(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
2024年2月29日(29.02.2024)

WIPO | PCT

(10) 国際公開番号
WO 2024/042736 A1

(51) 国際特許分類:
G06N 3/0455 (2023.01)

(21) 国際出願番号： PCT/JP2023/005451

(22) 国際出願日： 2023年2月16日(16.02.2023)

(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 2022-133545 2022年8月24日(24.08.2022) JP

(71) 出願人: 株式会社日立製作所(**HITACHI, LTD.**) [JP/JP]; 〒1008280 東京都千代田区丸の内一丁目６番６号 Tokyo (JP).

(72) 発明者: 関根 理敏 (**SEKINE, Masatoshi**); 〒1008280 東京都千代田区丸の内一丁目６番６号 株式会社日立製作所内 Tokyo (JP).

(74) 代理人: 弁理士法人サンネクスト国際特許事務所(**SUNNEXT INTERNATIONAL PATENT OFFICE**); 〒1400002 東京都品川区東品川二丁目３番１２号 シーフォートスクエア センタービルディング１６階 Tokyo (JP).

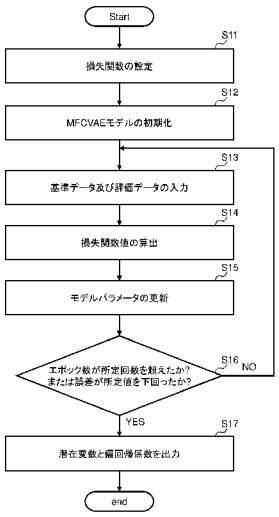
(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, 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, 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,

(54) **Title:** INFORMATION PROCESSING METHOD, INFORMATION PROCESSING SYSTEM, AND INFORMATION PROCESSING PROGRAM

(54) 発明の名称： 情報処理方法、情報処理システム、及び情報処理プログラム

図９



S11 Set loss function
S12 Initialize MFCVAE model
S13 Input reference data and assessment data
S14 Calculate loss function value
S15 Update model parameter
S16 Has number of epochs exceeded prescribed number of times or has error become less than prescribed value?
S17 Output latent variable and partial regression coefficient

(57) **Abstract:** The present invention increases interpretability of attribute values and attributes of training data and test data for a learning model. Accordingly, regression models, in which latent variables outputted from MFCVAE in response to input of input data including reference data having a plurality of attributes assigned with attribute values are defined as explanatory variables and the attribute values are defined as objective variables, are set for the respective attributes. From the latent variables and the attribute values, prediction values of the attribute values at which prediction errors with respect to the attribute values become minimum and regression coefficients of the regression models are calculated for the respective attributes. On the basis of the prediction values and the regression coefficients for the respective attributes, function values of loss functions, of the MFCVAE, obtained by adding additional terms based on indices which are for the respective attributes and of which the values become smaller as the adaptation of the latent variables and the attribute values to the regression models becomes better are calculated. By error backpropagation based on the function values, a model parameter of the MFCVAE is updated. As described above, model training of MFCVAE is executed.

WO 2024/042736 A1

［続葉有］

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)）

(57) 要約：学習モデルの訓練データ及びテストデータの属性と属性値の解釈性を高める。このため、データの複数の属性に属性値が付与されている基準データを含んだ入力データの入力に対してＭＦＣＶＡＥから出力された潜在変数を説明変数とし、属性値を目的変数とする回帰モデルを属性毎に設定する。潜在変数と属性値とから、属性値に対する予測誤差が最小となる属性値の予測値及び回帰モデルの回帰係数を属性毎に算出する。属性毎の予測値及び回帰係数に基づいて、潜在変数及び属性値の回帰モデルへの適合が良いほど小さい値を取る属性毎の指標に基づく追加項を追加したＭＦＣＶＡＥの損失関数の関数値を算出する。関数値に基づく誤差逆伝搬によってＭＦＣＶＡＥのモデルパラメータを更新する。以上のようにＭＦＣＶＡＥのモデル学習を実行する。

明　細　書

発明の名称 ：

情報処理方法、情報処理システム、及び情報処理プログラム

技術分野

[0001] 本発明は、情報処理方法、情報処理システム、及び情報処理プログラムに関する。

背景技術

[0002] ＡＩ（Artificial Intelligence）モデルの品質は、利用するデータの品質に依存する。ＡＩモデルの品質を保証するためには、ＡＩモデルを構築する際の訓練データや推論を行う際のテストデータが有する属性に関する情報を評価することが有用である。そのために、変分オートエンコーダ技術を用いて、学習モデルのエンコーダから潜在変数（特徴量）を抽出し、データの属性情報の内容を明らかにしたり、お互い類似する属性情報をもつデータを抽出したりすること等が行われている。

[0003] 例えば特許文献１では、訓練データから同一のセマンティック特徴に対応する３つの画像を抽出し、３つの画像の各画像について、セマンティック特徴に対応する潜在変数の損失関数を最小化するように変分オートエンコーダのパラメータを更新する。これにより、同一のセマンティック特徴を有する異なる画像の識別性を高めている。

[0004] また例えば非特許文献１では、各潜在変数が入力値に対して与える情報が一意となるように潜在変数の独立性を高めることで、潜在変数の変化に対応する属性の内容やその大きさの変化の解釈性を高めている。その結果、例えば手書き文字データにおいて、ある潜在変数の変化に対して、文字の角度が左斜めから右斜めへ連続変化することが分かる。

[0005] また例えば非特許文献２では、直行する基底の線形結合で潜在変数を表現し、学習によって得られた基底の係数とデータの属性の変化とを対応付けることで、基底の係数の変化に対応する属性の内容やその大きさの変化の解釈

性を高めている。その結果、例えば顔画像データにおいて、ある基底の係数の変化に対して、髪の毛の色が金色から黒色に連続変化することが分かる。

先行技術文献

特許文献

[0006] 特許文献1：特開２０１９－７５１０８号公報

非特許文献

[0007] 非特許文献1：Shuyang Gao, Rob Brekelmans, Greg Ver Steeg, Aram Galstyan, “Auto-Encoding Total Correlation Explanation,” Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics pages 1157-1166., [online], Proceedings of Machine Learning Research (PMLR) 2019., ［令和４年８月１日検索］, インターネット＜URL：https://arxiv.org/abs/1802.05822＞

非特許文献2：Jin-Young Kim, Sung-Bae Cho, “BasisVAE: Orthogonal Latent Space for Deep Disentangled Representation,” [online], International Conference on Learning Representations (ICLR) 2020., ［令和４年８月１日検索］, インターネット＜URL：https://arxiv.org/abs/1802.05822＞

発明の概要

発明が解決しようとする課題

[0008] しかしながら上述の従来技術では、潜在変数と、それに対応する属性や属性値がユーザの解釈に依存し、定性的にしか評価できないため、データの属性と属性値の解釈性が依然として低いという問題があった。

[0009] 本願の開示の一側面では、訓練データやテストデータの潜在変数に対応する属性と属性値の解釈性を高めることを目的とする。

課題を解決するための手段

[0010] 本願の開示の一側面では、処理部と記憶部とを有する情報処理システムが実行する情報処理方法であって、前記処理部が、データの複数の属性に属性

値が付与されている基準データを含んだ入力データを、該データの前記複数の属性のそれぞれに関する潜在変数を出力するＭＦＣＶＡＥ（Multi-Facet Clustering Variational Auto-Encoder）に入力する第１ステップと、前記入力データの入力に対して前記ＭＦＣＶＡＥから出力された前記潜在変数を説明変数とし、前記属性値を目的変数とする回帰モデルを前記属性毎に設定する第２ステップと、前記潜在変数と前記属性値とから、該属性値に対する予測誤差が最小となる前記属性値の予測値及び前記回帰モデルの回帰係数を前記属性毎に算出する第３ステップと、前記第３ステップによって算出された前記属性毎の前記予測値及び前記回帰係数に基づいて、前記潜在変数及び前記属性値の前記回帰モデルへの適合が良いほど小さい値を取る指標を前記属性毎に算出する第４ステップと、前記ＭＦＣＶＡＥによるデータ再構成の誤差を表す再構成誤差項と、前記潜在変数の分布に制約を与える正則化項と、を有する前記ＭＦＣＶＡＥの損失関数に、前記属性毎の前記指標に基づく追加項を追加した損失関数の関数値を算出する第５ステップと、前記第５ステップによって算出された前記関数値に基づく誤差逆伝搬によって前記ＭＦＣＶＡＥのモデルパラメータを更新する第６ステップと、を実行し、前記第１ステップから前記第６ステップまでを、前記予測誤差又はエポック回数が所定条件を充足するまでこの順序で繰り返すことで前記ＭＦＣＶＡＥのモデル学習を実行する、ことを特徴とする。

発明の効果

[0011] 本願の開示の一側面によれば、学習モデルの訓練データやテストデータの属性と属性値の解釈性を高めることができる。前述した以外の課題、構成及び効果は、以下の実施形態の説明により明らかにされる。

図面の簡単な説明

[0012] [図1]従来技術（ＭＦＣＶＡＥ）の問題点を説明するための図。

[図2]基準データと評価データ（文字データの場合）を示す図。

[図3]基準データと評価データ（一般データの場合）を示す図。

[図4]実施形態１に係る情報処理システムのモデル学習時の動作を説明するた

めの図。

[図5]実施形態１に係る情報処理システムの評価データに対する属性値付与時の動作を説明するための図。

[図6]実施形態１に係る情報処理システムの属性値を指定したデータ生成時の動作を説明するための図。

[図7]実施形態１に係る情報処理システムのモデル学習時の決定係数の推移を説明するための図。

[図8]実施形態１に係る情報処理システムの構成を示すブロック図。

[図9]実施形態１に係る特徴量抽出処理を示すフローチャート。

[図10]実施形態１に係る属性値付与処理を示すフローチャート。

[図11]実施形態１に係るデータ生成処理を示すフローチャート。

[図12]実施形態１に係るデータ品質評価処理を示すフローチャート。

[図13]実施形態１に係る属性及び属性値の出力例１（データに対する属性及び属性値）を示す図。

[図14]実施形態１に係るデータ、属性、及び属性値の出力例２（各属性及び属性値に対するデータ数）を示す図。

[図15]実施形態２に係る情報処理システムの構成を示すブロック図。

[図16]コンピュータのハードウェア構成を示す図。

発明を実施するための形態

[0013] 以下、図面を参照して本願の開示に係る実施形態を説明する。実施形態は、図面も含めて本願を説明するための例示である。実施形態では、説明の明確化のため、適宜、省略及び簡略化がされている。特に限定しない限り、実施形態の構成要素は単数でも複数でもよい。また、ある実施形態と他の実施形態を組み合わせた形態も、本願に係る実施形態に含まれる。

[0014] 同一又は類似の構成要素には、同一の符号を付与し、既出に対する後出の実施形態での説明を省略する、又は差分を中心とした説明のみを行う場合がある。また、同一又は類似の構成要素が複数ある場合には、同一の符号に異なる添字を付して説明する場合がある。また、これらの複数の構成要素を区

別する必要がない場合には、添字を省略して説明する場合がある。

[0015] 以下の実施形態では、各種情報をテーブル形式で説明するが、各種情報はテーブル形式以外のデータ形式であってもよい。また、例えば、「ＸＸ情報」「ＸＸテーブル」「ＸＸリスト」「ＸＸキュー」等の各種呼称は、これらは互換可能である。例えば「ＸＸテーブル」は、「ＸＸリスト」と呼んでもよい。また、識別情報について説明する際に、「識別情報」「識別子」「名」「ＩＤ」「番号」等の表現を用いるが、これらは互換可能である。

[0016] （従来技術の問題点）

実施形態の説明に先立ち、実施形態が前提とする従来技術（ＭＦＣＶＡＥ：Multi-Facet Clustering Variational Autoencoders）の問題点について説明する。図１は、従来技術の問題点を説明するための図である。ＭＦＣＶＡＥは、複数の観点での潜在変数を出力可能な、拡張された変分オートエンコーダ（ＶＡＥ：Variational Auto-Encoder）である。変分オートエンコーダとは、ニューラルネットワークを使い、潜在変数の空間として確率分布を仮定した生成モデルである。ＭＦＣＶＡＥにおける観点とは、ＭＦＣＶＡＥが出力する潜在変数（ベクトル）の種類であり、文字データの例では「文字の種類」「字形（太さ、角度等）」等が該当する。

[0017] なお、変分オートエンコーダは、文献１「Diederik P Kingma, Max Welling, “Auto-Encoding Variational Bayes,” May 2014., ［令和４年８月１日検索］, インターネット＜URL：https://arxiv.org/abs/1312.6114＞」に開示されている。ＭＦＣＶＡＥは、文献２「Fabian Falck et.al, “Multi-Facet Clustering Variational Autoencoders, Oct. 2021.［令和４年８月１日検索］, インターネット＜URL：https://arxiv.org/abs/2106.05241＞」に開示されている。

