

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2009-110341

(P2009-110341A)

(43) 公開日 平成21年5月21日(2009.5.21)

(51) Int.Cl.	F I	テーマコード (参考)
G06Q 10/00 (2006.01)	G06F 19/00 100	
G06N 1/00 (2006.01)	G06N 1/00 510	
G06N 5/04 (2006.01)	G06N 5/04 580A	

審査請求 未請求 請求項の数 7 O L (全 26 頁)

(21) 出願番号	特願2007-282843 (P2007-282843)	(71) 出願人	000004226 日本電信電話株式会社 東京都千代田区大手町二丁目3番1号
(22) 出願日	平成19年10月31日(2007.10.31)	(71) 出願人	504132272 国立大学法人京都大学 京都府京都市左京区吉田本町36番地1
		(74) 代理人	100064414 弁理士 磯野 道造
		(74) 代理人	100127720 弁理士 大石 恵
		(72) 発明者	岩田 具治 東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内
		(72) 発明者	田中 利幸 京都府京都市左京区吉田本町 国立大学法人京都大学大学院情報学研究科内

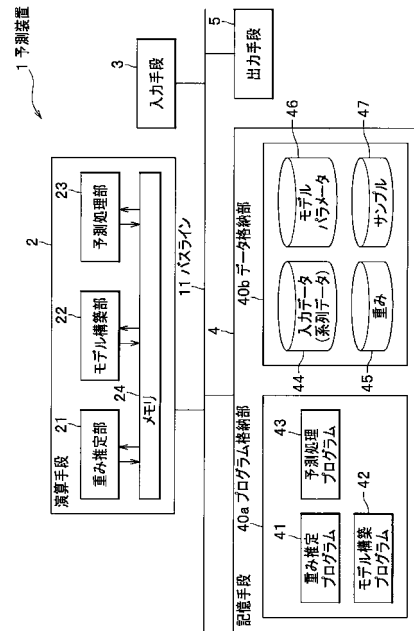
(54) 【発明の名称】 時間情報を用いた予測装置、予測方法、予測プログラムおよびそのプログラムを記録した記録媒体

(57) 【要約】

【課題】 時間とともに性質が変化する時系列データにおいて最新の時刻におけるデータを高い精度で予測することができる予測装置を提供すること。

【解決手段】 予測装置1は、入力データ44から得られる、最新時刻における入出力サンプルデータに対する期待誤差が、入出力サンプルデータが与えられたときのモデルの誤差関数と、それぞれの入出力サンプルデータの時間情報の重みとを用いて近似された期待誤差を最小化させるように、重みを推定する重み推定部21と、推定された重みを用いた期待誤差に基づく目的関数と入力データ44とを用いて、サンプルデータ47から、最新時刻における未知の出力サンプルを予測するモデルパラメータ46を構築するモデル構築部22と、モデルパラメータ46と、サンプル47とに基づいて、最新時刻におけるサンプル47の未知の出力サンプルを予測する処理を行う予測処理部23とを備える。

【選択図】 図1



【特許請求の範囲】**【請求項 1】**

予測対象に関する基準側の第 1 観測データと、予測側の第 2 観測データと、前記第 2 観測データが示す性質の変化を離散的に知覚するための所定の基準に立脚した離散的な指標を示す指標情報とを有する前記予測対象に関する学習用サンプルデータについての前記指標情報に関する系列を示す系列データを学習することで、前記第 1 観測データおよび前記指標情報が既知で前記第 2 観測データが未知である予測用サンプルデータから、前記未知の第 2 観測データを予測する予測装置であって、

前記系列データから得られる、前記指標情報の予め定められた離散値における学習用サンプルデータの前記第 1 および第 2 観測データに対する期待誤差が、当該第 1 および第 2 観測データが与えられたときのモデルの誤差関数と、前記それぞれの学習用サンプルデータの前記指標情報の重みと、を用いて近似された期待誤差を最小化させるように、前記指標情報の重みを推定する重み推定手段と、

前記推定された指標情報の重みを用いた期待誤差に基づく目的関数と前記系列データとを用いて、前記予測用サンプルデータから、前記指標情報の前記予め定められた離散値における前記未知の第 2 観測データを予測するモデルを構築するモデル構築手段と、

前記構築されたモデルと、前記予測用サンプルデータとに基づいて、前記指標情報の前記予め定められた離散値における当該予測用サンプルデータの前記未知の第 2 観測データを予測する処理を行う予測処理手段とを備えることを特徴とする予測装置。

【請求項 2】

前記重み推定手段は、

前記予め定められた離散値における前記系列データの真の分布を近似するために、前記指標情報の各離散値における経験分布に混合される予め定められた混合比と、前記系列データにおける所定の学習用サンプルデータの前記第 1 および第 2 観測データとを条件として、前記第 1 および第 2 観測データの前記指標情報に対する事後確率をそれぞれ推定する事後確率推定手段と、

前記推定されたそれぞれの事後確率を利用して、前記予め定められた離散値における前記系列データに対する尤度を最大化するように、前記混合比を推定し、前記尤度が最大化されたときの前記混合比から前記重みを決定する混合比推定手段とを備えることを特徴とする請求項 1 に記載の予測装置。

【請求項 3】

前記重み推定手段は、

前記予め定められた離散値における前記系列データの真の分布を近似するために、前記指標情報の各離散値における経験分布に混合される予め定められた混合比と、前記系列データにおける所定の学習用サンプルデータの前記第 2 観測データを条件として、前記第 2 観測データの前記指標情報に対する事後確率をそれぞれ推定する事後確率推定手段と、

前記推定されたそれぞれの事後確率を利用して、前記予め定められた離散値における前記系列データに対する尤度を最大化するように、前記混合比を推定し、前記尤度が最大化されたときの前記混合比から前記重みを決定する混合比推定手段とを備えることを特徴とする請求項 1 に記載の予測装置。

【請求項 4】

前記モデル構築手段は、

前記推定された指標情報の重みと、前記重みのハイパーパラメータと、前記近似された期待誤差における誤差関数とに基づく目的関数を最小化させるように、前記系列データおよび前記推定された重みに基づいてモデルを学習することで、前記予め定められた離散値におけるモデルを推定するモデルパラメータ推定手段と、

前記推定されたモデルに対して適用される前記誤差関数を最小化させるように、前記ハイパーパラメータを推定するハイパーパラメータ推定手段と、を備え、

前記モデルパラメータ推定手段は、

前記誤差関数が最小化されたときの前記ハイパーパラメータを用いて推定されたモデル

10

20

30

40

50

を構築すべきモデルとして決定することを特徴とする請求項 1 ないし請求項 3 のいずれか一項に記載の予測装置。

【請求項 5】

予測対象に関する基準側の第 1 観測データと、予測側の第 2 観測データと、前記第 2 観測データが示す性質の変化を離散的に知覚するための所定の基準に立脚した離散的な指標を示す指標情報とを有する前記予測対象に関する学習用サンプルデータについての前記指標情報に関する系列を示す系列データを学習することで、前記第 1 観測データおよび前記指標情報が既知で前記第 2 観測データが未知である予測用サンプルデータから、前記未知の第 2 観測データを予測する予測装置の予測方法であって、

前記予測装置の演算手段は、

前記系列データから得られる、前記指標情報の予め定められた離散値における学習用サンプルデータの前記第 1 および第 2 観測データに対する期待誤差が、当該第 1 および第 2 観測データが与えられたときのモデルの誤差関数と、前記それぞれの学習用サンプルデータの前記指標情報の重みとを用いて近似された期待誤差を最小化させるように、前記指標情報の重みを推定する重み推定ステップと、

前記推定された指標情報の重みを用いた期待誤差に基づく目的関数と前記系列データとを用いて、前記予測用サンプルデータから、前記指標情報の前記予め定められた離散値における前記未知の第 2 観測データを予測するモデルを構築するモデル構築ステップと、

前記構築されたモデルと、前記予測用サンプルデータとに基づいて、前記指標情報の前記予め定められた離散値における当該予測用サンプルデータの前記未知の第 2 観測データを予測する処理を行う予測処理ステップと、を含んで実行することを特徴とする予測方法

【請求項 6】

請求項 5 に記載の予測方法をコンピュータに実行させることを特徴とする予測プログラム。

【請求項 7】

請求項 6 に記載の予測プログラムが記録されたことを特徴とするコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、生起順序が予め与えられている事象を示す学習データを学習することで将来生じる事象を予測する予測装置、予測方法、予測プログラムおよびそのプログラムを格納した記録媒体に関する。

【背景技術】

【0002】

従来、生起順序が予め与えられている事象を示す学習データを学習することで、将来生じる事象を予測する予測問題に関する技術が知られている。この予測問題では、例えば、式(101)で示す入力サンプル(入力データ) x と、式(102)で示す出力サンプル(出力データ) y とを想定している。ここで、 X は入力サンプル集合、 Y は出力サンプル集合を示す。入力サンプル x と出力サンプル y とには必ずしも因果関係がある訳ではない。そして、式(103)で示す入出力サンプル (x, y) が生成したときの時刻 t を式(104)で定義する。ここで、 T は例えば最新の時刻を示す。そして、予測装置は、学習段階において、例えば式(105)で示す学習データを学習する。ここで、 n はインデックスを示し、 N は学習データ数を示す。そして、予測装置は、予測段階において、式(106)で示すように、出力 y が未知のサンプル (x, t) についてその出力 y を予測する。

【0003】

【数 1】

入力サンプル： $x \in X$ …式 (101)出力サンプル： $y \in Y$ …式 (102)入出力サンプル： (x, y) …式 (103)入出力サンプルの生成した時刻： $t \in \{1, \dots, T\}$ …式 (104)与えられた学習データ： $\{(x_n, y_n, t_n)\}_{n=1}^N$ …式 (105)出力 y が未知のサンプル： (x, t) …式 (106)

