

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4888038号
(P4888038)

(45) 発行日 平成24年2月29日 (2012.2.29)

(24) 登録日 平成23年12月22日 (2011.12.22)

(51) Int.Cl.		F I		
G06F	19/00	(2011.01)	G06F	19/00 130
G06Q	30/02	(2012.01)	G06F	17/60 170Z
G06Q	10/00	(2012.01)	G06F	17/60 504

請求項の数 6 (全 38 頁)

(21) 出願番号	特願2006-282896 (P2006-282896)	(73) 特許権者	000005223 富士通株式会社
(22) 出願日	平成18年10月17日 (2006.10.17)		神奈川県川崎市中原区上小田中4丁目1番1号
(65) 公開番号	特開2008-102621 (P2008-102621A)	(74) 代理人	110000040 特許業務法人池内・佐藤アンドパートナーズ
(43) 公開日	平成20年5月1日 (2008.5.1)	(72) 発明者	渡辺 理 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内
審査請求日	平成21年7月9日 (2009.7.9)	(72) 発明者	神田 陽治 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

最終頁に続く

(54) 【発明の名称】 時間種特性生成システム、時間種特性生成方法および時間種特性生成プログラム

(57) 【特許請求の範囲】

【請求項1】

ネットワークに接続されたコンピュータにより提供される少なくとも1つのサイトの利用状況を示すログデータから、所定期間における前記サイトの利用度の変動を示す遷移データを抽出する利用状況把握部と、

前記遷移データに基づいて、前記所定期間内の前記サイトの利用度の変動を、前記サイトの利用度を示す値の集合Yで表す変動データを生成する変動データ生成部と、

複数の時間種を示すデータと、前記複数の時間種それぞれで特定される期間を表す時間種因子値とが記録された時間種データ記録部と、

時間種データ記録部に記録された複数の時間種から、予め記録された手順に従って時間種を抽出し、前記抽出した時間種に対応するi個の時間種因子値の集合D_iを生成する時間種因子抽出部と、

前記集合D_iを説明変数として、集合Yを被説明変数とする下記(数1)において、誤差bの二乗が最小になるようなi個の係数a_iを計算することによって、前記集合D_iに対応する時間種および前記集合D_iそれぞれの前記係数a_iを含むデータを、前記サイトの利用状況の時間種特性を示すデータとして生成する時間種特性生成部とを備え、

【数1】

$$Y = \sum_i a_i D_i + b$$

10

20

前記変動データ生成部は、前記遷移データが示す利用度の変動を、利用度の多変量データで表し、当該多変量データを主成分分析することにより、複数の特徴基底を抽出し、抽出した前記複数の特徴基底それぞれの主成分得点を集合 Y として、前記複数の特徴基底ごとに集合 Y を生成し、

前記時間種因子抽出部は、前記複数の特徴基底それぞれについて、前記集合 D_i を生成し、

前記時間種特性生成部は、前記複数の特徴基底ごとの集合 Y について、それぞれの前記集合 D_i を用いて、係数 a_i を計算し、前記複数の特徴基底ごとの集合 D_i および係数 a_i を合成することにより、前記サイトの利用状況の時間種特性を示すデータとして生成する、時間種特性生成システム。

10

【請求項 2】

前記利用状況把握部は、少なくとも 1 つのサイトが提供する特定の広告にアクセスし、かつ当該サイトまたは他のサイトが提供する前記特定の広告に関するサービスを利用する行動の発生数の変動を表す遷移データを抽出する、請求項 1 に記載の時間種特性生成システム。

【請求項 3】

利用者の属性を示すデータと、当該属性を持つ利用者による利用状況の時間種特性を示すデータとを対応付けて記録する利用者時間種特性記録部と、

前記時間種特性生成部が生成した時間種特性を示すデータと、前記利用者時間種特性記録部に記録された時間種特性を示すデータとを比較することにより、前記サイトを利用する利用者の属性を特定する利用者解析部とをさらに備える、請求項 1 に記載の時間種特性生成システム。

20

【請求項 4】

前記変動データ生成部は、前記遷移データが示す利用度の変動を、利用度の多変量データで表し、当該多変量データを主成分分析することにより、特徴基底を抽出して当該特徴基底の主成分得点を集合 Y とし、

前記時間種因子抽出部は、前記時間種データ記録部に記録された複数の時間種から抽出した時間種に対応する時間種因子値の多変量データを生成し、当該多変量データを主成分分析することにより、特徴基底を抽出して当該特徴基底の主成分得点を集合 D_i とする、請求項 1 に記載の時間種特性生成システム。

30

【請求項 5】

コンピュータが、

ネットワークに接続されたサーバにより提供される少なくとも 1 つのサイトの利用状況を示すログデータから、所定期間における前記サイトの利用度の変動を示す遷移データを抽出し、

前記遷移データに基づいて、前記所定期間内の前記サイトの利用度の変動を、前記サイトの利用度を示す値の集合 Y で表す変動データを生成し、

複数の時間種を示すデータと、前記複数の時間種それぞれで特定される期間を表す時間種因子値とが記録された時間種データ記録部に記録された複数の時間種から、予め記録された手順に従って時間種を抽出し、前記抽出した時間種に対応する i 個の時間種因子値の集合 D_i を生成し、

40

前記集合 D_i を説明変数として、集合 Y を被説明変数とする下記（数 1）において、誤差 b の二乗が最小になるような i 個の係数 a_i を計算することによって、前記集合 D_i に対応する時間種および前記集合 D_i それぞれの前記係数 a_i を含むデータを、前記サイトの利用状況の時間種特性を示すデータとして生成し、

【数 1】

$$Y = \sum_i a_i D_i + b$$

前記変動データの生成において、前記コンピュータは、前記遷移データが示す利用度の

50

変動を、利用度の多変量データで表し、当該多変量データを主成分分析することにより、複数の特徴基底を抽出し、抽出した前記複数の特徴基底それぞれの主成分得点を集合 Y として、前記複数の特徴基底ごとに集合 Y を生成し、

前記集合 D_i の生成において、前記コンピュータは、前記複数の特徴基底それぞれについて、前記集合 D_i を生成し、

前記サイトの利用状況の時間種特性を示すデータの生成において、前記コンピュータは、前記複数の特徴基底ごとの集合 Y について、それぞれの前記集合 D_i を用いて、係数 a_i を計算し、前記複数の特徴基底ごとの集合 D_i および係数 a_i を合成することにより、前記サイトの利用状況の時間種特性を示すデータとして生成する、時間種特性生成方法。

【請求項 6】

ネットワークに接続されたコンピュータにより提供される少なくとも 1 つのサイトの利用状況を示すログデータから、所定期間における前記サイトの利用度の変動を示す遷移データを抽出する処理と、

前記遷移データに基づいて、前記所定期間内の前記サイトの利用度の変動を、前記サイトの利用度を示す値の集合 Y で表す変動データを生成する処理と、

複数の時間種を示すデータと、前記複数の時間種それぞれで特定される期間を表す時間種因子値とが記録された時間種データ記録部に記録された複数の時間種から、予め記録された手順に従って時間種を抽出し、前記抽出した時間種に対応する i 個の時間種因子値の集合 D_i を生成する処理と、

前記集合 D_i を説明変数として、集合 Y を被説明変数とする下記（数 1）において、誤差 b の二乗が最小になるような i 個の係数 a_i を計算することによって、前記集合 D_i に対応する時間種および前記集合 D_i それぞれの前記係数 a_i を含むデータを、前記サイトの利用状況の時間種特性を示すデータとして生成する処理とをコンピュータに実行させ、

【数 1】

$$Y = \sum_i a_i D_i + b$$

前記変動データを生成する処理は、前記遷移データが示す利用度の変動を、利用度の多変量データで表し、当該多変量データを主成分分析することにより、複数の特徴基底を抽出し、抽出した前記複数の特徴基底それぞれの主成分得点を集合 Y として、前記複数の特徴基底ごとに集合 Y を生成する処理であり、

前記集合 D_i を生成する処理は、前記複数の特徴基底それぞれについて、前記集合 D_i を生成する処理であり、

前記サイトの利用状況の時間種特性を示すデータを生成する処理は、前記コンピュータは、前記複数の特徴基底ごとの集合 Y について、それぞれの前記集合 D_i を用いて、係数 a_i を計算し、前記複数の特徴基底ごとの集合 D_i および係数 a_i を合成することにより、前記サイトの利用状況の時間種特性を示すデータとして生成する処理である、時間種特性生成プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、例えば、EバンクやEコマース等のサービスを提供するWebサイト等のようなサイトの利用状況の時間種特性を解析するための時間種特性生成システム、方法、プログラムに関する。

【背景技術】

【0002】

例えば、EバンクやEコマース等のようなサービスが、コンピュータネットワーク上のWebサイトで提供されている。また、サービス提供者は、サービス利用者が特定のサービスを利用、成約または購買することの促進を期待して、様々な電子広告を、ネットワー

10

20

30

40

50

クを通じて利用者に配信したり提示したりする。

【0003】

電子広告は、広告提示期間や提示場所を任意に変更することができる。また、該電子広告に対する利用者の反応特性について、Web上でのサービス提供サイトのログデータ等から推定することが可能である。

【0004】

近年、IT技術の進展（モバイル機器／センサー・アクチュエータの普及等）により、個々の利用者の利用シーンに応じた細かいサービスおよび電子広告の制御が可能になっている。そのため、従来以上にきめ細かく効率的なサイトにおけるサービスおよび電子広告の制御方法の確立が要請されている。

10

【0005】

利用者のサイト利用行動は時間種特性を持っていることが多い。例えば、週末は利用が少ない傾向にあることや、一部の利用者は五十日（ごとうび）に活発であり、別の利用者群は第一月曜と月末の金曜日に活発である、という現象がある。また、一日の中でも、利用者は深夜の利用を控えることが多い、といった傾向もある。例えば、交通情報サービスの分野において、曜日／祝祭日／季節／五十日等の時間種による道路の混雑度合いの変化を加味して、目的地までの所要時間を計算する方法が開示されている（例えば、非特許文献1参照）。

【0006】

また、利用者の行動特性を分析する例として、利用者による商品の閲覧、購買等の行動が生じた時刻情報を利用者ごとに記録しておき、複数の利用者の時刻情報に基づいて利用者の行動を分析するシステム及び方法が開示されている（例えば、特許文献1）。このシステムでは、同様の行動をする複数の利用者がセグメント化され、セグメントの利用者に狙いを定めて、利用者の行動に先行する宣伝が実行される。しかしながら、前記従来のシステムは、利用者ごとに時刻情報を記録することにより、利用者の行動を分析するものであり、例えば、サイトへアクセスする利用者の行動がどのような時間種特性を有するのかを分析することはできない。すなわち、サイトの利用状況の時間種特性を定量的に分析することはできない。

20

【非特許文献1】熊谷正俊、伏木匠、横田孝義、佐野豊、鈴木研二「全国規模の交通情報サービスを目的とする所要時間長期予測技術の開発」、情報処理学会論文誌、財団法人情報処理学会、平成16年12月15日、第45巻、第12号、p.2696-2706

30

【特許文献1】特開号公報2000-285175号公報

【発明の開示】

【発明が解決しようとする課題】

【0007】

そこで、本発明は、サイトの利用状況の時間種特性を定量的に示すデータを生成することができる時間種特性生成システム、時間種特性生成方法および時間種特性生成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】

【0008】

本発明にかかる時間種特性生成システムは、ネットワークに接続されたコンピュータにより提供される少なくとも1つのサイトの利用状況を示すログデータから、所定期間における前記サイトの利用度の変動を示す遷移データを抽出する利用状況把握部と、前記遷移データに基づいて、前記所定期間内の前記サイトの利用度の変動を、前記サイトの利用度を示す値の集合Yで表す変動データを生成する変動データ生成部と、複数の時間種を示すデータと、前記複数の時間種それぞれで特定される期間を表す時間種因子値とが記録された時間種データ記録部と、時間種データ記録部に記録された複数の時間種から、予め記録された手順に従って時間種を抽出し、前記抽出した時間種に対応するi個の時間種因子値の集合D_iを生成する時間種因子抽出部と、前記集合D_iを説明変数として、集合Yを被説明変数とする下記（数1）において、誤差bの二乗が最小になるようなi個の係数a_i

40

50

を計算することによって、前記集合 D_i に対応する時間種および前記集合 D_i それぞれの前記係数 a_i を含むデータを、前記サイトの利用状況の時間種特性を示すデータとして生成する時間種特性生成部とを備える。

【0009】

【数1】

$$Y = \sum_i a_i D_i + b$$

【0010】

時間種特性生成部は、サイト利用度の変動を示す遷移データから得られる集合 Y を被説明変数、時間種因子値の集合 D_i を説明変数とする上記(数1)において誤差 b の二乗が最小になるような i 個の係数 a_i を計算する。その結果、集合 Y を、時間種因子値の関数で近似することができる。すなわち、集合 Y が表すサイトの利用度の変動の時間種特性が、時間種因子抽出部が抽出した時間種に対応する時間種因子値の集合 D_i および D_i それぞれの係数 a_i により表される。したがって、時間種特性生成部は、 D_i に対応する時間種および D_i それぞれの係数 a_i を用いて、サイトの利用状況の時間種特性を定量的に示すデータを生成することができる。このため、サイトの利用者群の行動の特徴を定量的に表す時間種特性が得られる。

10

【0011】

時間種特性生成部により生成された時間種特性を示すデータにより、サイト利用状況の時間種による変化を予測することが可能になる。また、サイトの利用状況の時間種特性を示すデータを利用してサイト運用することで、例えば、サイト利用状況の時間種特性に応じたタイムリーな広告掲載、キャンペーン実施、サービス提供等が可能になる。

20

【0012】

本発明にかかる時間種特性生成システムにおいて、前記利用状況把握部は、前記サイトが提供するサービスの利用度の変動を表す遷移データを抽出することが好ましい。

【0013】

これにより、時間種特性生成部は、前記サイトが提供するサービスの利用状況の時間種特性を示すデータを出力することができる。そのため、サイト管理者は、生成されたデータを用いて、例えば、前記サイトが提供するサービスの利用状況の時間種特性に応じたタイムリーなサービス提供、広告掲載等することが可能になる。

30

【0014】

本発明にかかる時間種特性生成システムにおいて、前記利用状況把握部は、前記サイトが提供する広告に対するアクセス量の変動を表す遷移データを抽出することが好ましい。

【0015】

これにより、時間種特性生成部は、前記サイトが提供する広告に対するアクセス量の時間種特性を示すデータを生成することができる。そのため、サイト管理者は、生成されたデータを用いて、例えば、前記サイトが提供する広告に対するアクセス量の時間種特性に応じたタイムリーな広告掲載、サービス提供等することが可能になる。

【0016】

本発明にかかる時間種特性生成システムにおいて、前記利用状況把握部は、少なくとも1つのサイトが提供する特定の広告にアクセスし、かつ当該サイトまたは他のサイトが提供する前記特定の広告に関するサービスを利用する行動の発生数の変動を表す遷移データを抽出することが好ましい。

40

【0017】

これにより、時間種特性生成部は、前記サイトが提供する広告に反応してサービスを利用する利用者の行動の時間種特性を示すデータを生成することができる。そのため、サイト管理者は、生成されたデータを用いて、例えば、前記サイトが提供する広告に対する反応行動の時間種特性に応じたタイムリーな広告掲載、サービス提供等することが可能になる。

50

【0018】

本発明にかかる時間種特性生成システムにおいて、前記時間種データ記録部は、連続する時間を日単位で複数の日に区切った場合の当該複数の日のうち、1つの時間種で特定される日を表す値である時間種因子値を、複数の時間種について記録する態様とすることができる。

【0019】

本発明にかかる時間種特性生成システムにおいて、前記時間種データ記録部は、連続する時間を時間単位で複数の期間に区切った場合の当該複数の時間帯のうち、1つの時間種で特定される時間帯を表す値である時間種因子値を、複数の時間種について記録する態様とすることができる。

