

(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
2018年7月26日(26.07.2018)



(10) 国際公開番号

WO 2018/135281 A1

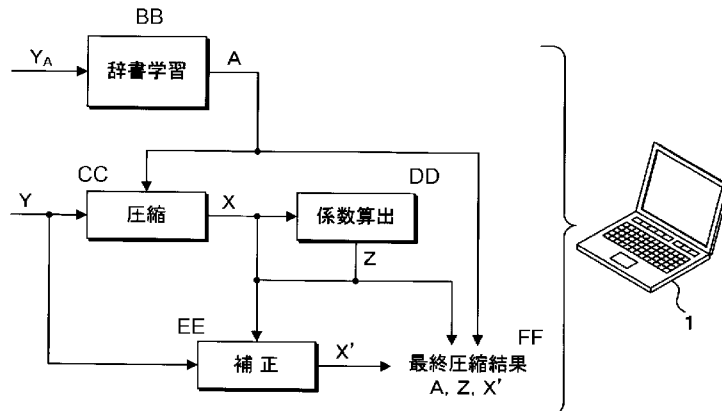
- (51) 国際特許分類:
H03M 7/30 (2006.01) *G06N 99/00* (2010.01)
G06F 17/16 (2006.01)
- (21) 国際出願番号: PCT/JP2017/047073
- (22) 国際出願日: 2017年12月27日(27.12.2017)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2017-005754 2017年1月17日(17.01.2017) JP
- (71) 出願人: 富士通株式会社 (FUJITSU LIMITED)
[JP/JP]; 〒2118588 神奈川県川崎市中原区上小田中4丁目1番1号 Kanagawa (JP).
- (72) 発明者: 此島 真喜子 (KONOSHIMA, Makiko);
〒2118588 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内 Kanagawa (JP).
- (74) 代理人: 向山 直樹 (MUKOUYAMA Naoki);
〒2118588 神奈川県川崎市中原区上小田中4丁目1番1号 Kanagawa (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,

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

(54) 発明の名称: データ圧縮装置、データ圧縮プログラム、データ圧縮方法及びプログラム

[図1]

AA
処理概要を示す説明図



- AA Explanatory diagram of process outline
BB Dictionary learning
CC Compression
DD Coefficient calculation
EE Correction
FF Final compression result

(57) Abstract: [Problem] To provide a data compression device and the like in which it is possible to efficiently compress data. [Solution] Provided are: a dictionary generating unit for generating, on the basis of inputted learning matrix data, dictionary matrix data used in compression; a compression unit for generating, on the basis of the inputted matrix data to be compressed and the generated dictionary matrix data, compression matrix data that is the result of compressing the matrix data to be compressed and position information of each matrix element of the compression matrix data in which the matrix

NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT,
QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA,
UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, 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：

— 国際調査報告（条約第21条(3)）

element is not zero; a coefficient calculation unit for comparing, for each column in the generated compression matrix data, a threshold and an indicator calculated for a specific value data row that is a data row linked with a specific value specified from the matrix element that is not zero, the coefficient calculation unit specifying a specific constant as the respective coefficients when the indicator is larger than the threshold and calculating the inverse of the respective specific values as the respective coefficients when the indicator is equal to or less than the threshold; and a element value operation unit.

(57) 要約：【課題】データを効率よく圧縮することが可能なデータ圧縮装置等を提供する。【解決手段】入力された学習用行列データに基づき、圧縮に用いる辞書行列データを生成する辞書生成部と、入力された圧縮対象行列データと生成された前記辞書行列データとに基づき、前記圧縮対象行列データを圧縮した結果である圧縮行列データと、前記圧縮行列データの各行列要素のうち行列要素がゼロではない位置情報とを生成する圧縮部と、生成された前記圧縮行列データの各行において、ゼロ値ではない要素値から特定される特定値を連結したデータ列である特定値データ列について算出された指標値と閾値とを比較し、前記指標値が前記閾値よりも大きい場合、所定の定数を各係数として特定するとともに、前記指標値が前記閾値以下である場合、各特定値の逆数を前記各係数として算出する係数算出部と、要素値操作部とを備える。

明 細 書

発明の名称：

データ圧縮装置、データ圧縮プログラム、データ圧縮方法及びプログラム
技術分野

[0001] 本発明は、データ圧縮装置、データ圧縮プログラム、データ圧縮方法及びプログラムに関する。

背景技術

[0002] 従来、スパース表現クラス分類に要する計算量を削減すると共に分類精度を低下させずに高速にクラス分類を行う装置が知られている（例えば特許文献1）。

先行技術文献

特許文献

[0003] 特許文献1：特開2012-173795号公報

発明の概要

発明が解決しようとする課題

[0004] しかしながら、従来の技術では数多くのデータを効率よく圧縮することができないという問題がある。

[0005] 一つの側面では、データを効率よく圧縮することが可能なデータ圧縮装置等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

[0006] 一つの案では、入力された学習用行列データに基づき、圧縮に用いる辞書行列データを生成する辞書生成部と、入力された圧縮対象行列データと生成された前記辞書行列データとに基づき、前記圧縮対象行列データを圧縮した結果である圧縮行列データと、前記圧縮行列データの各行列要素のうち行列要素がゼロではない位置情報とを生成する圧縮部と、生成された前記圧縮行列データの各行において、ゼロ値ではない要素値から特定される特定値を連

結したデータ列である特定値データ列について算出された指標値と閾値とを比較し、前記指標値が前記閾値よりも大きい場合、所定の定数を各係数として特定するとともに、前記指標値が前記閾値以下である場合、各特定値の逆数を前記各係数として算出する係数算出部と、前記圧縮行列データの各要素について、前記各要素と算出された前記各係数との積について、対応する各要素における小数部の有効数字の桁数に基づいて丸めた結果である操作後の行列データを圧縮結果として出力する要素値操作部と、を備える。

発明の効果

[0007] 一つの側面では、データを効率よく圧縮することが可能となる。

図面の簡単な説明

[0008] [図1]処理概要を示す説明図である。

[図2]コンピュータのハードウェア群を示すブロック図である。

[図3]辞書学習処理の手順を示すフローチャートである。

[図4]辞書学習処理の手順を示すフローチャートである。

[図5]残差 r_t 及び変数 A_t の数値例を示す説明図である。

[図6]辞書行列データの数値例を示す説明図である。

[図7]圧縮処理の手順を示すフローチャートである。

[図8]係数算出処理の手順を示すフローチャートである。

[図9]圧縮行列データのデータ例を示す説明図である。

[図10]係数のデータ例を示す説明図である。

[図11]補正処理及び端数処理の手順を示すフローチャートである。

[図12]数値例を示す説明図である。

[図13]数値例を示す説明図である。

[図14]補正処理及び端数処理の手順を示すフローチャートである。

[図15]数値例を示す説明図である。

[図16]補正処理及び端数処理の手順を示すフローチャートである。

[図17]数値例を示す説明図である。

[図18] $k = 2$ の場合の誤差とファイルサイズとの関係を示すグラフである。

[図19] $k = 5$ の場合の誤差とファイルサイズとの関係を示すグラフである。

[図20] 上述した形態のコンピュータの動作を示す機能ブロック図である。

[図21] 実施の形態4に係るコンピュータのハードウェア群を示すブロック図である。

発明を実施するための形態

[0009] 実施の形態1

以下実施の形態を、図面を参照して説明する。図1は処理概要を示す説明図である。係数の算出等の各種処理を実行する情報処理装置1は、サーバコンピュータまたはパーソナルコンピュータ等である。以下では情報処理装置1をコンピュータ1と読み替えて説明する。コンピュータ1は、学習用行列データ Y_A に基づき学習を行い、圧縮に用いる辞書Aを生成する。次いでコンピュータ1は圧縮対象の行列データYと辞書Aとを用いて、圧縮行列データXを算出する。

[0010] コンピュータ1は圧縮行列データXの各要素値に基づき、圧縮行列データXを補正するための係数Zを算出する。コンピュータ1は圧縮行列データX及び係数Zに基づき補正を行い、その後端数処理を行った上で最終的な補正圧縮行列データX'を求める。コンピュータ1は辞書A、係数Z及び補正圧縮行列データX'を最終結果として出力する。以下詳細を説明する。

[0011] 図2はコンピュータ1のハードウェア群を示すブロック図である。コンピュータ1は制御部としてのCPU (Central Processing Unit) 11、RAM (Random Access Memory) 12、入力部13、表示部14、記憶部15、時計部18、及び、通信部16等を含む。CPU 11は、バス17を介してハードウェア各部と接続されている。CPU 11は記憶部15に記憶された制御プログラム15Pに従いハードウェア各部を制御する。なお、CPU 11は複数のプロセッサコア

を搭載したマルチコアプロセッサであっても良い。RAM 12は例えばSRAM (Static RAM)、DRAM (Dynamic RAM)、フラッシュメモリ等である。RAM 12は、記憶部としても機能し、CPU 11による各種プログラムの実行時に発生する種々のデータを一時的に記憶する。

[0012] 入力部13はマウス、キーボード、タッチパネル、ボタン等の入力デバイスであり、受け付けた操作情報をCPU 11へ出力する。表示部14は液晶ディスプレイまたは有機EL (Electro Luminescence) ディスプレイ等であり、CPU 11の指示に従い各種情報を表示する。通信部16は通信モジュールであり、他のコンピュータ (図示せず) 等と間で情報の送受信を行う。時計部18は日時情報をCPU 11へ出力する。記憶部15は大容量メモリまたはハードディスクであり、制御プログラム15P等を記憶している。

[0013] 図3及び図4は辞書学習処理の手順を示すフローチャートである。コンピュータ1のCPU 11は、辞書行列データAの初期値に標準正規乱数を設定する (ステップS31)。

なお、入力として、データの圧縮単位 (要素数) を n 、学習用のデータの個数 (圧縮単位のベクトルの個数) を N 、学習用行列データ Y_A (以下、場合により Y で代表する) を n 行 N

列の行列とする。また出力を n 行 m 列の辞書行列データ A とする。 m 行 N 列の圧縮行列デー

タ X 、残差 r_t (残差ベクトル r_t)、変数 A_t 、ループ変数 i 等は変数とする。

[0014] CPU 11は、変数 i に1を入力する (ステップS32)。CPU 11は、サンプル y

に学習用行列データ Y の i 列目のデータを入力する (ステップS33)。CPU 11は、残差 r_t の初期値としてサンプル y を入力し、変数 A_t の初期値としてNULLを入力する (ステップS34)。

[0015] 図5は、残差 r_t 及び変数 A_t の数値例を示す説明図である。図6は、辞書行列データ A の

数値例を示す説明図である。図5に示すように残差 r_t には2.6、2.6、2.6、2.

