

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B1)

(11) 特許番号

**特許第6158464号
(P6158464)**

(45) 発行日 平成29年7月5日(2017.7.5)

(24) 登録日 平成29年6月16日(2017.6.16)

(51) Int. Cl. F 1
G06N 99/00 (2010.01) G06N 99/00 1 5 3
G06Q 30/02 (2012.01) G06Q 30/02 3 0 0

請求項の数 5 (全 20 頁)

<p>(21) 出願番号 特願2017-516189 (P2017-516189)</p> <p>(86) (22) 出願日 平成28年2月19日(2016.2.19)</p> <p>(86) 国際出願番号 PCT/JP2016/054935</p> <p>審査請求日 平成29年3月23日(2017.3.23)</p> <p>早期審査対象出願</p>	<p>(73) 特許権者 500461549 株式会社アイズファクトリー 東京都千代田区神田錦町1-2-3</p> <p>(74) 代理人 100105212 弁理士 保坂 延寿</p> <p>(72) 発明者 大場 智康 東京都千代田区神田錦町1-2-3 株式会 社アイズファクトリー内</p> <p>(72) 発明者 佐藤 晴正 東京都千代田区神田錦町1-2-3 株式会 社アイズファクトリー内</p> <p>審査官 多賀 実</p>
--	---

最終頁に続く

(54) 【発明の名称】 予測システム

(57) 【特許請求の範囲】

【請求項1】

第1の複数の対象をそれぞれ特定する第1の対象特定データと、前記第1の複数の対象それぞれの属性を示す属性データと、を対応付けた第1のリストの入力を第1の主体から受付け、

第2の複数の対象をそれぞれ特定する第2の対象特定データと、前記第2の複数の対象それぞれについての所定事象の評価値を示す評価データと、を対応付けた第2のリストの入力を第2の主体から受付ける

第1のモジュールと、

前記第1の対象特定データと前記第2の対象特定データとに基づいて、前記第1の複数の対象に含まれ且つ前記第2の複数の対象に含まれる第3の複数の対象を抽出する第2のモジュールと、

前記第1のリストに含まれる前記第3の複数の対象それぞれの前記属性データと、前記第2のリストに含まれる前記第3の複数の対象それぞれについての前記評価データと、に基づいて、前記評価データが未知である対象について前記属性データから前記所定事象を予測するための関数のパラメータを算出する第3のモジュールと、

前記第1の複数の対象に含まれ且つ前記第2の複数の対象に含まれない第4の複数の対象それぞれの前記属性データと、前記パラメータと、に基づいて、前記第4の複数の対象それぞれについての前記所定事象を予測する第4のモジュールと、を含む、予測システム。

10

20

【請求項 2】

請求項 1 に記載された予測システムであって、

前記第 1 の複数の対象を前記第 1 の対象特定データと異なる形式でそれぞれ特定する第 3 の対象特定データの入力を前記第 1 の主体から受付け、

前記第 2 の複数の対象を前記第 2 の対象特定データと異なる形式でそれぞれ特定する第 4 の対象特定データの入力を前記第 2 の主体から受付ける

第 5 のモジュールと、

前記第 3 の対象特定データと前記第 4 の対象特定データとに基づいて、

前記第 3 の複数の対象を前記第 3 の対象特定データと異なる形式で且つ前記第 4 の対象特定データと異なる形式でそれぞれ特定する第 5 の対象特定データと、

10

前記第 4 の複数の対象を前記第 3 の対象特定データと異なる形式でそれぞれ特定する第 6 の対象特定データと、

前記第 2 の複数の対象に含まれ且つ前記第 1 の複数の対象に含まれない第 5 の複数の対象を前記第 4 の対象特定データと異なる形式でそれぞれ特定する第 7 の対象特定データと

、

を生成する第 6 のモジュールと、

前記第 5 の対象特定データと、前記第 6 の対象特定データとを、前記第 3 の対象特定データと対応付けて前記第 1 の主体に送信し、

前記第 5 の対象特定データと、前記第 7 の対象特定データとを、前記第 4 の対象特定データと対応付けて前記第 2 の主体に送信する

20

第 7 のモジュールと、

をさらに含み、

前記第 1 のリストが、前記第 5 の対象特定データと前記第 6 の対象特定データとのそれぞれに基づいて第 1 のルールで生成された前記第 1 の対象特定データを含み、

前記第 2 のリストが、前記第 5 の対象特定データと前記第 7 の対象特定データとのそれぞれに基づいて前記第 1 のルールで生成された前記第 2 の対象特定データを含む、

予測システム。

【請求項 3】

請求項 2 に記載された予測システムであって、

前記第 1 の対象特定データが、前記第 5 の対象特定データと前記第 6 の対象特定データとのそれぞれを第 1 のハッシュ関数で処理した第 1 の複数のハッシュ値を含み、

30

前記第 2 の対象特定データが、前記第 5 の対象特定データと前記第 7 の対象特定データとのそれぞれを前記第 1 のハッシュ関数で処理した第 2 の複数のハッシュ値を含む、

予測システム。

【請求項 4】

請求項 3 に記載された予測システムであって、

前記第 3 の対象特定データは、前記第 1 の複数の対象それぞれを特定するための表現が正規化されたデータを第 2 のハッシュ関数で処理した第 3 の複数のハッシュ値を含み、

前記第 4 の対象特定データは、前記第 2 の複数の対象それぞれを特定するための表現が正規化されたデータを前記第 2 のハッシュ関数で処理した第 4 の複数のハッシュ値を含む

40

、

予測システム。

【請求項 5】

請求項 1 に記載された予測システムであって、

前記第 4 のモジュールが、前記第 3 の複数の対象及び前記第 4 の複数の対象それぞれの前記属性データと、前記パラメータと、に基づいて、前記第 3 の複数の対象及び前記第 4 の複数の対象それぞれについての前記所定事象を予測し、予測結果を前記第 1 の主体に送信する、

予測システム。

【発明の詳細な説明】

50

【技術分野】

【0001】

本発明は予測システムに関する。

【背景技術】

【0002】

既知の入力データとその入力データに対する既知の応答との組み合わせを教師データとして収集しておき、この教師データに基づいて機械学習を行い、新しい入力データに対する応答を予測するシステムが研究されている。このような予測システムは、医療や検索エンジンなど幅広い技術分野での利用が期待されている。

【発明の概要】

【0003】

本発明の1つの観点に係る予測システムは、

第1の複数の対象をそれぞれ特定する第1の対象特定データと、前記第1の複数の対象それぞれの属性を示す属性データと、を対応付けた第1のリストの入力を第1の主体から受付け、

第2の複数の対象をそれぞれ特定する第2の対象特定データと、前記第2の複数の対象それぞれについての所定事象の評価値を示す評価データと、を対応付けた第2のリストの入力を第2の主体から受付ける

第1のモジュールと、

前記第1の対象特定データと前記第2の対象特定データとに基づいて、前記第1の複数の対象に含まれ且つ前記第2の複数の対象に含まれる第3の複数の対象を抽出する第2のモジュールと、

前記第1のリストに含まれる前記第3の複数の対象それぞれの前記属性データと、前記第2のリストに含まれる前記第3の複数の対象それぞれについての前記評価データと、に基づいて、前記評価データが未知である対象について前記属性データから前記所定事象を予測するための関数のパラメータを算出する第3のモジュールと、

前記第1の複数の対象に含まれ且つ前記第2の複数の対象に含まれない第4の複数の対象それぞれの前記属性データと、前記パラメータと、に基づいて、前記第4の複数の対象それぞれについての前記所定事象を予測する第4のモジュールと、を含む。

