

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5104496号
(P5104496)

(45) 発行日 平成24年12月19日(2012.12.19)

(24) 登録日 平成24年10月12日(2012.10.12)

(51) Int. Cl. F 1
G06Q 10/00 (2012.01) G06F 17/60 150
G06Q 10/04 (2012.01) G06F 19/00 100

請求項の数 36 (全 50 頁)

(21) 出願番号	特願2008-101872 (P2008-101872)	(73) 特許権者	000004237
(22) 出願日	平成20年4月9日(2008.4.9)		日本電気株式会社
(65) 公開番号	特開2009-238193 (P2009-238193A)		東京都港区芝五丁目7番1号
(43) 公開日	平成21年10月15日(2009.10.15)	(74) 代理人	100103090
審査請求日	平成23年3月10日(2011.3.10)		弁理士 岩壁 冬樹
(31) 優先権主張番号	特願2008-57281 (P2008-57281)	(74) 代理人	100124501
(32) 優先日	平成20年3月7日(2008.3.7)		弁理士 塩川 誠人
(33) 優先権主張国	日本国(JP)	(72) 発明者	黒岩 由希子
			東京都港区芝五丁目7番1号 日本電気株式会社社内
		審査官	青柳 光代

最終頁に続く

(54) 【発明の名称】 普及予測システム、方法およびプログラム、並びに影響度推定システム、方法およびプログラム

(57) 【特許請求の範囲】

【請求項1】

顧客が商品またはサービスを利用し始めている場合には前記商品またはサービスの利用開始時期を表し、前記商品またはサービスを未利用の場合には未利用である旨を表す利用開始情報と、前記利用開始情報以外の顧客の属性を表す一つ以上の項目とを含む顧客毎の個別データの集合である顧客データを記憶する顧客データベースと、

前記顧客データベースを用いて、判定対象時刻を定める現時刻で商品またはサービスの利用を開始していない顧客の個別データを含むデータであるテストデータを生成するテストデータ生成部と、

前記現時刻で前記商品またはサービスの利用を開始している顧客の個別データに第1ラベルをラベル付け、前記商品またはサービスの利用を開始していない顧客の個別データに第2ラベルをラベル付け、第1ラベルをラベル付けた個別データ数を、ある期間の中で商品またはサービスの顧客が他の者にその商品またはサービスの利用を喚起させる度合いである影響度に応じて変動させたデータである学習データを生成する学習データ生成部と、

前記顧客の属性を表す項目から当該顧客の個別データに第1ラベルと第2ラベルのいずれをラベル付けるかを判定するルールである分類器を、前記学習データに基づいて生成する分類器生成部と、

前記分類器と前記テストデータ内の各個別データの項目とから、前記テストデータ内の各個別データに対するラベルを判定するテストデータラベル判定部とを備える

ことを特徴とする普及予測システム。

【請求項 2】

前記現時刻を一定の時間分増加した時刻を新たな現時刻とする現時刻更新部を備え、
テストデータラベル判定部は、第 1 ラベルがラベル付けられると判定した個別データに対して、利用開始時期を示す情報を付与し、

テストデータ生成部は、現時刻が更新された際に、更新後の現時刻におけるテストデータを生成し、

学習データ生成部は、現時刻が更新された際に、更新後の現時刻における学習データを生成し、

分類器生成部は、現時刻が更新されて新たに学習データが生成された際に、当該学習データに基づいて新たに分類器を生成し、

テストデータラベル判定部は、現時刻が更新されて新たにテストデータが生成された際に、前記分類器と前記テストデータ内の各個別データの項目とから、前記テストデータ内の各個別データに対するラベルを判定する

請求項 1 に記載の普及予測システム。

【請求項 3】

テストデータ生成部は、顧客データベースに記憶されている顧客データから、現時刻で商品またはサービスの利用を開始している顧客の個別データを除外したテストデータを生成する

請求項 1 または請求項 2 に記載の普及予測システム。

【請求項 4】

学習データ生成部は、商品またはサービスの利用を開始した顧客数と未だ開始していない顧客数との比率を影響度の係数とする関数の関数値倍に、第 1 ラベルをラベル付けた個別データ数を増加させて学習データを生成する

請求項 1 から請求項 3 のうちのいずれか 1 項に記載の普及予測システム。

【請求項 5】

学習データ生成部は、商品またはサービスの利用を開始した顧客数と全顧客数との比率を影響度の係数とする関数の関数値倍に、第 1 ラベルをラベル付けた個別データ数を増加させて学習データを生成する

請求項 1 から請求項 3 のうちのいずれか 1 項に記載の普及予測システム。

【請求項 6】

分類器生成部は、学習データ内で第 1 ラベルをラベル付けられた個別データの割合が高くなるほど、テストデータ内で第 2 ラベルをラベル付けられている個別データのうち、前記第 1 ラベルをラベル付けられた個別データの項目値と類似する項目値を持つ個別データに対して、第 1 ラベルをラベル付けると判定する頻度が高くなる分類器を生成する

請求項 1 から請求項 5 のうちのいずれか 1 項に記載の普及予測システム。

【請求項 7】

学習データ生成部は、当該普及予測システムの外部から入力された影響度を用いて学習データを生成する

請求項 1 から請求項 6 のうちのいずれか 1 項に記載の普及予測システム。

【請求項 8】

判定対象時刻を定めるとともに影響度の推定対象時刻としても用いられる現時刻と、前記現時刻の一定時間前を指定するための時刻間隔と、影響度の候補である複数の仮影響度とが入力され、前記現時刻から前記時刻間隔前の時刻である前時刻を計算し、顧客データベースを用いて、前記現時刻で前記商品またはサービスの利用を開始している顧客の個別データに第 1 ラベルをラベル付け、前記商品またはサービスの利用を開始していない顧客の個別データに第 2 ラベルをラベル付けた現時刻データを生成する現時刻データ生成部と、

個々の仮影響度毎に、前時刻で前記商品またはサービスの利用を開始している顧客の個別データに第 1 ラベルをラベル付け、前記商品またはサービスの利用を開始していない顧客の個別データに第 2 ラベルをラベル付け、第 1 ラベルをラベル付けた個別データ数を前

10

20

30

40

50

記仮影響度に応じて変動させたデータである前時刻データを生成する前時刻データ群生成部と、

個々の前時刻データ毎に、前記顧客の属性を表す項目から当該顧客の個別データに第1ラベルと第2ラベルのいずれをラベル付けるかを判定するルールである分類器を前記前時刻に基づいて生成する前時刻分類器群生成部と、

前時刻データ毎に生成された個々の分類器毎に、当該分類器と前記現時刻データ内の各個別データの項目とから前記現時刻データ内の個別データにラベル付けられるラベルを予測し、予測結果と前記現時刻データとの誤差を算出する誤差群算出部と、

個々の分類器毎に算出された誤差のうち、最小の誤差を特定し、最小の誤差に対応する仮影響度が1つである場合には、当該仮影響度を影響度として定め、最小の誤差に対応する仮影響度が複数個存在する場合には、前記複数の仮影響度に基づいて影響度を定める影響度算出部とを備える

10

請求項1から請求項6のうちのいずれか1項に記載の普及予測システム。

【請求項9】

現時刻データ生成部は、顧客データベースに記憶されている顧客データのうち、前時刻で商品またはサービスの利用を開始していた顧客の個別データを除外した個別データの集合から現時刻データを生成する

請求項8に記載の普及予測システム。

【請求項10】

前時刻データ群生成部は、前時刻で商品またはサービスの利用を開始している顧客数と前時刻で未だ開始していない顧客数との比率を仮影響度の係数とする関数の関数値倍に、第1ラベルをラベル付けた個別データ数を増加させて前時刻データを生成する

20

請求項8または請求項9に記載の普及予測システム。

【請求項11】

前時刻データ群生成部は、前時刻で商品またはサービスの利用を開始している顧客数と全顧客数との比率を仮影響度の係数とする関数の関数値倍に、第1ラベルをラベル付けた個別データ数を増加させて前時刻データを生成する

請求項8または請求項9に記載の普及予測システム。

【請求項12】

誤差群算出部は、前時刻データ毎に生成された個々の分類器毎に、当該分類器と前記現時刻データ内の各個別データの項目とから前記現時刻データ内の個別データにラベル付けられるラベルを予測し、予測したラベルと、現時刻データで実際にラベル付けられているラベルとが異なっている個別データの数を誤差として算出する

30

請求項8から請求項11のうちのいずれか1項に記載の普及予測システム。

【請求項13】

誤差群算出部は、前時刻データ毎に生成された個々の分類器毎に、当該分類器と前記現時刻データ内の各個別データの項目とから前記現時刻データ内の個別データにラベル付けられるラベルを予測し、第1ラベルがラベル付けられると予測した個別データ数と、現時刻データで第1ラベルがラベル付けられている個別データ数との差の絶対値を誤差として算出する

40

請求項8から請求項11のうちのいずれか1項に記載の普及予測システム。

【請求項14】

顧客が商品またはサービスを利用し始めている場合には前記商品またはサービスの利用開始時期を表し、前記商品またはサービスを未利用の場合には未利用である旨を表す利用開始情報と、前記利用開始情報以外の顧客の属性を表す一つ以上の項目とを含む顧客毎の個別データの集合である顧客データを記憶する顧客データベースと、

ある期間の中で商品またはサービスの顧客が他の者にその商品またはサービスの利用を喚起させる度合いである影響度の推定対象時刻として用いられる現時刻と、前記現時刻の一定時間前を指定するための時刻間隔と、影響度の候補である複数の仮影響度とが入力され、前記現時刻から前記時刻間隔前の時刻である前時刻を計算し、前記顧客データベース

50

を用いて、前記現時刻で前記商品またはサービスの利用を開始している顧客の個別データに第1ラベルをラベル付け、前記商品またはサービスの利用を開始していない顧客の個別データに第2ラベルをラベル付けた現時刻データを生成する現時刻データ生成部と、

個々の仮影響度毎に、前時刻で前記商品またはサービスの利用を開始している顧客の個別データに第1ラベルをラベル付け、前記商品またはサービスの利用を開始していない顧客の個別データに第2ラベルをラベル付け、第1ラベルをラベル付けた個別データ数を前記仮影響度に応じて変動させたデータである前時刻データを生成する前時刻データ群生成部と、

個々の前時刻データ毎に、前記顧客の属性を表す項目から当該顧客の個別データに第1ラベルと第2ラベルのいずれをラベル付けるかを判定するルールである分類器を前記前時刻に基づいて生成する前時刻分類器群生成部と、

前時刻データ毎に生成された個々の分類器毎に、当該分類器と前記現時刻データ内の各個別データの項目とから前記現時刻データ内の個別データにラベル付けられるラベルを予測し、予測結果と前記現時刻データとの誤差を算出する誤差群算出部と、

個々の分類器毎に算出された誤差のうち、最小の誤差を特定し、最小の誤差に対応する仮影響度が1つである場合には、当該仮影響度を影響度として定め、最小の誤差に対応する仮影響度が複数個存在する場合には、前記複数の仮影響度に基づいて影響度を定める影響度算出部とを備える

ことを特徴とする影響度推定システム。

【請求項15】

現時刻データ生成部は、顧客データベースに記憶されている顧客データのうち、前時刻で商品またはサービスの利用を開始していた顧客の個別データを除外した個別データの集合から現時刻データを生成する

請求項14に記載の影響度推定システム。

【請求項16】

前時刻データ群生成部は、前時刻で商品またはサービスの利用を開始している顧客数と前時刻で未だ開始していない顧客数との比率を仮影響度の係数とする関数の関数値倍に、第1ラベルをラベル付けた個別データ数を増加させて前時刻データを生成する

請求項14または請求項15に記載の影響度推定システム。

【請求項17】

前時刻データ群生成部は、前時刻で商品またはサービスの利用を開始している顧客数と全顧客数との比率を仮影響度の係数とする関数の関数値倍に、第1ラベルをラベル付けた個別データ数を増加させて前時刻データを生成する

請求項14または請求項15に記載の影響度推定システム。

【請求項18】

誤差群算出部は、前時刻データ毎に生成された個々の分類器毎に、当該分類器と前記現時刻データ内の各個別データの項目とから前記現時刻データ内の個別データにラベル付けられるラベルを予測し、予測したラベルと、現時刻データで実際にラベル付けられているラベルとが異なっている個別データの数を誤差として算出する

請求項14から請求項17のうちのいずれか1項に記載の影響度推定システム。

【請求項19】

誤差群算出部は、前時刻データ毎に生成された個々の分類器毎に、当該分類器と前記現時刻データ内の各個別データの項目とから前記現時刻データ内の個別データにラベル付けられるラベルを予測し、第1ラベルがラベル付けられると予測した個別データ数と、現時刻データで第1ラベルがラベル付けられている個別データ数との差の絶対値を誤差として算出する

請求項14から請求項17のうちのいずれか1項に記載の影響度推定システム。

【請求項20】

顧客が商品またはサービスを利用し始めている場合には前記商品またはサービスの利用開始時期を表し、前記商品またはサービスを未利用の場合には未利用である旨を表す利用

10

20

30

40

50

開始情報と、前記利用開始情報以外の顧客の属性を表す一つ以上の項目とを含む顧客毎の個別データの集合である顧客データを記憶する顧客データベースと、

前記顧客データベースを用いて、判定対象時刻を定める現時刻で商品またはサービスの利用を開始していない顧客の個別データを含むデータであるテストデータを生成するテストデータ生成部と、

前記現時刻で前記商品またはサービスの利用を開始している顧客の個別データに第1ラベルをラベル付け、前記商品またはサービスの利用を開始していない顧客の個別データに第2ラベルをラベル付けたラベル付けデータを生成し、前記ラベル付けデータ中の第1ラベルでラベル付けられた個別データ数の割合を前記商品またはサービスの普及率として計算するラベル付けデータ生成部と、

10

前記顧客の属性を表す項目から当該顧客の個別データに第1ラベルがラベル付けられる確からしさのスコアを定めるルールである分類器を、前記ラベル付けデータに基づいて生成する分類器生成部と、

普及率に対するスコアの閾値の関数である閾値関数にラベル付けデータ生成部が計算した普及率を代入して閾値を算出し、前記分類器と前記テストデータ内の各個別データの項目とから、前記テストデータ内の各個別データに第1ラベルがラベル付けられる確からしさのスコアを定め、前記テストデータ内の個別データのうちスコアが前記閾値以上である個別データに対するラベルが第1ラベルであると判定し、スコアが前記閾値未満の個別データに対するラベルが第2ラベルであると判定するテストデータラベル判定部とを備えることを特徴とする普及予測システム。

20

【請求項21】

前記現時刻を一定の時間分増加した時刻を新たな現時刻とする現時刻更新部を備え、

テストデータラベル判定部は、第1ラベルがラベル付けられると判定した個別データに対して、利用開始時期を示す情報を付与し、

テストデータ生成部は、現時刻が更新された際に、更新後の現時刻におけるテストデータを生成し、

ラベル付けデータ生成部は、現時刻が更新された際に、更新後の現時刻におけるラベル付けデータを生成し、当該ラベル付けデータにおける普及率を計算し、

分類器生成部は、現時刻が更新されて新たにラベル付けデータが生成された際に、当該ラベル付けデータに基づいて新たに分類器を生成し、

30

テストデータラベル判定部は、現時刻が更新されて新たにテストデータが生成された際に、前記分類器と前記テストデータ内の各個別データの項目とから、前記テストデータ内の各個別データに第1ラベルがラベル付けられる確からしさのスコアを定め、閾値関数にラベル付けデータ生成部が計算した普及率を代入して新たに閾値を算出し、前記テストデータ内の個別データのうちスコアが前記閾値以上である個別データに対するラベルが第1ラベルであると判定し、スコアが前記閾値未満の個別データに対するラベルが第2ラベルであると判定する

請求項20に記載の普及予測システム。

【請求項22】

テストデータ生成部は、顧客データベースに記憶されている顧客データから、現時刻で商品またはサービスの利用を開始している顧客の個別データを除外したテストデータを生成する

40

請求項20または請求項21に記載の普及予測システム。

【請求項23】

閾値の推定対象時刻を定めるための複数の基準時刻と、各基準時刻の一定時間前を指定するための時刻間隔と、閾値の候補である複数の仮閾値とが入力され、前記基準時刻毎に、基準時刻から前記時刻間隔前の時刻である閾値推定前時刻を計算し、顧客データベースを用いて、前記基準時刻で前記商品またはサービスの利用を開始している顧客の個別データに第1ラベルをラベル付け、前記商品またはサービスの利用を開始していない顧客の個別データに第2ラベルをラベル付けた基準時刻データを生成する基準時刻データ群生成部

50

と、

個々の閾値推定前時刻毎に、閾値推定前時刻で前記商品またはサービスの利用を開始している顧客の個別データに第1ラベルをラベル付け、前記商品またはサービスの利用を開始していない顧客の個別データに第2ラベルをラベル付けたデータである閾値推定前時刻データを生成し、前記閾値推定前時刻データ中の第1ラベルでラベル付けられた個別データ数の割合を前記商品またはサービスの普及率として計算する閾値推定前時刻データ群生成部と、

個々の閾値推定前時刻データ毎に、前記顧客の属性を表す項目から当該顧客の個別データに第1ラベルがラベル付けられる確からしさのスコアを定めるルールである分類器を、閾値推定前時刻データに基づいて生成する閾値推定前時刻分類器群生成部と、

閾値推定前時刻データ毎に生成された個々の分類器毎に、当該分類器と前記基準時刻データ内の各個別データの項目とから、前記基準時刻データ内の各個別データに第1ラベルがラベル付けられる確からしさのスコアを定め、個々の仮閾値について、スコアが仮閾値以上である個別データのラベルが第1ラベルであると予測し、スコアが閾値未満の個別データのラベルが第2ラベルであると予測し、予測結果と基準時刻データとの誤差を算出する誤差群算出部と、

個々の分類器毎に各仮閾値について算出された誤差のうちの最小の誤差を特定し、最小の誤差に対応する仮閾値が1つである場合には、当該仮閾値を分類器に対応する閾値推定前時刻での閾値とし、最小の誤差に対応する仮閾値が複数個存在する場合には前記複数の仮閾値に基づいて分類器に対応する閾値推定前時刻での閾値を定め、各閾値推定前時刻に対応する閾値および普及率から閾値関数を推定する関数推定部とを備える

請求項20から請求項22のうちのいずれか1項に記載の普及予測システム。

【請求項24】

テストデータラベル判定部は、当該普及予測システムの外部から入力された閾値関数を用いて閾値を算出する

請求項20から請求項22のうちのいずれか1項に記載の普及予測システム。

【請求項25】

顧客が商品またはサービスを利用し始めている場合には前記商品またはサービスの利用開始時期を表し、前記商品またはサービスを未利用の場合には未利用である旨を表す利用開始情報と、前記利用開始情報以外の顧客の属性を表す一つ以上の項目とを含む顧客毎の個別データの集合である顧客データを記憶する顧客データベースを用いて、判定対象時刻を定める現時刻で商品またはサービスの利用を開始していない顧客の個別データを含むデータであるテストデータを生成し、

前記現時刻で前記商品またはサービスの利用を開始している顧客の個別データに第1ラベルをラベル付け、前記商品またはサービスの利用を開始していない顧客の個別データに第2ラベルをラベル付け、第1ラベルをラベル付けた個別データ数を、ある期間の中で商品またはサービスの顧客が他の者にその商品またはサービスの利用を喚起させる度合いである影響度に応じて変動させたデータである学習データを生成し、

前記顧客の属性を表す項目から当該顧客の個別データに第1ラベルと第2ラベルのいずれをラベル付けるかを判定するルールである分類器を、前記学習データに基づいて生成し、

前記分類器と前記テストデータ内の各個別データの項目とから、前記テストデータ内の各個別データに対するラベルを判定する

ことを特徴とする普及予測方法。

【請求項26】

第1ラベルがラベル付けられると判定した個別データに対して、利用開始時期を示す情報を付与し、

前記現時刻を一定の時間分増加した時刻を新たな現時刻に更新し、

現時刻を更新した際に、更新後の現時刻におけるテストデータを生成し、

現時刻を更新した際に、更新後の現時刻における学習データを生成し、

10

20

30

40

50

現時刻を更新して新たに学習データを生成した際に、当該学習データに基づいて新たに分類器を生成し、

現時刻を更新して新たにテストデータを生成した際に、前記分類器と前記テストデータ内の各個別データの項目とから、前記テストデータ内の各個別データに対するラベルを判定する

請求項 25 に記載の普及予測方法。

【請求項 27】

顧客が商品またはサービスを利用し始めている場合には前記商品またはサービスの利用開始時期を表し、前記商品またはサービスを未利用の場合には未利用である旨を表す利用開始情報と、前記利用開始情報以外の顧客の属性を表す一つ以上の項目とを含む顧客毎の個別データの集合である顧客データを記憶する顧客データベースを用いて、判定対象時刻を定める現時刻で商品またはサービスの利用を開始していない顧客の個別データを含むデータであるテストデータを生成し、

10

前記現時刻で前記商品またはサービスの利用を開始している顧客の個別データに第 1 ラベルをラベル付け、前記商品またはサービスの利用を開始していない顧客の個別データに第 2 ラベルをラベル付けたラベル付けデータを生成し、前記ラベル付けデータ中の第 1 ラベルでラベル付けられた個別データ数の割合を前記商品またはサービスの普及率として計算し、

前記顧客の属性を表す項目から当該顧客の個別データに第 1 ラベルがラベル付けられる確からしさのスコアを定めるルールである分類器を、前記ラベル付けデータに基づいて生成し、

20

普及率に対するスコアの閾値の関数である閾値関数に計算した普及率を代入して閾値を算出し、前記分類器と前記テストデータ内の各個別データの項目とから、前記テストデータ内の各個別データに第 1 ラベルがラベル付けられる確からしさのスコアを定め、前記テストデータ内の個別データのうちスコアが前記閾値以上である個別データに対するラベルが第 1 ラベルであると判定し、スコアが前記閾値未満の個別データに対するラベルが第 2 ラベルであると判定する

