

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、2021年11月9日に日本国特許庁に出願された日本国特許出願第2021-182537号及び2022年3月24日に日本国特許庁に出願された日本国特許出願第2022-048893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2021-182537号及び日本国特許出願第2022-048893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] 従来、商品の販売データに基づき顧客の購買行動を分析することが行われている。顧客によるマスメディアやネットワークコンテンツへの接触行動を分析することも行われている。アンケート形式や対面での質問形式により、顧客に関する多様な情報を収集することも行われている。

[0004] 異なる手段で収集した複数のデータを共通変数に基づいて結合するデータフュージョン技術も知られている。特には、複数の第一の顧客に関して、顧客毎の第一の特徴データを備える第一のデータセットと、複数の第二の顧客に関して、顧客毎の第二の特徴データを備える第二のデータセットとの間のデータフュージョンに関する技術が既に開示されている（例えば、特許文献1参照）。データフュージョンは、第一のデータセットと第二のデータセットとの間で共通する変数、例えば、顧客のデモグラフィック属性に基づいて、近しい顧客の第一の特徴データと第二の特徴データとを結合するように行われる。

先行技術文献

特許文献

[0005] 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

[0006] 従来のデータフュージョン技術では、共通変数を用いて近しい顧客を判別するために、結合対象の第一のデータセットと第二のデータセットとの間に、顧客に関する共通変数が必要である。従って、共通変数のないデータ同士を結合することができない。

[0007] そこで、本開示の一側面によれば、共通変数によらずに、複数の第一のエンティティに関する第一のデータセットと、複数の第二のエンティティに関する第二のデータセットとに基づいて、第一のエンティティと第二のエンティティとの対応付けを実現可能な技術を提供することが望ましい。

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、次元削減部と、対応付け部と、を備える。第一取得部は、複数の第一のエンティティに関する第一のデータセットを取得するように構成される。第一のデータセットは、複数の第一のエンティティのそれぞれの特徴を記述し得る。

[0009] 第二取得部は、複数の第二のエンティティに関する第二のデータセットを取得するように構成される。第二のデータセットは、複数の第二のエンティティのそれぞれの特徴を記述し得る。

[0010] 次元削減部は、第一のデータセットから特定される第一の特徴ベクトルの一群、及び、第二のデータセットから特定される第二の特徴ベクトルの一群に対する次元削減処理を実行することによって、第一の特徴ベクトルの一群に対応する第一の低次元特徴ベクトルの一群、及び、第二の特徴ベクトルの一群に対応する第二の低次元特徴ベクトルの一群を生成するように構成される。第二の低次元特徴ベクトルの一群は、第一の低次元特徴ベクトルの一群と同次元数の特徴ベクトルの一群であり得る。

[0011] 第一の特徴ベクトルのそれぞれは、複数の第一のエンティティのうちの対

応する一つのエンティティの特徴を表し得る。第二の特徴ベクトルのそれぞれは、複数の第二のエンティティのうちの対応する一つのエンティティの特徴を表し得る。

[0012] 対応付け部は、第一の低次元特徴ベクトルの一群、及び、第二の低次元特徴ベクトルの一群に基づき、複数の第一のエンティティのそれぞれを、複数の第二のエンティティの少なくとも一つに対応付けるように構成される。

[0013] 第一のエンティティの集合及び第二のエンティティの集合が互いに共通する母集団又は互いに関係する母集団からの部分集合である場合には、第一の特徴ベクトルと第二の特徴ベクトルとの間に共通変数が存在しなくても、次元削減によって、第一のエンティティの特徴及び第二のエンティティの特徴を、互いに共通する又は関係する成分の組合せで表現することができる。

[0014] すなわち、次元削減によれば、第一の特徴ベクトル及び第二の特徴ベクトルから互いに共通する又は関係する主要な特徴成分を抽出することができる。従って、低次元特徴ベクトルの比較によれば、第一のエンティティと第二のエンティティとの間の適合度を、適切に判別することができる。

[0015] 従って、本開示の一側面によれば、共通変数によらずに、複数の第一のエンティティに関する第一のデータセットと、複数の第二のエンティティに関する第二のデータセットとに基づいて、第一のエンティティと第二のエンティティとの対応付けを適切に行うことができる。

[0016] 本開示の一側面によれば、対応付け部は、第一の低次元特徴ベクトルの一群から特定される第一のエンティティ間の類似度、及び、第二の低次元特徴ベクトルの一群から特定される第二のエンティティ間の類似度に基づき、類似度に関する第一のエンティティ間の相互関係が第二のエンティティ間の相互関係に適合するように、複数の第一のエンティティのそれぞれを、複数の第二のエンティティの少なくとも一つに対応付けることができる。

[0017] 第一のエンティティの集合、及び、第二のエンティティの集合が共通する母集団又は互いに関係する母集団からの部分集合である場合には、エンティティ間の類似度に関する相互関係が、母集団と同様に、第一のエンティティ

の集合と、第二のエンティティの集合との間で、およそ互いに共通する又は関係する。

[0018] 従って、類似度に関する第一のエンティティ間の相互関係が第二のエンティティ間の相互関係に適合するように、複数の第一のエンティティのそれぞれを、複数の第二のエンティティの少なくとも一つに対応付けることによれば、第一のエンティティのそれぞれを、同一性の高い又は関係性の強い適切な第二のエンティティに対応付けることができる。

[0019] 本開示の一側面によれば、第一の低次元特徴ベクトルの一群は、第一の特徴空間によって定義され得る。第二の低次元特徴ベクトルの一群は、第二の特徴空間によって定義され得る。

[0020] 対応付け部は、第一の低次元特徴ベクトルの一群から特定される第一の特徴空間における複数の第一のエンティティの分布が、第二の低次元特徴ベクトルの一群から特定される第二の特徴空間における複数の第二のエンティティの分布に適合するように、第一の特徴空間上の複数の第一のエンティティを第二の特徴空間にマッピングするための写像を探索し得る。

[0021] 対応付け部は、写像に基づいて、複数の第一のエンティティのそれぞれを、複数の第二のエンティティの少なくとも一つに対応付けるように構成され得る。

[0022] 本開示の一側面によれば、対応付け部は、行列K、行列L、及び行列Hを含む式

[0023] [数1]

$$\begin{aligned}\Omega^* &= \operatorname{argmax}_{\Omega} Z(\Omega) \\ Z(\Omega) &= \operatorname{trace}(K' \Omega^T L' \Omega) \\ K' &= HKH \\ L' &= HLH\end{aligned}$$

[0024] に従う値Z（ Ω ）を最大化する行列 Ω を行列 Ω^* として探索し、行列 Ω^* に基づいて、複数の第一のエンティティのそれぞれを、複数の第二のエンティティの少なくとも一つに対応付けるように構成され得る。Tは、転置記号で

ある。trace は、行列 X の対角和である。

[0025] 行列 K は、 N 行 N 列の行列であり得る。第一のエンティティの数は、 N であり得る。第二のエンティティの数は、第一のエンティティと同じであり得る。行列 K は、第 i 行第 j 列の要素の値が、複数の第一のエンティティのうちの i 番目のエンティティと j 番目のエンティティとの間の類似度を表す第一の類似度行列であり得る。

[0026] 行列 K における第 i 行第 j 列の要素の値は、複数の第一のエンティティのうちの i 番目のエンティティの第一の低次元特徴ベクトルと、複数の第一のエンティティのうちの j 番目のエンティティの第一の低次元特徴ベクトルと、に基づいて算出され得る。

[0027] 行列 L は、 N 行 N 列の行列であり得る。行列 L は、第 i 行第 j 列の要素の値が、複数の第二のエンティティのうちの i 番目のエンティティと j 番目のエンティティとの間の類似度を表す第二の類似度行列である。

[0028] 行列 L における第 i 行第 j 列の要素の値は、複数の第二のエンティティのうちの i 番目のエンティティの第二の低次元特徴ベクトルと、複数の第二のエンティティのうちの j 番目のエンティティの第二の低次元特徴ベクトルと、に基づいて算出され得る。

[0029] 行列 H は、 N 行 N 列の行列であり得る。行列 H は、第 i 行第 j 列の要素の値が、 $i = j$ であるとき値 $1 - 1/N$ を示し、 $i \neq j$ であるとき値 0 を示す行列であり得る。

[0030] 本開示の一側面によれば、対応付け部は、行列 Ω^* に基づいて、次元削減処理での次元削減方式を変更してもよい。例えば、対応付け部は、第一の低次元特徴ベクトルの一群及び第二の低次元特徴ベクトルの一群のうち、互いに対応する第一の低次元特徴ベクトルと第二の低次元特徴ベクトルとの間の特徴空間上の距離が短くなるように、次元削減処理での次元削減方式を変更してもよい。

[0031] 本開示の一側面によれば、対応付け部は、行列 Ω^* に関する再探索処理を、所定条件が満足されるまで繰返し実行することにより、行列 Ω^* を改善し、改

善された行列 Ω^* に基づいて、複数の第一のエンティティのそれぞれを、複数の第二のエンティティの少なくとも一つに対応付けるように構成されてもよい。

[0032] 再探索処理は、行列 Ω^* に基づいて、次元削減処理での次元削減方式を変更することを含み得る。再探索処理は、変更後の次元削減方式での次元削減処理を、次元削減部に実行させ、それにより新たに得られた第一の低次元特徴ベクトルの一群及び第二の低次元特徴ベクトルの一群に基づいて、行列 Ω^* を再探索することを含み得る。

[0033] 対応付け部がこのように構成される情報処理システムによれば、第一のエンティティと第二のエンティティとの間の対応付けを、高精度に実行することができる。

[0034] 本開示の一側面によれば、第一のデータセットは、複数の第一の特徴データを含み得る。複数の第一の特徴データのそれぞれは、複数の第一のエンティティのうちの対応する一つのエンティティの特徴を表し得る。第二のデータセットは、複数の第二の特徴データを含み得る。複数の第二の特徴データのそれぞれは、複数の第二のエンティティのうちの対応する一つのエンティティの特徴を表し得る。

[0035] 本開示の一側面によれば、情報処理システムは、データフュージョン部を更に備え得る。データフュージョン部は、対応付け部による複数の第一のエンティティと複数の第二のエンティティとの間の対応付けに基づき、複数の第一の特徴データのそれぞれに、複数の第二の特徴データのうちの一つを結合することによって、拡張データセットを生成するように構成され得る。拡張データセットは、複数の拡張データを備え得る。複数の拡張データのそれぞれは、対応する一つの第一の特徴データと第二の特徴データとの結合データであり得る。