10

【0020】

本発明にかかる時間種特性生成システムは、利用者の属性を示すデータと、当該属性を持つ利用者による利用状況の時間種特性を示すデータとを対応付けて記録する利用者時間種特性記録部と、前記時間種特性生成部が生成した時間種特性を示すデータと、前記利用者時間種特性記録部に記録された時間種特性を示すデータとを比較することにより、前記サイトを利用する利用者の属性を特定する利用者解析部とをさらに備えることが好ましい。

【0021】

利用者解析部により、サイトの利用者の属性が特定されるので、サイト管理者は、サイトの利用者群の属性を把握することができる。例えば、ある広告を提供するサイトについて、サイト管理社は、その広告がアピールした利用者群の属性を把握することができる。

20

【0022】

本発明にかかる時間種特性生成システムにおいて、前記時間種因子抽出部は、前記時間種データ記録部に記録された複数の時間種から、時間種の組み合わせを複数通り抽出して、各組み合わせに対応する時間種因子値の集合 D_i を複数通り生成し、前記時間種特性生成部は、前記複数通りの時間種因子値の集合 D_i それぞれについて、前記係数 a_i を計算することによって複数組の係数 a_i および時間種因子値の集合 D_i を生成し、当該複数組から所定基準に基づき選択した組の係数 a_i と、時間種因子の集合 D_i に対応する時間種を含むデータを、前記サイトの利用状況の時間種特性を示すデータとして生成することが好ましい。

30

【0023】

時間種特性生成部は、複数組の、係数 a_i および時間種因子値の集合 D_i から、所定基準に基づき選択した組を基に、時間種特性を示すデータを生成するので、複数組の係数 a_i および時間種因子値の集合 D_i の中から、前記サイトの利用状況の特徴をよく表す組を選択することができる。その結果、時間種特性生成部は、前記サイトの利用状況の特徴をよく表す時間種特性を示すデータを生成することができる。

【0024】

本発明にかかる時間種特性生成システムにおいて、前記変動データ生成部は、前記遷移データが示す利用度の変動を、利用度の多変量データで表し、当該多変量データを主成分分析することにより、特徴基底を抽出して当該特徴基底の主成分得点を集合 Y とし、前記時間種因子抽出部は、前記時間種データ記録部に記録された複数の時間種から抽出した時間種に対応する時間種因子値の多変量データを生成し、当該当該多変量データを主成分分析することにより、特徴基底を抽出して当該特徴基底の主成分得点を集合 D_i とすることが好ましい。

40

【0025】

これにより、被説明変数の集合 Y および説明変数の集合 D_i 双方のデータ量が圧縮され、時間種特性生成部の計算量が減少する。また、時間種特性生成部において、それぞれの特徴基底に基づく主成分得点同士で回帰分析が行われるので、独立性の高い時間種因子値の集合 D_i により被説明変数の特徴をよく説明する時間種特性を示すデータが生成される。

50

【0026】

本発明にかかる時間種特性生成システムにおいて、前記変動データ生成部は、前記遷移データが示す利用度の変動を、利用度の多変量データで表し、当該多変量データを主成分分析することにより、特徴基底を抽出し、当該抽出した特徴基底のうち1の特徴基底が、他の特徴基底に対して所定量以上異なっている特徴区間をさらに抽出し、当該特徴区間における前記特徴基底の主成分得点を集合 Y とし、前記時間種因子抽出部は、前記特徴区間内に含まれる時間種因子値の集合を、前記集合 D_i として生成することが好ましい。

【0027】

これにより、変動データ生成部は、独自性が高く特徴的な利用度の変動が現れている区間のみについての変動を表す主成分得点を集合 Y とすることができる。また、前記時間種因子抽出部は、独自性が高く特徴的な利用度の変動が現れている区間内から時間種因子値の集合を生成することができる。すなわち、特徴的な時間種因子値によって集合 D_i が生成される。そのため、時間種特性生成部は、特徴的な区間についての時間種特性を示すデータを効率よく生成することができる。

10

【0028】

本発明にかかる時間種特性生成システムにおいて、前記変動データ生成部は、前記遷移データが示す利用度の変動を、利用度の多変量データで表し、当該多変量データを主成分分析することにより、複数の特徴基底を抽出し、抽出した前記複数の特徴基底それぞれの主成分得点を集合 Y として、前記複数の特徴基底ごとに集合 Y を生成し、前記時間種因子抽出部は、前記複数の特徴基底それぞれについて、前記集合 D_i を生成し、前記時間種特性生成部は、前記複数の特徴基底ごとの集合 Y について、それぞれの前記集合 D_i を用いて、係数 a_i を計算し、前記複数の特徴基底ごとの集合 D_i および係数 a_i を合成することにより、前記サイトの利用状況の時間種特性を示すデータとして生成することが好ましい。

20

【0029】

時間種特性生成部は、特徴基底ごとの特徴的な時間種特性を表す集合 D_i および係数 a_i を合成することで、高い精度の、サイトの利用状況の時間種特性を示すデータを生成することができる。

【0030】

本発明にかかる時間種特性生成システムにおいて、前記変動データ生成部は、前記遷移データが示す利用度の変動を、利用度の多変量データで表し、当該多変量データを主成分分析することにより、特徴基底を抽出して当該特徴基底の主成分得点を集合 Y とし、前記時間種因子抽出部は、前記変動データ生成部が抽出した特徴基底の主成分得点と、前記時間種データ記録部に記録された複数の時間種のうち少なくとも1つの時間種に対応する時間種因子値との相関性を示す相関係数を計算し、当該相関係数に基づいて、時間種を抽出することが好ましい。

30

【0031】

時間種因子抽出部は、変動データ生成部が抽出した特徴基底の主成分得点との相関を示す相関係数を基に、集合 D_i に含める時間種因子値に対応する時間種を抽出するので、特徴基底の特徴をよく説明する説明変数として妥当な時間種因子値を集合 D_i に含めることができる。

40

【発明の効果】

【0032】

本発明によれば、サイトの利用状況の時間種特性を定量的に示すデータを生成することができる時間種特性生成システム、時間種特性生成方法および時間種特性生成プログラムを提供することができる。

【発明を実施するための最良の形態】

【0033】

[第1の実施形態]

図1は、本実施形態にかかる時間種特性生成システムの構成を示す機能ブロック図であ

50

る。図 1 に示すように、時間種特性生成システム 1 は、Webサーバ 2 に接続されている。Webサーバ 2 は、インターネット 5 に接続された他のコンピュータからアクセス可能な Web サイトを提供する。利用者は他のコンピュータを介して Web サイトにアクセスすることにより、Webサーバ 2 が提供する情報またはサービスを楽しむことができる。サーバ 2 は、ログデータ記録部 3、サービス場構造記録部 4 を備えている。ログデータ記録部 3 は、Web サイトの利用状況を示すログデータを記録する。サービス場構造記録部 4 は、Web サイトが提供するサービスや広告等に関するデータを記録する。

【0034】

時間種特性生成システム 1 は、Webサーバ 2 のログデータ記録部 3、サービス場構造記録部 4 に記録されたデータを用いて、Webサーバ 2 が提供するサービスや広告の利用者による利用状況の時間種特性を解析するためのシステムである。時間種特性生成システム 1 は、利用状況把握部 6、利用構造 DB 7、変動データ生成部 8、時間種因子抽出部 9、時間種データ記録部 11、抽出手順記録部 12、時間種特性生成部 13、評価部 14、時間種特性記録部 15、利用者解析部 16、利用者時間種特性記録部 17、利用者情報記録部 18 および広告挿入部 19 を備える。

10

【0035】

時間種特性生成システム 1 は、サーバマシン、パーソナルコンピュータ、ワークステーションなどのコンピュータ上に構築される。時間種特性生成システム 1 の利用状況把握部 6、変動データ生成部 8、時間種因子抽出部 9、時間種特性生成部 13、評価部 14、利用者解析部 16、および広告挿入部 19 のそれぞれは、1 台のコンピュータ上に構成されてもよいし、複数のコンピュータに分散して構成されてもよい。また、利用状況把握部 6、変動データ生成部 8、時間種因子抽出部 9、時間種特性生成部 13、評価部 14、利用者解析部 16、および広告挿入部 19 のそれぞれ機能は、コンピュータの CPU が所定のプログラムを実行することによって実現される。したがって、上記各機能を実行するためのプログラムおよびプログラムを記録した記録媒体も本発明の一実施形態である。また、時間種データ記録部 11、抽出手順記録部 12、時間種特性記録部 15、利用者時間種特性記録部 17、利用者情報記録部 18 は、コンピュータが備えるメモリ、ハードディスク等の記録媒体により具現化される。

20

【0036】

利用状況把握部 6 は、ログデータ記録部 3 およびサービス場構造記録部 4 から読み込んだデータを基に、所定期間における Webサーバ 2 が提供するサイトの利用度の変動を示す遷移データを生成して、利用構造 DB 7 へ記録する。利用状況把握部 6 がサービス場構造記録部 4 から読み込むデータには、例えば、Webサーバ 2 の Web サイトで提供するサービスに関する情報等が含まれる。

30

【0037】

利用状況把握部 6 がログデータ記録部 3 から読み込むデータには、例えば、Webサーバ 2 の Web サイトへアクセス履歴等が含まれる。利用構造 DB 7 に記録される遷移データには、例えば、Web サイトが提供するサービスの利用度、または Web サイトが提供する広告に対するアクセス量の変動を表すデータが含まれる。上記のサービス利用度または広告に対するアクセス量の変動は、例えば、特定の行動パターンの発生数の変動により表される。また、他の行動パターンの例として、Web サイトが提供する特定の広告にアクセスし、かつ当該 Web サイトまたは他の Web サイトが提供する特定の広告に関するサービスを利用する行動パターン（広告反応行動）等が挙げられる。

40

【0038】

変動データ生成部 8 は、利用構造 DB 7 に記録された遷移データに基づいて、所定期間内における Webサーバ 2 の Web サイトの利用度の変動を表す変動データを生成する。変動データは、前記 Web サイトの利用度を示す値の集合 Y で表される。前記 Web サイトの利用度を示す値は、例えば、所定の URL へのアクセス数、所定ボタンのクリック数、所定行動パターンの発生数、開始セッション数、セッション長、平均滞在時間等である。

50

【 0 0 3 9 】

変動データ生成部 8 は、例えば、N 日間の Web サイトの利用度の変動を、 $N \times M$ 行列 X で表すことができる。この場合、1 つの行が、1 日における利用度の変動を示す M 個の利用度を値で構成され、その行が日数分 (N 行) 並んで行列 X が構成される。図 2 (a) は、1 日あたり $M (= 144)$ 個の測定時刻における値を、 $N (= 365)$ 日分並べたデータの例を示す図である。図 2 (a) に示すデータは、0 : 00 から 23 : 50 まで 10 分ごとの時刻における利用度を示す値を $M (= 144)$ 個並べた行が、1 年分 ($N = 365$) 並んで構成される。

【 0 0 4 0 】

なお、変動データ生成部 8 は、計算量を減らすために、主成分分析を用いて $N \times M$ 行列 X の列または行の数を圧縮した主成分得点行列を変動データ (集合 Y) として生成することが好ましい。図 2 (b) は、図 2 (a) に示すデータを表す行列を、主成分分析を用いて圧縮した行列の主成分得点を表すデータの例を示す図である。図 2 (b) に示すデータは、1 つの特徴基底 (主成分) の N 日分の主成分得点を含む列が、第 1 主成分から第 p 主成分まで並んで構成されている。

10

【 0 0 4 1 】

ここで、変動データ生成部 8 が、主成分得点行列を計算する具体例を説明する。変動データ生成部 8 は、上記行列 X における各行の平均値 (= 各時刻の平均値、すなわち平均変動) を表す M 次元行列ベクトル x 、および全要素が 1 である N 次元ベクトル e を用いて、 X を列中心化した行列 X^* を下記式 (1) により求める。

20

【 0 0 4 2 】

$$X^* = X - e \cdot x \quad (1)$$

【 0 0 4 3 】

変動データ生成部 8 は、上記 $N \times M$ 行列を列中心化した行列 X^* に対して主成分分析を施す。その結果、行列 X^* の主成分 (= 特徴基底) を表すベクトルと、個々の特徴基底の寄与率が得られる。変動データ生成部 8 は、個々の特徴基底のうち、寄与率が閾値を超える特徴基底を p 個取り出し、特徴基底行列 P ($M \times p$ 行列) を生成する。

【 0 0 4 4 】

変動データ生成部 8 は、特徴基底行列 P に基づいて、Web サイトの利用度の変動を表す行列 X^* の主成分得点を計算する。主成分得点を表す行列 T は、例えば、特徴基底行列 P を用いて、下記式 (2) により計算される。

30

【 0 0 4 5 】

$$T = X^* \cdot P \quad (2)$$

【 0 0 4 6 】

この場合、行列 T は、 $N \times p$ 行列となる。行列 T の各列は、各特徴基底の主成分得点を表す。以上が、主成分得点行列の計算例である。ここでは、Web サイトの利用度の変動を表す行列 X が 2 次元の行列 ($N \times M$ 行列) であることを前提にしているが、行列 X が 1 次元ベクトルである場合、変動データ生成部 8 は、主成分分析による圧縮処理は行わなくてもよい。

【 0 0 4 7 】

40

時間種因子抽出部 9 は、時間種データ記録部 11 に記録された複数の時間種から予め記録された手順に従って時間種を抽出し、抽出した時間種に対応する所定数の時間種因子値の集合を生成する。

【 0 0 4 8 】

「時間種 (time species)」は、ある特定の属性を有する期間の呼び名を表す。特に、1 日または複数の日で区切ることができる期間であって、特定の属性を有する期間の呼び名を「日種 (day species)」と称する。日種の例として、「月曜日」、「火曜日」、「祝日」、「寒の入り」、「五十日 (ごとび : 日付に 5 または 10 が含まれる日)」、「週末」、「盆」、「暮れ」、「正月」、「クリスマス」、「啓蟄」、「春」、「残暑」、「花粉症の季節」、「厳冬期」、「積雪期」、「ゴールデンウィーク

50

」、「給料日」、「確定申告」、「新月」、「満月」、「晴」、「雨」、「熱帯夜」、「大安」、「仏滅」等が挙げられる。

【0049】

なお、時間種には、1日または複数の日で区切ることができる期間の他、例えば、時間または分単位で表される時間帯や、週、月、年またはその他の単位で表される期間の呼び名も含まれる。時間または分単位で表される時間帯の呼び名の例として、「深夜」、「早朝」、「昼食時」、「毎時0分」、「通勤ラッシュ」等が挙げられる。

【0050】

時間種データ記録部11には、複数の時間種を示すデータと、前記複数の時間種それぞれで特定される期間を表す時間種因子値とが記録されている。時間種因子値は、例えば、連続する時間を、所定単位で複数の期間に区切った場合の当該複数の期間のうち、時間種で特定される期間を表す値とすることができる。図3は、時間種データ記録部11に記録される日種データの例を示す図である。図3に示す日種データでは、最上段の行に「月曜日」、「火曜日」、「水曜日」、「祝日/振替」、「寒の入り」、「五十日」、「一月」等の日種を示すデータが記録されている。それぞれの日種で特定される日を表す日種因子値が各列に記録されている。日種因子値は、行方向に2006年1月1日から12月31日まで日ごとに記録された離散的な値であり、日種に該当する日に「1」、それ以外の日に「0」が記録されている。

10

【0051】

ここで、時間種因子抽出部9が、図3に示す日種データから日種を抽出して日種因子値の集合を生成する場合の計算例について説明する。時間種因子抽出部9が生成する日種因子の集合は、例えば、日種因子行列Dで表される。下記(数2)(式(3))は、図3に示す日種データに含まれる時刻種のうち「月曜日」、「火曜日」、「祝日/振替」、「五十日」が抽出された場合に生成される日種因子行列Dの具体例である。

20

【0052】

【数2】

$$D = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{————— (3)}$$

30

【0053】

上記(数2)の日種因子行列Dにおいて、各列は、時刻種「月曜日」、「火曜日」、「祝日/振替」、「五十日」の日種因子値をそれぞれ表している。各列において、「0」または「1」の要素が、2006年1月1日から12月31日まで日ごと順に並んでいる。なお、時間種因子抽出部9は、主成分分析を用いて行列Dの列または行の数を圧縮してもよい。また、時間種因子抽出部9は、2006年1月1日から12月31日のうちの一部の日の日種因子値を含むように日種因子行列Dを生成してもよい。

