

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2023年9月7日(07.09.2023)



(10) 国際公開番号

WO 2023/166579 A1

(51) 国際特許分類:
G06F 16/75 (2019.01) G16Y 40/20 (2020.01)

(21) 国際出願番号: PCT/JP2022/008750

(22) 国際出願日: 2022年3月2日(02.03.2022)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(71) 出願人: 日本電気株式会社 (NEC CORPORATION) [JP/JP]; 〒1088001 東京都港区芝五丁目7番1号 Tokyo (JP).

(72) 発明者: 山下 哲孝 (YAMASHITA Noritaka); 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP). 鹿嶋 卓郎 (KASHIMA Takuroh); 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP). 大井 憲人 (OI Norihito); 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP). 藤井 秋紗子 (FUJII Asako); 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP).

(74) 代理人: 岩壁 冬樹, 外 (IWAKABE Fuyuki et al.); 〒1040031 東京都中央区京橋一丁目14

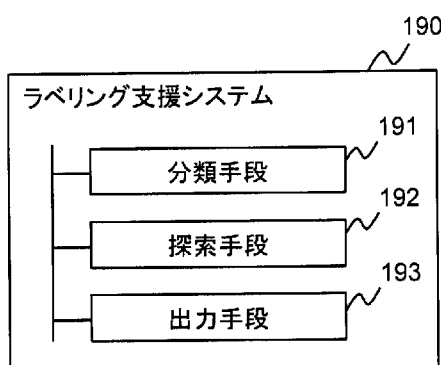
番4号 京橋TSビル8階 サンライズ
国際特許事務所 Tokyo (JP).

(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, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, 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, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS,

(54) Title: LABELLING ASSISTANCE SYSTEM, LABELLING ASSISTANCE METHOD, AND LABELLING ASSISTANCE PROGRAM

(54) 発明の名称: ラベリング支援システム、ラベリング支援方法およびラベリング支援プログラム



190 Labelling assistance system
191 Sorting means
192 Searching means
193 Output means

(57) Abstract: In the present invention, a sorting means 191 generates a plurality of clusters by sorting data to be labelled by learning without a teacher. A searching means 192 searches, in each generated cluster, for commonalities among data items contained in such clusters. An output means 193 outputs, for each cluster, information related to the commonalities that were searched for.

(57) 要約: 分類手段191は、ラベリング対象のデータを教師なし学習により分類することで複数のクラスタを生成する。探索手段192は、生成されたクラスタごとに、そのクラスタに含まれるデータの共通点を探索する。出力手段193は、クラスタごとに、探索された共通点に関する情報を出力する。

[続葉有]

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)）

明 細 書

発明の名称：

ラベリング支援システム、ラベリング支援方法およびラベリング支援プログラム

技術分野

[0001] 本発明は、ラベル付けされていないデータに対するラベリングを支援するラベリング支援システム、ラベリング支援方法およびラベリング支援プログラムに関する。

背景技術

[0002] I o T (Internet of Things) 社会において、様々な機器からデータを収集することが可能になっている。そして、収集される大量のデータを用いて行われるデータの検索やA I (Artificial Intelligence) の学習には、データの分類が重要になる。

[0003] このような状況において、データの分類を支援する方法が各種提案されている。例えば、特許文献1には、多数のセンサにより得られるセンサデータをその特徴に応じて分類するセンサデータ分類装置が記載されている。特許文献1に記載された装置は、予め設定した時間区間ごとに分割されたセンサデータの集合をセンサ識別子および分割区間識別子と関連付け、分割データの集合に含まれるデータからその複数種の特徴パラメータを算出する。

先行技術文献

特許文献

[0004] 特許文献1：特開2016-99888号公報

発明の概要

発明が解決しようとする課題

[0005] 例えば、データの特徴量を抽出してクラスタに分類した場合、クラスタへの意味付け（ラベリング）を行うことが重要である。しかし、クラスタ化さ

れた個々のデータを見てラベリングするのは非常にコストのかかる作業であり、特にそのクラスタに含まれるデータ量が大量である場合、その影響は顕著である。

[0006] さらに、例えば、分類対象のデータが動画であるような場合、そのデータの確認には時間がかかる。また、分類対象のデータに複数のセンサデータが含まれる場合、どのデータに着目すべきか判断することも煩雑な作業になる。

[0007] 特許文献1に記載された装置では、分類を行うための特徴パラメータの計算方法や、分割区間を予め定められる。しかし、何らかの基準に基づいて算出された数値からデータを分類したとしても、ラベル付けされていないデータに対して意味のあるラベリング作業を行うには、やはりコストがかかってしまうという問題がある。

[0008] そこで、本発明は、ラベル付けされていないデータが分類されたクラスタに対するラベリング作業を支援できるラベリング支援システム、ラベリング支援方法およびラベリング支援プログラム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

[0009] 本発明によるラベリング支援システムは、ラベリング対象のデータを教師なし学習により分類することで複数のクラスタを生成する分類手段と、生成されたクラスタごとに、そのクラスタに含まれるデータの共通点を探索する探索手段と、クラスタごとに、探索された共通点に関する情報を出力する出力手段とを備えたことを特徴とする。

[0010] 本発明によるラベリング支援方法は、コンピュータが、ラベリング対象のデータを教師なし学習により分類することで複数のクラスタを生成し、コンピュータが、生成されたクラスタごとに、そのクラスタに含まれるデータの共通点を探索し、コンピュータが、クラスタごとに、探索された共通点に関する情報を出力することを特徴とする。

[0011] 本発明によるラベリング支援プログラムは、コンピュータに、ラベリング対象のデータを教師なし学習により分類することで複数のクラスタを生成す

る分類処理、生成されたクラスタごとに、そのクラスタに含まれるデータの共通点を探索する探索処理、および、クラスタごとに、探索された共通点に関する情報を出力する出力処理を実行させることを特徴とする。

発明の効果

[0012] 本発明によれば、ラベル付けされていないデータが分類されたクラスタに対するラベリング作業を支援できる。

図面の簡単な説明

[0013] [図1]本発明によるラベリング支援システムの一実施形態の構成例を示すブロック図である。

[図2]ラベリング支援システムで利用されるデータの例を示す説明図である。

[図3]特徴量の例を示す説明図である。

[図4]次元削減されたデータをグラフで可視化した例を示す説明図である。

[図5]センサごとの寄与度をグラフ表示した例を示す説明図である。

[図6]クラスタ内におけるセンサ値の分布の例を示す説明図である。

[図7]クラスタ内における統計量の例を示す説明図である。

[図8]ラベリング支援システムの動作例を示すフローチャートである。

[図9]本発明によるラベリング支援システムの概要を示すブロック図である。

[図10]少なくとも1つの実施形態に係るコンピュータの構成を示す概略ブロック図である。

発明を実施するための形態

[0014] 以下、本発明の実施形態を図面を参照して説明する。以下の説明では、ラベル付けされていないデータの一例として、動画（映像データ）を例示する。ただし、ラベル付けされていないデータは、動画に限られず、例えば、静止画や、音楽データ、テキストデータなどであってもよい。また、ラベル付けされていないデータ（ラベリング対象のデータ）のことを、以下、未分類データと記すこともある。

[0015] 図1は、本発明によるラベリング支援システムの一実施形態の構成例を示すブロック図である。本実施形態のラベリング支援システム1は、データ取

得部 10 と、関連情報取得部 20 と、物体識別部 30 と、データ加工部 40 と、テキスト情報入力部 50 と、特徴抽出部 60 と、特徴記憶部 70 と、可視化処理部 80 と、入出力装置 90 とを備えている。

