

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6010133号

(P6010133)

(45) 発行日 平成28年10月19日(2016.10.19)

(24) 登録日 平成28年9月23日(2016.9.23)

(51) Int.Cl.	F I
GO6F 17/30 (2006.01)	GO6F 17/30 210D
GO6N 7/00 (2006.01)	GO6F 17/30 220Z
GO6N 99/00 (2010.01)	GO6N 7/00 150
	GO6N 99/00 153

請求項の数 20 (全 22 頁)

(21) 出願番号	特願2014-543637 (P2014-543637)	(73) 特許権者	510330264
(86) (22) 出願日	平成24年11月29日(2012.11.29)		アリババ・グループ・ホールディング・リミテッド
(65) 公表番号	特表2015-504564 (P2015-504564A)		ALIBABA GROUP HOLDING LIMITED
(43) 公表日	平成27年2月12日(2015.2.12)		英国領、ケイマン諸島、グランド・ケイマン、ジョージ・タウン、ワン・キャピタル・プレイス、フォース・フロア、ピー・オー・ボックス 847
(86) 国際出願番号	PCT/US2012/067090		
(87) 国際公開番号	W02013/082297		
(87) 国際公開日	平成25年6月6日(2013.6.6)	(74) 代理人	110001243
審査請求日	平成27年11月10日(2015.11.10)		特許業務法人 谷・阿部特許事務所
(31) 優先権主張番号	201110387773.6		
(32) 優先日	平成23年11月29日(2011.11.29)		
(33) 優先権主張国	中国 (CN)		

最終頁に続く

(54) 【発明の名称】 属性データ区間の分類

(57) 【特許請求の範囲】

【請求項 1】

コンピュータが実行可能な命令で構成される、1つ以上のプロセッサによって行われる方法であって、

1つ以上のユーザ属性に関する1人以上の分類されたメンバーの属性データを抽出することと、

前記1つ以上のユーザ属性の特定のユーザ属性に関して、

前記抽出された属性データからの前記分類されたメンバーの属性データおよび分類に基づいて、前記特定のユーザ属性に対応する1つ以上の初期ユーザ属性データ区間を決定することと、

前記特定のユーザ属性に関する属性データが欠損している分類されたメンバーに関して、前記分類されたメンバーの属性データを、予め設定された欠損値として設定することと、

前記予め設定された欠損値を、前記初期ユーザ属性データ区間のそれぞれにそれぞれマージすることと、

前記予め設定された欠損値が前記初期ユーザ属性データ区間のそれぞれにそれぞれマージされるときに、最大事後確率(MAP)ベイズ推定値を毎回算出することと、

最小のMAPベイズ推定値を伴う初期ユーザ属性データ区間を、前記特定のユーザ属性に対応する最終属性データ区間として決定することと、

を含む、方法。

【請求項 2】

前記最小のMA Pベイズ推定値を伴う前記初期ユーザ属性データ区間を、前記特定のユーザ属性に対応する前記最終属性データ区間として決定する前に、

前記特定のユーザ属性に関するユーザ属性データが欠損している分類されたメンバーの数が、第1の閾値未満であるかどうかを判定することと、

前記特定のユーザ属性に関するユーザ属性データが欠損している分類されたメンバーの前記数が、前記第1の閾値未満であると判定することに応じて、前記最小のMA Pベイズ推定値を伴う前記初期ユーザ属性データ区間を、前記特定のユーザ属性に対応する前記最終属性データ区間として決定することと、をさらに含む、請求項1に記載の方法。

【請求項 3】

10

前記特定のユーザ属性に関するユーザ属性データが欠損している分類されたメンバーの前記数が、前記第1の閾値以上であると判定することに応じて、前記予め設定された欠損値を、別個の属性データ区間として処理することと、

前記予め設定された欠損値が前記別個の属性データ区間として処理されるときMA Pベイズ値を算出することと、

前記予め設定された欠損値が前記別個の属性データ区間として処理されるとき前記MA Pベイズ値が、前記予め設定された欠損値が前記初期ユーザ属性データ区間のそれぞれにそれぞれマージされるときMA Pベイズ推定値のそれぞれ未満であるかどうかを判定することと、

前記予め設定された欠損値が前記別個の属性データ区間として処理されるとき前記MA Pベイズ値が、前記予め設定された欠損値が前記初期ユーザ属性データ区間のそれぞれにそれぞれマージされるとき前記MA Pベイズ推定値のそれぞれ未満であると判定することに応じて、前記判定された初期ユーザ属性区間および前記別個の属性データ区間を、前記特定のユーザ属性に対応する前記最終属性データ区間として使用することと、

20

前記予め設定された欠損値が前記別個の属性データ区間として処理されるとき前記MA Pベイズ値が、前記予め設定された欠損値が前記初期ユーザ属性データ区間のそれぞれにそれぞれマージされるときMA Pベイズ推定値のそれぞれ以上であると判定することに応じて、前記最小のMA Pベイズ推定値を伴う前記初期ユーザ属性データ区間を、前記特定のユーザ属性に対応する前記最終属性データ区間として決定することと、をさらに含む、請求項2に記載の方法。

30

【請求項 4】

分類対象メンバーの分類を決定するときに、前記分類対象メンバーの各ユーザ属性に関して、前記特定のユーザ属性に関する前記分類対象メンバーの属性データが欠損している場合に、前記特定のユーザ属性に関する前記分類対象メンバーの前記属性データを、予め設定された欠損値として設定することと、

前記特定のユーザ属性に対応する前記ユーザ属性データ区間の中で、前記予め設定された欠損値に関する属性データ区間を選択することと、

各ユーザ属性について選択された属性データ区間に基づいて、前記分類対象メンバーの前記分類を決定することと、をさらに含む、請求項1に記載の方法。

【請求項 5】

40

前記特定のユーザ属性に関する属性データが欠損している前記分類されたメンバーに関して、前記分類されたメンバーの属性データを、前記予め設定された欠損値として設定する前に、

各初期属性データ区間のそれぞれについて、前記初期属性データ区間のそれぞれの中に属性データを有する分類されたメンバーの数が、第2の閾値よりも高いかどうかを判定することと、

各初期属性データ区間の中に属性データを有する分類されたメンバーの前記数が、前記第2の閾値よりも高いと判定することに応じて、前記分類されたメンバーの属性データを、予め設定された欠損値として設定することと、

少なくとも1つの初期属性データ区間の中に属性データを有する分類されたメンバーの

50

前記数が、前記第2の閾値以下であると判定することに応じて、

前記少なくとも1つの区間データ区間の属性データ区間のそれぞれについて、

前記属性データ区間のそれぞれが前の属性データ区間とマージされるときにMAPベイズ値、および前記属性データ区間のそれぞれが次の属性データ区間とマージされるときにMAPベイズ値を算出することと、

より低いMAPベイズ値を伴うマージされた属性データ区間を、前記特定のユーザ属性の前記初期属性データ区間として使用することと、をさらに含む、請求項1に記載の方法。

【請求項6】

前記分類されたメンバーの属性データおよび分類に基づいて、前記特定のユーザ属性に対応する1つ以上の初期ユーザ属性データ区間を前記決定することは、

前記抽出されたデータからの前記特定のユーザ属性の属性データのそれぞれを、対応するコード値に変換することと、

変換されたコード値および前記分類されたメンバーの前記分類に基づいて、前記初期ユーザ属性データ区間を決定することと、を含む、請求項1に記載の方法。

【請求項7】

前記抽出されたデータからの前記特定のユーザ属性の各属性データを、対応するコード値に前記変換することは、

前記特定のユーザ属性の前記属性データのそれぞれが、数字タイプの属性データであると判定することと、

前記属性データのそれぞれの値を、前記対応するコード値として使用することと、含む、請求項6に記載の方法。

【請求項8】

前記抽出されたデータからの前記特定のユーザ属性の各属性データを、対応するコード値に前記変換することは、

前記特定のユーザ属性の前記属性データのそれぞれが、日付タイプの属性データであると判定することと、

前記属性データのそれぞれの前記日付と現在の日付との間の時間長を、前記対応するコード値として使用することと、を含む、請求項6に記載の方法。

【請求項9】

前記抽出されたデータからの前記特定のユーザ属性の各属性データを、対応するコード値に前記変換することは、

前記特定のユーザ属性の前記属性データのそれぞれが、設定タイプの属性データであると判定することと、

前記データのそれぞれが前記主分類に属する比率値を算出することと、

比率値に従って、前記特定のユーザ属性に関する各属性データをランク付けすることと、

前記属性データのそれぞれのランク付け番号を、前記属性データのそれぞれの前記対応するコード値として使用することと、を含む、請求項6に記載の方法。

【請求項10】

装置であって、

1つ以上のユーザ属性に関する1人以上の分類されたメンバーの属性データを抽出する、データ抽出ユニットと、

前記1つ以上のユーザ属性の特定のユーザ属性に関して、前記抽出された属性データからの前記分類されたメンバーの属性データおよび分類に基づいて、前記特定のユーザ属性に対応する1つ以上の初期ユーザ属性データ区間を決定する、第1の初期属性データ区間決定ユニットと、

