# 明　細　書

## 発明の名称 ：

## 推定システム、データ結合システム、及び関連する方法

## 関連出願の相互参照

[0001] 本国際出願は、２０１９年３月２８日に日本国特許庁に出願された日本国特許出願第２０１９－６２４５５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２０１９－６２４５５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、推定システム、データ結合システム、及び、関連する方法に関する。

## 背景技術

[0003] 消費者行動を分析することが従来行われている。例えば、消費者の購買行動を分析して、商品に対する潜在的な購買層を推定することが行われている。同一人物に関する複数種類のデータを、統一識別情報を用いて結びつける技術も知られている（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２０１８－５５３６号公報

## 発明の概要

## 発明が解決しようとする課題

[0005] しかしながら、複数種類のデータを収集して結びつける従来手法では、データが欠落している人の特徴を分析することができない。

[0006] そこで、本開示の一側面によれば、取得可能なデータから間接的に対象の特徴を推定するための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

[0007] 本開示の一側面に係る推定システムは、第一取得部と、第二取得部と、モデル構築部と、対象データ取得部と、推定部とを備える。第一取得部は、第一のデータを取得するように構成される。第一のデータは、複数の第一の対象体のそれぞれの第一の特徴を表すデータ、及び、第一の特徴が共通する複数の第一の対象体を表すデータの一方を含む。

[0008] 第二取得部は、複数の第一の対象体のそれぞれの第二の特徴を表す第二のデータを取得するように構成される。モデル構築部は、第一のデータ及び第二のデータから判別される複数の第一の対象体のそれぞれの第一の特徴及び第二の特徴に基づき、第二の特徴から第一の特徴を推定するための推定モデルを構築するように構成される。

[0009] 対象データ取得部は、第二の対象体の第二の特徴を表す対象データを取得するように構成される。推定部は、対象データに基づき、推定モデルを用いて、第二の対象体の第一の特徴を推定するように構成される。

[0010] この推定システムによれば、第二の対象体に関して、第一の特徴に関するデータを取得することができない環境でも、第一の対象体と第二の対象体とが共通して有する第二の特徴を表すデータに基づいて、第二の対象体の第一の特徴を推定することができる。従って、この推定システムによれば、直接的には判別できない第二の対象体の第一の特徴を間接的に推定することができる。

[0011] 本開示の一側面によれば、第一の対象体及び第二の対象体は、人であり得る。第二の特徴は、人の通信を伴う行動に関する特徴であってもよい。第二取得部は、データベースシステムから、第二のデータとして、複数の第一の対象体のそれぞれに関する第一の行動履歴データを取得してもよい。対象データ取得部は、第二の対象体に関する第二の行動履歴データを対象データとしてデータベースシステムから取得してもよい。データベースシステムは、第一の行動履歴データ及び第二の行動履歴データを含む、複数の人のそれぞれの通信を伴う行動に関する行動履歴データを記憶していてもよい。

[0012] 本開示の一側面によれば、通信を伴う行動は、対応する人の行動に関連す

る情報が通信によりデータベースシステムに送信され、関連する情報に基づき対応する人の行動が行動履歴データに記録される行動であり得る。

[0013] 本開示の一側面によれば、通信を伴う行動は、対応する人のオンライン空間上の行動であるオンライン行動であってもよい。データベースシステムにおける複数の人のそれぞれの行動履歴データは、対応する人のオンライン行動に関するオンライン行動履歴データを含んでいてもよい。オンライン行動履歴データは、対応する人のオンライン識別子と関連付けて記憶されていてもよい。

[0014] 本開示の一側面によれば、第一のデータは、複数の第一の対象体のそれぞれに対応するオンライン識別子を判別可能な情報を含んでいてもよい。第二取得部は、データベースシステムから、第二のデータとして、複数の第一の対象体のそれぞれに対応するオンライン識別子に関連付けられたオンライン行動履歴データを取得してもよい。

[0015] 本開示の一側面によれば、対象データ取得部は、第二の対象体に対応するオンライン識別子を判別可能なデータの入力に応じて、データベースシステムから、対象データとして、第二の対象体に対応するオンライン識別子に関連付けられたオンライン行動履歴データを取得してもよい。

[0016] 本開示の一側面によれば、オンライン行動は、複数のウェブページに対する閲覧行動を含んでいてもよい。

[0017] 本開示の一側面によれば、通信を伴う行動は、現実空間に配置された電子機器により検知される対応する人のオフライン行動であってもよい。電子機器により検知された対応する人のオフライン行動に関連する情報は、通信によりデータベースシステムに送信され得る。データベースシステムにおける複数の人のそれぞれの行動履歴データは、電子機器により検知された対応する人のオフライン行動に関連するオフライン行動履歴データを含んでいてもよい。

[0018] 本開示の一側面によれば、第一の特徴は、動的な特徴であってもよい。同様に、第二の特徴は、動的な特徴であってもよい。第一のデータは、第一の

特徴の変化を判別可能な情報を含んでいてもよい。第二のデータは、第二の特徴の変化を判別可能な情報を含んでいてもよい。対象データは、第二の特徴の変化を判別可能な情報を含んでいてもよい。

[0019] モデル構築部は、第一の特徴の変化及び第二の特徴の変化に基づき、第二の特徴の変化から第一の特徴の変化を推定するための推定モデルを構築するように構成されてもよい。推定部は、対象データに基づき、推定モデルを用いて、第二の対象体の第一の特徴の変化を推定するように構成されてもよい。動的な特徴に関するデータ収集の負荷は、静的な特徴に関するデータ収集より高い。従って、動的な特徴の変化を推定できることは有意義である。

[0020] 本開示の一側面によれば、複数の第一の対象体及び第二の対象体のそれぞれは、人であってもよい。人は、個人、及び、複数の個人を含むクラスタの一方であると理解されてもよい。本開示の一側面によれば、複数の第一の対象体及び第二の対象体のそれぞれは、消費者であってもよい。消費者が、個人、及び、複数の個人を含むクラスタの一方であってもよい。

[0021] 本開示の一側面によれば、第一の特徴は、消費者の第一の行動に関する特徴であってもよい。第二の特徴は、消費者の第二の行動に関する特徴であってもよい。

[0022] 本開示の一側面によれば、第一のデータは、複数の第一の対象体のそれぞれの第一の行動に関する行動履歴を表すデータであってもよい。第二のデータは、複数の第一の対象体のそれぞれの第二の行動に関する行動履歴を表すデータであってもよい。対象データは、第二の対象体の第二の行動に関する行動履歴を表すデータであってもよい。

[0023] モデル構築部は、第一のデータに基づいて判別される第一の対象体のそれぞれの第一の行動に関する行動履歴と、第二のデータに基づいて判別される第一の対象体のそれぞれの第二の行動に関する行動履歴とに基づいて、対象データから第二対象体の第一の行動に関する状態変化を推定するための推定モデルを構築するように構成されてもよい。

[0024] 本開示の一側面によれば、第一の特徴は、第一の行動の実行の有無に関す

る特徴であってもよい。モデル構築部は、複数の第一の対象体のそれぞれの第二の特徴に基づき、推定モデルとして、第二の特徴から第一の行動が行われる可能性を推定するための推定モデルを構築するように構成されてもよい。推定部は、推定モデルを用いて、第二の対象体が第一の行動を行う可能性を推定するように構成されてもよい。