[0036] こうした情報処理システムによれば、複数のデータセットを結合した情報量の多いデータセットを生成することが可能である。

[0037] 本開示の一側面によれば、第一のエンティティは、人であり得る。第二の

エンティティは、人であり得る。第一のデータセットは、第一の集団に属する複数の人のそれぞれの第一の特徴を記述するデータセットであり得る。第二のデータセットは、第二の集団の属する複数の人のそれぞれの第二の特徴を記述するデータセットであり得る。

[0038] 人の行動や関心等に関する特徴は、デモグラフィック属性に大きく影響し、異なる人の集団の間でも、デモグラフィック属性に応じた特徴分布は、大きく変わらないと考えられる。従って、本開示の一側面に係る情報処理システムによれば、共通変数なしで、異なる集団間の人に関する対応付けを、適切に行うことが可能である。

[0039] 本開示の一側面によれば、第一の特徴と第二の特徴との組合せは、購買行動に関する特徴と、オンライン空間及びオフライン空間の少なくとも一方の空間における移動に関する特徴、及び／又は、空間上の複数の地点への訪問に関する特徴と、の組合せであり得る。こうした特徴に係るデータセットに基づいたエンティティの対応付けは、更には、データフュージョンは、人の行動分析に役立つ。

[0040] 本開示の一側面によれば、第二のデータセットには、複数の第二のエンティティのそれぞれに対応する情報端末の識別情報が関連付けられていてもよい。

[0041] 本開示の一側面によれば、情報処理システムは、複数の第二のエンティティのうち、対応付け部により複数の第一のエンティティのいずれかと対応付けられた第二のエンティティの集合の少なくとも一部を、情報コンテンツの配信先に選択する選択部を備えていてもよい。

[0042] 本開示の一側面によれば、情報処理システムは、上記識別情報に基づいて、情報コンテンツの配信先に対応する情報端末の集合に、情報コンテンツを配信するように構成される配信部を備えていてもよい。

[0043] この情報処理システムは、第一のエンティティ及び第二のエンティティが人であるときに有意義に機能する。上述の配信方式によれば、第一のエンティティと情報端末との関係が不明であるときにも、第二のエンティティに関

連付けられた情報端末の識別情報を活用して、第一のエンティティに対応する第二のエンティティの情報端末に適切に情報コンテンツを配信することができる。

[0044] 本開示の一側面によれば、選択部は、対応付け部により複数の第一のエンティティのいずれかと対応付けられた第二のエンティティの集合である第一の集合と、複数の第二のエンティティのうち、第一の集合と特徴が類似する第二の集合と、を情報コンテンツの配信先を選択するように構成されてもよい。こうした配信先の選択によれば、第二のデータセットに基づいて、配信先を適切な範囲で拡張して、情報コンテンツを配信することができる。

[0045] 本開示の一側面によれば、第二のデータセットは、複数の第二のエンティティのそれぞれの行動に関する特徴を記述するデータセットであってもよい。この場合、情報処理システムは、一以上の注目エンティティに関して、注目エンティティ毎に、対応する注目エンティティの行動に関する推定値を算出する推定部を備えていてもよい。一以上の注目エンティティは、複数の第一のエンティティの少なくとも一部であり得る。推定値は、対応する注目エンティティに対応付けられた複数の第二のエンティティの少なくとも一つの行動に関する特徴に基づき算出され得る。第一のエンティティ及び第二のエンティティは、人であり得る。

[0046] 上述の推定部を備える情報処理システムによれば、第一のデータセットだけでは判別できない第一のエンティティの行動を、第二のデータセットを通じて推定することが可能である。推定は、予測であってもよい。

[0047] 本開示の一側面によれば、上述した情報処理システムが実行する方法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。情報処理方法は、複数の第一のエンティティに関する第一のデータセットであって、複数の第一のエンティティのそれぞれの特徴を記述する第一のデータセットを取得することを含み得る。

[0048] 情報処理方法は、複数の第二のエンティティに関する第二のデータセット

であって、複数の第二のエンティティのそれぞれの特徴を記述する第二のデータセットを取得することを含み得る。

[0049] 情報処理方法は、第一のデータセットから特定される第一の特徴ベクトルの一群、及び、第二のデータセットから特定される第二の特徴ベクトルの一群に対する次元削減処理を実行することによって、第一の特徴ベクトルの一群に対応する第一の低次元特徴ベクトルの一群、及び、第二の特徴ベクトルの一群に対応する第二の低次元特徴ベクトルの一群を生成することを含み得る。第二の低次元特徴ベクトルの一群は、第一の低次元特徴ベクトルの一群と同次元数の特徴ベクトルの一群であり得る。

[0050] 第一の特徴ベクトルのそれぞれは、複数の第一のエンティティのうちの対応する一つのエンティティの特徴を表し得る。第二の特徴ベクトルのそれぞれは、複数の第二のエンティティのうちの対応する一つのエンティティの特徴を表し得る。

[0051] 情報処理方法は、第一の低次元特徴ベクトルの一群、及び、第二の低次元特徴ベクトルの一群に基づき、複数の第一のエンティティのそれぞれを、複数の第二のエンティティの少なくとも一つに対応付けることを含み得る。

[0052] 本開示の一側面によれば、対応付けることは、第一の低次元特徴ベクトルの一群から特定される第一のエンティティ間の類似度、及び、第二の低次元特徴ベクトルの一群から特定される第二のエンティティ間の類似度に基づき、類似度に関する第一のエンティティ間の相互関係が第二のエンティティ間の相互関係に適合するように、複数の第一のエンティティのそれぞれを、複数の第二のエンティティの一つに対応付けることを含み得る。

[0053] こうした情報処理方法によれば、上述した情報処理システムと同様に、共通変数によらずに、複数の第一のエンティティに関する第一のデータセットと、複数の第二のエンティティに関する第二のデータセットとに基づいて、第一のエンティティと第二のエンティティとの対応付けを実現可能である。

[0054] 本開示の一側面によれば、コンピュータに上述した情報処理方法を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。本開示

の一側面によれば、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一時的でない有形の記録媒体が提供されてもよい。

図面の簡単な説明

[0055] [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]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3]図3 Aは、第一のデータセットの構成を例示する図であり、図3 Bは、第二のデータセットの構成を例示する図である。

[図4]図4 A及び図4 Bは、行列 Ω の探索手法を説明する図である。

[図5]プロセッサにより生成される対応表の構成を例示する図である。

[図6]プロセッサにより生成される拡張データセットの構成を例示する図である。

[図7]第二実施形態で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8]第三実施形態でプロセッサが実行する評価処理を表すフローチャートである。

[図9]第三実施形態でプロセッサが実行する選択処理を表すフローチャートである。

[図10]第四実施形態の配信システムの構成を表すブロック図である。

[図11]第四実施形態における内部データセットの構成を例示する図である。

[図12]第四実施形態でプロセッサが実行する配信制御処理を表すフローチャートである。

[図13]第五実施形態でプロセッサが実行する配信制御処理を表すフローチャートである。

[図14]第六実施形態でプロセッサが実行する予測処理を表すフローチャートである。

符号の説明

[0056] 1…情報処理システム、11、31…プロセッサ、13、33…メモリ、15、35…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、１５Ｃ…拡張データセット、１７…ユーザインタフェース、１９、３９…通信インタフェース、３０…配信システム、３５Ａ…外部データセット、３５Ｂ…内部データセット、３５Ｃ…拡張データセット、４０…利用企業側システム、Pr, Pr1…コンピュータプログラム。

発明を実施するための形態

[0057] 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。

[0058] <第一実施形態>

本実施形態の情報処理システム１は、汎用コンピュータに専用のコンピュータプログラムPrをインストールすることによって構成される。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。

[0059] 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り、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。

[0060] 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り、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。

[0061] ユーザインタフェース１７は、入力デバイスと、ディスプレイとを備える。入力デバイスは、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に入力するために設けられる。ディスプレイは、ユーザに対して各種情報を表示するために設けられる。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。

[0062] 通信インタフェース１９は、LAN（ローカルエリアネットワーク）インタフェース及びUSB（ユニバーサル・シリアル・シリアル）インタフェースを含み、外部装置との通信に使用される。情報処理システム１は、通信インタフェース１９を通じて外部装置との間でデータ送受する。

- [0063] 情報処理システム 1 におけるプロセッサ 11 は、コンピュータプログラム Pr に従う処理の実行により、通信インタフェース 19 を通じて外部装置から取得した第一のデータセット 15 A を、第二のデータセット 15 B を用いて拡張した拡張データセット 15 C を生成する。
- [0064] 拡張データセット 15 C は、第一のデータセット 15 A に対して、第二のデータセット 15 B が備える情報を付加したデータセットである。拡張により、第一のデータセット 15 A が記述する各エンティティの情報量は増大する。エンティティは、例えば人、特には個人である。情報量の増大は、拡張データセット 15 C に基づいた人の行動分析や広告配信のために実行される。
- [0065] 具体的に、情報処理システム 1 のプロセッサ 11 は、ユーザインタフェース 17 を通じてユーザからの実行指令が入力されると、図 2 に示す分析処理を実行する。図 2 に示す分析処理を開始すると、プロセッサ 11 は、データフュージョン対象の第一のデータセット 15 A と第二のデータセット 15 B とを取得する (S110, S120)。
- [0066] S110, S120 において、プロセッサ 11 は、ストレージ 15 に予め格納された第一のデータセット 15 A 及び第二のデータセット 15 B を、ストレージ 15 から読み出すことができる。これにより、プロセッサ 11 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 を取得することができる。
- [0067] 取得すべき第一のデータセット 15 A 及び第二のデータセット 15 B は、ユーザから指定され得る。ユーザは、データフュージョン対象の第一のデータセット 15 A 及び第二のデータセット 15 B を予め収集してストレージ 15 に格納することができる。
- [0068] あるいは、プロセッサ 11 は、通信インタフェース 19 を用いた通信により、第一の外部装置から第一のデータセット 15 A を取得し、第二の外部装置から第二のデータセット 15 B を取得することができる。
- [0069] 第一のデータセット 15 A は、複数の第一のエンティティに関するデータ

セットであって、第一のエンティティのそれぞれの第一の特徴を記述するデータセットである。第一のデータセット 15 A は、第一の特徴データの集合であり、第一の特徴データのそれぞれは、複数の第一のエンティティのうちの対応する一つのエンティティの第一の特徴を表す。

[0070] 第二のデータセット 15 B は、複数の第二のエンティティに関するデータセットであって、第二のエンティティのそれぞれの第二の特徴を記述するデータセットである。第二の特徴は、第一の特徴とは異なる特徴であり得る。具体的に、第二のデータセット 15 B は、第二の特徴データの集合であり、第二の特徴データのそれぞれは、複数の第二のエンティティのうちの対応する一つのエンティティの第二の特徴を表す。

