

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7247060号
(P7247060)

(45)発行日 令和5年3月28日(2023.3.28)

(24)登録日 令和5年3月17日(2023.3.17)

(51)国際特許分類

F I

G 0 6 Q 10/04 (2023.01)

G 0 6 Q 50/10 (2012.01)

G 0 6 F 16/907 (2019.01)

G 0 6 Q 10/04

G 0 6 Q 50/10

G 0 6 F 16/907

請求項の数 9 (全27頁)

(21)出願番号	特願2019-159980(P2019-159980)	(73)特許権者	000005108
(22)出願日	令和1年9月2日(2019.9.2)		株式会社日立製作所
(65)公開番号	特開2021-39523(P2021-39523A)		東京都千代田区丸の内一丁目6番6号
(43)公開日	令和3年3月11日(2021.3.11)	(74)代理人	110002365
審査請求日	令和4年3月16日(2022.3.16)		弁理士法人サンネクスト国際特許事務所
		(72)発明者	山本 秀典
			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	津野 高志
			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	齊藤 元伸
			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	関 博文

最終頁に続く

(54)【発明の名称】 データの利活用のためのデータ準備を支援するシステム、及び、その方法

(57)【特許請求の範囲】

【請求項1】

データの利活用のためのデータ準備を支援するシステムであって、
処理装置と、
記憶装置と、を備え、
前記処理装置は、前記記憶装置に記録されたプログラムを実行することによって、
複数の業務システムの夫々から業務データを収集し、当該業務データを前記記憶装置に
対象データとして少なくとも一時的に蓄積し、
ユーザ端末から前記対象データに対する分析目的を受信し、
前記分析目的と前記対象データとに基づいて、前記分析目的に対して推奨される分析ス
テップを決定し、
前記推奨される分析ステップに利用可能なデータの組合せを、前記対象データから抽出
し、
前記抽出されたデータの組合せの複数の夫々を評価し、
当該評価の結果に基づいて、前記データの組合せの複数の夫々順位付けしたリストを作
成し、
当該リストを前記ユーザ端末に出力させ、
ユーザが前記リストから所定のデータの組合せを選択できるようにし、
前記記憶装置は、前記対象データに対する基準情報を備え、
前記処理装置は、

10

前記推奨される分析ステップを前記基準情報に基づいて決定することと、
前記推奨される分析ステップを決定することを、前記対象データに対する分析の実施段階の複数タイプの中から、いずれかのタイプの分析の実施段階に決定することによって実行することと、
を有し、
前記処理装置は、
前記分析目的が、前記対象データ、又は、当該対象データ及び前記基準情報に適合する程度を算出し、
前記基準情報の前記分析目的に対する有効性の程度を算出し、
両方の算出結果に基づいて、前記いずれかのタイプの分析の実施段階を決定し、
前記分析目的はK P Iを含み、
前記基準情報は前記対象データに対する辞書データを含み、
前記処理装置は、
前記分析目的が前記対象データに適合する程度を算出することを、前記K P Iと前記対象データのカラム名とを比較することによって行い、
前記分析目的が前記基準情報に適合する程度を算出することを、前記K P Iと前記辞書データとを比較することによって行う、
システム。

10

【請求項 2】

前記分析目的は、前記複数のタイプの分析の実施段階のうち、ユーザが所望する分析の実施段階を含み、
前記処理装置は、
前記両方の算出結果に基づいて、当該ユーザが所望する分析の実施段階を評価し、この評価は当該分析の実施段階を維持するか、又は、他の分析の実施段階に変更することを含み、
前記評価の結果に基づいて、前記ユーザに、前記推奨される分析の実施段階を提示する、
請求項 1 記載のシステム。

20

【請求項 3】

データの利活用のためのデータ準備を支援するシステムであって、
処理装置と、
記憶装置と、を備え、
前記処理装置は、前記記憶装置に記録されたプログラムを実行することによって、
複数の業務システムの夫々から業務データを収集し、当該業務データを前記記憶装置に対象データとして少なくとも一時的に蓄積し、
ユーザ端末から前記対象データに対する分析目的を受信し、
前記分析目的と前記対象データとに基づいて、前記分析目的に対して推奨される分析ステップを決定し、
前記推奨される分析ステップに利用可能なデータの組合せを、前記対象データから抽出し、
前記抽出されたデータの組合せの複数の夫々を評価し、
当該評価の結果に基づいて、前記データの組合せの複数の夫々順位付けしたリストを作成し、
当該リストを前記ユーザ端末に出力させ、
ユーザが前記リストから所定のデータの組合せを選択できるようにし、
前記記憶装置は、前記対象データに関する基準情報を備え、
前記処理装置は、
前記推奨される分析ステップに利用可能なデータの組合せを、前記対象データから抽出することと、
前記対象データから前記分析目的に該当するデータを抽出することと、そして、
前記対象データから前記基準情報に該当するデータを抽出することと、から実行し、

30

40

50

前記分析目的に該当するデータと、前記基準情報に該当するデータと、に基づいて、前記データの組合せの複数を構成し、
当該データの組合せの複数の夫々を、データ量に基づいて評価し、
前記処理装置は、
前記対象データから前記分析目的に該当するデータを抽出することを、前記分析目的に該当するカラムを前記対象データから抽出することから実行し、
前記対象データから前記基準情報に該当するデータを抽出することを、前記対象データから前記基準情報に該当するカラムを前記対象データから抽出することから実行し、
前記データの組合せとして、前記分析目的に該当するカラムと前記基準情報に該当するカラムの組合せを複数構成し、
前記データ量として、当該複数の組合せ夫々のレコード数を算出する、
システム。

10

【請求項 4】

前記処理装置は、
前記分析の実施段階の複数タイプ毎に、前記抽出されたデータの組合せの複数の夫々を評価する、
請求項 1 記載のシステム。

【請求項 5】

前記処理装置は、
前記複数の組合せの夫々について、複数あるデータ値変動パタンのうちどのデータ値変動パターンを備えるかを判定し、
データ値変動パターン毎に前記レコード数を算出する、
請求項 3 記載のシステム。

20

【請求項 6】

前記処理装置は、
前記対象データのカラムを前記辞書データの 4 W 情報の夫々に基づいて抽出する、
請求項 1 記載のシステム。

【請求項 7】

ユーザによる分析目的情報の登録に対する、推奨する分析ステップ、または、作業項目に関する情報、さらに各分析ステップにおけるユーザが指定する K P I や着目データ項目に関連する利用可能なデータの組合せに関する情報を、ユーザに提示するための出力装置を有する、請求項 1 記載のシステム。

30

【請求項 8】

データの利活用のためのデータ準備を支援する方法であって、
コンピュータは、
複数の業務システムの夫々から業務データを収集し、当該業務データを記憶装置に対象データとして少なくとも一時的に蓄積し、
ユーザ端末から前記対象データに対する分析目的を受信し、
前記分析目的と前記対象データとに基づいて、前記分析目的に対して推奨される分析ステップを決定し、
前記推奨される分析ステップに利用可能なデータの組合せを、前記対象データから抽出し、
前記抽出されたデータの組合せの複数の夫々を評価し、
当該評価の結果に基づいて、前記データの組合せの複数を夫々順位付けしたリストを作成し、
当該リストを前記ユーザ端末に出力させ、
ユーザが前記リストから所定のデータの組合せを選択できるようにし、
前記記憶装置は、前記対象データに関する基準情報を備え、
前記コンピュータは、
前記推奨される分析ステップを前記基準情報に基づいて決定することと、

40

50

前記推奨される分析ステップを決定することを、前記対象データに対する分析の実施段階の複数タイプの中から、いずれかのタイプの分析の実施段階に決定することによって実行することを有し、

前記コンピュータは、

前記分析目的が、前記対象データ、又は、当該対象データ及び前記基準情報に適合した程度を算出し、

前記基準情報の前記分析目的に対する有効性の程度を算出し、

両方の算出結果に基づいて、前記いずれかのタイプの分析の実施段階を決定し、

前記分析目的はK P Iを含み、

前記基準情報は前記対象データに対する辞書データを含み、

前記コンピュータは、

前記分析目的が前記対象データに適した程度を算出することを、前記K P Iと前記対象データのカラム名を比較することによって行い、

前記分析目的が前記基準情報に適合する程度を算出することを、前記K P Iと前記辞書データとを比較することによって行う、

前記方法。

【請求項 9】

データの利活用のためのデータ準備を支援する方法であって、

コンピュータは、

複数の業務システムの夫々から業務データを収集し、当該業務データを記憶装置に対象データとして少なくとも一時的に蓄積し、

ユーザ端末から前記対象データに対する分析目的を受信し、

前記分析目的と前記対象データとに基づいて、前記分析目的に対して推奨される分析ステップを決定し、

前記推奨される分析ステップに利用可能なデータの組合せを、前記対象データから抽出し、前記抽出されたデータの組合せの複数の夫々を評価し、

当該評価の結果に基づいて、前記データの組合せの複数の夫々順位付けしたリストを作成し、

当該リストを前記ユーザ端末に出力させ、

ユーザが前記リストから所定のデータの組合せを選択できるようにし、

前記記憶装置は、前記対象データに関する基準情報を備え、

前記コンピュータは、

前記推奨される分析ステップに利用可能なデータの組合せを、前記対象データから抽出することを、

前記対象データから前記分析目的に該当するデータを抽出することと、そして、

前記対象データから前記基準情報に該当するデータを抽出することと、から実行し、

前記分析目的に該当するデータと、前記基準情報に該当するデータと、に基づいて、前記データの組合せの複数の構成し、

当該データの組合せの複数の夫々を、データ量に基づいて評価し、

前記コンピュータは、

前記対象データから前記分析目的に該当するデータを抽出することを、前記分析目的に該当するカラムを前記対象データから抽出することから実行し、

前記対象データから前記基準情報に該当するデータを抽出することを、前記対象データから前記基準情報に該当するカラムを前記対象データから抽出することから実行し、

前記データの組合せとして、前記分析目的に該当するカラムと前記基準情報に該当するカラムとの組合せを複数構成し、

前記データ量として、当該複数の組合せ夫々のレコード数を算出する、

前記方法。

【発明の詳細な説明】

【技術分野】

10

20

30

40

50

【 0 0 0 1 】

本発明は、データの利活用を支援するためのシステム、特に、データの利活用のためのデータ準備を支援するシステムに係り、例えば、複数の業務システムが有する多種、大量のデータの中から、データ分析の目的に沿ったデータを選出、抽出したりするのに好適なシステムに関する。

【背景技術】

【 0 0 0 2 】

昨今、企業内に蓄積された多種多量のデータを利用して様々な業務課題を改善することが行われている。例えば、特開2004-29971号公報には、生産工程における生産装置や処理時刻のデータの中から所望するデータ解析に必要なデータのみを容易に抽出して、歩留り向上に有効な解析結果を得るために、データの説明変数に対しデータ項目のカテゴリを識別する付加文字列を付加し、データの異常値を特定値に置換あるいは削除するデータクレンジングを行い、データの目的変数の変動に基づく特徴情報を得て、解析処理時にデータのカテゴリを認識しカテゴリに対応した条件設定及び解析手順によりデータ解析を効率的に自動実行するデータ解析方法が提案されている。