[0025] 推定システムは、複数の第二の対象体のそれぞれの第三の特徴を表す第三のデータを取得するように構成される第三取得部を備えていてもよい。対象データ取得部は、対象データとして、複数の第二の対象体のそれぞれの第二の特徴を表すデータを取得するように構成されてもよい。推定部は、対象データに基づき、推定モデルを用いて、複数の第二の対象体のそれぞれの第一の特徴を推定してもよい。推定部は、第三のデータに対する拡張データとして、複数の第二の対象体のそれぞれの第一の特徴及び第三の特徴を表すデータを出力するように構成されてもよい。

[0026] 本開示の一側面によれば、推定システムであって、第一の行動の実行の有無に関して共通する特徴を有する複数の第一の対象体を表す第一のデータを取得するように構成される第一取得部と、第一のデータから判別される複数の第一の対象体のそれぞれの第二の行動に関する特徴を表す第二のデータを取得するように構成される第二取得部と、第二のデータから判別される複数の第一の対象体のそれぞれの第二の行動に関する特徴に基づき、推定モデルとして、第二の行動に関する特徴から第一の行動が行われる可能性を推定するための推定モデルを構築するように構成されるモデル構築部と、第二の対象体の第二の行動に関する特徴を表す対象データを取得するように構成される対象データ取得部と、対象データに基づき、推定モデルを用いて、第二の対象体が第一の行動を行う可能性を推定するように構成される推定部と、を備える推定システムが提供されてもよい。

[0027] 本開示の一側面によれば、データ結合システムが提供されてもよい。データ提供システムは、第一取得部と、第二取得部と、第三取得部と、第四取得部と、データ結合部と、を備えてもよい。

[0028] データ結合システムにおいて、第一取得部は、複数の第一の対象体に関する複数の第一の特徴データを取得するように構成されてもよい。複数の第一の特徴データのそれぞれは、対応する第一の対象体の第一の特徴を表すデータであってもよい。

[0029] 第二取得部は、複数の第一の対象体に関する複数の第二の特徴データを取得するように構成されてもよい。第二の特徴データのそれぞれは、対応する第一の対象体の第二の特徴を表すデータであってもよい。

[0030] 第三取得部は、複数の第二の対象体に関する複数の第三の特徴データを取得するように構成されてもよい。第三の特徴データのそれぞれは、対応する第二の対象体の第三の特徴を表すデータであってもよい。

[0031] 第四取得部は、複数の第二の対象体に関する複数の第四の特徴データを取得するように構成されてもよい。第四の特徴データのそれぞれは、対応する第二の対象体の第二の特徴を表すデータであってもよい。

[0032] データ結合部は、複数の第二の特徴データから判別される複数の第一の対象体のそれぞれの第二の特徴と、複数の第四の特徴データから判別される複数の第二の対象体のそれぞれの第二の特徴と、に基づき、第二の特徴が少なくとも類似する第一の対象体と第二の対象体との組み合わせを判別するように構成されてもよい。

[0033] データ結合部は、組み合わせ毎に、対応する第一の対象体の第一の特徴データと、対応する第二の対象体の第三の特徴データとを関連付けて、組み合わせ毎の第一の特徴及び第三の特徴を表す結合データベースを生成するように構成されてもよい。

[0034] このデータ結合システムによれば、第二の特徴の類似性から、第一の特徴と第三の特徴とを関連づけることができる。従って、第二の対象体の第一の特徴をデータから直接的には判別できない環境において、第二の対象体の第一の特徴を、結合データベースから間接的に推定することができる。

[0035] 本開示の一側面によれば、データ結合部は、組み合わせ毎に、対応する第一の対象体の第一の特徴データと、対応する第二の対象体の第三の特徴デー

タとを、対応する第一の対象体の第二の特徴データ及び対応する第二の対象体の第四の特徴データを介して関連付けるように構成されてもよい。これにより、データ結合部は、結合データベースとして、組み合わせ毎に、第一の特徴、第二の特徴、及び第三の特徴を表す結合データベースを生成するように構成されてもよい。

[0036] 本開示の一側面によれば、データ結合システムにおける複数の第一の対象体及び第二の対象体のそれぞれは、人であってもよい。人は、個人、及び、複数の個人を含むクラスタの一方であると理解されてもよい。第二の特徴は、人の通信を伴う行動に関する特徴であってもよい。第二の特徴は、オンライン行動に関する特徴であってもよい。

[0037] 本開示の一側面によれば、第二取得部は、複数のオンラインユーザのそれぞれのオンライン行動に関する行動履歴データを、対応するオンラインユーザのオンライン識別子と関連付けて記憶するデータベースシステムから、複数の第二の特徴データとして、複数の第一の対象体のそれぞれに対応するオンライン識別子に関連付けられた行動履歴データを取得するように構成されてもよい。この場合、複数の第一の特徴データは、複数の第一の対象体のそれぞれに対応するオンライン識別子を判別可能な情報を含んでいてもよい。

[0038] 第四取得部は、上記データベースシステムから、複数の第四の特徴データとして、複数の第二の対象体のそれぞれに対応するオンライン識別子に関連付けられた行動履歴データを取得するように構成されてもよい。この場合、複数の第三の特徴データは、複数の第二の対象体のそれぞれに対応するオンライン識別子を判別可能な情報を含んでいてもよい。

[0039] 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推定方法が提供されてもよい。推定方法は、複数の第一の対象体のそれぞれの第一の特徴を表すデータ、及び、第一の特徴が共通する複数の第一の対象体を表すデータの一方を含む第一のデータを取得することと、複数の第一の対象体のそれぞれの第二の特徴を表す第二のデータを取得することと、第一のデータ及び第二のデータから判別される複数の第一の対象体のそれぞれの第一の特徴及び

第二の特徴に基づき、第二の特徴から第一の特徴を推定するための推定モデルを構築することと、第二の対象体の第二の特徴を表す対象データを取得することと、対象データに基づき、推定モデルを用いて、第二の対象体の第一の特徴を推定することと、を含んでいてもよい。

[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る推定方法であって、第一の行動の実行の有無に関して共通する特徴を有する複数の第一の対象体を表す第一のデータを取得することと、第一のデータから判別される複数の第一の対象体のそれぞれの第二の行動に関する特徴を表す第二のデータを取得することと、第二のデータから判別される複数の第一の対象体のそれぞれの第二の行動に関する特徴に基づき、推定モデルとして、第二の行動に関する特徴から第一の行動が行われる可能性を推定するための推定モデルを構築することと、第二の対象体の第二の行動に関する特徴を表す対象データを取得することと、対象データに基づき、推定モデルを用いて、第二の対象体が第一の行動を行う可能性を推定することと、を含む推定方法が提供されてもよい。

[0041] 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データ結合方法が提供されてもよい。データ結合方法は、複数の第一の対象体に関する複数の第一の特徴データであって、それぞれが対応する第一の対象体の第一の特徴を表す複数の第一の特徴データを取得することと、複数の第一の対象体に関する複数の第二の特徴データであって、それぞれが対応する第一の対象体の第二の特徴を表す複数の第二の特徴データを取得することと、複数の第二の対象体に関する複数の第三の特徴データであって、それぞれが対応する第二の対象体の第三の特徴を表す複数の第三の特徴データを取得することと、複数の第二の対象体に関する複数の第四の特徴データであって、それぞれが対応する第二の対象体の第二の特徴を表す複数の第四の特徴データを取得することと、複数の第二の特徴データから判別される複数の第一の対象体のそれぞれの第二の特徴と、複数の第四の特徴データから判別される複数の第二の対象体のそれぞれの第二の特徴と、に基づき、第二の特徴が少なくとも類

