

(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
2023年11月16日(16.11.2023)



(10) 国際公開番号

WO 2023/219037 A1

- (51) 国際特許分類:
G06Q 10/04 (2023.01) G16C 60/00 (2019.01)
- (21) 国際出願番号: PCT/JP2023/017086
- (22) 国際出願日: 2023年5月1日(01.05.2023)
- (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 2022-079269 2022年5月13日(13.05.2022) JP
- (71) 出願人: 株式会社レゾナック (RESONAC CORPORATION) [JP/JP]; 〒1057325 東京都港区東新橋一丁目9番1号 Tokyo (JP).
- (72) 発明者: 南 拓也(MINAMI, Takuya); 〒1058518 東京都港区芝大門一丁目13番9号 株式会社レ

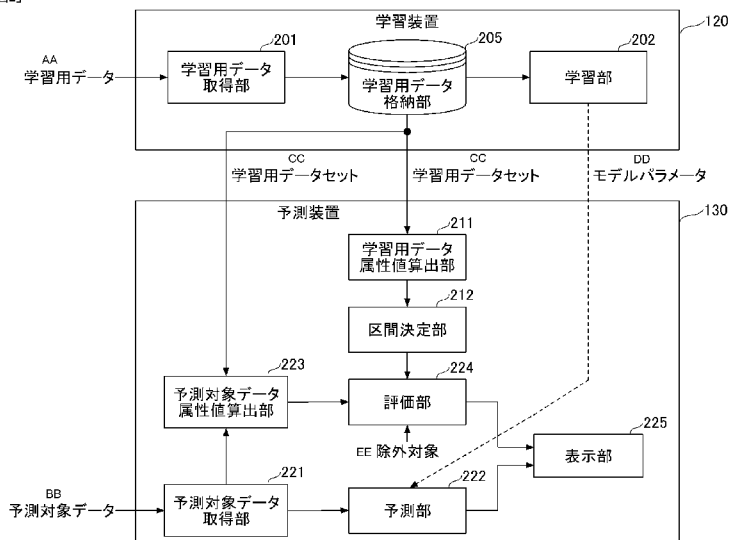
ゾナック内 Tokyo (JP). 藤森 貴大(FUJIMORI, Takahiro); 〒1058518 東京都港区芝大門一丁目13番9号 株式会社レゾナック内 Tokyo (JP). 李 海仁(LEE, Haein); 〒1058518 東京都港区芝大門一丁目13番9号 株式会社レゾナック内 Tokyo (JP). 奥野 好成(OKUNO, Yoshishige); 〒1058518 東京都港区芝大門一丁目13番9号 株式会社レゾナック内 Tokyo (JP).

- (74) 代理人: 伊東 忠重, 外(ITO, Tadashige et al.); 〒1000005 東京都千代田区丸の内二丁目1番1号 丸の内 M Y P L A Z A (明治安田生命ビル) 16階 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, CL, CN,

(54) Title: PREDICTION DEVICE, MATERIAL DESIGN SYSTEM, PREDICTION METHOD, AND PREDICTION PROGRAM

(54) 発明の名称: 予測装置、材料設計システム、予測方法及び予測プログラム

[図2]



- 120 Learning device
- 130 Prediction device
- 201 Learning data acquisition unit
- 202 Learning unit
- 205 Learning data storage unit
- 211 Learning data attribute value calculation unit
- 212 Segment determination unit
- 221 Predictive data acquisition unit
- 222 Prediction unit
- 223 Data subject to prediction attribute-value calculation unit
- 224 Evaluation unit
- 225 Display unit
- AA Learning data
- BB Data subject to prediction
- CC Learning data set
- DD Model parameter
- EE To be excluded

(57) Abstract: The present invention improves the efficiency of developing novel materials. This prediction device comprises: a segment determination unit for acquiring a learning data set to be used to generate a trained model, and determining, from a frequency distribution of attribute values calculated between a plurality of items of data included in the learning data set, a plurality of segments for classifying the attribute values; an evaluation unit for evaluating the suitability of data subject to prediction with respect to contradictory indicators by determining into which of the plurality of segments the attribute values calculated between the data subject to prediction and the plurality of items of data will be classified;

CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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))

and a display unit on which to display, in association with the result of evaluation by the evaluation unit, predicted values predicted by the trained model by inputting the data subject to prediction.

(57) 要約: 新規材料の開発効率を向上させる。予測装置は、学習済みモデルの生成に用いられる学習用データセットを取得し、前記学習用データセットに含まれる複数のデータ間で算出される属性値の度数分布から、該属性値を区分けするための複数の区間を決定する区間決定部と、予測対象のデータと、前記複数のデータとの間で算出される属性値が、前記複数の区間のいずれに区分けされるかを判定することで、相反する指標に対する前記予測対象のデータの適正を評価する評価部と、前記予測対象のデータを入力することで、前記学習済みモデルにより予測された予測値を、前記評価部による評価結果と対応付けて表示する表示部とを有する。

明 細 書

発明の名称：

予測装置、材料設計システム、予測方法及び予測プログラム

技術分野

[0001] 本開示は、予測装置、材料設計システム、予測方法及び予測プログラムに関する。

背景技術

[0002] 近年、新規材料の開発においては、例えば、機械学習により生成された学習済みの予測モデルを用いて、材料特性の予測を行う試み等がなされている。当該学習済みの予測モデルを用いることで、材料の生成、材料特性の検証実験等の工程を縮小することができ、新規材料の開発効率を向上させることが期待できるからである。

[0003] 一方で、機械学習により生成された学習済みの予測モデルの場合、予測対象データ（入力データ）が、当該機械学習の際に用いられた学習用データと著しく乖離していた場合、予測精度が低下することが知られている。

[0004] これに対して、例えば、下記非特許文献1等では、予測モデルの適用範囲（所望する予測精度が実現可能な予測対象データの範囲）を定義することが提案されている。

先行技術文献

非特許文献

[0005] 非特許文献1：Igor I. Baskin, Natalia Kireeva, and Alexandre Varnek, "The One-Class Classification Approach to Data Description and to Models Applicability Domain", NoI.Inf.2010, 29, 581p-587p

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、新規材料の開発において、学習用データに近い予測対象データについて材料特性の予測を行ったとしても、当該予測対象データは意外

性に乏しく、有益な新規材料となる可能性も低い。

[0007] 一方で、学習用データから著しく乖離した予測対象データの場合、そもそも予測精度が低いため、仮に所定の条件を満たす材料特性が予測されたとしても、実際に生成した際に予測したとおりの材料特性が得られることは少なく、有益な新規材料となる可能性も低い。

[0008] このようなことから、新規材料の開発においては、意外性と予測精度という相反する2つの指標のもとで予測モデルの適用範囲を定義することが、開発効率を向上させるうえで重要と考えられる。

[0009] 本開示は、新規材料の開発効率を向上させることを目的としている。

課題を解決するための手段

[0010] 本開示の第1の態様に係る予測装置は、

学習済みモデルの生成に用いられる学習用データセットを取得し、前記学習用データセットに含まれる複数のデータ間で算出される属性値の度数分布から、該属性値を区分けするための複数の区間を決定する区間決定部と、

予測対象のデータと、前記複数のデータとの間で算出される属性値が、前記複数の区間のいずれに区分けされるかを判定することで、相反する指標に対する前記予測対象のデータの適正を評価する評価部と、

前記予測対象のデータを入力することで、前記学習済みモデルにより予測された予測値を、前記評価部による評価結果と対応付けて表示する表示部とを有する。

[0011] 本開示の第2の態様は、第1の態様に記載の予測装置であって、

前記区間決定部は、前記複数のデータ間で算出される属性値について要約統計量を算出し、前記複数の区間を規定する、属性値の下限値または上限値を決定する。

[0012] 本開示の第3の態様は、第2の態様に記載の予測装置であって、

前記区間決定部は、互いに重複しない3つ以上の区間を決定する。

[0013] 本開示の第4の態様は、第3の態様に記載の予測装置であって、

前記評価部は、予測対象のデータと、前記複数のデータとの間で算出され

る属性値が、前記3つ以上の区間のうち、所定の要約統計量が含まれる区間に近い何番目の区間に分けられるかにより、前記予測対象のデータの適正を評価する。

[0014] 本開示の第5の態様は、第3の態様に記載の予測装置であって、
前記評価部は、予測対象のデータと、前記複数のデータとの間で算出される属性値が、前記3つ以上の区間のうち、所定の要約統計量が含まれる区間から最も離れた区間に分けされると判定した場合、前記予測対象のデータを、前記学習済みモデルに入力するデータから除外する。

[0015] 本開示の第6の態様は、第3の態様に記載の予測装置であって、
前記評価部は、予測対象のデータと、前記複数のデータとの間で算出される属性値が、前記3つ以上の区間のうち、所定の要約統計量が含まれる区間に近いR番目の区間に分けされると判定した場合、前記予測対象のデータを、前記学習済みモデルに入力するデータとして選択する。

[0016] 本開示の第7の態様は、第1の態様に記載の予測装置であって、
前記学習用データセットに含まれる複数のデータ間の距離を算出する学習用データ属性値算出部と、
前記複数のデータそれぞれについて算出された他のデータとの間の距離のうち、最小距離を抽出する学習用データ最小属性値抽出部と、を更に有し、
前記区間決定部は、抽出された前記最小距離の度数分布から、該最小距離を分けするための前記複数の区間を決定する。

[0017] 本開示の第8の態様は、第7の態様に記載の予測装置であって、
前記学習用データ属性値算出部は、前記学習用データセットに含まれるN個のデータ（Nは任意の整数）のうち、i番目（ $1 \leq i \leq N$ ）のデータと、i番目のデータを除く（ $N - 1$ ）個のデータとの間の距離をそれぞれ算出し、
前記学習用データ最小属性値抽出部は、前記i番目のデータについて算出された（ $N - 1$ ）個の距離の中から、最小距離を抽出する。

[0018] 本開示の第9の態様は、第7の態様に記載の予測装置であって、

前記予測対象のデータと、前記複数のデータとの間の距離を算出する予測対象データ属性値算出部と、

前記予測対象のデータと、前記複数のデータとの間の距離のうち、最小距離を抽出する予測対象データ最小属性値抽出部と、を更に有し、

前記評価部は、前記予測対象のデータについて抽出された前記最小距離が、前記複数の区間のうちのいずれに区分けされるかを判定することで、相反する指標に対する前記予測対象のデータの適正を評価する。

[0019] 本開示の第10の態様は、材料設計システムであって、

第1の態様に記載の予測装置と、

前記予測装置の前記評価部により、前記複数のデータとの間で算出された属性値が所定の区間に区分けされると判定された予測対象のデータであって、前記予測装置の前記学習済みモデルにより、所定の条件を満たす予測値が予測された予測対象のデータを選択して入力することで、材料設計データを生成する材料設計装置とを有する。

[0020] 本開示の第11の態様は、第10の態様に記載の材料設計システムであって、

学習用データセットに基づいて学習済みモデルを生成する学習装置を更に有し、

前記予測装置は、前記学習装置により生成された前記学習済みモデルに、前記予測対象のデータを入力することで前記予測値を予測する。

[0021] 本開示の第12の態様は、予測方法であって、

学習済みモデルの生成に用いられる学習用データセットを取得し、前記学習用データセットに含まれる複数のデータ間で算出される属性値の度数分布から、該属性値を区分けするための複数の区間を決定する区間決定工程と、

予測対象のデータと、前記複数のデータとの間で算出される属性値が、前記複数の区間のいずれに区分けされるかを判定することで、相反する指標に対する前記予測対象のデータの適正を評価する評価工程と、

前記予測対象のデータを入力することで、前記学習済みモデルにより予測

された予測値を、前記評価工程における評価結果と対応付けて表示する表示工程と

をコンピュータが実行する。

[0022] 本開示の第13の態様は、予測プログラムであって、

学習済みモデルの生成に用いられる学習用データセットを取得し、前記学習用データセットに含まれる複数のデータ間で算出される属性値の度数分布から、該属性値を区分けするための複数の区間を決定する区間決定工程と、

予測対象のデータと、前記複数のデータとの間で算出される属性値が、前記複数の区間のいずれに区分けされるかを判定することで、相反する指標に対する前記予測対象のデータの適正を評価する評価工程と、

前記予測対象のデータを入力することで、前記学習済みモデルにより予測された予測値を、前記評価工程における評価結果と対応付けて表示する表示工程と

をコンピュータに実行させる。

発明の効果

[0023] 本開示によれば、新規材料の開発効率を向上させることができる。

図面の簡単な説明

[0024] [図1]図1は、材料設計システムのシステム構成の一例を示す図である。

[図2]図2は、学習装置及び予測装置の機能構成の一例を示す図である。

[図3]図3は、学習装置及び予測装置のハードウェア構成の一例を示す図である。

[図4]図4は、学習装置の学習部による処理の具体例を示す図である。

[図5]図5は、予測装置の学習用データ属性値算出部による処理の具体例を示す図である。

[図6]図6は、予測装置の区間決定部による処理の具体例を示す図である。

[図7]図7は、予測装置の予測対象データ属性値算出部による処理の具体例を示す図である。

[図8]図8は、予測装置の評価部による処理の具体例を示す図である。

[図9]図9は、予測装置の予測部による処理の具体例を示す図である。

[図10]図10は、予測装置の表示部による処理の具体例を示す図である。

[図11]図11は、学習処理の流れを示すフローチャートである。

[図12]図12は、予測処理の流れを示すフローチャートである。

[図13]図13は、実施例1を示す図である。

[図14]図14は、実施例2を示す図である。

[図15]図15は、実施例3を示す図である。

[図16]図16は、予測特性の検証例を示す図である。

発明を実施するための形態

[0025] 以下、各実施形態について添付の図面を参照しながら説明する。なお、本明細書及び図面において、実質的に同一の機能構成を有する構成要素については、同一の符号を付することにより重複した説明を省略する。

[0026] [第1の実施形態]

