

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2015年4月2日(02.04.2015)



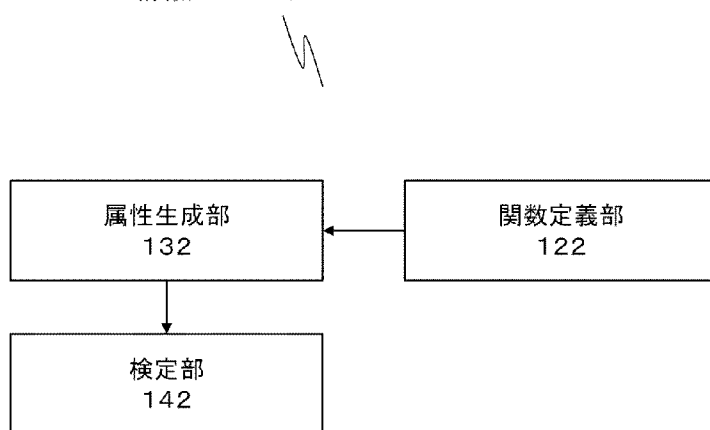
(10) 国際公開番号
WO 2015/045282 A1

- (51) 国際特許分類:
G06F 19/00 (2011.01)
- (21) 国際出願番号: PCT/JP2014/004520
- (22) 国際出願日: 2014年9月3日(03.09.2014)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
61/883,660 2013年9月27日(27.09.2013) US
- (71) 出願人: 日本電気株式会社(NEC CORPORATION) [JP/JP]; 〒1088001 東京都港区芝五丁目7番1号 Tokyo (JP).
- (72) 発明者: 森永 聡(MORINAGA, Satoshi); 〒1088001 東京都港区芝五丁目7番1号日本電気株式会社内 Tokyo (JP). 藤巻 遼平(FUJIMAKI, Ryohei); 〒1088001 東京都港区芝五丁目7番1号日本電気株式会社内 Tokyo (JP).
- (74) 代理人: 下坂 直樹(SHIMOSAKA, Naoki); 〒1088001 東京都港区芝五丁目7番1号日本電気株式会社内 Tokyo (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーロパ (AM, AZ, BY, KG, KZ, RU, TJ, TM), ユーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).
- 添付公開書類:
— 国際調査報告 (条約第21条(3))

(54) Title: INFORMATION PROCESSING SYSTEM, INFORMATION PROCESSING METHOD, AND RECORDING MEDIUM WITH PROGRAM STORED THEREON

(54) 発明の名称: 情報処理システム、情報処理方法およびプログラムを記憶する記録媒体

情報処理システム1002



1002 Information processing system
122 Function definition unit
132 Attribute generation unit
142 Evaluation unit

(57) Abstract: This invention helps improve the precision of data mining. This information processing device is provided with the following: a function-defining means that defines a new function by composing a plurality of functions; an attribute-generating means that applies said new function to an attribute to generate a new attribute that is the result of applying that function to that attribute; and a determining means that inputs the new attribute to an analysis engine, which executes an analysis process on the basis of the attribute, and determines whether or not information outputted by said analysis engine satisfies a prescribed requirement.

(57) 要約: データマイニングの精度向上に寄与する。情報処理装置は、複数の関数を合成することにより新しい関数を定義する関数定義手段と、前記新しい関数を、属性に対して適用することにより、属性に関数を適用した結果である新たな属性を生成する属性生成手段と、前記属性に基づき分析処理を実行する分析エンジンに、前記新たな属性を入力し、前記分析エンジンが出力する情報が所定の要件を満たすか否かを判定する判定手段と、を備える。

WO 2015/045282 A1

明 細 書

発明の名称：

情報処理システム、情報処理方法およびプログラムを記憶する記録媒体
技術分野

[0001] 本発明は、データマイニングを支援する技術に関する。

背景技術

[0002] データマイニングは、大量の情報の中から、これまで未知であった有用な知見を見つける技術である。データマイニングを用いて有用な知見が得られた実例として、大手スーパーマーケット・チェーンが所有する販売データを分析した例が知られている。販売データを分析した結果、「おむつを購入した顧客はビールも同時に購入する傾向がある」という知見が得られた。スーパーマーケット・チェーンは、当該知見を活かして、例えば、「おむつとビールとを同時に値下げしない」、などの措置をとることにより、売り上げの向上を図ることができる。

[0003] データマイニングを上述したような具体例に適用するプロセスは、下記に示す3つの段階に大別できる。

[0004] 1つ目の段階（工程）は、「前処理段階」である。「前処理段階」は、データマイニングアルゴリズムが効果的に機能するようにするために、データマイニングアルゴリズムに従って動作する装置などに入力する属性（feature）を加工することにより、その属性を新たな属性に変換する。

[0005] 2つ目の段階は、「分析処理段階」である。「分析処理段階」は、データマイニングアルゴリズムに従って動作する装置などに属性を入力し、係るデータマイニングアルゴリズムに従って動作する装置などの出力である分析結果を得る。

[0006] 3つ目の段階は、「後処理段階」である。「後処理段階」は、分析結果を、見やすいグラフや他の機器に入力するための制御信号等に変換する。

[0007] このように、データマイニングにより有用な知見を得るためには、「前処

理段階」が適切に行われる必要がある。「前処理段階」をどのような手順で実行すべきかを設計する作業は、分析技術の熟練技術者（データサイエンティスト）の知識に依存する。前処理段階の設計作業は、情報処理技術によって十分には支援されておらず、未だ熟練技術者の手作業による試行錯誤に依存する部分が多い。

- [0008] 非特許文献1は、データマイニングを実現するソフトウェアの一例を開示する。非特許文献1は、所望のタスク（分析処理）を実現するのに適した属性を選択することを支援する機能を提供する。この機能は、「属性選択（feature selection）」とも呼ばれる。

先行技術文献

非特許文献

- [0009] 非特許文献1：“WEKA”、[online]、[2013年9月5日検索]、インターネット<URL: <http://www.cs.waikato.ac.nz/ml/weka/>>

発明の概要

発明が解決しようとする課題

- [0010] オペレータが、非特許文献1が開示するソフトウェアを用いてデータマイニングを行う場合を想定する。この場合、オペレータは、必ずしも精度のよい分析結果を得ることができるとは限らない。なぜなら、非特許文献1が開示するソフトウェアは、精度のよい分析結果を得るための属性を、あらかじめ準備された属性のうちから選択するに過ぎないからである。このように、非特許文献1が開示するソフトウェアは、あらかじめ準備された属性の中から選択された解しか出力できないという制約がある。このため、あらかじめ準備された属性の中に精度のよい分析結果が得られる属性が含まれていないと、オペレータは、精度のよい分析結果を得ることができない。
- [0011] 本発明は、分析処理の精度向上に寄与する情報処理システム等を提供することを目的の1つとする。

課題を解決するための手段

[0012] 本発明の第１の側面は、複数の関数を合成することにより新しい関数を定義する関数定義手段と、前記新しい関数を、属性に対して適用することにより、属性に関数を適用した結果である新たな属性を生成する属性生成手段と、前記属性に基づき分析処理を実行する分析エンジンに、前記新たな属性を入力し、前記分析エンジンが出力する情報が所定の要件を満たすか否かを判定する検定手段と、を備える情報処理システムである。

[0013] 本発明の第２の側面は、複数の関数を記憶する関数記憶手段にアクセス可能なコンピュータを、複数の関数を合成することにより新しい関数を定義し、前記新しい関数を、属性に対して適用することにより、属性に関数を適用した結果である新たな属性を生成し、前記属性に基づき分析処理を実行する分析エンジンに、前記新たな属性を入力し、前記分析エンジンが出力する情報が所定の要件を満たすか否かを判定するよう制御する制御方法である。

[0014] 本発明の第３の側面は、複数の関数を記憶する関数記憶手段にアクセス可能なコンピュータに、複数の関数を合成することにより新しい関数を定義する処理と、前記新しい関数を、属性に対して適用することにより、属性に関数を適用した結果である新たな属性を生成する処理と、前記属性に基づき分析処理を実行する分析エンジンに、前記新たな属性を入力し、前記分析エンジンが出力する情報が所定の要件を満たすか否かを判定する処理と、を実行させるプログラムである。

[0015] また、本発明の目的は、上記のプログラムが格納されたコンピュータ読み取り可能な記憶媒体によっても達成される。

発明の効果

[0016] 本発明によれば、分析処理の精度向上に寄与する情報処理システム等を提供することができる。

図面の簡単な説明

[0017] [図１]図１は、本発明における第１の実施形態にかかる情報処理システム１００の構成を説明するブロック図である。

[図２]図２は、本発明における第１の実施形態にかかるデータセットの一例を

示す図である。

[図3]図3は、本発明における第1の実施形態にかかる関数記憶部110が記憶するデータの一例を示す図である。

[図4]図4は、本発明における第1の実施形態にかかる関数定義部120の動作を説明する図である。

[図5]図5は、本発明における第1の実施形態にかかる属性生成部130の詳細を説明する図である。

[図6]図6は、本発明における第1の実施形態にかかる検定部140の詳細を説明する図である。

[図7]図7は、本発明における第1の実施形態にかかる検定部140の詳細を説明する図である。

[図8]図8は、本発明における第1の実施形態にかかる検定部140の詳細を説明する図である。

[図9]図9は、本発明における第1の実施形態にかかる情報処理システム1000の動作を説明するフローチャートである。

[図10]図10は、本発明における第2の実施形態にかかる情報処理システム1001の構成を説明するブロック図である。

[図11]図11は、本発明における第2の実施形態にかかるデータセットの一例を示す図である。

[図12]図12は、本発明における第2の実施形態にかかる関数記憶部111が記憶するデータの一例を示す図である。

[図13]図13は、本発明における第2の実施形態にかかる関数定義部121の詳細を説明する図である。

[図14]図14は、本発明における第2の実施形態にかかる属性生成部131の詳細を説明する図である。

[図15]図15は、本発明における第2の実施形態にかかる検定部141の詳細を説明する図である。

[図16]図16は、本発明における第3の実施形態にかかる情報処理システム

1002の構成を説明するブロック図である。

[図17]図17は、本発明の各実施形態にかかる情報処理システムを実施可能なハードウェア構成の一例を示す図である。

発明を実施するための形態

[0018] はじめに、理解を容易にするため、本発明が適用され得る情報処理システム1000の詳細な説明に際して用いる用語を定義する。

[0019] (データセット)

