

- 〔続葉有〕

[图1]

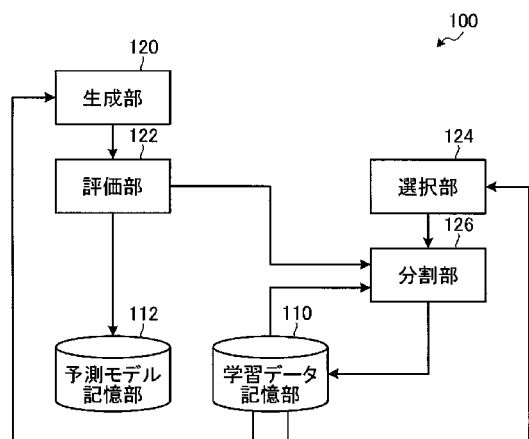


FIG. 1:
110 LEANING DATA STORAGE UNIT
112 PREDICTION MODEL STORAGE UNIT
120 GENERATION UNIT
122 EVALUATION UNIT
124 SELECTION UNIT
126 DIVISION UNIT

(57) Abstract: A learning data storage unit (110) stores learning data which is a set consisting of groups wherein a plurality of types of fixed first explanatory variates and time-series second explanatory variates are corresponded to an explained variate. A generation unit (120) learns the learning data, and generates a prediction model which explains a relationship of the plurality of types of first explanatory variates and second explanatory variates with the explained variate. An evaluation unit (122) evaluates the appropriateness of the prediction model, and registers the prediction model which has been evaluated as appropriate in a prediction model storage unit (112). A selection unit (124) calculates first evaluation values of the respective first explanatory variates when the prediction model is evaluated as inappropriate, and selects a selection explanatory variate which has the maximum first evaluation value among the first explanatory variates. A division unit (126) divides the learning data into sets each including groups having the same value regarding the selection explanatory variate, and registers the divided learning data in the learning data storage unit (110) as new learning data. A generation unit (120) generates a prediction model with respect to each of the new learning data by learning the new learning data.

(57) 要約:

〔続葉有〕



添付公開書類:

— 国際調査報告 (条約第 21 条(3))

学習データ記憶部 110 は、複数種類の固定的な第 1 説明変量と時系列的な第 2 説明変量と被説明変量とを対応付けた組の集合である学習データを記憶する。生成部 120 は、学習データを学習して、複数種類の第 1 説明変量及び第 2 説明変量と被説明変量との間の関係を説明する予測モデルを生成する。評価部 122 は、予測モデルの妥当性を評価し、妥当であると評価した予測モデルを予測モデル記憶部 112 に登録する。選択部 124 は、予測モデルが妥当でないと評価された場合に、第 1 説明変量それぞれの第 1 評価値を算出し、第 1 説明変量のうち最大の第 1 評価値を有する選択説明変量を選択する。分割部 126 は、同じ値の選択説明変量を有する組の集合に学習データを分割し、分割した学習データそれぞれを新たな学習データとして学習データ記憶部 110 に登録する。生成部 120 は、新たな学習データ毎に、当該新たな学習データを学習して予測モデルを生成する。

明 細 書

発明の名称： 予測モデル生成装置

技術分野

[0001] 本発明は、予測モデル生成装置に関する。

背景技術

[0002] 従来から、説明変量と被説明変量との間の関係をモデル化する技術が知られている。例えば特許文献 1 には、ユーザに選択された説明変量の組み合わせを用いて評価用のクラスターを仮分類し、仮分類された評価用のクラスターを代表する被説明変量の値を求めて、仮分類の継続・停止を判定する技術が開示されている。特許文献 1 に開示された技術によれば、説明変量と被説明変量との間の関係を説明する最適なクラスター分類を決定することができる。また、例えば非特許文献 1 には、回帰分析について記載されており、この回帰分析を利用することにより、固定的な説明変量及び時系列的な説明変量と被説明変量との間の関係を説明する予測モデルを生成することができる。

先行技術文献

特許文献

[0003] 特許文献1：特開 2 0 0 0 - 1 3 2 5 5 8 号公報

非特許文献

[0004] 非特許文献1：竹内啓著 「数理統計学」 東洋経済新報社 1 9 6 3 年 p . 3 0 5 - 3 2 6

発明の概要

発明が解決しようとする課題

[0005] しかしながら、特許文献 1 に開示された技術では、固定的な説明変量のみを対象としており、固定的な説明変量と時系列的な説明変量とを含む学習データの集合から、予測モデルを生成することはできなかった。また、非特許文献 1 に開示された回帰分析を利用する場合、学習データに含まれる全ての

固定的な説明変量及び時系列的な説明変量を一括して取り扱う必要があるため、説明変量に内在する複数の特徴を一括して処理することとなり、精度の高い予測モデルを構築することは困難であった。

[0006] 本発明は、上記事情に鑑みてなされたものであり、説明変量に内在する複数の特徴を考慮して、固定的な説明変量及び時系列的な説明変量と被説明変量との間の関係を説明する予測モデルを生成することができる予測モデル生成装置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

[0007] 本発明の一態様にかかる予測モデル生成装置は、複数種類の固定的な第1説明変量と時系列的な第2説明変量と被説明変量とを対応付けた組の集合である学習データを記憶する学習データ記憶部と、前記学習データを学習して、複数種類の前記第1説明変量及び前記第2説明変量と前記被説明変量との間の関係を説明する予測モデルを生成する生成部と、前記予測モデルの妥当性を評価し、妥当であると評価した前記予測モデルを予測モデル記憶部に登録する評価部と、前記予測モデルが妥当でないと評価された場合に、前記第1説明変量それぞれの第1評価値を算出し、前記第1説明変量のうち最大の前記第1評価値を有する選択説明変量を選択する選択部と、同じ値の前記選択説明変量を有する組の集合に前記学習データを分割し、分割した前記学習データそれぞれを新たな学習データとして前記学習データ記憶部に登録する分割部と、を備え、前記生成部は、前記新たな学習データ毎に、当該新たな学習データを学習して前記予測モデルを生成することを特徴とする。

発明の効果

[0008] 本発明によれば、説明変量に内在する複数の特徴を考慮して、固定的な説明変量及び時系列的な説明変量と被説明変量との間の関係を説明する予測モデルを生成することができるという効果を奏する。

