

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6800716号
(P6800716)

(45) 発行日 令和2年12月16日(2020.12.16)

(24) 登録日 令和2年11月27日(2020.11.27)

(51) Int.Cl.	F 1
GO6F 17/18 (2006.01)	GO6F 17/18 Z
GO6F 16/28 (2019.01)	GO6F 16/28
GO6Q 10/06 (2012.01)	GO6Q 10/06 300

請求項の数 8 (全 26 頁)

(21) 出願番号 特願2016-230439 (P2016-230439)
 (22) 出願日 平成28年11月28日 (2016.11.28)
 (65) 公開番号 特開2018-88087 (P2018-88087A)
 (43) 公開日 平成30年6月7日 (2018.6.7)
 審査請求日 平成31年3月11日 (2019.3.11)

(73) 特許権者 000005108
 株式会社日立製作所
 東京都千代田区丸の内一丁目6番6号
 (74) 代理人 110000062
 特許業務法人第一国際特許事務所
 (72) 発明者 小山 恒平
 東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内
 (72) 発明者 伊藤 信治
 東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内
 (72) 発明者 北原 圭
 東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内

最終頁に続く

(54) 【発明の名称】データ分析装置、データ分析方法、及びデータ分析プログラム

(57) 【特許請求の範囲】

【請求項1】

データ分析の対象となるデータに対して当該データの特徴を表す属性の集合である説明変数と、前記属性を選択するための制約情報と、前記データ分析の目的に関係する属性である目的変数とを入力とし、前記データ分析の結果を出力とするインターフェース部と、

前記制約情報を満たす前記説明変数中の前記属性の集合を選択パターンとして生成する生成部と、

生成した前記選択パターンに含まれる前記属性の集合と前記目的変数を構成する前記属性との相互の関係に基づいてデータ分析を実行するデータ分析部と、
を備えるデータ分析装置。

【請求項2】

前記制約情報は、少なくとも当該制約情報の有効性および当該制約情報の種別を含み、
前記種別として、前記目的変数の対象となる前記属性を選択しない除去制約、前記目的変数の対象となる前記属性から指定された当該属性を必ず選択する選択制約、前記目的変数の対象となる前記属性から最大1個を選択する排他制約および前記目的変数の対象となる前記属性が全て使用される場合に当該全ての前記属性を選択する付帯制約を有する、
請求項1に記載のデータ分析装置。

【請求項3】

統計分析部を備え、

前記統計分析部は、前記説明変数である前記属性の集合の中の各属性の出現頻度を、当

該説明変数を統計的に分析することで算出し、当該出現頻度に基づいて前記除去制約を生成する、

請求項 2 に記載のデータ分析装置

【請求項 4】

前記統計分析部は、前記説明変数である前記属性の集合から生成した当該属性のペアの相関係数に基づいて前記排他制約を生成する、

請求項 3 に記載のデータ分析装置。

【請求項 5】

前記生成部は、有効である前記制約情報から生成した論理式を用いて前記選択パターンを生成する、

請求項 2 から 4 のいずれか 1 項に記載のデータ分析装置。

【請求項 6】

履歴分析部を備え、

前記履歴分析部は、過去に生成した前記選択パターンの履歴から前記付帯制約および当該付帯制約に対応する前記制約情報を生成する、

請求項 5 に記載のデータ分析装置。

【請求項 7】

コンピュータによりコンピュータプログラムで実現されるユーザインターフェース部、選択パターン生成部およびデータ分析部が実行されてデータを分析するデータ分析方法であって、

前記ユーザインターフェース部が、補助記憶装置から、データ分析の対象となるデータに対して当該データの特徴を表す属性の集合である説明変数と、前記属性を選択するための制約情報と、前記データ分析の目的に関係する属性である目的変数とを入力する第 1 のステップと、

前記選択パターン生成部が、前記制約情報を満たす前記説明変数中の前記属性の集合を選択パターンとして生成する第 2 のステップと、

前記データ分析部が、前記選択パターンに含まれる前記属性の集合と前記目的変数を構成する前記属性との相互の関係に基づいてデータ分析を実行する第 3 のステップと、

前記ユーザインターフェース部が、前記データ分析の結果を表示画面に出力する第 4 のステップと

を有するデータ分析方法。

【請求項 8】

データ分析をコンピュータに実行させるためのデータ分析プログラムであって、

データ分析の対象となるデータに対して当該データの特徴を表す属性の集合である説明変数と、前記属性を選択するための制約情報と、前記データ分析の目的に関係する属性である目的変数とを入力する第 1 のステップと、

前記制約情報を満たす前記説明変数中の前記属性の集合を選択パターンとして生成する第 2 のステップと、

生成した前記選択パターンに含まれる前記属性の集合と前記目的変数を構成する前記属性との相互の関係に基づいてデータ分析を実行する第 3 のステップと、

前記データ分析の結果を出力する第 4 のステップと

を有するデータ分析プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、データ分析に掛かる工数を削減するためにデータ分析を自動化する装置に関する。

【背景技術】

【0002】

蓄積したデータを活用して現行業務を改善する動きが活発化している。具体的には、部

10

20

30

40

50

署ごとの業務履歴や小売店での売上の履歴などのデータに対して、統計学や人工知能などデータ解析技術を適用し、今まで気が付かなかった法則やルール(以下、仮説という)を発見し、業務の改善施策を検討するデータ分析が活発化している。

【0003】

データ分析は、一般的に、「業務の理解」、「説明変数の理解」、「説明変数の準備」、「モデリング(分析技術の適用)」、「仮説の評価」、「施策の運用」という6つのフェーズから構成されることが多い。

【0004】

「業務の理解」フェーズでは、分析の目的や要求を把握し、部署や小売店などの分析対象、及び、客単価や作業工数などの業務改善の指標を表す属性(以下、目的変数という)を決める。

10

【0005】

次に、「説明変数の理解」フェーズでは、分析に必要なデータ(以下、説明変数という)を収集し、具体的にどのような値が入っているかを確認する。ここで、説明変数とは、部署や小売店などの分析対象の特徴を表す属性の集合をいう。さらに、分析対象は、各属性に対応する個別の値(以下、属性値)を持っている。

【0006】

「説明変数の準備」フェーズでは、必要に応じて、欠損値の補完や既存の属性を組合せた新たな属性の生成、分析に使用する属性の選択などを行う。

【0007】

「モデリング」フェーズでは、準備した説明変数に対して、重回帰分析や決定木分析などデータ分析技術を適用する。このとき、データ分析技術は分析結果として、回帰式や決定木などの仮説を出力する。

20

【0008】

「仮説の評価」フェーズでは、得られた仮説に基づいて、業務の改善施策を検討する。このとき、施策に繋がる仮説(以下、有用な仮説という)が得られなかったら、作業内容を変更して、もう一度「説明変数の準備」フェーズと「モデリング」フェーズを実施する。

【0009】

「施策の運用」フェーズでは、「仮説の評価」フェーズで検討した課題を現場に適用し、業務改善できるかを評価する。

30

【0010】

データ分析技術の中には、属性の数によって処理時間が爆発的に増加するものがある。このような技術を使う場合には、分析者が、実用時間で処理が終わる範囲内で、属性を選択し、データ分析技術を適用するという試行錯誤を繰り返し実行する必要がある。この試行錯誤は有用な仮説が得られるまで続けるため、データ分析には多くの工数が必要になる。

【0011】

データ分析の工数を削減する別 の方法として、属性値を統計的に分析した結果を用いて、分析への有用さで属性をランク付けすることで属性の選択を支援し、さらに、ランク付けに基づいて分析者が選択した属性に対して、パラメータを変えながら、データ分析技術を繰り返し適用する技術が、たとえば、特許文献1で開示されている。

40

【先行技術文献】

【特許文献】

【0012】

【特許文献1】米国特許第7、627、620号明細書

【発明の概要】

【発明が解決しようとする課題】

【0013】

特許文献1に開示されている技術では、統計的に分析した結果に基づいて、人が選択し

50

た属性に対して、パラメータを変えながら繰り返しデータ分析技術を適用できる。しかし、属性の選択は自動化できていないという問題がある。

【0014】

そこで、本発明は、制約に基づいて属性の選択、及び、データ分析を自動化でき、データ分析に関わる工数を削減できるデータ分析装置の提供を目的とする。

【課題を解決するための手段】

【0015】

上記目的を達成するために、代表的な本発明のデータ分析装置の一つは、データ分析の対象となるデータに対して当該データの特徴を表す属性の集合である説明変数と、属性を選択するための制約情報と、データ分析の目的に関係する属性である目的変数とを入力とし、データ分析の結果を出力とするインターフェース部と、制約情報を満たす説明変数中の属性の集合を選択パターンとして生成する生成部と、生成した選択パターンに含まれる属性の集合と目的変数を構成する属性との相互の関係に基づいてデータ分析を実行するデータ分析部と、を備えるものである。

