

(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8月6日(06.08.2020)



(10) 国際公開番号

WO 2020/158628 A1

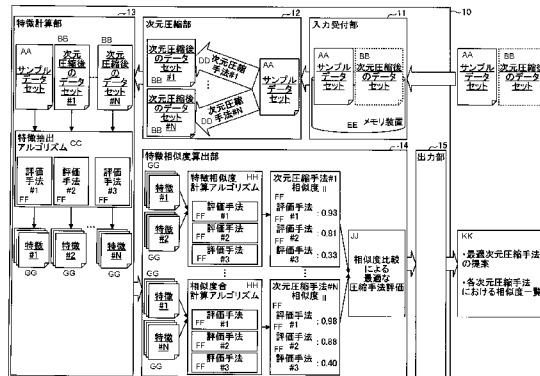
- (51) 国際特許分類:
G06F 17/15 (2006.01) G06F 17/16 (2006.01)
- (21) 国際出願番号: PCT/JP2020/002601
- (22) 国際出願日: 2020年1月24日(24.01.2020)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2019-016463 2019年1月31日(31.01.2019) JP
- (71) 出願人: 日本電信電話株式会社 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目5番1号 Tokyo (JP).
- (72) 発明者: 仲川 宜秀 (NAKAGAWA, Yoshihide); 〒1808585 東京都武蔵野市緑町3丁目9-1 1 N T T 知的財産センタ内 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, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

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

(54) 発明の名称: 評価装置、評価方法及びプログラム

[図4]



- 11 Input accepting unit
12 Dimensionality reduction unit
13 Feature calculating unit
14 Feature similarity computing unit
15 Output unit
AA Sample dataset
BB Dataset after dimensionality reduction
CC Feature extraction algorithm
DD Dimensionality reduction technique
EE Memory device
FF Evaluation technique
GG Feature
HH Feature similarity calculating algorithm
II Similarity for dimensionality reduction technique #1, N
JJ Evaluation of optimal reduction technique by comparison of similarities
KK Proposal of optimal dimensionality reduction technique
List of similarities using each dimensionality reduction technique

(57) Abstract: This evaluating device for evaluating a plurality of dimensionality reduction techniques includes: a feature calculating unit which employs a plurality of feature extraction algorithms to extract a first feature of a dataset before dimensionality reduction and a second feature of a dataset after dimensionality reduction, from the dataset before dimensionality reduction and the dataset after dimensionality reduction, for each of the plurality of dimensionality reduction techniques; a feature similarity computing unit which uses a plurality of feature similarity calculating algorithms corresponding respectively to the plurality of feature extraction algorithms to compute a similarity between the first feature and the second feature; and an output unit which outputs the similarities computed for each of the plurality of dimensionality reduction techniques.

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

添付公開書類:

一 国際調査報告 (条約第21条(3))

(57) 要約: 複数の次元圧縮手法を評価する評価装置は、複数の特徴抽出アルゴリズムを用いて、前記複数の次元圧縮手法のそれぞれに対して、次元圧縮前のデータセット及び次元圧縮後のデータセットから、当該次元圧縮前のデータセットの第1の特徴と、当該次元圧縮後のデータセットの第2の特徴とを抽出する特徴計算部と、前記複数の特徴抽出アルゴリズムにそれぞれ対応する複数の特徴相似度計算アルゴリズムを用いて、前記第1の特徴と前記第2の特徴との間の相似度を算出する特徴相似度算出部と、前記複数の次元圧縮手法のそれぞれに対して算出された前記相似度を出力する出力部とを有する。

明 細 書

発明の名称： 評価装置、評価方法及びプログラム

技術分野

[0001] 本発明は、次元圧縮手法を評価する評価装置、評価方法及びプログラムに関する。

背景技術

[0002] 機械学習を適用する分野において、学習の高速化やデータの可視化を行う上で、学習用にあらかじめ与えられるサンプルデータと答えデータのセットについて、サンプルデータの特徴量（例えば、人体に関するデータセットにおける身長や体重などのような、サンプルデータを特徴づけるプロパティ）の数（次元）を圧縮することが試みられている。

[0003] 特徴量の次元圧縮は、主に機械学習やビッグデータ分析に使われる。あるサンプルデータセットが膨大な特徴量を持つ場合、機械学習、分析に膨大な時間が必要であり、人間がサンプルデータセットのばらつきを目視できないといった課題がある。そこでデータセットの特徴をできるだけ保持し、特徴量を次元圧縮することにより、可視化や高速化を行うことが可能になる。特徴量の次元圧縮手法については種々あり、適切な次元圧縮手法を評価するための手法として、従来では次元圧縮後のデータセットをグラフなどを用いて定性的に評価する手法がある。図1に示すように、サンプルデータセットを次元圧縮手法#1～#3で次元圧縮すると、それぞれの次元圧縮手法について次元圧縮後のデータセットが得られる。説明を簡単にするために、図1の各データセットの下に、次元圧縮前のデータセットの特徴を3次元のグラフで示し、次元圧縮後のデータセットの特徴を2次元のグラフで示す。グラフによる定性的な評価とは、次元圧縮手法#1～#3のうち、どの手法が元のサンプルデータセットの特徴をよりとらえているかをそれぞれのグラフから視覚的に評価する手法である。

[0004] また、遺伝子解析について局所的な分布の相関に基づいて次元圧縮手法を

評価する技術が提案されている（非特許文献1）。

先行技術文献

非特許文献

- [0005] 非特許文献1: Samuel Kaski, et. al., “Trustworthiness and metrics in visualizing similarity of gene expression”, BMC Bioinformatics, 13 October 2003

発明の概要

発明が解決しようとする課題

- [0006] 上記のように特徴量の次元圧縮手法については種々あり、次元圧縮後のデータセットについてどの程度機械学習・分析に有意な情報が残されているかを評価することが望ましい。図1に示す従来の次元圧縮手法の評価は定性的な評価であるので、次元数が増加したときに評価が困難になることがあり、必ずしも適切な評価が行われるとは限らない。また、非特許文献1における次元圧縮手法の評価は、局所的な分布の相関を評価するものであり、局所的な分布の相関が小さい場合には適用が困難である。さらに、従来では、1つの手法による評価にとどまり、複数観点から評価することができないという課題がある。