似する第一の対象体と第二の対象体との組み合わせ毎に、第一の対象体の第一の特徴データと、第二の対象体の第三の特徴データとを関連付けて、組み合わせ毎の第一の特徴及び第三の特徴を表す結合データベースを生成することと、を含んでいてもよい。

[0042] 本開示の一側面によれば、コンピュータに上述の推定方法を実行させる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、コンピュータに上述のデータ結合方法を実行させる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。

[0043] 本開示の一側面によれば、上述の推定システムに対応する処理をコンピュータに実行させる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、上述のデータ結合システムに対応する処理をコンピュータに実行させる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。

[0044] 本開示の一側面によれば、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てもよい。

## 図面の簡単な説明

[0045] [図1]第一実施形態の推定手法を説明した図である。

[図2]第一実施形態の情報処理システムの構成及び周辺構成を説明した図である。

[図3]プロセッサが実行するモデル構築処理を表すフローチャートである。

[図4]推定モデルの構築に関する説明図である。

[図5]プロセッサが実行する推定処理を表すフローチャートトートである。

[図6]第二実施形態のモデル構築処理を表すフローチャートである。

[図7]図７Ａ及び図７Ｂは、拡張後の購買履歴データベースの構成を表す図である。

[図8]状態変化の判別に関する説明図である。

[図9]第二実施形態の推定処理を表すフローチャートである。

[図10]図１０Ａは拡張前後の顧客データベースの構成を表す図であり、図１

０Ｂは、クラスタ毎の顧客データを有する別例の顧客データベースの構成を表す図である。

[図11]第三実施形態の情報処理システムの構成及び周辺構成を説明した図である。

[図12]第三実施形態のデータ結合方法を説明した図である。

[図13]プロセッサが実行するデータ結合処理を表すフローチャートである。

[図14]拡張前後の第一行動データベースの構成を表す図である。

[図15]拡張前後の第二行動データベースの構成を表す図である。

[図16]結合データベースの構成を例示する図である。

[図17]図１７Ａは、オンライン行動に関するデータ収集方法を説明した図であり、図１７Ｂは、オフライン行動に関するデータ収集方法を説明した図である。

## 符号の説明

[0046] 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、１７…通信インタフェース、２０…データ収集システム、３０…利用者システム、４０…購買データ収集システム、５０…第一行動データ収集システム、６０…オンライン行動データ収集システム、７０…第二行動データ収集システム、Ｂ１，Ｂ２，Ｂ３…購買履歴データベース、Ｂ５，Ｂ６，Ｂ７…顧客データベース、Ｂ１１，Ｂ１２…第一行動データベース、Ｂ２１，Ｂ２２…第二行動データベース、Ｂ３０…結合データベース、Ｄ１…第一データ、Ｄ２…第二データ、Ｄ３…オンライン行動データ、Ｄ４…教師データ、Ｄ５…スコアデータ、Ｄ７…オフライン行動データ、Ｄ１１…第一行動データ、Ｄ１２…拡張第一行動データ、Ｄ２１…第二行動データ、Ｄ２２…拡張第二行動データ、Ｌ１…第一消費者リスト、Ｌ２…第二消費者リスト、Ｍ…推定モデル、ＥＬ…電子機器、ＴＭ…通信端末。

## 発明を実施するための形態

[0047] 以下に、本開示の例示的実施形態を、図面を参照しながら説明する。

[0048] ［第一実施形態］

本実施形態の情報処理システム１は、図１に示すように、第一消費者群Ｃ１の第一の行動に関する第一データＤ１、並びに、第一消費者群Ｃ１及び第二消費者群Ｃ２の第二の行動に関する第二データＤ２に基づいて、第二消費者群Ｃ２の第一の行動を推定するように構成される。

[0049] 情報処理システム１は、第一データＤ１から判別される第一消費者群Ｃ１の第一の行動と、第二データＤ２から判別される第一消費者群Ｃ１の第二の行動と、の関係に基づいて、第二の行動から、第一の行動を推定するための推定モデルＭを構築する。

[0050] 情報処理システム１は、推定モデルＭと、第二データＤ２から判別される第二消費者群Ｃ２の第二の行動とに基づいて、第二消費者群Ｃ２の第一の行動を推定する。推定は、過去の行動を推定すること、及び、未来の行動を予測することを含む。

[0051] 情報処理システム１は、図２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、通信インタフェース１７と、を備える。プロセッサ１１は、ストレージ１５が記憶するコンピュータプログラムに従う処理を実行するように構成される。

[0052] メモリ１３は、ＲＯＭ及びＲＡＭを含む。ストレージ１５は、コンピュータプログラム、及び各種データを記憶する。ストレージ１５の例には、ハードディスクドライブ及びソリッドステートドライブが含まれる。

[0053] 通信インタフェース１７は、ネットワークを通じて外部装置と通信可能に構成される。情報処理システム１は、通信インタフェース１７を通じて、データ収集システム２０と通信可能に構成される。データ収集システム２０は、消費者に関するデータを収集するように構成される。更に、情報処理システム１は、利用者システム３０と通信可能に構成される。利用者システム３０は、情報処理システム１が提供するサービスを利用する利用者の情報処理システムである。

[0054] 情報処理システム１は、通信により、利用者システム３０から、第一消費者群Ｃ１の第一の行動に関する第一データＤ１を取得する。情報処理システ

ム１は更に、データ収集システム２０から第一消費者群Ｃ１及び第二消費者群Ｃ２の第二の行動に関する第二データＤ２を取得する。情報処理システム１は、第一データＤ１及び第二データＤ２に基づき、第二消費者群Ｃ２の第一の行動を推定する。情報処理システム１は、推定結果を利用者システム３０に提供する。

[0055] 情報処理システム１が実行する処理の具体例を、図３に示すモデル構築処理及び図５に示す推定処理を用いて説明する。モデル構築処理及び推定処理は、プロセッサ１１により実行される。

[0056] この具体例において、第一の行動は、利用者システム３０に対応する企業が消費者に向けて提供する特定サービスに対する消費者の利用契約行動である。第二の行動は、消費者のオンライン行動、具体的には、ウェブページの閲覧行動である。オンライン行動は、消費者の通信を伴う行動に対応する。

[0057] 図３及び図４で説明される具体例では、情報処理システム１が、利用契約の実績を有する消費者のオンライン行動に基づき、推定モデルＭを構築し、利用契約の実績のない消費者のオンライン行動から、推定モデルＭを用いて、利用契約の可能性に関するスコアを算出するように構成される。

[0058] 現代では、多くの消費者がインターネットにアクセスし、ウェブページを閲覧している。このため、オンライン行動に関しては、多くの消費者のデータを収集できる。更に言えば、クッキー（Ｃｏｏｋｉｅ）技術を利用すれば、ウェブページの閲覧行動等のオンライン行動に関するデータを比較的容易に収集することができる。このような理由から、第二の行動としてオンライン行動を採用することは、第一の行動に関する高精度な推定モデルＭを構築するのに役立つ。