40

【0054】

行列Dの行数は、上記変動データ生成部8で生成されるN×M行列XのNと同じであることが好ましい。例えば、行列Dが、変動データが対象とする期間が2006年1月1日から12月31日とすると、時間種因子値が対象とする期間も同様に2006年1月1日から12月31日であることが好ましい。時間種因子抽出部9の詳細な処理は後述する。

【0055】

時間種特性生成部13は、変動データ生成部8が生成した変動データと、時間種因子抽

50

出部 9 が抽出した時間種およびその時間種に対応する時間種因子値の集合とに基づいて、時間種特性を示すデータを生成し、時間種特性記録部 15 に記録する。例えば、変動データ生成部 8 が変動データとして上述の主成分得点行列 T ($N \times p$ 行列) を、時間種因子抽出部 9 が時間種因子値の集合として上述の日種因子行列 D ($f \times N$ 行列) を生成した場合に、時間種特性生成部 13 が時間種特性を示すデータを生成する例について説明する。時間種特性生成部 13 は、日種因子行列 D を説明変数として、主成分特定行列 T を被説明変数とする下記式 (4) において、誤差 b の二乗が最小になるような係数行列 A ($f \times p$ 行列) を計算する。

【 0 0 5 6 】

$$T = D \cdot A + b \quad (4)$$

【 0 0 5 7 】

上記式 (4) は、Web サイトの利用度の変動を表す行列 T を、日種因子行列 D で表した式であるので、上記式 (4) の日種因子行列 D と、係数行列 A とによって、Web サイトの利用状況の日種特性が表されることになる。すなわち、日種因子行列 D と係数行列 A とは、日種による Web サイトの利用状況の変動を定量的に表すデータとなっている。また、日種因子行列 D に対応する日種と、係数行列 A を用いて、Web サイトの未来の利用状況を示す予測値を計算することができる。例えば、日種因子行列 D と主成分得点行列 T が、共に 2006 年 1 月 1 日から 12 月 31 日の期間の日種因子値、Web サイトの利用度の変動をそれぞれ表すデータである場合に、2007 年 1 月 1 日から 12 月 31 日の Web サイトの利用度の変動を予測する計算例を説明する。この場合、日種因子行列 D に対応する日種について、2007 年 1 月 1 日から 12 月 31 日の日種因子値を表した行列 D_{2007} が作成される。この行列 D_{2007} を用いて下記式 (5) により、2007 年 1 月 1 日から 12 月 31 日における Web サイトの利用度を変動の予測値を表す行列 T_{2007} を計算することができる。

【 0 0 5 8 】

$$T_{2007} = D_{2007} \cdot A \cdot P^T + e \cdot x \quad (5)$$

【 0 0 5 9 】

上記式 (5) において、 P^T は、行列 P の転置行列を表し、 $e \cdot x$ は、上記式 (2) の $e \cdot x$ と同様である。

【 0 0 6 0 】

このように、日種因子行列 D に対応する時間種と、係数行列 A によって、Web サイトの利用状況を予測することができる。したがって、時間種特性生成部 13 は、Web サイトの利用状況の時間種特性を示すデータとして、例えば、日種因子行列 D に対応する日種 (すなわち、時間種因子抽出部 9 が抽出した日種) と、日種因子行列 D の係数行列 A を、時間種特性記録部 15 に記録する。

【 0 0 6 1 】

なお、時間種因子抽出部 9 は、日種の組み合わせを複数通り (n 通りとする) 抽出して、各組み合わせに対応する n 個の日種因子行列 $D(n)$ を生成し、時間種特性記録部 15 は、 n 個の日種因子行列 $D(n)$ それぞれについて係数行列 A を計算して時間種特性を示すデータを生成することができる。

【 0 0 6 2 】

評価部 14 は、時間種特性生成部 13 が生成した時間種特性を示すデータを評価し、評価結果を時間種特性生成部 13 に通知する。評価部 14 は、例えば、時間種特性を示すデータによって、Web サイトの利用状況をどの程度正確に予測できるかを示す予測精度を生成する。また、評価部 14 は、時間種因子抽出部 9 が抽出した時間種の組み合わせの妥当性を示す値を生成してもよい。

【 0 0 6 3 】

時間種特性生成部 13 は、評価部 14 による評価に基づき、前記 n 個の日種因子行列 $D(n)$ のうち、最適な日種因子行列 D_{ideal} を選択することができる。時間種特性生成部 13 は、選択した日種因子行列 D_{ideal} に対応する時間種およびその係数行列 A を含むデ

10

20

30

40

50

ータが時間種特性を示すデータとして時間種特性記録部 15 に記録することができる。

【0064】

利用者解析部 16 は、時間種特性記録部 15 に記録された時間種特性を示すデータを用いて、Web サイトを利用した利用者の属性を解析する。利用者情報記録部 18 には、利用者の属性を示すデータと、その属性を持つ利用者による Web サイトの利用状況の時間種特性を示すデータとが対応付けられて記録されている。利用者解析部 16 は、時間種特性記録部 15 に記録された時間種特性を示すデータと、利用者時間種特性記録部 17 に記録された時間種特性を示すデータとを比較することにより、Web サイトを利用する利用者の属性を特定する。利用者解析部 16 は、特定した Web サイトの利用者の属性を、その利用状況の時間種特性とを関連付けて利用者情報記録部 18 に記録する。

10

【0065】

広告挿入部 19 は、時間種特性記録部 15 に記録された時間種特性を示すデータを用いて、Web サイトに広告を提供するのに最も適した時期を決定し、Web サイトに広告を挿入する。広告挿入部 19 は、Web サーバ 2 に記録された Web サイトのファイルを更新または追加することにより、広告を挿入する。

【0066】

例えば、広告挿入部 19 は、時間種特性を示すデータを用いて、Web サイトの利用が多い時期を予測し、その時期に Web サイトに広告を挿入する。時間種特性を示すデータに、例えば、ある Web サイトにおける 2006 年度の広告アクセス数または広告反応行動の日種因子行列 D に対応する日種と、日種因子行列 D の係数行列 A が含まれている場合、広告挿入部 19 は、これらのデータを上記式 (5) に代入して、2007 年度の Web サイトにおける広告アクセス数または広告反応行動の変動を予測することができる。これにより、Web サイトの利用状況の時間種特性に応じたタイムリーな広告の提供が可能になる。

20

【0067】

また、広告挿入部 19 は、利用者情報記録部 18 に記録されたデータを用いて、特定の属性を持つ利用者を対象とした広告の最適な提供時期を決定し、最適な提供時期に広告を挿入することもできる。

【0068】

(時間種特性生成システム 1 の動作)

30

次に、時間種特性生成システム 1 の動作について説明する。図 4 は、時間種特性生成システム 1 の動作例の大まかな流れを示すフローチャートである。図 4 に示すように、まず、利用状況把握部 6 が、ログデータ記録部 3 およびサービス場構造記録部 4 から必要なデータを読み込んで、遷移データを生成し利用構造 DB 7 へ記録する (Op 1)。

【0069】

次に、変動データ生成部 8 が、遷移データを基に、特定の Web サイトの利用度の変動データを生成する (Op 2)。その後、時間種因子抽出部 9 は、時間種データ記録部 11 から時間種およびその時間種因子値を抽出し、時間種特性生成部 13 が、変動データが示す Web サイトの利用度の変動の時間種特性を示すデータを生成する (Op 3)。

【0070】

40

その後、利用者解析部 16 は、時間種特性を示すデータを用いて、Web サイトの利用者の属性を特定する (Op 4)。また、広告挿入部 19 は、時間種特性を示すデータを用いて、Web サイトの広告を挿入する最適時期を決定し、その最適時期に広告を挿入する (Op 5)。

【0071】

図 4 に示した処理により、Web サーバ 2 の特定の Web サイトの利用度の変動の時間種特性を示すデータが生成される。また、その Web サイトの利用者の属性および Web サイトの広告を挿入する最適時期を示すデータが得られる。なお、図 4 に示す動作例における Op 1 ~ Op 3、Op 5 の処理の具体例は後述する。また、時間種特性生成システム 1 の動作は、図 4 に示す動作例に限られない。

50

【 0 0 7 2 】

(利用状況把握部 6 が、遷移データを生成する処理 (Op 1) の具体例)

図 5 は、利用状況把握部 6 が、遷移データを生成する処理 (Op 1) の具体的例を示すフローチャートである。図 5 に示すように、まず、利用状況把握部 6 は、Webサーバ 2 が Web サイト群により提供するサービス群および広告群の基礎情報を、サービス場構造記録部 4 から読み出して、利用構造 DB 7 に記録する (Op 1 1)。サービス群および広告群の基礎情報には、例えば、サービスまたは広告提供場 (サイト)、Web サイトを構成するページの URL、サイトにおける利用者のアクション等のサイトに関する情報が含まれる。

【 0 0 7 3 】

図 6 (a) は、利用構造 DB 7 に記録されるサービス群の基礎情報の一例であるサービス基礎情報テーブル 2 1 を示す図である。サービス基礎情報テーブル 2 1 には、サービスを識別するサービス ID、サービス場 (サイト)、Web サイトに含まれるページの URL、各ページにおける利用者のとりうるアクションがサービスごとに記録されている。図 6 (b) は、利用構造 DB 7 に記録される広告群の基礎情報の一例である広告基礎情報テーブル 2 2 を示す図である。広告基礎情報テーブル 2 2 には、広告を識別する広告 ID、広告場 (サイト)、Web サイトに含まれるページの URL、各ページにおける利用者のとりうるアクションが広告ごとに記録されている。

【 0 0 7 4 】

次に、利用状況把握部 6 は、ログデータ記録部 3 に記録されたアクセスログを基にして、Webサーバ 2 が提供するサービスを利用した利用者の少なくとも 1 つの Web サイトにおける行動パターンを示すデータを利用構造 DB 7 に記録する (Op 1 2)。

【 0 0 7 5 】

例えば、利用状況把握部 6 は、サービス基礎情報テーブル 2 1 から対象とする Web サイトの URL および利用者のアクションを抽出し、抽出した URL および利用者のアクションに該当するログをログデータ記録部 3 から抽出する。利用状況把握部 6 は、抽出したログを定量分析 (傾向分析) し、クラスタリングすることにより、行動パターンを示すデータを生成する。

【 0 0 7 6 】

図 7 は、サービス利用の行動パターンを示すデータの一例である行動パターンテーブル 2 3 を示す図である。行動パターンテーブル 2 3 には、行動パターン ID と、それぞれの行動パターンを示す情報とが対応付けられて記録される。行動パターンを示す情報は、行動パターンに含まれるアクションを「 1」、行動パターンに含まれないアクションを「 0」とする値で表されている。例えば、行動パターン ID = “ P - 0 0 0 1 ” の行動パターンは、A 社通信販売広告サイトで閲覧、A 社通信販売サイトで発注ボタン押下、B 社 e コマースサイト閲覧を順に行う行動パターンを表している。

【 0 0 7 7 】

また、利用状況把握部 6 は、Webサーバ 2 が提供するサービスの広告にアクセスした利用者の、少なくとも 1 つの Web サイトにおける広告アクセス行動パターンを示すデータを利用構造 DB 7 に記録する (Op 1 3)。広告アクセス行動パターンを示すデータは、例えば、図 7 に示した行動パターンテーブル 2 3 における行動パターンのアクションを、広告アクセスのアクションとしたテーブルとすることができる。

【 0 0 7 8 】

また、利用状況把握部 6 は、Webサーバ 2 が提供する広告にアクセスしてから、Webサーバ 2 が提供するサービスを利用した利用者の少なくとも 1 つの Web サイトにおける広告反応行動の行動パターンを示すデータを利用構造 DB 7 に記録する (Op 1 4)。広告反応行動の行動パターンを示すデータは、例えば、図 7 に示した行動パターンテーブル 2 3 における行動パターンのアクションを、広告反応行動に含まれるアクションとしたテーブルとすることができる。

【 0 0 7 9 】

10

20

30

40

50

利用状況把握部 6 は、ログデータ記録部 3 に記録されたアクセスログを基に、利用構造 DB 7 に記録された行動パターンの発生数の変動を表す遷移データを生成し、利用構造 DB 7 に記録する (Op 15)。図 8 は、遷移データの一例である遷移データテーブル 2 4 を示す図である。遷移データテーブル 2 4 には、行動パターン ID と、その行動パターンが発生した日時を示すデータとが対応付けられて記録されている。遷移データテーブル 2 4 は、変動データ生成部 8 が、特定の行動パターンの発生数の変動を表す変動データを生成する際に用いられる。

【 0 0 8 0 】

以上、利用状況把握部 6 が、利用構造 DB 7 に遷移データを記録する処理の例を説明したが、遷移データの取得方法および遷移データの構造は、上記例に限られない。

10

【 0 0 8 1 】

(変動データ生成部 8 が、変動データを生成する処理 (Op 2) の具体例)

図 9 は、変動データ生成部 8 が遷移データを基に変動データを生成する処理の具体例を示すフローチャートである。変動データ生成部 8 は、まず、時間種特性を解析する対象となる対象期間、対象 Web サイトおよび対象行動パターンを選択する (Op 21)。変動データ生成部 8 は、例えば、ユーザインタフェースを介してユーザからの選択指示を受け付けることによって上記選択を行ってもよいし、予め記録されたデータを基に上記選択を行ってもよい。

【 0 0 8 2 】

変動データ生成部 8 は、対象となる Web サイトを、例えば、図 6 (a)、(b) に示すサービス基礎情報テーブル 2 1、広告基礎情報テーブル 2 2 に記録されたサービス場または広告場から選択してもよい。また、対象行動パターンは、図 7 に示す行動パターンテーブル 2 3 から選択されてもよい。変動データ生成部 8 は、例えば、対象 Web サイトとして、ある特定の広告場を選択し、行動パターンとして、その広告場の広告へアクセスする行動、または、その広告に対する広告反応行動等を選択することができる。

20

【 0 0 8 3 】

なお、変動データ生成部 8 は、遷移データが示す全期間、全 Web サイトおよび全行動パターンをそれぞれ対象期間、対象 Web サイト、対象行動パターンとしてもよい。また、変動データ生成部 8 が解析の対象を決めるために選択する項目は、上記の期間、Web サイト、行動パターンに限られない。例えば、選択する項目に、解析対象とする利用者または利用者群が含まれてもよい。

30

【 0 0 8 4 】

変動データ生成部 8 は、Op 21 で選択した対象期間における対象 Web サイトでの対象行動パターンの一定時間ごとの発生数を、遷移データを基に集計し、行列 X を生成する (Op 22)。例えば、変動データ生成部 8 は、10 分ごとの対象行動パターンの発生数を、2006 年 1 月 1 日から 2006 年 12 月 31 日までの 365 日間の期間について集計した場合、1 日分の 10 分ごとの発生数をそれぞれ表す要素を 1 行として、365 日分の行 (365 行) を有する行列 X を生成する。この場合、行列 X の要素の並び方は、図 2 (a) に示したデータと同じ並び方になる。

【 0 0 8 5 】

次に、行列 X に変動データ生成部 8 は、行列 X に対して主成分分析を施して、主成分得点行列 T を生成する (Op 23)。主成分得点行列 T の計算例は上述した通りである。変動データ生成部 8 は、主成分得点行列 T を時間種特性生成部 1 3 に通知する (Op 24)。時間種特性生成部 1 3 は、主成分得点行列 T を用いて、対象期間における対象 Web サイトでの対象行動パターンの発生数の変動の時間種特性を示すデータを生成することができる。

40

【 0 0 8 6 】

(時間種因子抽出部 9 が時間種を抽出し、時間種特性生成部 1 3 が、時間種特性を示すデータを生成する処理 (Op 3) の具体例)