【図面の簡単な説明】

【0004】

【図1】本発明の実施形態に係る予測システム3と、予測システム3に接続される装置とを示す。

【図2A】第1の主体が持っているデータの例を示す。

【図2B】第2の主体が持っているデータの例を示す。

【図2C】図2A及び図2Bに示されるデータから生成可能な教師データの例を示す。

【図2D】図2A及び図2Bに示されるデータから生成可能な予測用データの例を示す。

【図3】図1に示される予測システム3による処理手順を示すフローチャートである。

【図4A】図3に示される名寄せの処理のために予測システム3が受付けるデータの第1の例を示す。

【図4B】図3に示される名寄せの処理において予測システム3が生成するデータの第1の例を示す。

【図5A】図3に示される名寄せの処理のために予測システム3が受付けるデータの第2の例を示す。

【図5B】図3に示される名寄せの処理において予測システム3が生成するデータの第2の例を示す。

【図6A】図3に示されるカテゴリー分割ルールの生成の処理のために予測システム3が受付けるデータの例を示す。

【図6B】図3に示されるカテゴリー分割ルールの生成の処理において予測システム3が

10

20

30

40

50

生成するデータの例を示す。

【図 7 A】図 3 に示されるカテゴリー分割の処理においてコンピュータ 1 に入力されるデータの例を示す。

【図 7 B】図 3 に示されるカテゴリー分割の処理においてコンピュータ 1 が生成するデータの例を示す。

【図 8 A】図 3 に示されるランダム化の処理においてコンピュータ 1 に設定される第 1 のデータの例を示す。

【図 8 B】図 3 に示されるランダム化の処理においてコンピュータ 1 が生成する第 1 のデータの例を示す。

【図 9 A】図 3 に示されるランダム化の処理においてコンピュータ 1 に設定される第 2 のデータの例を示す。 10

【図 9 B】図 3 に示されるランダム化の処理においてコンピュータ 1 が生成する第 2 のデータの例を示す。

【図 10 A】図 3 に示される教師データ及び予測用データの生成処理のために予測システム 3 がコンピュータ 1 から受付けるデータの例を示す。

【図 10 B】図 3 に示される教師データ及び予測用データの生成処理のために予測システム 3 がコンピュータ 2 から受付けるデータの例を示す。

【図 11 A】図 3 に示される教師データの生成処理において予測システム 3 が生成する教師データの例を示す。

【図 11 B】図 3 に示される予測用データの生成処理において予測システム 3 が生成する予測用データの第 1 の例を示す。 20

【図 11 C】図 3 に示される予測用データの生成処理において予測システム 3 が生成する予測用データの第 2 の例を示す。

【図 12】図 3 に示される予測モデルの構築の処理において予測システム 3 が生成するデータの例を示す。

【図 13 A】図 3 に示される予測結果の算出処理において予測システム 3 が生成するデータの第 1 の例を示す。

【図 13 B】図 3 に示される予測結果の算出処理において予測システム 3 が生成するデータの第 2 の例を示す。

【発明を実施するための形態】 30

【0005】

以下、本発明の実施形態を、図面を参照しながら詳細に説明する。以下に説明される各実施形態は、本発明の一例を示すものであって、本発明の内容を限定するものではない。また、各実施形態で説明される構成及び動作のすべてが本発明の構成及び動作として必須であるとは限らない。なお、同一の構成要素には同一の参照符号を付して、重複する説明を省略する。

【0006】

< 1 . 構成 >

図 1 に、本発明の実施形態に係る予測システム 3 と、予測システム 3 に接続される装置とを示す。予測システム 3 は、インターネット等のネットワークを介して、コンピュータ 1 及びコンピュータ 2 に対して接続されている。 40

コンピュータ 1 は、第 1 の主体によって管理される。以下の説明において、第 1 の主体を、A 社、A 機関などと称することがある。

コンピュータ 2 は、第 1 の主体とは異なる第 2 の主体によって管理される。以下の説明において、第 2 の主体を、B 社、B 病院などと称することがある。

予測システム 3 も、コンピュータで構成される。予測システム 3 は、第 1 の主体とも第 2 の主体とも異なる第 3 の主体によって管理される。以下の説明において、第 3 の主体を、C 社と称することがある。

【0007】

コンピュータ 1、コンピュータ 2、及び予測システム 3 を構成するコンピュータは、い 50

ずれも、図示しないプロセッサ、メモリ、ストレージ、入力装置、出力装置、通信装置等を備えている。各々のコンピュータは、1台の装置で構成されていてもよいし、各々のコンピュータが、複数の装置で構成されていてもよい。

【0008】

< 2 . 解析のためのデータ >

図2Aに、第1の主体が持っているデータの例を示す。

第1の主体は、第1の複数の対象Aについての生データRDaを持っている。生データRDaは、対象特定情報D(A, P)と、属性データD(A, E)と、を対応付けたものである。

対象特定情報D(A, P)は、第1の複数の対象Aの各々を特定するデータである。

属性データD(A, E)は、第1の複数の対象Aの各々の属性を示すデータである。

【0009】

図2Bに、第2の主体が持っているデータの例を示す。

第2の主体は、第2の複数の対象Bについての生データRDbを持っている。生データRDbは、対象特定情報D(B, P)と、評価データD(B, L)と、を対応付けたものである。

対象特定情報D(B, P)は、第2の複数の対象Bの各々を特定するデータである。

評価データD(B, L)は、第2の複数の対象Bの各々についての所定事象の評価値を示すデータである。

【0010】

対象特定情報D(A, P)は、第1の複数の対象Aの各々の電子メールアドレスでもよいし(図2A参照)、第1の複数の対象Aの各々の住所と氏名との組み合わせでもよい。対象特定情報D(A, P)は、個人を特定する情報であればよい。

対象特定情報D(B, P)は、第2の複数の対象Bの各々の電子メールアドレスでもよいし(図2B参照)、第2の複数の対象Bの各々の住所と氏名との組み合わせでもよい。対象特定情報D(B, P)は、個人を特定する情報であればよい。

【0011】

図2Cに、図2A及び図2Bに示されるデータから生成可能な教師データの例を示す。

図2Dに、図2A及び図2Bに示されるデータから生成可能な予測用データの例を示す。

【0012】

対象特定情報D(A, P)と対象特定情報D(B, P)とを照合することにより、次の(1)~(3)が可能である。

(1) 第1の複数の対象Aに含まれ、且つ第2の複数の対象Bに含まれる第3の複数の対象A - Bを抽出し得る。図2Cに示されるように、第3の複数の対象A - Bの各々について、属性データD(A - B, E)と評価データD(A - B, L)とを対応付けることもできる。

(2) 第1の複数の対象Aに含まれ、且つ第2の複数の対象Bに含まれない第4の複数の対象A - A - Bを抽出し得る。図2Dに示されるように、第4の複数の対象A - A - Bの各々については、属性データD(A - A - B, E)を抽出することもできる。但し、第4の複数の対象A - A - Bについて、評価データは存在しない。

(3) 第2の複数の対象Bに含まれ、且つ第1の複数の対象Aに含まれない第5の複数の対象B - A - Bも抽出し得る。

【0013】

本発明の第1の形態において、第1の主体は、営利企業であるA社であり得る。第1の複数の対象Aは、A社の複数の顧客であり得る。属性データD(A, E)は、第1の複数の対象Aそれぞれの消費行動に係るデータであり得る。消費行動に係るデータとして、例えば、性別、年齢、会員区分(例えば、有料会員(有)か無料会員(無)か)、利用回数などを挙げることができる。第1の形態において、図2Aのカテゴリ-Xaの欄に記載されたものが性別に相当し、カテゴリ-Xbの欄に記載されたものが年齢に相当し

10

20

30

40

50

、カテゴリー X c の欄に記載されたものが会員区分に相当し、カテゴリー X d の欄に記載されたものが利用回数に相当する。

この第 1 の形態において、第 2 の主体は、A 社とは別の営利企業である B 社であり得る。第 2 の複数の対象 B は、B 社の複数の顧客であり得る。評価データ D (B , L) は、第 2 の複数の対象 B の各々が B 社の特定商品を購入したか否かを示す情報であり得る (図 2 B 参照) 。

【 0 0 1 4 】

第 1 の形態においては、A 社の複数の顧客と B 社の複数の顧客とに共通する第 3 の複数の対象 A B のそれぞれについて、属性データ D (A B , E) と評価データ D (A B , L) との組み合わせを、教師データとして抽出し得る (図 2 C 参照) 。この教師データを用いれば、どのような属性の顧客が B 社の上記特定商品を購入しやすい傾向にあるかを示す予測モデルを構築できる。

この予測モデルを、A 社の複数の顧客に含まれ且つ B 社の複数の顧客に含まれない第 4 の複数の対象 A - A B の各々についての属性データ D (A - A B , E) に適用する (図 2 D 参照) 。これにより、第 4 の複数の対象 A - A B の各々が B 社の上記特定商品を購入しそうか否かを予測し得る。従って、A 社は B 社と提携し、A 社の顧客の内の購入可能性の高い顧客のみを対象にして B 社の上記特定商品を勧めるようにすれば、A 社及び B 社の双方にとって新たなビジネスチャンスが生まれ得る。

【 0 0 1 5 】

本発明の第 2 の形態において、第 1 の主体は、健康保険組合などの A 機関であり得る。第 1 の複数の対象 A は、A 機関の複数の加入者であり得る。属性データ D (A , E) は、第 1 の複数の対象 A それぞれの健康に関するデータであり得る。健康に関するデータとして、性別、年齢、アレルギーの有無、薬の投与回数などを挙げることができる。第 2 の形態において、図 2 A のカテゴリー X a の欄に記載されたものが性別に相当し、カテゴリー X b の欄に記載されたものが年齢に相当し、カテゴリー X c の欄に記載されたものがアレルギーの有無に相当し、カテゴリー X d の欄に記載されたものが薬の投与回数に相当する。