ことを特徴とする普及予測方法。

【請求項 28】

第 1 ラベルがラベル付けられると判定した個別データに対して、利用開始時期を示す情報を付与し、

30

前記現時刻を一定の時間分増加した時刻を新たな現時刻に更新し、

現時刻を更新した際に、更新後の現時刻におけるテストデータを生成し、

現時刻を更新した際に、更新後の現時刻におけるラベル付けデータを生成し、当該ラベル付けデータにおける普及率を計算し、

現時刻を更新して新たにラベル付けデータを生成した際に、当該ラベル付けデータに基づいて新たに分類器を生成し、

現時刻を更新して新たにテストデータを生成した際に、前記分類器と前記テストデータ内の各個別データの項目とから、前記テストデータ内の各個別データに第 1 ラベルがラベル付けられる確からしさのスコアを定め、計算した普及率を代入して新たに閾値を算出し、前記テストデータ内の個別データのうちスコアが前記閾値以上である個別データに対するラベルが第 1 ラベルであると判定し、スコアが前記閾値未満の個別データに対するラベルが第 2 ラベルであると判定する

40

請求項 27 に記載の普及予測方法。

【請求項 29】

顧客が商品またはサービスを利用し始めている場合には前記商品またはサービスの利用開始時期を表し、前記商品またはサービスを未利用の場合には未利用である旨を表す利用開始情報と、前記利用開始情報以外の顧客の属性を表す一つ以上の項目とを含む顧客毎の個別データの集合である顧客データを記憶する顧客データベースを備えるコンピュータに搭載される普及予測プログラムであって、

50

前記コンピュータに、

前記顧客データベースを用いて、判定対象時刻を定める現時刻で商品またはサービスの利用を開始していない顧客の個別データを含むデータであるテストデータを生成するテストデータ生成処理、

前記現時刻で前記商品またはサービスの利用を開始している顧客の個別データに第1ラベルをラベル付け、前記商品またはサービスの利用を開始していない顧客の個別データに第2ラベルをラベル付け、第1ラベルをラベル付けた個別データ数を、ある期間の中で商品またはサービスの顧客が他の者にその商品またはサービスの利用を喚起させる度合いである影響度に応じて変動させたデータである学習データを生成する学習データ生成処理、

前記顧客の属性を表す項目から当該顧客の個別データに第1ラベルと第2ラベルのいずれをラベル付けるかを判定するルールである分類器を、前記学習データに基づいて生成する分類器生成処理、および、

前記分類器と前記テストデータ内の各個別データの項目とから、前記テストデータ内の各個別データに対するラベルを判定するテストデータラベル判定処理

を実行させるための普及予測プログラム。

【請求項30】

前記コンピュータに、

テストデータラベル判定処理で、第1ラベルがラベル付けられると判定した個別データに対して、利用開始時期を示す情報を付与させ、

前記現時刻を一定の時間分増加した時刻を新たな現時刻とする現時刻更新処理を実行させ、

現時刻を更新した際に、テストデータ生成処理で、更新後の現時刻におけるテストデータを生成させ、

現時刻を更新した際に、学習データ生成処理で、更新後の現時刻における学習データを生成させ、

現時刻を更新して新たに学習データを生成した際に、分類器生成処理で、当該学習データに基づいて新たに分類器を生成させ、

現時刻を更新して新たにテストデータを生成した際に、テストデータラベル判定処理で、前記分類器と前記テストデータ内の各個別データの項目とから、前記テストデータ内の各個別データに対するラベルを判定させる

請求項29に記載の普及予測プログラム。

【請求項31】

顧客が商品またはサービスを利用し始めている場合には前記商品またはサービスの利用開始時期を表し、前記商品またはサービスを未利用の場合には未利用である旨を表す利用開始情報と、前記利用開始情報以外の顧客の属性を表す一つ以上の項目とを含む顧客毎の個別データの集合である顧客データを記憶する顧客データベースを備えるコンピュータに搭載される普及予測プログラムであって、

前記コンピュータに、

前記顧客データベースを用いて、判定対象時刻を定める現時刻で商品またはサービスの利用を開始していない顧客の個別データを含むデータであるテストデータを生成するテストデータ生成処理、

前記現時刻で前記商品またはサービスの利用を開始している顧客の個別データに第1ラベルをラベル付け、前記商品またはサービスの利用を開始していない顧客の個別データに第2ラベルをラベル付けたラベル付けデータを生成し、前記ラベル付けデータ中の第1ラベルでラベル付けられた個別データ数の割合を前記商品またはサービスの普及率として計算するラベル付けデータ生成処理、

前記顧客の属性を表す項目から当該顧客の個別データに第1ラベルがラベル付けられる確からしさのスコアを定めるルールである分類器を、前記ラベル付けデータに基づいて生成する分類器生成処理、

普及率に対するスコアの閾値の関数である閾値関数にラベル付けデータ生成処理で計算

10

20

30

40

50

した普及率を代入して閾値を算出し、前記分類器と前記テストデータ内の各個別データの項目とから、前記テストデータ内の各個別データに第1ラベルがラベル付けられる確からしさのスコアを定め、前記テストデータ内の個別データのうちスコアが前記閾値以上である個別データに対するラベルが第1ラベルであると判定し、スコアが前記閾値未満の個別データに対するラベルが第2ラベルであると判定するテストデータラベル判定処理

を実行させるための普及予測プログラム。

【請求項32】

前記コンピュータに、

テストデータラベル判定処理で、第1ラベルがラベル付けられると判定した個別データに対して、利用開始時期を示す情報を付与させ、

10

前記現時刻を一定の時間分増加した時刻を新たな現時刻とする現時刻更新処理を実行させ、

現時刻を更新した際に、テストデータ生成処理で、更新後の現時刻におけるテストデータを生成させ、

現時刻を更新した際に、ラベル付けデータ生成処理で、更新後の現時刻におけるラベル付けデータを生成させ、当該ラベル付けデータにおける普及率を計算させ、

現時刻を更新して新たにラベル付けデータを生成した際に、分類器生成処理で、当該ラベル付けデータに基づいて新たに分類器を生成させ、

現時刻を更新して新たにテストデータを生成した際に、テストデータラベル判定処理で、前記分類器と前記テストデータ内の各個別データの項目とから、前記テストデータ内の各個別データに第1ラベルがラベル付けられる確からしさのスコアを定めさせ、閾値関数にラベル付けデータ生成処理で計算した普及率を代入して新たに閾値を算出させ、前記テストデータ内の個別データのうちスコアが前記閾値以上である個別データに対するラベルが第1ラベルであると判定し、スコアが前記閾値未満の個別データに対するラベルが第2ラベルであると判定させる

20

請求項31に記載の普及予測プログラム。

【請求項33】

ある期間の中で商品またはサービスの顧客が他の者にその商品またはサービスの利用を喚起させる度合いである影響度の推定対象時刻として用いられる現時刻と、前記現時刻の一定時間前を指定するための時刻間隔と、影響度の候補である複数の仮影響度とが入力され、前記現時刻から前記時刻間隔前の時刻である前時刻を計算し、

30

顧客が商品またはサービスを利用し始めている場合には前記商品またはサービスの利用開始時期を表し、前記商品またはサービスを未利用の場合には未利用である旨を表す利用開始情報と、前記利用開始情報以外の顧客の属性を表す一つ以上の項目とを含む顧客毎の個別データの集合である顧客データを記憶する顧客データベースを用いて、前記現時刻で前記商品またはサービスの利用を開始している顧客の個別データに第1ラベルをラベル付け、前記商品またはサービスの利用を開始していない顧客の個別データに第2ラベルをラベル付けた現時刻データを生成し、

個々の仮影響度毎に、前時刻で前記商品またはサービスの利用を開始している顧客の個別データに第1ラベルをラベル付け、前記商品またはサービスの利用を開始していない顧客の個別データに第2ラベルをラベル付け、第1ラベルをラベル付けた個別データ数を前記仮影響度に応じて変動させたデータである前時刻データを生成し、

40

個々の前時刻データ毎に、前記顧客の属性を表す項目から当該顧客の個別データに第1ラベルと第2ラベルのいずれをラベル付けるかを判定するルールである分類器を前記前時刻に基づいて生成し、

前時刻データ毎に生成された個々の分類器毎に、当該分類器と前記現時刻データ内の各個別データの項目とから前記現時刻データ内の個別データにラベル付けられるラベルを予測し、予測結果と前記現時刻データとの誤差を算出し、

個々の分類器毎に算出された誤差のうち、最小の誤差を特定し、最小の誤差に対応する仮影響度が1つである場合には、当該仮影響度を影響度として定め、最小の誤差に対応す

50

る仮影響度が複数個存在する場合には、前記複数の仮影響度に基づいて影響度を定めることを特徴とする影響度推定方法。

【請求項 3 4】

顧客データベースに記憶されている顧客データのうち、前時刻で商品またはサービスの利用を開始していた顧客の個別データを除外した個別データの集合から現時刻データを生成する

請求項 3 3 に記載の影響度推定方法。

【請求項 3 5】

顧客が商品またはサービスを利用し始めている場合には前記商品またはサービスの利用開始時期を表し、前記商品またはサービスを未利用の場合には未利用である旨を表す利用開始情報と、前記利用開始情報以外の顧客の属性を表す一つ以上の項目とを含む顧客毎の個別データの集合である顧客データ記憶する顧客データベースを備えるコンピュータに搭載される影響度推定プログラムであって、

影響度推定対象時刻として用いられる現時刻と、前記現時刻の一定時間前を指定するための時刻間隔と、影響度の候補である複数の仮影響度とが入力され、前記現時刻から前記時刻間隔前の時刻である前時刻を計算し、前記顧客データベースを用いて、前記現時刻で前記商品またはサービスの利用を開始している顧客の個別データに第 1 ラベルをラベル付け、前記商品またはサービスの利用を開始していない顧客の個別データに第 2 ラベルをラベル付けた現時刻データを生成する現時刻データ生成処理、

個々の仮影響度毎に、前時刻で前記商品またはサービスの利用を開始している顧客の個別データに第 1 ラベルをラベル付け、前記商品またはサービスの利用を開始していない顧客の個別データに第 2 ラベルをラベル付け、第 1 ラベルをラベル付けた個別データ数を前記仮影響度に応じて変動させたデータである前時刻データを生成する前時刻データ群生成処理、

個々の前時刻データ毎に、前記顧客の属性を表す項目から当該顧客の個別データに第 1 ラベルと第 2 ラベルのいずれをラベル付けるかを判定するルールである分類器を前記前時刻に基づいて生成する前時刻分類器群生成処理、

前時刻データ毎に生成された個々の分類器毎に、当該分類器と前記現時刻データ内の各個別データの項目とから前記現時刻データ内の個別データにラベル付けられるラベルを予測し、予測結果と前記現時刻データとの誤差を算出する誤差群算出処理、および、

個々の分類器毎に算出された誤差のうち、最小の誤差を特定し、最小の誤差に対応する仮影響度が 1 つである場合には、当該仮影響度を影響度として定め、最小の誤差に対応する仮影響度が複数個存在する場合には、前記複数の仮影響度に基づいて影響度を定める影響度算出処理

を実行させるための影響度推定プログラム。

【請求項 3 6】

前記コンピュータに、

現時刻データ生成処理で、顧客データベースに記憶されている顧客データのうち、前時刻で商品またはサービスの利用を開始していた顧客の個別データを除外した個別データの集合から現時刻データを生成させる

請求項 3 5 に記載の影響度推定プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、商品またはサービスの普及に関する予測を行う普及予測システム、普及予測方法および普及予測プログラムと、ある期間の中で商品またはサービスの顧客が他の者にその商品またはサービスの利用を喚起させる度合いを推定する影響度推定システム、影響度推定方法および影響度推定プログラムに関する。

【背景技術】

【0002】

10

20

30

40

50

顧客に商品またはサービスを提供する場合、商品またはサービスの費用対効果や資源を調達すべき時期や量を推定するため、普及の推移を予測することが必要である。ここで、商品とは、製品などの完成品のみならず、製品の機能を維持するための最小単位も含む。商品やサービスの普及の推移は、時間に対する普及率（または販売数）で示すことができる。

【0003】

このような普及予測に適用可能な予測装置の一例が、特許文献1に記載されている。特許文献1に記載された予測装置は、入力部と、事象内容記憶部と、累積件数集計部と、予測条件記憶部と、回帰分析部と、分布推定部と、出力部とを備えている。入力部は、注目する現象で予測対象となる事象が発生した期日やその発生件数などの過去の履歴、およびその予測対象となる事象において予測すべき事象の上限や下限などの値の範囲や予測時点などに代表される予測条件を入力する入力手段である。回帰分析部は、ロジスティック曲線モデル（あるいは別の曲線モデル）を用いて、実績となる普及の推移に当てはまるように注目する現象を記述するパラメタを推定するとともに、推定したパラメタに基づいて、普及の推移である予測曲線を得る。

10

【0004】

また、特許文献2には、決定木学習用データを生成する時系列データ分類・予測装置が記載されている。特許文献2に記載の時系列データ分類・予測装置は、複数の時刻の情報を含む決定木学習用データを生成する。

20

【0005】

また、特許文献3には、顧客属性データをマスタ・データとする分析装置が記載されている。

【0006】

【特許文献1】特開2004-78780号公報

【特許文献2】特開平6-96052号公報

【特許文献3】特開2001-229150号公報

【発明の開示】

【発明が解決しようとする課題】

【0007】

顧客が商品やサービスを利用し始める時期は、顧客の特徴によって異なる。例えば、顧客は、イノベータ、アーリーアダプタ、アーリーマジョリティ等のカテゴリに分類することができる。

30

【0008】

上記のように、顧客が商品やサービスを利用し始める時期は顧客の特徴によって異なるので、個々の顧客毎に商品やサービスの利用開始時期を予測しようとしても、ロジスティック曲線等の曲線モデルを用いてパラメタフィッティングを行う予測手法では、個々の顧客の特徴が反映されず、顧客毎の商品やサービスの利用開始時期を予測することはできなかった。

【0009】

また、商品やサービスの提供者にとって、既に商品やサービスを利用している者が他の者に商品やサービスの利用を喚起させる度合いを知ることが好ましい。例えば、広告活動や販売促進活動などにより、いわゆる口コミが活発化し、既に商品やサービスの利用を開始している者からの影響によりどの程度、未利用者の利用開始意欲が喚起されたかを知ることができることが好ましい。ここで、ある期間の中で商品またはサービスの顧客が他の者にその商品またはサービスの利用を喚起させる度合いを影響度と呼ぶ。影響度は、ある期間において、商品やサービスを利用していなかった者が既に利用を開始した者からの影響でその商品を受け入れる受容性の程度と言うこともできる。

40

【0010】

そこで、本発明は、顧客が商品やサービスを利用し始める時期を顧客毎に予測することができる普及予測システム、普及予測方法および普及予測プログラムを提供することを目

50

的とする。また、本発明は、時期に応じて商品やサービスの影響度を推定することができる影響度推定システム、影響度推定方法および影響度推定プログラムを提供することを目的とする。

【課題を解決するための手段】

【0011】

本発明の普及予測システムは、顧客が商品またはサービスを利用し始めている場合には前記商品またはサービスの利用開始時期を表し、前記商品またはサービスを未利用の場合には未利用である旨を表す利用開始情報と、前記利用開始情報以外の顧客の属性を表す一つ以上の項目とを含む顧客毎の個別データの集合である顧客データを記憶する顧客データベースと、前記顧客データベースを用いて、判定対象時刻を定める現時刻で商品またはサービスの利用を開始していない顧客の個別データを含むデータであるテストデータを生成するテストデータ生成部と、前記現時刻で前記商品またはサービスの利用を開始している顧客の個別データに第1ラベルをラベル付け、前記商品またはサービスの利用を開始していない顧客の個別データに第2ラベルをラベル付け、第1ラベルをラベル付けた個別データ数を、ある期間の中で商品またはサービスの顧客が他の者にその商品またはサービスの利用を喚起させる度合いである影響度に応じて変動させたデータである学習データを生成する学習データ生成部と、前記顧客の属性を表す項目から当該顧客の個別データに第1ラベルと第2ラベルのいずれをラベル付けるかを判定するルールである分類器を、前記学習データに基づいて生成する分類器生成部と、前記分類器と前記テストデータ内の各個別データの項目とから、前記テストデータ内の各個別データに対するラベルを判定するテストデータラベル判定部とを備えることを特徴とする。

10

20

【0012】

また、本発明の影響度推定システムは、顧客が商品またはサービスを利用し始めている場合には前記商品またはサービスの利用開始時期を表し、前記商品またはサービスを未利用の場合には未利用である旨を表す利用開始情報と、前記利用開始情報以外の顧客の属性を表す一つ以上の項目とを含む顧客毎の個別データの集合である顧客データを記憶する顧客データベースと、ある期間の中で商品またはサービスの顧客が他の者にその商品またはサービスの利用を喚起させる度合いである影響度の推定対象時刻として用いられる現時刻と、前記現時刻の一定時間前を指定するための時刻間隔と、影響度の候補である複数の仮影響度とが入力され、前記現時刻から前記時刻間隔前の時刻である前時刻を計算し、前記顧客データベースを用いて、前記現時刻で前記商品またはサービスの利用を開始している顧客の個別データに第1ラベルをラベル付け、前記商品またはサービスの利用を開始していない顧客の個別データに第2ラベルをラベル付けた現時刻データを生成する現時刻データ生成部と、個々の仮影響度毎に、前時刻で前記商品またはサービスの利用を開始している顧客の個別データに第1ラベルをラベル付け、前記商品またはサービスの利用を開始していない顧客の個別データに第2ラベルをラベル付け、第1ラベルをラベル付けた個別データ数を前記仮影響度に応じて変動させたデータである前時刻データを生成する前時刻データ群生成部と、個々の前時刻データ毎に、前記顧客の属性を表す項目から当該顧客の個別データに第1ラベルと第2ラベルのいずれをラベル付けるかを判定するルールである分類器を前記前時刻に基づいて生成する前時刻分類器群生成部と、前時刻データ毎に生成された個々の分類器毎に、当該分類器と前記現時刻データ内の各個別データの項目とから前記現時刻データ内の個別データにラベル付けられるラベルを予測し、予測結果と前記現時刻データとの誤差を算出する誤差群算出部と、個々の分類器毎に算出された誤差のうち、最小の誤差を特定し、最小の誤差に対応する仮影響度が1つである場合には、当該仮影響度を影響度として定め、最小の誤差に対応する仮影響度が複数個存在する場合には、前記複数の仮影響度に基づいて影響度を定める影響度算出部とを備えることを特徴とする。

30

40

【0013】

また、本発明の普及予測システムは、顧客が商品またはサービスを利用し始めている場合には商品またはサービスの利用開始時期を表し、商品またはサービスを未利用の場合には未利用である旨を表す利用開始情報と、利用開始情報以外の顧客の属性を表す一つ以上

50

の項目とを含む顧客毎の個別データの集合である顧客データを記憶する顧客データベースと、顧客データベースを用いて、判定対象時刻を定める現時刻で商品またはサービスの利用を開始していない顧客の個別データを含むデータであるテストデータを生成するテストデータ生成部と、現時刻で商品またはサービスの利用を開始している顧客の個別データに第1ラベルをラベル付け、商品またはサービスの利用を開始していない顧客の個別データに第2ラベルをラベル付けたラベル付けデータを生成し、ラベル付けデータ中の第1ラベルでラベル付けられた個別データ数の割合を商品またはサービスの普及率として計算するラベル付けデータ生成部と、顧客の属性を表す項目から顧客の個別データに第1ラベルがラベル付けられる確からしさのスコアを定めるルールである分類器を、ラベル付けデータに基づいて生成する分類器生成部と、普及率に対するスコアの閾値の関数である閾値関数にラベル付けデータ生成部が計算した普及率を代入して閾値を算出し、分類器とテストデータ内の各個別データの項目とから、テストデータ内の各個別データに第1ラベルがラベル付けられる確からしさのスコアを定め、テストデータ内の個別データのうちスコアが閾値以上である個別データに対するラベルが第1ラベルであると判定し、スコアが閾値未満の個別データに対するラベルが第2ラベルであると判定するテストデータラベル判定部とを備えることを特徴とする。

10

【0014】

また、本発明の普及予測方法は、顧客が商品またはサービスを利用し始めている場合には前記商品またはサービスの利用開始時期を表し、前記商品またはサービスを未利用の場合には未利用である旨を表す利用開始情報と、前記利用開始情報以外の顧客の属性を表す一つ以上の項目とを含む顧客毎の個別データの集合である顧客データを記憶する顧客データベースを用いて、判定対象時刻を定める現時刻で商品またはサービスの利用を開始していない顧客の個別データを含むデータであるテストデータを生成し、前記現時刻で前記商品またはサービスの利用を開始している顧客の個別データに第1ラベルをラベル付け、前記商品またはサービスの利用を開始していない顧客の個別データに第2ラベルをラベル付け、第1ラベルをラベル付けた個別データ数を、ある期間の中で商品またはサービスの顧客が他の者にその商品またはサービスの利用を喚起させる度合いである影響度に応じて変動させたデータである学習データを生成し、前記顧客の属性を表す項目から当該顧客の個別データに第1ラベルと第2ラベルのいずれをラベル付けるかを判定するルールである分類器を、前記学習データに基づいて生成し、前記分類器と前記テストデータ内の各個別データの項目とから、前記テストデータ内の各個別データに対するラベルを判定することを特徴とする。