10

【0004】

この予測問題を、例えば、店舗側の購買予測に適用した場合には、出力サンプル y は購買商品を示し、入力サンプル x はそのときの購買履歴を示し、入出力サンプル (x, y) が生成したときの時刻 t は購買時刻を示すこととなる。なお、インデックス n は例えば購買番号や購買者（ユーザ）を示し、学習データ数 N は例えば購買数や購買者数（ユーザ数）を示すこととなる。購買予測の場合のデータ（購買データ）の具体例を表 1 に示す。ここで、購買履歴としては 1 つ前に購買した商品とした。購買履歴および購買商品（入出力サンプル）は時系列データである。なお、 x, y, t はすべて離散的なものとする。

20

【0005】

【表 1】

インデックス n 購買番号	入力 x 購買履歴	出力 y 購買商品	時刻 t 購買時刻
1	商品A	商品B	1 (2007/8/14)
2	商品B	商品C	1 (2007/8/14)
3	商品A	商品C	1 (2007/8/14)
4	商品B	商品C	2 (2007/8/15)
...
N	商品D	商品A	T (2007/8/20)

30

【0006】

このような時系列データの多くは、時間とともにその性質が変化するものである。時間とともに性質が変化する時系列データとは、式 (107) に示すように、時刻 t における入出力の（確率）分布 P と、時刻 t' における入出力の（確率）分布 P と、が異なるような時系列データである。なお、時系列データの一例として購買データを挙げたが、時系列データには、例えば、ニュース記事データ、論文データ、Web サーフィンデータ等が含まれる。

40

【0007】

【数 2】

$$P(x, y|t) \neq P(x, y|t') \quad \dots \text{式 (107)}$$

【0008】

時間とともにその性質が変化する時系列データの具体例を説明するために、ここでは、購買データを考える。例えば、商品販売の予測においては、現実にはメーカーや店舗の経営戦略や都合に起因して、新商品の発表や発売中止等の突発的な事象が生じる。そのため、このような最新の情報を予測に反映させることにより、予測される商品候補（出力される

50

商品候補)は日々変化することとなる。また、季節や流行、社会的、経済的環境等の大局的な事象の変化により、各商品の出力分布(予測される確率分布)も変化する。

【0009】

この購買データから、次にユーザが購入すると思われる商品を高い精度で予測できれば、その商品をユーザにリコメンドすることができる。従来、例えば、オンラインショッピング等の商品について購買順序を考慮した予測精度の高いリコメンド技術が知られている(例えば、非特許文献1参照)。

【非特許文献1】岩田具治、山田武士、上田修功、“購買順序を考慮した協調フィルタリング”、人工知能と知識処理研究会、AI2007-3,13-18,2007

【発明の開示】

【発明が解決しようとする課題】

【0010】

非特許文献1に開示された技術は、予測精度は高いものの、時間情報すなわち購買時間を考慮するものではなく、計算コストを低く抑えることを目的としており、最新データの予測精度を高めることを目的とするものでもない。しかしながら、時間とともに性質が変化する時系列データにおける予測問題では、全学習データに対する高い予測精度ではなく、最新の時刻におけるデータに対して高い予測精度が望まれることが多い。これは、例えば、性質変化が急激なデータの場合には、過去の時刻のデータに対する予測精度が高いモデルであっても、将来を予測するためには使えない場合があるからである。

【0011】

そこで、本発明は、以上のような問題点に鑑みてなされたものであり、時間とともに性質が変化する時系列データにおいて最新の時刻におけるデータを高い精度で予測することができる予測装置、予測方法、予測プログラムおよびそのプログラムを記録した記録媒体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】

【0012】

本発明は、前記課題を解決するために創案されたものであり、本発明に係る予測装置は、予測対象に関する基準側の第1観測データと、予測側の第2観測データと、前記第2観測データが示す性質の変化を離散的に知覚するための所定の基準に立脚した離散的な指標を示す指標情報とを有する前記予測対象に関する学習用サンプルデータについての前記指標情報に関する系列を示す系列データを学習することで、前記第1観測データおよび前記指標情報が既知で前記第2観測データが未知である予測用サンプルデータから、前記未知の第2観測データを予測する予測装置であって、前記系列データから得られる、前記指標情報の予め定められた離散値における学習用サンプルデータの前記第1および第2観測データに対する期待誤差が、当該第1および第2観測データが与えられたときのモデルの誤差関数と、前記それぞれの学習用サンプルデータの前記指標情報の重みと、を用いて近似された期待誤差を最小化させるように、前記指標情報の重みを推定する重み推定手段と、前記推定された指標情報の重みを用いた期待誤差に基づく目的関数と前記系列データとを用いて、前記予測用サンプルデータから、前記指標情報の前記予め定められた離散値における前記未知の第2観測データを予測するモデルを構築するモデル構築手段と、前記構築されたモデルと、前記予測用サンプルデータとに基づいて、前記指標情報の前記予め定められた離散値における当該予測用サンプルデータの前記未知の第2観測データを予測する処理を行う予測処理手段とを備えることを特徴とする。

【0013】

かかる構成によれば、予測装置は、予測対象に関する系列データから得られる期待誤差を近似した期待誤差を最小化させるように、学習用サンプルデータの指標情報の重みを推定し、推定された重みと系列データとを用いて、指標情報の予め定められた離散値におけるモデルを構築し、構築されたモデルと、予測用サンプルデータとに基づいて、指標情報の予め定められた離散値における未知の第2観測データを予測する処理を行う。ここで、所定の基準は、例えば、時間や地域等である。時間という基準に立脚した離散的な指標と

10

20

30

40

50

は、過去から未来へ向かう一方向性の特徴を有した時間における特定の時刻を示し、系列データは時系列データとなる。このときの離散値は、年月日や時分秒等所定の単位を用いることができ、予め定められた離散値とは、例えば最新時刻等に設定できる。また、所定の基準が地域である場合には、その基準の種類として、例えば、地域環境やその場所に居住する人やくらし等を挙げることができ、このときの指標は、例えば、位置や距離、各種統計量等となる。このときの離散値は、位置座標や各種距離単位、各種の統計や動向調査に用いられる所定単位、「0, 1」等のデジタル表記等を用いることができる。

【0014】

例えば、時間を基準にして時系列データを用いて予め定められた離散値として最新時刻を設定したときには、予測装置は、時系列データから得られる期待誤差を近似した期待誤差を最小化させるように、学習用サンプルデータの時間情報の重みを推定する。この時間情報の重みによって、最新時刻のデータに適合するモデルを学習するために有益な情報を最新時刻のデータ以外の過去のデータから取り込むことができる。また、この時間情報の重みによって、最新時刻での経験誤差を最新時刻での期待誤差に対する精度のよい近似として用いることができる。したがって、予測装置は、このように推定された重みを用いて構築されたモデルと、出力(第2観測データ)が未知のサンプルから、そのサンプルの最新の時刻における未知データを高い精度で予測することが可能となる。

10

【0015】

また、本発明に係る予測装置は、前記重み推定手段が、前記予め定められた離散値における前記系列データの真の分布を近似するために、前記指標情報の各離散値における経験分布に混合される予め定められた混合比と、前記系列データにおける所定の学習用サンプルデータの前記第1および第2観測データとを条件として、前記第1および第2観測データの前記指標情報に対する事後確率をそれぞれ推定する事後確率推定手段と、前記推定されたそれぞれの事後確率を利用して、前記予め定められた離散値における前記系列データに対する尤度を最大化するように、前記混合比を推定し、前記尤度が最大化されたときの前記混合比から前記重みを決定する混合比推定手段とを備えることが好ましい。

20

【0016】

かかる構成によれば、予測装置は、重み推定手段が、事後確率推定手段と、混合比推定手段とを備えているので、事後確率推定手段が、EMアルゴリズム(Expectation - Maximization algorithm)におけるEステップ(Expectation step)を行い、かつ、混合比推定手段がMステップ(Maximization step)を行うことで、混合比についての大域的最適解を求め、求めた混合比から重みを決定することができる。

30

【0017】

また、本発明に係る予測装置は、指標情報のある離散値における第1および第2観測データの条件付き分布と、別の離散値における第1および第2観測データの条件付き分布とが等しく、かつ、第2観測データの分布のみ離散値により異なる状況においては、簡易的に以下のように構成することができる。すなわち、この場合には、本発明に係る予測装置は、前記重み推定手段が、前記予め定められた離散値における前記系列データの真の分布を近似するために、前記指標情報の各離散値における経験分布に混合される予め定められた混合比と、前記系列データにおける所定の学習用サンプルデータの前記第2観測データを条件として、前記第2観測データの前記指標情報に対する事後確率をそれぞれ推定する事後確率推定手段と、前記推定されたそれぞれの事後確率を利用して、前記予め定められた離散値における前記系列データに対する尤度を最大化するように、前記混合比を推定し、前記尤度が最大化されたときの前記混合比から前記重みを決定する混合比推定手段とを備えることができる。

40

【0018】

かかる構成の予測装置によれば、予測装置は、重みを決定するためにEMアルゴリズムを用いて混合比を推定する際に、第1および第2観測データの条件付き分布の代わりに第2観測データの条件付き分布を用いればよいので、計算が比較的容易になり、処理負荷を低減できる。

50

【 0 0 1 9 】

また、本発明に係る予測装置は、前記モデル構築手段が、前記推定された指標情報の重みと、前記重みのハイパーパラメータと、前記近似された期待誤差における誤差関数とに基づき目的関数を最小化させるように、前記系列データおよび前記推定された重みに基づいてモデルを学習することで、前記予め定められた離散値におけるモデルを推定するモデルパラメータ推定手段と、前記推定されたモデルに対して適用される前記誤差関数を最小化させるように、前記ハイパーパラメータを推定するハイパーパラメータ推定手段と、を備え、前記モデルパラメータ推定手段が、前記誤差関数が最小化されたときの前記ハイパーパラメータを用いて推定されたモデルを構築すべきモデルとして決定することが好ましい。

