

(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)



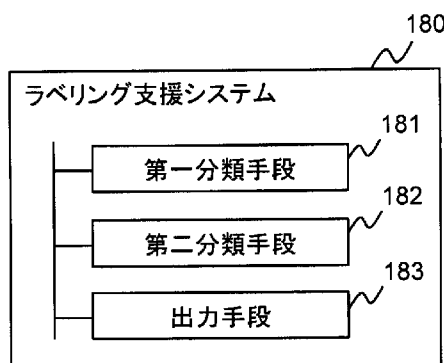
(10) 国際公開番号

WO 2023/166578 A1

- (51) 国際特許分類:
G06F 16/75 (2019.01)
- (21) 国際出願番号: PCT/JP2022/008749
- (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
- (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,

(54) Title: LABELING ASSISTANCE SYSTEM, LABELING ASSISTANCE METHOD, AND LABELING 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



- 180 Labeling assistance system
181 First classification means
182 Second classification means
183 Output means

(57) Abstract: A first classification means 181 generates a first plurality of clusters by classifying a first group of data, which is a group of data to be labeled, by unsupervised learning. A second classification means 182 generates a second plurality of clusters by classifying a second group of data, which is a group of data including at least a part of the group of data to be labeled. An output means 183 outputs data that is included in the second plurality of clusters and is classified into a different cluster for the first plurality of clusters.

(57) 要約: 第一分類手段 181 は、ラベリング対象のデータ群である第一のデータ群を教師なし学習により分類することで第一の複数のクラスタを生成する。第二分類手段 182 は、ラベリング対象のデータの少なくとも一部のデータを含むデータ群である第二のデータ群进行分类することで第二の複数のクラスタを生成する。出力手段 183 は、第二の複数のクラスタに含まれるデータのうち、第一の複数のクラスタでは異なるクラスタに分類されていたデータを出力する。

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, 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。一方、例えば、大量のデータの中から、目的とする映像を単純作業で見つけようとするのは、非常に困難である。そこで、収集されたデータを検索する仕組みが求められている。

[0003] データを検索するための仕組みとして、そのデータに対するラベリングを行う方法が挙げられる。ただし、大量のデータに対するラベリングを人手で行うには膨大な時間およびコストがかかってしまうため、データを分類するための方法が各種提案されている。

[0004] 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

[0005] 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

- [0006] 例えば、ルールベースで自動的にラベリングを行うことも考えられる。しかし、環境等の変化に応じてルールをメンテナンスする作業は煩雑であり、また、ルールの追加等の作業も容易ではない。
- [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、分割区間が予め定められる。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群である第一のデータ群を教師なし学習により分類することで第一の複数のクラスタを生成する第一分類手段と、ラベリング対象のデータの少なくとも一部のデータを含むデータ群である第二のデータ群を分類することで第二の複数のクラスタを生成する第二分類手段と、第二の複数のクラスタに含まれるデータのうち、第一の複数のクラスタでは異なるクラスタに分類されていたデータ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群である第一のデータ群を教師なし学習により分類することで第一の複数のクラスタを生成し、コンピュータが、ラベリング対象のデータの少なくとも一部のデータを含むデータ群である第二のデータ群を分類することで第二の複数のクラスタを生成し、コンピュータが、第二の複数のクラスタに含まれるデータのうち、第一の複数のクラスタでは異なるクラスタに分類されていたデータ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

対象のデータ群である第一のデータ群を教師なし学習により分類することで第一の複数のクラスタを生成する第一分類処理、ラベリング対象のデータの少なくとも一部のデータを含むデータ群である第二のデータ群を分類することで第二の複数のクラスタを生成する第二分類処理、および、第二の複数のクラスタに含まれるデータのうち、第一の複数のクラスタでは異なるクラスタに分類されていたデータ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]次元削減されたデータをグラフで可視化した他の例を示す説明図である。

[図6]クラスタ内のデータにラベル付けを行う処理の例を示す説明図である。

[図7]一部のクラスタを選択する処理の例を示す説明図である。

[図8]データの一部を除外する処理の例を示す説明図である。

[図9]精緻化前後の結果をオーバーレイ表示した例を示す説明図である。

[図10]精緻化前後の結果を並列窓で表示した例を示す説明図である。

[図11]精緻化前後の結果を並列窓で表示した例を示す説明図である。

[図12]精緻化前後で異なる結果になったデータを別窓でリスト表示した例を示す説明図である。

[図13]複数回の精緻化結果をオーバーレイ表示した例を示す説明図である。

[図14]複数回の精緻化により異なる結果になったデータを別窓でリスト表示

した例を示す説明図である。

[図15]各クラスタの統計情報を表示した例を示す説明図である。

[図16]各クラスタの統計情報を表示した他の例を示す説明図である。

[図17]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18]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19]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、データ精緻化部100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 inputs 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 inputs 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、inputs 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 inputs 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、まず初めに、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] 以下、特徴抽出部60が、ラベリング対象のデータ群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処理を、第一分類処理と記す。また、第一分類処理により生成される複数のクラスタを、第一の複数のクラスタと記し、第一の複数のクラスタに分類されるデータ群のことを、第一のデータ群と記す。また、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処理を行うことから、特徴抽出部60のことを分類手段と言うこともできる。

[0027] 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。

[0028] 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。

[0029] 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。

[0030] 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。また、特徴記憶部70は、後述するデータ精緻化部100によって付加されたラベルの情報を併せて記憶してもよい。なお、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。

[0031] 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。

[0032] 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。本実施形態の可視化処理部80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する。

[0033] 可視化処理部80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、可視化処理部80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。

[0034] 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。

[0035] 図5は、次元削減されたデータをグラフで可視化した他の例を示す説明図である。図5に例示するグラフは、映像データの種類ごとにプロットされる記号を変化させて表示したグラフである。また、図5に例示するように、可視化処理部80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、その範囲を点線で囲む表示をしてもよい。

[0036] さらに、グラフ描画の際、可視化処理部80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。可視化処理部80は、例えば、特定の条件を満たすクラスタ（例

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。

[0037] さらに、本実施形態の可視化処理部 80 は、後述する再学習処理の結果、異なるクラスタに属することとなったデータを出力する。なお、データの出力方法については後述される。

[0038] 入出力装置 90 は、可視化処理部 80 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。本実施形態では、入出力装置 90 の出力に対してユーザが指定したクラスタの入力に基づいて、後述するデータ精緻化部 100 の処理が行われる。

[0039] 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。

[0040] データ精緻化部 100 は、特徴抽出部 60 により生成されたクラスタに基づいて、ラベリング対象のデータ群に対する各処理を実行する。具体的には、データ精緻化部 100 は、ラベリング対象のデータ群の中から、生成された第一の複数のクラスタに応じて、第二のデータ群を生成する。本実施形態では、データ精緻化部 100 は、以下の 3 種類の処理を実行する場合について説明する。

[0041] まず、第一の処理について説明する。第一の処理は、クラスタ内のデータにラベル付けを行う処理である。第一の処理では、データ精緻化部 100 は、ラベリング対象のデータ群のうち、第一の複数のクラスタのいずれかに分類されたデータに対してクラスタごとのラベリングを行った第二のデータ群を生成する。データ精緻化部 100 がラベリングを行う対象とするクラスタは任意である。データ精緻化部 100 は、すべてのクラスタに対してラベリングを行ってもよく、入出力装置 90 を介して、ユーザに指定されたクラスタに対してラベリングを行ってもよい。

- [0042] また、クラスタ内のデータに同一のラベルが付加されるのであれば、そのラベルの内容は任意である。データ精緻化部100は、対象とするクラスタ内のデータに対し、任意の仮ラベルを付加してもよく、ユーザにより指定された内容のラベルを付加してもよい。そして、データ精緻化部100は、データ（より詳しくは、データの特徴量）と付加されたラベルとを対応付けて特徴記憶部70に記憶してもよい。
- [0043] 図6は、クラスタ内のデータにラベル付けを行う処理の例を示す説明図である。図6に示す例では、データ精緻化部100が、図5に例示するクラスタに対し、それぞれ仮ラベル「A」、「B」および「C」を付加したことを示す。なお、図5に例示するクラスタのうち、付加する対象のクラスタがユーザにより指定された場合、データ精緻化部100は、指定されたクラスタにのみ仮ラベルを付加すればよい。
- [0044] その後、特徴抽出部60は、ラベルが付加されたデータを用いた学習（教師あり学習）により、複数のクラスタを再度生成する。なお、特徴抽出部60は、ラベルが付加されていないデータを加えて学習（教師なし学習）を行ってもよい。以下、特徴抽出部60が、ラベリング対象のデータの少なくとも一部のデータを含むデータ群を分類することで複数のクラスタを生成する処理を、第二分類処理と記す。また、第二分類処理により生成される複数のクラスタを、第二の複数のクラスタと記し、第二の複数のクラスタに分類されるデータ群のことを、第二のデータ群と記す。
- [0045] このように、第二分類処理では、第一分類処理で用いたラベリング対象のデータの少なくとも一部のデータを用いて、再度複数のクラスタを生成して精緻化することから、第二分類処理のことを再学習処理または精緻化と言うことができる。これにより、教師なし学習を通じてラベル付けを半自動化でき、また、新規ラベルの発見にも寄与できる。
- [0046] 特徴抽出部60は、第二分類処理により生成されたクラスタ（第二の複数のクラスタ）に含まれる各データの特徴量を抽出し、抽出した特徴量を特徴記憶部70に記憶してもよい。

[0047] そして、第二分類処理の後、可視化処理部 80 は、第二の複数のクラスタに含まれるデータのうち、第一の複数のクラスタでは異なるクラスタに分類されていたデータを出力する。これは、再学習の結果、異なるクラスタに属することになったデータを可視化する処理に対応する。なお、可視化する具体的処理については、後述される。

[0048] 次に、第二の処理について説明する。第二の処理は、少なくとも一部のクラスタを選択して、再び学習（教師なし学習）をする処理である。データ精緻化部 100 は、ラベリング対象のデータ群のうち、第一の複数のクラスタの中から選択されたクラスタに分類されているデータ群を第二のデータ群として生成する。

[0049] まず、データ精緻化部 100 は、第一の複数のクラスタの中から、少なくとも一部のクラスタを選択する。データ精緻化部 100 は、入出力装置 90 を介して、ユーザに指定されたクラスタを選択してもよく、条件を満たすクラスタを自動で選択してもよい。ここでの条件は任意であり、例えば、データ数が予め定めた数以上のクラスタ、分類されたデータの割合が予め定めた閾値よりも大きい、などが挙げられる。ここで選択されたクラスタ内のデータ群が、上述する第二のデータ群に対応する。

