するデータセットを生成し、学習部は、入力データと、シミュレーション部により新たなパラメータについて生成された第2の事象を示すデータセットを参照して、新たなパラメータについて推定モデルを学習しても良い。

[0015] 第1の事象に関する複数のデータセットと、第2の事象に関する複数のデータセットを含む検証データを、学習部によって学習された推定モデルに入力し、検証データが示す事象と、推定モデルから得られた事象を比較して、推定モデルの不確実性を出力する検証部をさらに備えても良い。

[0016] 本発明の第2の特徴は、複数のデータセットを参照して、機械学習を行う学習方法に関する。本発明の第2の特徴に係る学習方法は、コンピュータが、第1の事象に関する複数のデータセットと、第2の事象に関する複数のデータセットを含む入力データであって、第2の事象に関するデータセットの数は、第1の事象に関するデータセットの数よりも少ない入力データを、記憶装置に記憶するステップと、コンピュータが、第2の事象に関するデータセットから、コピュラ関数とコピュラ関数で用いられるパラメータを推定するステップと、コンピュータが、コピュラ関数およびパラメータによるシミュレーションにより、第2の事象に関するデータセットを生成するステップと、コンピュータが、入力データと、生成された第2の事象に関するデータセットを参照して、第1の事象と第2の事象を区別する推定モデルを学習するステップを備える。

[0017] コンピュータが、推定するステップによって推定されたパラメータ以外の新たなパラメータを生成するステップと、コンピュータが、コピュラ関数および新たなパラメータによるシミュレーションにより、新たなパラメータに

ついて第2の事象に関するデータセットを生成するステップと、コンピュータが、入力データと、新たなパラメータについて生成された第2の事象を示すデータセットを参照して、新たなパラメータについて推定モデルを学習するステップをさらに備えても良い。

[0018] コンピュータが、第1の事象に関する複数のデータセットと、第2の事象に関する複数のデータセットを含む検証データを、推定モデルに入力し、検証データが示す事象と、推定モデルから得られた事象を比較して、推定モデルの不確実性を出力するステップをさらに備えても良い。

[0019] 本発明の第3の特徴は、コンピュータに、本発明の第1の特徴に記載の学習装置として機能させるための学習プログラムに関する。

発明の効果

[0020] 本発明によれば、コンピュータを用いてリサンプリングする学習装置、学習方法および学習プログラムを提供することができる。

図面の簡単な説明

[0021] [図1]本発明の実施の形態に係る学習装置のハードウェア構成と機能ブロックを説明する図である。

[図2]入力データを説明する図である。

[図3]コンピュータ関数推定部によるコンピュータ関数推定処理を説明するフローチャートである。

[図4]パラメータ生成部によるパラメータ生成処理を説明するフローチャートである。

[図5]シミュレーション部によるシミュレーションデータを説明する図である。

[図6]シミュレーション部によるシミュレーション処理を説明するフローチャートである。

[図7]学習部による学習処理を説明するフローチャートである。

[図8]検証部による検証処理を説明するフローチャートである。

[図9]実施例において用いられる入力データおよび検証データを説明する図で

ある。

[図 10]実施例においてシミュレーション部が生成した複数のデータセットの一例である。

[図 11]実施例において推定モデルに入力される複数のデータセットの一例である。

[図 12]実施例における検証結果の一例である。

発明を実施するための形態

[0022] 次 に、図面を参照して、本発明の実施の形態を説明する。以下の図面の記載において、同一または類似の部分には同一または類似の符号を付している。

[0023] (学習装置)

図 1 を参照して、本発明の実施の形態に係る学習装置 1 を説明する。学習装置 1 は、複数のデータセットを参照して、機械学習を行い、モデルを生成する。さらに学習装置 1 は、生成したモデルを検証する。

[0024] 学習装置 1 は、記憶装置 10、処理装置 20 および入出力インタフェース 30 を備える。学習装置 1 は、記憶装置 10、処理装置 20 および入出力インタフェース 30 を内蔵する一つのコンピュータであっても良いし、複数のハードウェアにより形成される仮想的なコンピュータであっても良い。このようなコンピュータが学習プログラムを実行することにより、図 1 に示す機能を実現する。

[0025] 記憶装置 10 は、ROM (Read Only Memory)、RAM (Random access memory)、ハードディスク等であって、処理装置 20 が処理を実行するための入力データ、出力データおよび中間データなどの各種データを記憶する。処理装置 20 は、CPU (Central Processing Unit) であって、記憶装置 10 に記憶されたデータを読み書きしたり、入出力インタフェース 30 とデータを入出力したりして、学習装置 1 における処理を実行する。入出力インタフェース 30 は、マウス、キーボード等の入力装置から入力されたデータを処理装置 20 に入力し、処理装置 20 から出力されたデータをプリンタ、表示

装置等の出力装置に出力する。また入出力インタフェース30は、ほかのコンピュータと通信するためのインタフェースであっても良い。

[0026] 記憶装置10は、入力データ11、パラメータデータ12、シミュレーションデータ13、推定モデルデータ14および検証データ15を記憶する。

[0027] 入力データ11は、第1の事象に関する複数のデータセットと、第2の事象に関する複数のデータセットを含む。入力データ11は、図2に示すように、複数のデータセットを含む。複数のデータセットのうち一部のデータセットは、第1の事象に関し、他のデータセットは、第2の事象に関する。各データセットは、複数の項目に対する値を含む。本発明の実施の形態において、各データセットは、変数六と変数巳の2つの変数に対する値を有する。

[0028] 図2に示すように、第2の事象に関するデータセットの数は、第1の事象に関するデータセットの数よりも少ない。第1の事象に関する複数のデータセットは、いわゆるメジャーデータで、第2の事象に関する複数のデータセットは、マイナーデータである。

[0029] 本発明の実施の形態において、第1の事象は、例えば、設備が故障していないことを意味し、第2の事象は、設備が故障していることを意味する。第1の事象に関するデータセットは、故障していない設備の2つのセンサからそれぞれ得られた2つのセンサの値を含む。第2の事象に関するデータセットは、故障している設備の2つのセンサからそれぞれ得られた2つのセンサの値を含む。なお、各データセットには、各データセットの値が取得された際の気温、湿度等の周辺状況のデータを含んでも良い。また屋外に設置される電柱等の設備は、周辺環境によって腐食等による劣化が生じる場合があるが、センサを設けることが困難な場合がある。そこで周辺環境によって故障が生じる設備のデータセットは、その設備の設置場所の周辺の気温、湿度等の周辺状況のデータを含んでも良い。このようにデータセットに含まれる値は、設備の故障に関連するデータであれば良く、センサ値、周辺状況のデータ等は一例である。

[0030] パラメータデータ12は、後述のパラメータ生成部22によって生成され

たコンピュータ関数のパラメータの値を含む。１つのコンピュータ関数について複数のパラメータがある場合、パラメータデータ１２は、各パラメータの値を対応づけて保持する。

[0031] シミュレーションデータ１３は、後述のシミュレーション部２３によって生成された第２の事象のデータセットである。シミュレーションデータ１３は、複数のデータセットを含んでも良い。

[0032] 推定モデルデータ１４は、後述の学習部２４によって得られるモデルを特定するデータである。本発明の実施の形態において推定モデルデータ１４は、第１の事象と第２の事象を区別するために用いられる。推定モデルデータ１４は、入力データ１１に対応するパラメータから生成された推定モデルを特定するデータを含む。推定モデルデータ１４はさらに、パラメータ生成部２２によって生成されたパラメータによって生成された推定モデルを特定するデータを含んでも良い。

[0033] 検証データ１５は、推定モデルデータ１４を検証するために参照されるデータである。検証データ１５は、入力データ１１と同様に、第１の事象に関する複数のデータセットと、第２の事象に関する複数のデータセットを含む。また検証データ１５に含まれるデータセットは、入力データ１１と同様に、変数六と変数巳の２つの変数に対する値を有する。また検証データ１５における第１の事象のデータセット数と第２の事象のデータセット数の割合は、入力データ１１における割合と同様である。入力データ１１と検証データ１５は、例えば、同一の母集団に属する複数のデータセットを、２分して生成されても良い。

[0034] 処理装置２０は、コンピュータ関数推定部２１、パラメータ生成部２２、シミュレーション部２３、学習部２４および検証部２５を備える。

[0035] コンピュータ関数推定部２１は、入力データ１１のうち、第２の事象に関するデータセットから、コンピュータ関数とコンピュータ関数で用いられるパラメータを推定する。コンピュータ関数は、変数八と変数巳の相関の構造を示す。コンピュータ関数で用いられるパラメータは、コンピュータ関数が示す相関の構造における様

相を示し、各変数の値のばらつきの程度等に関連する。コピュラ関数が複数のパラメータを含む場合、コピュラ関数推定部 2 1 は、各パラメータを推定する。

[0036] 本発明の実施の形態では、各データセットは変数六および変数巳の2つの変数を含むので、コピュラ関数推定部 2 1 は、2変数コピュラから、最適なコピュラを推定する。データセットに3つ以上の変数を含む場合、コピュラ関数推定部 2 1 は、多変数に対応したコピュラを推定しても良いし、2変数の組み合わせで変数全体の関係性を記述するヴァインコピュラのような方法を用いても良い。