図面の簡単な説明

[0009] [図1]本実施形態の予測モデル生成装置を示す図。

[図2]学習データを示す図。

[図3] 図2に示す学習データを「ライン」の値で分割した場合における、ラインの値がトップスとなる学習データを示す図。

[図4] 図2に示す学習データを「ライン」の値で分割した場合における、ラインの値がボトムとなる学習データを示す図。

[図5] 図2に示す学習データを「ライン」の値で分割した場合における、ラインの値が雑貨となる学習データを示す図。

[図6] 図4に示す学習データを「サイズ」の値で分割した場合における、サイズの値がSとなる学習データを示す図。

[図7] 図4に示す学習データを「サイズ」の値で分割した場合における、サイズの値がMとなる学習データを示す図。

[図8] 図6に示す学習データを「色」の値で分割した場合における、色の値が青となる学習データを示す図。

[図9] 図6に示す学習データを「色」の値で分割した場合における、色の値が黄となる学習データを示す図。

[図10] 本実施形態の処理を示すフローチャート。

[図11] 図2～9に示す学習データから生成される予測モデルを示す図。

発明を実施するための形態

[0010] 以下、添付図面を参照しながら、本発明の予測モデル生成装置の実施形態を説明する。なお、以下の実施形態では、アパレルショップの店内で販売されている各商品に付与されているRFID (Radio Frequency Identification) タグをRFIDリーダで読み込むことにより収集されるRFIDデータを学習データとして予測モデルを生成する例について説明するが、これに限定されるものではない。

[0011] 図1は、本実施形態の予測モデル生成装置100の構成の一例を示すブロック図である。図1に示すように、予測モデル生成装置100は、学習データ記憶部110と、生成部120と、評価部122と、予測モデル記憶部112と、選択部124と、分割部126と、を備える。

[0012] 学習データ記憶部110は、複数種類の固定的な第1説明変量と時系列的

な第2説明変量と被説明変量とを対応付けた組の集合である学習データを記憶する。本実施形態では、アパレルショップから収集されるRFIDデータが学習データとして学習データ記憶部110に記憶されるものとする。学習データ記憶部110は、例えば、HDD（Hard Disk Drive）、SSD（Solid State Drive）、メモ리카ード、光ディスクなどの磁氣的、光学的、又は電氣的に記憶可能な既存の記憶装置により実現できる。

[0013] 第1説明変量は、時系列が変化してもその値が変化しない固定的な説明変量である。本実施形態の第1説明変量としては、例えば、色（赤，青，…）、サイズ（S，M，…）、ライン（ボトム，トップス，…）、地域（関東，東北，…）、店舗業態（ショッピングモール内店舗，独立店舗，…）、素材（絹，ナイロン，…）、形状（丸首，V字，…）、又は対象季節（夏物，オールシーズン，…）などが挙げられる。

[0014] 第2説明変量は、時系列の変化に伴いその値が変化する時系列的な説明変量である。本実施形態の第2説明変量としては、例えば、ピックアップ数（非負の整数値）、試着数（非負の整数値）、販売数（非負の整数値）、温度（非負の整数値）、湿度（非負の整数値）、雨量（非負の整数値）、来店数（非負の整数値）、店頭在庫数（非負の整数値）、倉庫在庫数（非負の整数値）、又は店員数（非負の整数値）などが挙げられる。

[0015] 本実施形態の被説明変量としては、例えば、次期販売数（非負の整数値）、次期店頭在庫数（非負の整数値）、次期倉庫在庫数（非負の整数値）、次期店員数（非負の整数値）などが挙げられる。

[0016] なお、各説明変量の後の括弧書きは、当該説明変量を取りうる値を示している。

[0017] 図2は、学習データ記憶部110に記憶されている学習データの一例を示す図である。図2に示す例では、学習データは、複数種類の第1説明変量（色、サイズ、及びライン）と、複数種類の第2説明変量（ピックアップ数、試着数、及び販売数）と、被説明変量（次期販売数）とを対応付けた組の集

合となっている。ここでは、第1説明変量の「色」が取りうる値は、「赤」、「青」、又は「黄」であり、第1説明変量の「サイズ」が取りうる値は、「S」、「M」、又は「F」であり、第1説明変量の「ライン」が取りうる値は、「トップス」、「ボトム」、又は「雑貨」となっている。なお、各組に対応付けられている $t_1 \sim t_9$ は、各組の番号を示している。

[0018] 図1に戻り、生成部120は、学習データを学習して、複数種類の第1説明変量及び第2説明変量と被説明変量との間の関係を説明する予測モデルを生成する。

[0019] 具体的には、生成部120は、学習データに含まれる各組の第1説明変量及び第2説明変量に変量を設定する。生成部120は、第1説明変量については、当該第1説明変量が取りうる値（但し、最後の値は除く）毎に、その値である場合には1、その値でない場合には0となる変量を設定する。また、生成部120は、第2説明変量については、当該第2説明変量の値をそのまま変量に設定する。

[0020] そして、生成部120は、複数種類の第1説明変量及び第2説明変量と被説明変量との間の関係を説明可能な第1回帰式を設定し、学習データに含まれる各組の第1説明変量及び第2説明変量に設定した変量を第1回帰式に適用した場合における残差が最小となるように、第1回帰式に付随するパラメータを決定する。これにより、パラメータ決定後の第1回帰式が予測モデルとして生成される。

[0021] 例えば、生成部120は、図2に示す学習データに対して回帰分析を利用することにより、説明変量 x_i ($i = 1, 2, \dots, 9$) と、被説明変量 y との間の関係を、数式(1)に示す第1回帰式によってモデル化する。

[0022] [数1]

$$y = \alpha_L + \sum_{i=1}^9 \beta_{Li} \times x_i \quad \cdots(1)$$

[0023] ここで、 x_1 は、第1説明変量である「色」の値が「赤」であるか否かを識別する変量であり、 x_2 は、「色」の値が「青」であるか否かを識別する変量

であり、 x_3 は、第1説明変量である「サイズ」の値が「S」であるか否かを識別する変量であり、 x_4 は、「サイズ」の値が「M」であるか否かを識別する変量であり、 x_5 は、第1説明変量である「ライン」の値が「トップス」であるか否かを識別する変量であり、 x_6 は、「ライン」の値が「ボトム」であるか否かを識別する変量である。また、 x_7 は、第2説明変量である「ピックアップ数」の値を示す変量であり、 x_8 は、第2説明変量である「試着数」の値を示す変量であり、 x_9 は、第2説明変量である「販売数」の値を示す変量である。また、 y は、被説明変量である「次期販売数」の値を示す変量である。そして、 α_i 及び β_{Li} ($i = 1, 2, \dots, 9$) が、説明変量 x_i 及び被説明変量 y から決定されるパラメータである。

[0024] なお、数式(1)に示す第1回帰式では、 x_1 及び x_2 の値が決まると、「色」の値が「黄」であるか否かも自動的に決まるため、「色」の値が「黄」であるか否かを識別する変量は回帰式に組み込まれていない。同様の理由で、「サイズ」の値が「F」であるか否かを識別する変量、及び「ライン」の値が「雑貨」であるか否かを識別する変量も、回帰式に組み込まれていない。