4・・・と学習用行列データYのi列目のデータが入力されている。なお、図5及び図6の例では、ベクトルの長さを100としているが、紙面の都合上図5及び図6では20要素分までを記載している。

[0016] CPU 11は、図6に示す辞書行列データAの転置行列 A^T の一系列の中で、残差 r_t の転置行列 r_t^T との内積が最大となる列 j_i （位置情報）を抽出し、変数列 A_j とする（ステップS35）。すなわち、内積が最大となるものが最も似た方向を向いているベクトルであるとして抽出する。CPU 11は、変数 A_t に変数列 A_j を水平方向に連結する（ステップS36）。なお、初回は A_t の初期値がNULLであるため、連結対象の変数列 A_j がそのまま入力される。図5及び図6に示すとおり、変数 A_t の1列目は、辞書行列データAの13列目とベクトルの向きが類似し、変数 A_t の2列目は、辞書行列データAの292列目とベクトルの向きが類似している。また変数 A_t の3列目は、辞書行列データAの169列目とベクトルの向きが類似している。

[0017] CPU 11は、変数 A_t と残差 r_t との最小二乗解 X_{ti} を算出する（ステップS37）。具体的には、 $r_t = A_t * X_{ti}$ となる X_{ti} のうち、 $r_t - A_t * X_{ti}$ の二乗誤差を最小にする X_{ti} を算出する。図5に示すように X_{ti} は7.509095、6.965821、5.834313・・・と算出される。CPU 11は、変数 A_t に最小二乗解 X_{ti} を乗じて乗算値 y_t を算出する（ステップS38）。CPU 11は、残差 r_t から乗算値 y_t を減算した値を残差 r_t に置き換える（ステップS39）。図5には、 r_t が1.618316731、3.323110212・・・と置き換えられている。

- [0018] CPU 11は、記憶部15から閾値E1及び閾回数（例えば30回）を読み出す（ステップS41）。CPU 11は、残差 r_t の二乗誤差が閾値E1以下、または、ステップS35～S41の処理回数が閾回数以上か否かを判断する（ステップS42）。CPU 11は、残差 r_t の二乗誤差が閾値E1以下、または、閾回数以上のいずれでもないと判断した場合（ステップS42でNO）、処理をステップS35へ移行させる。CPU 11は、ステップS35への移行の際、ステップS35～S41の処理を実行した回数をインクリメントとし、RAM12に記憶する。CPU 11は、ステップS42において、閾回数とRAM12に記憶した回数とを比較する。
- [0019] CPU 11は、残差 r_t の二乗誤差が閾値E1以下、または、閾回数以上であると判断した場合（ステップS42でYES）、処理をステップS43へ移行させる。CPU 11は、変数 i に1を加える（ステップS43）。CPU 11は、変数 i が N より大きいと判断する（ステップS44）。CPU 11は、変数 i が N より大きくないと判断した場合（ステップS44でNO）、処理をステップS33に戻す。一方、CPU 11は、変数 i が N より大きいと判断した場合（ステップS44でYES）、処理をステップS45へ移行させる。
- [0020] CPU 11は、最小二乗解 X_{ti} 及び列 j_i から圧縮行列データ X を構成し、学習用行列データ Y -辞書行列データ A *圧縮行列データ X の最小二乗解 A を算出する（ステップS45）。具体的には、 X_{ti} は非0の成分からのみ構成されるため、CPU 11は、ステップS35で求めた列番号 j_i （図5の例では、13、292、169・・・）を使用し、非0の位置（13、292、169・・・）以外の要素値を0とし、圧縮行列データ X を構成する。次いで、CPU 11は、 $Y=A*X$ の最小二乗解を求めて新しい A とする。
- [0021] CPU 11は、記憶部15から第2閾値E2及び閾回数を読み出す（ステップS46）。CPU 11は、学習用行列データ Y -辞書行列データ A *圧縮行列

データ X の二乗誤差が第2

閾値以下、または、閾回数以上か否かを判断する（ステップS47）。具体的には、CPU11は、 $Y=AX$ の二乗誤差が第2閾値 $E2$ 以下か否かを判断する。またCPU11は、ステップS47の処理を行った回数が閾回数以上か否かを判断する。CPU11は、第2閾値 $E2$ 以下、または、閾回数以上でないと判断した場合（ステップS47でNO）、処理をステップS33に戻す。CPU11は、ステップS47の処理後、ステップS47を処理した回数をインクリメントし、RAM12に記憶する。CPU11は、ステップS47の処理において、閾回数と、RAM12に記憶した回数とを比較する。

[0022] CPU11は、第2閾値以下、または、閾回数以上と判断した場合（ステップS47でYES）、処理をステップS48へ移行させる。CPU11は、学習済みの辞書行列データ A を記憶部15に記憶する（ステップS48）。

[0023] 図7は圧縮処理の手順を示すフローチャートである。入力としての圧縮対象のデータ y

は圧縮単位（要素数） n 個の列ベクトルである。その他、圧縮行列データ X の非ゼロの個

数 K 、辞書行列データ A （ n 行 m 列の行列）を入力とする。また出力は圧縮結果 X_{tk} 及び非ゼロ

位置 jk （位置情報）とする。なお、残差 r_t 及び変数 A_t （辞書）は内部変数とする。CPU11は、残差 r_t の初期値としてデータ y を入力し、変数 A_t の初期値としてNULLを入力する（ステップS71）。CPU11は、変数 kk に1を入力する（ステップS72）。

[0024] CPU11は、辞書行列データ A の転置行列 A^T の一行の中で、残差 r_t の転置行列 r_t^T との内積が最大となる列 jk （位置情報）を抽出し、変数列 A_{jk} とする（ステップS73）。具

体的には、CPU11は、 A^T の列の中で、 r_t^T との内積が最大となる列 jk を抽出し、 A_{jk} とする。CPU11は、変数 A_t に変数列 A_{jk} を水平方向に連結する（ステップS74）。C

PU11は、変数 A_t と残差 r_t との最小二乗解を X_{tk} とする（ステップS75）。

[0025] CPU11は、変数 A_t に最小二乗解 X_{tk} を乗じて乗算値 y_t を算出する（ステップS76）。

CPU11は、残差 r_t から乗算値 y_t を減算した値を残差 r_t に置き換える（ステップS77）。CPU11は、変数 kk が非ゼロの個数 K より大きいかなんかを判断する（ステップS78）。CPU11は、大きくないと判断した場合（ステップS78でNO）、処理をステップS79へ移行させる。CPU11は、変数 kk の値をインクリメントする（ステップS79）。CPU11は、その後処理をステップS73へ移行させる。CPU11は、変数 kk が個数 K より大きいと判断した場合（ステップS78でYES）、処理を終了する。

CPU11は、ステップS76で求めた X_{tk} 及びステップS73で求めた非ゼロの位置 j_k ($k=1\sim K$)（位置情報）を出力する。

[0026] 図8は係数算出処理の手順を示すフローチャートである。CPU11は、圧縮行列データ X を取得する（ステップS81）。図9は圧縮行列データ X のデータ例を示す説明図である。圧縮行列データ X は m 行 N 列の行列であり、行が圧縮結果の要素の位置、列がサンプルの並びに対応する。行データの多くが0であり、非ゼロの要素が複数存在している。CPU11は、変数 i に初期値1を入力する（ステップS82）。

[0027] CPU11は、圧縮行列データ X の i 行の非ゼロの要素の特定値を取得する。特定値は i 行の非ゼロの要素に基づき特定される値であり、例えば複数の非ゼロの要素の、平均値、中央値、最大値、最小値、または、二乗平均等である。なお、実施形態では非ゼロの要素に基づく特定値を利用する例を示すが、ゼロの要素を一部に含む形態であっても良い。以下の実施形態では特定値を非ゼロの要素の絶対値の平均値であるものとして説明する。CPU11は、圧縮行列データ X の i 行の非ゼロの要

素の絶対値の平均値 M_i を算出する（ステップS83）。CPU11は、変数 i が m （行数 m ）より大きいかなんかを判断する（ステップS84）。CPU11は、変数 i が m より大きくないと判断した場合（ステップS84でNO）、処理をステップS85へ移行させる。

[0028] CPU11は、 i に1を加算する（ステップS85）。CPU11は、その後処理をステップS83へ戻す。

これにより平均値を連結したデータ列である特定値データ列（ M_i （ $i = 1 \sim m$ ））が算出される。CPU11は、 i が m より大きいと判断した場合（ステップS84でYES）、処理をステップS86へ移行させる。CPU11は、算出した複数の平均値に基づき、平均値のばらつきの度合いを示す指標値を算出する。実施形態では指標値を標準分散であるものとして説明するが、これに限るものではない。例えば不偏分散を用いても良い。

[0029] CPU11は、各列の平均値の標本分散を算出する（ステップS86）。CPU11は、平均値の最大値及び最小値を抽出する（ステップS87）。CPU11は、最大値から最小値を減算して減算値を求め、減算値を所定値（例えば6）で除して閾値（TX）を算出する（ステップS88）。なお所定値は、ユーザが適宜入力部13から値を設定すれば良い。CPU11は、標本分散が閾値より大きいかなんかを判断する（ステップS89）。CPU11は、標本分散が閾値より大きいと判断した場合（ステップS89でYES）、処理をステップS810へ移行させる。CPU11は、ばらつきが大きい場合、係数 Z_i を1として出力する。（ステップS810）。

[0030] CPU11は、標本分散が閾値より大きくない場合（ステップS89でNO）、処理をステップS811へ移行させる。CPU11は、係数 Z_i を平均値 M_i の逆数として出力する（ $Z_i = 1/M_i$ （ $i = 1 \sim m$ ））（ステップS811）。

[0031] 図10は係数のデータ例を示す説明図である。Y1及びY2は、圧縮前の圧縮対象行列データを示す。X1及びX2は、辞書行列データAに基づき圧縮処理され

た後の圧縮行列データXである。ここで圧縮行列データXの非ゼロの個数 $k=8$ である。係数Zは、説明を容易にするためにX1及びX2の絶対値の平均値の逆数としている。また係数Zの標本分散は4.1であり、閾値より大きくないものとしている。続いてCPU11は、係数Zを用いた圧縮行列データXの補正処理及び端数処理（丸め処理）を行う。

[0032] 図11は補正処理及び端数処理の手順を示すフローチャートである。図12及び図13は数値例を示す説明図である。CPU11は、記憶部15に記憶した第1係数及び第2係数を読み出す（ステップS111）。第1係数は、圧縮対象行列データYと圧縮行列データXの容量比率を示す。第1係数は例えば、圧縮行列Xの小数部の有効桁数（有効数字の桁数）を圧縮対象行列データYの小数点以下の有効桁数で除した値とすれば良い。本実施形態では小数部の有効桁数としたがこれに限るものではない。第1係数を圧縮行列Xの有効桁数としてもよい。さらに、第1係数を、圧縮行列データXの小数点を含んだ文字数を、圧縮対象行列データYの小数点を含んだ文字数で除した値としても良い。例えば、2.6の場合、小数点を含んで文字数は3となる。なお、実施形態では一例として第1係数を2であるものとして説明する。

[0033] 第2係数は、本実施形態では圧縮対象行列データYの有効桁数であり、一例として第2係数は2であるものとして説明する。CPU11は、第1係数に第2係数を乗じた乗算値から1を減じて桁数を算出する（ステップS112）。本実施形態では第1係数2に第2係数2を乗じて乗算値4を求め1を減じて、桁数は3となる。CPU11は、圧縮行列データXに係数Zを乗じて補正後の圧縮行列データXtempを算出する（ステップS113）。

図 1 2 には係数 Z を乗じた後の補正後の圧縮行列データ X_{temp} が示されている。

[0034] CPU 1 1 は、補正後の圧縮行列データ X_{temp} を、ステップ S 1 1 2 で求めた桁数で端数処理する（ステップ S 1 1 4）。図 1 3 に示すように小数点第 3 位の桁の値について端数処理されている。端数処理は小数点第 3 位以下の第 4 位の数値に基づく四捨五入、第 4 位の数値の切り捨て、または第 3 位の切り上げを行えば良い。本実施形態では四捨五入する例を示す。CPU 1 1 は、端数処理後の圧縮行列データ X' を記憶部 1 5 に記憶し、表示部 1 4 に出力または通信部 1 6 を介して外部に出力する（ステップ S 1 1 5）。これにより効率よくデータを圧縮することが可能となる。また辞書の学習により精度を向上させることが可能となる。また桁数の算出により、要求に応じたサイズに精度よく圧縮することが可能となる。さらに圧縮行列データ X のばらつきの度合いにより係数を適宜設定することで、効率よく圧縮することが可能となる。

[0035] 実施の形態 2

実施の形態 2 は、符号を考慮する形態に関する。図 1 4 は補正処理及び端数処理の手順を示すフローチャートである。図 1 5 は数値例を示す説明図である。CPU 1 1 は、記憶部 1 5 に記憶した第 1 係数及び第 2 係数を読み出す（ステップ S 1 4 1）。なお、第 1 係数は 2、第 2 係数は 2 であるものとして説明する。CPU 1 1 は、圧縮対象行列データ Y の符号の有無に応じて 1 または 0 を第 2 係数に加算する（ステップ S 1 4 2）。図 1 2 に示したように圧縮対象行列データ Y はマイナスの符号が存在しないため、加算されず第 2 係数は 2 のままである。