[0037] 図3を参照して、コピュラ関数推定部 2 1 によるコピュラ関数推定処理を説明する。

[0038] ステップ3 1 0 1においてコピュラ関数推定部 2 1 は、入力データ 1 1 から、第2の事象に関する複数のデータセットを抽出する。ステップ3 1 0 2においてコピュラ関数推定部 2 1 は、ステップ3 1 0 1で抽出したデータセットから、コピュラ関数およびそのコピュラ関数のパラメータを推定する。コピュラ関数推定処理は終了する。

[0039] パラメータ生成部 2 2 は、コピュラ関数推定部 2 1 によって推定されたパラメータ以外の新たなパラメータを生成する。パラメータ生成部 2 2 は、生成したパラメータを、パラメータデータ 1 2 に格納する。コピュラ関数推定部 2 1 が推定したコピュラ関数が複数のパラメータを含む場合、パラメータ生成部 2 2 は、生成した各パラメータを対応づけたパラメータセットを、パラメータデータ 1 2 に格納する。パラメータ生成部 2 2 は、1以上のパラメータまたはパラメータセットを生成する。

[0040] パラメータ生成部 2 2 は、各パラメータが取り得る範囲を等分割して、各パラメータの値を決定しても良い。あるいはパラメータ生成部 2 2 は、各パラメータが取り得る範囲の値をランダムに発生させて、各パラメータの値を決定しても良い。

[0041] 図4を参照して、パラメータ生成部 2 2 によるパラメータ生成処理を説明

する。

[0042] ステップ3 2 0 1においてパラメータ生成部2 2は、コンピュータ関数推定部2 1によって推定された関数について、複数のパラメータを生成する。ステップ3 2 0 2においてパラメータ生成部2 2は、ステップ3 2 0 1で生成した複数のパラメータを、パラメータデータ1 2に格納する。パラメータ生成処理は終了する。

[0043] シミュレーション部2 3は、コンピュータ関数推定部2 1によって推定されたコンピュータ関数およびパラメータを用いて、シミュレーションにより、第2の事象に関するデータセットを生成する。シミュレーション部2 3が生成するデータセットは、入力データ1 1における第2の事象に関するデータセットにおける変数間の相関構造を維持しつつ、相互依存性の強弱やばらつき等のデータの様相の異なるデータセットである。シミュレーション部2 3は、入力データ1 1においてデータセット数の少ない第2の事象に関するデータセット数を増やし、入力データ1 1における不均衡を軽減する。

[0044] シミュレーション部2 3は、シミュレーションにより、変数八および変数巳について、新たな値が設定された第2の事象に関するデータセットを生成する。ここでシミュレーション部2 3が新たに生成するデータセットの変数八および変数巳は、入力データ1 1の第2の事象に関するデータセットの変数八および変数巳と同じであっても良いし、異なっても良い。

[0045] シミュレーション部2 3はさらに、コンピュータ関数と、パラメータ生成部2 2によって生成された新たなパラメータによるシミュレーションにより、新たなパラメータについて第2の事象に関するデータセットを生成する。シミュレーション部2 3は、パラメータ生成部2 2によって生成されたパラメータまたはパラメータセットを用いて、コンピュータ関数推定部2 1によって推定されたコンピュータ関数を参照する。シミュレーション部2 3は、パラメータまたはパラメータセット毎に、シミュレーションにより、変数八および変数巳について、新たな値が設定された第2の事象に関するデータセットを生成する。シミュレーション部2 3が生成した第2の事象に関するデータセットは

、シミュレーションデータ 13 において、パラメータに対応づけられて格納される。

[0046] シミュレーション部 23 は、メジャーデータのデータセット数からマイナーデータのデータセット数を引いた数のデータセットを、シミュレーションにより生成するのが好ましい。これにより、図 5 に示すように、第 1 の事象を示すデータセット数と第 2 の事象を示すデータセット数が一致する。シミュレーション部 23 は、マイナーデータにおける変数間の相関構造を維持しつつ、相互依存性の強弱やばらつき等のデータの様相の異なる複数のデータセットを増やすことにより、メジャーデータとマイナーデータのデータセット数の不均衡に伴う不具合を解消することができる。

[0047] 図 6 を参照して、シミュレーション部 23 によるシミュレーション処理を説明する。

[0048] ステップ 3301 においてシミュレーション部 23 は、入力データ 11 における第 1 の事象のデータセット数と第 2 の事象のデータセット数の差分を、シミュレーションデータセット数として算出する。

[0049] 各パラメータについて、ステップ 3302 の処理を繰り返す。このパラメータは、コピュラ関数推定部 21 によって推定されたパラメータである。またパラメータは、パラメータ生成部 22 によって生成されたパラメータを含んでも良い。

[0050] ステップ 3302 において、コピュラ関数推定部 21 により推定されたコピュラ関数と処理対象のパラメータを用いて、ステップ 3301 で算出したシミュレーションデータセット数のデータセットを生成する。ここで生成されるデータセットは、第 2 の事象に関する。各パラメータについてステップ 3302 の処理が終了すると、シミュレーション処理は終了する。

[0051] 学習部 24 は、入力データ 11 と、シミュレーション部 23 によって生成された第 2 の事象に関するデータセットを参照して、第 1 の事象と第 2 の事象を区別する推定モデルを学習する。ここで学習部 24 は、入力データ 11 からコピュラ関数推定部 21 により推定されたパラメータについて、推定モ

デルを学習する。推定モデルは、データセットが入力されると、そのデータセットが示す事象を出力する。本発明の実施の形態において推定モデルは、変数 八および変数 巳を含むデータセットが入力されると、そのデータセットが第 1 の事象に関連すること、あるいはそのデータセットが第 2 の事象に関連することを判断する。

[0052] 学習部 24 は、さらに、パラメータ生成部 22 により生成されたパラメータについても、推定モデルを学習する。学習部 24 は、入力データ 11 と、シミュレーション部 23 により新たなパラメータについて生成された第 2 の事象を示すデータセットを参照して、新たなパラメータについて推定モデルを学習する。パラメータ生成部 22 が複数のパラメータを生成した場合、学習部 24 は、パラメータごとに、推定モデルを学習する。

[0053] 学習部 24 は、パラメータごとに学習した推定モデルを、推定モデルデータ 14 に格納する。本発明の実施の形態において、学習部 24 が採用する機械学習方法は制限がなく、既存の機械学習方法により機械学習を行えば良い。

[0054] 学習部 24 に入力される教師データは、第 1 の事象に関するデータセット数と同じ第 2 の事象に関するデータセット数を含む。学習部 24 は、第 1 の事象または第 2 の事象に傾向しない推定モデルを出力することができる。

[0055] 図 7 を参照して、学習部 24 による学習処理を説明する。

[0056] 学習部 24 は、各パラメータについて、ステップ 3401 の処理を繰り返す。ステップ 3401 において学習部 24 は、入力データ 11 のデータセットと、処理対象のパラメータについてシミュレーション部 23 により生成されたデータセットから、推定モデルを学習する。

[0057] 各パラメータについてステップ 3401 の処理が終了すると、学習部 24 は処理を終了する。

[0058] 検証部 25 は、検証データ 15 を、学習部 24 によって学習された推定モデルに入力し、検証データ 15 が示す事象と、推定モデルから得られた事象を比較して、推定モデルの不確実性を出力する。検証部 25 は、入力データ

1 1 の不均衡を補正したデータから導かれた推定モデルを用いて、不均衡を補正していない検証データ 1 5 の各データセットを判別し、その挙動を確認および検証する。検証部 2 5 が出力する推定モデルの不確実性は、シミュレーション部 2 3 により生成された第 2 の事象に関するデータセットに関する。

[0059] 学習部 2 4 は、コピュラ関数推定部 2 1 によって推定されたパラメータについて生成された推定モデルと、パラメータ生成部 2 2 によって生成されたパラメータについて生成された推定モデルの複数の推定モデルを生成する。パラメータ生成部 2 2 が生成したパラメータが複数の場合、学習部 2 4 によって 3 つ以上の推定モデルが生成される場合もある。

[0060] 検証部 2 5 は、このように生成された複数の推定モデルのそれぞれに、検証データ 1 5 を入力し、各推定モデルが示す事象が、検証データ 1 5 に示す事象と一致するか否かを評価する。例えば、検証データ 1 5 において第 1 の事象に関連するデータセットを推定モデルに入力し、推定モデルが第 1 の事象を示す場合、推定モデルは、正しい結果を出力したことになる。また検証データ 1 5 において第 1 の事象に関連するデータセットを推定モデルに入力し、推定モデルが第 2 の事象を示す場合、推定モデルは、誤った結果を出力したことになる。このように検証部 2 5 は、推定モデルが出力した事象と、検証データ 1 5 が示す事象を比較して、推定モデルの確からしさを出力する。