前記特定のユーザ属性に関する属性データが欠損している分類されたメンバーに関して、前記分類されたメンバーの属性データを、予め設定された欠損値として設定する、欠損値設定ユニットと、

10

20

30

40

50

前記予め設定された欠損値を前記初期ユーザ属性データ区間のそれぞれにそれぞれマージし、前記予め設定された欠損値が前記初期ユーザ属性データ区間のそれぞれにそれぞれマージされるときに、最大事後確率（MAP）ベイズ推定値を毎回算出する、第1の評価値算出ユニットと、

最小のMAPベイズ推定値を伴う初期ユーザ属性データ区間を、前記特定のユーザ属性に対応する最終属性データ区間として決定する、第1の最終属性データ区間決定ユニットと、

を備える、装置。

【請求項11】

前記第1の最終属性データ区間決定ユニットが、前記最小のMAPベイズ推定値を伴う前記初期ユーザ属性データ区間を、前記特定のユーザ属性に対応する前記最終属性データ区間として決定する前に、前記特定のユーザ属性に関するユーザ属性データが欠損している分類されたメンバーの数が、第1の閾値未満であるかどうかを判定する、第1の数判定ユニットと、

前記第1の数判定ユニットが、前記特定のユーザ属性に関するユーザ属性データが欠損している分類されたメンバーの前記数が、前記第1の閾値未満であると判定した後に、前記最小のMAPベイズ推定値を伴う前記初期ユーザ属性データ区間を、前記特定のユーザ属性に対応する前記最終属性データ区間として決定する、前記第1の最終属性データ区間決定ユニットと、をさらに備える、請求項10に記載の装置。

【請求項12】

前記第1の数判定ユニットが、前記特定のユーザ属性に関するユーザ属性データが欠損している分類されたメンバーの前記数が、前記第1の閾値以上であると判定するときに、前記予め設定された欠損値を、別個の属性データ区間としての処理し、前記予め設定された欠損値が前記別個の属性データ区間として処理されるときにMAPベイズ値を算出する、第2の評価値算出ユニットと、

前記第2の評価算出ユニットによって算出される前記MAPベイズ値が、前記予め設定された欠損値が前記初期ユーザ属性データ区間のそれぞれにそれぞれマージされるときに、前記第1の評価値算出ユニットによって算出されるMAPベイズ推定値のそれぞれ未満であるかどうかを判定する、評価値判定ユニットと、

前記第2の評価算出ユニットによって算出される前記MAPベイズ値が、前記予め設定された欠損値が前記初期ユーザ属性データ区間のそれぞれにそれぞれマージされるときに、前記第1の評価値算出ユニットによって算出されるMAPベイズ推定値のそれぞれ未満であると判定することに応じて、前記第1の初期属性データ区間決定ユニットによって決定される前記初期ユーザ属性区間、および前記別個の属性データ区間を、前記特定のユーザ属性に対応する前記最終属性データ区間として使用する、第2の最終属性データ区間決定ユニットと、

前記第2の評価算出ユニットによって算出される前記MAPベイズ値が、前記予め設定された欠損値が前記初期ユーザ属性データ区間のそれぞれにそれぞれマージされるときに、前記第1の評価値算出ユニットによって算出される前記MAPベイズ推定値のそれぞれ以上であると判定することに応じて、前記第1の評価値算出ユニットによって算出される前記最小のMAPベイズ推定値を伴う前記初期ユーザ属性データ区間を、前記特定のユーザ属性に対応する前記最終属性データ区間として決定する、前記第1の最終属性データ区間決定ユニットと、をさらに備える、請求項11に記載の装置。

【請求項13】

各初期属性データ区間のそれぞれについて、前記欠損値設定ユニットが、前記分類されたメンバーの属性データを、前記予め設定された欠損値として設定する前に、前記初期属性データ区間のそれぞれの中に属性データを有する分類されたメンバーの数が、第2の閾値よりも高いかどうかを判定する、第2の数判定ユニットと、

前記第2の数判定ユニットが、各初期属性データ区間の中に属性データを有する前記分類されたメンバーの数が、前記第2の閾値よりも高いと判定すると判定されるときに、前

10

20

30

40

50

記分類されたメンバーの属性データを、前記予め設定された欠損値として設定する、前記欠損値設定ユニットと、

前記第2の数判定ユニットが、少なくとも1つの初期属性データ区間の中に属性データを有する前記分類されたメンバーの数が、前記第2の閾値以下であると判定すると判定されるときに、前記少なくとも1つの区間データ区間の属性データ区間のそれぞれに関して、前記属性データ区間のそれぞれが前の属性データ区間とマージされるときにM A Pベイズ値、および前記属性データ区間のそれぞれが次の属性データ区間とマージされるときにM A Pベイズ値を算出する、第3の評価値算出ユニットと、

前記第3の評価値算出ユニットによって算出される、より低いM A Pベイズ値を伴うマージされた属性データ区間を、前記特定のユーザ属性の前記初期属性データ区間として決定する、第2の初期属性データ区間決定ユニットと、をさらに備える、請求項10に記載の装置。

10

【請求項14】

前記第1の初期属性データ区間決定ユニットは、

前記抽出されたデータからの前記特定のユーザ属性の属性データのそれぞれを、対応するコード値に変換する、コード値変換サブユニットと、

変換されたコード値および前記分類されたメンバーの前記分類に基づいて、前記初期ユーザ属性データ区間を決定する、初期属性データ区間決定サブユニットと、を備える、請求項10に記載の装置。

【請求項15】

20

前記コード値変換ユニットはさらに、前記特定のユーザ属性の前記属性データのそれぞれが、数字タイプの属性データであると判定し、前記属性データのそれぞれの値を、前記対応するコード値として使用する、請求項14に記載の装置。

【請求項16】

前記コード値変換ユニットはさらに、前記特定のユーザ属性の前記属性データのそれぞれが、日付タイプの属性データであると判定し、前記属性データのそれぞれの前記日付と現在の日付との間の時間長を、前記対応するコード値として使用する、請求項14に記載の装置。

【請求項17】

前記コード値変換ユニットはさらに、前記特定のユーザ属性の前記属性データのそれぞれが、設定タイプの属性データであると判定し、

30

前記データのそれぞれが前記主分類に属する比率値を算出し、

比率値に従って、前記特定のユーザ属性に関する各属性データをランク付けし、

前記属性データのそれぞれのランク付け番号を、前記属性データのそれぞれの前記対応するコード値として使用する、請求項14に記載の装置。

【請求項18】

アクションを行うために1つ以上のプロセッサによって実行可能であるコンピュータが実行可能な構成要素が記憶された、1つ以上のコンピュータ記憶媒体であって、該アクションは、

1つ以上のユーザ属性に関する1人以上の分類されたメンバーの属性データを抽出することと、

40

前記1つ以上のユーザ属性の特定のユーザ属性に関して、

前記抽出された属性データからの前記分類されたメンバーの属性データおよび分類に基づいて、前記特定のユーザ属性に対応する1つ以上の初期ユーザ属性データ区間を決定することと、

前記特定のユーザ属性に関する属性データが欠損している分類されたメンバーに関して、前記分類されたメンバーの属性データを、予め設定された欠損値として設定することと、

前記予め設定された欠損値を、前記初期ユーザ属性データ区間のそれぞれにそれぞれマージすることと、

50

前記予め設定された欠損値が前記初期ユーザ属性データ区間のそれぞれにそれぞれマージされるときに、最大事後確率（MAP）ベイズ推定値を毎回算出することと、

最小のMAPベイズ推定値を伴う初期ユーザ属性データ区間を、前記特定のユーザ属性に対応する最終属性データ区間として決定することと、
を含む、1つ以上のコンピュータ記憶媒体。

【請求項19】

前記アクションはさらに、

前記最小のMAPベイズ推定値を伴う前記初期ユーザ属性データ区間を、前記特定のユーザ属性に対応する前記最終属性データ区間として決定する前に、

前記特定のユーザ属性に関するユーザ属性データが欠損している分類されたメンバーの数が、第1の閾値未満であるかどうかを判定することと、

前記特定のユーザ属性に関するユーザ属性データが欠損している分類されたメンバーの前記数が、前記第1の閾値未満であると判定することに応じて、前記最小のMAPベイズ推定値を伴う前記初期ユーザ属性データ区間を、前記特定のユーザ属性に対応する前記最終属性データ区間として決定することと、をさらに含む、請求項18に記載の1つ以上のコンピュータ記憶媒体。

【請求項20】

前記アクションはさらに、

前記特定のユーザ属性に関するユーザ属性データが欠損している分類されたメンバーの前記数が、前記第1の閾値以上であると判定することに応じて、前記予め設定された欠損値を、別個の属性データ区間として処理することと、