[0050] 図 7 は、一部のクラスタを選択する処理の例を示す説明図である。図 7 に示す例では、生成された 3 つのクラスタのうち、2 つのクラスタが選択されたことを示す。なお、第二の処理においても、第一分類処理で分類されたクラスタを識別できるように、データ精緻化部 100 は、任意のクラスタ識別情報を各クラスタ内のデータに付与しておけばよい。

[0051] その後、特徴抽出部 60 は、選択されたクラスタ内のデータを対象とした学習（教師なし学習）により、複数のクラスタを再度生成する（すなわち、再学習処理を行う）。この処理が、上述する第二分類処理に対応し、生成された複数のクラスタが、第二の複数のクラスタに対応する。なお、特徴抽出部 60 は、新たなデータを別途加えて学習を行ってもよい。これにより、クラスタ内のデータを深掘りすることができるため、より詳細にデータを分類

することが期待できる。

[0052] そして、第二分類処理の後、可視化処理部 80 は、上記第一の処理と同様、第二の複数のクラスタに含まれるデータのうち、第一の複数のクラスタでは異なるクラスタに分類されていたデータを出力する。なお、選択されたクラスタが細分化される可能性があることから、可視化処理部 80 は、クラスタ内のデータのうち、クラスタ識別情報が少数派（最大の割合以外）のデータを、第一の複数のクラスタでは異なるクラスタに分類されていたデータとして出力してもよい。

[0053] 次に、第三の処理について説明する。第三の処理は、外れ値など、クラスタに分類されなかったデータの少なくとも一部を除外して、再び学習（教師なし学習または教師あり学習）をする処理である。データ精緻化部 100 は、ラベリング対象のデータ群のうち、第一の複数のクラスタのいずれにも分類されなかった一以上のデータを除外したデータ群を第二のデータ群として生成する

[0054] 図 8 は、データの一部を除外する処理の例を示す説明図である。図 8 に示す例では、実線の丸で囲まれた範囲のデータが外れ値として除外されることを示す。例えば、ラベリング対象のデータが映像データの場合、ノイズシーンを除外する処理に対応する。以降、上述する第一の処理と第二の処理の少なくとも一方、または、両方の処理が行われる。これにより、分類精度の向上が期待される。

[0055] 以上、データ精緻化部 100 が行う 3 種類の処理について説明した。ただし、データ精緻化部 100 が実行する処理は、上述する 3 種類の処理に限定されない。データ精緻化部 100 は、他にも、データのメンテナンス処理を行ってもよい。また、第一の処理、第二の処理、および、第三の処理の各処理の後で、再び同一の処理が行われてもよく、異なる処理が行われてもよい。

[0056] データをメンテナンスする処理の一例が、特徴抽出部 60 が学習に用いるためのデータをメンテナンスする処理である。データ精緻化部 100 は、ラ

ベルが付加されたデータ群や、外れ値が除外されたデータ群を含むファイル
を出力してもよい。

[0057] 例えば、上述する第一の処理で、ラベリング対象のデータ群に対してラベル付けが行われたとする。この場合、データ精緻化部100は、指定したラベルを記載したラベルファイルの作成、次回学習用フォルダにラベルが付与されたデータのためのコピー、ラベルに基づき、元データをラベルごとのフォルダに振り分ける処理（移動・コピー）などを行ってもよい。

[0058] また、例えば、上述する第二の処理で、クラスタが選別されたとする。この場合、データ精緻化部100は、選択したクラスタに属するデータのみを記載したデータリストファイルの作成、選択したクラスタに属するデータのみを次回学習用フォルダにコピーする処理などを行ってもよい。

[0059] また、例えば、上述する第三の処理で、外れ値を除外する処理が行われたとする。この場合、データ精緻化部100は、指定されたデータ（外れ値）以外のデータのみを記載したデータリストファイルの作成、指定されたデータ（外れ値）以外のデータを次回学習用フォルダにコピーする処理などを行ってもよい。

[0060] 以下、再学習の結果、異なるクラスタに属することになったデータを可視化処理部80が可視化する方法について具体的に説明する。まず、可視化処理部80は、ラベリング対象のデータ群を次元削減し、第一の複数のクラスタに含まれる次元削減されたデータ、および、第二の複数のクラスタに含まれる次元削減されたデータをクラスタごとに識別できる態様でグラフ描画する。そして、可視化処理部80は、第二の複数のクラスタに含まれる次元削減されたデータのうち、第一の複数のクラスタでは異なるクラスタに分類されていたデータを、他のデータと異なる態様で表示する。

[0061] 異なる態様の例として、例えば、色の濃淡を変化させたり、色そのものを変化させたり、外枠の線を変化させたり、点滅表示したりすることが挙げられる。

[0062] 図9は、精緻化前後の結果をオーバーレイ表示した例を示す説明図である。

図9に示す例では、可視化処理部80が、各精緻化のデータの分布を重ねて表示するとともに、注目するレイヤ（すなわち、精緻化）以外のデータを、注目するレイヤのデータとは異なる態様で表示していることを示す。具体的には、図9に示す例では、1回目の精緻化の結果と、2回目の精緻化の結果とを重ねて表示している。その際、1回目の精緻化の結果に注目している場合には、2回目の精緻化でのみ対象のクラスタに含まれているデータd1を、他のデータとは異なる態様で示している。同様に、2回目の精緻化の結果に注目している場合には、1回目の精緻化でのみ対象のクラスタに含まれているデータd2を、他のデータとは異なる態様で示している。

[0063] 図10および図11は、精緻化前後の結果を並列窓で表示した例を示す説明図である。図10に例示するように、可視化処理部80は、精緻化前後の結果を別々の窓で表示してもよい。その際、可視化処理部80は、図11に例示するように、精緻化の前後で変化したデータの態様を他のデータと異なる態様で表示してもよい。

[0064] さらに、可視化処理部80は、精緻化前後で異なる結果になったデータ（すなわち、異なるクラスタに分類されたデータ）を、リスト表示してもよい。図12は、精緻化前後で異なる結果になったデータd3を別窓でリスト表示した例を示す説明図である。図12に示す例では、精緻化前後で異なる結果になったデータが表示されている座標をリスト表示して結果を示す。

[0065] なお、図9から図12では、2つの精緻化結果を比較する場合を例示した。ただし、比較対象は、2つの結果に限定されず、3つ以上であってもよい。図13は、複数回の精緻化結果をオーバーレイ表示した例を示す説明図である。また、図14は、複数回の精緻化により異なる結果になったデータを別窓でリスト表示した例を示す説明図である。図13に示す例は、図9に示す例と比較し、精緻化結果が4つ存在する場合の例を示す。図14に示す例も同様に、図12に示す例と比較し、精緻化結果が4つ存在する場合の例を示す。

[0066] また、可視化処理部80は、上述するグラフとは別に、または、上述する

グラフと共に、データ群の分類処理（すなわち、精緻化）ごとにクラスタの統計情報を表示してもよい。なお、統計情報の作成は、可視化処理部 80 が行ってもよく、特徴抽出部 60 が行ってもよい。

[0067] 図 15 は、各クラスタの統計情報を表示した例を示す説明図である。図 15 に示す例では、クラスタの統計情報として、クラスタ内のデータ数、データの重心および分散（ x 方向および y 方向）を表示した例を示す。また、図 15 に例示するように、可視化処理部 80 は、精緻化ごとの統計情報を切替えて表示するようにしてもよく、並べて表示するようにしてもよい。

[0068] 図 16 は、各クラスタの統計情報を表示した他の例を示す説明図である。図 16 に例示するように、可視化処理部 80 は、クラスタの統計情報（例えば、誤検知率）をグラフおよび表形式で表示してもよい。図 16 に示す例では、教師あり学習を実施したときに、ラベルと振り分けられたクラスタとの一致度を表わす。なお、図 16 に示す例において、1 回目は教師なし学習を想定しており、評価結果は存在しない。

[0069] 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 と、データ精緻化部 100 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。

[0070] 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、可視化処理部 80、および、データ精緻化部 100 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。

[0071] 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 と、データ精緻化部 100 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。

[0072] 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。

[0073] 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 17 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 17 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。

[0074] 特徴抽出部 60 は、ラベリング対象のデータ群（第一のデータ群）から、第一の複数のクラスタを生成する（ステップ S11）。その後、特徴抽出部 60 は、ラベリング対象のデータの少なくとも一部のデータを含むデータ群（第二のデータ群）から第二の複数のクラスタを生成する（ステップ S12）。そして、可視化処理部 80 は、第二の複数のクラスタに含まれるデータのうち、第一の複数のクラスタでは異なるクラスタに分類されていたデータを出力する（ステップ S13）。

[0075] 以上のように、本実施形態では、特徴抽出部 60 が、第一のデータ群を教師なし学習により分類することで第一の複数のクラスタを生成する。また、特徴抽出部 60 が、第二のデータ群を分類することで第二の複数のクラスタ

を生成する。そして、可視化処理部 80 が、第二の複数のクラスタに含まれるデータのうち、第一の複数のクラスタでは異なるクラスタに分類されていたデータを出力する。よって、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。

[0076] また、本実施形態では、データ精緻化部 100 が、ラベリング対象のデータ群の中から、生成された第一の複数のクラスタに応じて、第二のデータ群を生成する。そのため、生成された第二のデータ群を用いた再学習の精度を向上させることが可能になる。

[0077] 次に、本発明の概要を説明する。図 18 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 180（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群である第一のデータ群を教師なし学習により分類することで第一の複数のクラスタを生成する第一分類手段 181（例えば、特徴抽出部 60）と、ラベリング対象のデータの少なくとも一部のデータを含むデータ群である第二のデータ群を分類（すなわち、再学習）することで第二の複数のクラスタを生成する第二分類手段 182（例えば、特徴抽出部 60）と、第二の複数のクラスタに含まれるデータのうち、第一の複数のクラスタでは異なるクラスタに分類されていたデータを出力する出力手段 183（例えば、可視化処理部 80）とを備えている。

[0078] 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。

[0079] また、ラベリング支援システム 180 は、ラベリング対象のデータ群の中から、生成された第一の複数のクラスタに応じて、第二のデータ群を生成するデータ精緻化手段（例えば、データ精緻化部 100）を備えていてもよい。

[0080] 具体的には、データ精緻化手段は、ラベリング対象のデータ群のうち、第一の複数のクラスタのいずれかに分類されたデータに対してクラスタごとのラベリングを行った第二のデータ群を生成してもよい（例えば、上記データ