10

【発明の効果】

【0016】

本発明によれば、制約に基づいて属性の選択、及び、データ分析を自動化でき、データ分析に関わる工数を削減できるデータ分析装置が実現できる。

【図面の簡単な説明】

【0017】

20

【図1】本実施例に関わるデータ分析自動化装置100の構成プロックの例を示す図。

【図2】本実施例に関わるコンビニの店舗情報200に重回帰分析を適用した例を示す図。

【図3】本実施例に関わるデータ記憶部112のデータクラス構造の例を示す図。

【図4】本実施例に関わる仮説記憶部113のデータクラスの構造の例を示す図。

【図5】本実施例に関わる選択パターン記憶部114のデータクラスの構造の例を示す図。

【図6】本実施例に関わる制約記憶部111のデータクラスの構造の例を示す図。

【図7】本実施例に関わるデータ分析設定画面700の例を示す図。

【図8】本実施例に関わるデータ読み込み画面800の例を示す図。

30

【図9】本実施例に関わる制約管理画面900の例を示す図。

【図10】本実施例に関わる制約編集画面1000の例を示す図。

【図11】本実施例に関わる分析結果表示画面1100の例を示す図。

【図12】本実施例に関わるデータ分析のメインのフローチャートの例を示す図。

【図13】本実施例に関わる統計情報からの制約生成の処理フローチャートの例を示す図。

【図14】本実施例に関わる選択パターン生成の処理フローチャートの例を示す図。

【図15】本実施例に関わる分析ツール適用の処理フローチャートの例を示す図。

【図16】本実施例に関わる履歴からの制約生成の処理フローチャートの例を示す図。

【発明を実施するための形態】

40

【0018】

以下、図面を用いて実施例を説明する。本実施例では、制約に基づいてデータ分析を自動化する装置の例を説明する。なお、以下に説明する実施例は特許請求の範囲にかかる発明を限定するものではなく、また実施例の中で説明されている各構成要素及びその組合せの全てが、課題を解決するための手段として必須であるとは限らない。

【実施例】

【0019】

図1は、本実施例に関わるデータ分析自動化装置100の構成プロックの例を示す図である。データ分析自動化装置100は、少なくともCPU(Central Processor Unit)等のプロセッサ121、RAM(Random Access

50

Memory)、ROM(Read Only Memory)等のメモリ122、補助記憶装置110、及び、キーボードや表示モニタ等の入出力装置123を備えたコンピュータである。データ分析自動化装置100は、ユーザインターフェース部101、統計分析部102、選択パターン生成部103、データ分析部104、履歴分析部105、矛盾検証部106などの機能ブロックを備える。これらの機能ブロックはコンピュータプログラムとして実現されており、メモリ122上でプロセッサ121により実行される。また、データ分析自動化装置100は、制約記憶部111、データ記憶部112、仮説記憶部113、選択パターン記憶部114などのデータを補助記憶装置110に保管している。これらのデータは、プロセッサ121が機能ブロックを実行する際に補助記憶装置110から読みだして使用される。

10

【0020】

図2は、本実施例に関わるコンビニの店舗情報200に重回帰分析を適用した例を示す図である。コンビニの店舗情報200としては、売上、気候、取扱商品などがある。

【0021】

ここで、分析対象は、コンビニの店舗情報200の1行に対応している。なお、本実施例では、データ分析技術として重回帰分析を用いたケースで説明するが、本発明自体は、決定木分析や相関ルール等、他のデータ分析ツールにも適用可能な技術である。

【0022】

重回帰分析を行うツール(以下、単に重回帰分析ツールという)は、目的変数と説明変数を入力として、目的変数と説明変数中の属性との関係を「目的変数 = $\alpha \times (\text{属性1}) + \beta \times (\text{属性2}) + \dots$ 」形式の回帰式で導出する。ここで、 α 及び β は定数である。以降では、仮説とは、重回帰分析ツールが出力する回帰式を指すこととする。

20

【0023】

図2の例では、売上を目的変数201とし、医療品取扱い、食料品取扱い、周辺人口、所在市面積、平均年収、平均気温、最低気温、最高気温、及び、降水量を説明変数202として、重回帰分析ツール204に与えることで、例えば「売上 = -0.03 × 周辺人口 - 18.27 × 所在市面積 + 0.25 × 平均年収 - 38709.4」や「売上 = 22,863.72 × 平均気温 + 200.81 × 降水量」を回帰式205として得ることができる。

【0024】

このとき、重回帰分析ツールによっては、各回帰式205の評価指標を出力できる。たとえば、相関係数206、決定係数207は、回帰式の評価指標の例である。相関係数206は、説明変数の実際の属性値と回帰式から求められる属性値の相関関係を数値化したものであり、この数値が大きいほど、相関が高いことを表している。決定係数207は、回帰式の当てはまりの良さを数値化したものであり、大きいほど目的変数と説明変数の実際の属性値によく当てはまる回帰式であることを表している。

30

【0025】

図3は、本実施例に関わるデータ記憶部112のデータクラス構造の例を示す図である。データ記憶部112は、属性情報301、分析対象情報302、及び、属性値情報303の各データクラスを有する。図3の各データクラス表記は、2つの四角形により構成されている。上の四角形内にはデータクラス名が表されており、下の四角形内には当該データクラスのフィールド名が表されている。

40

【0026】

また、データクラス間の関連は、矢印を結んで表している。矢印の始点にあるデータクラスは、終点にあるデータクラスを参照できることを表している。終点の文字列は参照されるデータクラスの役割を表している。さらに、矢印の始点及び終点の数字または「n」は、関連の多重度を表している。ここで、「n」は任意の数を意味する。以下に示す他のデータ構造の図についても同様の表記を用いている。

【0027】

属性情報301は、属性に関する情報として、属性名、型、目的フラグを有する。

【0028】

50

属性名は、属性間でユニークな識別子である。図2の例では、属性名は、「医療品取扱い」や「周辺人口」などである。

【0029】

型は、後述する属性値のデータ型である。図2の例では、属性「食料品取扱い」の「文字列」や属性「最高気温」の「整数」が型である。

【0030】

目的フラグは、属性が目的変数かどうかを表すフラグであり、目的フラグが「真（以下、T R U Eという）」の属性は目的変数であり、「偽（以下、F A L S Eという）」の属性は説明変数に含まれる属性である。図2の例では、属性「売上」が目的変数なので、目的フラグは「T R U E」になる。以降では、目的変数は、目的フラグが「T R U E」になっている属性を指すこととし、説明変数は、目的フラグが「F A L S E」になっている1以上の属性の属性すべてを指すこととする。なお、本実施例では、説明変数は、データ記憶部112に含まれる属性の中で、ひとつしか存在しないものとする。

10

【0031】

属性情報301は、母集団として、n個の後述する分析対象情報302と関連を持ち、さらに、母集団の属性値の集合として、n個の後述する属性値情報303と関連を持っている。このため、属性は、必要に応じて、分析対象に対応する1以上の属性値をすべて取得できる。

【0032】

分析対象情報302は、分析対象に関する情報として、分析対象IDを持つ。分析対象IDは、分析対象間でユニークな識別子である。図2の例では、分析対象IDは、「名古屋店」や「気仙沼店」などである。

20

【0033】

分析対象情報302は、保有属性として、n個の属性情報301と関連を持ち、保有属性の属性値として、n個の後述する属性値情報303と関連を持っている。このため、分析対象は、必要に応じて、属性に対応する1以上の属性値をすべて取得できる。

【0034】

属性値情報303は、属性に対する分析対象の値として、属性値を持つ。図2の例では、属性値は、属性「売上」、分析対象「名古屋店」の場合の「798,384」や、属性「平均気温」、分析対象「那覇店」の場合の「28」などである。

30

【0035】

図4は、本実施例に関わる仮説記憶部113のデータクラスの構造の例を示す図である。仮説記憶部113は、仮説情報401を有する。

【0036】

仮説情報401は、データを分析して導出された仮説に関する情報として、仮説ID、回帰式、相関係数、決定係数を有する。仮説IDは、仮説間でユニークな識別子である。回帰式は分析ツールの出力結果であり、図2の例では、回帰式205の「売上= 0.03 × 周辺人口 18.27 × 所在市面積 + 0.25 × 平均年収 38709.4」である。相関係数及び決定係数は、回帰式の評価指標であり、図2の例では、それぞれ、相関係数206の「0.83」、決定係数207の「0.69」である。

40

【0037】

仮説記憶部113から出力される1つの仮説情報401は、回帰式に含まれる属性として、n個の属性情報301と関連を持っているため、仮説は、必要に応じて、回帰式に含まれる1以上の属性をすべて取得できる。

【0038】

図5は、本実施例に関わる選択パターン記憶部114のデータクラスの構造の例を示す図である。選択パターン記憶部114は、選択パターン情報501を有する。

【0039】

選択パターン情報501は、分析に使用する説明変数中の属性の集合（以下、選択パターンという）に関する情報として、パターンIDを有する。パターンIDは、選択パター

50

ン間でユニークな識別子である。

【0040】

選択パターン記憶部114から出力される1つの選択パターン情報501は、使用する属性として、n個の属性情報301と関連を持っているため、選択パターンは、必要に応じて、自身に含まれている1以上の属性をすべて取得できる。さらに、選択パターン記憶部114から出力される1つの選択パターン情報501は、導出された仮説として、n個の仮説情報401とも関連を持っているため、選択パターンは、必要に応じて、導出された1以上の仮説をすべて取得できる。

