

明 細 書

発明の名称： 情報処理システム及び情報処理方法

関連出願の相互参照

- [0001] 本国際出願は、2022年3月25日に日本国特許庁に出願された日本国特許出願第2022-049442号に基づく優先権を主張するものであり、日本国特許出願第2022-049442号の全内容を本国際出願に参照により援用する。

技術分野

- [0002] 本開示は、情報処理システム及び情報処理方法に関する。

背景技術

- [0003] データフュージョン技術が既に知られている。出願人は、消費者の第一のグループに関する第一のデータセットと、消費者の第二のグループに関する第二のデータセットとを対応付ける技術を既に開示している。第一のデータセットは、消費者の第一のグループをクラスタリングして定義される複数のクラスタの特徴データを含む。第二のデータセットは、消費者の第二のグループをクラスタリングして定義される複数のクラスタの特徴データを含む。クラスタの類似性に基づいて、第一のデータセットにおける複数のクラスタの特徴データのそれぞれが、第二のデータセットにおける複数のクラスタの特徴データのいずれかに対応付けられる（例えば、特許文献1参照）。

先行技術文献

特許文献

- [0004] 特許文献1：特開2016-126609号公報

発明の概要

発明が解決しようとする課題

- [0005] 消費者個人の特徴データを、クラスタの特徴データに置き換えることは、個人情報保護しながら、消費者データを授受する目的で役立つ。
- [0006] しかしながら、クラスタの類似性を特徴データの内容に基づき評価するだ

けでは、高精度に特徴データ間の対応付けを行うことができない場合がある。すなわち、従来技術には、対応付けの精度について改善の余地があった。

[0007] そこで、本開示の一側面によれば、エンティティの第一のグループをクラスタリングして定義される複数の第一のクラスタ、及び、エンティティの第二のグループをクラスタリングして定義される複数の第二のクラスタに関して、複数の第一のクラスタの特徴データを、複数の第二のクラスタの特徴データに適切に対応付けることが可能な新規技術を提供できることが望ましい。

課題を解決するための手段

[0008] 本開示の一側面によれば、情報処理システムが提供される。情報処理システムは、第一取得部と、第二取得部と、対応付け部と、を備える。

[0009] 第一取得部は、第一のデータセットを取得するように構成される。第一のデータセットは、複数の第一のクラスタに関するデータセットである。複数の第一のクラスタは、エンティティの第一のグループをクラスタリングして定義される。第一のデータセットは、第一のクラスタ毎に、対応する第一のクラスタの特徴を表す第一のクラスタ特徴データを備える。

[0010] 第二取得部は、第二のデータセットを取得するように構成される。第二のデータセットは、複数の第二のクラスタに関するデータセットである。複数の第二のクラスタは、エンティティの第二のグループをクラスタリングして定義される。第二のデータセットは、第二のクラスタ毎に、対応する第二のクラスタの特徴を表す第二のクラスタ特徴データを備える。

[0011] 対応付け部は、一以上の共通エンティティに関して、共通エンティティのそれぞれが属する第一のクラスタ及び第二のクラスタを識別可能な情報に基づき、第一のデータセットが備える複数の第一のクラスタ特徴データのそれぞれを、第二のデータセットが備える複数の第二のクラスタ特徴データのうちの一つ以上に対応付けるように構成される。一以上の共通エンティティは、第一のグループと第二のグループとの間で共通する一以上のエンティティである。

- [0012] このように構成された情報処理システムによれば、第一のグループと第二のグループとの間に共通するエンティティの情報を活用し、クラスタに関する特徴データ間の対応付けを精度よく行うことができる。
- [0013] 本開示の一側面によれば、対応付け部は、互いに共通するエンティティを有する第一のクラスタ及び第二のクラスタのペアである第一のペア、及び、互いに共通するエンティティを有さない第一のクラスタ及び第二のクラスタのペアである第二のペアのうち、第一のペアに対応する第一のクラスタ特徴データと第二のクラスタ特徴データとを、第二のペアより優先的に対応付けるように、複数の第一のクラスタ特徴データのそれぞれを、複数の第二のクラスタ特徴データのうちの一つ以上に対応付けるように構成され得る。
- [0014] 本開示の一側面によれば、対応付け部は、少なくとも第二のペアに関しては、第二のペアに対応する第一のクラスタ特徴データ及び第二のクラスタ特徴データから判別される第一のクラスタと第二のクラスタとの間の特徴空間上の位置関係（例えば相対位置）に基づき、互に対応付ける第一のクラスタと第二のクラスタとのペアを判別し、第一のクラスタ特徴データと第二のクラスタ特徴データとを対応付けるように構成され得る。
- [0015] 本開示の一側面によれば、対応付け部は、複数の第一のクラスタのうちの二以上の特定クラスタに関しては、特定クラスタのそれぞれの第一のクラスタ特徴データを、対応する特定クラスタと同じエンティティを含む第二のクラスタの第二のクラスタ特徴データと対応付けるように構成され得る。二以上の特定クラスタは、複数の第一のクラスタのうち、それぞれが共通エンティティの少なくとも一つを含む二以上の第一のクラスタである。
- [0016] このように構成された情報処理システムによれば、共通エンティティが属する第一のクラスタと第二のクラスタとの関係に整合するように、第一のクラスタ特徴データと、第二のクラスタ特徴データとを対応付けることができる。
- [0017] 本開示の一側面によれば、対応付け部は、複数の第一のクラスタのうち、特定クラスタ以外の非特定クラスタに関しては、第一のクラスタ特徴データ

及び第二のクラスタ特徴データから判別される第一のクラスタと第二のクラスタとの間の特徴空間上の位置関係（例えば相対位置）に基づき、互いに対応付ける第一のクラスタと第二のクラスタとのペアを判別し、第一のクラスタ特徴データと第二のクラスタ特徴データとを対応付けるように構成され得る。

[0018] 本開示の一側面によれば、対応付け部は、一以上の非共通第一クラスタ及び一以上の非共通第二クラスタに関しては、一以上の非共通第一クラスタのそれぞれの第一のクラスタ特徴データと一以上の非共通第二クラスタのそれぞれの第二のクラスタ特徴データとに基づき、非共通第一クラスタのそれぞれと非共通第二クラスタのそれぞれとの間の関係を判別するように構成され得る。対応付け部は、判別された関係に基づいて、非共通第一クラスタのそれぞれの第一のクラスタ特徴データを、非共通第二クラスタの少なくとも一つの第二のクラスタ特徴データに対応付けるように構成され得る。一以上の非共通第一クラスタは、複数の第一のクラスタのうち、特定クラスタ以外の残りの一以上のクラスタである。一以上の非共通第二クラスタは、複数の第二のクラスタのうち、特定クラスタと対応付けられるクラスタ以外の残りの一以上のクラスタである。

[0019] このように構成された情報処理システムによれば、共通エンティティの有無と整合するように、第一のクラスタ特徴データと、第二のクラスタ特徴データとを対応付けることができる。

[0020] 本開示の一側面によれば、第一のクラスタ特徴データは、対応する第一のクラスタの第一の特徴空間上の点を定義するデータであり得る。第二のクラスタ特徴データは、対応する第二のクラスタの第二の特徴空間上の点を定義するデータであり得る。

[0021] 本開示の一側面によれば、対応付け部は、第一の特徴空間上の複数の第一のクラスタのそれぞれの点を第二の特徴空間に写す写像を、評価関数を用いて探索する処理を実行し得る。

[0022] 本開示の一側面によれば、評価関数は、同じエンティティを含む第一のク

ラストと第二のクラスタとのペアである特定ペアに関しては、特定ペアを構成する第一のクラスタと第二のクラスタとの間の第二の特徴空間上の距離が短いほど、特定方向に変化する値を出力するように設計され得る。

[0023] 本開示の一側面によれば、評価関数は、特定ペア以外の第一のクラスタと第二のクラスタとのペアに関しては、第一のクラスタと第二のクラスタとの間の第二の特徴空間上の距離が短いほど、特定方向とは逆方向に変化する値を出力するように設計され得る。この場合、対応付け部は、評価関数の出力を特定方向の限界値にする写像を探索するように構成され得る。

[0024] 本開示の一側面によれば、評価関数は、特定ペアに関しては、特定ペアを構成する第一のクラスタと第二のクラスタとの間の第二の特徴空間上の距離が短いほど、小さい値を出力するように設計され得る。本開示の一側面によれば、評価関数は、特定ペア以外の第一のクラスタと第二のクラスタとのペアに関しては、第一のクラスタと第二のクラスタとの間の第二の特徴空間上の距離が短いほど、大きい値を出力するように設計され得る。この場合、対応付け部は、評価関数の出力を最小にする写像を探索するように構成され得る。