10

【 0 0 0 3 】

そして、特開2016-181150号公報には、要求される分析の変更に対応自在のデータ処理を行うために、入力されたすべての入力データを格納するデータウェアハウスと、入力データを統合して統合データを生成した後、統合データを格納する統合レイヤと、統合データを、不加算項目の1つ以上の組合せ毎に、少なくとも加算項目の数量又は不加算項目の数を集計して複数の集計データを生成した後、複数の集計データを格納する集計レイヤと、設定部で設定された分析データの生成に必要な条件に基づき、複数の集計データから1つの集計データを選択し、さらに1つの集計データから分析データを抽出した後、分析データを格納する分析レイヤと、を有するデータ処理システムが提案されている。

20

【先行技術文献】

【特許文献】

【 0 0 0 4 】

【文献】特開2016-181150

【発明の概要】

【発明が解決しようとする課題】

30

【 0 0 0 5 】

近年、交通、電力、産業等の多くの分野において、多様化、多角化する様々な問題を解決するために、部署や業務を横断して収集された業務データを活用することが求められている。一方、夫々別々に構築、運用されてきた、複数の業務システム毎に異なる多種、大量のデータを扱うために、ユーザにはデータの理解や業務に対する経験や知識が必要であるものの、そのレベルは人によって異なり、そのことが、データを用いた課題の分析を進める上で妨げになっている。例えば、あるユーザは、自身が属する部署や自身の業務に係るデータを熟知していても、他の部署や他の業務に係るデータを把握できていないことが多く、その結果、複数の部署や業務を跨いで、ユーザが所望する分析に必要なデータを選出したり、データを加工したりする等、分析のための準備は困難になる。

40

【 0 0 0 6 】

そこで、複数の業務毎の違い、業務毎のデータやシステムの違いに対する理解や知識がユーザに不足していても、ユーザが、円滑にデータの分析に着手できるように、データの準備のための作業負担を低減させることが望まれる。

【 0 0 0 7 】

上記特許文献1に開示された発明は、ユーザが対象データを理解していることを前提とするものであって、そのために、分析目的、分析手段、及び、分析のために用意すべきデータの内容を、ユーザが事前に整理しておくことが必要であり、特定種類のデータに対して、ユーザが想定した目的の下で、当該データが活用できるに過ぎない。一方、特許文献2に開示された発明に活用できるデータは、統合データになり得るデータに限られる。複

50

数の業務システムからの多種多様なデータを一様に統合できるとは限らない。また統合データ、集計データから目的に合った分析データを作成するためには、ユーザは、元のデータを全て理解していることが必要となる。

【 0 0 0 8 】

本発明は、複数の業務システムから収集したデータの利活用を促進するために、データの利活用のためのデータの準備を支援するシステム、及び、その方法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】

【 0 0 0 9 】

前記目的を解決するために、本発明は、データの利活用のためのデータ準備を支援するシステムであって、処理装置と、記憶装置と、を備え、前記処理装置は、前記記憶装置に記録されたプログラムを実行することによって、複数の業務システムの夫々から業務データを収集し、当該業務データを前記記憶装置に対象データとして少なくとも一時的に蓄積し、ユーザ端末から前記対象データに対する分析目的を受信し、前記分析目的と前記対象データとに基づいて、前記分析目的に対して推奨される分析ステップを決定し、前記推奨される分析ステップに利用可能なデータの組合せを、前記対象データから抽出し、前記抽出されたデータの組合せの複数の夫々を評価し、当該評価の結果に基づいて、前記データの組合せの複数の夫々順位付けしたリストを作成し、当該リストを前記ユーザ端末に出力させ、ユーザが前記リストから所定のデータの組合せを選択できるようにした。

【発明の効果】

【 0 0 1 0 】

本発明によれば、複数の業務システムの多種多様、大量のデータに基づいて、課題解決などの分析等のために、これらデータの利活用を進める際のデータ準備作業に於けるユーザの負荷を軽減し、以って、データの利活用を迅速、かつ、高品質に進めることができる。

【図面の簡単な説明】

【 0 0 1 1 】

【図 1】データの利活用のためのデータ準備を支援するシステム構成の一例を示すブロック図である。

【図 2】サーバがデータの利活用のためのデータ準備を支援する動作を実現するためのシーケンスの一例である。

【図 3】サーバのモジュール構成の一例である。

【図 4】ユーザが作成する分析目的情報、サーバにて保持する基準情報の構成の一例である。

【図 5 A】分析ステップとして全体像把握をユーザが指定すると、サーバが作成する、データ準備テーブルの一例である。

【図 5 B】分析ステップとして、“特異事象抽出”をユーザが指定すると、サーバが作成する、データ準備テーブルの一例である。

【図 5 C】分析ステップとして、“要因分析 / 予測”をユーザが指定すると、サーバが作成する、データ準備テーブルの一例である。

【図 5 D】サーバから、ユーザに提示、又は、提案される、データの組合せを管理する管理テーブルの一例を示す。

【図 6】サーバがユーザに分析ステップを提案するための動作の一例に係る、フローチャートである。

【図 7 A】ユーザが登録した分析ステップの適否を判定するためのマトリックスの第 1 の例である。

【図 7 B】ユーザが登録した分析ステップの適否を判定するためのマトリックスの第 2 の例である。

【図 7 C】ユーザが登録した分析ステップの適否を判定するためのマトリックスの第 3 の例である。

10

20

30

40

50

【図 8】サーバがユーザに、分析ステップとしての全体像把握に利用可能なデータの組合せを、ユーザに提示するための動作の一例を示すフローチャートである。

【図 9】サーバが、分析ステップとしての特異事象抽出に利用可能なデータの組合せを、ユーザに提示するための動作の一例を示すフローチャートである。

【図 10】サーバが、分析ステップとしての要因分析 / 予測に利用可能なデータの組合せを、ユーザに提示するための動作の一例を示すフローチャートである。

【図 11 A】サーバからユーザ端末に対して提供された、データ利活用を支援するためのグラフィカルユーザインターフェースの一例である。

【図 11 B】サーバからユーザ端末に対して提供された、データ利活用を支援するためのグラフィカルユーザインターフェースの他の例である。

【図 11 C】サーバからユーザ端末に対して提供された、データ利活用を支援するためのグラフィカルユーザインターフェースのさらに他の例である。

【図 12】図 8 のフローチャートにおいて、複数のカラムの組合せを説明するためのテーブルである。

【図 13】図 10 のフローチャートにおいて、複数のカラムの組合せを説明するためのテーブルである。

【発明を実施するための形態】

【0012】

次に、本発明の実施形態について説明する。図 1 は、本発明を実現するための計算機システムの一例の構成図である。この計算機システムは、複数の業務システムから収集したデータの利活用のためのデータ準備を支援するサーバ 101（コンピュータ）を備える。サーバ 101 は、システムを利用するユーザが入力した、データ分析の目的と、サーバ 101 が収集した、複数の業務システムのデータとを照合して、ユーザに対して、最適な分析ステップを提案し、合わせて、この分析ステップに利用可能な有用度の高いデータを提案することによって、データの利活用のためのデータ準備を支援する。

【0013】

ユーザは、システムに分析目的を登録するだけで、そして、システムから提案、又は、提供されたデータに基づいて、課題解決のため等の分析を実行することにより、ユーザ自身がデータを収集したり、数多くのデータの中から有用なデータを抽出する等のデータ準備に係る作業を強いられることなく、ユーザが意図している、データ分析を行うことができる。

【0014】

サーバ 101 には、複数の業務システム 104、105、106 と、複数のユーザ端末 102、103 とが接続されている。ユーザ端末は、ユーザが希望する分析目的をサーバ 101 に登録し、次いで、分析目的を実現する上での分析ステップ、及び、この分析ステップに利用可能なデータの提案を、サーバ 101 から受け、次いで、これらを確認後、データ分析を実施する、もしくは、サーバ 101 に対してデータ分析を実施させることができる。

【0015】

複数の業務システムの夫々から、サーバ 101 に、ユーザが所望するデータ分析の対象として、サーバ 101 が参照するデータ（以下、対象データ、という。）が提供される。なお、複数の業務システムの夫々からサーバ 101 へのデータ提供はネットワーク 108 を介した通信により実施する場合だけでなく、ネットワーク 108 を介さず、例えば、人手を介してのサーバ 101 へのデータ格納を行うことでもよい。複数の業務システムとしては、鉄道分野を例とすると、例えば、列車の運行管理システム、駅の管理システム、そして、保守管理システムを例示することができる。

【0016】

業務システムからサーバ 101 に送られるデータは、特に、制限されなくてよく、例えば、データテーブル、その他、バイナリデータとしてのセンサ情報等、データの種類、データの方式等が制限されない。ユーザは、複数の業務システムを横断して、所定の課題や

10

20

30

40

50

問題点の原因を総合的に探るために、サーバ101に、データ分析の目的を登録する。この目的としては、輸送業を例にすると、例えば、“列車の遅延”があり、即ち、遅延の原因を、複数の業務システムを横断して特定し、以って、遅延の予防、抑制に役立てようとすることがある。

【0017】

サーバ101と、複数のユーザ端末102、103、・・・の夫々とはネットワーク107を介して接続されている。サーバ101と、複数の業務システム104、105、106、・・・の夫々とはネットワーク108を介して相互接続されている。

【0018】

データ利活用、及び、データ準備のための基盤としてのサーバ101の主なハードウェア構成は、記憶装置（メインメモリ、ハードディスク、外部ストレージシステムの記憶領域）111、処理装置（CPU）112、そして、通信装置113である。ユーザ端末102、103も同様である。処理装置112は、非一時的記録媒体（記憶装置）111に対象データを少なくとも一時的に記録する。

10

【0019】

図2は、サーバ101がデータの利活用のためのデータ準備を支援する動作を実現するためのシーケンスの一例である。なお、ユーザ201は、複数の部署を横断して収集された様々なデータに対して、所定の分析ステップに基づいて、問題発見、解決策の立案、等を行おうとしている者、又は、そのためのアプリケーションであってよい。

【0020】

20

複数の業務システム104 - 106の夫々は、サーバ101に業務データを登録する(211)。サーバ101は、業務データの記憶領域にアクセスして、業務データを参照し、データカタログ、データ定義を設定する(221)。サーバ101がデータカタログ等を設定するとは、データカタログ等を作成すること、又は、更新することを含む。サーバ101は、後述のとおり、ユーザが入力した分析目的と、自身が作成した基準情報（データカタログ、データ定義）とに基づいて、分析態様（分析ステップ、候補データリスト）を作成して、これをユーザに提示する。