【0041】

図6は、本実施例に関わる制約記憶部111のデータクラスの構造の例を示す図である
10。制約記憶部111は、制約情報601を有する。

【0042】

制約情報601は、選択パターンへの制約に関する情報として、制約ID、有効性、自動生成フラグ、制約種別602、選択数を有する。制約IDは、制約間でユニークな識別子である。有効性は、制約が有効化どうかを表すフラグであり、「有効」のとき、その制約は有効な制約といい、「無効」のとき、その制約は無効な制約という。自動生成フラグは、制約が分析者から、入力されたのかそれとも、装置が生成したのかを表すフラグであり、「TRUE」のとき、装置が生成した制約を意味し、「FALSE」のとき、分析者が入力した制約を意味する。以降では、自動生成フラグが「TRUE」となっている制約を特に、自動生成された制約という。
20

【0043】

制約種別602は、制約の内容を表し、4つの制約「除去」、「選択」、「排他」、「付帯」の何れかである。以降では、制約種別602が「除去」の制約を除去制約といい、「選択」の制約を選択制約といい、「排他」の制約を排他制約といい、「付帯」の制約を付帯制約ということとする。

【0044】

選択数は、対象属性の中から選択する属性の数を表し、選択制約のみが使用するフィールドである。ここで、対象属性とは、目的変数の対象になる属性をいう。

【0045】

制約記憶部111から出力される1つの制約601は、対象属性として、n個の属性情報301と関連を持っているため、制約は、必要に応じて、1以上の対象属性をすべて取得できる。さらに、前提属性としても、n個の属性情報301と関連を持っているが、この関連は、付帯制約のみが使用する。すなわち、前提属性とは、対象属性と同じ属性をいう。
30

【0046】

また、各制約は、属性が選択パターンに含まれるかどうかを表す命題変数Piを用いて論理式としても表現できる。ここで、「i」は属性情報301の属性名に対応している。さらに、Piが「TRUE」の場合、Piに対応する属性が選択パターンに含まれることを意味し、「FALSE」の場合、Piに対応する属性が選択パターンに含まれないことを意味する。
40

【0047】

除去制約は「対象属性に含まれる属性を選択しない」ことを意味しており、以下に示す対象属性に含まれる属性に対応する命題変数が「TRUE」にならないことを意味する論理式(1)で表現できる。なお、以降では、除去制約は「n t {対象制約}」とも記述する

【0048】

【数1】

$$\bigwedge_{\text{属性名} \in \text{対象属性}} \neg^{P_{\text{属性名}}} \quad (1)$$

選択制約は「対象属性に含まれる属性のうち、必ず指定された分の属性を選択する」ことを意味しており、以下に示す、対象属性に含まれる属性に対応する命題変数のうち、選択数分の命題変数が「TRUE」となることを意味する論理式(2)として表現できる。なお、以降では、選択制約は、「Select(選択数){対象属性}」とも記述する

【0049】

【数2】

$$at-most(\text{対象属性}, \text{選択数}) \wedge at-least(\text{対象属性}, \text{選択数}) \quad (2)$$

ここで、命題変数の集合において、TRUEとなる命題変数の数に関する論理式は、一般的に基数制約と呼ばれており、「at-most(対象属性、選択数)」は、「対象属性」に対応する命題変数の中で最大「選択数」個の命題変数が「TRUE」になることを表す基数制約であり、「at-least(対象属性、選択数)」は、「対象属性」に対応する命題変数の中で少なくとも「選択数」個の命題変数がTRUEになることを表す基数制約である。

【0050】

なお、基数制約の具体的な表現方法については、例えば「Olivier Bailleux、Yacine Boufkhad、"Efficient CNF encoding of Boolean Cardinality Constraints"、Principles and Practice of Constraint Programming - CP 2003、pp 108 - 122、2003」などに詳しく述べられているので、そちらを参照されたい。

【0051】

また、基数制約に対応したSATソルバも存在し、具体的な論理式に変換するのではなく、それらのSATソルバを活用することも可能である。なお、基数制約に対応したSATソルバについては、例えば、「山根裕二らの“基数制約の概念を持つSATソルバの設計と評価”、人工知能学会全国大会論文集25、pp 1 - 4、2011」などに詳しく述べられているので、そちらを参照されたい。

【0052】

排他制約は、「対象属性から0～1個の属性を選択する」ことを意味しており、以下に示す、対象属性に含まれる属性に対応する命題変数のうち、最大1個の命題変数が「TRUE」となることを意味する論理式(3)として表現できる。なお、以降では、排他制約は「at-most(対象属性, 1)」とも記述する

【0053】

【数3】

$$at-most(\text{対象属性}, 1) \quad (3)$$

付帯制約は、「前提属性がすべて使用されるとき、対象属性も一緒にすべて選択する」ことを意味しており、以下に示す、前提属性に含まれる属性に対応する命題変数がすべてTRUEのとき、対象属性に含まれる属性に対応する命題変数がすべてTRUEになることを意味する論理式(4)に対応する。なお以降では、付帯制約は、「implyn(前提属性){対象属性}」とも記述する

【0054】

【数4】

$$\left(\bigwedge_{\text{属性名} \in \text{前提属性}} P_{\text{属性名}} \right) \rightarrow \left(\bigwedge_{\text{属性名} \in \text{対象属性}} P_{\text{属性名}} \right) \quad (4)$$

以降において、制約は、除去制約、選択制約、排他制約、付帯制約の何れかを指すこととする。

10

20

30

40

50

【0055】

図7は、本実施例に関わるデータ分析設定画面700の例を示す図である。この図は、ユーザインタフェース部101が入出力装置123に表示し、分析者が入力した自動データ分析の設定情報を受け付ける画面の例を示している。すなわち、ユーザインタフェース部が、ユーザからの入力を受け付ける受付部としての機能を有する。

【0056】

データ分析設定画面700は、分析パラメータ設定エリア710、制約リスト表示エリア720、データ表示エリア730、データ読み込みボタン740、制約管理ボタン750、分析実行ボタン760を含む。

【0057】

分析パラメータ設定エリア710は、自動分析時のパラメータを表示・入力する領域であり、最大属性数、最小属性数、終了条件、制約自動生成対象、分析周期、分析カバレッジを表示し、統計情報チェックボックス、履歴チェックボックス、終了条件の単位指定コンボボックスを含む。

【0058】

ユーザインタフェース部101が、「最大属性数」、「最小属性数」、「終了条件」、「分析カバレッジ」、「制約自動生成対象」、「分析周期」のタイトル(項目名)、統計情報チェックボックス、履歴チェックボックス、終了条件の単位指定コンボボックスを表示し、分析者が最大属性数、最小属性数、終了条件、制約自動生成対象、分析周期、終了条件の単位を入力する。

【0059】

最大属性数は、選択パターンに含まれる属性の最大数を意味している。最小属性数は、選択パターンに含まれる属性数の最小数を意味している。終了条件は、制約を満たす全選択パターンに分析を適用した場合以外の分析の終了条件を意味している。

【0060】

図7では、終了条件の単位として、「時間」が入力されているが、それ以外に、分析回数を表す「回」を指定することも可能である。制約自動生成対象は、統計情報から制約を自動生成するか、また、履歴から制約を自動生成するかを意味しており、それぞれ該当するチェックボックスにチェックを入れることで、自動生成を有効にできる。

【0061】

分析周期は、履歴からの制約自動生成を行う間隔を意味している。分析カバレッジは、履歴分析部105において、分析していない属性の組合せを特定する際に使用する。なお、分析周期及び分析カバレッジは、履歴からの制約自動生成が有効な場合のみ、入力可能である。

【0062】

図7は、最大属性数として「5属性」、最小属性数として「3属性」、終了条件として「1時間」、が入力され、制約自動生成対象として「統計情報」、「履歴」のチェックボックスにチェックが入り、さらに、分析周期として「10回」、分析カバレッジとして「2属性」が入力された例である。

【0063】

制約リスト表示エリア720は、制約記憶部111に記憶されている制約情報601を表示する領域であり、ユーザインタフェース部101が制約記憶部111に記憶されている制約情報601のうち、制約表示条件設定エリア721で指定された条件を満たす制約一覧を表示し、制約表示条件設定エリア721を含む。分析者が、制約一覧中のいずれかの制約情報601を押下すると、ユーザインタフェース部101は、制約情報601を選択された制約として記憶し、さらに、分析者に制約が選択されていることを示すため、制約の背景色を変更する。

【0064】

図7は、制約情報601の一覧表として、制約ID1「付帯制約：implc ati n(周辺人口){平均年収}」、制約ID2「選択制約：select(1){周辺人

10

20

30

40

50

口、最高気温}」、制約ID3「{排他制約：at-m st-ne {平均年収、最高気温}}」を表示させた例である。

【0065】

制約表示条件設定エリア721は、制約の一覧表に表示させる制約情報601の条件を指定する領域であり、有効性、種別、対象／前提属性を表示し、「指定しない」、「有効」、「無効」の何れかを選択するラジオボタン、表示させる制約の種別を選択するコンボボックスを含む。ユーザインタフェース部101が「有効性」、「種別」、「使用属性」のタイトル（項目名）、ラジオボタン、コンボボックスを表示し、分析者が表示させる制約の有効性、種別、分析に使用する属性（使用属性）を入力する。なお、種別は、「除去制約」「排他制約」「選択制約」「付帯制約」の何れかを選択できる。

10

【0066】

ユーザインタフェース部101は、表示させる制約の条件が分析者によって変更されたことを検知すると、分析者が入力した制約の有効性、種別、使用属性を全て満たす制約の制約情報601を制約記憶部111から取得し、制約の一覧表に表示させる。なお、有効性の「指定しない」は、表示させる制約の条件に有効性を使用しないことを意味している。