前記予め設定された欠損値が前記別個の属性データ区間として処理されるときにMAPベイズ値を算出することと、

前記予め設定された欠損値が前記別個の属性データ区間として処理されるときに前記MAPベイズ値が、前記予め設定された欠損値が前記初期ユーザ属性データ区間のそれぞれにそれぞれマージされるときにMAPベイズ推定値のそれぞれ未満であるかどうかを判定することと、

前記予め設定された欠損値が前記別個の属性データ区間として処理されるときに前記MAPベイズ値が、前記予め設定された欠損値が前記初期ユーザ属性データ区間のそれぞれにそれぞれマージされるときに前記MAPベイズ推定値のそれぞれ未満であると判定することに応じて、前記判定された初期ユーザ属性区間および前記別個の属性データ区間を、前記特定のユーザ属性に対応する前記最終属性データ区間として使用することと、

前記予め設定された欠損値が前記別個の属性データ区間として処理されるときに前記MAPベイズ値が、前記予め設定された欠損値が前記初期ユーザ属性データ区間のそれぞれにそれぞれマージされるときにMAPベイズ推定値のそれぞれ以上であると判定することに応じて、前記最小のMAPベイズ推定値を伴う前記初期ユーザ属性データ区間を、前記特定のユーザ属性に対応する前記最終属性データ区間として決定することと、をさらに含む、請求項19に記載の1つ以上のコンピュータ記憶媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、メンバー分類技術の分野に関し、より具体的には、属性データ区間を分類する方法および装置に関する。

【0002】

本出願は、2011年11月29日に出願された中国特許出願第201110387773.6号、名称「Method and Apparatus of Classifying Attribute Data Intervals」に対する外国優先権を主張するものであり、該出願は、参照によりその全体が本明細書に組み込まれる。

【背景技術】

【0003】

従来の手法は、ユーザがウェブサイトにおいて登録するかどうかに基づいて、ウェブサイトのユーザをメンバーおよび非メンバーに分類する。メンバーがウェブサイトにおいて登録するときに、ウェブサイトは、メンバーに、年齢、登録日、性別、所在地、登録ソース、業種等の、ユーザ属性データを提出することを要求し得る。ウェブサイトは、データベースでのメンバーの識別情報に対応するユーザ属性データを記憶する。一般に、レコードは、表1で示されるように、メンバーの種々の属性のユーザ属性データを記憶する。

【0004】

【表1】

表1

ユーザ識別情報	年齢	登録日	性別	所在地
メンバーA	29	2011年11月17日	女性	北京
メンバーB	36	2010年5月1日	男性	上海
メンバーC	19	2009年3月5日	女性	天津

10

【0005】

表1において、各横列は、レコードを表す。レコードの各フィールドは、メンバーが1つのユーザ属性について提出する、属性データを記憶する。例えば、各レコードの中の「年齢」フィールドは、メンバーがユーザ属性「年齢」について提出する、属性データを記憶する。

20

【0006】

メンバーのユーザ属性についてメンバーが提出する属性データの間には非常に大きい違いがあり得るので、ウェブサイトは、ユーザ属性の属性データに基づいて、メンバーを分類し得る。一般に、メンバーは、2つの分類に分類される。1つは主分類であり、もう1つは副分類である。例えば、メンバーは、アクティブメンバーおよび非アクティブメンバーとして分類され得る。アクティブメンバーは主分類であり、非アクティブメンバーは副分類である。次いで、メンバーの分類に基づいて、対応するサービスがメンバーに提供され得る。

【0007】

従来の手法は、メンバーを分類するときに、既に分類されたメンバーの大量の属性データに基づいて、複数の各ユーザ属性の属性データ区間を取得する。例えば、ユーザ属性「年齢」の属性データ区間は、[10、20]、(20、40]、(40、60]等の3つであり得る。ユーザ属性「所在地」の属性データ区間は、{北京、上海、天津、重慶}、{河北、河南、山西}、{福建、江西、浙江}、および{安徽、甘肅、山東}等の4つであり得る。ユーザ属性「登録日」の属性データ区間は、[2001年1月1日、2005年12月31日]、(2006年1月1日、2010年12月31日]、(2011年1月1日、2015年12月31日]等の3つであり得る。各ユーザ属性に関する複数の属性日付区間が取得された後に、各ユーザ属性に関して、ブル特性が各属性データ区間に割り当てられる。各ブル特性は、その固有の特性識別情報を有する。

30

【0008】

従来の手法は、分類対象メンバーを分類するとき、またはメンバーがリアルタイムで分類されるときに、各ユーザ属性に関する分類対象メンバーの属性データの属性データ区間を決定する。決定された属性データの対応するブル特性には、値1が割り当てられ、値が1であるブル特性の特性識別情報が記憶される。各ユーザ属性について、ブル特性の対応する特性識別情報が抽出された後に、ブル特性のそれぞれの重み値に基づいて、分類対象メンバーが主分類に分類される確率が算出される。確率が50%より高い場合は、分類対象メンバーが主分類に分類される。確率が50%以下である場合は、分類対象メンバーが副分類に分類される。

40

【0009】

50

図1は、従来の手法に従ってユーザ属性の属性データ区間を決定する、例示的な方法のフローチャートを図示する。

【0010】

102で、分類されたメンバーの大量の属性データが、トレーニングデータとして抽出される。104で、各ユーザ属性について、ユーザ属性に対応する各属性データが、別個の属性データ区間として処理される。106で、最大事後確率(MAP)ベイズ推定ルールに基づいて、104で分類された属性データ区間の評価値が算出される。108で、複数のデータ属性区間を取得するために隣接する区間がマージされ、そして、マージされた属性データ区間の別の評価値が算出される。

【0011】

110で、106で取得した評価値が108で取得した評価値よりも小さい場合は、104で取得した属性データ区間が、ユーザ属性の最終属性データ区間として決定される。

【0012】

112で、106で取得した評価値が108で取得した評価値以上である場合は、104で取得した属性データ区間が維持され、最小の評価値を伴う属性データ区間の分類に到達するまで、隣接する区間が連続的にマージされる。最小の評価値を伴う分類から取得した属性データ区間が、ユーザ属性の最終属性データ区間として決定される。

【0013】

従来の手法は、トレーニングデータからの属性データに基づいて、ユーザ属性の属性データ区間进行分类する。従来の手法は、次いで、分類対象メンバーの属性データの属性データ区間を決定する。しかしながら、メンバーがウェブサイトにおいて登録するときに、該メンバーは、いくつかのユーザ属性の属性データを入力しない場合がある。例えば、メンバーがユーザ属性「年齢」の属性データを提出しない場合、ウェブサイトにおいて記憶されるユーザのレコードには、ユーザ属性「年齢」の属性データが欠損している。将来の分類では、そのようなユーザ属性に関するユーザの属性データの属性データ区間を正確に決定することができず、したがって、メンバーを正確に分類することができない。したがって、属性データ区間进行分类するための従来の手法の精度率は低い。

【発明の概要】

【0014】

この要約は、詳細な説明において以下でさらに説明される概念のいくつかを選択して、簡素化された形態で紹介するために提示される。この要約は、請求される主題の主要な特徴または基本的特徴を特定することを目的としておらず、請求される主題の範囲を決定する一助として使用されるものでもない。例えば、「技術」という用語は、上の文脈によって、および本開示の全体を通して容認される、デバイス(複数可)、システム(複数可)、方法(複数可)、および/またはコンピュータが読み出し可能な命令を指し得る。

【0015】

本開示は、属性データ区間进行分类する方法および装置を提供する。1つ以上のユーザ属性に関する分類されたメンバーの複数の属性データが抽出される。各ユーザ属性について、以下の操作が行われる。特定のユーザ属性の属性データおよび分類されたメンバーのメンバー分類に基づいて、特定のユーザ属性の初期属性データ区間が決定される。特定のユーザ属性に関する属性データが欠損している分類されたメンバーに関して、特定のユーザ属性に関する分類されたメンバーの属性データが、予め設定された欠損値として設定される。予め設定された欠損値は、次いで、決定された初期属性データ区間のそれぞれにマージされる。初期属性データ区間のそれぞれが予め設定された欠損値とそれぞれマージされるときに、各分類について、MAPベイズ評価値がそれぞれ算出される。最小のMAPベイズ評価値を伴う属性データ区間が、特定のユーザ属性に対応する最終属性区間として決定される。