- [0007] 本発明は、複数の観点から次元圧縮手法を評価する技術を提供することを目的とする。

課題を解決するための手段

- [0008] 本発明の一形態に係る評価装置は、
複数の次元圧縮手法を評価する評価装置であって、
複数の特徴抽出アルゴリズムを用いて、前記複数の次元圧縮手法のそれぞれに対して、次元圧縮前のデータセット及び次元圧縮後のデータセットから、当該次元圧縮前のデータセットの第1の特徴と、当該次元圧縮後のデータセットの第2の特徴とを抽出する特徴計算部と、
前記複数の特徴抽出アルゴリズムにそれぞれ対応する複数の特徴相類似度計

算アルゴリズムを用いて、前記第 1 の特徴と前記第 2 の特徴との間の相似度を算出する特徴相似度算出部と、

前記複数の次元圧縮手法のそれぞれに対して算出された前記相似度を出力する出力部と、

を有することを特徴とする。

[0009] また、本発明の一形態に係る評価方法は、

複数の次元圧縮手法を評価する評価装置が実行する評価方法であって、

複数の特徴抽出アルゴリズムを用いて、前記複数の次元圧縮手法のそれぞれに対して、次元圧縮前のデータセット及び次元圧縮後のデータセットから、当該次元圧縮前のデータセットの第 1 の特徴と、当該次元圧縮後のデータセットの第 2 の特徴とを抽出するステップと、

前記複数の特徴抽出アルゴリズムにそれぞれ対応する複数の特徴相似度計算アルゴリズムを用いて、前記第 1 の特徴と前記第 2 の特徴との間の相似度を算出するステップと、

前記複数の次元圧縮手法のそれぞれに対して算出された前記相似度を出力するステップと、

を有することを特徴とする。

[0010] また、本発明の一形態に係るプログラムは、

上記の評価装置の各部としてコンピュータを機能させることを特徴とする。

発明の効果

[0011] 本発明によれば、複数の観点から次元圧縮手法を評価することが可能になる。

図面の簡単な説明

[0012] [図1]従来技術における次元圧縮手法の評価手法を示す図である。

[図2]本発明の実施の形態におけるネットワーク構成例を示す図である。

[図3]本発明の実施の形態における評価装置を構成するコンピュータのハードウェア構成例を示す図である。

[図4]本発明の実施の形態における評価装置の機能構成例を示す図である。

[図5]特徴計算部及び特徴相似度算出部の処理を示すフローチャートである。

発明を実施するための形態

[0013] 以下、図面に基づいて本発明の実施の形態を説明する。

[0014] 図2は、本発明の実施の形態におけるネットワーク構成例を示す図である。図2において、評価装置10は、インターネット又はLAN (Local Area Network) 等のネットワークを介して1以上のユーザ端末20と接続される。

[0015] 評価装置10は、種々の次元圧縮手法に依存せず、次元圧縮前後のデータセットの特徴の類似性を、複数の観点から定量的に評価できるサーバ等の装置である。特徴の類似性を数値化するために、後述する特徴計算部により各データセットの特徴を算出し、後述する特徴相似度算出部により相似度を定量化し、最適な次元圧縮手法及び相似度の一覧をユーザ端末20に返却する。

[0016] ユーザ端末20は、評価装置10に対するデータや評価条件の入力をユーザから受け付けたり、評価装置10による評価結果を出力（表示）したりする端末である。例えば、PC (Personal Computer)、スマートフォン、タブレット端末等がユーザ端末20として利用されてもよい。

[0017] 図3は、本発明の実施の形態における評価装置10を構成するコンピュータのハードウェア構成例を示す図である。評価装置10を構成するコンピュータは、それぞれバスBで相互に接続されているドライブ装置100、補助記憶装置102、メモリ装置103、CPU104、及びインタフェース装置105等を有する。

[0018] 評価装置10での処理を実現するプログラムは、CD-ROM等の記録媒体101によって提供される。プログラムを記憶した記録媒体101がドライブ装置100にセットされると、プログラムが記録媒体101からドライブ装置100を介して補助記憶装置102にインストールされる。但し、プログラムのインストールは必ずしも記録媒体101より行う必要はなく、ネットワークを介して他のコンピュータよりダウンロードするようにしてもよ

い。補助記憶装置１０２は、インストールされたプログラムを格納すると共に、必要なファイルやデータ等を格納する。

[0019] メモリ装置１０３は、プログラムの起動指示があった場合に、補助記憶装置１０２からプログラムを読み出して格納する。ＣＰＵ１０４は、メモリ装置１０３に格納されたプログラムに従って評価装置１０に係る機能を実行する。インタフェース装置１０５は、ネットワークに接続するためのインタフェースとして用いられる。

[0020] 図４は、本発明の実施の形態における評価装置１０の機能構成例を示す図である。図４において、評価装置１０は、入力受付部１１、次元圧縮部１２、特徴計算部１３、特徴相似度算出部１４及び出力部１５等を有する。これら各部は、評価装置１０にインストールされた１以上のプログラムが、ＣＰＵ１０４に実行させる処理により実現される。

[0021] 入力受付部１１は、ユーザ端末２０において入力されたサンプルデータセット（次元圧縮前のデータセット）と次元圧縮後のデータセットと評価条件を当該ユーザ端末２０から受信し、サンプルデータセットと次元圧縮後のデータセットとをメモリ装置１０３等に格納する。

[0022] 入力受付部１１がユーザ端末２０から受信するサンプルデータセットは、トラフィックデータやセンサーデータのようなデータの集合である。例えば、各トラフィックデータは、ＩＰ、ポート、プロトコル、パケット数、長さ等の複数の特徴量で構成される。次元圧縮後のデータセットとは、サンプルデータセットの特徴量の数（次元）を圧縮した後のデータセットである。なお、後述する次元圧縮部１２において次元圧縮する場合には、入力受付部１１は次元圧縮後のデータセットを受け付ける必要はない。

[0023] 評価条件は、後述する複数の評価手法のうちどの評価手法を使用して次元圧縮手法を評価するか（複数選択可）を含み、また、次元圧縮部１２において次元圧縮する場合には、評価条件は評価対象の次元圧縮手法（複数選択可）を含む。