10

【 0 0 2 0 】

かかる構成によれば、予測装置は、モデル構築手段が、モデルパラメータ推定手段と、ハイパーパラメータ推定手段とを備えているので、誤差関数を最小化させるハイパーパラメータによって、重み推定手段で推定された重みを安定化させた精度のよいモデルを構築することができる。

【 0 0 2 1 】

また、前記課題を解決するために、本発明に係る予測方法は、予測対象に関する基準側の第1観測データと、予測側の第2観測データと、前記第2観測データが示す性質の変化を離散的に知覚するための所定の基準に立脚した離散的な指標を示す指標情報とを有する前記予測対象に関する学習用サンプルデータについての前記指標情報に関する系列を示す系列データを学習することで、前記第1観測データおよび前記指標情報が既知で前記第2観測データが未知である予測用サンプルデータから、前記未知の第2観測データを予測する予測装置の予測方法であって、前記予測装置の演算手段が、前記系列データから得られる、前記指標情報の予め定められた離散値における学習用サンプルデータの前記第1および第2観測データに対する期待誤差が、当該第1および第2観測データが与えられたときのモデルの誤差関数と、前記それぞれの学習用サンプルデータの前記指標情報の重みとを用いて近似された期待誤差を最小化させるように、前記指標情報の重みを推定する重み推定ステップと、前記推定された指標情報の重みを用いた期待誤差に基づく目的関数と前記系列データとを用いて、前記予測用サンプルデータから、前記指標情報の前記予め定められた離散値における前記未知の第2観測データを予測するモデルを構築するモデル構築ステップと、前記構築されたモデルと、前記予測用サンプルデータとに基づいて、前記指標情報の前記予め定められた離散値における当該予測用サンプルデータの前記未知の第2観測データを予測する処理を行う予測処理ステップと、を含んで実行することを特徴とする。

20

30

【 0 0 2 2 】

かかる手順の予測方法によれば、予測装置は、予測対象に関する系列データから得られる期待誤差を近似した期待誤差を最小化させるように、学習用サンプルデータの指標情報の重みを推定し、推定された重みと系列データとを用いて、指標情報の予め定められた離散値におけるモデルを構築し、構築されたモデルと、予測用サンプルデータとに基づいて、指標情報の予め定められた離散値における未知の第2観測データを予測する処理を行う。例えば、時間を基準にして時系列データを用いて予め定められた離散値として最新時刻を設定したときには、予測装置は、時系列データから得られる期待誤差を近似した期待誤差を最小化させるように、学習用サンプルデータの時間情報の重みを推定する。この時間情報の重みによって、最新時刻のデータに適合するモデルを学習するために有益な情報を最新時刻のデータ以外の過去のデータから取り込むことができる。また、この時間情報の重みによって、最新時刻での経験誤差を最新時刻での期待誤差に対する精度のよい近似として用いることができる。したがって、予測装置は、このように推定された重みを用いて構築されたモデルと、出力(第2観測データ)が未知のサンプルから、そのサンプルの最新の時刻における未知データを高い精度で予測することが可能となる。

40

【 0 0 2 3 】

50

また、本発明に係る予測プログラムは、前記した予測方法をコンピュータに実行させることを特徴とする。このように構成されることにより、このプログラムをインストールされたコンピュータは、このプログラムに基づいた各機能を実現することができる。

【0024】

また、本発明に係るコンピュータ読み取り可能な記録媒体は、前記した予測プログラムが記録されたことを特徴とする。このように構成されることにより、この記録媒体を装着されたコンピュータは、この記録媒体に記録されたプログラムに基づいた各機能を実現することができる。

【発明の効果】

【0025】

本発明によれば、時間情報等を示す指標情報の重みによって、指標情報の予め定められた離散値のデータに適合するモデルを学習するために有益な情報を、予め定められた離散値のデータ以外のデータから取り込むことで、予め定められた離散値での経験誤差を期待誤差に対する精度のよい近似として用いることができる。そのため、時間とともに性質が変化する時系列データにおいて、最新の時刻におけるデータを高い精度で予測することが可能となる。

【発明を実施するための最良の形態】

【0026】

以下、本発明の予測装置および予測方法を実施するための最良の形態（以下「実施形態」という）について詳細に説明する。説明の都合上、本発明の概要を説明した上で、実施形態について図面を参照して説明する。

【0027】

〔本発明の概要〕

本発明は、予測対象に関する入力サンプル（基準側の第1観測データ）と、出力サンプル（予測側の第2観測データ）と、出力サンプルが示す性質の変化を離散的に知覚するための所定の基準に立脚した離散的な指標を示す指標情報とを有する予測対象に関する入出力サンプル（学習用サンプルデータ）についての指標情報に関する系列を示す系列データを学習することで、入力サンプルおよび指標情報が既知で出力サンプルが未知であるサンプル（予測用サンプルデータ）から、未知の出力サンプルを予測する予測問題に関するものである。特に、本発明は、時間情報に応じて変化するデータ（時系列データ）を予測対象とすることが好ましい。その性質が時間的に変化するデータの種類としては、例えば、購買データ、ニュース記事データ、論文データ、Webサーフィンデータ等を挙げることができる。特に、予測対象とするデータが購買データである場合には、購買される対象物（商品）は、実体としてのモノに限らずデータ等のサービスであってもよい。このようなサービスデータは、例えば、画像データ、映像データ、音楽データ等である。

【0028】

本実施形態では、本発明の説明を単純化させるために、時間情報に応じて変化する購買データを予測対象とすることを念頭に説明する。この場合に、本発明は、全学習データに対する予測精度が高いモデルを学習するのではなく、最新のデータに対する予測精度が高いモデルを学習する予測手法であると言える。このような学習を行うためには、最新時刻で最小にすべき期待誤差を精度よく推定する必要がある。そこで、本発明では、大要、サンプルが生成された時刻（時間情報）を重み付ける重みを推定する処理（重み推定処理）と、推定された重みを安定に取り込んだ予測モデルを構築する処理（モデル構築処理）とを行うこととした。以下、これらの重み推定処理およびモデル構築処理の原理について順次説明する。

【0029】

< 重み推定処理の原理 >

最新時刻におけるデータ（サンプル）に対する期待誤差 E_T は、式（1）で示される。ここで、時間の単位は、任意であり、最新時刻とは、最新の時分秒を示すほか、最新日、最新月、最新年等を示すことができる。このような最新のデータに対する予測精度が高い

10

20

30

40

50

モデルを学習するときには、式(1)に示す期待誤差 E_T が最小になるようなモデルが望ましい。式(1)において、 T は最新時刻を示し、 $J(x, y; M)$ は、サンプル (x, y) が与えられたときの予測モデル(以下、単にモデルという) M の誤差関数を示す。また、 $P(x, y; T)$ は、最新時刻におけるサンプル (x, y) の確率分布(以下、単に分布という)を示す。なお、モデル M としては、例えば、1次マルコフモデル、2次マルコフモデル等、任意のモデルを用いることができる。ここで、入力サンプル (x) は、第1観測データであり、単に入力ともいう。また、出力サンプル (y) は、第2観測データであり、単に出力ともいう。

【0030】

【数3】

$$E_T[J(x, y; M)] = \sum_{x, y} P(x, y | T) J(x, y; M) \quad \dots \text{式(1)}$$

10

【0031】

誤差関数 $J(x, y; M)$ は、例えば、式(2)に示す負の対数尤度や、式(3)に示す0-1損失関数等が考えられる。なお、以下では、対数は自然対数、すなわち、対数 \log の底は「 e 」であるものとする。

【0032】

【数4】

$$J(x, y; M) = -\log P(y | x; M) \quad \dots \text{式(2)}$$

$$J(x, y; M) = \begin{cases} 0 & \text{if } f(x; M) = y \\ 1 & \text{otherwise} \end{cases} \quad \dots \text{式(3)}$$

20

【0033】

最新時刻のサンプル数が無限大の極限である場合には、最新時刻での経験誤差は、最新時刻での期待誤差 E_T と一致する。しかし、最新時刻のデータのみを用いて学習した場合には、全データを用いて学習する場合と比べて、サンプル数が少なくなるため、適切なモデルが得られない可能性がある。そのため、最新時刻のデータのみを用いて学習することは好ましくはない。また、最新時刻のデータより前の過去のデータにも、最新時刻のデータに適合するモデルを学習するために有益な情報がある程度は含まれていると考えられる。そこで、重み推定処理においては、サンプル (x, y) が与えられたときのモデル M の誤差関数 $J(x, y; M)$ を、サンプル (x, y) が生成された時刻における重みによって重み付けしたときの誤差 $E(M)$ を最小化させるようにモデル M を学習することとした。このときの誤差 $E(M)$ は、式(4)で示される。式(4)において、 $w(t_n)$ は時刻 t_n における重みであり、最新時刻 T におけるモデル(重みに基づいて後に構築するモデル)を学習するために時刻 t_n のデータがどのくらい参考になるかを表す。

30

【0034】

【数5】

$$E(M) = \sum_n w(t_n) J(x_n, y_n; M) \quad \dots \text{式(4)}$$

40

【0035】

前記した式(1)の代わりに式(4)を用いることの正当性を以下に説明する。

まず、時刻が t であるサンプル数を $N(t)$ とする。また、時刻が t 、入力が x 、出力が y のサンプル数を $N(t, x, y)$ とする。また、最新時刻 T におけるサンプル (x, y) の分布(真の分布) $P(x, y | T)$ を、式(5)に示すように、各時刻における経