【0021】

データカタログ、そして、データ定義等の基準情報は、サーバ101が、ユーザ201が意図する分析目的と、業務データとに基づいて、ユーザ201に、最適な分析ステップを提案し、そして、この分析ステップに利用可能な有用度の高いデータを提案するために、業務データを分別する、評価する等、業務データの解析のための基準、又は、指標等の一例として理解されればよい。データカタログ、データ定義の詳細は後述する。

30

【0022】

ユーザ201は、ユーザ端末102を介して、所望する、データ分析の目的をサーバ101に登録する(231)。サーバ101は、業務システムデータ、そして、基準情報（データカタログ、データ定義）とに基づいて、ユーザが登録した分析目的を評価して、ユーザに推奨する分析ステップを判定、判別、決定、そして、選定する等して確定し、これをユーザ端末102に報知する(222)。

【0023】

40

ユーザ201は、サーバ101から提示され、推奨された分析ステップを確認する、又は、複数の分析ステップの中から所定の分析ステップを選択する等して、分析ステップを確定させて、これをサーバ101に報知する(232)。

【0024】

サーバ101は、複数の業務システムから収集したデータの中から、分析ステップに適した、複数の候補データを抽出して、これをユーザ201に提示する(223)。ユーザ201は、サーバ101から送信された、複数の候補データのリストを参照して、分析対象とするデータを決定、或いは、選択等によって設定して、確定した分析ステップ(232)に基づいて、分析対象データを分析する。また、これをサーバ101に報知する(233)。サーバ101は、ステップ222、ステップ223、そして、ステップ233の結果に

50

基づいて、基準情報（データカタログ／データ定義）を更新する（224）。

【0025】

データカタログ、データ定義等の基準情報は、サーバ101の管理者、又は、運営者によって作成、又は、更新されてよい。サーバ101は、ユーザの分析目的、業務データに基づく機械学習を行うことによって、データカタログ等を作成、又は、更新してもよい。

【0026】

分析目的（後述の図4：401）はデータ分析のターゲットを端的に示す情報であってよく、例えば、“列車遅延”である。ユーザ201は、分析目的を、例えば、“分析ステップ”、“KPI”、そして、“着目データ項目”によって記述してサーバに登録することができる。“分析ステップ”、そして、“KPI”は必須情報とし、“着目データ項目”は任意情報とし10てよい。KPIは、データの分析において、ユーザが注目する要素であり、例えば、既述の“列車遅延”である。KPIは、キーワードの他、算出式によって定義されてもよい。

【0027】

着目データ項目は、例えば、KPIを評価するための軸として着目されるべき要素、又は、KPIと同時に監視すべき名称である。例えば、KPIが“列車遅延”として、着目データ項目は、“発車時刻”、“到着時刻”、“運行日時”、そして、“遅延時分”の少なくとも一つである。

【0028】

分析ステップは、業務上の問題解決等のために分析を進める上での実施段階を表したものであり、複数のタイプがあり、主として、対象データの全体像を把握するための態様、対象データから特異事象を抽出するための態様、そして、対象データから要因を分析して、結果を予測する態様が例示される。これらを、以後、(1)全体像把握、(2)特異事象抽出、(3)要因分析／予測、ということとする。

20

【0029】

ユーザ201は、複数の分析ステップを分析目的に含ませることができる。複数のステップには、優先度、順番等の優劣の区別があってもよい。

【0030】

前記分析ステップの夫々の態様について、ユーザ201が用い得る計算手法として、例えば、全体像把握には傾向分析、特異事象抽出には外れ値検出、及び／又は、変化点抽出、要因分析／予測には、回帰分析／重回帰分析、及び／又は、分類／クラスタ分析がある。

30

【0031】

サーバ101は、記憶装置111に、基準情報（“定義情報”と称してもよい）として、データカタログ、データ定義（221）を有する。データカタログ（図4：402）には、サーバ101が複数の業務システム104から収集した対象データのデータカタログと、対象データを加工した加工データのデータカタログとがある。加工データのデータカタログには、対象データから加工データを得るための算出式に関する情報を含んでよい。

【0032】

対象データに関するカタログとして、例えば、“カラム名”、“名称・意味”、“補足情報”、等の項目からなるものがある。一例として、{arrival time,到着時刻,・・・}、{KRT,キロ程,・・・}がある。なお、カラム名とは、対象データ（データテーブル）のカラムに記録された名称である。

40

【0033】

加工データに関するカタログとして、“データ名”、“算出式”、“名称・意味”、“補足情報”等の項目からなるものがあり、一例として、{delay,“実績時刻”-“計画時刻”,遅延時分,・・・}がある。

【0034】

既述のデータ定義221は、例えば、4W辞書（図4：403）と、データ推移パターン（図4：404）と、を備える。4W辞書は、対象データの全体像の傾向を4W(When、Where、What、Who)の観点から把握するためのものであり、4Wのカテゴリ毎にキーワードを含む。例えば、4W辞書データは、“キーワード”、“カテゴリ”の組合せからなり、

50

一例として、{キーワード,カテゴリ}={時刻,when}、{date,when}、{キロ程,where}、{Kilometrage,where}、{駅コード,where}である。

【0035】

データ推移ボタンは、4W辞書に加えて値の推移を含むものであり、例えば、“データ名・種別”、“カテゴリ”、“データ型”、“値変化幅”、“下限”、そして、“上限”とからなる。一例として、{キロ程,where,integer,1,0,300}、{到着時刻(分),when,integer,5,0,1440}がある。

【0036】

図3は、サーバ101のモジュール構成の一例である。サーバ101には、既述の図2において説明した、諸機能を実現するためのミドルウェア301が実装されている。ミドルウェア301は、サーバ101の記憶装置111に保存されたプログラムを実行する処理装置112によって実現される、複数の機能モジュールを備えている。そして、ミドルウェア301は、業務システム104の対象データ311、対象データ311から特徴量として抽出される等して加工された加工データ312とを、記憶装置111に備えている。

10

【0037】

ミドルウェア301は、対象データ311、加工データ312を管理するデータ管理モジュール321を備え、データカタログ402、データ定義315、そして、ユーザ201が登録した分析目的の適否を判定するためのマトリックス313を記憶装置111に備えている。

【0038】

ミドルウェア301は、既述の基準情報(データカタログ、データ定義)を管理する基準情報管理モジュール322、サーバ101にアクセスしてデータ準備に係る作業を行うユーザを管理するユーザ管理モジュール323、定義情報を参照し、そして、ユーザの分析目的情報401に基づいて、ユーザに分析ステップ、及び/又は、作業項目の提案を行う分析ステップ提案モジュール324を備える。

20

【0039】

ミドルウェア301は、データ定義315を参照して、ユーザの分析目的情報401に基づいて分析用データを提案する分析用データ提案モジュール325と、ユーザに対する提案情報をデータ準備テーブル314として更新、管理する提案実行管理モジュール326と、ユーザ端末102、103に、ミドルウェア301の機能にアクセスするためのインタフェースを提供するインタフェースモジュール327と、ネットワーク107、108を介して、ユーザ端末102、103、そして、業務システム104・・・と通信を行うデータ通信モジュール328と、分析ステップが全体像把握である場合の分析用データ提案のための、例えば、傾向分析等の処理を実行する全体像把握モジュール331と、分析ステップが特異事象である場合の分析用データ提案のための、例えば、特異点(例えば、外れ値、大きな変化点、大きな偏り)を抽出する処理を実行する特異事象抽出モジュール332と、分析ステップが要因分析/予測である場合の分析用データ提案のための、例えば、回帰分析、重回帰分析等の処理を実行する要因分析・予測モジュール333と、を備える。

30

【0040】

既述の基準情報管理モジュール322は、基準情報(データカタログ402、データ定義315、マトリックス313)を管理する。例えば、この情報は、産業分野毎(鉄道、電力等)に存在してよく、基準情報管理モジュール322は、産業分野毎に基準情報を管理し、業務データ、又は、分析目的等に応じて、所定分野の基準情報を選択するようにしてもよい。

40

【0041】

なお、既述のモジュールとは、処理装置がプログラムを実行することにより実現される機能であって、例えば、部、手段、要素、回路、又は、ユニットと言い換えられてもよい。

【0042】

次に、サーバ101によって作成され、ユーザに提示される分析用データの候補リスト

50

について説明する。図 5 A の 5 0 1 は、分析ステップとして“全体像把握”をユーザが指定すると、サーバ 1 0 1 が作成する、前記データ準備テーブル 3 1 4（候補リスト）の一例である。分析用データ提案モジュール 3 2 5 が、このテーブルを作成、又は、更新する。このテーブルは、順位 5 1 1、識別情報 5 1 2、K P I 5 1 3、着目データ 1 (When) 5 1 4、着目データ 2 (Where) 5 1 5、着目データ 3 (What) 5 1 6、着目データ 4 (Who) 5 1 7、レコード数 5 1 8、そして、出力ファイル 5 1 9 を備える。

【 0 0 4 3 】

順位 5 1 1 には、識別情報 5 1 2 により特定されるデータ組合せ（行）に対して、優先度に基づき割り振られる順位に関する情報が格納される。データ組合せとは、対象データ 3 1 1 のテーブルが“K P I データ”カラム名として含む、当該カラムのレコードと、“when”に該当する着目データ 1（例えば、既述の“時刻”）を、対象データ 3 1 1 のテーブルがカラム名として含む、当該カラムのレコードと、“where”に該当する着目データ 2（例えば、既述の“キロ程”）に係る同様なレコードと、“what”に該当する着目データ 3（例えば、既述の“列番”）に係る同様なレコードと、そして、“who”に該当する着目データ 4（例えば、“運転士 ID”）に係る同様なレコードとの組合せをいう。

【 0 0 4 4 】

なお、着目データ 1 - 4（5 1 4 - 5 1 7）のそれぞれには、分析目的情報 4 0 1 を構成するユーザ入力、“着目データ項目”と同一又は類似のものがあれば、これが記載される。着目データ 1 - 4（5 1 4 - 5 1 7）のうち、該当するものがない、即ち、4 W 辞書に登録が無く、かつ、着目データ項目も規定されていない場合には、空白が記録される。

【 0 0 4 5 】

K P I 5 1 3 には、分析目的情報 4 0 1 を構成する、ユーザ入力の“K P I”と、同一、又は、類似のものが格納される。

【 0 0 4 6 】

レコード数 5 1 8 には、識別情報 5 1 2 により特定されるデータ組合せにおける、有効状態にあるレコード数が格納される。出力ファイル 5 1 9 には、識別情報 5 1 2 により特定されるデータ組合せのレコードを出力するファイルのパスが格納される。

【 0 0 4 7 】

図 5 B の 5 0 2 は、分析ステップとして、“特異事象抽出”をユーザが指定すると、サーバ 1 0 1 が作成する、データ準備テーブル 3 1 4 の一例である。特異事象リストでもあるテーブル 5 0 2 において、テーブル 5 0 1 と同一の符号についての説明は同じである。テーブル 5 0 2 は、特異点(外れ値)数 5 2 0、特異点(変化点)数 5 2 1、特異点(偏り)数 5 2 2 を備える。

