

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号  
特許第7619470号  
(P7619470)

(45)発行日 令和7年1月22日(2025.1.22)

(24)登録日 令和7年1月14日(2025.1.14)

(51)国際特許分類 F I  
G 0 6 Q 10/10 (2023.01) G 0 6 Q 10/10

請求項の数 9 (全28頁)

(21)出願番号	特願2023-546585(P2023-546585)	(73)特許権者	000004237 日本電気株式会社 東京都港区芝五丁目7番1号
(86)(22)出願日	令和3年9月7日(2021.9.7)	(74)代理人	110000338 弁理士法人 HARAKENZO WOR LD PATENT & TRADEMA RK
(86)国際出願番号	PCT/JP2021/032767	(72)発明者	野澤 拓磨 東京都港区芝五丁目7番1号 日本電気 株式会社内
(87)国際公開番号	WO2023/037399	(72)発明者	董 于洋 東京都港区芝五丁目7番1号 日本電気 株式会社内
(87)国際公開日	令和5年3月16日(2023.3.16)	(72)発明者	榎本 昌文 東京都港区芝五丁目7番1号 日本電気 株式会社内
審査請求日	令和6年2月16日(2024.2.16)		最終頁に続く

(54)【発明の名称】 情報処理装置、情報処理方法及びプログラム

(57)【特許請求の範囲】

【請求項1】

文字列を構成要素として含む評価用データセットと、コンテキストデータとを記憶装置から読み出す又はインターフェースを介して取得する取得手段と、

前記コンテキストデータと前記評価用データセットに含まれる構成要素である文字列との関連度を、当該コンテキストデータと当該構成要素との類似度又は共起度を算出することにより算出する関連度算出手段と、

前記評価用データセットと前記関連度とを参照して、当該評価用データセットに含まれる評価用データの可視化結果を表すデータ、及び当該評価用データを可視化するために用いられるデータ、の少なくともいずれか一方を含むインサイトサブジェクトであって、前記関連度が所定の条件を満たす構成要素を含むインサイトサブジェクトを、複数の可視化パターンのそれぞれについて生成する第1の生成手段と、

複数の前記インサイトサブジェクトに対して、ユーザが求めるインサイトを与える可能性を示す評価値を、評価値を出力する評価モデルを用いて算出することにより、評価を行う評価手段と、

前記インサイトサブジェクト、及び前記インサイトサブジェクトを用いて前記評価用データを可視化した可視化結果、の少なくともいずれか一方を表示する表示手段とを備える情報処理装置。

【請求項2】

前記第1の生成手段は、

前記評価用データセットの構成要素のうち、前記関連度がゼロではない構成要素を含むインサイトサブジェクトを生成する、請求項 1 に記載の情報処理装置。

【請求項 3】

前記評価用データセットには、評価用データと、当該評価用データの集計結果を示す集計データ、当該集計データの統計量、及び、当該評価用データの可視化に利用される情報の少なくともいずれかを含む関連情報とが含まれており、

前記評価手段は、前記複数のインサイトサブジェクトの各々に対して、前記関連情報毎に前記評価値を算出する、請求項 1 又は 2 に記載の情報処理装置。

【請求項 4】

前記表示手段は、

前記評価手段が算出した評価値と共に、又は前記評価手段が算出した評価値が大きいほど優先的に若しくは大きなサイズで、前記複数のインサイトサブジェクトの少なくとも何れかを表示する、請求項 1 から 3 の何れか 1 項に記載の情報処理装置。

【請求項 5】

前記評価手段は、複数のインサイトサブジェクトに対して、前記コンテキストデータを前記評価モデルに入力して前記評価値を算出する、請求項 1 から 4 の何れか 1 項に記載の情報処理装置。

【請求項 6】

参照情報に対し自然言語処理を実行して前記コンテキストデータを生成する第 2 の生成手段を更に備えている、請求項 1 から 5 の何れか 1 項に記載の情報処理装置。

【請求項 7】

前記評価手段の評価結果に対するユーザからのフィードバックを受け付ける受付手段を更に備え、

前記評価手段は、前記ユーザからのフィードバックを参照して、前記複数のインサイトサブジェクトに対して、再度の前記評価値の算出を行う、請求項 1 から 6 の何れか 1 項に記載の情報処理装置。

【請求項 8】

少なくとも 1 つのプロセッサが、文字列を構成要素として含む評価用データセットと、コンテキストデータとを記憶装置から読み出す又はインターフェースを介して取得すること、

前記コンテキストデータと前記評価用データセットに含まれる構成要素である文字列との関連度を、当該コンテキストデータと当該構成要素との類似度又は共起度を算出することにより算出すること、

前記評価用データセットと前記関連度とを参照して、当該評価用データセットに含まれる評価用データの可視化結果を表すデータ、及び当該評価用データを可視化するために用いられるデータ、の少なくともいずれか一方を含むインサイトサブジェクトであって、前記関連度が所定の条件を満たす構成要素を含むインサイトサブジェクトを、複数の可視化パターンのそれぞれについて生成することと、

複数の前記インサイトサブジェクトに対して、ユーザが求めるインサイトを与える可能性を示す評価値を、評価値を出力する評価モデルを用いて算出することにより、評価を行うこと、及び

前記インサイトサブジェクト、及び前記インサイトサブジェクトを用いて前記評価用データを可視化した可視化結果、の少なくともいずれか一方を表示することを含む情報処理方法。

【請求項 9】

コンピュータに、

文字列を構成要素として含む評価用データセットと、コンテキストデータとを記憶装置から読み出す又はインターフェースを介して取得する処理と、

前記コンテキストデータと前記評価用データセットに含まれる構成要素である文字列との関連度を、当該コンテキストデータと当該構成要素との類似度又は共起度を算出するこ

10

20

30

40

50

とにより算出する処理と、

前記評価用データセットと前記関連度とを参照して、当該評価用データセットに含まれる評価用データの可視化結果を表すデータ、及び当該評価用データを可視化するために用いられるデータ、の少なくともいずれか一方を含むインサイトサブジェクトであって、前記関連度が所定の条件を満たす構成要素を含むインサイトサブジェクトを、複数の可視化パターン<sup>10</sup>のそれぞれについて生成する処理と、

複数の前記インサイトサブジェクトに対して、ユーザが求めるインサイトを与える可能性を示す評価値を、評価値を出力する評価モデルを用いて算出することにより、評価を行う処理と、

前記インサイトサブジェクト、及び前記インサイトサブジェクトを用いて前記評価用データを可視化した可視化結果、の少なくともいずれか一方を表示する処理と  
を実行させるプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、情報処理装置、情報処理方法及びプログラムに関する。

【背景技術】

【0002】

データ分析作業においては、「仮説設定、分析・可視化、仮説検証」のサイクルを回すことが一般的であるが、その作業は非常に時間と労力を要するものである。インサイト自動発見技術は、人が有益と考える可視化候補をデータの特徴に基づいて自動的に発見する技術である。これによって、データ分析作業における作業負荷を大幅に低減することが可能である。例えば、下記の特許文献1には、データの分析結果を可視化する方法を表現したキーワードを有するテンプレートデータに基づいて、可視化対象データを可視化したインスタンスデータを生成し、インスタンスデータをインスタンスメタデータの評価値に基づいて再生成する方法が記載されている。

【先行技術文献】

【特許文献】

【0003】

【文献】国際公開第2018/173251号

【発明の概要】

【発明が解決しようとする課題】

【0004】

しかしながら、ユーザの求めるデータの可視化結果は、データの内容及びユーザのニーズ等により様々であり、画一的に定まるものではない。特許文献1に記載の技術では、テンプレートデータがユーザコンテキストを捉えていない場合には、提示された可視化候補が必ずしもユーザの求める可視化結果とは限らない、という問題があった。

【0005】

本発明の一態様は、上記の問題に鑑みてなされたものであり、その目的の一例は、ユーザの求めるインサイトを与える情報を表示する技術を提供することである。

【課題を解決するための手段】

【0006】

本発明の一側面に係る情報処理装置は、評価用データセット及びコンテキストデータを取得する取得手段と、前記コンテキストデータと前記評価用データセットの構成要素との関連度を算出する関連度算出手段と、前記評価用データセットと前記関連度とを参照して生成された複数のインサイトサブジェクトに対して、評価を行う評価手段と、前記インサイトサブジェクトに関連する情報を表示する表示手段とを備える。

【0007】

本発明の一側面に係る情報処理方法は、少なくとも1つのプロセッサが、評価用データセット及びコンテキストデータを取得すること、前記コンテキストデータと前記評価用デ

10

20

30

40

50

ータセットの構成要素との関連度を算出すること、前記評価用データセットと前記関連度とを参照して生成された複数のインサイトサブジェクトに対して、評価を行うこと、及び前記インサイトサブジェクトに関連する情報を表示することを含む。

【0008】

本発明の一側面に係るプログラムは、コンピュータに、評価用データセット及びコンテキストデータを取得する処理と、前記コンテキストデータと前記評価用データセットの構成要素との関連度を算出する処理と、前記評価用データセットと前記関連度とを参照して生成された複数のインサイトサブジェクトに対して、評価を行う処理と、前記インサイトサブジェクトに関連する情報を表示する処理とを実行させる。