「データセット」とは、情報処理システム1000に入力されるデータである。「データセット」は、1つまたは複数の属性を含む。「属性」は、「変量」と言い換えることもできる。

[0020] <関数(function)>

「関数」は、ある属性(feature)から新たな属性を生成(construct)する処理(processing)を定義する。「関数」は、データセットに含まれる属性に対して適用(apply)される。すなわち、「関数」をある属性に適用すると、ある属性に対して当該関数が定義する処理が実行され、その結果として新たな属性が生成される。

[0021] 言い換えると、「関数」は、属性に対して適用する演算を定義する。関数は、ある属性を他の属性に変換(transform)する処理を定義する、と言い換えてもよい。「関数」は、データセットに含まれる属性に対して適用する写像であってもよい。さらに言い換えると、関数は、その関数に関連付けられている上述の演算を表す。さらに言い換えると、関数は、その関数に関連付けられている上述の処理を表す。

[0022] 「関数」が定義する処理は、例えば、単項演算である。「関数」は、例えば、三角関数($\sin(X)$, $\cos(X)$, $\tan(X)$)、自然対数、絶対値または符号反転などの演算を定義する。「関数」は例えば、 $\log_n X$ 、 X^n など、パラメータ n を含む演算を定義してもよい。

[0023] 「関数」が定義する処理は、例えば、多項演算である。多項演算とは、複数の被演算子(オペランド)を持つ演算である。「関数」は、例えば、属性

Xと属性Yとの算術演算（足し算、引き算、かけ算など）を定義する。属性X及び属性Yが論理値である場合、「関数」は、例えば、属性Xのビット値と属性Yのビット値とに適用する論理演算（論理積(AND)、論理和(OR)、排他的論理和(XOR)など）を定義する。

[0024] 「関数」が定義する処理は、データに応じて処理が決まる「データに依存する処理」であってもよい。データに依存する処理の1つの具体例は、標準化(normalization)処理である。

[0025] 「データに依存する処理」を、具体例を挙げて説明する。例えば、100人分の名前の値と身長値とが関連づけられた情報を含むデータセットが、データマイニング装置に入力された場合を想定する。この場合、当該データセットには、「名前」という属性と、「身長」という属性との、2つの属性が含まれる。この例において、係る「名前」という属性は、100人分の名前の値を表す。「身長値」という属性は、100人分の身長値を表す。

[0026] データマイニング装置が、属性「身長」に対して、標準化処理を定義する関数を適用することにより、「標準化された身長」という新たな属性を生成する場合を想定する。この場合、データマイニング装置は、属性に含まれる1人分ずつのデータを、個別に標準化することはない。たとえば、データマイニング装置が、まずは、100人分の情報のうち1人目の情報「氏名：N、身長：174」のみを受け付けたとする。この場合、データマイニング装置は、1人目の情報に対する新たな属性「標準化された身長」を算出することはない。なぜなら、データマイニング装置は、100人分の情報が揃ってからでないと、標準化するパラメータとして必要な値（すなわち、100人分の「身長」の値の平均値、および、100人分の「身長」の標準偏差）を知り得ず、この結果、標準化するための関数が定まらないからである。

[0027] このような「データに依存する処理」の他の具体例としては、例えば、ヒストグラム生成、クラスタリング、及び、主成分分析等が挙げられる。

[0028] （関数の合成）

第1の関数が定義する処理と第2の関数が定義する処理とを、ある属性に

対して逐次的に適用することを、本願では「関数の合成」と記載する。例えば、第1の関数が $\sin(X)$ という関数を定義しており、第2の関数が X^2 という関数を定義している場合を想定する。第1の関数が定義する処理と、第2の関数が定義する処理とを合成すると、 $(\sin(X))^2$ という新しい関数または、 $\sin(X^2)$ という新しい関数が定義される。

[0029] このように、第1の関数と第2の関数とを合成すると、新しい第3の関数が定義される。この場合における第3の関数が定義する処理を説明する。対象とする属性に対して第3の関数が定義する処理が実行されると、以下に示すような新たな属性が生成される。すなわち、当該対象とする属性に対して、第1の関数が定義する処理と第2の関数が定義する処理とを逐次的に適用した場合に生成される新たな属性が、第3の関数の適用により生成される。

[0030] (分析エンジン)

「分析エンジン」は、属性に基づく分析処理である。すなわち、分析エンジンは、入力として属性を受け付け、該属性に基づき分析を行い(execute)、分析した結果を出力する。分析エンジンは、データマイニング装置が実行する分析アルゴリズムなどとも呼ばれる。分析エンジンは、例えば、回帰分析(Regression Analysis)、因子分析(Factor Analysis)、共分散構造分析(Covariance Structure Analysis)、主成分分析(Principal Factor Analysis)、判別分析(Discriminant Analysis)、カーネル分析、異種混合回帰分析、クラスター分析(Cluster Analysis)または異常検出などの処理を実行する分析エンジンである。「分析エンジンの種類の指定」とは、このような分析エンジンの種類の指定を受け付けることをいう。「分析エンジン」は、例えば、上述の分析処理を実行する主体(例えば装置)、又は、プロセッサが分析処理を実行するよう制御するプログラムなどを指すこともある。

[0031] (制約条件)

制約条件は、分析エンジンが出力する情報が満たすべき要件である。言い換えれば、制約条件は、分析エンジンが出力する分析結果が満たすべき要件である。分析エンジンの種類が単回帰分析である場合、制約条件の1つの具

体例は、「カイ二乗値が0.9以上」である。

[0032] (情報を取得する)

以降、情報を記憶装置から読み出すこと、情報を外部装置から受信すること、または、オペレータから情報の入力を受け付けることなどを、まとめて「情報を取得する」と記載する。

[0033] (情報を出力する)

以降、情報を記憶装置に書き込むこと、情報を外部装置へ送信すること、または、画面表示または音声などの形式でオペレータに対して情報を提示することなどを、まとめて「情報を出力する」と記載する。

[0034] 以下、上述した文言の定義を参酌しつつ、本発明の実施形態について、図面を参照して詳細に説明する。

[0035] <第1の実施形態>

第1の実施形態は、分析エンジンの種類として単回帰分析が指定された場合における、本発明の1つの具体例である。

[0036] 図1は、第1の実施形態にかかる情報処理システム1000の概要を説明するブロック図である。

[0037] 情報処理システム1000は、関数記憶部110と、関数定義部120と、属性生成部130と、検定部140と、出力部150と、を備える。

[0038] 関数記憶部110は、複数の関数を記憶することができる。関数記憶部110は、情報処理システム1000の内部に実装されていてもよいし、情報処理システム1000がアクセス可能な図示しない外部の装置に実装されていてもよい。

[0039] 関数定義部120は、関数記憶部110から複数の関数を取得する。関数定義部120は、取得した関数を合成することにより、新しい関数を定義する。

[0040] 属性生成部130は、対象とするデータセットを取得する。属性生成部130は、オペレータからデータセットの入力を受け付けてもよいし、図示しない記憶部からデータセットを読み出してもよい。属性生成部130は、情

報処理システム１０００の外部に備えられた図示しない装置から、データセットを受信してもよい。

[0041] 属性生成部１３０は、関数記憶部１１０があらかじめ記憶していた関数、または、関数定義部１２０が定義した関数を、データセットに含まれる属性に対して適用する。これにより属性生成部１３０は、属性に関数を適用した結果である新たな属性を生成する。

[0042] 検定部１４０は、分析エンジンの種類の指定および制約条件の指定を、例えばオペレータから、取得する。

[0043] 第１の実施形態においては、検定部１４０は、分析エンジンの種類として「単回帰分析」を取得する。また、検定部１４０は、データセットに含まれる複数の属性のうち、関数が予測する対象であるところの目的変数である属性の指定を取得する。

[0044] 検定部１４０は、単回帰分析エンジン（不図示）に、属性生成部１３０が生成する新たな属性を説明変数として入力する。検定部１４０は、単回帰分析エンジンが出力する回帰式を取得する。検定部１４０は、回帰式が制約条件を満たすか否かを判定(test)する。

[0045] 出力部１５０は、例えば、要件を満たす回帰式を出力する。

[0046] 以下、図１から図８までを用いて、関数記憶部１１０、関数定義部１２０、属性生成部１３０、検定部１４０および出力部１５０の詳細を説明する。

[0047] 図２は、図１に示す情報処理システム１０００に入力されるデータセットの一例を示す図である。図２に示すように、データセットは、例えば、複数人の、識別子（ＩＤ；Identifier）と、身長値と、体重値と、アイスクリームの年間消費量の値と、を関連付ける情報を含む。図２に示す、「身長」、「体重」および「アイスクリームの年間消費量」は、それぞれ、「属性」に相当する。

[0048] 図３は、図１に示す関数記憶部１１０が記憶する情報の一例を示す図である。図３に示すように、関数記憶部１１０には、複数の関数が記憶されている。

[0049] 図3に示すように、関数ID（識別子）が「関数1」である関数が定義する処理は、 X である。ここで、 X は恒等写像を表す。関数IDが「関数2」である関数が定義する処理は、 $\sin(X)$ である。ここで、 $\sin$ は正弦関数を表す。関数IDが「関数3」である関数が定義する処理は、 X^2 である。ここで、 X^2 は、 X の値を二乗する関数を表す。以下の説明において、関数を、その関数の関数IDによって表す。例えば、関数2は、関数IDが関数2である関数を表す。

[0050] 図1と図4とを用いて、図1に示す関数定義部120の詳細を説明する。図4は、関数定義部120が、図3に示す関数1ないし3を取得した場合に出力する、新しい関数4及び5を説明する図である。

[0051] 図4に示すように、関数定義部120は、関数1ないし3を取得して、新しい関数4及び5を生成する。

[0052] 関数定義部120は、例えば、関数2と関数3とを合成することにより、新しい関数4を定義する。図4に示すように、関数4が定義する処理は、 $(\sin(X^2))$ である。関数定義部120は、関数を合成する順序を入れ替えてもよい。関数定義部120は、例えば、関数2と関数3とを合成することにより、関数5を定義してもよい。図4に示すように、関数5が定義する処理は、 $(\sin(X))^2$ である。

[0053] 図1と図5とを用いて、図1に示す属性生成部130の詳細を説明する。図1に示すように、属性生成部130は、対象とするデータセットを取得する。属性生成部130は、目的変数である属性の指定を取得してもよい。