精緻化部１００による第一の処理）。

[0081] また、データ精緻化手段は、ラベリング対象のデータ群のうち、第一の複数のクラスタの中から選択されたクラスタに分類されているデータ群を第二のデータ群として生成してもよい（例えば、上記データ精緻化部１００による第二の処理）。

[0082] また、データ精緻化手段は、ラベリング対象のデータ群のうち、第一の複数のクラスタのいずれにも分類されなかった一以上のデータを除外したデータ群を第二のデータ群として生成してもよい（例えば、上記データ精緻化部１００による第三の処理）。

[0083] また、出力手段は、ラベリング対象のデータ群を次元削減し、第一の複数のクラスタに含まれる次元削減されたデータ、および、第二の複数のクラスタに含まれる次元削減されたデータをクラスタごとに識別できる態様でグラフ描画し、第二の複数のクラスタに含まれる次元削減されたデータのうち、第一の複数のクラスタでは異なるクラスタに分類されていたデータを、他のデータと異なる態様で表示してもよい。

[0084] また、出力手段は、データ群の分類処理ごとにクラスタの統計情報を表示してもよい。

[0085] 図１９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。

[0086] 上述のラベリング支援システム１８０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。

[0087] 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。

[0088] 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。

[0089] 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るが、以下には限られない。

[0090] （付記 1）ラベリング対象のデータ群である第一のデータ群を教師なし学習により分類することで第一の複数のクラスタを生成する第一分類手段と、

前記ラベリング対象のデータの少なくとも一部のデータを含むデータ群である第二のデータ群を分類することで第二の複数のクラスタを生成する第二分類手段と、

前記第二の複数のクラスタに含まれるデータのうち、前記第一の複数のクラスタでは異なるクラスタに分類されていたデータ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。

[0091] （付記 2）ラベリング対象のデータ群の中から、生成された第一の複数のクラスタに応じて、第二のデータ群を生成するデータ精緻化手段を備えた

付記 1 記載のラベリング支援システム。

[0092] （付記 3）データ精緻化手段は、ラベリング対象のデータ群のうち、第一の複数のクラスタのいずれかに分類されたデータに対してクラスタごとのラベリングを行った第二のデータ群を生成する

付記 1 または付記 2 記載のラベリング支援システム。

[0093] (付記4) データ精緻化手段は、ラベリング対象のデータ群のうち、第一の複数のクラスタの中から選択されたクラスタに分類されているデータ群を第二のデータ群として生成する

付記1または付記2記載のラベリング支援システム。

[0094] (付記5) データ精緻化手段は、ラベリング対象のデータ群のうち、第一の複数のクラスタのいずれにも分類されなかった一以上のデータを除外したデータ群を第二のデータ群として生成する

付記1から付記4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。

[0095] (付記6) 出力手段は、ラベリング対象のデータ群を次元削減し、第一の複数のクラスタに含まれる次元削減されたデータ、および、第二の複数のクラスタに含まれる次元削減されたデータをクラスタごとに識別できる態様でグラフ描画し、前記第二の複数のクラスタに含まれる次元削減されたデータのうち、第一の複数のクラスタでは異なるクラスタに分類されていたデータを、他のデータと異なる態様で表示する

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。

[0096] (付記7) 出力手段は、データ群の分類処理ごとにクラスタの統計情報を表示する

付記1から付記6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。

[0097] (付記8) コンピュータが、ラベリング対象のデータ群である第一のデータ群を教師なし学習により分類することで第一の複数のクラスタを生成し、

前記コンピュータが、前記ラベリング対象のデータの少なくとも一部のデータを含むデータ群である第二のデータ群を分類することで第二の複数のクラスタを生成し、

前記コンピュータが、前記第二の複数のクラスタに含まれるデータのうち、前記第一の複数のクラスタでは異なるクラスタに分類されていたデータを

出力する

ことを特徴とするラベリング支援方法。

[0098] (付記 9) ラベリング対象のデータ群の中から、生成された第一の複数のクラスタに応じて、第二のデータ群を生成する

付記 8 記載のラベリング支援方法。

[0099] (付記 10) コンピュータに、

ラベリング対象のデータ群である第一のデータ群を教師なし学習により分類することで第一の複数のクラスタを生成する第一分類処理、

前記ラベリング対象のデータの少なくとも一部のデータを含むデータ群である第二のデータ群を分類することで第二の複数のクラスタを生成する第二分類処理、および、

前記第二の複数のクラスタに含まれるデータのうち、前記第一の複数のクラスタでは異なるクラスタに分類されていたデータ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。

[0100] (付記 11) コンピュータに、

ラベリング対象のデータ群の中から、生成された第一の複数のクラスタに応じて、第二のデータ群を生成するデータ精緻化処理を実行させる

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

[0101] (付記 12) コンピュータに、

ラベリング対象のデータ群である第一のデータ群を教師なし学習により分類することで第一の複数のクラスタを生成する第一分類処理、

前記ラベリング対象のデータの少なくとも一部のデータを含むデータ群である第二のデータ群を分類することで第二の複数のクラスタを生成する第二分類処理、および、

前記第二の複数のクラスタに含まれるデータのうち、前記第一の複数のクラスタでは異なるクラスタに分類されていたデータ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。

[0102] (付記 1 3) コンピュータに、

ラベリング対象のデータ群の中から、生成された第一の複数のクラスタに応じて、第二のデータ群を生成するデータ精緻化処理を実行させる

付記 1 2 記載のラベリング支援プログラム。

[0103] 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

[0104] 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

9 0 入出力装置

1 0 0 データ精緻化部

請求の範囲

- [請求項1] ラベリング対象のデータ群である第一のデータ群を教師なし学習により分類することで第一の複数のクラスタを生成する第一分類手段と、
- 、
- 前記ラベリング対象のデータの少なくとも一部のデータを含むデータ群である第二のデータ群を分類することで第二の複数のクラスタを生成する第二分類手段と、
- 前記第二の複数のクラスタに含まれるデータのうち、前記第一の複数のクラスタでは異なるクラスタに分類されていたデータ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] ラベリング対象のデータ群の中から、生成された第一の複数のクラスタに応じて、第二のデータ群を生成するデータ精緻化手段を備えた
- 請求項1記載のラベリング支援システム。
- [請求項3] データ精緻化手段は、ラベリング対象のデータ群のうち、第一の複数のクラスタのいずれかに分類されたデータに対してクラスタごとのラベリングを行った第二のデータ群を生成する
- 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] データ精緻化手段は、ラベリング対象のデータ群のうち、第一の複数のクラスタの中から選択されたクラスタに分類されているデータ群を第二のデータ群として生成する
- 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5] データ精緻化手段は、ラベリング対象のデータ群のうち、第一の複数のクラスタのいずれにも分類されなかった一以上のデータを除外したデータ群を第二のデータ群として生成する
- 請求項1から請求項4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6] 出力手段は、ラベリング対象のデータ群を次元削減し、第一の複数の

のクラスタに含まれる次元削減されたデータ、および、第二の複数のクラスタに含まれる次元削減されたデータをクラスタごとに識別できる態様でグラフ描画し、前記第二の複数のクラスタに含まれる次元削減されたデータのうち、第一の複数のクラスタでは異なるクラスタに分類されていたデータを、他のデータと異なる態様で表示する

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、データ群の分類処理ごとにクラスタの統計情報を表示する

請求項1から請求項6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8] コンピュータが、ラベリング対象のデータ群である第一のデータ群を教師なし学習により分類することで第一の複数のクラスタを生成し、

前記コンピュータが、前記ラベリング対象のデータの少なくとも一部のデータを含むデータ群である第二のデータ群を分類することで第二の複数のクラスタを生成し、

前記コンピュータが、前記第二の複数のクラスタに含まれるデータのうち、前記第一の複数のクラスタでは異なるクラスタに分類されていたデータを出力する

ことを特徴とするラベリング支援方法。

[請求項9] ラベリング対象のデータ群の中から、生成された第一の複数のクラスタに応じて、第二のデータ群を生成する

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群である第一のデータ群を教師なし学習により分類することで第一の複数のクラスタを生成する第一分類処理、
前記ラベリング対象のデータの少なくとも一部のデータを含むデー

タ群である第二のデータ群を分類することで第二の複数のクラスタを生成する第二分類処理、および、

前記第二の複数のクラスタに含まれるデータのうち、前記第一の複数のクラスタでは異なるクラスタに分類されていたデータ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。

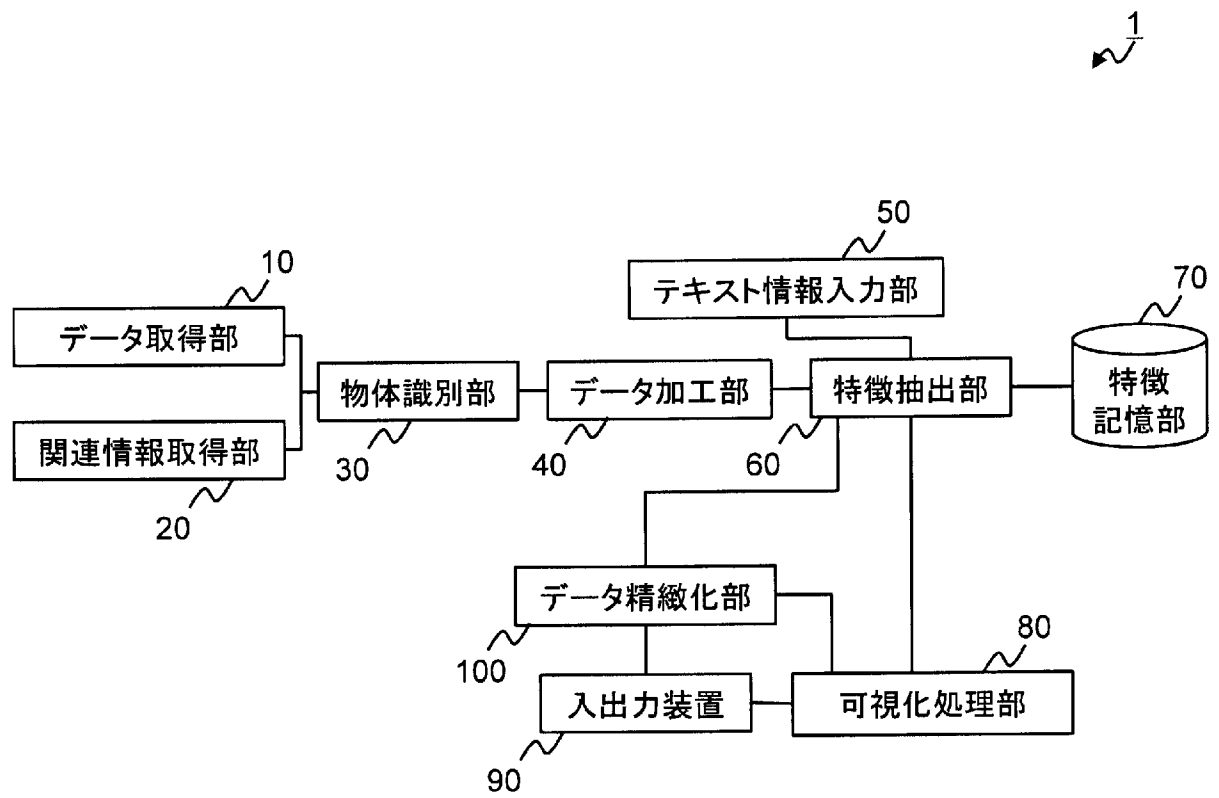
[請求項11]