[0018] 以下、特徴とは、データを特徴づける情報（属性、属性値、潜在変数等）である。特徴量とは、定量的に表現可能な特徴の値である。属性とは、データを特徴づける性質（文字データの例では「太さ」「傾き」「ノイズ量」「文字の砕け度合い」等）である。属性値とは、属性の度合いを示す値（文字

データの例では属性「太さ」に対する「１ｍｍ」、属性「傾き」に対する「１０度」、属性「ノイズ量」に対する「１０％」、属性「文字の砕け度合い」に対する「レベル２」等）である。属性値は、連続値でも離散値でもよい。潜在変数とは、変分オートエンコーダ関連技術において、エンコーダから出力される特徴量である。変分オートエンコーダ関連技術とは、ＶＡＥやＭＦＣＶＡＥを含む変分ベイズアルゴリズムを有する変分オートエンコーダ技術全般を指す。

[0019] ＭＦＣＶＡＥは、一つの潜在変数の変化に対して、複数の属性の属性値が変化する。このため、ある潜在変数の変化に対して変化する属性の対応付けが困難であった。図１を参照して、手書き文字の場合について属性及び属性値を例に説明する。

[0020] 図１に示すように、潜在変数１を横軸、潜在変数２を縦軸に取った座標系において、潜在変数１及び潜在変数２のグループ１０１は、属性「太さ」が細い文字に対応した潜在変数のグループである。グループ１０２は、属性「太さ」が中程度の文字に対応した潜在変数のグループである。グループ１０３は、属性「太さ」が太い文字に対応した潜在変数のグループである。グループ１０４は、属性「ノイズ量」が少ない文字に対応した潜在変数のグループである。グループ１０５は、属性「ノイズ量」が中程度の文字に対応した潜在変数のグループである。グループ１０６は、属性「ノイズ量」が多い文字に対応した潜在変数のグループである。ここで属性「太さ」の「太い」「中程度」「細い」、及び属性「ノイズ量」の「多い」「中程度」「少ない」は例示的表現に過ぎず、定量的な表現又はこれに付与したラベルの一例である。

[0021] 図１に示す例では、潜在変数の変化に対して、全ての属性値が一様に増加あるいは減少するものではない。よって、ある潜在変数の変化に対する属性値の変化の対応付けが困難である。例えば図１では、潜在変数１の値が小さいグループ１０１からグループ１０３、グループ１０２へと二段階にわたり増加すると、属性「太さ」の属性値が「細い」、「太い」、「中程度」と変

化する。しかし、潜在変数１の値の一様な増加に対して、属性値の変化が一様でない。潜在変数１の変化に関する属性「ノイズ量」の属性値の変化も同様である。その結果、ある潜在変数の変化に対する属性値の変化の解釈がしにくい。

[0022] 一方、潜在変数２の値が小さいグループ１０３からグループ１０２、グループ１０１へと二段階にわたり増加すると、属性「太さ」の属性値が「太い」、「中程度」、「細い」と変化するように、潜在変数１の値の一様な増加に対して、属性値が一様に減少する。潜在変数２に関する属性「ノイズ量」の属性値の変化も同様である。その結果、潜在変数の変化に対して、属性値が一様に減少するため、ある潜在変数の変化に対する属性値の変化の解釈がしやすい。

[0023]（基準データと評価データ）

先ず、文字データの基準データと評価データを説明する。図２は、基準データと評価データ（文字データの場合）を示す図である。図２において、各行をデータという。それぞれのデータに対して、「文字の種類」「太さ」「傾き」といった各属性について、属性値が格納されているデータと格納されていないデータがある。「文字の種類」は該当の文字のイメージデータである。

[0024] ある属性において、属性値が格納されているデータが、その属性に関する基準データであり、属性値が格納されていないデータが、その属性に関する評価データである。基準データは、評価データの属性値を求めるために用いられる属性値が既知のデータである。評価データは、属性値が未知の属性に関して属性値を求めて付与される対象のデータである。

[0025] 文字データの場合、手書き文字データに対して、一般に「Ａ」や「Ｂ」といった文字の種類の属性値を付与するのは容易であるが、「太さ」や「傾き」などの属性値を付与するのは容易ではない。そこでゴシック体といった「太さ」や「傾き」といった属性の属性値が変更可能である活字を用いて、変分オートエンコーダのモデルを学習する。

[0026] 図２の例では「太さ」や「傾き」等の属性値を持つ「データ属性」“活字”の「データ番号」“１”“２”“３”のデータが基準データ、属性値を持たない「データ属性」“手書き文字”「データ番号」“４”“５”のデータが評価データとなる。より一般には、基準データは活字及び手書き文字を含み、評価データは手書き文字を含む。

[0027] 次に、一般データの基準データと評価データを説明する。図３は、基準データと評価データ（一般データの場合）を示す図である。図２では、データ属性毎に基準データと評価データが分かれていた。図３では基準データと評価データを一般化し、それぞれのデータに対して、各属性に関して属性値が格納されているデータが該当の属性に関する基準データであり、各属性に関して属性値が格納されていないデータが該当の属性に関する評価データである。

[0028] 「属性１」に関して、基準データは「データ番号」“１”“２”“３”“４”のデータであり、評価データは「データ番号」“５”のデータである。同様に、「属性２」に関して、基準データは「データ番号」“１”“２”“５”のデータであり、評価データは「データ番号」“３”“４”のデータである。

[0029] 以下の実施形態の目的は、各属性において、基準データの属性値を利用して評価データの属性値を推定することである。

[0030] ［実施形態１］

（モデル学習時及びデータ再構成時の動作）

図４は、実施形態１に係る情報処理システム１のモデル学習時の動作を説明するための図である。本実施形態では、属性に対して属性値が予め付与された基準データ２０１を用い、基準データ２０１の潜在変数を基に属性値を重回帰モデルで予測できるように、変分オートエンコーダ関連技術において新規の損失関数を用いてモデル学習を実行する。

[0031] 情報処理システム１は、ＭＦＣＶＡＥ２を有する。ＭＦＣＶＡＥ２は、エンコーダ２０３とデコーダ２０５とを含んで構成される。

[0032] ＭＦＣＶＡＥ２のモデル学習時には、エンコーダ２０３に入力されるデータ（訓練データセット）は、基準データ２０１のみを含むか、あるいは基準データ２０１及び評価データ２０２の両方を含む。エンコーダ２０３は、ＭＦＣＶＡＥ２の中間出力である潜在変数２０４を出力する。ＭＦＣＶＡＥ２は、潜在変数２０４を説明変数とし、属性値の正解値（正解ラベル）２０８を目的変数とする重回帰モデルを設定する。なお重回帰モデルに限らず、線形回帰モデル及び非線形回帰モデルの何れでもよい。重回帰モデルは、モデル計算の負荷が少ないという利点があることから、本実施形態では重回帰モデルを採用する。

[0033] 情報処理システム１は、潜在変数２０４と属性値の正解値２０８とから、属性値の予測値と正解値との平均二乗誤差を最小化する偏回帰係数２０９及びその時の属性値の予測値２１０を求める。属性値の予測値２１０は、偏回帰係数２０９を各係数とする潜在変数２０４の一次結合として算出される。属性値の正解値２０８と属性値の予測値２１０とに基づいて、重回帰モデルへの適合（重回帰モデルの当てはまり）が良いほど小さい値を取る適合度を表す指標となり得る決定係数２１１や予測誤差２１２が求められる。決定係数２１１や予測誤差２１２は、重回帰モデルへの適合度を表す指標を含む損失関数２１３で学習される。このようにしてエンコーダ２０３が学習される。

[0034] 潜在変数２０４は、デコーダ２０５に入力される。デコーダ２０５は、デコーダ２０５によって再構成された基準データ２０１である再構成基準データ２０６と、デコーダ２０５によって再構成された評価データ２０２である再構成評価データ２０７とを出力する。再構成評価データ２０７は、属性値が付与されたデータとなっている。

[0035] ここで、偏回帰係数２０９、損失関数２１３、決定係数２１１、予測誤差２１２を説明する。

[0036] 従来技術のＭＦＣＶＡＥでは、目的関数である変分下限（Evidence Lower Bound（ＥＬＢＯ））は、式（１）のように表される。従来手法のＭＦＣＶ

ＡＥでは、式（１）の変分下限の符号をマイナスにした負の損失関数が最小化されるようにＭＦＣＶＡＥモデルのパラメータが学習される（上述の文献２参照）。式（１）において“Ｄ”は訓練データセット、“ｘ”が訓練データセットに含まれる訓練データ、“ｚ→”は潜在変数、“θ”はエンコーダのパラメータ、“φ”はデコーダのパラメータ、“ＫＬ（Ａ|Ｂ）”は分布Ａと分布ＢのＫＬダイバージェンスを表す。

[0037] [数1]

$$\mathcal{L}^{MFCVAE}(D;\theta,\phi) = \mathbb{E}_{x\sim D}\left[\mathbb{E}_{q_\phi(\vec{z}|\boldsymbol{x})}\log p_\theta(\boldsymbol{x}|\vec{z}) - \sum_{j=1}^{J}\left[\mathbb{E}_{q_\phi(c_j|\boldsymbol{x})}KL\left(q_\phi\left|(\boldsymbol{z_j}|\boldsymbol{x})\right|p_\theta(\boldsymbol{z_j}|c_j)\right) + KL\left(q_\phi\left|(c_j|\boldsymbol{x})\right|p_\theta(c_j)\right)\right]\right] \cdots(1)$$

[0038] これに対して本実施形態では、観点ｊでの自由度調整済み決定係数を$R_{f,j}^2$、各自由度調整済み決定係数$R_{f,j}^2$の重み係数をγ_j（＞０）として、式（２）のように目的関数を設定する。式（２）の右辺は、式（１）の右辺の期待値Ｅ［*］のカッコ内の式に第３項$\gamma_j R_{f,j}^2$が追加されたものである。

[0039] なお、式（２）の第３項では、自由度調整済み決定係数$R_{f,j}^2$ではなく、後述の決定係数R_j^2が採用されてもよい。決定係数R_j^2、自由度調整済み決定係数$R_{f,j}^2$は、決定係数２１１の一例である。

[0040] [数2]

$$\mathcal{L}^{Proposed}(D;\theta,\phi) = \mathbb{E}_{x\sim D}\left[\mathbb{E}_{q_\phi(\vec{z}|\boldsymbol{x})}\log p_\theta(\boldsymbol{x}|\vec{z}) - \sum_{j=1}^{J}\left[\mathbb{E}_{q_\phi(c_j|\boldsymbol{x})}KL\left(q_\phi\left|(\boldsymbol{z_j}|\boldsymbol{x})\right|p_\theta(\boldsymbol{z_j}|c_j)\right) + KL\left(q_\phi\left|(c_j|\boldsymbol{x})\right|p_\theta(c_j)\right)\right] + \gamma_j R_{f,j}^2\right] \cdots(2)$$

[0041] ここで、式（２）の期待値Ｅ［*］のカッコ内の第１項は、ＭＦＣＶＡＥによるデータ再構成の誤差を表す再構成誤差項である。式（２）の期待値Ｅ［*］のカッコ内の第２項は、ＭＦＣＶＡＥの潜在変数のバラつきを抑制する等の潜在変数の分布に制約を与える正則化項である。式（２）の期待値Ｅ［*］のカッコ内の第３項は、属性毎の潜在変数及び属性値に対する重回帰モデルへの適合が良いほど小さい値を取る指標に基づく追加項である。

[0042] 損失関数２１３（損失関数Ｌｏｓｓと表す）は負の目的関数であるから、式（２）の目的関数を用いて式（３）のように表される。

[0043] [数3]

$$Loss = -\mathcal{L}^{Proposed}(D;\theta,\phi)\cdots(3)$$

[0044] 本実施形態では、損失関数Ｌｏｓｓが最小化、すなわち決定係数$R_{f,i}^{2}$が最大化されるように、ＭＦＣＶＡＥモデルのパラメータが学習される。

[0045] 次に、観点ｊの決定係数R_{j}^{2}、及び観点ｊの自由度調整済み決定係数$R_{f,j}^{2}$の算出方法を説明する。

[0046] データ数がＮ個、データの属性の種類がｊ＝１，２，…，ＪのＪ個、ある属性ｊにおける潜在変数の次元数をK_{j}とする。またデータの各属性がＭＦＣＶＡＥのＪ個の各観点と一対一に対応しているものとする。インデックス番号ｎであるデータｎに属性ｊの属性値が付与されていれば、データｎは属性ｊにおいて基準データである。一方データｎに属性ｊの属性値が付与されていなければ、データｎは属性ｊにおいて評価データである。

[0047] ある属性ｊに関する基準データのインデックスの集合をB_{j}、集合B_{j}の要素数をM_{j}とする。集合$B_{j}=\{b_{j,1}, b_{j,2}, \cdots, b_{j,M_{j}}\}$とする。ある属性ｊに関する基準データｎにおける潜在変数を$z_{n,j}=\{z_{n,j,1}, z_{n,j,2}, \cdots, z_{n,j,K_{j}}\}$、属性値の正解値を$y_{n,j}$、潜在変数$z_{n,j}$を説明変数とする。また、属性値を目的変数とした重回帰モデルの偏回帰係数を$w_{j}=\{w_{j,0}, w_{j,1}, w_{j,2}, \cdots, w_{j,K_{j}}\}^{T}$、属性値の予測値を$\hat{y}_{n,j}$とする。属性値の予測値$\hat{y}_{n,j}$は、式（４）のように表される。