[0025] また、生成部120は、学習データに含まれる各組の第1説明変量及び第2説明変量に設定した変量を第1回帰式に適用することにより自由度調整済みの重回帰係数を算出する。

[0026] 評価部122は、予測モデルの妥当性を評価し、妥当であると評価した予測モデルを予測モデル記憶部112に登録する。

[0027] 具体的には、評価部122は、パラメータ決定後の第1回帰式の良さを評価する第2評価値が予め定められた第1閾値以上である場合に予測モデルが妥当であると評価し、第2評価値が閾値未満である場合に予測モデルが妥当でないと評価する。ここで、第1回帰式を評価する第2評価値としては多様なものが考えられるが、例えば、非特許文献1に記載されている自由度調整済の重相関係数を用いることができる。なお、自由度調整済の重相関係数は、生成部120により算出される。

- [0028] そして、評価部 122 は、第 2 評価値が第 1 閾値以上であると判定した場合、生成部 120 により生成された予測モデルを予測モデル記憶部 112 に登録する。具体的には、評価部 122 は、第 1 回帰式を構成する説明変量、及び第 1 回帰式を構成するパラメータを予測モデルとして予測モデル記憶部 112 に登録する。一方、評価部 122 は、第 2 評価値が第 1 閾値未満であると判定した場合、生成部 120 により生成された予測モデルの予測モデル記憶部 112 への登録は行わない。
- [0029] 予測モデル記憶部 112 は、評価部 122 により妥当であると評価された予測モデルを記憶するものであり、学習データ記憶部 110 同様、既存の記憶装置により実現できる。
- [0030] 選択部 124 は、予測モデルが妥当でないと評価された場合に、第 1 説明変量それぞれの第 1 評価値を算出し、算出した第 1 評価値が最大となる第 1 説明変量（選択説明変量）を選択する。第 1 評価値には、例えば、第 1 説明変量だけを説明変量として第 2 回帰式を求めた場合における自由度調整済の重相関係数を用いることができる。
- [0031] 例えば図 2 に示すように、第 1 説明変量が、「色」、「サイズ」、及び「ライン」である場合、選択部 124 は、第 1 説明変量が「色」でその値が「赤」である変量を x_1 、第 1 説明変量が「色」でその値が「青」である変量を x_2 に設定し、被説明変量である次期販売数 y との間の関係を説明する第 2 回帰式を設定する。また、選択部 124 は、第 1 説明変量が「サイズ」でその値が「S」である変量を x_1 、第 1 説明変量が「サイズ」でその値が「M」である変量を x_2 に設定し、被説明変量である次期販売数 y との間の関係を説明する第 2 回帰式を設定する。また、選択部 124 は、第 1 説明変量が「ライン」でその値が「トップス」である変量を x_1 、第 1 説明変量が「ライン」でその値が「ボトム」である変量を x_2 に設定し、被説明変量である次期販売数 y との間の関係を説明する第 2 回帰式を設定する。
- [0032] そして、選択部 124 は、設定した各第 2 回帰式に学習データを適用することにより、各第 2 回帰式に対応する自由度調整済の重相関係数を第 1 評価

値として算出し、算出した第 1 評価値が最大となる第 1 説明変量（第 2 回帰式）を選択する。例えば、「色」、「サイズ」、「ライン」の自由度調整済の重相関係数が、それぞれ、0.6、0.7、0.8である場合、選択部 124 は、重相関係数が最大となる第 1 説明変量として「ライン」を選択する。

[0033] 分割部 126 は、選択部 124 により選択された第 1 説明変量の値が同一である組の集合に学習データを分割し、分割した学習データそれぞれを新たな学習データとして学習データ記憶部 110 に登録する。

[0034] 例えば図 2 に示す学習データにおいて、選択部 124 により第 1 説明変量として「ライン」が選択された場合、分割部 126 は、図 2 に示す学習データを、「ライン」の値が「トップス」である組の集合（図 3 参照）、「ライン」の値が「ボトム」である組の集合（図 4 参照）、及び「ライン」の値が「雑貨」である組の集合（図 5 参照）に分割する。そして、分割部 126 は、分割した学習データそれぞれを新たな学習データとして学習データ記憶部 110 に登録する。

[0035] また例えば、生成部 120 により図 4 に示す学習データから予測モデルが生成され、評価部 122 により予測モデルが妥当でないと評価され、選択部 124 により第 1 説明変量として「サイズ」が選択されたとする。この場合、分割部 126 は、図 4 に示す学習データを、「サイズ」の値が「S」である組の集合（図 6 参照）、「サイズ」の値が「M」である組の集合（図 7 参照）に分割する。そして、分割部 126 は、分割した学習データそれぞれを新たな学習データとして学習データ記憶部 110 に登録する。

[0036] また例えば、生成部 120 により図 6 に示す学習データから予測モデルが生成され、評価部 122 により予測モデルが妥当でないと評価され、選択部 124 により第 1 説明変量として「色」が選択されたとする。この場合、分割部 126 は、図 6 に示す学習データを、「色」の値が「青」である組の集合（図 8 参照）、「色」の値が「黄」である組の集合（図 9 参照）に分割する。そして、分割部 126 は、分割した学習データそれぞれを新たな学習デ

ータとして学習データ記憶部 110 に登録する。

[0037] ここで、分割部 126 により学習データ記憶部 110 に登録された新たな学習データに対する生成部 120 及び評価部 122 の処理について説明する。

[0038] 例えば、生成部 120 は、図 3 に示す学習データを学習して予測モデルを生成する場合、図 3 に示す学習データに対して回帰分析を利用することにより、説明変量 x_i ($i = 1, 2, \dots, 7$) と、被説明変量 y との間の関係を、数式 (2) に示す第 1 回帰式によってモデル化する。

[0039] [数2]

$$y = \alpha_{LT} + \sum_{i=1}^7 \beta_{LTi} \times x_i \quad \dots(2)$$

[0040] ここで、 x_1 は、第 1 説明変量である「色」の値が「赤」であるか否かを識別する変量であり、 x_2 は、「色」の値が「青」であるか否かを識別する変量であり、 x_3 は、第 1 説明変量である「サイズ」の値が「S」であるか否かを識別する変量であり、 x_4 は、「サイズ」の値が「M」であるか否かを識別する変量である。また、 x_5 は、第 2 説明変量である「ピックアップ数」の値を示す変量であり、 x_6 は、第 2 説明変量である「試着数」の値を示す変量であり、 x_7 は、第 2 説明変量である「販売数」の値を示す変量である。また、 y は、被説明変量である「次期販売数」の値を示す変量である。そして、 α_{LT} 及び β_{LTi} ($i = 1, 2, \dots, 7$) が、説明変量 x_i 及び被説明変量 y から決定されるパラメータである。なお、図 3 に示す例では、第 1 説明変量である「ライン」の値は「トップス」のみとなっており、特定の第 1 説明変量の値が特定の値のみとなっているため、「ライン」は、第 1 回帰式に組み込まれていない。