コンピュータに、

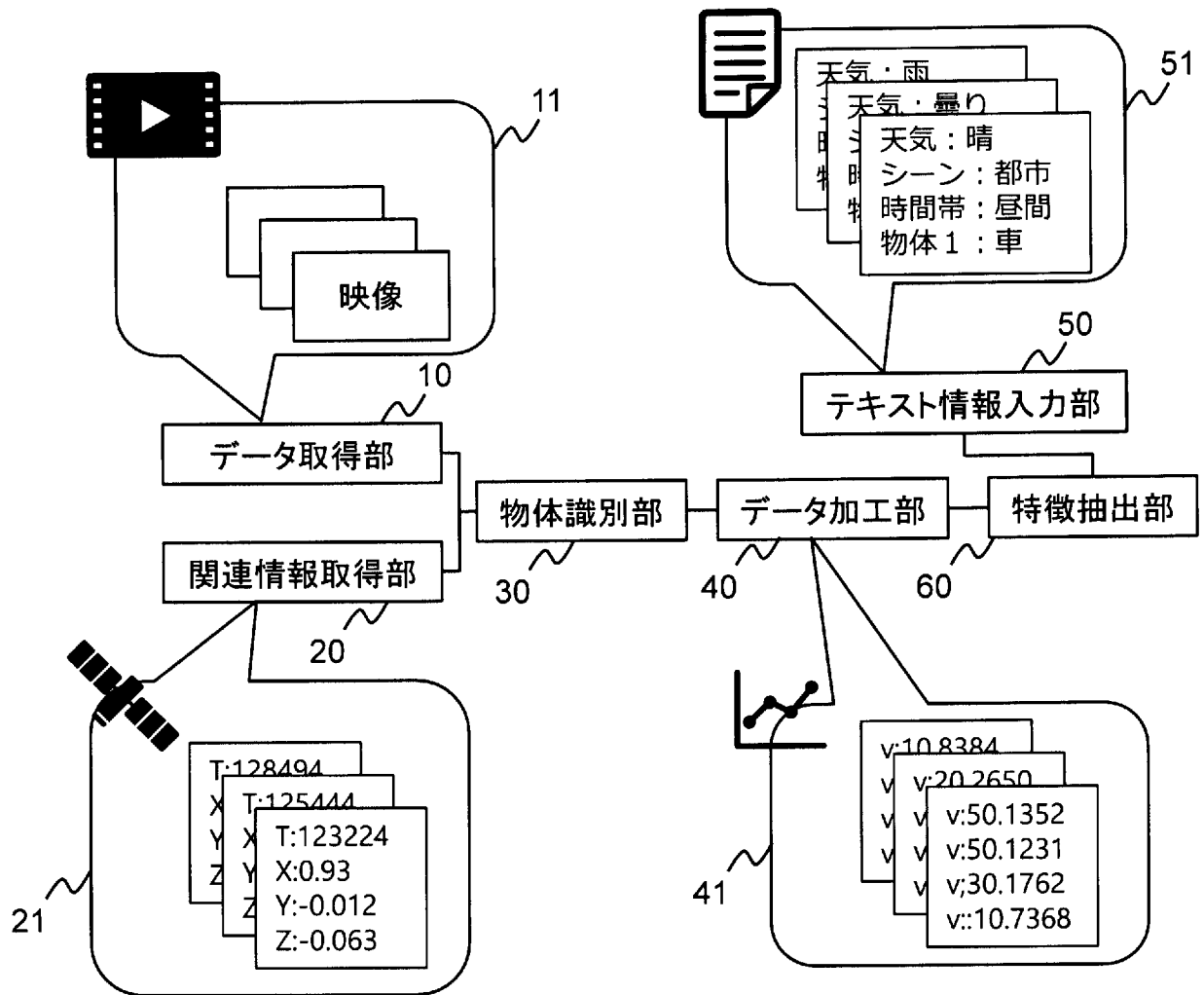
ラベリング対象のデータ群の中から、生成された第一の複数のクラスタに応じて、第二のデータ群を生成するデータ精緻化処理を実行させる

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]



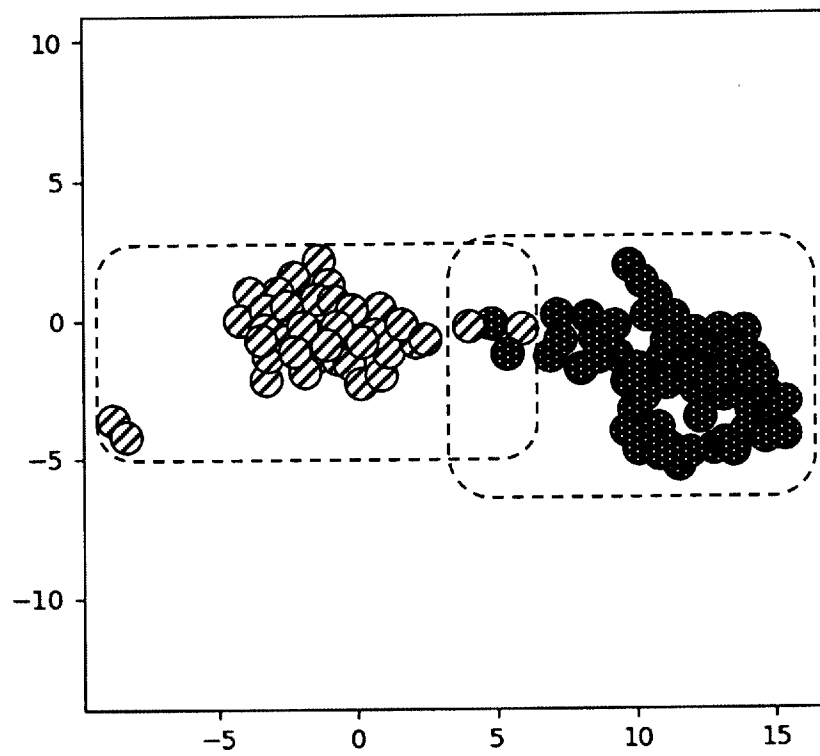
[図2]



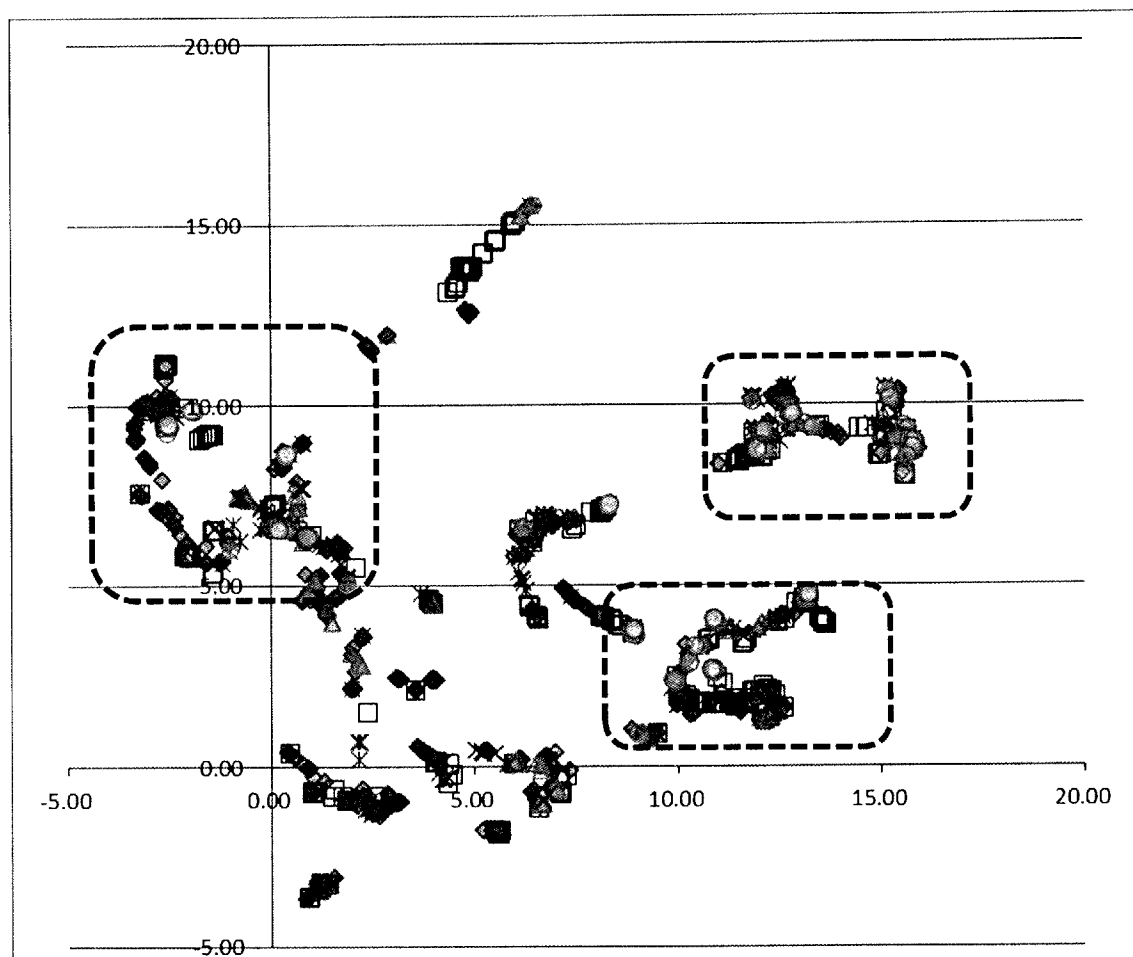
[図3]