【発明の効果】

10

【0009】

本発明の一態様によれば、ユーザの求めるインサイトを与える情報を表示することができる。

【図面の簡単な説明】

【0010】

【図1】本発明の例示的实施形態1に係る情報処理装置の構成を示すブロック図である。

【図2】本発明の例示的实施形態1に係る情報処理方法の流れを示すフロー図である。

【図3】本発明の例示的实施形態2に係る情報処理装置の構成を示すブロック図である。

【図4】本発明の例示的实施形態2に係る情報処理方法の流れを示すフロー図である。

【図5】本発明の例示的实施形態2に係る入力データの例を示す図である。

20

【図6】本発明の例示的实施形態2に係るコンテキストの生成例を示す図である。

【図7】本発明の例示的实施形態2に係るコンテキストと入力データとの関連度の計算方法の例を示す図である。

【図8】本発明の例示的实施形態2に係るコンテキストと入力データとの関連度の計算方法の例を示す図である。

【図9】本発明の例示的实施形態2に係る評価処理の例を示す図である。

【図10】本発明の例示的实施形態2に係る可視化結果の表示例を示す図である。

【図11】本発明の例示的实施形態2に係るインサイトサブジェクトを評価結果と共に表示する例を示す図である。

【図12】本発明の例示的实施形態2に係る可視化情報を評価結果と共に表示する例を示す図である。

30

【図13】本発明の例示的实施形態2に係るインサイトサブジェクトを評価結果と共に表示する例を示す図である。

【図14】本発明の例示的实施形態2に係る特徴ベクトルの生成例を示す図である。

【図15】本発明の例示的实施形態2に係る集計データ及び統計量の例を示す図である。

【図16】本発明の例示的实施形態2に係る評価モデルの例を示す図である。

【図17】本発明の例示的实施形態3に係る情報処理装置の構成を示すブロック図である。

【図18】上記情報処理装置の各機能を実現するソフトウェアであるプログラムの命令を実行するコンピュータの一例を示す図である。

【発明を実施するための形態】

40

【0011】

〔例示的实施形態1〕

本発明の第1の例示的实施形態について、図面を参照して詳細に説明する。本例示的实施形態は、後述する例示的实施形態の基本となる形態である。

【0012】

<情報処理装置の構成>

本例示的实施形態に係る情報処理装置1の構成について、図1を参照して説明する。図1は、情報処理装置1の構成を示すブロック図である。情報処理装置1は、データを可視化して表示する装置である。図示のように、情報処理装置1は、取得部11、関連度算出部12、評価部13及び表示部14を備える。取得部11は、評価用データセット及びコ

50

ンテキストデータを取得する。関連度算出部 1 2 は、コンテキストデータと評価用データセットの構成要素との関連度を算出する。評価部 1 3 は、評価用データセットと関連度を参照して生成された複数のインサイトサブジェクトに対して、評価を行う。表示部 1 4 は、インサイトサブジェクトに関連する情報を表示する。

**【 0 0 1 3 】**

( 評価用データセット )

評価用データセットは、データの可視化候補を情報処理装置 1 が評価するために用いるデータである。評価用データセットは、可視化の対象のデータである評価用データ、及び、当該評価用データに関連する関連データ、の少なくともいずれか一方を含む。ただし、評価用データセットに含まれるデータは上述した例に限られず、評価用データセットは他の情報を含んでいてもよい。

10

**【 0 0 1 4 】**

( 評価用データ )

評価用データは、可視化の対象のデータであり、一例として、複数のレコードを含む多次元データである。評価用データは一例として、ある店舗の月間の売上記録を示すデータ、店舗の規模及びエリアを示すデータ、店舗で販売されている商品の商品コード、商品名及び単価を示すデータ、及び/又は、顧客の性別、年齢、居住地、職業等を示すデータを含む。ただし、評価用データはこれに限られず、他のデータであってもよい。評価用データは、一例として、評価用データの内容を表すチャート(円グラフ、棒グラフ、線グラフ、等)として可視化される。

20

**【 0 0 1 5 】**

( 関連データ )

関連データは、評価用データに関連するデータである。関連データは、一例として、評価用データの集計結果を示す集計データ、集計データの統計量、及び/又は、評価用データの可視化に利用される各種情報の集合である関連情報を含む。関連情報は、一例として、評価用データの可視化に利用されるデータの名前、データ型、集計方法の種別、及びチャートデザインの種別、の一部又は全部を含む。なお、関連データに含まれるデータは上述した例に限られず、関連データは他のデータを含んでいてもよい。

**【 0 0 1 6 】**

( コンテキストデータ )

コンテキストデータは、ユーザがどのようなインサイトを求めるかを表すデータである。コンテキストデータは、一例として、ユーザが求めるインサイトに関するデータであるコンテキスト、及び、コンテキストをベクトル空間で表した特徴ベクトル、の少なくともいずれか一方を含む。なお、コンテキストデータに含まれるデータは上述した例に限られず、コンテキストデータは他のデータを含んでいてもよい。

30

**【 0 0 1 7 】**

( コンテキスト )

コンテキストは、ユーザが求めるインサイトに関するデータであり、一例として、ユーザクエリ又はメタデータから抽出される言語情報である。具体的には例えば、コンテキストは、「商品 A の顧客について」というユーザクエリから抽出される「商品 A 」及び「顧客」という単語である。また、他の例として、コンテキストは例えば、「売上推移について」というユーザクエリから抽出される「売上」及び「推移」という単語である。また、コンテキストは例えば、「検索履歴」が「商品 A の顧客」であるメタデータから抽出される「商品 A 」及び「顧客」という単語である。また、コンテキストは例えば、「検索履歴」が「売上推移」であるメタデータから抽出される「売上」及び「推移」という単語である。ただし、コンテキストは言語情報に限られず、他の情報であってもよい。コンテキストは例えば、ユーザの位置を示す位置情報、単語間の関連度を表す情報、又はサイトの閲覧履歴を示す情報であってもよい。

40

**【 0 0 1 8 】**

( 関連度 )

50

コンテキストデータと評価用データセットの構成要素との関連度は、コンテキストデータと評価用データセットの構成要素との関連の度合いを示す情報である。関連度は一例として、コンテキストである文字列と評価用データセットに含まれる構成要素である文字列との類似度であってもよい。文字列の類似度としては例えば、ハミング距離、レーベンシュタイン距離、ジャロ・ウィンクラー距離が用いられてもよい。また、関連度は一例として、文字列の意味的な類似度を表す情報であってもよい。文字列の意味的な類似度としては例えば、文字列の意味表現をベクトル空間で表した際のユークリッド距離、内積、コサイン類似度などを用いてもよい。また、関連度は一例として、文字列の共起度を表す情報であってもよい。文字列の共起度としては例えば、文字列の共起関係をベクトル空間で表した際のユークリッド距離、内積、コサイン類似度などを用いてもよい。また、関連度は一例として、コンテキストに対応するデータパターンと、評価用データセットの構成要素のデータパターンとの類似度を表す情報であってもよい。

10

【0019】

(インサイトサブジェクト)

インサイトサブジェクトは、評価用データセットと関連度とを参照して生成されたデータである。インサイトサブジェクトは、一例として、評価用データの可視化結果を表すデータ、及び、評価用データを可視化するために用いられるデータ、の少なくともいずれか一方を含む。評価用データを可視化した可視化結果は、例えば評価用データの内容を表すチャート(円グラフ、棒グラフ、線グラフ、等)である。また、インサイトサブジェクトは、一例として、上述した関連データの一部、例えば関連データに含まれる関連情報であってもよい。換言すると、インサイトサブジェクトは、評価用データセットの一部であってもよい。ただし、インサイトサブジェクトは上述した例に限られず、他のデータであってもよい。

20

【0020】

(インサイト)

また、本明細書において、インサイトとは、人が有益と認識する可視化結果、及び、そのような可視化結果を表すデータ、のことをいう。換言すれば、インサイトとは、人が有益と認識するインサイトサブジェクトのことをいう。

【0021】

取得部11が評価用データセット及びコンテキストデータを取得する方法は特に限定されない。例えば、取得部11は、外部の記憶装置又は内部の記憶装置から評価用データセット及びコンテキストデータを読み出すことにより取得してもよく、また、通信IF又は入出力IFを介して評価用データセット及びコンテキストデータを取得してもよい。

30

【0022】

関連度算出部12が関連度を算出する手法は特に限定されない。関連度算出部12は一例として、コンテキストである文字列と評価用データセットに含まれる構成要素である文字列との類似の度合いを表す関連度を算出する。関連度算出部12は一例として、コンテキストである文字列と評価用データセットに含まれる構成要素である文字列との意味的な類似度合いを表す関連度を算出してもよい。関連度算出部12は一例として、コンテキストである文字列と評価用データセットに含まれる構成要素である文字列との共起度合いを表す関連度を算出してもよい。また、関連度算出部12は一例として、コンテキストデータに対応するデータパターンと、評価用データセットの構成要素のデータパターンとの類似の度合いを表す関連度を算出してもよい。

40

【0023】

また、評価部13が複数のインサイトサブジェクトを評価する手法は特に限定されない。評価部13は、一例として、複数のインサイトサブジェクトのそれぞれについて、ユーザが求めるインサイトを与えるかの評価結果である評価値を算出する。以下では、この評価値をインサイトスコアともいう。インサイトスコアは、そのまま出力してもユーザが求めるインサイトを与えるインサイトサブジェクトを発見する大きな助けとなる。また、インサイトスコアを用いることにより、インサイトスコアが高い、すなわちユーザが求める

50

インサイトを与える可能性が高いインサイトサブジェクトを自動で検出することも可能になる。

【0024】

評価部13は一例として、関連データとコンテキストデータとが入力され、評価値を出力する評価モデルを用いて、複数のインサイトサブジェクトに対して評価を行う。評価モデルは、予め規定されたスコア関数であってもよく、また、機械学習により構築された学習済モデルであってもよい。スコア関数を用いる場合、評価部13は一例として、関連データとコンテキストデータとの関連性が高いほど、高い評価値を出力するスコア関数を用いて、複数のインサイトサブジェクトに対して評価を行う。ただし、評価部13が行う評価の手法はこれらに限られず、他の手法が用いられてもよい。

10

【0025】

評価用データを可視化した可視化結果は、可視化に利用される関連情報等の内容によってそれぞれ異なる。評価用データを複数の異なるパターンで可視化した複数の可視化結果のそれぞれを、以下では「可視化候補」ともいう。評価用データの複数の可視化候補がユーザに与える視覚的な特徴は、複数の可視化候補のそれぞれで異なる。