[0036] CPU 1 1 は、第 1 係数に第 2 係数を乗じた乗算値から 1 を減じて桁数を算出する（ステップ S 1 4 3）。本実施形態では第 1 係数 2 に第 2 係数 2 を乗じて乗算値 4 を求め 1 を減じて、桁数は 3 となる。CPU 1 1 は、圧縮行

列データXの符号の有無に応じて桁数に

1または0を減算する（ステップS 1 4 4）。圧縮行列データXは図1 2に示すように符

号を有するため、桁数3から1を減じて桁数2を算出する。CPU 1 1は、圧縮行列データXに係数Zを乗じて補正後の圧縮行列データXtempを算出する（ステップS 1 4 5）。

[0037] CPU 1 1は、補正後の圧縮行列データXtempを、ステップS 1 4 4の処理を経て得た

桁数で端数処理する（ステップS 1 4 6）。図1 5に示すように小数点第2位の桁の値について端数処理されている。本実施形態では四捨五入する例を示している。CPU 1 1は、端数処理後の圧縮行列データX'（圧縮結果）を記憶部1 5に記憶し、表示部1 4に出

力または通信部1 6を介して外部に出力する（ステップS 1 4 7）。これにより圧縮行列データXの符号の有無をも考慮した精度の良いデータ圧縮が可能となる。

[0038] 本実施の形態2は以上の如きであり、その他は実施の形態1と同様であるので、対応する部分には同一の参照番号を付してその詳細な説明を省略する。

[0039] 実施の形態3

実施の形態3は小数点の有無をも考慮する形態に関する。図1 6は補正処理及び端数処理の手順を示すフローチャートである。図1 7は数値例を示す説明図である。CPU 1 1は、記憶部1 5に記憶した第1係数及び第2係数を読み出す（ステップS 1 6 1）。なお、第1係数は2、第2係数は2であるものとして説明する。CPU 1 1は、圧縮対象行列データYの符号の有無に応じて1または0を第2係数に加算する（ステップS 1 6 2）。

図1 2に示したように圧縮対象行列データYはマイナスの符号が存在しないため、加算さ

れず第2係数は2のままである。

- [0040] CPU 11は、圧縮対象行列データYの小数点の有無に応じて1または0を第2係数に加算する（ステップS 1 6 3）。図 1 2に示したように、圧縮対象行列データYは小数点が存在するため1加算され第2係数は3となる。CPU 11は、第1係数に第2係数を乗じた乗算値から1を減じて桁数を算出する（ステップS 1 6 4）。本実施形態では第1係数2に第2係数3を乗じて乗算値6を求め1を減じて、桁数は5となる。CPU 11は、圧縮行列データXの符号の有無に応じて桁数に1または0を減算する（ステップS 1 6 5）。圧縮行列データXは図 1 2に示すように符号を有するため、桁数5から1を減じて桁数4を算出する。
- [0041] CPU 11は、圧縮行列データXの小数点の有無に応じて桁数に1または0を減算する（ステップS 1 6 6）。図 1 2に示すように圧縮行列データXは小数点を有するため、桁数4から1を減じて桁数3を算出する。CPU 11は、圧縮行列データXに係数Zを乗じて補正後の圧縮行列データXtempを算出する（ステップS 1 6 7）。CPU 11は、補正後の圧縮行列データXtempを、ステップS 1 6 6の処理を経て得た桁数で端数処理する（ステップS 1 6 8）。図 1 7に示すように小数点第3位の桁の値について端数処理されている。本実施形態では四捨五入する例を示している。CPU 11は、端数処理後の圧縮行列データX'を記憶部15に記憶し、表示部14に出力または通信部16を介して外部に出力する（ステップS 1 6 9）。これにより小数点の有無をも考慮した精度の良いデータ圧縮が可能となる。

- [0042] 図 1 8はk = 2の場合の誤差とファイルサイズとの関係を示すグラフであ

り、図 19 は $k = 5$ の場合の誤差とファイルサイズとの関係を示すグラフである。図 18 における横軸は圧縮行列データ X のファイルサイズ（バイト）であり、縦軸は圧縮誤差の最大値である

。圧縮対象行列データ Y のファイルサイズは 963,000 バイト、小数点あり、符号なし、有効桁数を 2 桁、圧縮対象行列データ Y の圧縮単位 $n=223$ である。圧縮行列データ X の非ゼロの

個数 k が 2 の場合が図 18、5 の場合が図 19 である。丸印で示す系列は比較例であり端

数処理を行わないものである。四角で示す系列は第 1 係数を 2.0 としたものであり、三角

で示す系列は第 1 係数を 1.6 としたものである。

[0043] 圧縮後のファイルサイズは概ね、圧縮前ファイルサイズ*第 1 係数* k/n となっている。

また端数処理を行うことで、テキストファイルを保管する際に誤差を拡大させることもなくデータの更なる圧縮が可能であることが理解できる。 $k = 2$ 、 $k = 5$ いずれの場合にも

、同程度の圧縮誤差であり、第 1 係数 2.0 の場合で比較例に対し約 1/2、第 1 係数 1.6 の場

合で比較例に対し約 1/3 程度の圧縮を行うことが可能となる。

[0044] 本実施の形態 3 は以上の如きであり、その他は実施の形態 1 または 2 と同様であるので、対応する部分には同一の参照番号を付してその詳細な説明を省略する。

[0045] 実施の形態 4

図 20 は上述した形態のコンピュータ 1 の動作を示す機能ブロック図である。CPU 11 が制御プログラム 15 P を実行することにより、コンピュータ 1 は以下のように動作する。辞書生成部 201 は、入力された学習用行列データに基づき、圧縮に用いる辞書行列データを生成する。圧縮部 202 は、入力された圧縮対象行列データと生成された前記辞書行列データとに基づ

き、前記圧縮対象行列データを圧縮した結果である圧縮行列データと、前記圧縮行列データの各行列要素のうち行列要素がゼロではない位置情報とを生成する。係数算出部203は、生成された前記圧縮行列データの各行において、ゼロ値ではない要素値から特定される特定値を連結したデータ列である特定値データ列について算出された指標値と閾値とを比較し、前記指標値が前記閾値よりも大きい場合、所定の定数を各係数として特定するとともに、前記指標値が前記閾値以下である場合、各特定値の逆数を前記各係数として算出する。

[0046] 要素値操作部204は、前記圧縮行列データの各要素について、前記各要素と算出された前記各係数との積について、対応する各要素における小数部の有効数字の桁数に基づいて丸めた結果である操作後の行列データを圧縮結果として出力する。抽出部205は、前記学習用行列データの列データに基づき、該列データとの内積が最大となる前記辞書行列データの列データ及び位置情報を抽出する。算出部206は、抽出した前記列データ及び前記学習用行列データの列データに基づき最小二乗解を算出する。生成部207は、算出した最小二乗解、前記位置情報及び前記学習用行列データに基づき、前記辞書行列データを生成する。

[0047] 第2抽出部208は、圧縮対象行列データの列データに基づき、該列データとの内積が最大となる前記辞書行列データの列データを抽出する。第2算出部209は、抽出した前記列データ及び前記圧縮対象行列データの列データに基づき最小二乗解を算出する。積算出部2010は、前記圧縮行列データの各要素と前記各係数との積を算出する。桁数算出部2011は、第1係数と、前記圧縮対象行列データの有効桁数及び符号の有無に基づき求まる第2係数と、前記圧縮行列データの各要素の符号の有無とに基づき算出される桁数を算出する。処理部2012は、該桁数算出部で算出した桁数に基づき、前記積算出部で算出した積の端数を処理する。

[0048] 図21は実施の形態4に係るコンピュータ1のハードウェア群を示すブロック図である。コンピュータ1を動作させるためのプログラムは、ディスク

ドライブ、メモリーカードスロット等の読み取り部 10A に CD-ROM、DVD ディスク、メモリーカード、または USB メモリ等の可搬型記録媒体 1A を読み取らせて記憶部 15 に記憶しても良い。また当該プログラムを記憶したフラッシュメモリ等の半導体メモリ 1B をコンピュータ 1 内に実装しても良い。さらに、当該プログラムは、インターネット等の通信網 N を介して接続される他のサーバコンピュータ（図示せず）からダウンロードすることも可能である。以下に、その内容を説明する。

[0049] 図 21 に示すコンピュータ 1 は、上述した各種ソフトウェア処理を実行するプログラムを、可搬型記録媒体 1A または半導体メモリ 1B から読み取り、或いは、通信網 N を介して他のサーバコンピュータ（図示せず）からダウンロードする。当該プログラムは、制御プログラム 15P としてインストールされ、RAM 12 にロードして実行される。これにより、上述したコンピュータ 1 として機能する。その他、上述した処理を実行するプログラムを記憶したメモリとプロセッサとを有するハードウェアデバイスをデータ圧縮装置として、図示しないサーバコンピュータ、パーソナルコンピュータ、または、スマートフォン等を実装するようにしても良い。

[0050] 本実施の形態 4 は以上の如きであり、その他は実施の形態 1 から 3 と同様であるので、対応する部分には同一の参照番号を付してその詳細な説明を省略する。なお、以上述べた各実施形態は適宜組み合わせることが可能である。

[0051] 以上の実施の形態 1 から 4 を含む実施形態に関し、さらに以下の付記を開示する。

[0052] （付記 1）

入力された学習用行列データに基づき、圧縮に用いる辞書行列データを生成する辞書生成部と、

入力された圧縮対象行列データと生成された前記辞書行列データとに基づき、前記圧縮対象行列データを圧縮した結果である圧縮行列データと、前記圧縮行列データの各行列要素のうち行列要素がゼロではない位置情報とを生

成する圧縮部と、

生成された前記圧縮行列データの各行において、ゼロ値ではない要素値から特定される特定値を連結したデータ列である特定値データ列について算出された指標値と閾値とを比

較し、前記指標値が前記閾値よりも大きい場合、所定の定数を各係数として特定するとともに、前記指標値が前記閾値以下である場合、各特定値の逆数を前記各係数として算出する係数算出部と、

前記圧縮行列データの各要素について、前記各要素と算出された前記各係数との積について、対応する各要素における小数部の有効数字の桁数に基づいて丸めた結果である操作後の行列データを圧縮結果として出力する要素値操作部と、

を備えるデータ圧縮装置。

(付記 2)

前記辞書生成部は、

前記学習用行列データの列データに基づき、該列データとの内積が最大となる前記辞書行列データの列データ及び位置情報を抽出する抽出部と、

抽出した前記列データ及び前記学習用行列データの列データに基づき最小二乗解を算出する算出部と、

算出した最小二乗解、前記位置情報及び前記学習用行列データに基づき、前記辞書行列データを生成する生成部と、

を含む付記 1 に記載のデータ圧縮装置。

(付記 3)

前記圧縮部は、

圧縮対象行列データの列データに基づき、該列データとの内積が最大となる前記辞書行列データの列データ及び位置情報を抽出する第 2 抽出部と、

抽出した前記列データ及び前記圧縮対象行列データの列データに基づき最小二乗解を算出する第 2 算出部と、

を含む付記 1 または 2 に記載のデータ圧縮装置。

(付記 4)

前記要素値操作部は、

前記圧縮行列データの各要素と前記各係数との積を算出する積算出部と、

第 1 係数と、前記圧縮対象行列データの有効桁数及び符号の有無に基づき求まる第 2 係数と、前記圧縮行列データの各要素の符号の有無とに基づき算出される桁数を算出する桁数算出部と、

該桁数算出部で算出した桁数に基づき、前記積算出部で算出した積の端数进行处理する処理部と、

を備える付記 1 から 3 のいずれか一つに記載のデータ圧縮装置。

(付記 5)