[0024] 次元圧縮部１２は、入力受付部１１において次元圧縮後のデータセットを

受け付けない場合に、入力受付部 1 1 において受け付けた評価対象の次元圧縮手法を用いて、サンプルデータセットの次元圧縮を実施し、次元圧縮後のデータセットを生成する。

[0025] 特徴計算部 1 3 は、入力受付部 1 1 又は次元圧縮部 1 2 からサンプルデータセット及び次元圧縮後のデータセットを受け取り、複数の特徴抽出アルゴリズムを用いて、各次元圧縮手法に対してサンプルデータセットの特徴と、次元圧縮後のデータセットの特徴とを抽出する。特徴計算部 1 3 は、これらの特徴を行列又はベクトルに変換してもよい。

[0026] 特徴相似度算出部 1 4 は、特徴計算部 1 3 で用いた複数の特徴抽出アルゴリズムにそれぞれ対応する複数の特徴相似度計算アルゴリズムを用いて、サンプルデータセットの特徴を表す行列又はベクトルと次元圧縮後のデータセットの特徴を表す行列又はベクトルとの相似度を算出する。相似度が高いほど、次元圧縮前後のデータセットの特徴が似ているとすることができる。特徴相似度算出部 1 4 は、それぞれの次元圧縮手法において算出された相似度に基づいて、最適な次元圧縮手法を決定することができる。

[0027] 出力部 1 5 は、最適な次元圧縮手法の提案、各次元圧縮手法における相似度一覧を出力する。

[0028] 以下に、次元圧縮手法を評価するための具体的な 3 つの評価手法に関して、特徴計算部 1 3 及び特徴相似度算出部 1 4 の機能を説明する。3 つの評価手法のうちどの評価手法を使用するかは、入力受付部 1 1 で受け付けた評価条件に従う。

[0029] (1) 評価手法 # 1 : 各点の大局的な分布の特徴を抽出し、相似度を算出する手法

評価手法 # 1 の特徴抽出アルゴリズムは以下の通りである。

[0030] 特徴計算部 1 3 は、次元圧縮前のデータセット $A=[a_1, a_2, \dots, a_n]$ 及び次元圧縮後のデータセット $B=[b_1, b_2, \dots, b_n]$ について、それぞれのデータセットの各点（各データ）間の関係を行列表す。各点間の関係は距離及び内積があり、必要に応じてユーザが選択可能とする。各点間の関係を距離で表す場合、次元

圧縮前のデータセットの各点間の距離は以下の行列 R^A で表される。

[0031] [数1]

$$R^A = \begin{bmatrix} r_{11}^A & \cdots & r_{1n}^A \\ \vdots & \ddots & \vdots \\ r_{n1}^A & \cdots & r_{nn}^A \end{bmatrix}, \quad r_{ij}^A = \|a_i - a_j\|_2$$

また、各点間の関係を内積で表す場合、次元圧縮前の各データセットの各点間の内積は以下の行列 R^A で表される。

[0032] [数2]

$$R^A = \begin{bmatrix} r_{11}^A & \cdots & r_{1n}^A \\ \vdots & \ddots & \vdots \\ r_{n1}^A & \cdots & r_{nn}^A \end{bmatrix}, \quad r_{ij}^A = \frac{a_i \cdot a_j}{|a_i| |a_j|}$$

次元圧縮後データの行列 R^B についても同様に算出することができる。

[0033] 評価手法 # 1 の特徴相似度計算アルゴリズムは以下の通りである。

[0034] 特徴相似度算出部 1 4 は、行列 R^A と行列 R^B との間の相関係数を算出する。具体的には、ピアソンの積率相関係数を用いて相似度を算出する。

[0035] (2) 評価手法 # 2 : 各点の局所的な分布の特徴を抽出し、相似度を算出する手法

本評価手法はTrustworthiness計算式（非特許文献 1）を用いる。評価手法 # 2 の特徴抽出アルゴリズムは以下の通りである。

[0036] 特徴計算部 1 3 は、次元圧縮前のデータ及び次元圧縮後のデータをTrustworthiness計算式を用いて分類し、その分類予測を特徴とする。具体的には、

次元圧縮前のデータセット $A=[a_1, a_2, \dots, a_n]$ について、 a_i から最も近いものから順に並べた際の a_j の順位を算出する。また、次元圧縮後のデータセット $B=[b_1, b_2, \dots, b_n]$ について、各点（各データ）に近い方から k 番目の点までを抽出する。

[0037] 評価手法 # 2 の特徴相似度計算アルゴリズムは以下の通りである。

[0038] 特徴相似度算出部 1 4 は、次元圧縮前のデータセット $A=[a_1, a_2, \dots, a_n]$ 及び次元圧縮後のデータセット $B=[b_1, b_2, \dots, b_n]$ について、以下の特徴ベクトル R を計算する。

[0039] [数3]

$$R = [r_1, r_2, \dots, r_n], \quad r_i = \sum_{j \in U_{(b_i)}^k} (r(a_j, a_i) - k)$$

ここで、

[0040] [数4]

$$U_{(b_i)}^k$$

は b_i に最も近いものから k 番目までの点のインデックスの集合（特徴計算部 1 3 において抽出した点の集合）であり、 $r(a_j, a_i)$ は、 a_i から最も近いものから順に並べた際の a_j の順位（特徴計算部 1 3 において算出した順位）を表す。

[0041] 特徴相似度算出部 1 4 は、下記の式によって相似度を算出する。

[0042]

[数5]

$$r = 1 - \frac{2}{nk(2n - 3k - 1)} \sum_{i=1}^n (r_i)$$

(3) 手法 # 3 : 機械学習結果により相似度を算出する手法

評価手法 # 3 の特徴抽出アルゴリズムは以下の通りである。

[0043] 予め、次元圧縮前のデータセット（学習用のデータセット）及び次元圧縮後のデータセット（学習用のデータセット）を機械学習で分類して、次元圧縮前のデータセットの特徴を表すベクトル R^A 及び次元圧縮後のデータセットの特徴を表す R^B を出力する機械学習モデルが構築されているものとする。特徴計算部 1 3 は、次元圧縮前のデータセット $A=[a_1, a_2, \dots, a_n]$ 及び次元圧縮後のデータセット $B=[b_1, b_2, \dots, b_n]$ のそれぞれについて、学習済みの機械学習モデルを用いて分類して得られたベクトル $R^A=[r_1^A, r_2^A, \dots, r_n^A]$ 及び $R^B=[r_1^B, r_2^B, \dots, r_n^B]$ を抽出する。