【0067】

図7は、有効性として「有効」を選択し、種別を選択せず、対象／前提属性を入力しない例である。

【0068】

データ表示エリア730は、データ記憶部112から読み込んだ属性情報301、分析対象情報302、及び属性値情報303を表示させる領域であり、ユーザインタフェース部101がデータ一覧を表示させる。データ一覧は、各属性情報301のうち目的変数としていずれかを選択するラジオボタンを含む。

20

【0069】

分析者が、ラジオボタンで、ある属性を目的変数として選択すると、ユーザインタフェース部101は、選択された属性に対応する属性情報301の目的フラグを「TRUE」に変更し、選択されなかった属性に対応する属性情報301の目的フラグを「FALSE」に変更する。ここでは、「売上」が目的変数になり、「医療品取扱い」、「食料品取扱い」、「周辺人口」、「所在市面積」、「平均年収」、「平均気温」、「最低気温」、「最高気温」、及び、「降水量」は説明変数になる。

30

【0070】

データ読み込みボタン740は、CSVファイルや外部DBからデータ読み込みを行うためのボタンである。分析者がデータ読み込みボタン740を押下すると、ユーザインタフェース部101は、後述するデータ読み込み画面800を表示する。

【0071】

制約管理ボタン750は、制約記憶部111に記憶された制約の管理を行うためのボタンである。分析者が制約管理ボタン750を押下すると、ユーザインタフェース部101は、後述する制約管理画面900を表示する。

【0072】

40

分析実行ボタン760は、分析パラメータ設定エリア710で入力された分析の設定及び制約記憶部111に記憶された制約に基づいて、データ記憶部112に記憶されたデータに対して、自動的にデータ分析を適用するためのボタンである。分析者が分析実行ボタン760を押下すると、ユーザインタフェース部101は、後述するデータ分析全体の流れS1200以降に従って、自動的に分析を実施し、分析終了後、後述する分析結果表示画面1100を表示させる。

【0073】

図8は、本実施例に関わるデータ読み込み画面800の例を示す図である。この図では、ユーザインタフェース部101が入出力装置123に表示する、CSVファイルやデータベースなどからデータを読み込むためのデータ読み込み画面800の例を表している。

50

【0074】

データ読み込み画面800は、取得元設定エリア810、キャンセルボタン820、読み込み実行ボタン830を含む。

【0075】

取得元設定エリア810は、データ取得元の情報を表示する領域であり、取得元を表示し、取得元の種類がデータベースかCSVファイルかを選択するラジオボタン（以下、取得元種類選択のラジオボタンという）、制約を生成するかどうかを選択するラジオボタン（以下、制約生成選択のラジオボタンという）を含む。ユーザインタフェース部101が「取得元」のタイトル（項目名）、取得元種類選択のラジオボタン、制約生成選択のラジオボタンを表示させ、分析者が、取得元、取得元の種類、制約を生成するかどうかをする。

10

【0076】

図8は、取得元として「C:\$data\$service_store_data.csv」、種類として「CSVファイル」、制約の生成を「する」と入力された例である。

【0077】

なお、本装置への入力ファイル及びデータベースには、属性情報301の各フィールド及び分析対象情報302の各フィールド、属性値情報303の各フィールドに対応する情報が機械解釈可能な形で含まれていることを前提とする。

【0078】

キャンセルボタン820は、データ読み込みを実施せずに、データ読み込み画面800を閉じるためのボタンである。分析者がキャンセルボタン820を押下すると、ユーザインタフェース部101は、データ読み込み画面800を閉じる。

20

【0079】

読み込み実行ボタン830は、データ読み込みを実施し、その後データ読み込み画面800を閉じるためのボタンである。分析者が読み込み実行ボタン830を押下すると、ユーザインタフェース部101は、取得元に指定されたファイルまたはデータベースから、属性、分析対象、属性値を読み込み、データ記憶部112に、属性情報301、分析対象情報302、属性値情報303を追加する。さらに、分析者が制約の生成を「する」と選択している場合は、データ分析部104を用いて、読み込んだデータの統計情報から排他制約、もしくは、除去制約を生成し、生成した制約に対応する制約情報601を制約記憶部111に追加する。データの読み込み及び制約の生成後、ユーザインタフェース部101は、データ読み込み画面800を閉じる。さらに、ユーザインタフェース部101は、データの読み込み及び制約の生成の結果をデータ分析設定画面700の制約一覧及びデータ一覧に反映させる。

30

【0080】

図9は、本実施例に関わる制約管理画面900の例を示す図である。この図では、ユーザインタフェース部101が入出力装置123に表示する、本装置が記憶している制約情報601を管理するための制約管理画面900の例を表している。

【0081】

制約管理画面900は、制約リスト表示エリア910、削除ボタン920、修正ボタン930、追加ボタン940、選択パターン例示エリア950、終了ボタン960を含む。

40

【0082】

制約リスト表示エリア910は、図7の制約リスト表示エリア720と同一のため、説明を省略する。

【0083】

削除ボタン920は、制約情報601を削除するためのボタンである。分析者が削除ボタン920を押下すると、ユーザインタフェース部101は、押下前に選択された制約情報601があれば、選択された制約情報601を制約一覧と制約記憶部111から削除する。押下前に選択された制約情報601がなければ、ユーザインタフェース部101は、選択された制約がない旨を出力する。

50

【0084】

修正ボタン930は、既存の制約情報601を修正するためのボタンである。分析者が修正ボタン930を押下すると、ユーザインタフェース部101は、押下前に選択された制約情報601があれば、制約情報601を編集対象制約情報として、後述する制約編集画面1000を表示し、押下前に選択された制約情報601がなければ、編集対象制約が選択されていない旨を出力する。

【0085】

追加ボタン940は、新たな制約情報601を追加するためのボタンである。分析者が追加ボタン940を押下すると、ユーザインタフェース部101は、編集対象制約情報として、制約IDと有効性が「有効」、自動生成フラグが「FALSE」、制約種別602が「選択」の制約情報601を生成し、後述する制約編集画面1000を表示する。このとき、ユーザインタフェース部101は、制約記憶部111に制約情報601が1以上、記憶されていない場合は「0」を、1以上、記憶されている場合は、制約記憶部111に記憶されている制約情報601の制約IDの最大値に「1」を加えた値を制約IDとする。なお、追加ボタン940を押下した時に、生成される制約の種別は、「選択」以外でも構わない。

【0086】

選択パターン例示エリア950は、制約を満たす選択パターンの例を表示するための領域であり、最大属性数、最小属性数、生成個数、選択パターン一覧を表示し、選択パターン例示ボタン951を含む。

10

20

【0087】

ユーザインタフェース部101が、「最大属性数」、「最小属性数」、「生成個数」のタイトル(項目名)、選択パターン一覧、選択パターン例示ボタン951を表示し、分析者が最大属性数、最小属性数、生成個数を入力する。

【0088】

図9は、最大属性数として「5属性」、最少属性数として「3属性」、生成個数として「1個」を入力し、選択パターンの例示として、「平均年収」、「周辺人口」、「所在市面積」の3つの属性からなる選択パターンを表示させた例である。

【0089】

選択パターン例示ボタン951は、分析者が入力した最大属性数、最小属性数、生成個数に基づいて、制約を満たす選択パターンを生成するためのボタンである。分析者が選択パターン例示ボタン951を押下すると、ユーザインタフェース部101は、選択パターン生成部103を用いて、制約記憶部111に記憶された制約情報601を満たし、最少属性数以上、最大属性数以下の属性を含む選択パターンを生成個数、生成し、選択パターン一覧に追加する。このとき、選択パターン一覧に既に選択パターンが追加されている場合、ユーザインタフェース部101は、既に追加されている選択パターンを削除してから追加する。

30

【0090】

終了ボタン960は、制約情報601の管理を終了するためのボタンである。分析者が終了ボタン960を押下すると、ユーザインタフェース部101は、制約管理画面900を閉じる。

40

【0091】

図10は、本実施例に関わる制約編集画面1000の例を示す図である。この図では、ユーザインタフェース部101が入出力装置123に表示する、編集対象制約を編集するための制約編集画面1000の例を表している。

【0092】

ユーザインタフェース部101は、制約表示エリア1010、矛盾制約表示エリア1020、終了ボタン1030を含む。

【0093】

制約表示エリア1010は、編集対象制約情報の表示、または編集を行うための領域で

50

あり、制約 ID、制約種別、対象属性、選択数、前提属性、有効性、自動生成フラグを表示し、制約種別を選択するためのコンボボックス、有効性を選択するためのラジオボタン、編集反映ボタン 1011 を含む。

【0094】

ユーザインターフェース部 101 が、「制約 ID」のタイトル(項目名)と実際の制約 ID、「制約種別」、「対象属性」、「選択数」、「前提属性」、「有効性」のタイトル(項目名)、「自動生成フラグ」のタイトル(項目名)、実際の自動生成フラグ、コンボボックス、ラジオボタンを表示し、分析者が制約種別、対象属性、選択数、前提属性、有効性を入力する。