図 10 は、時間種因子抽出部 9 が時間種を抽出し、時間種特性生成部 1 3 が、時間種特

50

性を示すデータを生成する処理（Op3）の具体的例を示すフローチャートである。図10に示すように、まず、時間種因子抽出部9は、抽出手順記録部12から抽出手順データを読み込む（Op301）。抽出手順データは、時間種データ記録部11に記録された複数の時間種から、時間種を抽出する手順または条件を表すデータであり、予め抽出手順記録部12に記録されている。抽出データには、例えば、後述する日種因子母群作成の条件を表すデータ、抽出リスト作成の手順を表すデータ等が含まれる。

【0087】

時間種因子抽出部9は、抽出手順データを基に、時間種因子母群を作成する（Op302）。時間種因子母群は、時間種因子抽出部9が抽出対象とする時間種群である。時間種因子母群は、時間種データ記録部11に記録された時間種を基に作成される。例えば、時間種データ記録部11に記録された時間種のうちの一部が時間種因子母群になる。また、時間種因子母群には、時間種複数の時間種を複合した時間種、複数の時間種を主成分化した時間種等も含まれてもよい。ここでは、図3に示した日種データから時間種因子母群を生成する場合を例に挙げて説明する。

10

【0088】

図11は、抽出手順データに含まれるデータの一例である母群抽出条件テーブル25を示す図である。母群抽出条件テーブル25には、日種の種別を示すデータと、各種別に含まれる日種、各種別における日種の複合の有無と複合条件、日種の主成分化による複合の有無を示すデータが、各種別について記録されている。ここで、日種の複合とは、種別に含まれる日種のうち2以上の日種を組み合わせて複合日種を作成することである。また、日種因子の主成分化による複合は、種別に含まれる複数の日種因子に対して主成分分析を行い、複合日種因子を生成することである。例えば、時間種因子抽出部9は、主成分分析により抽出された主要な主成分毎に、その主成分の中で固有ベクトルの絶対値成分が大きな日種因子を抽出し、抽出した日種因子のみを使って複合日種因子を生成することができる。

20

【0089】

ここで、母群抽出条件テーブル25に基づいて時間種因子母群を作成する処理の例を説明する。まず、時間種因子抽出部9は、母群抽出条件テーブル25の種別「曜日」、「月」、「季節」、「祝日/振替」、「期末」に含まれる日種を母群抽出における抽出範囲に決定し、抽出範囲の日種を全て時間種因子母群に追加する。

30

【0090】

次に、時間種因子抽出部9は、母群抽出条件テーブル25の日種の複合の有無と複合条件を参照して、各種別において日種の複合をするか否かを判断し、複合する場合は複合条件に従って、複合日種を作成し、時間種因子母群に追加する。例えば、「曜日」種別について、3連続までの複合とあるので、例えば、「土日」のように2日連続する曜日の複合した複合日種と、「土日月」のように3日連続の曜日を複合した複合日種とを、2連続する曜日および3連続する曜日の全ての組み合わせについて作成される。

【0091】

さらに、時間種因子抽出部9は、主成分化による複合の有無を参照して、次に各種別において日種の主成分化による複合をするか否かを判断する。主成分化による複合をする場合は、時間種因子抽出部9は、種別に含まれる日種に対して主成分分析を行い、各主成分に対応する複合日種を作成し、時間種因子母群に追加する。

40

【0092】

図12は、時間種因子抽出部9が、図11に示す母群抽出条件テーブル25に基づいて作成した時間種因子母群を表す母群データ26の例を示す図である。母群データ26は、時間種因子母群に含まれる日種ごとに、抽出対象日種ID、種別、日種、日種因子が記録される。抽出対象日種IDは、時間種因子母群に含まれる日種を識別するための識別子である。母群データ26の日種には、例えば、「月火」、「月火」等のような複合日種も含まれる。日種因子は、図3に示す日種データと同様に、日種に該当する日を「1」、それ以外の日を「0」で表したデータである。

50

【 0 0 9 3 】

なお、時間種因子母群の作成は、上述の母群抽出条件テーブル 2 5 のように予め記録されたデータに基づいて行われる場合に限られない。時間種因子抽出部 9 は、例えば、ユーザインタフェースを介して受け付けた、Web サイト管理者等のユーザから指示に従って作成してもよい。なお、後述する抽出リストの作成も同様にユーザからの指示に従って作成されてもよい。

【 0 0 9 4 】

以上のようにして、時間種因子母群が作成されると、時間種因子抽出部 9 は、時間種因子母群から抽出する日種の組み合わせを示すデータである抽出リストを作成する (Op 3 0 3)。時間種因子抽出部 9 は、例えば、母群データ 2 6 に記録された時間種因子母群から日種の組み合わせを選択し、選択した日種の組み合わせを示すデータを抽出リストに追加する処理を繰り返すことによって、複数通りの日種の組み合わせを抽出リストに記録する。

10

【 0 0 9 5 】

時間種因子抽出部 9 は、例えば、時間種因子母群から、一定数の日種を選択する処理を、すべての日種の組み合わせパターンについて行うことができる。これにより、時間種因子母群から選択できる一定数の日種の組み合わせのすべてのパターンが抽出リストに記録される。また、時間種因子抽出部 9 は、上記選択処理の繰り返しにおいて、重要な日種 (例えば、種別が「曜日」の日種) を毎回選択するようにしてもよい。なお、抽出リストの作成手順は、これに限られない。例えば、時間種因子抽出部 9 は、一定数だけランダムに日種を選択する処理を繰り返すことで、複数通りの日種の組み合わせを抽出リストに記録することができる。このような、抽出リストの作成手順を示すデータは、例えば、抽出手順記録部 1 2 に抽出手順データとして予め記録することができる。

20

【 0 0 9 6 】

図 1 3 は、時間種因子抽出部 9 が作成した抽出リストを表すデータの一例である抽出リストテーブル 2 7 を示す図である。抽出リストテーブル 2 7 には、リスト番号と、各リスト番号に対応する日種の組み合わせを示すデータが記録される。日種の組み合わせは、例えば、母群データ 2 6 に記録された抽出候補日種 ID の組み合わせにより表される。

【 0 0 9 7 】

このように抽出リストが作成されると、時間種特性生成部 1 3 は、抽出リストに記録されたそれぞれの日種の組み合わせを用いて、時間種特性を示すデータを生成する。図 1 0 に示すフローチャートにおいて、以下に述べる Op 3 0 5 ~ Op 3 0 9 の処理は、抽出リストに記録されたそれぞれの日種の組み合わせについて実行される。

30

【 0 0 9 8 】

まず、時間種因子抽出部 9 は、変数 n を「0」に初期化する (Op 3 0 4)。時間種因子抽出部 9 は、変数 n に 1 加算し (Op 3 0 5)、抽出リストテーブル 2 7 の n 番目の日種の組み合わせについて、日種因子行列 $D(n)$ を生成する (Op 3 0 6)。

【 0 0 9 9 】

例えば、 $n = 1$ の場合、時間種因子抽出部 9 は、抽出リストテーブル 2 7 の 1 番目の日種の組み合わせ (リスト番号 = “1” の日種の組み合わせ) について、日種因子行列 $D(1)$ を生成する。日種因子行列 $D(1)$ は、例えば、抽出候補日種 ID が “0 0 0 1” ~ “0 0 0 9”、“0 0 1 1” ~ “0 0 2 1”、・・・の日種それぞれの日種因子値を各列に配置した行列である。

40

【 0 1 0 0 】

時間種特性生成部 1 3 は、Op 3 0 6 で生成された日種因子行列 $D(n)$ を説明変数として、変動データ生成部 8 が生成した主成分特定行列 T を被説明変数とする下記式 (6) において、誤差 b の二乗が最小になるような係数行列 $A(n)$ を計算する (Op 3 0 7)。

【 0 1 0 1 】

$$T = D(n) \cdot A(n) + b \quad (6)$$

50

【 0 1 0 2 】

上記式 (6) において、主成分特定行列 T は、上記式 (4) における主成分特定行列 T と同じであり、Webサイトの利用度の変動を表す行列である。

【 0 1 0 3 】

評価部 1 4 は、日種因子行列 $D (n)$ に含まれる日種間の相関関係の有無を判定する (Op 3 0 8)。評価部 1 4 は、公知の統計的方法を用いて相関関係の有無を判定することができる。日種間の相関関係の有無は、時間種因子抽出部 9 が、時間種因子母群から日種の組み合わせを選択して抽出リスト作成する際の指針とすることができる。

【 0 1 0 4 】

図 1 4 は、日種因子行列 $D (n)$ と主成分特定行列 T との相関関係の有無を判定する処理の一例を示すフローチャートである。図 1 4 に示す例では、評価部 1 4 は、まず、日種因子行列 $D (n)$ に含まれる日種間の相関係数行列の逆行列の要素 $R_{i i}$ (= 分散拡大係数) を計算する (Op 3 0 8 1)。評価部 1 4 は、要素 $R_{i i}$ のうち 1 つの要素について、閾値より小さいか否かを判断する (Op 3 0 8 2)。前記 1 つの要素が閾値より小さい場合 (Op 3 0 8 2 で Yes)、評価部 1 4 は、前記 1 つの要素に対応する日種間は相関関係ありと判定する (Op 3 0 8 3)。前記 1 つの要素が閾値より小さくない場合 (Op 3 0 8 2 で No)、評価部 1 4 は、前記 1 つの要素に対応する日種間は相関関係なしと判定する (Op 3 0 8 4)。評価部 1 4 は、Op 3 0 8 2 ~ Op 3 0 8 4 の処理を、全ての要素 $R_{i i}$ について行う。

【 0 1 0 5 】

以上、図 1 4 に示した処理により、日種因子行列 $D (n)$ に含まれる日種間で相関関係がある日種ペアが特定される。なお、相関関係の判定方法は図 1 4 に示した例に限られない。例えば、評価部 1 4 は、日種因子行列 $D (n)$ に含まれる日種因子の独自性を示す値 (例えば t 検定値) を求め、その値が閾値を超えているか否かにより相関関係の有無を判定することができる。

【 0 1 0 6 】

さらに、評価部 1 4 は、 $D (n)$ および $A (n)$ を用いて計算されるWebサイトの利用度変動の予測値の予測精度を計算する (図 1 0 の Op 3 0 9)。評価部 1 4 は、例えば、実測データが得られている期間について、 $D (n)$ および $A (n)$ を用いて予測データを算出し、実測データと予測データとを比較することで、予測精度を示すデータを得ることができる。予測精度を示すデータは、すなわち、 $D (n)$ および $A (n)$ の予測に対する利用適性度合いを示すデータとなる。

【 0 1 0 7 】

ここで、一例として、 $D (n)$ および $A (n)$ が、2006年1月1日から12月31日におけるWebサイトの利用度の変動について計算されたデータであり、2007年1月1日から12月31日におけるWebサイトの利用度の変動を示す実績データが得られている場合について説明する。この場合、評価部 1 4 は、 $D (n)$ に含まれる日種について、2007年1月1日から12月31日の日種因子を表した行列 D_{2007} を作成する。この行列 D_{2007} を用いて上述した式 (5) により、2007年1月1日から12月31日におけるWebサイトの利用度を変動の予測値を表す行列 T_{2007} を計算することができる。この行列 T_{2007} が表すWebサイトの利用度を変動の予測値と、実績データとを比較する。図 1 5 は、行列 T_{2007} が表す変動の予測値と、実績データが表す変動の実測値とを、横軸を時間 t 、縦軸を利用度 Y とするグラフ上で表した図である。図 1 5 中、実線で示す曲線 f は実測値の変動を、破線で示す曲線 g は予測値の変動を示している。評価部 1 4 は、例えば、直線 $Y = 0$ と曲線 f との間の面積 S_f と、曲線 f と曲線 g との間の部分の間の面積 S_{fg} とを用いて下記式 (7) により、指標 H を計算する。

【 0 1 0 8 】

$$H = 1 - (S_{fg} / S_f) \quad (7)$$

【 0 1 0 9 】

指標 H は、予測値が実測値にどの程度近いかを表す指標であり、平均説明率と呼ぶこと

ができる。平均説明率は、0～1の値となるので、指標として利用するのに適している。評価部14は、この指標Hの値を、予測精度を示すデータとして記録媒体へ記録することができる。

【0110】

なお、予測精度を示すデータを計算する方法は、上述の例に限られない。例えば、評価部14は、上述の指標Hのように、実績値にどの程度近いかを表す指標ではなく、予測のための実行する計算がどの程度簡潔かを表す指標（例えば、情報量基準）を計算してもよい。これは、予測のために実行する計算は簡潔な方がよい予測分布が得られるという考え方もあるからである。

【0111】

時間種因子抽出部9は、抽出リストテーブル27が示す抽出リストに含まれる全ての日種の組み合わせについて処理が終了したか否かを判断する（Op310（図10参照））。抽出リストに含まれる全ての日種の組み合わせについて処理が終了していない場合（Op310でNoの場合）は、Op305～Op309の処理を再度繰り返す。

【0112】

これにより、上述したOp305～Op309の処理は、抽出リストテーブル27が示す抽出リストの日種の組み合わせ全てについて実行される。その結果、抽出リストの日種の組み合わせそれぞれについて、 $D(n)$ 、 $A(n)$ および予測精度を示すデータが得られる。

【0113】

時間種特性生成部13は、抽出リストの日種の組み合わせ全てについての $D(n)$ 、 $A(n)$ の組の中から、最適な組を選択する。時間種特性生成部13は、例えば、Op309において求められた予測精度が最も高い組を、最適な組として選択することができる。選択された最適な $D(n)$ 、 $A(n)$ の組は、変動データが示すWebサイトの利用度の変動の時間種特性を示すデータとして、時間種特性記録部15に記録される。なお、最適な組を選択する方法は、上記例に限られない。

【0114】

図16は、時間種特性記録部15に記録された時間種特性を示すデータの一例である時間種特性データ28を示す図である。時間種特性データ28には、解析IDごとに、解析の対象範囲、日種の組み合わせ（固有日種群）および係数行列が記録されている。解析の対象は、例えば、時間種特性の対象を示すデータであり、例えば、対象Webサイト、対象行動パターン等が含まれる。日種の組み合わせは、上記の最適な $D(n)$ に対応する日種群を示すデータであり、固有日種群と称される。係数行列は、上記の最適な $A(n)$ である。

【0115】

以上、図10に示した処理により、変動データが示す対象Webサイトの利用度の変動を、精度よく表現するための日種の組み合わせ（ $D(n)$ に対応する日種群（固有日種群））と日種の係数（ $A(n)$ ）が得られる。これらの値を用いて、Webサイトの利用者の属性を特定する利用者の属性解析（図4のOp4）や、Webサイトの広告を挿入する最適時期を決定し、その最適時期に広告を挿入する処理（図4のOp5）が行われる。

【0116】

（利用者解析部16が、利用者属性を特定する処理（Op4）の具体例）

図17は、利用者解析部16が、遷移データを生成する処理（図4のOp4参照）の具体的例を示すフローチャートである。ここでは、一例として、図16に示す時間種特性データ28の解析ID="1"のデータを用いて、Webサイトを利用した利用者の属性を解析する処理を説明する。図17に示すように、利用者解析部16は、利用者時間種特性記録部17からの中から特定の利用者属性に対応する時間種特性を示すデータを抽出する（Op41）。

【0117】

利用者時間種特性記録部17には、例えば、図16に示した固有日種群と係数行列との

10

20

30

40

50

組が、様々な利用者属性について記録されている。利用者属性は、例えば、利用者の性質、体質、習慣、習性、帰属する組織、社会的地位等である。利用者属性を表すデータの具体例としては、年齢、性別、職業、役職、年収、居住地域、勤務地、学校、勤務先、趣味または家族構成を表すデータまたはそのようなデータの組み合わせが挙げられる。このように様々な利用者属性について、過去に解析された固有日種群と係数行列が利用者時間種特性記録部17に記録されている。Op41で、利用者解析部16は、ある1つの利用者属性を選んで、その固有日種群と係数行列を抽出する。