桁数算出部は、

第 1 係数と、前記圧縮対象行列データの有効桁数、符号の有無及び小数点の有無に基づき求まる第 2 係数と、前記圧縮行列データの各要素の符号の有無及び小数点の有無とに基づき桁数を算出する

付記 4 に記載のデータ圧縮装置。

(付記 6)

前記処理部は、

前記積算出部で算出した積に対し、前記桁数算出部で算出した桁数以降の端数进行处理する

付記 4 または 5 に記載のデータ圧縮装置。

(付記 7)

前記閾値は前記特定値の最大値及び最小値に基づき決定される

付記 1 から 6 のいずれか一つに記載のデータ圧縮装置。

(付記 8)

前記特定値は、生成された前記圧縮行列データの各行のゼロ値ではない要素値の平均値、中央値、最大値または二乗平均であり、

前記指標値は、前記特定値を連結したデータ列である特定値データ列の分散である

付記 1 から 7 のいずれか一つに記載のデータ圧縮装置。

(付記 9)

入力された学習用行列データに基づき、圧縮に用いる辞書行列データを生成し、

入力された圧縮対象行列データと生成された前記辞書行列データとに基づき、前記圧縮対象行列データを圧縮した結果である圧縮行列データと、前記圧縮行列データの各行列要素のうち行列要素がゼロではない位置情報とを生成し、

生成された前記圧縮行列データの各行において、ゼロ値ではない要素値から特定される特定値を連結したデータ列である特定値データ列について算出された指標値と閾値とを比較し、前記指標値が前記閾値よりも大きい場合、所定の定数を各係数として特定するとともに、前記指標値が前記閾値以下である場合、各特定値の逆数を前記各係数として算出し、

前記圧縮行列データの各要素について、前記各要素と算出された前記各係数との積について、対応する各要素における小数部の有効数字の桁数に基づいて丸めた結果である操作後の行列データを圧縮結果として出力する、

処理をコンピュータに実行させるデータ圧縮プログラム。

(付記 10)

入力された学習用行列データに基づき、圧縮に用いる辞書行列データを生成し、

入力された圧縮対象行列データと生成された前記辞書行列データとに基づき、前記圧縮対象行列データを圧縮した結果である圧縮行列データと、前記圧縮行列データの各行列要素のうち行列要素がゼロではない位置情報とを生成し、

生成された前記圧縮行列データの各行において、ゼロ値ではない要素値から特定される特定値を連結したデータ列である特定値データ列について算出された指標値と閾値とを比較し、前記指標値が前記閾値よりも大きい場合、所定の定数を各係数として特定するとともに、前記指標値が前記閾値以下で

ある場合、各特定値の逆数を前記各係数として算出し、

前記圧縮行列データの各要素について、前記各要素と算出された前記各係数との積について、対応する各要素における小数部の有効数字の桁数に基づいて丸めた結果である操作後の行列データを圧縮結果として出力する、
処理をコンピュータに実行させるデータ圧縮方法。

(付記 1 1)

圧縮対象行列データ及び辞書行列データにより得られた圧縮行列データを取得し、

取得した前記圧縮行列データの各行データに基づき特定値を取得し、

取得した複数の特定値に基づき指標値を算出し、

算出した前記指標値及び閾値に基づき、前記圧縮行列データを補正する係数を特定する

処理をコンピュータに実行させるプログラム。

(付記 1 2)

前記圧縮行列データに前記係数を乗じて乗算値を算出し、

前記乗算値を前記圧縮対象行列データの有効桁数に基づき端数処理する

付記 1 1 に記載のプログラム。

(付記 1 3)

取得した複数の特定値に基づき、分散を算出する

付記 1 1 または 1 2 に記載のプログラム。

(付記 1 4)

前記指標値が、前記閾値を超える場合、前記係数を 1 に特定し、前記閾値を超えない場合、前記係数を前記特定値の逆数とする

付記 1 1 から付記 1 3 に記載のプログラム。

(付記 1 5)

第 1 係数と、圧縮対象行列データの有効桁数に基づき決定される第 2 係数とに基づき桁数を算出し、

前記乗算値を、算出した前記桁数に基づき端数処理する

付記 1 2 から 1 4 のいずれか一つに記載のプログラム。

(付記 1 6)

第 1 係数と、圧縮対象行列データの有効桁数及び符号の有無に基づき決定される第 2 係数と、前記圧縮行列データの符号の有無とに基づき桁数を算出し、

前記乗算値を、算出した前記桁数に基づき端数処理する

付記 1 2 から 1 4 のいずれか一つに記載のプログラム。

(付記 1 7)

第 1 係数と、圧縮対象行列データの有効桁数、符号の有無及び小数点の有無に基づき決定される第 2 係数と、前記圧縮行列データの符号の有無及び小数点の有無とに基づき桁数を算出し、

前記乗算値を、算出した前記桁数に基づき端数処理する

付記 1 2 から 1 4 のいずれか一つに記載のプログラム。

(付記 1 8)

圧縮対象行列データ及び辞書行列データにより得られた圧縮行列データを取得し、

取得した前記圧縮行列データの各行データに基づき特定値を取得し、

取得した複数の特定値に基づき指標値を算出し、

算出した前記指標値及び閾値に基づき、前記圧縮行列データを補正する係数を特定する

処理をコンピュータに実行させる情報処理方法。

(付記 1 9)

圧縮対象行列データ及び辞書行列データにより得られた圧縮行列データを取得する第 1 取得部と、

取得した前記圧縮行列データの各行データに基づき特定値を取得する第 2 取得部と、

取得した複数の特定値に基づき指標値を算出する算出部と、

算出した前記指標値及び閾値に基づき、前記圧縮行列データを補正する係

数を特定する特定部と
を備える情報処理装置。

符号の説明

[0053]	1	コンピュータ（情報処理装置、データ圧縮装置）
	1 A	可搬型記録媒体
	1 B	半導体メモリ
	1 O A	読み取り部
	1 1	C P U
	1 2	R A M
	1 3	入力部
	1 4	表示部
	1 5	記憶部
	1 5 P	制御プログラム
	1 6	通信部
	1 8	時計部
	2 0 1	辞書生成部
	2 0 2	圧縮部
	2 0 3	係数算出部
	2 0 4	要素値操作部
	2 0 5	抽出部
	2 0 6	算出部
	2 0 7	生成部
	2 0 8	第2抽出部
	2 0 9	第2算出部
	2 0 1 0	積算出部
	2 0 1 1	桁数算出部
	2 0 1 2	処理部
	N	通信網

請求の範囲

[請求項1] 入力された学習用行列データに基づき、圧縮に用いる辞書行列データを生成する辞書生成部と、

 入力された圧縮対象行列データと生成された前記辞書行列データとに基づき、前記圧縮対象行列データを圧縮した結果である圧縮行列データと、前記圧縮行列データの各行列要素のうち行列要素がゼロではない位置情報とを生成する圧縮部と、

 生成された前記圧縮行列データの各行において、ゼロ値ではない要素値から特定される特定値を連結したデータ列である特定値データ列について算出された指標値と閾値とを比

較し、前記指標値が前記閾値よりも大きい場合、所定の定数を各係数として特定するとともに、前記指標値が前記閾値以下である場合、各特定値の逆数を前記各係数として算出する係数算出部と、

 前記圧縮行列データの各要素について、前記各要素と算出された前記各係数との積について、対応する各要素における小数部の有効数字の桁数に基づいて丸めた結果である操作後の行列データを圧縮結果として出力する要素値操作部と、

 を備えるデータ圧縮装置。

[請求項2] 前記辞書生成部は、

 前記学習用行列データの列データに基づき、該列データとの内積が最大となる前記辞書行列データの列データ及び位置情報を抽出する抽出部と、

 抽出した前記列データ及び前記学習用行列データの列データに基づき最小二乗解を算出する算出部と、

 算出した最小二乗解、前記位置情報及び前記学習用行列データに基づき、前記辞書行列データを生成する生成部と、

 を含む請求項1に記載のデータ圧縮装置。

[請求項3] 前記圧縮部は、

圧縮対象行列データの列データに基づき、該列データとの内積が最大となる前記辞書行列データの列データ及び位置情報を抽出する第2抽出部と、

抽出した前記列データ及び前記圧縮対象行列データの列データに基づき最小二乗解を算出する第2算出部と、

を含む請求項1または2に記載のデータ圧縮装置。

[請求項4]

前記要素値操作部は、

前記圧縮行列データの各要素と前記各係数との積を算出する積算出部と、

第1係数と、前記圧縮対象行列データの有効桁数及び符号の有無に基づき求まる第2係数と、前記圧縮行列データの各要素の符号の有無とに基づき算出される桁数を算出する桁数算出部と、

該桁数算出部で算出した桁数に基づき、前記積算出部で算出した積の端数进行处理する処理部と、

を備える請求項1から3のいずれか一つに記載のデータ圧縮装置。

[請求項5]

桁数算出部は、

第1係数と、前記圧縮対象行列データの有効桁数、符号の有無及び小数点の有無に基づき求まる第2係数と、前記圧縮行列データの各要素の符号の有無及び小数点の有無とに基づき桁数を算出する

請求項4に記載のデータ圧縮装置。

[請求項6]

前記処理部は、

前記積算出部で算出した積に対し、前記桁数算出部で算出した桁数以降の端数进行处理する

請求項4または5に記載のデータ圧縮装置。

[請求項7]

前記閾値は前記特定値の最大値及び最小値に基づき決定される

請求項1から6のいずれか一つに記載のデータ圧縮装置。

[請求項8]

前記特定値は、生成された前記圧縮行列データの各行のゼロ値ではない要素値の平均値、中央値、最大値または二乗平均であり、

前記指標値は、前記特定値を連結したデータ列である特定値データ列の分散である

請求項 1 から 7 のいずれか一つに記載のデータ圧縮装置。

[請求項9] 入力された学習用行列データに基づき、圧縮に用いる辞書行列データを生成し、

入力された圧縮対象行列データと生成された前記辞書行列データとに基づき、前記圧縮対象行列データを圧縮した結果である圧縮行列データと、前記圧縮行列データの各行列要素のうち行列要素がゼロではない位置情報とを生成し、

生成された前記圧縮行列データの各行において、ゼロ値ではない要素値から特定される特定値を連結したデータ列である特定値データ列について算出された指標値と閾値とを比較し、前記指標値が前記閾値よりも大きい場合、所定の定数を各係数として特定するとともに、前記指標値が前記閾値以下である場合、各特定値の逆数を前記各係数として算出し、

前記圧縮行列データの各要素について、前記各要素と算出された前記各係数との積について、対応する各要素における小数部の有効数字の桁数に基づいて丸めた結果である操作後の行列データを圧縮結果として出力する、

処理をコンピュータに実行させるデータ圧縮プログラム。

[請求項10] 入力された学習用行列データに基づき、圧縮に用いる辞書行列データを生成し、

入力された圧縮対象行列データと生成された前記辞書行列データとに基づき、前記圧縮対象行列データを圧縮した結果である圧縮行列データと、前記圧縮行列データの各行列要素のうち行列要素がゼロではない位置情報とを生成し、

生成された前記圧縮行列データの各行において、ゼロ値ではない要素値から特定される特定値を連結したデータ列である特定値データ列

について算出された指標値と閾値とを比較し、前記指標値が前記閾値よりも大きい場合、所定の定数を各係数として特定するとともに、前記指標値が前記閾値以下である場合、各特定値の逆数を前記各係数として算出し、

前記圧縮行列データの各要素について、前記各要素と算出された前記各係数との積について、対応する各要素における小数部の有効数字の桁数に基づいて丸めた結果である操作後の行列データを圧縮結果として出力する、