[0041] そして、評価部 122 は、図 3 に示す学習データから生成した予測モデルを妥当であると評価する場合、図 3 に示す学習データに分割される際に選択部 124 により選択された第 1 説明変量である「ライン」、及び図 3 に示す学習データにおける「ライン」の値である「トップス」を、予測モデルに対

応付けて予測モデル記憶部 112 に登録する。

[0042] また例えば、生成部 120 は、図 5 に示す学習データを学習して予測モデルを生成する場合、図 5 に示す学習データに対して回帰分析を利用することにより、説明変量 x_i ($i = 1, 2, \dots, 7$) と、被説明変量 y との間の関係を、数式 (3) に示す第 1 回帰式によってモデル化する。

[0043] [数3]

$$y = \alpha_{L0} + \sum_{i=1}^7 \beta_{L0i} \times x_i \quad \dots(3)$$

[0044] ここで、 x_1 は、第 1 説明変量である「色」の値が「赤」であるか否かを識別する変量であり、 x_2 は、「色」の値が「青」であるか否かを識別する変量であり、 x_3 は、第 1 説明変量である「サイズ」の値が「S」であるか否かを識別する変量であり、 x_4 は、「サイズ」の値が「M」であるか否かを識別する変量である。また、 x_5 は、第 2 説明変量である「ピックアップ数」の値を示す変量であり、 x_6 は、第 2 説明変量である「試着数」の値を示す変量であり、 x_7 は、第 2 説明変量である「販売数」の値を示す変量である。また、 y は、被説明変量である「次期販売数」の値を示す変量である。そして、 α_{L0} 及び β_{L0i} ($i = 1, 2, \dots, 7$) が、説明変量 x_i 及び被説明変量 y から決定されるパラメータである。なお、図 5 に示す例では、第 1 説明変量である「ライン」の値は「雑貨」のみとなっているため、「ライン」は、第 1 回帰式に組み込まれていない。

[0045] そして、評価部 122 は、図 5 に示す学習データから生成した予測モデルを妥当であると評価する場合、図 5 に示す学習データに分割される際に選択部 124 により選択された第 1 説明変量である「ライン」、及び図 5 に示す学習データにおける「ライン」の値である「雑貨」を、予測モデルに対応付けて予測モデル記憶部 112 に登録する。

[0046] また例えば、生成部 120 は、図 7 に示す学習データを学習して予測モデルを生成する場合、図 7 に示す学習データに対して回帰分析を利用することにより、説明変量 x_i ($i = 1, 2, \dots, 5$) と、被説明変量 y との間の関係

を、数式（４）に示す第１回帰式によってモデル化する。

[0047] [数4]

$$y = \alpha_{s_M} + \sum_{i=1}^5 \beta_{s_M i} \times x_i \quad \cdots(4)$$

[0048] ここで、 x_1 は、第１説明変量である「色」の値が「赤」であるか否かを識別する変量であり、 x_2 は、「色」の値が「青」であるか否かを識別する変量である。また、 x_3 は、第２説明変量である「ピックアップ数」の値を示す変量であり、 x_4 は、第２説明変量である「試着数」の値を示す変量であり、 x_5 は、第２説明変量である「販売数」の値を示す変量である。また、 y は、被説明変量である「次期販売数」の値を示す変量である。そして、 α_{s_M} 及び $\beta_{s_M i}$ （ $i = 1, 2, \dots, 5$ ）が、説明変量 x_i 及び被説明変量 y から決定されるパラメータである。なお、図７に示す例では、第１説明変量である「ライン」、「サイズ」の値は、それぞれ「ボトム」、「M」のみとなっているため、「ライン」及び「サイズ」は、第１回帰式に組み込まれていない。

[0049] そして、評価部１２２は、図７に示す学習データから生成した予測モデルを妥当であると評価する場合、図７に示す学習データに分割されるまでに選択部１２４により選択された第１説明変量である「ライン」、「サイズ」、及びその値である「ボトム」、「M」を予測モデルに対応付けて予測モデル記憶部１１２に登録する。

[0050] また例えば、生成部１２０は、図９に示す学習データを学習して予測モデルを生成する場合、図９に示す学習データに対して回帰分析を利用することにより、説明変量 x_i （ $i = 1, 2, 3$ ）と、被説明変量 y との間の関係を、数式（５）に示す第１回帰式によってモデル化する。

[0051] [数5]

$$y = \alpha_{C_Y} + \sum_{i=1}^3 \beta_{C_Y i} \times x_i \quad \cdots(5)$$

[0052] ここで、 x_1 は、第２説明変量である「ピックアップ数」の値を示す変量であり、 x_2 は、第２説明変量である「試着数」の値を示す変量であり、 x_3 は

、第2説明変数である「販売数」の値を示す変数である。また、 y は、被説明変数である「次期販売数」の値を示す変数である。そして、 α_{CY} 及び β_{CYi} ($i = 1, 2, 3$) が、説明変数 x_i 及び被説明変数 y から決定されるパラメータである。なお、図9に示す例では、第1説明変数である「ライン」、「サイズ」、「色」の値は、それぞれ「ボトム」、「S」、「黄」のみとなっているため、「ライン」、「サイズ」、及び「色」は、第1回帰式に組み込まれていない。

[0053] そして、評価部122は、図9に示す学習データから生成した予測モデルを妥当であると評価する場合、図9に示す学習データに分割されるまでに選択部124により選択された第1説明変数である「ライン」、「サイズ」、「色」及びその値である「ボトム」、「S」、「黄」を予測モデルに対応付けて予測モデル記憶部112に登録する。

[0054] なお、生成部120、評価部122、選択部124、及び分割部126については、例えば、CPU (Central Processing Unit) などの既存の制御装置により実現できる。

[0055] 図10は、本実施形態の予測モデル生成装置100で行われる処理の一例を示すフローチャートである。

[0056] ステップS100では、生成部120は、学習データ記憶部110に記憶されている学習データの読み込みを行う。

[0057] ステップS102では、生成部120は、ノードを生成して、読み込んだ学習データに割り当てる。

[0058] ステップS103では、生成部120は、学習データに割り当てたノードをスタックに格納する。

[0059] ステップS104では、生成部120は、スタックに格納されている1のノードを取り出す。但し、スタックにノードが格納されておらず、生成部120がノードを取り出せない場合には（ステップS104でNo）、処理は終了となる。一方、生成部120がノードを取り出した場合には（ステップS104でYes）、ステップS106に進む。