【0026】

インサイトサブジェクトは、評価用データの可視化候補と1対1で対応する。そのため、評価部13が複数のインサイトサブジェクトに対してコンテキストデータに応じた評価を行うことにより、複数の可視化候補に対してコンテキストデータに応じた評価が行われる。

20

【0027】

表示部14が表示するインサイトサブジェクトに関する情報の表示態様は特に限定されない。表示部14は一例として、評価部13による評価が相対的に高いインサイトサブジェクトを、評価が相対的に低いインサイトサブジェクトよりも優先的に表示してもよい。また、表示部14は一例として、関連データに含まれる関連情報と、評価部13による評価結果とを対応付けて表示してもよい。

【0028】

以上のように、本例示的实施形態に係る情報処理装置1においては、評価用データセット及びコンテキストデータを取得する取得部11と、コンテキストデータと評価用データセットの構成要素との関連度を算出する関連度算出部12と、評価用データセットと関連度とを参照して生成された複数のインサイトサブジェクトに対して、評価を行う評価部13と、インサイトサブジェクトに関連する情報を表示する表示部14と、を備える構成が採用されている。このため、本例示的实施形態に係る情報処理装置1によれば、ユーザの求めるインサイトを与える情報を表示することが可能になる、という効果が得られる。

30

【0029】

上述の情報処理装置1の機能は、プログラムによって実現することもできる。本例示的实施形態に係るプログラムは、コンピュータに、評価用データセット及びコンテキストデータを取得する処理と、コンテキストデータと評価用データセットの構成要素との関連度を算出する処理と、評価用データセットと関連度とを参照して生成された複数のインサイトサブジェクトに対して、評価を行う処理と、インサイトサブジェクトに関連する情報を表示する処理とを実行させる。したがって、本例示的实施形態に係るプログラムによれば、ユーザの求めるインサイトを与える情報を表示することが可能になる、という効果が得られる。

40

【0030】

<情報処理方法の流れ>

本例示的实施形態に係る情報処理方法S1の流れについて、図2を参照して説明する。図2は、情報処理方法S1の流れを示すフロー図である。

【0031】

ステップS11では、少なくとも1つのプロセッサが、評価用データセット及びコンテキストデータを取得する。そして、ステップS12では、少なくとも1つのプロセッサが

50

、コンテキストデータと評価用データセットの構成要素との関連度を算出する。ステップ S 1 3 では、少なくとも 1 つのプロセッサが、評価用データセットと関連度とを参照して生成された複数のインサイトサブジェクトに対して、評価を行う。ステップ S 1 4 では、少なくとも 1 つのプロセッサが、インサイトサブジェクトに関連する情報を表示する。これにより、図 2 の情報処理方法 S 1 は終了する。

【 0 0 3 2 】

なお、1 つのプロセッサに S 1 1 ~ S 1 4 の処理を実行させてもよいし、S 1 1 ~ S 1 4 の処理を複数のプロセッサに分担させて実行させてもよい。後者の場合、各プロセッサは、1 つの情報処理装置が備えているものであってもよいし、それぞれ異なる情報処理装置が備えているものであってもよい。また、S 1 1 ~ S 1 4 の処理を実行する少なくとも 1 つのプロセッサは、情報処理装置 1 が備えているものであってもよい。

10

【 0 0 3 3 】

以上のように、本例示的实施形態に係る情報処理方法 S 1 においては、少なくとも 1 つのプロセッサが、評価用データセット及びコンテキストデータを取得すること、コンテキストデータと評価用データセットの構成要素との関連度を算出すること、評価用データセットと関連度とを参照して生成された複数のインサイトサブジェクトに対して、評価を行うこと、及びインサイトサブジェクトに関連する情報を表示することを含む、という構成が採用されている。このため、本例示的实施形態に係る情報処理方法 S 1 によれば、ユーザの求めるインサイトを与える情報を表示することが可能になる、という効果が得られる。

【 0 0 3 4 】

〔例示的实施形態 2〕

本発明の第 2 の例示的实施形態について、図面を参照して詳細に説明する。なお、例示的实施形態 1 にて説明した構成要素と同じ機能を有する構成要素については、同じ符号を付し、その説明を適宜省略する。

20

【 0 0 3 5 】

< 情報処理装置の構成 >

図 3 は、情報処理装置 1 A の構成を示すブロック図である。情報処理装置 1 A は、情報処理装置 1 A の各部を統括して制御する制御部 1 0 A と、情報処理装置 1 A が使用する各種データを記憶する記憶部 1 7 とを備える。また、情報処理装置 1 A は、情報処理装置 1 A が他の装置と通信するための通信部 1 8、情報処理装置 1 A がデータを表示出力するための表示部 1 9、及び情報処理装置 1 A に対する入力を受け付ける入力部 2 0 を備える。以下では、表示部 1 9 がデータを表示出力する例を説明するが、情報処理装置 1 A は、例えば印字出力や音声出力等の態様でデータを出力してもよい。また、表示部 1 9 と入力部 2 0 とは、情報処理装置 1 A に外付けされた、情報処理装置 1 A の外部の機器であってもよい。

30

【 0 0 3 6 】

制御部 1 0 A は、取得部 1 1、関連度算出部 1 2、評価部 1 3、表示部 1 4、第 1 の生成部 1 5 - 1、及び第 2 の生成部 1 5 - 2 を含む。また、記憶部 1 7 には、評価用データセット D S、コンテキストデータ C D、評価モデルパラメータ E M P、評価結果 E R、及び表示用データ D D が記憶される。

40

【 0 0 3 7 】

( 評価用データセット D S )

評価用データセット D S には、評価用データと、当該評価用データに関連する関連データ V D とが含まれている。評価用データは、可視化の対象であるデータであり、一例として、ある店舗の月間の売上記録を示すデータ、店舗の規模及びエリアを示すデータ、店舗で販売されている商品の商品コード、商品名及び単価を示すデータ、及び / 又は、顧客の性別、年齢、居住地、職業等を示すデータである。

【 0 0 3 8 】

( 関連データ V D )

関連データ V D は、評価用データに関連するデータである。関連データ V D には、

50

- ・評価用データに関連する関連情報 $V$
  - ・関連情報 $V$ をベクトル空間で表した特徴ベクトル $d^V$
  - ・評価用データに含まれるデータであって関連情報 $V$ に対応するデータを集計して得られる集計データ $s^V$ 、及び、
  - ・集計データ $s^V$ の統計量 $t^V$
- の少なくともいずれかが含まれる。

## 【0039】

(関連情報 $V$ )

関連情報 $V$ は、一例として、評価用データの可視化に利用される各種情報の集合であり、例えば以下の情報を含む。

- ・評価用データに含まれる各データの属性情報
- ・集計方法(フィルタ、集計関数、集計のキーとなる列名、等)に関する情報(評価用データに対して適用するフィルタに関する情報、等)
- ・チャートデザイン(x軸、y軸、チャートタイプ、プロットタイプ、等)に関する情報(チャートの各軸と項目との関係に関する情報、等)

## 【0040】

(特徴ベクトル $d^V$ )

関連情報の特徴ベクトル $d^V$ は、関連情報 $V$ をベクトル空間で表現したものである。ベクトル化の方法は任意であるが、例えば単語の分散表現などを用いてもよい。

## 【0041】

(集計データ $s^V$ )

集計データ $s^V$ は、評価用データから関連情報 $V$ に対応する数値を集計したデータである。集計データ $s^V$ は、関連情報 $V$ の可視化結果としてチャートにプロットされる。

## 【0042】

(統計量 $t^V$ )

集計データ $s^V$ の統計量 $t^V$ は、集計データ $s^V$ についての各種統計量を並べたものである。利用する統計量は任意であるが、例えば以下のものが統計量 $t^V$ として利用可能である。

- ・最大値、最小値、中央値
- ・平均値、標準偏差、分散
- ・カーディナリティ
- ・ゼロ値の割合、欠損値の割合
- ・尖度、歪度
- ・エントロピー
- ・ジニ係数

## 【0043】

(コンテキストデータ $CD$ )

コンテキストデータ $CD$ には、

- ・コンテキスト $C$ 、及び、
  - ・コンテキストをベクトル空間で表した特徴ベクトル $d^C$
- の少なくとも何れかが含まれる。

## 【0044】

(コンテキスト $C$ )

コンテキスト $C$ は、ユーザが求めるインサイトに関するデータである。コンテキスト $C$ は、一例として、ユーザが求めるインサイトを自然言語で表すデータであり、ユーザが求めるインサイトの質及び量に関するデータを含む。コンテキスト $C$ は、後述のユーザクエリ $Q$ 及び/又はメタデータ $M$ から抽出されてもよい。コンテキスト $C$ は一例として、「商品 $A$ 」及び「顧客」という単語を含む。

## 【0045】

(特徴ベクトル $d^C$ )

10

20

30

40

50

コンテキストCの特徴ベクトル $d^C$ は、コンテキストCをベクトル空間で表したものである。ベクトル化の方法は任意であるが、一例として、単語の分散表現などを用いてもよい。

【0046】

(ユーザクエリQ)

ユーザクエリQは、ユーザが求めるインサイトに関するクエリであり、ユーザによって自然言語で与えられる。ユーザクエリQは、例えば以下の情報を含む。

- ・分析対象とするデータについての情報(例:「商品A」、「売上」)
- ・インサイトに対する仮説(例:「～が増加している」、「～が突出している」)
- ・想定するチャートの特徴(例:地域別集計、円グラフ)

10

【0047】

(メタデータM)

メタデータMは、ユーザの求めるインサイトが推定できる情報である。メタデータMは、一例として、所定のシステムにより自動的に収集される。メタデータMは、例えば以下の情報を含む。

- ・ユーザの検索履歴(例:「商品A、顧客」で検索している)
- ・ユーザの分析履歴(例:商品Aの顧客分析を過去に実施)
- ・ユーザの評価履歴(例:商品Aの顧客についてのチャート等を高く評価していた)
- ・ユーザの行動履歴(例:商品Aのサイト又は販売店舗をxx分滞在していた)