[0016] データ取得部 10 は、ラベリング対象のデータ（すなわち、未分類データ）を取得する。例えば、カメラ（図示せず）によって走行する車両が撮像されている場合、データ取得部 10 は、ラベリング対象のデータとして、そのカメラが撮影した車両の動画を取得してもよい。なお、データ取得部 10 が取得するデータは、リアルタイムで取得されるデータに限られない。データ取得部 10 は、例えば、ラベリング対象のデータが記憶されたストレージサーバ（図示せず）から、ラベリング対象のデータを取得してもよい。

[0017] 関連情報取得部 20 は、ラベリング対象のデータに関連する情報（以下、関連情報と記す。）を取得する。本実施形態では、関連情報は、ラベリング対象のデータの生成された状況を示す情報であり、例えば、データが生成された場所（撮像された場所）や時間を表わす情報、センサにより取得されたデータ（以下、センサデータと記す。）である。

[0018] 例えば、ラベリング対象のデータが、車載カメラ（ドライブレコーダ）で撮像された映像データである場合、関連情報として車両位置を表わす GPS（Global Positioning System）情報や、CAN（Controller Area Network）に基づいて取得される情報などが挙げられる。この場合に取得されるセンサデータの例が、速度や加速度、位置（緯度、経度、高度など）である。

[0019] また、ラベリング対象のデータとして火力発電所の稼働状況を示す映像が用いられる場合、センサデータとして、例えば、燃料の流量、圧力、温度、回転数、発電量などが挙げられる。他にも、ラベリング対象のデータとして農場の状況を示す映像が用いられる場合、センサデータとして、時間や温度、湿度、pH、土壌水分量、日射量、風向・風速、水位などが挙げられる。

[0020] 物体識別部 30 は、取得されたデータに含まれる物体を識別し、識別した物体を特定する情報（以下、オブジェクトリストと記す。）を生成する。例えば、識別対象の物体が車両の場合、物体識別部 30 は、データ取得部 10

が取得したデータから、車両を識別し、その車両を特定する情報（例えば、画像中の位置を示す座標等）をオブジェクトリストとして生成してもよい。なお、画像や映像から物体を識別する方法は広く知られており、ここでは詳細な説明は省略する。

[0021] データ加工部40は、後述する特徴抽出部60が処理を行う際に用いることができる態様にデータ（より具体的には、オブジェクトリスト）を加工する。具体的には、データ加工部40は、特徴抽出やクラスタリングの精度を向上させられるようにデータを加工する。データ加工部40は、例えば、データの間引きや、欠損値の補間、外れ値の除外、不要なデータ項目の削除などを行う。また、例えば、ラベリング対象のデータが映像データの場合、データ加工部40は、映像データを数値時系列データへ変換してもよい。

[0022] テキスト情報入力部50は、ラベリング対象の各データに付加する情報（以下、付加情報と記す。）を含むテキストデータの入力を受け付ける。付加情報は、関連情報以外で取得し得るラベリング対象のデータの内容を示す情報である。付加情報を示すカテゴリとして、例えば、天気や植物の種類、交通参加者などが挙げられる。天気のカテゴリ値の例として、晴れ・曇り・雨・雪などが挙げられ、植物の種類のカテゴリ値の例として、米・小麦・大麦などが挙げられ、交通参加者の例として、自動車・自転車・歩行者などが挙げられる。

[0023] なお、テキストデータの入力は任意である。すなわち、ラベリング対象のデータに対する付加情報が入力されていなくてもよい。ただし、ラベリング対象のデータに付加情報が増えるほど、分類の精度を向上できるため、入力されることが好ましい。以下の説明では、付加情報が対応付けられたラベリング対象のデータも、単にラベリング対象のデータと記す。

[0024] 図2は、本実施形態のラベリング支援システム1で利用されるデータの例を示す説明図である。図2に示す例では、データ取得部10がラベリング対象のデータとして映像11を取得し、関連情報取得部20は、映像11が撮影された場所等に関する関連情報21を取得したことを示す。また、図2に

示す例では、データ加工部40が、映像11および関連情報21（より具体的には、物体識別部30により生成されたオブジェクトリスト）を加工して数値時系列データ41を生成したことを示す。さらに、図2に示す例では、テキスト情報入力部50が、付加情報として、天気、シーン、時間帯および物体に関する情報を含むテキストデータ51の入力を受け付けたことを示す。

[0025] 特徴抽出部60は、ラベリング対象の各データから特徴を抽出する。本実施形態の特徴抽出部60は、付加情報を含むラベリング対象の各データを教師なし学習により自動的に分類することで複数のクラスタを生成する。教師なし学習によりクラスタを生成する方法は任意であり、例えば、*k-means*法や、混合ガウスモデルなどが挙げられる。

[0026] そして、特徴抽出部60は、生成したクラスタに含まれる各データの特徴量を抽出する。特徴抽出部60は、例えば、テキストデータに含まれている付加情報を特徴量として抽出してもよい。他にも、特徴抽出部60は、数値時系列データが示す特徴量を抽出してもよい。具体的には、特徴抽出部60は、ラベリング対象のデータ（より具体的には、数値時系列データ）に含まれるセンサ値に基づいて特徴量を抽出してもよい。

[0027] なお、数値時系列データから特徴量を抽出する方法は任意である。例えば、*k-means*法により生成された各クラスタについて、特徴抽出部60は、クラスタに含まれる数値時系列データの重心点から各データまでの距離（cluster distance feature）という特徴量を抽出してもよい。

[0028] このように、特徴抽出部60は、ラベリング対象のデータを教師なし学習により分類する処理を行うことから、分類手段ということもできる。また、本実施形態では、データ取得部10と関連情報取得部20により取得された情報から物体識別部30が物体を識別し、識別結果に対してデータ加工部40が、特徴抽出部60が用いる形式にデータを加工する場合について説明した。ただし、データ取得部10が、直接、特徴抽出部60が用いる形式のデータを取得し、取得したデータを特徴抽出部60に入力してもよい。この場

合、ラベリング支援システム１は、関連情報取得部２０、物体識別部３０およびデータ加工部４０を備えていなくてもよい。

[0029] 特徴記憶部７０は、特徴抽出部６０が抽出した各データの特徴量を記憶する。特徴記憶部７０がデータごとの特徴量を記憶する態様は任意である。図３は、特徴記憶部７０が記憶する特徴量の例を示す説明図である。図３に示す例では、縦方向が１つの特徴点を表わし、横方向が各カテゴリ（例えば、天気、交通参加者、植物の種類など）の特徴量（カテゴリ値）を表わしている。特徴記憶部７０は、例えば、磁気ディスク等により実現される。

[0030] 可視化処理部８０は、生成されたクラスタに対するラベリング作業に寄与する情報を可視化するための処理を行う。可視化処理部８０は、探索部８１と、出力部８２とを含む。

[0031] 探索部８１は、生成されたクラスタごとに、そのクラスタに含まれるラベリング対象の各データの共通点を探索する。具体的には、探索部８１は、生成されたクラスタに含まれる各データの特徴量を抽出し、抽出された各データの特徴量の共通点を探索する。探索部８１は、特徴量として、抽出された各カテゴリにおけるカテゴリ値の共通点を探索してもよく、数値時系列データに基づいて抽出される特徴量の共通点を探索してもよい。