[0048] [数4]

$$\hat{y}_{n,j} = w_{j,0} + w_{j,1}z_{n,j,1} + w_{j,2}z_{n,j,2} \ldots + w_{j,K_{j}}z_{n,j,K_{j}}$$
$$= \left(1,\ z_{n,j,1}, z_{n,j,2}, \ldots z_{n,j,K_{j}}\right)\left(w_{j,0}, w_{j,1} \ldots w_{j,K_{j}}\right)^{T} = \boldsymbol{Z}_{n,j}\boldsymbol{w}_{j}\cdots(4)$$

[0049] ただし、式（５）のように潜在変数ベクトル$Z_{n,j}$を定義した。

[数5]

$$\boldsymbol{Z}_{n,j} = \left(1,\ z_{n,j,1}, z_{n,j,2}, \ldots z_{n,j,K_{j}}\right)\cdots(5)$$

[0050] ある属性ｊに関する属性値ｙ$_{n,j}$（ただしｎ∈Ｂ$_j$）と重回帰モデルによる属性値の予測値ｙ^$_{n,j}$との予測誤差である平均二乗誤差ＭＳＥ$_j$は、式（６）のように表される。平均二乗誤差ＭＳＥｊは、予測誤差２１２の一例である。

[数6]

$$MSE_j = \frac{1}{M_j}\sum_{n\in B_j}\left(\hat{y}_{n,j} - y_{n,j}\right)^2 = \frac{1}{M_j}\sum_{n\in B_j}\left(\boldsymbol{Z}_{n,j}\boldsymbol{w}_j - y_{n,j}\right)^2 \cdots(6)$$

[0051] ここで平均二乗誤差ＭＳＥ$_j$を最小化するｗ$_j$は、式（６）の右辺をｗ$_j$で偏微分してゼロとおく（∇ｗ$_j$＝０）ことで、式（７）のように、偏回帰係数ｗ$_j$は、潜在変数Ｚ$_j$と属性値ｙ$_j$の関数となる。偏回帰係数ｗ$_j$は、偏回帰係数２０９の一例である。

[数7]

$$\boldsymbol{w}_j = \left(\boldsymbol{Z}_j^T\boldsymbol{Z}_j\right)^T\boldsymbol{Z}_j^T\boldsymbol{y}_j \cdots(7)$$

[0052] ただし式（７）において、潜在変数Ｚ$_j$と属性値ｙ$_j$を、式（８）と式（９）に示すようにおいた。

[数8]

$$\boldsymbol{Z}_j = \left\{\boldsymbol{Z}_{b_{j,1},j}, \boldsymbol{Z}_{b_{j,2},j}, \ldots, \boldsymbol{Z}_{b_{j,M_j},j}\right\}^T \cdots(8)$$

[数9]

$$\boldsymbol{y}_j = \left\{y_{b_{j,1},j}, y_{b_{j,2},j}, \ldots, y_{b_{j,M_j},j}\right\}^T \cdots(9)$$

[0053] なお、ある属性ｊに関する属性値ｙ$_{n,j}$（ただしｎ∈Ｂ$_j$）と重回帰モデルによる属性値の予測値ｙ^$_{n,j}$との予測誤差は、平均二乗誤差に限らず、平均誤差、平均絶対誤差、平均平方二乗誤差、平均誤差率、平均絶対誤差率等を採用することもできる。

[0054] ある属性ｊに関する決定係数Ｒ$_j^2$は、説明変数が目的変数をどれくらい説明しているかを表す。決定係数Ｒ$_j^2$は、属性値の平均値$\bar{y}_{n,j}$を用いて式（

１０）のように表される。

[数10]

$$R_j^2 = 1 - \frac{\sum_{n \in B_j} \left(y_{n,j} - \hat{y}_{n,j}\right)^2}{\sum_{n \in B_j} \left(y_{n,j} - \bar{y}_{n,j}\right)^2} \cdots (10)$$

[0055] また決定係数は、説明変数の数が増えるほど１に近づくという性質を持っているため、説明変数の数が多い場合には、この性質を補正した自由度調整済み決定係数$R_{f,j}^2$が採用されてもよい。自由度調整済み決定係数$R_{f,j}^2$は、説明変数の数をｐとし、基準データのサンプル数はM_jであるので、式（１１）のように表される。

[数11]

$$R_{f,j}^2 = 1 - \frac{\frac{\sum_{n \in B_j} \left(y_{n,j} - \hat{y}_{n,j}\right)^2}{M_j - p - 1}}{\frac{\sum_{n \in B_j} \left(y_{n,j} - \bar{y}_{n,j}\right)^2}{M_j - 1}} \cdots (11)$$

[0056] なお、重み係数γ_jは、再構成誤差項、正則化項、及び決定係数R_j^2（$R_{f,j}^2$）の絶対値の比較から求めることができる。具体的には、決定係数R_j^2に対する重み係数γ_jは、$|\gamma_j R_j^2|$のオーダーが式（２）の右辺の期待値Ｅ［＊］のカッコ内の式の再構成誤差項と正則化項の各絶対値のオーダーと同じになるように定められる。同様に、自由度調整済み決定係数$R_{f,j}^2$に対する重み係数γ_jも、$|\gamma_j R_{f,j}^2|$のオーダーが、式（２）の右辺の期待値Ｅ［＊］のカッコ内の式の再構成誤差項と正則化項の各絶対値のオーダーと同じになるように定められる。

[0057] 基準データ２０１及び評価データ２０２の再構成時には、デコーダ２０５は、基準データ２０１及び評価データ２０２の属性及び属性値（評価データ２０２の場合は付与された属性値）と、エンコーダ２０３の学習の最後のエポックで得た偏回帰係数２０９を用いる。そして、基準データ２０１及び評価データ２０２の属性及び属性値と、偏回帰係数２０９とを用いて、式（４）から、潜在変数（潜在変数ベクトル$Z_{n,j}$）を算出する。そして、デコーダ２０５は、算出した潜在変数を入力として、入力された基準データ２０１及

び評価データ２０２をそれぞれ再構成した再構成基準データ２０６及び再構成評価データ２０７を出力する。

[0058]（属性値付与時の動作）

図５は、実施形態１に係る情報処理システム１の評価データ２０２に対する属性値付与時の動作を説明するための図である。情報処理システム１は、評価データ２０２への属性値付与時には、先ず評価データ２０２を学習済みのエンコーダ２０３に入力し、潜在変数２０４を得る。情報処理システム１は、ＭＦＣＶＡＥ２（図４）の学習の最終エポックで得た偏回帰係数２０９を用い、属性値の予測値２１０を、偏回帰係数２０９を各係数とする潜在変数２０４の一次結合式で算出する。情報処理システム１は、属性値の予測値２１０を評価データ２０２に付与する。

[0059]（データ生成時の動作）

図６は、実施形態１に係る情報処理システム１の属性値を指定したデータ生成時の動作を説明するための図である。データ生成とは、変分オートエンコーダ関連技術において、潜在変数を入力として、デコーダからデータを出力することをいう。情報処理システム１は、生成させたい属性値４０１を持つデータ４０５の生成時には、ユーザが生成させたい属性及び属性値４０１と、ＭＦＣＶＡＥ２（図４）の学習の最後で得た偏回帰係数２０９とから潜在変数２０４を算出する。そして情報処理システム１は、算出した潜在変数２０４をデコーダ２０５に入力することで、生成させたい属性値４０１を持つデータ４０５を生成する。

[0060] なお、データ生成の際に、指定された属性及び属性値４０１に該当する基準データ２０１が存在する場合には、この基準データ２０１に対応するデータを再構成したデータ４０５として採用する。指定された属性及び属性値４０１に該当する基準データ２０１が存在しない場合に、生成させたい属性及び属性値４０１と、偏回帰係数２０９とから潜在変数２０４を算出する。そして、算出した潜在変数２０４をデコーダ２０５に入力することで、生成させたい属性値４０１を持つデータ４０５を生成する。

[0061]（モデル学習時の決定係数の推移）

図７は、実施形態１に係る情報処理システム１のモデル学習時の決定係数の推移を説明するための図である。図７のグラフでは、潜在変数を横軸、属性値を縦軸に取り、属性値の実際の値を点で表し、属性値の予測値を直線で表している。情報処理システム１は、ＭＦＣＶＡＥ２の損失関数に、潜在変数を説明変数、属性値を目的関数とする重回帰モデルの決定係数を含む追加項を追加し、決定係数が高くなるようにＭＦＣＶＡＥ２を学習させる。その結果、学習の初期では決定係数は低い（図７（ａ））が、学習のエポック数が進行して学習の中期（図７（ｂ））、後期（図７（ｃ））と推移するに従って、決定係数は高くなり、潜在変数に基づく属性値の予測精度が高くなる。

[0062]（実施形態１に係る情報処理システム１の構成）

図８は、実施形態１に係る情報処理システム１の構成を示すブロック図である。情報処理システム１は、データ記憶部６０２、特徴量抽出部６０３、属性値付与部６０８、データ生成部６１４、及びデータ品質評価部６１２を有する。

[0063] データ記憶部６０２は、メモリ又はストレージであり、基準データ２０１と評価データ２０２の入力を受け付け、蓄積する。データ記憶部６０２は、情報処理システム１に含まれる装置であっても、情報処理システム１の外部装置であっても何れでもよい。

[0064] 特徴量抽出部６０３は、ＭＦＣＶＡＥ２のデータ記憶部６０２に格納されている基準データ２０１を元に、ＭＦＣＶＡＥ２のモデル学習を実行する。また特徴量抽出部６０３は、ＭＦＣＶＡＥ２のデータ記憶部６０２に格納されている評価データ２０２の属性推定を行う。また特徴量抽出部６０３は、属性値を指定したデータ生成を行う。特徴量抽出部６０３は、回帰モデル適合度評価部６０４、損失算出部６０５、モデル更新部６０６、及びエンコーダ部６０７を有する。特徴量抽出部６０３の処理機能は、図９を参照して後述する。

[0065] 属性値付与部６０８は、評価データ２０２の属性推定を行い、評価データ２０２の属性及び属性値６１１を出力する。属性値付与部６０８は、属性値推定部６０９と、属性及び属性値出力部６１０とを有する。属性値付与部６０８の処理機能は、図１０を参照して後述する。

[0066] データ生成部６１４は、属性値付与部６０８によって出力された対象データの属性及び属性値６１１を入力として、属性値を指定したデータ生成を行い、生成したデータ４０５を出力する。データ生成部６１４は、潜在変数算出部６１５、デコード部６１６、及びデータ出力部６１７を有する。データ生成部６１４の処理機能は、図１１を参照して後述する。

[0067] データ品質評価部６１２は、属性値付与部６０８によって出力された対象データ（基準データ２０１、評価データ２０２）の属性及び属性値６１１に基づいて対象データの品質評価を行い、データ品質評価結果６１３を出力する。

[0068] データ品質評価部６１２は、評価の対象データの属性及び属性値６１１を用いて、一例として、下記のような観点で対象データの品質評価を行う。品質評価については“機械学習品質マネジメントガイドライン”、国立研究開発法人産業技術総合研究所、［令和４年８月１日検索］、インターネット＜URL：https://www.aist.go.jp/aist_j/press_release/pr2020/pr20200630_2/pr20200630_2.html＞を参照すればよい。

（１）データ設計の十分性：データを用いる対象のシステムが対応すべき様々な状況に対して十分な訓練データやテストデータを確保していること。

（２）データセットの被覆性：基準を定めて網羅したそれぞれのケースに対してそれぞれのケースに対応する入力の可能性に対して抜け漏れなく、レアケース及び通常ケースそれぞれに正しく推論できる学習に必要な十分な量のデータが与えられていること。

（３）データの均一性：全体として推論性能の期待値を最大化するように、訓練データを偏り無く用意すること。

[0069] なお、データ品質評価部６１２の処理機能は、図１２を参照して後述する

。

[0070] また、特徴量抽出部６０３、属性値付与部６０８、データ生成部６１４、及びデータ品質評価部６１２は、１つのコンピュータ上に実現されていてもよいし、異なるコンピュータ上に実現されてもよく、これらの統合分散の形態は適宜変更可能である。

[0071] （実施形態１に係る特徴量抽出処理）

図９は、実施形態１に係る特徴量抽出処理を示すフローチャートである。特徴量抽出処理は、特徴量抽出部６０３（図８）によって、ユーザ指示を契機として実行される。

[0072] 先ずステップＳ１１では、回帰モデル適合度評価部６０４は、基準データ２０１の潜在変数を説明変数とし、属性値を目的変数としたＭＦＣＶＡＥモデル（本実施形態では重回帰モデル）の当てはまりの良さを示す指標を損失関数Ｌｏｓｓに設定する。損失関数Ｌｏｓｓに設定される指標は、本実施形態では、観点ｊでの自由度調整済み決定係数$R_{f,j}^{2}$である。