[0044] 評価手法 # 3 の特徴相似度計算アルゴリズムは以下の通りである。

[0045] 特徴相似度算出部 1 4 では、ベクトル R_A 及び R_B の各成分が一致しているか否かに基づいて下記の式で相似度を算出する。

[0046] [数6]

$$r = \frac{\sum_{i=1}^n f(r_i^A, r_i^B)}{n}, \quad f(x, y) = \begin{cases} 1 & (x = y) \\ 0 & (x \neq y) \end{cases}$$

上記の 3 つの手法によって、図 4 に示す次元圧縮手法 # 1 ~ # N において

、評価手法#1～#3の相似度の一覧が得られる。

[0047] 図5は、特徴計算部13及び特徴相似度算出部14の処理を示すフローチャートである。ここでは、評価装置10において上記の評価手法#1～#3が利用可能であるものとし、評価手法#1～#3のうちどの評価手法を使用するかは、入力受付部11において受け付けているものとする。

[0048] ステップS101において、特徴計算部13は評価手法#1を使用するか否かを判断する。評価手法#1を使用する場合、ステップS102に進み、評価手法#1を使用しない場合、ステップS105に進む。

[0049] ステップS102において、特徴計算部13は、上記の評価手法#1に従ってサンプルデータセットの特徴 R^A を算出する。

[0050] ステップS103において、特徴計算部13は、上記の評価手法#1に従って次元圧縮後のデータセットの特徴 R^B を算出する。

[0051] ステップS104において、特徴相似度算出部14は、上記の評価手法#1に従って相似度を算出する。

[0052] ステップS105において、特徴計算部13は評価手法#2を使用するか否かを判断する。評価手法#2を使用する場合、ステップS106に進み、評価手法#2を使用しない場合、ステップS109に進む。

[0053] ステップS106において、特徴計算部13は、上記の評価手法#2に従って $r(a_j, a_i)$ を抽出する。

[0054] ステップS107において、特徴計算部13は、上記の評価手法#2に従って b_i に最も近いものから k 番目までの点のインデックスの集合を抽出する。

[0055] ステップS108において、特徴相似度算出部14は、上記の評価手法#2に従って相似度を算出する。

[0056] ステップS109において、特徴計算部13は評価手法#3を使用するか否かを判断する。評価手法#3を使用する場合、ステップS110に進み、評価手法#3を使用しない場合、処理を終了する。

[0057] ステップS110において、特徴計算部13は、上記の評価手法#3に従ってサンプルデータセットの特徴 R^A を算出する。

- [0058] ステップS 1 0 3において、特徴計算部 1 3は、上記の評価手法 # 3に従って次元圧縮後のデータセットの特徴 R^b を算出する。
- [0059] ステップS 1 0 4において、特徴相似度算出部 1 4は、上記の評価手法 # 3に従って相似度を算出する。
- [0060] さらに、特徴相似度算出部 1 4は、それぞれの次元圧縮手法において算出された相似度に基づいて、最適な次元圧縮手法を決定する。例えば、特徴相似度算出部 1 4は、得られた相似度と閾値とを比較して、全ての相似度が閾値より高いものが最適な次元圧縮手法であると決定してもよく、また、複数のサンプルデータセットを用いて評価した結果、相似度のばらつきが小さいものが最適な次元圧縮手法であると決定してもよい。
- [0061] 本実施の形態によれば、次元圧縮手法の比較及び選定を行うにあたり、複数の観点から次元圧縮手法を定量的に評価することができ、最適な次元圧縮手法を提案することが可能となる。また、評価手法 # 1は大局的なデータの分布の相似度を計算することができる手法であり、評価手法 # 2は局所的な分布の相関を計算することができる手法であり、評価手法 # 3は実際のデータを用いた評価結果を反映することができる手法である。これらの評価手法を組み合わせることで、様々な観点から次元圧縮手法を評価することが可能になる。
- [0062] 以上、本発明の実施の形態について詳述したが、本発明は斯かる特定の実施形態に限定されるものではなく、特許請求の範囲に記載された本発明の要旨の範囲内において、種々の変形・変更が可能である。

符号の説明

- [0063] 1 0 評価装置
- 1 1 入力受付部
- 1 2 次元圧縮部
- 1 3 特徴計算部
- 1 4 特徴相似度算出部
- 1 5 出力部

20 ユーザ端末

請求の範囲

[請求項1]

複数の次元圧縮手法を評価する評価装置であって、

複数の特徴抽出アルゴリズムを用いて、前記複数の次元圧縮手法のそれぞれに対して、次元圧縮前のデータセット及び次元圧縮後のデータセットから、当該次元圧縮前のデータセットの第1の特徴と、当該次元圧縮後のデータセットの第2の特徴とを抽出する特徴計算部と、

前記複数の特徴抽出アルゴリズムにそれぞれ対応する複数の特徴相類似度計算アルゴリズムを用いて、前記第1の特徴と前記第2の特徴との間の相類似度を算出する特徴相類似度算出部と、

前記複数の次元圧縮手法のそれぞれに対して算出された前記相類似度を出力する出力部と、

を有する評価装置。

[請求項2]

前記複数の特徴抽出アルゴリズムのうち第1の特徴抽出アルゴリズムは、前記次元圧縮前のデータセット内の各データ間の距離又は内積を表す行列を前記第1の特徴として抽出し、前記次元圧縮後のデータセット内の各データ間の距離又は内積を表す行列を前記第2の特徴として抽出するアルゴリズムであり、

前記第1の特徴抽出アルゴリズムに対応する第1の特徴相類似度計算アルゴリズムは、前記第1の特徴と前記第2の特徴との相関係数を算出するアルゴリズムである、請求項1に記載の評価装置。