[0061] 本発明の実施の形態において検証部 2 5 は、複数の推定モデルについて検証する場合を説明するが、これに限らない。検証部 2 5 は、入力データ 1 1 のマイナーデータから得られたパラメータについての推定モデルについてのみ、検証しても良い。

[0062] 検証部 2 5 が不確実性を出力する指標は、適宜設定される。例えば指標は、全体正解率、劣化正解率、見逃し率、空振り率等が考えられる。全体正解率は、第 1 の事象（未故障）および第 2 の事象（故障）を問わない正解率であって、推定モデルが出力した事象が、検証データ 1 5 のデータセットが示

す事象と一致する確率である。劣化正解率は、検証データ 15 のうち第 2 の事象（故障）を示すデータセットについてのみの正解率である。見逃し率は、検証データ 15 のデータセットのうち、検証データ 15 において第 2 の事象に関するデータセットが推定モデルによって第 1 の事象と推定されたデータセットの数の確率である。空振り率は、検証データ 15 のデータセットのうち、検証データ 15 において第 1 の事象に関するデータセットが推定モデルによって第 2 の事象と推定されたデータセットの数の確率である。

[0063] 検証部 25 は、これらの必要な指標を設定し、あらかじめ設定された計算方法で算出して出力する。

[0064] 図 8 を参照して、検証部 25 による検証処理を説明する。

[0065] まず、各パラメータについて、ステップ 3401 およびステップ 3402 の処理を行う。ステップ 3401 において検証部 25 は、処理対象のパラメータで算出した推定モデルを取得する。ステップ 3402 において検証部 25 は、検証データ 15 の各データセットをステップ 3401 で取得した推定モデルに適用して、各データセットについて、推定モデルが推定した事象を取得する。

[0066] 各パラメータについて、ステップ 3401 およびステップ 3402 の処理が終了すると、ステップ 3403 において、ステップ 3402 において推定モデルに適用した結果を評価する。検証部 25 は、パラメータ毎に、推定モデルに適用した結果を評価しても良いし、各パラメータで得られた結果をまとめて評価しても良い。

[0067] 検証部 25 は、ステップ 3403 で得られた評価を出力して、処理を終了する。

[0068] (コピュラ)

ここで、コピュラについて説明する。コピュラの説明において周辺分布とは、同時分布を構成する各分布のことであり、データセットに含まれる変数八および変数巳のことである。

[0069] コピュラの基礎的な理論は、Sklarの定理に従って展開される。任意のづ次

元分布関数をドとすると、式 (1) となる d 次元接合関数 C が存在する。 d 次元接合関数 C を、コピュラと呼ぶ。

[0070] [数 1]

$$C(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))$$

• . . 式 (1)

C : コピュラ

d : 変数の次数

u_i : 変数

F_i : F の i 番目の 1 次元周辺分布関数 ($i = 1, \dots, d$)

F^{-1} : F の逆関数

[0071] ドが連続である場合、 C は一意的に決まり、 C は、 D の接合関数と呼ばれる。この場合、 C は、式 (2) により与えられる。

[0072] [数 2]

$$C(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))$$

• . . 式 (2)

[0073] コピュラは、分布関数から与えられることから、一様分布同士をつなぐ。すなわちコピュラは、元の周辺分布が持つ情報を失う一方、周辺分布の分布関数間の相関および関係性のみを残したものと言える。

[0074] コピュラが持つ周辺分布の分布関数間の相関および関係性の強さ、すなわち相互依存性の強さを表す指標として、多くの場合、ケンドールの τ が用いられる。 τ は、ケンドールの順位相関係数である。 τ は -1 から 1 の間の値を取り、値の増加は相互依存性が強いことを意味する。順位が完全に一致している場合、 τ は 1 を示し、順位が完全に独立している場合、 τ は 0 を示し、順位が完全に一致していない場合、 τ は -1 を示す。

[0075] コピュラ関数はいくつかの種類が示されており、2 次元コピュラや 3 次元以上の多次元コピュラが存在する。各コピュラ関数はそれぞれパラメータを持っており、パラメータによって分布が変化する。パラメータの数はコピュラ関数の種類によって異なる。また、各コピュラ関数のパラメータとケンド

ールの 7:は関係を持つ。

[0076] コピュラ関数推定部 2 1 は、入力データ 1 1 のマイナーデータについて、コピュラ関数の複数の種類のうち、変数 六と変数 巳の関係性を表すコピュラ関数を特定する。コピュラ関数推定部 2 1 はさらに、特定したコピュラ関数で用いられるパラメータの値を特定する。

[0077] (実施例)

本発明の実施の形態に係る学習装置 1 における実施例を説明する。

[0078] 入力データ 1 1 および検証データ 1 5 に含まれるデータセットは、非特許文献 2 および非特許文献 3 に開示されている中性子星の観測データからランダムに抽出した 1 万件のデータセットである。実施例において、中性子星の観測データの「クラスデータ」に記録されている 0 の値を、ある設備の未故障の事象を示す識別子と読み替え、1 の値を、ある設備の故障の事象を示す識別子と読み替える。なお、観測データの「クラスデータ」において、0 の値のデータセットは、1 の値のデータセットよりも多い。

[0079] 非特許文献 2 および 3 の観測データにおいて 8 項目の値が記録されているが、実施例において、8 項目から選択した 2 項目を、それぞれ変数 八および変数 巳の値とする。これにより、変数 六および変数 巳から、故障または未故障を判別するための、複数のデータセットが得られる。

[0080] まず、複数のデータセットを、推定モデルを生成するための入力データ 1 1 と、推定モデルを検証するための検証データ 1 5 に区分する。入力データ 1 1 に分類された複数のデータセットと検証データ 1 5 に分類された複数のデータセットに隔たりがなければ、どのような方法で分類されても良い。例えば、ランダムに分類する方法がある。また実施例において、入力データ 1 1 に分類されたデータセットの数と検証データ 1 5 に分類されたデータセットの数は、1 対 1 となるようにしたが、異なる比率であっても良い。

[0081] 実施例において、1 万件のデータセットから区分した入力データ 1 1 と検証データ 1 5 の内訳を、図 9 に示す。入力データ 1 1 および検証データ 1 5 とともに、未故障を示すデータセット数と故障を示すデータセット数の比率は

、約 10 : 1 と不均衡な状態である。実施例において、入力データ 11 のうち、未故障を示すデータセットを含むデータは、メジャーデータであって、故障を示すデータセットを含むデータは、マイナーデータである。

[0082] このように、入力データ 11 と検証データ 15 が決定されると、コピュラ関数推定部 21 が、コピュラ関数およびパラメータセットを推定する。コピュラ関数推定部 21 は、入力データ 11 のうちのマイナーデータ、すなわち故障を示すデータセットを参照して、コピュラ分析を行う。コピュラ分析は、一般的な方法で良い。実施例において、変数六と変数巳の相互依存性を表すコピュラと、そのコピュラのパラメータセットは以下のように推定された。実施例においてパラメータセットは、パラメータ 0 およびパラメータ 5 である。

[0083] コピュラ関数 :888 Copula

パラメータ 0 : 5.14

パラメータ 5 : 0.62

ケンドールの τ :0.41

BB8 Copula の定義式を、式 (3) に示す。

[0084] [数 3]

$$C(u, v; \theta, \delta) = \delta^{-1} \left(1 - \{1 - \eta^{-1} [1 - (1 - \delta u)^\theta] [1 - (1 - \delta v)^\theta]\}^{\frac{1}{\theta}} \right)$$

$$\theta \geq 1, 0 < \delta \leq 1, \text{ where } \eta = 1 - (1 - \delta)^\theta \text{ and } 0 \leq u, v \leq 1$$

・・・式 (3)

[0085] マイナーデータについて、コピュラ関数とパラメータセットが推定されると、パラメータ生成部 22 によって、パラメータセットを増やす。実施例においてパラメータ生成部 22 は、コピュラ関数推定部 21 によって推定されたパラメータセット (0, 5) = (5.14, 0.64) のほか、999 個のパラメータセットを生成し、合計 1000 個のパラメータセットを用意する。パラメータ生成部 22 は、0 と 5 の値をランダムに振って複数のパラメータセットを作成する。0 と 5 の各値の範囲は、コピュラ関数の各パラメータ

タが取りうる範囲が数学的に定められている場合、定められた範囲に従う。
コンピュータ関数の各パラメータが取りうる範囲が定められていない場合、ユーザが適宜設定しても良いし、あらかじめシステムに設定されても良い。実施例において、 $1 \leq \theta < 8$ かつ $0 < \delta \leq 1$ の範囲で、0および5について、1000個のパラメータセットが作成される。