【0118】

次に、利用者解析部16は、Op41で抽出された固有日種群および係数行列と、時間種特性データ28の解析ID="1"における固有日種群および係数行列との類似度を計算する(Op42)。利用者解析部16は、例えば、固有日種群N1と固有日種群N2との類似度は、N1およびN2の双方に含まれる日種の数とすることができる。また他の例として、N1の所定期間の日種因子行列D(N1)およびN1の係数行列A1と、N2の所定期間の日種因子行列D(N2)およびN2の係数行列A2とを用いて、N1とN2との類似度を求めることができる。この場合、利用者解析部16は、例えば、行列D1=(N1)・A1と、行列D2=D(N2)・A2とを計算する。そして、利用者解析部16は、行列D1の要素と行列D2の要素との相関係数を計算し、各要素間の相関係数のうち、閾値を超える値の割合を類似度とすることができる。なお、類似度の計算方法は、上記例に限定されない。

【0119】

類似度が閾値以上であれば(Op43でYes)、利用者情報記録部18に、利用者属性と、時間種特性データ28の解析ID="1"のデータとを対応付けて記録する(Op44)。Op41~Op44の処理は、利用者時間種特性記録部17に記録されている利用者属性全てについて終了するまで(Op45でYesと判断されるまで)繰り返される。

【0120】

例えば、解析ID="1"の固有日種属性と係数行列と、利用者属性="学生"の固有日種属性と係数行列との類似度が閾値以上であった場合、解析ID="1"の解析対象である「A社の通信販売サイトのサービス利用行動」の時間種特性に類似した時間種特性の利用者属性は「学生」であるという情報が得られる。そのため、「A社の通信販売サイトのサービス利用行動」をする利用者は「学生」が多いと推定することが可能になる。

【0121】

以上、図17に示した処理により、例えば、Webサイトのサービス利用行動と、利用者属性との関係を示すデータが利用者情報記録部18に記録される。また、同様に、Webサイトの広告アクセス行動や、広告反応行動についても、その行動を取り得る利用者の利用者属性との関係を示すデータが記録される。そのため、例えば、Webサイトによる広告が、どのような利用者に影響を与えたのが判定することが容易になる。このような効果は、個人情報保護法の成立により、利用者の個人プロファイルの利用は難しくなっている現状においてはより重要となる。

【0122】

(広告挿入部19が、広告を挿入する処理(Op5)の具体例)

図18は、広告挿入部19が、広告を挿入する時期を決定する処理の例を示すフローチャートである。ここでは、一例として、時間種特性生成部13が、あるWebサイトの広告にアクセスして、その広告に関するサービスを利用する利用者の広告反応行動の時間種特性を示すデータを生成した場合に、広告挿入部19が、別のWebサイトに前記サービスの広告を挿入する時期を決定する処理を例にあげて説明する。

【0123】

広告挿入部19は、まず、時間種特性生成部13が生成した時間種特性を示すデータを用いて、Webサイトの広告と前記サービスに対する広告反応行動の予測データを生成する(Op501)。

【 0 1 2 4 】

例えば、時間種特性を示すデータが、固有日種群および係数行列である場合について説明する。広告挿入部 19 は、固有日種群から、予測する期間における日種因子行列を生成する。例えば、予測する期間が 2007 年 1 月 1 日から 2007 年 12 月 31 日である場合、広告挿入部 19 は、固有日種群に含まれる各日種について、2007 年 1 月 1 日から 2007 年 12 月 31 日の該当する日を「1」、それ以外の日を「0」とする値の列である日種因子値を生成する。1 つの日種の日種因子値を 1 列として、固有日種群に含まれる全ての日種について列を並べた行列を、予測に用いる日種因子行列 D_f とする。広告挿入部 19 は、この日種因子行列 D_f と係数行列とを用いて、例えば、下記式 (8) により、2007 年 1 月 1 日から 2007 年 12 月 31 日における Web サイトの広告反応行動の発生数の予測値を示す行列 Y_f を計算することができる。

10

【 0 1 2 5 】

$$Y_f = D_f \cdot A \cdot P^T + e \cdot x \quad (8)$$

【 0 1 2 6 】

なお、上記式 (8) において、 P^T および $e \cdot x$ は、上記式 (5) と同様である。

【 0 1 2 7 】

次に、広告挿入部 19 は、行列 Y_f が示す予測値を参照して、Web サイトの広告反応行動の発生数が極大となる時期を特定する (Op 502)。そして、広告挿入部 19 は、Web サイトの広告反応行動の発生数が極大となる時期に基づいて、広告挿入時期を決定する (Op 503)。例えば、広告挿入部 19 は、極大となる時期の前後一定期間を含む時期を広告挿入時期と決定することができる。広告挿入部 19 は、決定した時期に、Web サーバ 2 に対して広告を挿入する処理を実行する。これにより、タイムリーな広告の挿入が可能になる。

20

【 0 1 2 8 】

なお、また、広告挿入時期の決定方法はこの例に限られない。例えば、広告挿入部 19 は、Web サイトで提供されるある 1 のサービスの広告反応行動の発生数の予測値が極大となる時期と、他のサービスの広告反応行動の発生数の予測値が極大となる時期とが重なる場合には、広告挿入部 19 は、1 のサービスの広告反応行動発生ピークが、前記他のサービスの広告反応行動発生ピークと重ならないように、誘導広告を挿入する時期を決定する。例えば、予測値が極大となるのが前記 1 のサービスも他のサービスも共に水曜日であった場合、広告挿入部 19 は、週ごとに、前記 1 のサービスの広告と前記他のサービスの広告とが入れ替わり Web サイトに掲載されるように、広告挿入時期を決定してもよい。これにより、例えば、今週の水曜日は前記 1 のサービスの広告、来週の水曜日は前記他のサービスの広告が挿入されるということになる。その結果、利用者は混雑を回避できるので、顧客満足度を高めるような誘導広告が可能となる。

30

【 0 1 2 9 】

また、広告挿入部 19 は、利用者者解析部 16 による解析結果を用いて、ある Web サイトの広告が訴求した利用者群の行動パターン、当該利用者群の広告非実行時期における行動パターン、当該利用者群の広告非実行時期における行動パターンに付随する行動の行動パターン等を抽出することができる。これにより、該広告商品に敏感な利用者が行う他の行動の共通特性を発見し、さらに効率的な広告の開発に活用することが可能となる。

40

【 0 1 3 0 】

例えば、ある Web サイト A の広告に対する広告反応行動の時間種特性と、同じ時間種特性 (例えば、「日曜」) を持つ利用者群が、利用者解析部 16 の解析によって特定されたとする。この利用者群は、該広告反応行動だけではなくて他の Web サイト B や Web サイト C も一緒に回遊して帰って行動をすること利用者群の行動パターンを示すデータとして記録されているとする。その場合、広告挿入部 19 は、この利用者群は、サービス場 B 内における特定の行動パターン X (例えば、サイト内をぐるぐる徘徊した後に、やっと購入する行動パターンとする) を抽出することができる。さらに、広告挿入部 19 は、Web サイトに該広告を掲載していない時期 (非実行時期) にも、例えば、この行

50

動パターンXがWebサイトAで一定数以上出現するかどうかを解析することができる。これにより、利用者群の、広告の非実行時期における、WebサイトAの利用状況を示すデータが得られる。

【0131】

また、広告挿入部19は、該広告の非実行時期において行動パターンXを引き起こす利用者群が、行動パターンXと同時に、他に起している行動パターンYが見つければ、それを、利用者群を特徴付ける別の行動パターンとして記録してもよい。広告挿入部19は、この行動パターンYを見張ることにより利用者群の利用状況を解析することができる。

【0132】

なお、図18においては、広告挿入時期を決定する処理を例示したが、同様に、サービス提供時期、キャンペーン実施時期等、Webサイトで実現されるイベントの発生時期を、利用者の時間種特性に応じて決定することができる。

【0133】

[第2の実施形態]

本実施形態は、上記第1の実施形態において、図10に示した動作の変形例である。図10に示した動作では、時間種因子抽出部9が時間種を抽出し、時間種特性生成部13が時間種特性を示すデータを生成する処理が、変動データ生成部8が生成した主成分得点行列Tについて行われる例である。本実施形態では、主成分得点行列Tに含まれる主成分(特徴基底)ごとに、図10に示した動作が行われる。

【0134】

図19は、本実施形態における、時間種因子抽出部9が時間種を抽出し、時間種特性生成部13が、時間種特性を示すデータを生成する処理の具体的例を示すフローチャートである。図19に示す処理は、図4におけるOp3に相当する処理である。

【0135】

図19に示すように、まず、時間種因子抽出部9は、変動データ生成部8が生成した主成分得点行列Tに含まれるp個の特徴基底それぞれについて、主成分得点を表すベクトル t_k をp個生成する(Op321)。ここで、主成分得点行列Tは、変動データ生成部8において、上述の通り、式(2)で計算される行列である。行列Tの各列が、主成分得点を表すベクトル t_k となる。

【0136】

次に、時間種因子抽出部9および時間種特性生成部13は、Op321で生成された主成分得点ベクトル t_k のうち1つについて、図10に示すOp301~Op311の処理を行う。そして、主成分得点行列Tに含まれるp個全ての特徴基底について処理が終了するまで(Op322でYesと判断されるまで)、Op301~Op311の処理が繰り返される。これにより、p個の特徴基底それぞれに対応する、日種因子行列 D_k およびその係数行列 A_k の組がp組計算される。

【0137】

なお、本実施形態では、Op307における上記式(6)は、下記式(9)のように表される。

【0138】

$$t_k = D_k(n) \cdot A_k(n) + b \quad (9)$$

【0139】

すなわち、上記式(6)における行列T、係数行列Aが、ベクトル t_k 、係数ベクトル $A_k(n)$ となる。日種因子行列 $D_k(n)$ は、特徴基底それぞれについて生成される日種因子行列である。

【0140】

各特徴基底に対応する日種因子行列 D_k およびその係数ベクトル A_k の組がp組計算されると、時間種特性生成部13は、p組の日種因子行列 D_k およびその係数ベクトル A_k を合成して、1組の合成日種因子行列 D_G およびその係数ベクトル A_G を生成する。時間種特性

10

20

30

40

50

生成部 13 は、例えば、各特徴基底の寄与率の値に基づく加重平均を p 組の日種因子行列 D_k およびその係数ベクトル A_k それぞれについて計算することで合成日種因子行列 D_G およびその係数ベクトル A_G を生成することができる。このような合成日種因子行列に対応する日種およびその係数ベクトルを用いて予測値を計算することで、特徴的な日種による精度の高い予測が可能になる。

【 0 1 4 1 】

図 20 は、各特徴基底に対応する日種因子行列 D_k およびその係数行列 A_k の組を合成する処理の例を示すフローチャートである。なお、これらを合成する方法は、図 20 に示す例に限られない。図 20 に示す例は、まず、時間種特性生成部 13 は、 p 個の特徴基底それぞれに対応する日種因子行列 D_k およびその係数ベクトル A_k の組 (p 組) を全て取得する (Op 3 2 3 1)。また、時間種特性生成部 13 は、 p 個の特徴基底それぞれの寄与率 r_k も取得する (Op 3 2 3 2)。そして、時間種特性生成部 13 は、 p 組みの日種因子行列 D_k およびその係数ベクトル A_k について、取得した寄与率 r_k を用いた加重平均をとり、合成日種因子行列 D_G およびその係数ベクトル A_G を計算する (Op 3 2 3 3)。合成日種因子行列 D_G およびその係数ベクトル A_G 、例えば、下記式 (10)、(11) により計算される。

【 0 1 4 2 】

$$D_G = (r_k \cdot D_k) / k \quad (10)$$

$$A_G = (r_k \cdot A_k) \quad (11)$$

【 0 1 4 3 】

以上のように、合成日種因子行列 D_G およびその係数ベクトル A_G が計算されると、評価部 14 が、それらを用いて計算される予測値の予測精度を計算する (Op 3 2 4)。予測精度の計算は、上述した計算例と同様であるので説明を省略する。時間種特性生成部 13 は、評価部 14 により計算された予測精度が、閾値以下である場合は、時間種の抽出条件を変えて、図 20 に示す処理を再度、実行し、合成日種因子行列 D_G およびその係数ベクトル A_G を求める処理を繰り返してもよい。

【 0 1 4 4 】

また、時間種因子抽出部 9 および時間種特性生成部 13 は、同じ Web サイトの利用度の変動を表す主成分得点行列 T に対して、図 20 に示す処理と、第 1 の実施形態で述べた図 10 に示す処理との両方を実行し、予測精度のよい方の計算結果を、時間種特性記録部 15 に記録するべき時間種特性を示すデータとして採用することができる。

【 0 1 4 5 】

また、図 20 に示した処理において、特徴基底ごとの日種因子行列 D_k およびその係数ベクトル A_k を合成する処理 (Op 3 2 3 3) は、必ずしも実行する必要はない。時間種特性生成部 13 は、例えば、特徴基底ごとの日種因子行列 D_k およびその係数ベクトル A_k を、時間種特性を示すデータとして時間種特性記録部 15 に記録してもよい。

【 0 1 4 6 】

[第 3 の実施形態]

本実施形態は、上記第 1 の実施形態において、図 9 に示した動作の変形例である。図 9 に示した動作は、変動データ生成部 8 が変動データとして、主成分得点行列 T を生成する動作例である。本実施形態では、変動データ生成部 8 は、特徴的な変動を示す特徴区間を抽出し、当該特徴区間についての主成分得点行列 T_b を生成する。

【 0 1 4 7 】

図 21 は、変動データ生成部 8 が、特徴的な変動を示す特徴区間を抽出し、当該特徴区間についての主成分得点行列 T_b を生成する処理の一例を示すフローチャートである。図 21 に示す処理は、図 9 に示す Op 2 3 の処理の変形例である。ここでは、一例として、上記第 1 の実施形態で説明した行列 X に対して主成分分析を施して特徴区間を抽出し、当該特徴区間についての主成分得点行列 T_b を生成する場合について説明する。

【 0 1 4 8 】

図 21 に示す処理では、まず、変動データ生成部 8 は、行列 X に対して主成分分析を施

10

20

30

40

50

し、特徴基底行列 P ($M \times p$ 行列) を生成する ($Op231$)。特徴基底行列 P ($M \times p$ 行列) を生成する処理の具体例は上述した通りである。特徴基底行列 P の各列は、 M 個の要素で構成されており、 M 個の要素は、1 日における Web サイトの利用度変動を示す値が時間遷移する順に M 個並んで構成されている。

【0149】

次に、変動データ生成部 8 は、特徴基底行列 P の第 2 主成分の要素列 (2 番目に寄与率が高い主成分の列) から、第 1 主成分の要素列 (最も寄与率が高い主成分の列) を引いた差分列を生成し、2 次元座標平面上にプロットする ($Op232$)。

【0150】

図 22 (a) は、主成分得点行列 T の第 1 主成分の要素列および第 2 主成分の要素列を 2 次元の XY 座標平面上にプロットしたグラフの例を示す図である。図 22 (a) において、 Y 軸は各要素の値、 X 軸は 1 日分の時刻を表す。図 22 (a) では、折れ線 w_1 が第 1 主成分の要素列を繋いだ折れ線を表し、折れ線 w_2 が第 2 主成分の要素列を繋いだ折れ線を表している。各折れ線 w_1 、 w_2 は、1 日分の Web サイトの利用度の変動を表している。図 22 (b) は、図 22 (a) に示す第 2 主成分の要素列から、第 1 主成分の要素列を引いた差分列を、2 次元の XY 座標平面上にプロットしたグラフの例を示す図である。図 22 (b) において、 Y 軸は差分値、 X 軸は 1 日分の時刻を表す。図 22 (b) において、折れ線 u_2 が、第 2 主成分の要素列の第 1 主成分の要素列に対する差分値を繋いだ折れ線を表している。

【0151】

変動データ生成部 8 は、差分値の符号が変化する点 (すなわち、 $Y = 0$ の軸と折れ線 u_2 との交点) を堺にして、座標平面を X 軸方向に複数の区間に分割する ($Op233$)。これにより、1 日分の差分値を示す折れ線 u_2 は、各区間に相当する複数の時間帯に分割される。図 22 (c) は、図 22 (b) に示す座標平面を X 方向に複数の区間に分割した場合の例を示す図である。図 22 (c) に示すように、 $Y = 0$ 軸と折れ線 u_2 との交点を通る Y 軸に平行な線 (点線で示される線) によって、座標平面が X 軸方向に複数の区間に分割されている。