<材料設計システムのシステム構成>

はじめに、第1の実施形態に係る材料設計システムのシステム構成について説明する。図1は、材料設計システムのシステム構成の一例を示す図である。図1に示すように、材料設計システム100は、材料特性実験装置110、学習装置120、予測装置130、材料設計装置140、材料生成装置150、材料特性実験装置160を有する。

[0027] 材料特性実験装置110は、“材料データ”が既知の構造式である材料についての“材料特性”の特性値を、実験によって求めるための装置である。実験者111が当該材料特性実験装置110を用いて求めた各材料の特性値は、対応する構造式とともに、学習用データとして、学習装置120に入力される。なお、学習用データは、公知のデータベースの情報を元に構成されてもよい。

[0028] 学習装置120は、取得した複数の学習用データを含む学習用データセットを用いて、予測モデル（材料の構造式から、当該材料の特性値を予測するモデル）について機械学習を行い、学習済みの予測モデルを生成する。また

、学習装置120は、学習済みの予測モデルを生成するのに用いた学習用データセットと、学習済みの予測モデルのモデルパラメータとを、予測装置130に通知する。

[0029] 予測装置130は、学習装置120より通知されたモデルパラメータが設定された学習済みの予測モデルを有し、設計者131により入力された予測対象データの予測特性（予測値の一例）を予測する。

[0030] また、予測装置130は、学習装置120より通知された学習用データセットを解析し、各学習用データの材料データが、他の学習用データの材料データからどの程度乖離しているかを定量的に算出することで、材料データ間の属性値を算出する。

[0031] また、予測装置130は、算出した材料データ間の属性値の度数分布を算出することで、該属性値を区分けするための複数の区間（具体的には、それぞれの区間を規定する属性値の下限値、上限値）を決定する。

[0032] また、予測装置130は、予測対象データの材料データが、各学習用データの材料データからどの程度乖離しているかを定量的に算出することで、予測対象データの材料データの属性値を算出する。更に、予測装置130は、予測対象データの材料データの属性値が、決定した複数の区間のいずれに区分けされるかを判定することで、予測対象データの適正を評価し、評価結果を予測特性と対応付けて出力する。

[0033] なお、ここでいう予測対象データの適正とは、新規材料の開発における、意外性と予測精度という相反する2つの指標に対する予測対象データの適正を指す。予測対象データの適正が高いとは、学習用データに対する予測対象データの意外性と、予測対象データの予測特性についての予測精度との間でバランスがとれていることを指す（本実施形態では、当該適正を「ランク」として表す（詳細は後述））。

[0034] 予測装置130により出力された、予測対象データの予測特性及び予測対象データの評価結果は、設計者131に報知される。これにより、設計者131は、予測対象データの予測特性を把握することができるとともに、当該

予測対象データに基づいて生成される材料が、有益な新規材料となる可能性が高いか低いかを把握することができる。

[0035] 材料設計装置 140 は、材料設計データを生成する装置である。設計者 131 は、有益な新規材料となる可能性が高い予測対象データを選択して、材料設計装置 140 に入力することで、材料設計データを生成する。

[0036] 材料生成装置 150 は、生成された材料設計データに基づいて、実際に材料を生成するための装置である。

[0037] 材料特性実験装置 160 は、材料生成装置 150 により実際に生成された新規材料の材料特性を、検証実験により求めるための装置である。実験者 161 が当該材料特性実験装置 160 を用いて求めた当該新規材料の材料特性は、設計者 131 に通知される。

[0038] <学習装置及び予測装置の機能構成>

次に、材料設計システム 100 を構成する各装置のうち、学習装置 120 及び予測装置 130 の機能構成について図 2 を参照しながら説明する。図 2 は、学習装置及び予測装置の機能構成の一例を示す図である。

[0039] 学習装置 120 には、学習プログラムがインストールされており、当該プログラムが実行されることで、学習装置 120 は、学習用データ取得部 201、学習部 202 として機能する。

[0040] 学習用データ取得部 201 は、実験者 111 により入力された学習用データを取得し、学習用データ格納部 205 に格納する。

[0041] 学習部 202 は、学習用データ格納部 205 に格納された複数の学習用データを含む学習用データセットを用いて、予測モデルに対して機械学習を行い、学習済みの予測モデルを生成する。学習部 202 により生成された学習済みの予測モデルのモデルパラメータは、学習済みの予測モデルの生成に用いられた学習用データセットとともに、予測装置 130 に通知される。

[0042] 予測装置 130 には、予測プログラムがインストールされている。予測装置 130 は、当該プログラムが実行されることで、学習用データ属性値算出部 211、区間決定部 212、予測対象データ取得部 221、予測部 222

、予測対象データ属性値算出部 2 2 3、評価部 2 2 4、表示部 2 2 5として機能する。

[0043] 学習用データ属性値算出部 2 1 1は、学習装置 1 2 0より通知された学習用データセットを解析し、各学習用データの材料データが、他の学習用データの材料データからどの程度乖離しているかを定量的に算出することで、材料データ間の属性値を算出する。

[0044] 区間決定部 2 1 2は、学習用データ属性値算出部 2 1 1により算出された材料データ間の属性値の度数分布を算出することで、該属性値を区分けするための複数の区間（具体的には、それぞれの区間を規定する属性値の下限值、上限値）を決定する。

[0045] 予測対象データ取得部 2 2 1は、設計者 1 3 1により入力された予測対象データを取得し、予測部 2 2 2及び予測対象データ属性値算出部 2 2 3に通知する。

[0046] 予測部 2 2 2は、学習部 2 0 2より通知されたモデルパラメータが設定された学習済みの予測モデルを有し、設計者 1 3 1により入力された予測対象データの予測特性を予測する。

[0047] 予測対象データ属性値算出部 2 2 3は、学習装置 1 2 0より通知された学習用データセットの各学習用データを取得するとともに、予測対象データ取得部 2 2 1より通知された予測対象データを取得する。

[0048] また、予測対象データ属性値算出部 2 2 3は、予測対象データの材料データが、各学習用データの材料データからどの程度乖離しているかを定量的に算出することで、予測対象データの材料データの属性値を算出する。

[0049] 評価部 2 2 4は、予測対象データ属性値算出部 2 2 3により算出された予測対象データの材料データの属性値が、区間決定部 2 1 2により決定された複数の区間のうちのいずれに区分けされるかを判定する。これにより、評価部 2 2 4は、予測対象データの適正を評価し、評価結果を表示部 2 2 5に通知する。

[0050] なお、評価部 2 2 4には予め除外対象が設定されており、算出された属性

値が所定の区間に区分けされた場合、当該予測対象データは、予測部 2 2 2 による予測対象から除外する（つまり、予測部 2 2 2 によって学習済みの予測モデルに入力されることはない）。あるいは、評価部 2 2 4 は、算出された属性値が所定の区間に区分けされた場合、当該予測対象データは、表示部 2 2 5 による表示対象から除外する（つまり、表示部 2 2 5 によって表示されることはない）。

[0051] 表示部 2 2 5 は、予測部 2 2 2 により予測された、予測対象データの予測特性と、評価部 2 2 4 により評価された予測対象データの評価結果とを対応付けて表示する。

[0052] このように、第 1 の実施形態に係る材料設計システム 1 0 0 は、

- ・ 予測装置 1 3 0 を有し、予測対象データの予測特性を予測する。これにより、第 1 の実施形態によれば、所定の条件を満たす予測特性が予測されなかった予測対象データを、材料の生成及び検証実験の対象から排除することができる。
- ・ 予測装置 1 3 0 を有し、予測対象データの評価結果を出力する。これにより、第 1 の実施形態によれば、予測装置 1 3 0 によって所定の条件を満たす予測特性が予測された場合であっても、有益な新規材料となる可能性が低い予測対象データを、材料の生成及び検証実験の対象から排除することができる。

[0053] この結果、第 1 の実施形態に係る材料設計システム 1 0 0 によれば、材料の生成及び検証実験を行う予測対象データを絞り込むことが可能となり、新規材料の開発効率を向上させることができる。

[0054] <学習装置及び予測装置のハードウェア構成>

次に、学習装置 1 2 0 及び予測装置 1 3 0 のハードウェア構成について説明する。なお、学習装置 1 2 0 及び予測装置 1 3 0 は、同様のハードウェア構成を有するため、ここでは、図 3 を用いて、学習装置 1 2 0 及び予測装置 1 3 0 のハードウェア構成をまとめて説明する。

[0055] 図 3 は、学習装置及び予測装置のハードウェア構成の一例を示す図である

。図3に示すように、学習装置120及び予測装置130は、プロセッサ301、メモリ302、補助記憶装置303、I/F (Interface) 装置304、通信装置305、ドライブ装置306を有する。なお、学習装置120及び予測装置130の各ハードウェアは、バス307を介して相互に接続されている。

[0056] プロセッサ301は、CPU (Central Processing Unit)、GPU (Graphics Processing Unit) 等の各種演算デバイスを有する。プロセッサ301は、各種プログラム (例えば、学習プログラム、予測プログラム等) をメモリ302上に読み出して実行する。

[0057] メモリ302は、ROM (Read Only Memory)、RAM (Random Access Memory) 等の主記憶デバイスを有する。プロセッサ301とメモリ302とは、いわゆるコンピュータを形成し、プロセッサ301が、メモリ302上に読み出した各種プログラムを実行することで、当該コンピュータは上記各種機能を実現する。

[0058] 補助記憶装置303は、各種プログラムや、各種プログラムがプロセッサ301によって実行される際に用いられる各種データを格納する。例えば、学習用データ格納部205は、補助記憶装置303において実現される。

[0059] I/F装置304は、ユーザインタフェース装置の一例である操作装置311、表示装置312と接続する接続デバイスである。通信装置305は、ネットワークを介して外部装置 (不図示) と通信するための通信デバイスである。

[0060] ドライブ装置306は記録媒体313をセットするためのデバイスである。ここでいう記録媒体313には、CD-ROM、フレキシブルディスク、光磁気ディスク等のように情報を光学的、電気的あるいは磁氣的に記録する媒体が含まれる。また、記録媒体313には、ROM、フラッシュメモリ等のように情報を電気的に記録する半導体メモリ等が含まれていてもよい。

[0061] なお、補助記憶装置303にインストールされる各種プログラムは、例えば、配布された記録媒体313がドライブ装置306にセットされ、該記録

媒体 3 1 3 に記録された各種プログラムがドライブ装置 3 0 6 により読み出されることでインストールされる。あるいは、補助記憶装置 3 0 3 にインストールされる各種プログラムは、通信装置 3 0 5 を介してネットワークからダウンロードされることで、インストールされてもよい。

[0062] <学習装置の各部の処理の具体例>

次に、学習装置 1 2 0 の各部（ここでは、学習部 2 0 2）の処理の具体例について説明する。

[0063] （１）学習部 2 0 2 による処理の具体例

はじめに、学習装置 1 2 0 の学習部 2 0 2 による処理の具体例について説明する。図 4 は、学習装置の学習部による処理の具体例を示す図である。図 4 に示すように、学習部 2 0 2 は、入力部 4 0 1、予測モデル 4 0 2、比較／変更部 4 0 3 を有する。

[0064] なお、図 4 において学習用データセット 4 0 0 は、情報の項目として、“データ番号”、“材料データ”、“材料特性”を含む。“データ番号”には、各学習用データを識別するための番号が格納され、“材料データ”には、各材料の構造式が格納され、“材料特性”には、各材料の特性値が格納される。図 4 の例は、N 個の学習用データ（N は任意の整数）が格納されていることを示している。

[0065] 入力部 4 0 1 は、学習用データセット 4 0 0 より各学習用データの“材料データ”に含まれる材料の構造式を読み出し、予測モデルに入力する。

[0066] 予測モデル 4 0 2 は、各学習用データの“材料データ”に含まれる材料の構造式が入力されることで、材料特性に相当する出力データを出力する。

[0067] 比較／変更部 4 0 3 は、予測モデル 4 0 2 より出力された出力データが、各学習用データの“材料特性”に含まれる材料の特性値に近づくように、予測モデル 4 0 2 のモデルパラメータを更新する。

[0068] これにより、学習部 2 0 2 では、予測対象データの材料データに基づいて、当該予測対象データの材料特性を予測可能な学習済みの予測モデルを生成することができる。

[0069] <予測装置の各部の処理の具体例>

次に、予測装置130の各部（ここでは、学習用データ属性値算出部211、区間決定部212、予測対象データ属性値算出部223、評価部224、予測部222、表示部225）の処理の具体例について説明する。

[0070] (1) 学習用データ属性値算出部211による処理の具体例

はじめに、学習用データ属性値算出部211による処理の具体例について説明する。図5は、予測装置の学習用データ属性値算出部による処理の具体例を示す図である。図5に示すように、学習用データ属性値算出部211は、相互属性値算出部501、学習用データ最小属性値抽出部502を有する。

[0071] 相互属性値算出部501は、N個の学習用データを含む学習用データセット400の”データ番号”及び”材料データ”から、各学習用データを識別するための番号及び各材料の構造式を読み出す。

[0072] また、相互属性値算出部501は、読み出したN個の学習用データの”材料データ”に含まれる材料の構造式のうち、

- ・ i 番目 ($1 \leq i \leq N$) の学習用データの”材料データ”に含まれる材料の構造式と、

- ・ i 番目の学習用データを除く ($N - 1$) 個の他の学習用データの”材料データ”に含まれる材料の構造式と、

の組み合わせを生成し、全ての組み合わせについて、材料データ間の属性値を算出する。ここでいう材料データ間の属性値とは、i 番目の学習用データの”材料データ”に含まれる材料の構造式と、i 番目の学習用データを除く他の学習用データの”材料データ”に含まれる材料の構造式とが、どの程度乖離しているかを示す値を指す。具体的には、i 番目の学習用データの”材料データ”に含まれる材料の構造式と、i 番目の学習用データを除く他の学習用データの”材料データ”に含まれる材料の構造式との間の距離を指す。あるいは、i 番目の学習用データの”材料データ”に含まれる材料の構造式と、i 番目の学習用データを除く他の学習用データの”材料データ”に含まれる材料の構造式との間の距離を指す。あるいは、i 番目の学習用データの”材料データ”に含まれる材料の構造式と、i 番目の学習用データを除く他の学習用データの”材料データ”に含まれる材料の構造式との間の距離を指す。