[0054] 例えば、属性生成部130が、目的変数である属性として「アイスクリームの年間消費量」という属性の指定を取得する場合を想定する。また、属性生成部130が、関数記憶部110から関数5（すなわち、 $(\sin(X))^2$ ）を取得する場合を想定する。属性生成部130は、係るデータセットに含まれる複数の属性のうち、目的変数として指定された属性以外の属性（すなわち、「身長」または「体重」）のうちから、関数に入力する属性を1つ選択する。

[0055] 属性生成部 130 は、例えば、「身長」という値を選択したとする。属性生成部 130 は、選択した属性「身長」に対して前記選択した関数 $(\sin(X))^2$ を適用して、新たな属性を生成する。この結果生成された新しい属性を、図 5 に示す。

[0056] 図 5 は、属性生成部 130 が、属性「身長」に対して関数 $(\sin(X))^2$ を適用することにより生成された、新たな属性を示す図である。

[0057] 属性生成部 130 は、例えば、属性を n 個受け付け、関数を m 個受け付けたとすると、 $n \times m$ 個の新たな属性を生成する。

[0058] 属性生成部 130 が、「身長」と「体重」という 2 つの属性を受け付け、関数 1 ないし 5 までの 5 つの関数を受け付けたとすると、属性生成部 130 は、 $2 \times 5 = 10$ 個の新しい属性を生成する。すなわち、属性生成部 130 は、以下に示す 10 個の新しい属性を生成する。

[0059] ・身長、
・ $(\text{身長})^2$ 、
・ $\sin(\text{身長})$ 、
・ $\sin(\text{身長}^2)$ 、
・ $(\sin(\text{身長}))^2$ 、
・体重、
・ $(\text{体重})^2$ 、
・ $\sin(\text{体重})$ 、
・ $\sin(\text{体重}^2)$ 、
・ $(\sin(\text{体重}))^2$ 。

[0060] ただし、属性生成部 130 は、必ずしも上述した 10 個の新しい属性のうち全てを生成する必要はない。属性生成部 130 は、生成した属性を出力する。

[0061] 図 1 に示す検定部 140 の詳細を、図 1、図 6、図 7 および図 8 を用いて説明する。以下の説明は、検定部 140 の動作の 1 つの具体例に過ぎず、検定部 140 の動作は限定的に解釈されない。

[0062] ここでは、検定部 140 は、分析エンジンの種類として「単回帰分析」を取得し、目的変数である属性として「アイスクリームの年間消費量」を取得し、制約条件として「カイ二乗値が 0.9 以上」という条件を取得したとする。

[0063] すなわち、検定部 140 は、 Y （アイスクリームの年間消費量） $= aX + b$ 、という式に従って回帰分析を行うことになる。ここで、 Y は目的変数である。 X は説明変数である。 a と b とは定数である。

[0064] 検定部 140 は、属性生成部 130 が生成する属性（説明変数）が、アイスクリームの年間消費量（目的変数）を、どの程度説明できるかについて分析する。

[0065] 検定部 140 は、属性生成部 130 が取得したデータセットに含まれる属性を取得する。また、検定部 140 は、属性生成部 130 が出力した属性を取得する。

[0066] 検定部 140 は、取得した複数の属性のうちから、一つの属性を選択する。例えば、検定部 140 は、「身長」という属性を選択したとする。

[0067] 図 6 は、検定部 140 が、「身長」という属性を説明変数として選択し、該説明変数に基づき単回帰分析を行った結果を表すグラフである。図 6 に示すように、単回帰分析の結果、 $a=0.0322$, $b=3.7137$ という結果が得られ、カイ二乗値は 0.031 であった。

[0068] 図 7 は、検定部 140 が、「 $(\sin(\text{身長}))^2$ 」という属性を説明変数として選択し、該説明変数に基づき単回帰分析を行った結果を表すグラフである。図 7 に示すように、単回帰分析の結果、 $a=11.179$, $b=3.0349$ という結果が得られ、カイ二乗値は 0.998 であった。

[0069] 検定部 140 は、取得した属性のそれぞれに対して、分析エンジン（上記の例では、単回帰分析エンジン）に属性を入力(input)する処理と、該分析エンジンが出力する分析結果（すなわち、回帰式とカイ二乗値）を取得する処理と、分析結果（すなわち、カイ二乗値）が制約条件を満たしているか否かを判定する処理と、を実行する。

[0070] 図8は、属性生成部130が生成した10種類の属性について、それぞれ検定部140が処理を実行した結果を説明する図である。図8に示すように、制約条件「カイ二乗値が0.9以上」を満たす説明変数は、「 $(\sin(\text{身長}))^2$ 」のみである。

[0071] 説明変数として「 $(\sin(\text{身長}))^2$ 」が選択された場合に、カイ二乗値が制約条件を満たすということは、すなわち、身長の値を正弦関数(sin)に代入して得られた値を二乗した値を用いて、 $Y=aX+b$ という関係式に従い、個人のアイスクリームの年間消費量を説明することができる、ということを表す。

[0072] これに対して図8の他の例に示すように、説明変数として他の属性が選択される場合に、カイ二乗値は、検定閾値を満たさない。これは、他の属性の値に基づき、 $Y=aX+b$ という関係式に従う場合に、個人のアイスクリームの年間消費量を説明することができない、ということを表す。

[0073] 出力部150は、例えば、要件を満たす回帰式を出力する。

[0074] 出力部150は、下記に示すように動作してもよい。例えば、例えば、以下に示すような属性Aを分析エンジンに入力して得られた分析結果が、制約条件を満たしているとする、

属性A：属性Bの値を正弦関数(sin)に代入して得られた値を二乗した値。

[0075] このとき出力部150は、「身長という属性の値を正弦関数(sin)に代入し、得られた値を更に二乗するような、前処理を実行すべきである」という情報を出力してもよい。あるいは、出力部150は、「身長という属性の値を正弦関数(sin)に代入し、得られた値を更に二乗した値を、指定された分析エンジンに入力すると、制約条件を満たす分析結果が得られる」という情報を出力してもよい。または、出力部150は、「身長という属性の値を正弦関数(sin)に代入し、得られた値を更に二乗した値」という情報を出力してもよい。出力部150は、これらの情報を、指定された分析エンジンの種類や、データセットのファイル名と共に出力してもよい。

[0076] 次に、第1の実施形態にかかる情報処理システム1000の動作を説明す

る。図9は、第1の実施形態にかかる情報処理システム1000の動作を説明するフローチャートである。

[0077] 関数定義部120は、関数記憶部110から関数を取得する（ステップS101）。関数定義部120は、取得した既存の関数を合成することによって新しい関数を定義する（ステップS102）。属性生成部130は、属性を新しい関数に入力し、該関数に従い出力される値を新たな属性として算出する。属性生成部130は、例えば関数と属性の全ての組み合わせについて、新たな属性を生成する（ステップS103）。ステップS103に示す動作は、取得した属性を関数に入力し、該関数に従い出力される値を新たな属性として算出する、と言い換えることもできる。

[0078] 検定部140は、複数の新たな属性から、特定の属性を選択する（ステップS104）。検定部140は、指定された目的変数を、特定の属性（説明変数）に基づき、どれくらい説明できるかを分析する。この結果、検定部140は、分析結果（すなわち、回帰式及び、カイ二乗値）を得る（ステップS105）。検定部140は、属性生成部130が生成した全ての属性について、ステップS105に示す動作を繰り返す（ステップS106）。

[0079] 検定部140は、制約条件を満たす分析結果が得られるか否かを検定する（ステップS107）。なお、ステップS104からステップS106までの繰り返しのうちにおいてステップS107に示す動作を実行してもよい。

[0080] 制約条件を満たす分析結果が得られた場合（ステップS107においてYES）、出力部150は、制約条件を満たす分析結果を出力する（ステップS108）。制約条件を満たす分析結果が得られない場合（ステップS107においてNO）、出力部150は、制約条件を満たす分析結果を出力しない。

[0081] 第1の実施形態にかかる情報処理システム1000が奏する作用効果を説明する。第1の実施形態によれば、分析処理の精度向上に寄与する情報処理システム1000を提供することができる。

[0082] その理由は、第1の実施形態にかかる属性生成部130が、属性に対して

関数を演算し、新たな属性を生成するからである。

[0083] かかる構成により、情報処理システム１０００は、「説明変数の候補である属性の数を増やす」ことができる。これは「仮説を検証するための属性の候補を増やす」ことができると言い換えることもできる。したがって本実施形態によれば、目的変数を十分に説明する説明変数が選択される可能性が高まり、データマイニングの精度が向上するという効果が実現する。

[0084] 上述した例において、オペレータ９００から入力された属性、すなわちデータセットに含まれる属性は、３種類（「身長」、「体重」、および、「アイスクリームの年間消費量」）である。上述した例においては、３種類の属性のうち１つ（すなわち、「アイスクリームの年間消費量」）は、目的変数として指定された。この場合、実質的な説明変数の候補は、アイスクリームの年間消費量以外の、２種類の属性（「身長」および「体重」）である。

[0085] 情報処理システム１０００は、上述したように、対象とするデータセットに含まれる２種類の属性と、関数記憶部１１０が記憶する関数（関数１ないし３）または関数定義部１２０が定義した関数（関数４または５）と、に基づいて、新たな１０個の属性を生成する。

[0086] このように、情報処理システム１０００は、説明変数の候補となる属性の数を増やすことにより、目的変数を十分に説明する属性を選択する可能性を高めるため、データマイニングの精度を向上することができる。

[0087] また、第１の実施形態にかかる関数定義部１２０は、複数の関数を合成することにより新しい関数を定義する。

[0088] かかる構成により、情報処理システム１０００は、あらかじめ準備された関数とは異なる関数を用いて、新たな属性を生成することができる。これにより、属性生成部１３０は、より多くの種類の属性を生成することができる。

[0089] また、第１の実施形態に係る情報処理システム１０００は、データマイニングの精度を向上するために、属性に対して実施すべき前処理の手順を出力することができる。その理由は、第１の実施形態にかかる出力部１５０が、

制約条件を満たす分析結果が得られた場合に、当該分析結果を得るために分析エンジンに入力した属性を出力するからである。または、出力部150が、制約条件を満たす分析結果を得るために、データセットに含まれる属性に対してどのような処理を行えばよいかを示す情報を出力するからである。