処理をコンピュータに実行させるデータ圧縮方法。

[請求項11] 圧縮対象行列データ及び辞書行列データにより得られた圧縮行列データを取得し、

取得した前記圧縮行列データの各行データに基づき特定値を取得し、

取得した複数の特定値に基づき指標値を算出し、

算出した前記指標値及び閾値に基づき、前記圧縮行列データを補正する係数を特定する

処理をコンピュータに実行させるプログラム。

[請求項12] 前記圧縮行列データに前記係数を乗じて乗算値を算出し、
前記乗算値を前記圧縮対象行列データの有効桁数に基づき端数処理する

請求項11に記載のプログラム。

[請求項13] 取得した複数の特定値に基づき、分散を算出する

請求項11または12に記載のプログラム。

[請求項14] 前記指標値が、前記閾値を超える場合、前記係数を1に特定し、前記閾値を超えない場合、前記係数を前記特定値の逆数とする

請求項11から請求項13に記載のプログラム。

[請求項15] 第1係数と、圧縮対象行列データの有効桁数に基づき決定される第2係数とに基づき桁数を算出し、

前記乗算値を、算出した前記桁数に基づき端数処理する

請求項 1 2 から 1 4 のいずれか一つに記載のプログラム。

[請求項16]

第 1 係数と、圧縮対象行列データの有効桁数及び符号の有無に基づき決定される第 2 係数と、前記圧縮行列データの符号の有無とに基づき桁数を算出し、

前記乗算値を、算出した前記桁数に基づき端数処理する

請求項 1 2 から 1 4 のいずれか一つに記載のプログラム。

[請求項17]

第 1 係数と、圧縮対象行列データの有効桁数、符号の有無及び小数点の有無に基づき決定される第 2 係数と、前記圧縮行列データの符号の有無及び小数点の有無とに基づき桁数を算出し、

前記乗算値を、算出した前記桁数に基づき端数処理する

請求項 1 2 から 1 4 のいずれか一つに記載のプログラム。

[請求項18]

圧縮対象行列データ及び辞書行列データにより得られた圧縮行列データを取得し、

取得した前記圧縮行列データの各行データに基づき特定値を取得し、

取得した複数の特定値に基づき指標値を算出し、

算出した前記指標値及び閾値に基づき、前記圧縮行列データを補正する係数を特定する

処理をコンピュータに実行させる情報処理方法。

[請求項19]

圧縮対象行列データ及び辞書行列データにより得られた圧縮行列データを取得する第 1 取得部と、

取得した前記圧縮行列データの各行データに基づき特定値を取得する第 2 取得部と、

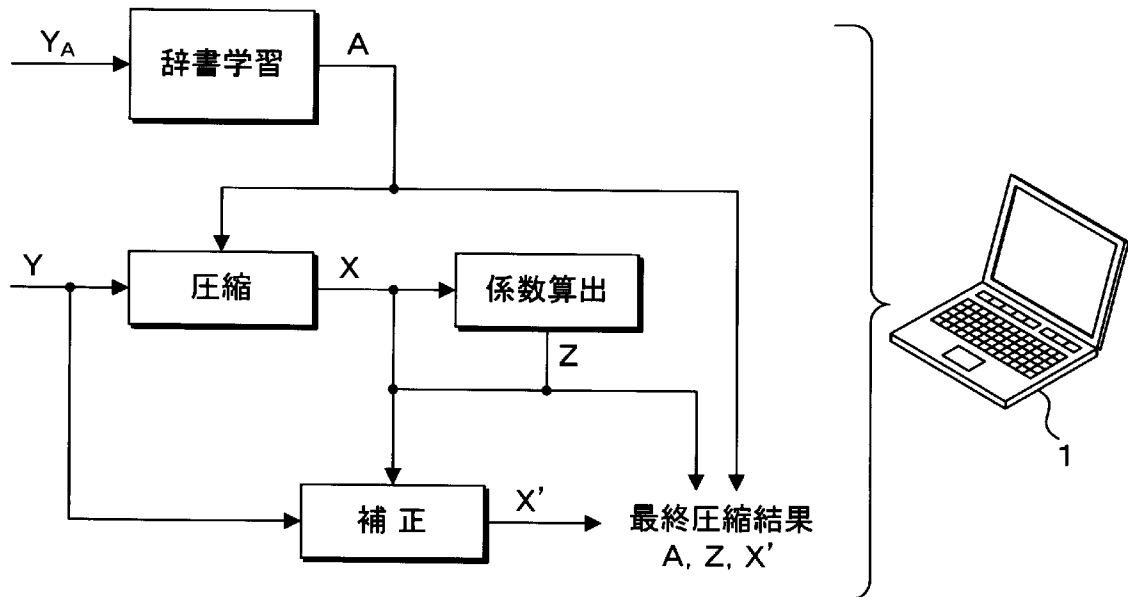
取得した複数の特定値に基づき指標値を算出する算出部と、

算出した前記指標値及び閾値に基づき、前記圧縮行列データを補正する係数を特定する特定部と

を備える情報処理装置。

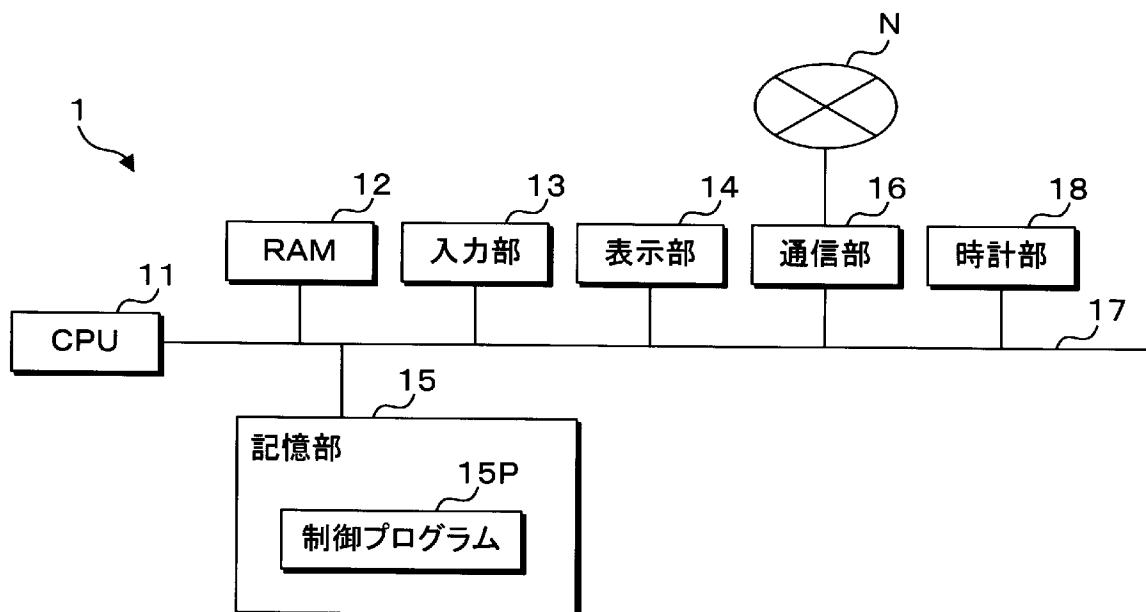
[図1]

処理概要を示す説明図



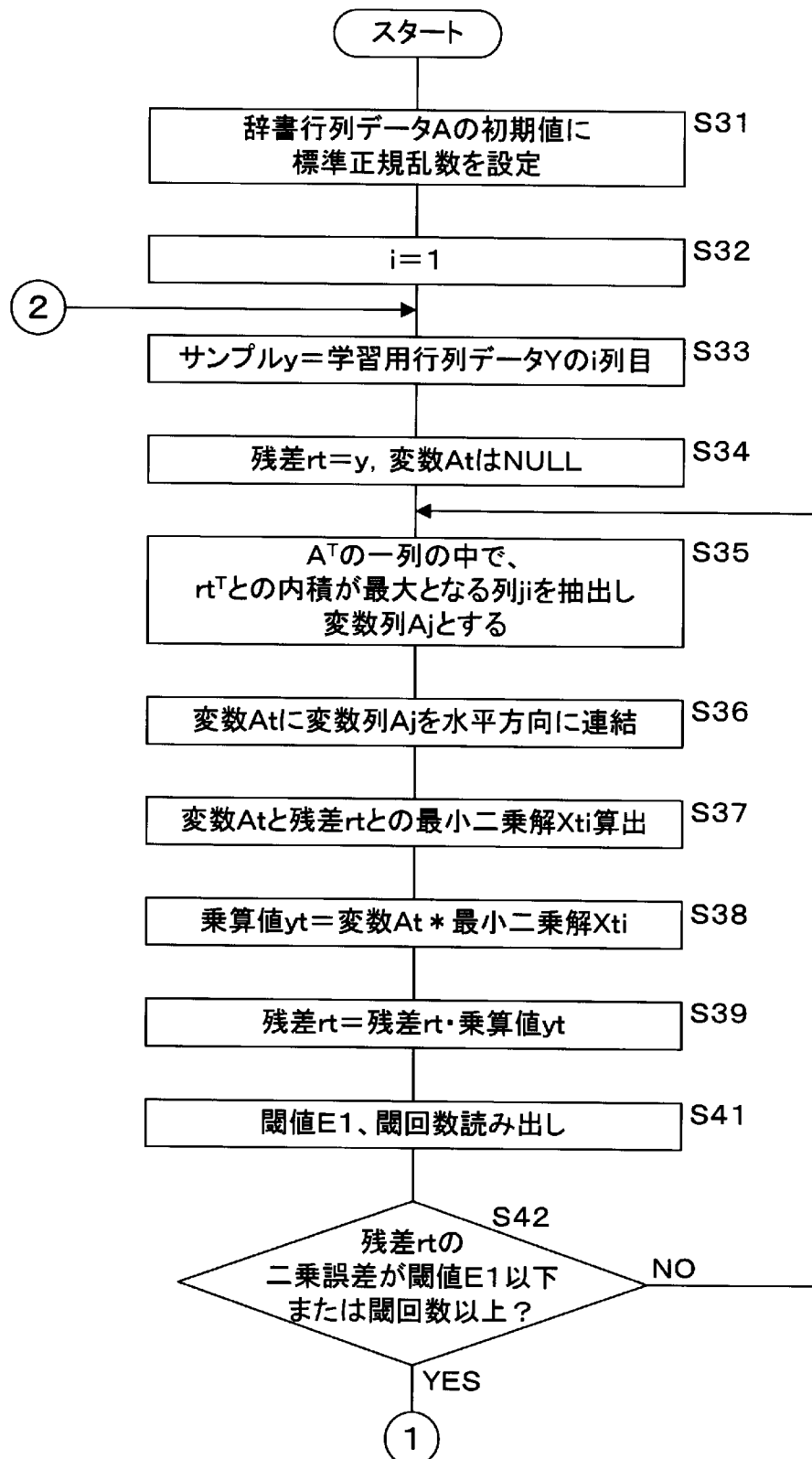
[図2]