の組み合わせを生成し、全ての組み合わせについて、材料データ間の属性値を算出する。ここでいう材料データ間の属性値とは、i 番目の学習用データの”材料データ”に含まれる材料の構造式と、i 番目の学習用データを除く他の学習用データの”材料データ”に含まれる材料の構造式とが、どの程度乖離しているかを示す値を指す。具体的には、i 番目の学習用データの”材料データ”に含まれる材料の構造式と、i 番目の学習用データを除く他の学習用データの”材料データ”に含まれる材料の構造式との間の距離を指す。あるいは、i 番目の学習用データの”材料データ”に含まれる材料の構造式と、i 番目の学習用データを除く他の学習用データの”材料データ”に含まれる材料の構造式との間の距離を指す。

の組み合わせを生成し、全ての組み合わせについて、材料データ間の属性値を算出する。ここでいう材料データ間の属性値とは、i 番目の学習用データの”材料データ”に含まれる材料の構造式と、i 番目の学習用データを除く他の学習用データの”材料データ”に含まれる材料の構造式とが、どの程度乖離しているかを示す値を指す。具体的には、i 番目の学習用データの”材料データ”に含まれる材料の構造式と、i 番目の学習用データを除く他の学習用データの”材料データ”に含まれる材料の構造式との間の距離を指す。あるいは、i 番目の学習用データの”材料データ”に含まれる材料の構造式と、i 番目の学習用データを除く他の学習用データの”材料データ”に含まれる材料の構造式との間の距離を指す。

の組み合わせを生成し、全ての組み合わせについて、材料データ間の属性値を算出する。ここでいう材料データ間の属性値とは、i 番目の学習用データの”材料データ”に含まれる材料の構造式と、i 番目の学習用データを除く他の学習用データの”材料データ”に含まれる材料の構造式とが、どの程度乖離しているかを示す値を指す。具体的には、i 番目の学習用データの”材料データ”に含まれる材料の構造式と、i 番目の学習用データを除く他の学習用データの”材料データ”に含まれる材料の構造式との間の距離を指す。あるいは、i 番目の学習用データの”材料データ”に含まれる材料の構造式と、i 番目の学習用データを除く他の学習用データの”材料データ”に含まれる材料の構造式との間の距離を指す。

の組み合わせを生成し、全ての組み合わせについて、材料データ間の属性値を算出する。ここでいう材料データ間の属性値とは、i 番目の学習用データの”材料データ”に含まれる材料の構造式と、i 番目の学習用データを除く他の学習用データの”材料データ”に含まれる材料の構造式とが、どの程度乖離しているかを示す値を指す。具体的には、i 番目の学習用データの”材料データ”に含まれる材料の構造式と、i 番目の学習用データを除く他の学習用データの”材料データ”に含まれる材料の構造式との間の距離を指す。あるいは、i 番目の学習用データの”材料データ”に含まれる材料の構造式と、i 番目の学習用データを除く他の学習用データの”材料データ”に含まれる材料の構造式との間の距離を指す。

式との間の密度等を指す。

[0073] 図5において、表511は、横軸及び縦軸に、各学習用データを識別するための番号をそれぞれ配した表である。表511には、学習用データセット400に含まれる各学習用データの材料データの全ての組み合わせが示されている。

[0074] また、表511において、横軸に示した番号と縦軸に示した番号とが交差する領域には、対応する学習用データの材料データの組み合わせについて算出された、材料データ間の属性値が示されている。

[0075] 例えば、横軸のデータ番号="2"と、縦軸のデータ番号="1"とが交差する領域には、

- ・データ番号="2"の学習用データの材料データ="構造式2"と、
- ・データ番号="1"の学習用データの材料データ="構造式1"と、

の組み合わせについて算出された、材料データ間の属性値(=0.1)が格納される。

[0076] 学習用データ最小属性値抽出部502は、相互属性値算出部501により算出された、全ての組み合わせについての材料データ間の属性値から、材料ごとに、最小の属性値(例えば、最小距離)を抽出する。

[0077] 図5において、表512は、各学習用データを識別するための番号に対応付けて、学習用データ最小属性値抽出部502により材料ごとに抽出された最小の属性値を示したものである。

[0078] 例えば、データ番号="1"により識別される学習用データの場合、当該材料の最小の属性値は、"0.1"となる。

[0079] なお、学習用データ最小属性値抽出部502により抽出された各材料の最小の属性値は、区間決定部212に通知される。

[0080] (2) 区間決定部212による処理の具体例

次に、区間決定部212による処理の具体例について説明する。図6は、予測装置の区間決定部による処理の具体例を示す図である。図6に示すように、区間決定部212は、度数分布生成部601、区間算出部602を有す

る。

[0081] 度数分布生成部601は、学習用データ最小属性値抽出部502より通知された、各材料の最小の属性値を取得し、度数分布を生成する。図6において、符号611は、度数分布生成部601により生成された度数分布の一例であり、横軸は属性値を、縦軸はデータ数をそれぞれ表している。

[0082] 区間算出部602は、度数分布生成部601により取得された、各材料の最小の属性値に基づいて、要約統計量を算出する。具体的には、区間算出部602は、

- ・各材料の最小の属性値の中央値、
- ・各材料の最小の属性値の第一四分位 (Q_{25})、
- ・各材料の最小の属性値の第三四分位 (Q_{75})、
- ・各材料の最小の属性値の四分位範囲 ($IQR = Q_{75} - Q_{25}$)、

を算出する。また、区間算出部602は、算出した上記要約統計量に基づいて、各属性値を区分けするための複数の区間を決定する際の、係数 (α_1 、 α_2) を決定する。

[0083] 図6の例は、係数 $\alpha_1 = "1.5"$ 、係数 $\alpha_2 = "3.0"$ と決定し、度数分布生成部601により生成された度数分布 (符号611) に対して、

- ・「属性値 < (第三四分位 (Q_{75}))」となる属性値が区分けされる区間の上限値、
 - ・「(第三四分位 (Q_{75})) \leq 属性値 < (四分位範囲 (IQR) を α_1 倍した値を第三四分位 (Q_{75}) に加算した値)」となる属性値が区分けされる区間の下限値、上限値、
 - ・「(四分位範囲 (IQR) を α_1 倍した値を第三四分位 (Q_{75}) に加算した値) \leq 属性値 < (四分位範囲 (IQR) を α_2 倍した値を第三四分位 (Q_{75}) に加算した値)」となる属性値が区分けされる区間の下限値、上限値、
 - ・「(四分位範囲 (IQR) を α_2 倍した値を第三四分位 (Q_{75}) に加算した値) \leq 属性値」となる属性値が区分けされる区間の下限値、
- を決定した様子を示している。

[0084] また、図6の例では、

- ・ 所定の要約統計量（ここでは、「中央値」）が含まれる区間を、「ランク1」、
 - ・ 所定の要約統計量が含まれる区間に近い1番目の区間を、「ランク2」、
 - ・ 所定の要約統計量が含まれる区間に近い2番目の区間を、「ランク3」、
 - ・ 所定の要約統計量が含まれる区間から最も離れた区間を、「ランク4」、
- と規定している。

[0085] ここで、上記ランクと新規材料の開発における相反する2つの指標（意外性、予測精度）との関係について考察すると、

- ・ 予測精度は、「ランク1」→「ランク2」→「ランク3」→「ランク4」の順に低下する。つまり、ランクは、予測精度の高さを表している。
- ・ 意外性は、「ランク1」→「ランク2」→「ランク3」→「ランク4」の順に上がる。つまり、ランクは、意外性の低さを表している。

[0086] このように、各ランクは、新規材料の開発における相反する2つの指標（意外性、予測精度）を同時に表しているといえる。

[0087] （3）予測対象データ属性値算出部223による処理の具体例

次に、予測対象データ属性値算出部223による処理の具体例について説明する。図7は、予測装置の予測対象データ属性値算出部による処理の具体例を示す図である。図7に示すように、予測対象データ属性値算出部223は、相互属性値算出部701、予測対象データ最小属性値抽出部702を有する。

[0088] 相互属性値算出部701は、学習装置120より通知されたN個の学習用データを含む学習用データセット400の”データ番号”及び”材料データ”から、各学習用データを識別するための番号及び各材料の構造式を読み出す。また、相互属性値算出部701は、予測対象データ取得部221より通知された予測対象データ700の”データ番号”及び”材料データ”から、予測対象データを識別するための番号及び材料の構造式を読み出す。

[0089] また、相互属性値算出部701は、読み出したN個の学習用データの”材料

データ”に含まれる材料の構造式のうち、

- ・ i 番目 ($1 \leq i \leq N$) の学習用データの”材料データ”に含まれる材料の構造式と、

- ・ 予測対象データ 700 の”材料データ”に含まれる材料の構造式と、

の組み合わせを生成し、全ての組み合わせについて、材料データ間の属性値を算出する。ここでいう材料データ間の属性値とは、 i 番目の学習用データの”材料データ”に含まれる材料の構造式と、予測対象データ 700 の”材料データ”に含まれる材料の構造式とが、どの程度乖離しているかを示す値を指す。具体的には、 i 番目の学習用データの”材料データ”に含まれる材料の構造式と、予測対象データ 700 の”材料データ”に含まれる材料の構造式との間の距離を指す。あるいは、 i 番目の学習用データの”材料データ”に含まれる材料の構造式と、予測対象データ 700 の”材料データ”に含まれる材料の構造式との間の密度等を指す。

[0090] 図 7 において、表 7 1 1 は、横軸に学習用データセット 400 に含まれる各学習用データを識別するための番号を、縦軸に予測対象データを識別するための番号をそれぞれ配した表である。表 7 1 1 には、学習用データセット 400 に含まれる各学習用データの材料データと、予測対象データの材料データとの間の全ての組み合わせが示されている。

[0091] また、表 7 1 1 において、横軸に示した番号と縦軸に示した番号とが交差する領域には、対応する学習用データの材料データと、対応する予測対象データの材料データとの組み合わせについて算出された材料データ間の属性値が示されている。

[0092] 例えば、横軸のデータ番号＝”2”と、縦軸のデータ番号＝”X”とが交差する領域には、

- ・ データ番号＝”2”の学習用データの材料データ＝”構造式 2”と、

- ・ データ番号＝”X”の予測対象データの材料データ＝”構造式 X”と、

の組み合わせについて算出された、材料データ間の属性値 ($= 0, 1$) が格納される。

[0093] 予測対象データ最小属性値抽出部702は、相互属性値算出部701により算出された、全ての組み合わせについての材料データ間の属性値から、最小の属性値を抽出する。

[0094] 図7において、表712は、予測対象データを識別するための番号に対応付けて、予測対象データ最小属性値抽出部702により抽出された最小の属性値を示したものである。図7の例は、最小の属性値が、“0.1”であったことを示している。

[0095] なお、予測対象データ最小属性値抽出部702により抽出された最小の属性値は、評価部224に通知される。

[0096] (4) 評価部224による処理の具体例

次に、評価部224による処理の具体例について説明する。図8は、予測装置の評価部による処理の具体例を示す図である。図8に示すように、評価部224は、属性値取得部801、評価結果出力部802を有する。

[0097] 属性値取得部801は、予測対象データ最小属性値抽出部702より通知された最小の属性値を取得し、区間決定部212より通知された複数の区間のいずれに区分けされるかを判定する。図8の例は、予測対象データ700の“材料データ”に含まれる材料の構造式（構造式X）について算出された最小の属性値が、所定の要約統計量（中央値）が含まれる区間に近いR番目の区間（ $R=2$ ）に区分けされたと判定された様子を示している。

[0098] 評価結果出力部802は、属性値取得部801により判定された区間が、R番目の区間（ $R=2$ ）であることから、新規材料の開発における予測対象データ700の“材料データ”に含まれる材料の構造式（構造式X）の適正について、

- ・意外性の低さがR番目（ $R=2$ ）、
- ・予測精度の高さがR番目（ $R=2$ ）、

である（ランク2である）と評価する。つまり、評価結果出力部802は、予測対象データの材料データの属性値が、所定の要約統計量（中央値）が含まれる区間に近い何番目の区間に区分けされるかにより、予測対象データの

適正（ランク）を評価する。

[0099] また、評価結果出力部802は、評価結果（ランク）を表示部225に通知する。

[0100] 更に、評価結果出力部802は、予測対象データ700の”材料データ”に含まれる材料の構造式（構造式X）について算出された最小の属性値が、所定の区間に分けられたか否かを判定する。具体的には、予測対象データ700の”材料データ”に含まれる材料の構造式（構造式X）の適正についての評価結果（ランク）が、除外対象に該当するか否かを判定する。評価結果出力部802は、除外対象に該当すると判定した場合には、予測部222及び／または表示部225に通知する。

[0101] （5）予測部222による処理の具体例

次に、予測部222による処理の具体例について説明する。図9は、予測装置の予測部による処理の具体例を示す図である。図9に示すように、予測部222は、予測対象データ入力部901、学習済み予測モデル902、予測特性出力部903を有する。

[0102] 予測対象データ入力部901は、予測対象データ取得部221より予測対象データ700が通知されると、予測対象データ700の”材料データ”に含まれる材料の構造式（構造式X）を、学習済み予測モデル902に入力する。

[0103] なお、予測対象データ入力部901は、評価結果出力部802より、除外対象に該当するとの判定結果が通知されていた場合、予測対象データ700の”材料データ”に含まれる材料の構造式（構造式X）を、入力しないように制御してもよい。

[0104] 学習済み予測モデル902は、学習部202により学習処理が行われることで算出されたモデルパラメータが設定された学習済みの予測モデルである。予測対象データ700の”材料データ”に含まれる材料の構造式（構造式X）が入力されると、学習済み予測モデル902は、予測対象データ700の予測特性を予測する。