20

【0048】

(評価モデルパラメータEMP)

評価モデルパラメータEMPは、評価モデル $f$ を規定するパラメータである。評価モデル $f$ は、関連データVD及びコンテキストデータCDを入力とし、入力された関連データVDに対応するインサイトサブジェクトを定量的に評価するモデルである。評価モデル $f$ としては、インサイトサブジェクトの評価結果の推定に利用できるものであれば任意のモデルが利用可能である。例えば、後述するようなルールベースのモデル、又は、機械学習により構築されたモデル等が評価モデル $f$ として利用可能である。評価モデル $f$ の出力は、一例として、評価結果を表すスコア、又は、ラベル確率である。評価モデル $f$ については後述する。

【0049】

(評価結果ER)

評価結果ERは、評価部13によるインサイトサブジェクトの評価結果を示すデータである。評価結果ERは、一例として、複数のインサイトサブジェクトのそれぞれについての評価結果を表すインサイトスコア $y^{\wedge}$ である。

30

【0050】

(インサイトスコア $y^{\wedge}$ )

インサイトスコア $y^{\wedge}$ は、評価モデル $f$ の出力値に基づいて計算される可視化のよさの定量的指標である。インサイトスコア $y^{\wedge}$ は例えば、評価モデル $f$ の出力値であってもよく、また、評価モデル $f$ の出力値に正規化及び/又は重み付け等の処理を加えた値であってもよい。インサイトスコア $y^{\wedge}$ の算出方法の具体例については後述する。

40

【0051】

(表示用データDD)

表示用データDDは、表示部14が表示する、インサイトサブジェクトに関する情報である。表示用データDDは一例として、コンテキストCと関連するインサイトオブジェクトを表す情報である。また、表示用データDDは一例として、インサイトオブジェクトの評価結果ERを含んでもよい。

【0052】

(取得部11)

取得部11は、評価用データセットDS及びコンテキストデータCDを取得する。取得部11は、一例として、評価用データセットDS及びコンテキストデータCDを記憶部1

50

7から読み出すことにより取得する。ただし、評価用データセットD S及びコンテキストデータC Dの取得方法は特に限定されない。例えば、取得部11は、情報処理装置1Aのユーザが入力部20を介して入力した評価用データセットD S及びコンテキストデータC Dを取得してもよい。また、例えば、取得部11は、通信部18を介した通信により、外部の装置から評価用データセットD S及びコンテキストデータC Dを取得してもよい。

【0053】

(関連度算出部12・評価部13)

関連度算出部12は、コンテキストデータと評価用データセットの構成要素との関連度を算出する。評価部13は、評価用データと関連度とを参照して生成された複数のインサイトサブジェクトに対して評価を行う。評価部13は、一例として、複数のインサイトサブジェクトのそれぞれについてインサイトスコア $y^{\wedge}$ を算出し、その算出結果を示す評価結果E Rを生成して記憶部17に記憶させる。

10

【0054】

(表示部14)

表示部14は、第1の生成部15-1が生成する表示用データD Dを用いて、インサイトサブジェクトに関する情報を表示部19に表示する。

【0055】

(第1の生成部15-1)

第1の生成部15-1は、評価用データセットと関連度とを参照して、複数のインサイトサブジェクトを生成する。インサイトサブジェクトの生成処理については後述する。また、第1の生成部15-1は、表示用データD Dを生成する。第1の生成部15-1は一例として、関連度算出部12が算出した関連度に基づき、コンテキストCと関連するインサイトサブジェクトを列挙した表示用データD Dを生成する。

20

【0056】

(第2の生成部15-2)

第2の生成部15-2は、参照情報を参照して、コンテキストデータを生成する。参照情報は例えば、ユーザクエリ又はメタデータである。また、第2の生成部15-2は、評価用データセットD Sを生成する。

【0057】

<情報処理方法の流れ>

本例示の実施形態に係る情報処理方法の流れについて、図面を参照して説明する。図4は、情報処理方法の流れを示すフロー図である。以下では、関連情報Vが評価用データの可視化に利用される可視化情報である場合について説明する。以下では、関連情報Vの一例である可視化情報を「可視化情報V」ともいう。

30

【0058】

(ステップS101)

ステップS101では、取得部11が、入力データD及びコンテキスト生成用データを取得する。入力データDは、本明細書に係る評価用データの一例である。入力データDは、チャートにプロットされるデータを含んでいればよく、入力データDのフォーマットとしては任意のフォーマットが利用可能である。取得部11は一例として、入力部20又は通信部18を介して入力データDを取得する。

40

【0059】

図5は、入力データDの例を示す図である。図5の例では、入力データDは、売上データ、店舗データ、商品データ、及び顧客データを含む。売上データ、店舗データ、商品データ及び顧客データは、何れも複数のレコードを含む多次元データのデータセットである。売上データは、「日付」、「商品コード」、「顧客コード」、「店舗コード」、及び「売上」のデータ項目を含む多次元データである。店舗データは、「店舗コード」、「店舗名」、「エリア」、及び「規模」のデータ項目を含む多次元データである。商品データは、「商品コード」、「商品名」、「分類」、及び「単価」のデータ項目を含む多次元データである。顧客データは、「顧客コード」、「年齢」、「性別」、「居住地」、「職業」

50

、及び「収入」のデータ項目を含む多次元データである。

【0060】

(コンテキスト生成用データ)

コンテキスト生成用データは、コンテキストCを生成するためのデータであり、本明細書に係る参照情報の一例である。コンテキスト生成用データは、一例として、ユーザクエリQ及びメタデータMの一方又は両方を含む。コンテキスト生成用データは、複数のユーザクエリを含んでいてもよく、また、複数のメタデータを含んでいてもよい。ただし、コンテキスト生成用データはユーザクエリ及びメタデータに限られず、他のデータであってもよい。また、コンテキスト生成用データは、そのままコンテキストCとして利用可能なデータであってもよい。取得部11は、一例として、入力部20又は通信部18を介してコンテキスト生成用データを取得してもよく、また、記憶部17からコンテキスト生成用データを読み出すことにより取得してもよい。

10

【0061】

(ステップS102)

ステップS102では、第2の生成部15-2が評価用データセットDS及びコンテキストデータCDを生成する。評価用データセットDSの生成及びコンテキストデータCDの生成の具体例について以下に説明する。

【0062】

(評価用データセットDSの生成)

第2の生成部15-2は、まず、可視化情報Vを取得する。第2の生成部15-2は、記憶部17の所定の記憶領域から可視化情報Vを読み出すことにより取得してもよく、また、入力部20又は通信部18を介して可視化情報Vを取得してもよい。このとき、第2の生成部15-2は、複数の可視化情報Vを取得する。可視化情報Vは、一例として、入力データDに含まれる各データの属性情報、チャートの各軸と項目との関係に関する情報、入力データDに対して適用するフィルタ、チャートタイプ、集計方法、等の情報を含む。

20

【0063】

また、第2の生成部15-2は、任意の言語モデルを用いて、取得した可視化情報Vをベクトル空間で表現する特徴ベクトル $d^V$ を生成する。特徴ベクトル $d^V$ は、複数の可視化情報Vのそれぞれについて生成される。また、第2の生成部15-2は、入力データDから可視化情報Vに対応する数値を集計した集計データ $s^V$ 、及び、集計データ $s^V$ につ

30

【0064】

第2の生成部15-2は、取得した可視化情報V、及び、生成した特徴ベクトル $d^V$ 、集計データ $s^V$ 、統計量 $t^V$ を含む関連データVDと、ステップS101で取得部11が取得した入力データDとを含む評価用データセットDSを生成する。関連データVDには、複数の可視化情報V及び複数の特徴ベクトル $d^V$ が含まれてもよく、また、1対の可視化情報V及び特徴ベクトル $d^V$ が含まれてもよい。

【0065】

(コンテキストデータCDの生成)

また、第2の生成部15-2は、ステップS101で取得部11が取得したコンテキスト生成用データに対し任意の自然言語処理を実行してコンテキストCを生成する。なお、第2の生成部15-2は、コンテキスト生成用データをそのままコンテキストCとして用いてもよい。

40

【0066】

図6は、コンテキストの生成例を示す図である。図6の例で、第2の生成部15-2は、「商品Aの顧客について」というユーザクエリQ1に対し自然言語処理を実行し、「商品A」、「顧客」というコンテキストC11を生成する。また、第2の生成部15-2は、「売上推移について」というユーザクエリQ2に対し自然言語処理を行い、「売上」、「推移」というコンテキストC12を生成する。また、第2の生成部15-2は、「検索履歴」が「商品Aの顧客」であるメタデータM1に対し自然言語処理を行って「商品A」

50

、「顧客」というコンテキストC 1 1を生成する。また、第2の生成部1 5 - 2は、「検索履歴」が「売上推移」であるメタデータM 2に対し自然言語処理を行った「売上」、「推移」というコンテキストC 1 2を生成する。

【0067】

第2の生成部1 5 - 2は、任意の言語モデルを用いて、生成したコンテキストCをベクトル空間で表現する特徴ベクトル $d^C$ を生成し、生成した特徴ベクトル $d^C$ とコンテキストCとを含むコンテキストデータCDを生成する。

【0068】

(ステップS 1 0 3)

図4のステップS 1 0 3では、関連度算出部1 2は、コンテキストデータCDと評価用データセットDSの構成要素との関連度を算出する。図7及び図8は、コンテキストデータCDと評価用データセットDSとの関連度の算出方法の例を示す図である。図7は、コンテキストデータCDが「商品A」及び「顧客」というコンテキストC 1 1を含み、評価用データセットDSに含まれる入力データDが、図5に示したような、売上データ、商品データ、店舗データ及び顧客データを含む例を示す。この場合、関連度算出部1 2は、コンテキストC 1 1の「商品A」及び「顧客」のそれぞれと、入力データDの複数の構成要素のそれぞれとの関連度を算出する。図7の例で、入力データDの構成要素は例えば、売上データ、店舗データ、商品データ、顧客データの各々のデータ、及び、各データのデータ項目を含む。