[0090] また、第1の実施形態に係る情報処理システム1000は、データ分析を行う分析技術者の工数を削減することができる。その理由は、第1の実施形態に係る情報処理システム1000の属性生成部130が、複数の属性に基づいて、新たな属性を生成するからである。そして、その情報処理システム1000の検定部140が、生成した新たな属性の中から、所定の基準を満たす属性を選択するからである。すなわち、検定部140は、例えば、生成した新たな属性を、入力された属性に基づき分析処理を実行する分析エンジンに入力する。そして、検定部140は、その分析エンジンが出力する情報が、所定の要件を満たすか否かを判定する。検定部140は、例えば、出力された情報が所定の要件を満たす場合、分析エンジンに入力された属性を選択する。前述の所定の要件（すなわち制約条件）は、例えば、目的変数に対する相関が、所定の基準より高いことである。すなわち、分析技術者が、複数の属性を情報処理システム1000に入力すれば、情報処理システム1000は、目的変数と相関の高い属性を自動的または半自動的に生成することができる。

[0091] 具体的には、例えば、第1の実施形態に係る情報処理システム1000によれば、分析技術者は、「個人のアイスクリームの年間消費量」と「 $(\sin(\text{身長}))^2$ 」との間に強い相関があるということを知らなくても、精度の良い分析結果を得ることができる。その理由は、情報処理システム1000が、「身長」という属性に基づいて、「 $(\sin(\text{身長}))^2$ 」という新たな属性を生成するからである。言い換えると、分析技術者が、「身長」という属性を情報処理システム1000に入力すれば、情報処理システム1000は、「 $(\sin(\text{身長}))^2$ 」という、目的変数と相関の高い属性を、ユーザにとって自動的または半自動的に生成することができる。

[0092] また、第1の実施形態に係る情報処理システム1000によれば、データ分析を行う分析技術者は、目的変数と、新たに生成される属性との間に、強い相関があることに気付くことができる。例えば、データ分析を行う分析技術者は、「個人のアイスクリームの年間消費量」と「 $(\sin(\text{身長}))^2$ 」との間に強い相関があるということに気が付くことができる。

[0093] (第1の実施形態の変形例)

関数定義部120は、関数記憶部110から、連続値パラメータ n を含む演算子を読み出して、 n に任意の値を代入することにより、新たな関数を定義してもよい。連続値パラメータ n を含む演算子は、例えば、 $\log_n X$ または X^n などである。関数定義部120が、例えば、 $\log_n X$ を定義する関数を読み出した場合、関数定義部120は、例えば、 $\log_2 X$ 、 $\log_3 X$ 、または、 $\log_5 X$ などの新しい関数を定義する。

[0094] 検定部140は、例えば、分析エンジンの種類として、重回帰分析の指定を受け付けてもよい。例えば、検定部140が、重回帰分析($Z = aX + bY + c$)の指定を受け付けるとする。ここで、 Z は目的変数である。 X は第1の説明変数である。 Y は第2の説明変数である。 a 、 b および c は、それぞれ定数である。

[0095] 検定部140は、例えば、属性生成部130から10個の属性を取得するとする。この場合、第1の説明変数 X と第2の説明変数 Y の選択の仕方の組み合わせは、45($= (10 \times 9) \div 2$)通りとなる。検定部140は、45通りの説明変数の組み合わせについて、図9に示したステップS104からステップS106に示した動作を繰り返す。

[0096] また検定部140は、分析エンジンの種類として曲線回帰分析を受け付けてもよい。この場合、検定部140は、曲線の種類、例えば、指数関数またはガウス関数などの指定を受け付ける。

[0097] 上述の変形例は、他の実施形態にも適用可能である。

[0098] <第2の実施形態>

第2の実施形態は、分析エンジンの種類として判別分析が指定された場合

における、本発明の１つの具体例である。

[0099] 図１０は、第２の実施形態にかかる情報処理システム１００１の構成を表わすブロック図である。図１０に示すように、第２の実施形態に係る情報処理システム１００１は、以下の構成を備え得る。

[0100] ・第１の実施形態にかかる関数記憶部１１０に代えて関数記憶部１１１を備える。

[0101] ・関数定義部１２０に代えて関数定義部１２１を備える。

[0102] ・属性生成部１３０に代えて属性生成部１３１を備える。

[0103] ・検定部１４０に代えて検定部１４１を備える。

[0104] 第１の実施形態と第２の実施形態とは、扱うデータセット、および指定される分析エンジンの種類が異なる。

[0105] 図１１は、図１０に示す情報処理システム１００１に入力されるデータセットの一例を説明する図である。図１１に示すデータセットは、多変量データと言い換えることもできる。図１１に示すように、データセットは、複数人の識別子の各々に対して、属性１ないし属性４を関連付ける情報を含む。図１１に示すデータセットは、例えば複数人分のアンケートの回答結果を表すデータである。各属性は、アンケートに含まれる質問事項に対する回答である。属性１ないし属性４の内容を、下記に示す。具体的には、各属性の、質問事項と、回答が表す値とを示す。

[0106] 属性１：犬と猫どちらが好き？ （犬を０と表す、猫を１と表す）

、

属性２：年齢は？ （４０歳以上を０と表す、４０歳未満を１と表す）、

属性３：性別は？ （男を０と表す、女を１と表す）

、

属性４：寿司と天麩羅どちらが好き？ （寿司を０と表す、天麩羅を１と表す）。

[0107] 図１２は、図１０に示す関数記憶部１１１が記憶する情報の一例を示す図

である。図 1 2 に示すように、関数記憶部 1 1 1 は、関数 1 ないし 4 を記憶している。関数 1 は、恒等写像 X を定義する。関数 2 は、2 つの属性の値の論理積 (AND) 演算を定義する。関数 3 は、2 つの属性の値の論理和 (OR) 演算を定義する。関数 4 は、ある属性の値の否定 (NOT) を定義する。

[0108] 図 1 0 に示す関数定義部 1 2 1 の詳細を、図 1 3 に示す例を用いて説明する。図 1 3 は、関数定義部 1 2 1 が、関数 1 ~ 4 を組み合わせて新しく定義した関数 5 を示す図である。関数 5 は排他的論理和 (XOR) を定義する。

[0109] 図 1 3 に示すように、関数定義部 1 2 1 は、関数 1 ないし 4 を組み合わせて、新しい関数を定義する。関数 1 ないし 4 の組み合わせ方には様々なバリエーションが考えられる。図 1 3 に示す一例は、組み合わせ方のバリエーションの 1 つである。図 1 3 は、関数 2 (AND) と、関数 3 (OR) と、関数 4 (NOT) を組み合わせることにより定義された関数 5 (XOR) を表す図である。関数定義部 1 2 1 は、関数 1 ないし 4 を組み合わせることにより、例えば、否定論理積 (NAND) や否定論理和 (NOR) などの新しい関数を定義してもよい。

[0110] 図 1 0 に示す属性生成部 1 3 1 の詳細を、図 1 4 に示す例を用いて説明する。図 1 4 は、属性生成部 1 3 1 が生成した新しい属性に関する 1 つの具体例を説明する図である。

[0111] 属性生成部 1 3 1 は、関数定義部 1 2 1 が定義した複数の新しい関数から、1 つの関数を選択する。属性生成部 1 3 1 は、入力されたデータセットに含まれる複数の属性から、1 つの属性または属性の組み合わせを選択する。例えば、属性生成部 1 3 1 が、関数として「否定論理積 (NAND)」を選択し、属性として、属性 1 および属性 2 を選択した場合を想定する。この結果、属性生成部 1 3 1 が生成する新しい属性を、図 1 4 に示す。

[0112] 属性生成部 1 3 1 は、例えば、関数定義部 1 2 1 が定義した新しい関数全てに対して、新しい属性を生成する。属性生成部 1 3 1 は、必ずしも新しい関数全てに対して、新しい属性を生成しなくてもよい。

- [0113] 図10を参照する説明に戻る。ここでは、検定部141は、分析エンジンの種類として「判別分析」を指定されたとする。さらに、検定部141は、目的変数として属性4（すなわち、「寿司と天麩羅どちらが好きか」）を指定されたとする。
- [0114] 検定部141は、制約条件（すなわち、分析エンジンが出力する情報が満たすべき要件）として、「一致率が95%以上」という条件を取得するとする。ここで、「一致率」とは、選択された属性の値と、予測対象として指定された属性の値とが、どの程度一致しているかを表す指標である。
- [0115] 検定部141は、属性生成部131が生成した新たな属性に基づき、「寿司と天麩羅どちらが好きか」を、を十分に説明できるかを分析する。
- [0116] 検定部141の詳細を説明する。検定部141は、属性生成部131が生成した新たな属性を取得する。検定部141は、取得した複数の属性から、一つの属性を選択する。例えば、検定部141は、「属性3」という属性を選択したとする。
- [0117] 検定部141は、選択された属性の値と、予測対象として指定された属性の値との、一致率を算出する。
- [0118] 図11を参照すると、図示した13人分のデータにおいて、属性3の値と属性4の値が一致するのは、5人分のデータである。よって、図示した13人分のデータにおいて、属性3の値と属性4の値の一致率は0.38（ $= 5 \div 13$ ）である。何人分のデータに対して一致率を算出するかは、例えば、予め指定されていてもよい。
- [0119] 検定部141は、取得した全ての属性に対して、目的変数「寿司と天麩羅どちらが好きか」の値との一致率を算出する。
- [0120] 図15は、属性生成部131が生成した属性について、検定部140が処理を実行した結果を説明する図である。図15に示すように、属性1と属性3とに対して、排他的論理和（XOR）を施した値と、属性4の値との一致率が100%であり、制約条件を満たす。これはつまり、「寿司」と「天麩羅」の好みは、アンケート結果における「属性1」と「属性3」との排他的

論理和X O Rの値に基づき、説明できることを表す。

[0121] 第2の実施形態にかかる情報処理システム1001が奏する作用効果を説明する。第2の実施形態によれば、分析処理の精度向上に寄与する情報処理システム1001を提供することができる。

[0122] その理由は、第2の実施形態にかかる属性生成部131が、属性に対して関数を適用し、新たな属性を生成するからである。

[0123] かかる構成により、情報処理システム1001は、「説明変数の候補である属性の数を増やす」ことができる。これは「仮説を検証するための属性の候補を増やす」ことができると言い換えることもできる。本実施形態によれば、目的変数を十分に説明する説明変数が選択される可能性が高まり、データマイニングの精度が向上するという効果が実現する。

[0124] また、第2の実施形態にかかる関数定義部121は、複数の関数を合成することにより新しい関数を定義する。

[0125] かかる構成により、情報処理システム1001は、あらかじめ準備された関数とは異なる関数を用いて、新たな属性を生成することができる。これにより、属性生成部131は、より多くの種類の属性を生成することができる。