【 0 0 4 8 】

特異点(外れ値)数 5 2 0 には、識別情報 5 1 2 により特定されるデータ組合せにおける、K P I から抽出した特異点の種別が“外れ値”であるものの数に関する情報が格納される。

【 0 0 4 9 】

特異点(変化点)数 5 2 1 には、識別情報 5 1 2 により特定されるデータ組合せにおける、K P I から抽出した特異点の種別が“大きな変化点”であるものの数に関する情報が格納される。特異点(偏り)数 5 2 2 には、識別情報 5 1 2 により特定されるデータ組合せにおける K P I から抽出した特異点の種別が“大きな偏り”であるものの数に関する情報が格納される。

【 0 0 5 0 】

図 5 C の 5 0 3 は、分析ステップとして、“要因分析 / 予測”をユーザが指定すると、サーバ 1 0 1 が作成する、データ準備テーブル 3 1 4 の一例である。目的変数・説明変数組合せリストでもあるテーブル 5 0 3 において、テーブル 5 0 1 と同一の符号についての説明は同じである。

【 0 0 5 1 】

目的変数・説明変数組合せリスト 5 0 3 は、目的変数 5 3 0、説明変数 5 3 1、レコード数 5 1 8、判定値 1（5 3 3）、判定値 2（5 3 4）、判定値 3（5 3 5）、判定値 4

10

20

30

40

50

(5 3 6) を備える。目的変数 5 3 0 には、分析目的に含まれる K P I が格納される。説明変数データ 5 3 1 には、K P I 名を含むカラム以外で、ユーザによって選択された、カラムの名称が記録される。

【 0 0 5 2 】

判定値 1 (5 3 3) には、識別情報 5 1 2 により特定される、データの組合せの順位 5 1 1 を算出するために用いる 1 つ目の判定値データが格納され、判定値 2 (5 3 4) には、2 つ目の判定値データが格納され、判定値 3 (5 3 5) には、3 つ目の判定値データに関する情報が格納され、判定値 4 (5 3 6) には、4 つ目の判定値データに関する情報が格納される。これら判定値については図 1 0 において、詳しく説明する。

【 0 0 5 3 】

図 5 D は、サーバ 1 0 1 から、ユーザに提示、又は、提案される、データの組合せを管理する管理テーブル 5 0 4 の一例である。このテーブルは、データ管理モジュール 3 2 1 によって、作成、又は、更新される。管理テーブル 5 0 4 は、サーバ 1 0 1 が、ユーザからの分析目的情報 4 0 1 と、基準情報 2 2 1 と、そして、対象データ 3 1 1 とに基づいてユーザに提案するデータ、に関する情報を格納する。

【 0 0 5 4 】

提案実行管理モジュール 3 2 6 は、ユーザからの分析目的情報 4 0 1 を受け付けて、分析ステップを推奨し、そして、分析に利用されるデータを提案する際に、テーブル 5 0 4 を作成、又は、更新する。このテーブル 5 0 4 は、識別情報 5 4 1、K P I (5 1 3)、データ組合せ 5 4 3、利用回数 5 4 4、利用人数 5 4 5、更新日時 5 4 6 を含む。

【 0 0 5 5 】

識別情報 5 4 1 には、ユーザへのデータ提案を識別するための情報が格納される。K P I (5 1 3) には、識別情報 5 4 1 によって特定されるデータ提案における K P I データが格納される。データ組合せ 5 4 3 には、識別情報 5 4 1 により特定されるユーザに提案されるデータの組合せが、分析ステップが全体像把握、又は、特異事象抽出である場合に格納され、分析ステップが要因分析・予測である場合、説明変数の組合せが格納される。

【 0 0 5 6 】

利用回数 5 4 4 は、識別情報 5 4 1 によって特定されるデータ提案におけるデータ組合せの利用回数に関する情報が格納され、ユーザによる利用の都度更新される。

【 0 0 5 7 】

利用人数 5 4 5 には、識別情報 5 4 1 によって特定されるデータ提案におけるデータ組合せの利用人数に関する情報が格納される。この利用人数は、延べ人数でもよいし、異なるユーザの人数でもよい。この情報は、識別情報 5 4 1 により特定されるデータ提案が、ユーザに利用される都度更新される。更新日時 5 4 6 には、候補リスト 5 0 1 のレコードが更新された日時が格納される。

【 0 0 5 8 】

図 6 は、サーバ 1 0 1 がユーザに分析ステップを推奨するための動作の一例に係るフローチャートである。サーバ 1 0 1 は、ユーザの入力情報 2 3 1 と、基準情報 2 2 1、対象データ 3 1 1 とに基づいて、ユーザの分析目的が適切か否かを判定し、判定結果を、ユーザに推奨する分析ステップとして出力する。

【 0 0 5 9 】

サーバ 1 0 1 は、ユーザから分析目的情報 4 0 1 の登録 2 3 1 があると、フローチャートを開始する。分析ステップ提案モジュール 3 2 4 は、ステップ 6 0 1 において、ユーザからの分析目的情報 4 0 1 に含まれる、K P I と着目データ項目夫々の名称を、対象データ 3 1 1 のテーブルのカラム名と比較して、比較の結果を、一致度として算出する。同一の名称が“一致”となることは勿論であるが、類似の名称を“一致”に含めてもよい。“類似”とは、例えば、類義語をいう。“一致度”とは、一致の累計、一致の割合、そして、閾値との比較等でよい。モジュール 3 2 4 は、対象データが複数ある場合には、対象データごとに一致度を算出し、例えば、対象データごとの一致度を累計して、複数の対象データの一致度としてよい。

10

20

30

40

50

【 0 0 6 0 】

次いで、分析ステップ提案モジュール 3 2 4 は、ステップ 6 0 2 において、K P I、そして、着目データ項目夫々の名称と、データカタログ 4 0 2 のデータ項目名とを比較して既述の一致度を計算する。

【 0 0 6 1 】

次いで、分析ステップ提案モジュール 3 2 4 は、ステップ 6 0 3 において、K P I、そして、着目データ項目夫々の名称と、データ定義 3 1 5 (4 W 辞書、データ推移パターン辞書) にあるデータ項目名とを比較し、一致度を求める。

【 0 0 6 2 】

ステップ 6 0 1 において、一致度が高いということは、ユーザの分析目的に対する、対象データのデータ量が不足していないこと、及び / 又は、対象データの質が高いことを示すものであり、そして、ステップ 6 0 2、6 0 3 において、一致度が高いということは、対象データを分別、分類、判別、又は、評価等するための情報が、ユーザの分析目的に適合していることを示す。一致度を、例えば、適合度、親和度、該当度等と言い換えてもよい。

10

【 0 0 6 3 】

分析ステップ提案モジュール 3 2 4 は、ステップ 6 0 4 において、ユーザからの分析目的情報 4 0 1 にある分析ステップが、“ 要因分析 / 予測 ” であるか否かを判定する。分析ステップ提案モジュール 3 2 4 が、ステップ 6 0 4 を否定判定すると、ステップ 6 0 5 において、基準情報 (データカタログ 4 0 2、データ定義 3 1 5) の充実度を算出する。充実度とは、基準情報の分析目的に対する有効性、有用性、又は、信頼性を表す指標、例えば、情報の豊富さの程度を示す指標であり、これは、例えば、データカタログ 4 0 2、データ定義 3 1 5 への情報登録件数、そして、参照回数、等のアクセス頻度に基づいて決定されてよい。充実度を有効度等と言い換えてもよい。

20

【 0 0 6 4 】

分析ステップ提案モジュール 3 2 4 がステップ 6 0 4 を肯定判定すると、ステップ 6 0 6 に移動し、分析ステップとしての“ 要因分析 / 予測 ” に於ける、説明変数候補となる加工データ 3 1 2 の充実度を算出する (ステップ 6 0 6)。この充実度は、加工データのカラム名の数、有効なレコードの数の多さから算出されてよい。

【 0 0 6 5 】

分析ステップ提案モジュール 3 2 4 は、ステップ 6 0 7 において、ステップ 6 0 5 ~ 6 0 6 の結果に基づいて、図 7 A 等 に示す、ユーザが登録した分析ステップ 2 3 1 の適否を判定するためのマトリックス 3 1 3 を参照して、ユーザに推奨すべき、分析ステップ、及び / 又は、作業項目を、判定、判別等を行うことによって決定する。

30

【 0 0 6 6 】

分析ステップ提案モジュール 3 2 4 は、ステップ 6 0 7 の決定内容をステップ 6 0 8 において、ユーザに提示して、ユーザの確認、選択等を求める。以上によって、サーバ 1 0 1 は、フローチャートを終了する。なお、分析ステップ、そして、作業項目を纏めて、例えば、分析態様、又は、分析手法等と呼んでよい。

【 0 0 6 7 】

分析ステップ提案モジュール 3 2 4 は、ユーザに提案すべき分析ステップを判定するための基準であるマトリックスとして、ユーザが分析目的情報 (分析ステップ) として、“ 全体像把握 ” を選択した場合には、マトリックス 7 0 1 (図 7 A) を採用し、“ 特異事象抽出 ” を選択した場合には、マトリックス 7 0 2 (図 7 B) を採用し、“ 要因分析 / 予測 ” を選択した場合には、マトリックス 7 0 3 (図 7 C) を採用する。

40

【 0 0 6 8 】

判定マトリックス 7 0 1 (図 7 A) は、ステップ 6 0 1 - 6 0 3 の分析目的情報との一致度と、ステップ 6 0 5 のデータカタログ、データ定義の充実度との相関を規定したものであり、分析ステップ提案モジュール 3 2 4 が、一致度、そして、充実度を夫々所定の閾値と比較して、その高低を決定してよい。

50

【 0 0 6 9 】

判定マトリックス 7 0 1 の“ 分析目的情報との一致度 ”は、ステップ 6 0 1 - 6 0 3 夫々の一致度を、例えば、加算したもの、平均したもの等でよい。ステップ 6 0 1 - 6 0 3 夫々の一致度のうち、所定のステップの一致度を優先させるようにしてもよい。判定マトリックス 7 0 1 の“ データカタログ/データ定義の充実度 ”は、夫々の充実度、例えば、加算したもの、平均したもの等でよい。一方の充実度を優先させるようにしてもよい。

【 0 0 7 0 】

一致度が“ 高い ”、かつ、充実度が“ 高い ”場合、そして、一致度が“ 高い ”、かつ、充実度が“ 低い ”場合、ユーザの分析目的に適する、対象データは十分量存在し得るから、分析ステップ提案モジュール 3 2 4 は、ユーザがサーバ 1 0 1 に登録した分析ステップとしての“ 全体像把握 ”を、そのまま推奨してユーザに提示する。 10

【 0 0 7 1 】

一方、一致度が“ 低い ”、かつ、充実度が“ 高い ”場合、対象データ量が不足している可能性があるから、“ 全体像把握 ”が直ちに実施されることは好ましくはないため、分析ステップ提案モジュール 3 2 4 は、“ 全体像把握 ”を維持しながらも、それを実現するのに必要な作業項目として、まずは、“ データ追加 ”を推奨する。