[0025] 本開示の一側面によれば、対応付け部は、写像により写される第二の特徴空間上の複数の第一のクラスタのそれぞれの点と、第二の特徴空間上の複数の第二のクラスタのそれぞれの点との間の位置関係（例えば相対位置）に基づいて、複数の第一のクラスタ特徴データのそれぞれを、複数の第二のクラスタ特徴データのうちの一つ以上に対応付ける処理を実行するように構成され得る。

[0026] このように構成された情報処理システムによれば、共通エンティティに注目して、適切な第一の特徴空間から第二の特徴空間への写像を求めた上で、第一のクラスタ特徴データと、第二のクラスタ特徴データとを対応付けることができるので、高精度な対応付けを実現することができる。

[0027] 本開示の一側面によれば、対応付け部は、特定クラスタのそれぞれの第一のクラスタ特徴データを、同じエンティティを含む第二のクラスタの第二の

クラスタ特徴データと対応付ける処理を実行するように構成され得る。対応付け部は更に、写像により写される第二の特徴空間上の非共通第一クラスタのそれぞれの点と、第二の特徴空間上の非共通第二クラスタのそれぞれの点との間の位置関係（例えば相対位置）に基づいて、非共通第一クラスタのそれぞれの第一のクラスタ特徴データを、非共通第二クラスタの少なくとも一つの第二のクラスタ特徴データに対応付ける処理を実行するように構成され得る。

[0028] 本開示の一側面によれば、一以上の共通エンティティは、複数の共通エンティティであり得る。複数の第一のクラスタは、複数の共通エンティティが複数の第一のクラスタに分散して配置されるように、第一のグループをクラスタリングして定義されるクラスタの集合であり得る。複数の第二のクラスタは、複数の共通エンティティが複数の第二のクラスタに分散して配置されるように、第二のグループをクラスタリングして定義されるクラスタの集合であり得る。

[0029] 共通エンティティを複数のクラスタに分散させることによれば、すなわち、一つのクラスタに多数の共通エンティティが含まれないように、第一のグループ及び第二のグループをクラスタリングすれば、共通エンティティを通じて第一のクラスタと第二のクラスタとの間の対応関係を精度よく判別することができる。従って、第一のクラスタ特徴データと、第二のクラスタ特徴データとを精度よく対応付けることができる。

[0030] 本開示の一側面によれば、複数の第一のクラスタは、複数の共通エンティティのそれぞれが互いに異なる第一のクラスタに属するように、第一のグループをクラスタリングして定義されるクラスタの集合であり得る。複数の第二のクラスタは、複数の共通エンティティのそれぞれが互いに異なる第二のクラスタに属するように、第二のグループをクラスタリングして定義されるクラスタの集合であり得る。

[0031] 本開示の一側面によれば、第一のクラスタ特徴データは、対応する第一のクラスタに属する複数のエンティティの特徴データを統計処理により統合し

て生成される特徴データであり得る。第二のクラスタ特徴データは、対応する第二のクラスタに属する複数のエンティティの特徴データを統計処理により統合して生成される特徴データであり得る。

[0032] エンティティ個々の情報を隠すために、エンティティの特徴が統計化されたクラスタ特徴データが、データ保持者から提供され得る。本開示の一側面によれば、こうした統計処理がなされたクラスタ特徴データ間の対応付けを、適切に行うことができる。

[0033] 本開示の一側面によれば、エンティティは、消費者であり得る。エンティティの特徴データは、消費者の特徴を説明する特徴データであり得る。

[0034] 本開示の一側面によれば、上述した情報処理システムが有する機能を少なくとも部分的に実現するためのコンピュータプログラムが提供されてもよい。本開示の一側面によれば、上述した情報処理システムにおける第一取得部、第二取得部、及び対応付け部として、コンピュータを機能させるためのコンピュータプログラムが提供されてもよい。コンピュータプログラムは、コンピュータ読取可能な一時的でない記録媒体に記録され得る。

[0035] 本開示の一側面によれば、上述した情報処理システムに対応する情報処理方法が提供されてもよい。本開示の一側面によれば、コンピュータにより実行される情報処理方法が提供されてもよい。

[0036] 本開示の一側面によれば、情報処理方法は、第一のデータセットを取得することを含み得る。第一のデータセットは、複数の第一のクラスタに関するデータセットであり得る。複数の第一のクラスタは、エンティティの第一のグループをクラスタリングして定義される。第一のデータセットは、第一のクラスタ毎に、対応する第一のクラスタの特徴を表す第一のクラスタ特徴データを備え得る。

[0037] 情報処理方法は、第二のデータセットを取得することを含み得る。第二のデータセットは、複数の第二のクラスタに関するデータセットである。複数の第二のクラスタは、エンティティの第二のグループをクラスタリングして定義され得る。第二のデータセットは、第二のクラスタ毎に、対応する第二

のクラスタの特徴を表す第二のクラスタ特徴データを備え得る。

[0038] 情報処理方法は、一以上の共通エンティティに関して、共通エンティティのそれぞれが属する第一のクラスタ及び第二のクラスタを識別可能な情報に基づき、第一のデータセットが備える複数の第一のクラスタ特徴データのそれぞれを、第二のデータセットが備える複数の第二のクラスタ特徴データのうちの一つ以上に対応付けることを含み得る。一以上の共通エンティティは、第一のグループと第二のグループとの間で共通する一以上のエンティティであり得る。

[0039] この情報処理方法によっても、上述した情報処理システムと同様に、第一のデータセットに対応するエンティティの第一のグループと、第二のデータセットに対応するエンティティの第二のグループとの間に共通するエンティティの情報を活用し、クラスタに関する特徴データ間の対応付けを精度よく行うことができる。

[0040] 本開示の一側面によれば、コンピュータにより実行されたときに、上述の情報処理方法をコンピュータに実行させるための命令を含むコンピュータプログラムが提供されてもよい。コンピュータプログラムを記憶するコンピュータ読取可能な記録媒体が提供されてもよい。

図面の簡単な説明

[0041] [図1]情報処理システムの構成を表すブロック図である。

[図2]図2 Aは、データ提供システムにおける動作を説明する図であり、図2 Bは、例示的なクラスタ特徴データの構成を表す図である。

[図3]共通消費者（白丸）及び非共通消費者（黒丸）を含む消費者グループのクラスタリングに関する説明図である。

[図4]情報処理システムのプロセッサが実行する取得関連処理を表すフローチャートである。

[図5]図5 Aは、第一のクラスタ消費者関係表を表す図であり、図5 Bは、第二のクラスタ消費者関係表を表す図であり、図5 Cは、クラスタ関係表を表す図である。

[図6]情報処理システムのプロセッサが実行する分析処理を表すフローチャートである。

[図7]図7 Aは、分析処理における一部手順の第一例を表すフローチャートであり、図7 Bは、分析処理における一部手順の第二例を表すフローチャートである。

[図8]図8 Aは、クラスタリング処理の第一例を表すフローチャートであり、図8 Bは、クラスタリング処理の第二例を表すフローチャートである。

[図9]変形例の分析処理を表すフローチャートである。

符号の説明

[0042] 1…情報処理システム、11…プロセッサ、13…メモリ、15…ストレージ、15 A…第一のデータセット、15 B…第二のデータセット、15 C…拡張データセット、17…ユーザインタフェース、19…通信インタフェース、31…第一のデータ提供システム、32…第二のデータ提供システム、311…プロセッサ、312…メモリ、321…プロセッサ、322…メモリ、Pr…コンピュータプログラム。

発明を実施するための形態

[0043] 以下に本開示の例示的实施形態を、図面を参照しながら説明する。

[0044] 本実施形態の情報処理システム1は、図1に示すように、プロセッサ11と、メモリ13と、ストレージ15と、ユーザインタフェース17と、通信インタフェース19とを備える。

[0045] プロセッサ11は、ストレージ15に格納されたコンピュータプログラムPrに従う処理を実行する。メモリ13は、RAMを備える一次記憶装置である。メモリ13は、プロセッサ11による処理の実行時に作業エリアとして使用される。

[0046] ストレージ15は、例えばハードディスクドライブ又はソリッドステートドライブを備える二次記憶装置である。ストレージ15は、コンピュータプログラムPrの他、コンピュータプログラムPrに従う処理の実行時に供される各種データを記憶する。