[0126] また、第2の実施形態に係る情報処理システム1001は、データマイニングの精度を向上するために、属性に対して実施すべき前処理の手順を出力することができる。その理由は、第2の実施形態にかかる出力部150が、制約条件を満たす分析結果が得られた場合に、当該分析結果を得るために分析エンジンに入力した属性を出力するからである。または、出力部150が、制約条件を満たす分析結果を得るために、データセットに含まれる属性に対してどのような処理を行えばよいかを示す情報を出力するからである。

[0127] <第3の実施形態>

図16は、第3の実施形態にかかる情報処理システム1002の構成を説明するブロック図である。図16に示すように、情報処理システム1002は、関数定義部122と、属性生成部132と、検定部142と、を備える

。

[0128] 関数定義部 1 2 2 は、複数の関数を合成することにより新しい関数を定義する。

[0129] 属性生成部 1 3 2 は、新しい関数を、属性に対して適用し、属性に関数を適用した結果である新たな属性を定義する。

[0130] 検定部 1 4 2 は、分析エンジンの選択を受け付け、分析エンジンが出力する情報が満たす要件の入力を受け付け、前記選択された分析エンジンに前記新たな属性を入力し、前記分析エンジンが出力する情報を取得し、前記取得した情報が前記要件を満たすか否かを判定する。

[0131] 第 3 の実施形態によれば、分析処理の精度向上に寄与する情報処理システム 1 0 0 2 を提供することができる。

[0132] <情報処理システムのハードウェア構成>

図 1 7 に示した情報処理システム（コンピュータ）1 0 0 0 を構成するハードウェアは、CPU（Central Processing Unit）1、メモリ 2、記憶装置 3、通信インターフェース（I/F）4 を備える。情報処理システム 1 0 0 0 は、入力装置 5 または出力装置 6 を備えていてもよい。情報処理 1 0 0 の機能は、例えば CPU 1 が、メモリ 2 に読み出されたコンピュータプログラム（ソフトウェアプログラム、以下単に「プログラム」と記載する）を実行することにより実現される。実行に際して、CPU 1 は、通信インターフェース 4、入力装置 5 および出力装置 6 を適宜制御する。

[0133] 尚、本実施形態および後述する各実施形態を例として説明される本発明は、係るプログラムが格納されたコンパクトディスク等の不揮発性の記憶媒体 8 によっても構成される。記憶媒体 8 が格納するプログラムは、例えばドライブ装置 7 により読み出される。

[0134] 情報処理システム 1 0 0 0 が実行する通信は、例えば OS（Operating System）が提供する機能を使ってアプリケーションプログラムが通信インターフェース 4 を制御することによって実現される。入力装置

5は、例えばキーボード、マウスまたはタッチパネルである。出力装置6は、例えばディスプレイである。情報処理システム1000は、2つ以上の物理的に分離した装置が、有線、無線、又はそれらの組み合わせにより、通信可能に接続されることによって構成されていてもよい。

[0135] 図17に示すハードウェア構成例は、前述した各実施形態にも適用可能である。なお、情報処理システム1000は専用の装置であってもよい。なお、情報処理システム1000およびその各機能ブロックのハードウェア構成は、上述の構成に限定されない。

[0136] <その他の変形例>

分析エンジンは、必ずしも情報処理システム1000と同一の装置に実装される必要はない。分析エンジンは、情報処理システム1000からアクセス可能であればよい。上述の変形例は、他の実施形態にも適用可能である。

[0137] 以上、分析エンジンの種類として単回帰分析、重回帰分析、および、判別分析を指定された場合を例に、本発明を説明した。

[0138] 本発明は上述した各実施の形態に限定されず、様々な態様で実施されることが可能である。本発明は、上記各実施形態に例示した種類以外の分析エンジンを用いるデータマイニングにも適用され得る。

[0139] また、上述した各実施の形態は、適宜組み合わせて実施されることが可能である。また、本発明は、上述した各実施の形態に限定されず、様々な態様で実施されることが可能である。

[0140] 各ブロック図に示したブロック分けは、説明の便宜上から表された構成である。各実施形態を例に説明された本発明は、その実装に際して、各ブロック図に示した構成には限定されない。

[0141] 以上、本発明を実施するための形態について説明したが、上記実施の形態は本発明の理解を容易にするためのものであり、本発明を限定して解釈するためのものではない。本発明はその趣旨を逸脱することなく変更、改良され得ると共に、本発明にはその等価物も含まれる。

[0142] この出願は、2013年9月27日に提出された米国出願US 61/88

3660を基礎とする優先権を主張し、その開示の全てをここに取り込む。

産業上の利用可能性

[0143] 上述した実施形態を例に説明した本発明は、例えばデータマイニングを支援するツールに用いることができる。

符号の説明

- [0144]
- 1 CPU
 - 2 メモリ
 - 3 記憶装置
 - 4 通信インターフェース
 - 5 入力装置
 - 6 出力装置
 - 7 ドライブ装置
 - 8 記憶媒体
 - 110 関数記憶部
 - 111 関数記憶部
 - 120 関数定義部
 - 121 関数定義部
 - 122 関数定義部
 - 130 属性生成部
 - 131 属性生成部
 - 132 属性生成部
 - 140 検定部
 - 141 検定部
 - 142 検定部
 - 150 出力部
 - 900 オペレータ
 - 1000 情報処理システム
 - 1001 情報処理システム

1 0 0 2 情報処理システム

請求の範囲

- [請求項1] 複数の関数を合成することにより新しい関数を定義する関数定義手段と、
前記新しい関数を、属性に対して適用することにより、属性に関数を適用した結果である新たな属性を生成する属性生成手段と、
前記属性に基づき分析処理を実行する分析エンジンに、前記新たな属性を入力し、前記分析エンジンが出力する情報が所定の要件を満たすか否かを判定する検定手段と、
を備える情報処理システム。
- [請求項2] 前記検定手段は、分析エンジンの選択を受け付け、その分析エンジンが出力する情報が満たす要件の入力を受け付け、前記選択された分析エンジンに前記新たな属性を入力する、
請求項1に記載の情報処理システム。
- [請求項3] 前記関数定義手段は、第1の関数と第2の関数とを取得し、前記第1の関数と前記第2の関数とを合成することにより、第3の関数を定義し、
前記第3の関数が定義する処理は、前記属性に対して、前記第1の関数が定義する処理と前記第2の関数が定義する処理とを逐次的に実行する処理である、
請求項1または2に記載の情報処理システム。
- [請求項4] 前記関数定義手段は、前記新しい関数を複数定義し、
前記属性生成手段は、前記複数の新しい関数のうちそれぞれの新しい関数に基づいて、複数の前記新たな属性をそれぞれ生成し、
前記検定手段は、前記分析エンジンに、複数の前記新たな属性のうち特定の属性を入力し、前記分析エンジンが出力する情報を取得し、前記分析エンジンが出力する情報が所定の要件を満たすか否かを判定する、
請求項1から3のいずれかに記載の情報処理システム。

- [請求項5] 前記検定手段は、複数の前記の新たな属性のそれぞれに対して、
前記分析エンジンに複数の前記新たな属性のうち特定の属性を入力
する処理と、
前記分析エンジンが出力する情報を取得する処理と、
前記取得した情報が所定の要件を満たすか否かを判定する処理と、
を実行する、
請求項4に記載の情報処理システム。
- [請求項6] 前記分析エンジンが出力する情報のうち、前記要件を満たす情報を
出力する、第1の出力手段を更に備える、請求項1から5のいずれか
に記載の情報処理システム。
- [請求項7] 前記分析エンジンが出力する情報が前記要件を満たした場合に、前
記分析エンジンが出力する情報を得るために当該分析エンジンに入力
された属性か、または、当該属性を生成するために、前記属性生成手
段が適用した関数および前記関数を適用した属性を出力する、第2の
出力手段を更に備える、請求項1から5のいずれかに記載の情報処理
システム。
- [請求項8] 前記関数定義手段は、複数の関数または写像を合成することによっ
て、前記新しい関数を定義する、
請求項1から7のいずれかに記載の情報処理システム。
- [請求項9] 前記検定手段は、分析エンジンとして回帰分析が選択された場合に
、更に、目的変数として前記属性のうちいずれかの属性の指定を受け
付け、前記要件として、説明変数の個数の指定を受け付ける、
請求項1から8のいずれかに記載の情報処理システム。
- [請求項10] 複数の関数を記憶する関数記憶手段にアクセス可能なコンピュータ
を、
前記複数の関数を合成することにより新しい関数を定義し、前記新
しい関数を、属性に対して適用することにより、その属性に関数を適
用した結果である新たな属性を生成し、前記属性に基づき分析処理を

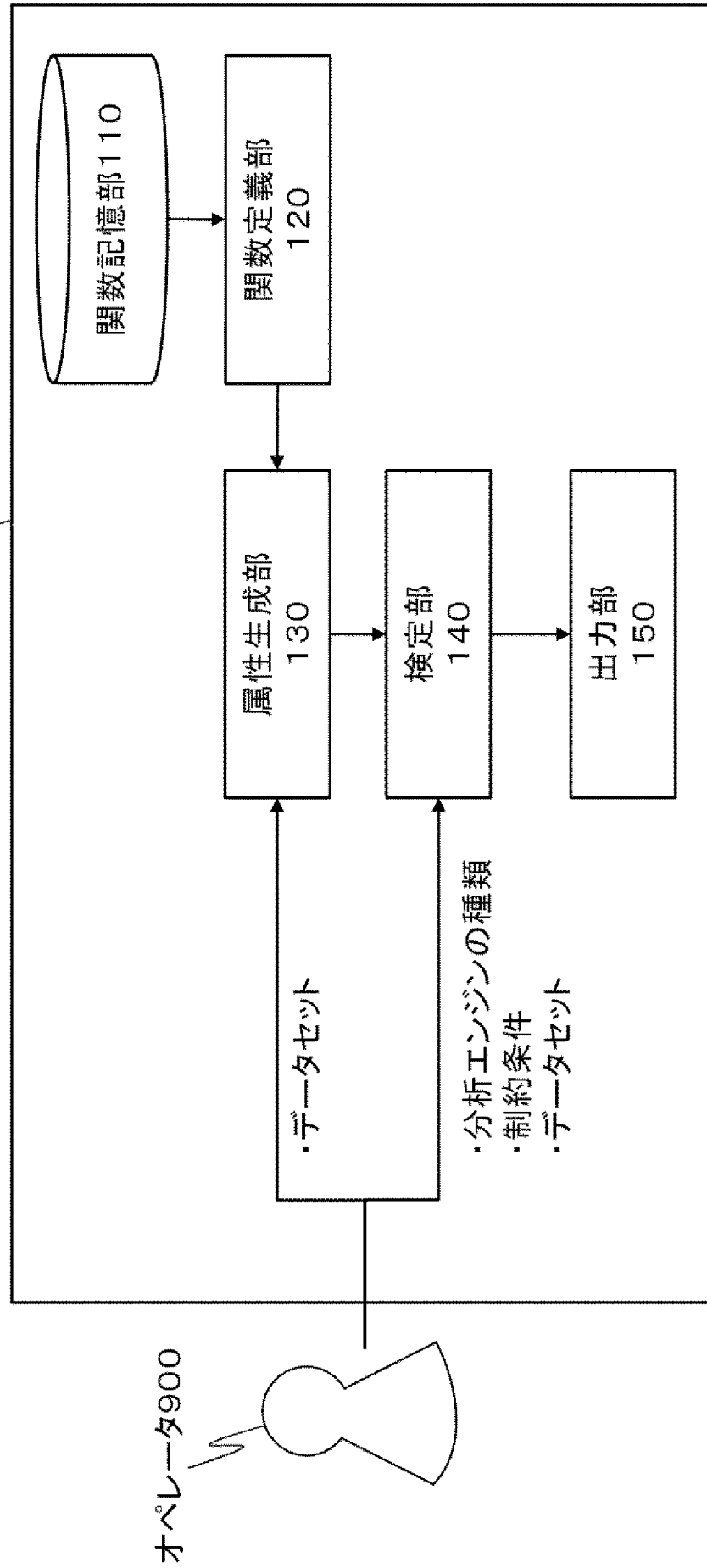
実行する分析エンジンに、前記新たな属性を入力し、前記分析エンジンが出力する情報が所定の要件を満たすか否かを判定するよう制御する制御方法。