【 0 0 7 2 】

また、一致度が“ 低い ”、かつ、充実度が“ 低い ”場合、ユーザの分析目的に適するデータは存在していたとしても、そもそも、データカタログ 4 0 2、そして、データ定義 3 1 5 の質が十分でない可能性があるため、分析ステップ提案モジュール 3 2 4 は、分析ステップとしての“ 全体像把握 ”の実施は好ましくないと判定して、作業項目として、統計によるデータ理解促進、データカタログ 4 0 2、データ定義 3 1 5 の拡充を推奨する。 20

【 0 0 7 3 】

判定マトリックス 7 0 2 (図 7 B) の“ 分析目的情報との一致度 ”、“ 高い ”、“ 低い ”の意義は、判定マトリックス 7 0 1 のものと同じである。“ 分析目的情報に一致のデータの充実度 ”は、ステップ 6 0 1 - 6 0 3 にてユーザからの分析目的情報 4 0 1 との一致度が高いと判定される、対象データ 3 1 1 におけるデータ、または、データカタログ 4 0 2、データ定義 3 1 5 のデータ項目に該当する対象データ 3 1 1 (または、加工データ 3 1 2) におけるデータの充実度を示す。

【 0 0 7 4 】

この充実度は、これらデータ (テーブル) の有効なレコードの数の多さから算出されてよい。一致度が、“ 高い ”、かつ、充実度が“ 高い ”場合、分析ステップ提案モジュール 3 2 4 は、ユーザの分析目的に適する、対象データ量は十分であり得るから、ユーザが登録した、分析ステップとしての“ 特異事象抽出 ”を、そのまま推奨する。 30

【 0 0 7 5 】

一方、一致度が“ 高い ”、かつ、充実度が“ 低い ”場合、ユーザの分析目的に適するデータが十分に存在し得ても、特異事象としてのデータレコード数が相対的に不足している可能性があるため、分析ステップ提案モジュール 3 2 4 は、分析ステップとしての“ 特異事象抽出 ”の実施は困難であると判定して、作業項目として、対象データを追加することを推奨する。 40

【 0 0 7 6 】

また、一致度が“ 低い ”、かつ、データの充実度が“ 高い ”場合、又は、一致度が“ 低い ”、かつ、充実度が“ 低い ”場合、分析ステップ提案モジュール 3 2 4 は、ユーザが望む分析目的に適する、対象データが不足している、もしくは、ユーザが、十分に、データを理解できていない可能性があると判断して、分析ステップとしての“ 特異事象抽出 ”を実施することは困難であり、ユーザに、まず、“ 全体像把握 ”に変更して、“ 全体像把握 ”に戻って、これからデータ分析を始めることを推奨する。この場合、分析ステップ提案モジュール 3 2 4 は、判定マトリックス 7 0 1 (図 7 A) に基づいて、“ 全体像把握 ”としての適否を判定する。

【 0 0 7 7 】

判定マトリックス 703 (図 7C) の“分析目的情報との一致度”について、“高い”、“低い”の意義は、判定マトリックス 701 のものと同じである。判定マトリックス 703 の“加工データの充実度”は加工データに対するものであって、“充実度”、そして、その“高い”、“低い”自体の意義は、既述のとおりである。

【0078】

一致度が“高い”、かつ、充実度が“高い”場合、ユーザの分析目的に適する対象データ量は十分であるから、分析ステップ提案モジュール 324 は、ユーザが指定した分析ステップとしての“要因分析/予測”をそのまま推奨する。

【0079】

一致度が“高い”、かつ、充実度が“低い”場合、分析ステップ提案モジュール 324 は、ユーザの分析目的に適する、対象データが十分に存在し得ても、要因分析/予測を実施するには説明変数となる、加工データのレコード数が不足し得るから、分析ステップとしての“要因分析/予測”の実施は困難であると判定して、作業項目として加工データの拡充をユーザに推奨する。 10

【0080】

一致度が“低い”、かつ、充実度が“高い”場合、分析ステップ提案モジュール 324 は、ユーザの分析目的に適する、対象データ自体が不足している可能性があるから、ユーザに、“全体像把握”に戻って検討することを推奨する。

【0081】

一致度が“低い”、かつ、充実度が“低い”場合、分析ステップ提案モジュール 324 は、ユーザの分析目的に適する、対象データ自体が不足している可能性がある、もしくは、データ理解が十分ではない可能性があるから、また要因分析/予測を実施するには説明変数となるデータのレコード数が不足している可能性があるから、ユーザに、“全体像把握”に戻ること、そして、作業項目として加工データを拡充することを推奨する。 20

【0082】

図 8 は、サーバ 101 がユーザに、分析ステップとしての“全体像把握”に利用可能なデータの組合せを、ユーザに提示するための動作の一例を示すフローチャートである。サーバ 101 は、判定マトリックス 701 に基づいて、分析ステップとして、“全体像把握”を決定すると、図 8 のフローチャートをスタートさせる。

【0083】

分析用データ提案モジュール 325 は、ステップ 801 において、ユーザ 201 から受け付けた分析目的情報 401 (231) から KPI と着目データ項目とを抽出する。次いで、分析用データ提案モジュール 325 は、ステップ 802 において、対象データ 311 のテーブルから前記 KPI に該当するカラム、即ち、KPI の名称と同一又は類似名のカラムを抽出する。図 12 において、対象データ 311 は 2 つのデータ (テーブル) 311 A, 311 B を備え、テーブル 311 A のカラム B が KPI と同一名のカラムである。 30

【0084】

分析用データ提案モジュール 325 は、ステップ 803 において、対象データ 311 から着目データ項目に該当するカラムを抽出する。分析用データ提案モジュール 325 は、ステップ 804 において、対象データ 311 から、4W 辞書 221 の 4W (When、Where、What、Who) 夫々のキーワードに該当する (キーワードに同一、又は、類似する)、対象データ 311 のカラムを抽出する。なお、分析用データ提案モジュール 325 は、ステップ 802、及び/又は、ステップ 803 において、抽出されたカラムが、ステップ 804 で再度抽出された場合には、これを除く。 40

【0085】

分析用データ提案モジュール 325 は、ステップ 805 において、対象データ 311 における各カラムのレコードを参照して、データ推移パターン定義 404 にある 4W 辞書 221 の 4W の夫々のキーワードのデータ値の推移パターンに該当するカラムを対象データ 311 から抽出する。なお、分析用データ提案モジュール 325 は、ステップ 802、及び/又は、ステップ 803 において、抽出されたカラムが、ステップ 805 で再度抽出された 50

場合には、これを除く。

【 0 0 8 6 】

分析用データ提案モジュール 3 2 5 は、ステップ 8 0 6 において、ステップ 8 0 3、ステップ 8 0 4、又は、ステップ 8 0 5 で抽出されたカラムの優先度を設定する。分析用データ提案モジュール 3 2 5 は、例えば、前記 3 つのステップにて抽出されたカラムの優先度を 2 つのステップもしくは 1 つのステップのみで抽出されたカラムより優先度を高くする。

【 0 0 8 7 】

分析用データ提案モジュール 3 2 5 は、ステップ 8 0 7 において、ステップ 8 0 4 - 8 0 5 のカラム (4 W 候補カラム) のリストを、ステップ 8 0 6 に係る優先度が分かるように、ユーザに提示し、ステップ 8 0 8 において、ユーザからの要求があれば、4 W 候補カラムの絞り込みを実施して、一部のカラムをデータ分析対象から除外することができる。

10

【 0 0 8 8 】

分析用データ提案モジュール 3 2 5 は、ステップ 8 0 9 において、ステップ 8 0 2 で抽出した、K P I に該当するカラム (K P I カラム) と、ステップ 8 0 3 で抽出した着目データ項目に該当するカラム (着目データ項目カラム)、ステップ 8 0 4 - 8 0 8 に係るカラム (4 W 候補カラム) と、からなる組合せを作成する。以後、この組合せを、“合成カラム組合せ”という。

【 0 0 8 9 】

ステップ 8 0 3 で抽出した着目データ項目に該当するカラムと、ステップ 8 0 4 - 8 0 8 に係る 4 W 候補カラムと、を纏めて、K P I カラムに対して軸候補となる“軸候補カラム”と称することとする。

20

【 0 0 9 0 】

図 1 2 において、テーブル 3 1 1 A のカラム C が軸候補カラム 1 であり、テーブル 3 1 1 A のカラム D が軸候補カラム 2 であり、テーブル 3 1 1 B のカラム G が軸候補カラム 2 であり、テーブル 3 1 1 A のカラム H が軸候補カラム 3 である (カラム D とカラム G の名称は同一または類似)。テーブル 3 1 1 C が合成カラム組合せである。

【 0 0 9 1 】

分析用データ提案モジュール 3 2 5 は、同一又は類似の名称の K P I カラムが複数のデータテーブルで重複する場合、そして、同一又は類似の名称の軸候補カラムが複数のデータテーブルで重複する場合、重複するカラムを交換するようにして合成カラム組合せを複数作成する。図 1 2 の例に即して説明すると、分析用データ提案モジュール 3 2 5 は、合成カラム組合せとしてのテーブル 3 1 1 C を、軸候補カラム 2 がテーブル 3 1 1 A のカラム C のものと、軸候補カラム 2 がテーブル 3 1 1 B のカラム G のものの二つ、作成する。即ち、分析用データ提案モジュール 3 2 5 は、合成カラム組合せ 1 (カラム B , C , D , H) と合成カラム組合せ 2 (カラム B , C , G , H) の 2 つを定義する。同一名の K P I カラムの数を “ k ” とし、同一名の軸候補 1 の数を “ m 1 ”、同一名の軸候補 2 の数を “ m 2 ” …… 同一名の軸候補 n の数を “ m n ” とすると、合成カラム組合せの数は “ k * m 1 * m 2 * …… * m n ” になる。もしくは同一または類似の名称の軸候補カラム同士をキーとしてテーブル 3 1 1 A、3 1 1 B を結合することにより合成カラム組合せを作成する。

30

40

【 0 0 9 2 】

分析用データ提案モジュール 3 2 5 は、ステップ 8 1 0 において、複数の合成カラム組合せから、有効なレコード数が零であるカラム組合せを除外する。有効なレコード数が零であるカラム組合せとは、カラム組合せの全てのレコードにデータが記録されていないものをいう。

【 0 0 9 3 】

次いで、分析用データ提案モジュール 3 2 5 は、ステップ 8 1 1 において、合成カラム組合せの夫々について、夫々の組合せのレコードの特性、属性、例えば、レコードに含まれる複数のデータ値の変動パターンを判定する。分析用データ提案モジュール 3 2 5 は、この判定結果を、分析用データの候補の作成に反映させる。

50

【 0 0 9 4 】

データ値の変動パターンは、例えば、“連続値”、“連続繰返し値”、又は、“離散値”である。“連続値”は、数値データで、全てのレコードでほとんど値が異なるパターンである。“連続繰返し値”は、数値データで、全てのレコードについて、データ値が一定範囲内で増減することを繰り返すパターンである。“離散値”は、文字列データ、または、数値データで、ユニークな値の数がレコード数の一定割合以下であるパターンである。