[0059] プロセッサ１１は、利用者システム３０からネットワークを通じてモデル構築処理の実行指令を受信すると、図３に示すモデル構築処理を開始する。

[0060] モデル構築処理において、プロセッサ１１は、利用者システム３０から、第一消費者群Ｃ１のリストである第一消費者リストＬ１を受信する（Ｓ１１０）。第一消費者リストＬ１は、利用者システム３０に対応する企業の顧客

のうち、特定サービスの利用契約の実績を有する複数の消費者を表す。第一消費者リストＬ１は例えば、複数の消費者の顧客ＩＤが列挙された構成にされる。顧客ＩＤは、例えば第一消費者リストＬ１を提供する企業が管理する顧客ＩＤである。

[0061] 続くＳ１２０において、プロセッサ１１は、第一消費者リストＬ１に列挙された複数の消費者の顧客ＩＤのそれぞれを、変換テーブルを参照して、クッキーＩＤに変換する（Ｓ１２０）。プロセッサ１１は、変換テーブルを、データ収集システム２０から取得することができる。クッキーＩＤは、クッキー技術に基づいてオンラインユーザ（具体的にはウェブブラウザ）に発行される識別子である。

[0062] 顧客ＩＤとクッキーＩＤとの対応付けは、例えば、第一消費者リストＬ１を提供する企業の協力によって実現することができる。対応付けは、企業のウェブページに、データ収集システム２０が管理するクッキーを発行するためのタグを埋め込むことで実現され得る。

[0063] 続くＳ１３０において、プロセッサ１１は、第一消費者リストＬ１に列挙された複数の消費者のそれぞれのオンライン行動データＤ３を、クッキーＩＤに基づいてデータ収集システム２０から取得する。データ収集システム２０は、複数のオンラインユーザのオンライン行動データＤ３を、対応するクッキーＩＤに関連付けて記憶する。

[0064] データ収集システム２０は、クッキーの発行及びクッキーによるオンラインユーザの行動分析によって、複数のオンラインユーザのオンライン行動データＤ３を生成及び記憶するように構成される。データ収集システム２０は、情報処理システム１から指定されたクッキーＩＤに関連付けられたオンライン行動データＤ３を、情報処理システム１に提供するように構成される。

[0065] オンライン行動データＤ３は、対応するオンラインユーザによるウェブページの閲覧履歴データと、対応するオンラインユーザのアフィニティカテゴリを示すデータと、を含む。閲覧履歴データは、対応するオンラインユーザのウェブページの閲覧行動に関する行動履歴データである。

[0066] 具体的に、閲覧履歴データは、対応するオンラインユーザにより閲覧されたウェブページ毎に、ウェブページの識別情報と、閲覧日時の情報と、を有する。アフィニティカテゴリは、対応するオンラインユーザが関心及び興味を有するカテゴリを示す。

[0067] 続くＳ１４０において、プロセッサ１１は、第一消費者リストＬ１における複数の消費者に対応する複数のオンライン行動データＤ３を用いて、これら複数のオンライン行動データＤ３から、特定サービスの利用契約の可能性を算出するための推定モデルＭを構築する。

[0068] 具体的に、プロセッサ１１は、Ｓ１３０において得られた利用契約の実績を有する複数の消費者のオンライン行動データＤ３に基づいて、複数の教師データＤ４を生成することができる。教師データＤ４のそれぞれは、図４に示すように、対応するオンライン行動データＤ３から判別される消費者のウェブ閲覧履歴及びアフィニティカテゴリと、利用契約の可能性に関するスコアとの組み合わせであり得る。

[0069] 例えば、教師データＤ４において、ウェブ閲覧履歴は、複数のウェブページのそれぞれ閲覧の有無を表す数値又は閲覧回数を示す配列で表され得る。スコアは、利用契約確率（％）で表され得る。第一消費者リストＬ１内の消費者はすべて、利用契約の実績を有する。このため、スコアは、値１００に設定され得る。複数の教師データＤ４には、第一消費者リストＬ１には含まれない消費者のオンライン行動データＤ３に基づいて生成される、利用契約の実績を有さない消費者に関する教師データが含まれていてもよい。

[0070] プロセッサ１１は、このような複数の教師データＤ４を用いた機械学習により、サイト閲覧履歴及びアフィニティカテゴリの入力に対して、利用契約確率に関するスコアを算出する推定モデルＭを構築することができる。推定モデルＭは、数理モデルである。推定モデルＭの例には、ロジスティック回帰モデル、ベイズ推定モデル、及びニューラルネットワークモデルなどが含まれる。

[0071] 続くＳ１５０において、プロセッサ１１は、構築した推定モデルＭを出力

する。その後、モデル構築処理を終了する。Ｓ１５０において、プロセッサ１１は、推定モデルＭを、メモリ１３又はストレージ１５に記憶することができる。あるいは、プロセッサ１１は、推定モデルＭを利用者システム３０に送信することができる。

[0072] プロセッサ１１は更に、利用者システム３０からネットワークを通じて推定処理の実行指令を受信すると、図５に示す推定処理を開始する。

[0073] 推定処理を開始すると、プロセッサ１１は、利用者システム３０から、第二消費者群Ｃ２のリストである第二消費者リストＬ２を受信する（Ｓ２１０）。第二消費者リストＬ２は、上記特定サービスの利用契約の可能性を推定する対象の複数の消費者を表す。第二消費者リストＬ２も、第一消費者リストＬ１と同様に、複数の消費者の顧客ＩＤが列挙された構成にされる。

[0074] 続くＳ２２０において、プロセッサ１１は、第二消費者リストＬ２に列挙された複数の消費者の顧客ＩＤのそれぞれを、Ｓ１２０での処理と同様に、クッキーＩＤに変換する。続くＳ２３０において、プロセッサ１１は、第二消費者リストＬ２に列挙された複数の消費者のそれぞれのオンライン行動データＤ３を、クッキーＩＤに基づいて、データ収集システム２０から取得する。

[0075] 続くＳ２４０において、プロセッサ１１は、推定モデルＭを取得する。Ｓ２４０において、プロセッサ１１は、利用者システム３０から推定モデルＭを取得することができる。あるいは、プロセッサ１１は、モデル構築処理において構築しストレージ１５に記憶した利用者システム３０の推定モデルＭを、ストレージ１５から読み出すことにより、推定モデルＭを取得することができる。

[0076] 続くＳ２５０において、プロセッサ１１は、推定モデルＭを用いて、第二消費者リストＬ２に列挙された複数の消費者のそれぞれの利用契約確率に関するスコアを算出する。プロセッサ１１は、消費者毎に、対応する消費者のオンライン行動データＤ３から判別されるサイト閲覧履歴及びアフィニティカテゴリの組み合わせを、推定モデルＭに入力することにより、対応する消

費者のスコアを算出することができる。これにより、プロセッサ１１は、第二消費者群Ｃ２に対応する複数の消費者のそれぞれについての特定サービスの利用契約の可能性を推定する。

[0077] 続くＳ２６０において、プロセッサ１１は、第二消費者リストＬ２の送信元に、スコアデータＤ５を送信する。スコアデータＤ５は、第二消費者リストＬ２内の各消費者に対応するスコアを第二消費者リストＬ２に対して追加したデータに対応する。その後、プロセッサ１１は、推定処理を終了する。

[0078] 以上に説明した本実施形態の情報処理システム１では、利用契約に関する特徴が共通する複数の消費者、具体的には利用契約実績のある複数の消費者を表す第一消費者リストＬ１が取得される（Ｓ１１０）。第一消費者リストＬ１内の複数の消費者のそれぞれのオンライン行動に関する特徴を示すオンライン行動データＤ３が取得される（Ｓ１３０）。