[0032] 例えば、上記に示すカテゴリに着目した場合、探索部８１は、クラスタ内のデータでカテゴリ値が共通する割合が予め定めた閾値を超えた場合、そのカテゴリ値を共通点としてもよい。具体的には、割合は、クラスタ内のデータ数に対する共通点を含むデータ数の比率に基づいて算出できる。なお、この場合、探索部８１は、全カテゴリのカテゴリ値について共通点を探索してもよく、任意の一部のカテゴリのカテゴリ値について共通点を探索してもよい。

[0033] 他にも、共通点を探索する処理として、探索部８１は、ラベリング対象のデータが示すカテゴリごとに、最も共通するカテゴリ値（例えば、数値であれば最頻値）を共通点として探索してもよい。そして、探索部８１は、最も共通するカテゴリ値の割合が最も高いカテゴリ値を共通点として特定しても

よい。

[0034] また、例えば、数値時系列データが示すセンサ値に基づいて特徴量が抽出される場合、探索部 81 は、特徴量に対するセンサ値の寄与度を算出してもよい。例えば、ラベリング対象のデータのセンサ値と特徴量との関係が、センサ値の線形式で表現される場合、探索部 81 は、線形式に含まれるセンサ値の重みを寄与度とし、最も重みが大きいセンサ値を共通点として特定してもよい。

[0035] 出力部 82 は、探索された共通点に関する情報を出力する。出力部 82 は、クラスタごとに探索された共通点に関する情報を、入出力装置 90 に出力して表示させてもよく、ラベリング支援システム 1 が備える記憶部（図示せず）に出力して記憶させてもよい。

[0036] 具体的には、出力部 82 は、探索された共通点のうち、最も共通性の度合いが高い共通点を 1 つ出力してもよい。例えば、カテゴリ値が共通点として特定された場合、出力部 82 は、そのカテゴリ値の名称とカテゴリ値（例えば、「天気：晴れ」など）を出力してもよい。また、例えば、センサ値が共通点として特定された場合、出力部 82 は、センサ値と、そのセンサ値を得たセンサの名称を出力してもよい。

[0037] さらに、特徴量に対するセンサ値の寄与度が算出されている場合、出力部 82 は、寄与度が最も大きいセンサ値を共通点として、そのセンサ値とセンサ値の名称を出力してもよい。

[0038] また、出力部 82 は、クラスタ内で探索された複数の共通点の候補を、その共通点の共通性の度合いに応じて出力してもよい。出力部 82 は、例えば、共通性の度合いそのものを出力してもよく、共通性の度合いが上位の共通点を、予め定めた順位までランキング形式でラベリング候補として出力してもよい。

[0039] また、出力部 82 は、各クラスタ内の未分類データ（すなわち、ラベリング対象のデータ）に対して、探索された共通点を示す情報を直接ラベリングして出力してもよい。この場合、出力部 82 は、最も共通性の度合いが高い

共通点を示す情報をラベリングして出力すればよい。

[0040] さらに、出力部 82 は、ラベリング対象のデータをクラスタ化した様子を人間が観察できるように、ラベリング対象のデータを次元削減（低次元化）したものを、入出力装置 90 にグラフ描画することで可視化してもよい。出力部 82 は、例えば、UMAP（Uniform Manifold Approximation and Projection）などにより、2次元または3次元にラベリング対象のデータを次元削減し、次元削減されたデータを、分布図などのグラフとして可視化してもよい。その際、出力部 82 は、同一のクラスタに分類されたデータを、他のクラスタと異なる態様（例えば、色を変える、記号を変える、など）で表示してもよい。

[0041] 図 4 は、次元削減されたデータをグラフで可視化した例を示す説明図である。図 4 に例示するグラフは、UMAP により 2次元に次元削減したデータを、属するクラスタごとに態様（斜線、黒塗り等）を変えて表示した例を示す。また、図 4 に例示するように、出力部 82 は、クラスタに含まれるデータの範囲を特定できるように、その範囲を囲む表示をしてもよい。

[0042] さらに、グラフ描画の際、出力部 82 は、全てのデータを表示してもよいし、特定の条件を満たすデータのみ表示する又は表示しないと決定してもよい。出力部 82 は、例えば、特定の条件を満たすクラスタ（例えば、データ数が所定数よりも多いクラスタ、など）や、未分類のデータ（すなわち、ラベリングされていないデータ）を対象に、表示するか表示しないか判断してもよい。

[0043] また、特徴量に対するセンサ値の寄与度が算出されている場合、出力部 82 は、クラスタ内におけるセンサごとの寄与度をグラフ表示してもよい。図 5 は、センサごとの寄与度をグラフ表示した例を示す説明図である。図 5 に示す例では、温度、湿度、および、水位を示すセンサ値を用いて各クラスタの特徴量が算出され、その特徴量の算出に用いられる各センサ値の寄与度を棒グラフで表示した例である。例えば、クラスタ 2 の特徴量は、他のクラスタと比較し、水位を示すセンサ値の寄与度が高いことを示す。

- [0044] なお、センサごとの寄与度の表示は、図5に例示する棒グラフに限定されず、例えば、集合縦棒グラフや折れ線グラフ、3次元面グラフなどであってもよい。
- [0045] また、出力部82は、クラスタ内におけるセンサ値の分布を出力してもよい。図6は、クラスタ内におけるセンサ値の分布の例を示す説明図である。図6に示す例では、ラベリング対象のデータにセンサ値として、温度、湿度および水位が含まれ、図6に例示するように、センサ値ごとに分布を示すグラフを表示していることを示す。なお、図6に例示するグラフの縦軸方向が要素数を示し、横軸方向がセンサ値を示す。なお、クラスタ内におけるセンサ値の分布の表示は、図6に例示する分布図に限定されず、例えば、度数分布表やヒストグラムであってもよい。
- [0046] さらに、出力部82は、クラスタ内における統計量を出力してもよい。図7は、クラスタ内における統計量の例を示す説明図である。図7に例示する統計量は、クラスタ内のデータに含まれる各センサ値の平均や分散、最大値や最小値を、クラスタごとに出力した例を示す。なお、出力した統計量は例示であり、他にも、中央値や最頻値など、任意の統計量が出力されればよい。
- [0047] 入出力装置90は、出力部82による出力結果を表示する。また、入出力装置90は、表示した結果に対するユーザからの入力を受け付け、入力に応じた処理を実行する。例えば、ユーザからクラスタを指定する入力を受け付けた場合、入出力装置90は、指定されたクラスタに関する詳細な情報を表示してもよい。具体的には、入出力装置90は、指定されたクラスタについて、出力部82により生成された統計情報を表示してもよい。
- [0048] 入出力装置90は、タブレット端末などにより実現されてもよい。他にも、入出力装置90は、ディスプレイ装置とポインティングデバイスを有する装置等により実現されてもよい。
- [0049] 例えば、図4に例示するようなクラスタの範囲が表示されている場合、入出力装置90は、対象とするクラスタの指定の入力をユーザから受け付け、

受け付けたクラスタに関する情報（例えば、図5、図6、図7に例示する情報）を表示してもよい。

[0050] データ取得部10と、関連情報取得部20と、物体識別部30と、データ加工部40と、テキスト情報入力部50と、特徴抽出部60と、可視化処理部80（より具体的には、探索部81と、出力部82）とは、プログラム（ラベリング支援プログラム）に従って動作するコンピュータのプロセッサ（例えば、CPU（Central Processing Unit））によって実現される。