【 0 0 9 5 】

分析用データ提案モジュール 3 2 5 は、ステップ 8 1 2 において、合成カラム組合せ毎にデータ値変動パターンを判別し、パターンが“連続値”であることを判別すると、ステップ 8 1 3 に移行して、合成カラム組合せの夫々について、全ての合成カラム組合せを結合して、全レコードの数を算出する。

10

【 0 0 9 6 】

分析用データ提案モジュール 3 2 5 は、ステップ 8 1 2 において、データ値変動パターンを判別し、パターンが“連続繰返し値”であることを判別すると、ステップ 8 1 4 に移行して、合成カラム組合せの夫々について、データの繰返しの範囲毎にレコードを分類し、さらに、ステップ 8 1 5 に移行して、繰返し範囲毎に、合成カラム組合せの夫々を結合して、レコード数を算出する。

【 0 0 9 7 】

分析用データ提案モジュール 3 2 5 は、ステップ 8 1 2 において、データ値変動パターンを判別し、パターンが“離散値”であることを判別すると、ステップ 8 1 6 に移行して、ステップ 8 1 1 において選択した軸候補カラムと 4 W の同分類にて“離散値”であることが判別された他の軸候補カラムを 1 つ以上選出し、ステップ 8 1 7 において、K P I カラムとステップ 8 1 6 で選出した複数の軸候補カラムを結合し、文字列データと数値データとの各分類でのレコード数を算出する。

20

【 0 0 9 8 】

分析用データ提案モジュール 3 2 5 は、ステップ 8 1 8 において、ステップ 8 1 1 - 8 1 7 の結果に基づいて、レコード数をデータ値変動パターン毎に集計し、ステップ 8 1 9 において、合成カラム組合せの全てに処理が完了しているか否かを判定し、これを否定すると、ステップ 8 1 1 にリターンし、これを肯定判定すると、ステップ 8 2 0 に移行する。

【 0 0 9 9 】

分析用データ提案モジュール 3 2 5 は、ステップ 8 2 0 において、データ値変動パターン毎に、全組合せの夫々の有用度を、例えば、有効状態のレコード数（データ量）をキーにして評価し、全組合せを評価値の降順にソートして候補リスト（図 5 A の 5 0 1）を作成してユーザに提示する。

30

【 0 1 0 0 】

分析用データ提案モジュール 3 2 5 は、候補リスト 5 0 1 の順位 5 1 1 に、ステップ 8 2 0 で決定された順位を格納し、K P I、そして、着目データ項目を、K P I 5 1 3、着目データ 1 - 4（5 1 4 - 5 1 7）に、有効レコード数をレコード数 5 1 8 に、データ組合せのファイル出力先を出力ファイル 5 1 9 にそれぞれ格納される。

【 0 1 0 1 】

分析用データ提案モジュール 3 2 5 は、ステップ 8 2 1 において、候補リスト 5 0 1 に含まれる、カラム組合せ夫々について、傾向分析を実施し、その結果をグラフとして作成し、ユーザに提示する。ユーザは、グラフを参照して、所定の組合せを指定、選択することができる。ユーザは、候補リストの中から所望の組合せを選択してもよい。この場合、選択された組合せについてのみ、グラフが作成されてもよい。

40

【 0 1 0 2 】

分析用データ提案モジュール 3 2 5 は、ステップ 8 2 2 において、ユーザによって指定された組合せに係るカラムのレコードを C S V ファイル等として出力する。C S V ファイルは、ユーザが改めて対象データの傾向分析を実施する際に使用できる。

【 0 1 0 3 】

50

分析用データ提案モジュール 3 2 5 は、ステップ 8 2 3 において、ステップ 8 2 0 で作成された候補リスト 5 0 1、ステップ 8 2 2 で指定された組合せに基づいて、データカタログ 4 0 2 とデータ定義 2 2 1 とを更新、又は、追加する。したがって、データ分析が進むにしたがって、データカタログとデータ定義とが改良されていく。

【 0 1 0 4 】

図 9 は、サーバ 1 0 1 が、分析ステップとしての“特異事象抽出”に利用可能なデータの組合せを、ユーザに提示するための動作の一例を示すフローチャートである。サーバ 1 0 1 は、判定マトリックス 7 0 2 に基づいて、分析ステップとして、“特異事象抽出”を決定すると、図 9 のフローチャートをスタートさせる。

【 0 1 0 5 】

図 9 のフローチャートにおいて、図 8 のフローチャートのステップ 8 0 1 8 1 9 の工程は同じである。分析用データ提案モジュール 3 2 5 は、ステップ 9 0 3 において、全カラム組合せの夫々について順位付けを行い、順位が閾値以上であるの組合せを抽出する。順位付けは、複数の組合せに夫々おける有効なレコードの多さに基づいたものでよい。

【 0 1 0 6 】

分析用データ提案モジュール 3 2 5 は、ステップ 9 0 4 において、ステップ 9 0 3 において抽出した、複数の組合せの一つの組合せのデータについて特異点の有無を判定し、ステップ 9 0 5 で特異点の判定を肯定すると、ステップ 9 0 6 において、特異点の数を算出する。特異点とは例えば、外れ値、大きな変化点、大きな偏り、等でよい。分析用データ提案モジュール 3 2 5 は、前記特異点に関する情報を特異事象リスト 5 0 2 に追加する。

【 0 1 0 7 】

分析用データ提案モジュール 3 2 5 は、ステップ 9 0 5 を否定すると、ステップ 9 0 6 を経ることなく、ステップ 9 0 7 に移行し、既述の組合せの全てについて、ステップ 9 0 4、9 0 5、9 0 6 を適用したか否かを判定し、否定判定するとステップ 9 0 4 に戻り、肯定判定するとステップ 9 0 8 に移行する。

【 0 1 0 8 】

分析用データ提案モジュール 3 2 5 は、ステップ 9 0 8 において、全ての組合せ夫々の有用度を、例えば、特異点の数に基づいて、昇順にソートする等して順位付けし、特異事象リスト 5 0 2 を作成して、これをユーザに提示する。特異点の数は、例えば、外れ値の数、変化点の数、そして、偏りの数を合計したもの、或いは、平均であってもよい。

【 0 1 0 9 】

分析用データ提案モジュール 3 2 5 は、特異事象リスト 5 0 2 の順位 5 1 1 に、組合せ毎の順位を格納し、K P I を K P I 5 1 3 に格納し、着目データ項目を着目データ 1 - 4 (5 1 4 - 5 1 7) に格納し、組合せ毎の有効レコード数をレコード数 5 1 8 に格納し、組合せ毎の特異点 (外れ値、大きな変化点、大きな偏り) の数を特異点数 5 2 0 - 5 2 2 に格納し、組合せ毎のデータファイルの出力先を出力ファイル 5 1 9 に格納する。

【 0 1 1 0 】

分析用データ提案モジュール 3 2 5 は、ステップ 9 0 9 において、全ての組合せ毎に特異事象算出グラフを作成し、ユーザに提示する。後は、図 8 のフローチャートのステップ 8 2 2 , 8 2 3 と同じである。

【 0 1 1 1 】

図 1 0 は、サーバ 1 0 1 が、分析ステップとしての要因分析 / 予測に利用可能なデータの組合せを、図 8 , 9 のフローチャートと同様にリストとして、ユーザに提示するための動作の一例を示すフローチャートである。サーバ 1 0 1 は、判定マトリックス 7 0 3 に基づいて、分析ステップとして、“要因分析 / 予測”を決定すると、図 1 0 のフローチャートをスタートさせる。

【 0 1 1 2 】

分析用データ提案モジュール 3 2 5 は、図 1 0 のフローチャートを開始すると、ステップ 1 0 0 1 において、ユーザより受け付けた分析目的情報 4 0 1 に基づいて、K P I (K P I 名称、K P I 算出式) を抽出する。

10

20

30

40

50

【 0 1 1 3 】

分析用データ提案モジュール 3 2 5 は、ステップ 1 0 0 2 において、K P I の名称に該当する、対象データ 3 1 1、及び / 又は、加工データ 3 1 2 のカラム名を抽出し、これを目的変数カラムとする。分析用データ提案モジュール 3 2 5 は、K P I 算出式に基づいて、目的変数カラムを作成してもよい。図 1 3 において、対象データ 3 1 1 C のカラム B が、K P I に該当した目的変数カラムである。

【 0 1 1 4 】

分析用データ提案モジュール 3 2 5 は、ステップ 1 0 0 3 において、対象データ 3 1 1 のテーブル、または、加工データ 3 1 2 のテーブルから説明変数の候補となるカラムの所定数を選択する。説明変数の候補となるカラムは、対象データ 3 1 1 のテーブル、または、加工データ 3 1 2 のテーブルのカラムのうち、ステップ 1 0 0 2 で選出された目的変数カラムを除いたカラムである。

10

【 0 1 1 5 】

分析用データ提案モジュール 3 2 5 は、目的変数カラムに対する、説明変数カラムの一つ又は複数の組合せを全て作成する。図 1 3 は、対象データ 3 1 1 (テーブル 3 1 1 A) のカラム B 以外の全てのカラム、加工データ 3 1 2 (テーブル 3 1 1 D) の全てのカラムが説明変数カラムになり得ることを示している。例えば、選択する説明変数カラムの数が 2 つであるとする、分析用データ提案モジュール 3 2 5 は、対象データ 3 1 1 A のカラム B 以外の全てのカラムと、加工データ 3 1 1 D の全てのカラムから 2 つのカラムの全ての組合せを複数作成し、夫々の組合せと目的変数カラムとを合わせて、目的変数カラムと説明変数カラムとの合成カラムからなるテーブルを複数構成する。図 1 3 の 3 1 1 E - 1 ~ 3 1 1 E - 4 の夫々が合成カラムからなるテーブルである。

20

【 0 1 1 6 】

分析用データ提案モジュール 3 2 5 は、ステップ 1 0 0 4 において、対象データのテーブル、そして、加工データのテーブルの全てのカラム間の相関係数を算出する。分析用データ提案モジュール 3 2 5 は、ステップ 1 0 0 5 において、ステップ 1 0 0 2 で抽出した目的変数カラムと、ステップ 1 0 0 3 で選出した数の説明変数の候補カラムとの全ての組合せに対して、回帰分析、または、重回帰分析等の学習を実施する。

【 0 1 1 7 】

分析用データ提案モジュール 3 2 5 は、ステップ 1 0 0 6 において、目的変数カラムと説明変数カラムとの全ての組合せの夫々について有用度を演算し、演算結果に基づいて、全ての組合せ夫々を順位付けしたリストを作成する。ユーザは所望の組合せを選択できる。