- [0047] ユーザインタフェース 17 は、情報処理システム 1 を操作するユーザからの操作信号をプロセッサ 11 に入力するための入力デバイスを備える。入力デバイスの例には、キーボード及びポインティングデバイスが含まれる。ユーザインタフェース 17 は更に、ユーザに対して各種情報を表示するためのディスプレイを備える。
- [0048] 通信インタフェース 19 は、広域ネットワークに接続される。情報処理システム 1 は、通信インタフェース 19 を通じて、第一のデータ提供システム 31 及び第二のデータ提供システム 32 と通信可能に構成される。
- [0049] プロセッサ 11 は、コンピュータプログラム P_r に従う処理の実行により、通信インタフェース 19 を通じて第一のデータ提供システム 31 から第一のデータセット 15A を取得し、第二のデータ提供システム 32 から第二のデータセット 15B を取得する。
- [0050] プロセッサ 11 は、取得した第一のデータセット 15A に、第二のデータセット 15B を結合することにより、データリッチ化した第一のデータセット 15A である拡張データセット 15C を生成する。
- [0051] 第一のデータセット 15A は、エンティティの第一のグループをクラスタリングして定義される複数の第一のクラスタに関し、第一のクラスタ毎に、対応する第一のクラスタの特徴を表すクラスタ特徴データを備えるデータセットである。
- [0052] 第二のデータセット 15B は、エンティティの第二のグループをクラスタリングして定義される複数の第二のクラスタに関し、第二のクラスタ毎に、対応する第二のクラスタの特徴を表すクラスタ特徴データを備えるデータセットである。
- [0053] 本実施形態によれば、エンティティは、消費者である。すなわち、第一のデータセット 15A は、消費者の第一のグループをクラスタリングして定義される複数の第一のクラスタに関し、第一のクラスタ毎に、クラスタ特徴データを備える。
- [0054] 図 2A に示すように、第一のデータ提供システム 31 は、第一のグループ

における複数の消費者の消費者特徴データを記憶する。各消費者特徴データは、対応する消費者の特徴を表す。第一のデータ提供システム 31 は、クラスタリング処理を実行することにより、第一のグループを、複数の第一のクラスタにクラスタリングする。

[0055] 第一のデータ提供システム 31 は更に、第一のクラスタ毎に、この第一のクラスタに関する消費者特徴データの一群を統計処理して、第一のクラスタ毎のクラスタ特徴データを生成する。

[0056] クラスタ特徴データのそれぞれは、対応するクラスタに属する複数の消費者の消費者特徴データを統計処理により統合して生成される。統計処理は、例えば個人情報保護のために行われる。第一のデータ提供システム 31 は、このように生成した第一のクラスタ毎のクラスタ特徴データを備える第一のデータセット 15 A を情報処理システム 1 に提供する。

[0057] 第一のデータ提供システム 31 は、プロセッサ 311 及びメモリ 312 を備える。プロセッサ 311 は、メモリ 312 が記憶するコンピュータプログラムに従って、クラスタリング処理及び統計処理を実行する。

[0058] 第二のデータ提供システム 32 もまた、プロセッサ 321 及びメモリ 322 を備える。第二のデータ提供システム 32 は、第二のグループにおける複数の消費者の消費者特徴データを記憶する。プロセッサ 321 は、第一のデータ提供システム 31 と同様に、メモリ 322 が記憶するコンピュータプログラムに従って、第二のデータ提供システム 32 が保持する消費者特徴データの一群に対し、クラスタリング処理及び統計処理を実行する。

[0059] これにより、第二のデータ提供システム 32 は、消費者の第二のグループをクラスタリングして定義される複数の第二のクラスタに関し、第二のクラスタ毎に、対応する第二のクラスタの特徴を表すクラスタ特徴データを備える第二のデータセット 15 B を生成し、情報処理システム 1 に提供する。

[0060] 消費者特徴データは、対応する消費者の特徴を、複数の要素の値によりベクトル表現する。クラスタ特徴データもまた、対応するクラスタの特徴を、複数の要素の値によりベクトル表現する。

- [0061] クラスタ特徴データにおける各要素の値は、対応するクラスタに属する消費者集合における、対応する要素の統計的な代表値である。代表値は、消費者集合における対応する要素の平均値、中央値、平均値と分散との組合せ、及び、最大値と最小値との組合せの一以上を含み得る。
- [0062] 図2Bに示すクラスタ特徴データは、クラスタの識別コードであるクラスタIDに関連付けて、対応するクラスタの特徴を表す各要素の値と、クラスタサイズと、を記述する。クラスタサイズは、対応するクラスタを構成する消費者の数に対応する。
- [0063] 本実施形態では更に、消費者の第一のグループと、消費者の第二のグループとの間に、共通する消費者が存在する可能性を考慮して、情報処理システム1が、第一のデータ提供システム31及び第二のデータ提供システム32と通信する。
- [0064] これにより、情報処理システム1は、第一のグループと第二のグループとの間に共通して存在する消費者の一群を、共通消費者の一群として判別し、共通消費者のリストを、第一のデータ提供システム31及び第二のデータ提供システム32に送信する。
- [0065] 第一のデータ提供システム31は、共通消費者のリストに基づいて、第一のグループを、共通消費者の一群が複数の第一のクラスタに分散して配置されるように、第一のグループをクラスタリングする。共通消費者の数がクラスタ数以下である場合には、複数の共通消費者のそれぞれが互いに異なるクラスタに属するように、第一のグループをクラスタリングする。
- [0066] 図3は、白丸及び黒丸で表される消費者の集合を、一つのクラスタに含まれる共通消費者が可能な限り少なくなるように、理想的には一つのクラスタに含まれる共通消費者が一人になるように、クラスタリングすることを説明している。丸い破線で囲われた部位が、一つのクラスタに対応する。白丸のそれぞれは、共通消費者の一人を示し、黒丸のそれぞれは、共通消費者ではない消費者、すなわち非共通消費者の一人を示す。
- [0067] 第二のデータ提供システム32も同様に、第二のグループを、共通消費者

の一群が複数の第二のクラスタに分散して配置されるように、第二のグループをクラスタリングする。共通消費者の数がクラスタ数以下である場合には、複数の共通消費者のそれぞれが互いに異なる第二のクラスタに属するように、第二のグループをクラスタリングする。

[0068] 情報処理システム 1 は、このようなクラスタリングにより生成された第一のクラスタ毎のクラスタ特徴データを含む第一のデータセット 15 A を、第一のデータ提供システム 31 から取得する。情報処理システム 1 は更に、第二のクラスタ毎のクラスタ特徴データを含む第二のデータセット 15 B を、第二のデータ提供システム 32 から取得する。

[0069] 図 4 には、第一のデータセット 15 A 及び第二のデータセット 15 B の取得のために、情報処理システム 1 のプロセッサ 11 が実行する取得関連処理の詳細を示す。プロセッサ 11 は、ユーザインタフェース 17 を通じたユーザからの指示に基づき、取得関連処理を開始する。

[0070] 取得関連処理を開始すると、プロセッサ 11 は、第一のデータ提供システム 31 から第一のグループに属する消費者のリストを、第一消費者リストとして取得する (S110)。第一消費者リストは、第一のグループに属する消費者毎に、対応する消費者の識別コードである消費者 ID の情報を有する。

[0071] プロセッサ 11 は更に、第二のデータ提供システム 32 から第二のグループに属する消費者のリストを、第二消費者リストとして取得する (S120)。第二消費者リストは、第二のグループに属する消費者毎に、消費者 ID の情報を有する。

[0072] 共通消費者の判別のために、第一消費者リスト及び第二消費者リストには、同種の消費者 ID が記述され得る。消費者 ID は、例えば、広告 ID であり得る。あるいは、消費者 ID は、非常に多く会員を有するリワードプログラムの会員 ID などに代表される、多数の消費者に付与された ID であり得る。

[0073] 第一消費者リストで使用される消費者 ID である第一消費者 ID は、第二

消費者リストで使用される消費者IDである第二消費者IDとは異なる種類の識別コードであってもよい。この場合には、同一人物を判別するために、第一消費者IDと第二消費者IDとの間の対応関係を示すテーブルが必要になり得る。

[0074] 続くS130において、プロセッサ11は、第一消費者リストに列挙された消費者IDと、第二消費者リストに列挙された消費者IDとの比較により、第一のグループと第二のグループとの間で共通する一以上の消費者である一以上の共通消費者を判別し、共通消費者のリストを生成する。

[0075] 続くS140において、プロセッサ11は、S130で作成した共通消費者のリストを、第一及び第二のデータ提供システム31、32に送信し、第一及び第二のデータ提供システム31、32に、このリストに基づき、上述した第一のデータセット15A及び第二のデータセット15Bを生成させる。

[0076] 続くS150において、プロセッサ11は、第一のデータ提供システム31から、第一のデータセット15Aを取得すると共に、第一のクラスタ消費者関係表を取得する。第一のクラスタ消費者関係表は、例えば、図5Aに示すように、共通消費者の消費者ID毎に、対応する消費者が属する第一のクラスタの識別コードである第一クラスタIDを記述する。第一のデータ提供システム31は、第一のデータセット15Aと共に、第一のクラスタ消費者関係表を情報処理システム1に送信するように構成される。