コンピュータのハードウェア群を示すブロック図



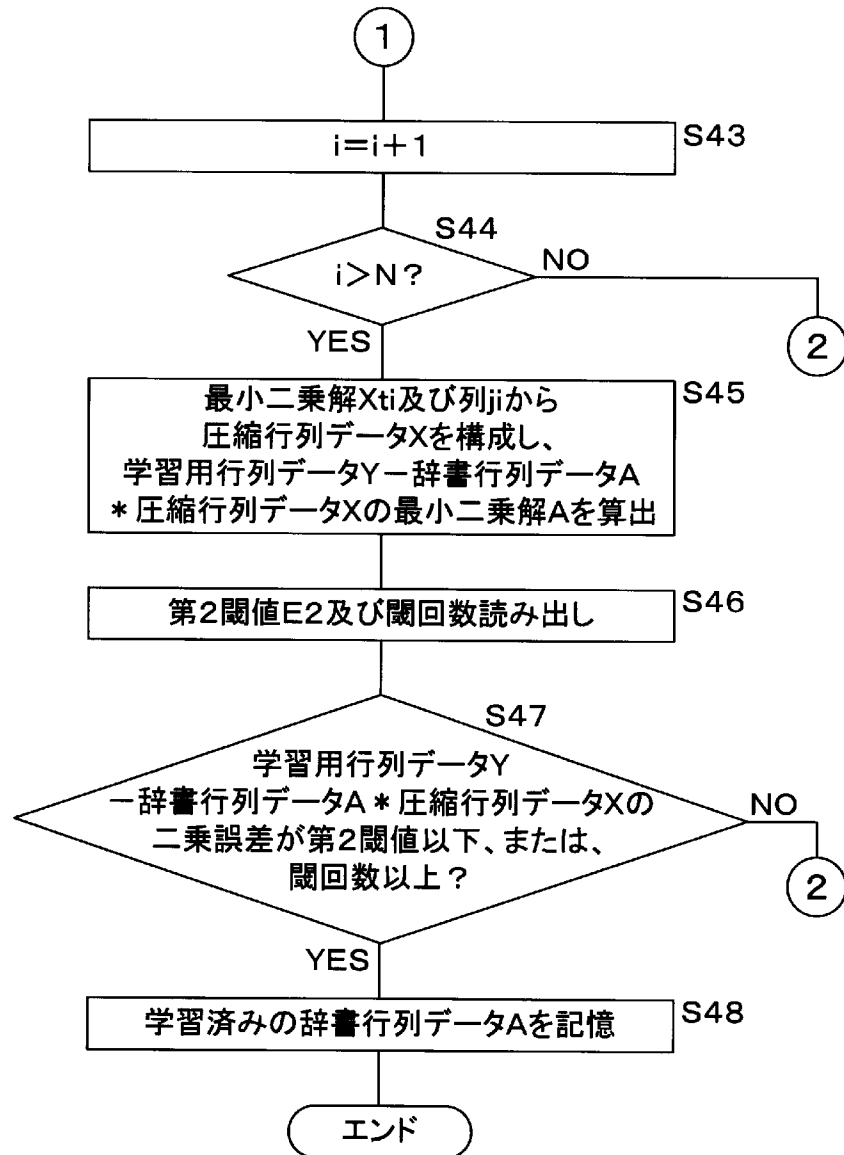
[図3]

辞書学習処理の手順を示すフローチャート



[図4]

辞書学習処理の手順を示すフローチャート



[図5]

残差 r_t 及び変数 A_t の数値例を示す説明図

	i=1			i=2		i=3	
	A_t (ステップ36)	X_{it} (ステップ37)	r_t (ステップ39)	A_t (ステップ36)2列目は、 A の292列目とベクトルの向きが類似している X_{it} (ステップ37)	X_{it} (ステップ37)	A_t (ステップ36)3列目は、 A の169列目とベクトルの向きが類似している X_{it} (ステップ37)	X_{it} (ステップ37)
2.6	0.128279876	7.652667745	1.618316731	0.128279876	7.509094505	0.085296072	-0.06100909
2.6	-0.094491259		3.323110212	-0.094491259	-0.007184447	-0.007184447	0.028966094
2.6	0.173789621		1.270045774	0.173789621	0.041643997	0.041643997	0.01982938
2.4	-0.035794966		2.673926978	-0.035794966	-0.021573164	-0.021573164	0.197872673
2.7	-0.138781861		3.762051471	-0.138781861	0.103897993	0.103897993	-0.15382105
2.5	-0.059144041		2.952609694	-0.059144041	-0.038707798	-0.038707798	0.142886426
2.5	0.089335362		1.816346158	0.089335362	-0.296293343	-0.296293343	0.140569256
2.6	0.100943526		1.827512734	0.100943526	-0.006369857	-0.006369857	0.106983676
2.6	0.015317225		2.482782364	0.015317225	-0.030554007	-0.030554007	0.032807458
2.6	0.027474569		2.389746248	0.027474569	0.189264316	0.189264316	0.100701121
2.4	0.060503789		1.936984608	0.060503789	0.180160113	0.180160113	-0.18074367
2.6	-0.139479828		3.66739278	-0.139479828	-0.044382842	-0.044382842	-0.06652218
2.7	-0.055609626		3.12556199	-0.055609626	0.175488839	0.175488839	0.108553293
2.7	-0.174956537		4.038884246	-0.174956537	0.102899993	0.102899993	0.130954935
2.5	-0.042932846		2.828550808	-0.042932846	0.093322052	0.093322052	-0.09513404
2.4	0.090747685		1.705538116	0.090747685	0.011399496	0.011399496	0.073893986
2.5	0.068585149		1.975140641	0.068585149	-0.039269941	-0.039269941	-0.055447
2.5	0.218709241		0.826290844	0.218709241	-0.037358078	-0.037358078	-0.15877413
2.4	0.015938642		2.27802687	0.015938642	0.044299202	0.044299202	0.163180034
2.5	-0.206153754		4.077626186	-0.206153754	0.056170976	0.056170976	-0.03774664

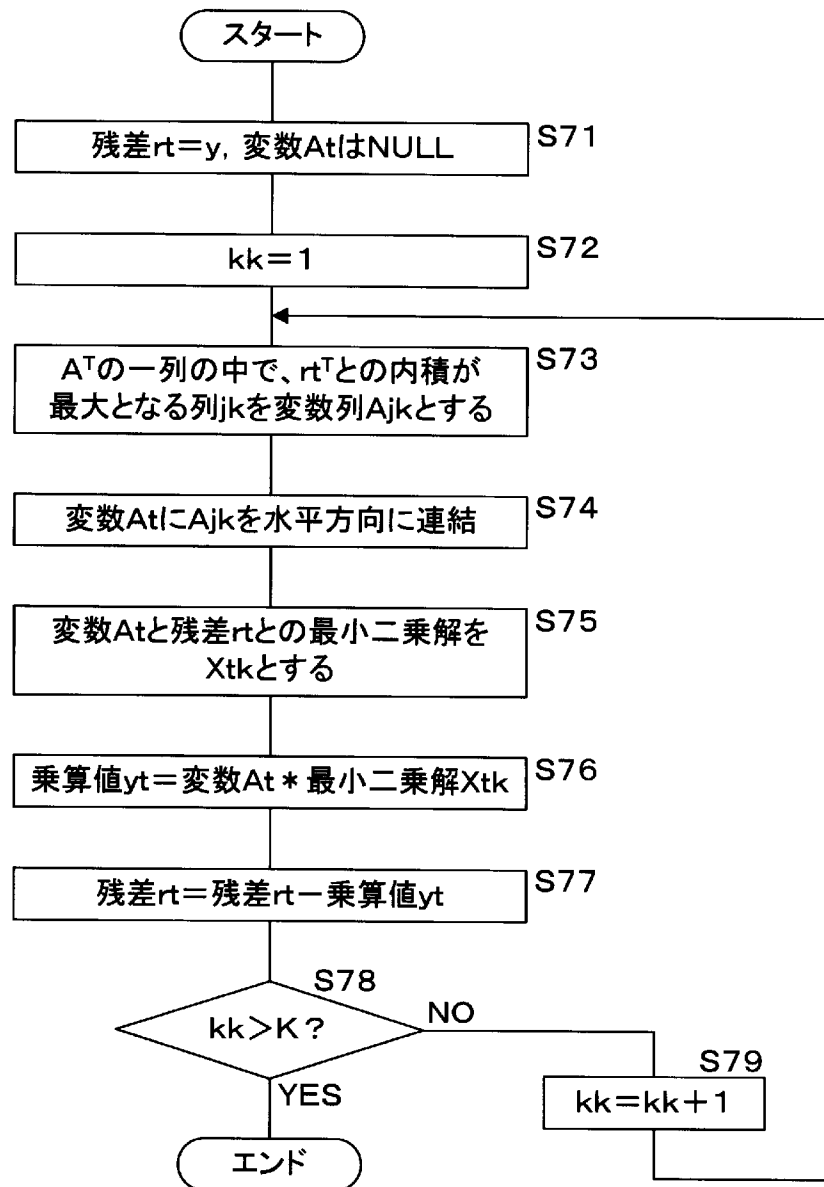
[図6]

辞書行列データの数値例を示す説明図

A(iのループ内では変更なし、初期値は乱数)				
1列目	13列目	169列目	292列目	
0.04622694	0.128279876	-0.061009094	0.085296072	
0.15767171	-0.09449126	0.028966094	-0.007184447	
-0.1942086	0.173789621	0.01982938	0.041643997	
0.07412697	-0.03579497	0.197872673	-0.021573164	
0.02740644	-0.13878186	-0.153821055	0.103897993	
-0.112431	-0.05914404	0.142886426	-0.038707798	
-0.0372789	0.089335362	0.140569256	-0.296293343	
0.02945778	0.100943526	0.106983676	-0.006369857	
0.3076594	0.015317225	0.032807458	-0.030554007	
0.23810755	0.027474569	0.100701121	0.189264316	
-0.1160591	0.060503789	-0.180743671	0.180160113	
0.26093324	-0.13947983	-0.066522184	-0.044382842	
0.06236799	-0.05560963	0.108553295	0.175488839	
-0.0054213	-0.17495654	0.130954935	0.102899993	
0.06145136	-0.04293285	-0.095134039	0.093322052	
-0.0176223	0.090747685	0.073893986	0.011399496	
-0.0106735	0.068585149	-0.055446998	-0.039269941	
0.12807955	0.218709241	-0.158774135	-0.037358078	
0.12114439	0.015938642	0.163180034	0.044299202	
0.12184578	-0.20615375	-0.037746639	0.056170976	

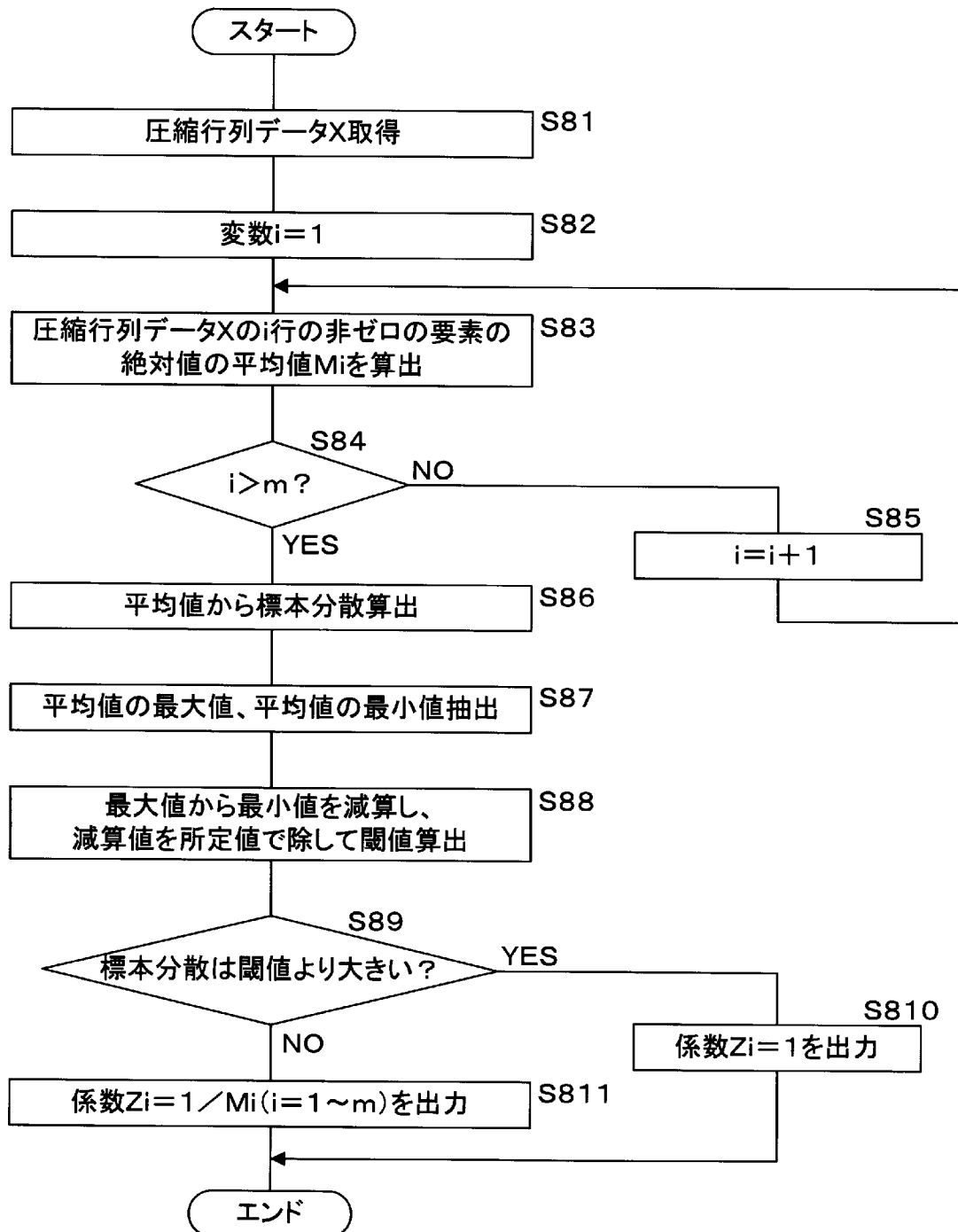
[図7]