【0095】

さらに、選択数は、制約種別として「選択制約」を選択しているときのみ入力可能であり、前提属性は、制約種別として「付帯制約」を選択しているときのみ入力可能である。

【0096】

また、制約編集画面 1000 が表示されたとき、制約表示エリア 1010 には、編集対象制約情報の制約 ID、制約種別、対象属性、選択数、前提属性、有効性、自動生成フラグが表示される。なお、編集対象制約情報が、選択制約以外の場合、選択数は空欄になり、編集対象が付帯制約以外の場合、前提属性は空欄になる。

【0097】

図 10 は、編集対象制約情報として、制約 ID に「2」、制約種別に「選択制約」、対象属性に、「周辺人口、最高気温」、有効性に「有効」、自動生成フラグに「FALSE」を表示し、分析者が選択数を「1」から「2」に変更した例である。

【0098】

編集反映ボタン 1011 は編集後の編集対象制約情報の内容を制約記憶部 111 に反映させるためのボタンである。分析者が編集反映ボタン 1011 を押下すると、ユーザインターフェース部 101 は、編集対象制約情報と同じ制約 ID を持つ制約記憶部 111 の制約情報 601 を編集対象の制約と一致させる。このとき、編集対象制約と同じ制約 ID を持つ制約情報 601 が制約記憶部 111 に存在しない場合、新たな制約として、制約情報 601 を制約記憶部 111 に追加する。さらに、ユーザインターフェース部 101 は、矛盾検証部 106 を用いて、制約記憶部 111 に記憶されている有効な制約情報 601 の中で、編集対象制約と矛盾している制約情報 601 を探し、後述する矛盾制約表示エリア 1020 の矛盾制約リストに表示する。

【0099】

本実施例では、1 以上の制約の集合において、制約の集合に含まれている制約を全て満たす属性の選択パターンが存在していないとき、制約の集合は、矛盾を含んでいるという。

【0100】

ここで、制約の集合が矛盾を含んでいるか判断するには、集合内に含まれる制約の論理式を論理積で結合した論理式(5)(以下、矛盾検証式という)の充足可能性を判定すれば良い

【0101】

【数 5】

$$\bigwedge_{\text{制約} \in \text{制約の集合}} \text{制約の論理式} \quad (5)$$

充足可能性とは、与えられた論理式を TRUE(真)にする命題変数の値の組合せ(以下、モデルという)が存在するか否かを表す性質である。矛盾検証式が充足可能とは、制約を満たす属性の選び方が存在することを意味している。

【0102】

充足可能性は、単純には、すべての変数の値の組合せをチェックすることで判定可能である。また、DPLL(Davis-Putnam-Logemann-Loveland)

10

20

30

40

50

) アルゴリズムなど、変数の値の組合せではなく、論理式の構造に着目して、効率的に充足可能性を判定するツールがある。たとえば、SAT (Satisfiability) ソルバは、充足可能性判定を判定するツールの一種であり、選択パターン生成用論理式の充足可能性を判定できる。また、SATソルバには、論理式が充足可能だった場合、モデルを取得する機能を備えるものがある。

【0103】

制約の集合の矛盾検証式が充足不能の場合、その制約の集合は矛盾を含んでおり制約の集合の矛盾検証式が充足可能の場合、その制約の集合は矛盾を含んでいない。さらに、矛盾検証式が充足不能な制約の集合において、どの属性を集合から除去しても、除去後の制約の集合の矛盾検証式が充足可能となる場合、そのような制約の集合を特に、矛盾検証式が充足不能となる極小な制約の集合という。以降において、矛盾とは、矛盾判定が充足不能となる極小な制約の集合を指すこととする。10

【0104】

ユーザインタフェース部101は、矛盾検証部106を用いて、編集対象制約情報と制約記憶部111に記憶されている有効な制約情報601の中で、編集対象制約情報を含む矛盾を探し、矛盾に含まれる編集対象制約情報以外の制約情報601を矛盾制約表示エリア1020に編集対象制約と矛盾している制約として表示する。

なお、矛盾を探すには、例えば「Thai Son Hoangらの“Validating the Consistency of Specification Rules”、<http://deploy-eprints.ecs.soton.ac.uk/465/1/ruleval-techrep.pdf>」に記載のMinimal Unsatisfiable Coreを探す方法を用いればよく、詳細はそちらを参照されたい。20

【0105】

矛盾制約表示エリア1020は、編集対象制約情報と矛盾している制約を表示するための領域であり、矛盾制約リスト1021、一括無効ボタン1022を含む。

【0106】

ユーザインタフェース部101は、制約記憶部111に記憶されている制約情報601の中で、制約表示エリア1010に表示されている制約と矛盾している制約情報601と一緒に一括無効ボタン1022を表示する。

【0107】

矛盾制約リスト1021は、編集対象制約情報と矛盾している制約を表示するためのリストであり、1行が、1つ以上の制約情報601の集合と対応する。分析者が、制約情報601の集合の中の制約情報601を押下すると、ユーザインタフェース部101は、押下された制約情報601を編集対象制約情報として、制約編集画面1000を表示する。30

【0108】

一括無効ボタン1022は、矛盾制約リスト1021中の全制約の集合の何れかに含まれる制約情報601の有効性を一括で無効にするためのボタンである。分析者が一括無効ボタン1022を押下すると、ユーザインタフェース部101は、制約記憶部111に記憶されている制約情報601の中で、矛盾制約リスト1021の制約情報601の集合に含まれる制約情報601の有効性を「無効」に変更する。40

【0109】

図10は、矛盾している制約の集合として、「付帯制約：implification(周辺人口){平均年収}」と「排他制約：at-mst-ne{平均年収、最高気温}」を持つ集合を表示させている例である。

【0110】

終了ボタン1030は、制約編集を終了するためのボタンである。分析者が終了ボタン1030を押下すると、ユーザインタフェース部101は、制約編集画面1000を閉じる。

【0111】

図11は、本実施例に関わる分析結果表示画面1100の例を示す図である。この図で50

は、ユーザインターフェース部 101 が入出力装置 123 に表示する、データ分析の結果、得られた仮説を表示するための分析結果表示画面 1100 の例を表している。

【0112】

分析結果表示画面 1100 は分析パラメータ表示エリア 1110、選択パターン表示エリア 1120、仮説リスト表示エリア 1130 を含む。

【0113】

分析パラメータ表示エリア 1110 は、データ分析設定画面 700 の分析パラメータ設定エリア 710 で分析者が入力した結果を表示する領域であり、分析パラメータ設定エリア 710 と内容が同一のため、説明を省略する。

【0114】

選択パターン表示エリア 1120 は、選択パターン記憶部 114 に記憶された選択パターン情報を表示するための領域であり、選択パターン一覧を表示し、選択パターン表示条件設定エリア 1121 を含む。分析者が選択パターン一覧中の選択パターン情報 501 を選択すると、ユーザインターフェース部 101 は、仮説記憶部 113 から、選択パターン情報 501 を用いて、導出された仮説情報 401 を探し、後述する仮説リスト表示エリア 1130 の仮説一覧に表示させる。

【0115】

図 11 は、選択パターンとして、No. 1 {周辺人口、所在市面積、平均年収} と No. 2 {最高気温、降水量、平均年収} を表示させた例である。

【0116】

選択パターン表示条件設定エリア 1121 は、選択パターン一覧に表示させる選択パターン情報 501 の条件を設定するための領域であり、使用属性を表示する。

【0117】

ユーザインターフェース部 101 が、「使用属性」のタイトル（項目名）を表示し、分析者が属性を入力する。分析者が属性名を入力すると、ユーザインターフェース部 101 は、選択パターン記憶部 114 から、分析者が入力した属性名を持つ属性を含む選択パターン情報 501 を探し、選択パターン一覧に表示させる。

【0118】

図 11 は、使用属性として「平均年収」を入力した例である。

【0119】

仮説リスト表示エリア 1130 は、データ分析によって導出された仮説を表示するための領域であり、仮説一覧を表示し、仮説表示条件設定エリア 1131 を含む。仮説一覧には、仮説表示条件設定エリア 1131 に入力された条件を満たす仮説情報 401 を表示する。

【0120】

図 11 は、No. 1 {周辺人口、所在市面積、平均年収} の場合の複数の回帰式として「 $\text{売上} = -0.03 \times \text{周辺人口} + 18.27 \times \text{所在市面積} + 0.25 \times \text{平均年収} - 38.709.4$ 」、相関係数として「0.83」、決定係数として「0.69」を持つ仮説情報 401 等を表示させた例である。

【0121】

仮説表示条件設定エリア 1131 は、仮説一覧に表示させる仮説の条件を設定するための領域であり、属性、相関係数、決定変数を表示させる。

【0122】

ユーザインターフェース部 101 が、「属性」、「相関係数」、「決定変数」のタイトル（項目名）を表示し、分析者が、属性、相関係数、決定変数を入力する。属性、相関係数、決定変数が入力されると、ユーザインターフェース部 101 は、選択パターン表示エリア 1120 の選択パターン一覧から分析者によって選択された選択パターン情報 501 を使って、導出された仮説情報 401 のうち、入力された属性を含む、相関係数が入力された範囲内であり、決定変数が入力された範囲内である仮説情報 401 を仮説記憶部 113 から探し、仮説一覧に表示させる。