[0077] プロセッサ11は、第二のデータ提供システム32から、第二のデータセット15Bを取得すると共に、第二のクラスタ消費者関係表を取得する（S160）。第二のクラスタ消費者関係表は、例えば、図5Bに示すように、共通消費者の消費者ID毎に、対応する消費者が属する第二のクラスタの識別コードである第二クラスタIDを記述する。第二のデータ提供システム32は、第二のデータセット15Bと共に、第二のクラスタ消費者関係表を情報処理システム1に送信するように構成される。

[0078] 続くS170において、プロセッサ11は、取得した第一及び第二のクラ

スタ消費者関係表に基づき、同じ消費者が含まれる第一のクラスタと第二のクラスタとの組合せを示した図5Cに示すクラスタ関係表を生成する。クラスタ関係表は、共通消費者のそれぞれが属する第一のクラスタ及び第二のクラスタを識別可能な情報に対応する。図5Cに例示されるクラスタ関係表は、同じ共通消費者が含まれる第一のクラスタと、第二のクラスタとの組合せを、第一のクラスタのクラスタIDと第二のクラスタのクラスタIDとの組み合わせにより示す。

[0079] 続くS180において、プロセッサ11は、S170で生成したクラスタ関係表を、第一のデータセット15A及び第二のデータセット15Bと共に、ストレージ15に保存して図4に示す処理を終了する。

[0080] 続いて、プロセッサ11が実行する分析処理の詳細を、図6を用いて説明する。分析処理は、例えばユーザインタフェース17を通じたユーザからの指示に基づいて実行される。

[0081] 分析処理を開始すると、プロセッサ11は、ユーザから指定された第一のデータセット15Aをストレージ15から読み出し、読み出した第一のデータセット15Aに含まれる第一のクラスタ毎のクラスタ特徴データに基づき、配列 D_f を生成する(S210)。

[0082] 例えば、クラスタがK1個であり、K1個のクラスタ特徴データのそれぞれが、L1個の要素を有するL1次元特徴ベクトルでクラスタの特徴を表すとき、配列 D_f は、K1行L1列の二次元配列である。

[0083] 配列 D_f の各行ベクトル $D_f[i] = (x_i[1], x_i[2], x_i[3], \dots, x_i[L1])$ は、対応するi番目の第一のクラスタのL1次元特徴ベクトルである。ここでのiは、配列 D_f の行番号に対応し、クラスタ数K1に対応する値1から値K1までの整数値を取る。

[0084] 以下では、第一のデータセット15Aに含まれるクラスタ特徴データによって表現される特徴ベクトル $D_f[i]$ に対応するL1次元空間のことを第一の特徴空間と表現する。特徴ベクトル $D_f[i]$ は、対応する第一のクラスタが第一の特徴空間上において配置される点を定義する。第一の特徴空間は、

特徴ベクトル $D_f[i]$ が有する L 1 個の要素のそれぞれに対応する次元を有する L 1 次元空間である。

[0085] 続く S 2 2 0 において、プロセッサ 1 1 は、ユーザから指定された第二のデータセット 1 5 B をストレージ 1 5 から読み出し、読み出した第二のデータセット 1 5 B に含まれる第二のクラスタ毎のクラスタ特徴データに基づき、配列 D_s を生成する。

[0086] 例えば、クラスタが K 2 個であり、 K 2 個のクラスタ特徴データのそれぞれが、 L 2 個の要素を有する L 2 次元特徴ベクトルでクラスタの特徴を表すとき、配列 D_s は、 K 2 行 L 2 列の二次元配列である。

[0087] 配列 D_s の各行ベクトル $D_s[j] = (y_j[1], y_j[2], y_j[3], \dots, y_j[L2])$ は、対応する j 番目の第二のクラスタの L 2 次元特徴ベクトルである。ここでの j は、配列 D_s の行番号に対応し、クラスタ数 K 2 に対応する値 1 から値 K 2 までの整数値を取る。

[0088] 以下では、第二のデータセット 1 5 B に含まれるクラスタ特徴データによって表現される特徴ベクトル $D_s[j]$ に対応する L 2 次元空間のことを第二の特徴空間と表現する。特徴ベクトル $D_s[j]$ は、対応する第二のクラスタが第二の特徴空間上で配置される点を定義する。第二の特徴空間は、特徴ベクトル $D_s[j]$ が有する L 2 個の要素のそれぞれに対応する次元を有する L 2 次元空間である。

[0089] 続く S 2 3 0 において、プロセッサ 1 1 は、次に示す評価関数 E を最小にする関数 g を探索する。これにより、プロセッサ 1 1 は、第一のデータセット 1 5 A から特定される第一の特徴空間上の共通消費者に対応する点を、第二のデータセット 1 5 B から特定される第二の特徴空間上の同一人物の点近くに、最もよく写す写像を探索する。

[0090] [数1]

$$E = \sum_{m,n \in M1} d(g(D_f[A_f^{-1}[m]]), D_s[A_s^{-1}[n]]) - \tau \sum_{m,n \in M2} d(g(D_f[A_f^{-1}[m]]), D_s[A_s^{-1}[n]])$$

上式における m は、第一のデータセット 15 A にクラスタ特徴データが記述される $K-1$ 個の第一のクラスタに対応する消費者群のうちの m 番目の消費者を意味する。

[0091] ここでの消費者群は、第一のクラスタのそれぞれには、クラスタサイズに対応する複数の消費者が存在するとみなしたときに解釈される $K-1$ 個の第一のクラスタに対応する消費者群であり得る。あるいは、消費者群は、クラスタのそれぞれに一人の代表する消費者が存在するとみなしたときに解釈される $K-1$ 個の第一のクラスタに対応する $K-1$ 人の消費者群であり得る。 $A_f^{-1}[m]$ は、 m 番目の消費者が所属する第一のクラスタが、 $K-1$ 個のクラスタのうち、 $A_f^{-1}[m]$ 番目のクラスタであることを意味する。

[0092] 上式における n は、第二のデータセット 15 B にクラスタ特徴データが記述される $K-2$ 個の第二のクラスタに対応する消費者群のうちの n 番目の消費者を意味する。ここでの消費者群は、第二のクラスタのそれぞれには、クラスタサイズに対応する複数の消費者が存在するとみなしたときに解釈される $K-2$ 個の第二のクラスタに対応する消費者群であり得る。あるいは、消費者群は、クラスタのそれぞれに一人の代表する消費者が存在するとみなしたときに解釈される $K-2$ 個の第二のクラスタに対応する $K-2$ 人の消費者群であり得る。 $A_s^{-1}[n]$ は、 n 番目の消費者が所属する第二のクラスタが、 $K-2$ 個のクラスタのうち、 $A_s^{-1}[n]$ 番目のクラスタであることを意味する。

[0093] $d(v_1, v_2)$ は、距離関数であり、ベクトル v_1 とベクトル v_2 との間の距離（例えばマハラノビス距離）を表す。 τ は、設計パラメータであり、ユーザにより、換言すれば分析者により定められる。

[0094] 評価関数 E の第一項及び第二項のそれぞれは、第一の特徴空間上の点 $D_f[A_f^{-1}[m]]$ を、写像としての関数 g により第二の特徴空間上の点に写したときの、第二の特徴空間上の点 $g(D_f[A_f^{-1}[m]])$ と、第二の特徴空間上の点 $D_s[A_s^{-1}[n]]$ との間の距離 $d(g(D_f[A_f^{-1}[m]]), D_s[A_s^{-1}[n]])$ の総和 $\sum d(g(D_f[A_f^{-1}[m]]), D_s[A_s^{-1}[n]])$ を表す。点 $D_f[A_f^{-1}[m]]$ は、第一のグループにおける m 番

目の消費者が所属する $A_f^{-1} [m]$ 番目の第一のクラスタの第一の特徴空間上の点 $D_f [A_f^{-1} [m]]$ である。点 $D_s [A_s^{-1} [n]]$ は、第二のグループにおける n 番目の消費者が所属する $A_s^{-1} [n]$ 番目の第二のクラスタの第二の特徴空間上の点 $D_s [A_s^{-1} [n]]$ である。

[0095] 但し、評価関数 E の第一項は、クラスタ関係表によれば、 $A_f^{-1} [m]$ 番目の第一のクラスタと $A_s^{-1} [n]$ 番目の第二のクラスタとの間に同じ消費者が存在する m, n の組合せの集合 $M1$ についての総和 $\sum d (g (D_f [A_f^{-1} [m]]), D_s [A_s^{-1} [n]])$ である。すなわち、同じ消費者の組合せ (m, n) についての総和 $\sum d (g (D_f [A_f^{-1} [m]]), D_s [A_s^{-1} [n]])$ である。