【0016】

本開示はまた、属性データ区間进行分类する装置を提供する。本装置は、属性データ抽出ユニットと、第1の初期属性データ区間決定ユニットと、欠損値設定ユニットと、第1の

10

20

30

40

50

評価値算出ユニットと、第1の最終属性データ区間決定ユニットとを含み得る。属性データ抽出ユニットは、1つ以上のユーザ属性に関する複数の分類されたメンバーの属性データを抽出する。第1の初期属性データ区間決定ユニットは、各ユーザ属性について、特定のユーザ属性の属性データおよび分類されたメンバーのメンバー分類に基づいて、特定のユーザ属性の初期属性データ区間を決定する。特定のユーザ属性に関する属性データが欠損している分類されたメンバーに関して、欠損値設定ユニットは、特定のユーザ属性に関する分類されたメンバーの属性データを、予め設定された欠損値として設定する。第1の評価値算出ユニットは、予め設定された欠損値を、第1の初期属性データ区間ユニットによって決定される初期属性データ区間のそれぞれにマージし、初期属性データ区間のそれぞれが予め設定された欠損値とそれぞれマージされるときに、各分類について、MAPベイズ評価値をそれぞれ算出する。最終属性データ区間決定ユニットは、最小のMAPベイズ評価値を伴う属性データ区間を、特定のユーザ属性に対応する最終属性データ区間として決定する。

10

【0017】

本開示によって開示される本手法により、ウェブサイトは、1つ以上のユーザ属性に関する1人以上の分類されたメンバーの属性データを抽出し得る。各ユーザ属性について、ウェブサイトは、特定のユーザ属性に対応する初期属性データ区間を決定する。特定の属性に関する属性データが欠損している分類されたメンバーに関して、ウェブサイトは、特定のユーザ属性に関する分類されたメンバーの属性データを、予め設定された欠損値として設定し、予め設定された欠損値を、決定された初期属性データ区間のそれぞれにそれぞれマージし、初期属性データ区間のそれぞれが予め設定された欠損値とそれぞれマージされるときに、MAPベイズ評価値を毎回算出し、そして、最小のMAPベイズ評価値を伴う属性データ区間を、特定のユーザ属性に対応する最終属性データ区間として決定する。本手法は、属性データ区間を分類するときに、特定のユーザ属性に関する1人以上の分類されたメンバーの属性データが欠損しているかもしれない可能性を考慮し、それによって、属性データ区間の分類の精度を向上させる。本手法は、メンバーの属性データが欠損しているときの不正確さの問題を回避し、それによって、メンバー分類の精度を向上させる。

20

【図面の簡単な説明】

【0018】

本開示の実施形態をより良好に例示するために、以下は、実施形態を説明する際に使用される図面の簡単な紹介である。以下の図面は、本開示のいくつかの実施形態に関連しているに過ぎないことは明白である。当業者は、創造的な努力をせずに、本開示の図に従って他の形態を取得することができる。

30

【図1】従来の手法に従ってユーザ属性の属性データ区間を決定する、例示的な方法のフローチャートを示す図である。

【図2】本開示の第1の例示的な実施形態に従って属性データ区間を分類する、例示的な方法のフローチャートを示す図である。

【図3】本開示の第2の例示的な実施形態に従ってオンラインでメンバーを分類する、例示的な方法のフローチャートを示す図である。

40

【図4】本開示の第3の例示的な実施形態に従って属性データ区間を分類する、別の例示的な方法のフローチャートを示す図である。

【図5】本開示の第4の例示的な実施形態に従ってオンラインで分類対象メンバーを分類する、別の例示的な方法のフローチャートを示す図である。

【図6】本開示の第5の例示的な実施形態に従って属性データ区間を分類する、例示的な装置を示す図である。

【発明を実施するための形態】

【0019】

以下は、図面を参照することによる、本手法の詳細な説明である。本明細書で説明される実施形態は、例示的な実施形態であり、本開示の範囲を制限するために使用されるべき

50

ではない。

【0020】

図2は、本開示の第1の例示的な実施形態に従って属性データ区間を分類する、例示的な方法のフローチャートを図示する。

【0021】

202で、1つ以上のユーザ属性に関する分類されたメンバーの属性データが抽出される。属性データ区間がオフラインで分類されるときに、またはトレーニング用であるときに、ユーザ属性に関する分類されたメンバーの属性データは、トレーニングデータとして使用される。本手法は、種々の属性データ区間を取得するために、トレーニングデータから学習するための種々の方法を使用し得る。

10

【0022】

例えば、ユーザ属性に関する分類されたメンバーの属性データは、データベースに事前に記憶され得る。1つのレコードは、種々のユーザ属性に関するメンバーの属性データを記憶し得る。ユーザ属性としては、年齢、登録日、性別、所在地、登録ソース、業種等が挙げられ得るが、それらに限定されない。

【0023】

データベースから属性データを抽出するための、種々の方法がある。例えば、データベースに記憶された、所定数の分類されたメンバーのレコードは、ランダムに抽出され得る。所定数は、予め設定され得る。例えば、所定数は、デバイスメモリのサイズを考慮することによって設定され得る。十分なメモリがあるときには、所定数は、できる限り大きくすべきである。Mメガバイトのデバイスメモリの例において、属性データの各フィールドの長さは、4バイトであり、所定数は、以下の条件を満たすように設定され得る。

20

【0024】

【数1】

$$\text{レコード数} \times \text{フィールド数} < \frac{M}{4} \times 10^6$$

【0025】

抽出された属性データが統計的な意味を有することを確実にするために、メンバーの各分類について抽出されたレコードは、所定数よりも大きくすべきである。例えば、メンバーは、主分類および副分類に分類され得る。レコードの所定数は、100である。主分類および副分類のそれぞれから抽出されるレコードは、それぞれ、100よりも大きくなり得る。

30

【0026】

204で、各特定のユーザ属性について、特定のユーザ属性および分類されたメンバーの分類に対応するユーザ属性データに基づいて、特定のユーザ属性に対応する初期属性データ区間が決定される。

【0027】

特定のユーザ属性に対応する初期属性データ区間を決定するための、種々の方法がある。例えば、特定のユーザ属性に対応する初期属性データ区間を決定するために、MAPベイズ評価値を算出する方法が使用され得る。

40

【0028】

例えば、分類された属性データ区間がa1、a2、・・・、aIであり、Iが任意の正の数であるときに、MAPベイズ評価値は、以下を使用することによって算出され得る。

【0029】

【数2】

$$\log(n) + \log\left(\frac{n+I-1}{I-1}\right) + \sum_{i=1}^I \log\left(\frac{n_i+c-1}{c-1}\right) + \sum_{i=1}^I \log(n_i!/n_{i,1}!n_{i,2}!\dots n_{i,c}!)$$

【0030】

50

n は、データトレーニングに関する抽出されたレコードの数を表す（主分類のメンバーのレコードの数、および副分類のメンバーのレコードの数を含み得る）。

【0031】

【数3】

$$\begin{pmatrix} - \\ - \end{pmatrix}$$

【0032】

は、

【0033】

【数4】

$$\begin{pmatrix} b \\ a \end{pmatrix} = C_b^a$$

【0034】

等の2項係数を表す。 I は、分類されたデータ区間の数を表す。 n_i は、属性データ区間 a_i ($1 \leq i \leq I$) のレコードの数を表す。 c は、メンバー分類の数を表す（例えば、メンバー分類が主分類および副分類を含むときに、 c は、2である）。 $n_{i,j}$ は、属性データ区間 a_i 内のメンバー j の数を表す。

【0035】

特定のユーザ属性に対応する初期属性データ区間が決定されるときに、特定のユーザ属性の属性データの各値は、別個の属性データ区間として処理され、最小のMAPベイズ評価値を伴う属性データ区間が取得されるまで、隣接する区間が連続的にマージされる。例えば、以下のステップを行い得る。

【0036】

第1のステップで、特定のユーザ属性の各属性データは、別個の属性データ区間として処理される。第2のステップで、MAPベイズ推定ルールに基づいて、第1のステップでの属性データ区間の分類について、MAPベイズ評価値が算出される。第3のステップで、複数の属性データ区間を取得するために、隣接する区間がマージされ、そして、第3のステップの属性データ区間の分類について、MAPベイズ評価値が算出される。

【0037】

第4のステップで、第2のステップで取得したMAPベイズ評価値が、第3のステップで取得したMAPベイズ評価値よりも小さい場合は、第1のステップで取得した複数の属性データ区間が、特定のユーザ属性に対応する最終属性データ区間として決定される。第5のステップで、第2のステップで取得したMAPベイズ評価値が、第3のステップで取得したMAPベイズ評価値以上である場合は、第3のステップで取得した複数の属性データ区間が維持され、属性データ区間の分類に関する最小のMAPベイズ評価値が取得されるまで、隣接する区間が連続的にマージされる。最小のMAPベイズ評価値を伴う属性データ区間が、特定のユーザ属性に対応する最終属性データ区間として決定される。

【0038】

本開示の第1の例示的な実施形態では、特定のユーザ属性に対応する複数の初期属性データ区間が決定されるときに、抽出された属性データの中で、特定のユーザ属性の属性データのそれぞれが、対応するコード値に変換され得、特定のユーザ属性に対応する複数の初期属性データ区間が、変換した後の複数のコード値および分類されたメンバーの分類に従って決定される。