[請求項3]

前記複数の特徴抽出アルゴリズムのうち第2の特徴抽出アルゴリズムは、学習用の次元圧縮前のデータセット及び学習用の次元圧縮後のデータセットを用いて機械学習によって構築された機械学習モデルを用いて、前記次元圧縮前のデータセット及び前記次元圧縮後のデータセットから、前記第1の特徴を表すベクトルと前記第2の特徴を表すベクトルとを抽出するアルゴリズムであり、

前記第2の特徴抽出アルゴリズムに対応する第2の特徴相類似度計算アルゴリズムは、前記第1の特徴を表すベクトルの各成分と前記第2

の特徴を表すベクトルの各成分とが一致しているか否かに基づいて、相似度を算出するアルゴリズムである、請求項 1 又は 2 に記載の評価装置。

[請求項4] 前記特徴相似度算出部は、前記複数の次元圧縮手法のそれぞれに対して算出された前記相似度に基づいて、最適な次元圧縮手法を決定し、

前記出力部は、前記決定された最適な次元圧縮手法を出力する、請求項 1 乃至 3 のうちいずれか 1 項に記載の評価装置。

[請求項5] 次元圧縮前のデータセットを受け取る入力受付部と、
前記複数の次元圧縮手法を用いて、前記次元圧縮前のデータセットの次元を圧縮した次元圧縮後のデータセットを生成する次元圧縮部と、

を更に有する、請求項 1 乃至 4 のうちいずれか 1 項に記載の評価装置。

[請求項6] 複数の次元圧縮手法を評価する評価装置が実行する評価方法であって、

複数の特徴抽出アルゴリズムを用いて、前記複数の次元圧縮手法のそれぞれに対して、次元圧縮前のデータセット及び次元圧縮後のデータセットから、当該次元圧縮前のデータセットの第 1 の特徴と、当該次元圧縮後のデータセットの第 2 の特徴とを抽出するステップと、

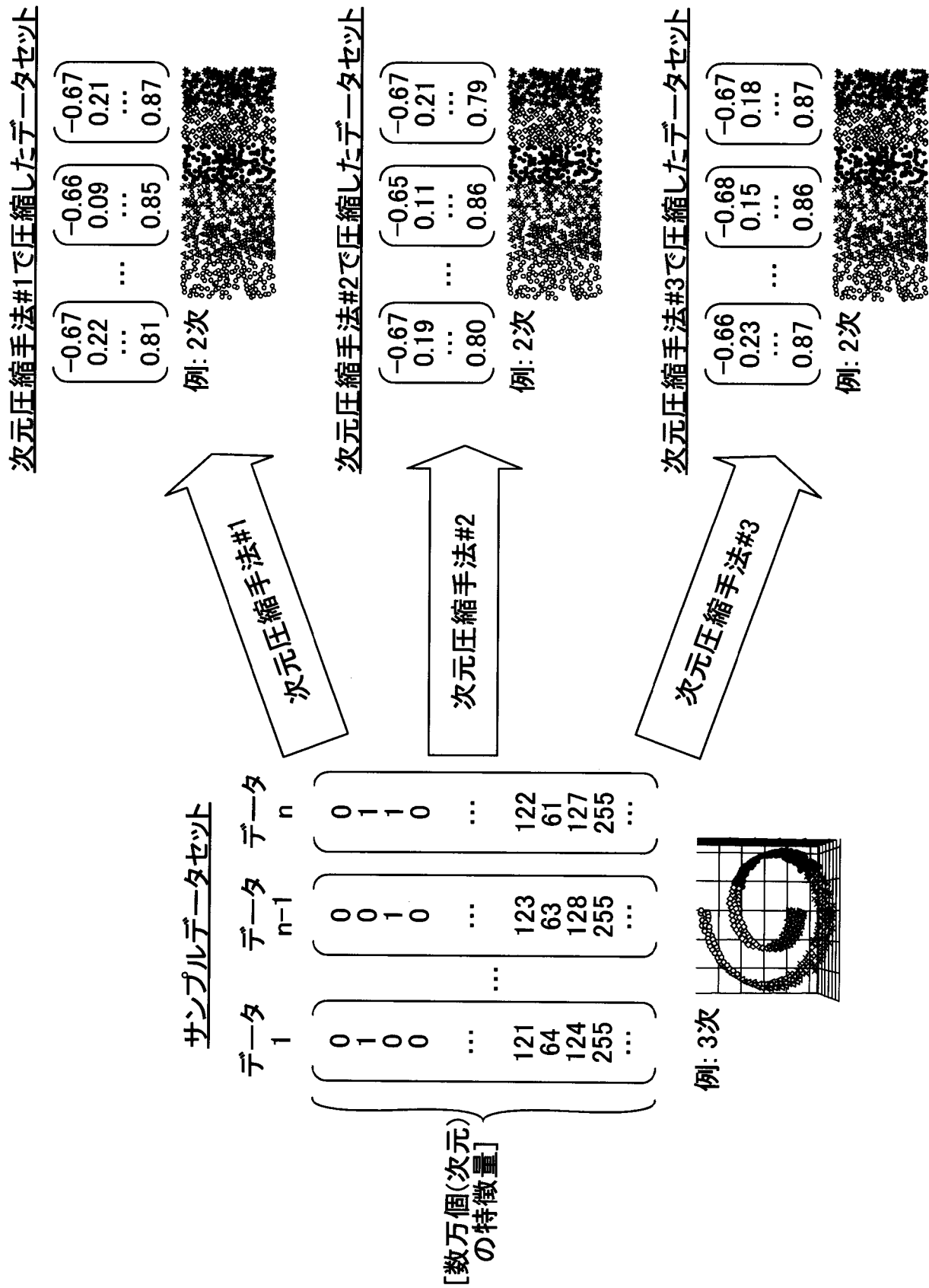
前記複数の特徴抽出アルゴリズムにそれぞれ対応する複数の特徴相似度計算アルゴリズムを用いて、前記第 1 の特徴と前記第 2 の特徴との間の相似度を算出するステップと、