[0086] パラメータセットが生成されると、シミュレーション部23が、パラメータセット毎に周辺分布のシミュレーションを行う。シミュレーション部23は、入力データ11における不均衡を是正するためにマイナーデータのデータセット数を増やす。図9に示すように、入力データ11において、メジャーデータは、4564件のデータセットを含み、マイナーデータは、436件のデータセットを含む。従ってシミュレーション部23は、パラメータセットごとに、メジャーデータのデータセット数4564件からマイナーデータのデータセット数436を引いた4128件のデータセットを、シミュレーションにより生成する。

[0087] 図10に、シミュレーション部23が生成したデータセットの例を示す。図10(3)は、コンピュータ関数推定部21によって推定されたパラメータセット $(\ominus, 8) = (5.14, 0.64)$ についてシミュレーションされた変数六および変数巳の周辺分布である。図10(ヒ)は、パラメータ生成部22によって生成されたパラメータセット $(0, 5) = (1.0, 0.64)$ についてシミュレーションされた変数八および変数巳の周辺分布である。図10(○)は、パラメータ生成部22によって生成されたパラメータセット $(0, 5) = (8.0, 0.64)$ についてシミュレーションされた変数八および変数巳の周辺分布である。

[0088] なお、図10(3)に示す周辺分布は、左下から右上にかけて帯状に形成され、右上よりも左下の方が、密度が濃い傾向がある。従って、コンピュータ関数推定部21は、このような変数の関係性を表現可能なコンピュータ関数を推定する。またパラメータセットに因って、分布の分散度が異なるが、図10(㊦)および図10(○)の各分布においても、図10(3)と同様に、左下

から右上にかけて帯状に形成され、右上よりも左下の方が、密度が濃い傾向がある。

[0089] シミュレーション部 2 3 により、各パラメータセットについて、メジャーデータのデータセット数とマイナーデータのデータセット数が、同じになり、教師データの不均衡が解消される。教師データは、入力データ 1 1 の各データセットと、シミュレーション部 2 3 により生成された各データセットである。

[0090] 図 1 1 を参照して、教師データの分布を説明する。図 1 1 (3) は、入力データ 1 1 のデータセットと、コピュラ関数推定部 2 1 によって推定されたパラメータセット $(\ominus, 8) = (5.14, 0.64)$ についてシミュレーションされたデータセットの、変数 六および変数 巳の周辺分布である。図 1 1 (し) は、入力データ 1 1 のデータセットと、パラメータ生成部 2 2 によって生成されたパラメータセット $(0, 5) = (1.0, 0.64)$ についてシミュレーションされたデータセットの、変数 六および変数 巳の周辺分布である。図 1 1 (○) は、入力データ 1 1 のデータセットと、パラメータ生成部 2 2 によって生成されたパラメータセット $(0, 5) = (8.0, 0.64)$ についてシミュレーションされたデータセットの、変数 八および変数 巳の周辺分布である。

[0091] 図 1 1 の各図において、黒点が、未故障を示すデータセットで、白点が、故障を示すデータセットである。白点のデータセットは、入力データ 1 1 に含まれるデータセットのほか、シミュレーション部 2 3 により生成されたデータセットを含む。実施例において、1000 個のパラメータセットのそれぞれについて、図 1 1 の各図に示すデータセット群が生成される。

[0092] 学習部 2 4 は、各パラメータセットについて、不均衡が解消された教師データから、推定モデルを生成する。実施例では、1000 個の推定モデルが生成される。実施例において学習部 2 4 は、サポートベクターマシンにより、事象を区別可能な推定モデルを導出する。

[0093] 検証部 2 5 は、学習部 2 4 によって生成された各推定モデルについて、不

確実性に関する指標を出力する。

[0094] 一般的に、機械学習により得られた推定結果のみを提示しても、実際の設備等の保守において充分ではないと考えられる。多くの場合、機械学習による推定行為は不確実性を持ち、推定結果は、潜在的に幅を持ちうる。すなわち、推定を用いて保守計画を立案する場合、推定が持つ不確実性を考慮することが求められる。

[0095] 本発明の実施の形態に係る学習装置 1 は、パラメータセット毎に、マイナーデータのデータセットを生成して、パラメータセット毎に異なる集団に対する推定モデルを生成する。パラメータセットは、コンピュータ関数のパラメータが数学的に規定される範囲、あるいは取り得ると想定した範囲で設定される。従って、各パラメータセットは、それぞれ、マイナーデータが属する可能性のある異なる母集団を、規定する。これにより、学習装置 1 で生成される推定モデル群は、マイナーデータが属する可能性のある異なる母集団に対応した推定モデルで構成される。検証部 25 は、このように生成した推定モデル群について各種の指標を出力する。これら推定モデル群を検証に用いることで、マイナーデータのリサンプリングに伴う機械学習結果の不確実性の情報を得ることができる。

[0096] 図 12 は、検証部 25 が出力する検証結果の一例である。図 12 は、実施例において、1000 個の推定モデルに検証データ 15 を適用した際の、劣化正解率と空振り率の関係を示す。図 12 に示される 1 つの黒マーク 70 は、1 つのパラメータセットに対応する推定モデルに検証データ 15 を適用した際の、劣化正解率および空振り率を示す。

[0097] 図 12 に示す検証結果は、劣化正解率は約 0.80 - 0.85 の範囲を、空振り率は約 0.03 - 0.06 の範囲で値を取りうるということがわかる。図 12 の検証結果は、保守計画者に対して、本発明の実施の形態に係る推定モデルは、図 12 に示された程度のブレが発生しうることを前提にして、推定モデルを活用した保守計画を立てるべきである、と示すことができる。

[0098] なお、検証部 25 が示す検証結果は、図 12 に示すように指標間の関係性

をグラフで表されても良いし、近似関数で表されても良い。また、保守で目標とする指標値や指標値の範囲が決まっている場合、検証部 25 は、学習部 24 で生成された複数の推定モデルのうち、その目標に合う推定モデルに関する検証結果のみを示しても良い。

[0099] このような本発明の実施の形態にかかる学習装置 1 によれば、コンピュータ関数のシミュレーションにより、入力データ 11 におけるマイナーデータにおける変量間の相互依存性を反映したデータセットを増やすことができる。従って学習装置 1 は、入力データ 11 において不均衡が生じる場合でも、各事象を示すデータセットのそれぞれの数を、同じにすることができる。これにより学習装置 1 が出力する推定モデルは、メジャーデータの不正解率を最小にする傾向が抑制され、メジャーデータおよびマイナーデータの不正解率を最小にすることができる。

[0100] また学習装置 1 は、コンピュータ関数のパラメータセットを複数生成して、各パラメータセットについて推定モデルを生成する。これにより、入力データ 11 から得られた傾向を有する多数の推定モデルを生成することができる。

[0101] 学習装置 1 はさらに、各パラメータセットについて生成された各推定モデルについて検証する。これにより学習装置 1 は、期待できる成績の範囲や想定すべき推定のはずれ具合を事前に把握できることで、各推定モデルの不確実性を数値化することができる。また学習装置 1 が出力する推定モデルの範囲が正確になるので、この推定モデルを利用した予測精度が向上し、不均衡データのリサンプリングによって発生する不確実性を考慮した保守計画立案が可能となる。

[0102] (その他の実施の形態)

上記のように、本発明の実施の形態とその実施例によって記載したが、この開示の一部をなす論述および図面はこの発明を限定するものであると理解すべきではない。この開示から当業者には様々な代替実施の形態、実施例および運用技術が明らかとなる。

[0103] 例えば、本発明の実施の形態に記載した学習装置は、図 1 に示すように一

つのハードウェア上に構成されても良いし、その機能や処理数に応じて複数のハードウェア上に構成されても良い。また、ほかの機能を実現する既存の情報処理装置上に実現されても良い。

[01 04] 本発明はここでは記載していない様々な実施の形態等を含むことは勿論である。従って、本発明の技術的範囲は上記の説明から妥当な請求の範囲に係る発明特定事項によってのみ定められるものである。

符号の説明

- [01 05] 1 学習装置
- 1 0 記憶装置
- 1 1 入力データ
- 1 2 パラメータデータ
- 1 3 シミュレーションデータ
- 1 4 推定モデルデータ
- 1 5 検証データ
- 2 0 処理装置
- 2 1 コピュラ関数推定部
- 2 2 パラメータ生成部
- 2 3 シミュレーション部
- 2 4 学習部
- 2 5 検証部
- 3 0 入出力インタフェース