30

【 0 1 1 8 】

複数の組合せのうちの夫々の組合せの有用度は、例えば、(1) 目的変数と説明変数の相関係数の絶対値 (ステップ 1 0 0 4)、(2) ステップ 1 0 0 5 での、学習結果の良さ (決定変数、正解率) (3) 目的変数のカラムと説明変数のカラムとに於ける、有効な値を含むレコードの全数、そして、(4) 説明変数間の相関係数の絶対値 (ステップ 1 0 0 4) の少なくとも一つの優劣であってよく、分析用データ提案モジュール 3 2 5 は、有用度の昇順、又は、降順によって、組合せを順位付ける。

【 0 1 1 9 】

分析用データ提案モジュール 3 2 5 は、有用度 (1) - 有用度 (4) 夫々の順位を総合した順位を決定して、これを目的変数・説明変数組合せリスト 5 0 3 の順位 5 1 1 に、K P I を目的変数 (K P I) 5 3 0 に、一つ以上の説明変数の名称を説明変数 5 3 1 に、全ての組合せの有効レコード数をレコード数 5 1 8 に、有用度 (1) - (4) を有用度 1 - 4 (5 3 3 5 3 6) に、組合せ夫々のレコードのファイルの出力先を出力ファイル 5 1 9 に、それぞれ格納される。

40

【 0 1 2 0 】

分析用データ提案モジュール 3 2 5 は、ステップ 1 0 0 7 において、ステップ 1 0 0 6 で作成したリストに含まれる目的変数・説明変数の組合せ毎に学習結果グラフを作成し、ユーザに提示して、ユーザが所望の組合せを選択できるようにする。ここでリストの上位

50

から指定された数の目的変数・説明変数の組合せのみのグラフを作成することも可能である。

【 0 1 2 1 】

そして、分析用データ提案モジュール 3 2 5 は、ステップ 8 2 2 において、ステップ 1 0 0 7 の結果を参照したユーザにより指定された説明変数カラム及び目的変数カラムに関するデータレコードを CSV ファイル等へ出力する。このファイルは、ユーザが自ら要因分析 / 予測を実施する際に使用される。

【 0 1 2 2 】

分析用データ提案モジュール 3 2 5 は、ステップ 1 0 0 6 でのリスト 5 0 3 の作成結果、このリストからのユーザによる選択の結果を基に、データカタログ、データ定義 3 1 5 の変更、情報の追記を行う。

10

【 0 1 2 3 】

図 1 1 A - 図 1 1 C は、サーバ 1 0 1 から、ユーザ端末 1 0 2、1 0 3 に対して提供された、データ利活用を支援するためのグラフィックインターフェースの例である。提案実行管理モジュール 3 2 6 は、分析用データ提案モジュール 3 2 5 と、分析手法提案モジュール 3 2 4 に基づいて、分析目的をリクエストしたユーザ端末に、グラフィックインターフェース用画面を表示させる。

【 0 1 2 4 】

図 1 1 A の画面 1 1 0 1 において、ユーザが指定した分析ステップとしての“全体像把握”と、これに対する、サーバ 1 0 1 が“データ追加”を推奨することが第 1 の領域 1 1 1 1 に表示されている。

20

【 0 1 2 5 】

さらに、分析目的情報 4 0 1 と対象データとの一致度と、データカタログ、データ定義情報の充実度とからなる第 2 の領域 1 1 1 2 と、データ追加のための不足データと、データカタログの格納先、そして、データ定義の格納先とからなる第 3 の領域 1 1 1 3 とが補足情報として表示されている。

【 0 1 2 6 】

図 1 1 B の画面 1 1 0 2 は、推奨分析ステップが全体像把握である、前記候補リスト 5 0 1 の一例である。複数のカラムを組合せたデータについて、その複数（ステップ 8 2 0）が有用度順に羅列されている（1 1 2 1）。領域 1 1 2 2 は、データ組合せ毎で、傾向分析の結果得られたグラフの例である（ステップ 8 2 1）。

30

【 0 1 2 7 】

図 1 1 C の画面 1 1 0 3、推奨分析ステップが要因分析 / 予測である、前記目的変数・説明変数組合せリスト 5 0 3 の一例である。複数のカラムを組合せたデータについて、その複数（ステップ 1 0 0 6）が有用度順に羅列されている（1 1 3 1）。領域 1 1 3 2 は、データ組合せ毎で、学習の結果を示すグラフである。

【 0 1 2 8 】

既述の実施形態によれば、サーバ 1 0 1 は、ユーザ 2 0 1 が複数の業務システムの多種多様、大量のデータを活用して課題分析等を行う際、ユーザが分析目的をサーバに登録すれば、ユーザの負荷を軽減しながら、分析に有用なデータのリストをユーザに提示できる。ユーザは、データ利活用を行う前のデータの準備作業を迅速かつ容易に実現できるようになる。

40

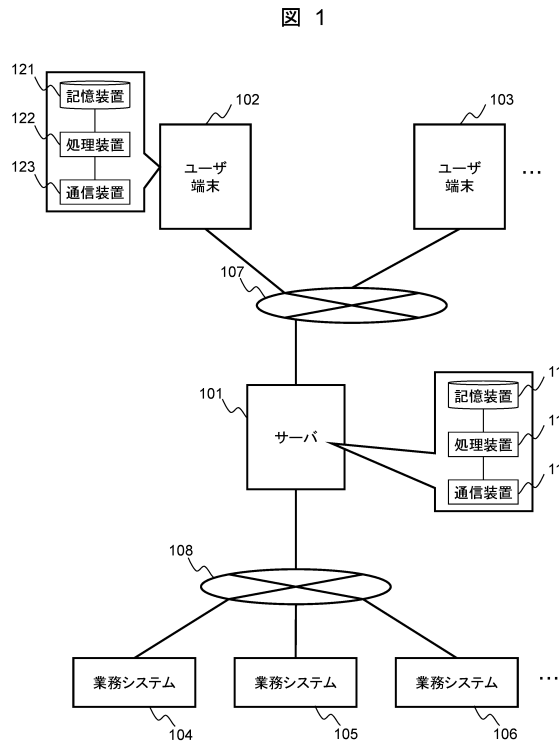
【 0 1 2 9 】

サーバ 1 0 1 は、ユーザに推奨した分析ステップを実行しようとしている当該ユーザに対して、分析目的に関連するデータを提示しようとする際、分析目的（K P I、着目データ項目）に関連するカラムと、分析目的と適合性があり、かつ、充実度も高い基準情報に関連するカラムと、を組合せ、両カラムのレコードをユーザに提示するため、ユーザは、分析目的をサーバに登録しさえすれば、分析目的に関連するデータの組合せを広範に取得することができる。

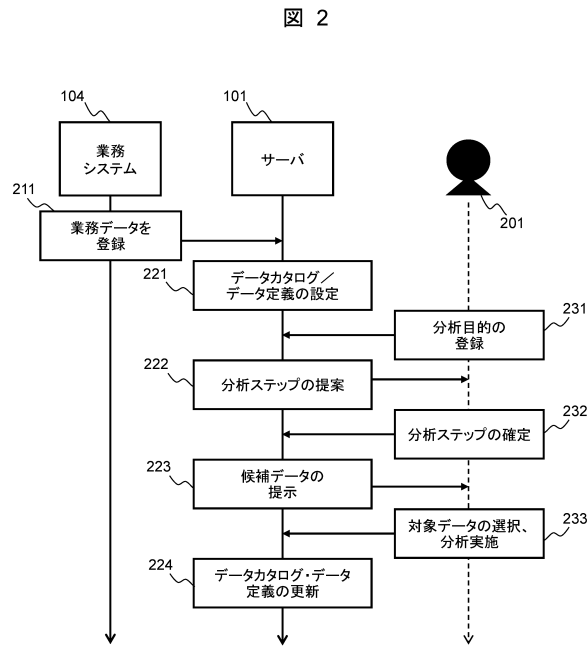
【 符号の説明 】

50

【 0 1 3 0 】
1 0 1 サーバ
1 0 2、1 0 3 ユーザ端末
1 0 4 - 1 0 6 業務システム
【 図 面 】
【 図 1 】



【 図 2 】



10

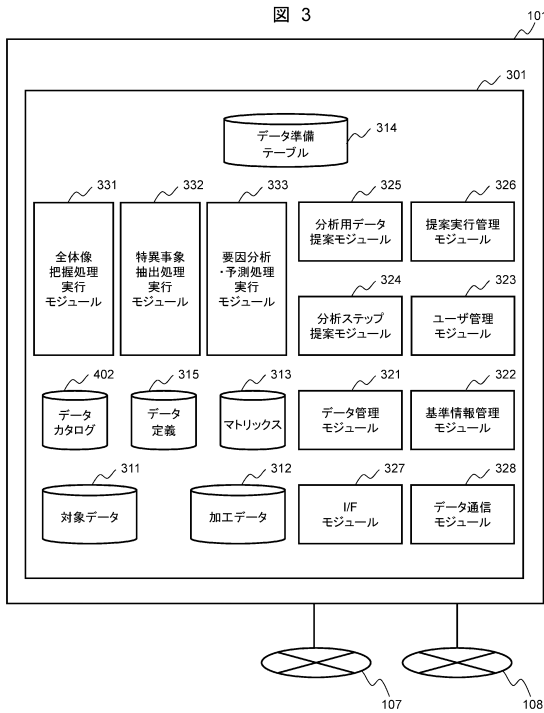
20

30

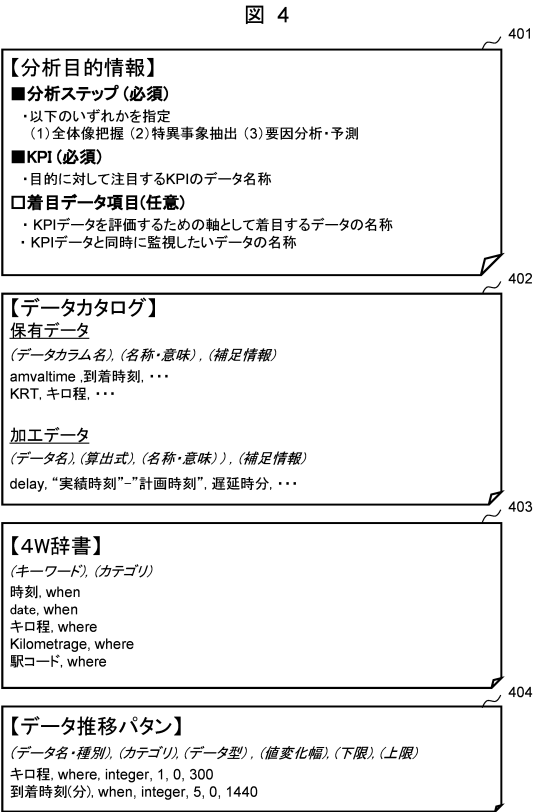
40

50

【図 3】



【図 4】



【図 5 A】

図 5A

501	順位	識別情報	KPI	着目データ1 (When)	着目データ2 (Where)	着目データ3 (What)	着目データ4 (Who)	レコード数	出力ファイル
511
512
513
514
515
516
517
518
519

【図 5 B】

図 5B

502	出力ファイル	特異点 (偏り) 数	特異点 (変化点) 数	特異点 (外れ値) 数	レコード数	着目データ4 (Who)	着目データ3 (What)	着目データ2 (Where)	着目データ1 (When)	KPI	識別情報	順位
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550

10

20

30

40

50

【図 5 C】

図 5C

511	512	530	531	518	533	534	535	536	519
順位	識別 情報	目的変数 (KPI)	説明変数	レコード数	有用度1	有用度2	有用度3	有用度4	出力 ファイル
...
...