【0039】

例えば、ユーザ属性の属性データは、数字、日付、および設定の3つの形を有し得る。例えば、ユーザ属性「年齢」の属性データは、数字タイプの属性データである。ユーザ属性「登録日」の属性データは、日付タイプの属性データである。ユーザ属性「都市」または「性別」の属性データは、設定タイプの属性データである。以下は、異なるタイプの属

10

20

30

40

50

性データの例示的な符号化方法をそれぞれ説明する。

【0040】

数字タイプの属性データの1つの例示的な符号化方法は、以下の通りである。特定のユーザ属性の属性データが数字タイプの属性データである場合は、属性データから抽出される特定のユーザ属性の各属性データに関して、属性データの値が、その対応するコード値として使用される。例えば、ユーザ属性「年齢」の属性データは20であり、その対応するコード値も20である。

【0041】

日付タイプの属性データの1つの例示的な符号化方法は、以下の通りである。特定のユーザ属性の属性データが日付タイプの属性データである場合は、属性データから抽出される特定のユーザ属性の各属性データに関して、属性データの時間と現在の時間との間の時間長が、属性データの対応するコード値として使用され得る。例えば、属性データの日付と現在の日付との間の日数が、属性データの対応するコード値として使用され得る。例えば、現在の日付は、2011年11月18日であり、したがって、属性データ2011年11月1日の対応するコード値は、17である。

【0042】

設定タイプの属性データの1つの例示的な符号化方法は、以下の通りである。特定のユーザ属性の属性データが設定タイプの属性データである場合は、属性データから抽出される特定のユーザ属性の各属性データに関して、属性データのそれぞれが主分類に属する比率値が算出される。高～低の比率値に基づいて、特定の属性データの属性データがランク付けされる。属性データから抽出される特定のユーザ属性の各属性データに関して、ランク付け後のランク付け順序番号が、属性データのそれぞれの対応するコード値として使用される。例えば、比率値のランク付け順序番号および複数の属性データの対応するコード値は、表2に示される。

【0043】

【表2】

表2

属性データ	属性データ1	属性データ2	属性データ3
比率値のランク付け順序番号	2	3	1
コード値	2	3	1

【0044】

第1の例示的な実施形態において、属性データ区間が分類されるときに、ユーザ属性の属性データが設定タイプの属性データである場合、最初に、属性データが符号化され、そして、符号化された属性データが、数字タイプの属性データを分類するための方法と同じ方法を使用することによって分類される。対照的に、従来の手法の下では、MAPベイズ評価値算出プロセスは、最大浮動小数点数を超える数をもたらす得、それによって、属性データ区間の不正確な分類およびメンバーの不正確な分類を生じさせる。したがって、本手法は、属性データ区間の分類およびメンバーの分類の精度を効果的に向上させる。

【0045】

206で、特定のユーザ属性に関する属性データが欠損している分類されたメンバーに関して、分類されたメンバーの特定のユーザ属性の属性データは、予め設定された欠損値として設定される。例えば、予め設定された欠損値としては、-5等の負の数が挙げられ得るが、それに限定されない。

【0046】

208で、予め設定された欠損値が、複数の初期属性データ区間のそれぞれにそれぞれマージされ、各マージの後に、複数の初期属性データ区間のMAPベイズ評価値が算出される。

【0047】

例えば、a1、a2、およびa3である、3つの初期属性データ区間があり得る。予め設定された欠損値が初期属性データ区間a1にマージされた後に、MAPベイズ評価値は、MAPベイズ評価値1である。予め設定された欠損値が初期属性データ区間a2にマージされた後に、MAPベイズ評価値は、MAPベイズ評価値2である。予め設定された欠損値が初期属性データ区間a3にマージされた後に、MAPベイズ評価値は、MAPベイズ評価値3である。結果は、表3に示される。

【0048】

【表3】

表3

10

予め設定された欠損値とマージされた後の複数の属性データ区間	MAPベイズ評価値
a1+予め設定された欠損値, a2, a3	MAPベイズ評価値1
a1, a2+予め設定された欠損値, a3	MAPベイズ評価値2
a1, a2, a3+予め設定された欠損値	MAPベイズ評価値3

【0049】

20

210で、最小のMAPベイズ評価値を伴う属性データ区間が、特定のユーザ属性に対応する最終属性データ区間として決定される。

【0050】

表3で示されるように、MAPベイズ評価値2が最小である場合、予め設定された欠損値とマージされた後の、a1、a2+予め設定された欠損値、a3である複数の属性データ区間が、特定のユーザ属性に対応する複数の最終属性データ区間として決定される。

【0051】

別の実施例では、最小のMAPベイズ評価値を伴う複数の属性データ区間が、特定のユーザ属性に対応する最終属性データ区間として決定される前に、以下の操作が行われ得る。

30

【0052】

特定のユーザ属性に関する属性データが欠損している分類されたメンバーの数が、第1の閾値未満であるかどうか判定される。判定の結果が第1の閾値未満である場合は、最小のMAPベイズ評価値を伴う複数の属性データ区間が、特定のユーザ属性に対応する最終属性データ区間として決定される。判定の結果が第1の閾値以上である場合は、予め設定された欠損値が、別個の属性データ区間として処理され、現在のMAPベイズ評価値が算出される。次いで、予め設定された欠損値が各初期属性データ区間にそれぞれマージされた後に、現在のMAPベイズ評価値が全てのMAPベイズ評価値未満であるかどうか判定される。判定の結果が、予め設定された欠損値が各初期属性データ区間にそれぞれマージされた後の現在のMAPベイズ評価値が全てのMAPベイズ評価値未満である場合は、決定された初期属性データ区間および別個の属性データ区間が、特定のユーザ属性に対応する最終属性データ区間として決定される。

40

【0053】

本開示によって開示される本手法により、ウェブサイトは、1つ以上のユーザ属性に関する1人以上の分類されたメンバーの属性データを抽出し得る。各ユーザ属性について、ウェブサイトは、特定のユーザ属性に対応する初期属性データ区間を決定する。特定の属性に関する属性データが欠損している分類されたメンバーに関して、ウェブサイトは、特定のユーザ属性に関する分類されたメンバーの属性データを、予め設定された欠損値として設定し、予め設定された欠損値を、決定された初期属性データ区間のそれぞれにそれぞれマージし、初期属性データ区間のそれぞれが予め設定された欠損値とそれぞれマージさ

50

れるときに、M A P ベイズ評価値を毎回算出し、そして、最小のM A P ベイズ評価値を伴う属性データ区間を、特定のユーザ属性に対応する最終属性データ区間として決定する。本手法は、属性データ区間を分類するときに、特定のユーザ属性に関する1人以上の分類されたメンバーの属性データが欠損しているかもしれない可能性を考慮し、それによって、属性データ区間の分類の精度を向上させる。本手法は、メンバーの属性データが欠損しているときの不正確さの問題を回避し、それによって、メンバー分類の精度を向上させる。

【0054】

第1の例示的な実施形態で説明されるような、属性データ区間を分類する方法に対応して、以下は、分類対象メンバーの分類のオンライン処理を例示する。図3は、本開示の第2の例示的な実施形態に従ってオンラインでメンバーを分類する、例示的な方法のフローチャートを図示する。

10

【0055】

302で、種々のユーザ属性に関する分類対象メンバーの属性データの中で、欠損している属性データが、予め設定された欠損値として設定される。例えば、予め設定された欠損値は、属性データ区間を分類するときの予め設定された欠損値に等しい。

【0056】

304で、分類対象メンバーの各ユーザ属性について、複数の属性データ区間の中で、特定のユーザ属性に対応する分類対象メンバーの属性データのユーザ属性区間が決定される。

20

【0057】

例えば、特定のユーザ属性の複数の初期属性データ区間がオフラインで決定される場合、特定のユーザ属性に関する抽出された属性データの中の各属性データが、対応するコード値に変換され、次いで、特定のユーザ属性に対応する複数の初期属性データ区間が決定される。次いで、304での操作の前に、分類対象メンバーの各ユーザ属性について、特定のユーザ属性に対応する属性データのそれぞれが、属性データに対応するコード値と置き換えられる。

【0058】

306で、決定された属性データ区間に対応するブル特性に値1が割り当てられ、値1を伴うブル特性の特性識別情報が記憶される。例えば、オフラインで分類された各属性データ区間には、1つのブル特性が割り当てられ得る。

30

【0059】

308で、複数のユーザ属性の中で、各ユーザ属性について値1を伴うブル特性に対応する特性識別情報が抽出された後に、各ユーザ属性からの各ブル特性の重み値に基づいて、分類対象メンバーが主分類に分類される確率が算出される。

【0060】

310で、特定の分類対象メンバーが主分類に分類される確率に基づいて、特定の分類対象メンバーの分類が算出される。確率が50%より高い場合は、特定の分類対象メンバーが主分類に分類される。確率が50%以下である場合は、特定の分類対象メンバーが副分類に分類される。

