

(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2012-194894

(P2012-194894A)

(43) 公開日 平成24年10月11日(2012.10.11)

(51) Int.Cl.
G06F 17/30 (2006.01)F I
G06F 17/30 220Z

テーマコード (参考)

審査請求 未請求 請求項の数 9 O L (全 60 頁)

(21) 出願番号 特願2011-59561 (P2011-59561)
(22) 出願日 平成23年3月17日 (2011.3.17)(71) 出願人 000005223
富士通株式会社
神奈川県川崎市中原区上小田中4丁目1番
1号
(74) 代理人 100104190
弁理士 酒井 昭徳
(72) 発明者 松本 和宏
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

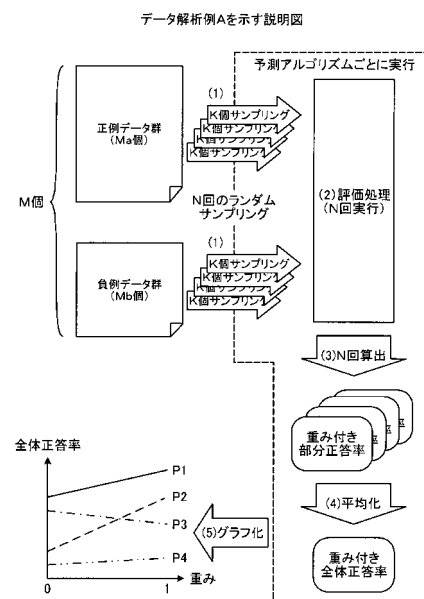
(54) 【発明の名称】 データ解析プログラム、データ解析方法、およびデータ解析装置

(57) 【要約】

【課題】 データ解析の予測精度の向上を図ること。

【解決手段】 データ群は、Ma個の正例データである正例データ群と、Mb個の負例データである負例データ群と、で構成される。(1) 正例データ群からK個の正例データをランダムサンプリングする。負例データ群からK個の負例データをランダムサンプリングする。(2) K個の正例データとK個の負例データとを用いて評価処理を実行する。(3) 評価処理を実行することで、重み付き部分正答率が得られる。(1)～(3)は、N回実行される。(1)～(3)をN回実行することで、重み付き部分正答率がN個算出される。(4) N個の重み付き部分正答率を平均化することで、重み付き全体正答率を算出する。(5) 重み付き全体正答率は予測アルゴリズムごとに算出されるため、予測アルゴリズムごとに、重み付き全体正答率をグラフ化する。

【選択図】 図1



【特許請求の範囲】**【請求項 1】**

少なくとも 1 つ以上の説明変数である説明変数群に対する目的変数が正例である正例実測値の集合と前記説明変数群に対する前記目的変数が負例である負例実測値の集合を有する実測値集合を記憶する記憶装置を有するコンピュータに、

前記実測値集合から、前記正例実測値と前記負例実測値とが同数となるように正例実測値群および負例実測値群をランダムに抽出する抽出工程と、

前記抽出工程によって抽出された正例実測値群および負例実測値群に基づいて、予測アルゴリズムにより前記目的変数を予測する予測式を生成する生成工程と、

を実行させることを特徴とするデータ解析プログラム。

10

【請求項 2】

前記生成工程によって生成された予測式に、前記正例実測値の集合から前記正例実測値群を除いた残余の正例実測値群を各々与えることにより、第 1 の予測値群を算出する第 1 の算出工程と、

前記生成工程によって生成された予測式に、前記負例実測値の集合から前記負例実測値群を除いた残余の負例実測値群を各々与えることにより、第 2 の予測値群を算出する第 2 の算出工程と、

前記第 1 の算出工程によって算出された第 1 の予測値群のうち前記予測式への入力元となる前記正例実測値と一致した第 1 の一致件数と、前記正例実測値と不一致となった第 1 の不一致件数を特定する第 1 の特定工程と、

20

前記第 2 の算出工程によって算出された第 2 の予測値群のうち前記予測式への入力元となる前記負例実測値と一致した第 2 の一致件数と、前記負例実測値と不一致となった第 2 の不一致件数を特定する第 2 の特定工程と、

前記第 1 および第 2 の特定工程によって特定された特定結果に基づいて、前記予測アルゴリズムについて実測値と予測値との一致度を示す正答率が重み変数に応じて変動する重み付き正答率を演算する演算工程と、

前記演算工程によって演算された演算結果を出力する出力工程と、

を前記コンピュータに実行させることを特徴とする請求項 1 に記載のデータ解析プログラム。

30

【請求項 3】

前記抽出工程、前記生成工程、前記第 1 の算出工程、前記第 2 の算出工程、前記第 1 の特定工程、前記第 2 の特定工程、および前記演算工程を、前記コンピュータに複数回実行させ、

前記演算工程によって演算された複数の重み付き正答率を平均化することにより、前記予測アルゴリズムについて実測値と予測値との一致度を示す正答率が重み変数に応じて変動する重み付き平均正答率を出力する平均化工程を前記コンピュータに実行させることを特徴とする請求項 2 に記載のデータ解析プログラム。

【請求項 4】

予測アルゴリズムごとに、前記抽出工程、前記生成工程、前記第 1 の算出工程、前記第 2 の算出工程、前記第 1 の特定工程、前記第 2 の特定工程、および前記演算工程を、前記コンピュータに複数回実行させ、

40

前記演算工程によって前記予測アルゴリズムごとに演算された複数の重み付き正答率を平均化することにより、前記予測アルゴリズムについて実測値と予測値との一致度を示す正答率が重み変数に応じて変動する重み付き平均正答率を出力する平均化工程を、前記予測アルゴリズムごとに前記コンピュータに実行させ、

前記出力工程は、前記予測アルゴリズムごとに出力された重み付き平均正答率をグラフ表示することを特徴とする請求項 2 に記載のデータ解析プログラム。

【請求項 5】

前記説明変数群の中から未選択の説明変数を選択する選択工程を前記コンピュータに実行させ、

50

前記選択工程によって選択された説明変数ごとに、前記抽出工程、前記生成工程、前記第1の算出工程、前記第2の算出工程、前記第1の特定工程、前記第2の特定工程、および前記演算工程を、前記コンピュータに複数回実行させ、

前記演算工程によって前記説明変数ごとに演算された複数の重み付き正答率を平均化することにより、前記予測アルゴリズムについて実測値と予測値との一致度を示す正答率が重み変数に応じて変動する重み付き平均正答率を出力する平均化工程を、前記説明変数ごとに前記コンピュータに実行させ、

前記出力工程は、前記説明変数ごとに出力された重み付き平均正答率をグラフ表示することを特徴とする請求項2に記載のデータ解析プログラム。

【請求項6】

前記説明変数群の中から未選択の説明変数を選択する選択工程を前記コンピュータに実行させ、

前記選択工程によって選択されなかった残余の説明変数群ごとに、前記抽出工程、前記生成工程、前記第1の算出工程、前記第2の算出工程、前記第1の特定工程、前記第2の特定工程、および前記演算工程を、前記コンピュータに複数回実行させ、

前記演算工程によって前記残余の説明変数群ごとに演算された複数の重み付き正答率を平均化することにより、前記予測アルゴリズムについて実測値と予測値との一致度を示す正答率が重み変数に応じて変動する重み付き平均正答率を出力する平均化工程を、前記残余の説明変数群ごとに前記コンピュータに実行させ、

前記出力工程は、前記残余の説明変数群ごとに出力された重み付き平均正答率をグラフ表示することを特徴とする請求項2に記載のデータ解析プログラム。

【請求項7】

前記生成工程によって生成された予測式に、前記正例実測値の集合から前記正例実測値群を除いた残余の正例実測値群を各々与えることにより、第1の予測値群を算出する第1の算出工程と、

前記生成工程によって生成された予測式に、前記負例実測値の集合から前記負例実測値群を除いた残余の負例実測値群を各々与えることにより、第2の予測値群を算出する第2の算出工程と、

前記第1の算出工程によって算出された第1の予測値群のうち前記予測式への入力元となる前記正例実測値と一致した第1の一致件数と、前記正例実測値と不一致となった第1の不一致件数を特定する第1の特定工程と、

前記第2の算出工程によって算出された第2の予測値群のうち前記予測式への入力元となる前記負例実測値と一致した第2の一致件数と、前記負例実測値と不一致となった第2の不一致件数を特定する第2の特定工程と、

前記第1および第2の特定工程によって特定された特定結果に基づいて、前記予測アルゴリズムについて実測値と予測値との不一致度を示す誤答率が重み変数に応じて変動する重み付き誤答率を演算する演算工程と、

前記演算工程によって演算された演算結果を出力する出力工程と、

を前記コンピュータに実行させることを特徴とする請求項1に記載のデータ解析プログラム。

【請求項8】

少なくとも1つ以上の説明変数である説明変数群に対する目的変数が正例である正例実測値の集合と前記説明変数群に対する前記目的変数が負例である負例実測値の集合を有する実測値集合を記憶する記憶装置を有するコンピュータが、

前記実測値集合から、前記正例実測値と前記負例実測値とが同数となるように正例実測値群および負例実測値群をランダムに抽出する抽出工程と、

前記抽出工程によって抽出された正例実測値群および負例実測値群に基づいて、予測アルゴリズムにより前記目的変数を予測する予測式を生成する生成工程と、

を実行することを特徴とするデータ解析方法。

【請求項9】

少なくとも1つ以上の説明変数である説明変数群に対する目的変数が正例である正例実測値の集合と前記説明変数群に対する前記目的変数が負例である負例実測値の集合を有する実測値集合を記憶する記憶装置を有するデータ解析装置であって、

前記実測値集合から、前記正例実測値と前記負例実測値とが同数となるように正例実測値群および負例実測値群をランダムに抽出する抽出手段と、

前記抽出手段によって抽出された正例実測値群および負例実測値群に基づいて、予測アルゴリズムにより前記目的変数を予測する予測式を生成する生成手段と、

を備えることを特徴とするデータ解析装置。

【発明の詳細な説明】

【技術分野】

10

【0001】

本発明は、データを解析するデータ解析プログラム、データ解析方法、およびデータ解析装置に関する。

【背景技術】

【0002】

従来から、入力データについて、回帰式により近似するなどのデータ解析がおこなわれている。たとえば、装置の故障原因についてのデータ解析の場合には、説明変数を装置の故障原因候補とし、目的変数を、装置が正常であるか故障であるかとする。そして、当該データ解析では、説明変数をもとに目的変数の予測値を計算する重回帰式を作成し、重回帰式の係数の大きさをもとに複数の故障の原因の候補について、どの故障原因候補が装置の正常、故障に関係するかを分析する。

20

【先行技術文献】

【特許文献】

【0003】

【特許文献1】特開2004-152205号公報

【特許文献2】特開2004-234302号公報

【特許文献3】特開2007-293889号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

30

解析対象については、正例（または負例）となる実測値が負例（または正例）となる実測値に対して多い。たとえば、上述した装置の故障原因についてのデータ解析の場合は、正常である件数の方が多い。

【0005】

しかしながら、上述した従来技術では、正例となる実測値（正常件数）と負例となる実測値（故障件数）とがほぼ同等の数であることを前提として、回帰式などの予測式を生成している。したがって、正例となる実測値が負例となる実測値に対して圧倒的に多いと予測式の予測精度が低下するという問題がある。負例となる実測値が正例となる実測値に対して圧倒的に多い場合も同様である。

【0006】

40

また、上述した従来技術では、評価対象のデータ全体を用いて予測式を生成して、予測式の生成元である解析対象のデータ全体を予測式で評価していたため、既知である評価対象データで予測値を求めてもそのとおりになる。したがって、予測値が正しいことを過剰に評価してしまうという問題がある。

【0007】

1つの側面では、本発明は、データ解析の予測精度の向上を図ることができるデータ解析プログラム、データ解析方法、およびデータ解析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】

【0008】

本発明の一側面によれば、少なくとも1つ以上の説明変数である説明変数群に対する目

50

的変数が正例である正例実測値の集合と前記説明変数群に対する前記目的変数が負例である負例実測値の集合を有する実測値集合を記憶する記憶装置を有するコンピュータが、前記実測値集合から、前記正例実測値と前記負例実測値とが同数となるように正例実測値群および負例実測値群をランダムに抽出し、抽出された正例実測値群および負例実測値群に基づいて、予測アルゴリズムにより前記目的変数を予測する予測式を生成するデータ解析プログラム、データ解析方法、およびデータ解析装置を用いる。

【発明の効果】

【0009】

本発明の一側面によれば、データ解析の予測精度の向上を図ることができるという効果を奏する。

10

【図面の簡単な説明】

【0010】

【図1】図1は、データ解析例Aを示す説明図である。

【図2】図2は、図1のデータ解析例Aで示した評価処理の詳細を示す説明図である。

【図3】図3は、データ解析例Bを示す説明図である。

【図4】図4は、図3のデータ解析例Bで示した評価処理の詳細を示す説明図である。

【図5】図5は、実施の形態にかかるデータ解析装置のハードウェア構成例を示すブロック図である。

【図6】図6は、回路の故障解析に関するデータ群を示す説明図である。

【図7】図7は、レコメンデーションに関するデータ群を示す説明図である。

20

【図8】図8は、融資の与信に関するデータ群を示す説明図である。

【図9】図9は、実施例1Aにおける予測アルゴリズムごとの重み付き全体正答率を示すグラフである。

【図10】図10は、実施例1Aにかかるデータ解析処理手順を示すフローチャートである。

【図11】図11は、実施例1Aにかかる、図10に示した評価処理（ステップS1003）の詳細な処理手順を示すフローチャートである。

【図12】図12は、実施例1Bにおける予測アルゴリズムごとの重み付き全体誤答率を示すグラフである。

【図13】図13は、実施例1Bにかかる、図10に示した評価処理（ステップS1003）の詳細な処理手順を示すフローチャートである。

30

【図14】図14は、実施例2Aにおける説明変数ごとの重み付き全体正答率を示すグラフである。

【図15】図15は、実施例2Aにかかるデータ解析処理手順を示すフローチャートである。

【図16】図16は、実施例2Bにおける説明変数ごとの重み付き全体誤答率を示すグラフである。

【図17】図17は、実施例3Aにおける残余の説明変数群ごとの重み付き全体正答率を示すグラフである。

【図18】図18は、実施例3Aにかかるデータ解析処理手順を示すフローチャートである。

40

【図19】図19は、実施例3Bにおける説明変数ごとの重み付き全体誤答率を示すグラフである。

【図20】図20は、実施例4Aにおける予測アルゴリズムと説明変数との組み合わせごとの全体正答率を示すグラフである。

【図21】図21は、実施例4Aにかかるデータ解析処理手順を示すフローチャートである。

【図22】図22は、実施例4Bにおける予測アルゴリズムと説明変数との組み合わせごとの全体誤答率を示すグラフである。

【図23】図23は、実施例5Aにおける予測アルゴリズムと残余の説明変数群との組み

50

合わせごとの全体正答率を示すグラフである。

【図 2 4】図 2 4 は、実施例 5 A にかかるデータ解析処理手順を示すフローチャートである。

【図 2 5】図 2 5 は、実施例 5 B における予測アルゴリズムと残余の説明変数群との組み合わせごとの全体誤答率を示すグラフである。

【図 2 6】図 2 6 は、実施例 6 A における予測アルゴリズムごとの全体正答率を示すグラフである。

【図 2 7】図 2 7 は、実施例 6 B における予測アルゴリズムごとの全体誤答率を示すグラフである。

【図 2 8】図 2 8 は、実施例 7 A における説明変数ごとの全体正答率を示すグラフである。

【図 2 9】図 2 9 は、実施例 7 B における説明変数ごとの全体誤答率を示すグラフである。

【図 3 0】図 3 0 は、実施例 8 A における残余の説明変数群ごとの全体正答率を示すグラフである。

【図 3 1】図 3 1 は、実施例 8 B における残余の説明変数群ごとの全体誤答率を示すグラフである。

【図 3 2】図 3 2 は、実施例 9 A における予測アルゴリズムごとの部分正答率のばらつきを示すグラフである。

【図 3 3】図 3 3 は、実施例 9 A にかかるデータ解析処理手順を示すフローチャートである。

【図 3 4】図 3 4 は、実施例 9 A にかかる、図 3 3 に示した評価処理（ステップ S 3 3 0 4）の詳細な処理手順を示すフローチャートである。

【図 3 5】図 3 5 は、実施例 9 B における予測アルゴリズムごとの部分誤答率のばらつきを示すグラフである。

【図 3 6】図 3 6 は、実施例 9 B にかかる評価処理（ステップ S 3 3 0 4）の詳細な処理手順を示すフローチャートである。

【図 3 7】図 3 7 は、実施例 1 0 A における説明変数ごとの部分正答率のばらつきを示すグラフである。

【図 3 8】図 3 8 は、実施例 1 0 A にかかるデータ解析処理手順を示すフローチャートである。

【図 3 9】図 3 9 は、実施例 1 0 B における説明変数ごとの部分誤答率のばらつきを示すグラフである。

【図 4 0】図 4 0 は、実施例 1 1 A における残余の説明変数群ごとの部分正答率のばらつきを示すグラフである。

【図 4 1】図 4 1 は、実施例 1 1 A にかかるデータ解析処理手順を示すフローチャートである。

【図 4 2】図 4 2 は、実施例 1 1 B における残余の説明変数群ごとの部分誤答率のばらつきを示すグラフである。

【図 4 3】図 4 3 は、データ解析装置の機能的構成例を示すブロック図である。

【発明を実施するための形態】

【0 0 1 1】

以下に添付図面を参照して、本発明にかかるデータ解析プログラム、データ解析方法、およびデータ解析装置の実施の形態を詳細に説明する。本実施の形態において、正例データとは、学習したい概念に含まれる事例データ（実測値）であり、負例データとは、学習したい概念に含まれない事例データ（実測値）である。たとえば、同一設計データから得られた回路ごとの故障解析では、解析対象箇所について正常と判断された結果が正例データとなり、故障と判断された結果が負例データとなる。正例データおよび負例データともに、1 以上の説明変数と目的変数を有する。上記の故障解析の例では、たとえば、解析対象箇所の配線幅が説明変数となり、正常／故障が目的変数となる。

10

20

30

40

50

【0012】

<データ解析例 A>

データ解析例 A では、データ解析結果として正答率を求める例である。正答率とは、正例データの実測値と当該正例データを統計学的な予測式に与えた結果得られる予測値との一致度と、負例データの実測値と当該負例データを統計学的な予測式に与えた結果得られる予測値との一致度と、を示す割合である。すなわち、正例を正例と予測し、負例を負例と予測する確率である。

【0013】

図 1 は、データ解析例 A を示す説明図である。図 1 において、データ解析例 A では、M 個のデータであるデータ群を用いる。データ群は、M a 個の正例データである正例データ群と、M b 個の負例データである負例データ群と、で構成される。

10

【0014】

(1) まず、正例データ群から K 個の正例データをランダムサンプリングする。同様に、負例データ群から K 個の負例データをランダムサンプリングする。(2) つぎに、データ解析例 A では、K 個の正例データと K 個の負例データとを用いて評価処理を実行する。評価処理は、ある予測アルゴリズムにしたがって実行される。予測アルゴリズムとは、統計学的な予測式による予測値の算出プログラムである。予測アルゴリズムとしては、たとえば、回帰分析、ステップワイズ回帰分析、SVM (Support Vector Machine) 回帰分析、回帰木分析、ランダムフォレスト回帰分析などの周知の近似アルゴリズムが適用可能である。評価処理の詳細は図 2 で説明する。

20

【0015】

(3) 評価処理を実行することで、重み付き部分正答率が得られる。重み付き部分正答率とは、正例データ群および負例データ群からの 1 回のランダムサンプリングで評価処理されたときの正答率であり、重み変数 w ($0 \leq w \leq 1$) により変動する正答率である。

【0016】

上記 (1) ~ (3) は、N 回実行される。N は、ユーザにより設定された回数である。上記 (1) ~ (3) を N 回実行することで、重み付き部分正答率が N 個算出される。(4) データ解析例 A では、N 個の重み付き部分正答率を平均化することで、重み付き全体正答率を算出する。図 1 中、点線で囲まれた内容は、予測アルゴリズムごとに実行される。したがって、重み付き全体正答率も、予測アルゴリズムごとに算出される。

30

【0017】

(5) 重み付き全体正答率は予測アルゴリズムごとに算出されるため、データ解析例 A では、予測アルゴリズムごとに、重み付き全体正答率をグラフ化する。これにより、利用者は、どの予測アルゴリズムの予測精度がよいか直感的に把握することができる。

【0018】

図 2 は、図 1 のデータ解析例 A で示した評価処理の詳細を示す説明図である。図 2 では、図 1 の (1) ~ (3) を 1 回実行した例を示している。図 2 において、正例データ群から K 個サンプリングされた正例データ群を根拠用正例データ群と称し、残余を評価用正例データ群と称す。また、負例データ群から K 個サンプリングされた負例データ群を根拠用負例データ群と称し、残余を評価用負例データ群と称す。

40

【0019】

(21) 評価処理では、根拠用正例データ群と根拠用負例データ群とを組み合わせた根拠用データ群を用いて予測式を生成する。予測式とは、予測アルゴリズムを表現する関数であり、予測アルゴリズムごとに生成される。予測式としては、たとえば、回帰式や SVM (サポートベクターマシン)、決定木など既存の予測式が適用可能である。単回帰式 $y = \alpha x + \beta$ を例にとると、根拠用データ群の各々の説明変数の値と目的変数の値を、単回帰式の説明変数 x および目的変数 y に与え、最小二乗法により単回帰式の係数 α 、 β の値をそれぞれ求める。

【0020】

(22) つぎに、評価用データ群の各々の説明変数の値 (実測値) を、(21) で生成

50

された予測式の説明変数に与えることで、評価用データ群の各々について目的変数の予測値を算出する。ここで、評価用データ群の目的変数の実測値は、上記の故障解析の例では、「正常」の場合が“0”、「故障」の場合が“1”とする。目的変数の予測値は、0から1の範囲内となるため、正規化することとなる。たとえば、目的変数の予測値の小数点第1位の数値で四捨五入することで、目的変数の予測値が“0”または“1”となる。

【0021】

ここで、評価用正例データ群の各々について説明変数の実測値を予測式に与えることで得られた目的変数の実測値群を、評価用正例データの予測値群と称す。同様に、評価用負例データ群の各々について説明変数の実測値を予測式に与えることで得られた目的変数の実測値群を、評価用負例データの予測値群と称す。

10

【0022】

(23) このあと、評価用データの実測値群と評価用データの予測値群とで突き合わせをおこなう。図2の表は、評価用データについての実測値と予測値との一致・不一致を示している。ここで、ある正例の評価用データについて正例の実測値が予測値と一致する場合は、当該正例の評価用データは、正例・正例件数にカウントされる。すなわち、正例・正例件数とは、正例の実測値を正しく予測できた評価用データの件数である。たとえば、上述した故障解析では、正常であると実測されたデータを正常と予測した件数となる。

【0023】

同様に、ある負例の評価用データについて負例の実測値が予測値と一致する場合は、当該負例の評価用データは、負例・負例件数にカウントされる。すなわち、負例・負例件数とは、負例の実測値を正しく予測できた評価用データの件数である。たとえば、上述した故障解析では、故障であると実測されたデータを故障と予測した件数となる。

20

【0024】

これに対し、ある正例の評価用データについて正例の実測値が予測値と不一致である場合は、当該正例の評価用データは、正例・負例件数にカウントされる。すなわち、正例・負例件数とは、正例の実測値を誤って予測した評価用データの件数である。たとえば、上述した故障解析では、正常であると実測されたデータを故障と予測した件数となる。

【0025】

同様に、ある負例の評価用データについて負例の実測値が予測値と不一致である場合は、当該負例の評価用データは、負例・正例件数にカウントされる。すなわち、負例・正例件数とは、負例の実測値を誤って予測した評価用データの件数である。たとえば、上述した故障解析では、故障であると実測されたデータを正常と予測した件数となる。評価処理では、正例・正例件数、負例・負例件数、正例・負例件数、負例・正例件数をそれぞれカウントする。

30

【0026】

(24) そして、評価処理では、カウントされた各件数に基づいて、下記式(1)に示す重み付き部分正答率を算出する。

【0027】

$$PCA_w = w \times N_{pp} / (N_{pp} + N_{pn}) + (1 - w) \times N_{nn} / (N_{nn} + N_{np}) \quad \dots (1)$$

40

ただし、 PCA_w は重み付き部分正答率、 w は重み変数、 N_{pp} は正例・正例件数、 N_{pn} は正例・負例件数、 N_{nn} は負例・負例件数、 N_{np} は負例・正例件数である。重み変数 w は、 $0 \leq w \leq 1$ である。すなわち、図2においてハッチングを施してある正例・正例件数が右辺第1項の分子となり、図2においてハッチングを施してある負例・負例件数が右辺第2項の分子となる。

【0028】

評価処理では、図1の(3)に示したように、重み付き部分正答率 PCA_w を、図1の(1)のランダムサンプリングの都度算出する。算出されたN個の重み付き部分正答率 PCA_w は平均化されることで、重み付き全体正答率となる。