【0069】

図7の例では、売上データの「商品コード」の列、及び、商品データの「商品名」の項目が「商品A」の行は、コンテキストC 1 1の「商品A」との文字列の類似度が高いため、関連度算出部1 2が算出する関連度は他の構成要素よりも大きくなる。また、売上データの「顧客コード」の列、及び顧客データは、コンテキストC 1 1の「顧客」との文字列の類似度が高いため、関連度算出部1 2が算出する関連度は他の構成要素よりも大きくなる。

【0070】

図8は、コンテキストデータCDが「売上」及び「推移」というコンテキストC 1 2を含み、評価用データセットDSに含まれる入力データDが、図5に示したような、売上データ、商品データ、店舗データ及び顧客データを含む例を示す。この場合、関連度算出部1 2は、コンテキストC 1 2の「売上」及び「推移」のそれぞれと、入力データDの複数の構成要素のそれぞれとの関連度を算出する。

【0071】

図8の例では、売上データの「売上」の列は、コンテキストC 1 2の「売上」との文字列の類似度が高いため、関連度算出部1 2が算出する関連度は他の構成要素よりも大きくなる。また、売上データの「日付」の列は、コンテキストC 1 2の「推移」に対応するデータパターンを有しているため、関連度算出部1 2が算出する関連度が他の構成要素よりも大きくなる。

【0072】

(ステップS 1 0 4)

図4のステップS 1 0 4では、第1の生成部1 5 - 1は、評価用データセットDSと、ステップS 1 0 3で算出した関連度とを参照して複数のインサイトサブジェクトを生成する。第1の生成部1 5 - 1は一例として、評価用データセットDSの構成要素のうち、関連度がゼロではない構成要素を含むインサイトサブジェクトを生成する。なお、第1の生成部1 5 - 1が行うインサイトサブジェクトの生成方法は上述したものに限られない。第1の生成部1 5 - 1は例えば、関連度が所定の条件を満たす(関連度が閾値以上である)構成要素を含むインサイトサブジェクトを生成してもよい。

【0073】

また、ステップS 1 0 4において、第1の生成部1 5 - 1は、コンテキストが与えられていない、又は、関連のあるデータの構成要素が存在しなかった場合には、データテーブル

10

20

30

40

50

ルにおいて考えられる全ての可視化情報Vについてインサイトサブジェクトを生成してもよい。

【0074】

より具体的には、第1の生成部15-1は例えば、関連データVDに含まれる集計データ $s^V$ を、可視化情報Vが表す表示態様のチャートにプロットした可視化結果を表すインサイトサブジェクトを生成する。このとき、第1の生成部15-1が、複数の可視化情報Vのそれぞれについてインサイトサブジェクトを生成することにより、複数のインサイトサブジェクトが生成される。また、1つの可視化情報Vについて1つのインサイトサブジェクトが生成されるため、可視化情報Vとインサイトサブジェクトとは1対1で対応する。なお、インサイトサブジェクトは可視化候補を表すデータに限られず、例えば、可視化情報Vがそのままインサイトサブジェクトとして扱われてもよい。

10

【0075】

(ステップS105)

ステップS105では、評価部13は、インサイトサブジェクトに対して評価を実行する。評価部13は一例として、複数のインサイトサブジェクトに対して、コンテキストデータCDに応じた評価を行う。

【0076】

より具体的には、評価部13は一例として、複数のインサイトサブジェクトの各々に対して、関連データVDとコンテキストデータCDとを参照した評価を行う。このとき、複数のインサイトサブジェクトは関連情報Vと1対1で対応しているため、評価部13は、可視化情報Vのそれぞれについて評価を行う。換言すると、評価部13は、複数のインサイトサブジェクトの各々に対して、関連データVDに含まれる関連情報V毎に評価を行う。

20

【0077】

評価部13は例えば、関連データ及びコンテキストデータの少なくとも一部が入力され、評価値を出力する評価モデルを用いて、複数のインサイトサブジェクトに対して評価を行う。図9は、評価部13が行う評価処理の例を示す図である。図9の例で、評価部13は、可視化情報VとコンテキストデータCD、入力データDが入力され、インサイトスコア $y^A$ を出力する評価モデルfを用いて、複数のインサイトサブジェクトに対して評価を行う。評価モデルfは、予め規定されたスコア関数であってもよく、また、機械学習により構築された学習済モデルであってもよい。評価モデルfを用いた評価方法については後述する。評価モデルfによる可視化情報V毎のインサイトスコア $y^A$ は記憶部17に評価結果ERとして記憶される。

30

【0078】

(ステップS106)

ステップS106では、表示部14は、インサイトサブジェクトに関連する情報を表示する。表示部14は、一例として、関連度算出部12が算出した関連度に基づき、コンテキストと関連する可視化情報Vを列挙した情報を表示する。ここで、表示部14は、コンテキストが存在しない場合又は関連するインサイトサブジェクトがない場合、第1の生成部15-1が生成した全てのインサイトサブジェクトを列挙した情報を表示してもよい。

【0079】

より具体的には、例えば、表示部14は、評価部13による評価結果と共に、又は評価部13による評価結果に応じた表示態様にて、複数のインサイトサブジェクトの少なくとも何れかを表示する。評価結果に応じた表示態様とは、例えば、表示順、又は、表示の大きさを含む。このとき、表示部19は、評価部13による評価が相対的に高いインサイトサブジェクトを、評価部13による評価が相対的に低いインサイトサブジェクトよりも優先的に表示してもよい。

40

【0080】

評価結果の表示例について、図10~図13を参照しつつ説明する。図10は、可視化結果の表示例1を示す図である。図10の例では、「1.商品Aの顧客数の推移」、「2.商品Aの顧客数(年齢)」及び「3.商品Aの売上高(年齢)」の3つの可視化結果で

50

あるチャートC101~C103が、インサイトスコア $y^A$ の高い順に表示される。また、チャートC101~C103のそれぞれに対応するインサイトスコア $y^A$ 及び可視化情報Vを含む情報V101~V103が、チャートC101~C103に対応付けて表示される。

【0081】

また、表示例1では、チャートC101~C103のそれぞれについて、可視化結果をユーザが評価するための評価ボタンC111~C113が表示される。また入力データDの再評価を行うための検索窓C114及び再評価ボタンC115が表示される。入力データDの再評価については後述する。

【0082】

チャートC101が表示されることにより、ユーザは例えば「商品Aの顧客数は年々減少してしまっている」といった情報を把握することができる。また、チャートC102が表示されることにより、ユーザは例えば、「商品Aの主要顧客の年代は20代である」といった情報を把握することができる。

【0083】

図11は、インサイトサブジェクトを評価結果と共に表示する例を示す図である。図11の例で、インサイトサブジェクトV7、V3、V8、...は、入力データDの可視化結果を表すチャートであり、インサイトサブジェクトV7、V3、V8、...の視覚的特徴はそれぞれ異なっている。インサイトサブジェクトV7、V3、V8、...のそれぞれに隣接して各インサイトサブジェクトのインサイトスコア $y^A$ が表示される。また、複数のインサイトサブジェクトV7、V3、V8、...は、インサイトスコア $y^A$ の降順で表示される。

【0084】

図11の例によれば、複数のインサイトサブジェクトがインサイトスコア $y^A$ の降順で表示されることにより、どのインサイトサブジェクトの評価が高いかをユーザが把握しやすい。

【0085】

図12は、可視化情報Vを評価結果と共に表示する例を示す図である。図12の例では、表示部14は、関連データに含まれる各関連情報Vと、評価部13による評価とを対応付けて表示する。具体的には、表示部19は、可視化情報V11~V18と、可視化情報V11~V18のそれぞれに対応するインサイトスコア $y^A$ とを対応付けて表示する。

【0086】

図13は、インサイトサブジェクトを評価結果と共に表示する例を示す図である。図13の例では、表示部14は、入力データDの可視化結果であるチャート(棒グラフ)を表示するとともに、表示したチャートに対応するインサイトスコア $y^A$ をチャートと共に表示する。

【0087】

ユーザは、表示部14の表示により、ユーザの求めるインサイトを与える可能性の高い可視化結果を優先的に視認することができる。このように、本例示の実施形態によれば、ユーザの求めるインサイトをユーザに与えやすい表示出力を行うことができる。

【0088】

(ステップS105の評価の具体例)

次いで、評価部13がステップS105で行う評価の具体例について、図14~16を参照しつつ説明する。評価部13は、一例として、特徴ベクトル $d^C$ 、特徴ベクトル $d^V$ 、集計データ $s^V$ 及び統計量 $t^V$ を用いてインサイトサブジェクトに対する評価を行う。

【0089】

図14は、特徴ベクトル $d^C$ 及び特徴ベクトル $d^V$ の生成例を示す図である。図14に示すように、可視化情報Vから特徴ベクトル $d^V$ が生成されるとともに、コンテキストCから特徴ベクトル $d^C$ が生成される。

【0090】

図15は、第2の生成部15-2が生成する集計データ $s^V$ 及び統計量 $t^V$ の例を示す

10

20

30

40

50

図である。図 15 の例で、集計データ  $s^V$  は、入力データ  $D$  に含まれるデータであって可視化情報  $V$  に対応するデータを集計して得られるデータである。統計量  $t^V$  は、集計データ  $s^V$  の統計量を表すデータである。

【0091】

以下では、評価部 13 が実行する評価の具体例として、ルールベースでの評価と学習ベースでの評価とについて説明する。

【0092】

(ルールベースでの評価)