圧縮処理の手順を示すフローチャート



[図8]

係数算出処理の手順を示すフローチャート



[図9]

圧縮行列データのデータ例を示す説明図

$$X = \begin{bmatrix} 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & \dots \\ 16.1397 & 0 & 0 & \dots \\ 0 & -2.11269 & 0 & \dots \\ 0 & 0 & 6.53215 & \dots \\ 0 & 0 & -2.63519 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

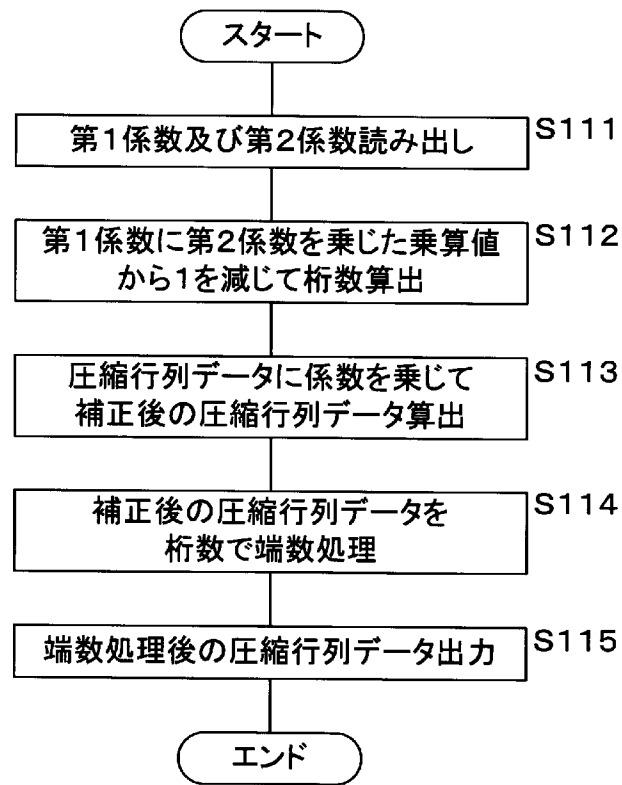
[図10]

係数のデータ例を示す説明図

Y1	Y2	X1	X2	Z
2.6	2.5	16.1397	15.0513	0.0641210
2.6	2.5	-3.98809	-2.11269	0.3278269
2.6	2.5	2.69214	1.55938	0.4704199
2.4	2.5	3.76721	1.54149	0.3767400
2.7	2.5	-1.84536	0.761282	0.7672706
2.5	2.4	0.18113	0.163425	5.8045885
2.5	2.4	0.530031	0.13557	3.0048031
2.6	2.5	7.39294	7.36217	0.1355462
2.6	2.4			
2.6	2.4			
2.4	2.4			
2.6	2.4			
2.7	2.5			
2.7	2.5			
2.5	2.5			

[図11]

補正処理及び端数処理の手順を示すフローチャート



[図12]

数値例を示す説明図

Y1	Y2	X1	X2	Z	Xtemp1 = X1*Z	Xtemp2 = X2*Z
2.6	2.5	16.1397	15.0513	0.0641210	1.0348946	0.9651053
2.6	2.5	-3.98809	-2.11269	0.3278269	-1.3074033	-0.6925966
2.6	2.5	2.69214	1.55938	0.4704199	1.2664364	0.7335635
2.4	2.5	3.76721	1.54149	0.3767400	1.4192589	0.5807410
2.7	2.5	-1.84536	0.761282	0.7672706	-1.4158906	0.5841093
2.5	2.4	0.18113	0.163425	5.8045885	1.051385	0.948614
2.5	2.4	0.530031	0.13557	3.0048031	1.5926388	0.4073611
2.6	2.5	7.39294	7.36217	0.1355462	1.0020853	0.9979146
2.6	2.4					
2.6	2.4					
2.4	2.4					
2.6	2.4					
2.7	2.5					
2.7	2.5					
2.5	2.5					

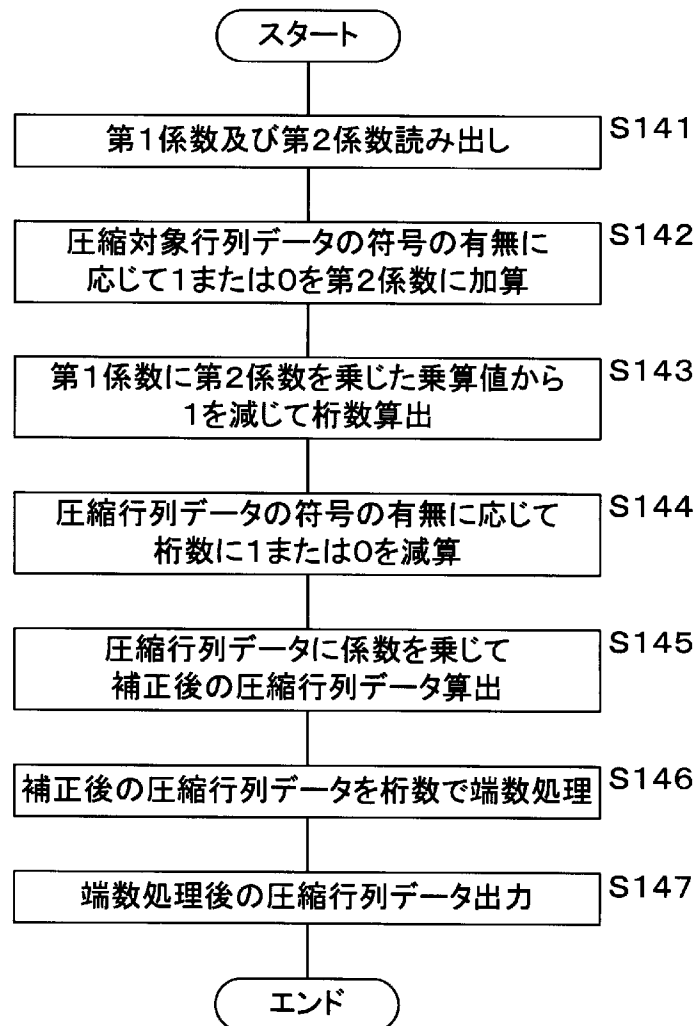
[図13]

数値例を示す説明図

X1'	X2'
1.035	0.965
-1.307	-0.693
1.266	0.734
1.419	0.581
-1.416	0.584
1.051	0.949
1.593	0.407
1.002	0.998

[図14]

補正処理及び端数処理の手順を示すフローチャート



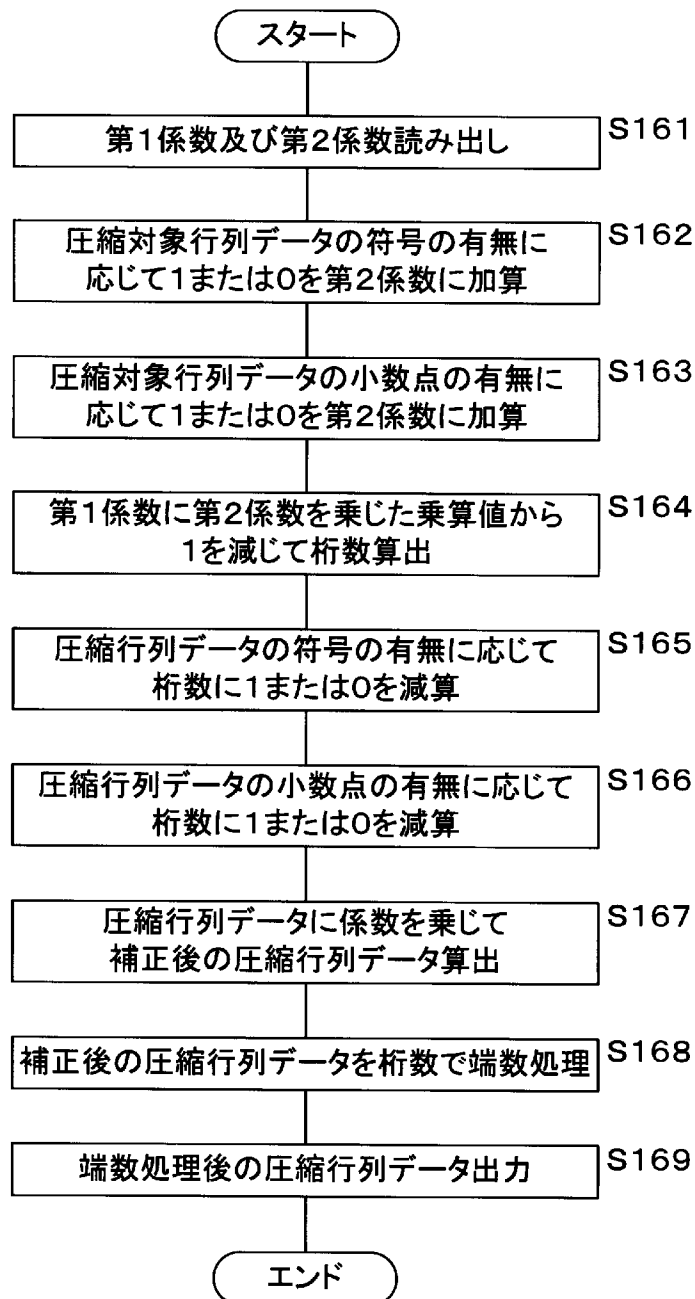
[図15]

数値例を示す説明図

X1'	X2'
1.03	0.97
-1.31	-0.69
1.27	0.73
1.42	0.58
-1.42	0.58
1.05	0.95
1.59	0.41
1.00	1.00

[図16]