前記複数の次元圧縮手法のそれぞれに対して算出された前記相似度を出力するステップと、

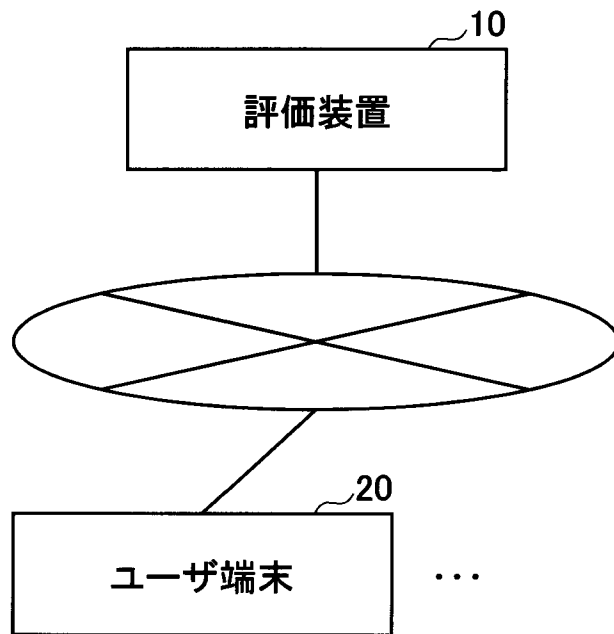
を有する評価方法。

[請求項7] 請求項 1 乃至 5 のうちいずれか 1 項に記載の評価装置の各部としてコンピュータを機能させるためのプログラム。

[図1]

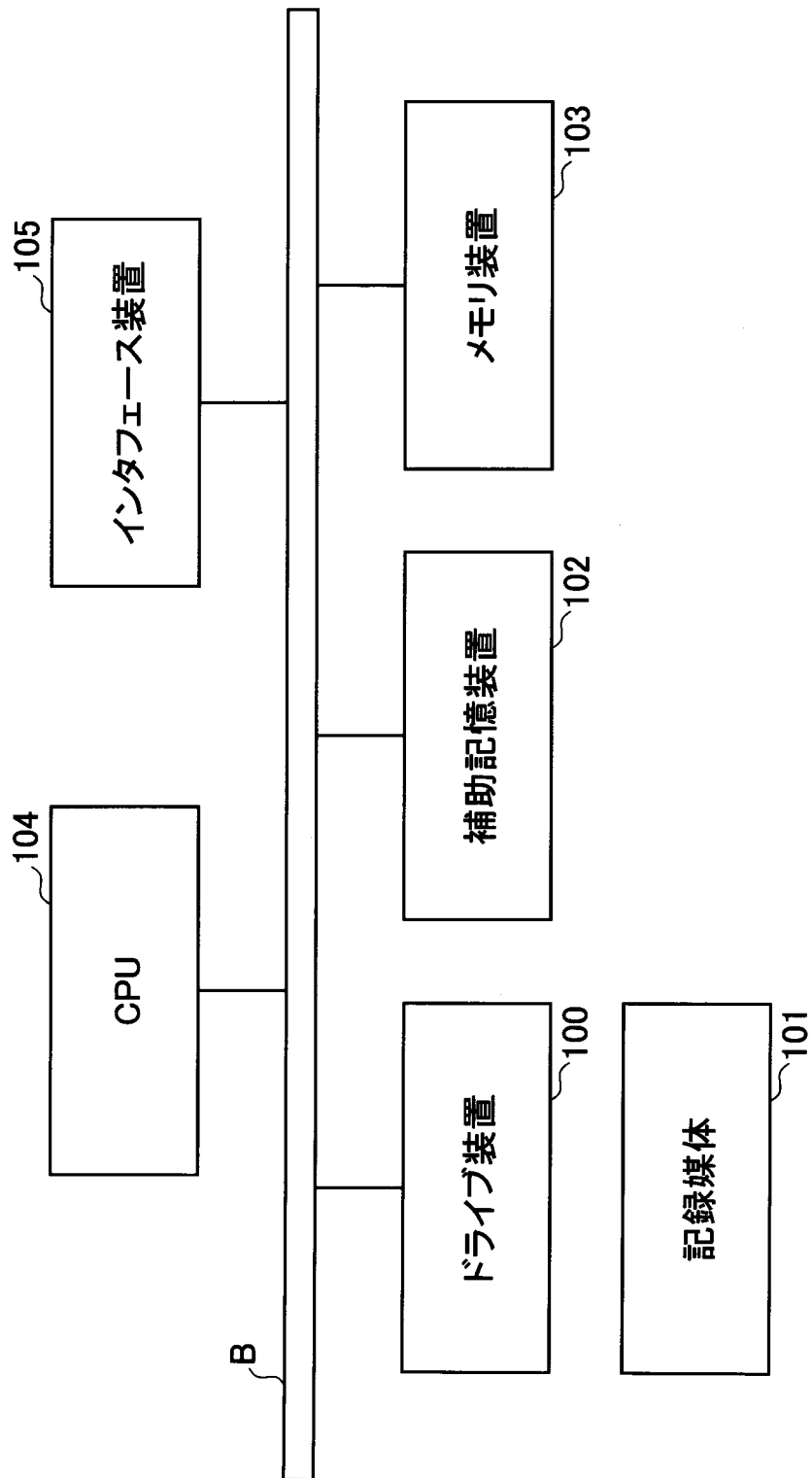


[図2]