ルールベースの場合、評価部 13 は関連データ  $V D$  を用いてスコア  $y_0^{\wedge}$  を計算し、スコア  $y_0^{\wedge}$  を用いてインサイトスコア  $y^{\wedge}$  を計算する。このとき、評価部 13 は、スコア  $y_0^{\wedge}$  をそのままインサイトスコア  $y^{\wedge}$  として用いてもよいし、スコア  $y_0^{\wedge}$  に正規化又は重み付けなどの処理を加えてインサイトスコア  $y^{\wedge}$  を算出してもよい。

10

【0093】

スコア  $y_0^{\wedge}$  の計算方法は限定されないが、評価部 13 は例えば、インサイトの種別ごとにルールベースで定義されたスコア関数を用いてもよく、また、インサイトを与えるチャートの特徴量を学習するモデルを用いてスコア  $y_0^{\wedge}$  を算出してもよい。

【0094】

スコア関数を用いる場合、スコア関数は、一例として、関連データ  $V D$  とコンテキストデータ  $C D$  との関連性が高い程、高い評価値を出力する関数である。換言すると、評価部 13 は、予め規定されたスコア関数であって、関連データ  $V D$  とコンテキストデータ  $C D$  との関連性が高い程、高い評価値を出力するスコア関数を用いて、複数のインサイトサブジェクトに対して評価を行う。

20

【0095】

(ルールベースでの評価の例 1)

評価部 13 は例えば、コンテキストデータ  $C D$  との関連性が低い関連データ  $V D$  についてのインサイトスコア  $y^{\wedge}$  をゼロ、又はマイナス値にするなどして、評価結果が低くなるようにする。コンテキストデータ  $C D$  と関連データ  $V D$  との関連性の程度(類似度)の計算方法は限定されないが、評価部 13 は例えば、集合の類似度(Jaccard、Dice、Simpson、等)、文字列の類似度(ハミング距離、レーベンシュタイン距離、ジャロ・ウィンクラー距離、等)、分散表現(word2vec、fastText、BERT、等)の類似度を用いる。

30

【0096】

(ルールベースでの評価の例 2)

また、評価部 13 は、コンテキストデータ  $C D$  と関連データ  $V D$  の類似度によって重み付けされたスコアを用いて、インサイトスコア  $y^{\wedge}$  を算出してもよい。より具体的には、例えば、関連データ  $V D$  を用いて計算されたスコア  $y_0^{\wedge}$  と、類似度  $sim(CD, DV)$  の積をインサイトスコア  $y^{\wedge}$  としてもよい。

【0097】

(学習ベースでの評価)

学習ベースの場合、評価部 13 は、予め学習された評価モデルであって、関連データ  $V D$  とコンテキストデータ  $C D$  とが入力され、評価値を出力する評価モデル  $f$  を用いて、複数のインサイトサブジェクトに対して評価を行う。評価モデル  $f$  の機械学習の手法は限定されず、一例として、決定木ベース、線形回帰、又はニューラルネットワークの手法が用いられてもよく、また、これらのうちの 1 以上の手法が用いられてもよい。決定木ベースとしては、例えば、LightGBM(Light Gradient Boosting Machine)、及びXGBoostが挙げられる。線形回帰としては、例えば、サポートベクター回帰、Ridge回帰、Lasso回帰、及びElasticNetが挙げられる。ニューラルネットワークとしては、例えばディープラーニングが挙げられる。

40

【0098】

評価モデル  $f$  の学習においては、インサイトがあるとみなせる任意の教師データが利用可能である。例えば、過去にデータ分析者によって作成されたチャートがインサイトを与

50

える特徴を含んでいるとみなして、それらの可視化情報Vをポジティブサンプルとして学習に用いてもよい。また、インサイトがないと考えられるチャートの可視化情報Vをネガティブサンプルとして学習に用いてもよい。

【0099】

図16は、評価モデルfの一例を示す図である。図16の例で、評価モデルfの入力は、特徴ベクトル $d^V$ 、特徴ベクトル $d^C$ 、集計データ $s^V$ 、及び統計量 $t^V$ を含む。評価モデルfの出力は、評価結果であり、一例として、ユーザの求めるインサイトを与えるかを示すラベル確率である。

【0100】

(学習ベースにおける評価モデルの例1)

可視化情報Vのインサイトに関する教師ラベルyが与えられる場合には、分類モデルとして評価モデルを学習させることができる。例えば、y {0, 1}が1の場合はインサイトあり、0の場合はインサイトなし、を示すラベルとして与えられているとき、2クラス分類タスクとして、例えば以下の式(1)により与えられる損失関数E( )を最小化する機械学習モデルを学習させればよい。式(1)において、Nは学習データの数である。

【数1】

$$E(\theta) = -\sum_{i=1}^N [y_i \log f(VD_i, CD_i; \theta) + (1 - y_i) \log \{1 - f(VD_i, CD_i; \theta)\}] \quad \dots (1)$$

【0101】

上記損失関数を最小化する機械学習モデルの出力は、 $p(y = 1 | VD_i, CD_i)$ 、すなわちインサイトありと判定される確率と解釈することができ、これをインサイトスコア $y^{\wedge}$ として用いることができる。

【0102】

(学習ベースにおける評価モデルの例2)

可視化情報Vごとの可視化のよさを表すスコアやランキングが教師データとして与えられる場合、回帰モデルとして評価モデルを学習させることができる。例えば、yを教師データによって与えられたスコアであるとした場合、例えば以下の式(2)により与えられる損失関数E( )を最小化する機械学習モデルを学習させればよい。式(2)において、Nは学習データの数である。

【数2】

$$E(\theta) = \frac{1}{2} \sum_{i=1}^N \|y_i - f(VD_i, CD_i; \theta)\|^2 \quad \dots (2)$$

【0103】

上記の損失関数を最小化する機械学習モデルの出力は、教師データのスコアと同様に可視化のよさを表すスコアであり、これをインサイトスコア $y^{\wedge}$ として用いてもよい。

【0104】

[例示的实施形態3]

本発明の第3の例示的实施形態について、図面を参照して詳細に説明する。なお、例示的实施形態1にて説明した構成要素と同じ機能を有する構成要素については、同じ符号を付記し、その説明を繰り返さない。

【0105】

図17は、本例示的实施形態に係る情報処理装置1Bの構成を示すブロック図である。図17に示すように、情報処理装置1Bは、例示的实施形態2に係る情報処理装置1Aの制御部10Aに代えて制御部10Bを備える。制御部10Bは、取得部11、関連度算出部12、評価部13、表示部14、第1の生成部15-1及び第2の生成部15-2に加えて、学習部16を備える。

10

20

30

40

50

## 【0106】

本例示的实施形態において、入力部20は、評価部13の評価結果に対するユーザからのフィードバックを受け付ける。また、評価部13は、ユーザからのフィードバックを参照して、複数のインサイトサブジェクトに対して、再度の評価を行う。

## 【0107】

例えば、入力部20が受け付けたユーザからのフィードバックが、ユーザからの評価が高いことを示している場合に、取得部11は、フィードバックが反映されたコンテキストデータを取得する。関連度算出部12は、フィードバックが反映されたコンテキストデータと評価用データセットの構成要素との関連度を算出する。評価部13は、評価用データセットと関連度とを参照して生成された複数のインサイトサブジェクトに対して、評価を行う。

10

## 【0108】

コンテキスト及びユーザクエリは任意のタイミングで更新することが可能である。コンテキスト及びユーザクエリは例えば、図10の検索窓C114にユーザが文字列を入力して再評価ボタンC115を選択することにより更新される。コンテキスト及びユーザクエリの更新後にはコンテキストデータの取得処理、及びインサイトサブジェクトに関連する情報の表示処理を情報処理装置1Bが再度実行し、これにより、表示部19に表示される情報が切り替わる。

## 【0109】

具体的には例えば、図10の表示例において、ユーザが評価ボタンC111~C113のいずれかを選択する操作、又は、検索窓C114への入力操作及び再評価ボタンC115の選択操作を行うと、取得部11は、ユーザの操作内容に基づくコンテキストデータを生成し、生成したコンテキストデータを用いて、図4に示す情報処理方法S1Aを実行する。これにより、ユーザのフィードバックが反映された情報の表示が行われる。

20

## 【0110】

また、学習部16は、ユーザからのフィードバックを参照して、評価モデルfを再学習してもよい。この場合、学習部16は例えば、表示部19が表示したインサイトサブジェクトに関連する情報(インサイトスコア $y^h$ 、可視化情報V、チャート、等)に関するユーザの操作履歴を、ユーザからのフィードバックとして記憶部17等に記録する。ユーザの操作履歴は、一例として、インサイトサブジェクトに関連する情報の表示時間、インサイトサブジェクトに関連する情報に対する評価ボタンの押下、等を含む。

30

## 【0111】

学習部16は、ユーザからのフィードバックを反映した評価モデルfの再学習を行う。学習部16は例えば、評価の高い可視化情報Vをポジティブサンプルとし、評価の低い可視化情報をネガティブサンプルとして、評価モデルfの再学習を行う。

## 【0112】

本例示的实施形態に係る情報処理装置1Bにおいては、入力部20が評価結果に対するユーザからのフィードバックを受け付け、評価部13が、ユーザからのフィードバックを参照して、複数のインサイトサブジェクトに対して、再度の評価を行う。このため、本例示的实施形態に係る情報処理装置1Bによれば、例示的实施形態1に係る情報処理装置1の奏する効果に加えて、インサイトサブジェクトに対する評価の精度をより高くすることができるという効果が得られる。

40

## 【0113】

## 〔変形例〕

上述の例示的实施形態1において、1つの情報処理装置1が行っていた処理は、複数の情報処理装置に分担させてもよい。言い換えれば、情報処理装置1が行う処理の一部を、少なくとも1つの他の情報処理装置に実行させてもよい。さらに言い換えれば、上述の各処理を少なくとも1つのプロセッサに行わせる場合、その少なくとも1つのプロセッサは、1つの情報処理装置1が備えているものであってもよいし、それぞれ異なる情報処理装置が備えているものであってもよい。これは、上述の例示的实施形態2における情報処理