- [0060] ステップS 1 0 6では、生成部 1 2 0は、取り出したノードに割り当てられた学習データを学習して、複数種類の第 1 説明変量及び第 2 説明変量と被説明変量との間の関係を説明する予測モデルを生成する。
- [0061] ステップS 1 0 8では、評価部 1 2 2は、生成部 1 2 0により生成された予測モデルの妥当性を評価し、妥当であると評価した場合には（ステップS 1 0 8でY e s）、ステップS 1 1 0に進み、妥当でないと評価した場合には（ステップS 1 0 8でN o）、ステップS 1 1 2に進む。
- [0062] ステップS 1 1 0では、評価部 1 2 2は、妥当であると評価した予測モデルを予測モデル記憶部 1 1 2に登録し、ステップS 1 0 4に戻る。なお、評価部 1 2 2は、妥当であると評価した予測モデルが分割部 1 2 6により分割された新たな学習データから生成された予測モデルである場合には、新たな学習データに割り当てたノードに到達するまでに選択部 1 2 4により選択された第 1 説明変量及びその値を当該予測モデルに対応付けて予測モデル記憶部 1 1 2に登録する。
- [0063] ステップS 1 1 2では、選択部 1 2 4は、評価部 1 2 2により予測モデルが妥当でないと評価された場合、取り出されたノードに割り当てられた学習データに含まれる第 1 説明変量それぞれの第 1 評価値を算出し、最大の第 1 評価値を有する第 1 説明変量（選択説明変量）を選択する。但し、学習データに選択すべき第 1 説明変量が存在せず、第 1 説明変量を選択できない場合には（ステップS 1 1 2でN o）、ステップS 1 2 4へ進む。一方、選択部 1 2 4が第 1 説明変量を選択した場合には（ステップS 1 1 2でY e s）、ステップS 1 1 4に進む。
- [0064] ステップS 1 1 4では、分割部 1 2 6は、取り出されたノードに割り当てられた学習データを、選択部 1 2 4により選択された第 1 説明変量の値が同一である組の集合に分割する。
- [0065] ステップS 1 1 5では、分割部 1 2 6は、分割した学習データそれぞれを新たな学習データとして学習データ記憶部 1 1 0に登録する。
- [0066] ステップS 1 1 6では、生成部 1 2 0は、学習データ記憶部 1 1 0に記憶

されている新たな学習データの読み込みを行う。但し、学習データ記憶部 110 に新たな学習データが記憶されておらず、生成部 120 が新たな学習データを読み込めない場合には（ステップ S 116 で No）、ステップ S 104 に戻る。一方、生成部 120 が新たな学習データを読み込んだ場合には（ステップ S 116 で Yes）、ステップ S 117 に進む。

[0067] ステップ S 117 では、生成部 120 は、ノードを生成して、読み込んだ新たな学習データに割り当てる。

[0068] ステップ S 118 では、生成部 120 は、読み込んだ新たな学習データに割り当てたノードをスタックに格納するか否かを判定する。具体的には、生成部 120 は、読み込んだ新たな学習データに含まれる組の個数から第 3 評価値を算出し、第 3 評価値が予め指定された第 2 閾値以上となるか否かを判定する。そして、生成部 120 は、第 3 評価値が第 2 閾値以上となる場合にはノードをスタックに格納すると判定して（ステップ S 118 で Yes）、ステップ S 120 に進み、第 3 評価値が第 2 閾値未満となる場合にはノードをスタックに格納しないと判定して（ステップ S 118 で No）、ステップ S 122 に進む。ここで、第 3 評価値としては多様なものが考えられるが、例えば、新たな学習データに含まれる組の個数を、当初の学習データに含まれる組の個数で割った値を用いることができる。

[0069] ステップ S 120 では、生成部 120 は、読み込んだ新たな学習データに割り当てたノードをスタックに格納する。

[0070] ステップ S 122 では、評価部 122 は、読み込んだ新たな学習データに割り当てたノードに、当該新たな学習データから予測モデルが生成できず、予測モデルが無いことを示すフラグを付与し、予測モデル記憶部 112 に登録し、ステップ S 116 に戻る。

[0071] ステップ S 124 では、評価部 122 は、評価部 122 により妥当でないと評価された予測モデルの生成元の学習データに割り当てたノードに、予測モデルが無いことを示すフラグを付与し、予測モデル記憶部 112 に登録し、ステップ S 104 に戻る。

- [0072] そして、本実施形態では、まず、図2に示す学習データから予測モデルが生成され、この予測モデルが妥当でないと評価され、第1説明変量として「ライン」が選択され、図2に示す学習データが図3～図5に示す学習データに分割されるものとする。
- [0073] 続いて、図3に示す学習データから予測モデルが生成され、この予測モデルが妥当であると評価され、予測モデル記憶部112に登録されるものとする。
- [0074] 続いて、図4に示す学習データから予測モデルが生成され、この予測モデルが妥当でないと評価され、第1説明変量として「サイズ」が選択され、図4に示す学習データが図6～図7に示す学習データに分割されるものとする。
- [0075] 続いて、図5に示す学習データから予測モデルが生成され、この予測モデルが妥当であると評価され、予測モデル記憶部112に登録されるものとする。
- [0076] 続いて、図6に示す学習データから予測モデルが生成され、この予測モデルが妥当でないと評価され、第1説明変量として「色」が選択され、図6に示す学習データが図8～図9に示す学習データに分割されるものとする。
- [0077] 続いて、図7に示す学習データから予測モデルが生成され、この予測モデルが妥当であると評価され、予測モデル記憶部112に登録されるものとする。
- [0078] 続いて、図8に示す学習データから予測モデルが生成され、この予測モデルが妥当でないと評価され、第1説明変量が選択されず、予測モデルが予測モデル記憶部112に登録されないものとする。
- [0079] 続いて、図9に示す学習データから予測モデルが生成され、この予測モデルが妥当であると評価され、予測モデル記憶部112に登録されるものとする。
- [0080] この結果、予測モデル記憶部112には、図11に示す予測モデルが生成される。なお、図11に示す例では、各丸印がノードに対応しており、末端

のノード以外には、選択された第 1 説明変量が割り当てられている。また、末端のノードには、第 1 説明変量の値の組み合わせに対応する第 1 回帰式、又は第 1 回帰式が割り当てられていないことを示すフラグ「予測モデル無し」が割り当てられている。また、ノード間を結ぶ枝には、上位のノードにおける第 1 説明変量の取りうる値が割り当てられている。なお、最上位のノードから末端のノードに至るまでに表れる第 1 説明変量及びその値が、末端ノードに割り当てられている回帰式及びフラグに対応する条件を表しているとする。