[0051] 例えば、プログラムは、ラベリング支援システム1の記憶部（図示せず）に記憶され、プロセッサは、そのプログラムを読み込み、プログラムに従って、データ取得部10、関連情報取得部20、物体識別部30、データ加工部40、テキスト情報入力部50、特徴抽出部60、および、可視化処理部80（より具体的には、探索部81と、出力部82）として動作してもよい。また、ラベリング支援システム1の機能がSaaS（Software as a Service）形式で提供されてもよい。

[0052] データ取得部10と、関連情報取得部20と、物体識別部30と、データ加工部40と、テキスト情報入力部50と、特徴抽出部60と、可視化処理部80（より具体的には、探索部81と、出力部82）とは、それぞれが専用のハードウェアで実現されていてもよい。また、各装置の各構成要素の一部又は全部は、汎用または専用の回路（circuitry）、プロセッサ等やこれらの組合せによって実現されもよい。これらは、単一のチップによって構成されてもよいし、バスを介して接続される複数のチップによって構成されてもよい。各装置の各構成要素の一部又は全部は、上述した回路等とプログラムとの組合せによって実現されてもよい。

[0053] また、ラベリング支援システム1の各構成要素の一部又は全部が複数の情報処理装置や回路等により実現される場合には、複数の情報処理装置や回路等は、集中配置されてもよいし、分散配置されてもよい。例えば、情報処理装置や回路等は、クライアントサーバシステム、クラウドコンピューティングシステム等、各々が通信ネットワークを介して接続される形態として実現

されてもよい。

[0054] 次に、本実施形態のラベリング支援システム 1 の動作を説明する。図 8 は、ラベリング支援システム 1 の動作例を示すフローチャートである。図 8 に例示する動作例は、データ取得部 10 が、直接、特徴抽出部 60 が用いる形式のデータを取得し、取得したデータを特徴抽出部 60 に入力した場合の動作例である。

[0055] 特徴抽出部 60 は、ラベリング対象のデータから複数のクラスタを生成する（ステップ S 51）。探索部 81 は、生成されたクラスタごとにデータの共通点を探索する（ステップ S 52）。そして、出力部 82 は、探索された共通点に関する情報をクラスタごとに出力する（ステップ S 53）。

[0056] 以上のように、本実施形態では、特徴抽出部 60 が、ラベリング対象のデータを教師なし学習により分類することで複数のクラスタを生成し、探索部 81 が、生成されたクラスタごとに、そのクラスタに含まれるデータの共通点を探索する。そして、出力部 82 が、クラスタごとに、探索された共通点に関する情報を出力する。そのような構成により、ラベル付けされていないデータが分類されたクラスタに対するラベリング作業を支援できる。

[0057] また、出力部 82 がラベリング対象のデータに自動的にラベリングを行ったり、ラベリング候補を出力したりすることで、人によるラベリングのコストを低減させ、また、人がそのラベルが付与される理由を把握できる。

[0058] 次に、本発明の概要を説明する。図 9 は、本発明によるラベリング支援システムの概要を示すブロック図である。本発明によるラベリング支援システム 190（例えば、ラベリング支援システム 1）は、ラベリング対象のデータを教師なし学習により分類することで複数のクラスタを生成する分類手段 191（例えば、特徴抽出部 60）と、生成されたクラスタごとに、そのクラスタに含まれるデータの共通点を探索する探索手段 192（例えば、特徴抽出部 60）と、クラスタごとに、探索された共通点に関する情報を出力する出力手段 193（例えば、出力部 82）とを備えている。

[0059] そのような構成により、ラベル付けされていないデータが分類されたクラ

スタに対するラベリング作業を支援できる。

[0060] また、分類手段１９１は、生成したクラスタに含まれる各データの特徴量を抽出し、探索手段１９２は、クラスタ内の各データについて抽出された特徴量の共通点を探索してもよい。

[0061] また、分類手段１９１は、ラベリング対象のデータに含まれるセンサ値に基づいて特徴量を抽出し、探索手段１９２は、特徴量に対するセンサ値の寄与度を算出し、出力手段１９３は、寄与度が最も大きいセンサ値を共通点として出力してもよい。

[0062] また、出力手段１９３は、クラスタ内におけるセンサごとの寄与度をグラフ表示してもよい。

[0063] また、出力手段１９３は、各クラスタ内のラベリング対象のデータに対して、探索された共通点を示す情報をラベリングして出力してもよい。

[0064] また、出力手段１９３は、クラスタ内で探索された複数の共通点を、その共通点の共通性の度合いに応じて出力してもよい。

[0065] また、出力手段１９３は、共通性の度合いが上位の共通点を、予め定めた順位までランキング形式でラベリング候補として出力してもよい。

[0066] 図１０は、少なくとも１つの実施形態に係るコンピュータの構成を示す概略ブロック図である。コンピュータ１０００は、プロセッサ１００１、主記憶装置１００２、補助記憶装置１００３、インタフェース１００４を備える。

[0067] 上述のラベリング支援システム１９０は、コンピュータ１０００に実装される。そして、上述した各処理部の動作は、プログラム（ラベリング支援プログラム）の形式で補助記憶装置１００３に記憶されている。プロセッサ１００１は、プログラムを補助記憶装置１００３から読み出して主記憶装置１００２に展開し、当該プログラムに従って上記処理を実行する。

[0068] なお、少なくとも１つの実施形態において、補助記憶装置１００３は、一時的でない有形の媒体の一例である。一時的でない有形の媒体の他の例としては、インタフェース１００４を介して接続される磁気ディスク、光磁気デ

ィスク、ＣＤ－ＲＯＭ（Compact Disc Read-only memory）、ＤＶＤ－ＲＯＭ（Read-only memory）、半導体メモリ等が挙げられる。また、このプログラムが通信回線によってコンピュータ１０００に配信される場合、配信を受けたコンピュータ１０００が当該プログラムを主記憶装置１００２に展開し、上記処理を実行してもよい。

[0069] また、当該プログラムは、前述した機能の一部を実現するためのものであっても良い。さらに、当該プログラムは、前述した機能を補助記憶装置１００３に既に記憶されている他のプログラムとの組み合わせで実現するもの、いわゆる差分ファイル（差分プログラム）であってもよい。