[0071] 第一のエンティティの集合、及び、第二のエンティティの集合は、例えば、互いに共通する母集団における異なる部分集合である。母集団は、人の集合、又は、消費者の集合であり得る。例えば第一のエンティティの集合は、第一の企業の顧客に対応する人の集合であり得る。例えば第二のエンティティの集合は、第一の企業とは異なる第二の企業の顧客に対応する人の集合であり得る。

[0072] あるいは、第一のエンティティの集合は、第一の行動の収集対象とされる人の集合であり得る。第二のエンティティの集合は、第二の行動の収集対象とされる人の集合であり得る。

[0073] 図 3 A に示される第一のデータセット 15 A は、第一の人の集合に関するデータであり、人毎の購買行動に関する特徴データを備える。各特徴データは、対応する人の ID に関連付けて、複数の商品 P 1, P 2, P 3, …のそれぞれを、対応する人が購入したかを 1 又は 0 の 2 値で表す。

[0074] 図 3 B に示される第二のデータセット 15 B は、第二の人の集合に関するデータであり、人毎のウェブコンテンツの閲覧行動に関する特徴データを備える。各特徴データは、対応する人の ID に関連付けて、複数のウェブサイト S 1, S 2, S 3, …のそれぞれについて、対応するウェブサイトを、対応する人が訪問してウェブコンテンツを閲覧したか否かを 1 又は 0 の 2 値で

表す。

[0075] プロセッサ11は、S110において、取得した第一のデータセット15Aに含まれる第一のエンティティ毎の第一の特徴データに基づいて、第一のエンティティ毎のM1次元特徴ベクトル $x = (x_1, x_2, x_3, \dots)$ を生成する。一例によれば、特徴ベクトル x の要素 $x_1, x_2, x_3, \dots$ は、それぞれ、対応する人の商品P1, P2, P3, ...の購買の有無を表し得る。

[0076] 同様に、プロセッサ11は、S120において、取得した第二のデータセット15Bに含まれる第二のエンティティ毎の第二の特徴データに基づいて、第二のエンティティ毎のM2次元特徴ベクトル $y = (y_1, y_2, y_3, \dots)$ を生成する。一例によれば、特徴ベクトル y の要素 $y_1, y_2, y_3, \dots$ は、それぞれ、対応する人のウェブサイトS1, S2, S3, ...でのウェブコンテンツの閲覧有無を表し得る。

[0077] その後、プロセッサ11は、特徴ベクトル x の一群に対する次元削減処理(S130)により、各特徴ベクトル x を、M1次元特徴ベクトルから、それより小さいM次元の特徴ベクトルである低次元特徴ベクトル $Dx = (Dx_1, Dx_2, \dots)$ に変換する。これにより、プロセッサ11は、特徴ベクトル x の一群に対応する低次元特徴ベクトル Dx の一群を生成する。図3Aの右下領域は、低次元特徴ベクトル Dx の例を、テーブルにより示す。

[0078] プロセッサ11は更に、特徴ベクトル y の一群に対する次元削減処理(S140)により、各特徴ベクトル y を、M2次元特徴ベクトルから、それより小さいM次元の特徴ベクトルである低次元特徴ベクトル $Dy = (Dy_1, Dy_2, \dots)$ に変換する。これにより、プロセッサ11は、特徴ベクトル y の一群に対応する低次元特徴ベクトル Dy の一群を生成する。低次元特徴ベクトル Dy は、低次元特徴ベクトル Dx と同次元数Mの特徴ベクトルである。図3Bの右下領域は、低次元特徴ベクトル Dy の例を、テーブルにより示す。

[0079] 低次元空間への写像を実現するためのアルゴリズムの例としては、非負値行列分解 (Nonnegative Matrix Factorization)

ion)、潜在的ディリクレ分配 (latent dirichlet allocation)、特異値分解 (singular value decomposition)、及び、確率的潜在意味解析 (Probabilistic Latent Semantic Analysis) が知られている。S130, S140における次元削減処理は、これらのアルゴリズムの一つを用いて実行され得る。

[0080] 上述したアルゴリズムによれば、特徴ベクトルは、エンティティの個々を強く特徴付ける主要な特徴成分が抽出されるように、低次元化され得る。あるいは、特徴ベクトルは、エンティティの個々を区別するための情報の損失が少ない形式で、低次元化され得る。

[0081] その後、プロセッサ11は、低次元特徴ベクトル D_x の一群と、低次元特徴ベクトル D_y の一群と、に基づいて、第一のエンティティと第二のエンティティとの間の関係を計算するアライメント処理を行う (S150-S180)。

[0082] アライメント処理は、カーネライズドソーティング (Kernelized Sorting) の技術を用いて行われる。以下には、カーネライズドソーティングを用いたアライメント処理の詳細を説明する。但し、アライメント処理は、敵対的学習、Gromov-Wasserstein Alignment技術、又は、不均衡最適輸送 (Unbalanced Optimal Transport) 技術を用いて実現されてもよい。

[0083] S150において、プロセッサ11は、低次元特徴ベクトル D_x の一群を用いて、第一のエンティティの集合に関する類似度行列 K を生成する。類似度行列 K は、 N 行 N 列の正方行列である。ここで、 N は、低次元特徴ベクトル D_x の個数、換言すれば、第一のエンティティの数である。

[0084] 類似度行列 K は、第 i 行第 j 列の要素の値 K_{ij} が、第一のエンティティの集合における i 番目のエンティティと j 番目のエンティティとの間の類似度を表す行列として定義される。

[0085] すなわち、類似度行列 K は、第一のエンティティの集合に関して、エンテ

ィティ間の類似度の分布を説明する行列として定義される。換言すれば、類似度行列Kは、第一のエンティティの集合に関して、特徴空間上のエンティティの分布を、エンティティ間の近しさの尺度を用いて説明する行列として定義される。

[0086] 具体的に、類似度は、 i 番目のエンティティの低次元特徴ベクトル $D \times$ である低次元特徴ベクトル $D \times [i]$ と、 j 番目のエンティティの低次元特徴ベクトル $D \times$ である低次元特徴ベクトル $D \times [j]$ と、をカーネル関数 $k(a, b)$ に代入した値 $k(D \times [i], D \times [j])$ として算出される。すなわち、 $K_{ij} = k(D \times [i], D \times [j])$ である。

[0087] カーネル関数 $k(a, b)$ の例には、次式で表されるガウシアンRBF（動径基底関数）カーネルが含まれる。このカーネル関数 $k(a, b)$ を用いて算出される類似度は、値0から値1までの範囲の値を採る。

[0088] [数2]

$$k(a, b) = \exp\left(\frac{-\|a - b\|^2}{\sigma^2}\right)$$

[0089] 上記カーネル関数 $k(a, b)$ によれば、類似度行列Kの要素の値 K_{ij} は、 $0 < K_{ij} \leq 1$ である。

[0090] S160において、プロセッサ11は、低次元特徴ベクトル D_y の一群を用いて、第二のエンティティの集合に関する類似度行列Lを生成する。類似度行列Lは、N行N列の正方行列である。ここで、Nは、低次元特徴ベクトル D_y の個数、換言すれば、第二のエンティティの数である。すなわち、第一のエンティティの数と、第二のエンティティの数は、同一である。

[0091] 類似度行列Lは、類似度行列Kと同様に、第 i 行第 j 列の要素の値 L_{ij} が第二のエンティティの集合のうち、 i 番目のエンティティと j 番目のエンティティとの間の類似度を表す行列として定義される。すなわち、第 i 行第 j 列の要素の値 $L_{ij} = k(D_y[i], D_y[j])$ である。

[0092] 続くS170において、プロセッサ11は、類似度行列K及び類似度行列

Lを用いて、次式に従う値Z(Ω)を最大化する行列Ωを行列Ω*として探索する。

[0093] [数3]

$$\begin{aligned}\Omega^* &= \operatorname{argmax}_{\Omega} Z(\Omega) \\ Z(\Omega) &= \operatorname{trace}(K' \Omega^T L' \Omega) \\ K' &= HKH \\ L' &= HLH\end{aligned}$$

[0094] ここで、行列Hは、N行N列の行列であって、第i行第j列の要素の値が、 $i = j$ であるとき値 $1 - 1/N$ を示し、 $i \neq j$ であるとき値0を示す対角行列である。Tは、転置記号である。 $\operatorname{trace}(X)$ は、行列Xの対角和である。類似度行列K、Lは、対称行列である。行列 $\Omega^T L' \Omega$ が行列K'の転置行列となるような理想的なΩが見つかるとき、値Z(Ω)は、最大化する。

[0095] 行列Ω*を探索することは、低次元特徴ベクトルD_xの一群から特定される第一のエンティティ間の類似度、及び、低次元特徴ベクトルD_yの一群から特定される第二のエンティティ間の類似度に基づき、類似度に関する第一のエンティティ間の相互関係が第二のエンティティ間の相互関係に適合するように、複数の第一のエンティティのそれぞれを、複数の第二のエンティティの少なくとも一つに対応付けることに対応する。

[0096] 換言すれば、行列Ω*を探索することは、低次元特徴ベクトルD_xの一群から特定される第一のM次元特徴空間における第一のエンティティの分布であって、エンティティ間の類似度で定義される第一のエンティティの分布が、低次元特徴ベクトルD_yの一群から特定される第二のM次元特徴空間における第二のエンティティの分布に適合するように、第一のM次元特徴空間上の複数の第一のエンティティを第二のM次元特徴空間にマッピングするための写像を探索することに対応する。

[0097] 図4Aの左グラフは、第一のエンティティの分布を概念的に表し、図4Bの左グラフは、第二のエンティティの分布を概念的に表す。図4A及び図4

Bに示す例は、技術説明のためだけに、2次元の低次元特徴ベクトル D_x 、 D_y を定義している。符号 E_{11} 、 E_{12} 、 E_{13} 、 E_{14} 、 E_{15} 、 E_{16} 、 E_{17} が付された各点は、第一のエンティティのそれぞれの特徴空間上の位置を示す。符号 E_{21} 、 E_{22} 、 E_{23} 、 E_{24} 、 E_{25} 、 E_{26} 、 E_{27} が付された各点は、第二のエンティティのそれぞれの特徴空間上の位置を示す。

[0098] 図4Bから理解できるように、この例によれば、低次元特徴ベクトル D_y の成分 D_{y1} は、低次元特徴ベクトル D_x の成分 D_{x2} に対応し、低次元特徴ベクトル D_y の成分 D_{y2} は、低次元特徴ベクトル D_x の成分 D_{x1} に対応する。