[0079] 更に、第一消費者リストＬ１内の消費者それぞれの利用契約に関する特徴及びオンライン行動に関する特徴に基づき、オンライン行動に関する特徴から利用契約に関する特徴としての利用契約の可能性を推定するための推定モデルＭが構築される（Ｓ１４０）。

[0080] 更に、利用契約の可能性を推定する対象の複数の消費者を表す第二消費者リストＬ２が取得される（Ｓ２１０）。第二消費者リストＬ２内の複数の消費者のそれぞれのオンライン行動データＤ３が取得される（Ｓ２３０）。

[0081] その後、第二消費者リストＬ２内の複数の消費者のそれぞれについて、対応する消費者のオンライン行動データＤ３に基づき、対応する消費者の利用契約の可能性が推定される（Ｓ２５０）。本実施形態の情報処理システム１では、このようにして第一消費者群Ｃ１の利用契約実績から、第二消費者群Ｃ２の利用契約の可能性が推定される。

[0082] 従って、本実施形態の情報処理システム１では、取得可能なデータを有意義に活用して、そのデータから直接的には判別できない消費者行動を推定することができる。従って、この情報処理システム１は、大変有意義である。

[0083] ［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３及び図５に示す処理に代えて、図６に示すモデル構築処理及び図９に示す推定処理を実行するように構成される点で、第一実施形態とは異なる。この点を除いて、第二実施形態の情報処理システム１は、第一実施形態と同様に構成される。従って、以下では、プロセッサ１１が実行するモデル構築処理及び推定処理を選択的に説明する。

[0084] プロセッサ１１は、所定条件が満足されると、図６に示すモデル構築処理を開始する。プロセッサ１１は、情報処理システム１の操作者からの指令に基づいて、モデル構築処理を開始することができる。

[0085] 図６に示すモデル構築処理を開始すると、プロセッサ１１は、購買データ収集システム４０から、第一消費者群Ｃ１の購買行動に関する購買履歴データベースＢ１を取得する（Ｓ３１０）。購買履歴データベースＢ１は、第一消費者のそれぞれの購買履歴データを有する。

[0086] 購買履歴データは、対応する消費者の識別情報である消費者ＩＤに関連付けて、対応する消費者のオンライン空間及びオフライン空間での購買毎に、購入した商品の識別情報と、購入日時の情報とを有する。

[0087] 購買データ収集システム４０は、情報処理システム１を運営する企業のシステムであってもよいし、別の企業のシステムであってもよい。プロセッサ１１は、ネットワークを通じて購買履歴データベースＢ１を取得することができる。

[0088] 続くＳ３２０において、プロセッサ１１は、各消費者の購買履歴データに含まれる消費者ＩＤに基づいて、各消費者のクッキーＩＤを判別する。クッキーＩＤの判別は、第一実施形態と同様に、データ収集システム２０から変換テーブルを取得して実現できる。消費者ＩＤとクッキーＩＤとの対応付けは、消費者ＩＤを取り扱うウェブサイトにタグを埋め込むことにより実現することができる。

[0089] 続くＳ３３０において、プロセッサ１１は、第一消費者のそれぞれに対応するオンライン行動データＤ３を、クッキーＩＤに基づいてデータ収集シス

テム２０から取得し、購買履歴データベースＢ１を拡張する。拡張後の購買履歴データベースＢ２は、図７Ａに示される。拡張後の購買履歴データベースＢ２は、第一消費者のそれぞれについて、購買履歴、ウェブ閲覧履歴、アフィニティカテゴリを表すレコードを有する。

[0090] 続くＳ３４０において、プロセッサ１１は、第一消費者のそれぞれについて、対応するレコードが示す消費者の購買履歴から、消費者の状態変化を判別する。その後、プロセッサ１１は、第一消費者のそれぞれのレコードに状態変化を表すデータを追加するように、購買履歴データベースＢ２を拡張する。状態変化を表すデータは、対応する消費者の各時刻における状態を判別可能な情報を含む。

[0091] 拡張後の購買履歴データベースＢ３は、図７Ｂに示される。拡張後の購買履歴データベースＢ３は、第一消費者のそれぞれについて、対応する消費者の購買履歴、状態変化、ウェブ閲覧履歴、アフィニティカテゴリを表すレコードを有する。

[0092] 本実施形態では、第二消費者群Ｃ２の特定状態への変化確率を推定することが考えられる。この場合、Ｓ３４０において、プロセッサ１１は、第一消費者のそれぞれについて、対応する消費者の非特定状態Ｚ１から特定状態Ｚ２への変化を判別することができる。例えば、非特定状態Ｚ１は、特定種類の商品を購入していない状態に対応し、特定状態Ｚ２は、特定種類の商品を購入している状態に対応する。特定種類の商品としては、出産に関係する商品を一例に挙げることができる。

[0093] 図８には、状態変化の判別手法を概念的に説明する。図８に示す例は、消費者の購買履歴が、時間軸において時刻Ｔ１より前では、丸印に対応する第一商品を購入していることを示し、時刻Ｔ１より後では、四角印に対応する第二商品を購入していることを示す。

[0094] 購買履歴データが、このような購買の変化を示す場合、プロセッサ１１は、時刻Ｔ１より前では、第二商品を購入していない非特定状態Ｚ１にあると判別し、時刻Ｔ１より後では、第二商品を購入している特定状態Ｚ２にある

と判別する。そして、購買履歴データベースＢ３に、状態変化を表すデータとして、時刻Ｔ１より前が非特定状態Ｚ１であることを示し、時刻Ｔ１より後が特定状態Ｚ２であることを示すデータを、各消費者のレコードに追加する。

[0095] 続くＳ３５０において、プロセッサ１１は、オンライン行動の変化から消費者の状態変化を推定するための推定モデルＭを構築する。具体的に、プロセッサ１１は、購買履歴データベースＢ３内の各レコードに対応する教師データとして、ウェブ閲覧履歴及びアフィニティカテゴリと、消費者の状態変化とを表す教師データを作成する。

[0096] 教師データにおいてウェブ閲覧履歴は、ウェブの閲覧履歴の収集期間を分割して定義される区間毎に、複数のウェブページのそれぞれの閲覧の有無を表す数値又は閲覧回数を示す配列で表現され得る。教師データにおいて、消費者の状態変化は、区間毎の消費者の状態を表す値を有する配列であり得る。

[0097] プロセッサ１１は、このような教師データの一群に基づき、消費者のウェブ閲覧履歴及びアフィニティカテゴリから、消費者の状態変化を推定する推定モデルＭを構築することができる。推定モデルＭは、ウェブ閲覧履歴及びアフィニティカテゴリの入力に対し、消費者の特定状態への状態変化確率に関するスコアを出力する関数であり得る。

[0098] 続くＳ３６０において、プロセッサ１１は、構築した推定モデルＭをストレージ１５に記憶する。その後、図６に示すモデル構築処理を終了する。

[0099] プロセッサ１１は更に、利用者システム３０からネットワークを通じて推定処理の実行指令を受信すると、図９に示す推定処理を開始する。推定処理を開始すると、プロセッサ１１は、利用者システム３０から、第二消費者群Ｃ２に関する顧客データベースＢ５を受信する（Ｓ４１０）。