[0105] 予測特性出力部903は、学習済み予測モデル902により予測された、予測対象データ700の予測特性を、表示部225に通知する。

[0106] (6) 表示部225による処理の具体例

次に、表示部225による処理の具体例について説明する。図10は、予測装置の表示部による処理の具体例を示す図である。図10に示すように、表示部225は、表示情報取得部1001を有する。表示情報取得部1001は、評価部224より通知された評価結果と、予測部222より通知された予測特性とを取得し、表示データ1011を生成する。

[0107] 図10に示すように、表示データ1011は、情報の項目として、“材料データ”、“予測特性”、“評価結果”を含む。“材料データ”には、予測対象データ700の“材料データ”に含まれる材料の構造式である“構造式X”が格納され、“予測特性”には、予測対象データ700について予測された予測特性である“特性値2”が格納される。また、“評価結果”には、新規材料の開発における予測対象データ700の“材料データ”に含まれる材料の構造式（構造式X）の適正を評価した評価結果である“ランク2”が格納される。

[0108] <学習装置による学習処理の流れ>

次に、学習装置120による学習処理の流れについて説明する。図11は、学習処理の流れを示すフローチャートである。

[0109] ステップS1101において、学習装置120は、実験者111が材料特性実験装置110を用いて、実験によって求めた各材料の材料特性を含む学習用データセットを取得する。

[0110] ステップS1102において、学習装置120は、学習用データセットを用いて予測モデルに対して学習処理を行い、学習済みの予測モデルを生成する。

[0111] <予測装置による予測処理の流れ>

次に、予測装置130による予測処理の流れについて説明する。図12は、予測処理の流れを示すフローチャートである。

[0112] ステップS1201において、予測装置130は、学習装置120が学習

済みの予測モデルを生成する際に用いた学習用データセットと、学習済みの予測モデルのモデルパラメータとを取得する。

- [0113] ステップS 1 2 0 2において、予測装置 1 3 0は、学習用データセットを解析し、材料データ間の属性値を算出する。
- [0114] ステップS 1 2 0 3において、予測装置 1 3 0は、算出した材料データ間の属性値の度数分布を算出し、属性値を区別するための複数の区間を決定する。
- [0115] ステップS 1 2 0 4において、予測装置 1 3 0は、除外対象の設定を参照する。
- [0116] ステップS 1 2 0 5において、予測装置 1 3 0は、予測対象データを取得する。
- [0117] ステップS 1 2 0 6において、予測装置 1 3 0は、予測対象データの材料データの属性値を算出する。
- [0118] ステップS 1 2 0 7において、予測装置 1 3 0は、算出した予測対象データの材料データの属性値が、決定した複数の区間のいずれに区別されるかを判定することで、相反する2つの指標に対する予測対象データの適正（ランク）を評価する。
- [0119] ステップS 1 2 0 8において、予測装置 1 3 0は、評価結果が除外対象に該当するか否かを判定する。ステップS 1 2 0 8において除外対象に該当しないと判定した場合には（ステップS 1 2 0 8においてNOの場合には）、ステップS 1 2 1 1に進む。
- [0120] 一方、ステップS 1 2 0 8において、除外対象に該当すると判定した場合には（ステップS 1 2 0 8においてYESの場合には）、ステップS 1 2 1 0に進む。
- [0121] ステップS 1 2 0 9において、予測装置 1 3 0は、予測対象データを、予測対象または表示対象から除外する。
- [0122] ステップS 1 2 1 0において、予測装置 1 3 0は、予測対象データを学習済みの予測モデルに入力することで、予測特性を予測する。

[0123] ステップS 1 2 1 1において、予測装置 1 3 0は、予測対象データと、予測特性と、評価結果とを対応付けた表示データを表示する。

[0124] <各実施例における指標と評価結果との関係>

次に、予測装置 1 3 0を用いて、複数の予測対象データの組について、予測処理を行った結果（予測特性、評価結果）を以下に示すとともに、相反する2つの指標と評価結果との関係及び予測特性の検証例について説明する。

[0125] 具体的には、後述の非特許文献に開示されている水溶解度データセットのうちの100分子を学習用データセットとして学習し、残りの未学習の分子を予測対象データとして、水溶解度を予測するとともに、評価結果を得た。学習用データセットの分子を変更して3回行った結果を、以下に実施例1～3として記す。

[0126] 手順としては、RDKitを使用して分子構造を示すSMILES (Simplified Molecular Line Entry System) から記述子を計算し、計算された記述子を標準化した値を説明変数として使用した。予測モデルとしてはリッジ回帰を用いた。標準化及びリッジ回帰は、scikit-learnにより実行した。

[0127] (1) 実施例1

図13は、実施例1を示す図である。図13に示すように、実施例1では、複数の予測対象データとして、

- ・ニトロメタン、
- ・メタノール、
- ・スクロース、
- ・ジゴキシン、

について、それぞれ、予測特性を予測するとともに、属性値が分けられる区間を判定することで、適正を評価した（表1310参照）。なお、上記予測対象データの材料特性（ここでは、水溶解度）は、いずれも既知であるため、予測精度を検証する趣旨で表1320に示した。

[0128] 表1310によれば、予測対象データごとの指標と評価結果との関係は、

以下のとおりである。

- ・ ランク 1 及びランク 2 と評価された予測対象データについては、水溶解度の高い予測特性が得られており、かつ、良好な予測精度が得られているが、構造式が単純であり、意外性が低い。

- ・ ランク 3 と評価された予測対象データについては、水溶解度が比較的高い予測特性が得られており、かつ、良好な予測精度が得られており、かつ、構造式が比較的複雑であり、意外性が高い。

- ・ ランク 4 と評価された予測対象データについては、水溶解度が非常に高い予測特性が得られているが、実際には、水溶解度の低い材料であり、予測精度が著しく低い。なお、構造式は複雑であり、意外性は高い。

[0129] (2) 実施例 2

図 1 4 は、実施例 2 を示す図である。図 1 4 に示すように、実施例 2 では、複数の予測対象データとして、

- ・ エリトリトール、
- ・ メタノール、
- ・ ラクトース、
- ・ ラフィノース、

について、それぞれ、予測特性を予測するとともに、属性値が分けられる区間を判定することで、適正を評価した（表 1 4 1 0 参照）。なお、上記予測対象データの材料特性（ここでも水溶解度）は、いずれも既知であるため、予測精度を検証する趣旨で表 1 4 2 0 に示した。

[0130] 表 1 4 1 0 によれば、予測対象データごとの指標と評価結果との関係は、以下のとおりである。

- ・ ランク 1 及びランク 2 と評価された予測対象データについては、水溶解度の高い予測特性が得られており、かつ、良好な予測精度が得られているが、構造式が単純であり、意外性が低い。

- ・ ランク 3 と評価された予測対象データについては、水溶解度が比較的高い予測特性が得られており、かつ、良好な予測精度が得られており、かつ、構

造式が比較的複雑であり、意外性が高い。

・ランク4と評価された予測対象データについては、水溶解度が比較的高い予測特性が得られており、かつ、良好な予測精度が得られており、かつ、構造式が複雑であり、意外性が高い。

[0131] (3) 実施例3

図15は、実施例3を示す図である。図15に示すように、実施例3では、複数の予測対象データとして、

- ・尿素、
- ・メタノール、
- ・カフェイン、
- ・ジゴキシン、

について、それぞれ、予測特性を予測するとともに、属性値が分けられる区間を判定することで、適正を評価した(表1510参照)。なお、上記予測対象データの材料特性(ここでも水溶解度)は、いずれも既知であるため、予測精度を検証する趣旨で表1520に示した。

[0132] 表1510によれば、予測対象データごとの指標と評価結果との関係は、以下のとおりである。

・ランク1及びランク2と評価された予測対象データについては、水溶解度の高い予測特性が得られており、かつ、良好な予測精度が得られているが、構造式が単純であり、意外性が低い。

・ランク3と評価された予測対象データについては、水溶解度が比較的高い予測特性が得られており、かつ、良好な予測精度が得られており、かつ、構造式が比較的複雑であり、意外性が高い。

・ランク4と評価された予測対象データについては、水溶解度が高い予測特性が得られているが、実際には、水溶解度の低い材料であり、予測精度が著しく低い。なお、構造式は複雑であり、意外性は高い。

[0133] (4) 指標と評価結果との関係及び予測特性の検証例

上記実施例1～実施例3によれば、良好な予測精度のもとで、水溶解度の

高い予測特性が得られ、比較的複雑な構造式を有する予測対象データ（つまり、有益な新規材料となる可能性が高い予測対象データ）は、ランク3と評価される可能性が高いことがわかる。

[0134] そこで、以下では、更に、実施例1～実施例3においてランク3と評価された予測対象データ（スクロース、ラクトース、カフェイン）の水溶解度が、一般的な材料と比較して、どの程度高いのか（つまり、客観的に高い予測特性といえるのか）を検証する。

[0135] 具体的には、下記の非特許文献に開示されている水溶解度データセットに基づいて、水溶解度の度数分布を生成し、実施例1～実施例3においてランク3と評価された予測対象データの水溶解度を検証した。

[0136] [非特許文献] J.S.Delaney, "Estimating Aqueous Solubility Directly from Molecular Structure", Journal of chemical information and computer sciences, p1000-p1005, May 24, 2004

図16は、予測特性の検証例を示す図である。図16に示すように、実施例1～実施例3においてランク3と評価された予測対象データの水溶解度は、客観的にも高いことがわかる。つまり、予測装置130により出力される、予測対象データの予測特性及び予測対象データの評価結果は、有益な新規材料となる可能性を的確に表しているといえる。

[0137] <まとめ>

以上の説明から明らかなように、第1の実施形態に係る予測装置は、

- ・学習済み予測モデルの生成に用いた学習用データセットを取得し、当該学習用データセットに含まれる複数の材料データ間で算出される属性値の度数分布から、該属性値を区分けするための複数の区間を決定する。

- ・予測対象データの材料データと、学習用データセットに含まれる複数の材料データとの間で算出される属性値が、複数の区間のいずれに区分けされるかを判定することで、相反する2つの指標に対する予測対象データの適正を評価する。

- ・予測対象データを入力することで、学習済みモデルにより予測された予測

特性を、評価結果と対応付けて表示する。

[0138] これにより、第1の実施形態によれば、学習済み予測モデルを用いて所定の条件を満たす予測特性が予測された場合であっても、有益な新規材料となる可能性が低い予測対象データを排除することができる。

[0139] この結果、第1の実施形態に係る予測装置によれば、材料の生成及び検証実験を行う予測対象データを絞り込むことが可能となり、新規材料の開発効率を向上させることができる。

[0140] [第2の実施形態]

上記第1の実施形態では、要約統計量として、中央値、第一四分位、第三四分位、四分位範囲を算出するとともに、係数 α_1 、 α_2 を決定することで、各区間を決定した。しかしながら、各区間の決定方法は、これに限定されない。

[0141] 例えば、要約統計量として、

- ・各材料の最小の属性値の平均値、
- ・各材料の最小の属性値の標準偏差、

を算出し、複数の区間を、

- ・「属性値 $<$ (平均値 $+$ β_1 \times 標準偏差)」となる属性値が分けられる区間、
- ・「(平均値 $+$ β_1 \times 標準偏差) \leq 属性値 $<$ (平均値 $+$ β_2 \times 標準偏差)」となる属性値が分けられる区間、
- ・「(平均値 $+$ β_2 \times 標準偏差) \leq 属性値 $<$ (平均値 $+$ β_3 \times 標準偏差)」となる属性値が分けられる区間、
- ・「(平均値 $+$ β_3 \times 標準偏差) \leq 属性値」となる属性値が分けられる区間、

と決定してもよい。

[0142] また、上記第1の実施形態では、互いに重複しない4つの区間を決定するものとして説明したが、決定する区間の数は、互いに重複しない区間であれば、4つに限定されず、3つ以上であればよい。

[0143] また、上記第1の実施形態では、予測対象データを1つずつ予測装置130に入力する場合について説明したが、複数の予測対象データを予測装置130に入力してもよい。この場合、予測装置130は、例えば、複数の予測対象データのうち、有益な新規材料となる可能性の高い予測対象データの予測特性、評価結果のみを表示するように構成してもよい。

[0144] また、上記第1の実施形態では、除外対象を設定する場合について説明したが、例えば、選択対象を設定するように構成してもよい。具体的には、予測対象データ入力部901は、

- ・予測対象データ700の”材料データ”に含まれる材料の構造式（構造式X）について算出された最小の属性値が、選択対象の区間（例えば、所定の要約統計量が含まれる区間に近い3番目の区間）に区分けされると判定された場合、または、

- ・予測対象データ700の”材料データ”に含まれる材料の構造式（構造式X）が、選択対象の評価結果（例えば、ランク3）であると評価された場合、当該材料の構造式（構造式X）を選択して、学習済み予測モデル902に入力するように制御してもよい。

[0145] また、上記第1の実施形態では、学習装置と予測装置とを別体の装置により実現するものとして説明したが、学習装置と予測装置とは一体の装置として実現してもよい。

[0146] なお、上記実施形態に挙げた構成等に、その他の要素との組み合わせ等、ここで示した構成に本発明が限定されるものではない。これらの点に関しては、本発明の趣旨を逸脱しない範囲で変更することが可能であり、その応用形態に応じて適切に定めることができる。

[0147] 本出願は、2022年5月13日に提出された日本国特許出願第2022-79269号に基づきその優先権を主張するものであり、同日本国特許出願の全内容を参照することにより本願に援用する。

符号の説明

[0148] 100 : 材料設計システム

- 1 1 0 : 材料特性実験装置
- 1 2 0 : 学習装置
- 1 3 0 : 予測装置
- 1 4 0 : 材料設計装置
- 1 5 0 : 材料生成装置
- 1 6 0 : 材料特性実験装置
- 2 0 1 : 学習用データ取得部
- 2 0 2 : 学習部
- 2 1 1 : 学習用データ属性値算出部
- 2 1 2 : 区間決定部
- 2 2 1 : 予測対象データ取得部
- 2 2 2 : 予測部
- 2 2 3 : 予測対象データ属性値算出部
- 2 2 4 : 評価部
- 2 2 5 : 表示部
- 4 0 0 : 学習用データセット
- 7 0 0 : 予測対象データ
- 9 0 2 : 学習済み予測モデル