50

験分布の混合で近似することとする。式(5)を満たす混合比を式(6)で示す。以下では、個々の混合比を混合分布 $P(t)$ と呼ぶ。つまり、式(5)は、最新時刻 T における真の分布 $P(x, y | T)$ が、各時刻における経験分布と混合分布 $P(t)$ との積を時刻 t について加算した結果で近似することを示すことになる。また、各時刻における経験分布には、サンプル数 $N(t)$ 、 $N(t, x, y)$ との間で、式(7)の関係がある。なお、式(7)において P に付した記号「 \hat{P} (ハット)」は、その P が経験分布であることを示すものである。

【0036】

【数6】

$$P(x, y | T) \approx \sum_t P(t) \hat{P}(x, y | t) \quad \dots \text{式 (5)}$$

10

$$\mathbf{P} = \{P(t)\}_{t=1}^T \quad \dots \text{式 (6)}$$

$$\hat{P}(x, y | t) = \frac{N(t, x, y)}{N(t)} \quad \dots \text{式 (7)}$$

【0037】

そして、式(8)に示すように、重みは、混合分布 $P(t)$ を、時刻が t であるサンプル数 $N(t)$ で除した値として定義される。なお、式(8)に式(7)の関係を適用すると、重みは式(9)で示される。

20

【0038】

【数7】

$$w(t) = \frac{P(t)}{N(t)} \quad \dots \text{式 (8)}$$

$$w(t) = \frac{P(t) \hat{P}(x, y | t)}{N(t, x, y)} \quad \dots \text{式 (9)}$$

30

【0039】

そして、このとき、前記した式(4)に示す誤差 $E(M)$ は、式(10)のように書き換えられる。式(10)において、右辺の2行目は、サンプル数 $N(t, x, y)$ の定義による変形結果を示し、右辺の3行目は、前記した式(9)による変形結果を示す。また、右辺の4行目は、前記した式(5)の近似を用いた変形結果を示し、右辺の5行目は、前記した式(1)による変形結果を示す。その結果、前記した式(4)に示す誤差 $E(M)$ は、最新時刻 T での期待誤差 E_T の近似となることが理解される。

【0040】

40

【数 8】

$$\begin{aligned}
E(M) &= \sum_{x,y} \left(\sum_{n: x_n=x \wedge y_n=y} w(t_n) \right) J(x, y; M) \\
&= \sum_{x,y} \left(\sum_t N(t, x, y) w(t) \right) J(x, y; M) \\
&= \sum_{x,y} \left(\sum_t P(t) \hat{P}(x, y|t) \right) J(x, y; M) \\
&\approx \sum_{x,y} P(x, y|T) J(x, y; M) \\
&= E_T [J(x, y; M)] \quad \dots \text{式 (10)}
\end{aligned}$$

10

【0041】

次に、時刻 t における重み $w(t)$ を推定する。前記した式 (8) の関係から、重み $w(t)$ を推定することは、混合分布 $P(t)$ を推定することと同義である。つまり、重み $w(t)$ を推定するためには、混合比を決定すればよい。前記した式 (6) に示す混合比は、式 (11) に示す最新時刻 T のサンプル集合に対する尤度 $L(P)$ を最大化することにより推定することができる。なお、過学習を回避するため leave-one-out 法を用いる。式 (11) の右辺の 2 行目は、式 (12) に示す関係式を用いて書き換えたものを示す。式 (12) は、 n 番目のサンプルを除いたときに前記した式 (5) より推定される最新時刻 T における分布の推定値の定義式である。

20

【0042】

【数 9】

$$\begin{aligned}
L(P) &= \sum_{n: t_n=T} \log P_{-n}(x_n, y_n|T) \\
&= \sum_{n: t_n=T} \log P(t) \hat{P}_{-n}(x_n, y_n|t) \quad \dots \text{式 (11)}
\end{aligned}$$

30

$$P_{-n}(x, y|T) = \sum_t P(t) \hat{P}_{-n}(x, y|t) \quad \dots \text{式 (12)}$$

ここで、 $\hat{P}_{-n}(x, y|t)$: n 番目のサンプルを除いたときの時刻 t での経験分布

40

【0043】

式 (12) の右辺における n 番目のサンプルを除いたときの時刻 t での経験分布は、式 (13) で示される。式 (13) において、 α は、ゼロ確率問題を回避するため導入されたものであり、Dirichlet 事前分布のパラメータを示す。また、 $\delta_{t,T}$ はクロネッカーのデルタを示す。

【0044】

【数 1 0】

$$\hat{P}_{-n}(x, y|t) = \frac{N(t, x, y) + \alpha - \delta_{t,T}}{N(t) + \alpha|\mathbf{X}||\mathbf{Y}| - \delta_{t,T}} \quad \dots\text{式 (13)}$$

ここで、 $x \in \mathbf{X}$, $y \in \mathbf{Y}$, $|\mathbf{X}|$: \mathbf{X} の要素数, $|\mathbf{Y}|$: \mathbf{Y} の要素数

$$\delta_{t,T} = \begin{cases} 1 & \text{if } t = T \\ 0 & \text{otherwise} \end{cases}$$

10

【0 0 4 5】

ここで、対数関数は上に凸な関数であるため、前記した式(11)に示す $L(P)$ の関数形も上に凸であり、 $L(P)$ を最大化することにより、 P についての大域的最適解(P^* と表記する)を得ることができる。 $L(P)$ を最大化させるための最適化手法は、任意であり、例えば、EMアルゴリズムを用いることができる。本実施形態では、EMアルゴリズムを用いて説明する。個々の完全データの対数尤度を式(14)で示す。ここで、式(14)は前記した式(11)に示す $L(P)$ の右辺第2行に示したものである。また、最大化すべき対数尤度は、 $t_n = T$ となる n の個数だけの完全データの対数尤度である。

20

【0 0 4 6】

【数 1 1】

$$\log \sum_t P(t) \hat{P}_{-n}(x_n, y_n|t) \quad \dots\text{式 (14)}$$

【0 0 4 7】

したがって、最大化すべき完全データの対数尤度の条件付き期待値は、式(15)で示される。ここで、 τ は、Eステップ(Expectation step)とMステップ(Maximization step)との2つの手順を繰り返した回数($\tau = 0, 1, 2, \dots$)を指す。なお、 $\tau = 0$ のときには推定値の予め定められた初期値を示す。

30

【0 0 4 8】

【数 1 2】

$$Q(\mathbf{P} | \mathbf{P}^{(\tau)}) = \sum_{n:t_n=T} \sum_t P(t | x_n, y_n; \mathbf{P}^{(\tau)}) \log P(t) \hat{P}_{-n}(x_n, y_n|t)$$

…式 (15)

ただし、 $\mathbf{P}^{(\tau)}$: EMアルゴリズムの第 τ ステップでの混合比の推定値

40

【0 0 4 9】

EMアルゴリズムでは、まず、Eステップにおいて、式(15)に示す条件付き期待値 $Q(\mathbf{P} | \mathbf{P}^{(\tau)})$ を計算する際に、式(17)を用いて推定した $P(t)$ を用いて式(16)の計算を行う。式(16)の計算においては、全サンプル(x, y)の全時刻 t に対する事後確率を推定する。次に、Mステップにおいて、式(15)に示す条件付き期待値 $Q(\mathbf{P} | \mathbf{P}^{(\tau)})$ を最大化する際に、式(16)を用いて求めた結果を用いて式(17)の計算を行い、混合分布についての新しい推定値 $\mathbf{P}^{(\tau+1)}(t)$ を求める。すなわち、混合比を推定する。

【0 0 5 0】

【数 1 3】

$$P(t|x_n, y_n; \mathbf{P}^{(\tau)}) = \frac{P^{(\tau)}(t) \hat{P}_{-n}(x_n, y_n|t)}{\sum_{t'} P^{(\tau)}(t') \hat{P}_{-n}(x_n, y_n|t')} \quad \dots \text{式 (16)}$$

$$P^{(\tau+1)}(t) = \frac{1}{N(t)} \sum_{n:t_n=T} P(t|x_n, y_n; \mathbf{P}^{(\tau)}) \quad \dots \text{式 (17)}$$

【0051】

10

そして、EステップとMステップとの2つの手順を収束条件が満たされるまで繰り返すことにより、各混合分布 $P(t)$ すなわち前記した式(6)に示す混合比が得られる。ここで、前記した式(15)に示す条件付き期待値 $Q(P|P^{(\tau)})$ が最大化するときには、前記した式(11)に示す尤度 $L(P)$ が最大化することになる。すなわち、前記した式(11)に示す尤度 $L(P)$ が収束する。このときに収束条件が満たされる。したがって、本実施形態においては、Eステップで、式(17)を用いて推定した $P^{(\tau)}(t)$ を利用して式(16)の計算を行い、また、Mステップで、式(16)を用いて求めた結果を利用して式(17)の計算を行い、EステップとMステップとの組み合わせが終了した時点で、前記した式(11)に示す尤度 $L(P)$ が収束したか否かを判別する。判別の結果、収束していなければ収束するまでEステップとMステップとを交互に繰り返す。そして、収束条件が満たされたときの混合分布 $P(t)$ を、前記した式(8)に示す重みの定義式に用いることで、サンプル (x, y) が生成された時刻 t における重み $w(t)$ が推定されたこととなる。

20

【0052】

<モデル構築処理の原理>

重み推定処理では、サンプル (x, y) が生成された時刻 t における重み $w(t)$ を推定することで、前記した式(1)に示す最新時刻 T におけるデータ(サンプル)に対する期待誤差 E_T を、前記した式(4)に示す誤差 $E(M)$ で近似することとした。この推定された重み $w(t)$ を安定化させるために、モデル構築処理では、ハイパーパラメータを導入して、式(18)に示す誤差 $E(M, \lambda)$ を、最小とすべき目的関数とすることとした。つまり、誤差 $E(M, \lambda)$ を最小化させるようにモデル M を学習する