[0100] 顧客データベースＢ５は、利用者の顧客である第二消費者のそれぞれの顧客データを備える。顧客データのそれぞれは、対応する消費者の顧客ＩＤと、対応する消費者の特徴データとを備える。顧客ＩＤは、購買履歴データベ

ースＢ１が有する消費者ＩＤとは異なる、利用者システム３０側の企業が管理するＩＤである。

[0101] 続くＳ４２０においてプロセッサ１１は、第二消費者のそれぞれに対応するクッキーＩＤを、顧客データに含まれる顧客ＩＤに基づいて判別する。続くＳ４３０においてプロセッサ１１は、第二消費者のそれぞれのオンライン行動データＤ３を、判別したクッキーＩＤに基づいて、データ収集システム２０から取得する。

[0102] 続くＳ４４０においてプロセッサ１１は、モデル構築処理によって構築及び記憶された推定モデルＭを、ストレージ１５から読み出すことにより、推定モデルＭを取得する。続くＳ４５０において、プロセッサ１１は、取得した推定モデルＭに基づき、第二消費者のそれぞれの状態変化を推定する。

[0103] 具体的には、プロセッサ１１は、状態変化の特徴量として、特定状態への状態変化確率に関するスコアを算出することができる。プロセッサ１１は、消費者毎に、対応する消費者のオンライン行動データＤ３から特定されるウェブ閲覧履歴及びアフィニティカテゴリの組み合わせを、推定モデルＭに入力することにより、対応する消費者のスコアを算出することができる。

[0104] その後、プロセッサ１１は、Ｓ４１０で受信した顧客データベースＢ５に含まれる顧客データに、対応する第二消費者のウェブ閲覧履歴、アフィニティカテゴリ、及び推定された状態変化を表すデータを追加することにより、顧客データベースＢ５を拡張する。拡張後の顧客データベースＢ６は、図１０Ａに示される。

[0105] 拡張後の顧客データベースＢ６は、第二消費者のそれぞれについて、第二消費者の特定状態への状態変化確率に関するスコアを有することができる。プロセッサ１１は、拡張後の顧客データベースＢ６を、利用者システム３０に送信する（Ｓ４６０）。

[0106] 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、利用者の顧客データベースＢ５に対して、各消費者の状態変化に関する情報（例えば特定状態への変化確率の情報）を追加することができる。

[0107] 上述したように本実施形態では、顧客ＩＤをクッキーＩＤに変換して、顧客に対応するオンライン行動データＤ３を取得する。このオンライン行動データＤ３と、予め構築した推定モデルＭとに基づいて、消費者の状態変化を推定する。推定モデルＭは、予めデータ収集により得られた購買履歴から判別される状態変化と、対応する消費者のオンライン行動との関係に基づいて構築される。

[0108] このように、本実施形態の情報処理システム１では、購買行動の変化から判別可能な消費者の状態変化を、購買履歴よりもデータ収集が容易なオンライン行動に関するデータから推定することができる。

[0109] 上述した実施形態では、個人情報が利用者システム３０と情報処理システム１との間でやりとりされる。しかしながら、個人情報保護のため、システム間では、個人単位のデータではなく、個人の集合であるクラスタ単位のデータがやりとりされてもよい。

[0110] 例えば、利用者システム３０から情報処理システム１に提供される顧客データベースＢ５は、個人毎の顧客データではなく、クラスタ毎の顧客データであってもよい。この場合、各クラスタの顧客データは、クラスタに属する複数の顧客の顧客ＩＤと共に、クラスタに属する顧客の特徴が統計量として表現されたデータであり得る。統計量は、例えば平均値であり得る。

[0111] この場合、情報処理システム１のプロセッサ１１は、クラスタ毎に、顧客データが示す顧客ＩＤ群に対応するクッキーＩＤ群に関連付けられたオンライン行動データＤ３を取得することができる。更にプロセッサ１１は、これらのオンライン行動データＤ３と推定モデルＭとに基づいて、クラスタ毎に、状態変化に関する統計量を算出することができる。

[0112] プロセッサ１１は、図１０Ｂに示すように、クラスタ毎に、統計化された顧客データと、ウェブページの閲覧履歴に関する統計量と、アフェリエイトカテゴリに関する統計量と、状態変化に関する統計量と、を有した顧客データベースＢ７を、Ｓ４６０において、利用者システム３０に送信するように構成され得る。

[0113] ［第三実施形態］

続いて、第三実施形態の情報処理システム１を説明する。第三実施形態の情報処理システム１は、図１１に示すように、第一実施形態と同一のハードウェア構成を有する。一方、情報処理システム１は、第一行動データ収集システム５０から得られる第一消費者群Ｃ３の第一の行動に関する第一行動データベースＢ１１と、第二行動データ収集システム７０から得られる第二消費者群Ｃ４の第二の行動に関する第二行動データベースＢ２１とを、結合するように構成されている点で、第一実施形態及び第二実施形態とは異なる。

[0114] 結合は、図１２に示すように、第一消費者群Ｃ３には、第二の行動に関するデータが不足し、第二消費者群Ｃ４には、第一の行動に関するデータが不足している環境の中で、第一消費者群Ｃ３及び第二消費者群Ｃ４のオンライン行動を根拠に行われる。すなわち、本実施形態の情報処理システム１は、第一消費者の第一の行動に関するデータと、第二消費者の第二の行動に関するデータと、をオンライン行動の類似性に基づき結び付けて、新たな結合データベースＢ３０を生成するように構成される。

[0115] 第一実施形態及び第二実施形態では、推定モデルＭを用いることでオンライン行動から間接的に消費者の特徴を推定した。しかしながら、本実施形態では、オンライン行動の類似性から複数のデータベースを結びつけることで、複数のデータベース間で相互に不足している消費者の行動に関するデータを補足し、消費者の特徴を間接的に広く推定できるようにする。

[0116] 結合データベースＢ３０の生成は、プロセッサ１１が図１３に示すデータ結合処理を実行することにより実現される。以下では、本実施形態の情報処理システム１の説明として、コンピュータプログラムに従って、プロセッサ１１が実行するデータ結合処理の内容を選択的に説明する。

[0117] プロセッサ１１は、情報処理システム１の操作者からの指令に従って、図１３に示すデータ結合処理を開始することができる。データ結合処理を開始すると、プロセッサ１１は、第一行動データ収集システム５０から第一行動データベースＢ１１を取得する（Ｓ５１０）。第一行動データ収集システム

５０は、第一消費者群Ｃ３の第一の行動に関するデータを収集するシステムである。

[0118] 第一行動データベースＢ１１は、第一消費者のそれぞれについて、第一消費者の第一の行動に関する特徴を、対応する消費者の識別情報である消費者ＩＤと関連付けて表す第一行動データＤ１１（図１４参照）を有する。

[0119] 以下では、第一の行動が購買行動であり、第一行動データＤ１１が、第一消費者の購買履歴データであるものとして話を進める。第一行動データＤ１１は、第二実施形態の購買履歴データと同様に、消費者ＩＤに関連付けて、対応する第一消費者の購買毎に、購入した商品の識別情報と、購入日時の情報とを有する。

[0120] 続くＳ５２０において、プロセッサ１１は、第一行動データベースＢ１１内の各第一行動データＤ１１が示す消費者ＩＤに対応するクッキーＩＤを判別する。クッキーＩＤの判別は、第一実施形態と同様に、オンライン行動データ収集システム６０から変換テーブルを取得して実現できる。