この第 2 の形態において、第 2 の主体は、B 病院であり得る。第 2 の複数の対象 B は、B 病院の複数の患者であり得る。評価データ D (B , L) は、第 2 の複数の対象 B それぞれに対して特定薬品が薬効を有したか否かを示す情報であり得る (図 2 B 参照) 。

【 0 0 1 6 】

この第 2 の形態においては、A 機関の複数の加入者と B 病院の複数の患者とに共通する第 3 の複数の対象 A B のそれぞれについて、属性データ D (A B , E) と評価データ D (A B , L) との組み合わせを、教師データとして抽出し得る (図 2 C 参照) 。この教師データを用いれば、どのような属性の加入者に対して上記特定薬品が薬効を有するかを示す予測モデルを構築できる。

この予測モデルを、A 機関の複数の加入者に含まれ且つ B 病院の複数の患者に含まれない第 4 の複数の対象 A - A B それぞれについての属性データ D (A - A B , E) に適用する (図 2 D 参照) 。これにより、第 4 の複数の対象 A - A B の各々に対して、上記特定薬品が薬効を有するか否かを予測し得る。従って、A 機関は、A 機関の加入者の内の薬効を有する可能性の高い加入者のみを対象にして上記特定薬品を勧めることができ、治療効果の向上及び健康の増進に寄与し得る。

【 0 0 1 7 】

しかしながら、以上のように第 1 の主体と第 2 の主体とが提供する情報を組み合わせて予測を行うためには、第 1 の主体と第 2 の主体とが、それぞれ自分が持っている情報を誰かに開示しなければならない。

例えば、上述の第 3 の複数の対象 A B を特定し、抽出するためには、対象特定情報 D (A , P) と対象特定情報 D (B , P) とを照合する作業が必要である。少なくともこの作業をする者は、対象特定情報 D (A , P) と対象特定情報 D (B , P) との両方を知ってしまうことになる。

10

20

30

40

50

また、例えば、上述の予測モデルを構築するためには、共通する第3の複数の対象A、Bのそれぞれについて、属性データD(A、B、E)と評価データD(A、B、L)との組み合わせを教師データとして用いる必要がある。少なくとも予測モデルを構築する者は、属性データD(A、B、E)と評価データD(A、B、L)との組み合わせを知るだけでなく、第1の複数の対象Aに含まれ且つ第2の複数の対象Bに含まれる第3の複数の対象A、Bがどれなのかも知ってしまう。

【0018】

以下では、情報を既に持っている者以外の者には、なるべく情報を知られないようにしつつ、第1の主体と第2の主体とが提供する情報を組み合わせて、予測を行うシステムについて説明する。

【0019】

<3. 手順の概要>

図3は、図1に示される予測システム3による処理手順を示すフローチャートである。

S1において、第3の主体によって管理される予測システム3は、第1の主体によって管理されるコンピュータ1から、対象特定情報D(A、P)を受付ける。予測システム3は、第2の主体によって管理されるコンピュータ2から、対象特定情報D(B、P)を受付ける。S1の処理を行うモジュールは、本開示の第5のモジュールに相当し得る。

S2において、予測システム3は、対象特定情報D(A、P)及び対象特定情報D(B、P)に基づいて名寄せの処理を行う。S2の処理を行うモジュールは、本開示の第6のモジュールに相当し得る。

S3において、予測システム3は、名寄せの結果を、それぞれコンピュータ1とコンピュータ2とに送信する。S3の処理を行うモジュールは、本開示の第7のモジュールに相当し得る。

【0020】

S4において、予測システム3は、コンピュータ1から属性データD(A、E)を受付ける。

S5において、予測システム3は、属性データD(A、E)に基づいて、カテゴリー分割ルールFcatを生成する。

S6において、予測システム3は、カテゴリー分割ルールFcatをコンピュータ1に送信する。

【0021】

S7は予測システム3による処理ではないが、予測システム3による処理と密接に関係するためここで説明する。

S7において、第1の主体が、コンピュータ1を用いて、属性データD(A、E)をカテゴリー分割及びランダム化する。カテゴリー分割及びランダム化した属性データ $Tcr(j) * Tnr(j)(Fcat[D(A, E)])$ は、後述のように、属性が具体的にわからないように数値化されたデータである。以下の説明において、カテゴリー分割及びランダム化した属性データ $Tcr(j) * Tnr(j)(Fcat[D(A, E)])$ のことを、列配置ランダム化した属性データと称することがある。

【0022】

S8において、予測システム3は、コンピュータ1から、カテゴリー分割及びランダム化した属性データ $Tcr(j) * Tnr(j)(Fcat[D(A, E)])$ を受付ける。予測システム3は、コンピュータ2から、評価データD(B、L)を受付ける。S8の処理を行うモジュールは、本開示の第1のモジュールに相当し得る。

S9において、予測システム3は、S8で受付けたデータをマージして、教師データ及び予測用データを生成する。S9の処理を行うモジュールは、本開示の第2のモジュールに相当し得る。

【0023】

S10において、予測システム3は、教師データに基づいて予測モデルを構築する。S10の処理を行うモジュールは、本開示の第3のモジュールに相当し得る。

10

20

30

40

50

S 1 1において、予測システム3は、この予測モデルを予測用データに適用することによって、予測結果を算出する。S 1 1の処理を行うモジュールは、本開示の第4のモジュールに相当し得る。

S 1 2において、予測システム3は、予測結果をコンピュータ1に送信する。

以下、それぞれの処理について詳細に説明する。

【0024】

< 4 . S 1 - S 3 : 名寄せの処理 >

< 4 - 1 . 第1の例 >

図4Aは、図3に示される名寄せの処理のために予測システム3が受付けるデータの第1の例を示す。第1の例においては、名寄せのためのデータとして、氏名と住所との組み合わせが用いられる。氏名同一の別人は存在し得るので氏名だけでは名寄せできないが、氏名と住所との組み合わせなら名寄せが可能である。但し、住所の表記にはしばしば表記ゆれがみられるので、表記ゆれを考慮して名寄せの処理をするのが望ましい。

【0025】

図4Aに示されるように、予測システム3は、第1の主体によって管理されるコンピュータ1から、第1の複数の対象Aについての対象特定情報D(A, P)を受付ける。第1の例において受信する対象特定情報D(A, P)は、本開示における第3の対象特定データの一例に相当し得る。

予測システム3は、第2の主体によって管理されるコンピュータ2から、第2の複数の対象Bについての対象特定情報D(B, P)を受付ける。第1の例において受信する対象特定情報D(B, P)は、本開示における第4の対象特定データの一例に相当し得る。

第1の例において、対象特定情報D(A, P)は、例えば、住所氏名a1から住所氏名a5までの5名の住所氏名を含んでいる。対象特定情報D(B, P)は、例えば、住所氏名b1から住所氏名b4までの4名の住所氏名を含んでいる。

【0026】

図4Bは、図3に示される名寄せの処理において予測システム3が生成するデータの第1の例を示す。予測システム3は、図4Aに示される住所氏名を照合し、名寄せを行う。その結果、住所氏名a1と住所氏名b1とが一致し、住所氏名a2と住所氏名b2とが一致し、住所氏名a3と住所氏名b3とが一致したとする。住所氏名a4、住所氏名a5、及び住所氏名b4は、いずれも他の住所氏名と一致しなかったとする。

【0027】

その場合、予測システム3は、住所氏名a1及び住所氏名b1で特定される対象に第1の共通ID(C001)を付与し、住所氏名a2及び住所氏名b2で特定される対象に第2の共通ID(C002)を付与し、住所氏名a3及び住所氏名b3で特定される対象に第3の共通ID(C003)を付与する。住所氏名a4、住所氏名a5、及び住所氏名b4の各々で特定される対象には、別々のID(C004~C006)を付与する。

【0028】

第1の複数の対象Aの各々について予測システム3が付与したID(C001~C005)を、Dc(A, I)とする。Dc(A, I)は、第3の複数の対象A-Bの各々について予測システム3が付与した共通ID(C001~C003)と、第4の複数の対象A-A-Bの各々について予測システム3が付与したID(C004、C005)とを含む。

第2の複数の対象Bの各々について予測システム3が付与したID(C001~C003、C006)を、Dc(B, I)とする。Dc(B, I)は、第3の複数の対象A-Bの各々について予測システム3が付与した共通ID(C001~C003)と、第5の複数の対象B-A-Bの各々について予測システム3が付与したID(C006)とを含む。

【0029】

第3の複数の対象A-Bの各々について付与した共通ID(C001~C003)を、Dc(A-B, I)とする。Dc(A-B, I)は、本開示における第5の対象特定デー