【0029】

50

このように、図 1 および図 2 に示した処理をおこなうことで、予測アルゴリズムごとに重み付き全体正答率を得ることができ、予測アルゴリズムごとの重み付き全体正答率の違いを、直感的にわかりやすく表示することができる。なお、予測アルゴリズムごとの重み付き全体正答率の表示例については後述する。

【0030】

また、重み変数 w は利用者が自由に設定することができる。たとえば、正例データ群の正例データ数 M_a と負例データ群の負例データ数 M_b において、 $M_a > M_b$ の場合は、負例・負例件数を強調するため重み変数 w を低く設定すればよい。 $M_a < M_b$ の場合は、正例・正例件数を強調するため重み変数 w を高く設定すればよい。

【0031】

<データ解析例 B>

つぎに、データ解析例 B について説明する。データ解析例 B は、データ解析結果として誤答率を求める例である。誤答率とは、正例データの実測値と当該正例データを統計学的な予測式に与えた結果得られる予測値との不一致度と、負例データの実測値と当該負例データを統計学的な予測式に与えた結果得られる予測値との不一致度と、を示す割合である。すなわち、正例を負例と予測し、負例を正例と予測する確率である。

【0032】

図 3 は、データ解析例 B を示す説明図である。図 3 では、図 1 と同一処理内容については同一番号を付して説明を省略し、相違点について説明する。具体的には、データ解析例 A では重み付き部分正答率と重み付き全体正答率を求めるのに対し、データ解析例 B では重み付き部分誤答率と重み付き全体誤答率を求める点が相違する。

【0033】

図 4 は、図 3 のデータ解析例 B で示した評価処理の詳細を示す説明図である。図 4 でも、図 2 と同一処理内容については同一番号を付して説明を省略し、相違点について説明する。図 4 において、評価処理では、カウントされた各件数に基づいて、下記式 (2) に示す重み付き部分誤答率を算出する。

【0034】

$$PWA_w = w \times N_{pn} / (N_{pp} + N_{pn}) + (1 - w) \times N_{np} / (N_{nn} + N_{np}) \quad \dots (2)$$

【0035】

ただし、 PWA_w は重み付き部分誤答率である。図 4 では、(23) 突き合わせにおいて、図 4 においてハッチングを施してある正例・負例件数が右辺第 1 項の分子となり、図 4 においてハッチングを施してある負例・正例件数が右辺第 2 項の分子となる。評価処理では、図 3 の (3) に示したように、重み付き部分誤答率 PWA_w を、図 3 の (1) のランダムサンプリングの都度算出する。算出された N 個の重み付き部分誤答率 PWA_w は平均化されることで、重み付き全体誤答率となる。

【0036】

このように、図 3 および図 4 に示した処理をおこなうことで、予測アルゴリズムごとに重み付き全体誤答率を得ることができ、予測アルゴリズムごとの重み付き全体誤答率の違いを、直感的にわかりやすく表示することができる。なお、予測アルゴリズムごとの重み付き全体誤答率の表示例については後述する。

【0037】

また、重み変数 w は利用者が自由に設定することができる。たとえば、正例データ群の正例データ数 M_a と負例データ群の負例データ数 M_b において、 $M_a > M_b$ の場合は、負例・負例件数を強調するため重み変数 w を低く設定すればよい。 $M_a < M_b$ の場合は、正例・正例件数を強調するため重み変数 w を高く設定すればよい。

【0038】

<データ解析装置のハードウェア構成例>

図 5 は、実施の形態にかかるデータ解析装置のハードウェア構成例を示すブロック図である。図 5 において、データ解析装置は、CPU (Central Processing

10

20

30

40

50

g Unit) 501と、ROM (Read Only Memory) 502と、RAM (Random Access Memory) 503と、磁気ディスクドライブ504と、磁気ディスク505と、光ディスクドライブ506と、光ディスク507と、ディスプレイ508と、I/F (Interface) 509と、キーボード510と、マウス511と、スキャナ512と、プリンタ513と、を備えている。また、各構成部はバス500によってそれぞれ接続されている。

【0039】

ここで、CPU 501は、データ解析装置の全体の制御を司る。ROM 502は、ブートプログラムなどのプログラムを記憶している。RAM 503は、CPU 501のワークエリアとして使用される。磁気ディスクドライブ504は、CPU 501の制御にしたがって磁気ディスク505に対するデータのリード／ライトを制御する。磁気ディスク505は、磁気ディスクドライブ504の制御で書き込まれたデータを記憶する。

【0040】

光ディスクドライブ506は、CPU 501の制御にしたがって光ディスク507に対するデータのリード／ライトを制御する。光ディスク507は、光ディスクドライブ506の制御で書き込まれたデータを記憶したり、光ディスク507に記憶されたデータをコンピュータに読み取らせたりする。

【0041】

ディスプレイ508は、カーソル、アイコンあるいはツールボックスをはじめ、文書、画像、機能情報などのデータを表示する。このディスプレイ508は、たとえば、CRT、TFT液晶ディスプレイ、プラズマディスプレイなどを採用することができる。

【0042】

インターフェース（以下、「I/F」と略する。）509は、通信回線を通じてLAN (Local Area Network)、WAN (Wide Area Network)、インターネットなどのネットワーク514に接続され、このネットワーク514を介して他の装置に接続される。そして、I/F 509は、ネットワーク514と内部のインターフェースを司り、外部装置からのデータの入出力を制御する。I/F 509には、たとえばモデムやLANアダプタなどを採用することができる。

【0043】

キーボード510は、文字、数字、各種指示などの入力のためのキーを備え、データの入力をおこなう。また、タッチパネル式の入力パッドやテンキーなどであってもよい。マウス511は、カーソルの移動や範囲選択、あるいはウィンドウの移動やサイズの変更などをおこなう。ポインティングデバイスとして同様に機能を備えるものであれば、トラックボールやジョイスティックなどであってもよい。

【0044】

スキャナ512は、画像を光学的に読み取り、データ解析装置内に画像データを取り込む。なお、スキャナ512は、OCR (Optical Character Reader) 機能を持たせてもよい。また、プリンタ513は、画像データや文書データを印刷する。プリンタ513には、たとえば、レーザプリンタやインクジェットプリンタを採用することができる。

【0045】

<データ群のサンプル例>

つぎに、図1や図3に示したデータ群（正例データ群および負例データ群）のサンプル例について説明する。本実施の形態では、回路の故障解析、レコメンデーションに関するデータ解析、融資の与信に関するデータ解析を例に挙げて説明する。以下に示す図6～図8のデータ群のサンプル例は、具体的には、たとえば、図5に示したROM 502、RAM 503、磁気ディスク505などの記憶装置に記憶されている。

【0046】

図6は、回路の故障解析に関するデータ群を示す説明図である。図6では、データごとに、サンプルID、間隔A～D、故障フラグを有する。サンプルIDとは、データを一意

10

20

30

40

50

に特定する識別子であり、この場合は、同一設計データにより作製されたM個の回路の各々を一意に特定される回路である。間隔A～Dとは、回路内の特定の箇所での配線間隔である。図6では、例として間隔A～Dの4箇所とする。この間隔A～Dが説明変数群となる。図6では、4個の説明変数となるが、1以上であればよい。図6では、間隔A～Dの単位は $[\mu\text{m}]$ とする。

【0047】

故障フラグとは、データごとに正常または故障を示すフラグである。“0”が「正常」（正例）、“1”が「故障」（不例）とする。故障フラグは、説明変数群に対する目的変数となる。間隔Aが正常となる範囲は、たとえば、 $8 \leq A \leq 12$ とする。間隔Bが正常となる範囲は、たとえば、 $18 \leq B \leq 22$ とする。間隔Cが正常となる範囲は、たとえば、 $28 \leq C \leq 32$ とする。間隔Dが正常となる範囲は、たとえば、 $38 \leq D \leq 42$ とする。図6では例として、データごとに、間隔A～Dのいずれか1つでも正常の範囲外となると、故障とする。なお、以降では、一例として、間隔Aを説明変数E1、間隔Bを説明変数E2、間隔Cを説明変数E3、間隔Dを説明変数E4とする。

【0048】

図7は、レコメンデーションに関するデータ群を示す説明図である。レコメンデーションに関するデータ群とは、ある店舗において顧客に対して推薦した商品を、顧客が購入したか否かを示すデータである。図7では、データごとに、サンプルID、年齢、性別、購入回数、平均購入金額、推薦商品の購入有無を有する。サンプルIDとは、データを一意に特定する識別子であり、この場合は、M人の顧客の各々を一意に特定する識別子である。年齢とは、顧客の年齢である。性別とは顧客の性別であり、ここでは、“0”を女性、“1”を男性とする。

【0049】

購入回数とは、顧客がこれまで店舗で商品を購入した回数である。平均購入金額とは、顧客が1回の購入機会での購入金額であり、これまでの総購入金額を購入回数で割った金額である。年齢～平均購入金額が説明変数群となる。説明変数群は1以上であればよい。

【0050】

推薦商品の購入有無とは、店舗での推薦商品を顧客が購入したか否かを示すフラグである。“1”が「購入」（正例）、“0”が「購入せず」（不例）を示す。1度でも推薦商品を購入すれば“1”となる。推薦商品の購入有無は、説明変数群に対する目的変数となる。なお、以降では、一例として、年齢を説明変数E1、性別を説明変数E2、購入回数を説明変数E3、平均購入金額を説明変数E4とする。

【0051】

図8は、融資の与信に関するデータ群を示す説明図である。融資の与信に関するデータ群とは、融資をした場合の返済の可否を示すデータである。図8では、データごとに、サンプルID、年齢、性別、年収、世帯年収、返済の可否を有する。サンプルIDとは、データを一意に特定する識別子であり、この場合は、M人の顧客の各々を一意に特定する識別子である。年齢とは、顧客の年齢である。性別とは顧客の性別であり、ここでは、“0”を女性、“1”を男性とする。

【0052】

年収とは、顧客の年間の収入である。世帯年収とは、顧客が属する世帯の年間の収入である。年収および世帯年収は、融資の申込時の収入とする。年齢～世帯年収は、説明変数群となる。説明変数群は1以上であればよい。

【0053】

返済の可否とは、融資を返済できたか否かを示すフラグである。“1”が「返済できた」こと（正例）を示し、“0”が「返済できなかった」こと（不例）を示す。返済の可否は、説明変数群に対する目的変数となる。なお、以降では、一例として、年齢を説明変数E1、性別を説明変数E2、年収を説明変数E3、世帯年収を説明変数E4とする。

【0054】

（実施例1A）

10

20

30

40

50

まず、実施例 1 A について図 9～図 11 を用いて説明する。実施例 1 A は、予測アルゴリズムごとに重み付き全体正答率を算出してグラフ表示する例である。これにより、どの予測アルゴリズムの予測精度（正答率）がよいかを利用者にわかりやすく提示することができる。

【0055】

図 9 は、実施例 1 A における予測アルゴリズムごとの重み付き全体正答率を示すグラフである。図 9 において、横軸は重み変数 w 、縦軸は全体正答率を示す。グラフは図 5 に示したディスプレイに表示される。図 9 では、4 種類の予測アルゴリズム（予測式）P1～P4 を用いている。

【0056】

図 9 では、重み変数 w がどのような値をとっても、予測アルゴリズム P1～P4 の中で予測アルゴリズム P1 の正答率が最も高い。したがって、予測アルゴリズム P1 を採用すると最も予測精度がよくなる。また、予測アルゴリズム P2, P3 については、重み変数 w が $0 \leq w < w_x$ では予測アルゴリズム P3 のほうが全体正答率がよく、 $w_x < w \leq 1$ では予測アルゴリズム P2 のほうが全体正答率がよい。

【0057】

したがって、たとえば、予測アルゴリズム P1 について重み付き全体正答率を求めているときは、利用者は、負例・負例件数よりも正例・正例件数を重視する場合は、重み変数 w の値が大きいときに正答率がよくなる予測アルゴリズム P2 を利用すればよい。一方、正例・正例件数よりも負例・負例件数を重視する場合は、重み変数 w の値が小さいときに正答率（予測精度）がよくなる予測アルゴリズム P3 を採用すればよい。

【0058】

なお、予測アルゴリズム P4 は、予測アルゴリズム P1～P4 の中で最も正答率が低い予測アルゴリズムであるため、利用者は予測アルゴリズム P4 を採用しない方がよいことがわかる。

【0059】

図 10 は、実施例 1 A にかかるデータ解析処理手順を示すフローチャートである。図 10 において、まず、データ解析装置は、未選択の予測アルゴリズムがあるか否かを判断する（ステップ S1001）。未選択の予測アルゴリズムがある場合（ステップ S1001：Yes）、データ解析装置は、未選択の予測アルゴリズムを選択し（ステップ S1002）、評価処理を実行する（ステップ S1003）。評価処理（ステップ S1003）の詳細については図 11 で説明する。

【0060】

評価処理（ステップ S1003）のあと、ステップ S1001 に戻り、未選択の予測アルゴリズムがない場合（ステップ S1001：No）、データ解析装置は、図 9 に示したように、予測アルゴリズムごとに得られた重み付き全体正答率のグラフ表示処理を実行する（ステップ S1004）。これにより、実施例 1 A でのデータ解析処理を終了する。

【0061】

図 11 は、実施例 1 A にかかる、図 10 に示した評価処理（ステップ S1003）の詳細な処理手順を示すフローチャートである。図 11 において、データ解析装置は、正例データ群および負例データ群のランダムサンプリング回数 i を $i=1$ とする（ステップ S1101）。なお、上限は N とする。すなわち、 N 回のランダムサンプリングをおこなうこととなる。

【0062】

つぎに、データ解析装置は、 $i > N$ であるか否かを判断する（ステップ S1102）。 $i > N$ でない場合（ステップ S1102：No）、図 1 および図 2 の（1）に示したように、正例データ群から K 個の正例データをランダムサンプリングし、負例データ群から K 個の負例データをランダムサンプリングする（ステップ S1103）。

【0063】

そして、データ解析装置は、図 2 の（21）に示したように、ランダムサンプリングさ

10

20

30

40

50

れた根拠用データ群を用いて予測アルゴリズムに基づく予測式を生成する（ステップS 1 1 0 4）。このあと、データ解析装置は、図2の（2 2）に示したように、生成された予測式に評価用データ（実測値）を与えることにより、評価用データごとに予測値を算出する（ステップS 1 1 0 5）。そして、データ解析装置は、算出された各予測値の正規化をおこなう（ステップS 1 1 0 6）。

【0064】

このあと、データ解析装置は、図2の（2 3）に示したように、正例・正例件数、正例・負例件数、負例・正例件数、および負例・負例件数を特定する（ステップS 1 1 0 7）。そして、データ解析装置は、図2の（2 4）に示したように、選択中の予測アルゴリズムでの重み付き部分正答率を算出する（ステップS 1 1 0 8）。このあと、 i をインクリメントし（ステップS 1 1 0 9）、ステップS 1 1 0 2に移行する。データ解析装置は、ステップS 1 1 0 3～S 1 1 0 9を $i > N$ となるまで繰り返し実行する。これにより、 N 個の重み付き部分正答率が算出される。

【0065】

そして、 $i > N$ となった場合（ステップS 1 1 0 2：Yes）、データ解析装置は、 N 個の重み付き部分正答率を平均化して重み付き全体正答率を算出する（ステップS 1 1 1 0）。これにより、選択中の予測アルゴリズムについて重み付き全体正答率が算出される。この評価処理（ステップS 1 0 0 3）を予測アルゴリズムごとに実行することで、予測アルゴリズムごとの重み付き全体正答率が算出される。したがって、図9に示したようなグラフ表示をおこなうことができる。

【0066】

このように、実施例1Aによれば、利用者の正例データについての関心度の高さに応じて、利用者が重み変数 w を設定することができる。したがって、実施例1Aでは、利用者が設定した重み変数 w の値での予測アルゴリズムの予測精度（正答率）を、利用者にわかりやすく提示することができる。

【0067】

たとえば、回路の故障解析に適用する場合、『どの予測アルゴリズムを採用すれば、正常らしい回路を正常であると予測してくれるか、また、どの予測アルゴリズムを採用すれば、故障らしい回路を故障であると予測してくれるか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。

【0068】

また、『正常らしい回路を正常と予測する』と『故障らしい回路を故障と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（故障）データ数 M_b よりも正例（正常）データ数 M_a が多い場合、『故障らしい回路を故障と予測する』ことを重視するほうがよい。この場合は、利用者は重み変数 w を低く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムごとの全体正答率を比較すればよい。

【0069】

これに対し、正例（正常）データ数 M_a よりも負例（故障）データ数 M_b が多い場合、『正常らしい回路を正常と予測する』ことを重視するほうがよい。この場合は、利用者は重み変数 w を高く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムごとの全体正答率を比較すればよい。

【0070】

また、推薦商品のレコメンデーションに関するデータ解析に適用する場合、『どの予測アルゴリズムを採用すれば、推薦商品を購入しそうな顧客を「購入する」と予測してくれるか、また、どの予測アルゴリズムを採用すれば、推薦商品を購入しそうな顧客を「購入しない」と予測してくれるか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。また、『推薦商品を購入しそうな顧客を「購入する」と予測する』と『推薦商品を購入しそうな顧客を「購入しない」と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。

【0071】

したがって、たとえば、負例（購入しなかった）データ数M_bよりも正例（購入した）データ数M_aが多い場合、『推薦商品を購入しそうにない顧客を「購入しない」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数wを低く設定することで、利用者の設定した重み変数wの値での予測アルゴリズムごとの全体正答率を比較すればよい。

【0072】

これに対し、正例（購入した）データ数M_aよりも負例（購入しなかった）データ数M_bが多い場合、『推薦商品を購入しそうな顧客を「購入する」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数wを高く設定することで、利用者の設定した重み変数wの値での予測アルゴリズムごとの全体正答率を比較すればよい。

10

【0073】

また、融資の与信に関するデータ解析に適用する場合、『どの予測アルゴリズムを採用すれば、返済できそうな顧客を「返済できる」と予測してくれるか、また、どの予測アルゴリズムを採用すれば、返済できそうにない顧客を「返済できない」と予測してくれるか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。また、『返済できそうな顧客を「返済できる」と予測する』と『返済できそうにない顧客を「返済できない」と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。

【0074】

20

したがって、たとえば、負例（返済できなかった）データ数M_bよりも正例（返済できた）データ数M_aが多い場合、『返済できそうにない顧客を「返済できない」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数wを低く設定することで、利用者の設定した重み変数wの値での予測アルゴリズムごとの全体正答率を比較すればよい。

【0075】

これに対し、正例（返済できた）データ数M_aよりも負例（返済できなかった）データ数M_bが多い場合、『返済できそうな顧客を「返済できる」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数wを高く設定することで、利用者の設定した重み変数wの値での予測アルゴリズムごとの全体正答率を比較すればよい。

30

【0076】

このように、実施例1Aによれば、利用者の関心の程度に応じた全体正答率で予測アルゴリズムを比較することで、利用者にとって予測精度のよい予測アルゴリズムをわかりやすく提示することができる。

【0077】

（実施例1B）

まず、実施例1Bについて図12および図13を用いて説明する。実施例1Bは、予測アルゴリズムごとに重み付き全体誤答率を算出してグラフ表示する例である。これにより、どの予測アルゴリズムの予測精度（誤答率）がよいかを利用者にわかりやすく提示することができる。

40

【0078】

図12は、実施例1Bにおける予測アルゴリズムごとの重み付き全体誤答率を示すグラフである。図12において、横軸は重み変数w、縦軸は全体誤答率を示す。グラフは図5に示したディスプレイ508に表示される。図12では、4種類の予測アルゴリズム（予測式）P1～P4を用いている。

【0079】

図12では、重み変数wがどのような値をとっても、予測アルゴリズムP1～P4の中で予測アルゴリズムP4の誤答率が最も低い。すなわち、最も予測精度がよい。また、予測アルゴリズムP2、P3については、重み変数wが $0 \leq w < w_x$ では予測アルゴリズムP2のほうが全体誤答率がよく、 $w_x < w \leq 1$ では予測アルゴリズムP3のほうが全体誤

50

答率がよい。

【0080】

したがって、たとえば、予測アルゴリズム P 4 について重み付き全体誤答率を求めているときは、負例・正例件数よりも正例・負例件数を重視する場合は、重み変数 w の値が大きいときに誤答率がよくなる予測アルゴリズム P 3 を利用すればよい。正例・負例件数よりも負例・正例件数を重視する場合は、重み変数 w の値が小さいときに誤答率がよくなる予測アルゴリズム P 2 を利用すればよい。

【0081】

なお、予測アルゴリズム P 1 は、予測アルゴリズム P 1 ～ P 4 の中で最も誤答率が高い予測アルゴリズムであるため、利用者は予測アルゴリズムを利用しない方がよいことがわかる。

10

【0082】

つぎに、実施例 1 B にかかるデータ解析処理手順について説明する。なお、図 10 に示した処理手順については実施例 1 A と同一であるため省略する。実施例 1 B では、実施例 1 A の評価処理（ステップ S 1003）の処理内容が異なるため、評価処理（S 1003）のかわりに、評価処理（ステップ S 1303）を実行する。

【0083】

図 13 は、実施例 1 B にかかる、図 10 に示した評価処理（ステップ S 1003）の詳細な処理手順を示すフローチャートである。図 13 において、データ解析装置は、正例データ群および負例データ群のランダムサンプリング回数 i を $i = 1$ とする（ステップ S 1301）。なお、上限は N とする。すなわち、 N 回のランダムサンプリングをおこなうこととなる。

20

【0084】

つぎに、データ解析装置は、 $i > N$ であるか否かを判断する（ステップ S 1302）。 $i > N$ でない場合（ステップ S 1302：No）、図 3 および図 4 の（1）に示したように、正例データ群から K 個の正例データをランダムサンプリングし、負例データ群から K 個の負例データをランダムサンプリングする（ステップ S 1303）。

【0085】

そして、データ解析装置は、図 4 の（21）に示したように、ランダムサンプリングされた根拠用データ群を用いて予測アルゴリズムに基づく予測式を生成する（ステップ S 1304）。このあと、データ解析装置は、図 4 の（22）に示したように、生成された予測式に評価用データ（実測値）を与えることにより、評価用データごとに予測値を算出する（ステップ S 1305）。そして、データ解析装置は、算出された各予測値の正規化をおこなう（ステップ S 1306）。

30

【0086】

このあと、データ解析装置は、図 4 の（23）に示したように、正例・正例件数、正例・負例件数、負例・正例件数、および負例・負例件数を特定する（ステップ S 1307）。そして、データ解析装置は、図 4 の（24）に示したように、選択中の予測アルゴリズムでの重み付き部分誤答率を算出する（ステップ S 1308）。このあと、 i をインクリメントし（ステップ S 1309）、ステップ S 1302 に移行する。データ解析装置は、ステップ S 1303 ～ S 1309 を $i > N$ となるまで繰り返し実行する。これにより、 N 個の重み付き部分誤答率が算出される。

40

【0087】

そして、 $i > N$ となった場合（ステップ S 1302：Yes）、データ解析装置は、 N 個の重み付き部分誤答率を平均化して重み付き全体誤答率を算出する（ステップ S 1310）。これにより、選択中の予測アルゴリズムについて重み付き全体誤答率が算出される。この評価処理（ステップ S 1003）を予測アルゴリズムごとに実行することで、予測アルゴリズムごとの重み付き全体誤答率が算出される。したがって、図 12 に示したようなグラフ表示をおこなうことができる。

【0088】

50

このように、実施例 1 B によれば、利用者の負例データについての関心度の高さに応じて、利用者が重み変数 w を設定することができる。したがって、実施例 1 B では、利用者が設定した重み変数 w の値での予測アルゴリズムの予測精度（誤答率）を、利用者にわかりやすく提示することができる。

【0089】

たとえば、回路の故障解析に適用する場合、『どの予測アルゴリズムを採用すれば、正常らしい回路を故障であると予測してしまうか、また、どの予測アルゴリズムを採用すれば、故障らしい回路を正常であると予測してしまうか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。

【0090】

また、『正常らしい回路を故障と予測する』と『故障らしい回路を正常と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（故障）データ数 M_b よりも正例（正常）データ数 M_a が多い場合、『故障らしい回路を正常と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w を低く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムごとの全体誤答率を比較すればよい。

【0091】

これに対し、正例（正常）データ数 M_a よりも負例（故障）データ数 M_b が多い場合、『正常らしい回路を故障と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w を高く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムごとの全体誤答率を比較すればよい。

【0092】

また、推薦商品のレコメンデーションに関するデータ解析に適用する場合、『どの予測アルゴリズムを採用すれば、推薦商品を購入しそうな顧客を「購入しない」と予測してしまうか、また、どの予測アルゴリズムを採用すれば、推薦商品を購入しそうな顧客を「購入する」と予測してしまうか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。

【0093】

また、『推薦商品を購入しそうな顧客を「購入しない」と予測する』と『推薦商品を購入しそうな顧客を「購入する」と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（購入しなかった）データ数 M_b よりも正例（購入した）データ数 M_a が多い場合、『推薦商品を購入しそうな顧客を「購入する」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w を低く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムごとの全体誤答率を比較すればよい。

【0094】

これに対し、正例（購入した）データ数 M_a よりも負例（購入しなかった）データ数 M_b が多い場合、『推薦商品を購入しそうな顧客を「購入しない」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w を高く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムごとの全体誤答率を比較すればよい。