[0121] オンライン行動データ収集システム６０は、第一実施形態のデータ収集システム２０と同様に、クッキーの発行及びクッキーによるオンラインユーザの行動分析によって、消費者ＩＤとクッキーＩＤとの間の変換テーブルを生成するように構成される。オンライン行動データ収集システム６０は更に、複数のオンラインユーザのオンライン行動データＤ３を生成して、オンライン行動データＤ３をクッキーＩＤに関連付けて記憶するように構成される。

[0122] 続くＳ５３０において、プロセッサ１１は、第一消費者のそれぞれに対応するオンライン行動データＤ３を、クッキーＩＤに基づいてオンライン行動データ収集システム６０から取得し、第一行動データベースＢ１１を拡張する。

[0123] 拡張後の第一行動データベースＢ１２は、図１４に示される。拡張後の第一行動データベースＢ１２は、第一消費者のそれぞれについて、対応する消費者ＩＤに関連付けて、第一行動及びオンライン行動の特徴を表すレコードを有する。図１４に示す例によれば、第一行動データベースＢ１２は、購買

履歴、ウェブ閲覧履歴、及びアフィニティカテゴリを表すレコードを有する。以下では、このレコードのことを、拡張第一行動データＤ１２と表現する。

[0124] 続くＳ５４０において、プロセッサ１１は、第二行動データ収集システム７０から第二行動データベースＢ２１を取得する。第二行動データ収集システム７０は、第二消費者群Ｃ４の第二の行動に関するデータを収集するシステムである。

[0125] 第二行動データベースＢ２１は、第二消費者のそれぞれについて、第二消費者の第二の行動の特徴を、対応する消費者ＩＤと関連付けて表す第二行動データＤ２１（図１５参照）を有する。第二行動データＤ２１が有する消費者ＩＤは、第一行動データＤ１１が有する消費者ＩＤとは別の独立したＩＤであると理解されてよい。

[0126] 以下では、第二の行動がテレビジョン放送の視聴行動であり、第二行動データＤ２１が、第二消費者のテレビジョン放送の視聴行動の特徴を表すデータである例を用いて話を進める。第二行動データＤ２１は、消費者ＩＤに関連付けて、視聴履歴として、第二消費者が視聴した番組の識別情報を有することができる。

[0127] 続くＳ５５０において、プロセッサ１１は、第二行動データベースＢ２１内の各第二行動データＤ２１が示す消費者ＩＤに対応するクッキーＩＤを判別する。クッキーＩＤの判別は、Ｓ５２０での処理と同様、オンライン行動データ収集システム６０から変換テーブルを取得して実現できる。

[0128] 続くＳ５６０において、プロセッサ１１は、第二消費者のそれぞれに対応するオンライン行動データＤ３を、クッキーＩＤに基づいてオンライン行動データ収集システム６０から取得し、第二行動データベースＢ２１を拡張する。

[0129] 拡張後の第二行動データベースＢ２２は、図１５に示される。拡張後の第二行動データベースＢ２２は、第二消費者のそれぞれについて、対応する消費者ＩＤに関連付けて、第二行動及びオンライン行動の特徴を表すレコード

を有する。図１５に示す例によれば、第二行動データベースＢ２２は、視聴履歴、ウェブ閲覧履歴、及びアフィニティカテゴリを表すレコードを有する。以下では、このレコードのことを、拡張第二行動データＤ２２と表現する。

[0130] プロセッサ１１は、上述のＳ５１０からＳ５６０までの処理を実行することにより、第一行動データベースＢ１１及び第二行動データベースＢ２１に、消費者の特徴に関する共通変数として、オンライン行動に関する変数、具体的にはウェブ閲覧履歴及びアフィニティカテゴリを追加する。

[0131] この共通変数は、第一行動データＤ１１と、第二行動データＤ２１とを、適切に結合するために使用される。本実施形態によれば、共通変数の存在しない第一行動データベースＢ１及び第二行動データベースＢ２のそれぞれに、クッキー技術を用いて、オンライン行動に関する共通変数を追加する点が特徴的である。

[0132] 続くＳ５７０において、プロセッサ１１は、共通変数の類似度、具体的にはオンライン行動の類似度に基づき、互いに類似する第一消費者及び第二消費者の第一行動データＤ１１及び第二行動データＤ２１を結合するように、結合データベースＢ３０を生成する。ここでいう類似は、一致を含む広義の意味で理解されるべきである。

[0133] 例えば、プロセッサ１１は、第一消費者及び第二消費者のそれぞれに関して、共通変数としてのウェブ閲覧履歴及びアフィニティカテゴリをベクトル表現し、共通変数を要素とする特徴ベクトルを生成する。プロセッサ１１は、類似度に関する評価値として、第一消費者の特徴ベクトルと第二消費者の特徴ベクトルの間の特徴空間上の距離（例えばコサイン距離）を、第一消費者と第二消費者との組み合わせのすべてに関して算出する。

[0134] プロセッサ１１は、組み合わせ毎の距離に基づき、距離が最短の、すなわち類似度が最も高い第一消費者及び第二消費者の組み合わせに対応する第一行動データＤ１１及び第二行動データＤ２１を結合することにより、結合データベースＢ３０を生成することができる。

[0135] 結合データベースＢ３０は、結合対象の第一消費者及び第二消費者の組み合わせ毎に、対応する第一行動データＤ１１及び第二行動データＤ２１を統合した一つのレコードを有し得る。結合データベースＢ３０は、組み合わせ毎に、対応する拡張第一行動データＤ１２及び拡張第二行動データＤ２２を統合した一つのレコードを有してもよい。

[0136] あるいは、結合データベースＢ３０は、図１６に示すように、結合関係を表すデータＤ３１を、第一行動データベースＢ１１又は第一行動データベースＢ１２と、第二行動データベースＢ２１又は第二行動データベースＢ２２とに加えて有した構成にされてもよい。

[0137] 結合関係を表すデータＤ３１は、第一行動データＤ１１又は拡張第一行動データＤ１２と、第二行動データＤ２１又は拡張第二行動データＤ２２との間の結合関係を表すデータである。結合データベースＢ３０は、第一消費者の第一の行動に関するデータと、対応する第二消費者の第二の行動に関するデータとを相互参照可能なその他の構成を採り得る。

[0138] 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、第一の行動に関するデータと、第二の行動に関するデータとを結び付けて、更には、それらにオンライン行動に関するデータを結び付けて、より情報量の多い有意義な消費者に関する結合データベースＢ３０を生成することができる。

[0139] 付言すると、第一行動データベースＢ１１における第一行動データＤ１１及び第二行動データベースＢ２１における第二行動データＤ２１は、個人情報保護のため、クラスタ単位で消費者の行動に関する特徴を表すデータであってもよい。

[0140] 例えば第一行動データＤ１１及び第二行動データＤ２１は、対応するクラスタに属する複数の個人の行動を、統計量として表す構成にされてもよい。この場合、第一行動データＤ１１及び第二行動データＤ２１は、対応するクラスタの消費者ＩＤ群を含んだ構成にされ得る。同様に、クラスタ毎の拡張第一行動データＤ１２及び拡張第二行動データＤ２２は、オンライン行動に関する特徴を統計量で表すように構成され得る。