[0070] 上記の実施形態の一部又は全部は、以下の付記のようにも記載されうが、以下には限られない。

[0071] （付記１）ラベリング対象のデータを教師なし学習により分類することで複数のクラスタを生成する分類手段と、

生成された前記クラスタごとに、当該クラスタに含まれる前記データの共通点を探索する探索手段と、

前記クラスタごとに、探索された前記共通点に関する情報を出力する出力手段とを備えた

ことを特徴とするラベリング支援システム。

[0072] （付記２）分類手段は、生成したクラスタに含まれる各データの特徴量を抽出し、

探索手段は、クラスタ内の各データについて抽出された特徴量の共通点を探索する

付記１記載のラベリング支援システム。

[0073] （付記３）分類手段は、ラベリング対象のデータに含まれるセンサ値に基づいて特徴量を抽出し、

探索手段は、前記特徴量に対する前記センサ値の寄与度を算出し、

出力手段は、前記寄与度が最も大きいセンサ値を共通点として出力する

付記１または付記２記載のラベリング支援システム。

[0074] (付記4) 出力手段は、クラスタ内におけるセンサごとの寄与度をグラフ表示する

付記3記載のラベリング支援システム。

[0075] (付記5) 出力手段は、各クラスタ内のラベリング対象のデータに対して、探索された共通点を示す情報をラベリングして出力する

付記1または付記2記載のラベリング支援システム。

[0076] (付記6) 出力手段は、クラスタ内で探索された複数の共通点を、当該共通点の共通性の度合いに応じて出力する

付記1から付記5のうちのいずれか1つに記載のラベリング支援システム。

[0077] (付記7) 出力手段は、共通性の度合いが上位の共通点を、予め定めた順位までランキング形式でラベリング候補として出力する

付記6記載のラベリング支援システム。

[0078] (付記8) コンピュータが、ラベリング対象のデータを教師なし学習により分類することで複数のクラスタを生成し、

前記コンピュータが、生成された前記クラスタごとに、当該クラスタに含まれる前記データの共通点を探索し、

前記コンピュータが、前記クラスタごとに、探索された前記共通点に関する情報を出力する

ことを特徴とするラベリング支援方法。

[0079] (付記9) コンピュータが、生成したクラスタに含まれる各データの特徴量を抽出し、

前記コンピュータが、クラスタ内の各データについて抽出された特徴量の共通点を探索する

付記8記載のラベリング支援方法。

[0080] (付記10) コンピュータに、

ラベリング対象のデータを教師なし学習により分類することで複数のクラスタを生成する分類処理、

生成された前記クラスタごとに、当該クラスタに含まれる前記データの共通点を探索する探索処理、および、

前記クラスタごとに、探索された前記共通点に関する情報を出力する出力処理

を実行させるためのラベリング支援プログラムを記憶するプログラム記憶媒体。

[0081] (付記 1 1) コンピュータに、

分類処理で、生成したクラスタに含まれる各データの特徴量を抽出させ、
探索処理で、クラスタ内の各データについて抽出された特徴量の共通点を探索させる

ためのラベリング支援プログラムを記憶する付記 1 0 記載のプログラム記憶媒体。

[0082] (付記 1 2) コンピュータに、

ラベリング対象のデータを教師なし学習により分類することで複数のクラスタを生成する分類処理、

生成された前記クラスタごとに、当該クラスタに含まれる前記データの共通点を探索する探索処理、および、

前記クラスタごとに、探索された前記共通点に関する情報を出力する出力処理

を実行させるためのラベリング支援プログラム。

[0083] (付記 1 3) コンピュータに、

分類処理で、生成したクラスタに含まれる各データの特徴量を抽出させ、
探索処理で、クラスタ内の各データについて抽出された特徴量の共通点を探索させる

付記 1 2 記載のラベリング支援プログラム。

[0084] 以上、実施形態を参照して本願発明を説明したが、本願発明は上記実施形態に限定されるものではない。本願発明の構成や詳細には、本願発明のスコープ内で当業者が理解し得る様々な変更をすることができる。

符号の説明

- [0085] 1 ラベリング支援システム
- 1 0 データ取得部
- 2 0 関連情報取得部
- 3 0 物体識別部
- 4 0 データ加工部
- 5 0 テキスト情報入力部
- 6 0 特徴抽出部
- 7 0 特徴記憶部
- 8 0 可視化処理部
- 8 1 探索部
- 8 2 出力部
- 9 0 入出力装置

請求の範囲

- [請求項1] ラベリング対象のデータを教師なし学習により分類することで複数のクラスタを生成する分類手段と、
- 生成された前記クラスタごとに、当該クラスタに含まれる前記データの共通点を探索する探索手段と、
- 前記クラスタごとに、探索された前記共通点に関する情報を出力する出力手段とを備えた
- ことを特徴とするラベリング支援システム。
- [請求項2] 分類手段は、生成したクラスタに含まれる各データの特徴量を抽出し、
- 探索手段は、クラスタ内の各データについて抽出された特徴量の共通点を探索する
- 請求項1記載のラベリング支援システム。
- [請求項3] 分類手段は、ラベリング対象のデータに含まれるセンサ値に基づいて特徴量を抽出し、
- 探索手段は、前記特徴量に対する前記センサ値の寄与度を算出し、
- 出力手段は、前記寄与度が最も大きいセンサ値を共通点として出力する
- 請求項1または請求項2記載のラベリング支援システム。
- [請求項4] 出力手段は、クラスタ内におけるセンサごとの寄与度をグラフ表示する
- 請求項3記載のラベリング支援システム。
- [請求項5] 出力手段は、各クラスタ内のラベリング対象のデータに対して、探索された共通点を示す情報をラベリングして出力する
- 請求項1または請求項2記載のラベリング支援システム。
- [請求項6] 出力手段は、クラスタ内で探索された複数の共通点を、当該共通点の共通性の度合いに応じて出力する
- 請求項1から請求項5のうちのいずれか1項に記載のラベリング支