請求の範囲

- [請求項 1] 複数のデータセットを参照して、機械学習を行う学習装置であって、
- 、
- 第 1 の事象に関する複数のデータセットと、第 2 の事象に関する複数のデータセットを含む入力データであって、前記第 2 の事象に関するデータセットの数は、前記第 1 の事象に関するデータセットの数よりも少ない前記入力データを記憶する記憶装置と、
- 前記第 2 の事象に関するデータセットから、コンピュータ関数と前記コンピュータ関数で用いられるパラメータを推定するコンピュータ関数推定部と
- 、
- 前記コンピュータ関数および前記パラメータによるシミュレーションにより、前記第 2 の事象に関するデータセットを生成するシミュレーション部と、
- 前記入力データと、前記シミュレーション部によって生成された前記第 2 の事象に関するデータセットを参照して、前記第 1 の事象と前記第 2 の事象を区別する推定モデルを学習する学習部
- を備えることを特徴とする学習装置。
- [請求項 2] 前記コンピュータ関数推定部によって推定されたパラメータ以外の新たなパラメータを生成するパラメータ生成部をさらに備え、
- 前記シミュレーション部は、前記コンピュータ関数および前記新たなパラメータによるシミュレーションにより、前記新たなパラメータについて前記第 2 の事象に関するデータセットを生成し、
- 前記学習部は、前記入力データと、前記シミュレーション部により前記新たなパラメータについて生成された前記第 2 の事象を示すデータセットを参照して、前記新たなパラメータについて推定モデルを学習する
- ことを特徴とする請求項 1 に記載の学習装置。
- [請求項 3] 第 1 の事象に関する複数のデータセットと、第 2 の事象に関する複

数のデータセットを含む検証データを、前記学習部によって学習された推定モデルに入力し、前記検証データが示す事象と、前記推定モデルから得られた事象を比較して、前記推定モデルの不確実性を出力する検証部

をさらに備えることを特徴とする請求項 1 または 2 に記載の学習装置。

[請求項 4] 複数のデータセットを参照して、機械学習を行う学習方法であって、

コンピュータが、第 1 の事象に関する複数のデータセットと、第 2 の事象に関する複数のデータセットを含む入力データであって、前記第 2 の事象に関するデータセットの数は、前記第 1 の事象に関するデータセットの数よりも少ない前記入力データを、記憶装置に記憶するステップと、

前記コンピュータが、前記第 2 の事象に関するデータセットから、コピュラ関数と前記コピュラ関数で用いられるパラメータを推定するステップと、

前記コンピュータが、前記コピュラ関数および前記パラメータによるシミュレーションにより、前記第 2 の事象に関するデータセットを生成するステップと、

前記コンピュータが、前記入力データと、生成された前記第 2 の事象に関するデータセットを参照して、前記第 1 の事象と前記第 2 の事象を区別する推定モデルを学習するステップ

を備えることを特徴とする学習方法。

[請求項 5] 前記コンピュータが、前記推定するステップによって推定されたパラメータ以外の新たなパラメータを生成するステップと、

前記コンピュータが、前記コピュラ関数および前記新たなパラメータによるシミュレーションにより、前記新たなパラメータについて前記第 2 の事象に関するデータセットを生成するステップと、

前記コンピュータが、前記入力データと、前記新たなパラメータについて生成された前記第2の事象を示すデータセットを参照して、前記新たなパラメータについて推定モデルを学習するステップ

をさらに備えることを特徴とする請求項4に記載の学習方法。

[請求項6]

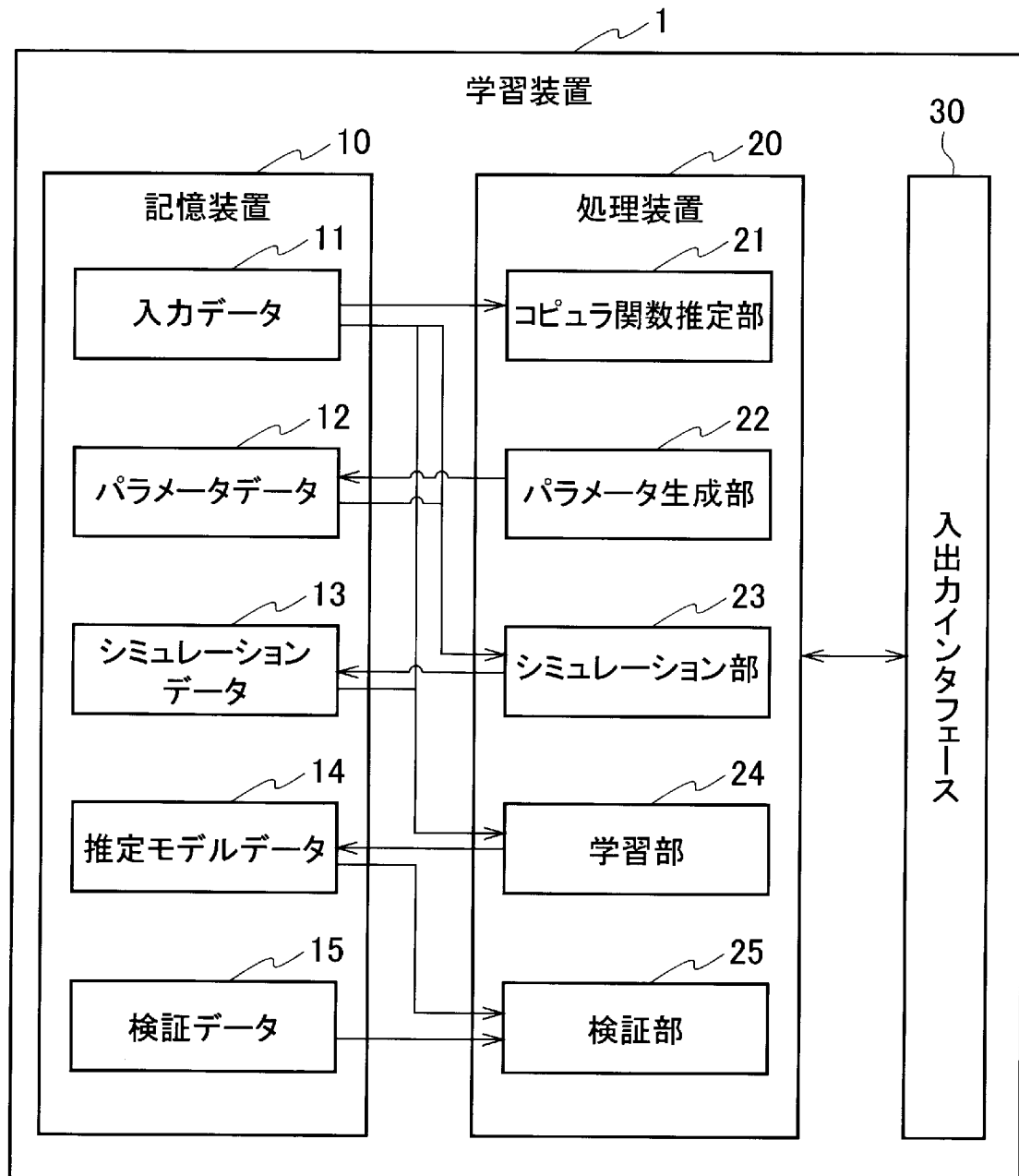
前記コンピュータが、第1の事象に関する複数のデータセットと、第2の事象に関する複数のデータセットを含む検証データを、前記推定モデルに入力し、前記検証データが示す事象と、前記推定モデルから得られた事象を比較して、前記推定モデルの不確実性を出力するステップ

をさらに備えることを特徴とする請求項4または5に記載の学習方法。

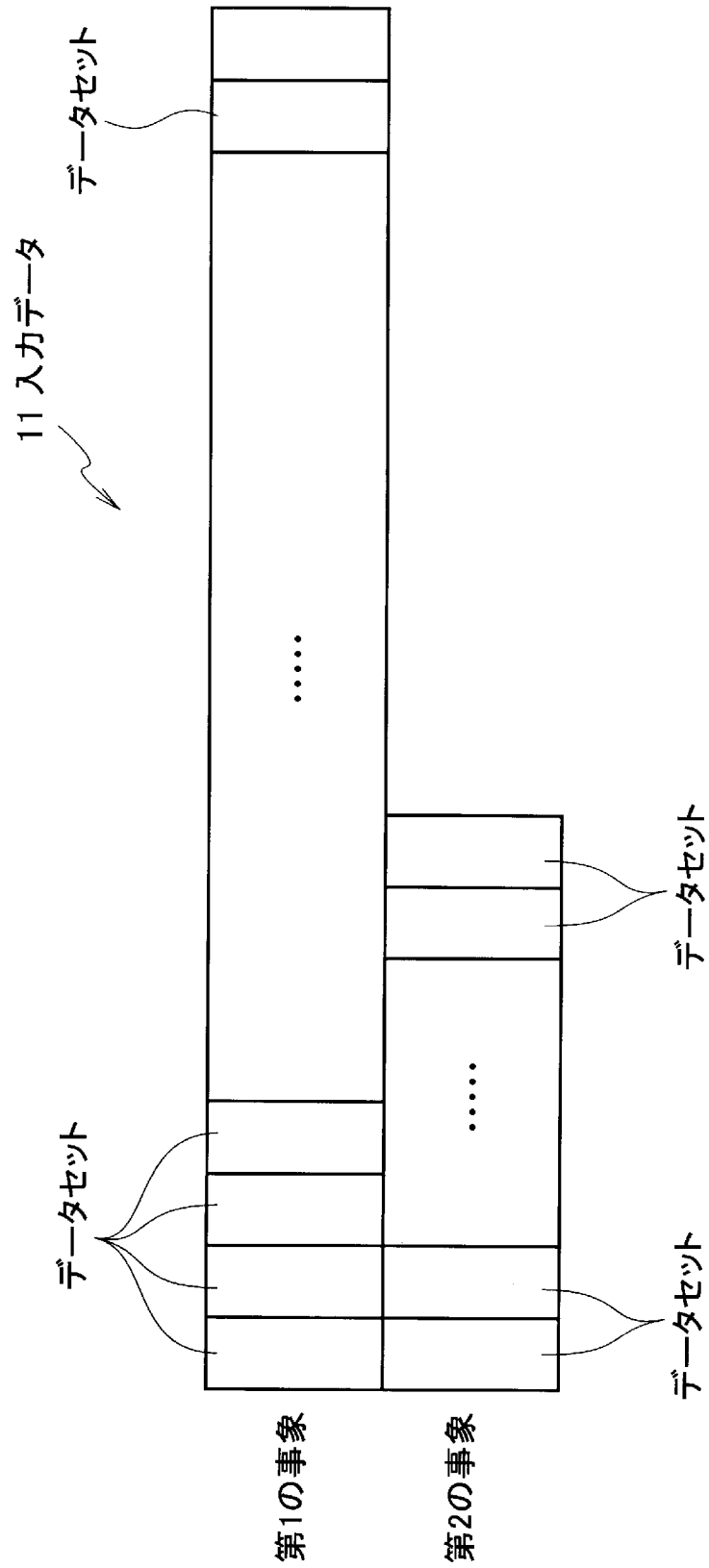
[請求項7]