[0099] すなわち、図4Aに示す例によれば、第一のエンティティの一群と、第二のエンティティの一群とは、エンティティの配列及び次元の順序が、類似度行列 K と類似度行列 L との間で異なる形で定義されているだけであり、実質、同じエンティティの集合の類似度分布を示す。

[0100] 第一のエンティティの一群と、第二のエンティティの一群とが、母集団が同じであるなどの理由により、相互に共通する又は関係する集団的性質を有するときには、特徴ベクトル x 、 y の低次元化により、情報源の第一のデータセット15Aと第二のデータセット15Bとの間に共通変数がなくとも、各エンティティに本質的な共通する特徴成分を抽出することができる。

[0101] 但し、このような低次元化によっても、低次元特徴ベクトル D_x 、 D_y が、同じ特徴成分を有するだけで、特徴成分の配列を揃えることはできない。また、第一のデータセット15Aと第二のデータセット15Bとの間でエンティティの配列は揃っていない。

[0102] 行列 Ω^* の探索は、エンティティの配列及び次元の配列に関して、不ぞろいな特徴ベクトル D_x 、 D_y の対応関係を、類似度分布の同一性を手掛かりに、探索する作業に対応する。

[0103] 続くS180において、プロセッサ11は、行列 Ω^* に基づいて、第一のエンティティのそれぞれを、第二のエンティティの少なくとも一つに対応付け

る。行列 Ω^* における第 i 行第 j 列の要素値は、類似度の分布によれば、第一のエンティティの集合のうち i 番目のエンティティと、第二のエンティティの集合のうちの j 番目のエンティティと、が対応する程度又は可能性の大きさを表す。

[0104] 行列 Ω^* の各要素は、理想的には0又は1を採り、各行について、一行の要素値の合計が1になり、各列について、一列の要素値の合計が1になる。行列 Ω^* が、こうした理想的な行列であるときには、値が1である要素の行番号の第一のエンティティと、列番号の第二のエンティティとが、互に対応する。

[0105] すなわち、行列 Ω^* における第 i 行第 j 列の要素が、値1であるとき、第一のエンティティの集合のうち i 番目のエンティティと、第二のエンティティの集合のうちの j 番目のエンティティと、が互に対応することを示す。

[0106] 但し、数値計算上において、行列 Ω^* が、こうした理想的な行列になることはまれである。従って、S180では、次のいずれかの手法で、複数の第一のエンティティのそれぞれを、第二のエンティティの少なくとも一つに対応付ける。

[0107] (手法1) 行列 Ω^* の第 i 行において、値が最大の要素を探索する。値が最大の要素が第 c 列である場合には、第一のエンティティの集合のうち i 番目のエンティティを、第二のエンティティの集合のうち c 番目のエンティティに対応付ける。これを全ての行について行う。

[0108] この手法では、第二のエンティティの一つに、複数の第一のエンティティが対応付けられる可能性がある。この可能性を抑制するために、近傍検索が行われてもよい。近傍検索の例としては、Contextual Dissimilarity measureが知られている。

[0109] (手法2) 厳密な一対一の対応付けを行うために、行列 Ω^* を入力とした最適割当問題を解くことにより、複数の第一のエンティティのそれぞれを、重複しない第二のエンティティの一つに対応付ける。

[0110] S180において、プロセッサ11は更に、第一のエンティティと第二の

エンティティと間の対応関係を説明するテーブルとして、図5に示す対応表を出力することができる。すなわち、第一のエンティティのそれぞれのIDに関連付けて、対応する第二のエンティティのIDを記述する対応表を出力して、ストレージ15に記憶することができる。

- [0111] 更にプロセッサ11は、データフュージョン処理を実行する(S190)。データフュージョン処理において、プロセッサ11は、上記対応付けの結果、又は上記対応表に基づいて、第一のデータセット15Aと、第二のデータセット15Bとを結合して、拡張データセット15Cを生成する。
- [0112] 拡張データセット15Cは、複数の拡張データを備える。図6に示すように、複数の拡張データのそれぞれは、対応する一つの第一の特徴データと第二の特徴データとの結合データである。
- [0113] すなわち、プロセッサ11は、対応表に基づき、第一のデータセット15Aに含まれる複数の第一の特徴データのそれぞれに、第二のデータセット15Bに含まれる複数の第二の特徴データのうちの一つを結合することによって、拡張データセット15Cを生成する。
- [0114] プロセッサ11は、対応表によって、第一のエンティティの集合のうちのi番目のエンティティと、第二のエンティティの集合のうちj番目のエンティティとが対応付けられているとき、第一のエンティティの集合のうちのi番目のエンティティの特徴を説明する第一の特徴データと、第二のエンティティの集合のうちのj番目のエンティティの特徴を説明する第二の特徴データとを結合して、上記i番目のエンティティの拡張データを生成する。
- [0115] このようにして生成された拡張データセット15Cは、ストレージ15に格納される。ストレージ15に格納された拡張データセット15Cは、例えばユーザインタフェース17を通じて入力されるユーザからの指令に基づき、通信インタフェース19を通じて別のシステムに転送される。
- [0116] 別のシステムは、例えば広告配信システムであり得る。広告配信システムは、拡張データセット15Cに基づき、広告配信先のエンティティを判別し、当該エンティティに広告を配信することができる。

- [0117] S 1 9 0において、データフュージョン処理を終了すると、プロセッサ 1 1 は、図 2 に示す分析処理を終了する。
- [0118] 以上に説明したように、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 1 5 A と、第二のデータセット 1 5 B との間に共通変数が存在しなくとも、類似度の分布に基づいて、第一のエンティティと第二のエンティティとを適切に対応付けることができる。
- [0119] 適切な対応付けのためには、第一のエンティティの集合と、第二のエンティティの集合との間において、類似度の分布が、相互に一致する、類似する、又は、関係するのが好ましい。
- [0120] 第一のエンティティの集合と、第二のエンティティの集合とが、同じ母集団からの部分集合であるとき、このような好ましい条件はおおよそ満足される。従って、第一のエンティティ及び第二のエンティティが人であるとき、すなわち、第一のデータセット 1 5 A 及び第二のデータセット 1 5 B として、人に関する特徴を表すデータセットが取り扱われるとき、本実施形態の技術は、有意義に機能する。
- [0121] 特に人の行動は、特にデモグラフィック属性に応じた傾向を示すことが多い。従って、第一のデータセット 1 5 A 及び第二のデータセット 1 5 B が、デモグラフィック属性の分布が互いに類似すると推定される集団からの収集データに基づいたデータセットであるとき、エンティティ間の適切な対応付けを実現可能である。
- [0122] 例えば、第一のデータセット 1 5 A 及び第二のデータセット 1 5 B が、共通変数の存在しない、互いに異なる集団に属する人の特徴を説明するデータセットであったり、異なる行動の特徴を説明するデータセットであったりしても、エンティティ間の対応付けを適切に行うことができる。従って、拡張データセット 1 5 C として、人の心理・行動分析に役立つデータセットを生成することができる。
- [0123] 上述した例によれば、第一のデータセット 1 5 A は、第一の集団に属する複数の人のそれぞれの購買行動に関する特徴を記述するデータセットであり

、第二のデータセット 15 B は、第二の集団に属する複数の人のそれぞれのウェブサイト訪問行動及び／又はウェブコンテンツ閲覧行動に関する特徴を記述するデータセットである。

[0124] 別例によれば、第一のデータセット 15 A 及び第二のデータセット 15 B の一方には、テレビ視聴行動などの人のメディア接触行動に関する特徴を記述するデータセットが用いられてもよい。第一のデータセット 15 A 及び第二のデータセット 15 B の一方には、スマートフォン等の携帯端末の使用状況に関する特徴を記述するデータセットが用いられてもよい。

[0125] 第一のデータセット 15 A 及び第二のデータセット 15 B の一方には、オフライン空間（すなわち現実空間）における人の移動に関する特徴を記述するデータセットが用いられてもよい。データセットは、オフライン空間における人の移動に関する特徴として、例えば複数の場所への訪問、移動経路、及び／又は、移動手段に関する特徴を記述し得る。

[0126] 第一のデータセット 15 A 及び第二のデータセット 15 B の一方には、オンライン空間における人の移動に関する特徴を記述するデータセットが用いられてもよい。データセットは、オンライン空間における人の移動に関する特徴として、仮想現実（VR）空間における人の移動やネットサーフィンに関する特徴を記述し得る。第一のデータセット 15 A 及び第二のデータセット 15 B の一方には、アンケートにより収集されたデータに基づくデータセットが使用されてもよい。

[0127] 第一のデータセット 15 A と第二のデータセット 15 B との組合せとして、アンケートにより収集されたデータセットと、テレビ視聴行動に関するデータセットとの組合せ、又は、移動履歴に関するデータセットと、購買に関するデータセットとの組合せが採用されてもよい。

[0128] 上記実施形態において、低次元特徴ベクトル D_x 、 D_y の一群に対しては、ZCA 白色化、正規化、及び、標準化などの処理が行われてもよい。

[0129] 上記実施形態では、低次元特徴ベクトル D_x 、 D_y の次元数 M が、設計者又はユーザにより定められるが、最適な次元数 M を探索するように、情報処

理システム 1 は構成されてもよい。例えば、情報処理システム 1 は、図 2 に示す分析処理を、同一のデータセット 15 A, 15 B について次元数 M を変更しながら繰返し実行して、Z (Ω) の最大値を指標に、最適な次元数 M を自動選定するように構成されてもよい。

[0130] <第二実施形態>

第二実施形態の情報処理システム 1 は、プロセッサ 11 が図 2 に示す分析処理に代えて、図 7 に示す分析処理を実行するように構成される。以下では、第二実施形態の説明として、プロセッサ 11 が実行する分析処理の詳細を選択的に説明する。本実施形態において言及されない情報処理システム 1 の構成は、第一実施形態と同じであると理解されてよい。

[0131] プロセッサ 11 は、図 7 に示す分析処理を開始すると、第一実施形態と同様に、データフュージョン対象の第一のデータセット 15 A と第二のデータセット 15 B とを取得する (S 310, S 320)。

[0132] プロセッサ 11 は、S 110 での処理と同様に、第一のデータセット 15 A に基づいて、第一のエンティティ毎の特徴ベクトル x を生成する (S 310)。プロセッサ 11 は、S 120 での処理と同様に、第二のデータセット 15 B に基づいて、第二のエンティティ毎の特徴ベクトル y を生成する (S 320)。

[0133] 更に、プロセッサ 11 は、S 130, S 140 での処理と同様に、次元削減処理によって、特徴ベクトル x の一群に対応する低次元特徴ベクトル D_x の一群を生成し、特徴ベクトル y の一群に対応する低次元特徴ベクトル D_y の一群を生成する (S 330)。