タの一例に相当し得る。

第4の複数の対象A - A Bの各々について付与したID (C 0 0 4、C 0 0 5)を、D c (A - A B, I)とする。D c (A - A B, I)は、本開示における第6の対象特定データの一例に相当し得る。

第5の複数の対象B - A Bの各々について付与したID (C 0 0 6)を、D c (B - A B, I)とする。D c (B - A B, I)は、本開示における第7の対象特定データの一例に相当し得る。

【0030】

名寄せ結果の送信は、以下のように行う。

予測システム3は、第1の複数の対象Aの各々について付与したD c (A, I)を、対象特定情報D (A, P)と対応付けて、第1の主体によって管理されるコンピュータ1に送信する。 10

予測システム3は、第2の複数の対象Bの各々について付与したD c (B, I)を、対象特定情報D (B, P)と対応付けて、第2の主体によって管理されるコンピュータ2に送信する。

【0031】

第1の主体は、対象特定情報D (A, P)と対応付けて、D c (A, I)を受信することになる。しかしながら、第1の主体は、第1の複数の対象Aの内のどれが第3の複数の対象A Bで、どれが第4の複数の対象A - A Bであるかを知ることができないようになっていてる。 20

第2の主体は、対象特定情報D (B, P)と対応付けて、D c (B, I)を受信することになる。しかしながら、第2の主体は、第2の複数の対象Bの内のどれが第3の複数の対象A Bで、どれが第5の複数の対象B - A Bであるかを知ることができないようになっていてる。

第3の主体は、対象特定情報D (A, P)及び対象特定情報D (B, P)の両方を知ることになる。しかしながら、第3の主体は、名寄せの処理においては属性データD (A, E)及び評価データD (B, L)を知ることができない。このように、第3の主体が知り得る情報も、わずかで済むようになっていてる。

【0032】

< 4 - 2 . 第2の例 >

図5 Aは、図3に示される名寄せの処理のために予測システム3が受付けるデータの第2の例を示す。第2の例においては、名寄せのためのデータとして、電子メールアドレスが用いられる。電子メールアドレスは、表現が正規化されており、大文字と小文字の表記ゆれ以外には表記ゆれが許されていないため、名寄せの処理を行うためのユニークキーとなり得る。従って、名寄せのために、大文字と小文字の表記ゆれ以外の表記ゆれを考慮する必要はない。 30

【0033】

このようなユニークキーがある場合にも、図4 A及び図4 Bを参照しながら説明した方法と同様の方法で名寄せの処理を行うことが可能である。さらに、ユニークキーがある場合には、以下のような処理を行うことにより、第3の主体が対象特定情報D (A, P)及び対象特定情報D (B, P)を知ることのないようにすることができる。 40

【0034】

予測システム3は、第1の主体によって管理されるコンピュータ1から第3の複数のハッシュ値Hash2(D (A, P))を受付ける。第3の複数のハッシュ値Hash2(D (A, P))は、第1の複数の対象Aの電子メールアドレス(図2 A参照)を、それぞれ第2のハッシュ関数Hash2でハッシュ化した値である。第2の例において受信する第3の複数のハッシュ値Hash2(D (A, P))は、本開示における第3の対象特定データの一例に相当し得る。

予測システム3は、第2の主体によって管理されるコンピュータ2から第4の複数のハッシュ値Hash2(D (B, P))を受付ける。第4の複数のハッシュ値Hash2(D (B, P)) 50

は、第2の複数の対象Bの電子メールアドレス(図2B参照)を、それぞれ上記第2のハッシュ関数Hash2でハッシュ化した値である。第2の例において受信する第4の複数のハッシュ値Hash2(D(B, P))は、本開示における第4の対象特定データの一例に相当し得る。

なお、第1のハッシュ関数、第1及び第2のハッシュ値については後述する。

【0035】

図5Bは、図3に示される名寄せの処理において予測システム3が生成するデータの第2の例を示す。第2のハッシュ関数Hash2としては、異なる電子メールアドレスに対しては異なるハッシュ値が得られるように、1対1の写像を生成するハッシュ関数が選択されることが望ましい。これによれば、対象特定情報D(A, P)に含まれる電子メールアドレスの1つと、対象特定情報D(B, P)に含まれる電子メールアドレスの1つとが一致する場合、これらを同じ第2のハッシュ関数Hash2でハッシュ化した第3のハッシュ値Hash2(D(A, P))及び第4のハッシュ値Hash2(D(B, P))は互いに一致する。

また、第2のハッシュ関数Hash2としては、ハッシュ値に基づいて元の値を復元することが困難なハッシュ関数が選択されることが望ましい。

【0036】

予測システム3は、第3の複数のハッシュ値Hash2(D(A, P))と第4の複数のハッシュ値Hash2(D(B, P))とを照合し、名寄せを行う。これにより、予測システム3は、第3の複数の対象A-B、第4の複数の対象A-A-B、及び第5の複数の対象B-A-Bを特定することができる。

他の処理については、図4A及び図4Bを参照しながら説明したものと同様である。

【0037】

第1の主体は、第3の複数のハッシュ値Hash2(D(A, P))と対応付けて、Dc(A, I)を受信することになる。しかしながら、第1の主体は、第1の複数の対象Aの内のどれが第3の複数の対象A-Bで、どれが第4の複数の対象A-A-Bであるかを知ることができないようになっている。

第2の主体は、第4の複数のハッシュ値Hash2(D(B, P))と対応付けて、Dc(B, I)を受信することになる。しかしながら、第2の主体は、第2の複数の対象Bの内のどれが第3の複数の対象A-Bで、どれが第5の複数の対象B-A-Bであるかを知ることができないようになっている。

第3の主体は、第3の複数のハッシュ値Hash2(D(A, P))及び第4の複数のハッシュ値Hash2(D(B, P))の両方を知ることになる。しかしながら、第3の主体は、これらのハッシュ値から対象特定情報D(A, P)及び対象特定情報D(B, P)を知ることができない。また、名寄せの処理においては属性データD(A, E)及び評価データD(B, L)を知ることができない。

【0038】

<5.S4-S6: カテゴリー分割ルールの生成>

図6Aは、図3に示されるカテゴリー分割ルールの生成の処理のために予測システム3が受付けるデータの例を示す。予測システム3は、第1の主体によって管理されるコンピュータ1から属性データD(A, E)を受付ける。

【0039】

図6Bは、図3に示されるカテゴリー分割ルールの生成の処理において予測システム3が生成するデータの例を示す。予測システム3は、属性データに基づいて、カテゴリー分割ルールFcatを生成する。カテゴリー分割ルールFcatは、以下のように設定されることができる。

カテゴリーXaについては、男(Xa1)と女(Xa2)に分ける。

カテゴリーXbについては、20才代(Xb1)、30才代(Xb2)、40才代(Xb3)に分ける。

カテゴリーXcについては、有(Xc1)と無(Xc2)に分ける。

カテゴリーXdについては、0回以上(Xd1)、10回以上(Xd2)、100回以

10

20

30

40

50

上 (X d 3) に分ける。

予測システム 3 は、生成されたカテゴリー分割ルール Fcat を、第 1 の主体によって管理されるコンピュータ 1 に送信する。

【 0 0 4 0 】

カテゴリー分割ルールの生成処理において、第 3 の主体は、属性データ D (A , E) を知るようになる。しかしながら、第 3 の主体は、対象特定情報 D (A , P) と属性データ D (A , E) との対応関係を知ることができない。このように、第 3 の主体が知り得る情報は、わずかで済むようになっている。

【 0 0 4 1 】

カテゴリー分割ルール Fcat の生成は、第 1 の主体が行ってもよい。カテゴリー分割ルール Fcat の生成を第 1 の主体が行う場合には、第 3 の主体が属性データ D (A , E) を知ることもなくなる。

10

但し、データ分布の偏りに応じて、適切なカテゴリー分割ルールが異なる場合がある。適切なカテゴリー分割ルールを生成できるような専門的知見を、第 1 の主体が有しない場合もあることを想定し、本開示ではカテゴリー分割ルールの生成を第 3 の主体が行う場合でも、第 3 の主体が知り得る情報がわずかで済むようにしている。

【 0 0 4 2 】

< 6 . S 7 : カテゴリー分割及びランダム化 >

< 6 - 1 . カテゴリー分割 >

図 7 A は、図 3 に示されるカテゴリー分割の処理においてコンピュータ 1 に入力されるデータの例を示す。第 1 の主体は、対象特定情報 D (A , P) と属性データ D (A , E) とを対応付けた生データ R D a を持っている (図 2 A 参照) 。また第 1 の主体は、名寄せの処理により、 D c (A , I) と対象特定情報 D (A , P) との対応付けが可能となっている (図 4 B 、 図 5 B 参照) 。そこで、第 1 の主体は、対象特定情報 D (A , P) をキーとして、 D c (A , I) と、属性データ D (A , E) と、を対応付けて、コンピュータ 1 に入力する。