-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

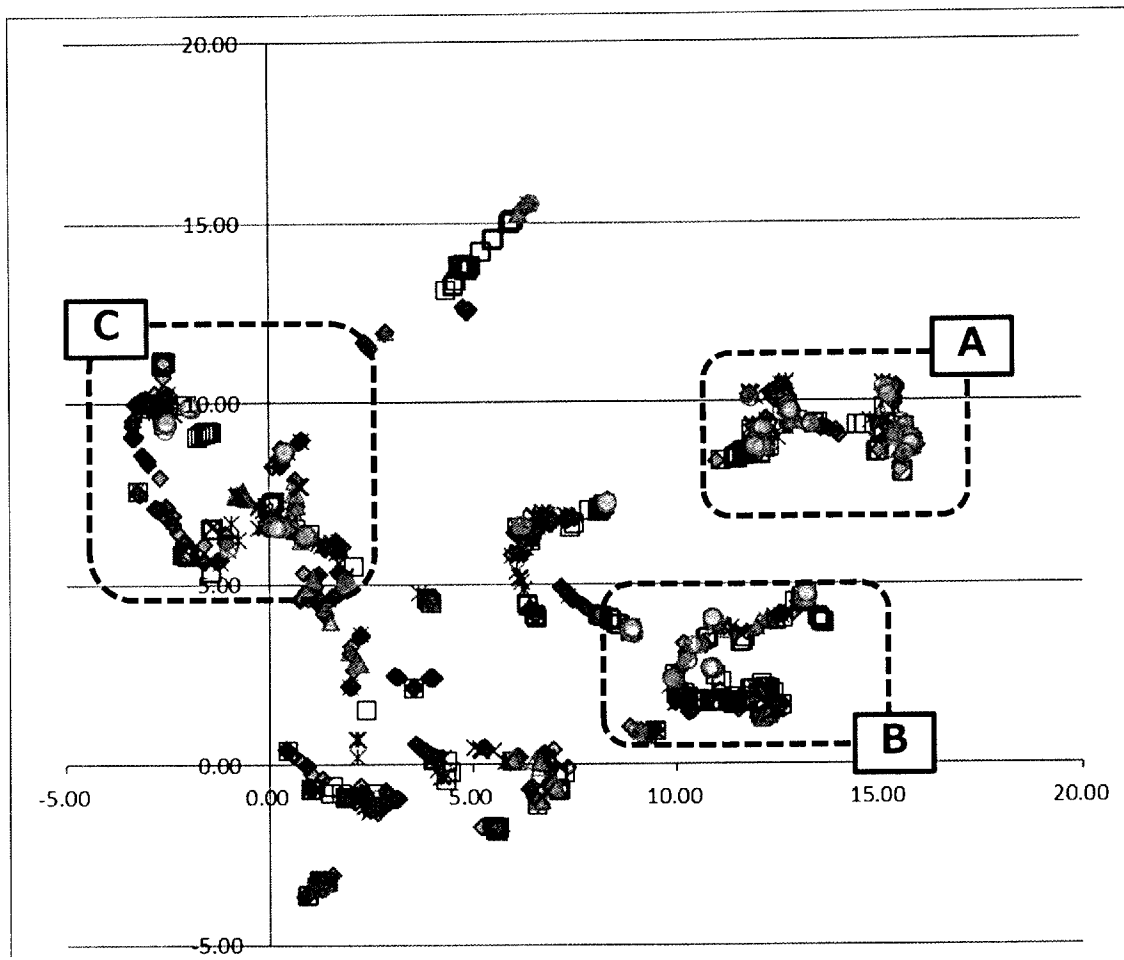
[図4]



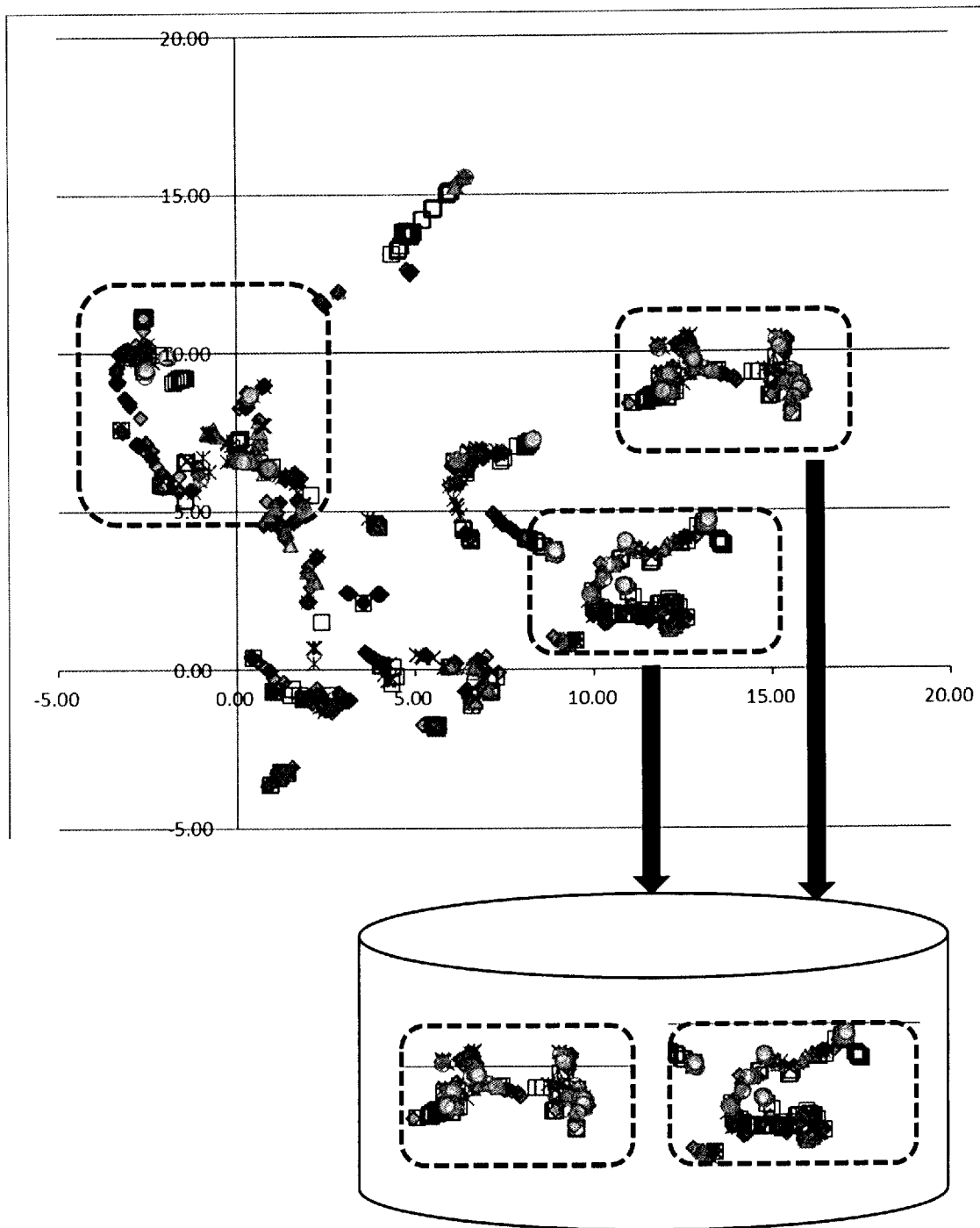
[図5]



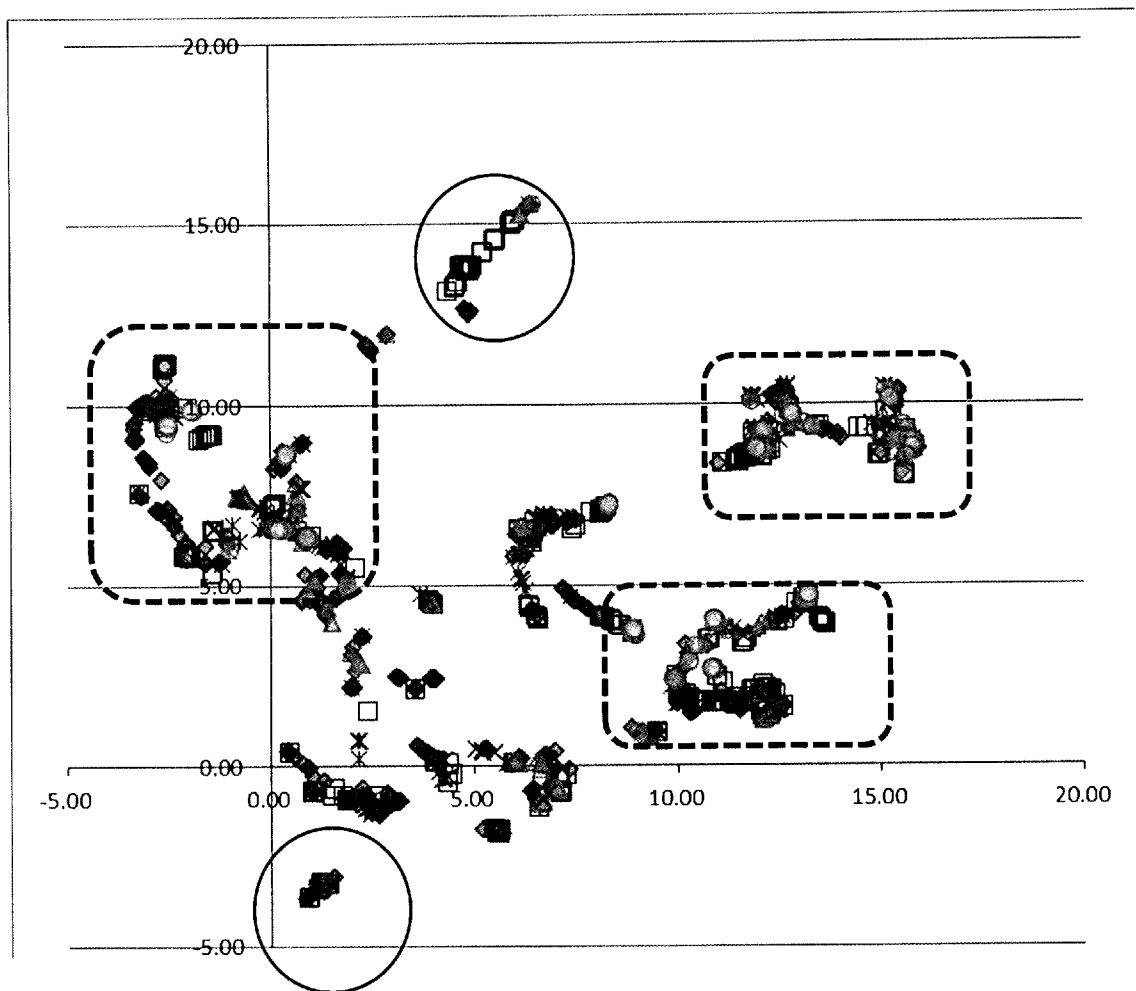
[図6]