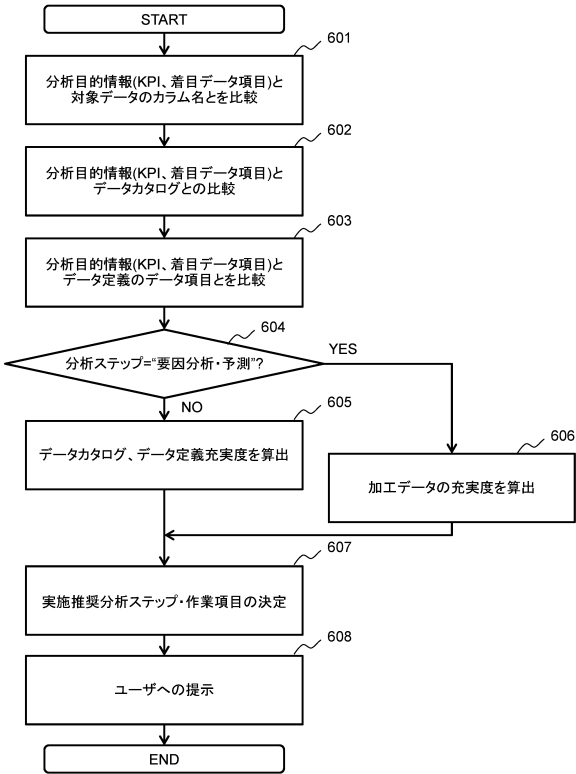
【図 5 D】

図 5D

541	513	543	544	545	546
識別情報	KPI	データ組合せ	利用回数	利用人数	更新日時
...
...

【図 6】

図 6



【図 7 A】

図 7A

		分析目的情報との一致度	
		高い	低い
データカタログ/ データ定義の 充実度	高い	全体像把握	データ追加
	低い	全体像把握	データの理解促進、 データカタログ/データ 定義拡充

10

20

30

40

50

【 図 7 B 】

図 7B

		分析目的情報との一致度	
		高い	低い
分析目的情報に一致のデータの充実度	高い	特異事象抽出	全体像把握
	低い	データ追加要求	全体像把握

【 図 7 C 】

図 7C

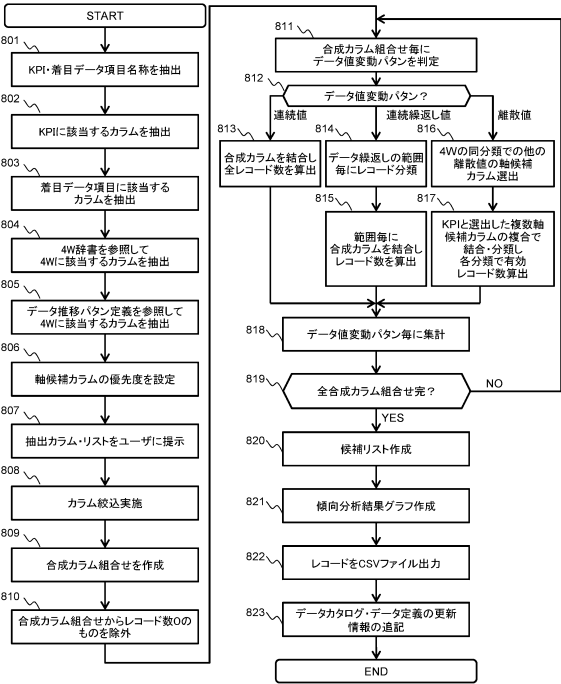
		分析目的情報との一致度	
		高い	低い
加工データの充実度	高い	要因分析・予測	全体像把握
	低い	加工データ拡充	全体像把握、加工データ拡充

10

20

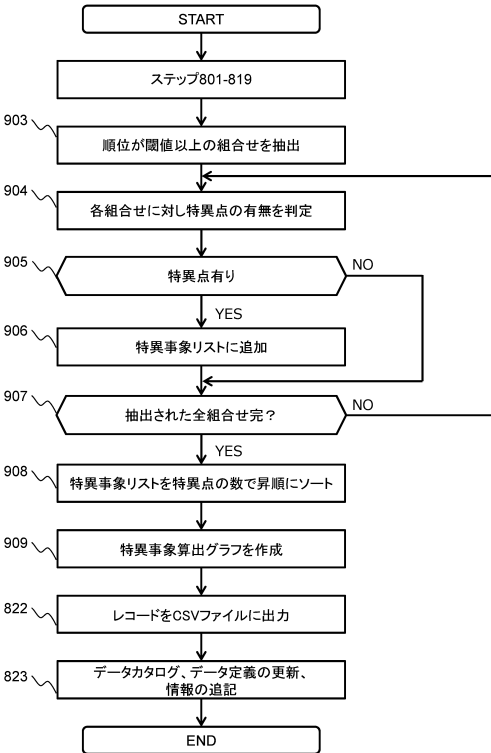
【 図 8 】

図 8



【 図 9 】

図 9

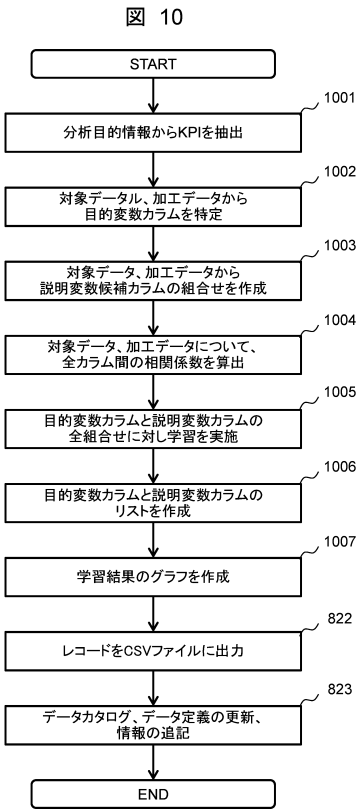


30

40

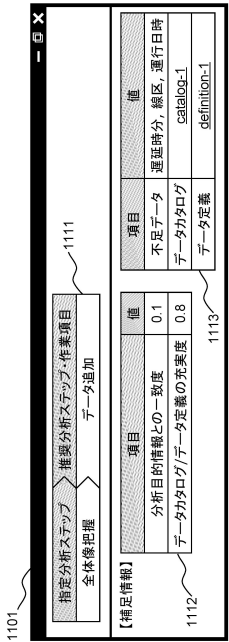
50

【図 1 0】



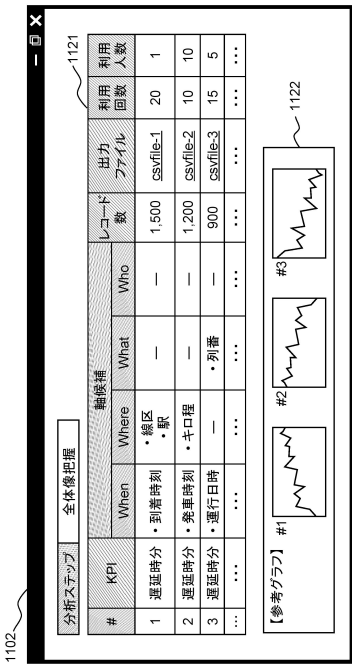
【図 1 1 A】

図 11A



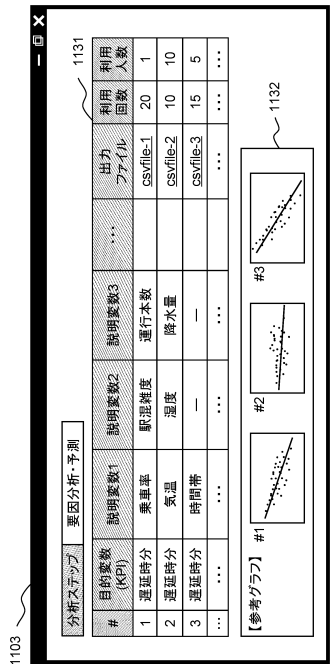
【図 1 1 B】

図 11B



【図 1 1 C】

図 11C



10

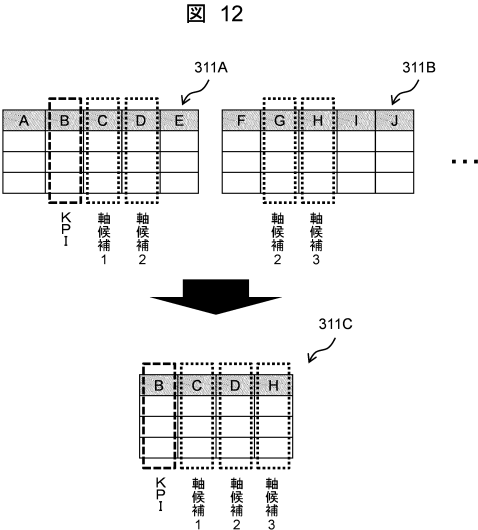
20

30

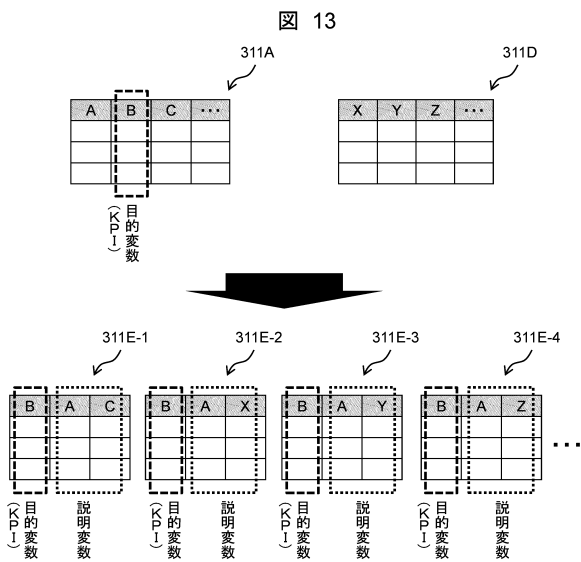
40

50

【図 1 2】



【図 1 3】



10

20

30

40

50

フロントページの続き

- (56)参考文献 特開 2 0 1 0 - 2 0 5 2 1 8 (J P , A)
国際公開第 2 0 1 8 / 1 5 9 0 4 2 (W O , A 1)
米国特許出願公開第 2 0 1 6 / 0 3 2 8 4 0 6 (U S , A 1)
- (58)調査した分野 (Int.Cl. , D B 名)
G 0 6 Q 1 0 / 0 0 - 9 9 / 0 0
G 0 6 F 1 6 / 9 0 7