請求の範囲

- [請求項1] 学習済みモデルの生成に用いられる学習用データセットを取得し、前記学習用データセットに含まれる複数のデータ間で算出される属性値の度数分布から、該属性値を区分けするための複数の区間を決定する区間決定部と、
- 予測対象のデータと、前記複数のデータとの間で算出される属性値が、前記複数の区間のいずれに区分けされるかを判定することで、相反する指標に対する前記予測対象のデータの適正を評価する評価部と、
- 前記予測対象のデータを入力することで、前記学習済みモデルにより予測された予測値を、前記評価部による評価結果と対応付けて表示する表示部と
- を有する予測装置。
- [請求項2] 前記区間決定部は、前記複数のデータ間で算出される属性値について要約統計量を算出し、前記複数の区間を規定する、属性値の下限値または上限値を決定する、請求項1に記載の予測装置。
- [請求項3] 前記区間決定部は、互いに重複しない3つ以上の区間を決定する、請求項2に記載の予測装置。
- [請求項4] 前記評価部は、予測対象のデータと、前記複数のデータとの間で算出される属性値が、前記3つ以上の区間のうち、所定の要約統計量が含まれる区間に近い何番目の区間に区分けされるかにより、前記予測対象のデータの適正を評価する、請求項3に記載の予測装置。
- [請求項5] 前記評価部は、予測対象のデータと、前記複数のデータとの間で算出される属性値が、前記3つ以上の区間のうち、所定の要約統計量が含まれる区間から最も離れた区間に区分けされると判定した場合、前記予測対象のデータを、前記学習済みモデルに入力するデータから除外する、請求項3に記載の予測装置。
- [請求項6] 前記評価部は、予測対象のデータと、前記複数のデータとの間で算

出される属性値が、前記3つ以上の区間のうち、所定の要約統計量が含まれる区間に近いR番目の区間に分けされると判定した場合、前記予測対象のデータを、前記学習済みモデルに入力するデータとして選択する、請求項3に記載の予測装置。

[請求項7]

前記学習用データセットに含まれる複数のデータ間の距離を算出する学習用データ属性値算出部と、

前記複数のデータそれぞれについて算出された他のデータとの間の距離のうち、最小距離を抽出する学習用データ最小属性値抽出部と、を更に有し、

前記区間決定部は、抽出された前記最小距離の度数分布から、該最小距離を分けするための複数の区間を決定する、請求項1に記載の予測装置。

[請求項8]

前記学習用データ属性値算出部は、前記学習用データセットに含まれるN個のデータ（Nは任意の整数）のうち、i番目（ $1 \leq i \leq N$ ）のデータと、i番目のデータを除く（ $N - 1$ ）個のデータとの間の距離をそれぞれ算出し、

前記学習用データ最小属性値抽出部は、前記i番目のデータについて算出された（ $N - 1$ ）個の距離の中から、最小距離を抽出する、

請求項7に記載の予測装置。

[請求項9]

前記予測対象のデータと、前記複数のデータとの間の距離を算出する予測対象データ属性値算出部と、

前記予測対象のデータと、前記複数のデータとの間の距離のうち、最小距離を抽出する予測対象データ最小属性値抽出部と、を更に有し、

前記評価部は、前記予測対象のデータについて抽出された前記最小距離が、前記複数の区間のうちのいずれに分けられるかを判定することで、相反する指標に対する前記予測対象のデータの適正を評価する、請求項7に記載の予測装置。

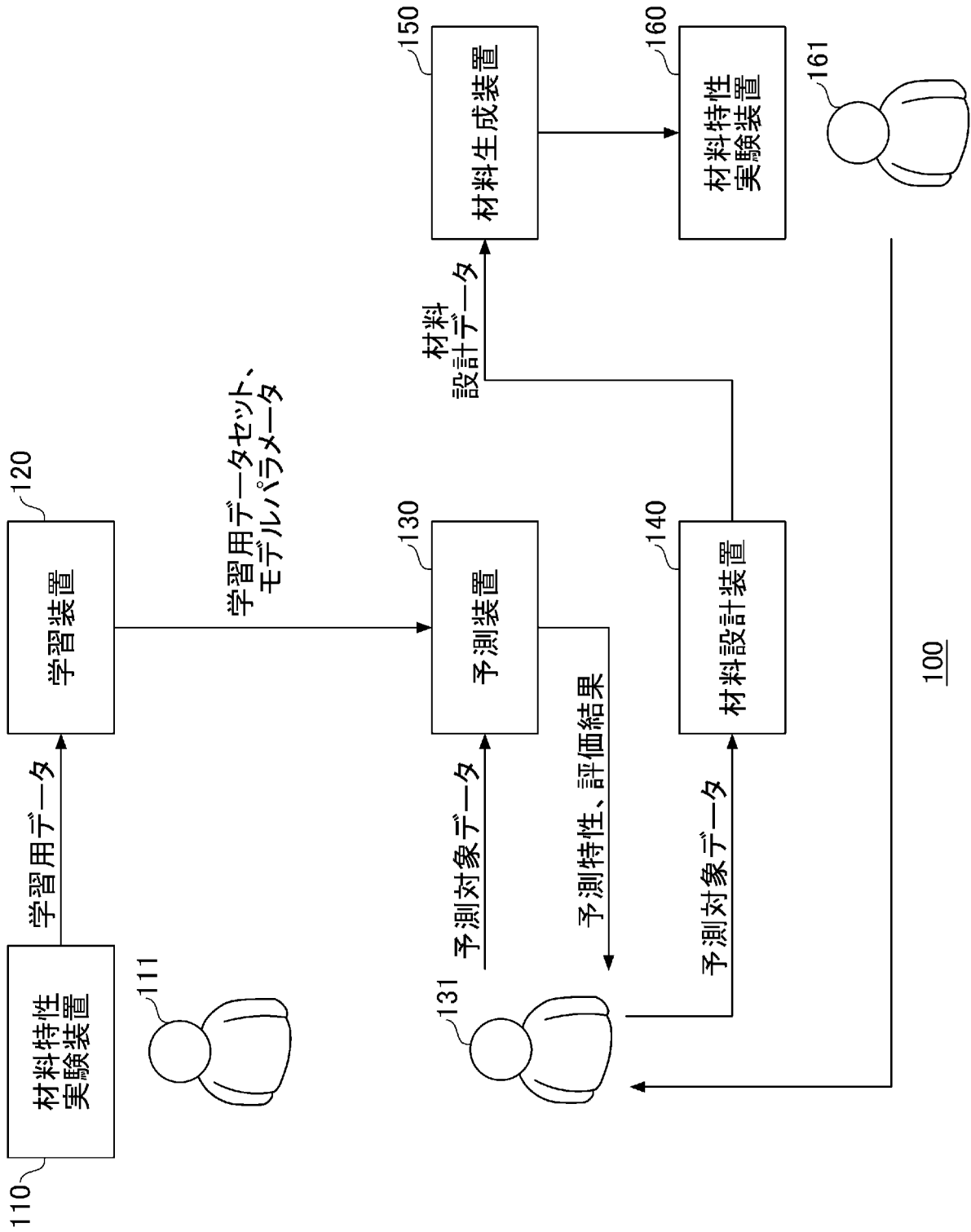
- [請求項10] 請求項1に記載の予測装置と、
前記予測装置の前記評価部により、前記複数のデータとの間で算出された属性値が所定の区間に分けられると判定された予測対象のデータであって、前記予測装置の前記学習済みモデルにより、所定の条件を満たす予測値が予測された予測対象のデータを入力し、材料設計データを生成する材料設計装置と、
を有する材料設計システム。
- [請求項11] 学習用データセットに基づいて学習済みモデルを生成する学習装置を更に有し、
前記予測装置は、前記学習装置により生成された前記学習済みモデルに、前記予測対象のデータを入力することで前記予測値を予測する、請求項10に記載の材料設計システム。
- [請求項12] 学習済みモデルの生成に用いられる学習用データセットを取得し、前記学習用データセットに含まれる複数のデータ間で算出される属性値の度数分布から、該属性値を分けするための複数の区間を決定する区間決定工程と、
予測対象のデータと、前記複数のデータとの間で算出される属性値が、前記複数の区間のいずれに分けられるかを判定することで、相反する指標に対する前記予測対象のデータの適正を評価する評価工程と、
前記予測対象のデータを入力することで、前記学習済みモデルにより予測された予測値を、前記評価工程における評価結果と対応付けて表示する表示工程と
をコンピュータが実行する予測方法。
- [請求項13] 学習済みモデルの生成に用いられる学習用データセットを取得し、前記学習用データセットに含まれる複数のデータ間で算出される属性値の度数分布から、該属性値を分けするための複数の区間を決定する区間決定工程と、

予測対象のデータと、前記複数のデータとの間で算出される属性値が、前記複数の区間のいずれに区分けされるかを判定することで、相反する指標に対する前記予測対象のデータの適正を評価する評価工程と、

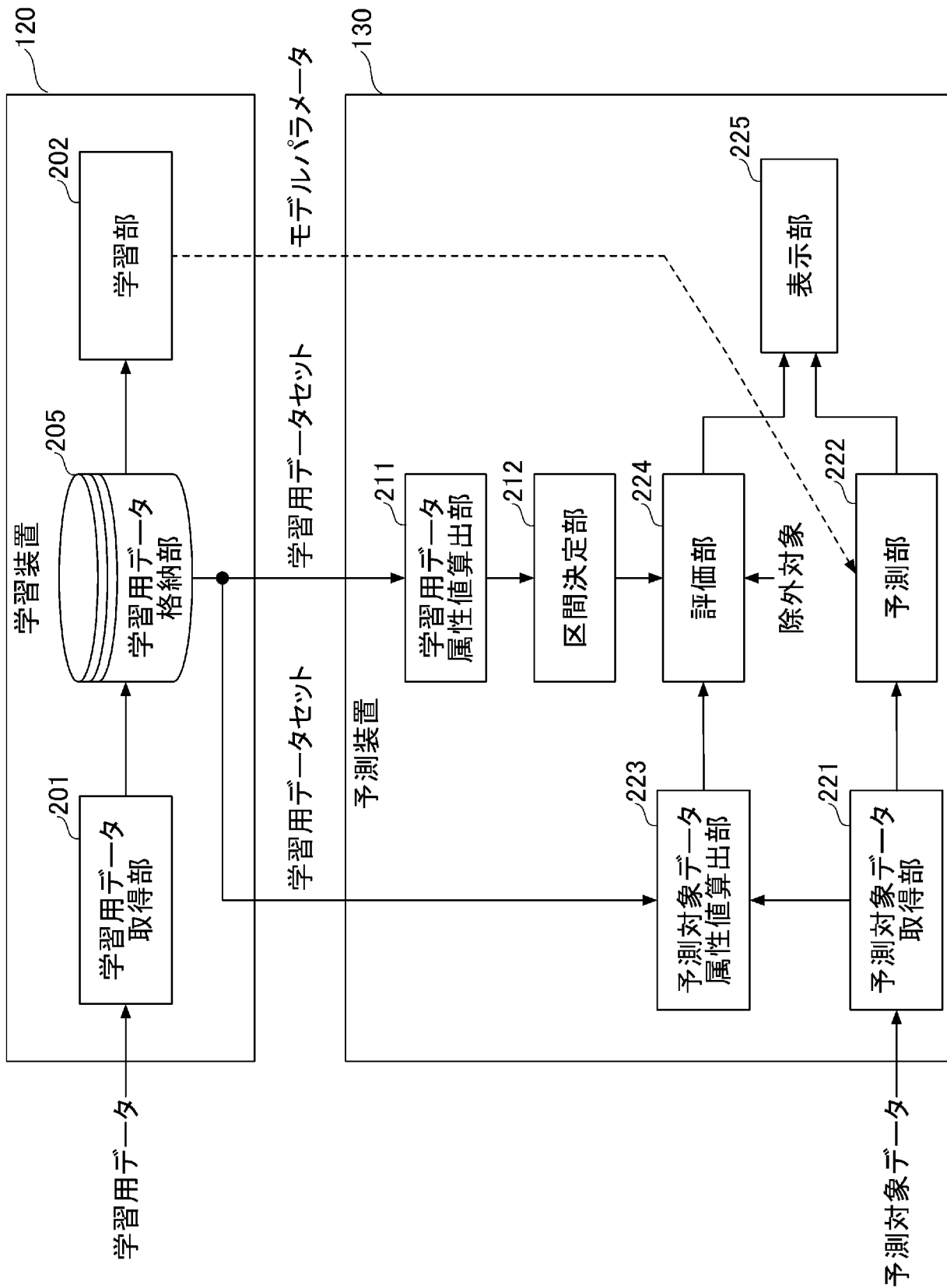
前記予測対象のデータを入力することで、前記学習済みモデルにより予測された予測値を、前記評価工程における評価結果と対応付けて表示する表示工程と