【0095】

また、融資の与信に関するデータ解析に適用する場合、『どの予測アルゴリズムを採用すれば、返済できそうな顧客を「返済できない」と予測してしまうか、また、どの予測アルゴリズムを採用すれば、返済できそうな顧客を「返済できる」と予測してしまうか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。

【0096】

また、『返済できそうな顧客を「返済できない」と予測する』と『返済できそうな顧客を「返済できる」と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（返済できなかった）データ数 M_b よりも正例（返済できた）データ数 M_a が多い場合、『返済できそうな顧客を「返済でき

10

20

30

40

50

る」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w を低く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムごとの全体誤答率を比較すればよい。

【0097】

これに対し、正例（返済できた）データ数 M_a よりも負例（返済できなかった）データ数 M_b が多い場合、『返済できそうな顧客を「返済できない」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w を高く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムごとの全体誤答率を比較すればよい。

【0098】

このように、実施例 1 B によれば、利用者の関心の程度に応じた全体誤答率で予測アルゴリズムを比較することで、利用者にとって予測精度のよい予測アルゴリズムをわかりやすく提示することができる。

【0099】

（実施例 2 A）

実施例 2 A は、ある 1 つの予測アルゴリズムを適用する場合に、説明変数群の各々について重み付き全体正答率を求めてグラフ表示する例である。これにより、どの説明変数が目的変数に関係しているかを利用者にわかりやすく提示することができる。

【0100】

図 1 4 は、実施例 2 A における説明変数ごとの重み付き全体正答率を示すグラフである。図 1 4 において、横軸は重み変数 w 、縦軸は全体正答率を示す。グラフは図 5 に示したディスプレイ 508 に表示される。図 1 4 では、4 種類の説明変数 $E_1 \sim E_4$ を用いている。

【0101】

図 1 4 では、重み変数 w がどのような値をとっても、説明変数 $E_1 \sim E_4$ の中で説明変数 E_1 の正答率が最も高い。したがって、説明変数 E_1 を採用すると最も予測精度がよくなる。また、説明変数 E_2 、 E_3 については、重み変数 w が $0 \leq w < w_x$ では説明変数 E_3 のほうが全体正答率がよく、 $w_x < w \leq 1$ では説明変数 E_2 のほうが全体正答率がよい。

【0102】

したがって、たとえば、説明変数 E_1 について重み付き全体正答率を求めているときは、利用者は、負例・負例件数よりも正例・正例件数を重視する場合は、重み変数 w の値が大きいときに正答率がよくなる説明変数 E_2 を利用すればよい。一方、正例・正例件数よりも負例・負例件数を重視する場合は、重み変数 w の値が小さいときに正答率（予測精度）がよくなる説明変数 E_3 を採用すればよい。

【0103】

なお、説明変数 E_4 は、説明変数 $E_1 \sim E_4$ の中で最も正答率が低い説明変数であるため、利用者は説明変数 E_4 が最も重要でない説明変数であることが分かる。

【0104】

図 1 5 は、実施例 2 A にかかるデータ解析処理手順を示すフローチャートである。図 1 5 において、まず、データ解析装置は、未選択の説明変数があるか否かを判断する（ステップ S1501）。未選択の説明変数がある場合（ステップ S1501：Yes）、データ解析装置は、未選択の説明変数を選択し（ステップ S1502）、評価処理を実行する（ステップ S1503）。評価処理（ステップ S1503）の詳細については図 11 で説明した評価処理（ステップ S1003）と同一処理内容であるため説明を省略する。評価処理（ステップ S1503）は、選択説明変数について実行される。

【0105】

評価処理（ステップ S1503）のあと、ステップ S1501 に戻り、未選択の説明変数がない場合（ステップ S1501：No）、データ解析装置は、図 1 4 に示したように、説明変数ごとに得られた重み付き全体正答率のグラフ表示処理を実行する（ステップ S1504）。これにより、実施例 2 A でのデータ解析処理を終了する。

【0106】

このように、実施例2Aによれば、利用者の正例データについての関心度の高さに応じて、利用者が重み変数 w を設定することができる。したがって、実施例2Aでは、利用者が設定した重み変数 w の値での説明変数ごとの予測アルゴリズムの予測精度（正答率）を、利用者にわかりやすく提示することができる。

【0107】

たとえば、回路の故障解析に適用する場合、『どの説明変数を採用すれば、正常らしい回路を正常であると予測アルゴリズムが予測してくれるか、また、どの説明変数を採用すれば、故障らしい回路を故障であると予測アルゴリズムが予測してくれるか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。

10

【0108】

また、『正常らしい回路を正常と予測する』と『故障らしい回路を故障と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（故障）データ数 M_b よりも正例（正常）データ数 M_a が多い場合、『故障らしい回路を故障と予測する』ことを重視するほうがよい。この場合は、利用者は重み変数 w を低く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体正答率を説明変数別に比較すればよい。

【0109】

これに対し、正例（正常）データ数 M_a よりも負例（故障）データ数 M_b が多い場合、『正常らしい回路を正常と予測する』ことを重視するほうがよい。この場合は、利用者は重み変数 w を高く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体正答率を説明変数別に比較すればよい。

20

【0110】

また、推薦商品のレコメンデーションに関するデータ解析に適用する場合、『どの説明変数を採用すれば、推薦商品を購入しそうな顧客を「購入する」と予測アルゴリズムが予測してくれるか、また、どの説明変数を採用すれば、推薦商品を購入しそうにない顧客を「購入しない」と予測アルゴリズムが予測してくれるか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。

【0111】

また、『推薦商品を購入しそうな顧客を「購入する」と予測する』と『推薦商品を購入しそうにない顧客を「購入しない」と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（購入しなかった）データ数 M_b よりも正例（購入した）データ数 M_a が多い場合、『推薦商品を購入しそうにない顧客を「購入しない」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w を低く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体正答率を説明変数別に比較すればよい。

30

【0112】

これに対し、正例（購入した）データ数 M_a よりも負例（購入しなかった）データ数 M_b が多い場合、『推薦商品を購入しそうな顧客を「購入する」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w を高く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体正答率を説明変数別に比較すればよい。

40

【0113】

また、融資の与信に関するデータ解析に適用する場合、『どの説明変数を採用すれば、返済できそうな顧客を「返済できる」と予測アルゴリズムが予測してくれるか、また、どの説明変数を採用すれば、返済できそうにない顧客を「返済できない」と予測アルゴリズムが予測してくれるか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。

【0114】

また、『返済できそうな顧客を「返済できる」と予測する』と『返済できそうにない顧客を「返済できない」と予測する』ということについて、いずれを重視すればよいかは、

50

利用者に依存する。したがって、たとえば、負例（返済できなかった）データ数 M_b よりも正例（返済できた）データ数 M_a が多い場合、『返済できそうにない顧客を「返済できない」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w を低く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体正答率を説明変数別に比較すればよい。

【0115】

これに対し、正例（返済できた）データ数 M_a よりも負例（返済できなかった）データ数 M_b が多い場合、『返済できそうな顧客を「返済できる」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w を高く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体正答率を説明変数別に比較すればよい。

10

【0116】

このように、実施例 2 A によれば、利用者の関心の程度に応じた全体正答率で予測アルゴリズムを説明変数別に比較することで、利用者にとって予測アルゴリズムの予測精度がよくなる説明変数をわかりやすく提示することができる。

【0117】

（実施例 2 B）

実施例 2 B は、ある 1 つの予測アルゴリズムを適用する場合に、説明変数群の各々について重み付き全体誤答率を求めてグラフ表示する例である。これにより、どの説明変数が目的変数に関係しているかを利用者にわかりやすく提示することができる。

【0118】

図 16 は、実施例 2 B における説明変数ごとの重み付き全体誤答率を示すグラフである。図 16 において、横軸は重み変数 w 、縦軸は全体誤答率を示す。グラフは図 5 に示したディスプレイ 508 に表示される。図 16 では、4 種類の説明変数 $E_1 \sim E_4$ を用いている。

20

【0119】

図 16 では、重み変数 w がどのような値をとっても、説明変数 $E_1 \sim E_4$ の中で説明変数 E_4 の誤答率が最も低い。したがって、説明変数 E_4 を採用すると最も予測精度がよくなる。説明変数 E_2 、 E_3 については、重み変数 w が $0 \leq w < w_x$ では説明変数 E_2 のほうが全体誤答率がよく、 $w_x < w \leq 1$ では説明変数 E_3 のほうが全体誤答率がよい。

【0120】

したがって、たとえば、説明変数 E_4 について重み付き全体誤答率を求めているときは、利用者は、負例・正例件数よりも正例・負例件数を重視する場合は、重み変数 w の値が大きいときに誤答率がよくなる説明変数 E_3 を利用すればよい。正例・負例件数よりも負例・正例件数を重視する場合は、重み変数 w の値が小さいときに誤答率がよくなる説明変数 E_2 を利用すればよい。

30

【0121】

なお、説明変数 E_1 は、説明変数 $E_1 \sim E_4$ の中で最も誤答率が高い説明変数であるため、説明変数 E_1 が最も重要でない説明変数であることが分かる。

【0122】

なお、実施例 2 B にかかるデータ解析処理手順は、評価処理（ステップ S1503）の詳細以外は、図 15 に示した処理手順と同じである。実施例 2 B では、評価処理（ステップ S1503）の詳細な処理が、図 13 に示した評価処理（ステップ S1003）になる。評価処理（ステップ S1503）は、選択説明変数について実行される。

40

【0123】

このように、実施例 2 B によれば、利用者の負例データについての関心度の高さに応じて、利用者が重み変数 w を設定することができる。したがって、実施例 2 B では、利用者が設定した重み変数 w の値での説明変数ごとの予測アルゴリズムの予測精度（誤答率）を、利用者にわかりやすく提示することができる。

【0124】

たとえば、回路の故障解析に適用する場合、『どの説明変数を採用すれば、正常らしい

50

回路を故障であると予測アルゴリズムが予測してしまうか、また、どの説明変数を採用すれば、故障らしい回路を正常であると予測アルゴリズムが予測してしまうか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。

【0125】

また、『正常らしい回路を故障と予測する』と『故障らしい回路を正常と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（故障）データ数 M_b よりも正例（正常）データ数 M_a が多い場合、『故障らしい回路を正常と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w を低く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体誤答率を説明変数別に比較すればよい。

10

【0126】

これに対し、正例（正常）データ数 M_a よりも負例（故障）データ数 M_b が多い場合、『正常らしい回路を故障と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w を高く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体誤答率を説明変数別に比較すればよい。

【0127】

また、推薦商品のレコメンデーションに関するデータ解析に適用する場合、『どの説明変数を採用すれば、推薦商品を購入しそうな顧客を「購入しない」と予測アルゴリズムが予測してしまうか、また、どの説明変数を採用すれば、推薦商品を購入しそうにない顧客を「購入する」と予測アルゴリズムが予測してしまうか』といった利用者の疑問に対し、

20

グラフ表示でわかりやすく提示することができる。

【0128】

また、『推薦商品を購入しそうな顧客を「購入しない」と予測する』と『推薦商品を購入しそうにない顧客を「購入する」と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（購入しなかった）データ数 M_b よりも正例（購入した）データ数 M_a が多い場合、『推薦商品を購入しそうにない顧客を「購入する」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w を低く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体誤答率を説明変数別に比較すればよい。

【0129】

これに対し、正例（購入した）データ数 M_a よりも負例（購入しなかった）データ数 M_b が多い場合、『推薦商品を購入しそうな顧客を「購入しない」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w を高く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体誤答率を説明変数別に比較すればよい。

30

【0130】

また、融資の与信に関するデータ解析に適用する場合、『どの説明変数を採用すれば、返済できそうな顧客を「返済できない」と予測アルゴリズムが予測してしまうか、また、どの説明変数を採用すれば、返済できそうにない顧客を「返済できる」と予測アルゴリズムが予測してしまうか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。

40

【0131】

また、『返済できそうな顧客を「返済できない」と予測する』と『返済できそうにない顧客を「返済できる」と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（返済できなかった）データ数 M_b よりも正例（返済できた）データ数 M_a が多い場合、『返済できそうにない顧客を「返済できる」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w を低く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体誤答率を説明変数別に比較すればよい。

【0132】

これに対し、正例（返済できた）データ数 M_a よりも負例（返済できなかった）データ

50

数 Mb が多い場合、『返済できそうな顧客を「返済できない」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w を高く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体誤答率を説明変数別に比較すればよい。

【0133】

このように、実施例 2 B によれば、利用者の関心の程度に応じた全体誤答率で予測アルゴリズムを説明変数別に比較することで、利用者にとって予測アルゴリズムの予測精度がよくなる説明変数をわかりやすく提示することができる。

【0134】

(実施例 3 A)

実施例 3 A は、ある 1 つの予測アルゴリズムを適用する場合に、説明変数群の中からいずれか 1 つの説明変数を除いた残余の説明変数群について重み付き全体正答率を求めてグラフ表示する例である。これにより、どの説明変数が目的変数との関係において意味のない説明変数であるかを利用者にわかりやすく提示することができる。

【0135】

図 17 は、実施例 3 A における残余の説明変数群ごとの重み付き全体正答率を示すグラフである。図 17 において、横軸は重み変数 w 、縦軸は全体正答率を示す。グラフは図 5 に示したディスプレイ 508 に表示される。図 17 では、4 種類の残余の説明変数群 $\{E2 \sim E4\}$ 、 $\{E1, E3, E4\}$ 、 $\{E1, E2, E4\}$ 、 $\{E1 \sim E3\}$ を用いている。

【0136】

図 17 では、重み変数 w がどのような値をとっても、残余の説明変数群 $\{E2 \sim E4\}$ 、 $\{E1, E3, E4\}$ 、 $\{E1, E2, E4\}$ 、 $\{E1 \sim E3\}$ の中で残余の説明変数群 $\{E2 \sim E4\}$ の正答率が最も高い。したがって、説明変数 $E1$ が最も意味のない説明変数であることがわかる。また、残余の説明変数群 $\{E1, E3, E4\}$ 、 $\{E1, E2, E4\}$ については、重み変数 w が $0 \leq w < w_x$ では残余の説明変数群 $\{E1, E2, E4\}$ のほうが全体正答率がよく、 $w_x < w \leq 1$ では残余の説明変数群 $\{E1, E3, E4\}$ のほうが全体正答率がよい。

【0137】

したがって、たとえば、残余の説明変数群 $\{E2 \sim E4\}$ について重み付き全体正答率を求めているときは、利用者は、負例・負例件数よりも正例・正例件数を重視する場合は、重み変数 w の値が大きいときに正答率がよくなる残余の説明変数群 $\{E1, E3, E4\}$ を利用すればよい。一方、正例・正例件数よりも負例・負例件数を重視する場合は、重み変数 w の値が小さいときに正答率（予測精度）がよくなる残余の説明変数群 $\{E1, E2, E4\}$ を採用すればよい。

【0138】

なお、残余の説明変数群 $\{E1 \sim E3\}$ は、残余の説明変数群 $\{E2 \sim E4\}$ 、 $\{E1, E3, E4\}$ 、 $\{E1, E2, E4\}$ 、 $\{E1 \sim E3\}$ の中で最も正答率が低い説明変数であるため、利用者は説明変数 $E4$ を採用しないのはよくないということがわかる。

【0139】

図 18 は、実施例 3 A にかかるデータ解析処理手順を示すフローチャートである。図 18 において、まず、データ解析装置は、未選択の説明変数があるか否かを判断する（ステップ S1801）。未選択の説明変数がある場合（ステップ S1801：Yes）、データ解析装置は、未選択の説明変数を選択し（ステップ S1802）、選択説明変数を除く残余の説明変数群を選択する（ステップ S1803）。そして、データ解析装置は、評価処理を実行する（ステップ S1804）。評価処理（ステップ S1804）の詳細については図 11 で説明した評価処理（ステップ S1003）と同一処理内容であるため説明を省略する。評価処理（ステップ S1804）は、選択された残余の説明変数群について実行される。

【0140】

評価処理（ステップ S1804）のあと、ステップ S1801 に戻り、未選択の説明変

10

20

30

40

50

数がない場合（ステップS1801：No）、データ解析装置は、図17に示したように、残余の説明変数群ごとに得られた重み付き全体正答率のグラフ表示処理を実行する（ステップS1805）。これにより、実施例3Aでのデータ解析処理を終了する。

【0141】

このように、実施例3Aによれば、利用者の正例データについての関心度の高さに応じて、利用者が重み変数 w を設定することができる。したがって、実施例3Aでは、利用者が設定した重み変数 w の値での残余の説明変数群ごとの予測アルゴリズムの予測精度（正答率）を、利用者にわかりやすく提示することができる。

【0142】

たとえば、回路の故障解析に適用する場合、『どの説明変数を採用しなければ、正常らしい回路を正常であると予測アルゴリズムが予測してくれるか、また、どの説明変数を採用しなければ、故障らしい回路を故障であると予測アルゴリズムが予測してくれるか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。

10

【0143】

また、『正常らしい回路を正常と予測する』と『故障らしい回路を故障と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（故障）データ数 M_b よりも正例（正常）データ数 M_a が多い場合、『故障らしい回路を故障と予測する』ことを重視するほうがよい。この場合は、利用者は重み変数 w を低く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体正答率を残余の説明変数群別に比較すればよい。

20

【0144】

これに対し、正例（正常）データ数 M_a よりも負例（故障）データ数 M_b が多い場合、『正常らしい回路を正常と予測する』ことを重視するほうがよい。この場合は、利用者は重み変数 w を高く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体正答率を残余の説明変数群別に比較すればよい。

【0145】

また、推薦商品のレコメンデーションに関するデータ解析に適用する場合、『どの説明変数を採用しなければ、推薦商品を購入しそうな顧客を「購入する」と予測アルゴリズムが予測してくれるか、また、どの説明変数を採用しなければ、推薦商品を購入しそうな顧客を「購入しない」と予測アルゴリズムが予測してくれるか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。

30

【0146】

また、『推薦商品を購入しそうな顧客を「購入する」と予測する』と『推薦商品を購入しそうな顧客を「購入しない」と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（購入しなかった）データ数 M_b よりも正例（購入した）データ数 M_a が多い場合、『推薦商品を購入しそうな顧客を「購入しない」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w を低く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体正答率を残余の説明変数群別に比較すればよい。

【0147】

これに対し、正例（購入した）データ数 M_a よりも負例（購入しなかった）データ数 M_b が多い場合、『推薦商品を購入しそうな顧客を「購入する」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w を高く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体正答率を残余の説明変数群別に比較すればよい。

40

【0148】

また、融資の与信に関するデータ解析に適用する場合、『どの説明変数を採用しなければ、返済できそうな顧客を「返済できる」と予測アルゴリズムが予測してくれるか、また、どの説明変数を採用しなければ、返済できそうな顧客を「返済できない」と予測アルゴリズムが予測してくれるか』といった利用者の疑問に対し、グラフ表示でわかりやす

50

く提示することができる。

【0149】

また、『返済できそうな顧客を「返済できる」と予測する』と『返済できそうにない顧客を「返済できない」と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（返済できなかった）データ数 M_b よりも正例（返済できた）データ数 M_a が多い場合、『返済できそうにない顧客を「返済できない」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w を低く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体正答率を残余の説明変数群別に比較すればよい。

【0150】

これに対し、正例（返済できた）データ数 M_a よりも負例（返済できなかった）データ数 M_b が多い場合、『返済できそうな顧客を「返済できる」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w を高く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体正答率を残余の説明変数群別に比較すればよい。

【0151】

このように、実施例 3 A によれば、利用者の関心の程度に応じた全体正答率で予測アルゴリズムを残余の説明変数群別に比較することで、利用者にとって予測アルゴリズムの予測精度がよくなるために不要な説明変数をわかりやすく提示することができる。

【0152】

（実施例 3 B）

実施例 3 B は、ある 1 つの予測アルゴリズムを適用する場合に、説明変数群の各々について重み付き全体誤答率を求めてグラフ表示する例である。これにより、どの説明変数が目的変数に関係しているかを利用者にわかりやすく提示することができる。

【0153】

図 19 は、実施例 3 B における説明変数ごとの重み付き全体誤答率を示すグラフである。図 19 において、横軸は重み変数 w 、縦軸は全体誤答率を示す。グラフは図 5 に示したディスプレイ 508 に表示される。図 19 では、4 種類の残余の説明変数群 $\{E_2 \sim E_4\}$ 、 $\{E_1, E_3, E_4\}$ 、 $\{E_1, E_2, E_4\}$ 、 $\{E_1 \sim E_3\}$ を用いている。

【0154】

図 19 では、重み変数 w がどのような値をとっても、残余の説明変数群 $\{E_2 \sim E_4\}$ 、 $\{E_1, E_3, E_4\}$ 、 $\{E_1, E_2, E_4\}$ 、 $\{E_1 \sim E_3\}$ の中で残余の説明変数群 $\{E_2 \sim E_4\}$ の誤答率が最も高い。したがって、説明変数 E_1 が最も重要な説明変数であることが分かる。また、残余の説明変数群 $\{E_1, E_3, E_4\}$ 、 $\{E_1, E_2, E_4\}$ については、重み変数 w が $0 \leq w < w_x$ では残余の説明変数群 $\{E_1, E_3, E_4\}$ のほうが全体誤答率がよく、 $w_x < w \leq 1$ では残余の説明変数群 $\{E_1, E_2, E_4\}$ のほうが全体誤答率がよい。

【0155】

したがって、たとえば、残余の説明変数群 $\{E_1 \sim E_3\}$ について重み付き全体誤答率を求めているときは、利用者は、負例・正例件数よりも正例・負例件数を重視する場合は、重み変数 w の値が大きいときに誤答率がよくなる残余の説明変数群 $\{E_1, E_2, E_4\}$ を利用すればよい。正例・負例件数よりも負例・正例件数を重視する場合は、重み変数 w の値が小さいときに誤答率がよくなる残余の説明変数群 $\{E_1, E_3, E_4\}$ を利用すればよい。

【0156】

なお、残余の説明変数群 $\{E_1 \sim E_3\}$ は、残余の説明変数群 $\{E_2 \sim E_4\}$ 、 $\{E_1, E_3, E_4\}$ 、 $\{E_1, E_2, E_4\}$ 、 $\{E_1 \sim E_3\}$ の中で最も誤答率が低い残余の説明変数群であるため、利用者は説明変数 E_4 が最も重要でない説明変数であることが分かる。

【0157】

10

20

30

40

50

なお、実施例 3 B にかかるデータ解析処理手順は、評価処理（ステップ S 1 8 0 4）の詳細以外は、図 1 8 に示した処理手順と同じである。実施例 3 B では、評価処理（ステップ S 1 8 0 4）の詳細な処理が、図 1 3 に示した評価処理（ステップ S 1 0 0 3）になる。評価処理（ステップ S 1 8 0 4）は、選択された残余の説明変数群について実行される。

【0158】

このように、実施例 3 B によれば、利用者の負例データについての関心度の高さに応じて、利用者が重み変数 w を設定することができる。したがって、実施例 3 B では、利用者が設定した重み変数 w の値での残余の説明変数群ごとの予測アルゴリズムの予測精度（誤答率）を、利用者にわかりやすく提示することができる。

10

【0159】

たとえば、回路の故障解析に適用する場合、『どの説明変数を採用しなければ、正常らしい回路を故障であると予測アルゴリズムが予測してしまうか、また、どの説明変数を採用しなければ、故障らしい回路を正常であると予測アルゴリズムが予測してしまうか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。

【0160】

また、『正常らしい回路を故障と予測する』と『故障らしい回路を正常と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（故障）データ数 M_b よりも正例（正常）データ数 M_a が多い場合、『故障らしい回路を正常と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w を低く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体誤答率を残余の説明変数群別に比較すればよい。

20

【0161】

これに対し、正例（正常）データ数 M_a よりも負例（故障）データ数 M_b が多い場合、『正常らしい回路を故障と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w を高く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体誤答率を残余の説明変数群別に比較すればよい。

【0162】

また、推薦商品のレコメンデーションに関するデータ解析に適用する場合、『どの説明変数を採用しなければ、推薦商品を購入しそうな顧客を「購入しない」と予測アルゴリズムが予測してしまうか、また、どの説明変数を採用しなければ、推薦商品を購入しそうな顧客を「購入する」と予測アルゴリズムが予測してしまうか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。

30

【0163】

また、『推薦商品を購入しそうな顧客を「購入しない」と予測する』と『推薦商品を購入しそうな顧客を「購入する」と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（購入しなかった）データ数 M_b よりも正例（購入した）データ数 M_a が多い場合、『推薦商品を購入しそうな顧客を「購入する」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w を低く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体誤答率を残余の説明変数群別に比較すればよい。

40

【0164】

これに対し、正例（購入した）データ数 M_a よりも負例（購入しなかった）データ数 M_b が多い場合、『推薦商品を購入しそうな顧客を「購入しない」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w を高く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体誤答率を残余の説明変数群別に比較すればよい。

【0165】

また、融資の与信に関するデータ解析に適用する場合、『どの説明変数を採用しなければ、返済できそうな顧客を「返済できない」と予測アルゴリズムが予測してしまうか、ま

50

た、どの説明変数を採用しなければ、返済できそうにない顧客を「返済できる」と予測アルゴリズムが予測してしまうか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。