[0081] 以上のように、本実施形態では、学習データから生成された予測モデルが妥当でないと評価された場合、選択された第 1 説明変量の値が同一である組の集合に学習データが分割され、分割された新たな学習データそれぞれから予測モデルが生成される。従って本実施形態によれば、説明変量に内在する複数の特徴を考慮して、固定的な第 1 説明変量及び時系列的な第 2 説明変量と被説明変量との間の関係を説明する予測モデルを生成することができる。また、従来よりも精緻な予測モデルを生成することができる。

[0082] 例えば図 2 に示すように、「ライン」が「ボトム」である商品は、通常は試着されるため試着回数が増えるが、「ライン」が「トップス」である商品は、通常は試着されないため試着回数が少なくなる傾向にある。

[0083] 従来の技術では、このような「ライン」の値による特徴の違いを考慮することができないため、「ライン」が「ボトム」である学習データと「ライン」が「トップス」である学習データをひとまとめにして、予測モデルを生成していた。このため、「ライン」の値による特徴の違いが平均化されてしまい、妥当な予測モデルを生成することができなかった。

[0084] これに対し、本実施形態では、「ライン」が「ボトム」である学習データ及び「ライン」が「トップス」である学習データから個別に予測モデルを生成できるため、「ライン」の値による特徴の違いを捉えた妥当な予測モデルを生成することができる。

[0085] また本実施形態によれば、固定的な第 1 説明変量及び時系列的な第 2 説明

変量と被説明変量との間の関係を説明する予測モデルを生成できるので、従来に比べ学習データとして取り扱える対象を拡大することができる。

[0086] なお、本実施形態の予測モデル生成装置は、CPUなどの制御装置と、ROMやRAMなどの記憶装置と、HDDやリムーバブルドライブ装置などの外部記憶装置と、ディスプレイなどの表示装置と、キーボードやマウスなどの入力装置を備えており、通常のコンピュータを利用したハードウェア構成となっている。

[0087] (変形例)

本発明は、上記実施形態そのままに限定されるものではなく、実施段階ではその要旨を逸脱しない範囲で構成要素を変形して具体化することができる。また、上記実施形態に開示されている複数の構成要素の適宜な組み合わせにより、種々の発明を形成することができる。例えば、実施形態に示される全構成要素からいくつかの構成要素を削除してもよい。

[0088] 例えば上記実施の形態では、線形な回帰式によって第1回帰式を構成しているが、 n 次式、 $\log$ 関数、又は指数関数などの非線形関数を利用して第1回帰式を構成することもできる。

[0089] また上記実施の形態では、第1説明変量として、「色」、「サイズ」、「ライン」を例に取り説明したが、第1説明変量の種類はこれに限定されるものではなく、また、その個数も2以上の任意の値とすることができる。同様に、第2説明変量として、「ピックアップ数」、「試着数」、「販売数」を例に取り説明したが、第2説明変量の種類及び個数はこれに限定されるものではない。同様に、被説明変量として、「次期販売数」を例に取り説明したが、これに限定されるものではない。

[0090] また、上記実施形態の予測モデル生成装置の機能を、予測モデル生成プログラムを実行することにより実現させるようにしてもよい。

[0091] この場合、上記実施形態の予測モデル生成装置で実行される予測モデル生成プログラムは、インストール可能な形式又は実行可能な形式のファイルでCD-ROM、CD-R、メモリカード、DVD (Digital Ver

s a t i l e D i s k)、フレキシブルディスク (FD) 等のコンピュータで読み取り可能な記憶媒体に記憶されてコンピュータプログラムプロダクトとして提供される。また、上記実施形態の予測モデル生成装置で実行される予測モデル生成プログラムを、ROM等に予め組み込んで提供するようにしてもよい。

[0092] 上記実施形態の予測モデル生成装置で実行される予測モデル生成プログラムは、上述した各部をコンピュータ上で実現させるためのモジュール構成となっている。実際のハードウェアとしては、CPUがROM等から予測モデル生成プログラムをRAM上に読み出して実行することにより、上記各部がコンピュータ上で実現されるようになっている。

産業上の利用可能性

[0093] 以上のように、本発明にかかる予測モデル生成装置は、説明変量と被説明変量との間の関係を説明する予測モデルの生成に適している。

符号の説明

- [0094] 1 0 0 予測モデル生成装置
 1 1 0 学習データ記憶部
 1 1 2 予測モデル記憶部
 1 2 0 生成部
 1 2 2 評価部
 1 2 4 選択部
 1 2 6 分割部

請求の範囲

[請求項1]

複数種類の固定的な第1説明変量と時系列的な第2説明変量と被説明変量とを対応付けた組の集合である学習データを記憶する学習データ記憶部と、

前記学習データを学習して、複数種類の前記第1説明変量及び前記第2説明変量と前記被説明変量との間の関係を説明する予測モデルを生成する生成部と、

前記予測モデルの妥当性を評価し、妥当であると評価した前記予測モデルを予測モデル記憶部に登録する評価部と、

前記予測モデルが妥当でないと評価された場合に、前記第1説明変量それぞれの第1評価値を算出し、前記第1説明変量のうち最大の前記第1評価値を有する選択説明変量を選択する選択部と、

同じ値の前記選択説明変量を有する組の集合に前記学習データを分割し、分割した前記学習データそれぞれを新たな学習データとして前記学習データ記憶部に登録する分割部と、を備え、

前記生成部は、前記新たな学習データ毎に、当該新たな学習データを学習して前記予測モデルを生成することを特徴とする予測モデル生成装置。

[請求項2]

前記生成部は、複数種類の前記第1説明変量及び前記第2説明変量と前記被説明変量との間の関係を説明可能な第1回帰式を設定し、前記第1回帰式に前記学習データを適用した場合における残差が最小となるように前記第1回帰式のパラメータを決定することにより、前記パラメータ決定後の前記第1回帰式を前記予測モデルとして生成することを特徴とする請求項1に記載の予測モデル生成装置。

[請求項3]