20

【 0 0 4 3 】

図 7 B は、図 3 に示されるカテゴリー分割の処理においてコンピュータ 1 が生成するデータの例を示す。コンピュータ 1 は、属性データ D (A , E) にカテゴリー分割ルール Fcat を適用することにより、カテゴリー分割した属性データ Fcat [D (A , E)] を生成する。この処理のために、一般的に普及している表計算プログラムが用いられてもよいし、専用のプログラムが用いられてもよい。

30

【 0 0 4 4 】

< 6 - 2 . 数値ランダム化 >

図 8 A は、図 3 に示されるランダム化の処理においてコンピュータ 1 に設定される第 1 のデータの例を示す。第 1 の主体は、コンピュータ 1 において数値ランダム化ルール Tnr (j) を設定する。

【 0 0 4 5 】

数値ランダム化ルール Tnr (j) は、カテゴリー分割ルール Fcat で定義された上述の X a 1 、 X a 2 、 X b 1 、 X b 2 等に対し、ランダムに選ばれた数値を割り当てるルールである。図 8 A に示される例においては、 X a 1 に数値 5 が割り当てられ、 X a 2 に数値 8 が割り当てられている。他の X b 1 、 X b 2 等に割り当てられた数値は図示の通りである。

40

【 0 0 4 6 】

図 8 B は、図 3 に示されるランダム化の処理においてコンピュータ 1 が生成する第 1 のデータの例を示す。コンピュータ 1 は、図 7 B に示されるカテゴリー分割した属性データ Fcat [D (A , E)] に、図 8 A に示される数値ランダム化ルール Tnr (j) を適用する。これにより、コンピュータ 1 は、図 8 B に示される数値ランダム化した属性データ Tnr (j) (Fcat [D (A , E)]) を生成する。この処理のために、一般的に普及している表計算プログラムが用いられてもよいし、専用のプログラムが用いられてもよい。

【 0 0 4 7 】

50

< 6 - 3 . 列配置ランダム化 >

図 9 A は、図 3 に示されるランダム化の処理においてコンピュータ 1 に設定される第 2 のデータの例を示す。第 1 の主体は、コンピュータ 1 において列配置ランダム化ルール $Tcr(j)$ を設定する。

【 0 0 4 8 】

列配置ランダム化ルール $Tcr(j)$ は、属性データ $D(A, E)$ に含まれるカテゴリー X_a 、 X_b 等を、ランダムに並び替えるルールである。図 9 A に示される例においては、カテゴリー X_a は 3 番目のカテゴリー X_3 に割り当てられ、カテゴリー X_b は 1 番目のカテゴリー X_1 に割り当てられ、カテゴリー X_c は 4 番目のカテゴリー X_4 に割り当てられ、カテゴリー X_d は 2 番目のカテゴリー X_2 に割り当てられている。

10

【 0 0 4 9 】

図 9 B は、図 3 に示されるランダム化の処理においてコンピュータ 1 が生成する第 2 のデータの例を示す。コンピュータ 1 は、図 8 B に示される数値ランダム化した属性データ $Tnr(j)(Fcat[D(A, E)])$ に、図 9 A に示される列配置ランダム化ルール $Tcr(j)$ を適用する。これにより、コンピュータ 1 は、図 9 B に示される列配置ランダム化した属性データ $Tcr(j)*Tnr(j)(Fcat[D(A, E)])$ を生成する。この処理のために、一般的に普及している表計算プログラムが用いられてもよいし、専用のプログラムが用いられてもよい。なお、数値ランダム化ルール $Tnr(j)$ 及び列配置ランダム化ルール $Tcr(j)$ における j は、ルールを識別するための番号である。1 回の解析において、教師データの生成時と予測用データの生成時とは同一の j で特定される数値ランダム化ルール $Tnr(j)$ 及び列配置ランダム化ルール $Tcr(j)$ が適用される。しかし、その次に解析を行うときには、別の j で特定される数値ランダム化ルール $Tnr(j)$ 及び列配置ランダム化ルール $Tcr(j)$ が適用されてもよい。

20

【 0 0 5 0 】

< 7 . S 8 - S 9 : 教師データ及び予測用データの生成 >

< 7 - 1 . 受付けるデータ >

図 10 A は、図 3 に示される教師データ及び予測用データの生成処理のために予測システム 3 がコンピュータ 1 から受付けるデータの例を示す。

【 0 0 5 1 】

予測システム 3 は、第 1 の主体によって管理されるコンピュータ 1 から、第 1 のリストを受付ける。第 1 のリストは、第 1 の複数のハッシュ値 $Hash1(Dc(A, I))$ の各々と、列配置ランダム化した属性データ $Tcr(j)*Tnr(j)(Fcat[D(A, E)])$ の各々と、が対応付けられたリストである。

30

第 1 の複数のハッシュ値 $Hash1(Dc(A, I))$ は、名寄せの処理においてコンピュータ 1 が受信した $Dc(A, I)$ を、第 1 のハッシュ関数 $Hash1$ でハッシュ化した値である。第 1 の複数のハッシュ値 $Hash1(Dc(A, I))$ は、本開示における第 1 の対象特定データの一例に相当し得る。また、第 1 のハッシュ関数 $Hash1$ は、本開示における第 1 のルールの一例に相当し得る。

【 0 0 5 2 】

図 10 B は、図 3 に示される教師データ及び予測用データの生成処理のために予測システム 3 がコンピュータ 2 から受付けるデータの例を示す。

40

【 0 0 5 3 】

予測システム 3 は、第 2 の主体によって管理されるコンピュータ 2 から、第 2 のリストを受付ける。第 2 のリストは、第 2 の複数のハッシュ値 $Hash1(Dc(B, I))$ の各々と、評価データ $D(B, L)$ の各々と、が対応付けられたリストである。

第 2 の複数のハッシュ値 $Hash1(Dc(B, I))$ は、名寄せの処理においてコンピュータ 2 が受信した $Dc(B, I)$ を、上記第 1 のハッシュ関数 $Hash1$ でハッシュ化した値である。第 2 の複数のハッシュ値 $Hash1(Dc(B, I))$ は、本開示における第 2 の対象特定データの一例に相当し得る。

【 0 0 5 4 】

50

第1のハッシュ関数Hash1としては、 $D_c(A, I)$ 及び $D_c(B, I)$ のうち異なるものに対しては異なるハッシュ値が得られるように、1対1の写像を生成するハッシュ関数が選択されることが望ましい。これによれば、 $D_c(A, I)$ 及び $D_c(B, I)$ のうち、これらに共通する $D_c(A, B, I)$ の各々については、第1のハッシュ値Hash1($D_c(A, I)$)及び第2のハッシュ値Hash1($D_c(B, I)$)は互いに一致する。

また、第1のハッシュ関数Hash1としては、ハッシュ値に基づいて元の値を復元することが困難なハッシュ関数が選択されることが望ましい。

【0055】

上述のように第3の主体が名寄せの処理をした場合には、第3の主体が $D_c(A, I)$ 及び $D_c(B, I)$ を知っていることになる。仮に、第3の主体が第1のハッシュ関数Hash1を知っている場合には、第3の主体は、既知の $D_c(A, I)$ 及び $D_c(B, I)$ をハッシュ化することにより、第1の複数のハッシュ値Hash1($D_c(A, I)$)及び第2の複数のハッシュ値Hash1($D_c(B, I)$)を生成できることになる。その場合、第3の主体は、第1のリストに含まれる第1の複数のハッシュ値Hash1($D_c(A, I)$)及び第2のリストに含まれる第2の複数のハッシュ値Hash1($D_c(B, I)$)から、逆に $D_c(A, I)$ 及び $D_c(B, I)$ を特定できてしまう。 $D_c(A, I)$ 及び $D_c(B, I)$ を特定できてしまうと、名寄せに用いられた $D(A, P)$ 及び $D(B, P)$ 、或いは、Hash2($D(A, P)$)及びHash2($D(B, P)$)を特定できてしまう。そして、第1のリストに含まれる列配置ランダム化した属性データ $Tcr(j)*Tnr(j)(Fcat[D(A, E)])$ との対応関係や、第2のリストに含まれる評価データ $D(B, L)$ との対応関係も特定できてしまう。このように、第3の主体が知り得る情報が増えてしまうのは、好ましいとは言えない。