50

装置 1 A、および例示的实施形態 3 における情報処理装置 1 B についても同様である。

【0114】

〔ソフトウェアによる実現例〕

情報処理装置 1、1 A、1 B の一部又は全部の機能は、集積回路（ICチップ）等のハードウェアによって実現してもよいし、ソフトウェアによって実現してもよい。

【0115】

後者の場合、情報処理装置 1、1 A、1 B は、例えば、各機能を実現するソフトウェアであるプログラムの命令を実行するコンピュータによって実現される。このようなコンピュータの一例（以下、コンピュータ C と記載する）を図 18 に示す。コンピュータ C は、少なくとも 1 つのプロセッサ C 1 と、少なくとも 1 つのメモリ C 2 と、を備えている。メモリ C 2 には、コンピュータ C を情報処理装置 1、1 A、1 B として動作させるためのプログラム P が記録されている。コンピュータ C において、プロセッサ C 1 は、プログラム P をメモリ C 2 から読み取って実行することにより、情報処理装置 1、1 A、1 B の各機能が実現される。

10

【0116】

プロセッサ C 1 としては、例えば、CPU（Central Processing Unit）、GPU（Graphic Processing Unit）、DSP（Digital Signal Processor）、MPU（Micro Processing Unit）、FPU（Floating point number Processing Unit）、PPU（Physics Processing Unit）、マイクロコントローラ、又は、これらの組み合わせなどを用いることができる。メモリ C 2 としては、例えば、フラッシュメモリ、HDD（Hard Disk Drive）、SSD（Solid State Drive）、又は、これらの組み合わせなどを用いることができる。

20

【0117】

なお、コンピュータ C は、プログラム P を実行時に展開したり、各種データを一時的に記憶したりするための RAM（Random Access Memory）を更に備えていてもよい。また、コンピュータ C は、他の装置との間でデータを送受信するための通信インタフェースを更に備えていてもよい。また、コンピュータ C は、キーボードやマウス、ディスプレイやプリンタなどの入出力機器を接続するための入出力インタフェースを更に備えていてもよい。

【0118】

また、プログラム P は、コンピュータ C が読み取り可能な、一時的でない有形の記録媒体 M に記録することができる。このような記録媒体 M としては、例えば、テープ、ディスク、カード、半導体メモリ、又はプログラマブルな論理回路などを用いることができる。コンピュータ C は、このような記録媒体 M を介してプログラム P を取得することができる。また、プログラム P は、伝送媒体を介して伝送することができる。このような伝送媒体としては、例えば、通信ネットワーク、又は放送波などを用いることができる。コンピュータ C は、このような伝送媒体を介してプログラム P を取得することもできる。

30

【0119】

〔付記事項 1〕

本発明は、上述した実施形態に限定されるものでなく、請求項に示した範囲で種々の変更が可能である。例えば、上述した実施形態に開示された技術的手段を適宜組み合わせ得られる実施形態についても、本発明の技術的範囲に含まれる。

40

【0120】

〔付記事項 2〕

上述した実施形態の一部又は全部は、以下のようにも記載され得る。ただし、本発明は、以下の記載する態様に限定されるものではない。

【0121】

（付記 1）

評価用データセット及びコンテキストデータを取得する取得手段と、  
前記コンテキストデータと前記評価用データセットの構成要素との関連度を算出する関

50

連度算出手段と、

前記評価用データセットと前記関連度とを参照して生成された複数のインサイトサブジェクトに対して、評価を行う評価手段と、

前記インサイトサブジェクトに関連する情報を表示する表示手段と、  
を備える情報処理装置。

【0122】

上記の構成によれば、ユーザの求めるインサイトを与える情報を表示することができる。

【0123】

(付記2)

前記評価用データセットと前記関連度とを参照して、前記複数のインサイトサブジェクトを生成する第1の生成手段を更に備える、付記1に記載の情報処理装置。

10

【0124】

上記の構成によれば、ユーザの求めるインサイトを与える情報を表示することができる。

【0125】

(付記3)

前記第1の生成手段は、

前記評価用データセットの構成要素のうち、前記関連度がゼロではない構成要素を含むインサイトサブジェクトを生成する、付記2に記載の情報処理装置。

【0126】

上記の構成によれば、コンテキストデータとの関連度がゼロでない構成要素を含むインサイトサブジェクトに関する情報を表示することができる。

20

【0127】

(付記4)

前記評価用データセットには、評価用データと、当該評価用データに関連する関連データが含まれており、

前記評価手段は、前記複数のインサイトサブジェクトの各々に対して、前記関連データに含まれる関連情報毎に評価を行う、付記1から3の何れか1つに記載の情報処理装置。

【0128】

上記の構成によれば、関連情報毎にインサイトサブジェクトを評価することができる。

【0129】

(付記5)

前記表示手段は、前記評価手段による評価が相対的に高いインサイトサブジェクトを、前記評価手段による評価が相対的に低いインサイトサブジェクトよりも優先的に表示する、付記4に記載の情報処理装置。

30

【0130】

上記の構成によれば、インサイトサブジェクトに対する評価の程度をユーザに提示することができる。

【0131】

(付記6)

前記表示手段は、前記関連データに含まれる各関連情報と、前記評価手段による評価結果とを対応付けて表示する、付記4又は5に記載の情報処理装置。

40

【0132】

上記の構成によれば、表示手段が表示する情報によりユーザが複数のインサイトサブジェクトのそれぞれに対する評価を把握することができる。

【0133】

(付記7)

前記表示手段は、

前記評価手段による評価結果と共に、又は前記評価手段による評価結果に応じた表示態様にて、前記複数のインサイトサブジェクトの少なくとも何れかを表示する、付記1から6の何れか1つに記載の情報処理装置。

50

## 【 0 1 3 4 】

上記の構成によれば、表示手段が表示するインサイトサブジェクトによりユーザがインサイトサブジェクトに対する評価をより把握し易くすることができる。

## 【 0 1 3 5 】

( 付記 8 )

前記評価手段は、複数のインサイトサブジェクトに対して、前記コンテキストデータに応じた評価を行う、付記 1 から 7 の何れか 1 つに記載の情報処理装置。

## 【 0 1 3 6 】

上記の構成によれば、データの可視化候補がユーザの求めるインサイトを与えるかについて評価することができる。

## 【 0 1 3 7 】

( 付記 9 )

参照情報を参照して、前記コンテキストデータを生成する第 2 の生成手段を更に備えている、付記 1 から 8 の何れか 1 つに記載の情報処理装置。

## 【 0 1 3 8 】

上記の構成によれば、参照情報に対応するインサイトを与える情報を表示することができる。

## 【 0 1 3 9 】

( 付記 1 0 )

前記評価手段の評価結果に対するユーザからのフィードバックを受け付ける受付手段を更に備え、

前記評価手段は、前記ユーザからのフィードバックを参照して、前記複数のインサイトサブジェクトに対して、再度の評価を行う、付記 1 から 9 の何れか 1 つに記載の情報処理装置。

## 【 0 1 4 0 】

上記の構成によれば、インサイトサブジェクトに対する評価の精度をより高くすることができる。

## 【 0 1 4 1 】

( 付記 1 1 )

前記受付手段が受け付けた前記ユーザからのフィードバックが、ユーザからの評価が高いことを示している場合に、

前記取得手段は、前記フィードバックが反映されたコンテキストデータを取得し、

前記関連度算出手段は、前記フィードバックが反映されたコンテキストデータと前記評価用データセットの構成要素との関連度を算出し、

前記評価手段は、前記評価用データセットと前記関連度とを参照して生成された複数のインサイトサブジェクトに対して、評価を行う、付記 1 0 に記載の情報処理装置。

## 【 0 1 4 2 】

上記の構成によれば、インサイトサブジェクトに対する評価の精度をより高くすることができる。

## 【 0 1 4 3 】

( 付記 1 2 )

少なくとも 1 つのプロセッサが、

評価用データセット及びコンテキストデータを取得すること、

前記コンテキストデータと前記評価用データセットの構成要素との関連度を算出すること、

前記評価用データセットと前記関連度とを参照して生成された複数のインサイトサブジェクトに対して、評価を行うこと、及び、

前記インサイトサブジェクトに関連する情報を表示すること、を含む情報処理方法。

## 【 0 1 4 4 】

10

20

30

40

50

( 付記 1 3 )

コンピュータに、  
評価用データセット及びコンテキストデータを取得する処理と、  
前記コンテキストデータと前記評価用データセットの構成要素との関連度を算出する処理と、

前記評価用データセットと前記関連度とを参照して生成された複数のインサイトサブジェクトに対して、評価を行う処理と、

前記インサイトサブジェクトに関連する情報を表示する処理と、  
を実行させるプログラム。

【 0 1 4 5 】

10

〔 付記事項 3 〕

上述した実施形態の一部又は全部は、更に、以下のように表現することもできる。

【 0 1 4 6 】

少なくとも1つのプロセッサを備え、前記プロセッサは、評価用データセット及びコンテキストデータを取得する取得処理と、前記コンテキストデータと前記評価用データセットの構成要素との関連度を算出する関連度算出処理と、前記評価用データセットと前記関連度とを参照して生成された複数のインサイトサブジェクトに対して、評価を行う評価処理と、前記インサイトサブジェクトに関連する情報を表示する表示処理とを実行する情報処理装置。

【 0 1 4 7 】

20

なお、この情報処理装置は、更にメモリを備えていてもよく、このメモリには、前記取得処理と、前記関連度算出処理と、前記評価処理と、前記表示処理とを前記プロセッサに実行させるためのプログラムが記憶されていてもよい。また、このプログラムは、コンピュータ読み取り可能な一時的でない有形の記録媒体に記録されていてもよい。