援システム。

[請求項7] 出力手段は、共通性の度合いが上位の共通点を、予め定めた順位までランキング形式でラベリング候補として出力する

請求項6記載のラベリング支援システム。

[請求項8] コンピュータが、ラベリング対象のデータを教師なし学習により分類することで複数のクラスタを生成し、

前記コンピュータが、生成された前記クラスタごとに、当該クラスタに含まれる前記データの共通点を探索し、

前記コンピュータが、前記クラスタごとに、探索された前記共通点に関する情報を出力する

ことを特徴とするラベリング支援方法。

[請求項9] コンピュータが、生成したクラスタに含まれる各データの特徴量を抽出し、

前記コンピュータが、クラスタ内の各データについて抽出された特徴量の共通点を探索する

請求項8記載のラベリング支援方法。

[請求項10] コンピュータに、

ラベリング対象のデータを教師なし学習により分類することで複数のクラスタを生成する分類処理、

生成された前記クラスタごとに、当該クラスタに含まれる前記データの共通点を探索する探索処理、および、

前記クラスタごとに、探索された前記共通点に関する情報を出力する出力処理

を実行させるためのラベリング支援プログラムを記憶するプログラム記憶媒体。

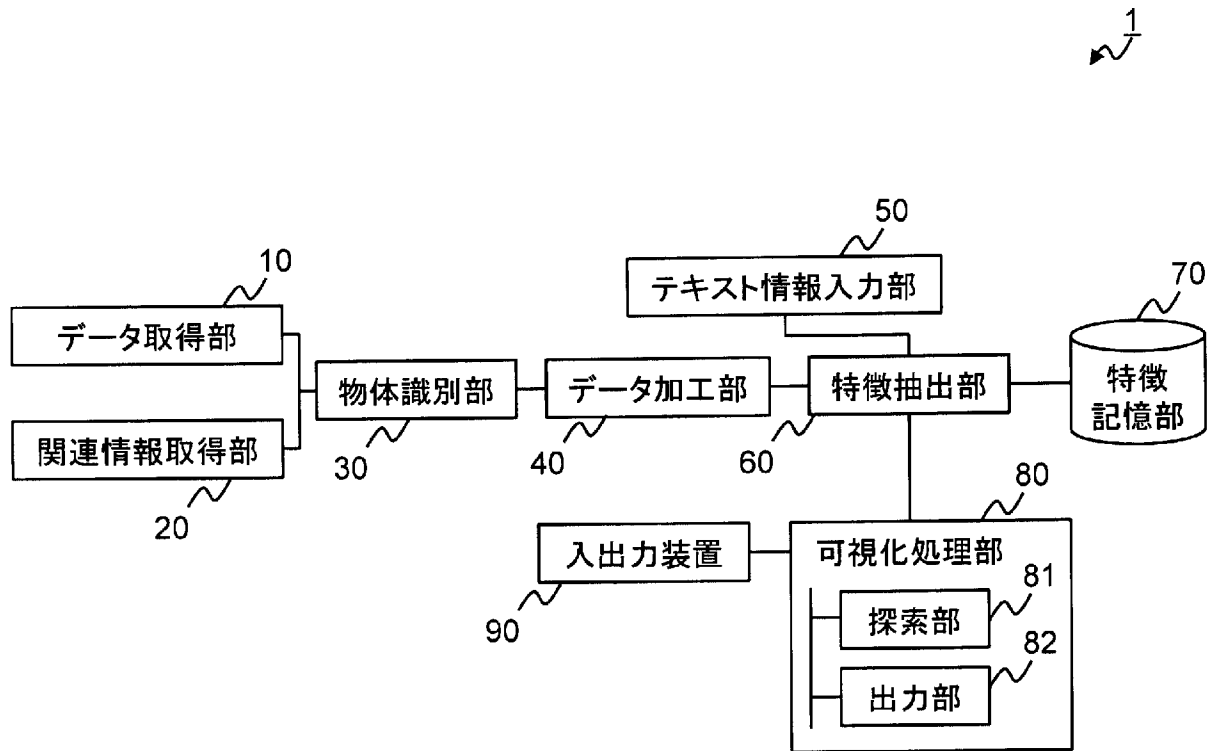
[請求項11] コンピュータに、

分類処理で、生成したクラスタに含まれる各データの特徴量を抽出させ、

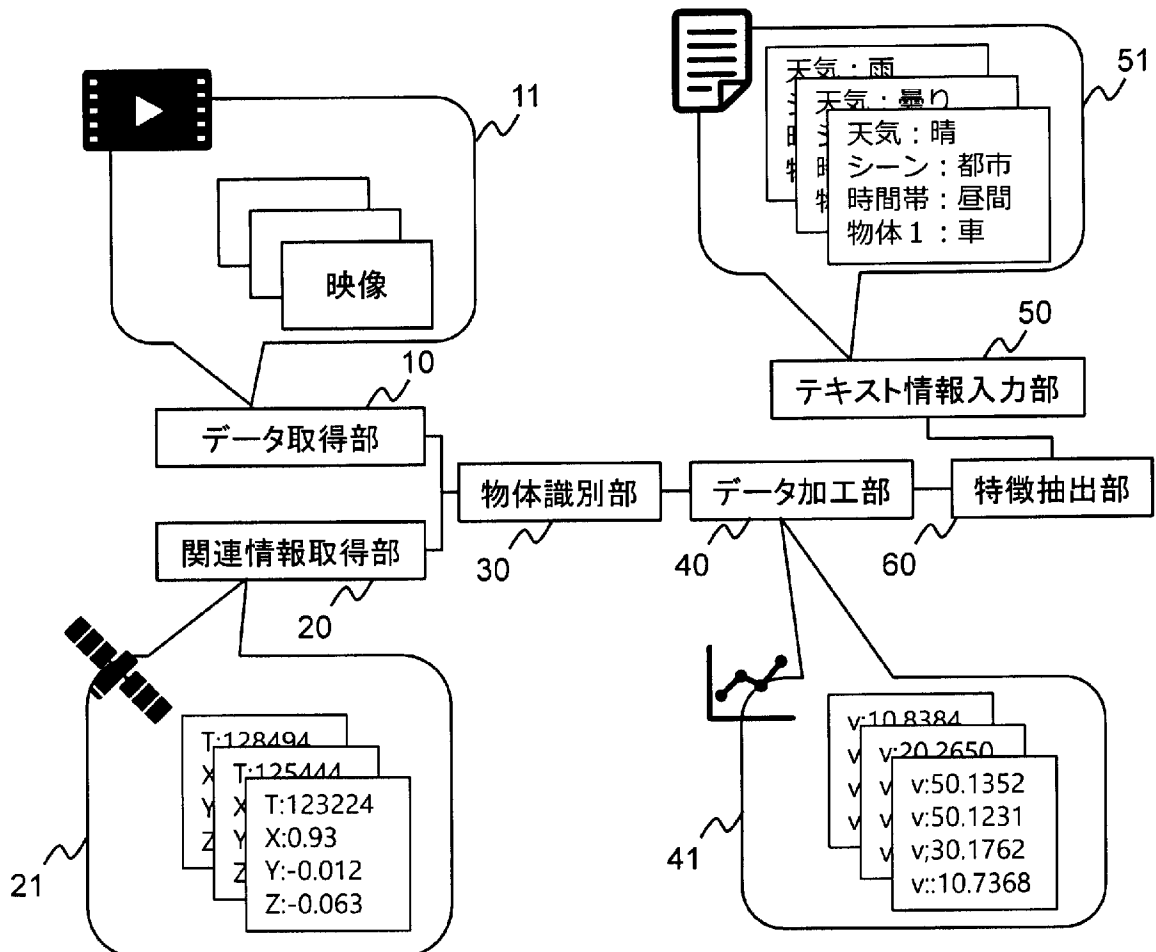
探索処理で、クラスタ内の各データについて抽出された特徴量の共通点を探索させる

ためのラベリング支援プログラムを記憶する請求項 10 記載のプログラム記憶媒体。

[図1]



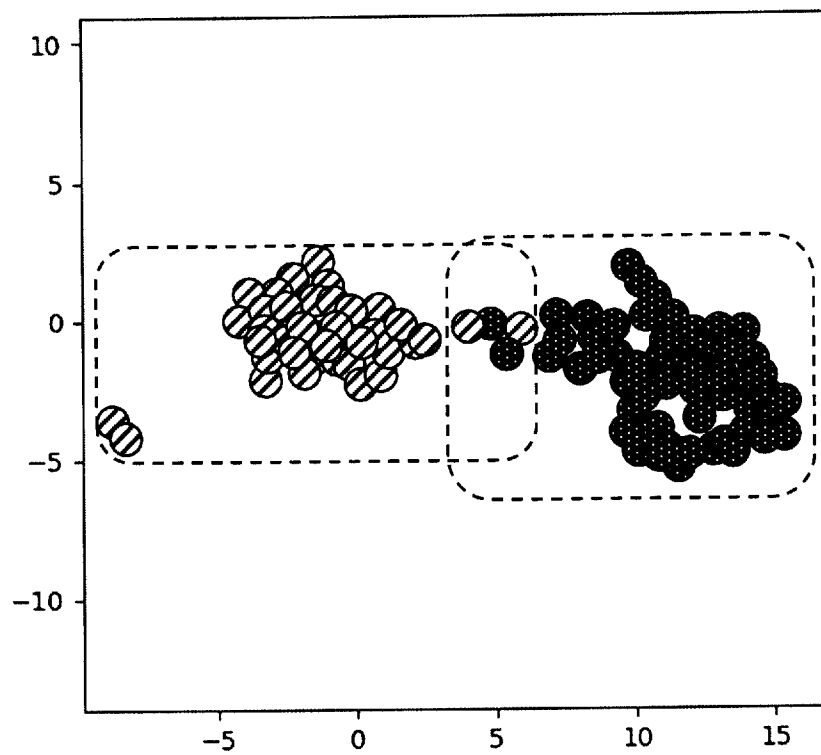
[図2]