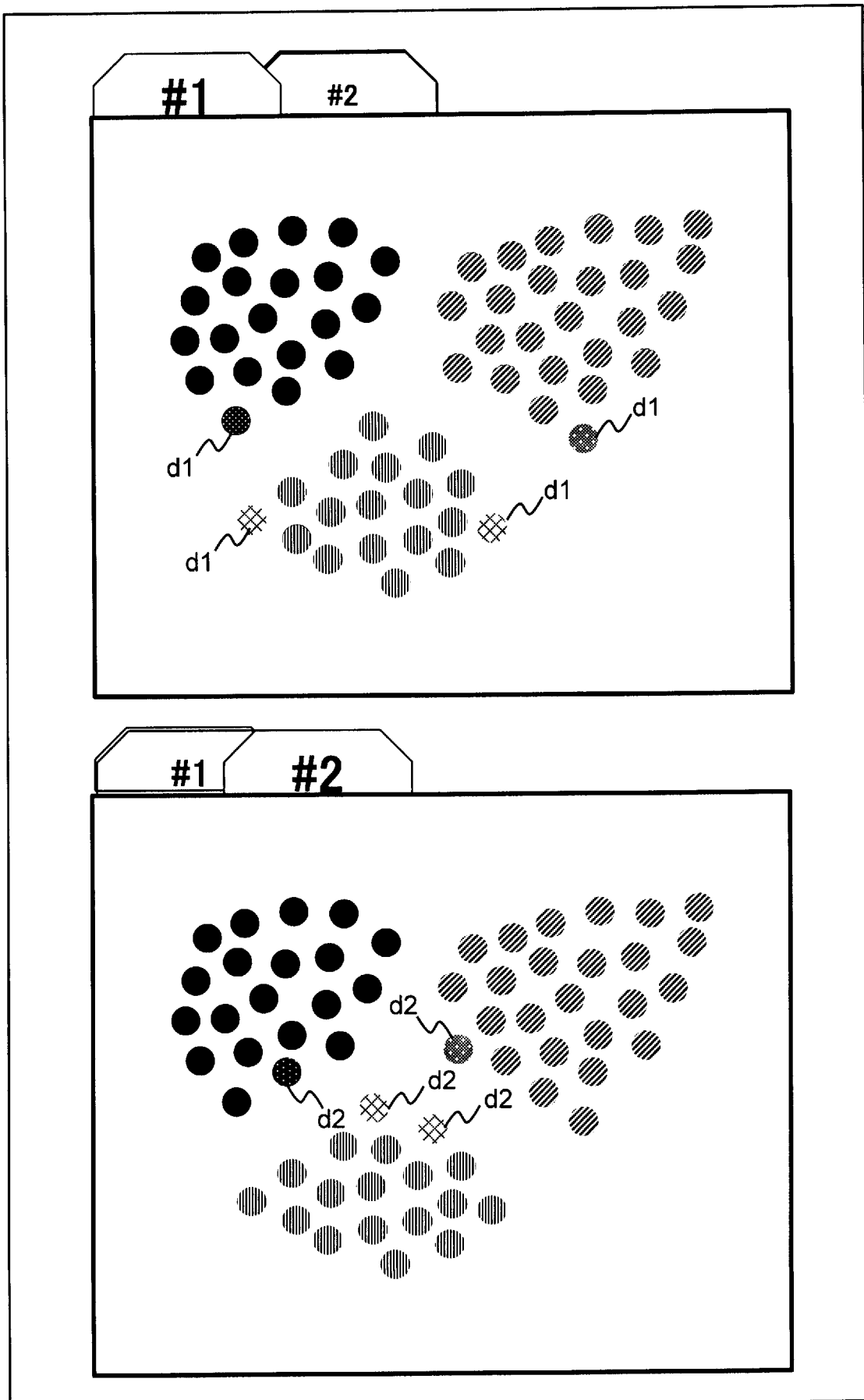
[図7]



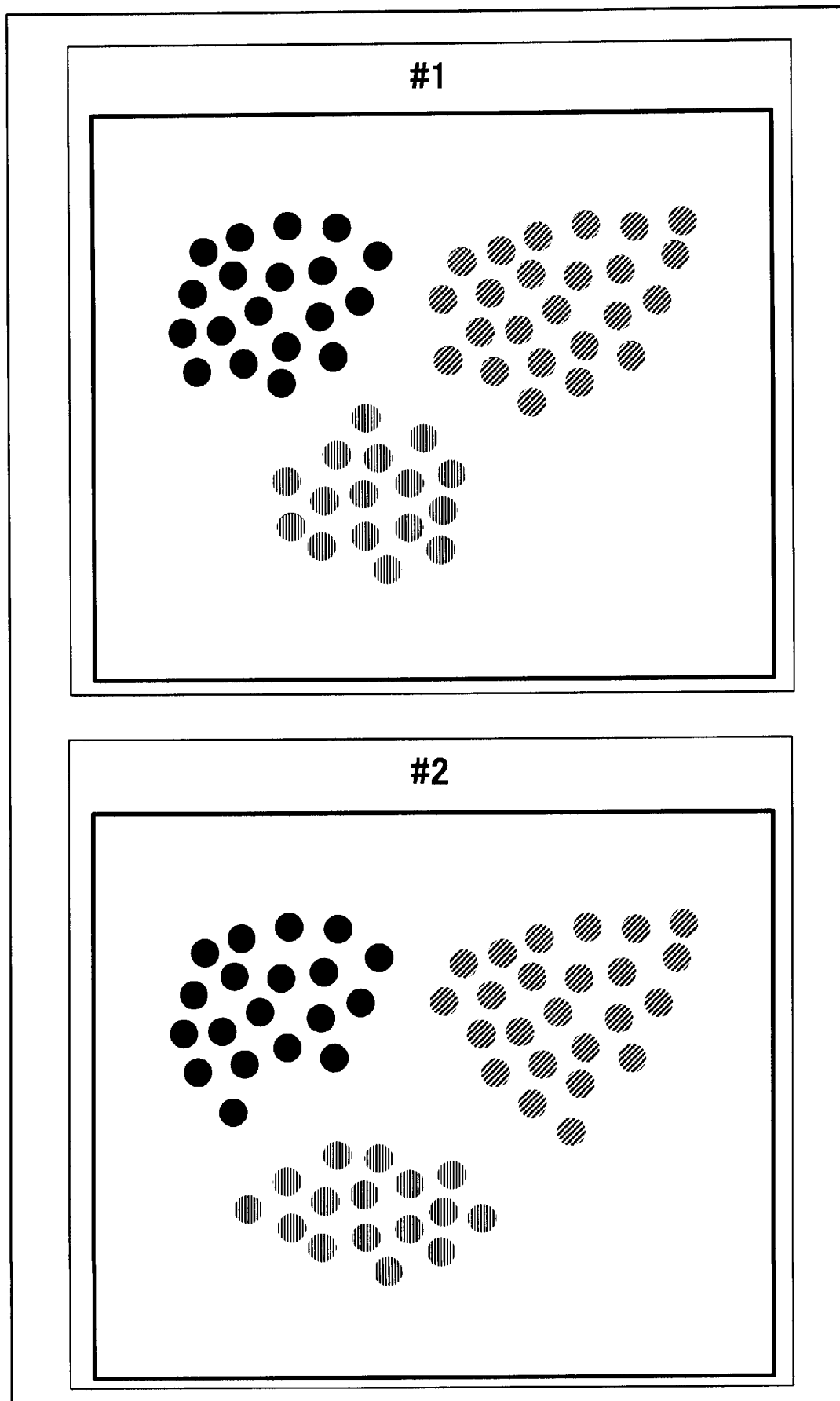
[図8]



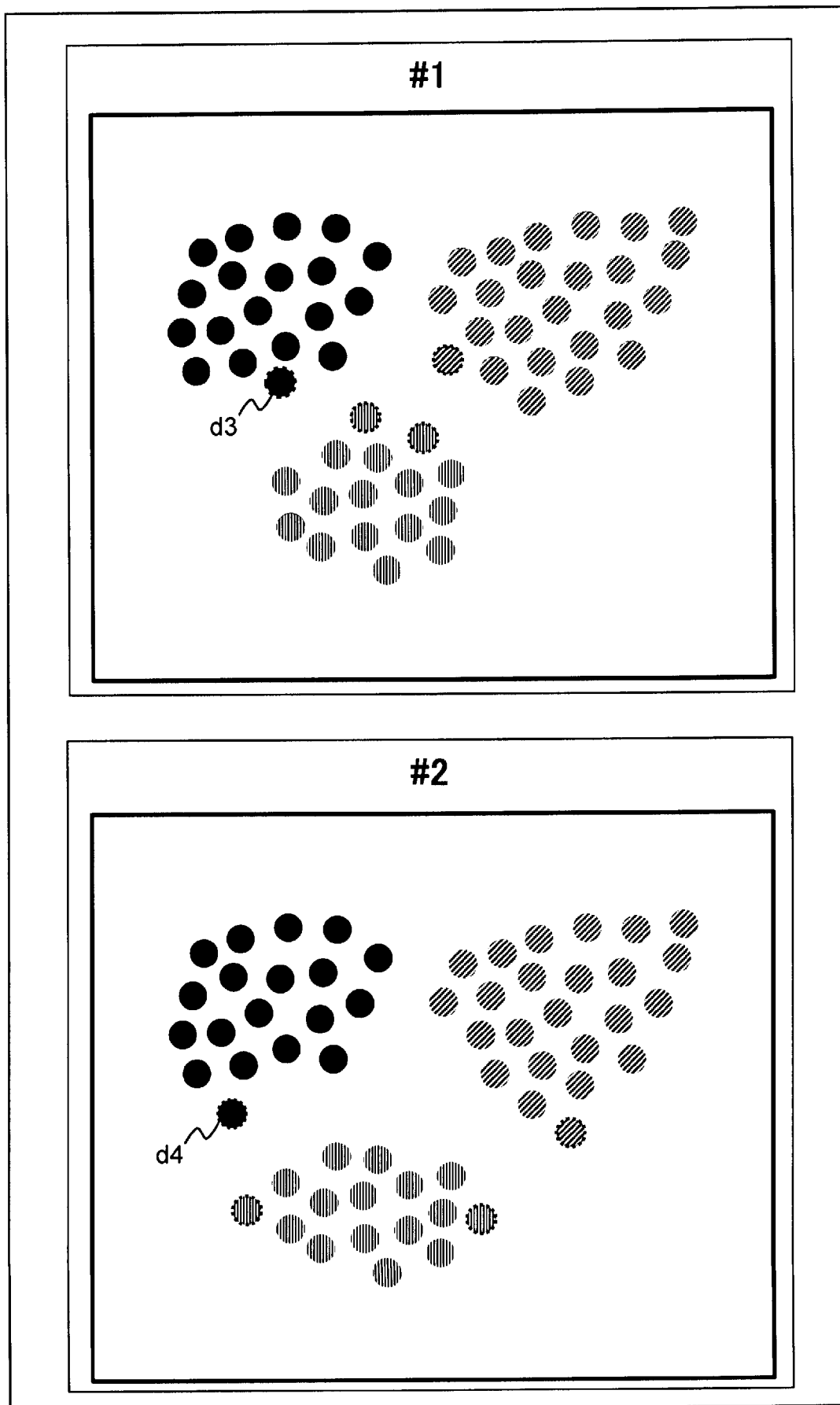
[図9]