40

【0061】

以下は、本開示に従う第3の例示的な実施形態を説明する。本開示の第1の例示的な実施形態で説明される方法に従って属性データ区間が分類されるときに、分類された属性データ区間に含まれるレコードの数が少ない、または閾値に到達しない場合、分類された属性データ区間は、統計的な意味を有しない。分類された属性データ区間に基づいてメンバーが分類されるときに、分類の精度率は低い。したがって、本開示の第3の例示的な実施形態は、属性データ区間を分類する、別の例示的な方法を提供する。

【0062】

図4は、本開示の第3の例示的な実施形態に従って属性データ区間を分類する、別の例示的な方法のフローチャートを図示する。

50

【0063】

402で、1つ以上のユーザ属性に関する複数の分類されたメンバーの属性データが抽出される。404で、各特定のユーザ属性について、特定のユーザ属性および分類されたメンバーの分類に対応するユーザ属性データに基づいて、特定のユーザ属性に対応する初期属性データ区間が決定される。本開示の第3の例示的な実施形態の402および404での操作は、本開示の第1の例示的な実施形態の202および204での操作と同一または類似する。402および404での操作の詳細は、本明細書で説明されない。

【0064】

406で、特定のユーザ属性に対応する各決定された初期属性データ区間について、属性データが特定の初期属性データ区間に分類される分類されたメンバーの数が、第2の閾値よりも高いかどうか判定される。初期属性データ区間のそれぞれに関する判定の結果が肯定的である場合は、412での操作が行われる。少なくとも1つの初期属性データに関する決定の結果が否定的である場合は、408での操作が行われる。

10

【0065】

第2の閾値は、予め設定され得るか、または分類された初期属性データ区間が統計的な意味を有するかどうかに基づいて設定され得る。例えば、第2の閾値は、30として設定され得るが、それに限定されない。

【0066】

408で、その判定の結果が否定的である各初期ユーザ属性データ区間について、以下の操作が行われる。初期属性データ区間を前の初期属性データ区間とマージしたときのMAPベイズ評価値、および初期属性データ区間を次の初期属性データ区間とマージしたときのMAPベイズ評価値が算出される。

20

【0067】

410で、より小さいMAPベイズ評価値を伴うマージした複数の属性データ区間が、特定のユーザ属性に対応する複数の初期ユーザ属性として使用される。決定された複数の初期属性データ区間は、特定の属性データ区間について、属性データが特定の属性データ区間に分類される、分類されたメンバーの数が第2の閾値よりも大きいという条件を少なくとも満たすように選択される。

【0068】

412で、特定のユーザ属性に関する属性データが欠損している分類されたメンバーについて、特定のユーザ属性に関する分類されたメンバーの属性データが、予め設定された欠損値として設定される。

30

【0069】

414で、欠損値が、決定された初期属性データ区間のそれぞれにそれぞれマージされ、決定された初期属性データ区間のそれぞれに欠損値がマージされた後に、MAPベイズ評価値がそれぞれ毎回算出される。

【0070】

416で、最小のMAPベイズ評価値を伴う複数の属性データ区間が、特定のユーザ属性に対応する最終属性データ区間として決定される。本開示の第3の例示的な実施形態の412～416の操作は、本開示の第1の例示的な実施形態の206～210の操作と同一または類似する。簡潔にするために、詳細は、本明細書で繰り返さない。

40

【0071】

本開示の第3の例示的な実施形態は、ユーザ属性データが欠損している可能性があるシナリオを考慮する。ユーザ属性に対応する初期属性データ区間が決定されるときに、初期属性データ区間は、特定のユーザ属性に対応する複数の属性データ区間として直接使用されない。特定のユーザ属性の各決定された初期属性データ区間について、属性データが特定の初期属性データ区間に分類される分類されたメンバーの数が、第2の閾値よりも高いかどうか判定される。少なくとも1つの初期属性データ区間に関する判定の結果が否定的である場合は、否定的な結果を伴う各初期属性データ区間について、特定の初期属性データ区間を前の初期属性データ区間とマージしたときのMAPベイズ評価値、および特定

50

の初期属性データ区間を次の初期属性データ区間とマージしたときのMAPベイズ評価値が算出される。より小さいMAPベイズ評価値を伴う、マージした複数の属性データ区間は、特定のユーザ属性に対応する複数の初期ユーザ属性として使用され、次いで、属性データにおける欠損データのシナリオに基づいて、特定のユーザ属性に対応する最終属性データ区間がさらに決定される。本開示の上の第3の例示的な実施形態で示されるように、特定のユーザ属性に対応する複数の初期属性データ区間が決定されるときに、属性データが各初期属性データ区間に分類される分類されたメンバーの数が、第2の閾値よりも高い場合、決定された複数の初期属性データ区間のそれぞれにおける分類されたメンバーの数が第2の閾値よりも高いことを確実にするために、属性データ区間がマージされる。したがって、分類された複数の属性データ区間は、統計的な意味を有し、それによって、属性データ区間を分類する精度が向上し、さらに、メンバーを分類する精度が向上する。

10

【0072】

以下は、本開示の第4の例示的な実施形態を説明する。分類対象メンバーの属性データの中に欠損データがある一方で、分類されたメンバーの属性データの中にはいかなる欠損データもない場合、トレーニングデータに基づく、分類された複数の属性データ区間は、欠損値を有しない。したがって、分類対象メンバーの属性データ区間を決定することができず、また、分類対象メンバーの分類を決定することができない。故に、本開示の第4の例示的な実施形態は、図5で示されるような、分類対象メンバーをオンラインで分類する例示的な方法を提供する。

【0073】

20

502で、分類対象メンバーの各ユーザ属性について、分類対象メンバーのユーザ属性の属性データが欠損している場合は、特定のユーザ属性の所定の複数の属性データ区間が、予め設定された欠損値を含む1つ以上の属性データ区間を含むかどうか判定される。判定の結果が肯定的である場合は、本開示の第2の例示的な実施形態で説明される方法に従って、分類対象メンバーが分類される。

【0074】

504で、502での判定の結果が否定的である場合は、各属性データ区間の主分類の比率が取得される。例えば、各属性データ区間における主分類の比率は、属性データ区間がオフラインで分類されるときに決定され得る。最終属性データ区間が決定された後に、各属性データ区間における主分類の比率または主分類のメンバーに属する属性データの比率が算出され、各属性データ区間とその主分類の比率との間の対応関係が記憶される。

30

【0075】

506で、主分類の最小の比率を伴う属性データ区間が、分類対象メンバーの属性データの特定の属性に関する属性データ区間として決定される。

【0076】

508で、決定された属性データ区間のブル特性が1として設定され、値1を伴うブル特性の特性識別情報が記憶される。例えば、オフラインで分類される属性データ区間のそれぞれには、ブル特性が割り当てられ得る。

【0077】

510で、ユーザ属性のそれぞれについて、ブル特性が1である特性識別情報がそれぞれ抽出された後に、ブル特性のそれぞれの重み値に従って、分類対象メンバーが主分類に分類される確率が算出される。

40

【0078】

512で、分類対象メンバーが主分類に分類される確率に基づいて、分類対象メンバーの分類が決定される。

【0079】

以下は、本開示の第5の例示的な実施形態を説明する。図6は、本開示の第5の例示的な実施形態に従って属性データ区間进行分类する、例示的な装置600を図示する。図6の実施例において、装置600は、1つ以上のプロセッサ602と、メモリ604とを含み得るが、それらに限定されない。メモリ604は、コンピュータ記憶媒体の例である。コ

50

コンピュータ記憶媒体としては、コンピュータが実行可能な命令、データ構造、プログラムモジュール、または他のデータ等の情報を記憶するための任意の方法または技術で実現される、揮発性および不揮発性媒体、リムーバブルおよび非リムーバブル媒体が挙げられる。コンピュータ記憶媒体の例としては、相変化メモリ（PRAM）、スタティックランダムアクセスメモリ（SRAM）、ダイナミックランダムアクセスメモリ（DRAM）、他のタイプのランダムアクセスメモリ（RAM）、リードオンリーメモリ（ROM）、電気的消去可能プログラマブルリードオンリーメモリ（EEPROM）、フラッシュメモリ、もしくはその他メモリ技術、読み出し専用コンパクトディスク（CD-ROM）、デジタル多用途ディスク（DVD）、もしくは他の光記憶装置、磁気カセット、磁気テープ、磁気ディスク記憶装置、もしくは他の磁気記憶デバイス、またはコンピューティングデバイスによってアクセスするための情報を記憶するために使用することができる、任意の他の非伝送媒体が挙げられるが、これらに限定されない。本明細書で定義されるように、コンピュータ記憶媒体は、変調データ信号および搬送波等の一時的媒体を含まない。

10

【0080】

メモリ604は、その中にプログラムユニットまたはモジュールおよびプログラムデータを記憶し得る。図6の実施例において、メモリ604は、その中に、属性データ抽出ユニット606と、第1の初期属性データ区間決定ユニット608と、欠損値設定ユニット610と、第1の評価値算出ユニット612と、第1の最終属性データ区間決定ユニット614とを記憶することができる。