30

【0053】

【数 1 4】

$$E(M, \lambda) = \sum_n w(t_n)^\lambda J(x_n, y_n; M) \quad \dots \text{式 (18)}$$

【0054】

モデル M を学習したデータでハイパーパラメータ λ を学習すると過学習を起こす可能性がある。そこで、 K 重交差検定法 (K -fold cross validation) を用いてハイパーパラメータ λ を学習する。なお、学習に用いなかったデータに対する汎化誤差が大きくなってしまふ現象は過学習と呼ばれている。具体的には、 K 重交差検定法では、まず、式(19)に示す全学習データ D を式(20)に示すように K 個の部分集合 D_k に分割する。式(20)において、 k は学習データ D 中の部分集合のインデックスを示す。そして、分割された部分集合のうち部分集合 D_j を学習に用いる。ここで、部分集合 D_j は、式(21)に示すように、所定の部分集合を除く $(K-1)$ 個の部分集合を示す。式(21)では、除かれる所定の部分集合のインデックスを k として、それ以外の部分集合のインデックスを j とした。そして、部分集合 D_j を学習に用いて、前記した式(18)に示す目的関数(誤差 $E(M, \lambda)$) を最小化させるモデルを、最新時刻 T におけるモデルとして構築する。構築すべきモデルとして、所定の部分集合 k に依存したモデルを式(22)に示す。 K 重交差検定法では、除かれる部分集合 D_k として分割したすべての種類 (K 種類) の集合を

40

50

考慮するので、式(22)に示すモデルを、除かれる所定の部分集合のインデックスk毎に学習し、K個のモデルを構築する。なお、式(22)においてMに付した記号「^ (ハット)」は、そのMがargmin関数の引数を最小化させることを示すものである。

【0055】
【数15】

$$D = \{n\}_{n=1}^N \equiv \{(x_n, y_n, t_n)\}_{n=1}^N \quad \dots \text{式(19)}$$

$$D = \{D_k\}_{k=1}^K \quad \dots \text{式(20)}$$

$$\{D_j\}_{j \neq k} \quad \dots \text{式(21)}$$

$$\hat{M}_{D_k}(\lambda) = \arg \min_M \sum_{n \in \{D_j\}_{j \neq k}} w(t_n)^\lambda J(x_n, y_n; M) \quad \dots \text{式(22)}$$

10

【0056】

ハイパーパラメータについては、部分集合D_j以外、すなわち、学習に用いなかった部分集合D_k毎に、その部分集合D_kの最新時刻Tのデータに対する誤差関数Jが最小になるように学習し、さらに、除かれる部分集合D_kとしてすべての種類(K種類)を考慮する。具体的には、式(23)に示す計算により、誤差関数Jを最小化させるハイパーパラメータを求める。なお、式(23)においてに付した記号「^ (ハット)」は、そのがargmin関数の引数を最小化させることを示すものである。

20

【0057】
【数16】

$$\hat{\lambda} = \arg \min_\lambda \sum_k \sum_{n \in D_k \wedge t_n = T} J(x_n, y_n; \hat{M}_{D_k}(\lambda)) \quad \dots \text{式(23)}$$

【0058】

ここで、前記した式(18)に示す目的関数(誤差E(M,))が最小化するときには、目的関数(誤差E(M,))が収束する。このときに収束条件が満たされる。したがって、本実施形態においては、式(23)を用いて推定したハイパーパラメータを利用して式(22)の計算を行い、また、式(22)を用いて求めたモデルを利用して式(23)の計算を行い、モデルの推定とハイパーパラメータの推定との組み合わせが終了した時点で、前記した式(18)に示す目的関数(誤差E(M,))が収束したか否かを判別する。判別の結果、収束していなければ収束するまでモデルの推定とハイパーパラメータの推定とを交互に繰り返す。そして、収束条件が満たされたときに、その時点のモデルが、最新時刻Tにおけるモデルとして推定されたこととなる。

30

【0059】

[予測装置の全体構成]

40

次に、本発明の実施形態について図面を参照して説明する。図1は、本発明の実施形態に係る予測装置の構成を示すブロック図である。予測装置1は、予測対象に関する入出力サンプル(学習用サンプルデータ)についての系列を示す入力データ(系列データ)を学習することで、出力が未知であるサンプルから、未知の出力サンプルを予測するものである。図1に示すように、予測装置1は、演算手段2と、入力手段3と、記憶手段4と、出力手段5とを備えている。各手段2~5はバスライン11に接続されている。

【0060】

演算手段2は、例えば、CPU(Central Processing Unit)およびRAM(Random Access Memory)から構成される主制御装置である。この演算手段2は、図1に示すように、重み推定部(重み推定手段)21と、モデル構築部(モデル構築手段)22と、予測処

50

理部（予測処理手段）23と、メモリ24とを含んで構成される。なお、各部21～23の説明は後記する。

【0061】

入力手段3は、例えば、キーボード、マウス、ディスクドライブ装置等から構成される。この入力手段3は、例えば、データとして入力データ（系列データ）およびサンプルを入力し、記憶手段4に格納する。本実施形態では、入力データ（系列データ）を前記した表1に示す購買予測の場合のデータ（購買データ）とする。

【0062】

記憶手段4は、例えば、一般的なハードディスク装置等から構成され、演算手段2で用いられる各種プログラムや各種データ等を記憶する。この記憶手段4は、プログラムとして、重み推定プログラム41と、モデル構築プログラム42と、予測処理プログラム43とをプログラム格納部40aに記憶する。そして、演算手段2は、これらのプログラム41～43を記憶手段4から読み込んでメモリ24に展開して実行することで、前記した重み推定部21、モデル構築部22、予測処理部23の各機能を実現する。

10

【0063】

また、記憶手段4は、入力データ（系列データ）44と、重み45と、モデルパラメータ46と、サンプル47とをデータ格納部40bに記憶する。ここで、入力データ（系列データ）44は、入力手段3から入力されるデータであり、例えば、前記した表1に示した購買データである。重み45は、演算手段2の重み推定部21の演算処理結果を示すデータである。モデルパラメータ46は、演算手段2のモデル構築部22の演算処理結果を示すデータである。サンプル47は、例えば、前記した表1に示した購買データに対応したデータであり、入力 x と時刻 t とが既知であり、出力 y が未知のデータである。

20

【0064】

出力手段5は、例えば、グラフィックボード（出力インタフェース）およびそれに接続されたモニタである。モニタは、例えば、液晶ディスプレイ等から構成され、演算処理結果（例えば、予測を行った結果＜予測される購買商品の情報等＞）を表示する。

【0065】

次に、演算手段2の各部の構成の詳細を説明する。

<重み推定部>

図2を参照しながら、重み推定部21の構成について説明する。図2は、重み推定部の構成を示す機能ブロック図である。重み推定部（重み推定手段）21は、入力データ（系列データ）44から得られる、最新時刻（指標情報の予め定められた離散値）における入出力サンプル（学習用サンプルデータ）の入力サンプル（第1観測データ）および出力サンプル（第2観測データ）に対する期待誤差が、入出力サンプルが与えられたときのモデルの誤差関数と、それぞれの入出力サンプルの時間情報の重みと、を用いて近似された期待誤差を最小化させるように、時間情報の重みを推定するものである。図2に示すように、重み推定部21は、入力データ読込部211と、事後確率推定部212と、混合比推定部213と、重み書込部214とを備えている。

30

【0066】

入力データ読込部211は、入力データ44を読み込み、事後確率推定部212に出力するものである。

40

事後確率推定部（事後確率推定手段）212は、最新時刻における入力データ44の真の分布を近似するために、各時刻における経験分布に混合される予め定められた混合比と、入力データ44における入出力サンプル（ x , y ）とを条件として、入出力サンプル（ x , y ）の時間情報 t に対する事後確率をそれぞれ推定するものである。本実施形態では、事後確率推定部212は、前記した式（13）で示される経験分布を予め計算しておく。また、事後確率推定部212は、前記した式（16）により全サンプルの全時刻に対する事後確率を推定する。

【0067】

混合比推定部（混合比推定手段）213は、事後確率推定部212で推定されたそれぞ

50

れの事後確率を利用して、最新時刻における入力データ 4 4 に対する尤度を最大化するように、混合比を推定し、尤度が最大化されたときの混合比から重みを決定するものである。本実施形態では、混合比推定部 2 1 3 は、前記した式 (1 7) により混合比を推定する。また、混合比推定部 2 1 3 は、式 (1 1) に示す尤度が収束するか否かを判別し、収束していなときには、その時点の混合比を事後確率推定部 2 1 2 に出力し、一方、収束したときには、その時点の混合比から前記した式 (8) に示す重み $w(t)$ を算出し、重み書込部 2 1 4 に出力する。

重み書込部 2 1 4 は、混合比推定部 2 1 3 で推定された重み $w(t)$ を、重み 4 5 として、記憶手段 4 (図 1 参照) に書き込む。なお、書き込まれた重み 4 5 は、モデル構築部 2 2 で利用される。

【 0 0 6 8 】

< モデル構築部 >

図 3 を参照しながら、モデル構築部 2 2 の構成について説明する。図 3 は、モデル構築部の構成を示す機能ブロック図である。モデル構築部 (モデル構築手段) 2 2 は、重み推定部 2 1 で推定された時間情報 (指標情報) の重みを用いた期待誤差に基づく目的関数と入力データ (系列データ) 4 4 とを用いて、サンプル (予測用サンプルデータ) 4 7 から、最新時刻 (指標情報の予め定められた離散値) における未知の出力サンプル (第 2 観測データ) を予測するモデルを構築するものである。図 3 に示すように、モデル構築部 2 2 は、入力データ読込部 2 2 1 と、重み読込部 2 2 2 と、モデルパラメータ推定部 2 2 3 と、ハイパーパラメータ推定部 2 2 4 と、モデルパラメータ書込部 2 2 5 とを備えている。