前記生成部は、前記第1回帰式に前記学習データを適用することにより自由度調整済みの重回帰係数を第2評価値として算出し、

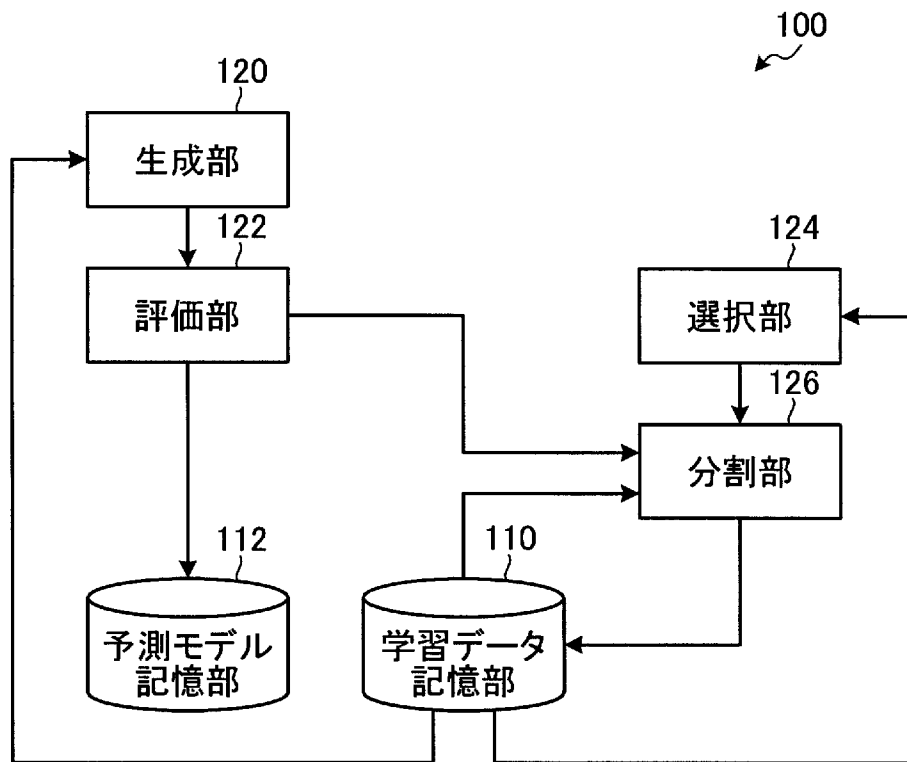
前記評価部は、前記第2評価値が閾値以上である場合に前記予測モデルが妥当であると評価し、前記第2評価値が閾値未満である場合に

前記予測モデルが妥当でないと評価することを特徴とする請求項 2 に記載の予測モデル生成装置。

[請求項 4] 前記選択部は、前記第 1 説明変量と前記被説明変量との間の関係を説明可能な第 2 回帰式を前記第 1 説明変量毎に設定し、前記第 2 回帰式それぞれに前記学習データを適用することにより、自由度調整済みの重回帰係数を前記第 1 評価値として算出することを特徴とする請求項 1 に記載の予測モデル生成装置。

[請求項 5] 前記評価部は、前記新たな学習データから生成された前記予測モデルを妥当であると評価した場合、前記予測モデルと前記選択説明変量及び前記新たな学習データにおける前記選択説明変量の値とを対応付けて前記予測モデル記憶部に登録することを特徴とする請求項 1 に記載の予測モデル生成装置。

[図1]



[図2]

	第1説明変量			第2説明変量			被説明変量
	色	サイズ	ライン	ピックアップ数	試着数	販売数	次期販売数
t1	赤	S	トップス	21	2	6	7
t2	青	S	トップス	17	3	4	3
t3	黄	M	トップス	3	0	1	2
t4	青	S	ボトム	20	10	4	5
t5	青	M	ボトム	32	12	5	4
t6	黄	S	ボトム	25	3	3	3
t7	黄	M	ボトム	6	2	1	2
t8	赤	F	雑貨	15	0	3	4
t9	赤	F	雑貨	12	0	2	1
...	...	...	...	...	...	...	...

[図3]

	第1説明変量			第2説明変量			被説明変量
	色	サイズ	ライン	ピックアップ数	試着数	販売数	次期販売数
t1	赤	S	トップス	21	2	6	7
t2	青	S	トップス	17	3	4	3
t3	黄	M	トップス	3	0	1	2
...	...	...	...	...	...	...	...

[図4]

	第1説明変量			第2説明変量			被説明変量
	色	サイズ	ライン	ピックアップ数	試着数	販売数	次期販売数
t4	青	S	ボトム	20	10	4	5
t5	青	M	ボトム	32	12	5	4
t6	黄	S	ボトム	25	3	3	3
t7	黄	M	ボトム	6	2	1	2
...	...	...	...	...	...	...	...

[図5]

	第1説明変量			第2説明変量			被説明変量
	色	サイズ	ライン	ピックアップ数	試着数	販売数	次期販売数
t8	赤	F	雑貨	15	0	3	4
t9	赤	F	雑貨	12	0	2	1
...	...	...	...	...	...	...	...

[図6]

	第1説明変量			第2説明変量			被説明変量
	色	サイズ	ライン	ピックアップ数	試着数	販売数	次期販売数
t4	青	S	ボトム	20	10	4	5
t6	黄	S	ボトム	25	3	3	3
...	...	...	...	...	...	...	...

[図7]

	第1説明変量			第2説明変量			被説明変量
	色	サイズ	ライン	ピックアップ数	試着数	販売数	次期販売数
t5	青	M	ボトム	32	12	5	4
t7	黄	M	ボトム	6	2	1	2
...	...	...	...	...	...	...	...