[0073] 次にステップＳ１２では、回帰モデル適合度評価部６０４は、ＭＦＣＶＡＥモデルの初期化を行う。次にステップＳ１３では、回帰モデル適合度評価部６０４は、基準データ２０１及び評価データ２０２をＭＦＣＶＡＥモデルに入力する。ステップＳ１３では、回帰モデル適合度評価部６０４は、少なくとも基準データ２０１をＭＦＣＶＡＥモデルに入力すればよい。

[0074] 次にステップＳ１４では、損失算出部６０５は、式（３）に基づいて損失関数Ｌｏｓｓの関数値を算出する。回帰モデル適合度評価部６０４は、損失算出部６０５による損失関数Ｌｏｓｓの関数値の算出の前段階として、次の処理を行う。すなわち、回帰モデル適合度評価部６０４は、ステップＳ１３の入力データの入力に対してＭＦＣＶＡＥモデルから出力された潜在変数を説明変数とし、属性値を目的変数とする重回帰モデルを属性毎に設定する。次に、回帰モデル適合度評価部６０４は、潜在変数と属性値とから、属性値に対する予測誤差が最小となる属性値の予測値及び重回帰モデルの回帰係数を属性毎に算出する。次に、回帰モデル適合度評価部６０４は、算出された

属性毎の予測値及び回帰係数に基づいて、潜在変数及び属性値に対する重回帰モデルへの適合が良いほど小さい値を取る指標を属性毎に算出する。その後、損失算出部６０５は、ステップＳ１４で、損失関数Ｌｏｓｓの関数値を算出する。

[0075] なお、入力データが基準データ２０１及び評価データ２０２を含む場合、ステップＳ１４では、回帰モデル適合度評価部６０４は、損失関数Ｌｏｓｓの追加項を、基準データ２０１を用いて計算する。一方、損失算出部６０５は、再構成誤差項及び正則化項を、基準データ２０１及び評価データ２０２の何れか一方又は両方を用いて計算する。これは、損失関数Ｌｏｓｓの追加項は、潜在変数と属性値との重回帰モデルへの適合度に基づくことから、属性値を含む基準データのみ損失関数Ｌｏｓｓの追加項を計算可能なためである。

[0076] 次にステップＳ１５では、モデル更新部６０６は、誤差逆伝搬によりＭＦＣＶＡＥモデルのパラメータを更新する。次にステップＳ１６では、モデル更新部６０６は、所定条件（エポック数が所定回数を超えたか、又はＭＦＣＶＡＥモデルによる推定値と実際の値の誤差が所定値を下回った）が充足されたかを判定する。モデル更新部６０６は、所定条件が充足された場合（ステップＳ１６ＹＥＳ）にステップＳ１７に処理を移し、所定条件が充足されていない場合（ステップＳ１６ＮＯ）にステップＳ１３に処理を戻す。

[0077] ステップＳ１７では、エンコーダ部６０７は、学習済みのＭＦＣＶＡＥモデルのエンコーダ２０３に評価データ２０２を入力し、潜在変数２０４と偏回帰係数２０９を出力する。

[0078]（実施形態１に係る属性値付与処理）

図１０は、実施形態１に係る属性値付与処理を示すフローチャートである。属性値付与処理は、属性値付与部６０８（図８）によって、ユーザ指示を契機として実行される。

[0079] 先ずステップＳ２１では、属性値推定部６０９は、特徴量抽出部６０３（エンコーダ２０３）から得られた潜在変数２０４と偏回帰係数２０９から、

評価データ２０２の属性値の予測値２１０を算出する。次にステップＳ２２では、属性及び属性値出力部６１０は、基準データ２０１と、ステップＳ２１で属性値の予測値２１０が算出された評価データ２０２との属性及び属性値から、属性毎の各属性値の出現頻度のヒストグラムを求める（後述の図１４参照）。そして属性及び属性値出力部６１０は、このヒストグラムをもとに各属性値が基準データ２０１及び評価データ２０２の各属性において出現する確率をデータ含有率として求め、結果を出力する。

[0080]（実施形態１に係るデータ生成処理）

図１１は、実施形態１に係るデータ生成処理を示すフローチャートである。データ生成処理は、データ生成部６１４（図８）によって、ユーザ指示を契機として実行される。

[0081] 先ずステップＳ３１では、データ生成部６１４は、ユーザによる生成させたい属性及び属性値４０１の入力を受け付ける。次にステップＳ３２では、潜在変数算出部６１５は、ステップＳ３１で入力を受け付けた属性及び属性値４０１と偏回帰係数２０９から、潜在変数２０４を計算して出力する。次にステップＳ３３では、デコード部６１６（デコーダ２０５）は、ステップＳ３２で計算された潜在変数２０４を入力として生成させたい属性及び属性値４０１を持つデータ４０５（例えば文字データ）を再構成する。次にステップＳ３４では、データ出力部６１７は、デコード部６１６（デコーダ２０５）によって再構成されたデータ４０５を出力する。

[0082] なお、潜在変数算出部６１５は、ステップＳ３１で入力を受け付けた属性及び属性値４０１に該当する基準データ２０１が存在する場合には、ステップＳ３２をスキップし、ステップＳ３３でこの基準データ２０１に対応するデータを再構成したデータ４０５とする。

[0083]（実施形態１に係るデータ品質評価処理）

図１２は、実施形態１に係るデータ品質評価処理を示すフローチャートである。データ品質評価処理は、データ品質評価部６１２（図８）によって、ユーザ指示を契機として実行される。

[0084] ステップＳ４１では、データ品質評価部６１２は、属性値付与部６０８によって出力された属性及び属性値６１１に関して、例えば上述の（１）データ設計の十分性、（２）データセットの被覆性、（３）データセットの均一性の少なくとも一つの観点で評価する。次にステップＳ４２では、データ品質評価部６１２は、ステップＳ４１のデータ品質評価結果６１３を出力する。

[0085]（属性及び属性値の出力例１）

図１３は、属性及び属性値の出力例１（データに対する属性及び属性値）を示す図である。図１３は、属性値付与部６０８の属性値推定部６０９（図８）によって、例えば図２又は図３に示す属性値が付与されていなかったデータに属性値が付与され、属性及び属性値出力部６１０によって出力されたものである。

[0086]（属性及び属性値の出力例２）

図１４は、データ、属性、及び属性値の出力例２（各属性及び属性値に対するデータ数）を示す図である。図１４は、図１３の表示方法を変えた出力例である。図１４は、属性値付与部６０８の属性及び属性値出力部６１０（図８）によって出力される属性毎の属性値のヒストグラムである。この表示によって、属性毎に例えば上述の（２）データセットの被覆性や（３）データの均一性を確認できる。（２）データセットの被覆性は、図１４のヒストグラムの各属性の属性値が所定の広い範囲に分布しかつ各度数が何れも所定数以上であることで充足されると考えられる。（３）データの均一性は、図１４のヒストグラムの各属性の属性値が所定の広い範囲に均等に分布していることで充足されると考えられる。このような分析によって、属性値に対して不足しているデータを確認することが可能となる。

[0087] 例えば図１４のヒストグラム１１０１は、属性１の度数分布を示す。ヒストグラム１１０１は、ヒストグラム１１０２、１１０３と比較してデータの分布範囲が広い又は同等であるが、この分布範囲に存在しない属性値がある。この点でヒストグラム１１０１は、（２）データセットの被覆性が充足さ

れていないと言える。またヒストグラム１１０１は、属性値の分布が均一でない。属性値の分布の均一性は、属性値の分散や標準偏差といったバラつきを表す統計値に基づいて判断できる。この点でヒストグラム１１０１は、（３）データの均一性が充足されていないと言える。

[0088] また図１４のヒストグラム１１０２は、属性２の度数分布を示す。ヒストグラム１１０２は、ヒストグラム１１０１、１１０３と比較してデータの分布範囲が狭く、この分布範囲に存在しない属性値がある。この点でヒストグラム１１０２は、（２）データセットの被覆性が充足されていないと言える。またヒストグラム１１０２は、属性値の分布が均一でない。この点でヒストグラム１１０２は、（３）データの均一性が充足されていないと言える。

[0089] また図１４のヒストグラム１１０３は、属性Ｊの度数分布を示す。ヒストグラム１１０３は、ヒストグラム１１０１、１１０２と比較してデータの分布範囲が広い又は同等であるが、この分布範囲に存在しない属性値がある。この点でヒストグラム１１０３は、（２）データセットの被覆性が充足されていないと言える。またヒストグラム１１０３は、ヒストグラム１１０１、１１０２と比較して属性値の分布が均一でない。この点でヒストグラム１１０２は、（３）データの均一性が充足されていないと言える。

[0090] なお、図１４の各グラフを、「データ数」に代えて「各属性値が基準データ２０１及び評価データ２０２の各属性において出現するデータ含有率」を縦軸とするグラフとしてもよい。

[0091] ［実施形態１の効果］

本実施形態では、訓練データやテストデータの属性をユーザが明示的に指定し、定量的な属性値で表すことにより、ユーザにとって解釈性の高い属性分析が可能となる。このため、訓練データやテストデータの中に不足しているデータや、誤判別の多いデータの特徴を発見しやすい。

[0092] また本実施形態では、従来技術のように、得られた潜在変数の持つ属性をユーザが解釈する（偏在変数は太さ又は角度に依存する等）ではなく、ユーザが潜在変数に持たせるべき属性を明示的に指定できるため、ユーザの意図

に従った属性分析が可能となる。

[0093] また、従来技術では属性が定性的にしか分からないため、異なるデータセットや異なるモデルで学習したデータ間の属性は比較できなかった。しかし、本実施形態では属性値が定量的に求まるため、異なるデータセットや異なるモデルで学習したデータ間の属性の比較が可能となる。

[0094] また、本実施形態では、属性値の予測に限定した場合であっても、教師あり学習による回帰モデルを使って属性値を推定する場合よりも少ないデータ量又は学習量で属性値を付与できる。

[0095] また、本実施形態では、データ生成では、複数の属性に対する属性値を指定してデータ生成できるため、必要とされるデータを容易に生成できる。また、データ生成の際に、指定された属性及び属性値に該当する基準データが存在する場合にはこの基準データに対応するデータを再構成したデータとして採用する。これにより、属性及び属性値と偏回帰係数から潜在変数を算出しデコーダでデコードしてデータを再構成する場合と比較して、速やかにデータを再構成できる。

[0096] ［実施形態２］

実施形態１では、１つの情報処理システム１を用いてモデル学習（図４、図９）、属性値付与及び属性と属性値の関係の出力（図５、図１０）、及びデータ品質評価処理（図１２）を実行する例を示した。しかし、モデル学習、属性値付与、及びデータ品質評価処理は、図１５に示す複数の情報処理システム１（１－１，１－２，…，１－ｎ）で並列に実行されてもよい。

[0097] 例えばモデル学習を複数の情報処理システム１で実行する場合、ステップＳ１３～Ｓ１６（図９）を、複数の情報処理システム１毎にそれぞれ異なる基準データ２０１を含む入力データを用いて実行してもよい。そして、複数の情報処理システム１の少なくとも１つが、各情報処理システム１によって得られたＭＦＣＶＡＥモデルの学習結果をマージして出力する。

[0098] また複数の情報処理システム１毎に得られたＭＦＣＶＡＥのモデルのそれぞれの学習結果に基づいて、情報処理システム１毎に各入力データへの属性

値付与及び属性と属性値の関係の出力（ステップＳ２１～Ｓ２３（図１０））を実行してもよい。そして、複数の情報処理システム１の少なくとも１つが、各情報処理システム１によって得られた属性と属性値の関係（図１４）をマージして出力する。

[0099] 本実施形態では、従来技術と比較して、潜在変数が定量的に求まるため、モデル毎に別システムで並列処理しても計算結果をマージできることから、複数システムでモデル学習、属性値付与、及び属性と属性値の関係の出力の各処理の負荷分散が可能となる。よって、従来と比較して短い時間で、これらの処理を完了させ、必要とされるデータを生成することができる。

[0100] （実施形態の適用例）

実施形態は、上述のように手書き文字の文字認識に適用できる。その他、実施形態は、全てのデータにラベル付与するのが困難であり、一部のデータのみ正確な属性値（ラベル）が付与されており、残りのデータに属性値（ラベル）を付与したいといったケースであれば適用できる。

[0101] 例えば、工場設備の振動データに対する回転数のラベルの付与がある。前提として過去に取得した工場設備の振動データには回転数のラベルが付与されておらず、新たに回転数が計測できる装置を導入し、過去に取得した工場設備のデータに対して回転数のラベルを付与するような場合である。

[0102] また、画像における被写体の角度予測がある。角度のラベル付与された少量のデータから、未知の画像の被写体の角度を予測する場合である。この適用例は、ロボットが物をつかむときの把持の方向制御等に利用できる。

[0103] また、楽曲の印象評価を行う場合がある。予めユーザが評価した楽曲の印象（楽しい、悲しい、うれしい、寂しいなど）から、未知の楽曲の印象のラベル付与を行うことができる。