20

30

【0015】

また、本発明の普及予測方法は、顧客が商品またはサービスを利用し始めている場合には商品またはサービスの利用開始時期を表し、商品またはサービスを未利用の場合には未利用である旨を表す利用開始情報と、利用開始情報以外の顧客の属性を表す一つ以上の項目とを含む顧客毎の個別データの集合である顧客データを記憶する顧客データベースを用いて、判定対象時刻を定める現時刻で商品またはサービスの利用を開始していない顧客の個別データを含むデータであるテストデータを生成し、現時刻で商品またはサービスの利用を開始している顧客の個別データに第1ラベルをラベル付け、商品またはサービスの利用を開始していない顧客の個別データに第2ラベルをラベル付けたラベル付けデータを生成し、ラベル付けデータ中の第1ラベルでラベル付けられた個別データ数の割合を商品またはサービスの普及率として計算し、顧客の属性を表す項目から顧客の個別データに第1ラベルがラベル付けられる確からしさのスコアを定めるルールである分類器を、ラベル付けデータに基づいて生成し、普及率に対するスコアの閾値の関数である閾値関数に計算した普及率を代入して閾値を算出し、分類器とテストデータ内の各個別データの項目とから、テストデータ内の各個別データに第1ラベルがラベル付けられる確からしさのスコアを定め、テストデータ内の個別データのうちスコアが閾値以上である個別データに対するラベルが第1ラベルであると判定し、スコアが閾値未満の個別データに対するラベルが第2ラベルであると判定することを特徴とする。

40

50

【 0 0 1 6 】

また、本発明の普及予測プログラムは、顧客が商品またはサービスを利用し始めている場合には前記商品またはサービスの利用開始時期を表し、前記商品またはサービスを未利用の場合には未利用である旨を表す利用開始情報と、前記利用開始情報以外の顧客の属性を表す一つ以上の項目とを含む顧客毎の個別データの集合である顧客データを記憶する顧客データベースを備えるコンピュータに搭載される普及予測プログラムであって、前記コンピュータに、前記顧客データベースを用いて、判定対象時刻を定める現時刻で商品またはサービスの利用を開始していない顧客の個別データを含むデータであるテストデータを生成するテストデータ生成処理、前記現時刻で前記商品またはサービスの利用を開始している顧客の個別データに第1ラベルをラベル付け、前記商品またはサービスの利用を開始していない顧客の個別データに第2ラベルをラベル付け、第1ラベルをラベル付けた個別データ数を、ある期間の中で商品またはサービスの顧客が他の者にその商品またはサービスの利用を喚起させる度合いである影響度に応じて変動させたデータである学習データを生成する学習データ生成処理、前記顧客の属性を表す項目から当該顧客の個別データに第1ラベルと第2ラベルのいずれをラベル付けるかを判定するルールである分類器を、前記学習データに基づいて生成する分類器生成処理、および、前記分類器と前記テストデータ内の各個別データの項目とから、前記テストデータ内の各個別データに対するラベルを判定するテストデータラベル判定処理を実行させることを特徴とする。

10

【 0 0 1 7 】

また、本発明の普及予測プログラムは、顧客が商品またはサービスを利用し始めている場合には前記商品またはサービスの利用開始時期を表し、前記商品またはサービスを未利用の場合には未利用である旨を表す利用開始情報と、前記利用開始情報以外の顧客の属性を表す一つ以上の項目とを含む顧客毎の個別データの集合である顧客データを記憶する顧客データベースを備えるコンピュータに搭載される普及予測プログラムであって、前記コンピュータに、前記顧客データベースを用いて、判定対象時刻を定める現時刻で商品またはサービスの利用を開始していない顧客の個別データを含むデータであるテストデータを生成するテストデータ生成処理、前記現時刻で前記商品またはサービスの利用を開始している顧客の個別データに第1ラベルをラベル付け、前記商品またはサービスの利用を開始していない顧客の個別データに第2ラベルをラベル付けたラベル付けデータを生成し、前記ラベル付けデータ中の第1ラベルでラベル付けられた個別データ数の割合を前記商品またはサービスの普及率として計算するラベル付けデータ生成処理、前記顧客の属性を表す項目から当該顧客の個別データに第1ラベルがラベル付けられる確からしさのスコアを定めるルールである分類器を、前記ラベル付けデータに基づいて生成する分類器生成処理、普及率に対するスコアの閾値の関数である閾値関数にラベル付けデータ生成処理で計算した普及率を代入して閾値を算出し、前記分類器と前記テストデータ内の各個別データの項目とから、前記テストデータ内の各個別データに第1ラベルがラベル付けられる確からしさのスコアを定め、前記テストデータ内の個別データのうちスコアが前記閾値以上である個別データに対するラベルが第1ラベルであると判定し、スコアが前記閾値未満の個別データに対するラベルが第2ラベルであると判定するテストデータラベル判定処理を実行させることを特徴とする。

20

30

40

【 0 0 1 8 】

また、本発明の影響度推定方法は、ある期間の中で商品またはサービスの顧客が他の者にその商品またはサービスの利用を喚起させる度合いである影響度の推定対象時刻として用いられる現時刻と、前記現時刻の一定時間前を指定するための時刻間隔と、影響度の候補である複数の仮影響度とが入力され、前記現時刻から前記時刻間隔前の時刻である前時刻を計算し、顧客が商品またはサービスを利用し始めている場合には前記商品またはサービスの利用開始時期を表し、前記商品またはサービスを未利用の場合には未利用である旨を表す利用開始情報と、前記利用開始情報以外の顧客の属性を表す一つ以上の項目とを含む顧客毎の個別データの集合である顧客データを記憶する顧客データベースを用いて、前記現時刻で前記商品またはサービスの利用を開始している顧客の個別データに第1ラベル

50

をラベル付け、前記商品またはサービスの利用を開始していない顧客の個別データに第2ラベルをラベル付けた現時刻データを生成し、個々の仮影響度毎に、前時刻で前記商品またはサービスの利用を開始している顧客の個別データに第1ラベルをラベル付け、前記商品またはサービスの利用を開始していない顧客の個別データに第2ラベルをラベル付け、第1ラベルをラベル付けた個別データ数を前記仮影響度に応じて変動させたデータである前時刻データを生成し、個々の前時刻データ毎に、前記顧客の属性を表す項目から当該顧客の個別データに第1ラベルと第2ラベルのいずれをラベル付けるかを判定するルールである分類器を前記前時刻に基づいて生成し、前時刻データ毎に生成された個々の分類器毎に、当該分類器と前記現時刻データ内の各個別データの項目とから前記現時刻データ内の個別データにラベル付けられるラベルを予測し、予測結果と前記現時刻データとの誤差を算出し、個々の分類器毎に算出された誤差のうち、最小の誤差を特定し、最小の誤差に対応する仮影響度が1つである場合には、当該仮影響度を影響度として定め、最小の誤差に対応する仮影響度が複数個存在する場合には、前記複数の仮影響度に基づいて影響度を定めることを特徴とする。

10

【0019】

また、本発明の影響度推定プログラムは、顧客が商品またはサービスを利用し始めている場合には前記商品またはサービスの利用開始時期を表し、前記商品またはサービスを未利用の場合には未利用である旨を表す利用開始情報と、前記利用開始情報以外の顧客の属性を表す一つ以上の項目とを含む顧客毎の個別データの集合である顧客データ記憶する顧客データベースを備えるコンピュータに搭載される影響度推定プログラムであって、影響度推定対象時刻として用いられる現時刻と、前記現時刻の一定時間前を指定するための時刻間隔と、影響度の候補である複数の仮影響度とが入力され、前記現時刻から前記時刻間隔前の時刻である前時刻を計算し、前記顧客データベースを用いて、前記現時刻で前記商品またはサービスの利用を開始している顧客の個別データに第1ラベルをラベル付け、前記商品またはサービスの利用を開始していない顧客の個別データに第2ラベルをラベル付けた現時刻データを生成する現時刻データ生成処理、個々の仮影響度毎に、前時刻で前記商品またはサービスの利用を開始している顧客の個別データに第1ラベルをラベル付け、前記商品またはサービスの利用を開始していない顧客の個別データに第2ラベルをラベル付け、第1ラベルをラベル付けた個別データ数を前記仮影響度に応じて変動させたデータである前時刻データを生成する前時刻データ群生成処理、個々の前時刻データ毎に、前記顧客の属性を表す項目から当該顧客の個別データに第1ラベルと第2ラベルのいずれをラベル付けるかを判定するルールである分類器を前記前時刻に基づいて生成する前時刻分類器群生成処理、前時刻データ毎に生成された個々の分類器毎に、当該分類器と前記現時刻データ内の各個別データの項目とから前記現時刻データ内の個別データにラベル付けられるラベルを予測し、予測結果と前記現時刻データとの誤差を算出する誤差群算出処理、および、個々の分類器毎に算出された誤差のうち、最小の誤差を特定し、最小の誤差に対応する仮影響度が1つである場合には、当該仮影響度を影響度として定め、最小の誤差に対応する仮影響度が複数個存在する場合には、前記複数の仮影響度に基づいて影響度を定める影響度算出処理を実行させることを特徴とする。

20

30

【発明の効果】

40

【0020】

本発明によれば、顧客が商品やサービスを利用し始める時期を顧客毎に予測することができる。また、本発明によれば、時期に応じて商品やサービスの影響度を推定することができる。

【発明を実施するための最良の形態】**【0021】**

以下、本発明の実施形態を図面を参照して説明する。

【0022】

なお、有料で商品やサービスを利用する者だけでなく、無料で商品やサービスを利用する者も顧客と呼ぶ。

50

【 0 0 2 3 】

また、以下に説明する影響度推定システムおよび普及予測システムは、一つの商品またはサービスに関して影響度推定や普及予測を行う。

【 0 0 2 4 】

実施形態 1 .

図 1 は、本発明の影響度推定システムの例を示すブロック図である。本発明の影響度推定システムは、影響度（ある期間の中で商品またはサービスの顧客が他の者にその商品またはサービスの利用を喚起させる度合い）を推定する影響度推定装置 3 と、顧客データベース（以下、顧客 DB と記す）2 とを備える。製品などの完成品だけでなく、製品の機能を維持するための最小単位も商品の概念に含まれる。

10

【 0 0 2 5 】

顧客 DB 2 は、顧客データを記憶する記憶手段であり、顧客データ記憶手段と呼ぶことができる。顧客データは、個々の顧客毎に定められたデータの集合であり、以下、個々の顧客毎のデータを個別データと呼ぶ。顧客 DB 2 において、各顧客毎の個別データは、利用開始情報と、利用開始情報以外の顧客の属性を表す一つ以上の項目とを含む。利用開始情報とは、顧客が商品またはサービスを利用し始めている場合には、その商品またはサービスを利用し始めた時期を表し、顧客が商品またはサービスを未利用である（すなわち、まだ利用開始していない）場合には未利用である旨を表す情報である。利用開始情報以外の顧客の属性を表す項目として、例えば、性別、年齢、住所の地域区分、勤務地の地域区分、他の商品の利用履歴、様々な嗜好性などの特徴を用いることができるが、これらに限定されるわけではない。個々の個別情報は、これらの項目を一つ以上含む。顧客の属性を表す項目は、例えば、店頭での店員による入力、商品を利用開始時点での登録、顧客会員カード、アンケートなど、様々な方法で収集することができる。

20

【 0 0 2 6 】

影響度推定装置 3 は、現時刻データ生成部 3 1 と、前時刻データ群生成部 3 2 と、前時刻分類器群生成部 3 3 と、誤差群算出部 3 4 と、影響度算出部 3 5 とを備える。

【 0 0 2 7 】

現時刻データ群生成部 3 1 は、現時刻と、その現時刻の一定時間前を指定するための時刻間隔と、複数の仮影響度が入力される。後述するように、キーボード等の入力装置を介して、現時刻データ群生成部 3 1 に、現時刻、時刻間隔、および複数の仮影響度が入力されてもよい。本実施形態における現時刻とは、影響度の推定対象時刻（影響度推定対象時刻）として用いられる時刻であり、時刻間隔は、影響度を推定しようとする期間である。そして、本実施形態では、過去のある時点に影響度推定対象時刻として、その時点から時刻間隔分遡った時刻からその影響度推定対象時刻までの期間における影響度を推定する。従って、過去のある時点での時刻が現時刻として入力される。現時刻から時刻間隔前の時刻（現時刻から時刻間隔分遡った時刻）を、前時刻と記す。現時刻データ群生成部 3 1 は、現時刻から時刻間隔を減算して前時刻を計算する。

30

【 0 0 2 8 】

また、入力される仮影響度は、求めようとしている影響度の候補であり、影響度推定装置 3 は、複数の仮影響度から影響度を求める。仮影響度として現時刻データ群生成部 3 1 に入力される数値は、0 以上の数値である。例えば、影響度推定システムのユーザが、0 以上から 5 0 程度までの値の中から影響度の候補を複数選択し、仮影響度として入力する。候補として仮影響度を多く入力すれば、影響度の推定精度が高くなる。

40

【 0 0 2 9 】

また、現時刻データ群生成部 3 1 は、現時刻と顧客 DB 2 とに基づいて、1 つの現時刻データを生成する。現時刻データは、現時刻で商品またはサービスの利用を開始している顧客の個別データに第 1 ラベルをラベル付け、現時刻で商品またはサービスの利用を開始していない顧客の個別データに第 2 ラベルをラベル付けたデータである。第 1 ラベルは、商品またはサービスの利用を開始している状態を示す情報であり、第 2 ラベルは、商品またはサービスの利用を開始していない状態を示す情報である。以下、第 1 ラベルを正（ま

50

たは+)と記し、第2ラベルを負(または-)と記す。現時刻データ群生成部31は、顧客DB2を用いて(すなわち、顧客データDB2に記憶された顧客データに基づいて)、現時刻データを生成する。

【0030】

前時刻データ群生成部32は、前時刻と複数の仮影響度と顧客DB2に記憶されている顧客データとに基づいて、複数の前時刻データを生成する。前時刻データは、前時刻で商品またはサービスの利用を開始している顧客の個別データに第1ラベル(正)をラベル付け、前時刻で商品またはサービスの利用を開始していない顧客の個別データに第2ラベル(負)をラベル付け、仮影響度に応じて重み付けを行ったデータである。重み付けとは、第2ラベルをラベル付けた個別データ数に対して、第1ラベルをラベル付けた個別データ数を相対的に変動させることである。前時刻データ群生成部32は、仮影響度毎に前時刻データを生成する。仮影響度は複数あるので、複数の前時刻データを生成することになる。

10

【0031】

前時刻分類器群生成部33は、前時刻データ群生成部32によって生成された各前時刻データ毎に1つずつ分類器を生成する。従って、前時刻分類器群生成部33は、複数の分類器を生成することになる。分類器は、顧客の属性を表す項目からその顧客の個別データに正(第1ラベル)と負(第2ラベル)のいずれをラベル付けるかを判定するルールである。換言すれば、顧客の属性を表す項目を独立変数として、従属変数のとり得る値を正または負とし、その独立変数から従属変数を定めるルールである。

20

【0032】

誤差群算出部34は、前時刻分類器群生成部33で生成された各分類器を用いて、現時刻データのラベルを予測する。すなわち、誤差群算出部34は、分類器と、現時刻データ内の各個別データの項目(顧客の属性を示す項目)と照合し、現時刻データ内の各個別データ毎に、その個別データにラベル付けられるラベルを予測する。また、その現時刻データ内の各個別データには既に実際にラベル付けが行われている。誤差群算出部34は、分類器と項目とから予測したラベルと、実際に現時刻データでラベル付けられているラベルとの誤差を算出する。誤差群算出部34は、この処理を分類器毎に行う。したがって、誤差群算出部34は複数の誤差を算出する。

【0033】

影響度算出部35は、個々の分類器毎に算出された誤差のうちの最小の誤差を特定し、最小の誤差に対応する仮影響度を影響度として定める。具体的には、最小の誤差に対応する仮影響度の数が一つである場合、影響度算出部35は、その仮影響度を影響度として定める。また、最小の誤差に対応する仮影響度が複数個存在する場合には、その複数の仮影響度に基づいて影響度を定める。例えば、最小の誤差に対応する仮影響度が複数個存在する場合、その複数の仮影響度の平均値を計算し、その平均値を影響度として定める。以下、最小の誤差に対応する仮影響度が複数個存在する場合、その仮影響度の平均値を影響度と定める場合を例にして説明する。

30

【0034】

影響度推定装置3が備える現時刻データ生成部31、前時刻データ群生成部32、前時刻分類器群生成部33、誤差群算出部34、および影響度算出部35は、例えば、影響度推定プログラムに従って動作するCPUによって実現される。すなわち、影響度推定システムに設けられた記憶装置から影響度推定プログラムを読み込んだCPUが現時刻データ生成部31、前時刻データ群生成部32、前時刻分類器群生成部33、誤差群算出部34、および影響度算出部35として動作してもよい。

40

【0035】

なお、キーボード等の入力装置を介して、現時刻データ群生成部31に現時刻、時刻間隔、および複数の仮影響度が入力されてもよい。また、影響度算出部35が求めた影響度を出力するための出力装置が設けられていてもよい。図2は、入力装置および出力装置を備えた影響度推定システムの例を示すブロック図である。図1に示す構成要素と同様の構

50

成要素については、図 1 と同一の符号を付して説明を省略する。

【 0 0 3 6 】

入力装置 1 は、例えばキーボードなどの入力装置である。このような入力装置 1 を介して現時刻、時刻間隔、および複数の仮影響度が入力されてもよい。ただし、他の態様で現時刻、時刻間隔、および複数の仮影響度が現時刻データ群生成部 3 1 に入力されてもよい。

【 0 0 3 7 】

また、出力装置 4 は、例えばディスプレイ装置などの出力装置である。影響度算出部 3 5 は、求めた影響度を出力装置 4 に出力してもよい。例えば、出力装置 4 がディスプレイ装置である場合、出力装置 4 に影響度を表示させてもよい。また、出力装置 4 が印刷装置であってよい。

10

【 0 0 3 8 】

次に、動作について説明する。

図 3 は、本発明の影響度推定システムの処理経過の例を示すフローチャートである。例えば入力装置 1 (図 2 参照) を介して、現時刻データ生成部 3 1 に現時刻、時刻間隔、および複数の仮影響度が入力されると、影響度推定システムは以下のように動作する。

【 0 0 3 9 】

まず、現時刻データ生成部 3 1 は、前時刻を計算し、現時刻データを作成する (ステップ A 1)。さらに、ステップ A 1 において、現時刻データ生成部 3 1 は、誤差群算出部 3 4 が計算する誤差のうちの最小値を示す変数 (Error Min とする。) の初期値を設定し、変数 i , p に対して、 $i = 1$, $p = 0$ という初期値を設定する。Error Min の初期値は、誤差群算出部 3 4 が計算する誤差のとり得る値の最大値、あるいは、その誤差のとり得る値に比べて十分に大きな値であればよい。変数 i は、入力された複数の仮影響度を順番に指定するための変数である。例えば、 $i = 1$ であれば、1 番目の仮影響度を指定していることを意味する。また、 p は、最小の誤差に対応する仮影響度を指定するための変数である。例えば、最小の誤差に対応する仮影響度が p 個あったとすると、その最小誤差に対応する仮影響度の 1 番目から p 番目までをそれぞれ、 $Imp [1]$, \dots , $Imp [p]$ とする。

20

【 0 0 4 0 】

ステップ A 1 において、現時刻データ生成部 3 1 は、前時刻を、現時刻 - 時刻間隔と設定する。すなわち、現時刻から時刻間隔を減算した時刻を前時刻とする。

30

【 0 0 4 1 】

また、ステップ A 1 において、現時刻データ生成部 3 1 は、顧客 DB 2 に記憶された顧客データを読み込む。そして、現時刻で商品またはサービスの利用を開始している顧客の個別データ (利用開始情報が現時刻以前の時刻を表している個別データ) に正をラベル付け、現時刻で商品またはサービスの利用を開始していない顧客の個別データに負をラベル付けることによって、現時刻データを生成する。

【 0 0 4 2 】

このとき、現時刻データ生成部 3 1 は、顧客 DB 2 に記憶されている顧客データに属する個別データのうち、前時刻で商品またはサービスの利用を開始していた顧客の個別データを除外した個別データの集合から現時刻データを生成してもよい。すなわち、利用開始情報が前時刻以前の時刻となっている個別データを除外し、残りの個別データに対して正または負のラベル付けを行ってもよい。以下、このように、利用開始情報が前時刻以前の時刻となっている個別データが現時刻データに含まれないように除外する場合を例にして説明するが、現時刻データ生成部 3 1 は、顧客データに属する各個別データに対して正または負のラベル付けを行ってもよい。

40

【 0 0 4 3 】

顧客データから現時刻データを生成する動作を具体的に示す。図 4 は、顧客 DB 2 される顧客データの例を示す説明図である。図 4 に示す各行はそれぞれ一人の顧客の個別データとなっていて、その個別データの集合が顧客データである。図 4 示す例では、個々の個

50

別データの識別情報 (ID) が付与されている場合を示している。図4では、顧客が商品またはサービスの利用を開始した月を利用開始時期としている場合を例示しているが、利用開始時期を日単位や時間単位などで表してもよい。また、利用開始情報が“?”で表されている場合、その“?”は、「未利用(すなわち、利用開始していない状態)」を意味しているが、他の記号で「未利用」を表してもよい。また、図4では、それぞれの個別データが、利用開始情報以外の顧客の属性を表すN個の項目を含んでいる場合を例示している。本例では、項目1が性別である。各項目に関し、不明である場合は欠損値を示す記号で表せばよい。本例では、欠損値を表す記号として“?”を用いている。

【0044】

現時刻が2007年4月であり、計算した前時刻が2007年2月であるとする。現時刻データ生成部31は、各項目の項目値はそのままとして、利用開始情報を無くし、代わりに利用を開始しているか否かを示すクラスを作成し、個別データに正または負をラベル付ける。なお、後の分類器を作成するステップで必要となるならば、男性/女性を0, 1等の数値に変換してもよい。また、現時刻データ生成部31は、図4に示す顧客データ内の個別データのうち、利用開始情報が前時刻(2007年2月)以前の時刻となっている個別データ(ID=1の個別データ)を除外し、他の個別データに対して、現時刻(2007年4月)で商品またはサービスを利用開始しているならば正をラベル付け、そうでなければ負をラベル付ける。例えば、ID=2の個別データでは、2007年3月から利用開始となっているので、正とする。また、ID=3の個別データでは、2007年5月から利用開始となっていて、2007年4月では商品またはサービスの利用を開始していないので、負とする。また、利用開始情報が“(未利用)”となっている個別データについても負とする。図5は、このようにして図2に示す顧客データから生成された現時刻データの例を示す説明図である。前時刻以前の2007年1月から利用を開始していた顧客の顧客データ(ID=1)は、現時刻データでは除かれ、2007年4月以前に利用開始した顧客の顧客データ(ID=2)は正(+)とラベル付けられ、他の顧客データは負(-)とラベル付けられている。なお、図5では、分かりやすさのためにIDを残した現時刻データを示しているが、分類に無関係な項目は予め削除してもよい。