【 0 0 6 9 】

入力データ読込部 2 2 1 は、入力データ 4 4 を読み込み、モデルパラメータ推定部 2 2 3 に出力するものである。

重み読込部 2 2 2 は、重み 4 5 を読み込み、モデルパラメータ推定部 2 2 3 に出力するものである。

モデルパラメータ推定部 (モデルパラメータ推定手段) 2 2 3 は、重み読込部 2 2 2 で読み込んだ重み 4 5 と、重み 4 5 のハイパーパラメータと、近似された期待誤差における誤差関数とに基づく目的関数を最小化させるように、入力データ (系列データ) 4 4 および重み 4 5 に基づいてモデルを学習することで、最新時刻におけるモデルを推定するものである。このモデルパラメータ推定部 2 2 3 は、ハイパーパラメータ推定部 2 2 4 で推定されたハイパーパラメータをモデル推定に用いる。本実施形態では、モデルパラメータ推定部 2 2 3 は、前記した式 (2 2) によりモデルパラメータを推定する。

【 0 0 7 0 】

ハイパーパラメータ推定部 (ハイパーパラメータ推定手段) 2 2 4 は、モデルパラメータ推定部 2 2 3 で推定されたモデルに対して適用される誤差関数を最小化させるように、ハイパーパラメータを推定するものである。本実施形態では、ハイパーパラメータ推定部 2 2 4 は、前記した式 (2 3) によりハイパーパラメータを推定する。また、ハイパーパラメータ推定部 2 2 4 は、式 (1 8) に示す誤差が収束するか否かを判別し、収束するまで、その時点のハイパーパラメータをモデルパラメータ推定部 2 2 3 に出力し続ける。これにより、モデルパラメータ推定部 2 2 3 では、誤差関数が最小化されたときのハイパーパラメータを用いて推定されたモデルを構築すべきモデルとして決定する。また、ハイパーパラメータ推定部 2 2 4 は、式 (1 8) に示す誤差が収束したと判別したときに、その時点で推定されているモデルをモデルパラメータ書込部 2 2 5 に出力する。

モデルパラメータ書込部 2 2 5 は、推定されたモデルパラメータを、モデルパラメータ 4 6 として、記憶手段 4 (図 1 参照) に書き込む。なお、書き込まれたモデルパラメータ 4 6 は、予測処理部 2 3 で利用される。

【 0 0 7 1 】

< 予測処理部 >

図 4 を参照しながら、予測処理部 2 3 の構成について説明する。図 4 は、予測処理部の構成を示す機能ブロック図である。予測処理部 (予測処理手段) 2 3 は、モデル構築部 2

10

20

30

40

50

2で構築されたモデル(モデルパラメータ46)と、サンプル(予測用サンプルデータ)47とに基づいて、最新時刻(指標情報の予め定められた離散値)におけるサンプル47の未知の出力サンプル(第2観測データ)を予測する処理を行うものである。図4に示すように、予測処理部23は、サンプル読込部231と、モデルパラメータ読込部232と、予測結果出力部233とを備えている。

【0072】

サンプル読込部231は、出力が未知のサンプル47を読み込み、予測結果出力部233に出力するものである。

モデルパラメータ読込部232は、モデルパラメータ46を読み込み、予測結果出力部233に出力するものである。

予測結果出力部233は、サンプル47とモデルパラメータ46とを使って出力サンプル(y)の予測結果を計算し、計算結果(予測結果)を出力手段5(図1参照)に出力するものである。

【0073】

[予測装置の動作]

<処理の流れ>

図1に示した予測装置1の動作について図5を参照(適宜図1参照)して説明する。図5は、予測装置の処理の流れを示す説明図である。まず、予測装置1は、重み推定部21によって、記憶手段4(図1参照)に予め格納された入力データ44に基づいて重みを推定する(ステップS1:重み推定ステップ)。推定された重みは、重み45として記憶手段4に格納される。次に、予測装置1は、モデル構築部22によって、記憶手段4(図1参照)に予め格納された入力データ44および重み45に基づいてモデルを構築する(ステップS2:モデル構築推定ステップ)。構築されたモデルは、モデルパラメータ46として記憶手段4に格納される。続いて、予測装置1は、予測処理部23によって、記憶手段4(図1参照)に予め格納された出力が未知であるサンプル47に対して、モデルパラメータ46に基づいて、対応する出力を予測する予測処理を行う(ステップS3:予測処理ステップ)。

【0074】

次に、前記したステップS1の重み推定ステップと、前記したステップS2のモデル構築推定ステップについて図6および図7をそれぞれ参照(適宜図1ないし図5参照)して説明する。図6は、重み推定処理を示すフローチャートであり、図7は、モデル構築処理を示すフローチャートである。

【0075】

<重み推定ステップ>

前記したステップS1の重み推定ステップでは、図6に示すように、重み推定部21は、入力データ読込部211によって、記憶手段4(図1参照)から、入力データ44を読み込む(ステップS11)。具体的には、入力データ読込部211は、前記した表1で示される構造をした入力データ(系列データ)を読み込む。そして、重み推定部21は、事後確率推定部212によって、経験分布の計算を行う(ステップS12)。具体的には、前記した式(13)で示される経験分布を計算する。そして、重み推定部21は、事後確率推定部212によって、初期化を行う(ステップS13)。具体的には、事後確率推定部212は、EMアルゴリズムのEステップとMステップとの2つの手順の繰り返し回数を0に設定し($\tau = 0$)、例えば所定の乱数を発生させて前記した式(6)で示す混合比の混合分布 $P(t)$ をランダムに設定する。

【0076】

そして、初期化終了後に、重み推定部21は、事後確率推定部212によって、EMアルゴリズムのEステップを実行する(ステップS14)。具体的には、事後確率推定部212は、前記した式(16)により、全サンプル(x, y)の全時刻tに対する事後確率を推定する。続いて、重み推定部21は、混合比推定部213によって、EMアルゴリズムのMステップを実行する(ステップS15)。具体的には、混合比推定部213は、前

10

20

30

40

50

記した式(17)により、第($i + 1$)ステップにおける混合分布 $P^{(i+1)}(t)$ すなわち混合比を推定する。次に、重み推定部21は、混合比推定部213によって、収束条件が満たされたか否かを判別する(ステップS16)。具体的には、混合比推定部213は、前記した式(11)に示す尤度 $L(P)$ が収束したか否かを判別する。

【0077】

収束条件が満たされた場合、すなわち前記した式(11)に示す尤度 $L(P)$ が収束した場合(ステップS16: Yes)、混合比推定部213は、重み $w(t)$ を計算する(ステップS17)。具体的には、混合比推定部213は、収束した時点における混合分布 $P(t)$ とサンプル数 $N(t)$ とに基づいて前記した式(8)に示す重み $w(t)$ を全時刻 t について算出する。そして、重み推定部21は、重み書込部214によって、前記した式(8)に示す重み $w(t)$ を、重み45として、記憶手段4(図1参照)に書き込み(ステップS18)、処理を終了する。この重み45は、最新時刻 T におけるモデルを構築するために時刻 t に対して推定された重みを示す。

10

【0078】

一方、ステップS16において、収束条件が満たされていない場合、すなわち前記した式(11)に示す尤度 $L(P)$ が収束していない場合(ステップS16: No)、重み推定部21は、EステップおよびMステップの繰り返し回数に「1」を加算し($i = i + 1$)、繰り返し回数を更新し(ステップS19)、ステップS14に戻る。

【0079】

<モデル構築ステップ>

20

前記したステップS2のモデル構築ステップでは、図7に示すように、モデル構築部22は、まず、入力データ(系列データ)と重みとを読み込む(ステップS21)。具体的には、モデル構築部22は、入力データ読込部221によって、記憶手段4(図1参照)から、入力データ44を読み込む。また、モデル構築部22は、重み読込部222によって、記憶手段4(図1参照)から、重み推定処理により推定した重み45を読み込む。なお、入力データ44の読み込みと、重みの読み込みとの実行順序は、任意であり、処理を並列に実行してもよい。

【0080】

そして、モデル構築部22は、モデルパラメータ推定部223によって、初期化を行う(ステップS22)。具体的には、モデルパラメータ推定部223は、学習データ D を K 個の部分集合に分割し、ハイパーパラメータを0に設定する($\alpha = 0$)。そして、モデル構築部22は、モデルパラメータ推定部223によって、モデルの学習を行う(ステップS23)。具体的には、モデルパラメータ推定部223は、 K 重交差検定法にしたがって、前記した式(22)で最新時刻 T におけるモデルを学習する。そして、モデル構築部22は、ハイパーパラメータ推定部224によって、ハイパーパラメータの学習を行う(ステップS24)。具体的には、ハイパーパラメータ推定部224は、前記した式(23)でハイパーパラメータを学習する。そして、モデル構築部22は、ハイパーパラメータ推定部224によって、収束条件が満たされたか否かを判別する(ステップS25)。具体的には、ハイパーパラメータ推定部224は、前記した式(18)に示す誤差 $E(M, \alpha)$ が収束したか否かを判別する。

30

40

【0081】

収束条件が満たされた場合、すなわち前記した式(18)に示す誤差 $E(M, \alpha)$ が収束した場合(ステップS25: Yes)、モデル構築部22は、モデルパラメータ書込部225によって、その時点のモデルをモデルパラメータ46として記憶手段4(図1参照)に書き込み(ステップS26)、処理を終了する。このモデルパラメータ46は、最新時刻 T におけるモデルとして推定されたモデルを示す。一方、ステップS25において、収束条件が満たされていない場合、すなわち前記した式(18)に示す誤差 $E(M, \alpha)$ が収束していない場合(ステップS25: No)、モデル構築部22は、ステップS23に戻る。