[0141] ［その他］

本開示は、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ことは言うまでもない。上記実施形態では、例えば図１７Ａに示すように、インターネット上のウェブページＷＰにタグＧが埋め込まれることで、通信端末ＴＭを通じてウェブページＷＰにアクセスしたオンラインユーザに対してデータ収集システム２０からクッキーＩＤが発行され得る。更には、オンラインユーザに対応する顧客ＩＤが、アクセス元の通信端末ＴＭから通信によりデータ収集システム２０に提供され得る。

[0142] クッキーＩＤが既に発行されたオンラインユーザが通信端末ＴＭを通じて別のウェブページＷＰにアクセスしたとき、そのアクセス行動を表す情報がクッキーＩＤと共に、通信端末ＴＭから通信によりデータ収集システム２０に送信され得る。この情報の受信により、データ収集システム２０は、新たなウェページの閲覧履歴が記録されるように、対応するオンラインユーザのオンライン行動データＤ３を更新し得る。

[0143] このような通信を用いた消費者行動の記録は、オフライン空間、すなわち現実空間においても、タグＧに代えて、通信機能を有した電子機器ＥＬを用いることで実現可能である。

[0144] 例えば、図１７Ｂに示すように、現実空間には、電子機器ＥＬとして、ＩＣカードリーダが配置され得る。ＩＣカードリーダは、例えば、公共交通機関を利用するためのチケットに対応するＩＣカード１００を読み取るために、改札に配置され得る。この種のＩＣカード１００は、例えば日本国等のいくつかの国において、市中の小売店舗における支払時に電子マネーカードとして利用可能である。ＩＣカード１００が小売店舗で利用される場合、利用者システム３０に対応する企業は、顧客ＩＤに代えて又は加えてＩＣカード１００の識別情報を用いて、第一消費者リストＬ１及び第２消費者リストＬ２を提供可能である。

[0145] 電子機器ＥＬは、消費者が公共交通機関を利用するときに、ＩＣカード１００を電子機器ＥＬにかざす行動に応じて、公共交通機関の利用を表す情報

をＩＣカード１００の識別情報と共に、データ収集システム２０に送信することができる。この情報の受信により、データ収集システム２０は、公共交通機関の利用行動が記録されるように、対応する消費者の公共交通機関を利用した移動に関するオフライン行動データＤ７を更新し得る。データ収集システム２０は、消費者毎のオフライン行動データＤ７を、オンライン行動データＤ３に代えて又は加えて記憶し得る。

[0146] こうして、データ収集システム２０は、消費者毎の公共交通機関を利用した移動に関するオフライン行動データＤ７を更新するように構成される。データ収集システム２０は、情報処理システム１から指定されたＩＣカード１００の識別情報に関連付けられたオフライン行動データＤ７を、情報処理システム１に提供するように構成される。

[0147] 情報処理システム１において、プロセッサ１１は、第一消費者リストＬ１に記されるＩＣカード１００の識別情報に基づき、オンライン行動データＤ３に代えて又は加えて、オフライン行動データＤ７をデータ収集システム２０から取得することができる（Ｓ１３０）。プロセッサ１１は、取得したオフライン行動データＤ７に基づいて、推定モデルＭを構築することができる（Ｓ１４０）。

[0148] プロセッサ１１は、第二消費者リストＬ２に記されるＩＣカード１００の識別情報に基づき、オフライン行動データＤ７をデータ収集システム２０から取得することができる（Ｓ２３０）。プロセッサ１１は、取得したオフライン行動データＤ７及び推定モデルＭに基づいて、スコアを算出することができる（Ｓ２５０）。

[0149] 以上の文脈から理解できるように、本開示の技術は、推定モデルの構築に、オンライン行動データ及びオフライン行動データの一方又は両方を用いることができ、特には、通信を伴う行動に関するデータを用いることができる。

[0150] 電子機器ＥＬは、ＰＯＳ（Ｐｏｉｎｔ－Ｏｆ－Ｓａｌｅ）端末であってもよく、オフライン行動データＤ７は、ＰＯＳデータであってもよい。上述し

た通信を伴う行動の具体例としては、更に次のものが挙げられる。

・電子マネー及び／又はクレジットカードを用いた電子決済による消費者の購買行動

・位置座標をセンタに送信する車両に乗車する消費者の移動

・位置座標をセンタに送信する通信端末を携帯する消費者の移動

・インターネット上の購買行動、検索行動、コンテンツ閲覧行動、及び、ソーシャルネットワークサービスの利用行動

従って、オンライン行動データＤ３及びオフライン行動データＤ７に代えて、上述したオンライン行動及びオフライン行動に関する複数種類の行動履歴を説明する消費者行動データが用いられてもよい。この消費者行動データに基づいて、推定モデルＭが構築されてもよい。この消費者行動データから特定されるオンライン行動及びオフライン行動に関する複数種類のパラメータを共通変数として用いて、データ拡張及び／又はデータ結合が行われてもよい。

[0151] この他、推定モデルＭは、決定木又は回帰木を用いて設計されてもよい。推定モデルＭは、重回帰などの一般化線形モデルを用いて設計されてもよい。推定モデルＭは、ｋ近傍法を用いて設計されてもよい。

[0152] 推定モデルＭは、サポートベクタマシン（ＳＶＭ）を用いて設計されてもよい。推定モデルＭは、状態空間モデルなどの時系列解析を通じて設計されてもよい。推定モデルＭは、アンサンブル学習又は深層学習を通じて設計されてもよい。アンサンブル学習は、ランダムフォレスト等のバギング、及び、勾配ブースティング等のブースティングにより実現されてもよい。

[0153] 第三実施形態において、共通変数としてのオンライン行動に関する特徴と結び付けられる消費者の特徴は、上述した、利用契約行動、購買行動、及び視聴行動に限定されない。例えば、消費者の特徴として、結婚、出産、卒業、就職、退職、及び家を建てるなどのライフスタイルに係る特徴が、共通変数と結びつけられてもよい。また、消費者の状態変化として、消費者のライフスタイルの変化が消費者の購買履歴から判別されてもよい。

[0154] 上述の実施形態における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。特許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。

[0155] 最後に、用語間の対応関係を説明する。第一及び第二実施形態の情報処理システム１は、推定システムの一例に対応し、第三実施形態の情報処理システム１は、データ結合システムの一例に対応する。

[0156] プロセッサ１１が実行するＳ１１０；Ｓ３１０；Ｓ５１０の処理は、第一取得部により実現される処理の一例に対応する。プロセッサ１１が実行するＳ１２０，Ｓ１３０；Ｓ３２０，Ｓ３３０；Ｓ５２０，Ｓ５３０の処理は、第二取得部により実現される処理の一例に対応する。

[0157] プロセッサ１１が実行するＳ１４０；Ｓ３４０，Ｓ３５０の処理は、モデル構築部により実現される処理の一例に対応する。プロセッサ１１が実行するＳ２２０，Ｓ２３０；Ｓ４２０，Ｓ４３０の処理は、対象データ取得部により実現される処理の一例に対応する。プロセッサ１１が実行するＳ２５０，Ｓ２６０；Ｓ４５０，Ｓ４６０の処理は、推定部により実現される処理の一例に対応する。

[0158] プロセッサ１１が実行するＳ４１０，Ｓ５４０の処理は、第三取得部により実現される処理の一例に対応する。プロセッサ１１が実行するＳ５５０，Ｓ５６０の処理は、第四取得部により実現される処理の一例に対応する。プロセッサ１１が実行するＳ５７０の処理は、データ結合部により実現される処理の一例に対応する。