[0104] また、学会論文の研究分野の可視化を行う場合がある。予め各分野との関係度（画像認識分野との関連度が３０、強化学習分野との関連度が５０、・・・）が分かっている論文を基に、未知の論文の各分野との関係度を推定する場合である。

[0105] （コンピュータ１０００のハードウェア）

図１６は、コンピュータ１０００の構成を示すハードウェア図である。例えば、情報処理システム１、あるいは特徴量抽出部６０３、属性値付与部６０８、データ生成部６１４、及びデータ品質評価部６１２等の情報処理システム１を適宜分散した各システムは、コンピュータ１０００によって実現される。

[0106] コンピュータ１０００は、バス等の内部通信線１００９を介して相互に接続されたＣＰＵをはじめとするプロセッサ１００１、主記憶装置１００２、補助記憶装置１００３、ネットワークインタフェース１００４、入力装置１００５、及び出力装置１００６を備える。

[0107] プロセッサ１００１は、コンピュータ１０００全体の動作制御を司る。また主記憶装置１００２は、例えば揮発性の半導体メモリから構成され、プロセッサ１００１のワークメモリとして利用される。補助記憶装置１００３は、非一時的記憶媒体の一例であり、ハードディスク装置、ＳＳＤ（Solid State Drive）、又はフラッシュメモリ等の大容量の不揮発性の記憶装置から構成され、各種プログラムやデータを長期間保持するために利用される。

[0108] 補助記憶装置１００３に格納された実行可能プログラム１１００がコンピュータ１０００の起動時や必要時に主記憶装置１００２にロードされ、主記憶装置１００２にロードされた実行可能プログラム１１００をプロセッサ１００１が実行する。これにより、各種処理を実行するシステムが実現される。

[0109] なお、実行可能プログラム１１００は、非一時的記録媒体に記録され、媒体読み取り装置によって非一時的記録媒体から読み出されて、主記憶装置１００２にロードされてもよい。または、実行可能プログラム１１００は、ネットワークを介して外部のコンピュータから取得されて、主記憶装置１００２にロードされてもよい。

[0110] ネットワークインタフェース１００４は、コンピュータ１０００をシステム内の各ネットワークに接続する、あるいは他のコンピュータと通信するた

めのインタフェース装置である。ネットワークインタフェース１００４は、例えば、有線ＬＡＮ（Local Area Network）や無線ＬＡＮ等のＮＩＣ（Network Interface Card）から構成される。

[0111] 入力装置１００５は、キーボードや、マウス等のポインティングデバイス等から構成され、ユーザがコンピュータ１０００に各種指示や情報を入力するために利用される。出力装置１００６は、例えば、液晶ディスプレイ又は有機ＥＬ（Electro Luminescence）ディスプレイ等の表示装置や、スピーカ等の音声出力装置から構成され、必要時に必要な情報をユーザに提示するために利用される。

[0112] なお、本発明は前述した実施形態に限定されるものではなく、添付した特許請求の範囲の趣旨内における様々な変形例及び同等の構成が含まれる。例えば、前述した実施形態は本発明を分かりやすく説明するために詳細に説明したものであり、必ずしも説明した全ての構成を備えるものに本発明は限定されない。また、ある実施形態の構成の一部を他の実施形態の構成に置き換えてもよい。また、ある実施形態の構成に他の実施形態の構成を加えてもよい。また、各実施形態の構成の一部について、他の構成の追加、削除、又は置換をしてもよい。

[0113] また、前述した各構成、機能、処理部、処理手段等は、それらの一部又は全部を、例えば集積回路で設計する等により、ハードウェアで実現してもよい。あるいは、プロセッサがそれぞれの機能を実現するプログラムを解釈し実行することにより、ソフトウェアで実現してもよい。

[0114] 各機能を実現するプログラム、テーブル、ファイル等の情報は、メモリ、ハードディスク、ＳＳＤ（Solid State Drive）等の記憶装置、又は、ＩＣ（Integrated Circuit）カード、ＳＤカード、ＤＶＤ（Digital Versatile Disc）の非一時的記録媒体に格納することができる。

[0115] また、制御線や情報線は説明上必要と考えられるものを示しており、実装上必要な全ての制御線や情報線を示しているとは限らない。実際には、ほとんど全ての構成が相互に接続されていると考えてよい。

符号の説明

[0116] １：情報処理システム、２０１：基準データ、２０２：評価データ、２０４：潜在変数、６０３：特徴量抽出部、６０８：属性値付与部、６１２：データ品質評価部、６１４：データ生成部、１０００：コンピュータ。

請求の範囲

[請求項1] 処理部と記憶部とを有する情報処理システムが実行する情報処理方法であって、

前記処理部が、

データの複数の属性に属性値が付与されている基準データを含んだ入力データを、該データの前記複数の属性のそれぞれに関する潜在変数を出力するＭＦＣＶＡＥ（Multi-Facet Clustering Variational Auto-Encoder）に入力する第１ステップと、

前記入力データの入力に対して前記ＭＦＣＶＡＥから出力された前記潜在変数を説明変数とし、前記属性値を目的変数とする回帰モデルを前記属性毎に設定する第２ステップと、

前記潜在変数と前記属性値とから、該属性値に対する予測誤差が最小となる前記属性値の予測値及び前記回帰モデルの回帰係数を前記属性毎に算出する第３ステップと、

前記第３ステップによって算出された前記属性毎の前記予測値及び前記回帰係数に基づいて、前記潜在変数及び前記属性値の前記回帰モデルへの適合が良いほど小さい値を取る指標を前記属性毎に算出する第４ステップと、

前記ＭＦＣＶＡＥによるデータ再構成の誤差を表す再構成誤差項と、前記潜在変数の分布に制約を与える正則化項と、を有する前記ＭＦＣＶＡＥの損失関数に、前記属性毎の前記指標に基づく追加項を追加した損失関数の関数値を算出する第５ステップと、

前記第５ステップによって算出された前記関数値に基づく誤差逆伝搬によって前記ＭＦＣＶＡＥのモデルパラメータを更新する第６ステップと、を実行し、

前記第１ステップから前記第６ステップまでを、前記予測誤差又はエポック回数が所定条件を充足するまでこの順序で繰り返すことで前記ＭＦＣＶＡＥのモデル学習を実行する、ことを特徴とする情報処理

方法。

[請求項2] 請求項１に記載の情報処理方法であって、

前記回帰モデルは、重回帰モデルである、ことを特徴とする情報処理方法。

[請求項3] 請求項１に記載の情報処理方法であって、

前記属性毎の前記指標は、前記回帰モデルの決定係数である、ことを特徴とする情報処理方法。

[請求項4] 請求項１に記載の情報処理方法であって、

前記属性毎の前記指標は、前記予測誤差である、ことを特徴とする情報処理方法。

[請求項5] 請求項４に記載の情報処理方法であって、

前記予測誤差は、平均二乗誤差である、ことを特徴とする情報処理方法。

[請求項6] 請求項１に記載の情報処理方法であって、

前追加項は、前記属性毎の前記指標に、前記属性毎の重み係数を乗算した項であり、

前記処理部が、

前記属性毎の前記重み係数を、前記属性毎に、前記指標と前記再構成誤差項及び前記正則化項との各絶対値のオーダーが等しくなるように決定する、ことを特徴とする情報処理方法。

[請求項7] 請求項１に記載の情報処理方法であって、

前記入力データは、前記基準データ及び前記属性に前記属性値が付与されていない評価データを含み、

前記処理部が、前記第５ステップにおいて、

前記追加項を、前記基準データを用いて計算し、

前記再構成誤差項及び前記正則化項を、前記基準データ及び前記評価データの何れか一方又は両方を用いて計算する、ことを特徴とする情報処理方法。

[請求項8] 請求項１に記載の情報処理方法であって、

前記入力データは、前記基準データ及び前記属性に前記属性値が付与されていない評価データを含み、

前記処理部が、

前記第１ステップから前記第６ステップまでを繰り返すことでモデル学習済みの前記ＭＦＣＶＡＥに前記評価データを入力し、前記評価データに関する前記潜在変数を取得する第７ステップと、

前記第７ステップによって取得された前記潜在変数と、前記ＭＦＣＶＡＥのモデル学習の最終エポック時における前記回帰係数とに基づいて、前記評価データの前記属性値が付与されていない前記属性の前記属性値の予測値を算出して該評価データに該属性値として付与する第８ステップと、を実行することを特徴とする情報処理方法。

[請求項9] 請求項８に記載の情報処理方法であって、

前記処理部が、

前記基準データと前記第８ステップによって前記予測値が付与された前記評価データとの前記属性及び前記属性値に関する情報を出力する第９ステップ、を実行することを特徴とする情報処理方法。

[請求項10] 請求項８に記載の情報処理方法であって、

前記処理部が、

前記基準データと前記第８ステップによって前記予測値が付与された前記評価データとを用いて、データの設計の十分性、データの被覆性、又はデータの均一性を含む観点に従って前記入力データを評価する第１０ステップ、を実行することを特徴とする情報処理方法。

[請求項11] 請求項１に記載の情報処理方法であって、

複数の前記情報処理システムの各前記処理部が、

前記第１ステップから前記第６ステップまでを、前記予測誤差又はエポック回数が所定条件を充足するまで繰り返すことで前記ＭＦＣＶＡＥをモデル学習することを、それぞれ異なる前記入力データを用い

て実行し、

各前記処理部によって得られた前記ＭＦＣＶＡＥのモデルの学習結果をマージして出力する第１１ステップ、を実行することを特徴とする情報処理方法。

[請求項12] 請求項８に記載の情報処理方法であって、

複数の前記情報処理システムの各前記処理部が、

前記第１ステップから前記第８ステップまでを、それぞれ異なる前記入力データを用いて実行し、

各前記処理部によって得られた、前記基準データと前記第８ステップによって前記予測値が付与された前記評価データの前記属性及び前記属性値に関する情報をマージして出力する第１２ステップ、を実行することを特徴とする情報処理方法。

[請求項13] 請求項１に記載の情報処理方法であって、

前記処理部が、

指定された前記属性及び前記属性値を、前記第１ステップから前記第６ステップまでを繰り返すことでモデル学習済みの前記ＭＦＣＶＡＥに入力し、入力された前記属性及び前記属性値と前記回帰係数とから前記潜在変数を算出し、該潜在変数を基に、入力された前記属性及び前記属性値に対応する前記データを再構成する第１３ステップ、を実行することを特徴とする情報処理方法。

[請求項14] 請求項１３に記載の情報処理方法であって、

前記処理部が、

前記指定された前記属性及び前記属性値に該当する前記基準データが存在する場合には、該基準データに対応する前記データを再構成したデータとして採用し、

前記指定された前記属性及び前記属性値に該当する前記基準データが存在しない場合に、前記第１３ステップを実行する、ことを特徴とする情報処理方法。

[請求項15] 請求項１に記載の情報処理方法であって、

前記基準データは活字及び手書き文字を含み、前記評価データは手書き文字を含む、ことを特徴とする情報処理方法。

[請求項16] データの複数の属性に属性値が付与されている基準データを含んだ入力データを、該データの前記複数の属性のそれぞれに関する潜在変数を出力するＭＦＣＶＡＥ（Multi-Facet Clustering Variational Auto-Encoder）に入力し、

前記入力データの入力に対して前記ＭＦＣＶＡＥから出力された前記潜在変数を説明変数とし、前記属性値を目的変数とする回帰モデルを前記属性毎に設定し、

前記潜在変数と前記属性値とから、該属性値に対する予測誤差が最小となる前記属性値の予測値及び前記回帰モデルの回帰係数を前記属性毎に算出し、

算出された前記属性毎の前記予測値及び前記回帰係数に基づいて、前記潜在変数及び前記属性値の前記回帰モデルへの適合が良いほど小さい値を取る指標を前記属性毎に算出する回帰モデル適合度評価部と、

前記ＭＦＣＶＡＥによるデータ再構成の誤差を表す再構成誤差項と、前記潜在変数の分布に制約を与える正則化項と、を有する前記ＭＦＣＶＡＥの損失関数に、前記属性毎の前記指標に基づく追加項を追加した損失関数の関数値を算出する損失算出部と、

前記損失算出部によって算出された前記関数値に基づく誤差逆伝搬によって前記ＭＦＣＶＡＥのモデルパラメータを更新するモデル更新部と、を有し、

前記回帰モデル適合度評価部、前記損失算出部、及び前記モデル更新部は、前記予測誤差又はエポック回数が所定条件を充足するまでこの順序で処理を順次繰り返すことで前記ＭＦＣＶＡＥのモデル学習を実行する、ことを特徴とする情報処理システム。

[請求項17] コンピュータを情報処理システムとして機能させるための情報処理プログラムであって、

前記コンピュータを、

データの複数の属性に属性値が付与されている基準データを含んだ入力データを、該データの前記複数の属性のそれぞれに関する潜在変数を出力するＭＦＣＶＡＥ（Multi-Facet Clustering Variational Auto-Encoder）に入力し、