[0134] 続く S 340 において、プロセッサ 11 は、S 150, S 160, S 170 での処理と同様の処理を実行する。すなわち、プロセッサ 11 は、低次元特徴ベクトル D_x の一群を用いて、第一のエンティティの集合に関する類似度行列 K を生成し、低次元特徴ベクトル D_y の一群を用いて、第二のエンティティの集合に関する類似度行列 L を生成する。

[0135] 更に、プロセッサ 11 は、類似度行列 K 及び類似度行列 L を用いて、第一

実施形態で説明した値 $Z(\Omega)$ を最大化する行列 Ω を行列 Ω^* として探索する (S340)。ここでは、探索された行列 Ω^* のことを、対応関係行列 Ω^* と表現する。

- [0136] その後、プロセッサ11は、繰返し終了条件が満足されたか否かを判断する (S350)。繰返し終了条件が満足されていないと判断すると (S350でNo)、プロセッサ11は、S360の処理を実行する。
- [0137] S360において、プロセッサ11は、S340で探索された対応関係行列 Ω^* を固定した状態で、Gromov-Wasserstein距離のコストを最小化する次元削減方式を探索する。
- [0138] 対応関係行列 Ω^* を固定した状態は、第一のエンティティと第二のエンティティとの間の対応関係を固定した状態に対応する。上述したように値 $Z(\Omega)$ を最大化する行列 Ω を対応関係行列 Ω^* として探索することは、第二の特徴空間における第二のエンティティの分布に適合するように、第一の特徴空間上の複数の第一のエンティティを第二の特徴空間にマッピングするための写像を探索することに対応する。
- [0139] Gromov-Wasserstein距離のコストは、第一のエンティティの集合を、第二の特徴空間にマッピングしたときの第一のエンティティと第二のエンティティとの間の最適輸送問題における輸送コストに対応する。
- [0140] Gromov-Wasserstein距離のコストは、類似度行列 K 、 L 及び対応関係行列 Ω^* を用いて算出可能である。類似度行列 K は、上述の通り、次元削減後の低次元特徴ベクトル D_x に基づいて算出された第一のエンティティ間の類似度を要素に含む行列である。類似度行列 L は、次元削減後の低次元特徴ベクトル D_y に基づいて算出された第二のエンティティ間の類似度を要素に含む行列である。
- [0141] Gromov-Wasserstein距離のコストを最小化する次元削減方式を探索することは、対応関係行列 Ω^* で示される第一のエンティティと第二のエンティティとの間の対応関係を最もよく正当化する低次元特徴ベク

トル D_x , D_y を生成するための次元削減方式を探索することに対応する。

[0142] コストの最小化は、対応関係行列 Ω^* によれば、互いに対応する第一のエンティティと第二のエンティティとの間の特徴空間上の距離、換言すれば、第一のエンティティの低次元特徴ベクトル D_x と、第二のエンティティの低次元特徴ベクトル D_y との間の特徴空間上の距離が短くなるように、次元削減方式を探索することに対応する。

[0143] 例えば、 M_1 次元の特徴ベクトル x を、 M 次元の低次元特徴ベクトル D_x に変換する場合には、特徴ベクトル x に M 行 M_1 列の変換行列 T_x を作用させる。 M_2 次元の特徴ベクトル y を、 M 次元の低次元特徴ベクトル D_y に変換する場合には、特徴ベクトル y に M 行 M_2 列の変換行列 T_y を作用させる。このとき、変換行列 T_x , T_y を構成するパラメータ m の数は、 $(M * M_1 + M * M_2)$ 個である。

[0144] 次元削減方式の探索は、例えば変換行列 T_x , T_y のパラメータ m として、上述のコストを最小化するパラメータ m を、勾配法等を用いて探索することにより実現される。

[0145] その後、プロセッサ 11 は、探索された次元削減方式（例えば変換行列 T_x , T_y ）で特徴ベクトル x , y を低次元化し、新たな低次元特徴ベクトル D_x , D_y を算出する（S370）。

[0146] プロセッサ 11 は、新たな低次元特徴ベクトル D_x に基づく類似度行列 K 、及び、新たな低次元特徴ベクトル D_y に基づく類似度行列 L を用いて、値 $Z(\Omega)$ を最大化する行列 Ω を、新たな対応関係行列 Ω^* として探索する（S340）。

[0147] プロセッサ 11 は、このように S360, S370, S340 の処理を繰り返し実行することによって、マッチング精度の高い対応関係行列 Ω^* を、より良い次元削減方式と共に再探索する。

[0148] プロセッサ 11 は、繰り返し終了条件が満足されると（S350でYes）、S380の処理を実行する。繰り返し終了条件は、例えば、S340の処理が所定回実行された場合に、あるいは、再探索による対応関係行列 Ω^* の変化

量が一定未満になった場合に満足される。

[0149] S 3 8 0において、プロセッサ 1 1 は、第一実施形態における S 1 8 0 の処理と同様に、繰返し処理の最後に算出された対応関係行列 Ω^* に基づいて、第一のエンティティのそれぞれを、第二のエンティティの少なくとも一つに対応付ける。プロセッサ 1 1 は更に、第一のエンティティと第二のエンティティとの対応関係を説明する対応表を記憶及び出力することができる。

[0150] その後、プロセッサ 1 1 は、S 1 9 0での処理と同様に、データフュージョン処理を実行することにより、第一のデータセット 1 5 A と、第二のデータセット 1 5 B とを結合して、拡張データセット 1 5 C を生成し、生成した拡張データセット 1 5 C をストレージ 1 5 に格納する (S 3 9 0)。

[0151] 以上に説明した第二実施形態の情報処理システム 1 は、上述の繰返し処理によって、更に精度よく、第一のエンティティと第二のエンティティとの間の対応付けを行うことができる。従って、精度の良い拡張データセット 1 5 C を生成することが可能である。

[0152] <第三実施形態>

第三実施形態の情報処理システム 1 は、ユーザインタフェース 1 7 を通じたユーザからの実行指示に基づき、プロセッサ 1 1 が図 8 に示す評価処理を実行するように構成される。以下では、第三実施形態の説明として、プロセッサ 1 1 が実行する評価処理の詳細を説明する。本実施形態において言及されない情報処理システム 1 の構成は、第一又は第二実施形態と同じであると理解されてよい。

[0153] 評価処理は、評価対象のデータセットが、図 2 又は図 7 に示す分析処理での対応付け及びデータフュージョンを高精度に実行可能な優良なデータセットであるか否かを評価するために実行される。評価対象のデータセットは、分析処理で、第一のデータセット 1 5 A 又は第二のデータセット 1 5 B として使用され得るデータセットに対応する。

[0154] プロセッサ 1 1 は、評価処理を開始すると、ユーザから実行指示と共に指定された評価対象のデータセットを取得する (S 4 1 0)。プロセッサ 1 1

は、ストレージ 15 から指定された評価対象のデータセットを取得することができる。

[0155] その後、プロセッサ 11 は、評価対象のデータセットに基づき、エンティティ毎に、第一の特徴ベクトル x_1 と、第二の特徴ベクトル x_2 と、を生成する (S420)。評価対象のデータセットは、エンティティ毎に、対応するエンティティの特徴を $(Q1 + Q2)$ 個の要素で表す特徴データを備えることができる。

[0156] プロセッサ 11 は、 $(Q1 + Q2)$ 個の要素を、 $Q1$ 個の要素からなる第一の要素群と、 $Q2$ 個の要素からなる第二の要素群と、に分割することができる。 $(Q1 + Q2)$ 個の要素のそれぞれは、ランダムに、第一の要素群及び第二の要素群のいずれかに分類され得る。

[0157] プロセッサ 11 は、評価対象のデータセットに基づいて、エンティティ毎に、対応するエンティティの第一の要素群に関する特徴を記述した第一の特徴ベクトル x_1 と、対応するエンティティの第二の要素群に関する特徴を記述した第二の特徴ベクトル x_2 と、を生成することができる。

[0158] 例えば、評価対象のデータセットが、S110, S120, S310, 又は S320 で特徴ベクトル $v = (v[1], v[2], v[3], \dots, v[Q])$ が生成され得る要素数 $Q = (Q1 + Q2)$ の特徴データをエンティティ毎に備える場合、 $Q1$ 個の要素を含む第一の特徴ベクトル $x_1 = (v[1], v[2], \dots, v[Q1])$ 及び $Q2$ 個の要素を含む第二の特徴ベクトル $x_2 = (v[Q1 + 1], v[Q1 + 2], \dots, v[Q1 + Q2])$ が生成され得る。

[0159] 第一の特徴ベクトル x_1 は、第一のエンティティの集合におけるエンティティ毎の特徴ベクトル x に対応し、第二の特徴ベクトル x_2 は、第一のエンティティの集合と同一の第二のエンティティの集合におけるエンティティ毎の特徴ベクトル y に対応する。

[0160] その後、プロセッサ 11 は、S130～S170 で実行される処理と同様の処理を、S430, S440 において、第一の特徴ベクトル x_1 及び第

二の特徴ベクトル $\times_2$ に対して実行する。

- [0161] S 4 3 0において、プロセッサ11は、S 1 3 0, S 1 4 0での処理と同様に、第一のエンティティ毎の第一の特徴ベクトル $\times_1$ 及び第二のエンティティ毎の第二の特徴ベクトル $\times_2$ に対する次元削減処理を実行して、同次元数の低次元特徴ベクトル $D \times_1$ 及び低次元特徴ベクトル $D \times_2$ を生成する。
- [0162] プロセッサ11は、第一のエンティティ毎の低次元特徴ベクトル $D \times_1$ に基づき、類似度行列 K に対応する第一のエンティティ間の低次元特徴ベクトル $D \times_1$ の類似度を表す類似度行列を生成する。プロセッサ11は更に、第二のエンティティ毎の低次元特徴ベクトル $D \times_2$ に基づき、類似度行列 L に対応する第二のエンティティ間の低次元特徴ベクトル $D \times_2$ の類似度を表す類似度行列を生成する。
- [0163] プロセッサ11は、これらの類似度行列に基づき、値 $Z(\Omega)$ を最大化する行列 Ω を対応関係行列 Ω^* として探索する(S 4 4 0)。
- [0164] その後、プロセッサ11は、低次元特徴ベクトル $D \times_1$ の一群に対応する第一のエンティティの集合と、低次元特徴ベクトル $D \times_2$ の一群に対応する第二のエンティティの集合とに関して、対応関係行列 Ω^* が、第一のエンティティと第二のエンティティとの間の対応関係を正しく表している程度をスコアとして算出する(S 4 5 0)。
- [0165] これにより、プロセッサ11は、評価対象のデータセットが分析処理による対応付け及びデータフュージョンを高精度に実行可能な優良なデータセットであるか否かを評価する(S 4 5 0)。
- [0166] プロセッサ11は、予めS 4 2 0で第一のエンティティ毎の特徴ベクトル $\times_1$ 及び第二のエンティティ毎の特徴ベクトル $\times_2$ を生成する際に、第一のエンティティと第二のエンティティとの間の正しい対応関係を記憶しておくことができる。
- [0167] プロセッサ11は、このように対応関係の正解を記憶した環境で、S 4 3 0, S 4 4 0において分析処理と同様の処理を実行して対応関係行列 Ω^* を算