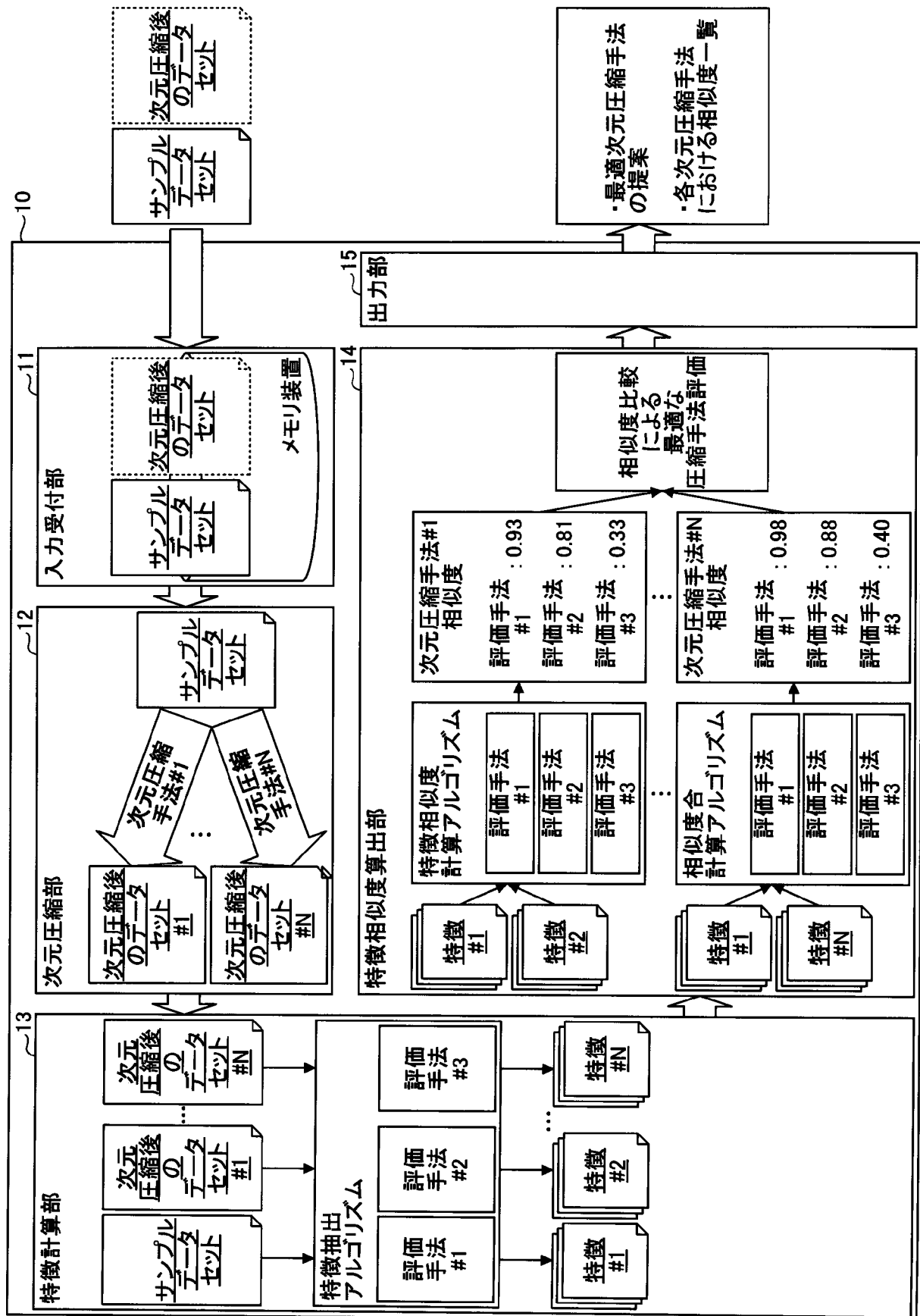


[図3]

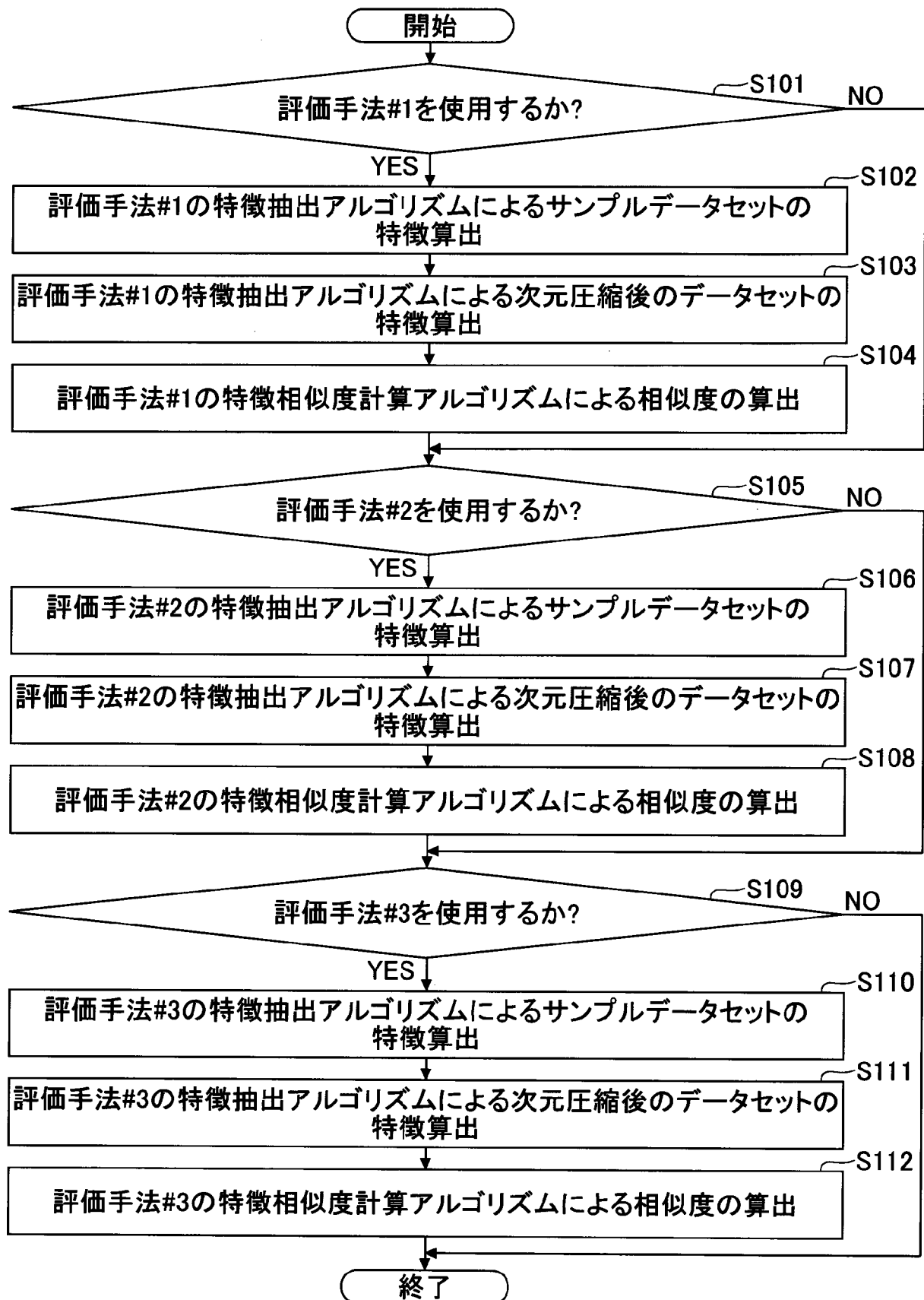
10



[図4]