20

【0081】

属性データ抽出ユニット606は、1つ以上のユーザ属性に関する複数の分類されたメンバーの属性データを抽出する。第1の初期属性データ区間決定ユニット608は、各特定のユーザ属性について、属性データ抽出ユニット606によって抽出される特定のユーザ属性に対応するユーザ属性データおよび分類されたメンバーの分類に基づいて、特定のユーザ属性に対応する初期属性データ区間を決定する。

【0082】

欠損値設定ユニット610は、特定のユーザ属性に関する属性データが欠損している分類されたメンバーについて、特定のユーザ属性に対応する分類されたメンバーの属性データを、予め設定された欠損値として設定する。

【0083】

第1の評価算出ユニット612は、予め設定された欠損値を、第1の初期属性データ区間決定ユニット608によってそれぞれ決定される各初期属性データ区間にマージし、予め設定された欠損値が1つの初期属性データ区間にそれぞれマージされた後に、初期ユーザ属性データ区間のMAPベイズ評価値を毎回算出する。

30

【0084】

第1の最終属性データ区間決定ユニット614は、特定のユーザ属性に対応する最終属性データ区間として、各マージャから、最小のMAPベイズ評価値を伴う初期属性データ区間を決定する。

【0085】

例えば、メモリ604はまた、その中に、第1の数判定ユニットも記憶し得る。第1の最終属性データ区間決定ユニット614が、特定のユーザ属性に対応する最終属性データ区間として、各マージャから、最小のMAPベイズ評価値を伴う初期属性データ区間を決定する前に、第1の数判定ユニットが、特定のユーザ属性に関する属性データが欠損している分類されたメンバーの数が、第1の閾値未満であるかどうかを判定する。

40

【0086】

特定のユーザ属性に関する属性データが欠損している分類されたメンバーの数が、第1の閾値未満である場合は、第1の最終属性データ区間決定ユニット614が、第1の評価値算出ユニット612によって算出される、各マージャからの最小のMAPベイズ評価値を伴う初期属性データ区間を、特定のユーザ属性に対応する最終属性データ区間として決定する。

50

【0087】

別の実施例について、メモリ604はまた、その中に、第2の評価値算出ユニット、評価値判定ユニット、および第2の最終属性データ区間決定ユニットも記憶し得る。

【0088】

特定のユーザ属性に関する属性データが欠損している分類されたメンバーの数が、第1の閾値以上である場合は、第2の評価値判定ユニットが、欠損値を別個の属性データ区間として処理し、現在のMAPベイズ評価値を算出する。

【0089】

評価値判定ユニットは、欠損値が第1の評価値算出ユニット612によって算出された各初期属性データ区間にマージされた後に、第2の評価値判定ユニットによって算出される現在のMAPベイズ評価値が、全てのMAPベイズ評価値未満であるかどうかを判定する。

10

【0090】

判定の結果が、欠損値が各初期属性データ区間にそれぞれマージされた後の現在のMAPベイズ評価値が全てのMAPベイズ評価値未満である場合は、第2の最終属性データ区間決定ユニットは、第1の初期属性データ区間決定ユニット608によって決定される初期属性データ区間および別個の属性データ区間を、ユーザ属性に対応する最終属性データ区間として決定する。

【0091】

判定の結果が、欠損値が各初期属性データ区間にマージされた後の現在のMAPベイズ評価値が全てのMAPベイズ評価値以上である場合は、第1の最終属性データ区間決定ユニット614は、第1の評価値算出ユニット612によって算出される最小のMAPベイズ評価値に伴う属性データ区間を、特定のユーザ属性に対応する最終属性データ区間として決定する。

20

【0092】

別の実施例について、メモリ604はまた、その中に、第2の数判定ユニット、第3の評価値算出ユニット、および第2の初期属性データ区間決定ユニットも記憶し得る。

【0093】

欠損値設定ユニット610の前に、特定のユーザ属性に関する属性データが欠損している分類されたメンバーに関して、特定のユーザ属性に対応する分類されたメンバーの属性データを、予め設定された欠損値として設定し、第2の数判定ユニットは、属性データが特定のユーザ属性に対応する各初期属性データ区間に分類される分類されたメンバーの数が、第2の閾値よりも高いかどうかを判定する。特定のユーザ属性の各初期属性データ区間は、第1の初期属性データ区間決定ユニット608によって決定される。

30

【0094】

全ての初期属性データ区間について第2の数判定ユニットによって行われる判定の結果が肯定的である場合、欠損値設定ユニット610は、特定のユーザ属性に関する属性データが欠損している分類されたメンバーに関して、特定のユーザ属性に対応する分類されたメンバーの属性データを、予め設定された欠損値として設定する。

【0095】

全ての初期属性データ区間について第2の数判定ユニットによって行われる少なくとも1つの判定の結果が否定的である場合、第3の評価値算出ユニットは、否定的な結果を伴う各初期ユーザ属性データ区間に関して、そのような初期ユーザ属性データ区間が前の初期属性データ区間とマージされるときMAPベイズ評価値、およびそのような初期ユーザ属性データ区間が次の初期属性データ区間とマージされるときMAPベイズ評価値を算出する。

40

【0096】

第2の初期属性データ区間決定ユニットは、第3の評価値判定ユニットによって算出される、より小さいMAPベイズ評価値を伴うマージされた属性データ区間を、特定のユーザ属性に対応する初期属性データ区間として選択する。

50

【0097】

一実施例において、第1の初期属性データ区間決定ユニット608は、コード値変換サブユニットと、初期属性データ区間決定サブユニットとを含み得る。コード値変換サブユニットは、属性データ抽出ユニット606によって抽出される特定のユーザ属性の各属性データを、対応するコード値に変換する。初期属性データ区間決定サブユニットは、コード評価変換サブユニットによって変換されるコード値および分類されたメンバーの分類に基づいて、特定のユーザ属性に対応する初期属性データ区間を決定する。

【0098】

例えば、特定のユーザ属性の属性データが数字タイプの属性データである場合、属性データから抽出される特定のユーザ属性の各属性データに関して、コード値変換ユニットは、属性データの値を、その対応するコード値として使用し得る。

10

【0099】

別の実施例について、特定のユーザ属性の属性データが日付タイプの属性データである場合、属性データから抽出される特定のユーザ属性の各属性データに関して、コード値変換ユニットは、属性データの時間と現在の時間との間の時間長を、属性データのコード値として使用し得る。

【0100】

別の実施例について、特定のユーザ属性の属性データが設定タイプの属性データである場合、属性データから抽出される特定のユーザ属性の各属性データに関して、コード値変換ユニットは、属性データのそれぞれが主分類に属する比率値を算出し、高～低の比率値に基づいて、特定の属性の属性データをランク付けし、そして、属性データから抽出される特定のユーザ属性の各属性データに関して、ランク付け後のランク付け順序番号を、属性データのそれぞれの対応するコード値として使用し得る。

20

【0101】

当業者は、本開示の実施形態が、方法、システム、またはコンピュータプログラミング製品であり得ることを理解すべきである。したがって、本開示は、ハードウェア、ソフトウェア、または両者の組み合わせによって実装され得る。さらに、本開示は、コンピュータ記憶媒体（ディスク、CD-ROM、光ディスク等を含むが、これらに限定されない）に実装され得る、コンピュータが実行可能な、またはプロセッサが実行可能な命令を含む1つ以上のコンピュータプログラムの形式であってもよい。

30

【0102】

本開示は、本開示の実施形態の方法、装置（システム）、およびコンピュータプログラムのフローチャート、および／またはブロック図を参照して説明される。フローチャートおよび／またはブロック図の、各フローおよび／またはブロック、ならびにフローおよび／またはブロックの組み合わせは、コンピュータプログラム命令によって実装され得ることを理解されたい。コンピュータ、または他のプログラム可能なデータプロセッサによって動作される命令を通して、フローチャートの1つ以上のフロー、および／またはブロック図の1つ以上のブロックを実装している装置が生成され得るように、これらのコンピュータプログラム命令は、マシンを生成するために、一般的なコンピュータ、特定のコンピュータ、内蔵プロセッサ、または他のプログラム可能なデータプロセッサに提供されることができる。

40

【0103】

これらのコンピュータプログラム命令は、コンピュータが読み出し可能な記憶装置内に記憶された命令が、フローチャートの1つ以上のフローおよび／またはブロック図の1つ以上のブロックに特定される機能を実装する命令装置を含む製品を生成するように、コンピュータ、または他のプログラム可能なデータプロセッサに、ある一定の動作をするよう命令することができる他のコンピュータ記憶媒体に記憶されることもできる。