10

20

30

40

50

【 0 1 2 3 】

図11は、表示させる仮説の相関係数の下限として、「0.7」を入力し、決定変数の上限として「0.7」を入力した例である。

【 0 1 2 4 】

すなわち、図11の分析結果表示画面1100が、選択パターンの一覧からデータ分析者が選択した選択パターンに基づいて、データ分析によって導出された仮説の一覧を表示する表示部としての機能を有する。

【 0 1 2 5 】

図12は、本実施例に関わるデータ分析のメインのフローチャートの例を示す図である。なお、図中の符号Sはステップを表す。

10

【 0 1 2 6 】

ユーザインタフェース部101は、分析者がデータ分析設定画面700の分析実行ボタン760を押下した場合に、処理を実行する(S1200)。

【 0 1 2 7 】

ユーザインタフェース部101は、分析パラメータ設定エリア710の統計情報チェックボックスにチェックが入っているかを確認する(S1201)。

【 0 1 2 8 】

当該チェックボックスにチェックが入っている場合、ユーザインタフェース部101は、S1202に進み、チェックが入っていない場合、ユーザインタフェース部101は、S1203に進む。

20

【 0 1 2 9 】

ユーザインタフェース部101は、統計分析部102を用いて、統計情報から制約を生成する(S1202)。

【 0 1 3 0 】

ユーザインタフェース部101は、選択パターン生成部103を用いて、選択パターンをひとつ生成する(S1203)。

【 0 1 3 1 】

ユーザインタフェース部101は、選択パターンの生成に成功したかどうかを確認する(S1204)。

【 0 1 3 2 】

生成に成功した場合、ユーザインタフェース部101は、S1205に進み、生成に失敗した場合、ユーザインタフェース部101は、S1209に進む。

30

【 0 1 3 3 】

ユーザインタフェース部101は、データ分析をするデータ分析部104を用いて、生成した選択パターン中に含まれる属性及び属性値に対して、データ分析を適用する(S1205)。

【 0 1 3 4 】

ユーザインタフェース部101は、分析パラメータ設定エリア710の履歴チェックボックスがチェックされており、かつ、当該エリアの分析周期を満たしているか確認する(S1206)。

40

【 0 1 3 5 】

満たしている場合、ユーザインタフェース部101は、S1207に進み、満たしていない場合、ユーザインタフェース部は、S1208に進む。

【 0 1 3 6 】

ユーザインタフェース部101は、履歴分析部105を用いて、過去に生成した選択パターンの履歴から、制約を生成する(S1207)。

【 0 1 3 7 】

ユーザインタフェース部101は、分析パラメータ設定エリア710の終了条件を満たしているか確認する(S1208)。

【 0 1 3 8 】

50

終了条件を満たす場合、ユーザインタフェース部は、S1209に進み、満たさない場合は、S1203に進む。

【0139】

ユーザインタフェース部101は、分析結果表示画面1100を表示し(S1209)、処理を終了する(S1210)。

【0140】

図13は、本実施例に関わる統計情報からの制約生成の処理フローチャートの例を示す図である。この図では、統計分析部102がデータの統計的な分析結果から制約を生成する処理の流れを表している。

【0141】

統計分析部102は、ユーザインタフェース部101から起動されることで、処理を開始する(S1300)。

【0142】

統計分析部102は、データ記憶部112から、説明変数を対象属性集合として取得する。(S1301)

統計分析部102は、対象属性集合中の各属性に対して、分析対象ごとの属性値の出現頻度を算出する(S1302)。ここで、出現頻度とは、ユーザインタフェース部101が受け付けた説明変数を統計的に分析した結果の値である。

【0143】

出現頻度は、「属性値iの出現頻度 = 属性値iを含む分析対象数 / 分析対象数」という式で算出できる。ここで、属性値iは、任意の属性値である。

【0144】

統計分析部102は、出現頻度が閾値を超える属性値を持つ属性(以下、除去対象制約)に対して、属性を選択しないことを意味する除去制約「n t {除去対象制約}」を生成する(S1303)。ここで、閾値は、固定値でもパラメータとして、分析者が事前に与える形でも良い。

【0145】

統計分析部102は、説明変数を統計的に分析した結果である出現頻度に基づいて生成した除去制約に対応する制約情報601を制約記憶部111に追加する(S1304)。このとき、制約IDは、制約記憶部111に記憶されている制約情報601が使用していない数値とし、有効性は「有効」、自動生成フラグは「TRUE」とする。

【0146】

統計分析部102は、除去対象属性を対象属性集合から除去する(S1305)。

【0147】

統計分析部102は、対象属性集合中に含まれる2つの属性の組合せ(以下、属性のペアという)を全て生成する(S1306)。

【0148】

統計分析部102は、生成した属性のペアごとに、相関係数を算出する(S1307)。

。

【0149】

統計分析部102は、相関係数が、閾値を超える属性のペア(排他対象属性ペア)に対して、排他対象属性ペアから、高々1個の属性しか選択しないことを意味する排他制約「a t - m o s t - n e {排他対象属性ペア}」を生成する(S1308)。すなわち、統計分析部102は、属性間の相関関係に基づいて、前記排他制約を生成する。

【0150】

統計分析部102は、生成した排他制約に対応する制約情報601を制約記憶部111に追加し(S1309)、処理を終了する(S1310)。このとき、制約IDは、制約記憶部111に記憶されている制約情報601が使用していない数値とし、有効性は「有効」、自動生成フラグは「TRUE」とする。

【0151】

10

20

30

40

50

図14は、本実施例に関わる選択パターン生成の処理フローチャートの例を示す図である。この図では、選択パターン生成部103が、制約を満たす選択パターンを生成する処理の流れを表している。

【0152】

選択パターン生成部103は、ユーザインターフェース部101から起動されることで、処理を開始する(S1400)。

【0153】

選択パターン生成部103は、制約記憶部111から、有効となっている制約情報601を全て取得する(S1401)。

【0154】

選択パターン生成部103は、有効となっている制約情報601を使って、選択パターンを生成するための選択パターン生成用論理式を生成する(S1402)。

【0155】

選択パターン生成用論理式(6)は、「説明変数から分析パラメータ設定エリア710の最大属性数以下の属性を選択することを意味する論理式(7)(以下、最大属性数制約という)」と「説明変数から分析パラメータ設定エリア710の最小属性数以上の属性を選択することを意味する論理式(8)(以下、最少属性数制約という)」と「既に生成した選択パターンと同一の選択パターンを生成しないことを意味する論理式(9)(以下、重複禁止制約という)」と「取得した制約情報601の論理式表現を論理積で繋げた式」を論理積で繋げた論理式である

【0156】

【数6】

$$\text{最大属性数制約} \wedge \text{最小属性数制約} \wedge \text{重複禁止制約} \wedge \bigwedge_{\text{制約} \in \text{取得した制約の集合}} \text{制約の論理式} \quad (6)$$

【0157】

【数7】

$$\text{at-most}(\text{属性}, \text{最大数}) \quad (7)$$

【0158】

【数8】

$$\text{at-least}(\text{属性}, \text{最小数}) \quad (8)$$

【0159】

【数9】

$$\bigwedge_{\text{選択パターン} \in \text{全選択パターンの集合}} \neg \left(\left(\bigwedge_{\text{属性名}_i \in \text{選択パターン}} {}^P \text{属性名}_i \right) \wedge \left(\bigwedge_{\text{属性名}_j \in (\text{全有効な属性}-\text{選択パターン})} \neg {}^P \text{属性名}_j \right) \right) \quad (9)$$

最大属性数制約は、属性に対応する命題変数のうち、分析パラメータ設定エリア710の最大属性数以下の数の命題変数を「TRUE」にすることを意味する基数制約である。

【0160】

最少属性数制約は、属性に対応する命題変数のうち、分析パラメータ設定エリア710の最小属性数以上の数の命題変数を「TRUE」にすることを意味する基数制約である。

【0161】

重複禁止制約は、選択パターン記憶部114に記憶されている選択パターンごとの、属性パターンに含まれる属性に対応する命題変数のみが「TRUE」となることを否定する論理式、を論理積で結合した式である。

【0162】

10

20

30

40

50

選択パターン生成部 103 は、選択パターン生成用論理式の充足可能性を判定する (S 1403)。

【0163】

充足不能の場合、選択パターン生成部 103 は、「生成失敗」として処理を終了し (S 1404)、充足可能の場合、選択パターン生成部 103 は、S 1405 に進む。

【0164】

選択パターン生成部 103 は、SAT ソルバからモデルを取得する (S 1405)。

【0165】

選択パターン生成部 103 は、モデルから、「TRUE」となっている命題変数を取得し、「TRUE」となっている命題変数と対応する属性の属性情報 301 の集合をデータ記憶部 112 から、選択パターン生成用論理式を満たす選択パターンを取得し (S 1406)、処理を終了する (S 1407)。 10

【0166】

すなわち、選択パターン生成部 103 は、有効となっている制約から、選択パターンを生成するための選択パターン生成用論理式を生成する生成部としての機能と、選択パターン生成用論理式を満たす選択パターンを取得する取得部としての機能を有する。

【0167】

図 15 は、本実施例に関わる分析ツール適用の処理フローチャートの例を示す図である。この図は、データ分析部 104 が、選択パターン生成部 103 で生成した選択パターンに対して、データ分析を適用するフローチャートを表している。 20