【0152】

変動データ生成部 8 は、 $Op233$ で分割した各区間について、独自性を示す値を計算する ($Op234$)。1 つの区間における独自性を示す値は、例えば、その区間の x 方向の長さ、その区間における折れ線 u_2 と $Y = 0$ との間の面積 S_k 、その区間における差分値の最大絶対値 UK_{max} 、その区間における折れ線 u_2 の突出率 TS_k が含まれる。突出率 TS_k は、上記 UK_{max} および S_k を用いて、下記式 (12) で計算される値である。

【0153】

$$TS_k = UK_{max} / S_k \quad (12)$$

【0154】

図 22 (d) は、突出率の異なる 2 つの区間例を示す図である。図 22 (d) に示す例では、区間 1 における突出率の方が、区間 2 の突出率より大きくなる。なぜなら、区間 1 において差分値を表す折れ線 u_2 は鋭角的な形をしており、区間 1 における折れ線 u_2 と $Y = 0$ 軸との間の面積に対する差分値の最大絶対値 UK_{max} の比率が、区間 2 より高いからである。

【0155】

変動データ生成部 8 は、1 つの区間について計算した上記独自性を示す値を、それぞれ閾値と比較し、独自性を示す値のうち少なくとも 1 つが閾値を超えるか否かを判断する。すべての区間について、この判断を行い、独自性を示す値のうち少なくとも 1 つが閾値を超える区間が存在すれば ($Op2351$)、その区間を特徴区間とする。変動データ生成部 8 は、特徴区間特徴区間における主成分得点を、 p 個の特徴基底それぞれについて抽出して各列とする主成分得点行列 T_b を生成する ($Op236$)。 $Op2351$ において、該当する区間が存在しない (No) と判断された場合は、全区間について主成分得点行列を生成する ($Op237$)。

10

20

30

40

50

【 0 1 5 6 】

以上、図 2 1 に示した処理により、変動データ生成部 8 は、特徴区間を抽出して、その特徴区間についての主成分得点行列 T_b を生成することができる。なお、特徴期間の抽出方法は、図 2 1 に示した方法に限られない。例えば、Op 2 3 4 において計算される独自性を示す値には、上記例で示した値以外の値が用いられてよいし、上記の値のうちの一部が用いられてもよい。

【 0 1 5 7 】

[第 4 の実施形態]

本実施形態では、図 1 の機能ブロック図に示す時間種因子抽出部 9 が、時間種データ記録部 1 1 から抽出した時間種に対応する時間種因子値を主成分分析することにより主成分得点行列を生成する機能をさらに備える。また、時間種特性生成部 1 3 は、時間種因子抽出部 9 が生成した主成分得点行列を説明変数として時間種特性を示すデータの生成に用いる。

【 0 1 5 8 】

具体的な例として、時間種因子抽出部 9 が、図 1 0 に示したフローチャート中の Op 3 0 6 において作成した日種因子行列 $D(n)$ に対して、主成分分析の処理を施すことにより主成分得点の行列に圧縮した日種因子行列 $D_A(n)$ を生成する処理をさらに行う例を挙げる。これにより、Op 3 0 7 において、時間種特性生成部 1 3 は、圧縮された日種因子行列 $D_A(n)$ を用いて計算を行うことができるので、計算量が減るといった効果が生じる。

【 0 1 5 9 】

図 2 3 は、本実施形態において、時間種因子抽出部 9 が、日種因子行列 $D(n)$ を圧縮する処理の一例を示すフローチャートである。時間種因子抽出部 9 は、日種因子行列 $D(n)$ を、主成分分析を施し、複数の特徴基底および寄与率を生成する (Op 6 0 1)。時間種因子抽出部 9 は、複数の特徴基底のうち、寄与率が閾値を超える特徴基底を主要な特徴基底として抽出する。そして、時間種因子抽出部 9 は、主要な特徴基底を用いて、主成分得点行列 $D_T(n)$ を生成する (Op 6 0 2)。主成分得点行列 $D_T(n)$ の各列の要素は、各特徴基底に対応する日種ごとの主成分得点が並んで構成されている。時間種因子抽出部 9 は、主要な特徴基底のうち、1 つの特徴基底を選択する (Op 6 0 3)。そして、時間種因子抽出部 9 は、選択した特徴基底に対応する主成分得点行列 $D_T(n)$ の列の主成分得点のうち、絶対値が閾値を超える主成分得点に対応する日種を抽出する (Op 6 0 4)。

【 0 1 6 0 】

図 2 4 (a) は、主成分得点行列 $D_T(n)$ の第 1 主成分 (寄与率が 1 番高い特徴基底) の列の主成分得点を、図 2 4 (b) は、第 2 主成分 (寄与率が 2 番目に高い特徴基底) の列の主成分得点を、図 2 4 (c) は、第 3 主成分 (寄与率が 3 番目に高い特徴基底) の列の主成分得点を、それぞれ 2 次元座標平面にプロットしたグラフの例を示す図である。図 2 4 (a) ~ (c) に示すグラフにおいて、縦軸は主成分得点を、横軸は日種 (第 1 日種 ~ 第 3 3 日種) を表す。図 2 4 (a) ~ (c) に示すグラフにおいて、ある主成分における 1 つの日種の主成分得点は、その日種の性質がどの程度、その主成分に受け継がれているかを示している。

【 0 1 6 1 】

図 2 4 (a) に示す第 1 主成分のグラフでは、第 1 7 日種、第 2 6 日種、第 3 2 日種において、主成分得点の高いピークが見られる。例えば、第 1 7 日種、第 2 6 日種、第 3 2 日種がそれぞれ、「8 月」、「夏」、「残暑」である場合、第 1 主成分は、8 月、夏、残暑の日種の性質を強く受け継いでいることになる。

【 0 1 6 2 】

時間種因子抽出部 9 が、例えば、図 2 4 (a) に示す第 1 主成分について、絶対値が閾値を超える主成分得点に対応する日種を抽出する場合、例えば、閾値が 0.3 であるとすると、第 1 7 日種、第 2 6 日種、第 3 2 日種、すなわち、「8 月」、「夏」、「残暑」の

10

20

30

40

50

日種が抽出される。

【0163】

時間種因子抽出部9は、抽出した日種の主成分得点が正の場合は「1」を、負の場合は「-1」を付与する(Op605)。そして、時間種因子抽出部9は、抽出した日種のうち、「-1」が付与された日種の意味的な距離を求め、当該意味的な距離が閾値以上の日種を抽出した日種から除く。例えば、図24(b)に示すグラフにおいて、第15日種、第26日種、第30日種が抽出された場合、第26日種には「-1」が付与される。時間種因子抽出部9は、第26日種の意味的な距離を求める。意味的な距離とは、日種が持つ性質がその主成分に与える影響の度合いを表す値である。例えば、意味的な距離の値は、予め記録されたルールに基づいて決定されてもよいし、ユーザインタフェースを介してユーザからの意味的な距離の値の入力を受け付けることによって決定されてもよい。

10

【0164】

ここで、時間種因子抽出部9が、予め記録されたルールに基づいて意味的な距離を決定する例を説明する。例えば、上記の第15日種 = 「火曜」、第26日種 = 「金曜」、第30日種 = 「木曜」であり、予め記録されたルールが下記の場合、時間種因子抽出部9は、第15日種と第26日種との意味的な距離を、相対距離グレード「C」(=遠い)と決定することができる。

(ルール)

週中 = 「火曜、水曜、木曜」

週末 = 「金曜」

「週中」と「週末」の相対距離グレード = C

ここで、相対距離グレードは、例えば、A = 類似、B = 中間、C = 遠いとする。

20

【0165】

時間種因子抽出部9は、抽出した日種を複合して複合日種を生成する(Op607)。例えば、図24(a)に示すように、第1主成分において、「8月」、「夏」、「残暑」の日種が抽出された場合、これら日種を複合した複合日種「8月&夏&残暑」を生成する。

【0166】

以上、Op603 ~ Op607の処理を、時間種因子抽出部9は、主成分得点行列 $D_T(n)$ (n)の全ての列(特徴基底)について終了するまで(Op608でYesと判断されるまで)繰り返す。これにより、主成分得点行列 $D_T(n)$ に含まれる各特徴基底についてそれぞれ複合日種が生成される。

30

【0167】

時間種因子抽出部9は、各特徴基底について生成された複合日種の主成分得点を計算し、複合日種の主成分得点で構成された日種因子行列 $D_A(n)$ を生成する(Op609)。これにより、日種因子行列 $D(n)$ において対象期間の日数分あった列の数は特徴基底の数に圧縮され、日種因子行列 $D(n)$ の行数は上記Op607の処理で生成された複合日種の数に圧縮される。このように圧縮された日種因子行列 $D_A(n)$ は、図10に示すOp307の計算において、時間種特性生成部13が、説明変数として用いる。これにより、時間種特性生成部13の計算負荷が軽減する。

40

【0168】

なお、時間種因子抽出部9が、日種因子行列 $D(n)$ を圧縮する処理は、図23に示した処理に限られない。例えば、Op603 ~ Op609の処理を省略して、Op602で生成した主成分得点行列 $D_T(n)$ を、圧縮された日種因子行列をして、時間種特性生成部13に渡してもよい。

【0169】

また、図23に示す日種因子行列 $D(n)$ の圧縮処理と同様の処理を用いて、図10におけるOp302で、時間種因子抽出部9が作成した時間種因子母群の日種因子を圧縮することもできる。

【0170】

50

[第 5 の実施形態]

本実施形態は、図 1 の機能ブロック図に示す時間種因子抽出部 9 が、変動データ生成部 8 が生成した主成分得点行列 T と、時間種データ記録部 11 に記録された複数の時間種のうち少なくとも 1 つの時間種に対応する時間種因子値との相関性を示す相関係数を計算し、当該相関係数に基づいて時間種を抽出する機能をさらに備える形態である。

【 0 1 7 1 】

具体例として、時間種因子抽出部 9 が、図 10 に示す Op 3 0 2 において、時間種因子母群を作成する際に、変動データ生成部 8 が生成した主成分得点行列 T と、時間種データ記録部 11 に記録された複数の時間種の時間種に対応する時間種因子値との相関性を示す相関係数を計算し、当該相関係数に基づいて時間種因子母群に追加する時間種を抽出する
10

【 0 1 7 2 】

図 2 5 は、本実施形態において、時間種因子抽出部 9 が、時間種因子母群に追加する時間種を抽出する処理の一例を示すフローチャートである。図 2 5 に示すように、時間種因子抽出部 9 は、変動データ生成部 8 が生成した主成分得点行列 T を取得する (Op 3 2 0 1)。主成分得点行列 T は、例えば、第 1 の実施形態で述べたように、図 9 に示す処理で計算される。主成分得点行列 T は、日数分 (N 日分) に相当する N 行と、第 1 主成分 ~ 第 p 主成分に相当する p 列とで構成される $N \times p$ 行列である。

【 0 1 7 3 】

時間種因子抽出部 9 は、主成分得点行列の 1 つの主成分 (例えば、第 1 主成分とする)
20 に対応する列の要素を、複数の区間に分割する (Op 3 2 0 2)。時間種因子抽出部 9 は、例えば、主成分得点の正負が変化する点を境として、複数の区間に分割することができる。図 2 6 (a) は、主成分得点行列の 1 つの主成分における主成分得点を 2 次元座標平面にプロットしたグラフの例を示す図である。図 2 6 (a) において、縦軸は主成分得点を表し、横軸は日を表す。折れ線 q は、日ごとの主成分得点を繋ぐ線である。

【 0 1 7 4 】

なお、ここでは、一例として、主成分得点行列の 1 つの主成分 (例えば、第 1 主成分とする) に対応する列の要素が、処理の対象となる場合を説明するが、処理の対象となる要素はこれに限られない。例えば、主成分得点行列の全ての主成分に対応する列の要素を処理の対象とすることもできる。
30

【 0 1 7 5 】

図 2 6 (b) は、時間種因子抽出部 9 が、図 2 6 (a) に示す折れ線 q の主成分得点 = 0 の線との交点を境として、折れ線 q を横軸方向に複数の区間に分割した場合の例を示す図である。図 2 6 (b) において、点線は、区間の境界を示している。

【 0 1 7 6 】

時間種因子抽出部 9 は、各区間において、主成分得点の絶対値が最大となる日の要素を、その主成分得点の値 (以下、ピーク値と称する) とし、その他の日の要素を「 0 」とするピークベクトルを生成する (Op 3 2 0 3)。図 2 6 (c) は、図 2 6 (b) に示す各区間におけるピーク値を矢印により示す図である。なお、ここで、ピークベクトルに含まれるピーク値のうち、絶対値が閾値に満たないピーク値は「 0 」にするように、ピークベクトルを修正することもできる。これにより、小さなピーク値が無視されるので、後述する相関係数の計算によって相関関係を見つけやすくなる場合もある。
40

【 0 1 7 7 】

時間種因子抽出部 9 は、さらに、ピークベクトルにおける負の値のピーク値を「 0 」にして、正の値のピーク値を残した正ピークベクトルと、ピークベクトルにおける正の値のピーク値を「 0 」にして、負の値のピーク値を残した負ピークベクトルとを生成する (Op 3 2 0 4)。

【 0 1 7 8 】

そして、時間種因子抽出部 9 は、時間種データ記録部 11 から、処理対象とする 1 つの日種を選択し、その日種の日種因子値を要素とする日種因子ベクトルを取得する (Op 3
50

205)。時間種因子抽出部9は、日種因子ベクトルと、ピークベクトルとの相関係数を計算する(Op3206)。時間種因子抽出部9は、例えば、日種因子ベクトルとピークベクトルとの内積を、日種因子ベクトルの大きさとピークベクトルの大きさの積で割った値(すなわち \cos)を相関係数として求めることができる。時間種因子抽出部9は、同様にして、日種因子ベクトルと正ピークベクトルとの相関係数、日種因子ベクトルと負ピークベクトルとの相関係数も計算する。これらの相関係数は、Webサイトの利用度の変動を表す主成分得点行列Tとの相関性を示す値である。

【0179】

時間種因子抽出部9は、上記の3つの相関係数それぞれを閾値と比較することにより、主成分得点行列Tとの相関性の高い日種を抽出することができる。例えば、上記3つの相関係数のうち少なくとも1つが閾値より高い場合(Op3207でYesの場合)、その日種を有望な候補として、例えば、図11に示した母群抽出条件テーブル25のような時間種因子母群作成のための抽出手順データに記録する(Op3208)。

10

【0180】

時間種因子抽出部9は、上記のOp3205~Op3208の処理を、例えば、時間種データ記録部11に記録された全ての日種について終了するまで(Op3209でYesと判断されるまで)、繰り返す。これにより、時間種データ記録部11に記録された日種から、主成分得点行列Tと相関性の高く、説明変数として有望な日種が時間種因子母群に追加される。

【0181】

20

なお、時間種因子抽出部9が、変動データと時間種因子値との相関性を用いて時間種候補を抽出する処理は、図25に示した処理に限られない。また、時間種因子抽出部9は、図25に示す時間種候補を抽出する処理と同様の処理を用いて、例えば、図10におけるOp303で、抽出リストに記録すべき日種の組み合わせを抽出してもよい。

【0182】

以上、上述した第1~第5の実施形態は、本発明の実施する場合の具体例であり、これに限定されない。例えば、時間種特性生成システム1において生成された時間種を示すデータは、上記の広告挿入部19、利用者解析部16による処理以外にも、Webサイト利用度の変動を予測する予測モデルとして広く利用することができる。また、時間種特性生成システム1による時間種特性解析の対象となるサイトは、Webサイトに限られず、例えば、LAN等、限られた地域におけるネットワークにおけるサイトも含まれる。

30

【0183】

以上の実施形態に関し、更に以下の付記を開示する。

(付記1)