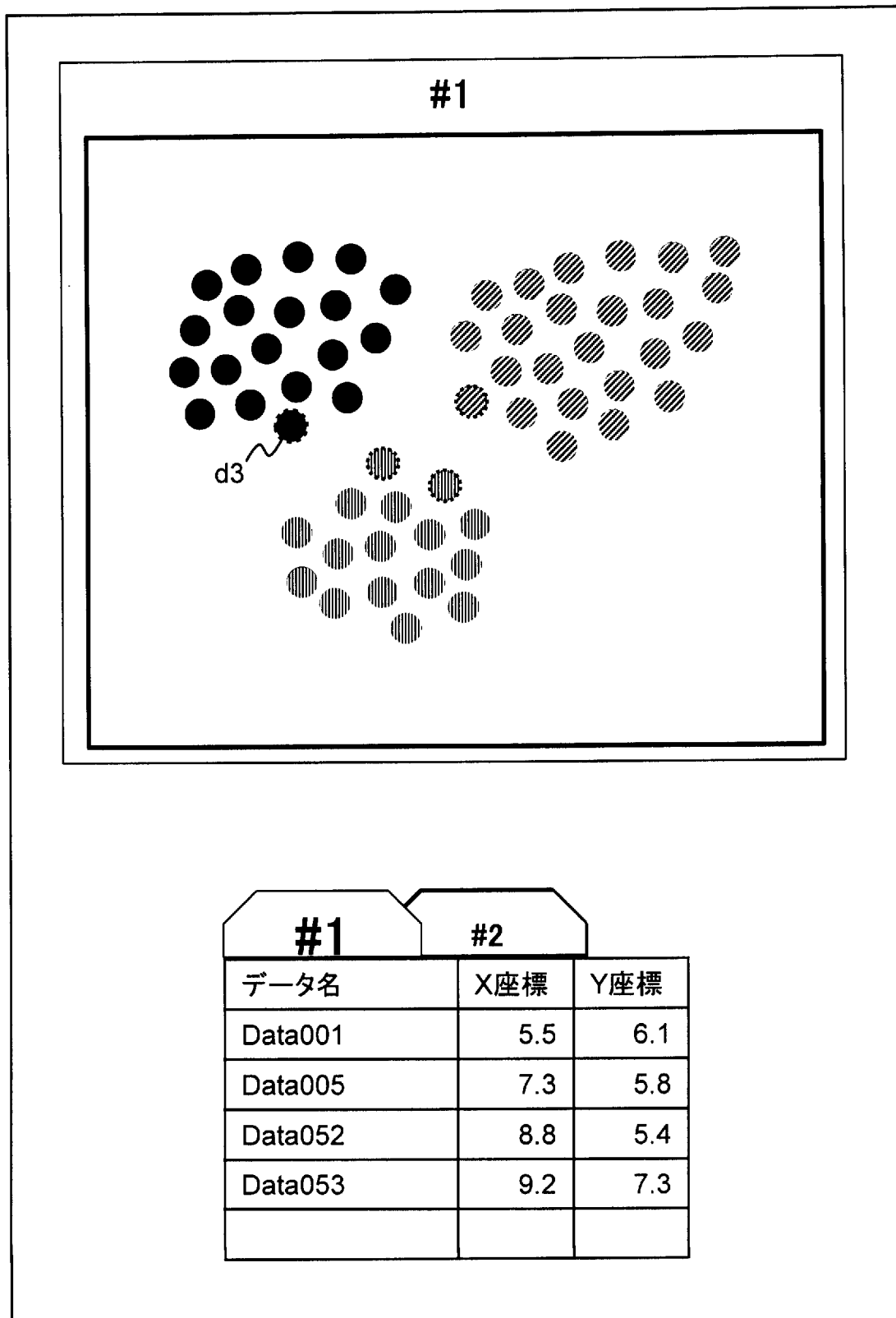
[図10]



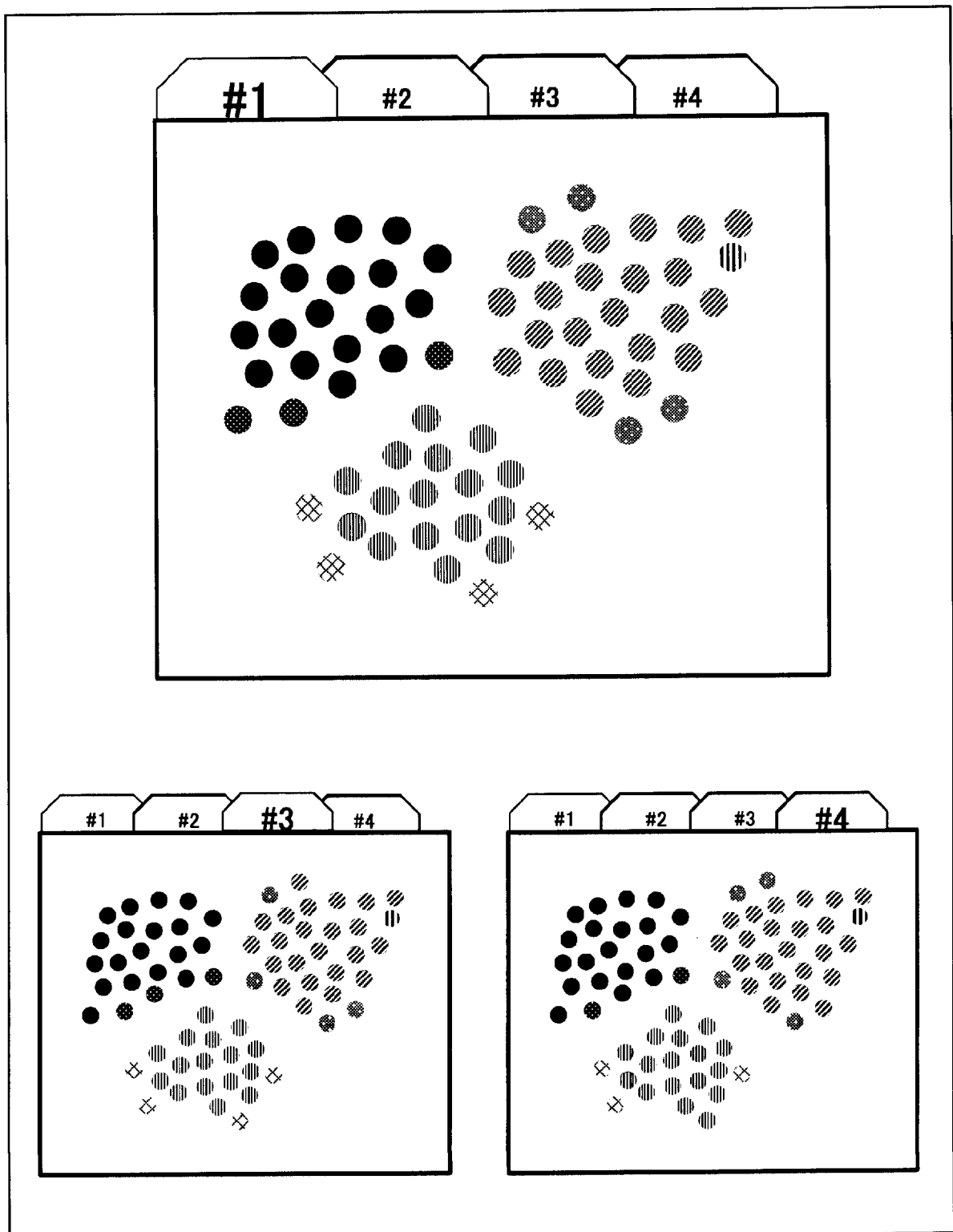
[図11]