[請求項11] 複数の関数を記憶する関数記憶手段にアクセス可能なコンピュータに、

前記複数の関数を合成することにより新しい関数を定義する処理と、前記新しい関数を、属性に対して適用することにより、その属性に関数を適用した結果である新たな属性を生成する処理と、前記属性に基づき分析処理を実行する分析エンジンに、前記新たな属性を入力し、前記分析エンジンが出力する情報が所定の要件を満たすか否かを判定する処理と、を実行させるプログラムを記憶するコンピュータ読み取り可能な記録媒体。

[図1]

情報処理システム1000



[図2]

ID (識別子)	身長 (cm)	体重 (kg)	アイスクリームの年間消費量 (リットル)
1	122	95	5.78
2	128	56	8.82
3	166	68	5.61
4	148	88	4.28
5	152	61	12.75
6	157	90	3.07
7	142	52	6.86
8	134	77	11.79
9	136	70	9.99
10	135	51	3.08
11	168	57	14.13
...	...	...	...

[図3]

関数ID	関数が定義する処理
関数1	X
関数2	$\sin(X)$
関数3	X^2

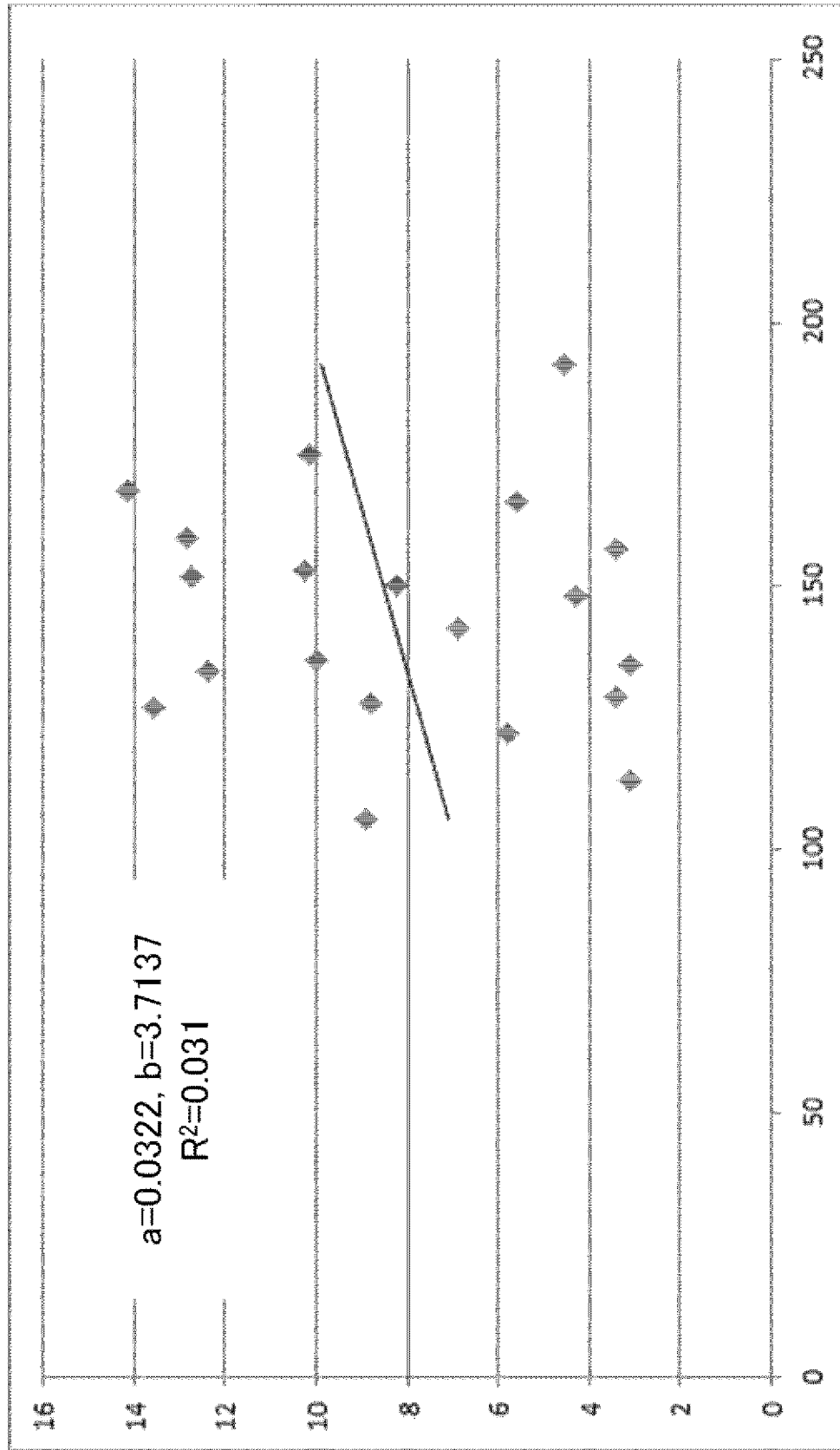
[図4]

関数ID	関数
関数4	$\sin(X^2)$
関数5	$(\sin(X))^2$

[図5]

ID (識別子)	$(\sin(\text{身長}))^2$
1	0.249
2	0.520
3	0.234
4	0.114
5	0.871
6	0.006
7	0.346
8	0.784
9	0.625
10	0.008
11	0.994
...	...

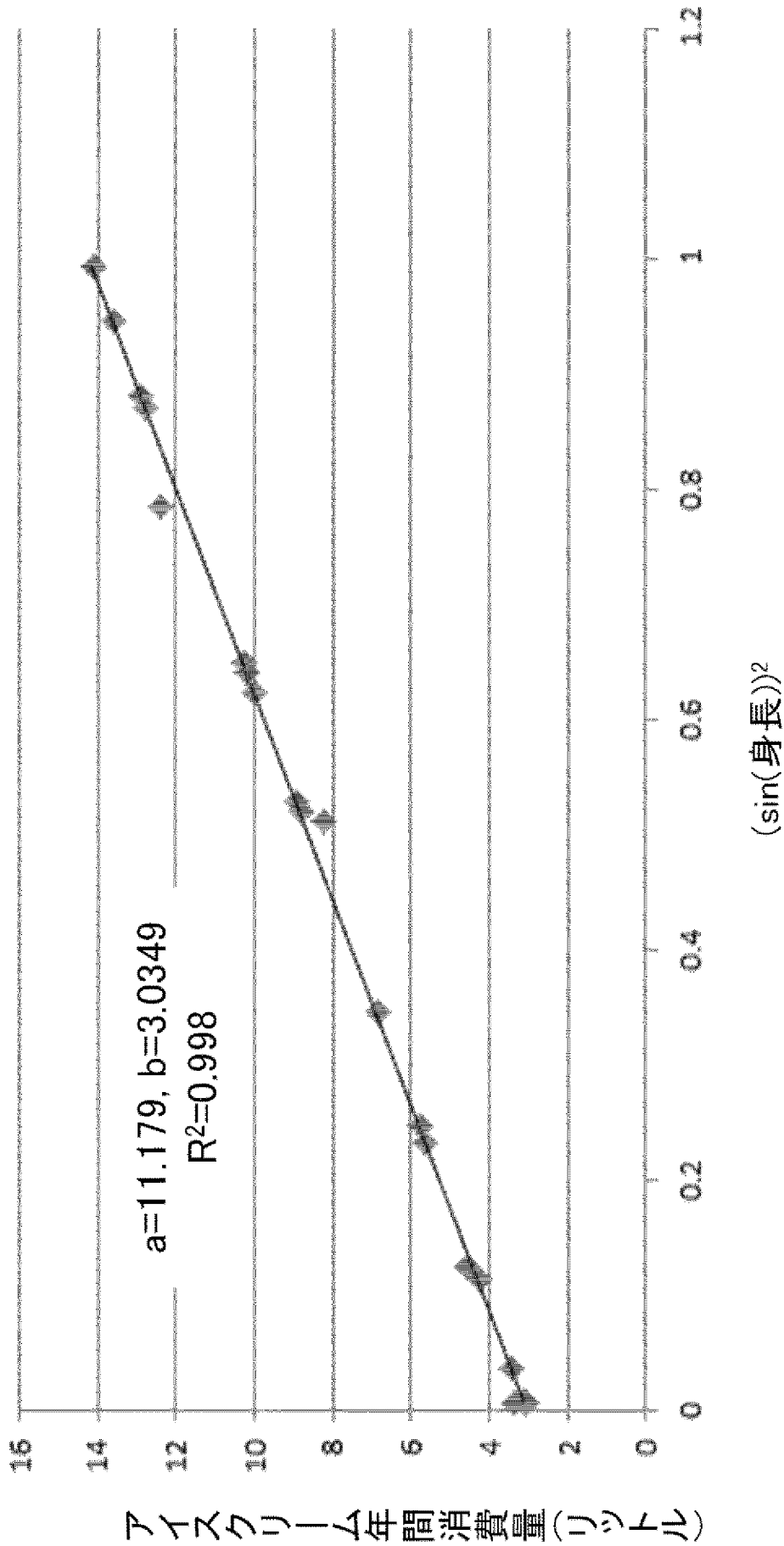
[図6]