をコンピュータに実行させるための予測プログラム。

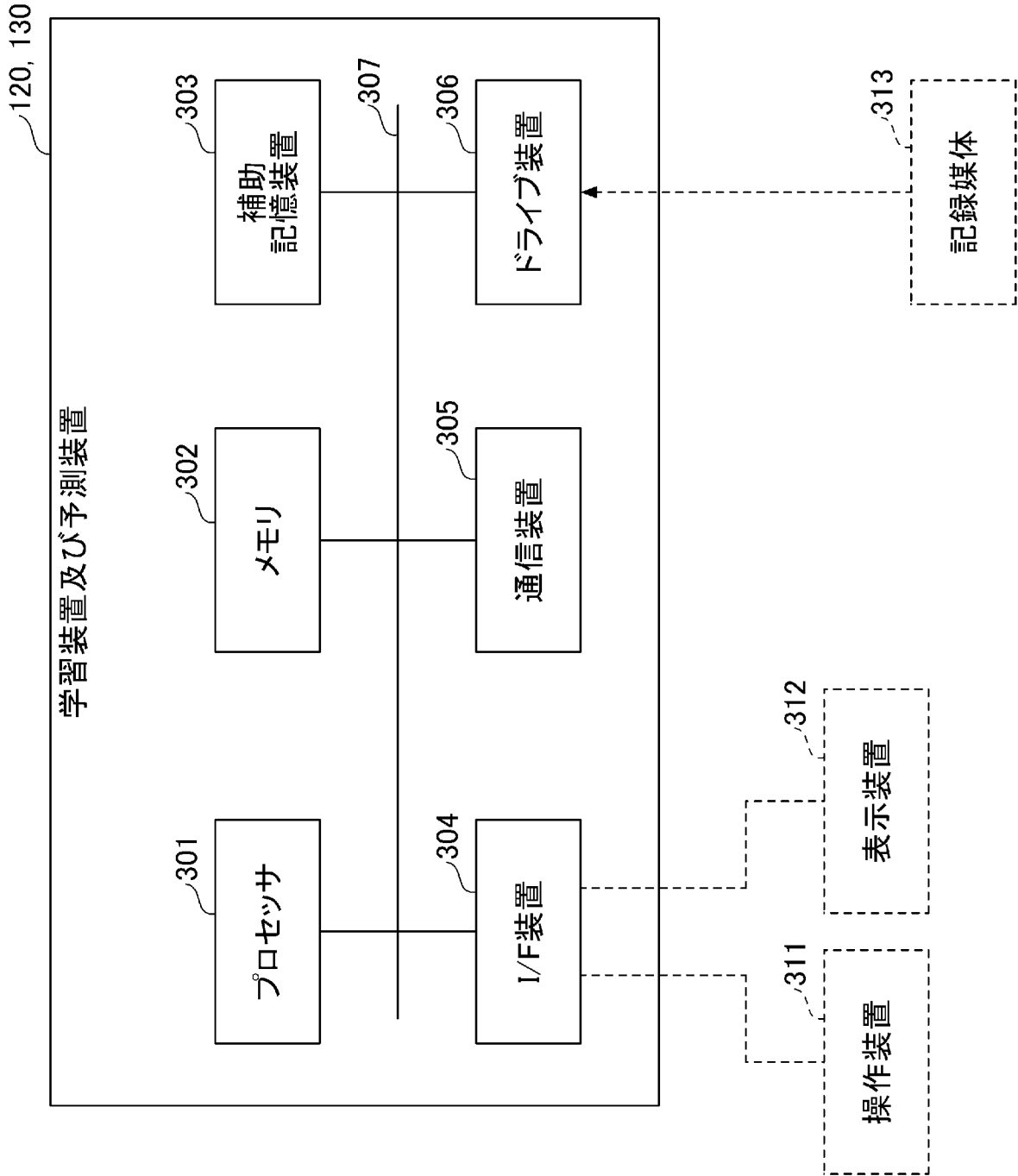
[図1]



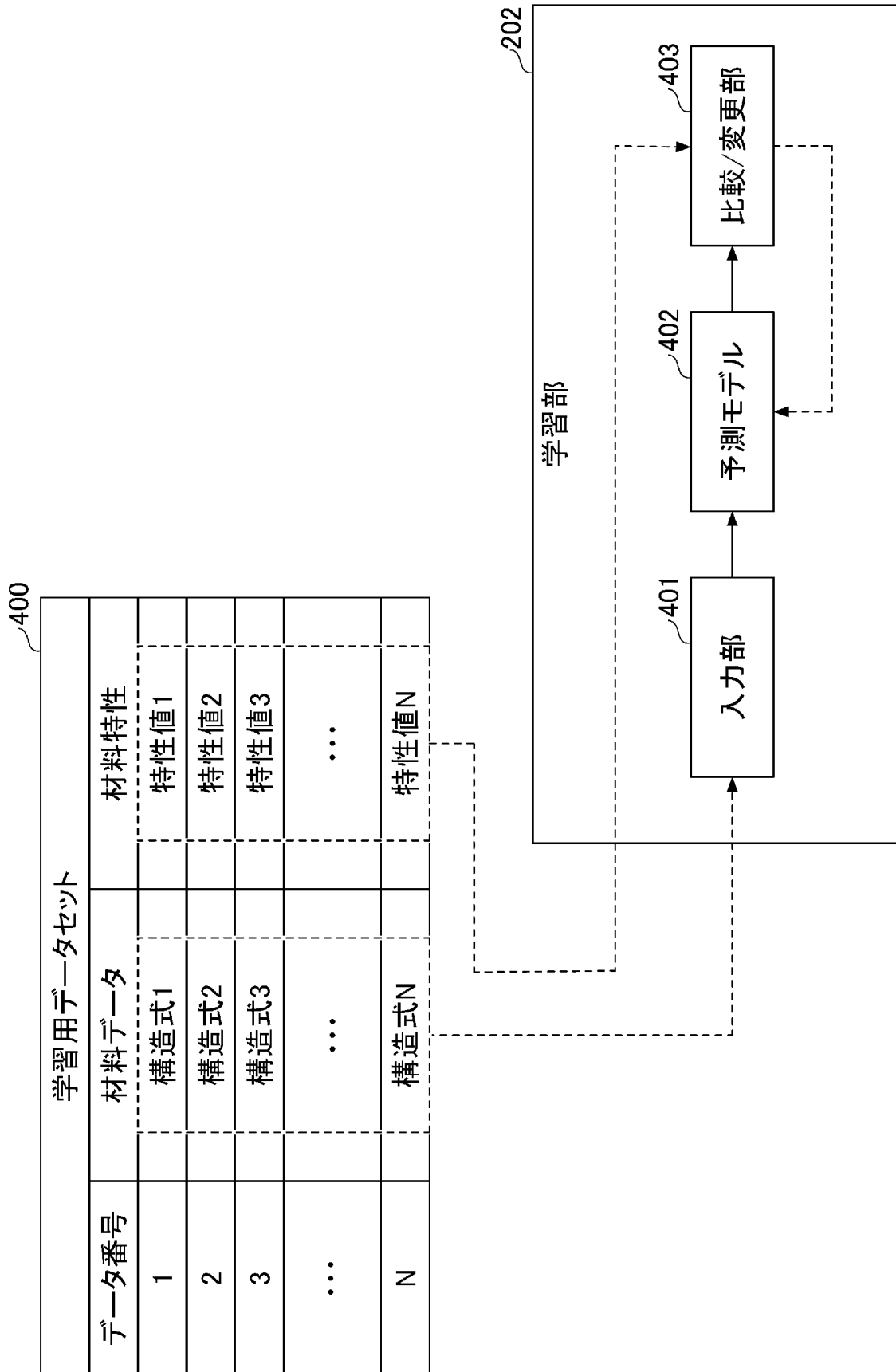
[図2]



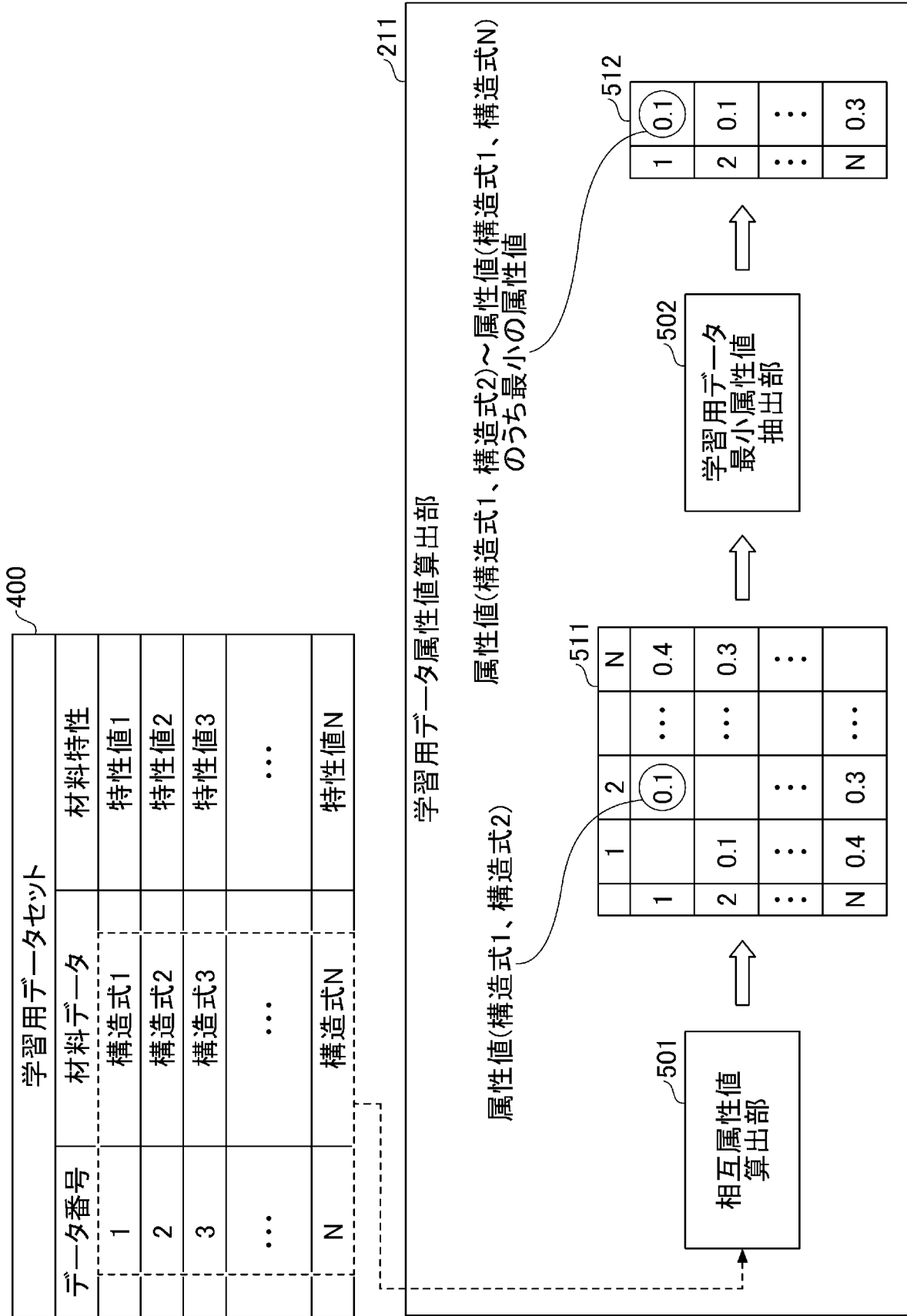
[図3]



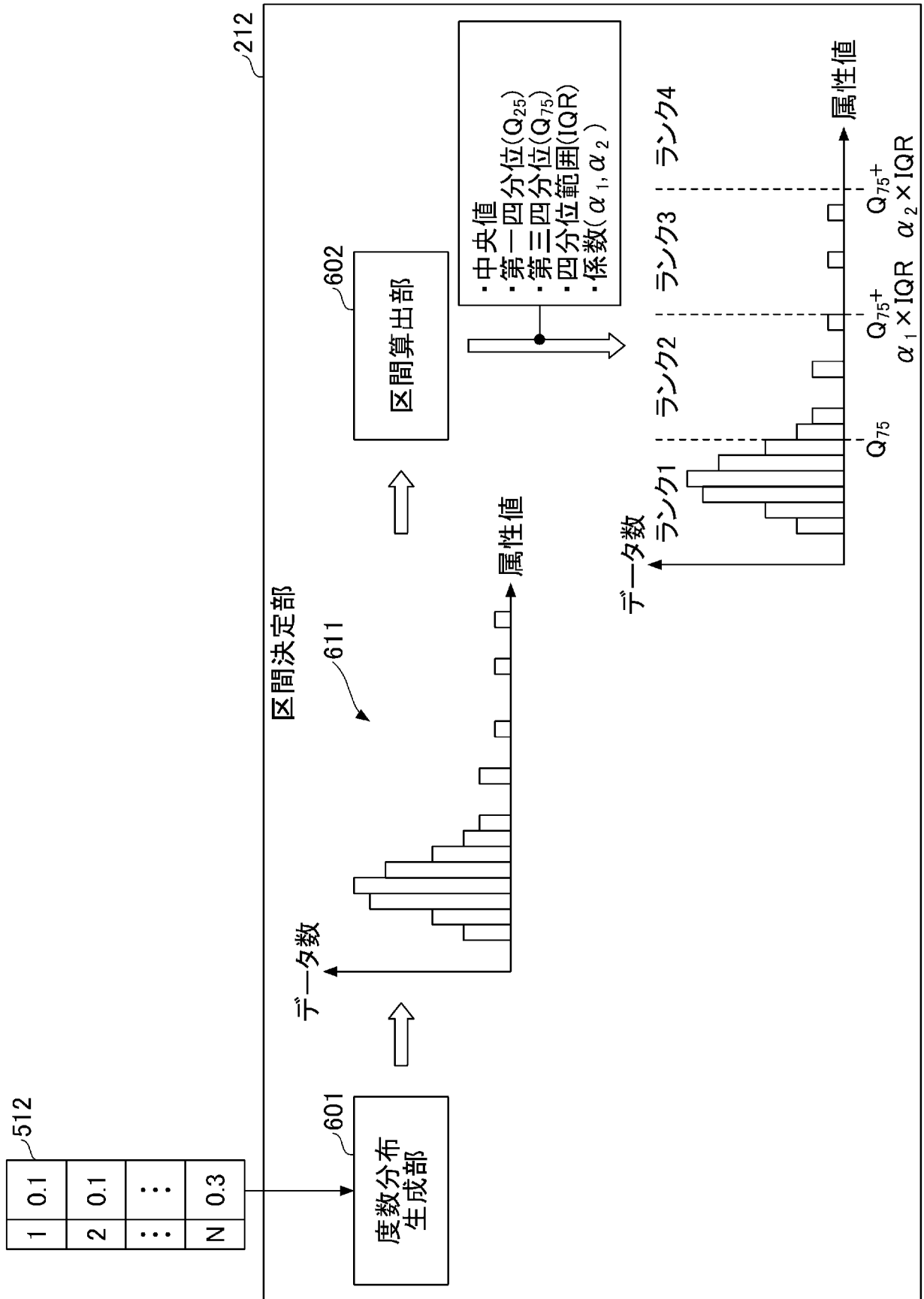
[図4]