出し、対応関係行列 Ω^* から特定される対応関係を正解と比較する。

- [0168] 例えば、プロセッサ 11 は、対応関係行列 Ω^* に基づいて、第一のエンティティのそれぞれを、第二のエンティティの一つと対応付ける処理を、S 180、S 380での処理と同様に実行する。
- [0169] プロセッサ 11 は、対応関係行列 Ω^* に基づいて対応付けられた第一のエンティティと第二のエンティティとが、評価対象のデータセットにおいて同一のエンティティである場合には、対応付けに成功したと判別し、同一のエンティティではない場合には、対応付けに失敗したと判別する。
- [0170] プロセッサ 11 は、エンティティ全体のうち、対応付けに成功した割合を、評価対象のデータセットのスコアとして算出することができる (S 450)。その後、プロセッサ 11 は、算出したスコアを評価結果として出力し (S 460)、評価処理を終了する。
- [0171] 一つのデータセットに基づく対応付け及びデータフュージョンを高精度に実行できない場合には、そのデータセットが、集合の特徴に関して高精度な対応付け及びデータフュージョンを実現するために十分な情報又はデータ構造を有していないと推測できる。
- [0172] この情報不足は、二つの異なるデータセットに関して分析処理を実行して、対応付け及びデータフュージョンを行う場合の精度にも影響する。従って、上記評価処理によれば、評価対象のデータセットが、共通変数なしのデータフュージョンを高精度に実行可能なデータセットであるかを、事前に推測することができる。
- [0173] プロセッサ 11 は、S 460において、スコアの出力により、評価対象のデータセットが優良なデータセットであるか否かを情報処理システム 1 のユーザに伝達することができる。これにより、ユーザは、分析処理に、適切な第一のデータセット 15 A 及び第二のデータセット 15 B の組合せを採用して、信頼性の高い拡張データセット 15 C を得ることができる。
- [0174] 所望の拡張データセット 15 C を得るために、第二のデータセット 15 B に結合する第一のデータセット 15 A として、互いに類似する複数のデータ

セットのいずれかを採用すれば十分である環境が考えられる。

[0175] 例えば、購買行動に関する第一のデータセット15Aと、ウェブサイト訪問行動／ウェブコンテンツ閲覧行動に関する第二のデータセット15Bとを、結合して、拡張データセット15Cを生成することを考える。この場合、第一のデータセット15Aとして、複数の流通組織のいずれか一組織の顧客の購買行動に関するデータセットを用いて、拡張データセット15Cを生成すれば十分であることが考えられる。

[0176] 複数の流通組織の例には、複数のコンビニエンスストアチェーンが含まれる。各コンビニストアチェーンの購買に関するデータセットには、消費者の購買行動として、他のコンビニストアチェーンと同種の購買行動に関する情報が含まれ得る。

[0177] 従って、第一のデータセット15Aとしては、複数のコンビニエンスストアチェーンのうちのいずれか一つの顧客の購買行動に関するデータセットを用いて、拡張データセット15Cを生成すれば十分であることが考えられる。

[0178] 上述の評価処理は、第一のデータセット15A（又は第二のデータセット15B）の候補として、複数のデータセットが存在する場合に、これらの複数のデータセットから、対応付け及びデータフュージョンの精度の観点で最適なデータセットを、第一のデータセット15A（又は第二のデータセット15B）として選択するために利用することができる。

[0179] 例えば、プロセッサ11は、S110、S120、S310、S320のいずれかの処理において、必要に応じて、図9に示す選択処理を実行することにより、データフュージョン対象のデータセットの複数の候補から、一つの候補を、データフュージョン対象のデータセットとして採用することができる。S110、S310におけるデータフュージョン対象のデータセットは、第一のデータセット15Aに対応し、S120、S320におけるデータフュージョン対象のデータセットは、第二のデータセット15Bに対応する。

- [0180] 図9に示す選択処理を開始すると、プロセッサ11は、データフュージョン対象のデータセットの複数の候補として、複数のデータセットを取得する(S510)。プロセッサ11は、ユーザから指定された複数のデータセットを、ストレージ15から取得することができる。
- [0181] その後、プロセッサ11は、複数のデータセットのうちの一つを、評価対象のデータセットに設定して(S520)、図8に示す評価処理を実行する(S530)。プロセッサ11は、複数のデータセットのすべてに関する評価処理を実行するまで(S540でYes)、データセット毎に、これを評価対象のデータセットに設定して(S520)、評価処理(S530)を実行する処理を繰り返す。これにより、データセット毎に、S450で算出されるスコアを取得する。
- [0182] 複数のデータセットのすべてに関して評価処理を実行し、スコアを取得すると(S540でYes)、プロセッサ11は、複数のデータセットのうち、最もスコアの高いデータセットを、データフュージョン対象のデータセットに採用する(S550)。その後、選択処理を終了する。S110, S120, S310, S320において、プロセッサ11は、採用されたデータフュージョン対象のデータセットに基づく特徴ベクトル(x又はy)を生成することができる。
- [0183] このように選択処理を実行して、複数の候補の中から最適なデータセットを選択することによれば、精度の高い拡張データセット15Cを生成することが可能である。
- [0184] 付言すると、購買行動の例において、データフュージョン対象のデータセットの複数の候補には、消費者の購買行動を異なるパラメータで表す複数のデータセットが含まれ得る。例えば、第一の候補は、エンティティとしての消費者毎に、商品毎の購入個数を要素に含む特徴ベクトルを生成可能なデータセットであり得る。第二の候補は、エンティティとしての消費者毎に、商品毎の購入金額を要素に含む特徴ベクトルを生成可能なデータセットであり得る。

[0185] こうした同種の特徴を異なるパラメータで説明する複数のデータセットを用意して、データフュージョンに適したデータセットを選択することは、より良い拡張データセット 15 C の生成に繋がる。

[0186] <第四実施形態>

図 10 に示す第四実施形態の配信システム 30 は、第一実施形態又は第二実施形態のデータフュージョン技術を用いて、配信システム 30 の外部から提供されるデータセットである外部データセット 35 A と、配信システム 30 の内部に保持するデータセットである内部データセット 35 B とを結合し、それにより生成される拡張データセット 35 C に基づいて、広告配信を行うシステムである。

[0187] 配信システム 30 は、図 10 に示すように、プロセッサ 31 と、メモリ 33 と、ストレージ 35 と、通信インタフェース 39 とを備える。プロセッサ 31 は、ストレージ 35 に格納されたコンピュータプログラム P r 1 に従う処理を実行する。ストレージ 35 は、更に、内部データセット 35 B を備える。

[0188] 内部データセット 35 B は、図 11 に示すように、ユーザ毎に、対応するユーザの広告 I D に関連付けて、対応するユーザのオンライン行動の特徴を説明する特徴データを備える。広告 I D は、良く知られるように、広告のために使用される識別コードであって、情報端末に固有の I D である。

[0189] 広告 I D に関連付けられた特徴データは、対応する広告 I D が割り当てられた情報端末を通じて観測されたユーザのオンライン行動の特徴を説明する。オンライン行動には、ウェブサイト訪問行動及びウェブコンテンツ閲覧行動が含まれる。

[0190] 配信システム 30 は、通信インタフェース 39 を通じて広域ネットワークと接続され、広域ネットワークを介して、広告配信サービスを提供する。広告配信サービスを利用する企業側のシステムである利用企業側システム 40 は、配信システム 30 に対して、配信対象の広告コンテンツと共に、配信指定情報を提供する。広告コンテンツは、広告用の情報コンテンツである。配

信指定情報には、配信ターゲットを指定するターゲット指定情報、及び、配信数を指定する配信数指定情報が含まれる。

[0191] 利用企業側システム40は更に、配信システム30に対し、外部データセット35Aとして、配信先候補に対応する顧客の特徴を説明するデータセットである顧客データセットを提供する。

[0192] 顧客データセットは、例えば、利用企業が運営する店舗を利用する顧客の購買行動に関する特徴を説明するデータセットであり得る。例えば、顧客データセットは、顧客毎の特徴データとして、複数の商品に関する、対応する顧客の商品毎の購買量を説明する特徴データを備えることができる。

[0193] プロセッサ31は、通信インタフェース39を通じて利用企業側システム40から配信要求が入力されると、コンピュータプログラムP1に基づいて図12に示す配信制御処理を実行する。

[0194] 配信制御処理を開始すると、プロセッサ31は、利用企業側システム40から、配信対象の広告コンテンツと共に、ターゲット指定情報及び配信数指定情報を含む配信指定情報、並びに、外部データセット35Aとしての顧客データセットを取得する(S610)。

[0195] その後、プロセッサ31は、第一のデータセット15Aとして外部データセット35Aを用いて、更には、第二のデータセット15Aとして内部データセット35Bを用いて、分析処理におけるS110～S190の処理と同様の処理を実行する。これにより、プロセッサ31は、外部データセット35Aと内部データセット35Bとを結合し、拡張データセット35Cを生成する(S620)。

[0196] 外部データセット35Aと内部データセット35Bとの結合により、外部データセット35Aに含まれる顧客毎の特徴データには、内部データセット35Bに含まれる顧客と同一人物である可能性の高いユーザの広告IDが関連付けられる。

[0197] 拡張データセット35Cは、エンティティ毎に、対応する顧客の外部データセット35Aが有する特徴データと、対応するユーザの内部データセット

35 Bが有する特徴データとが結合された拡張データを備える。各拡張データには、内部データセット35 Bが有する対応するユーザの広告IDが関連付けられる。

[0198] ここでいうエンティティは、データフュージョンにより互いに対応付けられた顧客とユーザとの組合せのことである。データフュージョンでは、顧客とユーザとが一对一对応付けられる。例えば、拡張データセット35 Cは、図6に示す拡張データセット15 Cにおいて図示される「ID2__1」「ID2__2」「ID2__3」を有する列に、各エンティティの広告IDが記述された構造を有するデータセットであり得る。