【0166】

また、『返済できそうな顧客を「返済できない」と予測する』と『返済できそうにない顧客を「返済できる」と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（返済できなかった）データ数 M_b よりも正例（返済できた）データ数 M_a が多い場合、『返済できそうにない顧客を「返済できる」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w を低く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体誤答率を残余の説明変数群別に比較すればよい。

10

【0167】

これに対し、正例（返済できた）データ数 M_a よりも負例（返済できなかった）データ数 M_b が多い場合、『返済できそうな顧客を「返済できない」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w を高く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体誤答率を残余の説明変数群別に比較すればよい。

【0168】

このように、実施例3Bによれば、利用者の関心の程度に応じた全体誤答率で予測アルゴリズムを残余の説明変数群別に比較することで、利用者にとって予測アルゴリズムの予測精度がよくなるために不要な説明変数をわかりやすく提示することができる。

20

【0169】

（実施例4A）

実施例4Aは、予測アルゴリズムごとに、説明変数の各々について全体正答率を求めてグラフ表示する例である。すなわち、実施例4Aは、実施例1Aと実施例2Aとを組み合わせた例である。実施例4Aでは、重み変数 w の値はあらかじめ指定されることになる。これにより、予測アルゴリズムと説明変数との組み合わせごとに、どの説明変数を採用するとどの予測アルゴリズムの予測精度（正答率）がよくなるかを利用者にわかりやすく提示することができる。

【0170】

図20は、実施例4Aにおける予測アルゴリズムと説明変数との組み合わせごとの全体正答率を示すグラフ508である。図20において、横軸は説明変数、縦軸は全体正答率を示す。グラフは図5に示したディスプレイに表示される。図20では、4種類の予測アルゴリズム $P_1 \sim P_4$ と4種類の説明変数 $E_1 \sim E_4$ を用いている。

30

【0171】

図20では、予測アルゴリズム $P_1 \sim P_4$ ごとに、説明変数 $E_1 \sim E_4$ を採用した場合の全体正答率を折れ線グラフで表示している。たとえば、予測アルゴリズム P_1 は、説明変数 $E_2 \sim E_4$ を採用したときは最も予測精度がよい予測アルゴリズムとなるが、説明変数 E_1 を採用したときに限り、最も予測精度が悪い予測アルゴリズムとなる。

【0172】

このように、予測アルゴリズムが Q 個、説明変数が R 個あるとすると、 $Q \times R$ の組み合わせを網羅した折れ線グラフを得ることができ、どの組み合わせにすると正答率がよくなるか、どの組み合わせにすると正答率が悪くなるか、といったことをわかりやすく提示することができる。

40

【0173】

図21は、実施例4Aにかかるデータ解析処理手順を示すフローチャートである。図21において、重み変数 w の値は指定済みとする。まず、データ解析装置は、未選択の予測アルゴリズムがあるか否かを判断する（ステップS2101）。未選択の予測アルゴリズムがある場合（ステップS2101：Yes）、データ解析装置は、未選択の予測アルゴリズムを選択する（ステップS2102）。

50

【0174】

つぎに、データ解析装置は、選択予測アルゴリズムにおいて未選択の説明変数があるかを判断する（ステップS2103）。未選択の説明変数がある場合（ステップS2103：Yes）、データ解析装置は、未選択の説明変数を選択し（ステップS2104）、評価処理を実行する（ステップS2105）。評価処理（ステップS2105）の詳細については図11で説明した評価処理（ステップS1003）と同一処理内容であるため説明を省略する。評価処理（ステップS2105）は、選択予測アルゴリズムと選択説明変数との組み合わせについて実行される。

【0175】

評価処理（ステップS2105）のあと、ステップS2103に戻り、未選択の説明変数がない場合（ステップS2103：No）、ステップS2101に戻り、データ解析装置は、未選択の予測アルゴリズムがあるかを判断する（ステップS2101）。そして、未選択の予測アルゴリズムがない場合（ステップS2101：No）、データ解析装置は、図20に示したように、予測アルゴリズムと説明変数との組み合わせごとに得られた全体正答率のグラフ表示処理を実行する（ステップS2106）。これにより、実施例4Aでのデータ解析処理を終了する。

【0176】

このように、実施例4Aによれば、利用者の正例データについての関心度の高さに応じて、利用者が設定した重み変数 w の値で、予測アルゴリズムごとに、説明変数ごとの予測アルゴリズムの予測精度（正答率）を、利用者にわかりやすく提示することができる。

【0177】

たとえば、回路の故障解析に適用する場合、『どの説明変数とどの予測アルゴリズムを採用すれば、正常らしい回路を正常であると予測アルゴリズムが予測してくれるか、また、どの説明変数とどの予測アルゴリズムを採用すれば、故障らしい回路を故障であると予測アルゴリズムが予測してくれるか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。

【0178】

また、『正常らしい回路を正常と予測する』と『故障らしい回路を故障と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（故障）データ数 M_b よりも正例（正常）データ数 M_a が多い場合、『故障らしい回路を故障と予測する』ことを重視するほうがよい。この場合は、利用者は重み変数 w の値をあらかじめ低く設定することで、利用者の設定した重み変数 w の値での全体正答率を、予測アルゴリズムと説明変数との組み合わせ別に比較すればよい。

【0179】

これに対し、正例（正常）データ数 M_a よりも負例（故障）データ数 M_b が多い場合、『正常らしい回路を正常と予測する』ことを重視するほうがよい。この場合は、利用者は重み変数 w の値をあらかじめ高く設定することで、利用者の設定した重み変数 w の値での全体正答率を、予測アルゴリズムと説明変数との組み合わせ別に比較すればよい。

【0180】

また、推薦商品のレコメンデーションに関するデータ解析に適用する場合、『どの説明変数とどの予測アルゴリズムを採用すれば、推薦商品を購入しそうな顧客を「購入する」と予測アルゴリズムが予測してくれるか、また、どの説明変数とどの予測アルゴリズムを採用すれば、推薦商品を購入しそうな顧客を「購入しない」と予測アルゴリズムが予測してくれるか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。

【0181】

また、『推薦商品を購入しそうな顧客を「購入する」と予測する』と『推薦商品を購入しそうな顧客を「購入しない」と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（購入しなかった）データ数 M_b よりも正例（購入した）データ数 M_a が多い場合、『推薦商品を購入しそうな顧客を「購入する」と予測する』と『推薦商品を購入しそうな顧客を「購入しない」と予測する』という

顧客を「購入しない」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w の値をあらかじめ低く指定することで、利用者の設定した重み変数 w の値での全体正答率を、予測アルゴリズムと説明変数との組み合わせ別に比較すればよい。

【0182】

これに対し、正例（購入した）データ数 M_a よりも負例（購入しなかった）データ数 M_b が多い場合、『推薦商品を購入しそうな顧客を「購入する」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w の値をあらかじめ高く設定することで、利用者の設定した重み変数 w の値での全体正答率を、予測アルゴリズムと説明変数との組み合わせ別に比較すればよい。

【0183】

また、融資の与信に関するデータ解析に適用する場合、『どの説明変数とどの予測アルゴリズムを採用すれば、返済できそうな顧客を「返済できる」と予測アルゴリズムが予測してくれるか、また、どの説明変数とどの予測アルゴリズムを採用すれば、返済できない顧客を「返済できない」と予測アルゴリズムが予測してくれるか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。

【0184】

また、『返済できそうな顧客を「返済できる」と予測する』と『返済できない顧客を「返済できない」と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（返済できなかった）データ数 M_b よりも正例（返済できた）データ数 M_a が多い場合、『返済できない顧客を「返済できない」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w の値をあらかじめ低く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体正答率を説明変数別に比較すればよい。

【0185】

これに対し、正例（返済できた）データ数 M_a よりも負例（返済できなかった）データ数 M_b が多い場合、『返済できそうな顧客を「返済できる」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w の値をあらかじめ高く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体正答率を説明変数別に比較すればよい。

【0186】

このように、実施例4Aによれば、利用者の関心の程度に応じた全体正答率を、予測アルゴリズムと説明変数との組み合わせ別に比較することで、利用者にとってどの組み合わせを採用すれば正答率がよくなるか、どの組み合わせを採用すれば正答率が悪くなるかをわかりやすく提示することができる。

【0187】

（実施例4B）

実施例4Bは、予測アルゴリズムごとに、説明変数の各々について全体誤答率を求めてグラフ表示する例である。すなわち、実施例4Bは、実施例1Bと実施例2Bとを組み合わせた例である。実施例4Bでは、重み変数 w の値はあらかじめ指定されることになる。これにより、予測アルゴリズムと説明変数との組み合わせごとに、どの説明変数を採用するとどの予測アルゴリズムの予測精度（誤答率）がよくなるかを利用者にわかりやすく提示することができる。

【0188】

図22は、実施例4Bにおける予測アルゴリズムと説明変数との組み合わせごとの全体誤答率を示すグラフである。図22において、横軸は説明変数、縦軸は全体誤答率を示す。グラフは図5に示したディスプレイ508に表示される。図22では、4種類の予測アルゴリズム $P_1 \sim P_4$ と4種類の説明変数 $E_1 \sim E_4$ を用いている。

【0189】

図22では、予測アルゴリズム $P_1 \sim P_4$ ごとに、説明変数 $E_1 \sim E_4$ を採用した場合の全体誤答率を折れ線グラフで表示している。たとえば、予測アルゴリズム P_1 は、説明

10

20

30

40

50

変数 E 2 ～ E 4 を採用したときは最も予測精度がよい予測アルゴリズムとなるが、説明変数 E 1 を採用したときに限り、最も予測精度が悪い予測アルゴリズムとなる。

【0190】

このように、予測アルゴリズムが Q 個、説明変数が R 個あるとすると、J × K の組み合わせを網羅した折れ線グラフを得ることができ、どの組み合わせにすると誤答率がよくなるか、どの組み合わせにすると誤答率が悪くなるか、といったことをわかりやすく提示することができる。

【0191】

なお、実施例 4 B にかかるデータ解析処理手順は、評価処理（ステップ S 2 1 0 5）の詳細以外は、図 2 1 に示した処理手順と同じである。実施例 4 B では、評価処理（ステップ S 2 1 0 5）の詳細な処理が、図 1 3 に示した評価処理（ステップ S 1 0 0 3）になる。評価処理（ステップ S 2 1 0 5）は、選択予測アルゴリズムと選択説明変数との組み合わせについて実行される。

【0192】

このように、実施例 4 B によれば、利用者の負例データについての関心度の高さに応じて、利用者が重み変数 w の値で、予測アルゴリズムごとに、説明変数ごとの予測アルゴリズムの予測精度（誤答率）を、利用者にわかりやすく提示することができる。

【0193】

たとえば、回路の故障解析に適用する場合、『どの説明変数とどの予測アルゴリズムを採用すれば、正常らしい回路を故障であると予測アルゴリズムが予測してしまうか、また、どの説明変数とどの予測アルゴリズムを採用すれば、故障らしい回路を正常であると予測アルゴリズムが予測してしまうか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。

【0194】

また、『正常らしい回路を故障と予測する』と『故障らしい回路を正常と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（故障）データ数 M b よりも正例（正常）データ数 M a が多い場合、『故障らしい回路を正常と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w の値をあらかじめ低く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体誤答率を説明変数別に比較すればよい。

【0195】

これに対し、正例（正常）データ数 M a よりも負例（故障）データ数 M b が多い場合、『正常らしい回路を故障と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w の値をあらかじめ高く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体誤答率を説明変数別に比較すればよい。

【0196】

また、推薦商品のレコメンデーションに関するデータ解析に適用する場合、『どの説明変数とどの予測アルゴリズムを採用すれば、推薦商品を購入しそうな顧客を「購入しない」と予測アルゴリズムが予測してしまうか、また、どの説明変数とどの予測アルゴリズムを採用すれば、推薦商品を購入しそうにない顧客を「購入する」と予測アルゴリズムが予測してしまうか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。

【0197】

また、『推薦商品を購入しそうな顧客を「購入しない」と予測する』と『推薦商品を購入しそうにない顧客を「購入する」と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（購入しなかった）データ数 M b よりも正例（購入した）データ数 M a が多い場合、『推薦商品を購入しそうにない顧客を「購入する」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w の値をあらかじめ低く指定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体誤答率を説明変数別に比較すればよい。

【0198】

これに対し、正例（購入した）データ数 M_a よりも負例（購入しなかった）データ数 M_b が多い場合、『推薦商品を購入しそうな顧客を「購入しない」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w の値をあらかじめ高く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体誤答率を説明変数別に比較すればよい。

【0199】

また、融資の与信に関するデータ解析に適用する場合、『どの説明変数とどの予測アルゴリズムを採用すれば、返済できそうな顧客を「返済できない」と予測アルゴリズムが予測してしまうか、また、どの説明変数とどの予測アルゴリズムを採用すれば、返済できそ

10

【0200】

うにない顧客を「返済できる」と予測アルゴリズムが予測してしまうか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。

また、『返済できそうな顧客を「返済できない」と予測する』と『返済できそうにない顧客を「返済できる」と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（返済できなかった）データ数 M_b よりも正例（返済できた）データ数 M_a が多い場合、『返済できそうにない顧客を「返済できる」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w の値をあらかじめ低く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体誤答率を説明変数別に比較すればよい。

20

【0201】

これに対し、正例（返済できた）データ数 M_a よりも負例（返済できなかった）データ数 M_b が多い場合、『返済できそうな顧客を「返済できない」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w の値をあらかじめ高く指定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体誤答率を説明変数別に比較すればよい。

【0202】

このように、実施例 4 B によれば、利用者の関心の程度に応じた全体誤答率を、予測アルゴリズムと説明変数との組み合わせ別に比較することで、利用者にとってどの組み合わせを採用すれば誤答率がよくなるか、どの組み合わせを採用すれば誤答率が悪くなるかをわかりやすく提示することができる。

30

【0203】

（実施例 5 A）

実施例 5 A は、予測アルゴリズムごとに、説明変数群からいずれか 1 つの説明変数を除いた残余の説明変数群の各々について全体正答率を求めてグラフ表示する例である。すなわち、実施例 5 A は、実施例 1 A と実施例 3 A とを組み合わせた例である。実施例 5 A では、重み変数 w の値はあらかじめ指定されることになる。これにより、予測アルゴリズムと残余の説明変数群との組み合わせごとに、どの説明変数を採用しないとどの予測アルゴリズムの予測精度（正答率）がよくなるかを利用者にわかりやすく提示することができる。

40

【0204】

図 2 3 は、実施例 5 A における予測アルゴリズムと残余の説明変数群との組み合わせごとの全体正答率を示すグラフである。図 2 3 において、横軸は残余の説明変数群、縦軸は全体正答率を示す。グラフは図 5 に示したディスプレイに表示される。図 2 3 では、4 種類の予測アルゴリズム $P_1 \sim P_4$ と 4 種類の残余の説明変数群 $\{E_2 \sim E_4\}$, $\{E_1, E_3, E_4\}$, $\{E_1, E_2, E_4\}$, $\{E_1 \sim E_3\}$ を用いている。

【0205】

図 2 3 では、予測アルゴリズム $P_1 \sim P_4$ ごとに、残余の説明変数群 $\{E_2 \sim E_4\}$, $\{E_1, E_3, E_4\}$, $\{E_1, E_2, E_4\}$, $\{E_1 \sim E_3\}$ を採用した場合の全体正答率を折れ線グラフで表示している。たとえば、予測アルゴリズム P_1 は、残余の説明変

50

数群 {E 1, E 3, E 4}, {E 1, E 2, E 4}, {E 1 ~ E 3} を採用したときは最も予測精度がよい予測アルゴリズムとなるが、残余の説明変数群 {E 2 ~ E 4} を採用したときに限り、最も予測精度が悪い予測アルゴリズムとなる。

【0206】

このように、予測アルゴリズムが Q 個、残余の説明変数群が R 個あるとすると、J × K の組み合わせを網羅した折れ線グラフを得ることができ、どの組み合わせにすると正答率がよくなるか、どの組み合わせにすると正答率が悪くなるか、といったことをわかりやすく提示することができる。

【0207】

図 2 4 は、実施例 5 A にかかるデータ解析処理手順を示すフローチャートである。図 2 4 において、重み変数 w の値は指定済みとする。まず、データ解析装置は、未選択の予測アルゴリズムがあるか否かを判断する（ステップ S 2 4 0 1）。未選択の予測アルゴリズムがある場合（ステップ S 2 4 0 1 : Y e s）、データ解析装置は、未選択の予測アルゴリズムを選択する（ステップ S 2 4 0 2）。

【0208】

つぎに、データ解析装置は、選択予測アルゴリズムにおいて未選択の説明変数があるか否かを判断する（ステップ S 2 4 0 3）。未選択の説明変数がある場合（ステップ S 2 4 0 3 : Y e s）、データ解析装置は、未選択の説明変数を選択し（ステップ S 2 4 0 4）、選択説明変数を除く残余の説明変数群を選択する（ステップ S 2 4 0 5）。そして、データ解析装置は、評価処理を実行する（ステップ S 2 4 0 6）。評価処理（ステップ S 2 4 0 6）の詳細については図 1 1 で説明した評価処理（ステップ S 1 0 0 3）と同一処理内容であるため説明を省略する。評価処理（ステップ S 2 4 0 6）は、選択予測アルゴリズムと選択された残余の説明変数群との組み合わせについて実行される。

【0209】

評価処理（ステップ S 2 4 0 6）のあと、ステップ S 2 4 0 3 に戻り、未選択の説明変数がない場合（ステップ S 2 4 0 3 : N o）、ステップ S 2 4 0 1 に戻り、データ解析装置は、未選択の予測アルゴリズムがあるか否かを判断する（ステップ S 2 4 0 1）。そして、未選択の予測アルゴリズムがない場合（ステップ S 2 4 0 1 : N o）、データ解析装置は、図 2 3 に示したように、予測アルゴリズムと残余の説明変数群との組み合わせごとに得られた全体正答率のグラフ表示処理を実行する（ステップ S 2 4 0 7）。これにより、実施例 4 A でのデータ解析処理を終了する。

【0210】

このように、実施例 5 A によれば、利用者の正例データについての関心度の高さに応じて、利用者が設定した重み変数 w の値で、予測アルゴリズムごとに、残余の説明変数群ごとの予測アルゴリズムの予測精度（正答率）を、利用者にわかりやすく提示することができる。

【0211】

たとえば、回路の故障解析に適用する場合、『予測アルゴリズムの各々についてどの説明変数を採用しなければ、正常らしい回路を正常であると予測アルゴリズムが予測してくれるか、また、予測アルゴリズムの各々についてどの説明変数を採用しなければ、故障らしい回路を故障であると予測アルゴリズムが予測してくれるか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。

【0212】

また、『正常らしい回路を正常と予測する』と『故障らしい回路を故障と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（故障）データ数 M b よりも正例（正常）データ数 M a が多い場合、『故障らしい回路を故障と予測する』ことを重視するほうがよい。この場合は、利用者は重み変数 w の値をあらかじめ低く設定することで、利用者の設定した重み変数 w の値での全体正答率を、予測アルゴリズムと残余の説明変数群との組み合わせ別に比較すればよい。

【0213】

これに対し、正例（正常）データ数 M_a よりも負例（故障）データ数 M_b が多い場合、『正常らしい回路を正常と予測する』ことを重視するほうがよい。この場合は、利用者は重み変数 w の値をあらかじめ高く設定することで、利用者の設定した重み変数 w の値での全体正答率を、予測アルゴリズムと残余の説明変数との組み合わせ別に比較すればよい。

【0214】

また、推薦商品のレコメンデーションに関するデータ解析に適用する場合、『予測アルゴリズムの各々についてどの説明変数を採用しなければ、推薦商品を購入しそうな顧客を「購入する」と予測アルゴリズムが予測してくれるか、また、予測アルゴリズムの各々についてどの説明変数を採用しなければ、推薦商品を購入しそうにない顧客を「購入しない」と予測アルゴリズムが予測してくれるか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。

10

【0215】

また、『推薦商品を購入しそうな顧客を「購入する」と予測する』と『推薦商品を購入しそうにない顧客を「購入しない」と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（購入しなかった）データ数 M_b よりも正例（購入した）データ数 M_a が多い場合、『推薦商品を購入しそうにない顧客を「購入しない」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w の値をあらかじめ低く設定することで、利用者の設定した重み変数 w の値での全体正答率を、予測アルゴリズムと残余の説明変数群との組み合わせ別に比較すればよい。

【0216】

20

これに対し、正例（購入した）データ数 M_a よりも負例（購入しなかった）データ数 M_b が多い場合、『推薦商品を購入しそうな顧客を「購入する」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w の値をあらかじめ高く設定することで、利用者の設定した重み変数 w の値での全体正答率を、予測アルゴリズムと残余の説明変数群との組み合わせ別に比較すればよい。

【0217】

また、融資の与信に関するデータ解析に適用する場合、『予測アルゴリズムの各々についてどの説明変数を採用しなければ、返済できそうな顧客を「返済できる」と予測アルゴリズムが予測してくれるか、また、予測アルゴリズムの各々についてどの説明変数を採用しなければ、返済できそうにない顧客を「返済できない」と予測アルゴリズムが予測してくれるか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。

30

【0218】

また、『返済できそうな顧客を「返済できる」と予測する』と『返済できそうにない顧客を「返済できない」と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（返済できなかった）データ数 M_b よりも正例（返済できた）データ数 M_a が多い場合、『返済できそうにない顧客を「返済できない」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w の値をあらかじめ低く指定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体正答率を残余の説明変数群別に比較すればよい。

40

【0219】

これに対し、正例（返済できた）データ数 M_a よりも負例（返済できなかった）データ数 M_b が多い場合、『返済できそうな顧客を「返済できる」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数 w の値をあらかじめ高く設定することで、利用者の設定した重み変数 w の値での予測アルゴリズムの全体正答率を残余の説明変数群別に比較すればよい。

【0220】

このように、実施例 5 A によれば、利用者の関心の程度に応じた全体正答率を、予測アルゴリズムと残余の説明変数群との組み合わせ別に比較することで、利用者にとって予測アルゴリズムの各々についてどの説明変数を採用しなければ正答率がよくなるか、予測ア

50

ルゴリズムの各々についてどの説明変数を採用しなければ正答率が悪くなるかをわかりやすく提示することができる。

【0221】

(実施例5B)

実施例5Bは、予測アルゴリズムごとに、残余の説明変数群の各々について全体誤答率を求めてグラフ表示する例である。すなわち、実施例5Bは、実施例1Bと実施例3Bとを組み合わせた例である。実施例5Bでは、重み変数 w の値はあらかじめ指定されることになる。これにより、予測アルゴリズムと残余の説明変数群との組み合わせごとに、どの説明変数を採用しないとどの予測アルゴリズムの予測精度（誤答率）がよくなるかを利用者により提示することができる。

10

【0222】

図25は、実施例5Bにおける予測アルゴリズムと残余の説明変数群との組み合わせごとの全体誤答率を示すグラフである。図25において、横軸は残余の説明変数群、縦軸は全体誤答率を示す。グラフは図5に示したディスプレイに表示される。図25では、4種類の予測アルゴリズム $P1 \sim P4$ と4種類の残余の説明変数群 $\{E2 \sim E4\}$ 、 $\{E1, E3, E4\}$ 、 $\{E1, E2, E4\}$ 、 $\{E1 \sim E3\}$ を用いている。

【0223】

図25では、予測アルゴリズム $P1 \sim P4$ ごとに、残余の説明変数群 $\{E2 \sim E4\}$ 、 $\{E1, E3, E4\}$ 、 $\{E1, E2, E4\}$ 、 $\{E1 \sim E3\}$ を採用した場合の全体誤答率を折れ線グラフで表示している。たとえば、予測アルゴリズム $P1$ は、残余の説明変数群 $\{E1, E3, E4\}$ 、 $\{E1, E2, E4\}$ 、 $\{E1 \sim E3\}$ を採用したときは最も予測精度がよい予測アルゴリズムとなるが、残余の説明変数群 $\{E2 \sim E4\}$ を採用したときに限り、最も予測精度が悪い予測アルゴリズムとなる。

20

【0224】

このように、予測アルゴリズムが Q 個、説明変数が R 個あるとすると、 $J \times K$ の組み合わせを網羅した折れ線グラフを得ることができ、どの組み合わせにすると誤答率がよくなるか、どの組み合わせにすると誤答率が悪くなるか、といったことをわかりやすく提示することができる。

【0225】

なお、実施例5Bにかかるデータ解析処理手順は、評価処理（ステップS2406）の詳細以外は、図24に示した処理手順と同じである。実施例5Bでは、評価処理（ステップS2406）の詳細な処理が、図13に示した評価処理（ステップS1003）になる。評価処理（ステップS2406）は、選択予測アルゴリズムと選択された残余の説明変数群との組み合わせについて実行される。

30

【0226】

このように、実施例5Bによれば、利用者の負例データについての関心度の高さに応じて、利用者が設定した重み変数 w の値で、予測アルゴリズムごとに、残余の説明変数群ごとの予測アルゴリズムの予測精度（誤答率）を、利用者により提示することができる。