[図5]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2020/002601

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/15 (2006.01) i; G06F 17/16 (2006.01) i
 FI: G06F17/16Q; G06F17/15

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)

G06F17/15; G06F17/16

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

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2017-97718 A (RICOH CO., LTD.) 01.06.2017 (2017-06-01) entire text, all drawings	1-7
A	JP 2006-227835 A (NIPPON TELEGRAPH AND TELEPHONE CORP.) 31.08.2006 (2006-08-31) entire text, all drawings	1-7
A	佐藤重雄, ほか 5 名, “時系列データ次元圧縮方式の評価および業務適用性の考察”, 情報処理学会第 66 回全国大会講演論文集(3), 09 March 2004, pp. 3-37 to 3-38, entire text, all drawings, non-official translation (SATO, Shigeo et al., “Evaluation of dimension compression method for time-series data and consideration of business applicability”, Lecture proceedings (3) of the 66th National Convention of IPSJ)	1-7



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
 16 March 2020 (16.03.2020)

Date of mailing of the international search report
 24 March 2020 (24.03.2020)

Name and mailing address of the ISA/
 Japan Patent Office
 3-4-3, Kasumigaseki, Chiyoda-ku,
 Tokyo 100-8915, Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2020/002601

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, A	仲川宜秀, ほか 2 名, "ネットワークデータへの次元圧縮手法の適用に関する一考察", 電子情報通信学会 2019 年総合大会講演論文集 通信 2, 05 March 2019, page 39, ISSN:1349-1369, entire text, all drawings, non-official translation (NAKAGAWA, Yoshihide et al., "A study on the application of dimensional compression method to network data", Lecture Proceedings of the 2019 general conference of IEICE: Communication 2)	1-7

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/JP2020/002601

Patent Documents
referred in the
ReportPublication
Date

Patent Family

Publication
Date

JP 2017-97718 A

01 Jun. 2017

(Family: none)

JP 2006-227835 A

31 Aug. 2006

(Family: none)

A. 発明の属する分野の分類（国際特許分類（IPC）） G06F 17/15(2006.01)i; G06F 17/16(2006.01)i FI: G06F17/16 Q; G06F17/15														
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）） G06F17/15; G06F17/16														
最小限資料以外の資料で調査を行った分野に含まれるもの														
<table border="0"> <tr> <td>日本国実用新案公報</td> <td>1922 - 1996年</td> </tr> <tr> <td>日本国公開実用新案公報</td> <td>1971 - 2020年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996 - 2020年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994 - 2020年</td> </tr> </table>			日本国実用新案公報	1922 - 1996年	日本国公開実用新案公報	1971 - 2020年	日本国実用新案登録公報	1996 - 2020年	日本国登録実用新案公報	1994 - 2020年				
日本国実用新案公報	1922 - 1996年													
日本国公開実用新案公報	1971 - 2020年													
日本国実用新案登録公報	1996 - 2020年													
日本国登録実用新案公報	1994 - 2020年													
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	JP 2017-97718 A（株式会社リコー）01.06.2017（2017-06-01） 全文，全図	1-7												
A	JP 2006-227835 A（日本電信電話株式会社）31.08.2006（2006-08-31） 全文，全図	1-7												
A	佐藤 重雄，ほか5名，“時系列データ次元圧縮方式の評価および業務適用性の考察”，情報処理学会第66回全国大会講演論文集（3），2004.03.09，第3-37頁－第3-38頁 全文，全図	1-7												
P, A	仲川 宜秀，ほか2名，“ネットワークデータへの次元圧縮手法の適用に関する一考察”，電子情報通信学会2019年総合大会講演論文集 通信2，2019.03.05，第39頁，ISSN:1349-1369 全文，全図	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。														
<table border="0"> <tr> <td>* 引用文献のカテゴリー</td> <td>“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの</td> </tr> <tr> <td>“A” 特に関連のある文献ではなく、一般的技術水準を示すもの</td> <td>“X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの</td> </tr> <tr> <td>“E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの</td> <td>“Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの</td> </tr> <tr> <td>“L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）</td> <td>“&” 同一パテントファミリー文献</td> </tr> <tr> <td>“O” 口頭による開示、使用、展示等に言及する文献</td> <td></td> </tr> <tr> <td>“P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献</td> <td></td> </tr> </table>			* 引用文献のカテゴリー	“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの	“A” 特に関連のある文献ではなく、一般的技術水準を示すもの	“X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの	“E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの	“Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの	“L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）	“&” 同一パテントファミリー文献	“O” 口頭による開示、使用、展示等に言及する文献		“P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献	
* 引用文献のカテゴリー	“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの													
“A” 特に関連のある文献ではなく、一般的技術水準を示すもの	“X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの													
“E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの	“Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの													
“L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）	“&” 同一パテントファミリー文献													
“O” 口頭による開示、使用、展示等に言及する文献														
“P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献														
国際調査を完了した日 16.03.2020	国際調査報告の発送日 24.03.2020													
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号	権限のある職員（特許庁審査官） 加藤 優一 5B 6296 電話番号 03-3581-1101 内線 3545													

国際調査報告
パテントファミリーに関する情報

国際出願番号

PCT/JP2020/002601

引用文献			公表日	パテントファミリー文献	公表日
JP	2017-97718	A	01.06.2017	(ファミリーなし)	
JP	2006-227835	A	31.08.2006	(ファミリーなし)	