[0096] 評価関数 E の第二項は、クラスタ関係表によれば、 $A_f^{-1} [m]$ 番目の第一のクラスタと $A_s^{-1} [n]$ 番目の第二のクラスタとの間に同じ消費者が存在しない m, n の組合せの集合 $M2$ についての総和 $\sum d (g (D_f [A_f^{-1} [m]]), D_s [A_s^{-1} [n]])$ である。すなわち、異なる消費者の組合せ (m, n) についての総和 $\sum d (g (D_f [A_f^{-1} [m]]), D_s [A_s^{-1} [n]])$ である。

[0097] この評価関数 E は、同じ消費者を含む第一のクラスタと第二のクラスタとのペアである特定ペアに関しては、特定ペアを構成する第一のクラスタと第二のクラスタとの間の第二の特徴空間上の距離が短いほど、小さい値を出力し、特定ペア以外の第一のクラスタと第二のクラスタとのペアに関しては、第一のクラスタと第二のクラスタとの間の第二の特徴空間上の距離が短いほど、大きい値を出力する関数である。

[0098] 従って、評価関数 E を最小にする関数 g を探索することによれば、第一のデータセット 15 A から特定される第一の特徴空間上の共通消費者に対応する点を、第二のデータセット 15 B から特定される第二の特徴空間上の同一人物の点に近い位置に、最もよく写す写像を探索することができる。この写像によれば、非同一人物間の距離は、第二の特徴空間上で遠くなる。関数 g は、変換直交行列であってもよいし、多層パーセプトロンであってもよい。

- [0099] S 2 3 0における評価関数Eを最小にする関数gの探索後、プロセッサ11は、探索された関数gにより写されるK 1個の第一のクラスタの第二の特徴空間上の点 $g(D_f[A_f^{-1}[m]])$ と、K 2個の第二のクラスタの第二の特徴空間上の点 $D_s[A_s^{-1}[n]]$ との間の位置関係（換言すれば相対位置）に基づいて、複数の第一のクラスタのそれぞれを、第二のクラスタの一以上に対応付ける（S 2 4 0）。
- [0100] 例えば、プロセッサ11は、各点には、対応するクラスタのクラスタサイズに一致する人数の消費者が存在するとみなして、最適輸送問題を解くことにより、複数の第一のクラスタのそれぞれを、第二のクラスタの一以上に対応付ける。
- [0101] 例えば、プロセッサ11は、S 2 4 0において図7 Aに示す処理を実行することができる。この処理によれば、プロセッサ11は、クラスタ関係表に基づき、同じ消費者を含む第一のクラスタと第二のクラスタとのペアに関しては、第一のクラスタと第二のクラスタとの間の距離を特定値 ε に設定する（S 2 4 1）。値 ε は、例えばゼロ又はゼロに近い値であり得る。
- [0102] プロセッサ11は、同じ消費者を含まない第一のクラスタと第二のクラスタとのペアに関しては、特徴ベクトル $g(D_f[A_f^{-1}[m]])$ 及び特徴ベクトル $D_s[A_s^{-1}[n]]$ から特定される第一のクラスタと第二のクラスタとの間の第二の特徴空間上の距離をそのまま用いて、上記最適輸送問題を解くことにより、複数の第一のクラスタのそれぞれを、第二のクラスタの一以上に対応付ける（S 2 4 3）。
- [0103] あるいは、プロセッサ11は、S 2 4 0において図7 Bに示す処理を実行することができる。この処理によれば、プロセッサ11は、同じ消費者を含む第一のクラスタと第二のクラスタとのペアに関しては、ペア毎に、クラスタ特徴データを参照せず、一律に、第一のクラスタをペアに対応する第二のクラスタに対応付ける（S 2 4 5）。
- [0104] 同じ消費者を含まない第一のクラスタと第二のクラスタとのペアの一群に関しては、ペア毎に、クラスタ特徴データから特定される第一のクラスタの

第二の特徴空間上の点 $g(D_f[A_f^{-1}[m]])$ と、第二のクラスタの第二の特徴空間上の点 $D_s[A_s^{-1}[n]]$ との間の位置関係に基づいて、複数の第一のクラスタのそれぞれを、第二のクラスタの一以上に対応付ける (S 2 4 7)。

[0105] 例えば、プロセッサ 11 は、同じ消費者を含まない第一のクラスタと第二のクラスタとのペアの一群に限って、これに対する最適輸送問題を解くことにより、複数の第一のクラスタのそれぞれを、第二のクラスタの一以上に対応付ける (S 2 4 7)。

[0106] このようにして S 2 4 0 での処理を終了すると、プロセッサ 11 は、第一のデータセット 15 A に含まれる第一のクラスタ毎のクラスタ特徴データを、第二のデータセット 15 B に含まれる、S 2 4 0 の処理で対応付けられた一以上の第二のクラスタのクラスタ特徴データと結合するように、第一のデータセット 15 A と第二のデータセット 15 B とを結合する。この結合（すなわち、データフュージョン）により、プロセッサ 11 は、拡張データセット 15 C を生成し、ストレージ 15 に保存する (S 2 5 0)。その後、プロセッサ 11 は、分析処理を終了する。

[0107] 以上に説明した本実施形態の情報処理システム 1 によれば、第一のデータセット 15 A に対応する消費者の第一のグループと、第二のデータセット 15 B に対応する消費者の第二のグループとの間に共通する消費者の情報を活用し、クラスタに関する特徴データ間の対応付けを精度よく行うことができる。特には、同じ消費者を有する第一のクラスタと第二のクラスタとを対応付けるように、第一のクラスタ特徴データと、第二のクラスタ特徴データとを対応付けることができる。

[0108] ところで、対応付けの精度向上のためには、第一のデータ提供システム 3 1 及び第二のデータ提供システム 3 2 において、一つのクラスタに含まれる共通消費者が可能な限り少なくなるように、理想的には一つのクラスタに含まれる共通消費者が一人になるように、クラスタリング処理が実行されることが好ましい。

[0109] このために、第一のデータ提供システム 31 及び第二のデータ提供システム 32 では、k-means 法を改良した図 8A に示すクラスタリング処理が実行され得る。

[0110] すなわち、第一のデータ提供システム 31 のプロセッサ 311 は、クラスタリング処理として、次のペナルティ付き評価関数 H が最小となるように、第一のグループを所定数 K 個のクラスタにクラスタリングする処理 (S300) を実行することができる。

[0111] [数2]

$$H = \sum_m \sum_k I(m \in A[k]) W[m] d(C[k], D[m]) + \lambda \sum_{i,j \in Q} I(A^{-1}[i] == A^{-1}[j])$$

評価関数 H の第一項は、k-means 法に従う評価関数であり、評価関数 H の第二項は、一つのクラスタに複数の共通消費者が存在するときに、ペナルティ λ を加算する項である。

[0112] 評価関数 H における m は、消費者のインデックスであり、k は、クラスタのインデックスである。A[k] は、第 k クラスタを構成する消費者集合である。指示関数 I(m ∈ A[k]) は、第 m 消費者が第 k クラスタを構成する消費者集合に存在するとき、値 1 を返し、それ以外の場合には値 0 を返す関数である。W[m] は、第 m 消費者の重みである。重み W は 1 と理解されてもよい。

[0113] C[k] は、第 k クラスタの中心ベクトルであり、D[m] は、第 m 消費者の消費者特徴データから特定される第 m 消費者の特徴ベクトルである。A⁻¹[i] は、第 i 消費者が所属するクラスタが、K 個のクラスタのうち、第 A⁻¹[i] クラスタであることを示す。

[0114] 指示関数 I(A⁻¹[i] == A⁻¹[j]) は、共通消費者の集合 Q のうち、第 i 消費者と第 j 消費者とが同じクラスタに存在するとき値 1 を返し、第 i 消費者と第 j 消費者とが同じクラスタに存在しないとき値 0 を返す関数である。