【0056】

そこで、上記第1のハッシュ関数Hash1は、第1の主体と第2の主体との間で予め取り決めた、第3の主体にとって不知のハッシュ関数であることが望ましい。

あるいは、第1のハッシュ値Hash1($D_c(A, I)$)及び第2のハッシュ値Hash1($D_c(B, I)$)は、 $D_c(A, I)$ 及び $D_c(B, I)$ のそれぞれに文字列wを付加して、上記第2のハッシュ関数Hash2でハッシュ化した値であってもよい。文字列wは、第1の主体と第2の主体との間で予め取り決めた、第3の主体にとって不知である共通の文字列である。第3の主体が上記第2のハッシュ関数Hash2を知っている場合でも、第3の主体が文字列wを知らない限り、第1の複数のハッシュ値Hash1($D_c(A, I)$)及び第2の複数のハッシュ値Hash1($D_c(B, I)$)から $D_c(A, I)$ 及び $D_c(B, I)$ を特定することは困難となる。

また、 $D_c(A, I)$ 及び $D_c(B, I)$ が数値部分を有する場合に、その数値部分に、第3の主体にとって不知である共通の数値を加算してもよい。第1のハッシュ値Hash1($D_c(A, I)$)及び第2のハッシュ値Hash1($D_c(B, I)$)は、上記共通の数値が加算された $D_c(A, I)$ 及び $D_c(B, I)$ を、上記第2のハッシュ関数Hash2でハッシュ化した値であってもよい。

【0057】

< 7 - 2 . 教師データの生成 >

予測システム3は、第1のリスト(図10A参照)に含まれる第1の複数のハッシュ値Hash1($D_c(A, I)$)と、第2のリスト(図10B参照)に含まれる第2の複数のハッシュ値Hash1($D_c(B, I)$)とを照合する。これにより、第3の複数の対象A-Bと、第4の複数の対象A-A-Bと、を特定する。

【0058】

図11Aは、図3に示される教師データの生成処理において予測システム3が生成する教師データの例を示す。

予測システム3は、第3の複数の対象A-Bについて、列配置ランダム化した属性データ $Tcr(j)*Tnr(j)(Fcat[D(A-B, E)])$ と、評価データ $D(A-B, L)$ とを対応付けた教師データを生成する。

【0059】

なお、第3の主体が、カテゴリー分割ルールFcatを生成する処理を行ったために属性データD(A, E)を知っていたとしても、列配置ランダム化した属性データTcr(j)*Tnr(j)(Fcat[D(A, B, E)])から属性データD(A, B, E)を再現することは困難となっている。従って、第3の主体は、第3の複数の対象A, Bの具体的な属性を知ることが困難である。また、第2の複数の対象Bがどのような傾向を有するかを推定することも困難となっている。

【0060】

< 7 - 3 . 予測用データの生成 >

図11Bは、図3に示される予測用データの生成処理において予測システム3が生成する予測用データの第1の例を示す。

予測システム3は、第4の複数の対象A, A, Bについて、第1の複数のハッシュ値Hash1(Dc(A, I))と、列配置ランダム化した属性データTcr(j)*Tnr(j)(Fcat[D(A, B, E)])とを抽出し、予測用データ(1)として生成する。

【0061】

図11Cは、図3に示される予測用データの生成処理において予測システム3が生成する予測用データの第2の例を示す。

予測システム3は、第3の複数の対象A, B及び第4の複数の対象A, A, Bの両方について、第1の複数のハッシュ値Hash1(Dc(A, I))と、列配置ランダム化した属性データTcr(j)*Tnr(j)(Fcat[D(A, B, E)])とを抽出し、予測用データ(2)として生成する。

【0062】

< 8 . S10 : 予測モデルの構築 >

図12は、図3に示される予測モデルの構築の処理において予測システム3が生成するデータの例を示す。予測モデルの構築は、図11Aを参照しながら説明した教師データに基づいて行われる。予測モデルは、列配置ランダム化した属性データTcr(j)*Tnr(j)(Fcat[D(A, B, E)])を変数 x_k としたとき、関数 $f(x_k)$ の値が評価データD(A, B, L)と対応するような関数 $f(x_k)$ として与えられる。

【0063】

説明を単純化するため、関数 $f(x_k)$ が以下の式で表されるものとする。

$$f(x_k) = (p(X_1) + p(X_2) + p(X_3) + p(X_4)) / 4$$

$p(X_1)$ は、列配置ランダム化した属性データTcr(j)*Tnr(j)(Fcat[D(A, B, E)])の内のカテゴリー X_1 に関して付与されるパラメータである。このパラメータの具体例が図12に示されており、例えば、 $p(X_{11}) = 0.9$ 、 $p(X_{12}) = 0.3$ 、 $p(X_{13}) = 0.8$ のように設定される。

$p(X_2)$ は、カテゴリー X_2 に関して付与されるパラメータである。 $p(X_3)$ は、カテゴリー X_3 に関して付与されるパラメータである。 $p(X_4)$ は、カテゴリー X_4 に関して付与されるパラメータである。これらのパラメータの具体例も図12に示されている。

【0064】

適切なパラメータを付与することにより、関数 $f(x_k)$ の値を評価データD(A, B, L)に対応させることができる。例えば、第3の複数の対象A, Bの内、評価データD(A, B, L)が「 \square 」である対象については、関数 $f(x_k)$ の値が大きくなるようにパラメータが設定される。また、第3の複数の対象A, Bの内、評価データD(A, B, L)が「 \times 」である対象については、関数 $f(x_k)$ の値が小さくなるようにパラメータが設定される。

【0065】

< 9 . S11 - S12 : 予測結果の算出 >

図13Aは、図3に示される予測結果の算出処理において予測システム3が生成するデータの第1の例を示す。予測システム3は、図11Bを参照しながら説明した予測用データ(1)を用いて、予測スコアを算出する。すなわち、予測システム3は、第4の複数の

対象 A - A B について、上述の関数 $f(x, k)$ を予測スコアとして算出する。予測スコアの値は、所定事象の発生確率に対応する。例えば、予測スコアの値が大きいほど、所定事象の発生確率が大きいことを示す。予測システム 3 は、算出された予測スコアを、第 1 の複数のハッシュ値 $\text{Hash1}(D_c(A - A B, I))$ と対応付けて、第 1 の主体によって管理されるコンピュータ 1 に送信する。

【0066】

第 3 の複数の対象 A B については、既にコンピュータ 2 から評価データ $D(A B, L)$ が提供されている。従って、未知の事項を予測するという目的だけを考えた場合には、第 3 の複数の対象 A B について予測スコアを算出する必要はない場合がある。

【0067】

予測システム 3 は、第 4 の複数の対象 A - A B の内の予測スコアが高いものを抽出し、あるいは予測スコアが低いものを抽出して、コンピュータ 1 に送信するようにしてもよい。すなわち、予測スコアの使用目的を達するのであれば、算出された予測スコアのすべてを漏れなくコンピュータ 1 に送信する必要はない場合がある。

【0068】

図 13 B は、図 3 に示される予測結果の算出処理において予測システム 3 が生成するデータの第 2 の例を示す。予測システム 3 は、図 11 C を参照しながら説明した予測用データ (2) を用いて、予測スコアを算出する。すなわち、予測システム 3 は、第 3 の複数の対象 A B 及び第 4 の複数の対象 A - A B の両方について、上述の関数 $f(x, k)$ を算出する。予測システム 3 は、算出された予測スコアを、第 1 の複数のハッシュ値 $\text{Hash1}(D_c(A, I))$ と対応付けて、第 1 の主体によって管理されるコンピュータ 1 に送信する。

【0069】

第 3 の複数の対象 A B については、既にコンピュータ 2 から評価データ $D(A B, L)$ が提供されている。しかしながら、予測システム 3 が、第 4 の複数の対象 A - A B についてだけ、予測スコアをコンピュータ 1 に送信する場合には、第 1 の主体が、第 1 の複数の対象 A の内のどれが第 3 の複数の対象 A B であるかを知ることになってしまう。そこで、第 2 の主体が第 3 の複数の対象 A B を第 1 の主体に知られることを望まない場合には、予測システム 3 は、第 3 の複数の対象 A B 及び第 4 の複数の対象 A - A B の両方について、予測スコアをコンピュータ 1 に送信する。

【0070】

< 10 . 効果 >

上述の処理によれば、第 1 の主体は、対象特定情報 $D(A, P)$ を第 3 の主体に開示する場合があります (S 1)、属性データ $D(A, E)$ も第 3 の主体に開示する場合があります (S 4、S 8)。しかし、上述のように、対象特定情報 $D(A, P)$ と属性データ $D(A, E)$ との対応関係は第 3 の主体にもわからないようになっている。また、これらのデータは、いずれも第 2 の主体にはわからないようになっている。