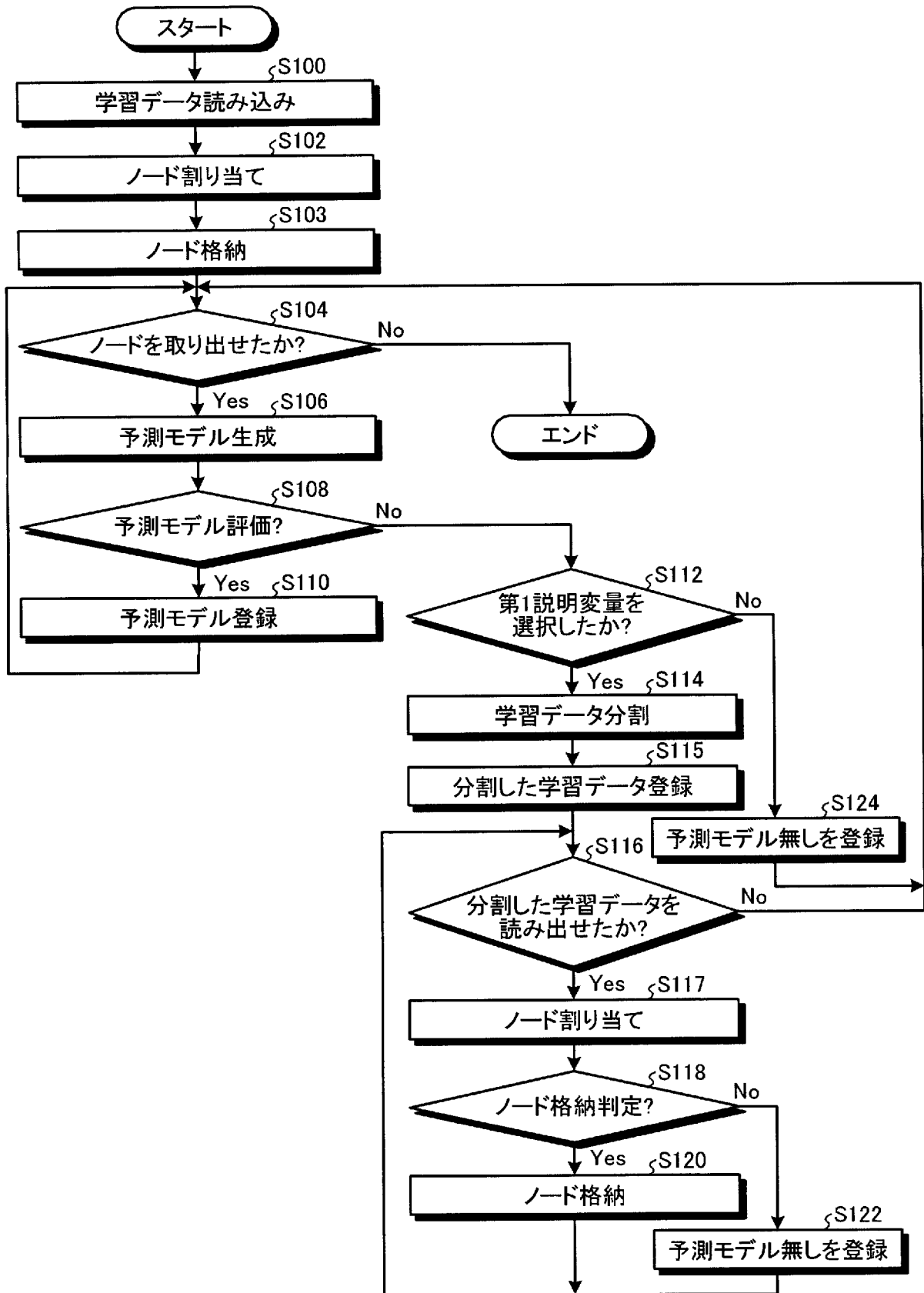
[図8]

	第1説明変量			第2説明変量			被説明変量
	色	サイズ	ライン	ピックアップ数	試着数	販売数	次期販売数
t4	青	S	ボトム	20	10	4	5
...	...	...	...	...	...	...	...

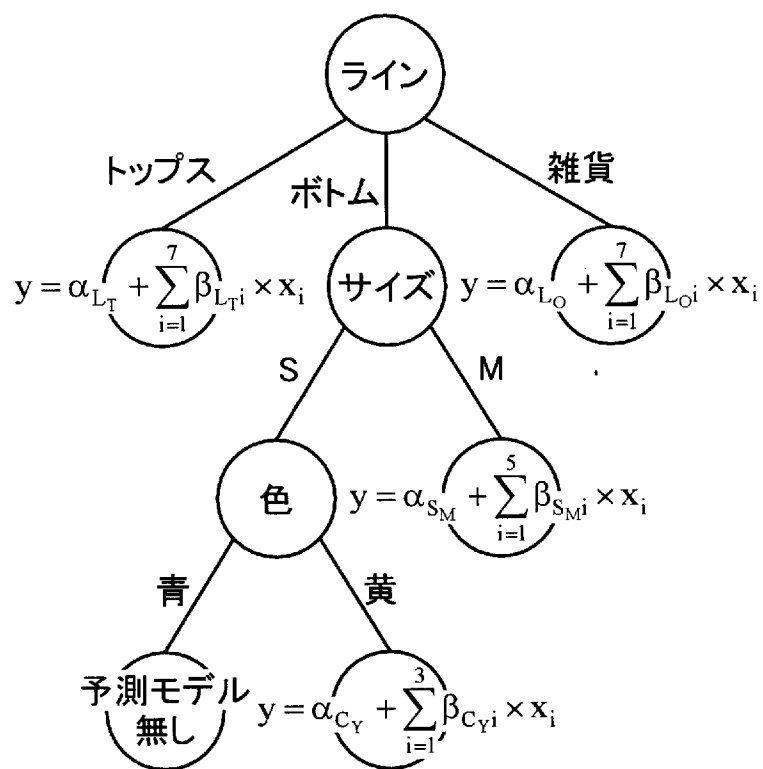
[図9]

	第1説明変量			第2説明変量			被説明変量
	色	サイズ	ライン	ピックアップ数	試着数	販売数	次期販売数
t6	黄	S	ボトム	25	3	3	3
...	...	...	...	...	...	...	...

[図10]



[図11]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2009/070013

A. CLASSIFICATION OF SUBJECT MATTER

G06F17/30 (2006.01) i, G06F19/00 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F17/30, G06F19/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2009
Kokai Jitsuyo Shinan Koho	1971-2009	Toroku Jitsuyo Shinan Koho	1994-2009

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2009-282926 A (Toshiba Corp.), 03 December 2009 (03.12.2009), entire text; all drawings (Family: none)	1-5
A	JP 2005-222445 A (NIFTY Corp. et al.), 18 August 2005 (18.08.2005), entire text; all drawings (Family: none)	1-5
A	JP 2008-117381 A (Omron Corp.), 22 May 2008 (22.05.2008), paragraph [0112] (Family: none)	1-5



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
16 December, 2009 (16.12.09)

Date of mailing of the international search report
28 December, 2009 (28.12.09)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

A. 発明の属する分野の分類（国際特許分類（I P C））
Int.Cl. G06F17/30(2006.01)i, G06F19/00(2006.01)i

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（I P C））

Int.Cl. G06F17/30, G06F19/00

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1 9 2 2 - 1 9 9 6 年
日本国公開実用新案公報	1 9 7 1 - 2 0 0 9 年
日本国実用新案登録公報	1 9 9 6 - 2 0 0 9 年
日本国登録実用新案公報	1 9 9 4 - 2 0 0 9 年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 2009-282926 A（株式会社東芝）2009.12.03, 全文, 全図（ファミリーなし）	1-5
A	JP 2005-222445 A（ニフティ株式会社ほか）2005.08.18, 全文, 全図（ファミリーなし）	1-5
A	JP 2008-117381 A（オムロン株式会社）2008.05.22, 第112段落等（ファミリーなし）	1-5

☐ C欄の続きにも文献が列挙されている。

☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの

「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの

「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）

「O」口頭による開示、使用、展示等に言及する文献

「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「&」同一パテントファミリー文献

国際調査を完了した日

1 6 . 1 2 . 2 0 0 9

国際調査報告の発送日

2 8 . 1 2 . 2 0 0 9

国際調査機関の名称及びあて先

日本国特許庁（I S A / J P）

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官（権限のある職員）

小原 正信

電話番号 03-3581-1101 内線 3562

5 L

3 8 0 3