[0199] プロセッサ31は、その後、拡張データセット35 C内の各エンティティが配信ターゲットである可能性に関するスコアを算出する(S630)。例えば、外部データセット35 Aが顧客の購買行動に関するデータセットであり、内部データセット35 Bがユーザのオンライン行動に関するデータセットである場合、プロセッサ31は、拡張データセット35 C内の各エンティティの購買行動に関する特徴データとオンライン行動に関する特徴データとを所定の関数に入力して、対応するエンティティが配信ターゲットである可能性を数値化したスコアを算出する。

[0200] 配信ターゲットは、性別、年齢、購買傾向、オンライン行動傾向、興味、及び関心等の消費者を特徴付けるパラメータにより絞り込まれる配信先の消費者群であり、ターゲット指定情報を通じて指定される。

[0201] S630におけるスコア算出後、プロセッサ31は、広告IDが関連付けられているエンティティの一群のうち、算出されたスコアが高い順に、利用企業側システム40から指定された配信数に対応する数のエンティティを、コンテンツ配信先に決定する(S640)。このようにして、プロセッサ31は、外部データセット35 Aに対応する複数の顧客のいずれかと対応付けられた内部データセット35 Bに対応する複数のユーザの少なくとも一部を、広告コンテンツの配信先に選択する。

[0202] その後、プロセッサ31は、決定したコンテンツ配信先の情報端末に、利

用企業側システム40から提供された広告コンテンツを、広域ネットワークを通じて送信する(S650)。広告コンテンツは、コンテンツ配信先の広告IDから識別される情報端末に配信される。その後プロセッサ31は、配信制御処理を終了する。

[0203] 以上に説明した第四実施形態の配信システム30によれば、共通変数なしのデータフォージョン技術を用いて、外部データセット35Aと内部データセット35Bとを結合することにより、広告IDが不明な顧客の特徴データに対して広告IDを関連付けることができる。これにより、広告IDが不明な外部データセット35Aの顧客に対して、広告コンテンツを適切に配信することができる。

[0204] <第五実施形態>

第五実施形態の配信システム30は、プロセッサ31が図12に示す配信制御処理に代えて、図13に示す配信制御処理を実行するように構成される。以下では、第五実施形態の説明として、プロセッサ31が実行する配信制御処理の詳細を選択的に説明する。本実施形態において言及されない配信システム30の構成は、第四実施形態と同じであると理解されてよい。

[0205] 本実施形態において、プロセッサ31は、通信インタフェース39を通じて利用企業側システム40から、配信要求が入力されると、図13に示す配信制御処理を実行する。

[0206] 配信制御処理を開始すると、プロセッサ31は、利用企業側システム40から、配信対象の広告コンテンツと共に、配信指定情報、外部データセット35Aとしての顧客データセットを取得する(S710)。

[0207] S710で取得される配信指定情報は、ターゲット指定情報を含まず、配信数指定情報のみを含む。外部データセット35Aとして取得される顧客データセットは、利用企業が絞り込んだ配信ターゲットに対応する顧客群の特徴を説明する特定顧客データセットである。

[0208] その後、プロセッサ31は、S620での処理と同様に、外部データセット35Aと内部データセット35Bとを結合し、拡張データセット35Cを

生成する（S 7 2 0）。拡張データセット 3 5 C は、エンティティ毎に、対応する顧客の外部データセット 3 5 A が有する特徴データと、対応するユーザの内部データセット 3 5 B が有する特徴データとが結合された拡張データを備える。

[0209] 本実施形態の S 7 2 0 の処理では、内部データセット 3 5 B のユーザのすべてに対して外部データセット 3 5 A の顧客が対応付けられる結果は生じない。本実施形態の拡張データセット 3 5 C は、利用企業側の顧客と対応付けられていないユーザの特徴データも、一つのエンティティの拡張データとして含む。この拡張データは、実質的には拡張されていない内部データセット 3 5 B が有する該当ユーザの特徴データである。

[0210] 本実施形態では、拡張データセット 3 5 C に対応するエンティティの一群のうち、外部データセット 3 5 A に対応する顧客群に対応付けられたエンティティ群のことをシードと表現し、それ以外のエンティティの一群のことを、非シードと表現する。

[0211] S 7 2 0 の処理後、プロセッサ 3 1 は、拡張データセット 3 5 C に基づいて、非シードの各エンティティと、シードの各エンティティとの間の内部データセット 3 5 B が示す特徴の類似度を算出する（S 7 3 0）。類似度は、非シードの各エンティティと、シードの各エンティティとの間の特徴空間上の距離によって算出され得る。

[0212] 類似度の算出後、プロセッサ 3 1 は、類似度の高い順に、配信指定情報で指定された配信数に対応する数のエンティティを配信先に決定する（S 7 4 0）。この際、シードに対応する全てのエンティティも配信先に決定される。

[0213] このようにして、プロセッサ 3 1 は、外部データセット 3 5 A に対応する複数の顧客と対応付けられたユーザの集合であるシードの集合と、内部データセット 3 5 B に対応する複数のユーザのうち、シードと特徴が類似するユーザの集合と、を広告コンテンツの配信先に選択する。

[0214] その後、プロセッサ 3 1 は、S 6 5 0 の処理と同様に、S 7 4 0 で決定し

たコンテンツ配信先の情報端末に、利用企業側システム40から提供された広告コンテンツを、広域ネットワークを通じて送信する（S750）。その後、配信制御処理を終了する。

[0215] 以上に説明した本実施形態の配信システム30によれば、利用企業側システム40から提供された顧客群のデータセットを基礎に、顧客群と類似した特徴を示すより大きな集合の消費者の情報端末に、広告コンテンツを配信することができる。従って、本実施形態によれば、多くの消費者に、効率的に広告配信を行うことが可能である。

[0216] <第六実施形態>

第六実施形態の配信システム30は、第四実施形態又は第五実施形態の配信システム30と同様の広告配信サービスと共に、予測サービスを提供するように構成される。

[0217] 本実施形態において、プロセッサ31は、利用企業側システム40からの実行要求に応じて、図14に示す予測処理を実行する。以下では、第六実施形態の説明として、プロセッサ31が実行する予測処理の詳細を選択的に説明する。本実施形態において言及されない配信システム30の構成は、第四実施形態又は第五実施形態と同じであると理解されてよい。

[0218] プロセッサ31は、予測処理を開始すると、通信インタフェース39を通じて分析対象のデータセットを、分析条件指定情報と共に、利用企業側システム40から取得する（S810）。分析対象のデータセットは、分析対象の顧客毎の特徴データを備えるデータセットである。

[0219] 分析条件指定情報は、顧客の購入可能性を評価する対象の商品を指定する情報であり得る。予測処理では、指定された対象商品进行分析対象の各顧客が購入する可能性が、対象商品の購入数の予測値を算出することによって予測される。ここでの予測は、顧客の行動を推定することに対応し、予測値は、行動に関する推定値に対応する。

[0220] S810の処理実行後、プロセッサ31は、第一のデータセット15Aとして分析対象のデータセットを用いて、更には、第二のデータセット15B

として内部データセット35Bを用いて、分析処理におけるS110～S170又はS310～S370の処理と同様の処理を実行することにより、分析対象の各顧客と内部データセット35Bに特徴データを有する各ユーザとの対応関係を示す対応関係行列 Ω^* を算出する(S820)。

[0221] プロセッサ31は更に、算出された対応関係行列 Ω^* に基づき、分析対象の顧客毎に、対応する顧客に近い所定数のユーザを抽出し、内部データセット35Bから特定可能な上記抽出したユーザの対象商品の購入数の重み付け平均により、対応する顧客の対象商品の購入数の予測値を算出する(S830)。このようにしてプロセッサ31は、顧客の購買行動を、対応付けられたユーザの購買行動から推定する。内部データセット35Bは、各ユーザの対象商品の購入数を特定可能な情報を含む。

[0222] 対応関係行列 Ω^* の各要素は、顧客とユーザとの間の類似度を0～1の値で示す。対応関係行列 Ω^* における第i行第j列の要素は、内部データセット35Bに対応するユーザの集合のうちi番目のユーザと、分析対象のデータセットに対応する顧客の集合のうちのj番目の顧客と、の間の類似度を値0～1で示す。

[0223] 重み付け平均は、例えば、類似度を重みとして用いて算出される。重み付け平均は、顧客と近い3人のユーザとして第一、第二、及び第三のユーザが抽出されたと仮定したとき、次のように算出され得る。

[0224] すなわち、顧客と第一のユーザとの類似度が w_1 、顧客と第二のユーザとの類似度が w_2 、顧客と第三のユーザとの類似度 w_3 であり、第一のユーザの対象商品購入数が p_1 、第二のユーザの対象商品購入数が p_2 であり、第三のユーザの対象商品購入数が p_3 であるとき、顧客の対象商品の購入数の予測値 p_e は、 $p_e = (w_1 \cdot p_1 + w_2 \cdot p_2 + w_3 \cdot p_3) / 3$ で算出され得る。

[0225] 対応関係行列 Ω^* からは、顧客毎に、すべてのユーザとの類似度（換言すれば対応付けの大きさ）を特定可能である。従って、近いユーザを抽出するプロセスなしに、すべてのユーザの対象商品の購入数の重み付け平均によっ

て、顧客の対象商品の購入数の予測値が算出されてもよい。

[0226] S 8 3 0 の処理実行後、プロセッサ 3 1 は、顧客毎の対応商品の購入数の予測値を記述した予測データを、予測処理の実行要求元へ出力する（S 8 4 0）。その後、プロセッサ 3 1 は、図 1 4 に示す予測処理を終了する。

[0227] 別例によれば、プロセッサ 3 1 は、S 8 3 0 の処理実行後、予測データを出力することに代えて又は加えて、顧客毎の対応商品の購入数の予測値に基づき、予測値が大きい順に、利用企業から指定された配信数に対応する人数の顧客に対して、対象商品の購入を進める広告コンテンツを配信する処理を実行してもよい（S 8 4 0）。

[0228] 以上に、第六実施形態の配信システム 3 0 を説明したが、本実施形態によれば、共通変数なしのデータフュージョン技術を用いて、有意義な広告配信サービスを提供でき、更には、有意義なマーケティングソリューションを提供可能である。

[0229] [その他]