【0045】

ステップA1の後、前時刻データ群生成部32は、i番目の仮影響度(以下、False[i])と記す。)を用いて、仮影響度False[i]に対応する前時刻データを生成する(ステップA2)。前時刻データ群生成部32は、顧客DB2に記憶された顧客データを読み込み、前時刻以前で商品またはサービスの利用を開始している顧客の個別データに正をラベル付け、前時刻で商品またはサービスの利用を開始していない顧客の個別データに負をラベル付ける。すなわち、利用開始情報(図4参照)が前時刻以前の時刻となっている個別データに正をラベル付け、利用開始情報が前時刻以前の時刻でない個別データに負をラベル付ける。この時点では、また仮影響度[i]に応じた重み付けは行っていない。図6は、前時刻データ生成過程において重み付けが行われる前のデータの例を示す説明図である。ID=1の個別データでは2007年1月から利用開始となっていて、前時刻(2007年2月)では利用が開始されているので、正(+)をラベル付けている。また、ID=2以降の個別データでは、利用開始の時期が前時刻(2007年2月)よりも後であり、前時刻では利用開始されていなかったため、ID=2以降の個別データにはいずれも負(-)をラベル付ける(図4、図6参照)。なお、前時刻データを生成する際には、個別データを除外しない。

【0046】

そして、前時刻データ群生成部32は、仮影響度False[i]に応じた重み付けを行う。すなわち、仮影響度False[i]に応じて、正とした個別データの数を、負とした個別データの数に対して相対的に変動させる。本例では、仮影響度False[i]の関数の関数値を計算し、正をラベル付けた個別データをその関数値倍に増加させる場合を例にする。

【0047】

10

20

30

40

50

仮影響度 $F a l s e [i]$ の関数として、例えば、前時刻で商品またはサービスの利用を開始している顧客数と前時刻で未だその商品またはサービスの利用を開始していない顧客数との比率を仮影響度の係数とする関数を用いてもよい。すなわち、以下に例示する関数を用いてもよい。

【 0 0 4 8 】

$F a l s e [i] \times (\text{前時刻の未利用者数}) / (\text{前時刻の利用者数})$

【 0 0 4 9 】

上記の関数において、「前時刻の未利用者数」は、前時刻で未だ商品またはサービスの利用を開始していない顧客の数である。すなわち、顧客データにおいて、利用開始情報が前時刻以前ではない個別データの数である。また、「前時刻の利用者数」は、前時刻で商品またはサービスの利用を開始している顧客の数である。すなわち、顧客データにおいて、利用開始情報が前時刻以前となっている個別データの数である。このように、「前時刻の未利用者数」を「前時刻の利用者数」で除算して得た比率を $F a l s e [i]$ の係数とする関数を用いてもよい。

【 0 0 5 0 】

また、仮影響度 $F a l s e [i]$ の関数として、例えば、前時刻で商品またはサービスの利用を開始している顧客数と全顧客数との比率を仮影響度の係数とする関数を用いてもよい。すなわち、以下に例示する関数を用いてもよい。

【 0 0 5 1 】

$F a l s e [i] \times (\text{顧客人数}) / (\text{前時刻の利用者数})$

【 0 0 5 2 】

上記の関数において、「顧客人数」は、全顧客数（顧客データ内の全個別データ数）であり、前時刻で利用を開始している顧客の数と、前時刻で未だ利用を開始していない顧客の数の和である。「前時刻の利用者数」は、上述の関数の場合と同様である。このように、「顧客人数」を「前時刻の利用者数」で除算して得た比率を $F a l s e [i]$ の係数とする関数を用いてもよい。

【 0 0 5 3 】

前時刻データ群生成部 3 2 は、このような関数の関数値倍に、正をラベル付けた個別データを増加させる。例えば、仮影響度が $F a l s e [i]$ のときの関数値が 5 であるとする。この場合、正をラベル付けた個別データの数が 5 倍になるように、正をラベル付けた個別データの複製を生成すればよい。この結果得られるデータが前時刻データである。図 7 は、重み付けを行って得た前時刻データの例を示す説明図である。図 6 に示す例では、 $I D = 1$ の個別データが正となっていて、他の個別データが負となっている。従って、前時刻データ群生成部 3 2 は、 $I D = 1$ の個別データと同内容のデータを 5 つ生成する。ただし、 $I D$ はそれぞれ $1'$, $1''$ 等として、区別している。

【 0 0 5 4 】

なお、正とラベル付けた個別データのコピーだけを生成する場合には、正とラベル付けた個別データに対する重み付けを行ってから、他の個別データに対して負のラベル付けを行ってもよい。

【 0 0 5 5 】

また、例示した 2 つの関数における仮影響度の係数の値は、いずれも時間が経過して利用者数が増えるに従って減少する。仮影響度の係数は、このような係数に限定されない。また、仮影響度 $F a l s e [i]$ の値をそのまま関数値として用いてもよい。この場合係数を 1 としていることになる。

【 0 0 5 6 】

また、上記の例では、正とラベル付けた個別データを関数値倍に増加させる場合を示しているが、重み付けでは、正とラベル付けた個別データ数を、負とラベル付けた個別データ数に対して相対的に変化させればよい。例えば、重み（関数値）が 5 で、正とラベル付けた個別データ数を 10 倍とするならば、負とラベル付けた個別データ数（未利用者数）を 2 倍にすればよい。また、重みが 0 . 5 で、正とラベル付けた個別データ数を 1 倍とす

10

20

30

40

50

るならば、負とラベル付けた個別データ数を2倍にすればよい。なお、ここでは、個別データ数を10倍とする等の例を挙げたが、このように、個別データ数をk倍にする等の条件を定める場合、そのk等は予め外部から入力しておけばよい。

【0057】

また、ここでは、実際に個別データの複製を作成することで個別データを増加させていたが、次のステップで分類器を作成する際に、仮影響度 $F a l s e [i]$ の関数値を重み付けパラメータとして用いてもよい。すなわち、正とラベル付けた個別データを何倍に増やすかということを決めておき、その結果をパラメータとして分類器を生成してもよい。

【0058】

ステップA2で前時刻データ群生成部32が $F a l s e [i]$ に応じた前時刻データを生成した後、前時刻分類器群生成部33は、その前時刻データを用いて、その前時刻データに対応する分類器（前時刻分類器 i と記す。）を生成する（ステップA3）。IDは分類器作成に不要であり、前時刻分類器群生成部33は、IDを削除してよい。前時刻分類器群生成部33は、データマイニングの様々な手法により分類器を生成すればよい。例えば、重回帰分析、決定木、ニューラルネットワーク、サポートベクタマシン、ベイジアンネットワークなどの手法のいずれかあるいはその組合せなどにより分類器を生成すればよい。

【0059】

以下、分類器として決定木を用いる場合を例にして、分類器の具体例を説明する。本例では、「ゴルフをするか否か」を従属変数とし、「天気」、「天候」、「湿度」、「風が強い
20
強い
か否か」を独立変数とする場合を例にして説明する。図8は、既知の独立変数および従属変数の組み合わせの例を示す説明図である。図9は、分類器の例を示し、ここでは、決定木を分類器としている。図10は、分類器から予測される従属変数の例を示す説明図である。図8に示すように、既知の独立変数および従属変数の組み合わせがあれば、その既知の独立変数および従属変数から決定木を生成することができる。そして、その決定木と既知の独立変数から従属変数を予測することができる。例えば、図9に示す決定木が得られていれば、図10に例示する独立変数から、図10に例示する従属変数を予測することができる。項目の項目値に応じて枝分かれさせることを分割という。

【0060】

図8に示す例では、各行が、「ゴルフをする／しない」とラベル付けられた個別データ
30
に相当する。以下、図8に示す各行を個別データとして説明する。また、「ゴルフをする」は、正（+）のラベルに相当し、「ゴルフをしない」は、負（-）のラベルに相当する。決定木では、従属変数のとり得るラベル毎（例えば、「する（正）」、「しない（負）」というラベル毎）に個別データ数をまとめた情報をノードとする。例えば、図8に示すルートのノードでは、「する：9、しない：5」という情報をノードとしている。

【0061】

前時刻分類器群生成部33は、前時刻データが与えられると、どの項目で最初に分割させるかを決定する。このとき、前時刻分類器群生成部33は、項目1～Nの各項目について、分割時の評価値を計算し、その評価値が最大の項目を、分割に最も適した項目として
40
選択する。ここでは、分割前のノードのエントロピーと、分割後のエントロピーの差を評価値とする場合を例にするが、他の計算方法で評価値を求めてもよい。ノードのエントロピーは、ラベルが正（+）の個別データの割合を q とし、ラベルが負（-）の個別データの割合を $1 - q$ とすると、 $-q \log q - (1 - q) \log (1 - q)$ で表される。分割後のノードのエントロピーは、分割後の各ノードのエントロピーの加重平均である。

【0062】

例えば、前時刻データにおいて、「正」が9データあり、「負」が5データあるとするとルートのノードは「正：9、負：5」となる。この場合、正（+）のデータが9データあり、負（-）のデータが5データあるので、ルートのノードのエントロピーは、 $-(9/14) \times \log (9/14) - (5/14) \times (5/14) = 0.940$ となる。

【0063】

10

20

30

40

50

前時刻分類器群生成部 33 は、一つの項目でルートのノードを分割して得られるノードを求める。すなわち、その項目の項目値毎に、正および負の個別データ数を表す情報（ノード）を生成する。例えば、その項目 1 のとり得る値が「0」または「1」であり、項目 1 の値が「0」のときには、正が 5 データあり、負が 2 データあるとし、項目 1 の値が「1」のときには、正が 0 データあり、負が 7 データあるとする。この場合、項目 1 の値が「0」か「1」かで分岐するノードとして、「正：5，負：2」というノードと、「正：0，負：7」というノードとを生成する。前時刻分類器群生成部 33 は、分割後の各ノードのエントロピーを計算し、分割後の各ノードにおける正または負としてカウントされる個別データ数に応じて各ノードのエントロピーの加重平均を求める。上記の例では「正：5，負：2」というノードにおいても、「正：0，負：7」というノードにおいても個別データの総数は 7 であるので、加重平均を行う際の重み付け係数は各ノードでいずれも（ $7/14$ ）となる。従って、本例の場合、前時刻分類器群生成部 33 は、分割後のエントロピーを以下のように計算する。

【0064】

$$(7/14) \times \{ -(5/7) \times \log(5/7) - (2/7) \times \log(2/7) \} + (7/14) \times \{ -(0/7) \times \log(0/7) - (7/7) \times \log(7/7) \} = 0.432$$

【0065】

従って、本例の場合、前時刻分類器群生成部 33 は、項目 1 で分割した場合の評価値を、 $0.940 - 0.432 = 0.508$ と計算する。

【0066】

前時刻分類器群生成部 33 は、項目 1 だけでなく、他の項目についても同様に、その項目で分割したときの評価値を計算し、評価値が最大となる項目で分割すると決定する。このようにして、ルートのノードを分割する項目を決定する。

【0067】

なお、上記の項目 1 の例では、項目 1 のとり得る値が「0」または「1」の二つだけである場合を示した。項目値が年齢であり、その値が 20，21，22 のように連続する値の場合には、どこで項目値で分割させるのかも決める。この場合、前時刻分類器群生成部 33 は、各項目値間の中間値をしきい値とし、各しきい値毎に、その「しきい値以下」および「そのしきい値より大」とに分割させた場合の評価値を求める。そして、評価値が最大となる場合を選択することによって、どの項目値で分割させるのかも決定する。例えば、項目値が 20，21，22，・・・と連続する場合は、「20.5 以下」および「20.5 より大」で分割した場合の評価値、「21.5 以下」および「21.5 より大」で分割した場合の評価値等をそれぞれ計算し、評価値が最も高くなるように分割すればよい。

【0068】

前時刻分類器群生成部 33 は、分割後の各ノードについても、上記と同様の処理を行い、次にどの項目で分割するのかを決定する処理を順次、繰り返す。また、前時刻分類器群生成部 33 は、所定の条件が満たされたときには、ノードの分割を停止する。所定の条件とは、例えば、「ノードにおける個別データのクラスが全て同じになる」という条件や、「ノードにおける正または負としてカウントされる個別データ数が所定数（例えば 2）以下になる」という条件を用いてよい。前者の条件を採用すると、ノードにおける個別データが全て正または負になると、そのノードの分割を継続しない。このように、前時刻分類器群生成部 33 は、ルートのノードから順次、分割を繰り返し、木構造の決定木を生成する。

【0069】

また、前時刻分類器群生成部 33 は、上記のように、木構造の決定木を生成した後、その決定木に対する枝刈りを行う。決定木において、分割されて生成された最終的なノードを葉と呼ぶ。ある葉に分類されたデータ数が N であるとする（すなわち、正または負としてカウントされる個別データ数が N であるとする）。この葉に分類された N データ中、E

10

20

30

40

50

データが誤りであるとする。この仮定では、N回の試行中、誤りという事象をE回観測したとみなし、大きさNの標本で、誤りという事象が起きる確率がrである二項分布と考えることができる。予め与えられた信頼度CFに対して、rの上限を $U_{CF}(E, N)$ と表すことにすると、Nデータでの誤りの発生する期待値は、 $N \times U_{CF}(E, N)$ となる。前時刻分類器群生成部33は、子のノードが全て葉である親のノードに対し、親における誤りの期待値(誤りの発生する期待値)と、子である葉の誤りの期待値の合計とを比較する。そして、子での期待値の合計の方が親の誤りの期待値よりも大きければ、前時刻分類器群生成部33は、葉を縮退して、その親を葉とする。前時刻分類器群生成部33は、この処理を順次繰り返すことで、決定木全体の葉の枝刈りを行う。

【0070】

10

葉を縮退する場合、前時刻分類器群生成部33は、葉を削除して、その削除した葉の親のノードを葉とすればよい。例えば、図9に例示する決定木において、「湿度」という項目の値に応じて分割したノードを縮退する場合、前時刻分類器群生成部33は、湿度の項目値が70%以下となっている個別データ数を表すノード「する:2,しない:0」と、湿度の項目値が70%より高くなっている個別データ数を表すノード「する:0,しない:3」とを削除して、その2つのノードの親ノード「する:2,しない:3」を葉とすればよい。

【0071】

分類器として決定木を生成する場合、例えば、上記のように、決定木を定めて枝刈りを行うことで、決定木を生成すればよい。

20

【0072】

ステップA3でFalse[i]に応じた前時刻データから分類器(前時刻分類器i)を生成すると、誤差群算出部34は、その前時刻分類器iを用いて、現時刻データのラベルを予測する(ステップA4)。すなわち、誤差群算出部34は、現時刻データ内の各個別データ毎に、その個別データにおける顧客の属性を示す項目と、前時刻分類器iとを照合し、その個別データにラベル付けられるラベルを予測する。

【0073】

例えば、分類器が決定木である場合、誤差群算出部34は、そのルートのノードの項目に関して現時刻データ内の個別データの項目値を参照し、その項目値に応じて子ノードを辿る。誤差群算出部34は、葉のノードまで辿ったならば、葉のノードでカウント数の多い方のラベルを予測結果とすればよい。例えば、葉のノードで、「正:3,負:0」となっていれば、「正」と予測すればよい。

30

【0074】

ステップA4の後、誤差群算出部34は、現時刻データ内の各個別データにラベル付けられるラベルの予測結果と、現時刻データとの誤差を算出する(ステップA5)。すなわち、現時刻データ内の各個別データには、現時刻データ生成時に正または負のラベルがラベル付けられているので、ステップA4での予測結果と、実際の現時刻データでラベル付けされているラベルとの誤差を予測する。この誤差をErr[i]と記す。

【0075】

誤差群算出部34は、ステップA5において、例えば、現時刻データ内の個別データ毎に予測したラベルと、実際に現時刻データの各個別データにラベル付けられているラベルとを比較し、両者が異なっている個別データ数をカウントして、そのカウント値をErr[i]として求めてもよい。

40

【0076】

あるいは、誤差群算出部34は、現時刻データ内の個別データ毎に予測したラベルのうち、正がラベル付けられると予測した個別データの数と、現時刻データの中で実際に正がラベル付けられた個別データの数との差を計算して、その差をErr[i]として求めてもよい。すなわち、正のラベルの予測数と、実際に現時刻データの中で正がラベル付けられた個別データ数(実際の利用者数)との差の絶対値をErr[i]としてもよい。

【0077】

50

ステップA5の後、影響度算出部35は、誤差 $Err[i]$ と最小誤差値 $ErrorMin$ とを比較し、 $Err[i]$ が $ErrorMin$ 未満であるか否かを判定する(ステップA6)。 $Err[i]$ が $ErrorMin$ 未満であるならば(ステップA6におけるYes)、影響度算出部35は、 p を0に初期化し、 $ErrorMin$ に $Err[i]$ を代入する(ステップA11)。 $Err[i]$ が $ErrorMin$ 未満ということは、これまで最小としていた誤差よりもさらに小さい最小値が見つかったことを意味する。この場合、ステップA11において、その最小値で $ErrorMin$ を更新し、また、その誤差に対応する仮影響度が複数ある場合に各仮影響度を個別に指定するための変数 p を初期化している。

【0078】

10

ステップA11の後、影響度算出部35は、 p の値を1インクリメントし、 $Imp[p]$ に $False[i]$ を代入する(ステップA12)。また、ステップA6において、 $Err[i]$ が $ErrorMin$ 以上であると判定した場合(ステップA6におけるNo)、影響度算出部35は、 $Err[i]$ が $ErrorMin$ と等しいか否かを判定する(ステップA7)。ここで等しいと判定した場合(ステップA7におけるYes)にも、影響度算出部35は、 p の値を1インクリメントし、 $Imp[p]$ に $False[i]$ を代入する(ステップA12)。

【0079】

ステップA11からステップA12に移行した場合、 $p=1$ であり、現在着目している仮影響度 $False[i]$ を、誤差が最小となる1番目の仮影響度とする。また、ステップA7において、着目している $False[i]$ に対応する $Err[i]$ が $ErrorMin$ と等しく、ステップA12に移行した場合、 p は2以上の値になる。この場合、既に最小の誤差に対応する仮影響度は1つ以上見つかり、影響度算出部35は、現在着目している仮影響度 $False[i]$ を、誤差が最小となる p 番目の仮影響度と定めることになる。

20

【0080】

ステップA12の後、あるいは、ステップA7で $Err[i]$ が $ErrorMin$ と等しくないと判定した場合(ステップA7におけるNo)、影響度算出部35は、 i の値をインクリメントする(ステップA8)。次に、影響度算出部35は、 i の値が、最初に入力された仮影響度の個数(M とする。)以下であるか否かを判定する(ステップA9)。 i が仮影響度数 M 以下であるならば(ステップA9におけるYes)、ステップA8でインクリメントされた i によって定まる仮影響度 $False[i]$ に関してステップA2以降の処理を行う。すなわち、着目する仮影響度を変更してステップA2以降の処理を行う。

30

【0081】

i の値が仮影響度数 M よりも大きければ(ステップA9におけるNo)、ステップA10に移行する。ここで、 i は、誤差算出が終了した仮影響度の数を表示している。ステップA10において、影響度算出部35は、最小誤差に対応する各仮影響度の平均値を計算し、その平均値を影響度として出力する(ステップA10)。最小誤差に対応する仮影響度が1つしかなければ、 $p=1$ となっている。このとき、影響度算出部35は、 $Imp[1]$ を影響度として出力すればよい。また、最小誤差に対応する仮影響度が複数ある場合、 p は2以上の値となっている。このとき、影響度算出部35は、 $Imp[1]$ から $Imp[p]$ までの各影響度の平均値を計算し、その平均値を影響度として出力すればよい。

40

【0082】

影響度算出部35は、例えば、出力装置4に影響度を表示させる。あるいは、他の出力態様で影響度を出力してもよい。例えば、出力装置4が印刷装置であって、出力装置4に影響度を印刷させてもよい。

【0083】

次に、第1の実施形態の効果について説明する。

第1の実施の形態では、入力された複数の仮影響度のそれぞれに対し予測誤差を求め、

50

その誤差が最小となる仮影響度の平均を影響度とする。従って、前時刻から現時刻までの期間における影響度を推定することができる。また、影響度を調べようとする期間にあわせて、入力する現時刻および時刻間隔を定めることにより、調べようとする期間における影響度を推定することができる。

【 0 0 8 4 】

また、誤差群算出部 3 4 が、予測したラベルと実際に現時刻データの各個別データにラベル付けられているラベルとを比較し、両者が異なっている個別データ数をカウントして、そのカウント値を誤差 ($Err[i]$) とする場合には、影響度の推定性能を高めることができる。

【 0 0 8 5 】

また、誤差群算出部 3 4 が、現時刻データ内の個別データ毎に予測したラベルのうち、正がラベル付けられると予測した個別データの数と、現時刻データの中で実際に正がラベル付けられた個別データの数との差を誤差 ($Err[i]$) とする場合には、現時刻で利用を開始すべき顧客が次の時刻で開始したり、次の時刻で利用を開始すべき顧客が現時刻で早めに利用を開始したりするといった細かな変動を均して影響度を推定することができる。

【 0 0 8 6 】

また、上記の実施形態では、入力された 1 つの現時刻における影響度を推定する場合を説明したが、複数種類の現時刻が入力され、その各現時刻に対して、それぞれステップ A 1 以降の処理を行い、各現時刻毎に影響度を求めてもよい。この場合、刻々と変動する影響度を推定することができる。