前記入力データの入力に対して前記ＭＦＣＶＡＥから出力された前記潜在変数を説明変数とし、前記属性値を目的変数とする回帰モデルを前記属性毎に設定し、

前記潜在変数と前記属性値とから、該属性値に対する予測誤差が最小となる前記属性値の予測値及び前記回帰モデルの回帰係数を前記属性毎に算出し、

算出された前記属性毎の前記予測値及び前記回帰係数に基づいて、前記潜在変数及び前記属性値の前記回帰モデルへの適合が良いほど小さい値を取る指標を前記属性毎に算出する回帰モデル適合度評価部と、

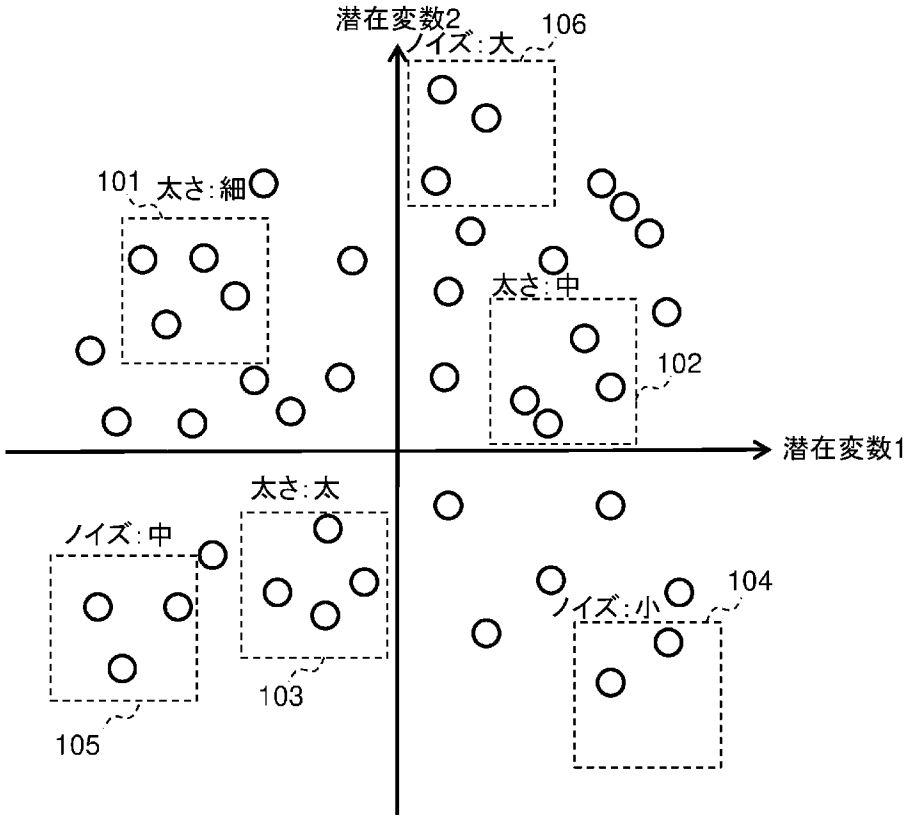
前記ＭＦＣＶＡＥによるデータ再構成の誤差を表す再構成誤差項と、前記潜在変数の分布に制約を与える正則化項と、を有する前記ＭＦＣＶＡＥの損失関数に、前記属性毎の前記指標に基づく追加項を追加した損失関数の関数値を算出する損失算出部と、

前記損失算出部によって算出された前記関数値に基づく誤差逆伝搬によって前記ＭＦＣＶＡＥのモデルパラメータを更新するモデル更新部と、して機能させ、

前記回帰モデル適合度評価部、前記損失算出部、及び前記モデル更新部は、前記予測誤差又はエポック回数が所定条件を充足するまでこの順序で処理を順次繰り返すことで前記ＭＦＣＶＡＥのモデル学習を実行する、ことを特徴とする情報処理プログラム。

[図1]

図 1



[図2]

図 2

データ番号 (データ属性)	属性				
	文字の種類	太さ(mm)	傾き(度)	…	砕け度合い (レベル)
1 (活字)	あ	0.3	3.0	…	3
2 (活字)	い	0.2	1.0	…	2
3 (活字)	う	0.5	5.0	…	4
4 (手書き文字)	あ			…	
5 (手書き文字)	か			…	
…					

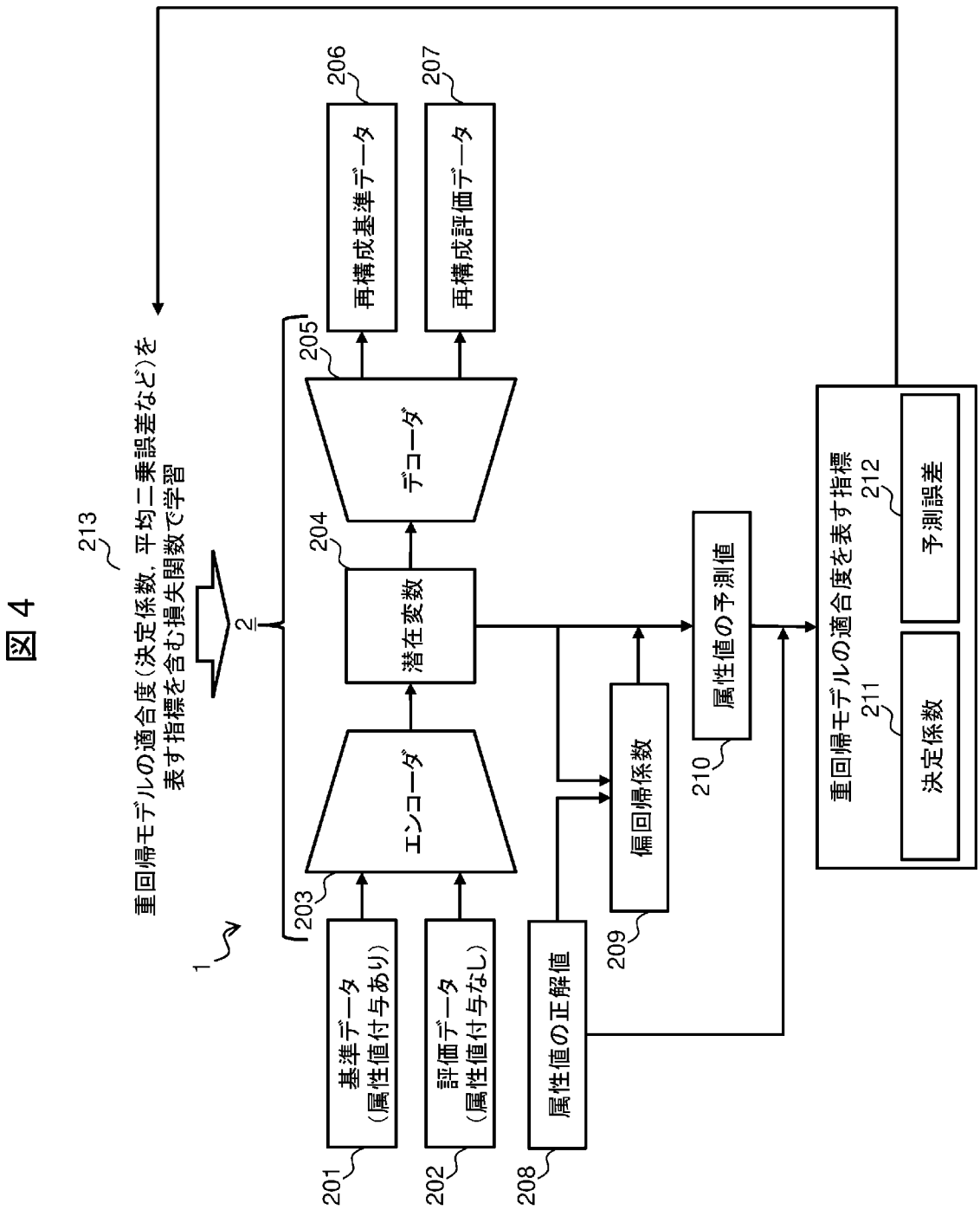
[図3]

図 3

データ番号	属性				
	属性1	属性2	属性3	…	属性J
1	0.1	0.3		…	
2	0.5	0.2		…	
3	0.6		3.0	…	
4	0.7		1.0		2
5		0.5	5.0	…	3
…					

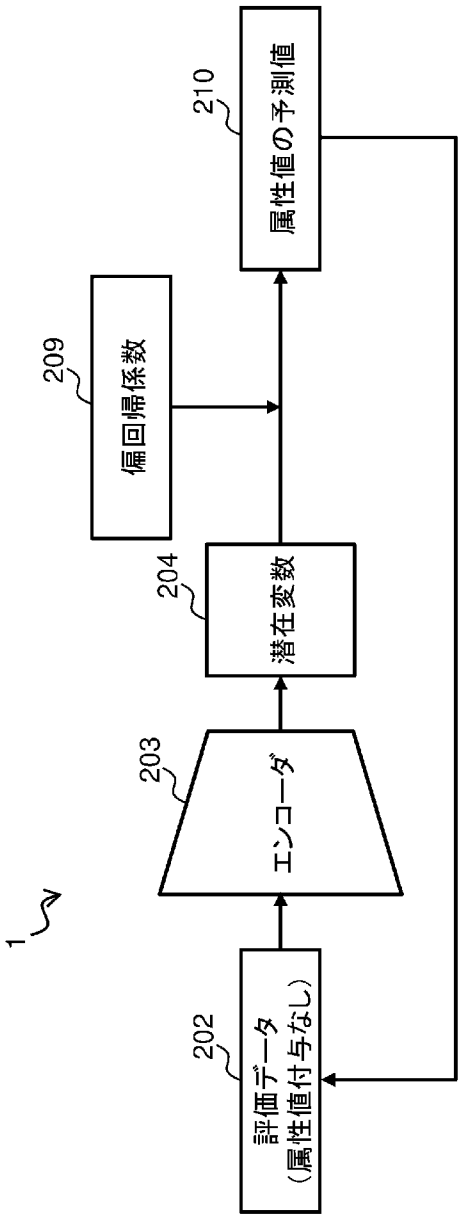
[図4]

図 4


重回帰モデルの適合度(決定係数, 平均二乗誤差など)を
表す指標を含む損失関数で学習
213
1
2
201
基準データ
(属性値付与あり)
202
評価データ
(属性値付与なし)
203
エンコーダ
204
潜在変数
205
デコーダ
206
再構成基準データ
207
再構成評価データ
208
属性値の正解値
209
偏回帰係数
210
属性値の予測値
211
決定係数
212
予測誤差
重回帰モデルの適合度を表す指標

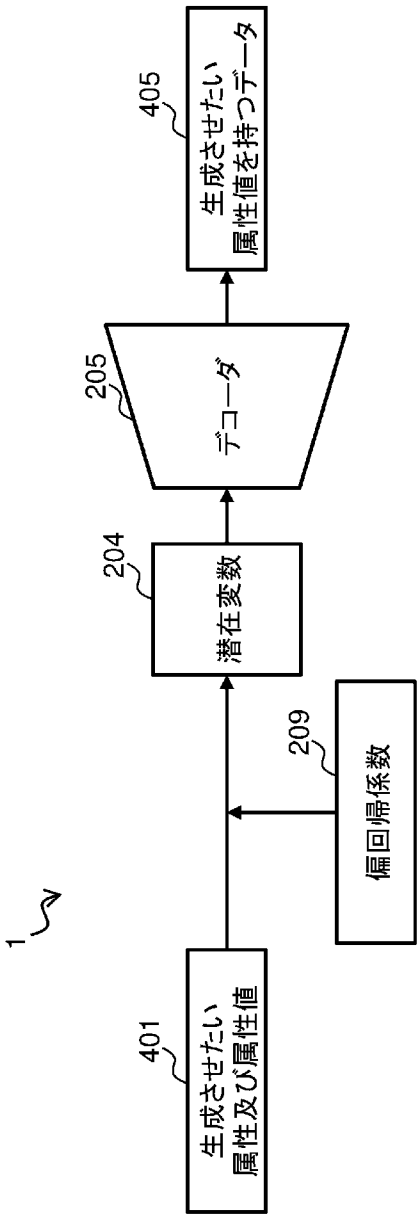
[図5]