本開示が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ことは言うまでもない。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] 複数の第一のエンティティに関する第一のデータセットであって、前記複数の第一のエンティティのそれぞれの特徴を記述す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、
- 複数の第二のエンティティに関する第二のデータセットであって、前記複数の第二のエンティティのそれぞれの特徴を記述す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、
- 前記第一のデータセットから特定される第一の特徴ベクトルの一群であって、前記第一の特徴ベクトルのそれぞれが、前記複数の第一のエンティティのうちの対応する一つのエンティティの特徴を表す第一の特徴ベクトルの一群、及び、前記第二のデータセットから特定される第二の特徴ベクトルの一群であって、前記第二の特徴ベクトルのそれぞれが、前記複数の第二のエンティティのうちの対応する一つのエンティティの特徴を表す第二の特徴ベクトルの一群に対する次元削減処理を実行することによって、前記第一の特徴ベクトルの一群に対応する第一の低次元特徴ベクトルの一群、及び、前記第二の特徴ベクトルの一群に対応する第二の低次元特徴ベクトルの一群であって前記第一の低次元特徴ベクトルの一群と同次元数の第二の低次元特徴ベクトルの一群を生成するように構成される次元削減部と、
- 前記第一の低次元特徴ベクトルの一群、及び、前記第二の低次元特徴ベクトルの一群に基づき、前記複数の第一のエンティティのそれぞれを、前記複数の第二のエンティティの少なくとも一つ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、前記第一の低次元特徴ベクトルの一群から特定される前記第一のエンティティ間の類似度、及び、前記第二の低次元特徴ベクトルの一群から特定される前記第二のエンティティ間の類似度に基づき、類似度に関する前記第一のエンティティ間の相互関係が

前記第二のエンティティ間の相互関係に適合するように、前記複数の第一のエンティティのそれぞれを、前記複数の第二のエンティティの少なくとも一つに対応付ける請求項 1 記載の情報処理システム。

[請求項3]

前記第一の低次元特徴ベクトルの一群は、第一の特徴空間によって定義され、

前記第二の低次元特徴ベクトルの一群は、第二の特徴空間によって定義され、

前記対応付け部は、前記第一の低次元特徴ベクトルの一群から特定される前記第一の特徴空間における前記複数の第一のエンティティの分布が、前記第二の低次元特徴ベクトルの一群から特定される前記第二の特徴空間における前記複数の第二のエンティティの分布に適合するように、前記第一の特徴空間上の前記複数の第一のエンティティを前記第二の特徴空間にマッピングするための写像を探索し、前記写像に基づいて、前記複数の第一のエンティティのそれぞれを、前記複数の第二のエンティティの少なくとも一つに対応付ける請求項 1 又は請求項 2 記載の情報処理システム。

[請求項4]

前記対応付け部は、行列 K、行列 L、及び行列 H を含む式

[数1]

$$\begin{aligned}\Omega^* &= \operatorname{argmax}_{\Omega} Z(\Omega) \\ Z(\Omega) &= \operatorname{trace}(K' \Omega^T L' \Omega) \\ K' &= HKH \\ L' &= HLH\end{aligned}$$

に従う値 $Z(\Omega)$ を最大化する行列 Ω を行列 Ω^* として探索し、前記行列 Ω^* に基づいて、前記複数の第一のエンティティのそれぞれを、前記複数の第二のエンティティの少なくとも一つに対応付けるように構成され、

前記第一のエンティティの数は、Nであり、

前記第二のエンティティの数は、前記第一のエンティティと同じで

あり、

前記行列 K は、第 i 行第 j 列の要素の値が、前記複数の第一のエンティティのうちの i 番目のエンティティと j 番目のエンティティとの間の類似度を表し、前記複数の第一のエンティティのうちの前記 i 番目のエンティティの第一の低次元特徴ベクトルと、前記複数の第一のエンティティのうちの前記 j 番目のエンティティの第一の低次元特徴ベクトルと、に基づいて算出される N 行 N 列の第一の類似度行列であり、

前記行列 L は、第 i 行第 j 列の要素の値が、前記複数の第二のエンティティのうちの i 番目のエンティティと j 番目のエンティティとの間の類似度を表し、前記複数の第二のエンティティのうちの前記 i 番目のエンティティの第二の低次元特徴ベクトルと、前記複数の第二のエンティティのうちの前記 j 番目のエンティティの第二の低次元特徴ベクトルと、に基づいて算出される N 行 N 列の第二の類似度行列であり、

前記行列 H は、第 i 行第 j 列の要素の値が、 $i = j$ であるとき値 $1 - 1/N$ を示し、 $i \neq j$ であるとき値 0 を示す N 行 N 列の行列である
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]

前記対応付け部は、前記行列 Ω^* に関する再探索処理を、所定条件が満足されるまで繰返し実行することにより、前記行列 Ω^* を改善し、改善された行列 Ω^* に基づいて、前記複数の第一のエンティティのそれぞれを、前記複数の第二のエンティティの少なくとも一つに対応付けるように構成され、

前記再探索処理は、

前記行列 Ω^* に基づいて、前記次元削減処理での次元削減方式を変更することと、

変更後の前記次元削減方式に基づく前記次元削減処理を、前記次元削減部に実行させ、

それにより新たに得られた前記第一の低次元特徴ベクトルの一群及び前記第二の低次元特徴ベクトルの一群に基づいて、前記行列 Ω^* を再探索することと、

を含む請求項4記載の情報処理システム。

[請求項6]

前記対応付け部は、前記次元削減方式を、前記第一の低次元特徴ベクトルの一群及び前記第二の低次元特徴ベクトルの一群のうち、互いに対応する第一の低次元特徴ベクトルと第二の低次元特徴ベクトルと間の特徴空間上の距離が短くなるように変更する請求項5記載の情報処理システム。

[請求項7]

前記第一のデータセットは、複数の第一の特徴データを含み、前記複数の第一の特徴データのそれぞれは、前記複数の第一のエンティティのうちの対応する一つのエンティティの特徴を表し、

前記第二のデータセットは、複数の第二の特徴データを含み、前記複数の第二の特徴データのそれぞれは、前記複数の第二のエンティティのうちの対応する一つのエンティティの特徴を表し、

前記情報処理システムは、

前記対応付け部による前記複数の第一のエンティティと前記複数の第二のエンティティとの間の対応付けに基づき、前記複数の第一の特徴データのそれぞれに、前記複数の第二の特徴データのうちの一つを結合することによって、複数の拡張データを含む拡張データセットであって、前記複数の拡張データのそれぞれが、対応する一つの第一の特徴データと第二の特徴データとの結合データである複数の拡張データを含む拡張データセットを生成するデータフュージョン部

を更に備える請求項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]

前記第一のエンティティ及び前記第二のエンティティは、人であり、

前記第一のデータセットは、第一の集団に属する複数の人のそれぞれ

れの第一の特徴を記述するデータセットであり、

前記第二のデータセットは、第二の集団の属する複数の人のそれぞれの第二の特徴を記述するデータセットである

請求項 1 ～請求項 7 のいずれか一項記載の情報処理システム。

[請求項9]

前記第一の特徴と前記第二の特徴との組合せは、購買行動に関する特徴と、オンライン空間及びオフライン空間の少なくとも一方の空間における移動に関する特徴、及び／又は、前記空間上の複数の地点への訪問に関する特徴と、の組合せである請求項 8 記載の情報処理システム。

[請求項10]

前記第一のエンティティ及び前記第二のエンティティは、人であり、

前記第二のデータセットには、前記複数の第二のエンティティのそれぞれに対応する情報端末の識別情報が関連付けられており、

前記情報処理システムは、

前記複数の第二のエンティティのうち、前記対応付け部により前記複数の第一のエンティティのいずれかと対応付けられた第二のエンティティの集合の少なくとも一部を、情報コンテンツの配信先に選択する選択部と、

前記識別情報に基づいて、前記情報コンテンツの配信先に対応する情報端末の集合に、前記情報コンテンツを配信するように構成される配信部と、

を備える請求項 1 ～請求項 9 のいずれか一項記載の情報処理システム。

[請求項11]

前記選択部は、前記対応付け部により前記複数の第一のエンティティのいずれかと対応付けられた前記第二のエンティティの集合である第一の集合と、前記複数の第二のエンティティのうち、前記第一の集合と特徴が類似する第二の集合と、を前記情報コンテンツの配信先に選択する請求項 10 記載の情報処理システム。

- [請求項12] 前記第一のエンティティ及び前記第二のエンティティは、人であり、
- 前記第二のデータセットは、前記複数の第二のエンティティのそれぞれの行動に関する特徴を記述し、
- 前記情報処理システムは、更に、
- 一以上の注目エンティティに関して、注目エンティティ毎に、対応する注目エンティティの行動に関する推定値を算出する推定部を備え、
- 前記一以上の注目エンティティは、前記複数の第一のエンティティの少なくとも一部であり、
- 前記推定値は、前記対応する注目エンティティに対応付けられた前記複数の第二のエンティティの少なくとも一つの行動に関する特徴に基づき算出される請求項1～請求項11のいずれか一項記載の情報処理システム。
- [請求項13] コンピュータにより実行される情報処理方法であって、
- 複数の第一のエンティティに関する第一のデータセットであって、前記複数の第一のエンティティのそれぞれの特徴を記述する第一のデータセットを取得することと、
- 複数の第二のエンティティに関する第二のデータセットであって、前記複数の第二のエンティティのそれぞれの特徴を記述する第二のデータセットを取得することと、
- 前記第一のデータセットから特定される第一の特徴ベクトルの一群であって、前記第一の特徴ベクトルのそれぞれが、前記複数の第一のエンティティのうちの対応する一つのエンティティの特徴を表す第一の特徴ベクトルの一群、及び、前記第二のデータセットから特定される第二の特徴ベクトルの一群であって、前記第二の特徴ベクトルのそれぞれが、前記複数の第二のエンティティのうちの対応する一つのエンティティの特徴を表す第二の特徴ベクトルの一群に対する次元削減

処理を実行することによって、前記第一の特徴ベクトルの一群に対応する第一の低次元特徴ベクトルの一群、及び、前記第二の特徴ベクトルの一群に対応する第二の低次元特徴ベクトルの一群であって前記第一の低次元特徴ベクトルの一群と同次元数の第二の低次元特徴ベクトルの一群を生成することと、

前記第一の低次元特徴ベクトルの一群、及び、前記第二の低次元特徴ベクトルの一群に基づき、前記複数の第一のエンティティのそれぞれを、前記複数の第二のエンティティの少なくとも一つに対応付けることと、

を含む情報処理方法。

[請求項14]

前記対応付けることは、

前記第一の低次元特徴ベクトルの一群から特定される前記第一のエンティティ間の類似度、及び、前記第二の低次元特徴ベクトルの一群から特定される前記第二のエンティティ間の類似度に基づき、類似度に関する前記第一のエンティティ間の相互関係が前記第二のエンティティ間の相互関係に適合するように、前記複数の第一のエンティティのそれぞれを、前記複数の第二のエンティティの一つに対応付けること

を含む請求項13記載の情報処理方法。

[請求項15]

コンピュータに請求項13又は請求項14記載の情報処理方法を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。