【 0 0 8 7 】

また、例えば、複数の現時刻が入力され、各現時刻毎に影響度を求める場合、以下の様に動作してもよい。すなわち、影響度推定システムのユーザから各現時刻毎に影響度を求めるように指定された場合に、各現時刻毎にステップ A 1 以降の処理を行って、各現時刻毎の影響度を求め、影響度算出部 3 5 は、それらの現時刻毎の影響度の平均値を影響度として求めてもよい。このように影響度を求めれば、広告・販売促進活動の影響を受けにくい期間で多くの時刻を用いると影響度の推定誤差を減少できると思われる場合に、推定誤差の少ない影響度を求めることができる。また、各現時刻毎の影響度の平均値を求めるときに、最近に近いほど（すなわち、現時刻が遅くなっているほうの影響度ほど）重みを付けて、重み付き平均の値を求め、その重み付き平均の値を影響度としてもよい。

【 0 0 8 8 】

また、影響度推定システムのユーザが複数の現時刻を指定する場合、現時刻の最小値、最大値、および現時刻を変化させる間隔を現時刻データ生成部 3 1 に入力してもよい。現時刻データ生成部 3 1 は、現時刻の最小値、最大値、および現時刻を変化させる間隔が入力されると、その現時刻の最小値から最大値までの期間をその間隔毎に区切って、入力された最小値および最大値の他に、その境界の時刻を現時刻としてもよい。そして、それらの現時刻毎に影響度を求め、その移動平均あるいは移動重み付き平均を求めてもよい。

【 0 0 8 9 】

また、複数の現時刻が入力され、各現時刻毎に各仮影響度における誤差を計算し、その各現時刻における各仮影響度の誤差から影響度を定めてもよい。例えば、複数の現時刻 $T_1 \sim T_x$ が入力されているとする。本例では、影響度推定システムは、現時刻 T_1 に関し、ステップ A 1 ~ A 5 を行い、 $Err[1]$ を求める。ステップ A 5 の後に誤差算出部 3 5 は、ステップ A 8 およびステップ A 9 を行い、さらに $i = 2$ 以上の場合の誤差 $Err[2] \sim Err[M]$ を順次求める。同様に、他の現時刻に関しても $Err[1] \sim Err[M]$ を求める。図 1 1 は、このように現時刻毎に求めた各仮影響度における誤差を示している。各現時刻における $Err[1] \sim Err[M]$ を求めた後、誤差算出部 3 5 は、 $1 \sim M$ までの各 i 毎に、各現時刻における $Err[i]$ の和を計算し、その和が最小となる i によって定められる仮影響度を影響度としてもよい。図 1 1 に示す例では、各行毎に、 $Err[i]$ の和を求め、その和が最小となる i によって定まる $False[i]$ を影

10

20

30

40

50

響度としてもよい。

【 0 0 9 0 】

また、影響度推定システムのユーザが複数の仮影響度を指定する場合、仮影響度の最大値、最小値、および仮影響度を变化させる間隔を現時刻データ生成部 3 1 に入力してもよい。現時刻データ生成部 3 1 は、仮影響度の最大値、最小値、および仮影響度を变化させる間隔が入力されると、その仮影響度の最小値から最大値までの値の範囲をその期間毎に区切って、入力された最小値および最大値の他に、その境界の値を仮影響度として定めてもよい。

【 0 0 9 1 】

また、第 1 の実施の形態では、影響度を出力するよう記載したが、影響度の信頼性をユーザーが知るため、算出した誤差などを出力してもよい。例えば、影響度算出手段 3 5 が、影響度の他に、Err [i] を出力装置 4 に出力させてもよい。

10

【 0 0 9 2 】

また、第 1 の実施形態では、影響度推定システムが顧客 DB 2 を備える場合を説明したが、顧客 DB 2 が影響度推定システムの外部に設けられ、影響度推定システムがその顧客 DB 2 を参照して処理を行う構成であってもよい。

【 0 0 9 3 】

実施形態 2 .

図 1 2 は、本発明の普及予測システムの例を示すブロック図である。第 1 の実施形態と同様の構成要素については、図 1 と同一の符号を付す。本発明の普及予測システムは、個々の顧客毎に、予測対象の時刻において商品またはサービスの利用を開始するか否かを判定する普及予測装置 5 と、顧客データベース（以下、顧客 DB と記す）2 とを備える。

20

【 0 0 9 4 】

本実施形態の顧客 DB 2 は、顧客データを記憶する記憶手段であり、第 1 の実施形態における顧客 DB 2（図 1 参照）と同様である。本実施形態の顧客 DB 2 が記憶する顧客データも、第 1 の実施形態における顧客データと同様である。すなわち、顧客データは、図 4 に示すように、個々の顧客毎に定められた個別データの集合である。顧客 DB 2 において、各個別データは、利用開始情報と、利用開始情報以外の顧客の属性を表す一つ以上の項目とを含む。顧客が商品またはサービスを利用し始めている場合には、その商品またはサービスを利用し始めた時期が利用開始情報となる。また、顧客が商品またはサービスを未利用である場合には、未利用である旨が利用開始情報となる。

30

【 0 0 9 5 】

普及予測装置 5 は、現時刻と影響度とが入力される。第 2 の実施形態における現時刻は、顧客が商品またはサービスを利用し始めたか否かの判定対象時刻を定める。また、影響度は、第 1 の実施形態の影響度推定システムが求めた影響度であることが好ましいが、普及予測システムのユーザが決めた影響度であってもよい。また、本実施形態における現時刻は、将来の時刻であってもよい。

【 0 0 9 6 】

普及予測装置 5 は、テストデータ生成部 5 1 と、学習データ生成部 5 2 と、分類器生成部 5 3 と、テストデータラベル判定部 5 4 とを備える。

40

【 0 0 9 7 】

テストデータ生成部 5 1 は、現時刻と顧客 DB 2 とに基づいて、現時刻に利用を開始していない顧客の個別データを含むテストデータを生成する。

【 0 0 9 8 】

テストデータは、現時刻において、商品またはサービスの利用を開始していない顧客の個別データを含むデータである。テストデータは、現時刻で商品またはサービスの利用を開始していない顧客の個別データのみを含むことが好ましいが、現時刻で商品またはサービスの利用を開始している顧客の個別データを含んでいてもよい。以下、テストデータが、現時刻で商品またはサービスの利用を開始していない顧客の個別データのみを含む場合を例にして説明する。

50

【 0 0 9 9 】

学習データ生成部 5 2 は、入力された現時刻および影響度と顧客 D B 2 に記憶されている顧客データとに基づいて、学習データを生成する。学習データは、現時刻で商品またはサービスの利用を開始している顧客の個別データに第 1 ラベルをラベル付け、現時刻で商品またはサービスの利用を開始していない顧客の個別データに第 2 ラベルをラベル付け、入力された影響度に応じて重み付けを行ったデータである。第 1 の実施形態で述べたように、重み付けとは、第 2 ラベルをラベル付けた個別データ数に対して、第 1 ラベルをラベル付けた個別データ数を相対的に変動させることである。また、第 1 の実施形態と同様に、第 1 ラベルを正（または + ）と記し、第 2 ラベルを負（または - ）と記す。

【 0 1 0 0 】

分類器生成部 5 3 は、学習データ生成部 5 2 が生成した学習データを用いて分類器を生成する。

【 0 1 0 1 】

テストデータラベル判定部 5 4 は、テストデータ内の各個別データのラベルを分類器によって予測する。すなわち、テストデータ内の各個別データの項目と分類器とを照合して、正または負のいずれにラベル付けられるかという判定を行う。正と判定された個別データの顧客は、現時刻の次の時刻で商品またはサービスの利用を開始すると予測されることになる。負と判定された個別データの顧客は、現時刻の次の時刻ではまだ商品またはサービスの利用を開始しないと予測されることになる。なお、現時刻の次の時刻とは、現時刻に、定められた一定時間を加算した時刻である。

【 0 1 0 2 】

なお、テストデータラベル判定部 5 4 は、テストデータ内の各個別データに対するラベルの予測結果（判定結果）を、出力装置（図 1 2 において図示せず。）に出力させてもよい。例えば、予測結果を、ディスプレイ装置などに表示させても、印刷装置などに印刷させてもよい。

【 0 1 0 3 】

また、テストデータ生成部 5 1 および学習データ生成部 5 2 は、外部から現時刻を入力されてもよい。また、学習データ生成部 5 2 は、外部から影響度を入力されてもよい。このとき、例えば、キーボードなどの入力装置（図 1 2 において図示せず。）を介して現時刻および影響度が入力されてもよい。あるいは、第 1 の実施形態の影響度推定システムがある現時刻における影響度を求めた場合、例えば、第 1 の実施形態の影響度推定システムの影響度算出手段 3 5 が、現時刻をテストデータ生成部 5 1 および学習データ生成部 5 2 に渡し、影響度を学習データ生成部 5 2 に渡してもよい。

【 0 1 0 4 】

普及予測装置 5 が備えるテストデータ生成部 5 1、学習データ生成部 5 2、分類器生成部 5 3、およびテストデータラベル判定部 5 4 は、例えば普及予測プログラムに従って動作する CPU によって実現される。すなわち、普及予測装置 5 に設けられた記憶装置から普及予測プログラムを読み込んだ CPU がテストデータ生成部 5 1、学習データ生成部 5 2、分類器生成部 5 3、およびテストデータラベル判定部 5 4 として動作してもよい。

【 0 1 0 5 】

次に、普及予測システムの動作について説明する。

図 1 3 は、本発明の普及予測システムの処理経過の例を示すフローチャートである。例えば、キーボードなどの入力装置（図 1 3 において図示せず。）を介して、普及予測装置 5 に現時刻および影響度が入力されると、普及予測システムは以下のように動作する。

【 0 1 0 6 】

まず、テストデータ生成部 5 1 は、テストデータを生成する（ステップ B 1）。テストデータ生成部 5 1 は、顧客データ D B 2 に記憶されている顧客データの中から、現時刻で商品またはサービスの利用を開始していない顧客の個別データのみを抜き出して、その個別データの集合をテストデータとすればよい。具体的には、利用開始情報（図 4 参照）が現時刻以前の時刻となっていない個別データのみを抜き出して、その個別データの集合を

10

20

30

40

50

テストデータとすればよい。ここでは、テストデータが、現時刻で利用を開始していない顧客の個別データのみを含む場合を例にしているが、現時刻で利用を開始している顧客の個別データをテストデータに含めるようにしてもよい。

【0107】

ステップB1の後、学習データ生成部52は、学習データを生成する(ステップB2)。学習データ生成部52は、顧客DB2に記憶された顧客データを読み込み、現時刻以前で商品またはサービスの利用を開始している顧客の個別データに正をラベル付け、現時刻で商品またはサービスの利用を開始していない顧客の個別データに負をラベル付ける。すなわち、利用開始情報が現時刻以前の時刻となっている個別データに正をラベル付け、利用開始情報が現時刻以前の時刻でない個別データに負をラベル付ける。そして、学習データ生成部52は、影響度に応じて、正とした個別データの数を、負とした個別データの数に対して相対的に変動させる。本例では、影響度の関数の関数値を計算し、その関数値倍に、正をラベル付けた個別データを増加させる場合を例にする。

10

【0108】

影響度の関数として、例えば、現時刻で商品またはサービスの利用を開始している顧客数と現時刻で未だその商品またはサービスの利用を開始していない顧客数との比率を影響度の係数とする関数を用いてもよい。すなわち、以下に例示する関数を用いてもよい。

【0109】

影響度 × (現時刻の未利用者数) / (現時刻の利用者数)

【0110】

上記の関数において、「現時刻の未利用者数」は、現時刻で未だ商品またはサービスの利用を開始していない顧客の数(より具体的には個別データの数)である。すなわち、顧客データにおいて、利用開始情報が現時刻以前でない個別データの数である。また、「現時刻の利用者数」は、現時刻で商品またはサービスの利用を開始している顧客の数である。すなわち、顧客データにおいて、利用開始情報が現時刻以前となっている個別データの数である。このように、「現時刻の未利用者数」を「現時刻の利用者数」で除算して得た比率を影響度の係数とする関数を用いてもよい。

20

【0111】

また、影響度の関数として、例えば、現時刻で商品またはサービスの利用を開始している顧客数と全顧客数との比率を影響度の係数とする関数を用いてもよい。すなわち、以下に例示する関数を用いてもよい。

30

【0112】

影響度 × (顧客人数) / (現時刻の利用者数)

【0113】

上記の関数において、「顧客人数」は、全顧客数(顧客データ内の全個別データ数)である。「現時刻の未利用者数」は、上述の関数の場合と同様である。このように、「顧客人数」を「現時刻の利用者数」で除算して得た比率を影響度の係数とする関数を用いてもよい。

【0114】

学習データ生成部52は、このような関数の関数値倍に、正をラベル付けた個別データを増加させる。例えば、入力された影響度を関数に代入して得た値がjであるとする。この場合、正をラベル付けた個別データの数がj倍になるように、正をラベル付けた個別データの複製を生成すればよい。

40

【0115】

また、例示した2つの関数における影響度の係数の値は、いずれも時間が経過して利用者数が増えるに従って減少する。影響度の係数は、このような係数に限定されない。また、影響度の値をそのまま関数値として用いてもよい。この場合、係数を1としていることになる。

【0116】

また、上記の例では、正とラベル付けた個別データを関数値倍に増加させる場合を示し

50

ているが、重み付けでは、正とラベル付けた個別データ数を、負とラベル付けた個別データ数に対して相対的に変化させればよい。この点は、第1の実施形態において前時刻データ群生成部32が行う重み付けと同様である。

【0117】

ステップB2の後、分類器生成部53は、学習データを用いて分類器を生成する(ステップB3)。分類器生成部53は、重回帰分析、決定木、ニューラルネットワーク、サポートベクターマシン、ベイジアンネットワークなどのデータマイニングによる手法のいずれかあるいはその組合せにより分類器を生成すればよい。

【0118】

また、第1の実施形態において前時刻データ群生成部32の分類器生成動作として例示したように、分類器として決定木を生成してもよい。すなわち、分類器生成部53は、学習データが与えられると、どの項目で最初に分割させるかを決定する。このとき、分類器生成部53は、項目1~Nの各項目について、分割時の評価値を計算し、その評価値が最大の項目を、分割に最も適した項目として選択すればよい。評価値として、例えば、分割前のノードのエントロピーと、分割後のエントロピーの差を用いればよい。分割前のノードのエントロピー、および分割後のエントロピーの計算方法は、第1の実施形態で既に説明した計算方法と同様であり、説明を省略する。分類器生成部53は、分割後の各ノードについても、上記と同様の処理を行い、次にどの項目で分割するのかを決定する処理を順次繰り返し、所定の条件が満たされたときには、ノードの分割を停止する。そして、分類器生成部53は、このようにして得た木構造の決定木に対して、枝刈りを行うことにより、分類器となる決定木を生成する。上記の所定の条件や枝刈りの処理も、第1の実施形態で示した所定の条件や枝刈り処理と同様であり、説明を省略する。このようにして生成した決定木(分類器)は、学習データ内で第1ラベルをラベル付けられた個別データの割合が高くなるほど、テストデータ内で第2ラベルをラベル付けられている個別データのうち、第1ラベルをラベル付けられた個別データの項目値と類似する項目値を持つ個別データに対して、第1ラベルをラベル付けると判定する頻度が高くなるという性質を有している。なお、分類器は決定木に限定されず、分類器生成部53は、上記以外の動作で分類器を生成してもよい。

【0119】

ステップB3の後、テストデータラベル判定部54は、ステップB3で生成された分類器を用いて、テストデータのラベルを予測する(ステップB4)。すなわち、テストデータラベル判定部54は、テストデータ内の各個別データ毎に、その個別データにおける顧客の属性を示す項目と、ステップB3で生成された分類器とを照合し、その個別データに対するラベルが正であるか負であるかを判定する。例えば、分類器が決定木である場合、テストデータラベル判定部54は、そのルートの子ノードの項目に関してテストデータ内の個別データの項目値を参照し、その項目値に応じて子ノードを辿る。テストデータラベル判定部54は、葉の子ノードまで辿ったならば、葉の子ノードでカウント数の多い方のラベルを判定結果とすればよい。

【0120】

テストデータラベル判定部54は、テストデータ内の個別データ毎に判定した正または負のラベルを、例えば、ディスプレイ装置などに表示させてもよい。あるいは、印刷装置に印刷させてもよい。

【0121】

本実施形態によれば、顧客の属性を示す項目を含む個別データの集合に対してラベル付けを行い、さらに重み付けをして得た学習データから分類器を生成する。そして、テストデータラベル判定部54がテストデータ内の各個別データ毎に、その項目と分類器とを照合して、正または負のいずれかを判定する。すなわち、顧客の属性を示す項目を利用して分類器を生成し、利用を開始していない顧客の項目とその分類器とから、その顧客が現時刻の次の時刻で利用を開始するか否かを予測している。従って、顧客の属性を利用して、顧客が現時刻の次の時刻で利用を開始するか否かを判定することができる。

10

20

30

40

50

【 0 1 2 2 】

また、個別データ毎にラベルの予測を行うので、個々の顧客毎に現時刻の次の時刻で商品またはサービスを利用開始するか否かを予測することができる。また、指定する現時刻を変更すれば、所望の時刻で各顧客毎に商品またはサービスを利用開始するか否かを予測することができる。すなわち、個別の顧客がある時刻で商品やサービスの利用を開始するか否かを予測できる。従って、その予測結果を、商品やサービスの販売促進に活用することができる。

【 0 1 2 3 】

また、ロジスティック曲線等にデータを当てはめて普及曲線を予測する場合には、立ち上がりからのデータ（すなわち、商品やサービスの提供開始時からの提供数データ）が必要であるが、本発明では、顧客データが一部の期間におけるデータである場合にも、現時刻の次の時刻で各顧客毎に商品またはサービスを利用開始するか否かを予測することができる。

10

【 0 1 2 4 】

次に、第2の実施形態の変形例を示す。以下に示す第2の実施形態の変形例では、テストデータに対する予測結果を顧客DB2が記憶する顧客データに反映させる。そして、現時刻を更新してテストデータに対する予測を繰り返して、商品またはサービスの普及の程度を予測する。

【 0 1 2 5 】

図14は、第2の実施形態の変形例を示すブロック図である。既に説明した構成要素と同様の構成要素は、図12と同一の符号を付し、説明を省略する。また、図12では、普及予測装置5に現時刻および影響度などを入力するための入力装置1と、予測結果を出力するための出力装置4とを備えている場合を例にして示している。入力装置1は、例えば、キーボードなどの入力装置であり、出力装置4は、例えば、ディスプレイ装置などの出力装置である。

20

【 0 1 2 6 】

普及予測装置5は、テストデータ生成部51と、学習データ生成部52と、分類器生成部53と、テストデータラベル判定部54と、現時刻更新部55とを備える。

【 0 1 2 7 】

本変形例では、テストデータラベル判定部54は、ステップB4での判定結果が実際の利用開始情報と異なっている個別データの利用開始情報を上書きして更新する。判定結果が実際の利用開始情報と異なっている場合とは、個別データの利用開始情報が未利用を示していて、判定結果が正である場合である。この場合、テストデータラベル判定部54は、顧客DB2に記憶されている顧客データ内におけるその個別データの利用開始情報を現時刻に一定の時間（Tとする。）分を加算した時刻に更新する。このように時刻を更新して、その個別データが、顧客がその時刻から商品またはサービスの利用を開始したことを表すようにする。

30

【 0 1 2 8 】

現時刻更新部55は、テストデータラベル判定部54がテストデータに対する判定を行った後、現時刻を一定の時間（T）分増加した時刻を新たな現時刻とするように、現時刻を更新する。現時刻更新部55も、例えば、普及予測プログラムに従って動作するCPUによって実現される。

40

【 0 1 2 9 】

テストデータ生成部51、学習データ生成部52、分類器生成部53、およびテストデータラベル判定部54は、更新後の現時刻が定められた期間内であることを条件に、更新後の顧客データおよび現時刻を用いて、第2の実施形態と同様の処理を繰り返す。

【 0 1 3 0 】

次に、本変形例の動作について説明する。図15は、第2の実施形態の変形例における処理経過の例を示すフローチャートである。本変形例では、普及予測装置5に現時刻、影響度の他に、普及予測の対象とする期間（以下、予測期間と記す。）が入力される。予測

50

期間として、例えば、普及予測の対象とする期間の最後の時刻が入力されてもよい。

【0131】

その後、ステップB1～B4の処理を行う。ステップB1～B4の処理は、第2の実施形態で既に説明したステップB1～B4（図12参照）と同様の処理である。

【0132】

ステップB4の後、テストデータラベル判定部54は、個別データの利用開始情報が未利用を示している、ステップB4での判定結果として正がラベル付けられるという結果が得られた個別データの利用開始情報を更新する（ステップB5）。テストデータラベル判定部54は、顧客データ内のその個別データの利用開始情報を、現時刻に一定の時間Tを加算した時刻に更新する。

10

【0133】

次に、現時刻更新部55は、現時刻を更新する（ステップB6）。ステップB6において、現時刻更新部55は、現時刻に一定の時間Tを加算した時刻を新たな現時刻とする。

【0134】

そして、現時刻更新部55は、更新後の現時刻が予測期間内であるか否かを判定する（ステップB7）。更新後の現時刻が予測期間内であると判定した場合（ステップB7におけるYes）、普及予測装置5はステップB1以降の処理を繰り返す。

【0135】

また、更新後の現時刻が予測期間を越えると判定した場合（ステップB7におけるNo）、現時刻更新部55は、予測結果を出力装置4に出力させる（ステップB8）。例えば、各顧客毎の個別データにおける利用開始情報を、個々の顧客の予測される利用開始時期として出力装置4に出力させてもよい。また、商品またはサービスの普及曲線を出力装置4に出力させてもよい。普及予測システムが出力装置4としてディスプレイ装置などの表示装置を備えていて、現時刻更新部55が出力装置4に予測結果を表示させてもよい。あるいは、普及予測システムが出力装置4として印刷装置を備えていて、現時刻更新部55が出力装置4に予測結果を印刷させてもよい。

20

【0136】

普及曲線を出力する場合、例えば、ステップB1に移行する毎に、現時刻更新部55（他の構成要素でもよい）が、現時刻において商品またはサービスの利用を開始している顧客の個別データ数を顧客DBの中からカウントし、そのカウント値と現時刻とを対応付けて記憶していけばよい。そして、ステップB8において、例えば、現時刻を横軸にとり、カウント値を縦軸にとって、現時刻の変化に伴うカウント値の推移を示すグラフを作成し、そのグラフを出力装置4に出力させてもよい。このカウント値は、利用を開始した顧客数に相当する。

30

【0137】

第2の実施形態の変形例では、顧客の属性を示す項目を含む個別データの集合に対してラベル付けを行い、さらに重み付けをして得た学習データから分類器を生成し、その分類器でテストデータに対する予測を行うことを、現時刻を更新しながら繰り返す。従って、顧客の属性（特徴）の違いを考慮して、商品やサービスの普及の程度を高い精度で予測することができる。

40

【0138】

また、顧客のデータを増やすほど普及の予測精度を向上させることができる。

【0139】

また、学習データ生成時に用いる関数として、現時刻で商品またはサービスの利用を開始している顧客数と現時刻で未だその商品またはサービスの利用を開始していない顧客数との比率や、あるいは、現時刻で商品またはサービスの利用を開始している顧客数と全顧客数との比率を、影響度の係数とする関数を用いることにより、影響度だけでなく商品やサービスの普及率に応じて関数値を定め、重み付けを行うことができる。その結果、予測の精度をより高めることができる。

【0140】

50

また、学習データ生成時に用いる関数として、影響度の係数として、顧客の属性に応じた係数も含む関数を用いてもよい。例えば、学習データ生成時に、以下のような関数を用いて重み付けを行ってもよい。

【0141】

影響度 × 顧客係数 × (現時刻の未利用者数) / (現時刻の利用者数)

【0142】

学習データ生成部52は、個別データにおける顧客の属性を示す項目の内容に応じて顧客係数を選択し、その顧客係数を用いて、例えば、上記のような関数により関数値を計算してもよい。このように、顧客毎に個別に重み付けを行ってもよい。

【0143】

また、上記の変形例では、顧客DB2に記憶されている個別データの利用開始情報を上書きして更新する場合を例にして説明したが、個別データに当初から含まれている利用開始情報のコピーを利用開始時期予測情報としてコピーしておき、そのコピーを更新するようにしてもよい。この場合、実績と予測値とをそれぞれ保持しておくことができる。

【0144】

実施形態3 .