【0227】

たとえば、回路の故障解析に適用する場合、『予測アルゴリズムの各々についてどの説明変数を採用しなければ、正常らしい回路を故障であると予測アルゴリズムが予測してしまうか、また、予測アルゴリズムの各々についてどの説明変数を採用しなければ、故障らしい回路を正常であると予測アルゴリズムが予測してしまうか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。

40

【0228】

また、『正常らしい回路を故障と予測する』と『故障らしい回路を正常と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（故障）データ数 Mb よりも正例（正常）データ数 Ma が多い場合、『故障らしい回路を正常と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数

50

wの値をあらかじめ低く設定することで、利用者の設定した重み変数wの値での予測アルゴリズムの全体誤答率を残余の説明変数群別に比較すればよい。

【0229】

これに対し、正例（正常）データ数M_aよりも負例（故障）データ数M_bが多い場合、『正常らしい回路を故障と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数wの値をあらかじめ高く設定することで、利用者の設定した重み変数wの値での予測アルゴリズムの全体誤答率を残余の説明変数群別に比較すればよい。

【0230】

また、推薦商品のレコメンデーションに関するデータ解析に適用する場合、『予測アルゴリズムの各々についてどの説明変数を採用しなければ、推薦商品を購入しそうな顧客を「購入しない」と予測アルゴリズムが予測してしまうか、また、予測アルゴリズムの各々についてどの説明変数を採用しなければ、推薦商品を購入しそうにない顧客を「購入する」と予測アルゴリズムが予測してしまうか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。

【0231】

また、『推薦商品を購入しそうな顧客を「購入しない」と予測する』と『推薦商品を購入しそうにない顧客を「購入する」と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（購入しなかった）データ数M_bよりも正例（購入した）データ数M_aが多い場合、『推薦商品を購入しそうにない顧客を「購入する」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数wの値をあらかじめ低く設定することで、利用者の設定した重み変数wの値での予測アルゴリズムの全体誤答率を残余の説明変数群別に比較すればよい。

【0232】

これに対し、正例（購入した）データ数M_aよりも負例（購入しなかった）データ数M_bが多い場合、『推薦商品を購入しそうな顧客を「購入しない」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数wの値をあらかじめ高く設定することで、利用者の設定した重み変数wの値での予測アルゴリズムの全体誤答率を残余の説明変数群別に比較すればよい。

【0233】

また、融資の与信に関するデータ解析に適用する場合、『予測アルゴリズムの各々についてどの説明変数を採用しなければ、返済できそうな顧客を「返済できない」と予測アルゴリズムが予測してしまうか、また、予測アルゴリズムの各々についてどの説明変数を採用しなければ、返済できそうにない顧客を「返済できる」と予測アルゴリズムが予測してしまうか』といった利用者の疑問に対し、グラフ表示でわかりやすく提示することができる。

【0234】

また、『返済できそうな顧客を「返済できない」と予測する』と『返済できそうにない顧客を「返済できる」と予測する』ということについて、いずれを重視すればよいかは、利用者に依存する。したがって、たとえば、負例（返済できなかった）データ数M_bよりも正例（返済できた）データ数M_aが多い場合、『返済できそうにない顧客を「返済できる」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数wの値をあらかじめ低く設定することで、利用者の設定した重み変数wの値での予測アルゴリズムの全体誤答率を残余の説明変数群別に比較すればよい。

【0235】

これに対し、正例（返済できた）データ数M_aよりも負例（返済できなかった）データ数M_bが多い場合、『返済できそうな顧客を「返済できない」と予測する』ことを重視したほうがよい。この場合は、利用者は重み変数wの値をあらかじめ高く設定することで、利用者の設定した重み変数wの値での予測アルゴリズムの全体誤答率を残余の説明変数群別に比較すればよい。

【0236】

10

20

30

40

50

このように、実施例 5 B によれば、利用者の関心の程度に応じた全体誤答率を、予測アルゴリズムと残余の説明変数群との組み合わせ別に比較することで、利用者にとって予測アルゴリズムの各々についてどの説明変数を採用しなければ正答率がよくなるか、予測アルゴリズムの各々についてどの説明変数を採用しなければ正答率が悪くなるかをわかりやすく提示することができる。

【0237】

(実施例 6 A)

実施例 6 A は、予測アルゴリズムごとに全体正答率を算出してグラフ表示する例である。実施例 1 A との相違点は、重み変数 w の値と説明変数とをあらかじめ指定しておく点である。それ以外は実施例 1 A と同一である。これにより、利用者が設定した重み変数 w の値および説明変数で、どの予測アルゴリズムの予測精度（正答率）がよいかを利用者にわかりやすく提示することができる。なお、実施例 6 A にかかるデータ解析処理手順は、実施例 1 A と同一であるため省略する。

10

【0238】

図 26 は、実施例 6 A における予測アルゴリズムごとの全体正答率を示すグラフである。図 26 において、横軸は予測アルゴリズム、縦軸は全体正答率を示す。図 26 では、棒グラフによりグラフ表示する例を示している。図 26 のグラフを表示することで、利用者は、自身が指定した重み変数 w の値および説明変数において、どの予測アルゴリズムの予測精度がよいかを把握することができる。

【0239】

(実施例 6 B)

実施例 6 B は、予測アルゴリズムごとに全体誤答率を算出してグラフ表示する例である。実施例 1 B との相違点は、重み変数 w の値と説明変数とをあらかじめ指定しておく点である。それ以外は実施例 1 B と同一である。これにより、利用者が設定した重み変数 w の値および説明変数で、どの予測アルゴリズムの予測精度（誤答率）がよいかを利用者にわかりやすく提示することができる。なお、実施例 6 B にかかるデータ解析処理手順は、実施例 1 B と同一であるため省略する。

20

【0240】

図 27 は、実施例 6 B における予測アルゴリズムごとの全体誤答率を示すグラフである。図 27 において、横軸は予測アルゴリズム、縦軸は全体誤答率を示す。図 27 では、棒グラフによりグラフ表示する例を示している。図 27 のグラフを表示することで、利用者は、自身が指定した重み変数 w の値および説明変数において、どの予測アルゴリズムの予測精度がよいかを把握することができる。

30

【0241】

(実施例 7 A)

実施例 7 A は、説明変数ごとに全体正答率を算出してグラフ表示する例である。実施例 2 A との相違点は、重み変数 w の値と予測アルゴリズムとをあらかじめ指定しておく点である。それ以外は実施例 2 A と同一である。これにより、利用者が設定した重み変数 w の値および予測アルゴリズムで、どの説明変数を採用したときの予測アルゴリズムの予測精度（正答率）がよいかを利用者にわかりやすく提示することができる。なお、実施例 7 A にかかるデータ解析処理手順は、実施例 2 A と同一であるため省略する。

40

【0242】

図 28 は、実施例 7 A における説明変数ごとの全体正答率を示すグラフである。図 28 において、横軸は説明変数、縦軸は全体正答率を示す。図 28 では、棒グラフによりグラフ表示する例を示している。図 28 のグラフを表示することで、利用者は、自身が指定した重み変数 w の値および予測アルゴリズムにおいて、どの説明変数を採用したときの予測アルゴリズムの予測精度がよいかを把握することができる。

【0243】

(実施例 7 B)

実施例 7 B は、説明変数ごとに全体誤答率を算出してグラフ表示する例である。実施例

50

2 Bとの相違点は、重み変数 w の値と予測アルゴリズムとをあらかじめ指定しておく点である。それ以外は実施例 2 Bと同一である。これにより、利用者が設定した重み変数 w の値および予測アルゴリズムで、どの説明変数を採用したときの予測アルゴリズムの予測精度（誤答率）がよいかを利用者にわかりやすく提示することができる。なお、実施例 7 Bにかかるデータ解析処理手順は、実施例 2 Bと同一であるため省略する。

【0244】

図 29 は、実施例 7 Bにおける説明変数ごとの全体誤答率を示すグラフである。図 29 において、横軸は説明変数、縦軸は全体誤答率を示す。図 29 では、棒グラフによりグラフ表示する例を示している。図 29 のグラフを表示することで、利用者は、自身が指定した重み変数 w の値および予測アルゴリズムにおいて、どの説明変数を採用したときの予測アルゴリズムの予測精度がよいかを把握することができる。

10

【0245】

（実施例 8 A）

実施例 8 Aは、残余の説明変数群ごとに全体正答率を算出してグラフ表示する例である。実施例 3 Aとの相違点は、重み変数 w の値と予測アルゴリズムとをあらかじめ指定しておく点である。それ以外は実施例 3 Aと同一である。これにより、利用者が設定した重み変数 w の値および予測アルゴリズムで、どの説明変数を採用しなかったときの予測アルゴリズムの予測精度（正答率）がよいかを利用者にわかりやすく提示することができる。なお、実施例 8 Aにかかるデータ解析処理手順は、実施例 3 Aと同一であるため省略する。

【0246】

図 30 は、実施例 8 Aにおける残余の説明変数群ごとの全体正答率を示すグラフである。図 30 において、横軸は残余の説明変数群、縦軸は全体正答率を示す。図 30 では、棒グラフによりグラフ表示する例を示している。図 30 のグラフを表示することで、利用者は、自身が指定した重み変数 w の値および予測アルゴリズムにおいて、どの説明変数を採用しなかったときの予測アルゴリズムの予測精度がよいかを把握することができる。

20

【0247】

（実施例 8 B）

実施例 8 Bは、残余の説明変数群ごとに全体誤答率を算出してグラフ表示する例である。実施例 3 Bとの相違点は、重み変数 w の値と予測アルゴリズムとをあらかじめ指定しておく点である。それ以外は実施例 3 Bと同一である。これにより、利用者が設定した重み変数 w の値および予測アルゴリズムで、どの説明変数を採用しなかったときの予測アルゴリズムの予測精度（誤答率）がよいかを利用者にわかりやすく提示することができる。なお、実施例 8 Bにかかるデータ解析処理手順は、実施例 3 Bと同一であるため省略する。

30

【0248】

図 31 は、実施例 8 Bにおける残余の説明変数群ごとの全体誤答率を示すグラフである。図 31 において、横軸は残余の説明変数群、縦軸は全体誤答率を示す。図 31 では、棒グラフによりグラフ表示する例を示している。図 31 のグラフを表示することで、利用者は、自身が指定した重み変数 w の値および予測アルゴリズムにおいて、どの説明変数を採用しなかったときの予測アルゴリズムの予測精度がよいかを把握することができる。

【0249】

（実施例 9 A）

実施例 9 Aは、予測アルゴリズムごとに部分正答率を算出してグラフ表示する例である。具体的には、実施例 9 Aでは、全体正答率を算出するのではなく、 N 個の部分正答率に関する箱ひげ図（箱型図ともいう）を表示する。また、実施例 6 Aと同様、重み変数 w の値と説明変数とをあらかじめ指定しておく。このように、 N 個の部分正答率に関する箱ひげ図を予測アルゴリズムごとに表示することで、予測アルゴリズムごとの部分正答率のばらつきを視覚的に把握することができる。

40

【0250】

図 32 は、実施例 9 Aにおける予測アルゴリズムごとの部分正答率のばらつきを示すグラフである。図 32 において、横軸は予測アルゴリズム、縦軸は部分正答率を示す。図 3

50

2では、箱ひげ図によりグラフ表示する例を示している。図32のグラフを表示することで、利用者は、自身が指定した重み変数 w の値および説明変数において、どの予測アルゴリズムを採用すると、部分正答率のばらつきが大きいかまたは小さいかを視覚的に把握することができる。

【0251】

図33は、実施例9Aにかかるデータ解析処理手順を示すフローチャートである。図33において、まず、データ解析装置は、利用者による入力装置（キーボードやマウス）の操作入力により、重み変数 w の値と説明変数とを設定する（ステップS3301）。つぎに、データ解析装置は、未選択の予測アルゴリズムがあるか否かを判断する（ステップS3302）。未選択の予測アルゴリズムがある場合（ステップS3302：Yes）、データ解析装置は、未選択の予測アルゴリズムを選択し（ステップS3303）、評価処理を実行する（ステップS3304）。評価処理（ステップS3304）の詳細については図34で説明する。

10

【0252】

評価処理（ステップS3304）のあと、ステップS3302に戻り、未選択の予測アルゴリズムがない場合（ステップS3302：No）、データ解析装置は、図32に示したように、予測アルゴリズムごとに得られた部分正答率のばらつきを示す箱ひげ図の表示処理を実行する（ステップS3305）。これにより、実施例9Aでのデータ解析処理を終了する。

【0253】

図34は、実施例9Aにかかる、図33に示した評価処理（ステップS3304）の詳細な処理手順を示すフローチャートである。図34において、ステップS1101～S1109は、図11と同一処理であるため、説明を省略する。ステップS1102において、 $i > N$ となった場合（ステップS1102：Yes）、 N 個の部分正答率から箱ひげ図の要素を算出して（ステップS3410）、ステップS3302に移行する。

20

【0254】

ここで、箱ひげ図の要素とは、最小値、中央値、最大値、第1四分位点、第3四分位点、平均値、IQR（interquartile range）である。ステップS3410では、これらの箱ひげ図の要素を N 個の部分正答率について算出することとなる。これにより、グラフ表示処理（ステップS3305）において、図32に示したような箱ひげ図を表示することができる。

30

【0255】

このように、実施例9Aによれば、予測アルゴリズムごとに部分正答率のばらつきを確認することができるため、利用者は、自身が設定した重み変数 w の値および説明変数で、どの予測アルゴリズムを用いれば部分正答率のばらつきが小さくなるかを、視覚的に把握することができる。

【0256】

（実施例9B）

実施例9Bは、予測アルゴリズムごとに部分誤答率を算出してグラフ表示する例である。具体的には、実施例9Bでは、全体誤答率を算出するのではなく、 N 個の部分誤答率に関する箱ひげ図を表示する。また、実施例6Bと同様、重み変数 w の値と説明変数とをあらかじめ指定しておく。このように、 N 個の部分誤答率に関する箱ひげ図を予測アルゴリズムごとに表示することで、予測アルゴリズムごとの部分誤答率のばらつきを視覚的に把握することができる。

40

【0257】

図35は、実施例9Bにおける予測アルゴリズムごとの部分誤答率のばらつきを示すグラフである。図35において、横軸は予測アルゴリズム、縦軸は部分誤答率を示す。図35では、箱ひげ図によりグラフ表示する例を示している。図35のグラフを表示することで、利用者は、自身が指定した重み変数 w の値および説明変数において、どの予測アルゴリズムを採用すると、部分誤答率のばらつきが大きいかまたは小さいかを視覚的に把握す

50

ることができる。

【0258】

つぎに、実施例9Bにかかるデータ解析処理手順について説明する。実施例9Bにかかるデータ解析処理手順では、評価処理（S3304）以外は、実施例9Aと同一であるため省略する。実施例9Bでは、実施例9Aの評価処理（ステップS3304）の処理内容が異なるため、評価処理（S3304）のかわりに、評価処理（ステップS3604）を実行する。

【0259】

図36は、実施例9Bにかかる評価処理（ステップS3304）の詳細な処理手順を示すフローチャートである。図36において、ステップS1301～S1309は、図13と同一処理であるため、説明を省略する。ステップS1302において、 $i > N$ となった場合（ステップS1302：Yes）、N個の部分誤答率から箱ひげ図の要素を算出して（ステップS3610）、ステップS3302に移行する。

10

【0260】

ここで、箱ひげ図の要素とは、実施例9Aと同様、最小値、中央値、最大値、第1四分位点、第3四分位点、平均値、IQR（interquartile range）である。ステップS3610では、これらの箱ひげ図の要素をN個の部分誤答率について算出することとなる。これにより、グラフ表示処理（ステップS3305）において、図35に示したような箱ひげ図を表示することができる。

【0261】

このように、実施例9Bによれば、予測アルゴリズムごとに部分誤答率のばらつきを確認することができるため、利用者は、自身が設定した重み変数 w の値および説明変数で、どの予測アルゴリズムを用いれば部分誤答率のばらつきが小さくなるかを、視覚的に把握することができる。

20

【0262】

（実施例10A）

実施例10Aは、ある予測アルゴリズムについて、説明変数ごとに部分正答率を算出してグラフ表示する例である。具体的には、実施例10Aでは、全体正答率を算出するのではなく、N個の部分正答率に関する箱ひげ図を表示する。また、実施例7Aと同様、重み変数 w の値と予測アルゴリズムとをあらかじめ指定しておく。このように、N個の部分正答率に関する箱ひげ図を説明変数ごとに表示することで、説明変数ごとの部分正答率のばらつきを視覚的に把握することができる。

30

【0263】

図37は、実施例10Aにおける説明変数ごとの部分正答率のばらつきを示すグラフである。図37において、横軸は説明変数、縦軸は部分正答率を示す。図37では、箱ひげ図によりグラフ表示する例を示している。図37のグラフを表示することで、利用者は、自身が指定した重み変数 w の値および予測アルゴリズムにおいて、どの説明変数を採用すると、部分正答率のばらつきが大きいかわたは小さいかを視覚的に把握することができる。

【0264】

図38は、実施例10Aにかかるデータ解析処理手順を示すフローチャートである。図38において、まず、データ解析装置は、利用者による入力装置（キーボードやマウス）の操作入力により、重み変数 w の値と予測アルゴリズムとを設定する（ステップS3801）。つぎに、データ解析装置は、未選択の説明変数があるか否かを判断する（ステップS3802）。未選択の説明変数がある場合（ステップS3802：Yes）、データ解析装置は、未選択の説明変数を選択し（ステップS3803）、評価処理を実行する（ステップS3804）。評価処理（ステップS3804）の詳細は、図34の評価処理（ステップS3304）と同一処理内容であるため、省略する。

40

【0265】

評価処理（ステップS3804）のあと、ステップS3802に戻り、未選択の予測ア

50

ルゴリズムがない場合（ステップ S 3 8 0 2 : N o）、データ解析装置は、図 3 7 に示したように、説明変数ごとに得られた部分正答率のばらつきを示す箱ひげ図の表示処理を実行する（ステップ S 3 8 0 5）。これにより、実施例 1 0 Aでのデータ解析処理を終了する。

【0 2 6 6】

このように、実施例 1 0 Aによれば、説明変数ごとに部分正答率のばらつきを確認することができるため、利用者は、自身が設定した重み変数 w の値および予測アルゴリズムで、どの説明変数を用いれば部分正答率のばらつきが小さくなるかを、視覚的に把握することができる。

【0 2 6 7】

（実施例 1 0 B）

実施例 1 0 Bは、ある予測アルゴリズムについて、説明変数ごとに部分誤答率を算出してグラフ表示する例である。具体的には、実施例 1 0 Bでは、全体誤答率を算出するのではなく、N個の部分誤答率に関する箱ひげ図を表示する。また、実施例 7 Bと同様、重み変数 w の値と予測アルゴリズムとをあらかじめ指定しておく。このように、N個の部分誤答率に関する箱ひげ図を説明変数ごとに表示することで、説明変数ごとの部分誤答率のばらつきを視覚的に把握することができる。

【0 2 6 8】

図 3 9 は、実施例 1 0 Bにおける説明変数ごとの部分誤答率のばらつきを示すグラフである。図 9 Bにおいて、横軸は説明変数、縦軸は部分誤答率を示す。図 3 9では、箱ひげ図によりグラフ表示する例を示している。図 3 9のグラフを表示することで、利用者は、自身が指定した重み変数 w の値および予測アルゴリズムにおいて、どの説明変数を採用すると、部分誤答率のばらつきが大きいかわたは小さいかを視覚的に把握することができる。

【0 2 6 9】

つぎに、実施例 1 0 Bにかかるデータ解析処理手順について説明する。実施例 1 0 Bにかかるデータ解析処理手順では、評価処理（S 3 8 0 4）以外は、実施例 1 0 Aと同一であるため省略する。実施例 1 0 Bでは、実施例 1 0 Aの評価処理（ステップ S 3 8 0 4）の処理内容が異なるため、図 3 6 に示した評価処理（ステップ S 3 3 0 4）を実行することとなる。

【0 2 7 0】

このように、実施例 1 0 Bによれば、説明変数ごとに部分誤答率のばらつきを確認することができるため、利用者は、自身が設定した重み変数 w の値および予測アルゴリズムで、どの説明変数を用いれば部分誤答率のばらつきが小さくなるかを、視覚的に把握することができる。

【0 2 7 1】

（実施例 1 1 A）

実施例 1 1 Aは、ある予測アルゴリズムについて、残余の説明変数群ごとに部分正答率を算出してグラフ表示する例である。具体的には、実施例 1 1 Aでは、全体正答率を算出するのではなく、N個の部分正答率に関する箱ひげ図を表示する。また、実施例 8 Aと同様、重み変数 w の値と予測アルゴリズムとをあらかじめ指定しておく。このように、N個の部分正答率に関する箱ひげ図を残余の説明変数群ごとに表示することで、残余の説明変数群ごとの部分正答率のばらつきを視覚的に把握することができる。

【0 2 7 2】

図 4 0 は、実施例 1 1 Aにおける残余の説明変数群ごとの部分正答率のばらつきを示すグラフである。図 4 0において、横軸は残余の説明変数群、縦軸は部分正答率を示す。図 4 0では、箱ひげ図によりグラフ表示する例を示している。図 4 0のグラフを表示することで、利用者は、自身が指定した重み変数 w の値および予測アルゴリズムにおいて、どの説明変数を採用しないと、部分正答率のばらつきが大きいかわたは小さいかを視覚的に把握することができる。

10

20

30

40

50

【0273】

図41は、実施例11Aにかかるデータ解析処理手順を示すフローチャートである。図41において、まず、データ解析装置は、利用者による入力装置（キーボードやマウス）の操作入力により、重み変数 w の値と予測アルゴリズムとを設定する（ステップS4101）。つぎに、データ解析装置は、未選択の説明変数があるか否かを判断する（ステップS4102）。未選択の説明変数がある場合（ステップS4102：Yes）、データ解析装置は、未選択の説明変数を選択し（ステップS4103）、選択説明変数を除く残余の説明変数群を選択する（ステップS4104）。

【0274】

このあと、データ解析装置は、評価処理を実行する（ステップS4105）。評価処理（ステップS4105）の詳細は、図34の評価処理（ステップS3304）と同一処理内容であるため、省略する。

【0275】

評価処理（ステップS4105）のあと、ステップS4102に戻り、未選択の説明変数がない場合（ステップS4102：No）、データ解析装置は、図40に示したように、残余の説明変数群ごとに得られた部分正答率のばらつきを示す箱ひげ図の表示処理を実行する（ステップS4106）。これにより、実施例11Aでのデータ解析処理を終了する。

【0276】

このように、実施例11Aによれば、残余の説明変数群ごとに部分正答率のばらつきを確認することができるため、利用者は、自身が設定した重み変数 w の値および予測アルゴリズムで、どの説明変数を用いなければ部分正答率のばらつきが小さくなるかを、視覚的に把握することができる。

【0277】

（実施例11B）

実施例11Bは、ある予測アルゴリズムについて、残余の説明変数群ごとに部分誤答率を算出してグラフ表示する例である。具体的には、実施例11Bでは、全体誤答率を算出するのではなく、 N 個の部分誤答率に関する箱ひげ図を表示する。また、実施例8Bと同様、重み変数 w の値と予測アルゴリズムとをあらかじめ指定しておく。このように、 N 個の部分誤答率に関する箱ひげ図を残余の説明変数群ごとに表示することで、残余の説明変数群ごとの部分誤答率のばらつきを視覚的に把握することができる。

【0278】

図42は、実施例11Bにおける残余の説明変数群ごとの部分誤答率のばらつきを示すグラフである。図42において、横軸は残余の説明変数群、縦軸は部分誤答率を示す。図42では、箱ひげ図によりグラフ表示する例を示している。図42のグラフを表示することで、利用者は、自身が指定した重み変数 w の値および予測アルゴリズムにおいて、どの説明変数を採用しないと、部分誤答率のばらつきが大きいまたは小さいかを視覚的に把握することができる。

【0279】

つぎに、実施例11Bにかかるデータ解析処理手順について説明する。実施例11Bにかかるデータ解析処理手順では、評価処理（S4105）以外は、実施例11Aと同一であるため省略する。実施例11Bでは、実施例11Aの評価処理（ステップS4105）の処理内容が異なるため、図36に示した評価処理（ステップS3304）を実行することとなる。

【0280】

このように、実施例11Bによれば、残余の説明変数群ごとに部分誤答率のばらつきを確認することができるため、利用者は、自身が設定した重み変数 w の値および予測アルゴリズムで、どの説明変数を用いなければ部分誤答率のばらつきが小さくなるかを、視覚的に把握することができる。

【0281】

(データ解析装置の機能的構成例)

図43は、データ解析装置の機能的構成例を示すブロック図である。データ解析装置は、DB(データベース)4300と、評価処理部4301と、平均化処理部4302と、出力部4303と、入力部4304と、選択部4305と、を含む。評価処理部4301～選択部4305は、具体的には、たとえば、図5に示したROM502、RAM503、磁気ディスク505、光ディスク507などの記憶装置に記憶されたプログラムをCPU501に実行させることにより、または、I/F509により、その機能を実現する。なお、評価処理部4301～選択部4305での処理結果は、記憶装置にその都度保持され、他の機能部の処理結果を使用する場合は、記憶装置に保持された処理結果を読み出して自機能部の処理を実行することとなる。