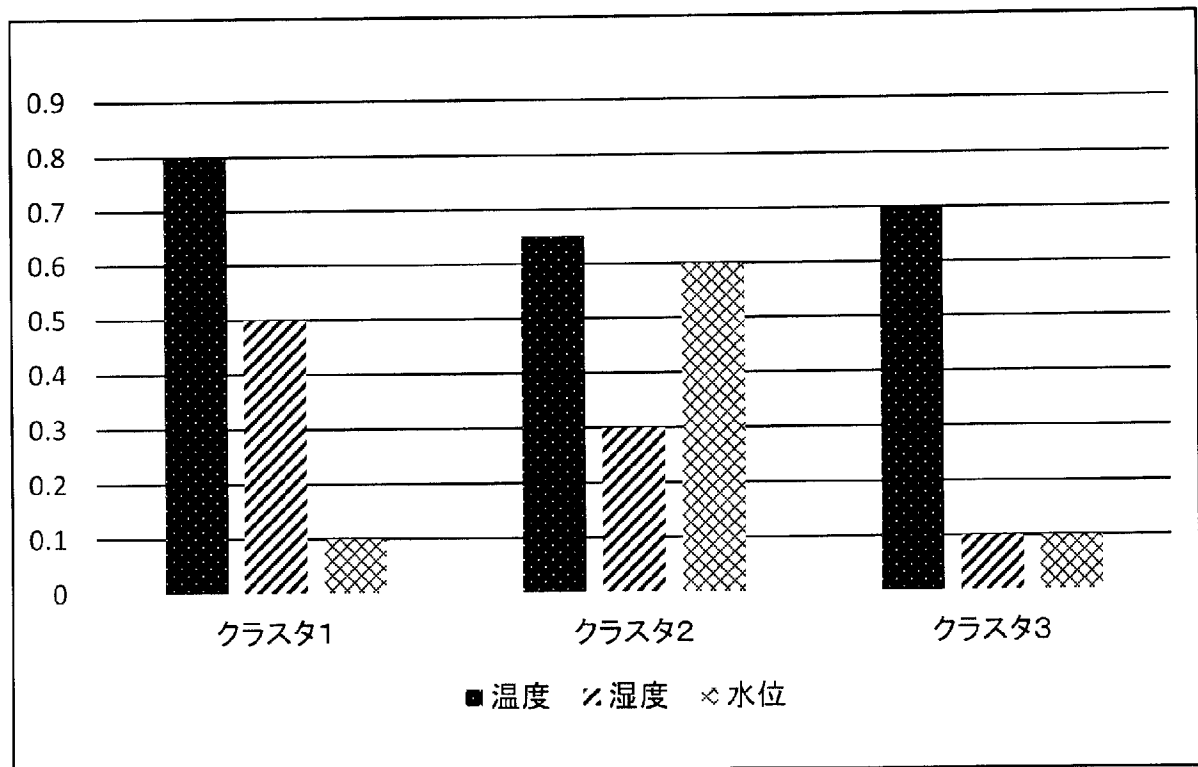
[図3]

-4.04E-01	7.74E-01	-3.25E-01	6.25E-01
-3.50E-01	5.55E-01	9.81E-01	-9.60E-02
-8.40E-01	7.51E-01	1.53E+00	-4.47E-01
-4.51E-01	6.42E-01	1.35E+00	-7.03E-01
-3.16E-01	6.34E-01	-3.01E+00	2.02E+00
-2.88E-01	6.32E-01	1.60E-01	-9.19E-02
1.34E+00	-8.01E-01	-8.85E-01	6.87E-01
-3.68E-01	6.26E-01	2.77E-01	1.18E-01

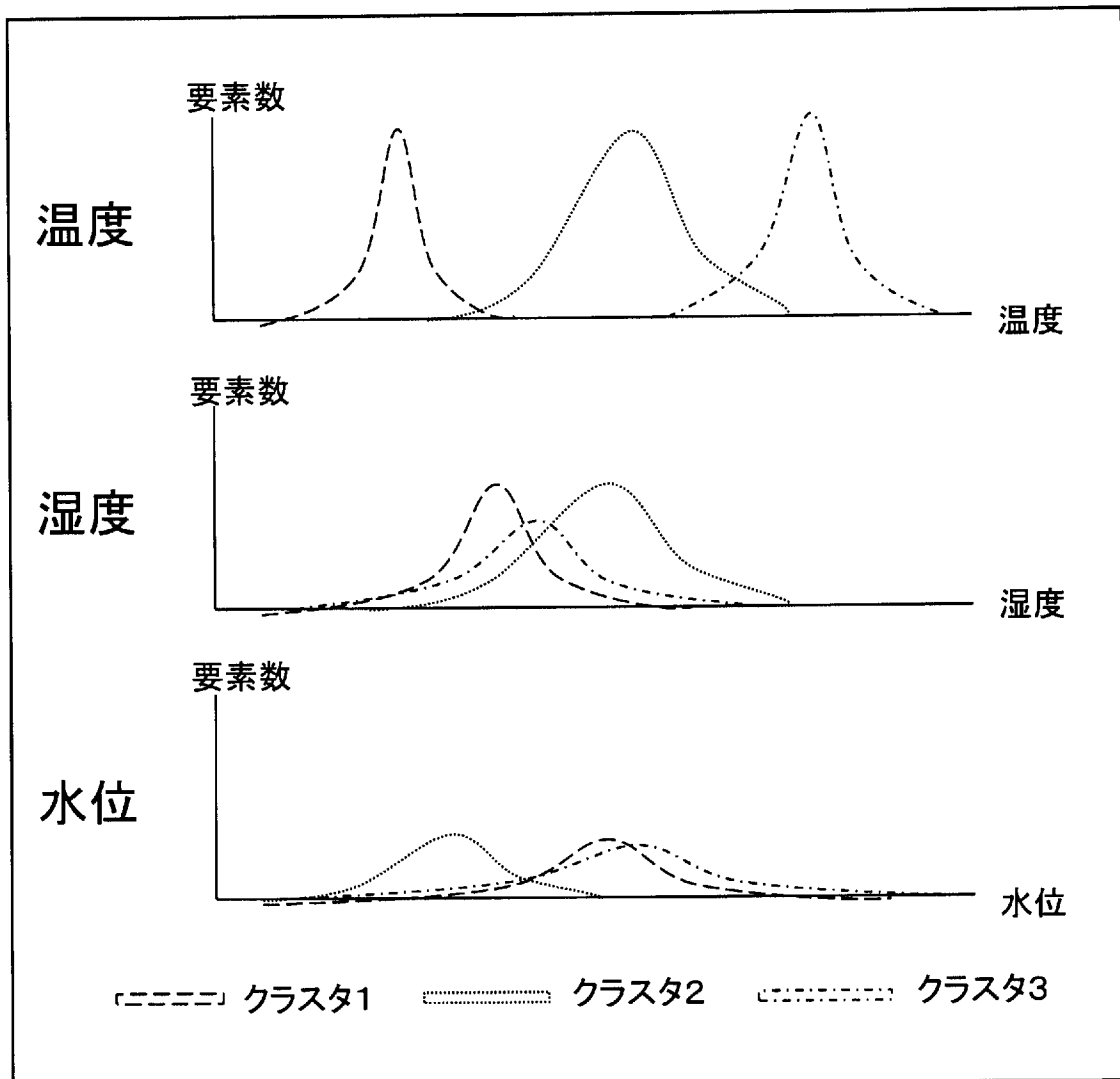
[図4]



[図5]



[図6]