本実施形態の普及予測システムは、影響度を算出し、その影響度を用いて、予測対象の時刻において各顧客が商品またはサービスの利用を開始するか否かを判定する。図16は、本実施形態の普及予測システムの例を示すブロック図である。本実施形態の普及予測システムは、顧客DB2と、影響度推定装置3と、普及予測装置5とを備える。また、入力装置1および出力装置4を備えていてもよい。影響度推定装置3は、第1の実施形態で説明した影響度推定装置3と同様に動作する。また、普及予測装置5は、第2の実施形態で説明した普及予測装置5と同様に動作する。顧客DB2、入力装置1、および出力装置4も、第1の実施形態および第2の実施形態で説明した顧客DB2、入力装置1、出力装置4と同様である。

【0145】

顧客DB2は、顧客データを記憶する記憶手段であり、顧客DB2が記憶する顧客データも、第1の実施形態および第2の実施形態における顧客データと同様である。すなわち、顧客データは、利用開始情報と、利用開始情報以外の顧客の属性を表す一つ以上の項目とを含む顧客毎の個別データの集合である。

【0146】

影響度推定装置3は、例えば入力装置1を介して、現時刻、時刻間隔、および複数の仮影響度を入力される。影響度推定装置3は、その現時刻、時刻間隔、および複数の仮影響度を用いて影響度を推定し、その影響度を普及予測装置5に渡す。

【0147】

普及予測装置5は、例えば入力装置1を介して、現時刻を入力される。普及予測装置5は、その現時刻と、影響度推定装置3が定めた影響度とを用いて、テストデータに対する予測を行う。

【0148】

本実施形態において、入力される現時刻は、影響度推定装置3において影響度の推定対象時刻としても用いられる。

【0149】

図17は、本実施形態の普及予測システムを詳細に示すブロック図である。ただし、入力装置1および出力装置4は省略している。

【0150】

影響度推定装置3は、現時刻データ生成部31と、前時刻データ群生成部32と、前時刻分類器群生成部33と、誤差群算出部34と、影響度算出部35とを備える。普及予測装置5は、テストデータ生成部51と、学習データ生成部52と、分類器生成部53と、テストデータラベル判定部54とを備える。

【0151】

10

20

30

40

50

現時刻データ生成部 3 1、前時刻データ群生成部 3 2、前時刻分類器群生成部 3 3、誤差群算出部 3 4、影響度算出部 3 5 は、第 1 の実施形態と同様に動作する。すなわち、ステップ A 1 ~ A 1 0 (図 3 参照。) を実行し、影響度を定める。影響度算出部 3 5 は、その影響度を学習データ生成部 5 2 に渡す。

【 0 1 5 2 】

テストデータ生成部 5 1、学習データ生成部 5 2、分類器生成部 5 3、テストデータラベル判定部 5 4 は、第 2 の実施形態と同様に動作する。すなわち、ステップ B 1 ~ B 4 (図 1 3 参照) を実行し、個々の顧客毎に、商品またはサービスの利用が開始されるか否かを判定する。

【 0 1 5 3 】

なお、影響度算出部 3 5 は、1 つの影響度を定めて、普及予測装置 5 は、その影響度を用いて、ステップ B 1 以降の処理を行えばよい。

【 0 1 5 4 】

第 1 の実施形態で説明した影響度推定装置 3 に関する種々の変形が本実施形態の影響度推定装置 3 に適用されてもよい。また、第 2 の実施形態で説明した普及予測装置 5 に関する種々の変形が本実施形態の普及予測装置 5 に適用されていてもよい。第 2 の実施形態の変形例 (図 1 4 参照) と同様に、普及予測装置 5 が現時刻更新部 5 5 を備え、ステップ B 1 ~ B 8 を実行する構成であってもよい。例えば、普及曲線を出力装置 4 に出力させてもよい。

【 0 1 5 5 】

現時刻データ生成部 3 1、前時刻データ群生成部 3 2、前時刻分類器群生成部 3 3、誤差群算出部 3 4、影響度算出部 3 5、テストデータ生成部 5 1、学習データ生成部 5 2、分類器生成部 5 3、およびテストデータラベル判定部 5 4 は、例えば、普及予測プログラムに従って動作する CPU によって実現される。普及予測システムが現時刻更新部 5 5 を備える場合、現時刻更新部 5 5 もその CPU によって実現されてよい。

【 0 1 5 6 】

本実施形態においても、第 2 の実施形態と同様の効果を得ることができる。特に、影響度推定装置 3 が影響度を用いて処理を行うので、予測精度をより高くすることができる。

【 0 1 5 7 】

また、第 2 の実施形態および第 3 の実施形態では、普及予測システムが顧客 DB 2 を備える場合を説明したが、顧客 DB 2 が普及予測システムの外部に設けられ、普及予測システムがその顧客 DB 2 を参照して処理を行う構成であってもよい。

【 0 1 5 8 】

実施形態 4 .

図 1 8 は、本発明の第 4 の実施形態の普及予測システムの例を示すブロック図である。第 4 の実施形態の普及予測システムは、顧客 DB 2 と、閾値関数推定装置 6 と、普及予測装置 7 とを備える。また、第 1 の実施形態から第 3 の実施形態で説明した入力装置 1 および出力装置 4 と同様の入力装置および出力装置を備えていてもよい。顧客 DB 2 は、第 1 の実施形態から第 3 の実施形態で説明した顧客 DB 2 と同様である。

【 0 1 5 9 】

顧客 DB 2 は、顧客データを記憶する記憶手段であり、顧客 DB 2 が記憶する顧客データも、第 1 の実施形態から第 3 の実施形態における顧客データと同様である。すなわち、顧客データは、利用開始情報と、利用開始情報以外の顧客の属性を表す一つ以上の項目とを含む顧客毎の個別データの集合である。

【 0 1 6 0 】

本実施形態において、閾値関数推定装置 6 は、商品またはサービスの普及率と閾値との関係を示す閾値関数を推定する。閾値関数推定装置 6 は、複数の時刻において、普及率および閾値を求め、それぞれの時刻毎の普及率および閾値から閾値関数を推定する。この閾値とは、個別データに第 1 ラベルがラベル付けられるか否かを判定するための閾値である。普及予測装置 7 は、顧客 DB 2 に記憶された顧客データからテストデータを生成し、テ

10

20

30

40

50

ストデータ内の個別データに第1のラベルがラベル付けられる確からしさのスコアを計算する。また、普及予測装置7は、商品またはサービスの普及率も計算し、閾値関数推定装置6によって推定された閾値関数にその普及率を代入して閾値を算出する。普及予測装置7は、スコアがその閾値以上である個別データに対するラベルは第1ラベルであり、スコアがその閾値未満である個別データに対するラベルは第2ラベルであると判定する。以下、既に説明した実施形態と同様に、第1ラベルを正（または+）と記し、第2ラベルを負（または-）と記す。

【0161】

閾値関数推定装置6は、基準時刻データ群生成部61と、閾値推定前時刻データ群生成部62と、閾値推定前時刻分類器群生成部63と、誤差群算出部64と、関数推定部65

10

【0162】

基準時刻データ群生成部61は、閾値の推定対象時刻を定めるための複数の基準時刻と、その各基準時刻の一定時間前を指定するための時刻間隔と、複数の仮閾値とが入力される。キーボード等の入力装置（図18において図示せず。）を介して基準時刻データ群生成部61に、複数の基準時刻、時刻間隔および複数の仮閾値が入力されてもよい。

【0163】

本実施形態における基準時刻とは、閾値の推定対象時刻となる後述の閾値推定前時刻を定めるための基準時刻であり、後述の基準時刻データを用いて各閾値推定前時刻毎の閾値が決定される。なお、それぞれの閾値推定前時刻毎に普及率も計算され、それぞれの閾値推定前時刻における普及率と閾値との関係から閾値関数が推定される。

20

【0164】

時刻間隔は、入力された基準時刻の一定時間前を指定するための時間である。入力される仮閾値は、閾値関数推定装置6で求めようとしている閾値の候補である。各仮閾値は、例えば、0以上1以下の数値である。

【0165】

基準時刻データ群生成部61に入力される基準時刻の数をU個とし、仮閾値の数をV個とする。

【0166】

基準時刻データ群生成部61は、1つの基準時刻に対して、1つの基準時刻データを生成する。従って、U個の基準時刻データが生成される。基準時刻データは、基準時刻で商品またはサービスの利用を介している顧客の個別データに正（第1ラベル）をラベル付け、基準時刻で商品またはサービスの資料を開始していない顧客の個別データに負（第2ラベル）をラベル付けたデータである。基準時刻データ群生成部61は、顧客DB2を用いて（すなわち、顧客DB2に記憶された顧客データに基づいて）、基準時刻データを生成する。個々の基準時刻データの生成処理は、第1の実施形態における現時刻データの生成処理と同様である。また、基準時刻データ群生成部61は、それぞれの基準時刻から時刻間隔前の時刻（基準時刻から時刻間隔分遡った時刻）を求める。この時刻を、閾値推定前時刻と記す。基準時刻データ群生成部61は、入力された各基準時刻毎に、基準時刻から時刻間隔を減算して閾値推定前時刻を計算する。従って、閾値推定前時刻もU個求められる。

30

40

【0167】

閾値推定前時刻データ群生成部62は、1つの閾値推定前時刻に対して、1つの閾値推定前時刻データを生成する。従って、U個の閾値推定前時刻データが生成される。閾値推定前時刻データ群生成部62は、閾値推定前時刻毎に、閾値推定前時刻と顧客DB2に記憶されている顧客データとに基づいて、閾値推定前時刻データを生成する。閾値推定前時刻データは、閾値推定前時刻で商品またはサービスの利用を開始している顧客の個別データベースに正をラベル付け、閾値推定前時刻で商品またはサービスの利用を開始していない顧客の個別データに負をラベル付けたデータである。閾値推定前時刻データは、重み付けが行われていない点で、第1の実施形態で説明した前時刻データと異なる。

50

【 0 1 6 8 】

また、閾値推定前時刻データ群生成部 6 2 は、閾値推定前時刻データを用いて普及率を計算する。普及率は、顧客数に対する、閾値推定前時刻までに商品またはサービスの利用を開始した顧客数の割合である。閾値推定前時刻データ群生成部 6 2 は、閾値推定前時刻データ毎に普及率を計算する。よって、普及率も U 個計算される。閾値推定前時刻データ群生成部 6 2 は、関数推定部 6 5 で用いるために閾値推定前時刻データ毎の U 個の普及率を記憶しておいてもよい。

【 0 1 6 9 】

閾値推定前時刻分類器群生成部 6 3 は、1 つの閾値推定前時刻データに基づいて、1 つの分類器を生成する。従って、U 個の分類器が生成される。本実施形態における分類器は、顧客の個別データに第 1 ラベルがラベル付けられる確からしさのスコアを定めるルールである。

10

【 0 1 7 0 】

誤差群算出部 6 4 は、閾値推定前時刻分類器群生成部 6 3 で生成された分類器を用いて、基準時刻データ内の各個別データのスコアを計算し、スコアと仮閾値とを比較することで、各個別データのラベルを予測する。このスコアは、個別データに第 1 ラベルがラベル付けられる確からしさのスコアである。誤差群算出部 6 4 は、このラベル予測を分類器毎に行う。また、一つの分類器でのラベル予測では、入力された V 個の仮閾値毎にラベル予測を行う。誤差群算出部 6 4 は、ある分類器および仮閾値を用いてラベル予測を行う場合、その分類器で基準時刻データ内の各個別データのスコアを計算し、スコアがその仮閾値以上となっている個別データに対するラベルが正であり、スコアがその仮閾値未満となっている個別データに対するラベルが負であると予測する。また、その基準時刻データ内の各個別データには、既に実際にラベル付けが行われている。誤差群算出部 6 4 は、分類器および仮閾値を用いて予測したラベルと、実際に基準時刻データでラベル付けられているラベルとの誤差を算出する。この予測したラベルと実際のラベルとの誤差の算出は、第 1 の実施形態における誤差群算出部 3 4 が予測したラベルと実際のラベルとの誤差を算出する処理と同様である。U 個の分類器が生成され、各分類器に関しそれぞれ V 個の仮閾値について誤差を算出するので、 $U \times V$ 個の誤差を算出することになる。

20

【 0 1 7 1 】

関数推定部 6 5 は、U 個の分類器それぞれに対し、V 個の仮閾値毎に算出された誤差のうち、最小の誤差を特定し、最小の誤差に対応する仮閾値を分類器に対応する閾値（すなわち、閾値推定前時刻に対応する閾値）として定める。具体的には、最小の誤差に対応する仮閾値の数が一つである場合、関数推定部 6 5 は、その仮閾値を閾値として定める。また、最小の誤差に対応する仮閾値が複数個存在する場合には、その複数の仮閾値に基づいて、閾値推定前時刻に対応する閾値を定める。例えば、最小の誤差に対応する仮閾値が複数個存在する場合、その複数の仮閾値の平均値を計算し、その平均値を閾値推定前時刻に対応する閾値と定める。以下、最小の誤差に対応する仮閾値が複数個存在する場合、その仮閾値の平均値を閾値と定める場合を例にして説明する。

30

【 0 1 7 2 】

分類器毎に閾値を定めるので、各閾値推定前時刻に対応する閾値が定まる。また、閾値推定前時刻データ群生成部 6 2 によって、閾値推定前時刻毎の普及率が計算されている。よって、閾値推定前時刻毎の閾値および普及率の組が U 組得られる。関数推定部 6 5 は、この閾値および普及率の組から閾値関数を推定する。

40

【 0 1 7 3 】

閾値関数推定装置 6 が備える基準時刻データ群生成部 6 1、閾値推定前時刻データ群生成部 6 2、閾値推定前時刻分類器群生成部 6 3、誤差群算出部 6 4、および関数推定部 6 5 は、例えばプログラムに従って動作する CPU によって実現される。すなわち、閾値関数推定装置 6 に設けられた記憶装置からプログラムを読み込んだ CPU が、基準時刻データ群生成部 6 1、閾値推定前時刻データ群生成部 6 2、閾値推定前時刻分類器群生成部 6 3、誤差群算出部 6 4、および関数推定部 6 5 として動作してもよい。

50

【 0 1 7 4 】

普及予測装置 7 は、テストデータ生成部 7 1 と、ラベル付けデータ生成部 7 2 と、分類器生成部 7 3 と、テストデータラベル判定部 7 4 とを備える。

【 0 1 7 5 】

普及予測装置 7 には現時刻が入力される。この現時刻は、第 2 の実施形態で説明した現時刻データと同様に、顧客が商品またはサービスを利用し始めたか否かの判定対象時刻を定める。入力装置（図 1 8 において図示略）を介して普及予測装置 7 に現時刻が入力されてもよい。

【 0 1 7 6 】

テストデータ生成部 7 1 は、現時刻と顧客 DB 2 とに基づいて、現時刻に利用を開始していない顧客の個別データを含むテストデータを生成する。以下、テストデータが、現時刻で商品またはサービスの利用を開始していない顧客の個別データのみを含む場合を例にして説明する。

【 0 1 7 7 】

ラベル付けデータ生成部 7 2 は、現時刻および顧客 DB 2 に記憶されている顧客データに基づいて、ラベル付けデータを生成する。ラベル付けデータは、現時刻で商品またはサービスの利用を開始している顧客の個別データに正（第 1 ラベル）をラベル付け、現時刻で商品またはサービスの利用を開始していない顧客の個別データに負（第 2 ラベル）をラベル付けたデータである。ラベル付けデータは、重み付けが行われていない点で、第 2 の実施形態で説明した学習データと異なる。重み付けを行うか否かという点以外では、本実施形態におけるラベル付けデータ生成処理と、第 2 の実施形態における学習データ生成処理は同様である。

【 0 1 7 8 】

また、ラベル付けデータ生成部 7 2 は、生成したラベル付けデータを用いて普及率を計算する。

【 0 1 7 9 】

分類器生成部 7 3 は、ラベル付けデータ生成部 7 2 が生成したラベル付けデータを用いて分類器を生成する。この分類器は、閾値推定前時刻分類器群生成部 6 3 が生成する分類器と同じく、顧客の個別データに第 1 ラベルがラベル付けられる確からしさのスコアを定めるルールである。

【 0 1 8 0 】

テストデータラベル判定部 7 4 は、分類器生成部 7 3 で生成された分類器を用いて、テストデータ内の各個別データのスコアを計算する。また、テストデータラベル判定部 7 4 は、関数推定部 6 5 で推定された閾値関数と、ラベル付けデータから計算された普及率とを用いて、閾値を計算する。すなわち、閾値関数に普及率を代入して閾値を計算する。そして、スコアとその閾値とを比較して、テストデータラベル内の各個別データのラベルを判定する。具体的には、テストデータラベル判定部 7 4 は、テストデータラベル内の各個別データのうち、スコアが閾値以上となっている個別データに対するラベルが正であると判定し、スコアが閾値未満の個別データに対するラベルが第 2 ラベルであると判定する。正と判定された個別データの顧客は、現時刻の次の時刻で商品またはサービスの利用を開始すると予測されることになる。負と判定された個別データの顧客は、現時刻の次の時刻ではまだ商品またはサービスの利用を開始しないと予測されることになる。なお、現時刻の次の時刻とは、現時刻に、定められた一定時間を加算した時刻である。

【 0 1 8 1 】

テストデータラベル判定部 7 4 は、テストデータ内の各個別データに対するラベルの予測結果（判定結果）を、出力装置（図 1 8 において図示せず。）に出力させてもよい。

【 0 1 8 2 】

普及予測装置 7 が備えるテストデータ生成部 7 1、ラベル付けデータ生成部 7 2、分類器生成部 7 3、およびテストデータラベル判定部 7 4 は、例えば普及予測プログラムに従って動作する CPU によって実現される。すなわち、普及予測装置 7 に設けられた記憶装

10

20

30

40

50

置から普及予測プログラムを読み込んだCPUがテストデータ生成部71、ラベル付けデータ生成部72、分類器生成部73、およびテストデータラベル判定部74として動作してもよい。

【0183】

また、閾値関数推定装置6および普及予測装置7が同一の情報処理装置によって実現されてもよい。そして、基準時刻データ群生成部61、閾値推定前時刻データ群生成部62、閾値推定前時刻分類器群生成部63、誤差群算出部64、関数推定部65、テストデータ生成部71、ラベル付けデータ生成部72、分類器生成部73、およびテストデータラベル判定部74が、普及予測プログラムに従って動作するCPUによって実現されている。

10

【0184】

次に、動作について説明する。

図19および図20は、本実施形態における閾値関数推定装置6の処理経過の例を示すフローチャートである。例えば、入力装置(図18において図示せず。)を介して基準時刻データ群生成部61に複数の基準時刻、時刻間隔、および複数の仮閾値が入力されると、普及予測システムは以下のように動作する。なお、複数の基準時刻を $CTime[t]$ と記す。変数 t は、入力された複数の基準時刻を順番に指定するための変数であり、 $t = 1, \dots, U$ である。例えば、 $t = 1$ であれば、1番目の基準時刻を指定していることを意味し、 $CTime[1]$ は1番目の基準時刻を表している。時刻間隔を Δt と記す。また、複数の仮閾値を $FalseThr[i]$ と記す。ここで、変数 i は、入力された複数の仮閾値を順番に指定するための変数であり、 $i = 1, \dots, V$ である。例えば、 $i = 1$ であれば、1番目の仮閾値を指定していることを意味し、 $FalseThr[1]$ は1番目の仮閾値を表している。入力される基準時刻の数を U 個とし、仮閾値の数を V 個としている。