【0071】

また、第 2 の主体は、対象特定情報 $D(B, P)$ を第 3 の主体に開示する場合があります (S 1)、評価データ $D(B, L)$ も第 3 の主体に開示する場合があります (S 8)。しかし、上述のように、対象特定情報 $D(B, P)$ と評価データ $D(B, L)$ との対応関係は第 3 の主体にもわからないようになっている。また、これらのデータは、いずれも第 1 の主体にはわからないようになっている。

【0072】

予測システム 3 は、第 1 の主体が持っている属性データ $D(A, E)$ と、第 2 の主体が持っている評価データ $D(B, L)$ とを組み合わせて、所定事象を予測することができる。このとき、情報を既に持っている者以外の者には、なるべく情報を知られないようにすることができる。

【0073】

< 11 . その他 >

第1の主体は、生データR D aの中に、第1の複数の対象Aとして架空の対象を加えておいてもよい。第2の主体は、生データR D bの中に、第2の複数の対象Bとして架空の対象を加えておいてもよい。これにより、第3の主体が第1の複数の対象Aの数及び第2の複数の対象Bの数も知らないようにすることができる。なお、このような架空の対象は、第3の複数の対象A Bの中に入ったり、教師データに加わったりすることはないと考えられる。従って、このような架空の対象が予測モデルに影響する可能性は低い。

【0074】

第1の主体又は第2の主体の少なくとも一方が第3の複数の対象A Bを既に知っている場合には、S 1からS 3までの名寄せの処理は行わなくてもよい。この場合に、S 8において第1の主体が第3の主体に送信する第1の対象特定データは、第1の主体によって識別可能なIDであり、S 8において第2の主体が第3の主体に送信する第2の対象特定データは、第2の主体によって識別可能なIDであってもよい。

第1の主体が第3の複数の対象A Bを知っている場合に、第1の主体は、第1の複数の対象Aの内の第3の複数の対象A Bについて、第2の複数の対象Bとの対応関係を記述した対応データを作成してもよい。この対応データは、S 8において第1の主体が第3の主体に送信する第1の対象特定データに含まれてもよい。これにより、第3の主体は、第1の対象特定データと第2の対象特定データとに基づいて、第3の複数の対象A Bを抽出することができる。

第2の主体が第3の複数の対象A Bを知っている場合に、第2の主体は、第2の複数の対象Bの内の第3の複数の対象A Bについて、第1の複数の対象Aとの対応関係を記述した対応データを作成してもよい。この対応データは、S 8において第2の主体が第3の主体に送信する第2の対象特定データに含まれてもよい。これにより、第3の主体は、第1の対象特定データと第2の対象特定データとに基づいて、第3の複数の対象A Bを抽出することができる。

【符号の説明】

【0075】

- 1 第1の主体によって管理されるコンピュータ、
- 2 第2の主体によって管理されるコンピュータ、
- 3 第3の主体によって管理される予測システム、
- A 第1の複数の対象、
- B 第2の複数の対象、
- A B 第3の複数の対象、
- A - A B 第4の複数の対象、
- B - A B 第5の複数の対象、
- D (A , P) 第1の複数の対象Aについての対象特定情報（第3の対象特定データの一例）、
- D (B , P) 第2の複数の対象Bについての対象特定情報（第4の対象特定データの一例）、
- D (A , E) 第1の複数の対象Aについての属性データ、
- D (B , L) 第2の複数の対象Bについての評価データ、
- D c (A , I) 第1の複数の対象Aの各々について予測システム3が付与したID、
- D c (B , I) 第2の複数の対象Bの各々について予測システム3が付与したID、
- Hash2(D (A , P)) 第3の複数のハッシュ値（第3の対象特定データの一例）、
- Hash2(D (B , P)) 第4の複数のハッシュ値（第4の対象特定データの一例）、
- D c (A B , I) 第3の複数の対象A Bの各々について付与した共通ID（第5の対象特定データの一例）、
- D c (A - A B , I) 第4の複数の対象A - A Bの各々について付与したID（第6の対象特定データの一例）、
- D c (B - A B , I) 第5の複数の対象B - A Bの各々について付与したID（第7の対象特定データの一例）、

10

20

30

40

50

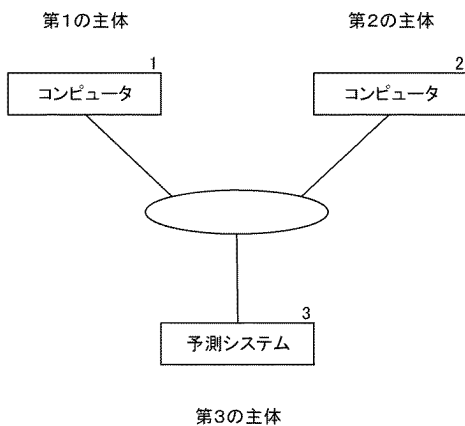
Fcat[D (A , E)] カテゴリ分割した属性データ、
 Tnr(j)(Fcat[D (A , E)]) 数値ランダム化した属性データ、
 Tcr(j)*Tnr(j)(Fcat[D (A , E)]) 列配置ランダム化した属性データ (カテゴリ分割及びランダム化した属性データ)、
 Hash1(D c (A , I)) 第 1 の複数のハッシュ値 (第 1 の対象特定データの一例)、
 Hash1(D c (B , I)) 第 2 の複数のハッシュ値 (第 2 の対象特定データの一例)

【要約】

予測システムは、第 1 の複数の対象のそれぞれについて、第 1 の対象特定データと属性データとを第 1 の主体から受付ける。予測システムは、第 2 の複数の対象のそれぞれについて、第 2 の対象特定データと評価データとを第 2 の主体から受付ける。予測システムは、前記第 1 の複数の対象に含まれ且つ前記第 2 の複数の対象に含まれる第 3 の複数の対象を抽出する。予測システムは、前記第 3 の複数の対象それぞれの前記属性データと前記評価データとに基づいて、所定事象を予測するための関数のパラメータを算出する。予測システムは、前記第 1 の複数の対象に含まれ且つ前記第 2 の複数の対象に含まれない第 4 の複数の対象のそれぞれについての前記所定事象を予測する。

10

【図 1】



【図 2 A】

第1の主体が持っている生データ RDa

第1の対象特定データ D(A, P)	属性データ D(A, E)			
	Xa	Xb	Xc	Xd
aa@bb.co.jp	男	43才	有	5回
cc@dd.jp	女	36才	有	72回
ee@ff.ne.jp	男	27才	無	171回
gg@hh.gr.jp	女	24才	有	125回
kk@ll.ac.jp	男	38才	無	39回

【図 2 B】

第2の主体が持っている生データ RDb

第2の対象特定データ D(B, P)	評価データ D(B, L)
aa@bb.co.jp	○
cc@dd.jp	x
ee@ff.ne.jp	○
ii@jj.com	x

【図 2 C】

生データから生成可能な教師データ

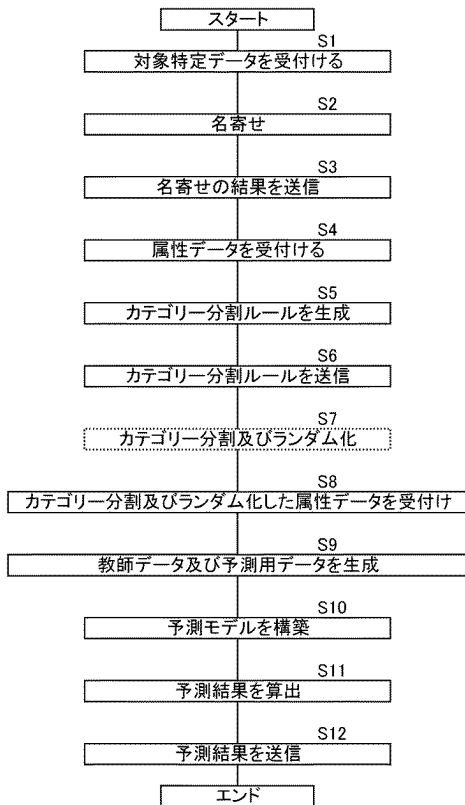
属性データ D(A∩B, E)				評価データ D(A∩B, L)
男	43才	有	5回	○
女	36才	有	72回	x
男	27才	無	171回	○

【図 2 D】

生データから生成可能な予測用データ

属性データ D(A-A∩B, E)				予測スコア
女	24才	有	125回	
男	38才	無	39回	

【 図 3 】



【 図 4 A 】

第1の主体から受付けるデータ		第2の主体から受付けるデータ	
D(A, P)		D(B, P)	
住所氏名a1		住所氏名b1	
住所氏名a2		住所氏名b2	
住所氏名a3		住所氏名b3	
住所氏名a4		住所氏名b4	
住所氏名a5			