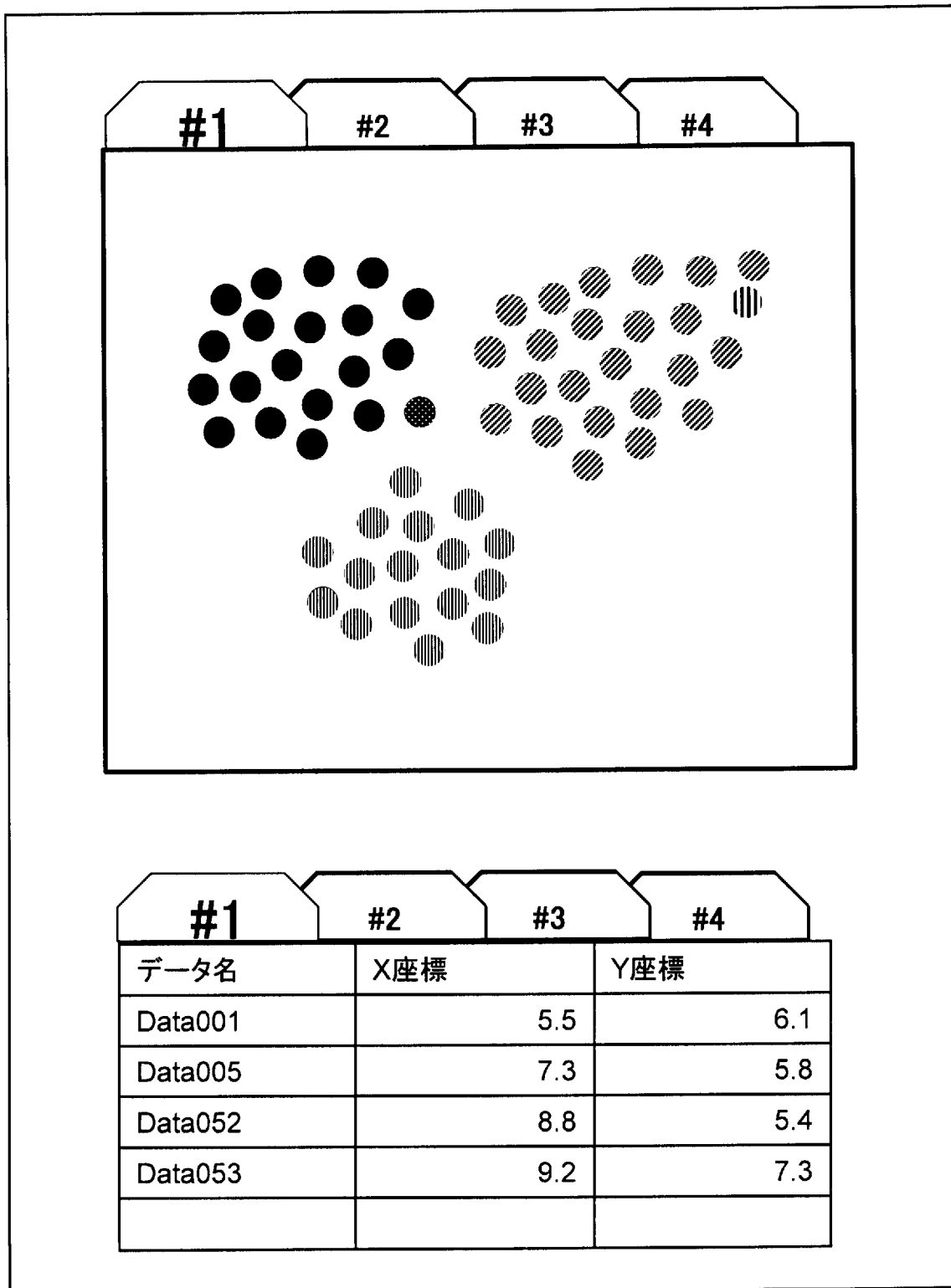
[図12]



[図13]



[図14]



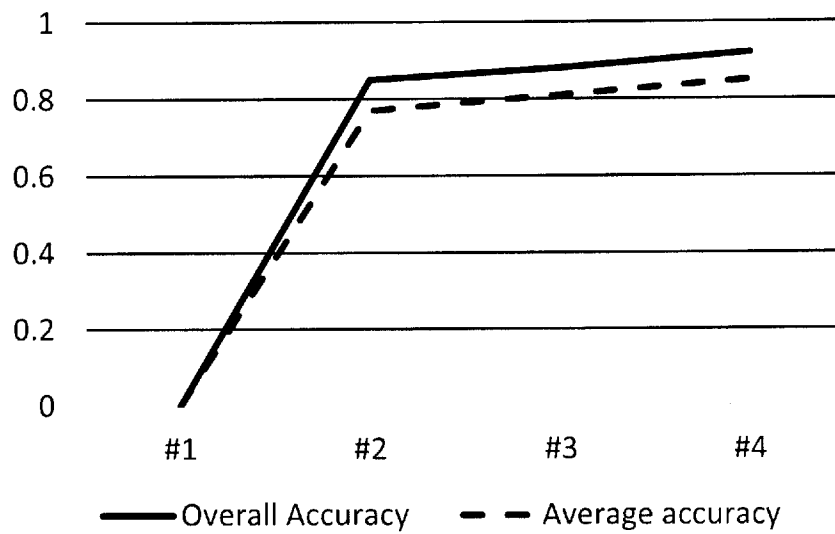
[図15]

#	クラスタ	項目名	数値
#1	クラスタ1	データ数	19
		重心	(5.5, 8.2)
		分散(x)	1.2
		分散(y)	1.1
		...	...
	クラスタ2	データ数	5.5
	...	...	...
#2	クラスタ1	データ数	19

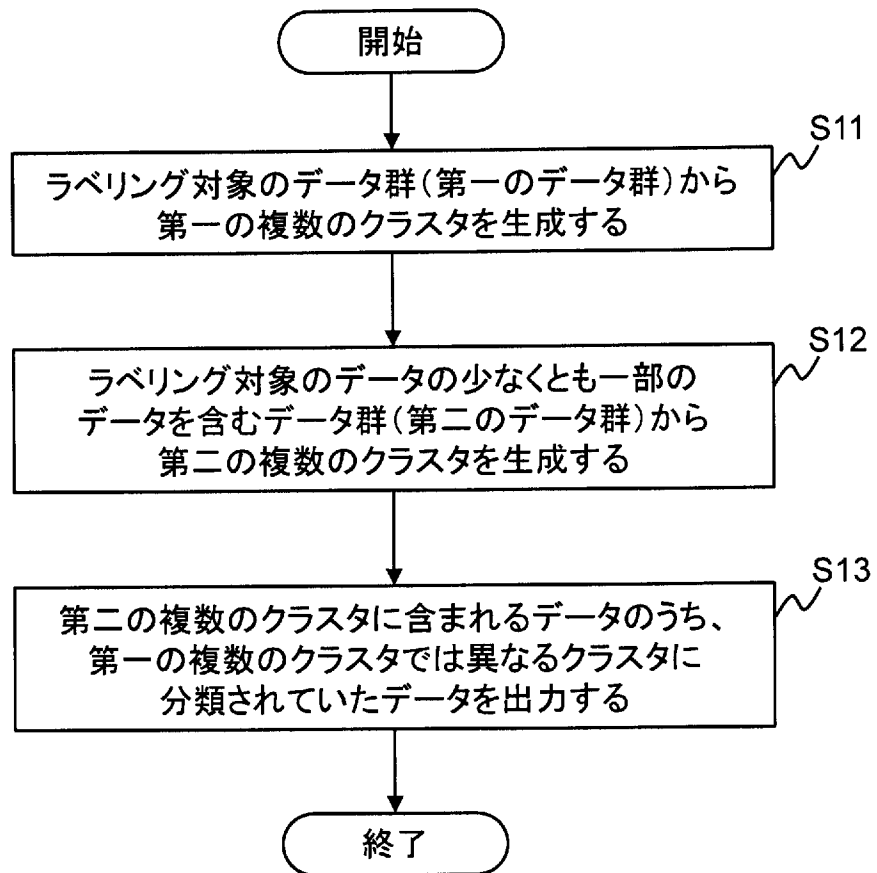
#1	#2	#3	#4
クラスタ	項目名	数値	
クラスタ1	データ数	19	
	重心	(5.5, 8.2)	
	分散(x)	1.2	
	分散(y)	1.1	
	...	...	
クラスタ2	データ数	5.5	

[図16]

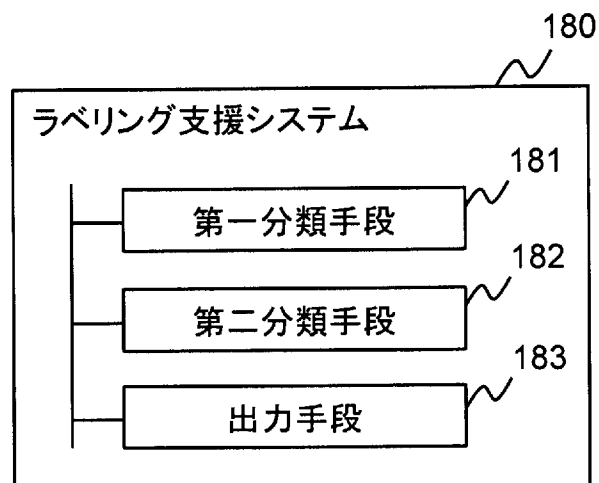
#	overall accuracy	average accuracy
#1	-	-
#2	0.85	0.77
#3	0.88	0.81
#4	0.92	0.85



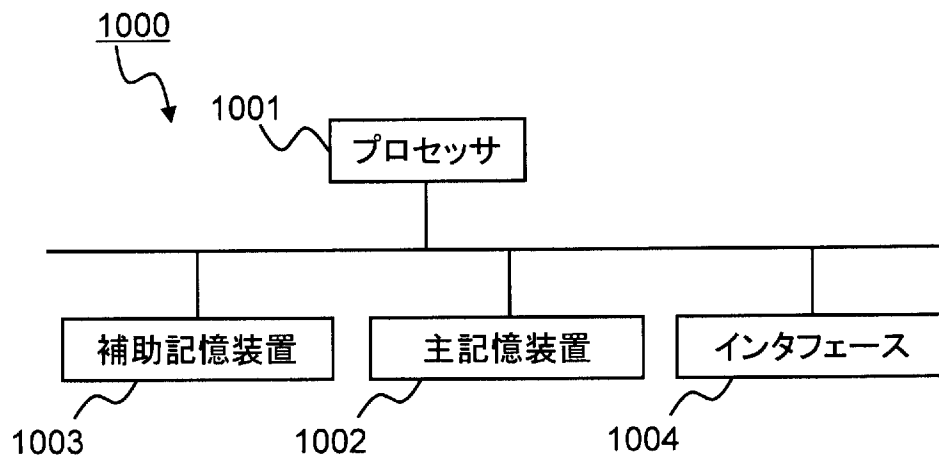
[図17]



[図18]



[図19]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2022/008749

A. CLASSIFICATION OF SUBJECT MATTER**G06F 16/75**(2019.01)i

FI: G06F16/75

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

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.
A	JP 2008-084151 A (JUST SYST CORP) 10 April 2008 (2008-04-10) paragraph [0066]	1-11
A	JP 2008-84203 A (NEC CORP) 10 April 2008 (2008-04-10) paragraph [0003]	1-11
A	JP 2014-63343 A (NIPPON TELEGR & TELEPH CORP <NTT>) 10 April 2014 (2014-04-10) paragraph [0003]	1-11

☐ 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/JP2022/008749

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
JP	2008-084151	A	10 April 2008	(Family: none)	
JP	2008-84203	A	10 April 2008	(Family: none)	
JP	2014-63343	A	10 April 2014	(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 FI: G06F16/75														
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														
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												
A	JP 2008-084151 A（株式会社ジャストシステム）10.04.2008（2008 - 04 - 10） [0066]	1-11												
A	JP 2008-84203 A（日本電気株式会社）10.04.2008（2008 - 04 - 10） [0003]	1-11												
A	JP 2014-63343 A（日本電信電話株式会社）10.04.2014（2014 - 04 - 10） [0003]	1-11												
☐ 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 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/JP2022/008749

引用文献			公表日	パテントファミリー文献	公表日
JP	2008-084151	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)	
JP	2014-63343	A	10.04.2014	(ファミリーなし)	