20

【0185】

基準時刻データ群生成部61は、基準時刻を指定する変数 t を初期化して $t = 1$ とする。 $t = 1$ であるので、1番目の基準時刻 $CTime[1]$ を選択していることになる。基準時刻データ群生成部61は、 $CTime[1]$ から時刻間隔 Δt を減算して、閾値推定前時刻を求める。すなわち、閾値推定前時刻を $CTime[1] - \Delta t$ に設定する(以上、ステップC1)。次に、基準時刻データ群生成部61は、 $t = U$ であるか否かを判定し(ステップC2)。 $t = U$ であれば(ステップC2におけるYes)、ステップC3に移行し、 $t > U$ であれば(ステップC2におけるNo)、ステップC18に移行する。

30

【0186】

ステップC2で $t = U$ と判定した場合、基準時刻データ群生成部61は、 t により定まる基準時刻 $CTime[t]$ と、顧客DB2に記憶された顧客データとに基づいて、基準時刻データを生成する(ステップC3)。基準時刻データ群生成部61は、顧客DB2に記憶された顧客データを読み込み、基準時刻 $CTime[t]$ で商品またはサービスの利用を開始している顧客の個別データ(利用開始情報が $CTime[t]$ 以前の時刻を表している個別データ)に正をラベル付け、基準時刻 $CTime[t]$ で商品またはサービスの利用を開始していない顧客の個別データに負をラベル付けることによって、基準時刻データを生成する。このとき、基準時刻データ群生成部61は、例えば、顧客DB2に記憶されている顧客データに属する個別データのうち、 $CTime[t]$ に対応する閾値推定前時刻($CTime[t] - \Delta t$)で商品またはサービスの利用を開始していた顧客の個別データを除外して基準時刻データを生成する。すなわち、利用開始情報が閾値推定前時刻以前の時刻となっている個別データを顧客データから除外し、残りの個別データに対して正または負のラベル付けを行って、基準時刻データを生成する。以下、閾値推定前時刻で利用開始していた顧客の個別データを除外して基準時刻データを生成する場合を例にして説明する。ステップC3の基準時刻データ生成処理は、第1の実施形態における現時刻データ生成処理と同様の処理である。

40

【0187】

50

ステップC3の次に、閾値推定前時刻データ群62は、変数 t により定まる閾値推定前時刻($CTime[t] - t$)と、顧客DB2に記憶された顧客データとに基づいて、閾値推定前時刻データを生成する(ステップC4)。閾値推定前時刻データ群62は、顧客DB2に記憶された顧客データを読み込み、閾値推定前時刻以前で商品またはサービスの利用を開始している顧客の個別データに正をラベル付け、閾値推定前時刻で商品またはサービスの利用を開始していない顧客の個別データに負をラベル付けることにより、閾値推定前時刻データを生成する。すなわち、使用開始情報が閾値推定前時刻以前の時刻となっている個別データに正をラベル付け、利用開始情報が前時刻以前の時刻でない個別データに負をラベル付けて、閾値推定前時刻データを生成する。

【0188】

10

次に、閾値推定前時刻データ群62は、その閾値推定前時刻データを用いて普及率を計算する(ステップC5)。普及率は、顧客数に対する、閾値推定前時刻までに商品またはサービスの利用を開始した顧客数の割合であり、閾値推定前時刻データ中の各個別データは各顧客に対応している。従って、閾値推定前時刻データ中の全個別データ数に対する、閾値推定前時刻データ中の正がラベル付けられた個別データの数の割合が普及率となる。閾値推定前時刻データ群62は、閾値推定前時刻データ中の正がラベル付けられた個別データの数を、閾値推定前時刻データ中の全個別データ数で除算して普及率を求めればよい。この普及率を変数 t を用いて $D[t]$ と記す。閾値推定前時刻データ群62は、計算した普及率 $D[t]$ を記憶しておく。

【0189】

20

ステップC5の後、閾値推定前時刻分類器群生成部63は、閾値推定前時刻データを用いて、分類器を生成する(ステップC6)。例えば、分類器生成部63は、重回帰分析、回帰木、ニューラルネットワーク、サポートベクタマシン、ベイジアンネットワーク、決定木のアンサンブル学習などの出力が連続値であるデータマイニング手法のいずれかあるいはその組合せにより分類器を生成すればよい。

【0190】

閾値推定前時刻分類器群生成部63は、例えば、アンサンブル学習の一種であるバギングを決定木に適用して、複数の決定木を生成し、その複数の決定木の組を分類器としてもよい。この場合の分類器作成処理の例を説明する。閾値推定前時刻分類器群生成部63は、閾値推定前時刻データから重複を許して一部の個別データを取り出して、その個別データの集合を用いて決定木を生成する。この処理を複数回繰り返すことにより、複数の決定木を生成し、その複数の決定木の組み合わせを一つの分類器とする。例えば、閾値推定前時刻データから取り出す個別データ数を K 個とすると、閾値推定前時刻分類器群生成部63は、閾値推定前時刻データから重複を許して K 個の個別データを取り出し、 K 個の個別データから決定木を生成することを繰り返せばよい。例えば、この繰り返し処理を500回行えば、500個の決定木が作成される。

30

【0191】

以下の説明では、上記のようにバギングにより複数の決定木を生成し、その複数の決定木を分類器とする場合を例にする。

【0192】

40

取り出した個別データから決定木を生成する処理は、第1の実施形態で前時刻データ群生成部32の分類器生成動作として例示した処理と同様である。すなわち、閾値推定前時刻分類器群生成部63は、どの項目(顧客の属性を表す項目)で最初に分割するのかを決定する。このとき、項目1~ N の各項目について、分割時の評価値を計算し、その評価値が最大の項目を、分割に最も適した項目として選択すればよい。評価値として、例えば、分割前のノードのエントロピーと、分割後のエントロピーの差を用いればよい。閾値推定前時刻分類器群生成部63は、分割後の各ノードについても、上記と同様の処理を行い、次にどの項目で分割するのかを決定する処理を順次繰り返し、所定の条件が満たされたときには、ノードの分割を停止する。閾値推定前時刻分類器群生成部63は、このようにして得た木構造の決定木に対して、枝刈りを行うことにより、決定木を生成する。上記の所

50

定の条件や枝刈りの処理も、第1の実施形態で示した所定の条件や枝刈り処理と同様である。

【0193】

また、閾値推定前時刻データから取り出すK個の個別データの組み合わせが異なることにより、生成される決定木も異なる。個別データに対して上記のように得られた複数の決定木を適用し、その個別データのスコア（ラベルが正となる確からしさ）を求める場合、個々の決定木毎に、個別データに対するラベルの判定を行い、正（利用する）と判定された回数を、判定回数で除算することによってスコアを求める。例えば、分類器として500個の決定木を生成した場合、個別データに対して、それぞれの決定木を用いてラベル判定を行う。その結果、正と判定された回数をX回とすると、スコアとして $X / 500$ を求める。

10

【0194】

ステップC6において分類器が生成されると、誤差群算出部64は、誤差の最小値を示す変数ErrorMinの初期値を設定し、変数 i, p に対して、 $i = 1, p = 0$ という初期値を設定する（ステップC7）。ErrorMinの初期値は、閾値算出部64が計算する誤差のとり得る値の最大値、あるいは、その誤差のとり得る値に比べて十分に大きな値であればよい。変数 p は、最小の誤差に対応する仮閾値を指定するための変数である。例えば、最小語の誤差に対応する仮閾値が p 個あったとすると、その最小誤差に対応する仮閾値の1番目から p 番目までをそれぞれ、 $Thres[1], \dots, Thres[p]$ とする。

20

【0195】

ステップC7の次に、誤差群算出部64は、仮閾値の番号 i が仮閾値の数 V 以下であるか否かを判定する（ステップC8）。 $i \leq V$ ならば（ステップC8のYes）、ステップC9に移行し、 $i > V$ ならば（ステップC8のNo）、ステップC16に移行する。

【0196】

ステップC9では、誤差群算出部64は、ステップC6で生成された閾値推定前時刻分類器と、仮閾値FalseThr $[i]$ とを用いて、ステップC3で生成された基準時刻データ内の各個別データのラベルを予測する（ステップC9）。本例では、誤差群算出部64は、基準時刻データ内の各個別データ毎に、その個別データにおける顧客の属性を示す項目と、分類器として生成されたそれぞれの決定木とを照合し、その個別データに対するラベルを判定する。決定木は複数存在するので、誤差群算出部64は、この判定を決定木毎に行い、その判定回数（決定木数）に対する正と判定された回数の割合を個別データのスコアとして求める。すなわち、誤差群算出部64は、一つの個別データに関し、「正と判定された回数 / 判定回数」を計算し、その値をスコアとする。さらに、誤差群算出部64は、変数 i により定まる仮閾値FalseThr $[i]$ とそのスコアとを比較し、スコアがFalseThr $[i]$ 以上であれば、個別データに対するラベルが正であると予測し、スコアがFalseThr $[i]$ 未満であれば、個別データに対するラベルが負であると予測する。

30

【0197】

次に、誤差群算出部64は、基準時刻データ内の各個別データにラベル付けられるラベルの予測結果と、基準時刻データとの誤差を算出する（ステップC10）。すなわち、基準時刻データ内の各個別データには、基準時刻データ生成時に正または負のラベルがラベル付けられているので、ステップC9での予測結果と、実際の基準時刻データでラベル付けられているラベルとの誤差を予測する。この誤差をErrorと記す。

40

【0198】

誤差群算出部64は、ステップC10において、例えば、基準時刻データ内の個別データ毎に予測したラベルと、実際に基準時刻データの各個別データにラベル付けられているラベルとを比較し、両者が異なっている個別データ数をErrorとしてもよい。

【0199】

あるいは、誤差群算出部64は、基準時刻データ内の個別データのうちステップC9で

50

正がラベル付けられると予測した個別データの数と、ステップC3で生成された基準時刻データにおいて実際に正がラベル付けられた個別データの数との差の絶対値をErrorとしてもよい。

【0200】

ステップC10の後、関数推定部65は、ステップC10で計算された誤差Errorと最小誤差値ErrorMinとを比較し、ErrorがErrorMin未満であるかを判定する(ステップC11)。ErrorがErrorMin未満である場合(ステップC11におけるYes)、関数推定部65は、pを0に初期化し、ErrorMinにErrorを代入する(ステップC13)。また、ステップC13において、関数推定部65は、変数Sumを0に初期化する。なお、変数Sumは、誤差が最小となる仮閾値の和を格納するための変数であり、その仮閾値の平均値を計算するために用いられる。ErrorがErrorMin未満ということは、これまで最小としていた誤差よりもさらに小さい最小値が見つかったことを意味する。この場合、ステップC13において、その最小値でErrorMinを更新し、また、その誤差に対応する仮閾値が複数ある場合に各仮閾値を個別に指定するための変数pを初期化している。

10

【0201】

ステップC13の後、関数推定部65は、pの値を1インクリメントし、Sumの値に仮閾値FalseThr[i]を加算することによって、Sumを更新する(ステップC14)。すなわち、Sum+FalseThr[i]の値を新たなSumの値とするようにSumを更新する。また、ErrorがErrorMin以上であると判定した場合(ステップC11におけるNo)、関数推定部65は、ErrorがErrorMinと等しいか否かを判定する(ステップC12)。ここで等しいと判定した場合(ステップC12におけるYes)にも、関数推定部65は、pの値を1インクリメントし、Sumの値にFalseThr[i]を加算してSumを更新する(ステップC14)。

20

【0202】

ステップC13からステップC14に移行した場合、p=1であり、現在着目している仮閾値FalseThr[i]を、誤差が最小となる1番目の仮閾値とする。また、ステップC12でError=ErrorMinとなり、ステップC14に移行した場合、pは2以上の値となる。この場合、既に最小の誤差に対応する仮閾値は1つ以上見つかり、関数推定部65は、現在着目している仮閾値FalseThr[i]を、誤差が最小となるp番目の仮閾値として定めることになる。

30

【0203】

ステップC14の後、あるいは、ステップC12でErrorがErrorMinと等しくない判定した場合、関数推定部65は、仮閾値を示す変数iをインクリメントする(ステップC15)。その後、ステップC15でインクリメントされたiを用いて、ステップC8移行の処理を繰り返す。

【0204】

また、ステップC8で $i > V$ と判定された場合、関数推定部65は、tに対応する閾値(Thres[t]と記す。)の値を、Sum/pとする(ステップC16)。すなわち、最小誤差に対応する仮閾値の合計となるSumをその仮閾値の個数pで除算し、平均値を閾値とする。最小誤差に対応する仮閾値が1つしかなければ、Sumはその仮閾値の値となっているので、その仮閾値をThres[t]とすることになる。

40

【0205】

ステップC16の後、関数推定部65は、基準時刻を順番に指定するための変数tをインクリメントする(ステップC17)。ステップC17の後、インクリメントされたtを用いて、ステップC2以降の処理を繰り返す。

【0206】

ステップC2で $t > U$ と判定された場合、関数推定部65は、各閾値推定前時刻毎に求めた普及率D[t]および閾値Thres[t]から、普及率に対する閾値の関数である閾値関数を求める(ステップC18)。ステップC18に移行した時点で、各tにおける

50

閾値推定間時刻が計算され、その閾値推定前時刻に対応する普及率 $D[t]$ および閾値 $Thres[t]$ が求められている。すなわち、普及率 $D[t]$ および閾値 $Thres[t]$ の組が U 組求められている。関数推定部 65 は、この普及率 $D[t]$ および閾値 $Thres[t]$ の関係を曲線近似して、閾値関数を求めればよい。例えば、 $Thres[t]$ が $D[t]$ を用いて、以下の式 (1) で表されるとみなし、式 (1) のパラメタ a, b を推定すればよい。

【0207】

$$Thres[t] = a \cdot D[t] + b \quad \text{式(1)}$$

【0208】

式 (1) において、 a, b は推定すべきパラメタである。普及率 $D[t]$ および閾値 $Thres[t]$ の組み合わせが U 個求められており、関数推定部 65 は、例えば、その U 個の組み合わせ (閾値および普及率の組み合わせ) を用いて、最小二乗法によってパラメタ a, b を計算してもよい。なお、パラメタ a の値は、商品またはサービスの普及率が時間の経過と共に急激に上昇するなら低く、普及率の上昇が緩やかなら高くなる。 a は、例えば、0.5 などの値となる。パラメタ b は、普及開始段階での普及率が急激に上昇するなら低く、普及率がほとんど上昇しないなら低くなる値であり、0 以上 1 以下で、例えば 0.05 等の小さな値となる。

【0209】

また、式 (1) は、パラメタが定まっていない閾値関数の例であり、推定すべきパラメタを含む閾値関数は式 (1) に限定されない。式 (1) に例示する推定すべきパラメタを含む閾値関数は、普及予測システムのユーザに入力されてもよい。あるいは、予め定められていてもよい。

【0210】

関数推定部 65 は、例えば最小二乗法などによりパラメタ a, b を計算することによって、閾値関数を決定する。関数推定部 65 は、閾値関数を普及予測装置 7 (テストデータラベル判定部 74) に入力する。

【0211】

図 21 は、本実施形態における普及予測装置 7 の処理経過の例を示すフローチャートである。普及予測装置 7 は、例えば、関数推定部 65 から閾値関数が入力され、また、キーボードなどの入力装置を介して現時刻が入力されると、以下のように動作する。

【0212】

テストデータ生成部 71 は、テストデータを生成する (ステップ D1)。この動作は、第 2 の実施形態におけるステップ B1 と同様である。ここでは、現時刻で商品またはサービスの利用を開始していない顧客の個別データのみを含むテストデータを生成する場合を例にして説明する。テストデータ生成部 71 は、顧客 DB 2 に記憶されている顧客データの中から、現時刻で商品またはサービスの利用を開始していない顧客の個別データのみ (すなわち、利用開始情報が現時刻以前の時刻でない個別データのみ) を読み込み、その個別データの集合をテストデータとする。

【0213】

次に、ラベル付けデータ生成部 72 は、ラベル付けデータを生成する (ステップ D2)。ラベル付けデータ生成部 72 は、顧客 DB 2 に記憶された顧客データを読み込み、現時刻以前で商品またはサービスの利用を開始している顧客の個別データに正をラベル付け、現時刻で商品またはサービスの利用を開始していない顧客の個別データに負をラベル付けることにより、ラベル付けデータを生成する。

【0214】

次に、ラベル付けデータ生成部 72 は、そのラベル付けデータから普及率を計算する (ステップ D3)。ラベル付けデータ中の各個別データは各顧客に対応しているので、ラベル付けデータ中の全個別データに対する、ラベル付けデータ中の正がラベル付けられた個別データの数の割合が普及率となる。ラベル付けデータ生成部 72 は、ラベル付けデータ中の正がラベル付けられた個別データの数を、ラベル付けデータ中の全個別データ数で除

10

20

30

40

50

算して普及率を求めればよい。

【0215】

次に、分類器生成部73は、ラベル付けデータを用いて分類器を生成する(ステップD4)。この分類器は、閾値関数推定装置6の閾値推定前時刻分類器群生成部63が生成する分類器と同様である。本例では、バギングにより複数の決定木を生成し、その複数の決定木を分類器とするものとする。すなわち、分類器生成部73は、ラベル付けデータから重複を許して一部の個別データを取り出して、その個別データの集合を用いて決定木を生成する処理を繰り返して、複数の決定木を生成する。ラベル付けデータから取り出す個別データの数をK個とすると、分類器生成部73は、ラベル付けデータから重複を許してK個の個別データを取り出し、K個の個別データから決定木を生成することを繰り返せばよい。

10

【0216】

取り出した個別データから決定木を生成する処理は、第1の実施形態で前時刻データ群生成部32の分類器生成動作として例示した処理と同様である。分類器生成部73は、どの項目(顧客の属性を表す項目)で最初に分割するのかを決定する。このとき、項目1~Nの各項目について、分割時の評価値を計算し、その評価値が最大の項目を、分割に最も適した項目として選択すればよい。評価値として、例えば、分割前のノードのエントロピーと、分割後のエントロピーの差を用いればよい。分類器生成部73は、分割後の各ノードについても、上記と同様の処理を行い、次にどの項目で分割するのかを決定する処理を順次繰り返し、所定の条件が満たされたときには、ノードの分割を停止する。また、分類器生成部73は、このようにして得た木構造の決定木に対して、枝刈りを行うことにより、決定木を生成する。上記の所定の条件や枝刈りの処理も、第1の実施形態で示した所定の条件や枝刈り処理と同様である。

20

【0217】

テストデータラベル判定部74は、関数推定部65が推定した閾値関数に、ステップD3で算出した普及率を代入して、閾値を計算する(ステップD5)。

【0218】

次に、テストデータラベル判定部74は、ステップD4で生成された分類器およびステップD5で計算された閾値を用いてテストデータのラベルを予測する(ステップD6)。テストデータラベル判定部74は、ステップD6において、テストデータ中の各個別データ毎に以下の処理を行う。テストデータラベル判定部74は、個別データにおける顧客の属性を示す項目と、分類器として生成されたそれぞれの決定木とを照合し、その個別データに対するラベルを判定する。決定木は複数存在するので、テストデータラベル判定部74は、この判定を決定木毎に行い、その判定回数(決定木数)に対する正と判定された回数の割合を個別データのスコアとして求める。すなわち、テストデータラベル判定部74は、一つの個別データに関し、「正と判定された回数/判定回数」を計算し、その値をスコアとする。さらに、テストデータラベル判定部74は、ステップD5で計算された閾値とそのスコアとを比較し、スコアが閾値以上であれば、その個別データに対するラベルが正であると判定し、スコアが閾値未満であれば、その個別データに対するラベルが負であると判定する。

30

40

【0219】

テストデータラベル判定部74は、テストデータ内の個別データ毎に判定した正または負のラベルを、ディスプレイ装置などに表示させたり、印刷装置に印刷させたりしてもよい。

【0220】

本実施形態によれば、第2の実施形態と同様に、顧客が商品やサービスを利用し始める時期を顧客毎に予測することができる。また、本実施形態では、閾値関数を求めるときには、入力された複数の基準時刻毎に、また、複数の仮閾値毎にそれぞれ予測誤差(Error)を求め、誤差が最小となる閾値を求めて、閾値関数を推定するので、刻々と変動する受容性を閾値関数として推定することができる。また、重み付けを行う場合には、重み

50

付けを行う毎に分類器を生成することになるが、本実施形態では、重み付けは行わないので、各仮閾値毎に誤差 (Error) を算出するときに、仮閾値毎に分類器を生成する必要はなく、一つの分類器を用いてそれぞれの仮閾値における誤差を求めることができる。よって、分類器生成のための計算時間を短縮することができる。

【0221】

次に、第4の実施形態の変形例を示す。以下に示す第4の実施形態の変形例では、第2の実施形態の変形例(図14、図15参照)と同様に、現時刻を更新してテストデータに対する予測を繰り返す。

【0222】

図22は、第4の実施形態の変形例を示すブロック図である。既に説明した構成要素と同様の構成要素は、図18と同一の符号を付し、詳細な説明を省略する。普及予測装置7は、テストデータ生成部71と、ラベル付けデータ生成部72と、分類器生成部73と、テストデータラベル判定部74と、現時刻更新部75とを備える。現時刻更新部75は、第2の実施形態の変形例で示した現時刻更新部55と同様である。

10

【0223】

本変形例では、テストデータラベル判定部74は、ステップD6での判定結果が実際の利用開始情報と異なっている個別データの利用開始情報を上書きして更新する。すなわち、個別データの利用開始情報が未利用を示していて、ステップD6で正と判定した場合、テストデータラベル判定部74は、顧客DB2に記憶されている顧客データ内におけるその個別データの利用開始情報を現時刻に一定の時間(Tとする。)分を加算した時刻に更新する。