図 5



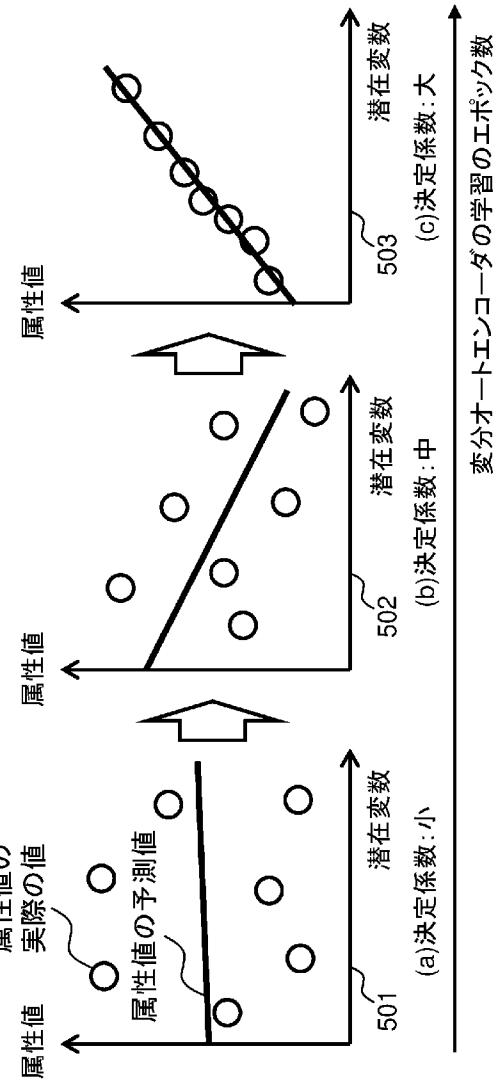
[図6]

図 6


1
401
生成させたい
属性及び属性値
209
偏回帰係数
204
潜在変数
205
デコーダ
405
生成させたい
属性値を持つデータ

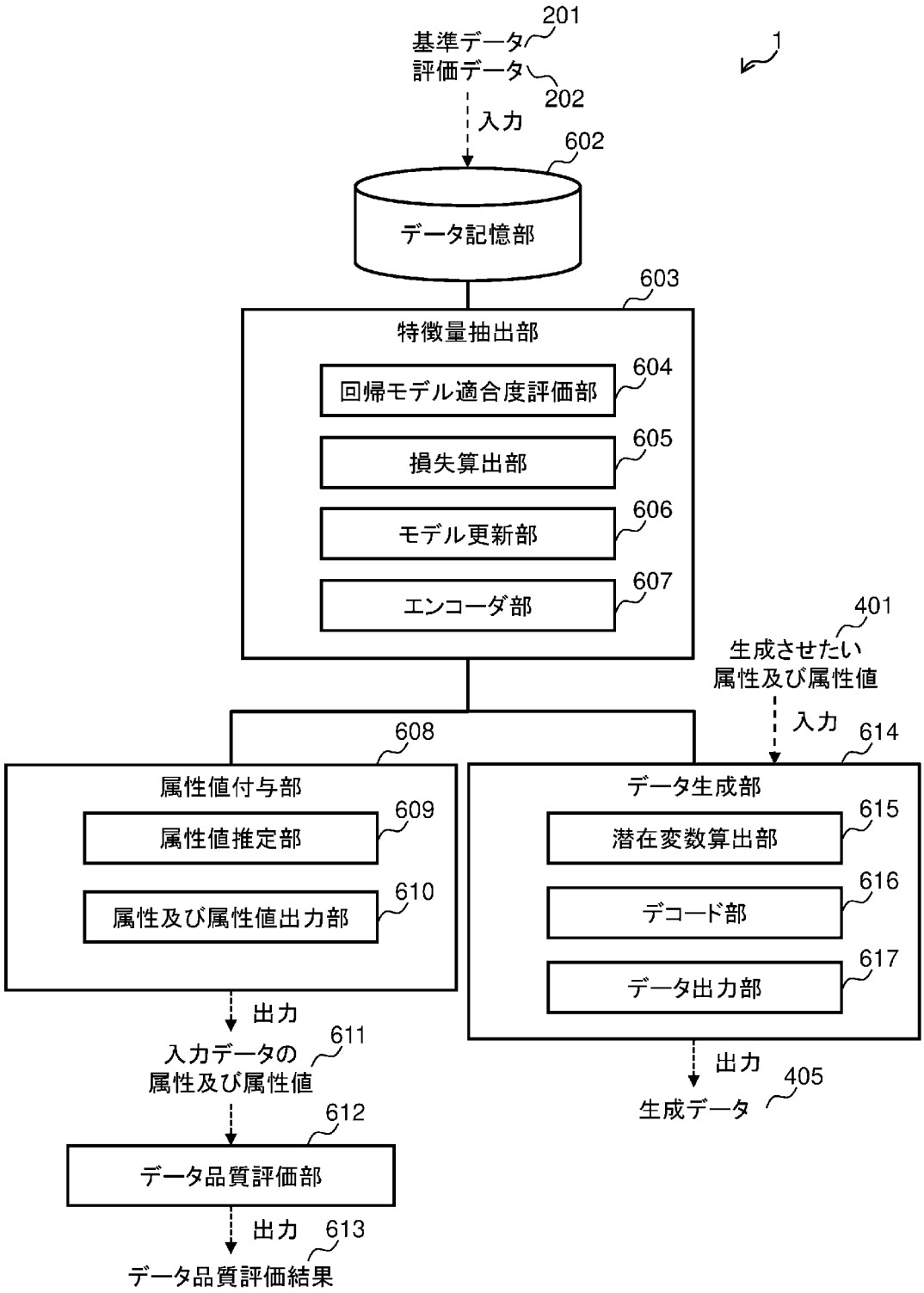
[図7]

図 7


属性値
属性値の
実際の値
属性値の予測値
潜在変数
501
(a)決定係数:小
属性値
潜在変数
502
(b)決定係数:中
属性値
潜在変数
503
(c)決定係数:大
変分オートエンコーダの学習のエポック数

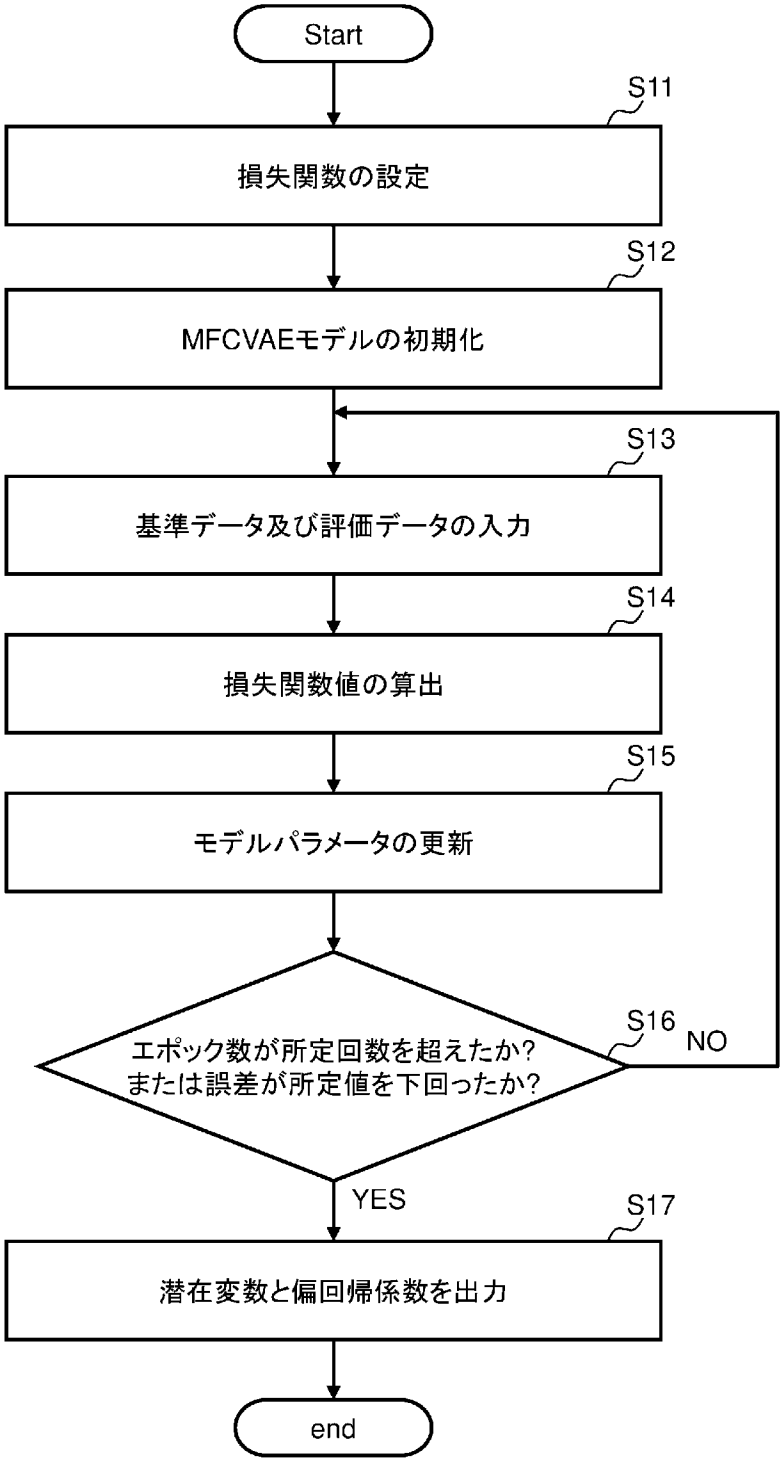
[図8]

図 8


201
基準データ
評価データ
202
入力
602
データ記憶部
603
特徴量抽出部
604
回帰モデル適合度評価部
605
損失算出部
606
モデル更新部
607
エンコーダ部
401
生成させたい
属性及び属性値
入力
608
属性値付与部
609
属性値推定部
610
属性及び属性値出力部
614
データ生成部
615
潜在変数算出部
616
デコード部
617
データ出力部
出力
611
入力データの
属性及び属性値
612
データ品質評価部
出力
613
データ品質評価結果
出力
405
生成データ
1

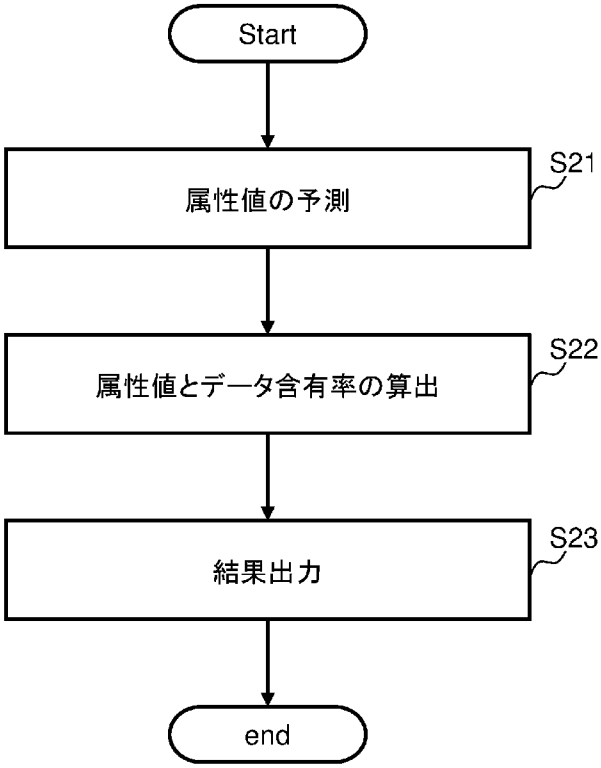
[図9]

図 9


Start
S11
損失関数の設定
S12
MFCVAEモデルの初期化
S13
基準データ及び評価データの入力
S14
損失関数値の算出
S15
モデルパラメータの更新
S16
エポック数が所定回数を超えたか?
または誤差が所定値を下回ったか?
NO
YES
S17
潜在変数と偏回帰係数を出力
end

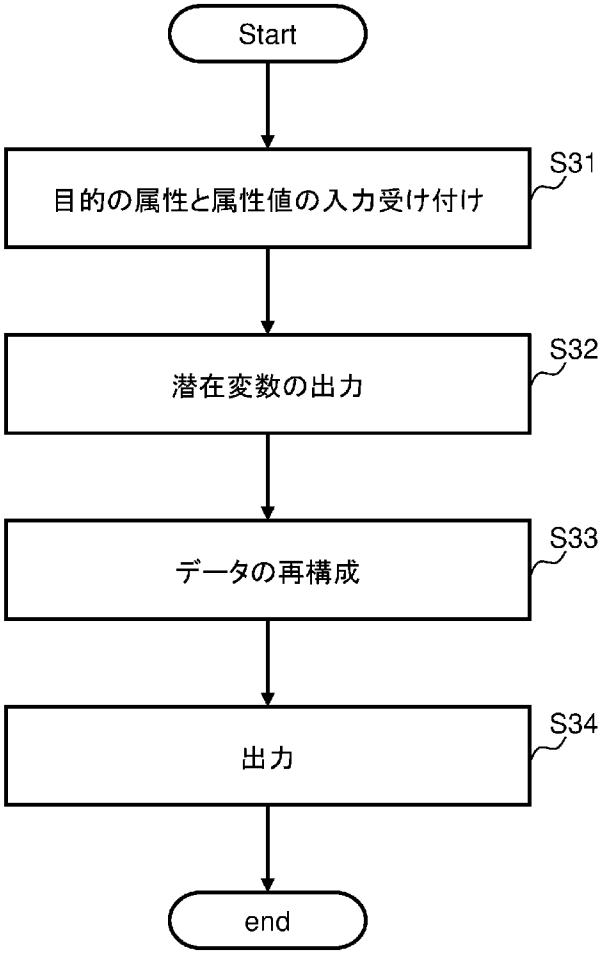
[図10]