身長

アイスクリーム年間消費量(リットル)

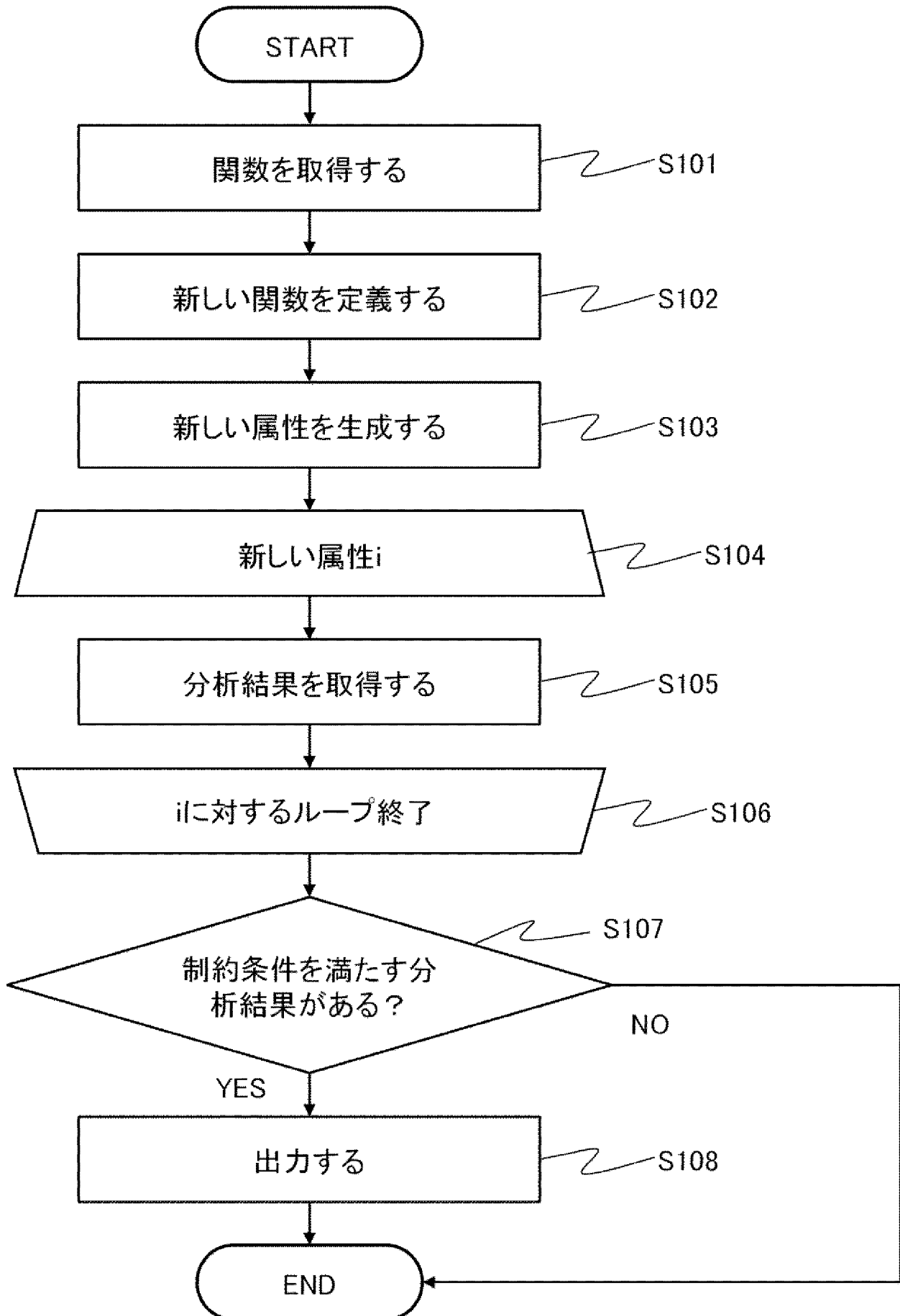
[図7]



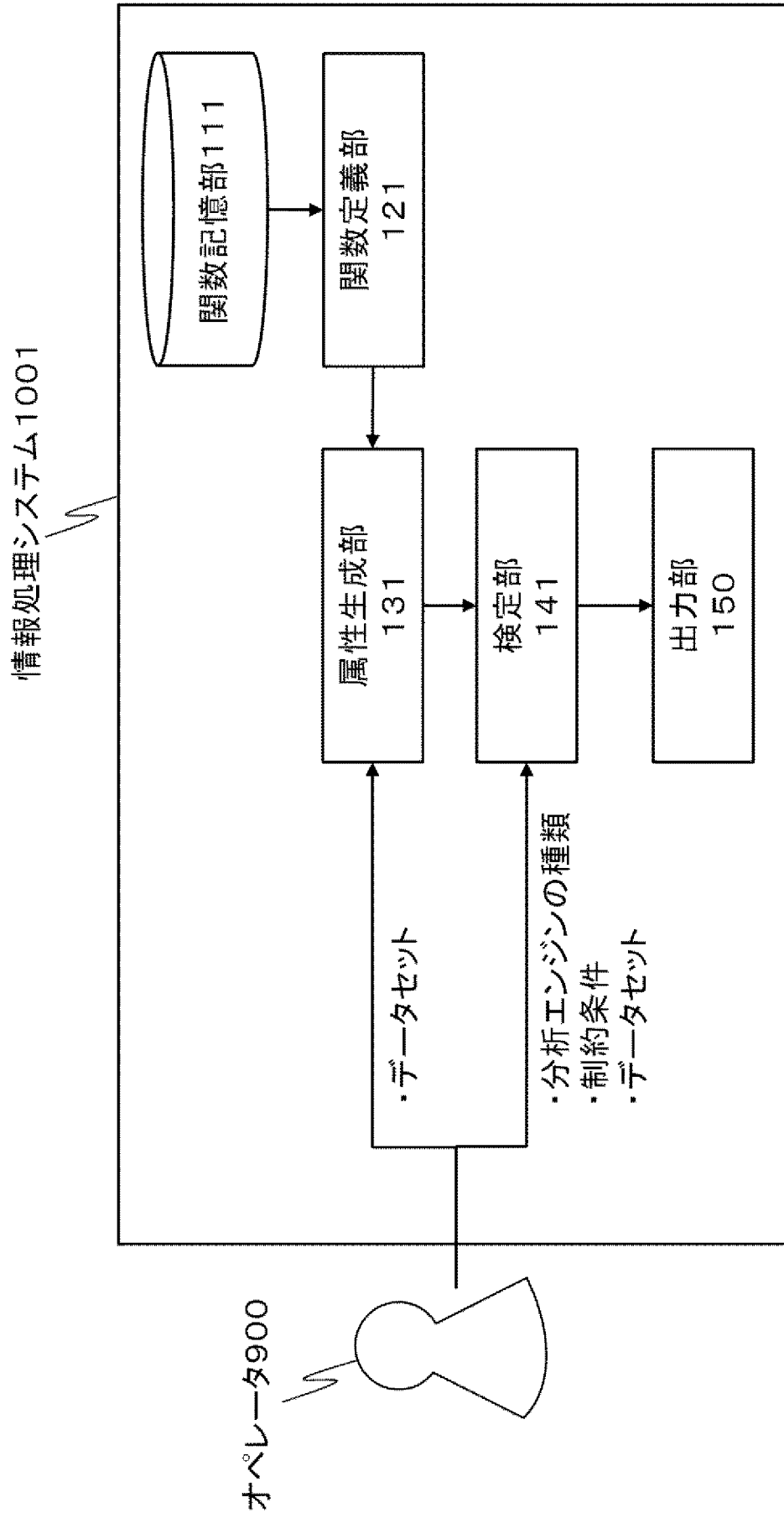
[図8]

説明変数(X)	$y=aX+b$	カイ二乗値
身長	$y=0.032X+3.714$	0.031
(身長) ²	$y=0.001X+6.271$	0.025
$\sin(\text{身長})$	$y=0.667X+8.411$	0.014
$\sin(\text{身長}^2)$	$y=-1.068X+8.365$	0.035
$(\sin(\text{身長}))^2$	$y=11.178X+3.035$	0.997
体重	$y=-0.0719X+13.31$	0.075
(体重) ²	$y=-0.0005X+11.021$	0.082
$\sin(\text{体重})$	$y=-1.3356X+8.6233$	0.055
$\sin(\text{体重}^2)$	$y=-0.339X+8.4431$	0.003
$(\sin(\text{体重}))^2$	$y=0.72X+8.053$	0.004

[図9]



[図10]