【0082】

50

なお、予測装置 1 は、一般的なコンピュータに、前記した各ステップを実行させる予測プログラムを実行することで実現することもできる。このプログラムは、通信回線を介して配布することも可能であるし、CD-ROM等の記録媒体に書き込んで配布することも可能である。

【0083】

本実施形態によれば、予測装置 1 は、時間情報の重みを推定することによって、最新時刻のデータに適合するモデルを学習するために有益な情報を、最新時刻のデータ以外の過去のデータから取り込むことで、最新時刻での経験誤差を、最新時刻での期待誤差に対する精度のよい近似として用いることができる。そのため、予測装置 1 は、時間とともに性質が変化する時系列データにおいて、最新の時刻におけるデータを高い精度で予測することができる。その結果、例えば、ユーザの履歴購買データから、ユーザの現時点（最新時刻）における購買傾向を予測し、予測された商品をユーザにリコメンドすることができる。これにより、ユーザが所望する商品の情報に対して迅速にアクセスできるように利便性を向上させる効果と、商品提供者の収益を増加させる効果とをもたらすことが可能となる。

10

【0084】

以上、本発明の実施形態について説明したが、本発明はこれに限定されるものではなく、その趣旨を変えない範囲で実施することができる。例えば、重み推定処理において、前記した式(5)による近似は一例であって、条件に応じて可変とすることができる。例えば、式(24)による近似を用いることもできる。すなわち、時刻 t における入出力サンプルの条件付き分布 $P(x|y, t)$ と、時刻 t とは異なる時刻 t' における入出力サンプルの条件付き分布 $P(x|y, t')$ とが等しく、かつ、出力分布のみ時刻により異なる状況においては、最新時刻 T での出力サンプル y の分布 $P(y|T)$ を各時刻の出力サンプル y の経験分布の混合で近似するように混合比を決定することができる。このとき、式(24)の関係式を満たす混合比は、前記した式(5)の関係式を満たす混合比の近似となる。

20

【0085】

【数17】

$$P(y|T) \approx \sum_t P(t) \hat{P}(y|t) \quad \dots \text{式(24)}$$

30

$$\text{ただし、} P(x|y, t) = P(x|y, t') \quad \wedge \quad P(y|t) \neq P(y|t')$$

【0086】

式(24)を用いる場合には、予測装置 1 の重み推定部 2 1 の事後確率推定部 2 1 2 は、混合比と、入力データ 4 4 における出力サンプル (y) とを条件として、出力サンプル (y) の時間情報 t に対する事後確率をそれぞれ推定する。具体的には、事後確率推定部 2 1 2 は、前記した式(13)の代わりに式(13A)で示される経験分布を予め計算しておく。式(13A)において、時刻が t 、出力が y のサンプル数を $N(t, y)$ とする。また、この場合、事後確率推定部 2 1 2 は、前記した式(16)の代わりに式(16A)により全サンプルの全時刻に対する事後確率を推定する。さらに、この場合、混合比推定部 2 1 3 は、前記した式(17)の代わりに式(17A)により混合比を推定する。

40

【0087】

【数 18】

$$\hat{P}_{-n}(y|t) = \frac{N(t, y) + \alpha - \delta_{t,T}}{N(t) + \alpha |Y| - \delta_{t,T}} \quad \dots \text{式 (13A)}$$

$$P(t | y_n; \mathbf{P}^{(\tau)}) = \frac{P^{(\tau)}(t) \hat{P}_{-n}(y_n | t)}{\sum_{t'} P^{(\tau)}(t') \hat{P}_{-n}(y_n | t')} \quad \dots \text{式 (16A)}$$

$$P^{(\tau+1)}(t) = \frac{1}{N(t)} \sum_{n:t_n=T} P(t | y_n; \mathbf{P}^{(\tau)}) \quad \dots \text{式 (17A)}$$

10

【0088】

また、本実施形態では、サンプルが与えられたときのモデルの誤差関数 J として前記した式 (2) および式 (3) で示すものを例示したが、任意の誤差関数を用いることが可能である。さらに、本実施形態では、例えば、前記した式 (1) において T を最新時刻であるものとして説明したが、T は最新時刻に限らず任意の時刻とすることができる。

【0089】

さらにまた、本実施形態では、時間とともに性質が変化する時系列データを扱い、例えば、前記した式 (4) において t を時間情報としたが、t は離散変数であれば時間以外の情報であってもよい。この場合には、t を例えば特定の場所や地域を指し示す地域情報 (例えば空間座標) とすることができる。地域情報 t に対応付けた購買データの一例を表 2 に示す。これにより、対象とする所定の地域に関するデータについて、その地域とは異なる他の地域に関するデータを重み付けて考慮することで、対象とする地域に関する未知データを高い精度で予測することができる。

20

【0090】

【表 2】

インデックス n 購買番号	入力 x 購買履歴	出力 y 購買商品	地域情報 t 購買地域
1	商品A	商品B	1(東京)
2	商品B	商品C	1(東京)
3	商品A	商品C	1(東京)
4	商品B	商品C	2(大阪)
...
N	商品D	商品A	T(神奈川)

30

【0091】

また、予測装置 1 を構成する装置は、1 台に限定されることはなく、複数の装置に機能を分散配置してもよい。例えば、演算手段 2 の重み推定部 2 1、モデル構築部 2 2、予測処理部 2 3 や、記憶手段 4 のデータ格納部 4 0 b を、別々の装置として構成してもよい。これにより、各装置への負荷が分散され、高速な処理が実現可能となる。

40

【実施例】

【0092】

本発明の効果を確認するために、本実施形態に係る予測装置 1 に、動画配信サービスの購買履歴を入力する場合の商品の予測精度と、携帯電話用漫画配信サービスの購買履歴を入力する場合の商品の予測精度とを実験により求めた。

【0093】

50

< 設定 >

動画配信サービスの購買履歴（以下、動画データという）は、2007年1月1日から2007年1月31日までの動画配信サービスにおける動画データを示す。この動画データにおいて、ユーザ数は「40,282」、商品数は「5,914」、購買数は「1,667,414」であった。携帯電話用漫画配信サービスの購買履歴（以下、漫画データという）は、2005年4月1日から2006年3月31日の携帯電話用漫画配信サービスにおける漫画データを示す。この漫画データにおいて、ユーザ数は「164,538」、商品数は「175」、購買数は「1,018,741」であった。ここで1つの漫画が複数巻あるものは同一の商品として扱った。

【0094】

なお、動画データおよび漫画データから、売上数が「1」以下の商品を省くと共に、購買数が「1」以下であるユーザを省いた。また、ある商品が2回以上同一ユーザに購入された場合、その商品に関する2回目以降の購買を購買履歴から省いた。

【0095】

< 定式化 >

この場合のモデルMとして1次マルコフモデルを用いた。また、1次マルコフモデルの誤差関数として、式(25)に示すように、負の対数尤度を用いた。式(25)において、 θ_{ij} は、商品iの次に商品jを購入する確率（ $0 \leq \theta_{ij} \leq 1$, $\sum_j \theta_{ij} = 1$ ）を意味するモデルパラメータである。また、商品 x_n は、あるユーザ U_n が商品 y_n を購入したときにその1つ前に購入した商品を表すものとする。

【0096】

【数19】

$$J(x_n, y_n; M) = -\log \theta_{x_n, y_n} \quad \dots \text{式 (25)}$$

【0097】

そして、この場合、前記した式(18)に示す誤差Eは、式(26)に示すように書き換えられる。つまり、式(26)に示すEが、最小にすべき目的関数である。

【0098】

【数20】

$$E(\theta, \lambda) = -\sum_n w(t_n)^\lambda \log \theta_{x_n, y_n} \quad \dots \text{式 (26)}$$

【0099】

ここで、前記した式(25)に示す誤差関数Jが最小となるモデルパラメータ θ_{ij} を式(27)に示す。式(27)において、 λ はスムージングパラメータである。以下の計算では、 $\lambda = 10^{-2}$ を用いた。また、Vは商品数を示す。なお、式(27)においてに付した記号「 $\hat{\cdot}$ （ハット）」は、 θ_{ij} が誤差関数を最小化させることを示すものである。

【0100】

【数21】

$$\hat{\theta}_{i,j} = \frac{\sum_{n: x_n=i \wedge y_n=j} w(t_n)^\lambda + \beta}{\sum_{n: x_n=i} w(t_n)^\lambda + \beta V} \quad \dots \text{式 (27)}$$

【0101】

(実施例1、実施例2)

実施例1(OurXY)では、前記した式(5)に基づいて、最新時刻Tの同時分布(入出力分布 x, y)を近似するように重み $w(t)$ を設定した。また、実施例2(OurY)では、前記した式(24)に基づいて、最新時刻Tの出力分布 y を近似するように重み $w(t)$

10

20

30

40

50

を)を設定した。これら実施例1(OurXY)および実施例2(OurY)では、前記した式(13)に示すスムージングパラメータの値を、 $\sigma = 10^{-8}$ とし、重みの最大値で各時刻tにおける重みの値を割った後に、ハイパーパラメータを推定した。ここで、重みの最大値が「1」となるようにハイパーパラメータを推定した。また、10重交差検定で汎化誤差が最小になるように黄金分割法によりを推定した。なお、このときにハイパーパラメータの取り得る区間は[0, 10]とした。

【0102】

(比較例1、比較例2)

実施例1(OurXY)および実施例2(OurY)を以下の2つのモデル(比較例1、比較例2)と比較した。

比較例1(NoWeight)：日にちによる重みの変動なし($w(t) = 1$)

比較例2(Present)：最新日のみ重みあり($w(T) = 1$ 、 $t < T$ で $w(t) = 0$)