【0104】

コンピュータ、または他のプログラム可能なデータプロセッサが、コンピュータによって実装されるプロセスを生成するための一連の動作ステップを実行し得るように、これら

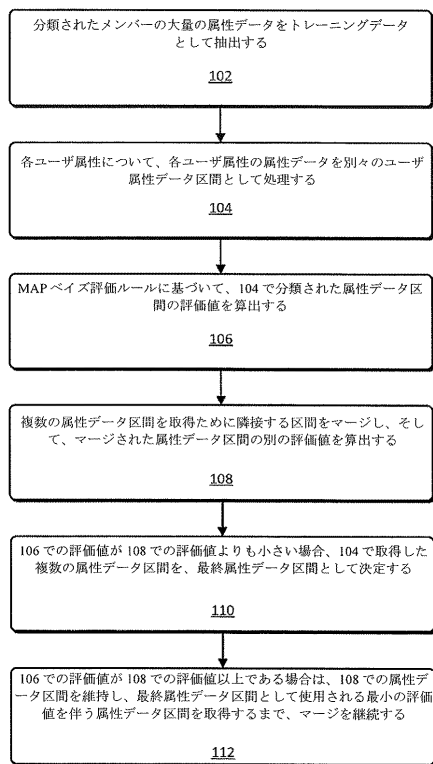
50

のコンピュータプログラム命令はまた、コンピュータ、または他のプログラム可能なデータプロセッサに読み込まれることができる。これにより、コンピュータ、または他のプログラム可能なデータプロセッサで実行される命令は、フローチャートの1つ以上のフローおよび／またはブロック図の1つ以上のブロックに特定される機能を実装するステップを提供し得る。

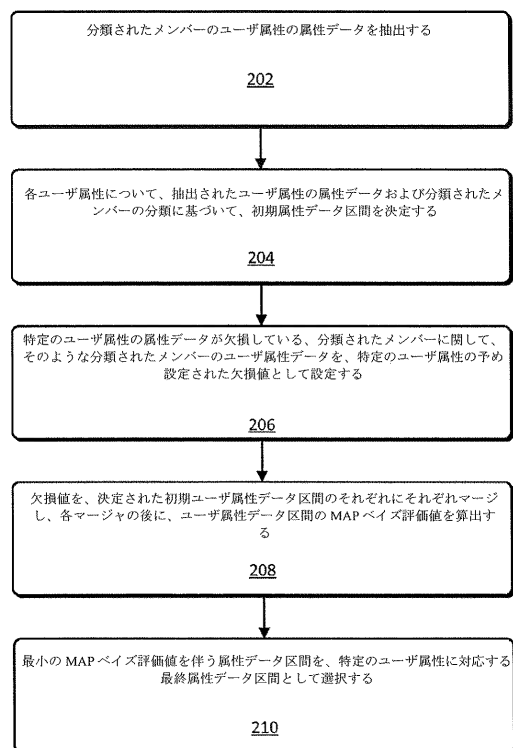
【0105】

これらの実施形態は、本開示を説明するに過ぎず、本開示の範囲を限定するものではない。当業者には、本開示の原則から逸脱することなく、若干の変更、および改良が行われてもよく、また本開示の範囲内にあるとみなされるべきであることを理解されたい。

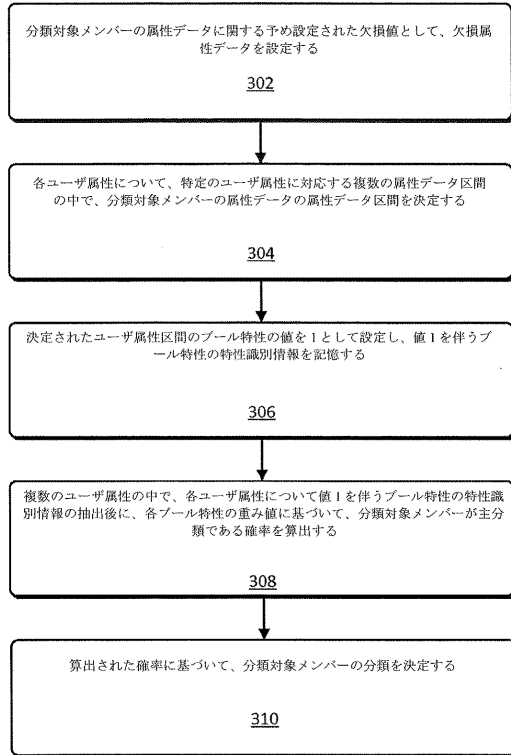
【図1】



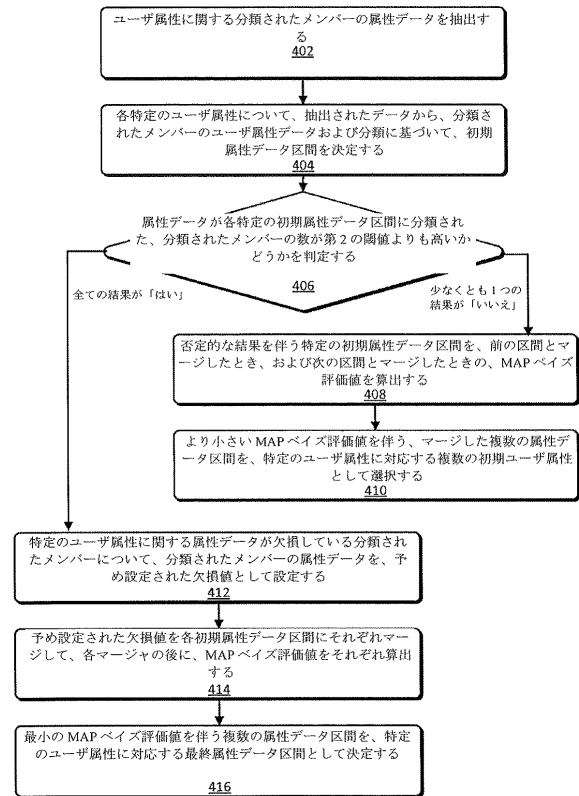
【図2】



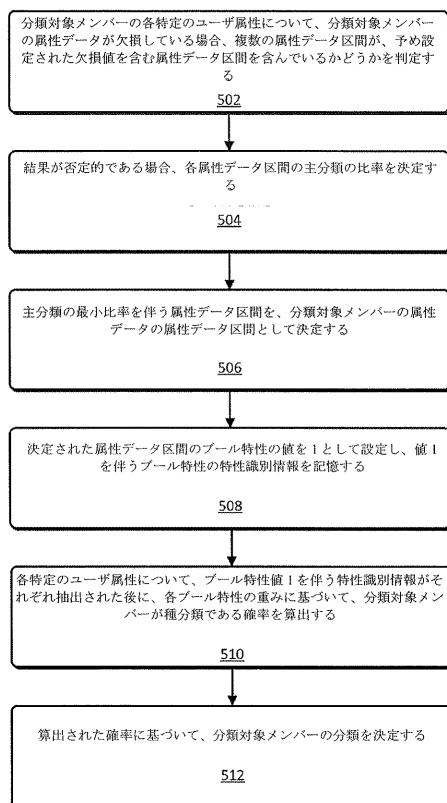
【図 3】



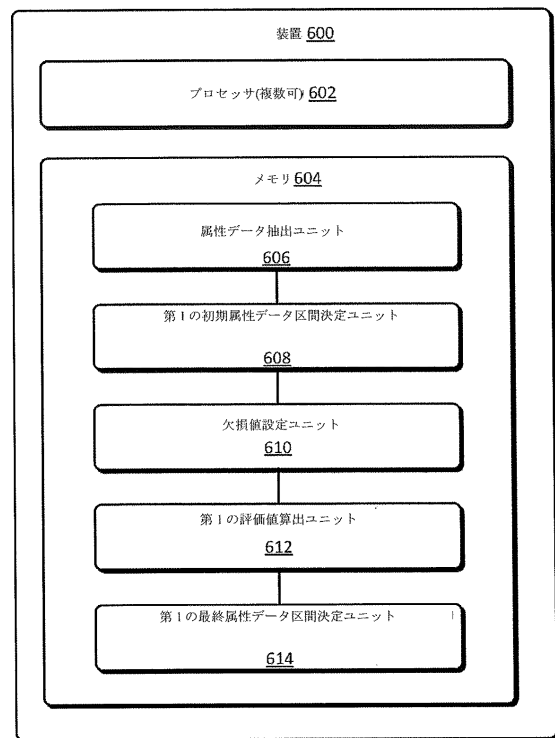
【図 4】



【図 5】



【図 6】



フロントページの続き

(72)発明者 シャオ ジードン

中華人民共和国 310013 ハンチョウ ウェン アル ロード ナンバー 391 ザ ウ
ェスト レイク インターナショナル プラザ オブ エスアンドティー ビルディング エー
10/エフ アリババ グループ リーガル デパートメント内

審査官 樋口 龍弥

(56)参考文献 特開2000-003281 (JP, A)

欧州特許出願公開第01014281 (EP, A1)

特表2003-527686 (JP, A)

(58)調査した分野(Int.Cl., DB名)

G06F 17/30

G06N 7/00

G06N 99/00