- [0115] すなわち、評価関数Hの第二項は、共通消費者の集合Qであって、 $i \neq j$ である消費者の組合せについての値 $\| (A^{-1} [i]) - (A^{-1} [j]) \|^2$ の和にペナルティ λ を掛けた項である。ペナルティ λ は正の値として設計者により定められ得る。
- [0116] 第二のデータ提供システム32のプロセッサ321も、第一のデータ提供システム31と同様にペナルティ付き評価関数Hに従って、第二のグループを複数の第二のクラスタにクラスタリングすることができる。
- [0117] 別例として、第一のデータ提供システム31のプロセッサ311は、図8Bに示すクラスタリング処理を実行してもよい。
- [0118] 図8Bに示すクラスタリング処理によれば、プロセッサ311は、所定数K個の空のクラスタを設定する(S410)。その後、プロセッサ311は、消費者の第一のグループのうち、第二のグループと共通する共通消費者の一人を、対象消費者としてランダムに選択する(S420)。
- [0119] 続いて、プロセッサ311は、設定されたK個のクラスタの中に、共通消費者が割り当てられていない空のクラスタが存在するかを判断する(S430)。プロセッサ311は、空のクラスタが存在すると判断すると(S430でYes)、対象消費者を、空のクラスタの一つに割り当てる(S440)。
- [0120] 続くS450において、プロセッサ311は、共通消費者の全てをK個のクラスタのいずれかに割り当てたかを判断する。ここで否定判断すると(S450でNo)、プロセッサ311は、S420において新たに共通消費者を一人、対象消費者に選択し、選択した対象消費者についてS430以降の処理を実行する。
- [0121] このようにして、プロセッサ311は、共通消費者毎にS430以降の処理を実行し、共通消費者のそれぞれを、K個のクラスタのいずれかに割り当てる。
- [0122] クラスタ数より共通消費者数が少ない場合には、S430で否定判断されることなく、共通消費者がK個のクラスタのいずれかに割り当てられる。一

方、クラスタ数より共通消費者数が多い場合、K人の共通消費者を空のクラスタに割り当てた後、残りの共通消費者を新たに対象消費者に選択するとき、プロセッサ311は、S430で空のクラスタが存在しないと判断する。

[0123] S430で空のクラスタが存在しないと判断すると、プロセッサ311は、対象消費者を、次の評価関数 $R = d(C[k], D[i])$ が最小となるクラスタに割り当てる(S445)。

[0124] 評価関数Rにおける、 $C[k]$ は、第kクラスタの中心ベクトルであり、 $D[i]$ は、対象消費者である第i消費者の消費者特徴データから特定される第i消費者の特徴ベクトルである。 $d()$ は、距離関数である。

[0125] S445におけるクラスタの割当後、プロセッサ311は、S450の処理を実行する。S450において共通消費者の全てをK個のクラスタのいずれかに割り当てたと判断すると、プロセッサ311は、第一のグループにおける残りの非共通消費者のそれぞれを、k-means法と同様に、当該消費者の特徴ベクトルとクラスタの中心ベクトルとの間の距離の総和が最も短くなるように、近傍のクラスタに割り当てる(S460)。

[0126] プロセッサ311は、このようにして、第一のグループに属する消費者のそれぞれをK個のクラスタのいずれかに割り当てることにより、消費者の第一のグループをK個のクラスタにクラスタリングする。その後、図8Bに示すクラスタリング処理を終了する。

[0127] このクラスタリング処理によれば、共通消費者の数がクラスタ数より少ないときには、一つのクラスタに含まれる共通消費者が一人になるように、共通消費者を複数のクラスタに分散して配置することができる。従って、情報処理システム1における第一のクラスタと第二のクラスタとの対応付けを、共通消費者の情報に基づいて、高精度に行うことが可能である。

[0128] 以上に説明した本開示の技術は、上述した実施形態に限定されるものではなく、種々の態様を採り得る。

[0129] 上述した実施形態は、第一のデータセット15Aに含まれるクラスタ特徴データと、第二のデータセット15Bに含まれるクラスタ特徴データとの間

に、クラスタの特徴を説明する要素として共通する要素が存在しない場合でも、写像（関数 g ）の探索により、第一のクラスタと第二のクラスタとを対応付け可能である。

[0130] 共通要素が存在する場合、すなわち共通変数が存在する場合には、この共通変数のみに基づいて最適輸送問題を解くことにより、第一のクラスタと第二のクラスタとの間の対応付けを行ってもよい。第一のデータセット 15 A に含まれるクラスタ特徴データと、第二のデータセット 15 B に含まれるクラスタ特徴データとの間に、クラスタの特徴を説明する要素として共通する要素が存在する場合が考えられる。すなわち、第一のデータセット 15 A に含まれるクラスタ特徴データの特徴ベクトル $(x_i[1], x_i[2], x_i[3], \dots, x_i[L1])$ と、第二のデータセット 15 B に含まれるクラスタ特徴データの特徴ベクトル $(y_j[1], y_j[2], y_j[3], \dots, y_j[L2])$ との間に共通する変数が存在する場合が考えられる。

[0131] この場合、プロセッサ 11 は、図 6 に示す分析処理に代えて、図 9 に示す分析処理を実行することができる。図 9 に示す分析処理を開始すると、プロセッサ 11 は、S510において、指定された第一のデータセット 15 A をストレージ 15 から読み出す。プロセッサ 11 は、読み出した第一のデータセット 15 A に含まれる第一のクラスタ毎のクラスタ特徴データに基づき、第一のクラスタ毎に共通変数のみの特徴ベクトル $D_i^*[i] = (x_i[1], x_i[2], x_i[3], \dots, x_i[L0])$ を生成する。

[0132] ここでは、第一のデータセット 15 A において第一のクラスタの特徴を説明する $L1$ 個の要素のうちの $L0$ 個の要素 $x_i[1], x_i[2], x_i[3], \dots, x_i[L0]$ と、第二のデータセット 15 B において第二のクラスタの特徴を説明する $L2$ 個の要素のうちの $L0$ 個の要素 $y_j[1], y_j[2], y_j[3], \dots, y_j[L0]$ と、が共通変数であるとの前提を置く。

[0133] 続く S520において、プロセッサ 11 は、指定された第二のデータセット 15 B をストレージ 15 から読み出す。プロセッサ 11 は、読み出した第二のデータセット 15 B に含まれる第二のクラスタ毎のクラスタ特徴データ

に基づき、第二のクラスタ毎に共通変数のみの特徴ベクトル $D_s^*[j] = (y_j[1], y_j[2], y_j[3], \dots, y_j[L0])$ を生成する。

[0134] 続く S540 において、プロセッサ 11 は、第一のクラスタと第二のクラスタとの組合せ毎の、特徴ベクトル $D_f^*[i]$ と特徴ベクトル $D_s^*[j]$ との間の特徴空間上の距離に基づいて、最適輸送問題を解くことにより、複数の第一のクラスタのそれぞれを、第二のクラスタの一つ以上に対応付ける。

[0135] S540 において、プロセッサ 11 は、図 7A に示す処理と同様に、同じ消費者を含む第一のクラスタと第二のクラスタとのペアに関しては、第一のクラスタと第二のクラスタとの間の距離を特定値 ε に設定して、最適輸送問題を解くことができる。

[0136] あるいは、プロセッサ 11 は、同じ消費者を含む第一のクラスタと第二のクラスタとのペアに関しては、ペア毎に、クラスタ特徴データを参照せず、一律に、第一のクラスタをペアに対応する第二のクラスタに対応付けることができる。プロセッサ 11 は、同じ消費者を含まない第一のクラスタと第二のクラスタとのペアの一群に関しては、特徴ベクトル $D_f^*[i]$ と特徴ベクトル $D_s^*[j]$ との間の特徴空間上の距離に基づいて、最適輸送問題を解くことにより、複数の第一のクラスタのそれぞれを、第二のクラスタの一つ以上に対応付けることができる。

[0137] 続く S550 においてプロセッサ 11 は、S540 での対応付けに従って、第一のデータセット 15A に含まれる第一のクラスタ毎のクラスタ特徴データを、第二のデータセット 15B に含まれる、対応付けられた一つ以上の第二のクラスタのクラスタ特徴データと結合するように、第一のデータセット 15A と第二のデータセット 15B とを結合する。この結合により、拡張データセット 15C を生成し、これをストレージ 15 に保存する。その後、プロセッサ 11 は、分析処理を終了する。

[0138] 図 9 に示す分析処理によれば、共通変数に注目して負荷の低い手順で、第一のデータセット 15A と第二のデータセット 15B との間のデータフュージョンを実現することができる。

- [0139] 以上には、消費者の特徴データを含むデータセットの結合に、本開示の技術を適用した例を説明したが、消費者以外の、情報端末及び自動車为例に含む様々なエンティティに対して、本開示の技術は適用され得る。すなわち、エンティティは、消費者以外及び人以外であってもよい。
- [0140] 上記実施形態における1つの構成要素が有する機能は、複数の構成要素に分散して設けられてもよい。複数の構成要素が有する機能は、1つの構成要素に統合されてもよい。上記実施形態の構成の一部は、省略されてもよい。上記実施形態の構成の少なくとも一部は、他の上記実施形態の構成に対して付加又は置換されてもよい。請求の範囲に記載の文言から特定される技術思想に含まれるあらゆる態様が本開示の実施形態である。