10

【0282】

DB4300は、各種情報を記憶する。具体的には、たとえば、図1や図3に示したデータ群を記憶する。また、DB4300は、各種予測アルゴリズム(予測プログラム)を記憶する。DB4300は、具体的には、たとえば、図5に示したROM502、RAM503、磁気ディスク505、光ディスク507などの記憶装置により実現される。

【0283】

評価処理部4301は、図1～図4に示した評価処理に対応する機能である。評価処理部4301による評価処理をN($N \geq 1$ の自然数)回実行することで、評価結果(たとえば、N個の重み付き部分正答率、N個の重み付き部分誤答率、N個の部分正答率、N個の部分誤答率)を算出することができる。

20

【0284】

平均化処理部4302は、評価処理部4301で算出されたN個の重み付き部分正答率、N個の重み付き部分誤答率、N個の部分正答率、N個の部分誤答率を平均化する。これにより、重み付き全体正答率、N個の重み付き全体誤答率、N個の全体正答率、N個の全体誤答率を得ることができる。

【0285】

出力部4303は、評価処理部4301からの評価結果や、平均化処理部4302による平均化結果(重み付き全体正答率、N個の重み付き全体誤答率、N個の全体正答率、N個の全体誤答率)を出力する。具体的には、たとえば、出力部4303は、評価結果や平均化結果を、実施例1A～11A、1B～11Bに示したように、グラフ表示する。特に、重み変数wに固定値を設定していない場合は、重み変数wの変化を示すグラフを表示することになる。このように、グラフ表示することで、利用者にわかりやすく予測アルゴリズムの予測精度の違いを提示することができる。

30

【0286】

入力部4304は、重み変数wの値の入力を受け付ける。具体的には、たとえば、入力部4304は、利用者が入力装置(キーボードやマウス)を操作することで、重み変数wの値を入力する。重み変数wの値が入力されることで、N個の部分正答率、N個の部分誤答率、N個の全体正答率、N個の全体誤答率が得られ、グラフ表示することができる。また、入力部4304では、利用者が入力装置を操作することで、説明変数や予測アルゴリズムを指定することもできる。これにより、指定された説明変数や予測アルゴリズムで評価処理部4301による評価処理が実行されることとなる。

40

【0287】

選択部4305は、データ群の各々のデータで定義されている説明変数群の中から未選択の説明変数を選択する。選択部4305により説明変数が選択された場合、選択説明変数について評価処理部4301により評価処理が実行されることとなる。また、選択部4305は、説明変数群の中から未選択の説明変数を選択し、選択説明変数を除く残余の説明変数群を、評価処理部4301に与える説明変数群として選択する。これにより、残余の説明変数群について評価処理部4301により評価処理が実行されることとなる。

【0288】

つぎに、評価処理部4301の詳細な機能について説明する。評価処理部4301は、

50

抽出部 4 3 1 1 と、生成部 4 3 1 2 と、第 1 の算出部 4 3 1 3 と、第 2 の算出部 4 3 1 4 と、第 1 の特定部 4 3 1 5 と、第 2 の特定部 4 3 1 6 と、演算部 4 3 1 7 と、を備える。

【0 2 8 9】

抽出部 4 3 1 1 は、DB 4 3 0 0 からデータ群を抽出する。具体的には、たとえば、抽出部 4 3 1 1 は、図 1 ～図 4 に示したように、正例データ群から K 個の正例データをランダムサンプリングする。同様に、負例データ群から同じく K 個の負例データをランダムサンプリングする。

【0 2 9 0】

このように、正例データ群および負例データ群ともに同数のデータ数を抽出することで、正例データ数と負例データ数とに差がある場合でも、同数（K 個）抽出できるため、正例データ数と負例データ数がほぼ同等であるという予測式生成の前提条件に適合する。したがって、生成される予測式の予測精度の低下を防止することができる。なお、正例データ群および負例データ群から同数（K 個）のデータをランダムサンプリングしているが、同数にかぎらず、予測精度に影響を与えない程度の差であれば、同数でなくてもよい。

【0 2 9 1】

生成部 4 3 1 2 は、予測アルゴリズムに基づく予測式を生成する。予測式は、図 2 および図 4 の（2 1）に示したように、抽出部 4 3 1 1 によるランダムサンプリングによって得られた根拠用データ群を用いて求められる。

【0 2 9 2】

第 1 の算出部 4 3 1 3 は、生成部 4 3 1 2 によって生成された予測式に、正例実測値の集合から正例実測値群を除いた残余の正例実測値群を各々与えることにより、第 1 の予測値群を算出する。具体的には、たとえば、第 1 の算出部 4 3 1 3 は、図 2 に示したように、正例データ群から根拠用正例データ群（ランダムサンプリングされた K 個の正例データ）を除いた評価用正例データ群を予測式に与えることで、評価用正例データの予測値群を得る。より具体的には、図 2 の（2 2）に示したように、予測式に評価用正例データ群の各々の評価用正例データを代入し、正規化をおこなうことで、評価用正例データの予測値群を得る。

【0 2 9 3】

第 2 の算出部 4 3 1 4 は、生成部 4 3 1 2 によって生成された予測式に、負例実測値の集合から負例実測値群を除いた残余の負例実測値群を各々与えることにより、第 2 の予測値群を算出する。具体的には、たとえば、第 2 の算出部 4 3 1 4 は、図 4 に示したように、負例データ群から根拠用負例データ群（ランダムサンプリングされた K 個の負例データ）を除いた評価用負例データ群を予測式に与えることで、評価用負例データの予測値群を得る。より具体的には、図 4 の（2 2）に示したように、予測式に評価用負例データ群の各々の評価用負例データを代入し、正規化をおこなうことで、評価用負例データの予測値群を得る。

【0 2 9 4】

第 1 の特定部 4 3 1 5 は、第 1 の算出部 4 3 1 3 によって算出された第 1 の予測値群のうち予測式への入力元となる正例実測値と一致した第 1 の一致件数と、正例実測値と不一致となった第 1 の不一致件数を特定する。具体的には、たとえば、図 2 および図 4 に示したように、第 1 の特定部 4 3 1 5 は、正例・正例件数（第 1 の一致件数）と、正例・負例件数（第 1 の不一致件数）を特定する。

【0 2 9 5】

第 2 の特定部 4 3 1 6 は、第 2 の算出部 4 3 1 4 によって算出された第 2 の予測値群のうち予測式への入力元となる負例実測値と一致した第 2 の一致件数と、負例実測値と不一致となった第 2 の不一致件数を特定する。具体的には、たとえば、図 2 および図 4 に示したように、第 2 の特定部 4 3 1 6 は、負例・負例件数（第 2 の一致件数）と、負例・正例件数（第 2 の不一致件数）を特定する。

【0 2 9 6】

演算部 4 3 1 7 は、第 1 および第 2 の特定部 4 3 1 6 によって特定された特定結果に基

10

20

30

40

50

づいて、予測アルゴリズムについて実測値と予測値との一致度を示す正答率が重み変数に応じて変動する重み付き正答率を演算する。重み付き部分正答率を演算する場合は、演算部4317は、上述した式(1)により、重み付き部分正答率PCAwを演算する。重み変数wが指定されている場合は、重み変数wの値を、式(1)に代入すればよい。

【0297】

また、重み付き部分誤答率を演算する場合は、演算部4317は、上述した式(2)により、重み付き部分正答率PWAwを演算する。重み変数wが指定されている場合は、重み変数wの値を、式(2)に代入すればよい。

【0298】

このように、評価処理部4301では、抽出部4311による1回のランダムサンプリングで重み付き部分正答率、部分正答率、重み付き部分誤答率、部分誤答率のいずれかを演算することができる。

【0299】

また、演算部4317は、実施例9A～11Aに示したように、箱ひげ図を表示する場合には、N個の部分正答率から箱ひげ図の要素を計算することで、出力部4303において箱ひげ図をグラフ表示することができる。同様に、演算部4317は、実施例9B～11Bに示したように、箱ひげ図を表示する場合には、N個の部分誤答率から箱ひげ図の要素を計算することで、出力部4303において箱ひげ図をグラフ表示することができる。

【0300】

以上説明したように、上述した実施の形態によれば、目的変数の値ごとのデータ件数が著しく異なる場合(換言すれば、正例データ数と負例データ数とが著しく異なる)であっても、目的変数の値ごとに同数のデータをサンプリングし、説明変数をもとに目的変数を予測する予測式を作成するため、予測式を精度良く作成することができる。

【0301】

また、予測式を作成するための根拠用データ群と、予測式を評価するための評価用データ群について、各々異なるデータをサンプリングすることになるため、予測式の予測精度を適切に評価することができる。すなわち、予測式を生成したデータと同一データで評価しても予測どおりになり意味がない。したがって、予測式の生成と評価を異なるデータでおこなうことで、予測式を適切に評価することができる。

【0302】

また、各ランダムサンプリングで根拠用データ群を抽出するということは、各ランダムサンプリングで抽出されなかった評価用データ群も、毎回異なるデータ群になる。このように、ランダムサンプリングを複数回おこなって、ランダムサンプリングの都度評価することで、予測式の予測精度のばらつきを適切に評価することができる。

【0303】

具体的には、同じ説明変数を採用した場合にどの予測アルゴリズムに予測精度のばらつきがあるか、どの説明変数を採用した場合に予測アルゴリズムに予測精度のばらつきがあるか、どの説明変数を採用しなかった場合に予測アルゴリズムに予測精度のばらつきがあるか、といった予測アルゴリズムの予測精度のばらつきを適切に評価することができる。

【0304】

また、本実施の形態では、説明変数と目的変数の関係の程度について重回帰式の係数で評価するのではなく、予測式を各々の説明変数を用いて作成したり、各々の説明変数を用いないで(残余の説明変数群で)作成したりすることで、予測式の予測精度を評価する。これにより、説明変数と目的変数のデータ傾向が複雑な場合に、重回帰分析以外の適切な予測アルゴリズム(たとえば、サポートベクターマシンや決定木)を用いて、どの説明変数が、目的変数に関係しているかについての解析を適切におこなうことができる。

【0305】

また、利用者によっては、正例(故障解析でいう「正常」、レコメンデーション解析でいう「購入した」、融資の与信に関するデータ解析でいう「返済できた」)のデータ数が多い場合は、負例に重きをおいた解析結果を得たいというニーズがある。

【0306】

また、同様に、利用者によっては、負例（故障解析でいう「故障」、レコメンデーション解析でいう「購入しなかった」、融資の与信に関するデータ解析でいう「返済できなかった」）のデータ数が多い場合は、正例に重きをおいた解析結果を得たいというニーズがある。

【0307】

このように、利用者のニーズによって正例、負例のいずれに重きを置くかが異なるため、重み変数 w が変化する重み付き全体正答率や重み付き全体誤答率を表示することで、『どの程度の重みにすれば、予測アルゴリズムの精度（正答率や誤答率）がよくなるか』といったことを、わかりやすく提示することができる。したがって、利用者によって使い勝手のよいデータ解析結果を提供することができる。

10

【0308】

また、利用者が重み変数 w の値を指定することで、その重み変数 w の値で、全体正答率や全体誤答率をグラフ表示したり、部分正答率や部分誤答率のばらつきをグラフ表示（たとえば、箱ひげ図）する。したがって、利用者が重きをおきたい状況を反映して、『どの予測アルゴリズムの精度（正答率や誤答率）がよいか』といったことを、わかりやすく提示することができる。

【0309】

なお、上述した実施の形態では、回路の故障解析やレコメンデーション解析、融資の与信に関するデータ解析について実施例を用いて説明したが、これらのデータ解析には限らず、各種装置の性能に関するデータ解析、各種商品やサービスに関するデータ解析に適用することができる。

20

【0310】

また、上述した実施の形態では、正例データ群および負例データ群からそれぞれ K 個サンプリングしているが、 K は、正例データ群および負例データ群のうちデータ数が少ない方のデータ群のデータ数よりも低い値となる。また、ランダムサンプリングの都度、サンプリングされるデータをなるべく前回と異なるデータをサンプリングするのが好ましい。このため、 K は、少ない方のデータ群のデータ数の $2/3$ 以下、さらに好ましくは $1/2$ 以下としてもよい。

【0311】

なお、本実施の形態で説明したデータ解析方法は、予め用意されたプログラムをパーソナル・コンピュータやワークステーション等のコンピュータで実行することにより実現することができる。本データ解析プログラムは、ハードディスク、フレキシブルディスク、CD-ROM、MO、DVD等のコンピュータで読み取り可能な記録媒体に記録され、コンピュータによって記録媒体から読み出されることによって実行される。

30

【0312】

上述した実施の形態に関し、さらに以下の付記を開示する。

【0313】

（付記1）少なくとも1つ以上の説明変数である説明変数群に対する目的変数が正例である正例実測値の集合と前記説明変数群に対する前記目的変数が負例である負例実測値の集合を有する実測値集合を記憶する記憶装置を有するコンピュータに、

40

前記実測値集合から、前記正例実測値と前記負例実測値とが同数となるように正例実測値群および負例実測値群をランダムに抽出する抽出工程と、

前記抽出工程によって抽出された正例実測値群および負例実測値群に基づいて、予測アルゴリズムについて前記目的変数を予測する予測式を生成する生成工程と、

を実行させることを特徴とするデータ解析プログラム。

【0314】

（付記2）前記生成工程によって生成された予測式に、前記正例実測値の集合から前記正例実測値群を除いた残余の正例実測値群を各々与えることにより、第1の予測値群を算出する第1の算出工程と、

50

前記生成工程によって生成された予測式に、前記負例実測値の集合から前記負例実測値群を除いた残余の負例実測値群を各々与えることにより、第2の予測値群を算出する第2の算出工程と、

前記第1の算出工程によって算出された第1の予測値群のうち前記予測式への入力元となる前記正例実測値と一致した第1の一致件数と、前記正例実測値と不一致となった第1の不一致件数を特定する第1の特定工程と、

前記第2の算出工程によって算出された第2の予測値群のうち前記予測式への入力元となる前記負例実測値と一致した第2の一致件数と、前記負例実測値と不一致となった第2の不一致件数を特定する第2の特定工程と、

前記第1および第2の特定工程によって特定された特定結果に基づいて、前記予測アルゴリズムについて実測値と予測値との一致度を示す正答率が重み変数に応じて変動する重み付き正答率を演算する演算工程と、

前記演算工程によって演算された演算結果を出力する出力工程と、

を前記コンピュータに実行させることを特徴とする付記1に記載のデータ解析プログラム。

【0315】

(付記3) 前記抽出工程、前記生成工程、前記第1の算出工程、前記第2の算出工程、前記第1の特定工程、前記第2の特定工程、および前記演算工程を、前記コンピュータに複数回実行させ、

前記演算工程によって演算された複数の重み付き正答率を平均化することにより、前記予測アルゴリズムについて実測値と予測値との一致度を示す正答率が重み変数に応じて変動する重み付き平均正答率を出力する平均化工程を前記コンピュータに実行させることを特徴とする付記2に記載のデータ解析プログラム。

【0316】

(付記4) 予測アルゴリズムごとに、前記抽出工程、前記生成工程、前記第1の算出工程、前記第2の算出工程、前記第1の特定工程、前記第2の特定工程、および前記演算工程を、前記コンピュータに複数回実行させ、

前記演算工程によって前記予測アルゴリズムごとに演算された複数の重み付き正答率を平均化することにより、前記予測アルゴリズムについて実測値と予測値との一致度を示す正答率が重み変数に応じて変動する重み付き平均正答率を出力する平均化工程を、前記予測アルゴリズムごとに前記コンピュータに実行させ、

前記出力工程は、前記予測アルゴリズムごとに出力された重み付き平均正答率をグラフ表示することを特徴とする付記2に記載のデータ解析プログラム。

【0317】

(付記5) 前記重み変数の値の入力を受け付ける入力工程を前記コンピュータに実行させ、

前記出力工程は、前記予測アルゴリズムごとに出力された重み付き平均正答率に、前記入力工程によって入力された前記重み変数の値が与えられた前記予測アルゴリズムごとの平均正答率を、グラフ表示することを特徴とする付記4に記載のデータ解析プログラム。

【0318】

(付記6) 前記説明変数群の中から未選択の説明変数を選択する選択工程を前記コンピュータに実行させ、

前記選択工程によって選択された説明変数ごとに、前記抽出工程、前記生成工程、前記第1の算出工程、前記第2の算出工程、前記第1の特定工程、前記第2の特定工程、および前記演算工程を、前記コンピュータに複数回実行させ、

前記演算工程によって前記説明変数ごとに演算された複数の重み付き正答率を平均化することにより、前記予測アルゴリズムについて実測値と予測値との一致度を示す正答率が重み変数に応じて変動する重み付き平均正答率を出力する平均化工程を、前記説明変数ごとに前記コンピュータに実行させ、

前記出力工程は、前記説明変数ごとに出力された重み付き平均正答率をグラフ表示する

10

20

30

40

50

ことを特徴とする付記 2 に記載のデータ解析プログラム。

【0319】

(付記 7) 前記重み変数の値の入力を受け付ける入力工程を前記コンピュータに実行させ、

前記出力工程は、前記説明変数ごとに出力された重み付き平均正答率に前記入力工程によって入力された前記重み変数の値が与えられた前記説明変数ごとの平均正答率を、グラフ表示することを特徴とする付記 6 に記載のデータ解析プログラム。

【0320】

(付記 8) 前記説明変数群の中から未選択の説明変数を選択する選択工程を前記コンピュータに実行させ、

前記選択工程によって選択されなかった残余の説明変数群ごとに、前記抽出工程、前記生成工程、前記第 1 の算出工程、前記第 2 の算出工程、前記第 1 の特定工程、前記第 2 の特定工程、および前記演算工程を、前記コンピュータに複数回実行させ、

前記演算工程によって前記残余の説明変数群ごとに演算された複数の重み付き正答率を平均化することにより、前記予測アルゴリズムについて実測値と予測値との一致度を示す正答率が重み変数に応じて変動する重み付き平均正答率を出力する平均化工程を、前記残余の説明変数群ごとに前記コンピュータに実行させ、

前記出力工程は、前記残余の説明変数群ごとに出力された重み付き平均正答率をグラフ表示することを特徴とする付記 2 に記載のデータ解析プログラム。

【0321】

(付記 9) 前記重み変数の値の入力を受け付ける入力工程を前記コンピュータに実行させ、

前記出力工程は、前記残余の説明変数群ごとに出力された重み付き平均正答率に、前記入力工程によって入力された前記重み変数の値が与えられた前記残余の説明変数群ごとの平均正答率を、グラフ表示することを特徴とする付記 8 に記載のデータ解析プログラム。

【0322】

(付記 10) 前記説明変数群の中から未選択の説明変数を選択する選択工程を前記コンピュータに実行させ、

前記選択工程によって選択された説明変数ごとに、前記抽出工程、前記生成工程、前記第 1 の算出工程、前記第 2 の算出工程、前記第 1 の特定工程、前記第 2 の特定工程、および前記演算工程を、前記コンピュータに複数回実行させ、前記平均化工程を前記説明変数ごとに前記コンピュータに実行させる処理を、予測アルゴリズムごとに前記コンピュータに実行させ、

前記出力工程は、前記説明変数ごとに出力された重み付き平均正答率を、前記予測アルゴリズムごとにグラフ表示することを特徴とする付記 3 に記載のデータ解析プログラム。

【0323】

(付記 11) 前記説明変数群の中から未選択の説明変数を選択する選択工程を前記コンピュータに実行させ、

前記選択工程によって選択されなかった残余の説明変数群ごとに、前記抽出工程、前記生成工程、前記第 1 の算出工程、前記第 2 の算出工程、前記第 1 の特定工程、前記第 2 の特定工程、および前記演算工程を、前記コンピュータに複数回実行させ、前記平均化工程を前記残余の説明変数群ごとに前記コンピュータに実行させる処理を、予測アルゴリズムごとに前記コンピュータに実行させ、

前記出力工程は、前記残余の説明変数群ごとに出力された重み付き平均正答率を、前記予測アルゴリズムごとにグラフ表示することを特徴とする付記 3 に記載のデータ解析プログラム。

【0324】

(付記 12) 予測アルゴリズム群の中から未選択の予測アルゴリズムを選択する選択工程と、

前記重み変数の値の入力を受け付ける入力工程と、を前記コンピュータに実行させ、

10

20

30

40

50

前記選択工程によって選択された予測アルゴリズムごとに、前記抽出工程、前記生成工程、前記第1の算出工程、前記第2の算出工程、前記第1の特定工程、前記第2の特定工程、および前記演算工程を、前記コンピュータに複数回実行させ、

前記出力工程は、前記予測アルゴリズムごとに複数個出力された重み付き正答率に前記入力工程によって入力された前記重み変数の値が与えられた、前記予測アルゴリズムごとに複数個存在する正答率について、それぞればらつきを示すグラフを表示することを特徴とする付記2に記載のデータ解析プログラム。

【0325】

(付記13) 前記説明変数群の中から未選択の説明変数を選択する選択工程と、

前記重み変数の値の入力を受け付ける入力工程と、を前記コンピュータに実行させ、

前記選択工程によって選択された説明変数ごとに、前記抽出工程、前記生成工程、前記第1の算出工程、前記第2の算出工程、前記第1の特定工程、前記第2の特定工程、および前記演算工程を、前記コンピュータに複数回実行させ、

前記出力工程は、前記説明変数ごとに複数個出力された重み付き正答率に前記入力工程によって入力された前記重み変数の値が与えられた、前記説明変数ごとに複数個存在する正答率について、それぞればらつきを示すグラフを表示することを特徴とする付記2に記載のデータ解析プログラム。

【0326】

(付記14) 前記説明変数群の中から未選択の説明変数を選択する選択工程と、

前記重み変数の値の入力を受け付ける入力工程と、を前記コンピュータに実行させ、

前記選択工程によって選択されなかった残余の説明変数群ごとに、前記抽出工程、前記生成工程、前記第1の算出工程、前記第2の算出工程、前記第1の特定工程、前記第2の特定工程、および前記演算工程を、前記コンピュータに複数回実行させ、

前記出力工程は、前記残余の説明変数群ごとに複数個出力された重み付き正答率に前記入力工程によって入力された前記重み変数の値が与えられた、前記残余の説明変数群ごとに複数個存在する正答率について、それぞればらつきを示すグラフを表示することを特徴とする付記2に記載のデータ解析プログラム。

【0327】

(付記15) 前記演算工程は、下記式(1)に基づいて、前記重み付き正答率を演算することを特徴とする付記2～14のいずれか一つに記載のデータ解析プログラム。

$$PCAw = w \times N_{pp} / (N_{pp} + N_{pn}) + (1 - w) \times N_{nn} / (N_{nn} + N_{np}) \cdots (1)$$

ただし、PCAwは前記重み付き正答率、wは前記重み変数、N_{pp}は前記第1の一致件数、N_{pn}は前記第1の不一致件数、N_{nn}は前記第2の一致件数、N_{np}は前記第2の不一致件数。

【0328】

(付記16) 前記生成工程によって生成された予測式に、前記正例実測値の集合から前記正例実測値群を除いた残余の正例実測値群を各々与えることにより、第1の予測値群を算出する第1の算出工程と、

前記生成工程によって生成された予測式に、前記負例実測値の集合から前記負例実測値群を除いた残余の負例実測値群を各々与えることにより、第2の予測値群を算出する第2の算出工程と、

前記第1の算出工程によって算出された第1の予測値群のうち前記予測式への入力元となる前記正例実測値と一致した第1の一致件数と、前記正例実測値と不一致となった第1の不一致件数を特定する第1の特定工程と、

前記第2の算出工程によって算出された第2の予測値群のうち前記予測式への入力元となる前記負例実測値と一致した第2の一致件数と、前記負例実測値と不一致となった第2の不一致件数を特定する第2の特定工程と、

前記第1および第2の特定工程によって特定された特定結果に基づいて、前記予測アルゴリズムについて実測値と予測値との不一致度を示す誤答率が重み変数に応じて変動する

重み付き誤答率を演算する演算工程と、

前記演算工程によって演算された演算結果を出力する出力工程と、

を前記コンピュータに実行させることを特徴とする付記 1 に記載のデータ解析プログラム。

【0329】

(付記 17) 前記抽出工程、前記生成工程、前記第 1 の算出工程、前記第 2 の算出工程、前記第 1 の特定工程、前記第 2 の特定工程、および前記演算工程を、前記コンピュータに複数回実行させ、

前記演算工程によって演算された複数の重み付き誤答率を平均化することにより、前記予測アルゴリズムについて実測値と予測値との不一致度を示す誤答率が重み変数に応じて変動する重み付き平均誤答率を出力する平均化工程を前記コンピュータに実行させることを特徴とする付記 16 に記載のデータ解析プログラム。

10

【0330】

(付記 18) 予測アルゴリズムごとに、前記抽出工程、前記生成工程、前記第 1 の算出工程、前記第 2 の算出工程、前記第 1 の特定工程、前記第 2 の特定工程、および前記演算工程を、前記コンピュータに複数回実行させ、

前記演算工程によって前記予測アルゴリズムごとに演算された複数の重み付き誤答率を平均化することにより、前記予測アルゴリズムについて実測値と予測値との不一致度を示す誤答率が重み変数に応じて変動する重み付き平均誤答率を出力する平均化工程を、前記予測アルゴリズムごとに前記コンピュータに実行させ、