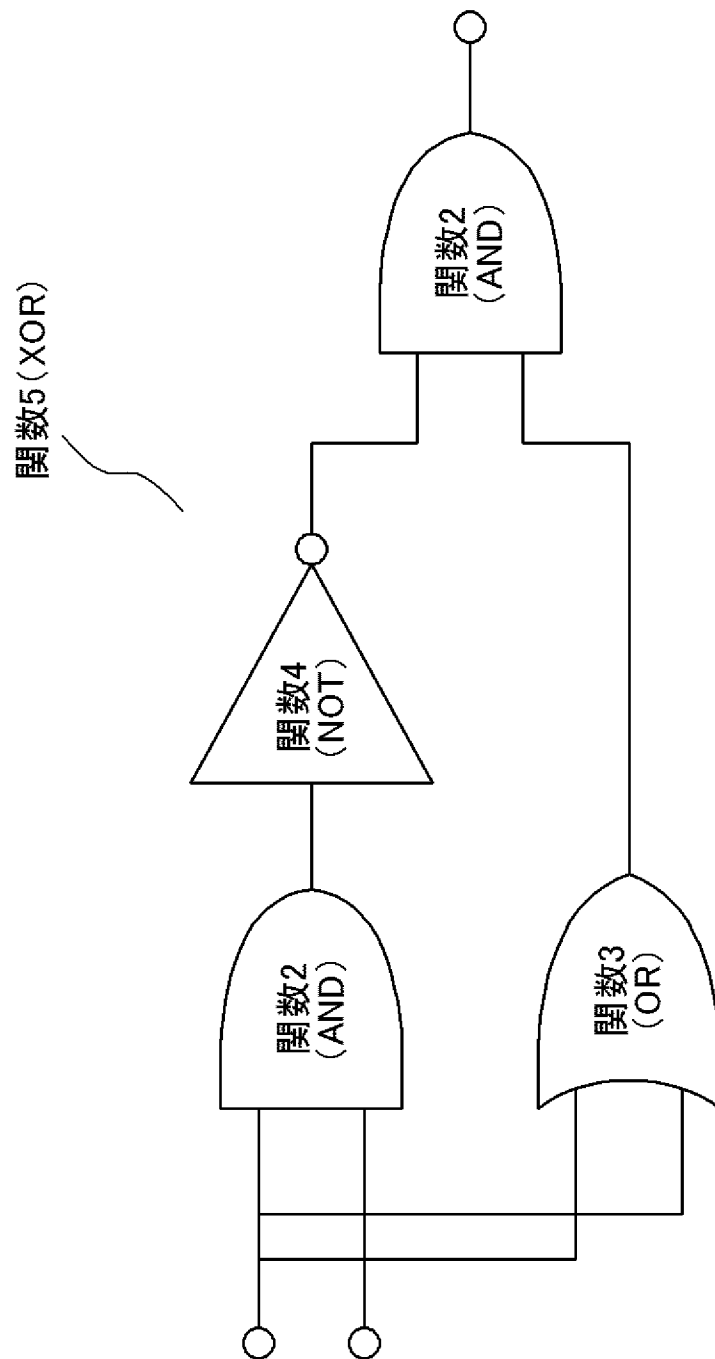
[図11]

ID (識別子)	属性1		属性2		属性3		属性4	
	犬と猫どちらが好き？ 犬→0 猫→1	年齢は？ 40歳以上→0 40歳未満→1	性別 男→0 女→1	寿司と天麩羅どちらが好き？ 寿司→0 天麩羅→1				
1	0	1	1					
2	1	1	1					
3	1	1	1					
4	1	0	0					
5	0	0	0					
6	0	1	1					
7	1	0	0					
8	1	1	1					
9	1	0	0					
10	1	1	1					
11	0	0	1					
12	1	0	1					
13	0	1	1					
...	...	...	...					...

[図12]

関数ID	関数
関数1	X
関数2	AND
関数3	OR
関数4	NOT

[図13]



[図14]

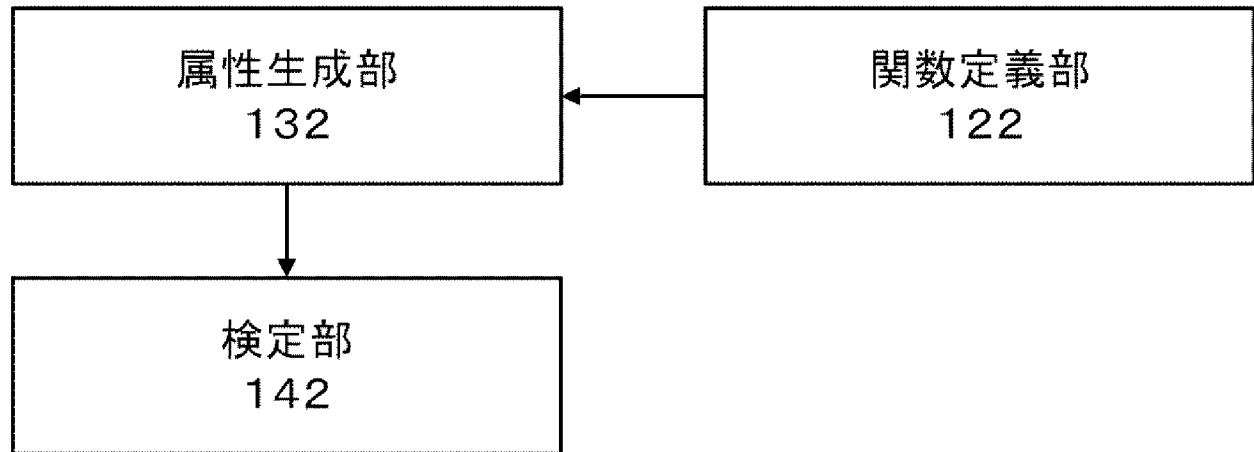
ID (識別子)	属性1 NAND 属性2
1	1
2	0
3	0
4	1
5	1
6	1
7	1
8	0
9	1
10	0
11	1
12	1
13	1
...	...

[図15]

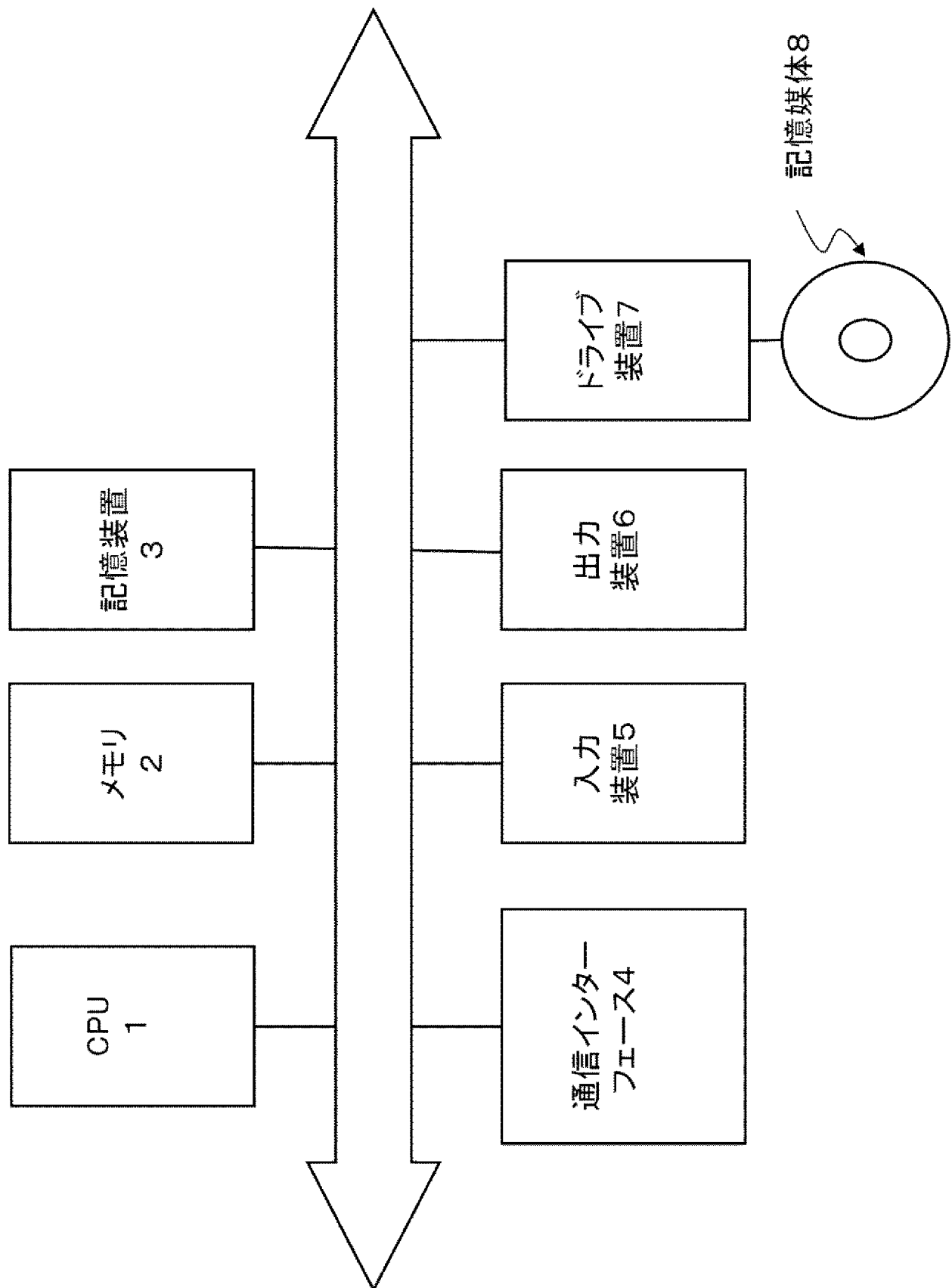
説明変数(X)	一致率
属性1	31%
属性2	38%
属性3	38%
属性1 AND 属性2	15%
属性1 AND 属性3	15%
属性2 AND 属性3	38%
属性1 OR 属性2	54%
属性1 OR 属性3	54%
属性2 OR 属性3	38%
属性1 XOR 属性2	85%
属性1 XOR 属性3	100%
属性2 XOR 属性3	31%
...	...

[図16]

情報処理システム1002



[図17]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2014/004520

A. CLASSIFICATION OF SUBJECT MATTER

G06F19/00(2011.01) 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)

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	1996-2014
Kokai Jitsuyo Shinan Koho	1971-2014	Toroku Jitsuyo Shinan Koho	1994-2014

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 2012-256182 A (Sharp Corp.), 27 December 2012 (27.12.2012), paragraph [0007] (Family: none)	1-11
A	JP 2006-048429 A (NEC Corp.), 16 February 2006 (16.02.2006), paragraphs [0019] to [0021] (Family: none)	1-11
A	JP 2010-204966 A (Nippon Telegraph and Telephone Corp.), 16 September 2010 (16.09.2010), paragraph [0012] (Family: none)	1-11

☐ 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
27 October, 2014 (27.10.14)

Date of mailing of the international search report
04 November, 2014 (04.11.14)

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. G06F19/00(2011.01)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. 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 1 4 年
日本国実用新案登録公報	1 9 9 6 - 2 0 1 4 年
日本国登録実用新案公報	1 9 9 4 - 2 0 1 4 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 2012-256182 A（シャープ株式会社）2012.12.27, 段落0007（ファミリーなし）	1-11
A	JP 2006-048429 A（日本電気株式会社）2006.02.16, 段落0019-0021（ファミリーなし）	1-11
A	JP 2010-204966 A（日本電信電話株式会社）2010.09.16, 段落0012（ファミリーなし）	1-11

☐ 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

2 7 . 1 0 . 2 0 1 4

国際調査報告の発送日

0 4 . 1 1 . 2 0 1 4

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）

塩田 徳彦

5 L

4 5 3 3

電話番号 0 3 - 3 5 8 1 - 1 1 0 1 内線 3 5 6 2