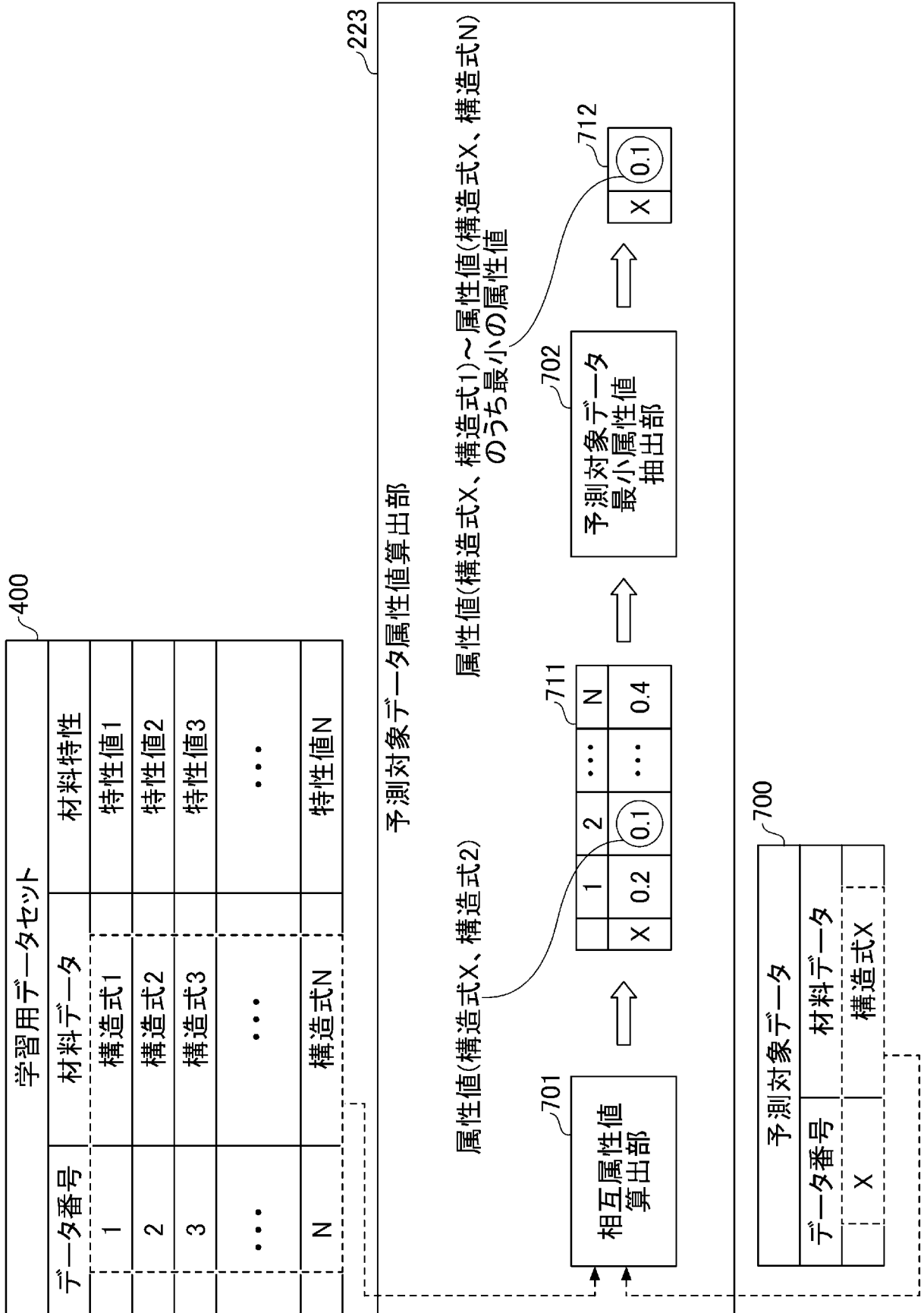
[図5]



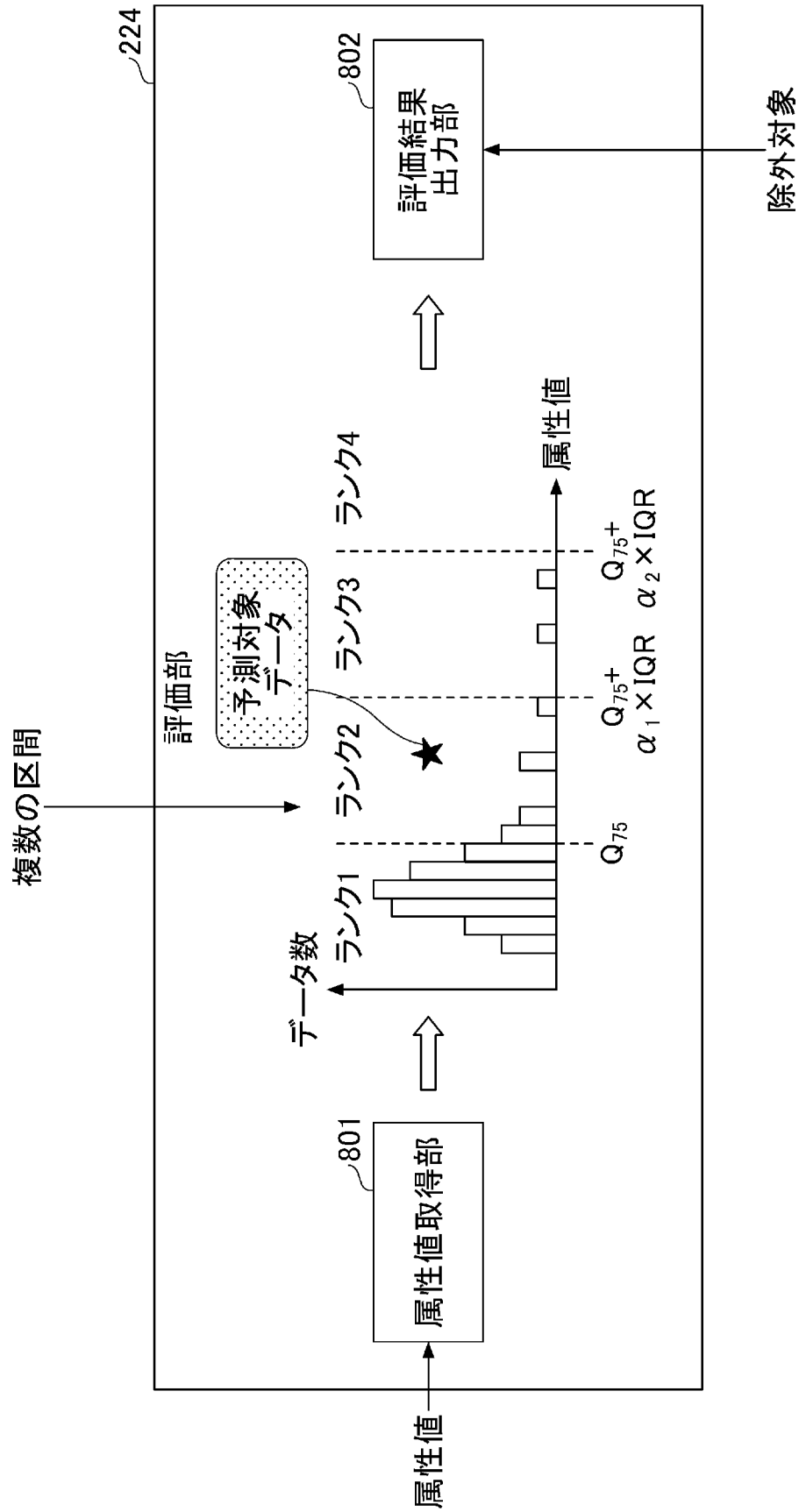
[図6]



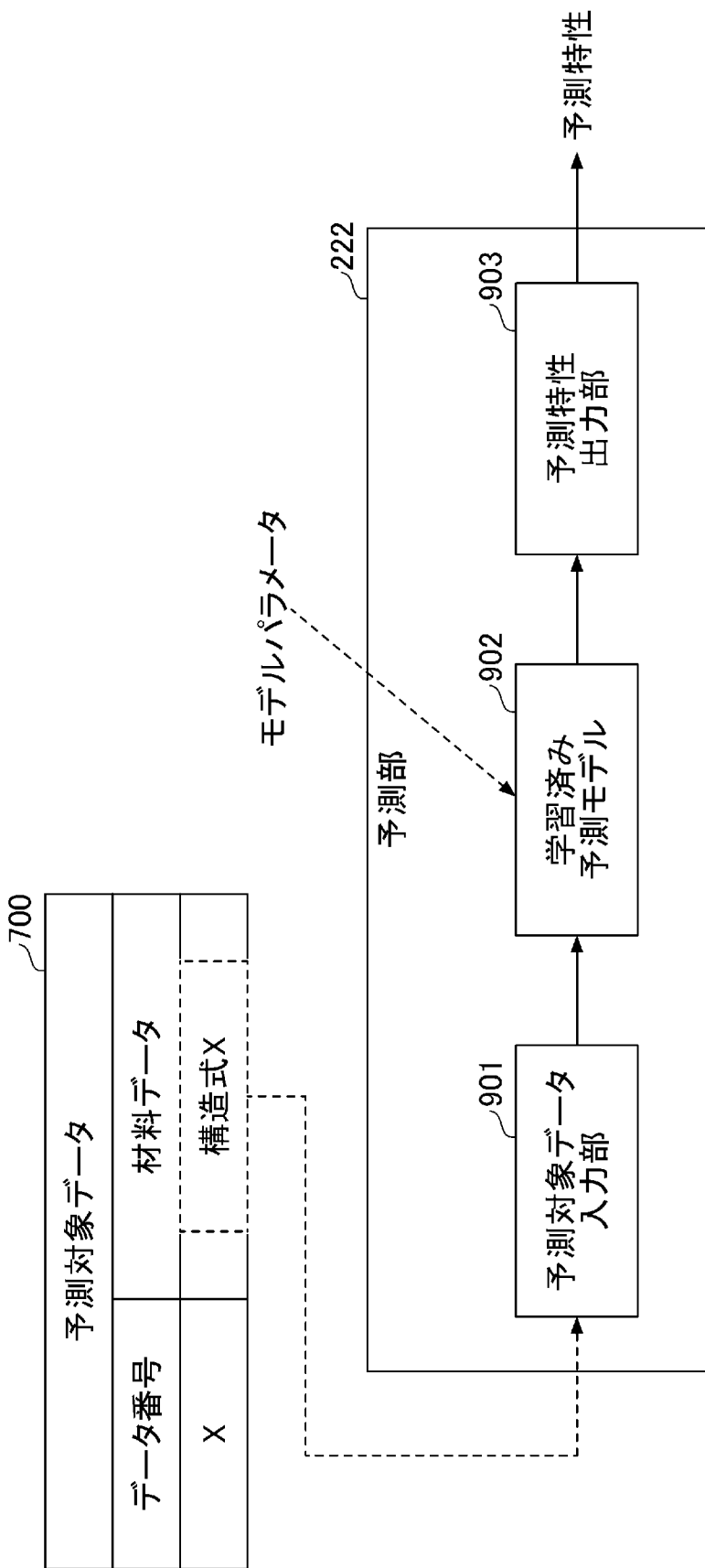
[図7]



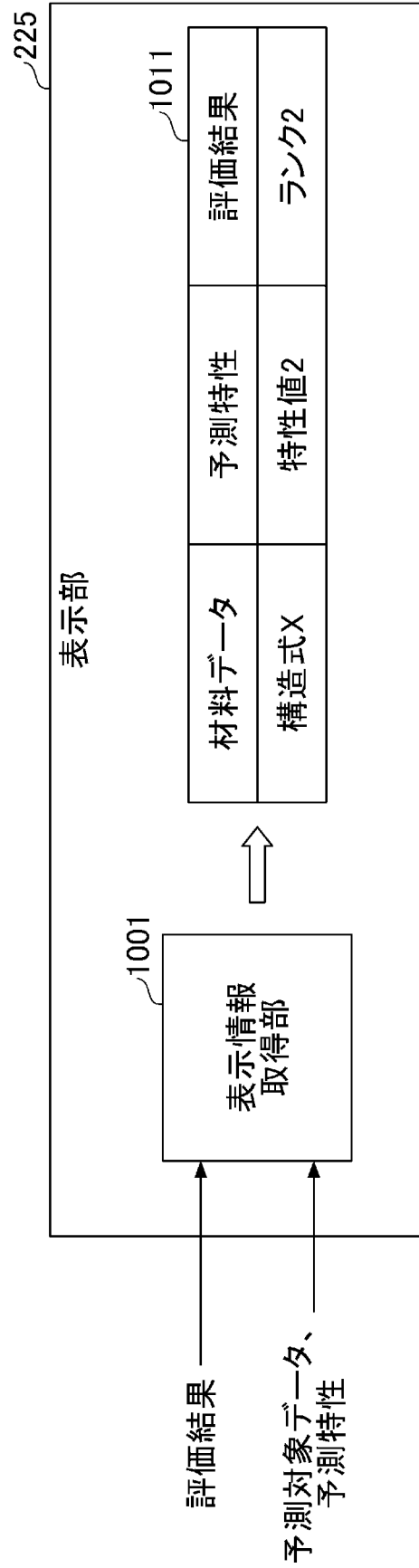
[図8]