【0103】

<実験方法>

実験では時間情報の単位時間を1日とした。つまり、最新時刻とは、最新日すなわち最終日を意味する。そして、実験では、まずはじめの段階、すなわち、1日目では、最終日Tを配信開始日に設定した。次の段階、すなわち、2日目では、最終日Tを1日ずらし、配信開始日の次の日に設定した。以下同様に、最終日Tを1日ずつずらしていった。各段階、すなわち、それぞれの日にち(それぞれの最終日T)では、その最終日T以前の購買履歴のみを用いて、その最終日Tにおけるデータを作成した。そして、それぞれの日にちTにおいて、その性能を式(28)に示すパープレキシティL(T)により評価した。

【0104】

【数22】

$$L(T) = \exp\left(-\frac{1}{N_T} \sum_{n: t_n=T} \log \theta_{x_n, y_n}\right) \quad \dots \text{式(28)}$$

【0105】

式(28)において、 N_T は、その日にちの日付T(最新日T)におけるサンプル数である。パープレキシティは、その値が低い場合に、予測性能が高いことを表す。それぞれの日にちTにおいて、まず、動画データについて10重交差検定により、学習およびテストデータを10セット作成し、その日にちTのパープレキシティをそれぞれ求めた。そして、式(29)に示すように、すべての日数Dで平均したパープレキシティ(平均パープレキシティ)Lを求めた。式(29)に示すDは、動画データにおいては、 $D = 31$ である。同様に漫画データについても平均パープレキシティLを求めた。ここで、漫画データにおいては、 $D = 365$ である。

【0106】

【数23】

$$L = \frac{1}{D} \sum_T L(T) \quad \dots \text{式(29)}$$

【0107】

<実験結果>

実施例および比較例についての実験結果を表2に示す。

【0108】

10

20

30

40

【表 3】

平均パープレキシティL

	実施例1 OurXY	実施例2 OurY	比較例1 NoWeight	比較例2 Present
動画	163.50	163.46	174.58	319.53
漫画	32.48	32.58	46.94	75.89

【0109】

10

表3に示すように、実施例1(OurXY)および実施例2(OurY)は、比較例1(NoWeight)および比較例2(Present)と比べて、平均パープレキシティLが低い。言い換えると、実施例1(OurXY)および実施例2(OurY)は、比較例1(NoWeight)および比較例2(Present)と比べて、予測性能が高い。つまり、本発明では、その特徴である時間情報の重み付けにより、最新日のデータをより正確に予測できるモデルが学習できていると言える。なお、比較例1(NoWeight)は、最新日とは性質の異なる過去の時刻のデータも学習に用いているため低い予測精度となっている。また、比較例2(Present)は、学習に用いるデータの数が少ないため、低い予測精度となっている。

【図面の簡単な説明】

【0110】

20

【図1】本発明の実施形態に係る予測装置の構成を示すブロック図である。

【図2】重み推定部の構成を示す機能ブロック図である。

【図3】モデル構築部の構成を示す機能ブロック図である。

【図4】予測処理部の構成を示す機能ブロック図である。

【図5】予測装置の処理の流れを示す説明図である。

【図6】重み推定処理を示すフローチャートである。

【図7】モデル構築処理を示すフローチャートである。

【符号の説明】

【0111】

1 予測装置

30

2 演算手段

3 入力手段

4 記憶手段

5 出力手段

11 バスライン

21 重み推定部(重み推定手段)

22 モデル構築部(モデル構築手段)

23 予測処理部(予測処理手段)

24 メモリ

40a プログラム格納部

40

41 重み推定プログラム

42 モデル構築プログラム

43 予測処理プログラム

40b データ格納部

44 入力データ(系列データ)

45 重み

46 モデルパラメータ

47 サンプル(予測用サンプルデータ)

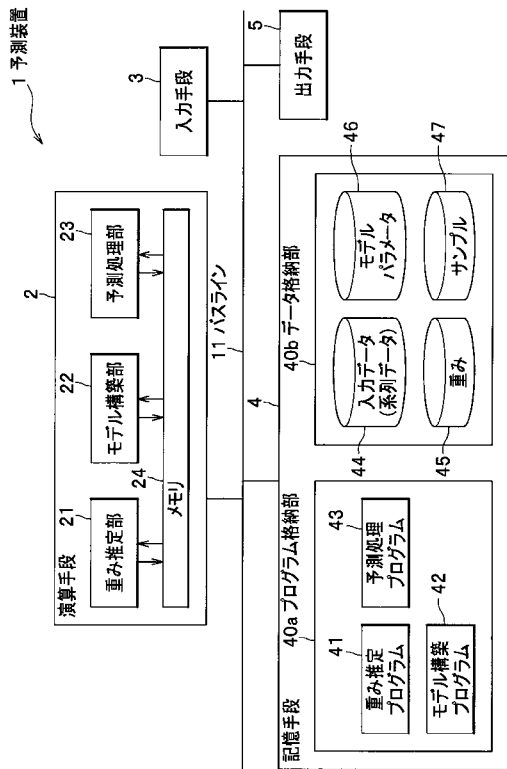
211 入力データ読込部

212 事後確率推定部(事後確率推定手段)

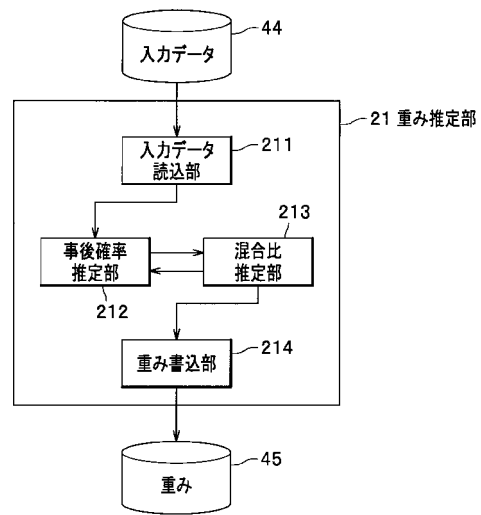
50

- 2 1 3 混合比推定部 (混合比推定手段)
- 2 1 4 重み書込部
- 2 2 1 入力データ読込部
- 2 2 2 重み読込部
- 2 2 3 モデルパラメータ推定部 (モデルパラメータ推定手段)
- 2 2 4 ハイパーパラメータ推定部 (ハイパーパラメータ推定手段)
- 2 2 5 モデルパラメータ書込部
- 2 3 1 サンプル読込部
- 2 3 2 モデルパラメータ読込部
- 2 3 3 予測結果出力部

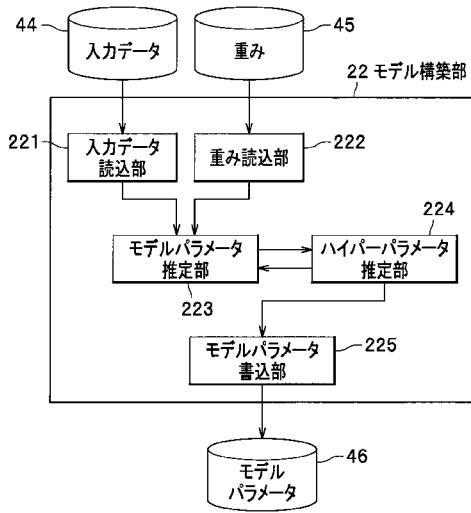
【 図 1 】



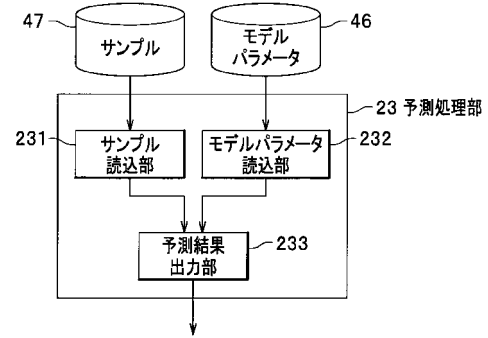
【 図 2 】



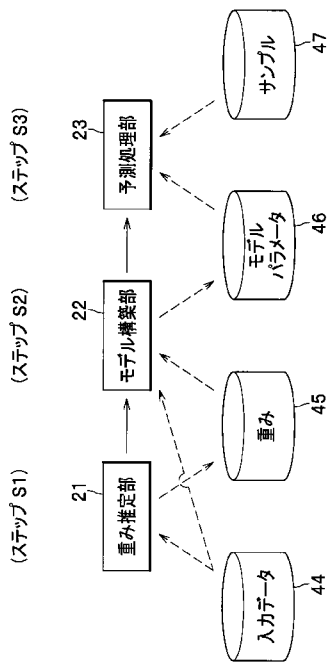
【 図 3 】



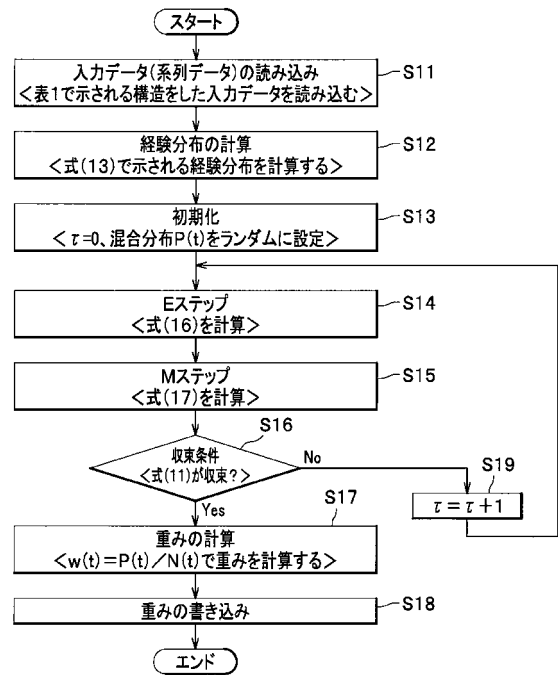
【 図 4 】



【 図 5 】



【 図 6 】



【 図 7 】