図 10



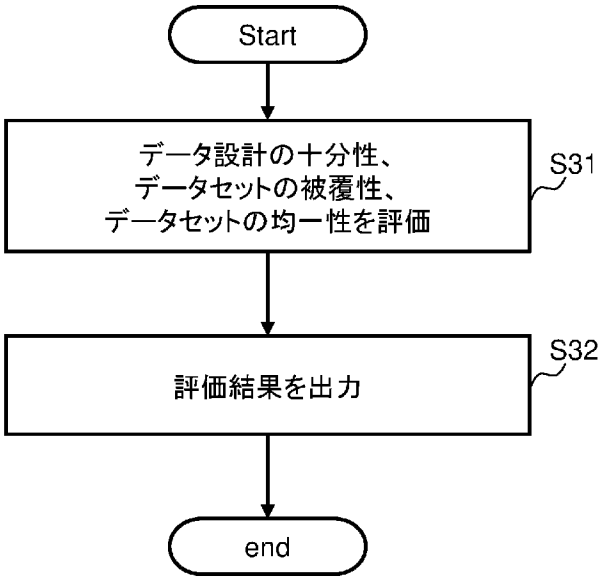
[図11]

図 11


Start
目的の属性と属性値の入力受け付け
S31
潜在変数の出力
S32
データの再構成
S33
出力
S34
end

[図12]

図 12


Start
データ設計の十分性、
データセットの被覆性、
データセットの均一性を評価
S31
評価結果を出力
S32
end

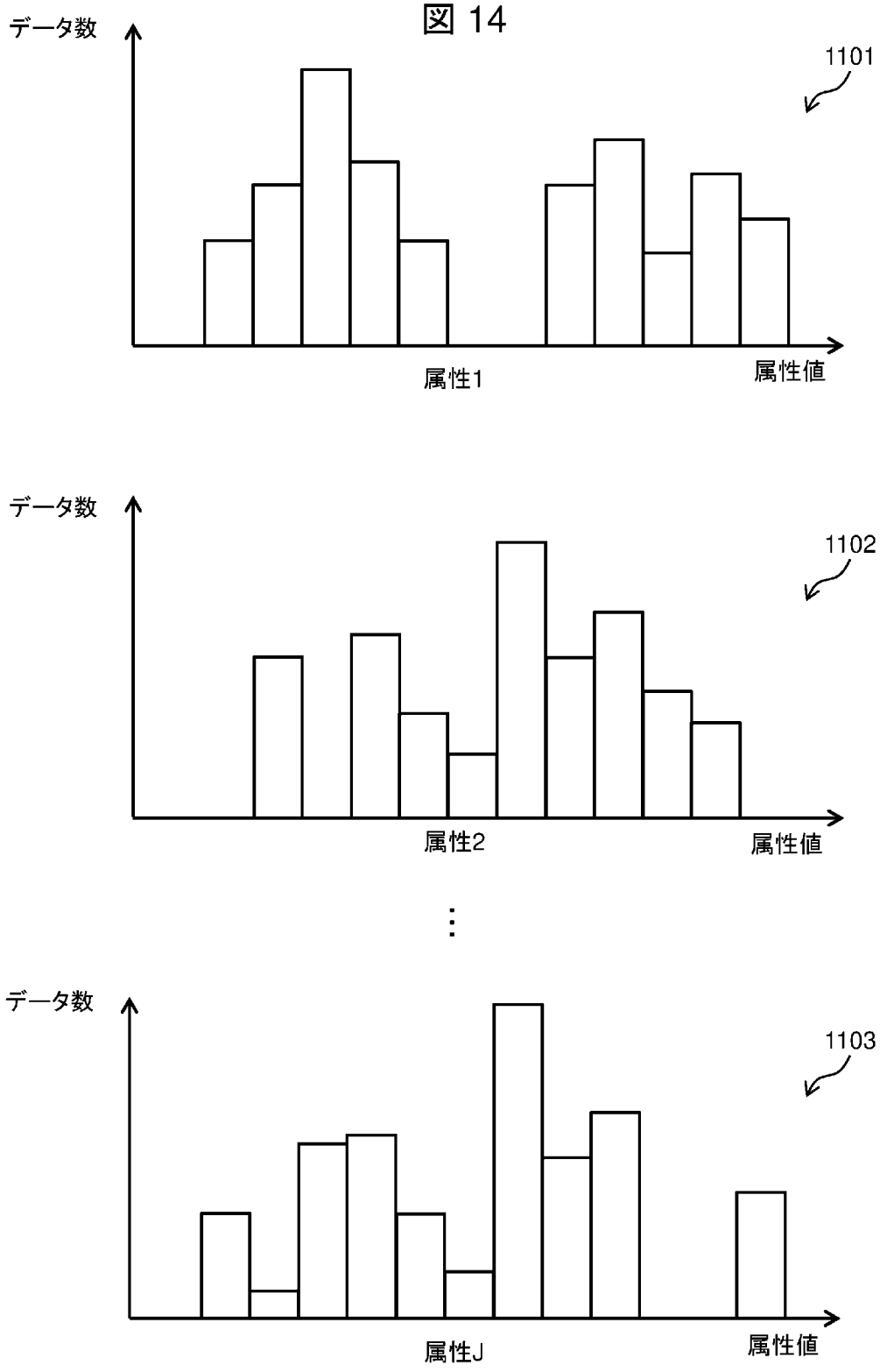
[図13]

図 13

データ番号/属性	属性1	属性2	属性3	…	属性J
1	0.1	0.3	2.0	…	4
2	0.5	0.2	1.0	…	2
3	0.6	0.5	3.0	…	3
4	0.7	0.3	1.0		2
5	0.2	0.5	5.0	…	3

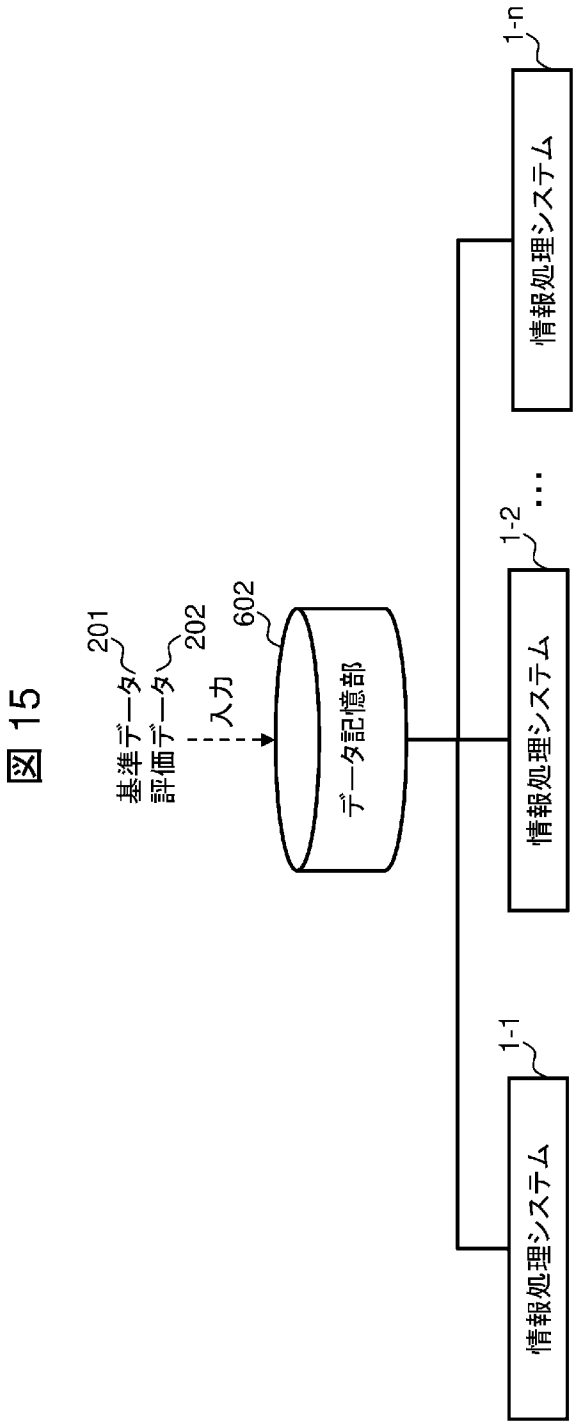
[図14]

図 14



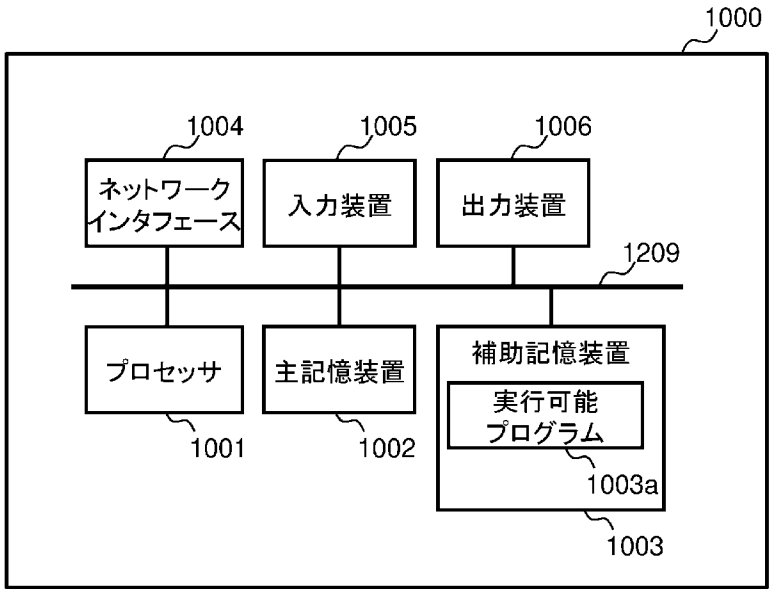
[図15]

図 15


201
基準データ
評価データ
202
入力
602
データ記憶部
1-1
情報処理システム
1-2
情報処理システム
...
1-n
情報処理システム

[図16]

図 16


1000
1004
1005
1006
ネットワーク
インタフェース
入力装置
出力装置
1209
プロセッサ
主記憶装置
補助記憶装置
実行可能
プログラム
1001
1002
1003a
1003

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2023/005451

A. CLASSIFICATION OF SUBJECT MATTER

G06N 3/0455(2023.01)i
FI: G06N3/0455

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)

G06N3/00-99/00

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	FALCK, Fabian et al. Multi-Facet Clustering Variational Autoencoders. arXiv [online]. 29 October 2021, [retrieval date 25 April 2023], Internet: <URL: https://arxiv.org/abs/2106.05241v2> abstract, section 3	1-17
A	JP 2020-144799 A (FUJITSU LTD) 10 September 2020 (2020-09-10) paragraphs [0026]-[0032], fig. 3-4	1-17
P, A	関根 理敏ほか, データセットの多種多様な属性情報抽出に向けた多面的クラスタリング変分オートエンコーダの手書き文字データへの適用, 情報処理学会 シンポジウム ソフトウェアエンジニアリングシンポジウム 2022, 29 August 2022, pp. 145-146 entire text, all drawings, (SEKINE, Masatoshi et al.), non-official translation (Application of multi-facet clustering variational autoencoders to handwritten character data, so as to extract a wide variety of attribute information from datasets. IPSJ symposium: Software Engineering Symposium 2022.)	1-17

☐ 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
"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	Date of mailing of the international search report
25 April 2023	**09 May 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/JP2023/005451

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
JP 2020-144799 A	10 September 2020	US 2020/0285690 A1 paragraphs [0039]-[0045], fig. 3-4 EP 3706029 A1 CN 111666708 A	

国際調査報告

国際出願番号
PCT/JP2023/005451

A. 発明の属する分野の分類（国際特許分類（ＩＰＣ））

G06N 3/0455(2023.01)i
FI: G06N3/0455

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））

G06N3/00-99/00

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	FALCK, Fabian et al., Multi-Facet Clustering Variational Autoencoders, arXiv [online], 2021.10.29, [検索日 2023.04.25], インターネット：<URL: https://arxiv.org/abs/2106.05241v2> Abstract, 第3節	1-17
A	JP 2020-144799 A (富士通株式会社) 10.09.2020 (2020-09-10) [0026]-[0032], 図3-4	1-17
P, A	関根 理敏ほか, データセットの多種多様な属性情報抽出に向けた多面的クラスタリング変分オートエンコーダの手書き文字データへの適用, 情報処理学会 シンポジウム ソフトウェアエンジニアリングシンポジウム ２０２２, 2022.08.29, pp.145-146 全文、全図	1-17

☐ 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	国際調査報告の発送日
25.04.2023	09.05.2023

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）
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/JP2023/005451

引用文献	公表日	パテントファミリー文献	公表日
JP 2020-144799 A	10.09.2020	US 2020/0285690 A1 [0039]-[0045], FIGs.3-4 EP 3706029 A1 CN 111666708 A	