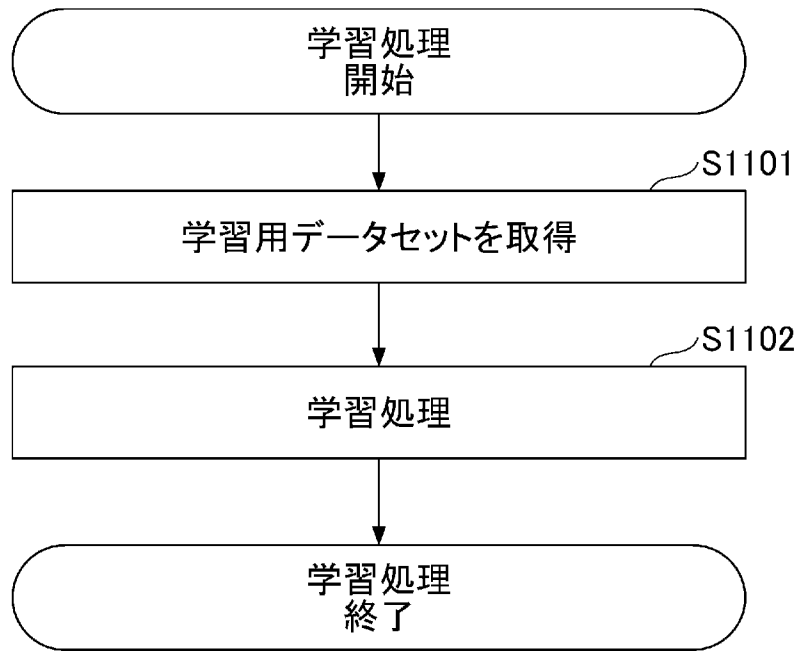
[図9]



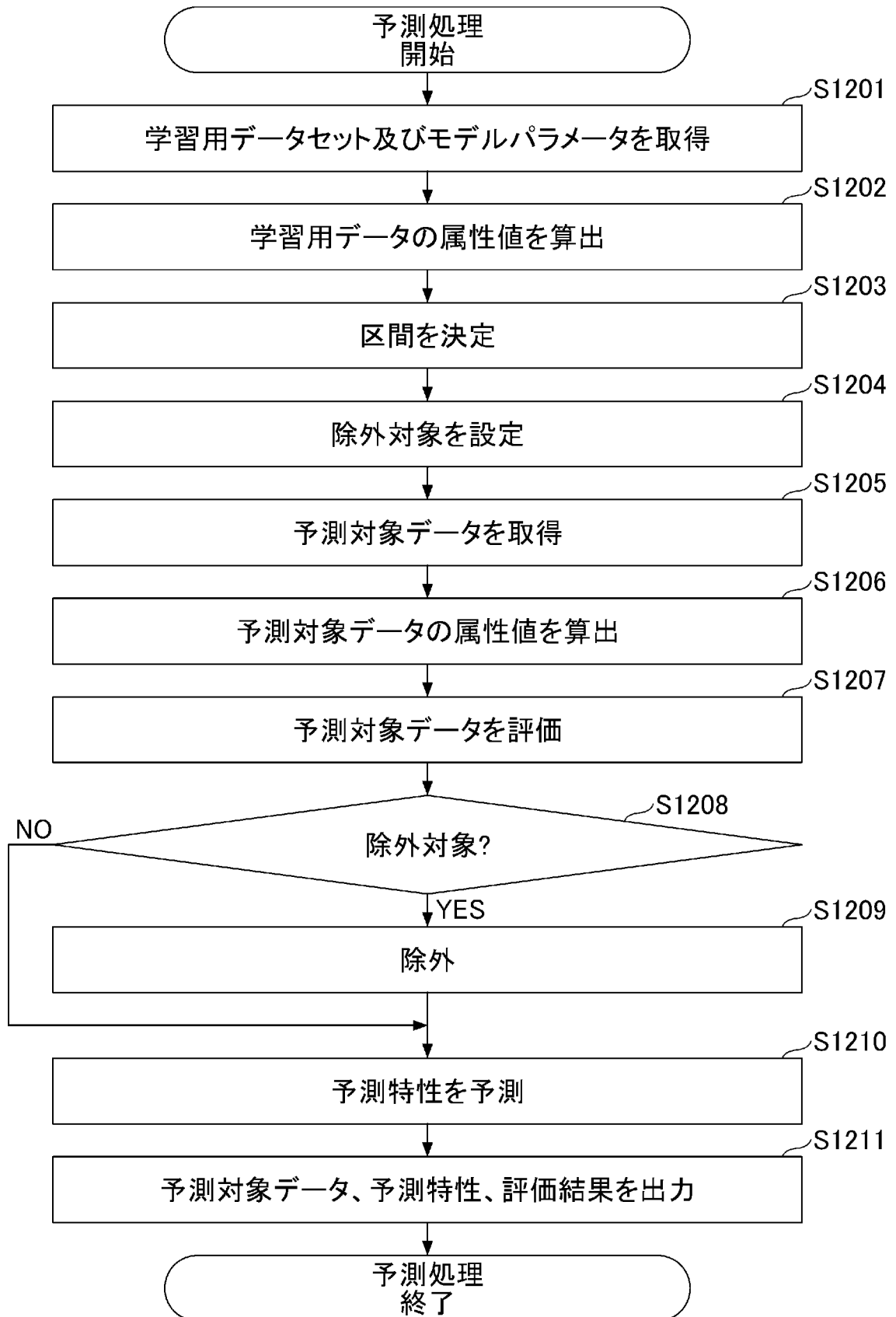
[図10]



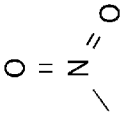
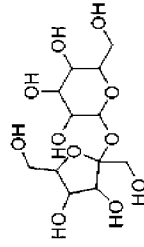
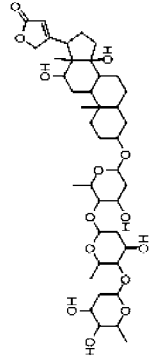
[図11]

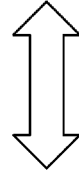


[図12]



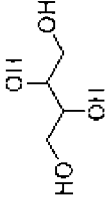
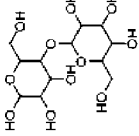
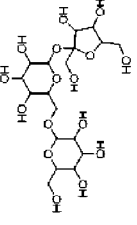
[図13]

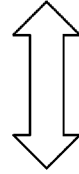
材料名	材料データ	予測特性	評価結果
ニトロメタン		1.3	ランク1
メタノール	—OH	2.7	ランク2
スクロース		-0.3	ランク3
ジゴキシジン		3758.2	ランク4



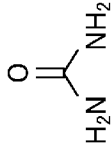
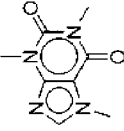
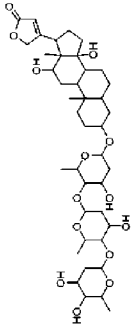
材料特性
0.3
1.6
0.8
-4.1

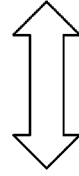
[図14]

1410				1420	
材料名	材料データ	予測特性	評価結果	材料特性	
エリトリトール		1.8	ランク1	0.7	
メタノール	—OH	2.7	ランク2	1.6	
ラクトース		3.4	ランク3	-0.2	
ラフィノース		0.8	ランク4	-0.4	

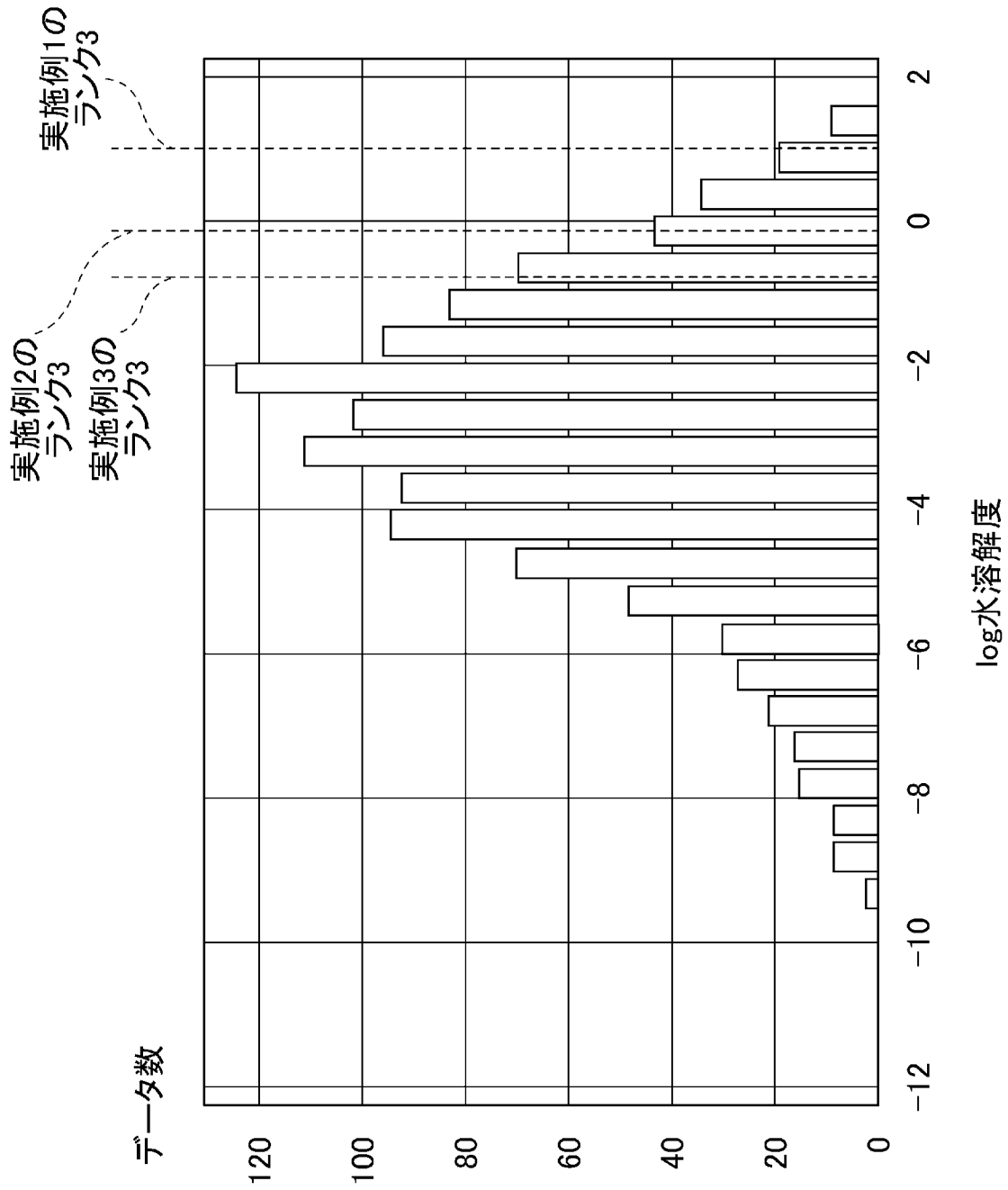


[図15]

1510				1520			
材料名	材料データ	予測特性	評価結果	材料特性			
尿素		1.2	ランク1	1.0			
メタノール	—OH	2.9	ランク2	1.6			
カフェイン		-0.8	ランク3	-0.9			
ジゴキシン		112.0	ランク4	-4.1			



[図16]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2023/017086

A. CLASSIFICATION OF SUBJECT MATTER		
<i>G06Q 10/04</i> (2023.01)i; <i>G16C 60/00</i> (2019.01)i FI: G06Q10/04; G16C60/00		
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) G06Q10/00-99/00; G16C10/00-99/00		
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 1971-2023 Registered utility model specifications of Japan 1996-2023 Published registered utility model applications of Japan 1994-2023		
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	WO 2019/171115 A1 (OMRON CORPORATION) 12 September 2019 (2019-09-12) p. 10, line 8 to p. 12, line 7	1-13
A	JP 2020-086778 A (TOSHIBA CORP) 04 June 2020 (2020-06-04) paragraphs [0024]-[0052]	1-13
A	JP 2020-166499 A (FANUC LTD) 08 October 2020 (2020-10-08) paragraphs [0030]-[0040]	1-13
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" 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 11 July 2023		Date of mailing of the international search report 18 July 2023
Name and mailing address of the ISA/JP Japan Patent Office (ISA/JP) 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/JP2023/017086

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
WO 2019/171115 A1	12 September 2019	(Family: none)	
JP 2020-086778 A	04 June 2020	(Family: none)	
JP 2020-166499 A	08 October 2020	US 2020/0311588 A1 paragraphs [0045]-[0060] CN 111831703 A	

A. 発明の属する分野の分類（国際特許分類（IPC）） G06Q 10/04(2023.01)i; G16C 60/00(2019.01)i FI: G06Q10/04; G16C60/00		
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）） G06Q10/00-99/00; G16C10/00-99/00 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 1971-2023年 日本国実用新案登録公報 1996-2023年 日本国登録実用新案公報 1994-2023年 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）		
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 2019/171115 A1 (OMRON CORPORATION) 12.09.2019 (2019-09-12) 第10頁第8行-第12頁第7行	1-13
A	JP 2020-086778 A (株式会社東芝) 04.06.2020 (2020-06-04) 段落[0024]-[0052]	1-13
A	JP 2020-166499 A (ファナック株式会社) 08.10.2020 (2020-10-08) 段落[0030]-[0040]	1-13
<input type="checkbox"/> C欄の続きにも文献が列挙されている。 <input checked="" type="checkbox"/> 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 “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	11.07.2023	国際調査報告の発送日 18.07.2023
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） 永野 一郎 5L 1204 電話番号 03-3581-1101 内線 3562	

国際調査報告
特許ファミリーに関する情報

国際出願番号

PCT/JP2023/017086

引用文献	公表日	特許ファミリー文献	公表日
WO 2019/171115 A1	12.09.2019	(ファミリーなし)	
JP 2020-086778 A	04.06.2020	(ファミリーなし)	
JP 2020-166499 A	08.10.2020	US 2020/0311588 A1 段落[0045]-[0060] CN 111831703 A	