【0168】

データ分析部 104 は、ユーザインタフェース部 101 から起動されることで、処理を開始する (S 1500)。

【0169】

データ分析部 104 は、データ記憶部 112 から、選択パターンに含まれる属性情報 301 と関連がある分析対象ごとの属性値情報 303 を取得する (S 1501)。

【0170】

データ分析部 104 は、データ記憶部 112 から、目的変数の属性情報 301 と目的変数と関連がある分析対象ごとの属性値情報 303 を取得する (S 1502)。 30

【0171】

データ分析部 104 は、目的変数の属性情報 301、目的変数と関連がある分析対象毎の属性値情報 303、選択パターンに含まれる属性の属性情報 301、属性と関連がある分析対象ごとの属性値情報に対して、データ分析ツールを適用する (S 1503)。このとき、使用するデータ分析ツールは、固定でも、パラメータとして分析者が事前に指定する形でも良い。

【0172】

データ分析部 104 は、データ分析ツールの分析結果を仮説として取得する (S 1504)。このとき、データ分析ツールが相関係数や決定係数など、仮説の定量的な評価指標を出力しない場合、別の評価用ツールを使って、仮説の定量的な評価指標を取得しても良い。 40

【0173】

データ分析部 104 は、仮説から、仮説情報 401 を生成し、仮説記憶部 113 に記憶する (S 1505)。

【0174】

データ分析部 104 は、選択パターンから、選択パターン情報 501 を生成し、選択パターン記憶部 114 に記憶し (S 1506)、処理を終了する (S 1507)。

【0175】

図 16 は、本実施例に関わる履歴からの制約生成の処理フローチャートの例を示す図。この図は、履歴分析部 105 が、選択パターン記憶部 114 に記憶されている、過去に生成した選択パターンの履歴から、新たな制約を生成するフローチャートを表している。 50

【0176】

履歴分析部105は、ユーザインターフェース部101から起動されることで、処理を開始する(S1600)。

【0177】

履歴分析部105は、自動生成された付帯制約に対応する制約情報601を制約記憶部111から削除する(S1601)。

【0178】

履歴分析部105は、まだ選択されていない選択パターンを生成するための論理式である未選択パターン生成用論理式を生成する(S1602)。ここで、未選択パターン生成用論理式は、「データ記憶部112内の属性のうち、分析パラメータ設定エリア710の分析カバレッジに指定された数だけ属性を選択することを意味する論理式(選択制約：select(分析カバレッジに指定された数){データ記憶部112内の属性}の論理式表現と等価)」、「過去の選択パターンから、分析パラメータ設定エリア710の分析カバレッジに指定された数の属性を選択してできる組合せを除去することを意味する論理式(以下、過去の組合せ除去制約という)」、「制約記憶部111に記憶された有効な制約の制約情報601の論理式表現を論理積で繋げた式」を論理積で結合した式である。過去の組合せ除去制約は、選択パターン記憶部114に記憶された各選択パターンに対する、選択パターン含まれる属性に対応する命題変数のうち、TRUEとなる命題変数の数が、分析パラメータ設定エリア710の分析カバレッジと等しいことを意味する基数制約(選択制約：select(分析カバレッジに指定された数){選択パターン含まれる属性}の論理式表現と等価)の否定、を論理積で結合した式である。10

【0179】

履歴分析部105は、未選択パターン生成用論理式の充足可能性を判定する(S1603)。

【0180】

充足不能の場合、履歴分析部105は処理を終了し(S1604)、充足可能の場合、S1605に進む。

【0181】

履歴分析部105は、未選択パターン生成用論理式のモデルを取得する(S1605)30。

【0182】

履歴分析部105は、モデルから、TRUEとなっている命題変数を取得し、命題変数と対応する全属性の属性情報301をデータ記憶部112から取得し、取得した属性情報301のうち、ひとつを前提属性とし、それ以外を対象属性とする付帯制約「import_cat_in(前提属性){対象属性}」を生成する(S1606)。すなわち、履歴分析部105は、まだ選択されていない選択パターンを生成するための未選択パターン生成用論理式のモデル(与えられた論理式をTRUE(真)にする命題変数の値の組合せ)から付帯制約を生成する。

【0183】

履歴分析部105は、付帯制約に対応する制約情報601を生成し、制約記憶部111に追加する(S1607)。このとき、制約IDは、制約記憶部111に記憶されている制約情報601が使用していない数値とし、有効性は「有効」、自動生成フラグは「TRUE」とする。40

【0184】

履歴分析部105は、未選択パターン生成用論理式に対して、論理積を用いて、すでに取得済みのモデルが以降出現しないことを意味するモデル重複禁止制約を結合する(S1608)。モデル重複禁止制約は、モデルでTRUEとなっている命題変数を論理積で結合しさらに全体を否定した論理式である。

【0185】

本実施例の構成によれば、制約に基づいて属性の選択、及び、データ分析を自動化でき50

、データ分析に関わる工数を削減できるという効果を有する。

【0186】

なお、本発明は上記した実施例に限定されるものではなく、マーケティング以外の金融、交通、公共機関などのエンタープライズ系のデータ分析に適用されうる様々な変形例が含まれる。例えば、上記した実施例は本発明を分かりやすく説明するために詳細に説明したものであり、必ずしも説明した全ての構成を備えるものに限定されるものではない。また、ある実施例の構成の一部を他の実施例の構成に置き換えることが可能であり、また、ある実施例の構成に他の実施例の構成を加えることも可能である。また、各実施例の構成の一部について、他の構成の追加・削除・置換をすることが可能である。

【0187】

10

また、上記の各構成は、それらの一部又は全部が、ハードウェアで構成されても、プロセッサでプログラムが実行されることにより実現されるように構成されてもよい。また、制御線や情報線は説明上必要と考えられるものを示しており、製品上必ずしも全ての制御線や情報線を示しているとは限らない。実際には殆ど全ての構成が相互に接続されていると考えてもよい。

【符号の説明】

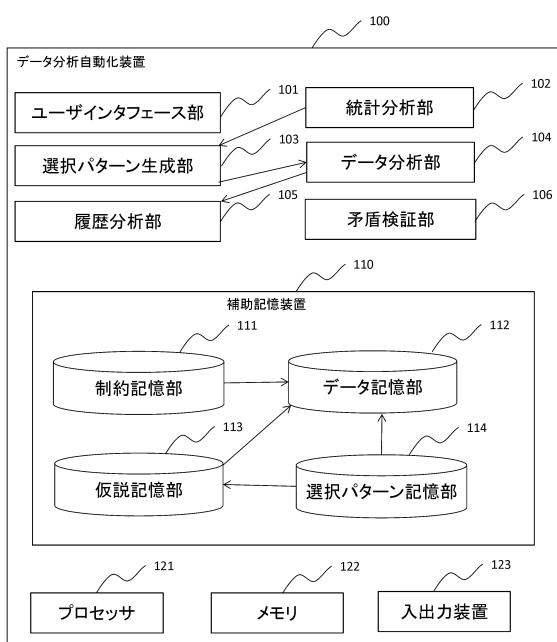
【0188】

- 101 ユーザインターフェース部
- 102 統計分析部
- 103 選択パターン生成部
- 104 データ分析部
- 105 履歴分析部
- 106 矛盾検証部
- 111 制約記憶部
- 112 データ記憶部
- 113 仮説記憶部
- 114 選択パターン記憶部