請求の範囲

〔請求項1〕 エンティティの第一のグループをクラスタリングして定義される複数の第一のクラスタに関する第一のデータセットであって、第一のクラスタ毎に、対応する第一のクラスタの特徴を表す第一のクラスタ特徴データを備える第一のデータセットを取得するように構成される第一取得部と、

エンティティの第二のグループをクラスタリングして定義される複数の第二のクラスタに関する第二のデータセットであって、第二のクラスタ毎に、対応する第二のクラスタの特徴を表す第二のクラスタ特徴データを備える第二のデータセットを取得するように構成される第二取得部と、

前記第一のグループと前記第二のグループとの間で共通する一以上のエンティティである一以上の共通エンティティに関して、前記共通エンティティのそれぞれが属する第一のクラスタ及び第二のクラスタを識別可能な情報に基づき、前記第一のデータセットが備える複数の第一のクラスタ特徴データのそれぞれを、前記第二のデータセットが備える複数の第二のクラスタ特徴データのうちの一つ以上に対応付けるように構成される対応付け部と、

を備える情報処理システム。

〔請求項2〕 前記対応付け部は、前記複数の第一のクラスタのうちの一以上の特定クラスタに関しては、前記特定クラスタのそれぞれの前記第一のクラスタ特徴データを、同じエンティティを含む前記第二のクラスタの前記第二のクラスタ特徴データと対応付けるように構成され、

前記一以上の特定クラスタは、前記複数の第一のクラスタのうち、それぞれが前記共通エンティティの少なくとも一つを含む一以上の第一のクラスタである請求項1記載の情報処理システム。

〔請求項3〕 前記対応付け部は、

一以上の非共通第一クラスタ及び一以上の非共通第二クラスタに関

しては、前記一以上の非共通第一クラスタのそれぞれの前記第一のクラスタ特徴データと前記一以上の非共通第二クラスタのそれぞれの前記第二のクラスタ特徴データとに基づき、前記非共通第一クラスタのそれぞれと前記非共通第二クラスタのそれぞれとの間の関係を判別し、

判別された前記関係に基づいて、前記非共通第一クラスタのそれぞれの前記第一のクラスタ特徴データを、前記非共通第二クラスタの少なくとも一つの前記第二のクラスタ特徴データに対応付けるように構成され、

前記一以上の非共通第一クラスタは、前記複数の第一のクラスタのうち、前記特定クラスタ以外の残りの一以上のクラスタであり、

前記一以上の非共通第二クラスタは、前記複数の第二のクラスタのうち、前記特定クラスタと対応付けられるクラスタ以外の残りの一以上のクラスタである請求項2記載の情報処理システム。

[請求項4]

前記第一のクラスタ特徴データは、前記対応する第一のクラスタの第一の特徴空間上の点を定義し、

前記第二のクラスタ特徴データは、前記対応する第二のクラスタの第二の特徴空間上の点を定義し、

前記対応付け部は、

前記第一の特徴空間上の前記複数の第一のクラスタのそれぞれの点を前記第二の特徴空間に写す写像を、評価関数を用いて探索する処理と、

前記写像により写される前記第二の特徴空間上の前記複数の第一のクラスタのそれぞれの点と、前記第二の特徴空間上の前記複数の第二のクラスタのそれぞれの点との間の位置関係に基づいて、前記複数の第一のクラスタ特徴データのそれぞれを、前記複数の第二のクラスタ特徴データのうちの一つ以上に対応付ける処理と、

を実行し、

前記評価関数は、

同じエンティティを含む前記第一のクラスタと前記第二のクラスタとのペアである特定ペアに関しては、前記特定ペアを構成する前記第一のクラスタと前記第二のクラスタとの間の前記第二の特徴空間上の距離が短いほど、特定方向に変化する値を出力し、

前記特定ペア以外の前記第一のクラスタと前記第二のクラスタとのペアに関しては、前記第一のクラスタと前記第二のクラスタとの間の前記第二の特徴空間上の距離が短いほど、前記特定方向とは逆方向に変化する値を出力する

ように設計される関数である請求項 1 又は請求項 2 記載の情報処理システム。

[請求項5]

前記第一のクラスタ特徴データは、前記対応する第一のクラスタの第一の特徴空間上の点を定義し、

前記第二のクラスタ特徴データは、前記対応する第二のクラスタの第二の特徴空間上の点を定義し、

前記対応付け部は、

前記第一の特徴空間上の前記特定クラスタを含む前記複数の第一のクラスタのそれぞれの点を前記第二の特徴空間に写す写像を、評価関数を用いて探索する処理と、

前記特定クラスタのそれぞれの前記第一のクラスタ特徴データを、同じエンティティを含む前記第二のクラスタの前記第二のクラスタ特徴データと対応付ける処理と、

前記写像により写される前記第二の特徴空間上の前記非共通第一クラスタのそれぞれの点と、前記第二の特徴空間上の前記非共通第二クラスタのそれぞれの点との間の位置関係に基づいて、前記非共通第一クラスタのそれぞれの前記第一のクラスタ特徴データを、前記非共通第二クラスタの少なくとも一つの前記第二のクラスタ特徴データに対応付ける処理と、

を実行し、

前記評価関数は、

同じエンティティを含む前記第一のクラスタと前記第二のクラスタとのペアである特定ペアに関しては、前記特定ペアを構成する前記第一のクラスタと前記第二のクラスタとの間の前記第二の特徴空間上の距離が短いほど、特定方向に変化する値を出力し、

前記特定ペア以外の前記第一のクラスタと前記第二のクラスタとのペアに関しては、前記第一のクラスタと前記第二のクラスタとの間の前記第二の特徴空間上の距離が短いほど、前記特定方向とは逆方向に変化する値を出力する

ように設計される関数である請求項3記載の情報処理システム。

[請求項6]

前記評価関数は、

前記特定ペアに関しては、前記特定ペアを構成する前記第一のクラスタと前記第二のクラスタとの間の前記第二の特徴空間上の距離が短いほど、小さい値を出力し、

前記特定ペア以外の前記第一のクラスタと前記第二のクラスタとのペアに関しては、前記第一のクラスタと前記第二のクラスタとの間の前記第二の特徴空間上の距離が短いほど、大きい値を出力する

ように設計され、

前記対応付け部は、前記評価関数の出力を最小にする前記写像を探索する請求項4又は請求項5記載の情報処理システム。

[請求項7]

前記一以上の共通エンティティは、複数の共通エンティティであり、

前記複数の第一のクラスタは、前記複数の共通エンティティが前記複数の第一のクラスタに分散して配置されるように、前記第一のグループをクラスタリングして定義されるクラスタの集合であり、

前記複数の第二のクラスタは、前記複数の共通エンティティが前記複数の第二のクラスタに分散して配置されるように、前記第二のグル

ープをクラスタリングして定義されるクラスタの集合である請求項 1～請求項 6 のいずれか一項記載の情報処理システム。

[請求項 8] 前記一以上の共通エンティティは、複数の共通エンティティであり、

前記複数の第一のクラスタは、前記複数の共通エンティティのそれぞれが互いに異なる第一のクラスタに属するように、前記第一のグループをクラスタリングして定義されるクラスタの集合であり、

前記複数の第二のクラスタは、前記複数の共通エンティティのそれぞれが互いに異なる第二のクラスタに属するように、前記第二のグループをクラスタリングして定義されるクラスタの集合である請求項 1～請求項 6 のいずれか一項記載の情報処理システム。

[請求項 9] 前記第一のクラスタ特徴データは、前記対応する第一のクラスタに属する複数のエンティティの特徴データを統計処理により統合して生成される特徴データであり、

前記第二のクラスタ特徴データは、前記対応する第二のクラスタに属する複数のエンティティの特徴データを統計処理により統合して生成される特徴データである請求項 1～請求項 8 のいずれか一項記載の情報処理システム。

[請求項 10] 前記エンティティは、消費者であり、前記エンティティの前記特徴データは、前記消費者の特徴を説明する特徴データである請求項 9 記載の情報処理システム。

[請求項 11] コンピュータにより実行される情報処理方法であって、

エンティティの第一のグループをクラスタリングして定義される複数の第一のクラスタに関し、第一のクラスタ毎に、対応する第一のクラスタの特徴を表す第一のクラスタ特徴データを備える第一のデータセットを取得することと、