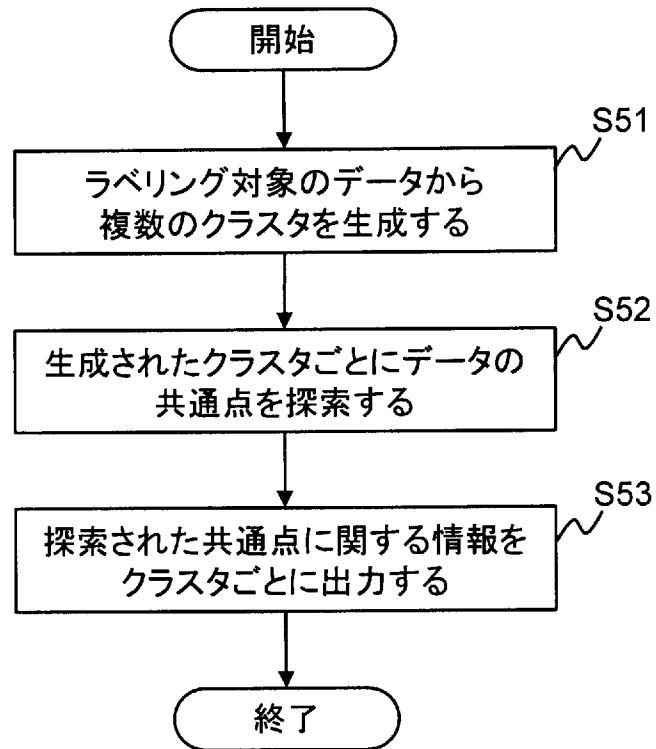
[図7]

クラス1	平均	分散	最大	最小
温度	15.5	1.5	17.8	13.1
湿度	48	2.3	68	31
水位	2.1	3.1	5.1	0.1

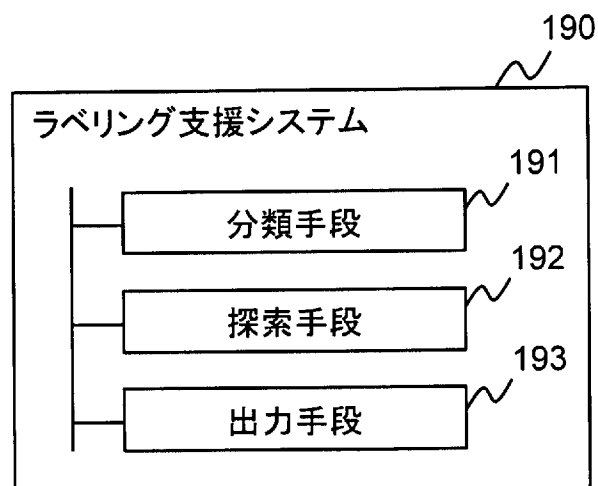
クラス2	平均	分散	最大	最小
温度	20.1	2.0	28.1	16.1
湿度	56	2.5	75	40
水位	1.4	2.7	2.9	0.1

クラス3	平均	分散	最大	最小
温度	28.2	1	33.1	24.8
湿度	52	4.8	73	33
水位	2.3	3.0	5.5	0.3

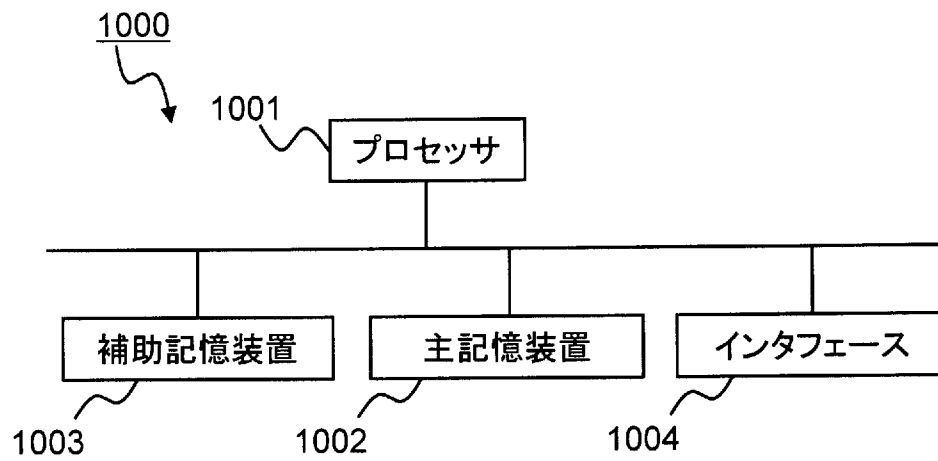
[図8]



[図9]



[図10]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2022/008750

A. CLASSIFICATION OF SUBJECT MATTER**G06F 16/75**(2019.01)i; **G16Y 40/20**(2020.01)i

FI: G06F16/75; G16Y40/20

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)

G06F16/75; G16Y40/20

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 1971-2022

Registered utility model specifications of Japan 1996-2022

Published registered utility model applications of Japan 1994-2022

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.
Y	JP 2008-84151 A (JUST SYST. CORP.) 10 April 2008 (2008-04-10) paragraph [0066]	1-11
Y	JP 2008-84203 A (NEC CORP.) 10 April 2008 (2008-04-10) paragraph [0003]	1-11
Y	服部傑外. 多数センサによる観測データの相関を手がかりとする人間行動判別. 電子情通学会技術研究報告. 23 November 2006, vol. 106, no. 396, pp. 29-34, (HATTORI, Takashi et al. Human Action Classification by Utilizing Correlation of Observation Data from Massive Sensors. IEICE Technical Report.) p. 34, left column to right column, fig. 6	3-4

☐ 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

28 April 2022

Date of mailing of the international search report

17 May 2022

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/JP2022/008750

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
JP	2008-84151	A	10 April 2008	(Family: none)	
JP	2008-84203	A	10 April 2008	(Family: none)	

A. 発明の属する分野の分類（国際特許分類（IPC）） G06F 16/75(2019.01)i; G16Y 40/20(2020.01)i FI: G06F16/75; G16Y40/20														
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）） G06F16/75; G16Y40/20														
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 - 2022年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996 - 2022年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994 - 2022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 - 2022年	日本国実用新案登録公報	1996 - 2022年	日本国登録実用新案公報	1994 - 2022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 - 2022年													
日本国実用新案登録公報	1996 - 2022年													
日本国登録実用新案公報	1994 - 2022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												
Y	JP 2008-84151 A (株式会社ジャストシステム) 10.04.2008 (2008 - 04 - 10) [0066]	1-11												
Y	JP 2008-84203 A (日本電気株式会社) 10.04.2008 (2008 - 04 - 10) [0003]	1-11												
Y	服部 傑 外, 多数センサによる観測データの相関を手がかりとする人間行動判別, 電子情報通信学会技術研究報告, 2006.11.23, Vol. 106 No. 396, pp. 29--34 p. 34 左欄～右欄, 図6	3-4												
☐ 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>* 参考文献のカテゴリー</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>			* 参考文献のカテゴリー	“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	
* 参考文献のカテゴリー	“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 28.04.2022	国際調査報告の発送日 17.05.2022													
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） 原 秀人 5N 9644 電話番号 03-3581-1101 内線 3586													

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/JP2022/008750

引用文献			公表日	パテントファミリー文献	公表日
JP	2008-84151	A	10.04.2008	(ファミリーなし)	
JP	2008-84203	A	10.04.2008	(ファミリーなし)	