【 図 4 B 】

第1の主体に送信するデータ		第2の主体に送信するデータ	
Dc(A, I)	D(A, P)	D(B, P)	Dc(B, I)
C001	住所氏名a1	住所氏名b1	C001
C002	住所氏名a2	住所氏名b2	C002
C003	住所氏名a3	住所氏名b3	C003
C004	住所氏名a4		
C005	住所氏名a5		
		住所氏名b4	C006

【 図 5 A 】

第1の主体から受付けるデータ		第2の主体から受付けるデータ	
Hash2(D(A, P))		Hash2(D(B, P))	
Hash2(aa@bb.co.jp)		Hash2(aa@bb.co.jp)	
Hash2(cc@dd.jp)		Hash2(cc@dd.jp)	
Hash2(ee@ff.ne.jp)		Hash2(ee@ff.ne.jp)	
Hash2(gg@hh.gr.jp)		Hash2(ii@jj.com)	
Hash2(kk@ll.ac.jp)			

【 図 5 B 】

第1の主体に送信するデータ		第2の主体に送信するデータ	
Dc(A, I)	Hash2(D(A, P))	Hash2(D(B, P))	Dc(B, I)
C001	Hash2(aa@bb.co.jp)	Hash2(aa@bb.co.jp)	C001
C002	Hash2(cc@dd.jp)	Hash2(cc@dd.jp)	C002
C003	Hash2(ee@ff.ne.jp)	Hash2(ee@ff.ne.jp)	C003
C004	Hash2(gg@hh.gr.jp)		
C005	Hash2(kk@ll.ac.jp)		
		Hash2(ii@jj.com)	C006

【 図 7 A 】

入力するデータ		S7			
Dc(A, I)	D(A, E)	Xa	Xb	Xc	Xd
C001	男	43	有	有	5回
C002	女	36	有	有	72回
C003	男	27	無	無	171回
C004	女	24	有	有	125回
C005	男	38	無	無	39回

【 図 6 A 】

第1の主体から受付けるデータ		S4			
D(A, E)		Xa	Xb	Xc	Xd
男	43才	有	有	有	5回
女	36才	有	有	有	72回
男	27才	無	無	無	171回
女	24才	有	有	有	125回
男	38才	無	無	無	39回

【 図 7 B 】

生成するデータ		S7			
Dc(A, I)	Fcat[D(A, E)]	Xa	Xb	Xc	Xd
C001	Xa1	Xb3	Xc1	Xd1	
C002	Xa2	Xb2	Xc1	Xd2	
C003	Xa1	Xb1	Xc2	Xd3	
C004	Xa2	Xb1	Xc1	Xd3	
C005	Xa1	Xb2	Xc2	Xd2	

【 図 6 B 】

第1の主体に送信するデータ		S5, S6	
カテゴリー	カテゴリー分割ルール	Fcat	
Xa	男(Xa1)と女(Xa2)に分ける		
Xb	20才代(Xb1)、30才代(Xb2)、40才代(Xb3)に分ける		
Xc	有(Xc1)と無(Xc2)に分ける		
Xd	0回以上(Xd1)、10回以上(Xd2)、100回以上(Xd3)に分ける		

【 図 8 A 】

設定するデータ		S7			
数値ランダム化ルール		Tnr(i)			
Xa1=5	Xb1=6	Xc1=7	Xd1=4		
Xa2=8	Xb2=3	Xc2=2	Xd2=6		
	Xb3=9		Xd3=1		

【図 8 B】

生成するデータ S7

Dc(A, I)	Tnr(j)(Fcat[D(A, E)])			
	Xa	Xb	Xc	Xd
C001	Xa1=5	Xb3=9	Xc1=7	Xd1=4
C002	Xa2=8	Xb2=3	Xc1=7	Xd2=6
C003	Xa1=5	Xb1=6	Xc2=2	Xd3=1
C004	Xa2=8	Xb1=6	Xc1=7	Xd3=1
C005	Xa1=5	Xb2=3	Xc2=2	Xd2=6

【図 1 0 A】

第1の主体から受付けるデータ (第1のリスト) S8

Hash1(Dc(A, I))	Tnr(j)*Tnr(j)(Fcat[D(A, E)])			
	X1	X2	X3	X4
Hash1(C001)	X13=9	X21=4	X31=5	X41=7
Hash1(C002)	X12=3	X22=6	X32=8	X41=7
Hash1(C003)	X11=6	X23=1	X31=5	X42=2
Hash1(C004)	X11=6	X23=1	X32=8	X41=7
Hash1(C005)	X12=3	X22=6	X31=5	X42=2

【図 9 A】

設定するデータ S7

列配置ランダム化ルール Tcr(j)
a=3, b=1, c=4, d=2

【図 1 0 B】

第2の主体から受付けるデータ (第2のリスト) S8

Hash1(Dc(B, I))	D(B, L)
Hash1(C001)	○
Hash1(C002)	×
Hash1(C003)	○
Hash1(C006)	×

【図 9 B】

生成するデータ S7

Dc(A, I)	Tcr(j)*Tnr(j)(Fcat[D(A, E)])			
	X1	X2	X3	X4
C001	X13=9	X21=4	X31=5	X41=7
C002	X12=3	X22=6	X32=8	X41=7
C003	X11=6	X23=1	X31=5	X42=2
C004	X11=6	X23=1	X32=8	X41=7
C005	X12=3	X22=6	X31=5	X42=2

【図 1 1 A】

生成するデータ 教師データ S9

Tcr(j)*Tnr(j)(Fcat[D(A∩B, E)])				D(A∩B, L)
X1	X2	X3	X4	
X13=9	X21=4	X31=5	X41=7	○
X12=3	X22=6	X32=8	X41=7	×
X11=6	X23=1	X31=5	X42=2	○

【図 1 2】

生成するデータ 予測関数のパラメータ S10

p(X11)=0.9	p(X21)=0.6	p(X31)=0.7	p(X41)=0.3
p(X12)=0.3	p(X22)=0.2	p(X32)=0.1	p(X42)=0.2
p(X13)=0.8	p(X23)=0.7		

【図 1 1 B】

生成するデータ 予測用データ(1) S9

Hash1(Dc(A-A∩B, I))	Tcr(j)*Tnr(j)(Fcat[D(A-A∩B, E)])			
	X1	X2	X3	X4
Hash1(C004)	X11=6	X23=1	X32=8	X41=7
Hash1(C005)	X12=3	X22=6	X31=5	X42=2

【図 1 3 A】

予測用データ(1)に予測関数のパラメータを適用した例 S11, S12

Hash1(Dc(A-A∩B, I))	p(Tcr(j)*Tnr(j)(Fcat[D(A-A∩B, E)]))				予測スコア
	p(X1)	p(X2)	p(X3)	p(X4)	
Hash1(C004)	0.9	0.7	0.1	0.3	0.500
Hash1(C005)	0.3	0.2	0.7	0.2	0.350

【図 1 1 C】

生成するデータ 予測用データ(2) S9

Hash1(Dc(A, I))	Tcr(j)*Tnr(j)(Fcat[D(A, E)])			
	X1	X2	X3	X4
Hash1(C001)	X13=9	X21=4	X31=5	X41=7
Hash1(C002)	X12=3	X22=6	X32=8	X41=7
Hash1(C003)	X11=6	X23=1	X31=5	X42=2
Hash1(C004)	X11=6	X23=1	X32=8	X41=7
Hash1(C005)	X12=3	X22=6	X31=5	X42=2

【図 1 3 B】

予測用データ(2)に予測関数のパラメータを適用した例 S11, S12

Hash1(Dc(A, I))	p(Tcr(j)*Tnr(j)(Fcat[D(A, E)]))				予測スコア
	p(X1)	p(X2)	p(X3)	p(X4)	
Hash1(C001)	0.8	0.6	0.7	0.3	0.600
Hash1(C002)	0.3	0.2	0.1	0.3	0.225
Hash1(C003)	0.9	0.7	0.7	0.2	0.625
Hash1(C004)	0.9	0.7	0.1	0.3	0.500
Hash1(C005)	0.3	0.2	0.7	0.2	0.350

フロントページの続き

(56)参考文献 特開2015-79331(JP,A)

特開2001-216369(JP,A)

伊藤 淳 外4名,「TwitterとBlogの共通ユーザプロフィールを利用したTwitterユーザ属性推定」,情報処理学会研究報告 2012(平成24)年度(5)[CD-ROM],一般社団法人情報処理学会,2013年,pp.1-8

(58)調査した分野(Int.Cl.,DB名)

G06N 3/00-99/00

G06F 17/30,19/00

G06Q 10/04,30/02