20

【0224】

現時刻更新部75は、テストデータラベル判定部74がテストデータに対する判定を行った後、現時刻を一定の時間(T)分増加した時刻を新たな現時刻とするように、現時刻を更新する。現時刻更新部75も、例えば、普及予測プログラムに従って動作するCPUによって実現される。

【0225】

テストデータ生成部71、ラベル付けデータ生成部72、分類器生成部73、テストデータラベル判定部74は、現時刻が更新されると、更新後の現時刻および更新後の顧客データを用いて、ステップD1以降の処理を繰り返す。

30

【0226】

なお、この繰り返し処理の終了条件は、例えば、第2の実施形態の変形例と同様に、更新後の現時刻が、定められた期間(予測期間)を越えることとすればよい。現時刻更新部75は、更新後の現時刻が予測期間内であることを条件に現時刻の更新を行えばよい。予測期間は、例えば、普及予測システムのユーザによって予め入力されればよい。

【0227】

現時刻更新部75は、更新後の現時刻が予測期間を越えると判定した場合、予測結果を出力装置(図22において図示せず。)に出力させてもよい。この出力態様は、図14に例示する第2の実施形態の変形例と同様である。例えば、普及曲線を出力してもよい。この場合、ステップD1に移行する毎に、現時刻更新部75(他の構成要素でもよい)が、現時刻において商品またはサービスの利用を開始している顧客の個別データ数を顧客DBの中からカウントし、そのカウント値と現時刻とを対応付けて記憶していけばよい。そして、現時刻更新部75は、現時刻を横軸にとり、カウント値を縦軸にとって、現時刻の変化に伴うカウント値の推移を示すグラフを作成して出力すればよい。

40

【0228】

なお、本変形例において、顧客DB2に記憶されている個別データの利用開始情報を上書きするのではなく、個別データに当初から含まれている利用開始情報のコピーを利用開始時期予測情報としてコピーしておき、そのコピーを更新してもよい。

【0229】

図23は、第4の実施形態の他の変形例を示すブロック図である。図18および図22

50

では、普及予測システムが閾値関数推定装置 6 を備える場合を説明したが、図 2 3 に示すように、閾値関数推定装置 6 を備えていない構成であってもよい。図 2 3 に示すテストデータ生成部 7 1、ラベル付けデータ生成部 7 2、分類器生成部 7 3、テストデータラベル判定部 7 4 の動作は、既に説明した動作と同様である。閾値関数推定装置 6 を備えない構成では、普及予測システムの外部から閾値関数が、テストデータラベル判定部 7 4 に入力されればよい。例えば、普及予測システムのユーザが、閾値関数を入力してもよい。本変形例においても、閾値関数推定装置 6 は現時刻更新部 7 5 を備えてよい。

【実施例】

【0230】

本実施例では、性別、年齢、既存の製品の購買履歴、イノベータ特性、接触メディアを、顧客の特徴を示す項目とし、それらの項目とともに利用開始時期を含む顧客データを顧客 DB 2 に予め記憶させた。特定製品の発売時期から 4 年間の間にその製品の利用を開始した顧客の個別データに関しては、顧客がその製品の利用開始時期を利用開始情報として定めた。その 4 年間の間にまだ製品の利用を開始していない顧客の個別データに関しては、利用開始情報として未利用を意味する「？」を定めた。利用開始情報として購買時期（利用開始時期）が記述された顧客は顧客 DB 2 に記憶された顧客数全体の 8.4% であり、利用開始情報として未利用を意味する「？」が記述された顧客は顧客 DB 2 に記憶された顧客数全体の 91.6% であった。また、各顧客毎に、顧客の特徴を示す上記の項目を定めた。項目の値が不明な場合には、その項目の値として、不明を意味する「？」を記述した。

【0231】

以上のような顧客データを顧客 DB 2 に予め記憶させ、第 1 の実施形態の影響度推定システムに適用した。その結果、影響度推定システムが影響度を推定した結果、影響度は 0.191 であった。

【0232】

また、この影響度 0.191 を、第 2 の実施形態の普及予測システム（図 1 4）に適用した。普及予測システムが出力した普及曲線の例を図 2 4 に示す。発売開始後 4 年目の普及率は 8.4% であり、このときの関数値は、 $0.191 \times (100 - 8.4) / 8.4$ であり、4 年目を現時刻とする予測では利用開始した顧客は未利用の顧客の 2 倍に重み付けた。以降の時刻を現時刻とするときにも影響度は 0.191 で一定としたが、関数値は普及率の増加に伴い減少し、重み付けも少なくなった。

【0233】

また、普及予測システムによって得られた普及曲線を検証するため、製品発売後 4 年目から 6 年目までの 2 年間の普及実績も調査した。

【0234】

図 2 4 に示す実線は、普及予測システムが出力した普及曲線である。また、+ でプロットした曲線は、製品発売後 4 年目までの普及の実績を示している。四角形でプロットした曲線は、製品発売後 4 年目から 6 年目までの 2 年間の普及の実績を示している。図 2 4 に示すように、普及予測システムが出力した普及曲線（図 2 4 に示す実線）は、普及の実績（図 2 4 に示す四角形のプロット）に、極めて近かった。

【産業上の利用可能性】

【0235】

本発明は、例えば、個別の顧客毎に商品やサービスの利用開始時期を予測する普及予測システムに好適に適用される。また、指定した時点での影響度を推定する影響度推定システムに好適に適用される。

【図面の簡単な説明】

【0236】

【図 1】本発明の影響度推定システムの例を示すブロック図である。

【図 2】入力装置および出力装置を備えた影響度推定システムの例を示すブロック図である。

10

20

30

40

50

【図3】本発明の影響度推定システムの処理経過の例を示すフローチャートである。

【図4】顧客データの例を示す説明図である。

【図5】現時刻データの例を示す説明図である。

【図6】前時刻データ生成過程において重み付けが行われる前のデータの例を示す説明図である。

【図7】前時刻データの例を示す説明図である。

【図8】既知の独立変数および従属変数の組み合わせの例を示す説明図である。

【図9】分類器の例を示す説明図である。

【図10】分類器から予測される従属変数の例を示す説明図である。

【図11】現時刻毎に求めた各仮影響度における誤差を示す説明図である。

10

【図12】本発明の普及予測システムの例を示すブロック図である。

【図13】本発明の普及予測システムの処理経過の例を示すフローチャートである。

【図14】第2の実施形態の変形例を示すブロック図である。

【図15】第2の実施形態の変形例における処理経過の例を示すフローチャートである。

【図16】第3の実施形態の普及予測システムの例を示すブロック図である。

【図17】第3の実施形態の普及予測システムを詳細に示すブロック図である。

【図18】第4の実施形態の普及予測システムの例を示すブロック図である。

【図19】第4の実施形態における閾値関数推定装置の処理経過の例を示すフローチャートである。

【図20】第4の実施形態における閾値関数推定装置の処理経過の例を示すフローチャートである。

20

【図21】第4の実施形態における普及予測装置の処理経過の例を示すフローチャートである。

【図22】第4の実施形態の変形例を示すブロック図である。

【図23】第4の実施形態の他の変形例を示すブロック図である。

【図24】普及予測システムが出力した普及曲線の例を示すグラフである。

【符号の説明】

【0237】

2 顧客データベース

3 1 現時刻データ生成部

3 2 前時刻データ群生成部

3 3 前時刻分類器群生成部

3 4 誤差群算出部

3 5 影響度算出部

5 1 テストデータ生成部

5 2 学習データ生成部

5 3 分類器生成部

5 4 テストデータラベル判定部

5 5 現時刻更新部

6 1 基準時刻データ群生成部

6 2 閾値推定前時刻データ群生成部

6 3 閾値推定前時刻分類器群生成部

6 4 誤差群算出部

6 5 関数推定部

7 1 テストデータ生成部

7 2 ラベル付けデータ生成部

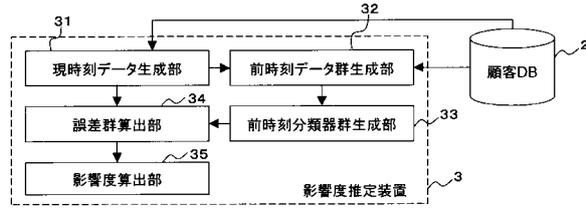
7 3 分類器生成部

7 4 テストデータラベル判定部

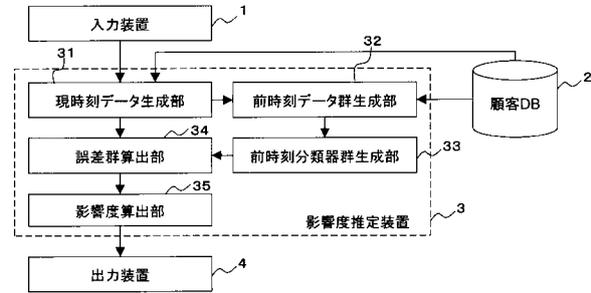
30

40

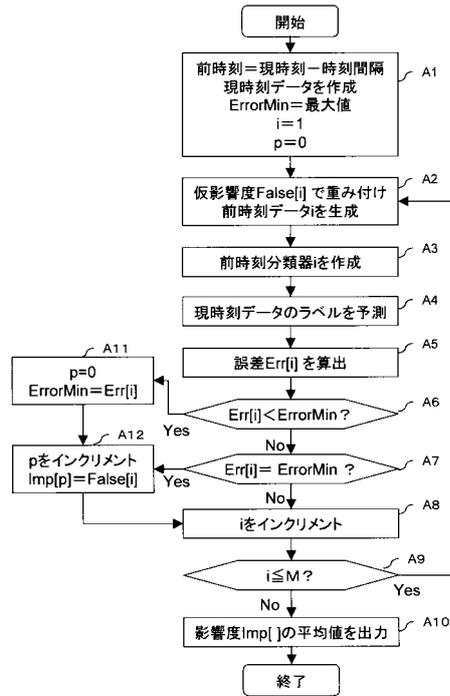
【図1】



【図2】



【図3】



【図4】

顧客データ

ID	項目1	...	項目N	利用開始情報
1	男性		1	2007年1月
2	女性		2	2007年3月
3	男性		1	2007年5月
4	男性		?	2007年5月
:				
M-1	女性		3	?
M	女性		3	?

【図6】

重み付け前の前時刻データ(2007年2月)

ID	項目1	...	項目N	クラス
1	0		1	+
2	1		2	-
3	0		1	-
4	0		?	-
:				
M-1	1		3	-
M	1		3	-

【図5】

現時刻データ (現時刻:2007年4月)

ID	属性1	...	属性N	クラス
2	1		2	+
3	0		1	-
4	0		?	-
:				
M-1	1		3	-
M	1		3	-

【図7】

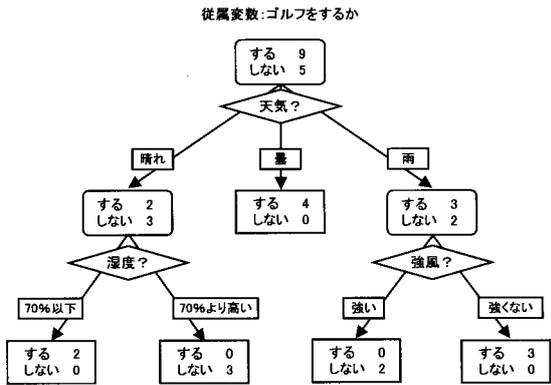
5倍に重み付けた前時刻データ(2007年2月)

ID	項目1	...	項目N	クラス
1	0		1	+
1'	0		1	+
1''	0		1	+
1'''	0		1	+
1''''	0		1	+
2	1		2	-
3	0		1	-
4	0		?	-
:				
M-1	1		3	-
M	1		3	-

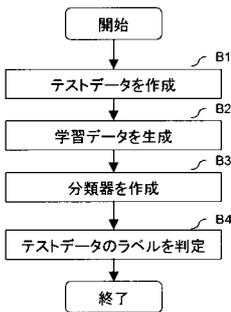
【図8】

独立変数				従属変数
天気	気温(度)	湿度(%)	風が強いか	ゴルフをするか
晴れ	29	85	強くない	しない
晴れ	27	90	強い	しない
曇	28	78	強くない	する
雨	21	96	強くない	する
雨	20	80	強くない	する
雨	18	70	強い	しない
曇	18	65	強い	する
晴れ	22	95	強くない	しない
晴れ	21	70	強くない	する
雨	24	80	強くない	する
晴れ	24	70	強い	する
曇	22	90	強い	する
曇	27	75	強くない	する
雨	22	80	強い	しない

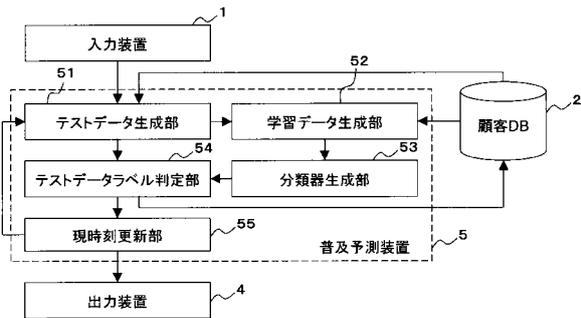
【図9】



【図13】



【図14】



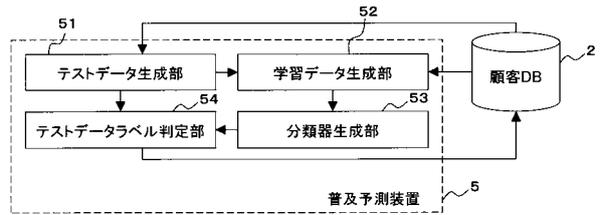
【図10】

独立変数				従属変数
天気	気温(度)	湿度(%)	風が強いか	ゴルフをするか
晴れ	18	85	強くない	しない
晴れ	20	90	強い	しない

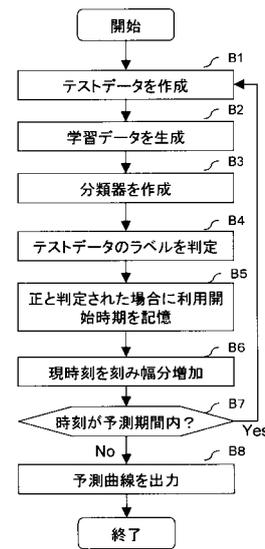
【図11】

i	現時刻			
	T1	T2	...	Tx
1	Err[1]	Err[1]		Err[1]
2	Err[2]	Err[2]		Err[2]
3	Err[3]	Err[3]		Err[3]
⋮				
M	Err[M]	Err[M]		Err[M]

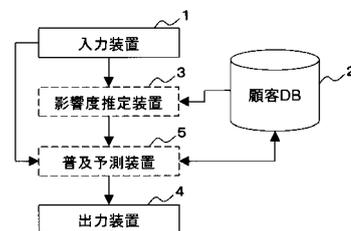
【図12】



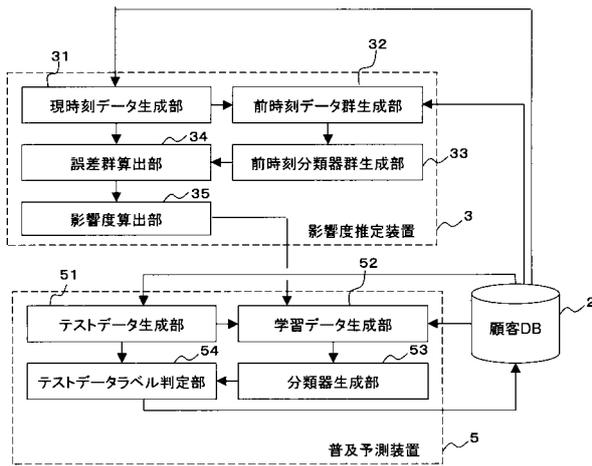
【図15】



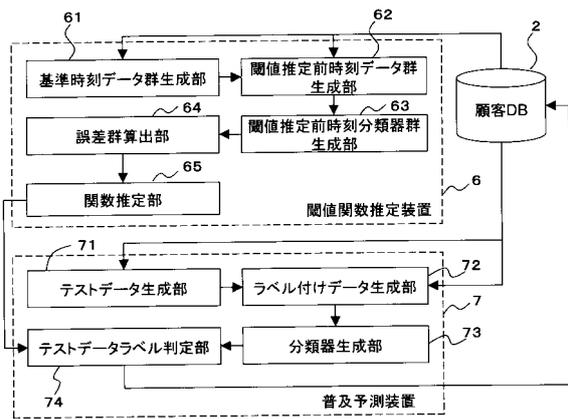
【図16】



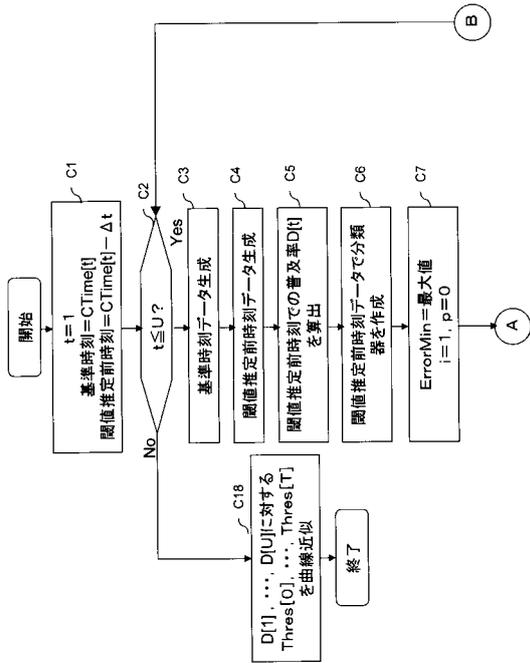
【図17】



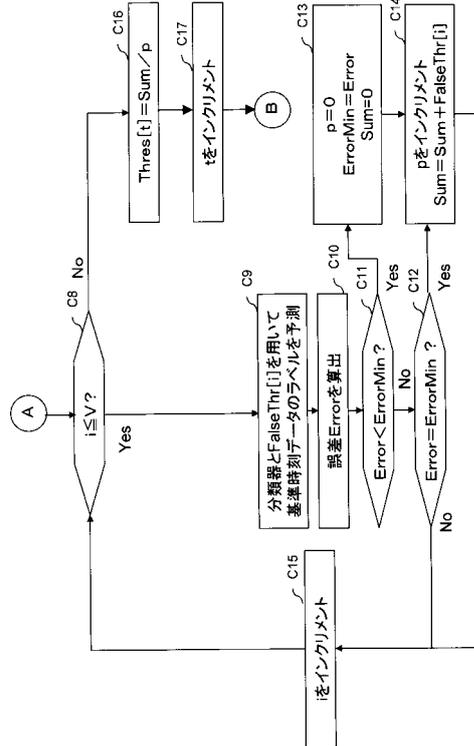
【図18】



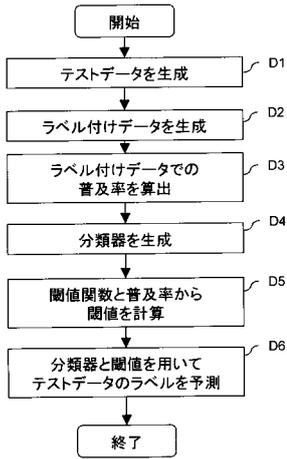
【図19】



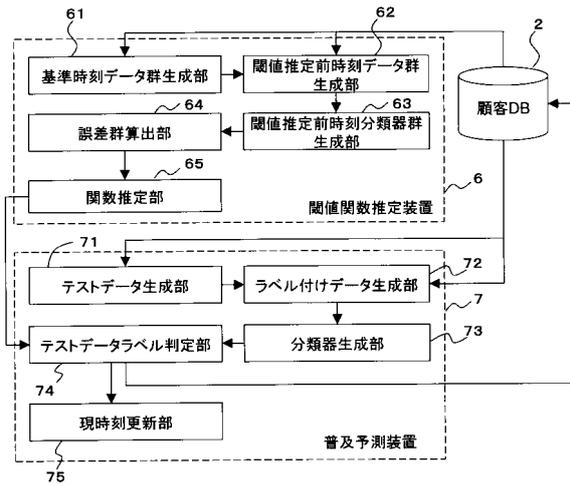
【図20】



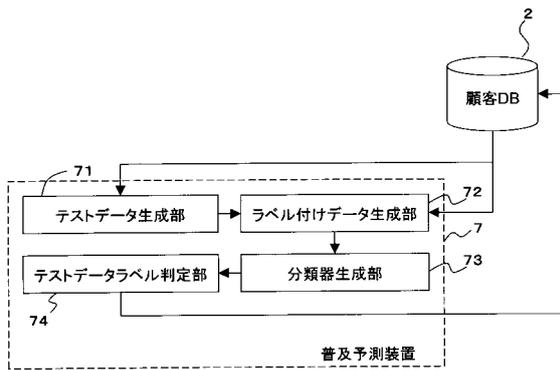
【図 2 1】



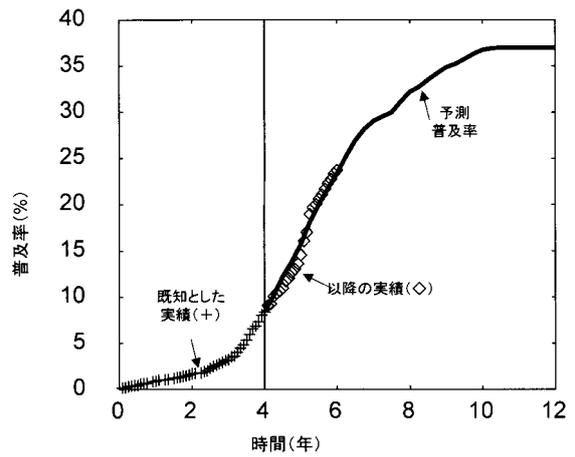
【図 2 2】



【図 2 3】



【図 2 4】



フロントページの続き

(56)参考文献 特開2005-301389(JP,A)
特開2008-299542(JP,A)
特開2004-70525(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06Q10/00
30/00
50/00