20

前記出力工程は、前記予測アルゴリズムごとに出力された重み付き平均誤答率をグラフ表示することを特徴とする付記 16 に記載のデータ解析プログラム。

【0331】

(付記 19) 前記重み変数の値の入力を受け付ける入力工程を前記コンピュータに実行させ、

前記出力工程は、前記予測アルゴリズムごとに出力された重み付き平均誤答率に、前記入力工程によって入力された前記重み変数の値が与えられた前記予測アルゴリズムごとの平均誤答率を、グラフ表示することを特徴とする付記 18 に記載のデータ解析プログラム。

【0332】

(付記 20) 前記説明変数群の中から未選択の説明変数を選択する選択工程を前記コンピュータに実行させ、

30

前記選択工程によって選択された説明変数ごとに、前記抽出工程、前記生成工程、前記第 1 の算出工程、前記第 2 の算出工程、前記第 1 の特定工程、前記第 2 の特定工程、および前記演算工程を、前記コンピュータに複数回実行させ、

前記演算工程によって前記説明変数ごとに演算された複数の重み付き誤答率を平均化することにより、前記予測アルゴリズムについて実測値と予測値との不一致度を示す誤答率が重み変数に応じて変動する重み付き平均誤答率を出力する平均化工程を、前記説明変数ごとに前記コンピュータに実行させ、

40

前記出力工程は、前記説明変数ごとに出力された重み付き平均誤答率をグラフ表示することを特徴とする付記 16 に記載のデータ解析プログラム。

【0333】

(付記 21) 前記重み変数の値の入力を受け付ける入力工程を前記コンピュータに実行させ、

前記出力工程は、前記説明変数ごとに出力された重み付き平均誤答率に前記入力工程によって入力された前記重み変数の値が与えられた前記説明変数ごとの平均誤答率を、グラフ表示することを特徴とする付記 20 に記載のデータ解析プログラム。

【0334】

(付記 22) 前記説明変数群の中から未選択の説明変数を選択する選択工程を前記コンピュータに実行させ、

50

前記選択工程によって選択されなかった残余の説明変数群ごとに、前記抽出工程、前記生成工程、前記第1の算出工程、前記第2の算出工程、前記第1の特定工程、前記第2の特定工程、および前記演算工程を、前記コンピュータに複数回実行させ、

前記演算工程によって前記残余の説明変数群ごとに演算された複数の重み付き誤答率を平均化することにより、前記予測アルゴリズムについて実測値と予測値との不一致度を示す誤答率が重み変数に応じて変動する重み付き平均誤答率を出力する平均化工程を、前記残余の説明変数群ごとに前記コンピュータに実行させ、

前記出力工程は、前記残余の説明変数群ごとに出力された重み付き平均誤答率をグラフ表示することを特徴とする付記16に記載のデータ解析プログラム。

【0335】

(付記23) 前記重み変数の値の入力を受け付ける入力工程を前記コンピュータに実行させ、

前記出力工程は、前記残余の説明変数群ごとに出力された重み付き平均誤答率に、前記入力工程によって入力された前記重み変数の値が与えられた前記残余の説明変数群ごとの平均誤答率を、グラフ表示することを特徴とする付記22に記載のデータ解析プログラム。

【0336】

(付記24) 前記説明変数群の中から未選択の説明変数を選択する選択工程を前記コンピュータに実行させ、

前記選択工程によって選択された説明変数ごとに、前記抽出工程、前記生成工程、前記第1の算出工程、前記第2の算出工程、前記第1の特定工程、前記第2の特定工程、および前記演算工程を、前記コンピュータに複数回実行させ、前記平均化工程を前記説明変数ごとに前記コンピュータに実行させる処理を、予測アルゴリズムごとに前記コンピュータに実行させ、

前記出力工程は、前記説明変数ごとに出力された重み付き平均誤答率を、前記予測アルゴリズムごとにグラフ表示することを特徴とする付記17に記載のデータ解析プログラム。

【0337】

(付記25) 前記説明変数群の中から未選択の説明変数を選択する選択工程を前記コンピュータに実行させ、

前記選択工程によって選択されなかった残余の説明変数群ごとに、前記抽出工程、前記生成工程、前記第1の算出工程、前記第2の算出工程、前記第1の特定工程、前記第2の特定工程、および前記演算工程を、前記コンピュータに複数回実行させ、前記平均化工程を前記残余の説明変数群ごとに前記コンピュータに実行させる処理を、予測アルゴリズムごとに前記コンピュータに実行させ、

前記出力工程は、前記残余の説明変数群ごとに出力された重み付き平均誤答率を、前記予測アルゴリズムごとにグラフ表示することを特徴とする付記17に記載のデータ解析プログラム。

【0338】

(付記26) 予測アルゴリズム群の中から未選択の予測アルゴリズムを選択する選択工程と、

前記重み変数の値の入力を受け付ける入力工程と、を前記コンピュータに実行させ、

前記選択工程によって選択された予測アルゴリズムごとに、前記抽出工程、前記生成工程、前記第1の算出工程、前記第2の算出工程、前記第1の特定工程、前記第2の特定工程、および前記演算工程を、前記コンピュータに複数回実行させ、

前記出力工程は、前記予測アルゴリズムごとに複数個出力された重み付き誤答率に前記入力工程によって入力された前記重み変数の値が与えられた、前記予測アルゴリズムごとに複数個存在する誤答率について、それぞればらつきを示すグラフを表示することを特徴とする付記16に記載のデータ解析プログラム。

【0339】

10

20

30

40

50

(付記 27) 前記説明変数群の中から未選択の説明変数を選択する選択工程と、
 前記重み変数の値の入力を受け付ける入力工程と、を前記コンピュータに実行させ、
 前記選択工程によって選択された説明変数ごとに、前記抽出工程、前記生成工程、前記第 1 の算出工程、前記第 2 の算出工程、前記第 1 の特定工程、前記第 2 の特定工程、および前記演算工程を、前記コンピュータに複数回実行させ、
 前記出力工程は、前記説明変数ごとに複数個出力された重み付き誤答率に前記入力工程によって入力された前記重み変数の値が与えられた、前記説明変数ごとに複数個存在する誤答率について、それぞればらつきを示すグラフを表示することを特徴とする付記 16 に記載のデータ解析プログラム。

【0340】

10

(付記 28) 前記説明変数群の中から未選択の説明変数を選択する選択工程と、
 前記重み変数の値の入力を受け付ける入力工程と、を前記コンピュータに実行させ、
 前記選択工程によって選択されなかった残余の説明変数群ごとに、前記抽出工程、前記生成工程、前記第 1 の算出工程、前記第 2 の算出工程、前記第 1 の特定工程、前記第 2 の特定工程、および前記演算工程を、前記コンピュータに複数回実行させ、
 前記出力工程は、前記残余の説明変数群ごとに複数個出力された重み付き誤答率に前記入力工程によって入力された前記重み変数の値が与えられた、前記残余の説明変数群ごとに複数個存在する誤答率について、それぞればらつきを示すグラフを表示することを特徴とする付記 16 に記載のデータ解析プログラム。

【0341】

20

(付記 29) 前記演算工程は、下記式 (2) に基づいて、前記重み付き誤答率を演算することを特徴とする付記 16 ~ 28 のいずれか一つに記載のデータ解析プログラム。

$$PWA w = w \times N_{pn} / (N_{pp} + N_{pn}) + (1 - w) \times N_{np} / (N_{nn} + N_{np}) \dots (2)$$

ただし、PWA w は前記重み付き誤答率、w は前記重み変数、N_{pp} は前記第 1 の一致件数、N_{pn} は前記第 1 の不一致件数、N_{nn} は前記第 2 の一致件数、N_{np} は前記第 2 の不一致件数。

【0342】

(付記 30) 前記説明変数群は、回路の故障原因に関する数値情報であり、前記説明変数は、前記故障原因による正常または故障を示す数値情報であることを特徴とする付記 1 ~ 29 のいずれか一つに記載のデータ解析プログラム。

30

【0343】

(付記 31) 前記説明変数群は、顧客の属性に関する数値情報であり、前記説明変数は、商品の購入実績状況を示す数値情報であることを特徴とする付記 1 ~ 29 のいずれか一つに記載のデータ解析プログラム。

【0344】

(付記 32) 前記説明変数群は、顧客の属性に関する数値情報であり、前記説明変数は、前記顧客への融資後の返済状況を示す数値情報であることを特徴とする付記 1 ~ 29 のいずれか一つに記載のデータ解析プログラム。

【0345】

40

(付記 33) 少なくとも 1 つ以上の説明変数である説明変数群に対する目的変数が正例である正例実測値の集合と前記説明変数群に対する前記目的変数が負例である負例実測値の集合を有する実測値集合を記憶する記憶装置を有するコンピュータが、

前記実測値集合から、前記正例実測値と前記負例実測値とが同数となるように正例実測値群および負例実測値群をランダムに抽出する抽出工程と、

前記抽出工程によって抽出された正例実測値群および負例実測値群に基づいて、予測アルゴリズムについて前記目的変数を予測する予測式を生成する生成工程と、

を実行することを特徴とするデータ解析方法。

【0346】

(付記 34) 少なくとも 1 つ以上の説明変数である説明変数群に対する目的変数が正例で

50

ある正例実測値の集合と前記説明変数群に対する前記目的変数が負例である負例実測値の集合を有する実測値集合を記憶する記憶装置を有するデータ解析装置であって、

前記実測値集合から、前記正例実測値と前記負例実測値とが同数となるように正例実測値群および負例実測値群をランダムに抽出する抽出手段と、

前記抽出手段によって抽出された正例実測値群および負例実測値群に基づいて、予測アルゴリズムについて前記目的変数を予測する予測式を生成する生成手段と、

を備えることを特徴とするデータ解析装置。

【符号の説明】

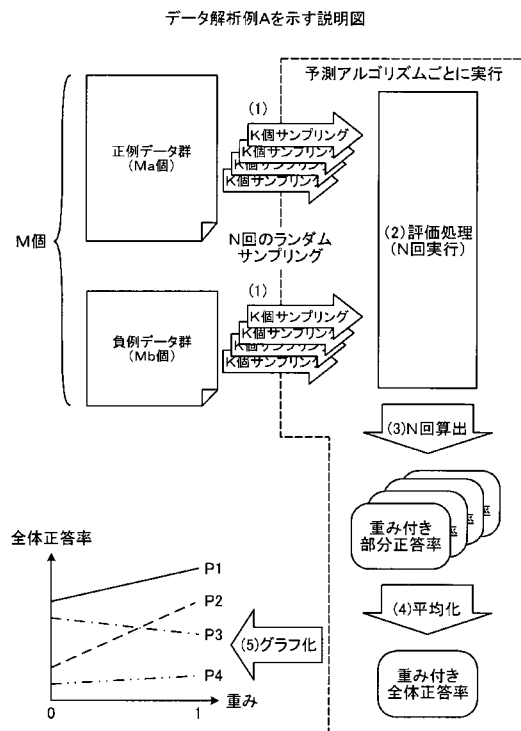
【0347】

- 4301 評価処理部
- 4302 平均化処理部
- 4303 出力部
- 4304 入力部
- 4305 選択部
- 4311 抽出部
- 4312 生成部
- 4313 第1の算出部
- 4314 第2の算出部
- 4315 第1の特定部
- 4316 第2の特定部
- 4317 演算部

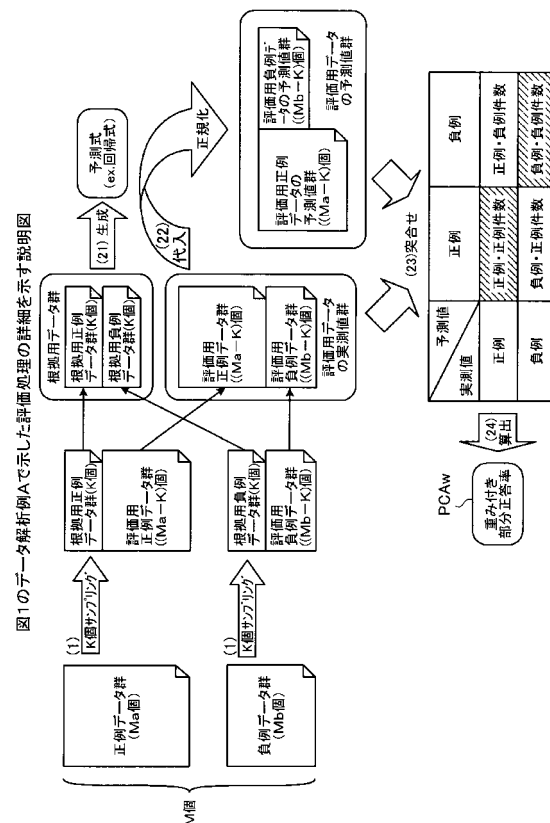
10

20

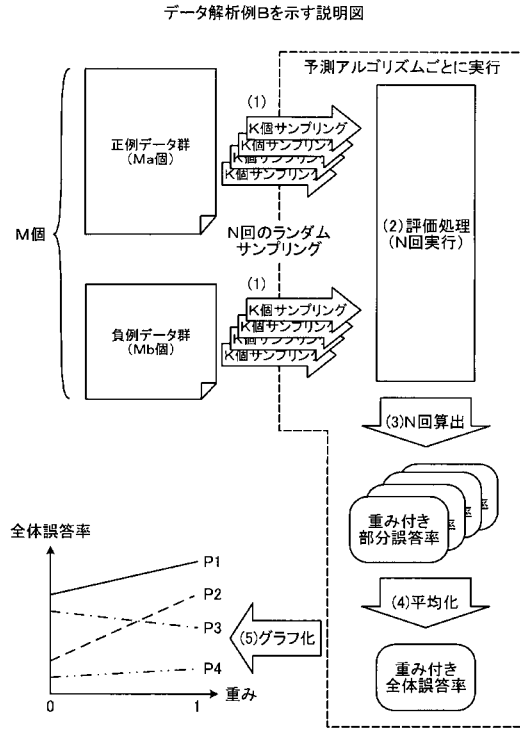
【図1】



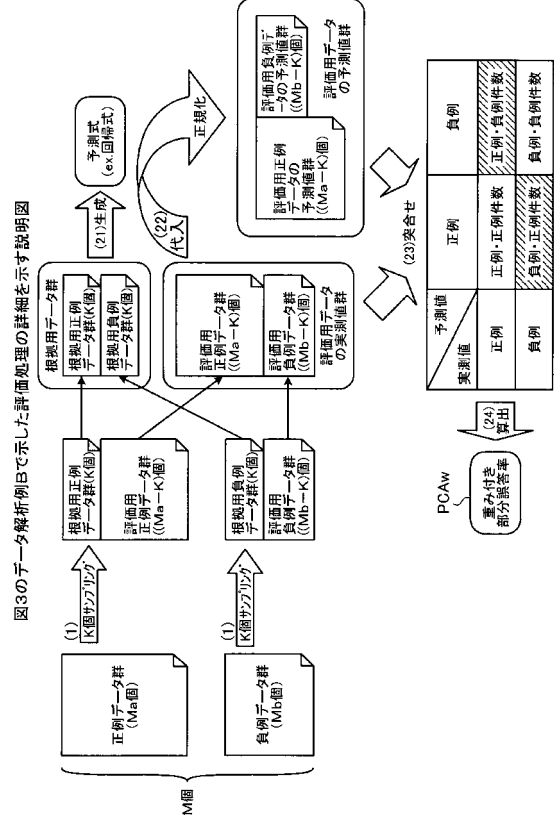
【図2】



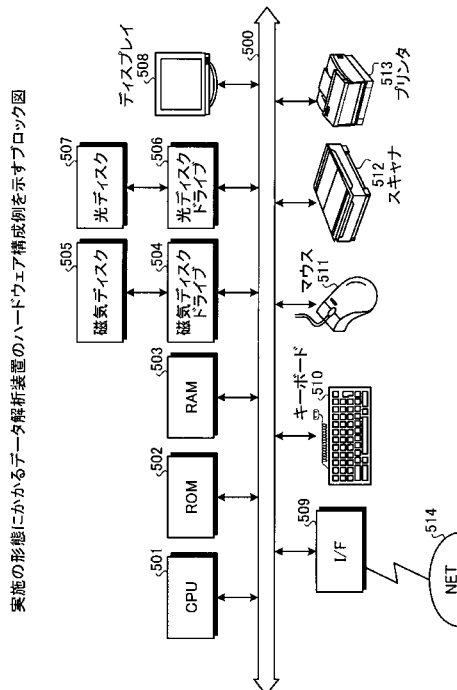
【図 3】



【図 4】



【図 5】



【図 6】

回路の故障解析に関するデータ群を示す説明図

サンプルID	間隔A	間隔B	間隔C	間隔D	故障フラグ
1	10	20	30	40	0
2	11	21	31	39	0
3	9	19	29	41	0
4	16	20	32	41	1
...	...	...	...	...	...
M	8	22	28	33	1

【図 7】

レコメンデーションに関するデータ群を示す説明図

サンプルID	年齢	性別	購入回数	平均購入金額	推薦商品の購入有無
1	23	0(女)	16	5000	1(購入)
2	34	1(男)	1	12000	0(購入せず)
3	47	0(女)	3	8000	0(購入せず)
4	51	1(男)	8	28000	1(購入)
...	...	...	...	...	...
M	29	0(女)	24	3700	1(購入)

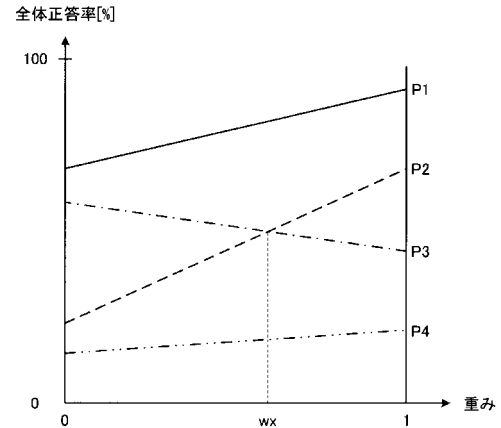
【図 8】

融資の与信に関するデータ群を示す説明図

サンプルID	年齢	性別	年収	世帯年収	返済の可否
1	23	0(女)	300	300	0(できなかった)
2	34	1(男)	560	820	0
3	47	0(女)	700	2000	1(できた)
4	51	1(男)	1000	1000	1
⋮	⋮	⋮	⋮	⋮	⋮
M	29	0(女)	410	960	0