補正処理及び端数処理の手順を示すフローチャート



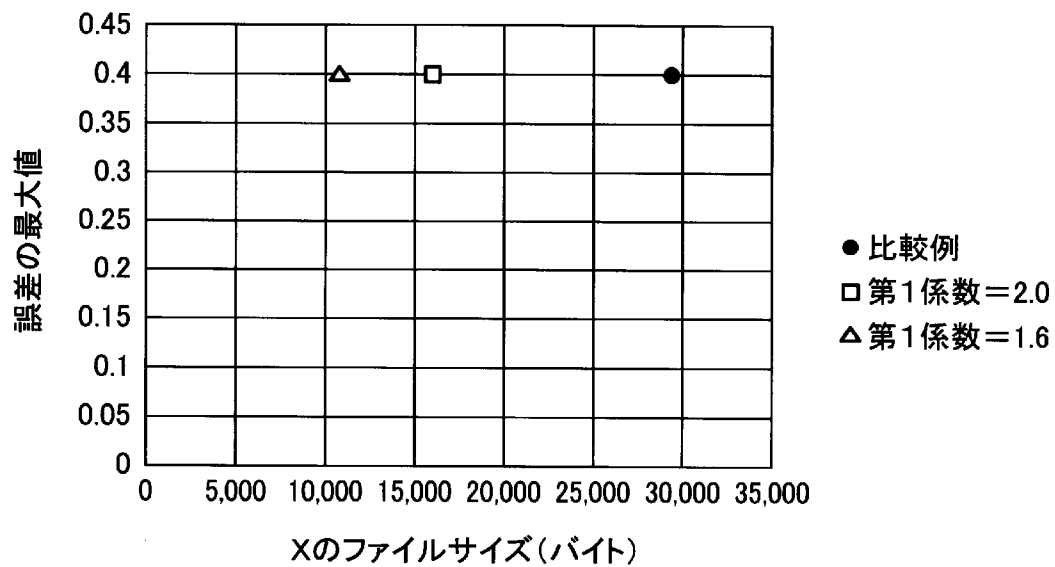
[図17]

数値例を示す説明図

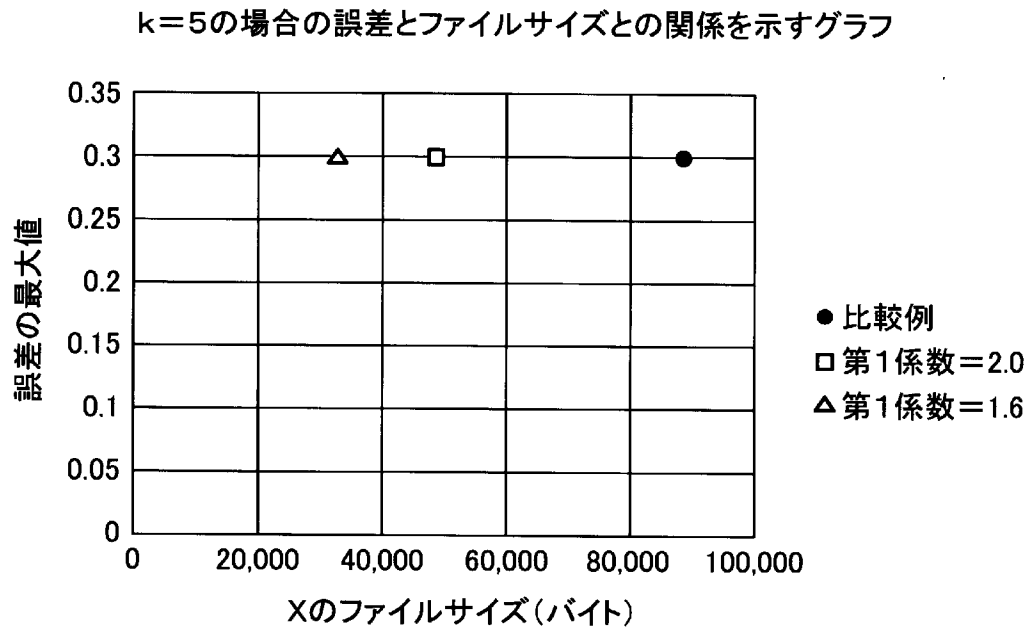
X1'	X2'
1.035	0.965
-1.307	-0.693
1.266	0.734
1.419	0.581
-1.416	0.584
1.051	0.949
1.593	0.407
1.002	0.998

[図18]

k=2の場合の誤差とファイルサイズとの関係を示すグラフ

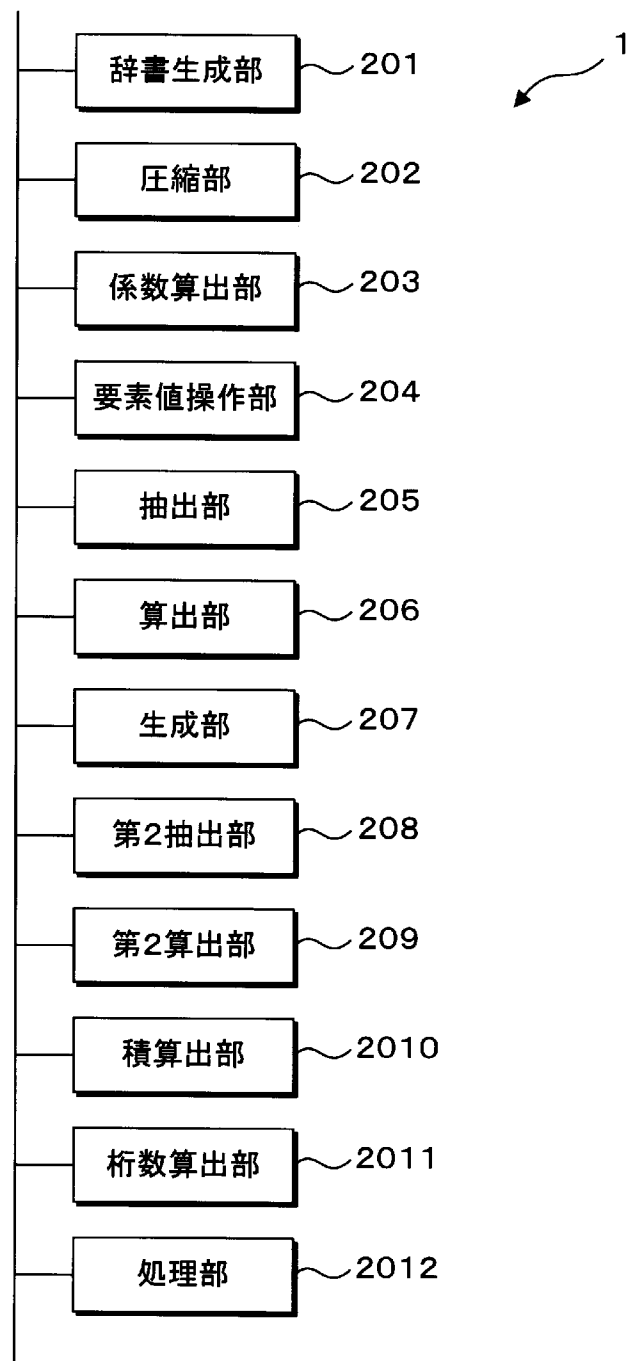


[図19]



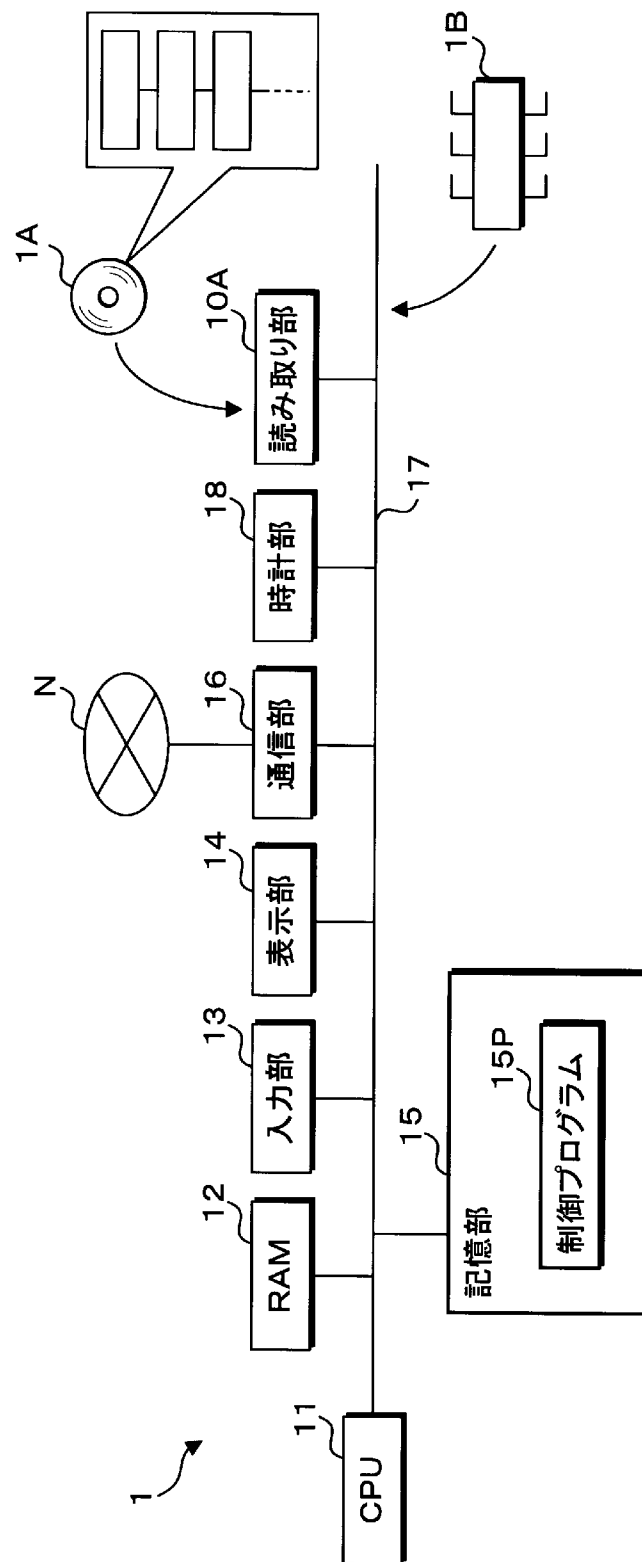
[図20]

上述した形態のコンピュータの動作を示す機能ブロック図



[図21]

実施の形態4に係るコンピュータのハードウェア群を示すブロック図



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2017/047073

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl. H03M7/30 (2006.01) i, G06F17/16 (2006.01) i, G06N99/00 (2010.01) i

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl. H03M7/30, G06F17/16, G06N99/00

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	1971-2018
Registered utility model specifications of Japan	1996-2018
Published registered utility model applications of Japan	1994-2018

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

IEEE Xplore

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2016-515732 A (TOSHIBA CORPORATION) 30 May 2016 & US 2016/0043884 A1 & WO 2014/155026 A1	1-19
A	US 2013/0318020 A1 (GEORGIA TECH RESEARCH CORPORATION) 28 November 2013 (Family: none)	1-19
A	坂田 緩香, 棒島 祥介, Dictionary Learning におけるサンプル複雑度の典型時解析, 電子情報通信学会技術研究報告, 31 October 2012, vol. 112, no. 279, pp. 113-120 (SAKATA, Ayaka, KABASHIMA, Yoshiyuki, Typical case analysis of sample complexity in Dictionary Learning, IEICE Technical Report)	1-19

☐ 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
15.03.2018

Date of mailing of the international search report
27.03.2018

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

Authorized officer

Telephone No.

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int.Cl. H03M7/30(2006.01)i, G06F17/16(2006.01)i, G06N99/00(2010.01)i

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int.Cl. H03M7/30, G06F17/16, G06N99/00

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	1971-2018年
日本国実用新案登録公報	1996-2018年
日本国登録実用新案公報	1994-2018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)

IEEE Xplore

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 2016-515732 A (株式会社東芝) 2016.05.30 & US 2016/0043884 A1 & WO 2014/155026 A1	1-19
A	US 2013/0318020 A1 (GEORGIA TECH RESEARCH CORPORATION) 2013.11.28 (ファミリーなし)	1-19
A	坂田 綾香, 樺島 祥介, Dictionary Learning におけるサンプル複雑度の典型時解析, 電子情報通信学会技術研 究報告, 2012.10.31, VOL.112, NO.279, pp.113-120	1-19

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矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「&」 同一パテントファミリー文献

国際調査を完了した日

15.03.2018

国際調査報告の発送日

27.03.2018

国際調査機関の名称及びあて先

日本国特許庁 (ISA/J P)

郵便番号 100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

北村 智彦

5 K

9297

電話番号 03-3581-1101 内線 3556