コンピュータに、請求項1ないし請求項3のいずれか1項に記載の学習装置として機能させるための学習プログラム。

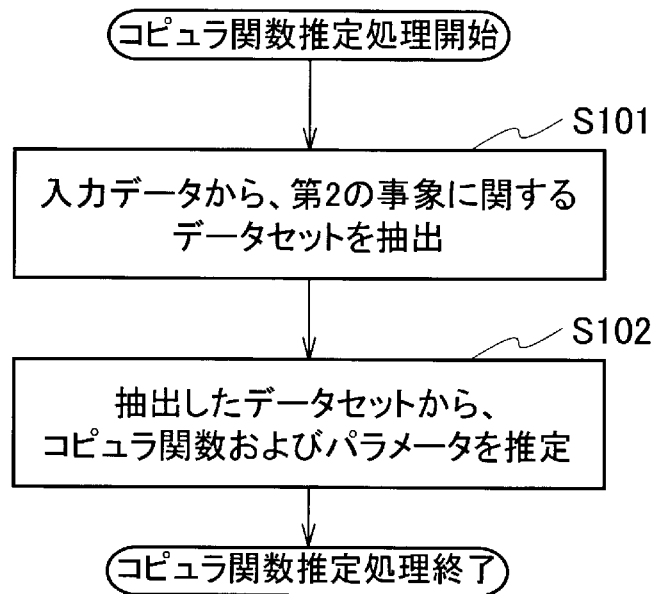
[図1]



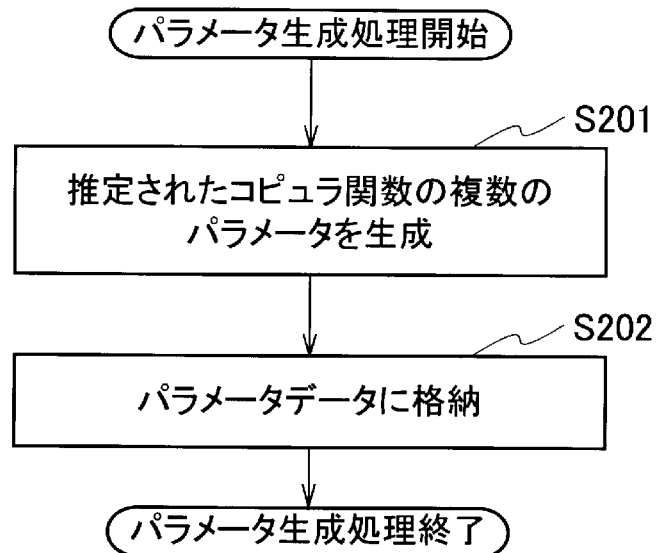
[図2]



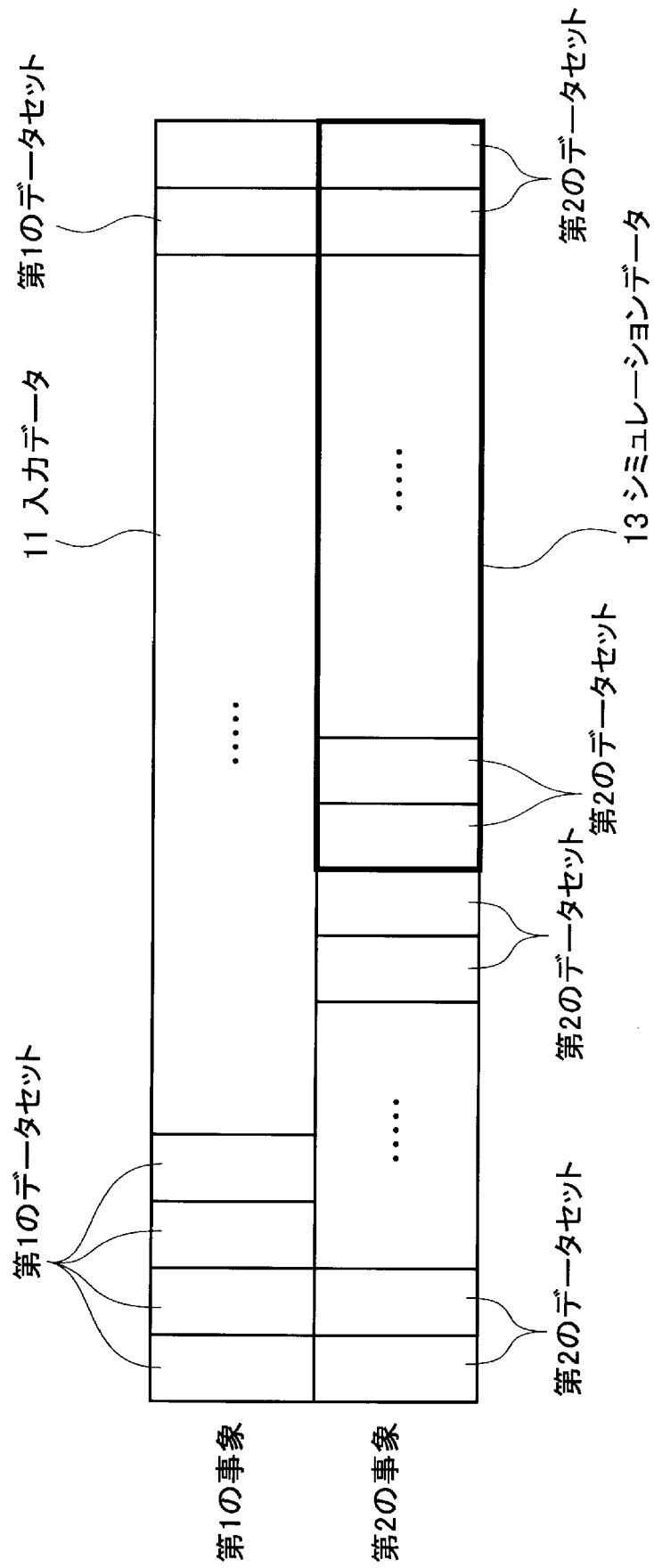
[図3]



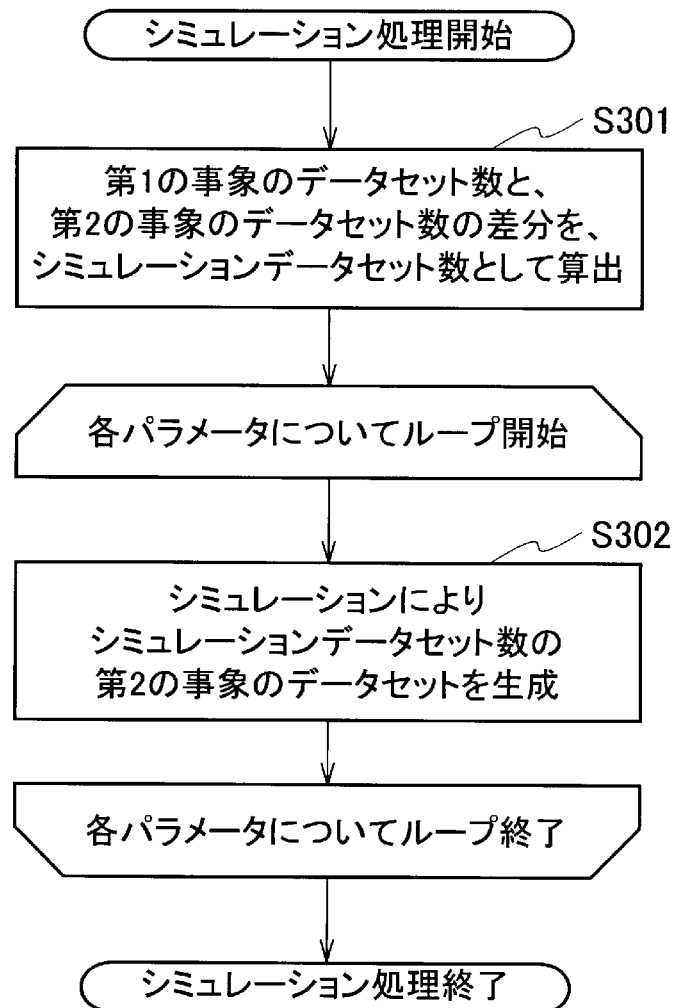
[図4]