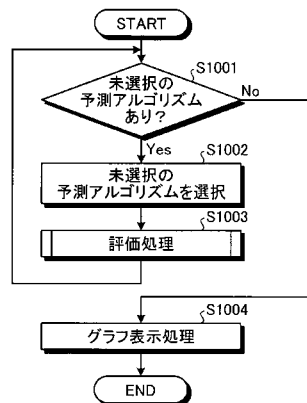
【図 9】

実施例1Aにおける予測アルゴリズムごとの重み付き全体正答率を示すグラフ



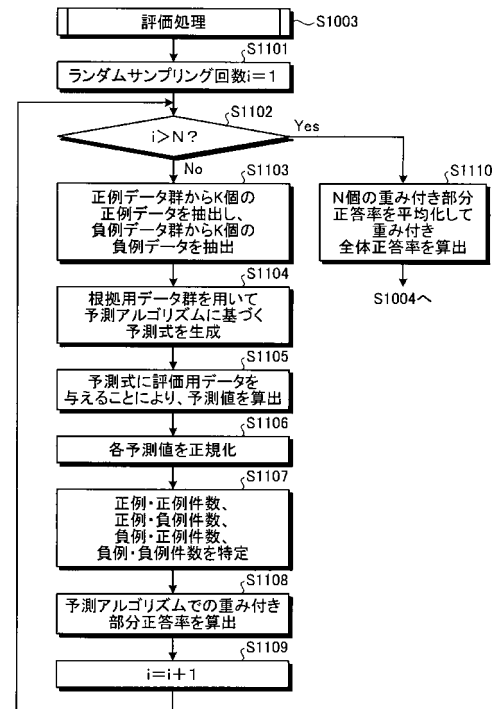
【図 10】

実施例1Aにかかるデータ解析処理手順を示すフローチャート



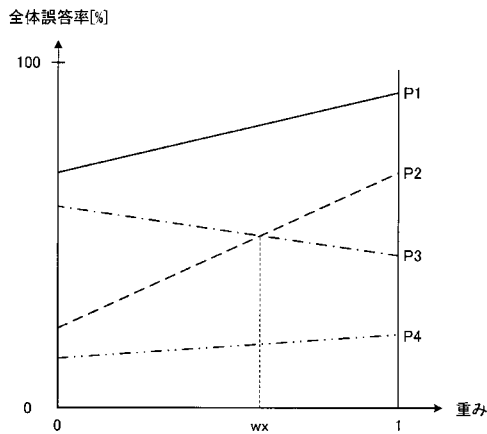
【図 11】

実施例1Aにかかる、図10に示した評価処理(ステップS1003)の詳細な処理手順を示すフローチャート



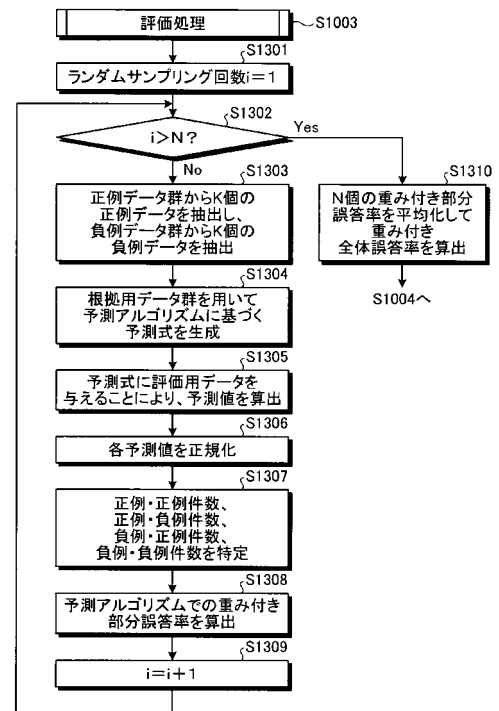
【図 1 2】

実施例1Bにおける予測アルゴリズムごとの重み付き全体誤答率を示すグラフ



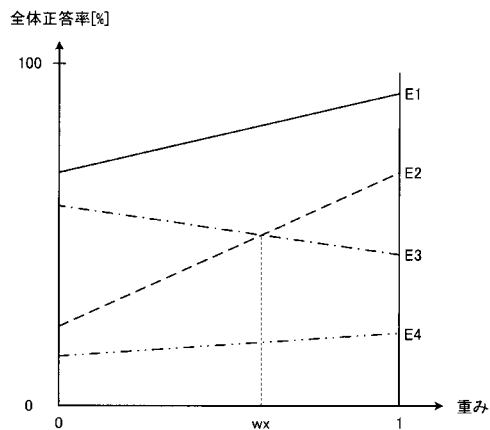
【図 1 3】

実施例1Bにかかる、図10Iに示した評価処理(ステップS1003)の詳細な処理手順を示すフローチャート



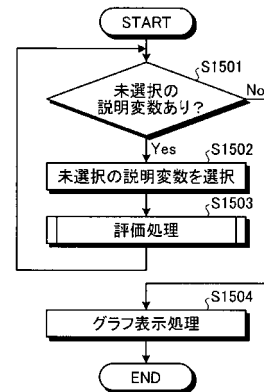
【図 1 4】

実施例2Aにおける説明変数ごとの重み付き全体正答率を示すグラフ



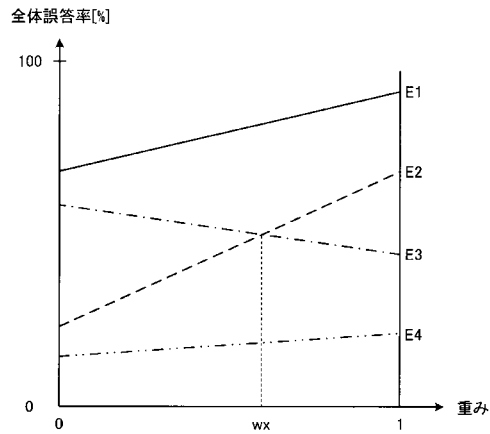
【図 1 5】

実施例2Aにかかるデータ解析処理手順を示すフローチャート



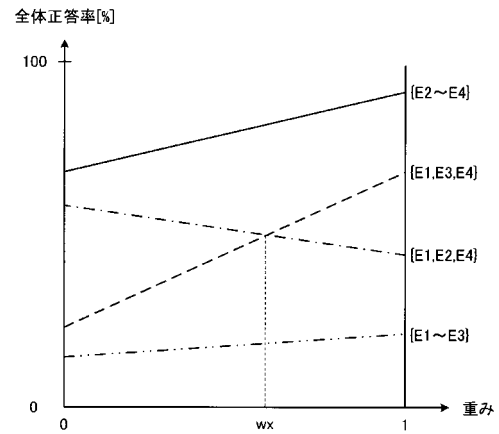
【図 16】

実施例2Bにおける説明変数ごとの重み付き全体誤答率を示すグラフ



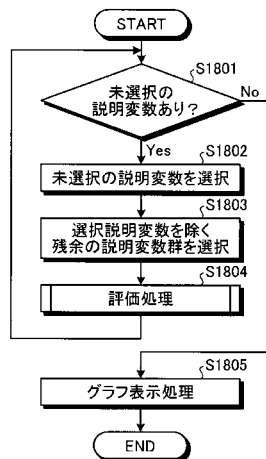
【図 17】

実施例3Aにおける残余の説明変数群ごとの重み付き全体正答率を示すグラフ



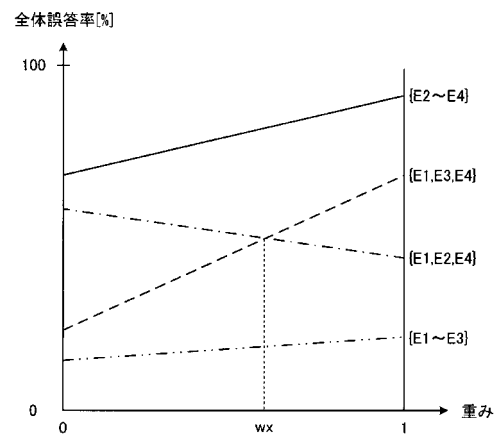
【図 18】

実施例3Aにかかるデータ解析処理手順を示すフローチャート



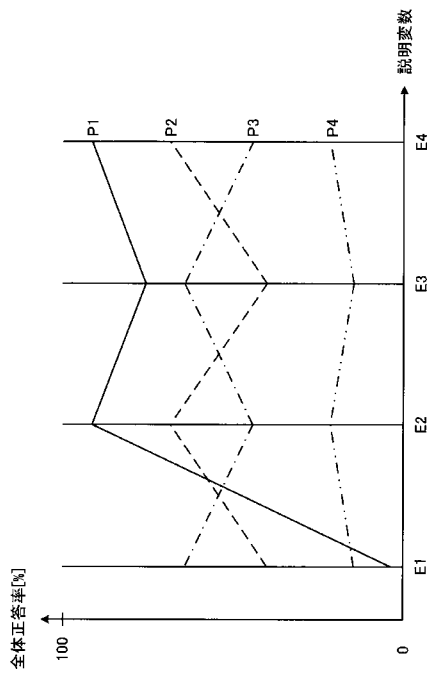
【図 19】

実施例3Bにおける説明変数ごとの重み付き全体誤答率を示すグラフ



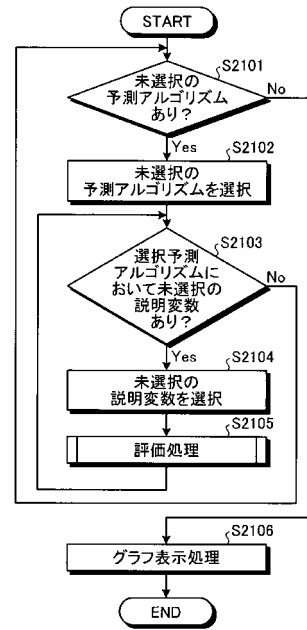
【図 20】

実施例4Aにおける予測アルゴリズムと説明変数との組み合わせごとの全体正答率を示すグラフ



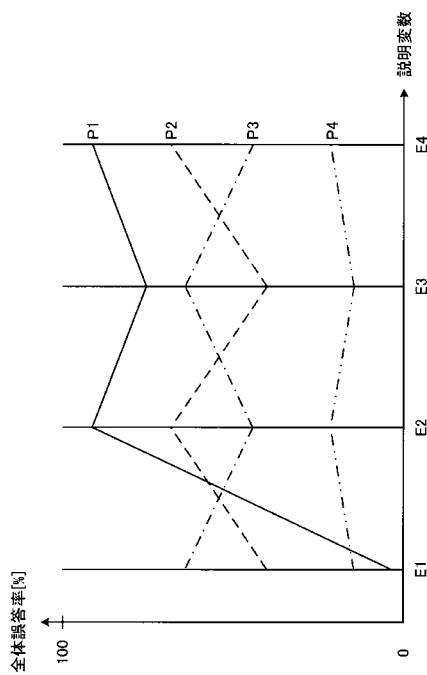
【図 21】

実施例4Aにかかるデータ解析処理手順を示すフローチャート



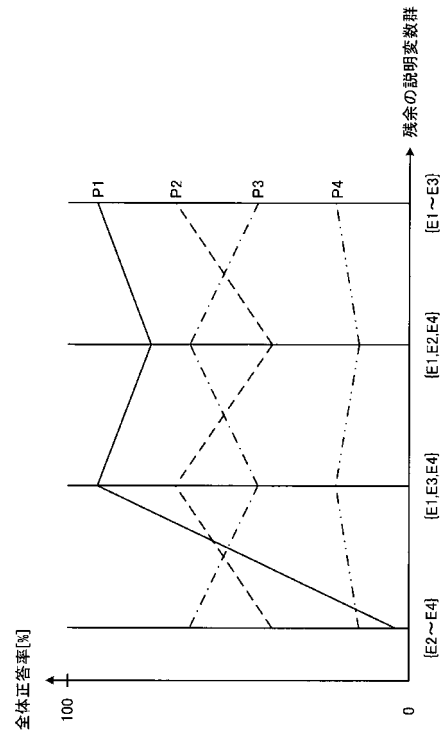
【図 22】

実施例4Bにおける予測アルゴリズムと説明変数との組み合わせごとの全体誤答率を示すグラフ



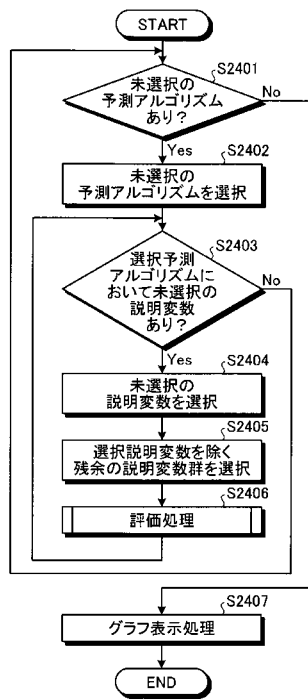
【図 23】

実施例5Aにおける予測アルゴリズムと残余の説明変数群との組み合わせごとの全体正答率を示すグラフ



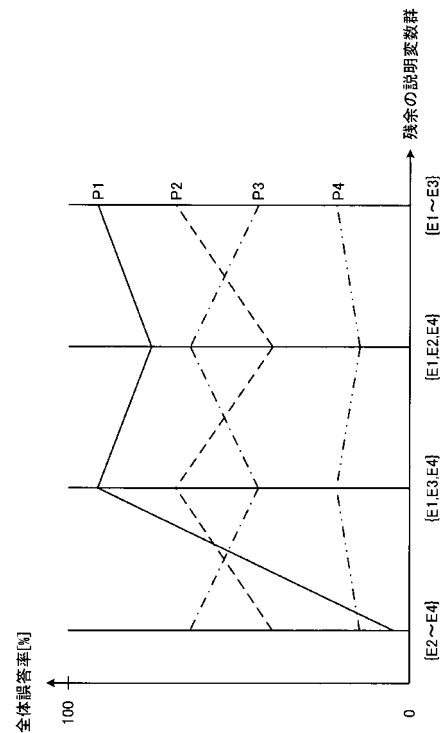
【図 24】

実施例5Aにかかるデータ解析処理手順を示すフローチャート



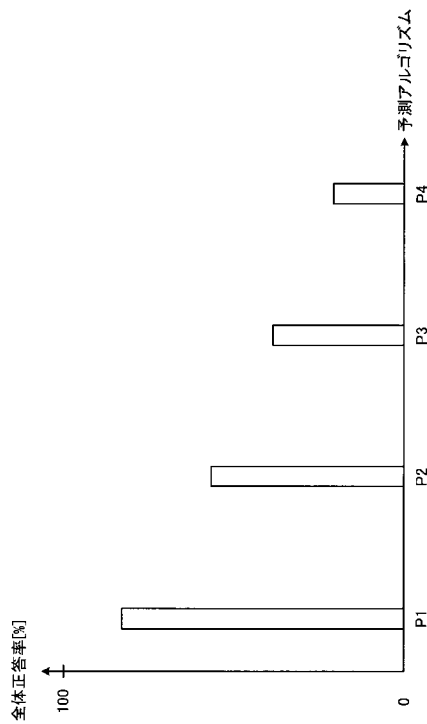
【図 25】

実施例5Bにおける予測アルゴリズムと残余の説明変数群との組み合わせごとの全体誤差率を示すグラフ



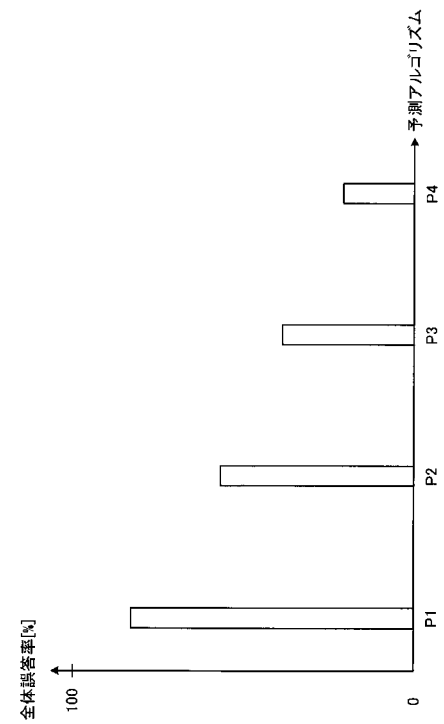
【図 26】

実施例6Aにおける予測アルゴリズムごとの全体正答率を示すグラフ

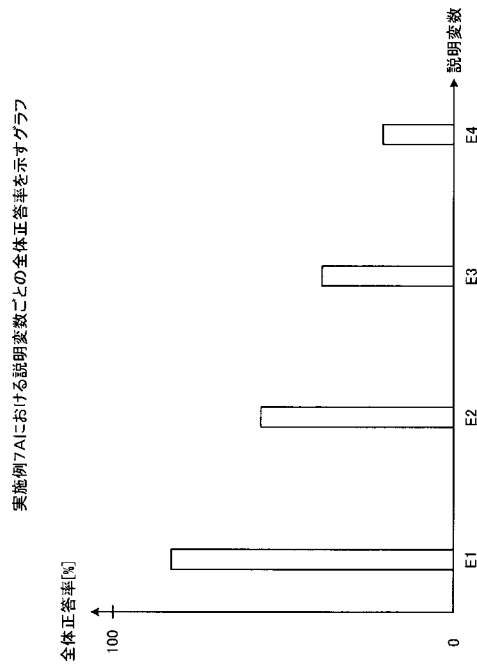


【図 27】

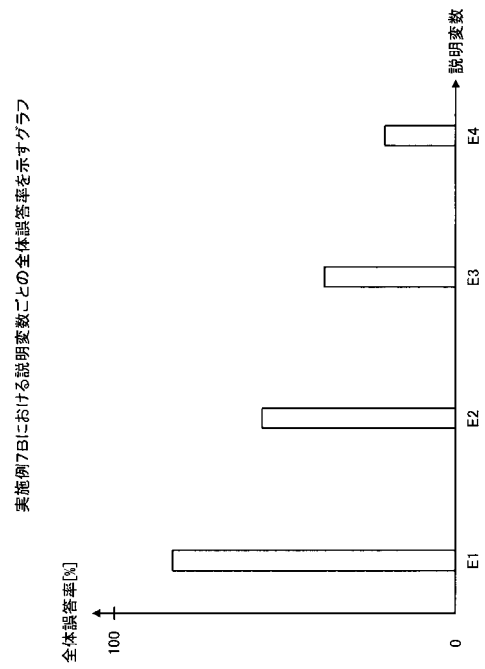
実施例6Bにおける予測アルゴリズムごとの全体誤差率を示すグラフ



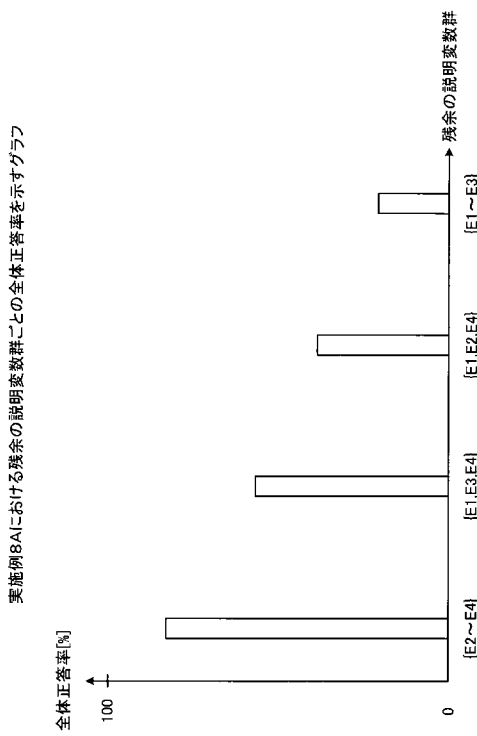
【図 28】



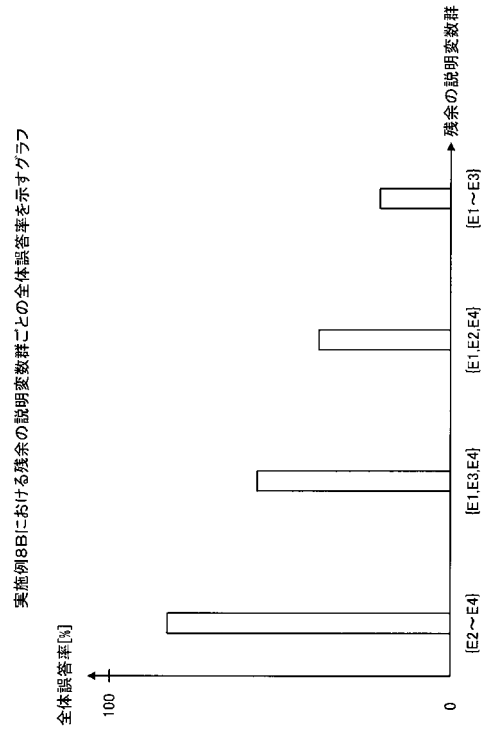
【図 29】



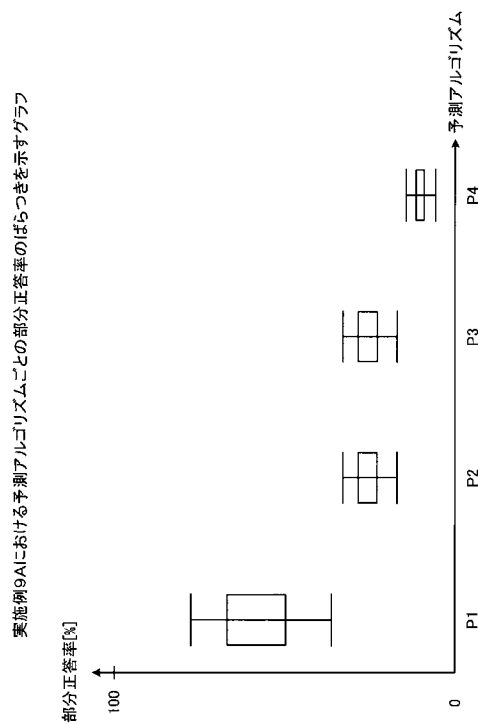
【図 30】



【図 31】

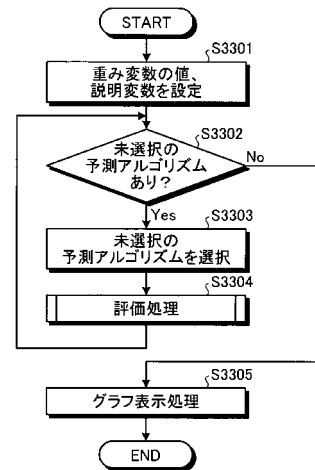


【図 3 2】



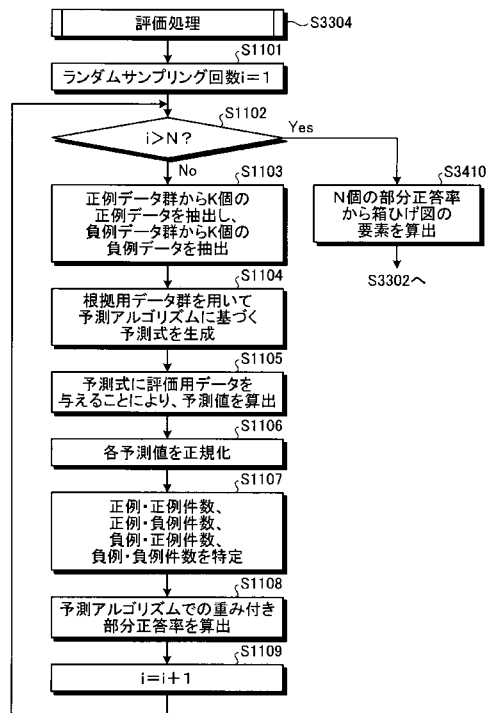
【図 3 3】

実施例9Aにかかるデータ解析処理手順を示すフローチャート



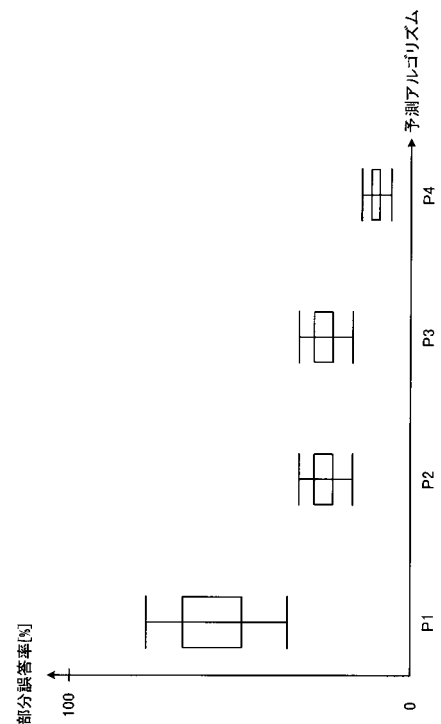
【図 3 4】

実施例9Aにかかる、図33に示した評価処理(ステップS3304)の詳細な処理手順を示すフローチャート



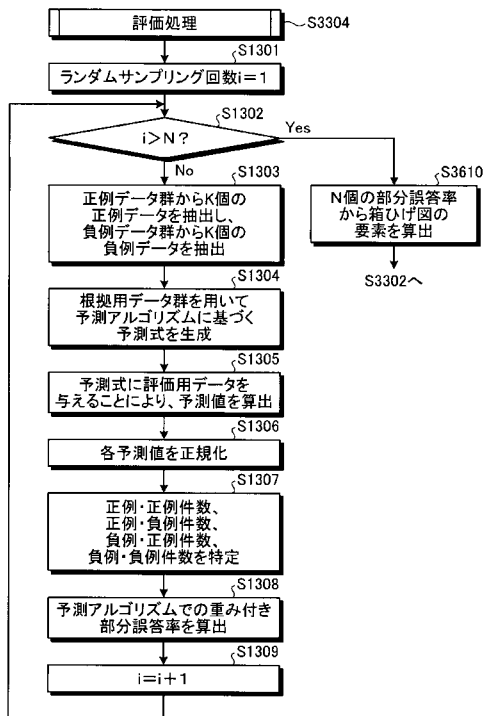
【図 3 5】

実施例9Bにおける予測アルゴリズムごとの部分誤答率のばらつきを示すグラフ



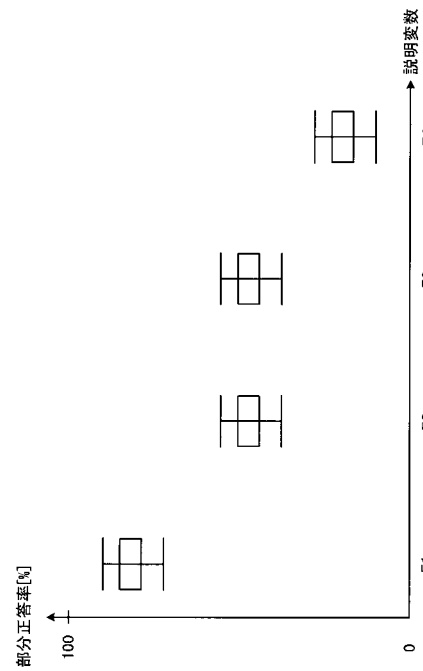
【図 36】

実施例9Bにかかる評価処理(ステップS3304)の
詳細な処理手順を示すフローチャート



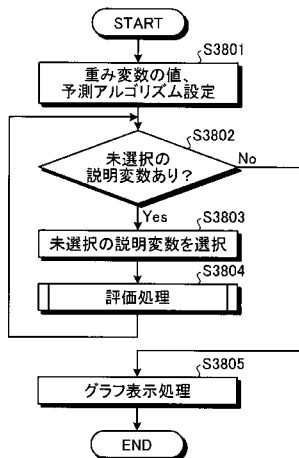
【図 37】

実施例10Aにおける説明変数ごとの部分正答率のばらつきを示すグラフ



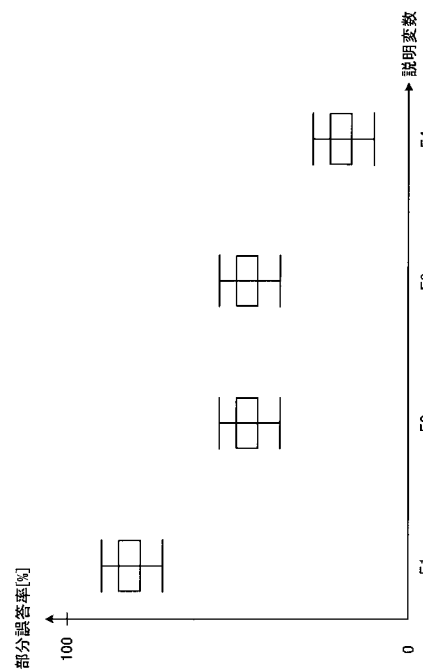
【図 38】

実施例10Aにかかるデータ解析処理手順を示すフローチャート

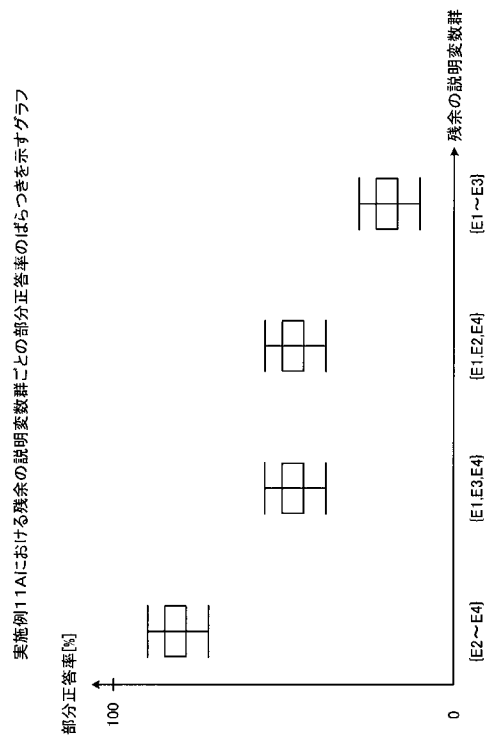


【図 39】

実施例10Bにおける説明変数ごとの部分誤答率のばらつきを示すグラフ

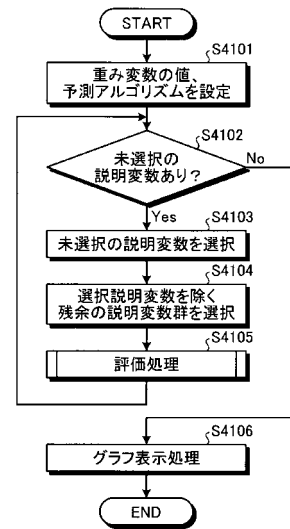


【図 40】

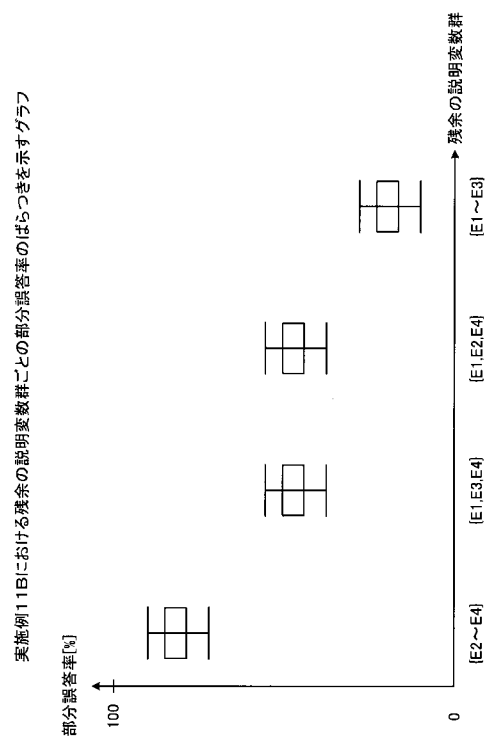


【図 41】

実施例11Aにかかるデータ解析処理手順を示すフローチャート



【図 42】



【図 43】

データ解析装置の機能的構成例を示すブロック図