20

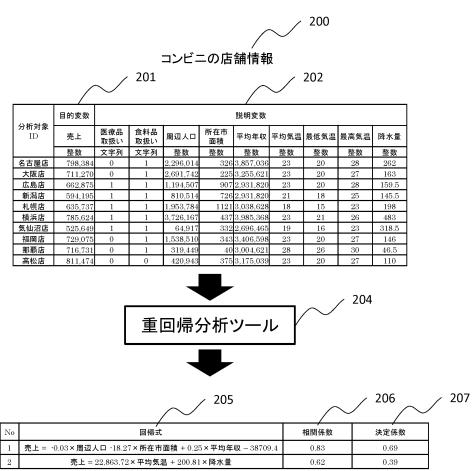
【図1】

図1



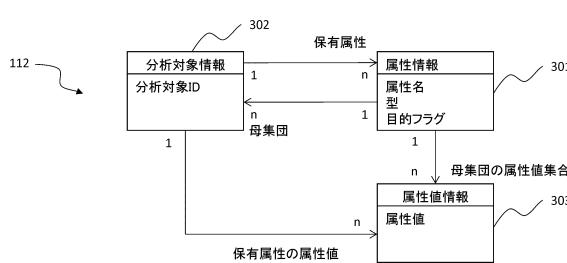
【図2】

図2



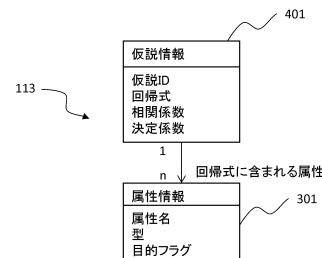
【図3】

図3



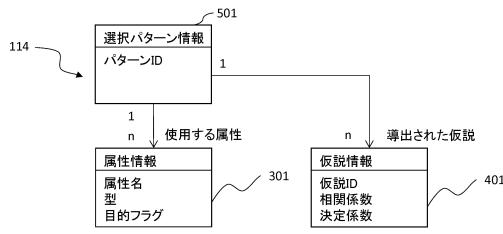
【図4】

図4



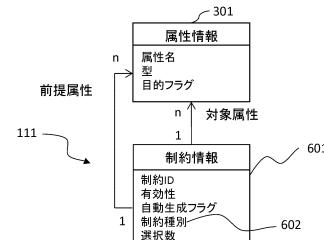
【図5】

図5



【図6】

図6



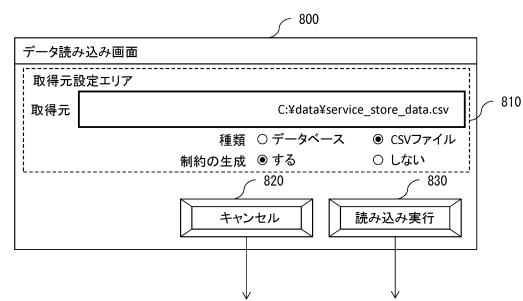
【図7】

図7



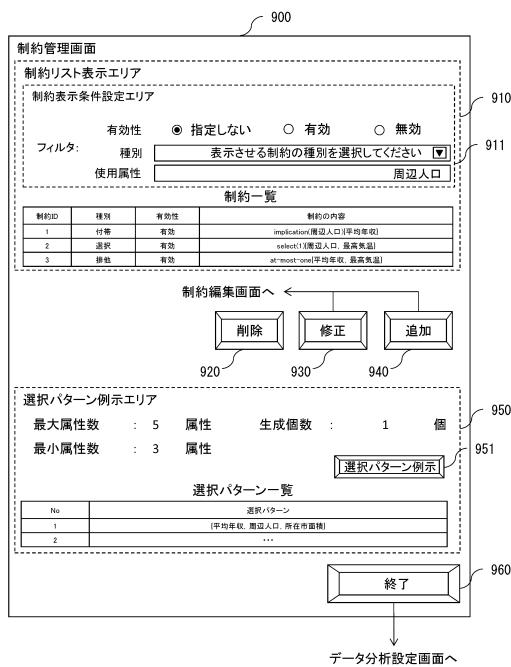
【図8】

図8

データ分析設定画面へ
読み込み後
データ分析設定画面へ

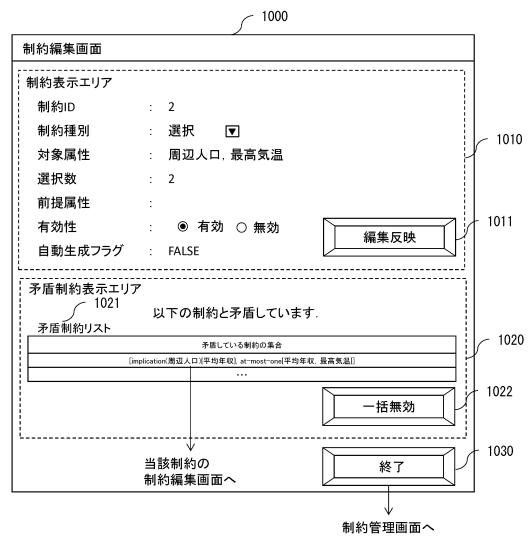
【図9】

図9



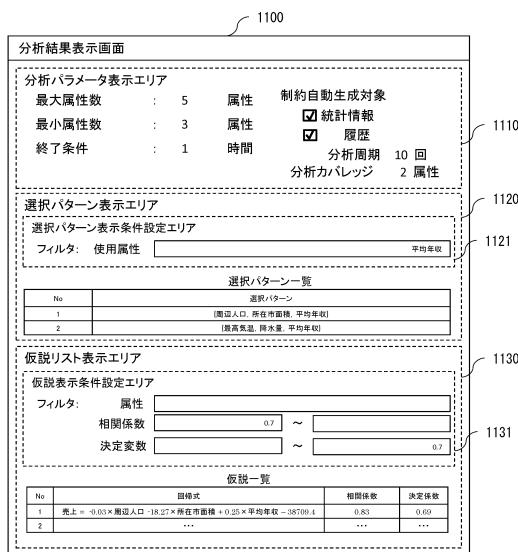
【図10】

図10



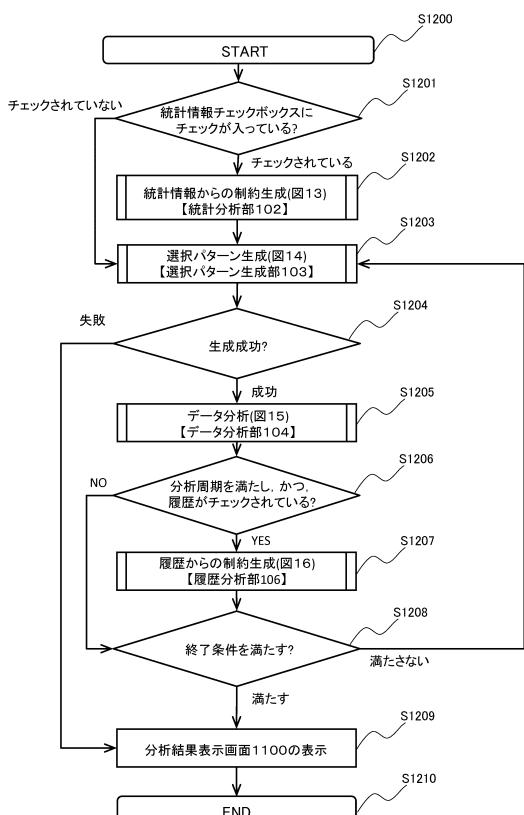
【図11】

図11



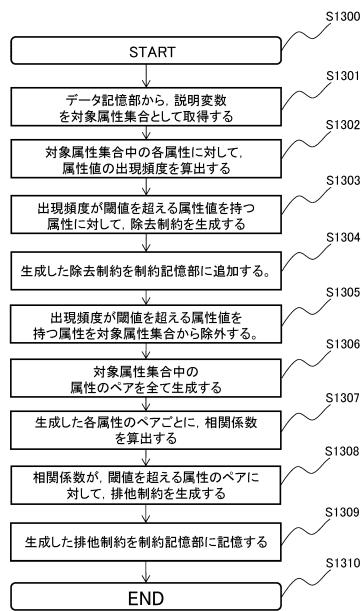
【図12】

図12



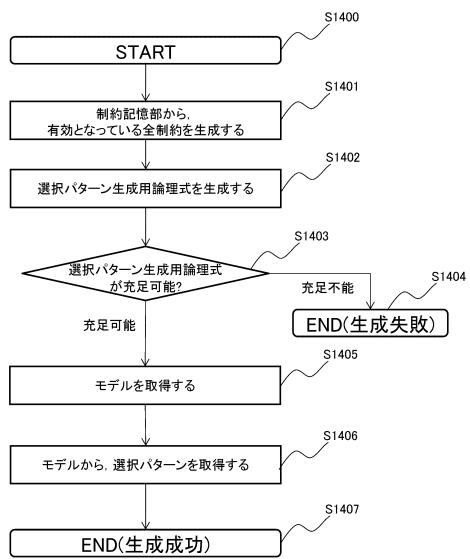
【図13】

図13



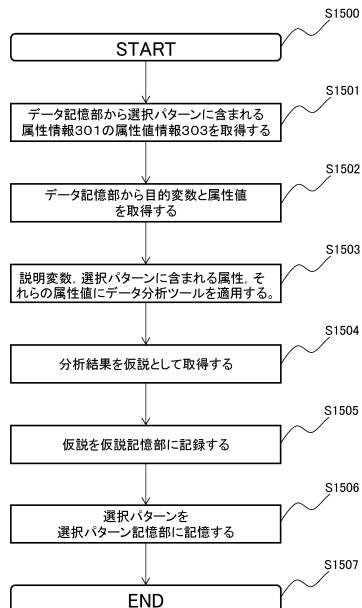
【図14】

図14



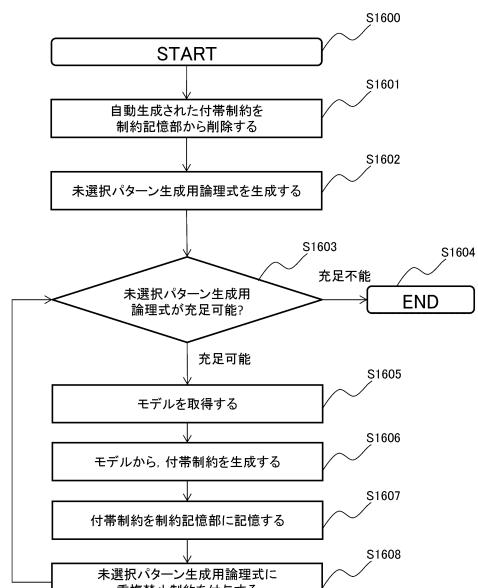
【図15】

図15



【図16】

図16



フロントページの続き

(72)発明者 安田 和矢
東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内

審査官 井上 宏一

(56)参考文献 特開2000-20504(JP,A)
特開平10-91612(JP,A)
特開2012-57195(JP,A)
斎 礼,外4名,「通販業における需要予測支援のためのエキスパートシステムの開発と評価」
,電気学会論文誌C,社団法人電気学会,1998年 2月 1日,第118-C巻,第2号,pp.242
-248

(58)調査した分野(Int.Cl., DB名)

G 06 F 17 / 18
G 06 F 16 / 28
G 06 Q 10 / 06