エンティティの第二のグループをクラスタリングして定義される複数の第二のクラスタに関し、第二のクラスタ毎に、対応する第二のク

ラストの特徴を表す第二のクラスタ特徴データを備える第二のデータセットを取得することと、

前記第一のグループと前記第二のグループとの間で共通する一以上のエンティティである一以上の共通エンティティに関して、前記共通エンティティのそれぞれが属する第一のクラスタ及び第二のクラスタを識別可能な情報に基づき、前記第一のデータセットが備える複数の第一のクラスタ特徴データのそれぞれを、前記第二のデータセットが備える複数の第二のクラスタ特徴データのうちの一つ以上に対応付けることと、

を含む情報処理方法。

[請求項12] 前記対応付けることは、前記複数の第一のクラスタのうちの一以上の特定クラスタに関しては、前記特定クラスタのそれぞれの前記第一のクラスタ特徴データを、同じエンティティを含む前記第二のクラスタの前記第二のクラスタ特徴データと対応付けることを含み、

前記一以上の特定クラスタは、前記複数の第一のクラスタのうち、それぞれが前記共通エンティティの少なくとも一つを含む一以上の第一のクラスタである請求項1記載の情報処理方法。

[請求項13] 前記対応付けることは、
一以上の非共通第一クラスタ及び一以上の非共通第二クラスタに関しては、前記一以上の非共通第一クラスタのそれぞれの前記第一のクラスタ特徴データと前記一以上の非共通第二クラスタのそれぞれの前記第二のクラスタ特徴データとに基づき、前記非共通第一クラスタのそれぞれと前記非共通第二クラスタのそれぞれとの間の関係を判別し、

判別された前記関係に基づいて、前記非共通第一クラスタのそれぞれの前記第一のクラスタ特徴データを、前記非共通第二クラスタの少なくとも一つの前記第二のクラスタ特徴データに対応付けること
を含み、

前記一以上の非共通第一クラスタは、前記複数の第一のクラスタのうち、前記特定クラスタ以外の残りの一以上のクラスタであり、

前記一以上の非共通第二クラスタは、前記複数の第二のクラスタのうち、前記特定クラスタと対応付けられるクラスタ以外の残りの一以上のクラスタである請求項 1 2 記載の情報処理方法。

[請求項14]

前記第一のクラスタ特徴データは、前記対応する第一のクラスタの第一の特徴空間上の点を定義し、

前記第二のクラスタ特徴データは、前記対応する第二のクラスタの第二の特徴空間上の点を定義し、

前記対応付けることは、

前記第一の特徴空間上の前記複数の第一のクラスタのそれぞれの点を前記第二の特徴空間に写す写像を、評価関数を用いて探索することと、

前記写像により写される前記第二の特徴空間上の前記複数の第一のクラスタのそれぞれの点と、前記第二の特徴空間上の前記複数の第二のクラスタのそれぞれの点との間の位置関係に基づいて、前記複数の第一のクラスタ特徴データのそれぞれを、前記複数の第二のクラスタ特徴データのうちの一つ以上に対応付けることと、

を含み、

前記評価関数は、

同じエンティティを含む前記第一のクラスタと前記第二のクラスタとのペアである特定ペアに関しては、前記特定ペアを構成する前記第一のクラスタと前記第二のクラスタとの間の前記第二の特徴空間上の距離が短いほど、特定方向に変化する値を出力し、

前記特定ペア以外の前記第一のクラスタと前記第二のクラスタとのペアに関しては、前記第一のクラスタと前記第二のクラスタとの間の前記第二の特徴空間上の距離が短いほど、前記特定方向とは逆方向に変化する値を出力する

ように設計される関数である請求項 1 1 又は請求項 1 2 記載の情報処理方法。

[請求項15] 前記第一のクラスタ特徴データは、前記対応する第一のクラスタの第一の特徴空間上の点を定義し、

前記第二のクラスタ特徴データは、前記対応する第二のクラスタの第二の特徴空間上の点を定義し、

前記対応付けることは、

前記第一の特徴空間上の前記特定クラスタを含む前記複数の第一のクラスタのそれぞれの点を前記第二の特徴空間に写す写像を、評価関数を用いて探索することと、

前記特定クラスタのそれぞれの前記第一のクラスタ特徴データを、同じエンティティを含む前記第二のクラスタの前記第二のクラスタ特徴データと対応付けることと、

前記写像により写される前記第二の特徴空間上の前記非共通第一クラスタのそれぞれの点と、前記第二の特徴空間上の前記非共通第二クラスタのそれぞれの点との間の位置関係に基づいて、前記非共通第一クラスタのそれぞれの前記第一のクラスタ特徴データを、前記非共通第二クラスタの少なくとも一つの前記第二のクラスタ特徴データに対応付けることと、

を含み、

前記評価関数は、

同じエンティティを含む前記第一のクラスタと前記第二のクラスタとのペアである特定ペアに関しては、前記特定ペアを構成する前記第一のクラスタと前記第二のクラスタとの間の前記第二の特徴空間上の距離が短いほど、特定方向に変化する値を出力し、

前記特定ペア以外の前記第一のクラスタと前記第二のクラスタとのペアに関しては、前記第一のクラスタと前記第二のクラスタとの間の前記第二の特徴空間上の距離が短いほど、前記特定方向とは逆方向に

変化する値を出力する

ように設計される関数である請求項 13 記載の情報処理方法。

[請求項16]

前記評価関数は、

前記特定ペアに関しては、前記特定ペアを構成する前記第一のクラスタと前記第二のクラスタとの間の前記第二の特徴空間上の距離が短いほど、小さい値を出力し、

前記特定ペア以外の前記第一のクラスタと前記第二のクラスタとのペアに関しては、前記第一のクラスタと前記第二のクラスタとの間の前記第二の特徴空間上の距離が短いほど、大きい値を出力する

ように設計され、

前記対応付けることは、前記評価関数の出力を最小にする前記写像を探索することを含む請求項 14 又は請求項 15 記載の情報処理方法。

[請求項17]

前記一以上の共通エンティティは、複数の共通エンティティであり、

前記複数の第一のクラスタは、前記複数の共通エンティティが前記複数の第一のクラスタに分散して配置されるように、前記第一のグループをクラスタリングして定義されるクラスタの集合であり、

前記複数の第二のクラスタは、前記複数の共通エンティティが前記複数の第二のクラスタに分散して配置されるように、前記第二のグループをクラスタリングして定義されるクラスタの集合である請求項 11～請求項 16 のいずれか一項記載の情報処理方法。

[請求項18]

前記一以上の共通エンティティは、複数の共通エンティティであり、

前記複数の第一のクラスタは、前記複数の共通エンティティのそれぞれが互いに異なる第一のクラスタに属するように、前記第一のグループをクラスタリングして定義されるクラスタの集合であり、

前記複数の第二のクラスタは、前記複数の共通エンティティのそれ

ぞれが互いに異なる第二のクラスタに属するように、前記第二のグループをクラスタリングして定義されるクラスタの集合である請求項 11～請求項 16 のいずれか一項記載の情報処理方法。

[請求項19]

前記対応付けることは、

互いに共通するエンティティを有する第一のクラスタ及び第二のクラスタのペアである第一のペア、及び、互いに共通するエンティティを有さない第一のクラスタ及び第二のクラスタのペアである第二のペアのうち、前記第一のペアに対応する第一のクラスタ特徴データと第二のクラスタ特徴データとを、前記第二のペアより優先的に対応付けるように、前記複数の第一のクラスタ特徴データのそれぞれを、前記複数の第二のクラスタ特徴データのうちの一つ以上に対応付け、

少なくとも前記第二のペアに関しては、前記第二のペアに対応する第一のクラスタ特徴データ及び第二のクラスタ特徴データから判別される前記第一のクラスタと前記第二のクラスタとの間の特徴空間上の位置関係に基づき、互いに対応付ける前記第一のクラスタと前記第二のクラスタとのペアを判別し、前記第一のクラスタ特徴データと前記第二のクラスタ特徴データとを対応付けること

を含む請求項 11 記載の情報処理方法。

[請求項20]

前記対応付けることは、前記複数の第一のクラスタのうち、前記特定クラスタ以外の非特定クラスタに関しては、前記第一のクラスタ特徴データ及び前記第二のクラスタ特徴データから判別される前記第一のクラスタと前記第二のクラスタとの間の特徴空間上の位置関係に基づき、互いに対応付ける前記第一のクラスタと前記第二のクラスタとのペアを判別し、前記第一のクラスタ特徴データと前記第二のクラスタ特徴データとを対応付けることを含む請求項 12 記載の情報処理方法。

[請求項21]

コンピュータにより実行されたときに、請求項 11～請求項 20 のいずれか一項記載の情報処理方法をコンピュータに実行させるための

命令を含むコンピュータプログラムを記憶するコンピュータ読取可能な記録媒体。