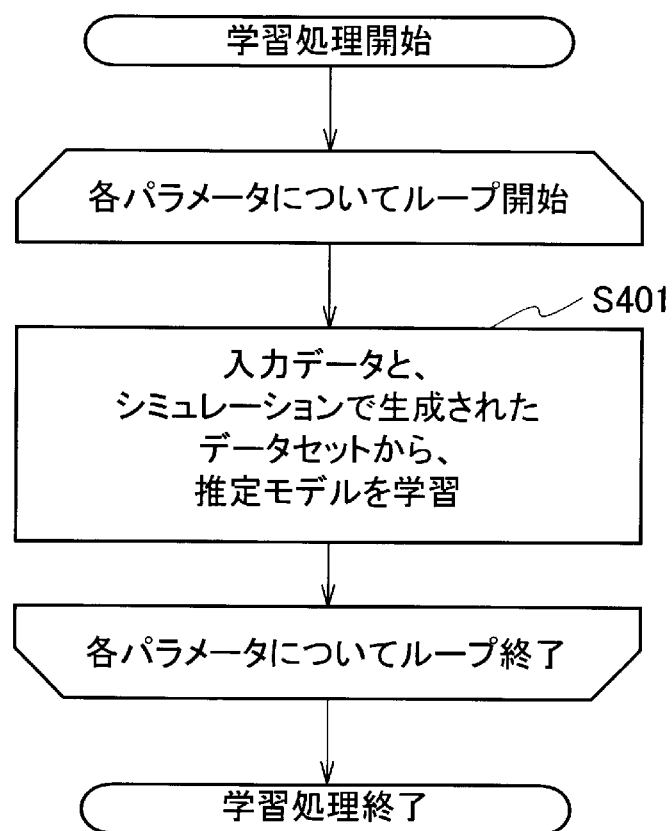
[図5]



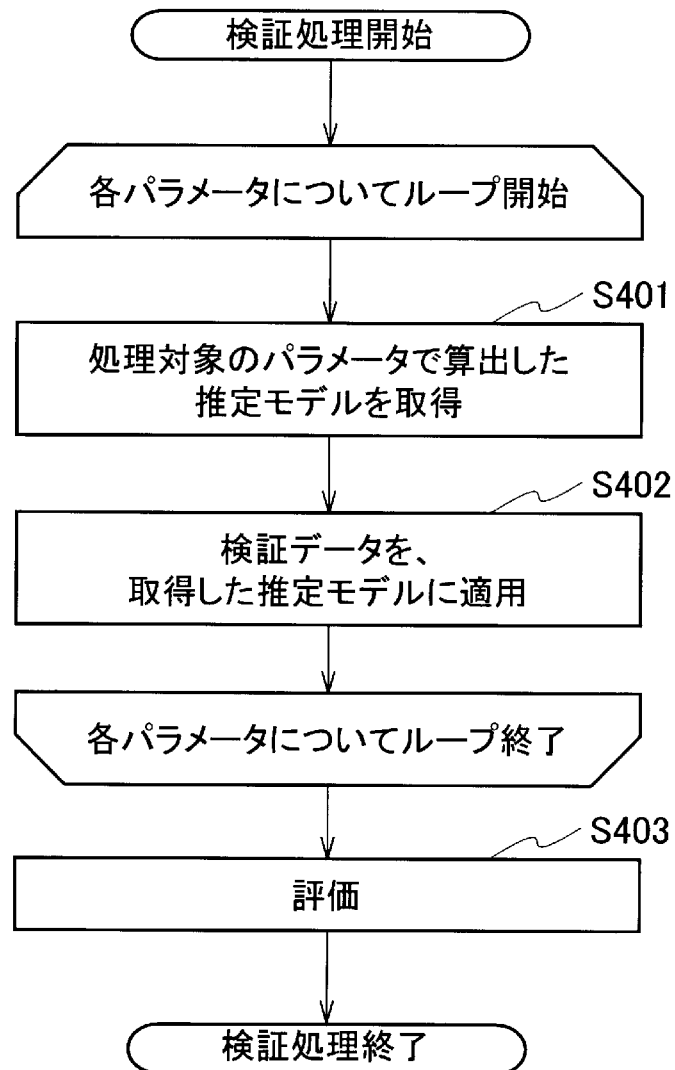
[図6]



[図7]



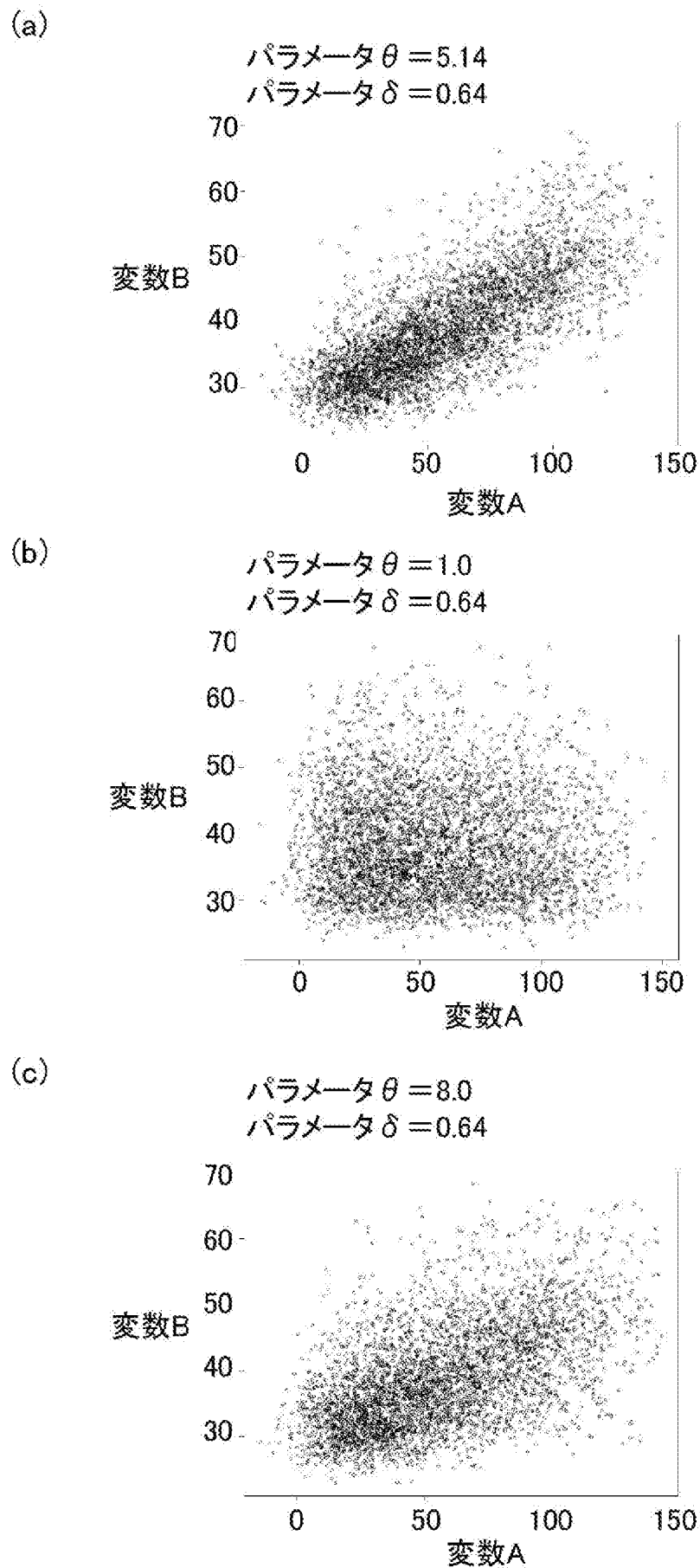
[図8]