ネットワークに接続されたコンピュータにより提供される少なくとも1つのサイトの利用状況を示すログデータから、所定期間における前記サイトの利用度の変動を示す遷移データを抽出する利用状況把握部と、

前記遷移データに基づいて、前記所定期間内の前記サイトの利用度の変動を、前記サイトの利用度を示す値の集合Yで表す変動データを生成する変動データ生成部と、

複数の時間種を示すデータと、前記複数の時間種それぞれで特定される期間を表す時間種因子値とが記録された時間種データ記録部と、

40

時間種データ記録部に記録された複数の時間種から、予め記録された手順に従って時間種を抽出し、前記抽出した時間種に対応するi個の時間種因子値の集合Diを生成する時間種因子抽出部と、

前記集合Diを説明変数として、集合Yを被説明変数とする下記(数1)において、誤差bの二乗が最小になるようなi個の係数aiを計算することによって、前記集合Diに対応する時間種および前記集合Diそれぞれの前記係数aiを含むデータを、前記サイトの利用状況の時間種特性を示すデータとして生成する時間種特性生成部とを備える、時間種特性生成システム。

【0184】

50

【数 1】

$$Y = \sum_i a_i D_i + b$$

【0185】

(付記 2)

前記利用状況把握部は、前記サイトが提供するサービスの利用度の変動を表す遷移データを抽出する、付記 1 に記載の時間種特性生成システム。

(付記 3)

前記利用状況把握部は、前記サイトが提供する広告に対するアクセス量の変動を表す遷移データを抽出する、付記 1 に記載の時間種特性生成システム。 10

(付記 4)

前記利用状況把握部は、少なくとも 1 つのサイトが提供する特定の広告にアクセスし、かつ当該サイトまたは他のサイトが提供する前記特定の広告に関するサービスを利用する行動の発生数の変動を表す遷移データを抽出する、付記 1 に記載の時間種特性生成システム。

(付記 5)

前記時間種データ記録部は、連続する時間を日単位で複数の日に区切った場合の当該複数の日のうち、1 つの時間種で特定される日を表す値である時間種因子値を、複数の時間種について記録する、付記 1 に記載の時間種特性生成システム。 20

(付記 6)

前記時間種データ記録部は、連続する時間を時間単位で複数の期間に区切った場合の当該複数の時間帯のうち、1 つの時間種で特定される時間帯を表す値である時間種因子値を、複数の時間種について記録する、付記 1 に記載の時間種特性生成システム。

(付記 7)

利用者の属性を示すデータと、当該属性を持つ利用者による利用状況の時間種特性を示すデータとを対応付けて記録する利用者時間種特性記録部と、

前記時間種特性生成部が生成した時間種特性を示すデータと、前記利用者時間種特性記録部に記録された時間種特性を示すデータとを比較することにより、前記サイトを利用する利用者の属性を特定する利用者解析部とをさらに備える、付記 1 に記載の時間種特性生成システム。 30

(付記 8)

前記時間種因子抽出部は、前記時間種データ記録部に記録された複数の時間種から、時間種の組み合わせを複数通り抽出して、各組み合わせに対応する時間種因子値の集合 D_i を複数通り生成し、

前記時間種特性生成部は、前記複数通りの時間種因子値の集合 D_i それぞれについて、前記係数 a_i を計算することによって複数組の係数 a_i および時間種因子値の集合 D_i を生成し、当該複数組から所定基準に基づき選択した組の係数 a_i と、時間種因子の集合 D_i に対応する時間種とを含むデータを、前記サイトの利用状況の時間種特性を示すデータとして生成する、付記 1 に記載の時間種特性生成システム。 40

(付記 9)

前記変動データ生成部は、前記遷移データが示す利用度の変動を、利用度の多変量データで表し、当該多変量データを主成分分析することにより、特徴基底を抽出して当該特徴基底の主成分得点を集合 Y とし、

前記時間種因子抽出部は、前記時間種データ記録部に記録された複数の時間種から抽出した時間種に対応する時間種因子値の多変量データを生成し、当該当該多変量データを主成分分析することにより、特徴基底を抽出して当該特徴基底の主成分得点を集合 D_i とする、付記 1 に記載の時間種特性生成システム。

(付記 10)

前記変動データ生成部は、前記遷移データが示す利用度の変動を、利用度の多変量デー 50

タで表し、当該多変量データを主成分分析することにより、特徴基底を抽出し、当該抽出した特徴基底のうち1の特徴基底が、他の特徴基底に対して所定量以上異なっている特徴区間をさらに抽出し、当該特徴区間における前記特徴基底の主成分得点を集合 Y とし、

前記時間種因子抽出部は、前記特徴区間内に含まれる時間種因子値の集合を、前記集合 D_i として生成する、付記1に記載の時間種特性生成システム。

(付記11)

前記変動データ生成部は、前記遷移データが示す利用度の変動を、利用度の多変量データで表し、当該多変量データを主成分分析することにより、複数の特徴基底を抽出し、抽出した前記複数の特徴基底それぞれの主成分得点を集合 Y として、前記複数の特徴基底ごとに集合 Y を生成し、

10

前記時間種因子抽出部は、前記複数の特徴基底それぞれについて、前記集合 D_i を生成し、

前記時間種特性生成部は、前記複数の特徴基底ごとの集合 Y について、それぞれの前記集合 D_i を用いて、係数 a_i を計算し、前記複数の特徴基底ごとの集合 D_i および係数 a_i を合成することにより、前記サイトの利用状況の時間種特性を示すデータとして生成する、付記1に記載の時間種特性生成システム。

(付記12)

前記変動データ生成部は、前記遷移データが示す利用度の変動を、利用度の多変量データで表し、当該多変量データを主成分分析することにより、特徴基底を抽出して当該特徴基底の主成分得点を集合 Y とし、

20

前記時間種因子抽出部は、前記変動データ生成部が抽出した特徴基底の主成分得点と、前記時間種データ記録部に記録された複数の時間種のうち少なくとも1つの時間種に対応する時間種因子値との相関性を示す相関係数を計算し、当該相関係数に基づいて、時間種を抽出する、付記1に記載の時間種特性生成システム。

(付記13)

ネットワークに接続されたコンピュータにより提供される少なくとも1つのサイトの利用状況を示すログデータから、所定期間における前記サイトの利用度の変動を示す遷移データを抽出するステップと、

前記遷移データに基づいて、前記所定期間内の前記サイトの利用度の変動を、前記サイトの利用度を示す値の集合 Y で表す変動データを生成するステップと、

30

複数の時間種を示すデータと、前記複数の時間種それぞれで特定される期間を表す時間種因子値とが記録された時間種データ記録部に記録された複数の時間種から、予め記録された手順に従って時間種を抽出し、前記抽出した時間種に対応する i 個の時間種因子値の集合 D_i を生成するステップと、

前記集合 D_i を説明変数として、集合 Y を被説明変数とする上記(数1)において、誤差 b の二乗が最小になるような i 個の係数 a_i を計算することによって、前記集合 D_i に対応する時間種および前記集合 D_i それぞれの前記係数 a_i を含むデータを、前記サイトの利用状況の時間種特性を示すデータとして生成するステップとを含む、時間種特性生成方法。

(付記14)

40

ネットワークに接続されたコンピュータにより提供される少なくとも1つのサイトの利用状況を示すログデータから、所定期間における前記サイトの利用度の変動を示す遷移データを抽出する処理と、

前記遷移データに基づいて、前記所定期間内の前記サイトの利用度の変動を、前記サイトの利用度を示す値の集合 Y で表す変動データを生成する処理と、

複数の時間種を示すデータと、前記複数の時間種それぞれで特定される期間を表す時間種因子値とが記録された時間種データ記録部に記録された複数の時間種から、予め記録された手順に従って時間種を抽出し、前記抽出した時間種に対応する i 個の時間種因子値の集合 D_i を生成する処理と、

前記集合 D_i を説明変数として、集合 Y を被説明変数とする上記(数1)において、誤

50

差 b の二乗が最小になるような i 個の係数 a_i を計算することによって、前記集合 D_i に対応する時間種および前記集合 D_i それぞれの前記係数 a_i を含むデータを、前記サイトの利用状況の時間種特性を示すデータとして生成する処理とをコンピュータに実行させる、時間種特性生成プログラム。

【産業上の利用可能性】

【0186】

本発明は、サイトの利用状況の時間種特性を定量的に示すデータを生成することができる時間種特性生成システムとして利用可能である。

【図面の簡単な説明】

【0187】

【図1】第1の実施形態にかかる時間種特性生成システムの構成を示す機能ブロック図

【図2】(a)は、1日あたり $M (= 144)$ 個の値を、 $N (= 365)$ 日分並べたデータの例を示す図である。(b)は、(a)に示すデータを、主成分分析を用いて圧縮した主成分得点を表すデータの例を示す図である。

【図3】時間種データ記録部11に記録される日種データの例を示す図

【図4】時間種特性生成システム1の動作例の大まかな流れを示すフローチャート

【図5】利用状況把握部が、遷移データを生成する処理の具体的例を示すフローチャート

【図6】(a)は、サービス群の基礎情報の一例であるサービス基礎情報テーブルを示す図である。(b)は、広告群の基礎情報の一例である広告基礎情報テーブルを示す図である。

【図7】サービス利用の行動パターンを示すデータの一例である行動パターンテーブルを示す図

【図8】遷移データの一例である遷移データテーブルを示す図

【図9】変動データ生成部が遷移データを基に変動データを生成する処理の具体例を示すフローチャート

【図10】時間種因子抽出部が時間種を抽出し、時間種特性生成部が、時間種特性を示すデータを生成する処理の具体的例を示すフローチャート

【図11】抽出手順データに含まれるデータの一例である母群抽出条件テーブルを示す図

【図12】時間種因子母群を表す母群データの例を示す図

【図13】抽出リストを表すデータの一例である抽出リストテーブルを示す図

【図14】日種因子行列 $D(n)$ と主成分特定行列 T との相関関係の有無を判定する処理の一例を示すフローチャート

【図15】行列 T_{2007} が表す変動の予測値と、実績データが表す変動の実測値とをグラフ上で表した図

【図16】時間種特性を示すデータの一例である時間種特性データを示す図

【図17】利用者解析部が、遷移データを生成する処理の具体的例を示すフローチャート

【図18】広告挿入部が、広告を挿入する時期を決定する処理の例を示すフローチャート

【図19】第2の実施形態において、時間種因子抽出部が時間種を抽出し、時間種特性生成部が時間種特性を示すデータを生成する処理の具体的例を示すフローチャート

【図20】各特徴基底に対応する日種因子行列 D_k およびその係数行列 A_k の組を合成する処理の例を示すフローチャート

【図21】変動データ生成部が、特徴的な変動を示す特徴区間を抽出し、当該特徴区間についての主成分得点行列 T_b を生成する処理の一例を示すフローチャート

【図22】(a)は、主成分得点行列 T の第1主成分および第2主成分をプロットしたグラフの例を示す図である。(b)は、(a)に示す第2主成分と第1主成分と差分列をプロットしたグラフの例を示す図である。(c)は、(b)に示す X 方向を複数の区間に分割した場合の例を示す図である。(d)は、突出率の異なる2つの区間の例を示す図である。

【図23】第4の実施形態において、時間種因子抽出部が、日種因子行列 $D(n)$ を圧縮する処理の一例を示すフローチャート

10

20

30

40

50

【図24】(a)は、主成分得点行列 $D_T(n)$ の第1主成分の列の主成分得点をプロットしたグラフを示す図である。(b)は、第2主成分の列の主成分得点をプロットしたグラフを示す図である。(c)は、第3主成分の列の主成分得点をプロットしたグラフを示す図である。

【図25】第5の実施形態において、時間種因子抽出部が、時間種因子母群に追加する時間種を抽出する処理の一例を示すフローチャート

【図26】(a)は、主成分得点行列の主成分得点のグラフを示す図である。(b)は、(a)に示す折れ線 q を複数の区間に分割した場合の例を示す図である。(c)は、(b)に示す各区間におけるピーク値を示す図である。

【符号の説明】

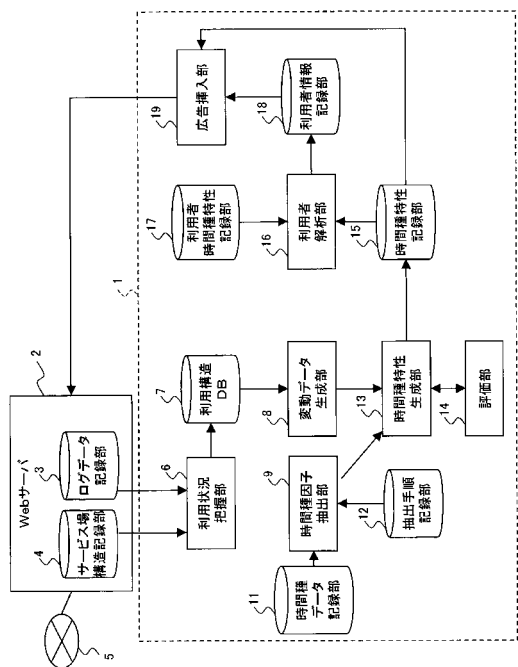
10

【0188】

- 1 時間種特性生成システム
- 2 Webサーバ
- 3 ログデータ記録部
- 4 サービス場構造記録部
- 5 インターネット
- 6 利用状況把握部
- 7 利用構造DB
- 8 変動データ生成部
- 9 時間種因子抽出部
- 11 時間種データ記録部
- 12 抽出手順記録部
- 13 時間種特性生成部
- 14 評価部
- 15 時間種特性記録部
- 16 利用者解析部
- 17 利用者時間種特性記録部
- 18 利用者情報記録部
- 19 広告挿入部

20

【 図 1 】



【 図 2 】

(a)

	M					
	0:00	0:10	0:20	...	23:50	
NB	2006/01/01	24	34	34	...	33
	2006/01/02	21	26	52	...	14
	2006/01/03	25	17	45	...	25
	2006/01/04	21	16	31	...	21

2006/12/31	20	27	36	...	26	

(b)

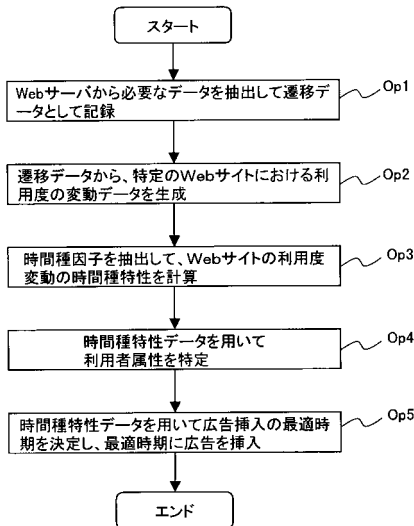
	P				
	第1主成分	第2主成分	...	第p主成分	
NB	2006/01/01	85	42	...	53
	2006/01/02	63	64	...	61
	2006/01/03	92	74	...	25
	2006/01/04	78	81	...	46

2006/12/31	47	84	...	56	

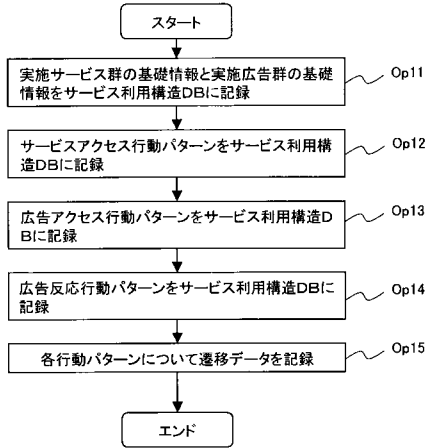
【 図 3 】

	月曜日	火曜日	水曜日	祝日/振替	寒の入り	五十日	1月	...
2006年1月1日	0	0	0	1	0	0	1	...
2006年1月2日	1	0	0	0	0	0	1	...
2006年1月3日	0	1	0	0	0	0	1	...
2006年1月4日	0	0	1	0	0	0	1	...
2006年1月5日	0	0	0	0	1	1	1	...
2006年1月6日	0	0	0	0	0	0	1	...
2006年1月7日	0	0	0	0	0	0	1	...
2006年1月8日	0	0	0	0	0	0	1	...
2006年1月9日	1	0	0	1	0	0	1	...
2006年1月10日	0	1	0	0	0	1	1	...
2006年1月11日	0	0	1	0	0	0	1	...
2006年1月12日	0	0	0	0	0	0	1	...
2006年1月13日	0	0	0	0	0	0	1	...
2006年1月14日	0	0	0	0	0	0	1	...
2006年1月15日	0	0	0	0	0	0	1	...
...
2006年12月31日	0	0	0	0	0	0	0	...