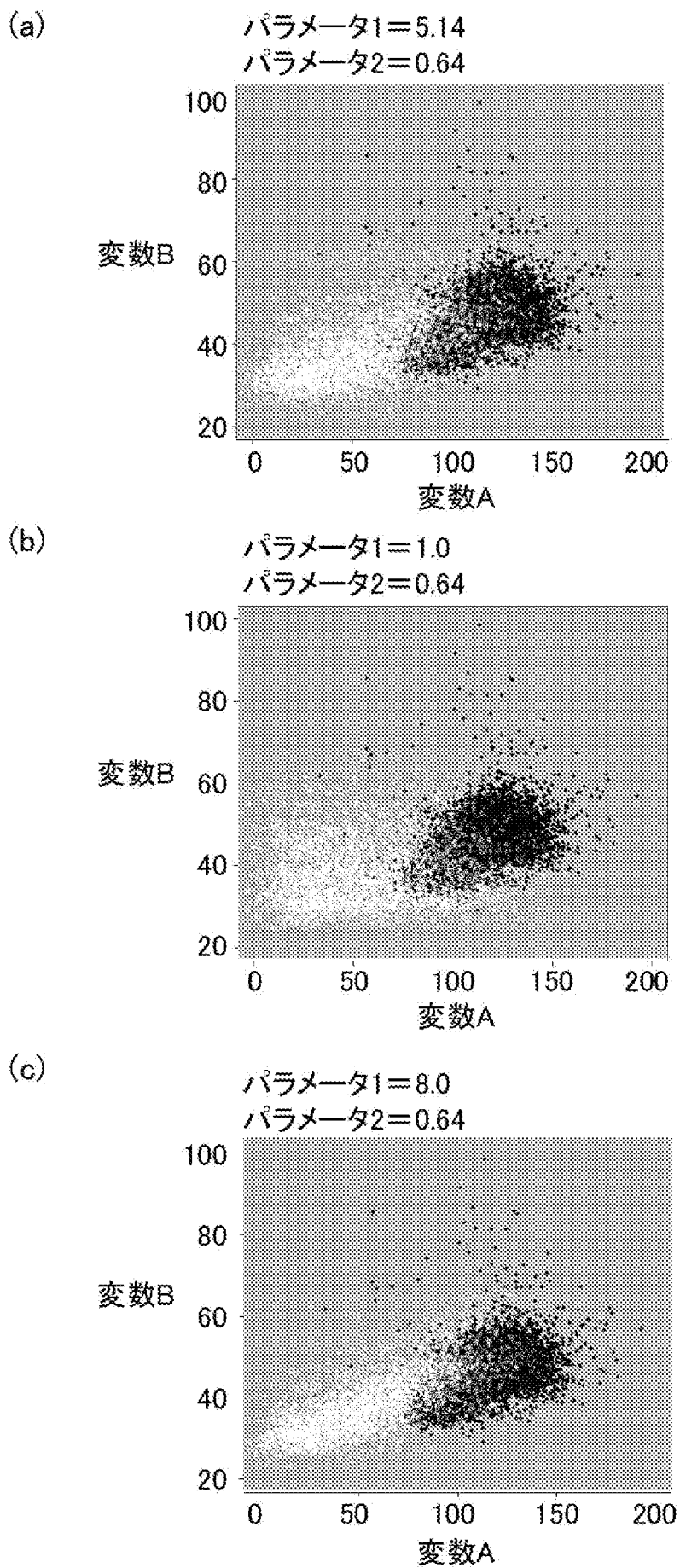
[図9]

	総データ セット数	未故障 データセット数	故障データ セット数
元データ	10000	9093	907
入力データ	5000	4564	436
検証データ	5000	4529	471

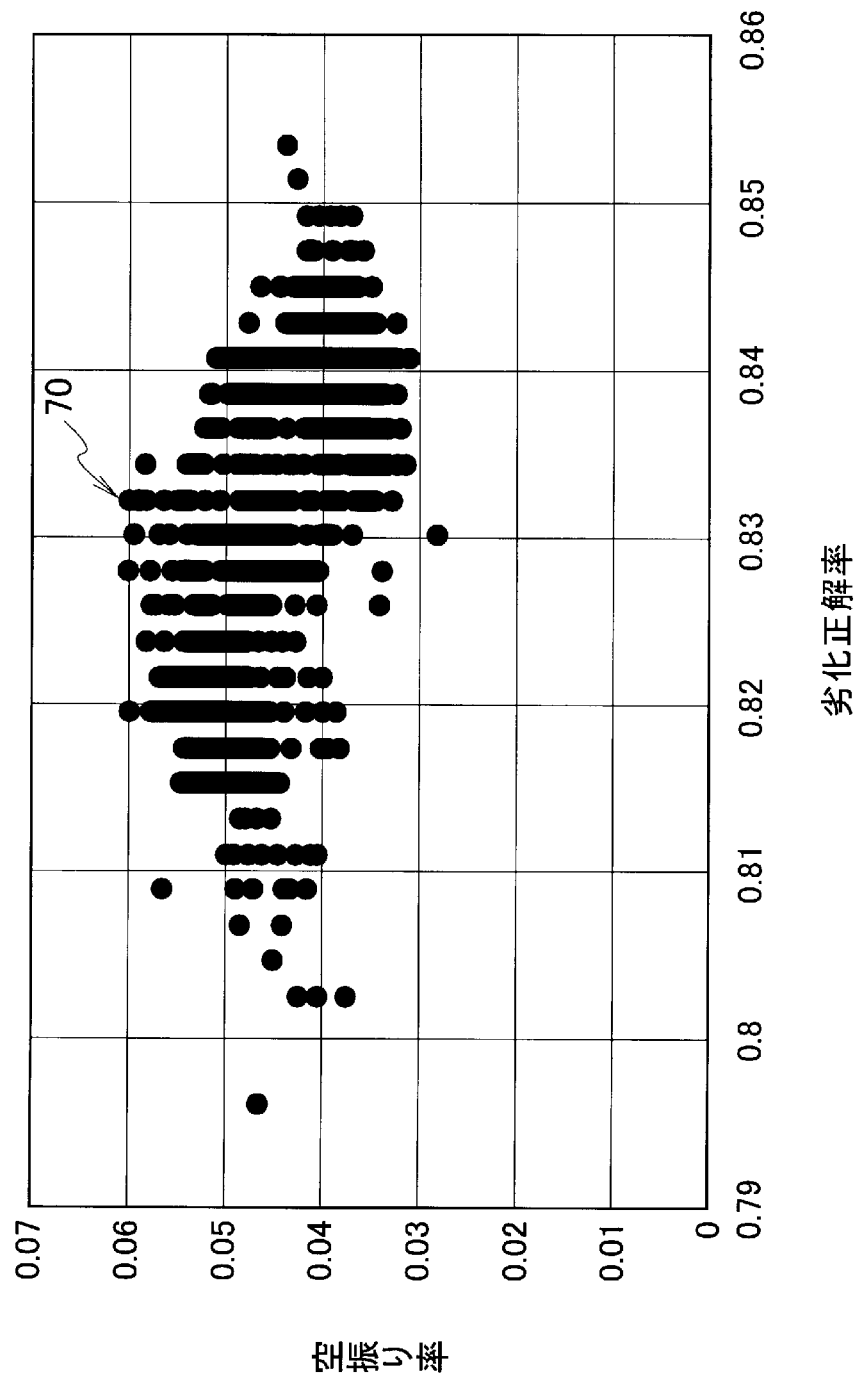
[図10]



[図11]



[図12]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2019/019547

A. CLASSIFICATION OF SUBJECT MATTER Int.Cl. G06N20/00 (2019.01) i		
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) Int.Cl. 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 1971-2019 Registered utility model specifications of Japan 1996-2019 Published registered utility model applications of Japan 1994-2019		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
C. DOCUMENTS CONSIDERED TO BE <u>RELEVANT</u>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2019/049688 A1 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) 14 March 2019, paragraphs [0114]-[0147] (Family: none)	1-7
A	US 2015/0088791 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 26 March 2015, paragraphs [0042]-[0052] (Family: none)	1-7
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> 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 "T" 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 12.06.2019		Date of mailing of the international search report 25.06.2019
Name and mailing address of the ISA/ Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan		Authorized officer Telephone No.

A. 発明の属する分野の分類（国際特許分類（I P C））

Int.Cl. G06N20/00 (2019.01) i

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（I P C））

Int.Cl. 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

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-2019年
日本国実用新案登録公報	1996-2019年
日本国登録実用新案公報	1994-2019年

国際調査で利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）

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	WO 2019/049688 A1（日本電信電話株式会社）2019.03.14, 段落[0114]-[0147]（ファミリーなし）	1-7
A	US 2015/0088791 A1（INTERNATIONAL BUSINESS MACHINES CORPORATION）2015.03.26, 段落[0042]-[0052]（ファミリーなし）	1-7

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。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的な技術水準を示すもの
「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
「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）
「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

「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矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「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

「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

「&」同一パテントファミリー文献

国際調査を完了した日

12.06.2019

国際調査報告の発送日

25.06.2019

国際調査機関の名称及びあて先

日本国特許庁（I S A/J P）

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官（権限のある職員）

多胡 滋

電話番号 03-3581-1101 内線 3545

5B

3562