【 図 4 】



【図5】



【図6】

(a)

サービスID	サービス場	URL	アクション
s0001	A社通信販売サイト	http://www.site1/ad.html	閲覧
		http://www.site1/go.html	選択商品入力
		http://www.site1/con.html	発注ボタン押下
S0002	B社eコマースサイト	http://www.site2/ad.html	閲覧
		http://www.site2/con.html	発注ボタン押下
...

(b)

広告ID	広告場	URL	アクション
k0001	A社通信販売広告サイト	http://www.site1/home.html	閲覧
k0002	B社eコマース広告サイト	http://www.site2/home.html	閲覧
...

【図7】

23

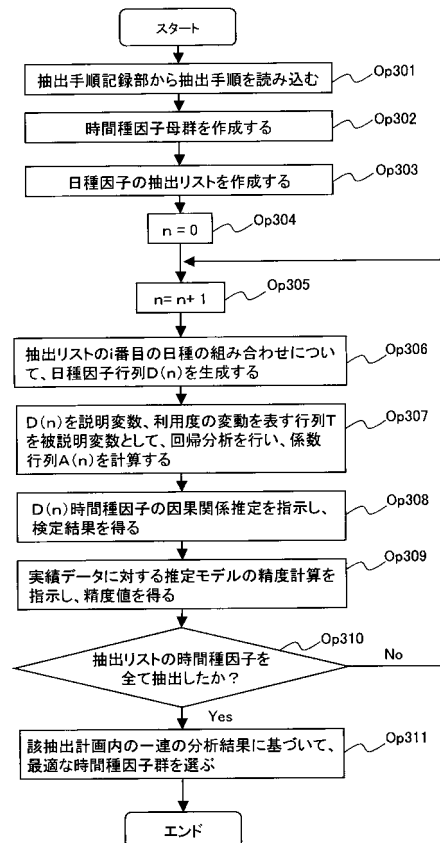
	行動パターンID=P-0001	行動パターンID=P-0002	行動パターンID=P-0003	-
A社通信販売広告サイトで閲覧	1	1	1	-
-	-	-	-	-
A社通信販売広告サイトで発注ボタン押下	1	0	1	-
-	-	-	-	-
B社eコマースサイト閲覧	1	0	0	-

【図8】

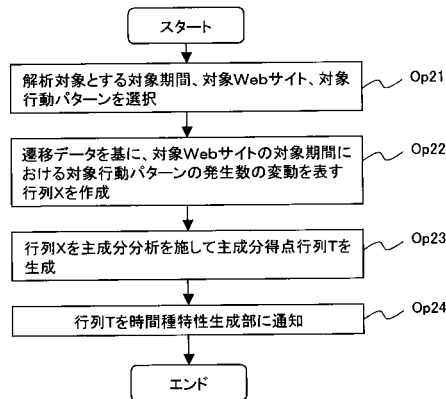
24

日時	行動パターンID
2006/03/04 14:13	P-0001
2006/03/04 14:24	P-0002
2006/03/04 14:25	P-0001
2006/03/04 15:09	P-0003
...	...

【図10】



【図9】



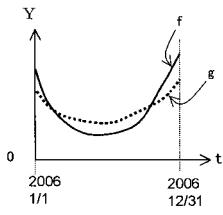
【図11】

種別	日種	複合の有無と複合条件	主成分化による複合の有無
曜日	土、日、月、～、金	土・日・月、のように連続の複合因子まで作成	しない
月	1、2、～、12	飛び飛びでもよいから3つの複合因子まで作成	する
季節	春、夏、秋、冬	作成しない	しない
祝日/振替	祝日/振替	なし	しない
期末	上期末、下期末、上半期末、下半期末	作成しない	しない

【図12】

抽出候補日種ID	種別	日種	日種因子
0001	曜日	月	010000...
0002	曜日	火	001000...
0003	曜日
0004	曜日	月火	011000...
0005	曜日	月水	010100...
...	曜日
0025	曜日	月火水	011100...
...	曜日
0061	月	1月	111111...
0062	月	2月	000000...
...	月
0073	月	月の第1主成分	111111...
0074	月	月の第2主成分	000000...
...	月
0078	祝日	祝日	100000...
0079	季節	春	000000...
...	季節
0083	期末	上期末	000000...
...

【図15】



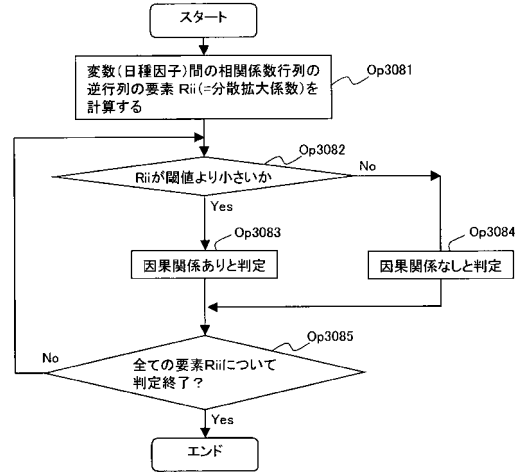
【図16】

解析ID	解析の対象	日種の組み合わせ (固有日種群)	係数行列
1	A社通信販売サイトのサービス利用行動	「月」、「火」、...、「1月」、...、「祝日」、「期末」、...	$\begin{pmatrix} A_{11} & \dots & A_{1p} \\ \vdots & \ddots & \vdots \\ A_{n1} & \dots & A_{np} \end{pmatrix}$
...

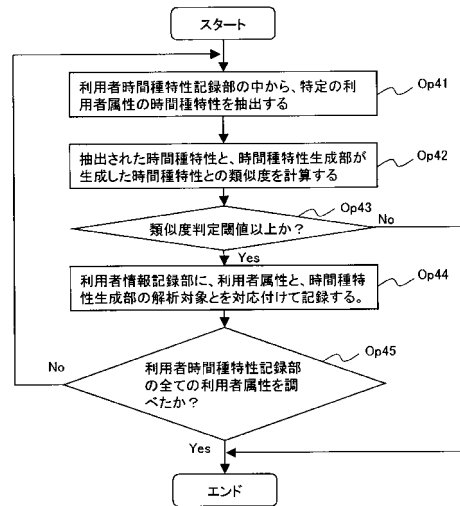
【図13】

リスト番号	日種の組み合わせ
1	0001~0009,0011~0021...
2	0001~0008,0010,0015~0021...
3	0001~0005,0012~0021...
...	...

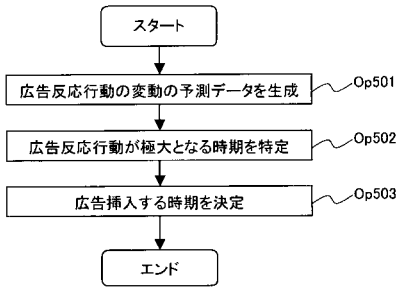
【図14】



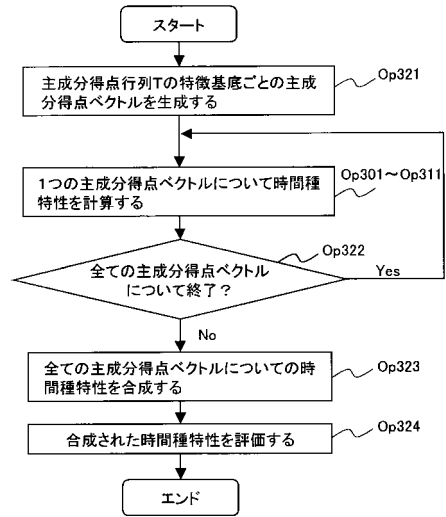
【図17】



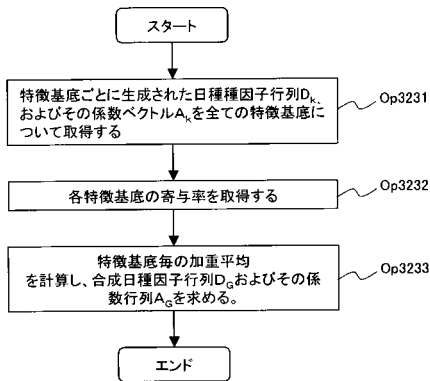
【図18】



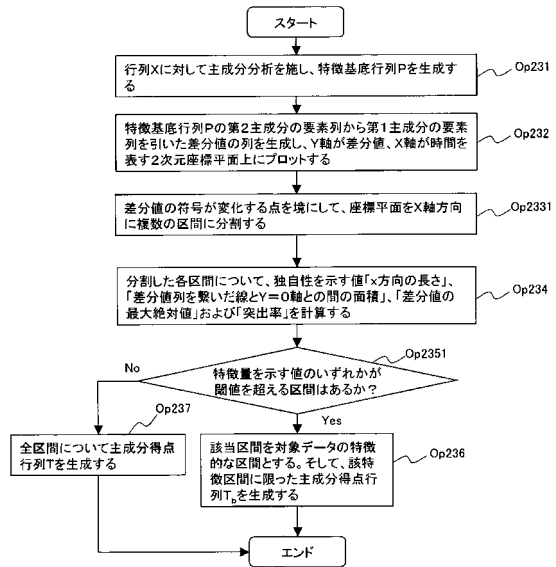
【図19】



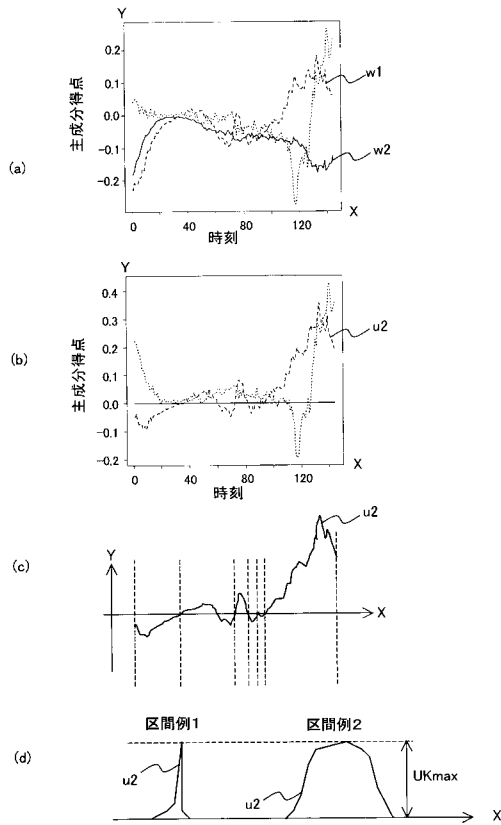
【図20】



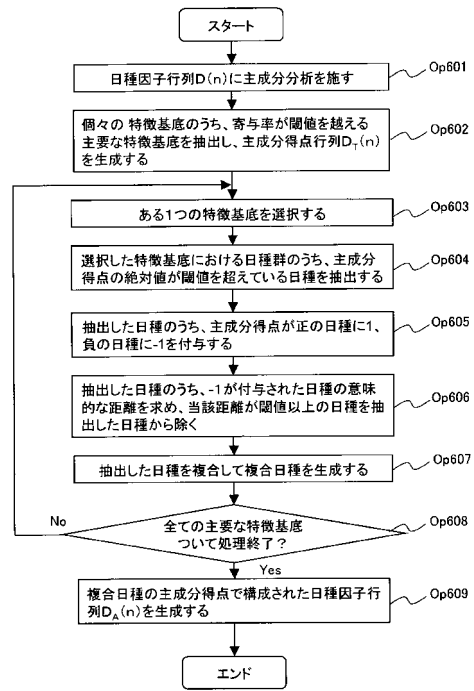
【図21】



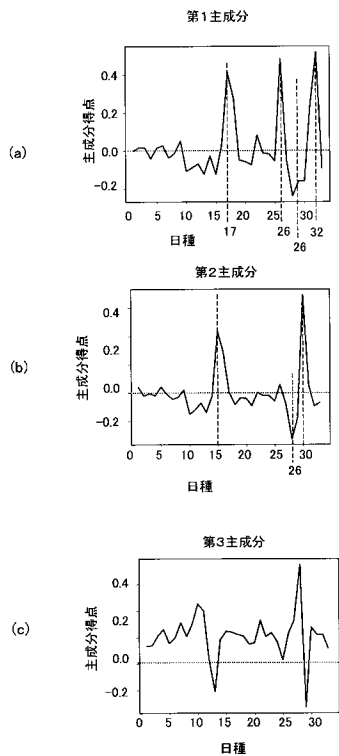
【図22】



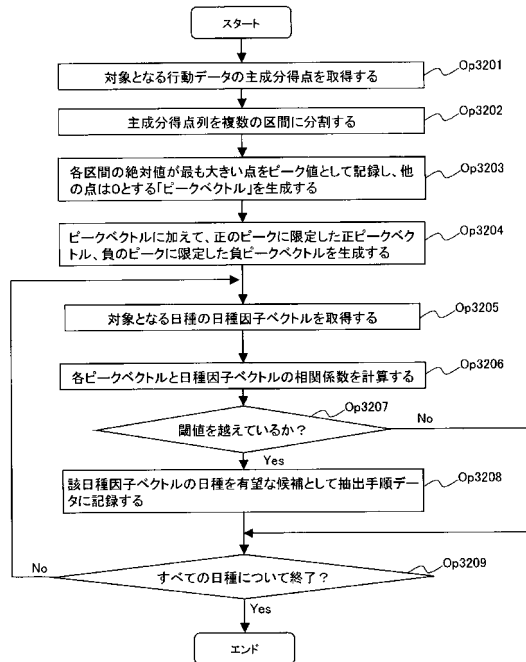
【図23】



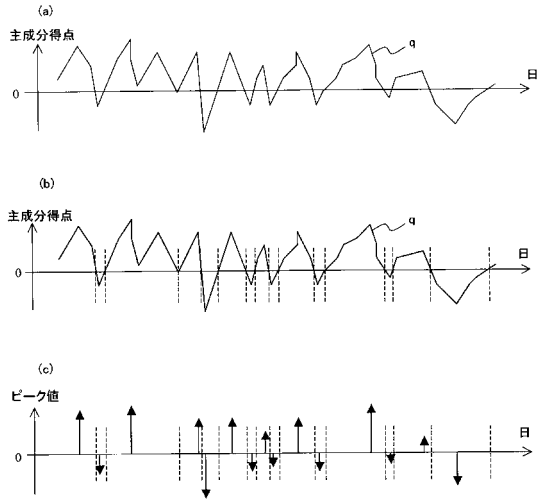
【図24】



【図25】



【 図 26 】



フロントページの続き

- (72)発明者 矢崎 昌朋
神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内
- (72)発明者 松本 安英
神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内
- (72)発明者 宇山 政志
神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内
- (72)発明者 織田 充
神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内
- (72)発明者 市來 宏基
神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

審査官 関 博文

- (56)参考文献 渡辺理, Eコマースサイトの利用者行動の解明: 日種因子による主成分回帰分析, マルチメディア、分散、協調とモバイル (D I C O M O) シンポジウム論文集 1997~2006年版, 日本, 社団法人情報処理学会, 2006年 8月31日, 第309~311頁
- 小松原 健, Webユーザをプロファイリング: ユーザの年齢や好みなどをアクセス履歴から分析、推測, 日経インターネットテクノロジー, 日本, 日経BP社, 1998年 9月22日, 第15号, 第19頁

(58)調査した分野(Int.Cl., DB名)

G 0 6 F 1 9 / 0 0

G 0 6 Q 1 0 / 0 0 - 5 0 / 0 0