【 符号の説明 】

【 0 1 4 8 】

1、 1 A、 1 B 情報処理装置

1 0 A、 1 0 B 制御部

1 1 取得部 ( 取得手段 )

1 2 関連度算出部 ( 関連度算出手段 )

30

1 3 評価部 ( 評価手段 )

1 4 表示部 ( 表示手段 )

1 5 - 1 第 1 の生成部

1 5 - 2 第 2 の生成部

1 6 学習部

1 7 記憶部

1 8 通信部

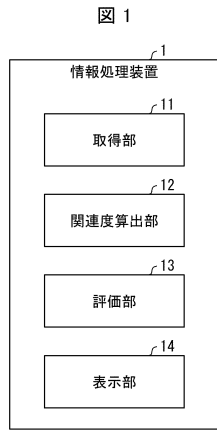
1 9 表示部

2 0 入力部 ( 受付手段 )

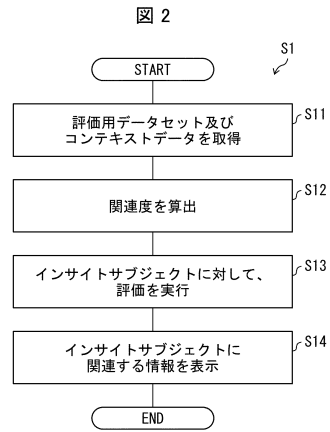
40

【図面】

【図 1】

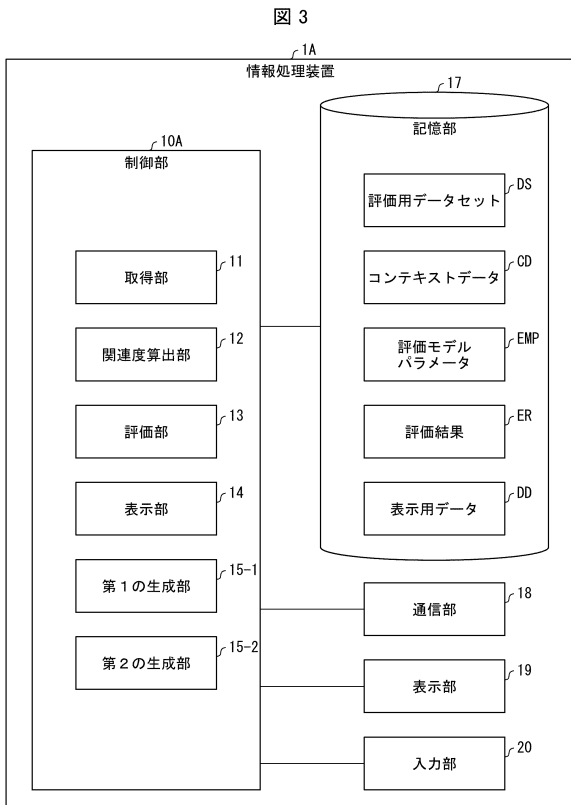


【図 2】

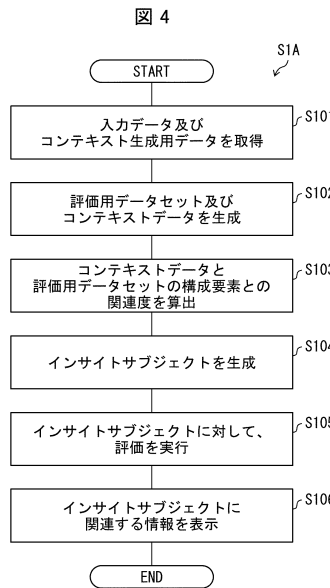


10

【図 3】



【図 4】



20

30

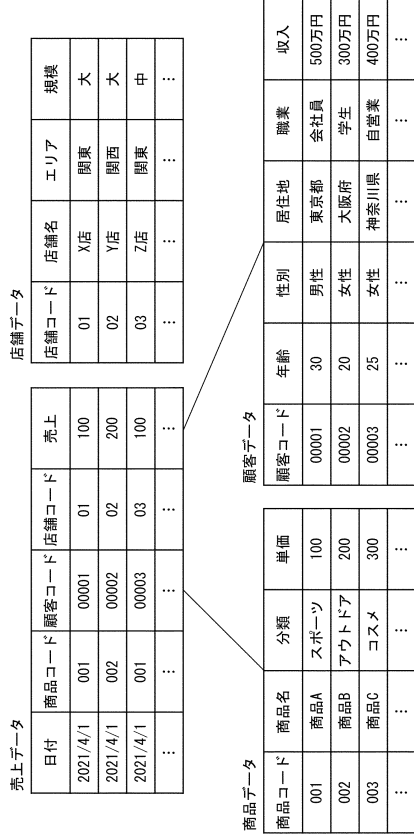
40

50

【 図 5 】

図 5

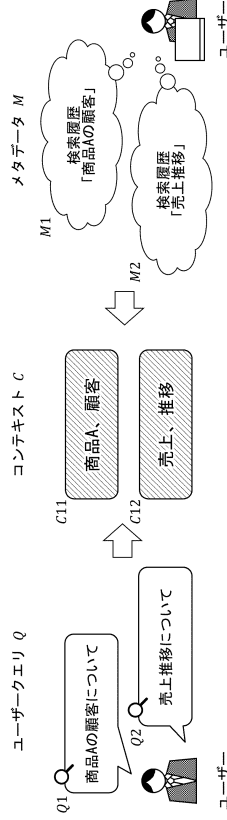
入力データの例



【 図 6 】

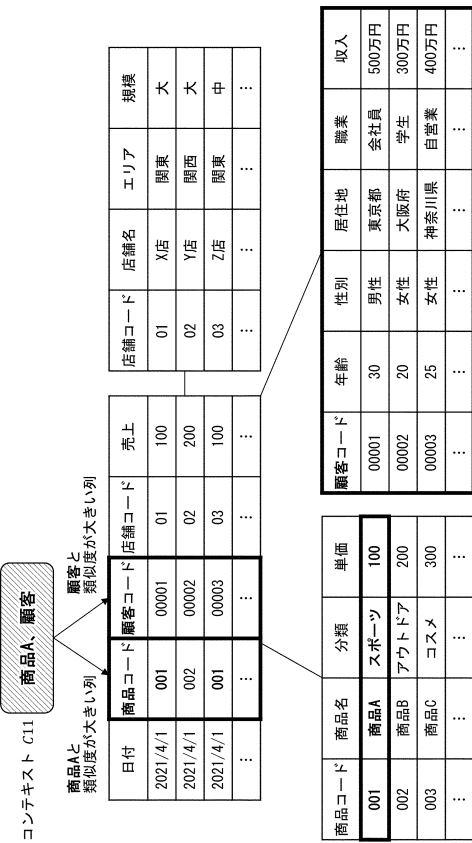
図 6

コンテキストの生成例



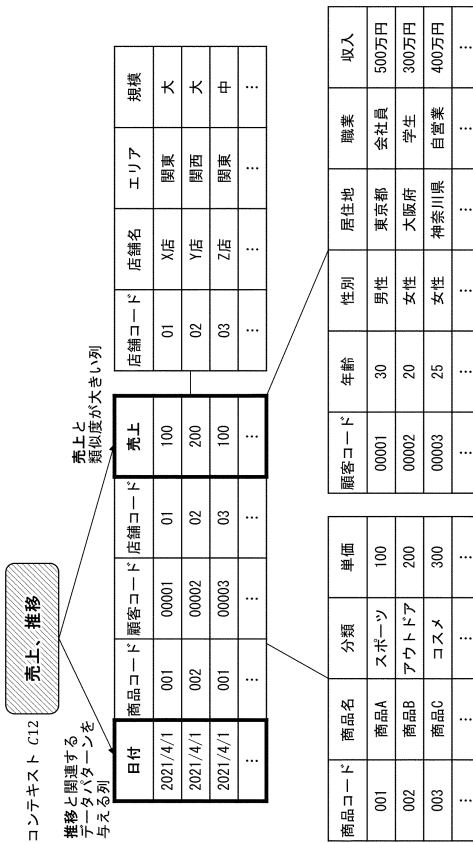
【 図 7 】

図 7



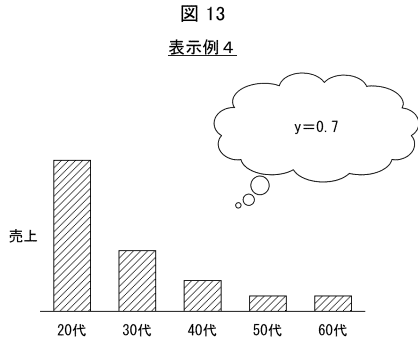
【 図 8 】

図 8

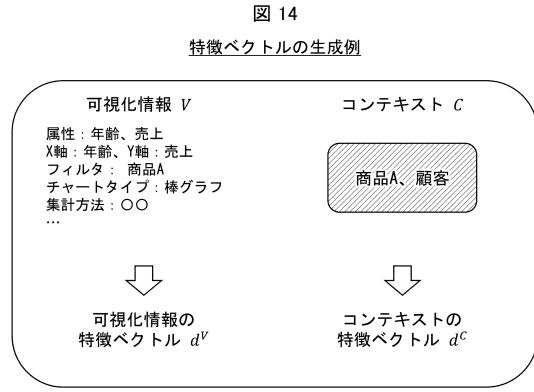




【 図 1 3 】



【 図 1 4 】

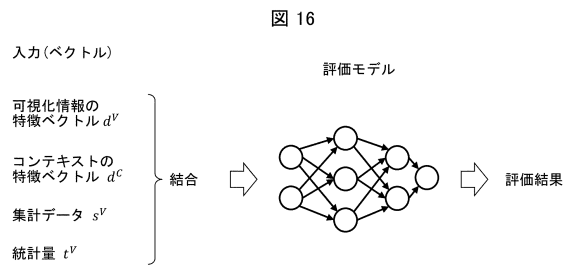


10

【 図 1 5 】



【 図 1 6 】



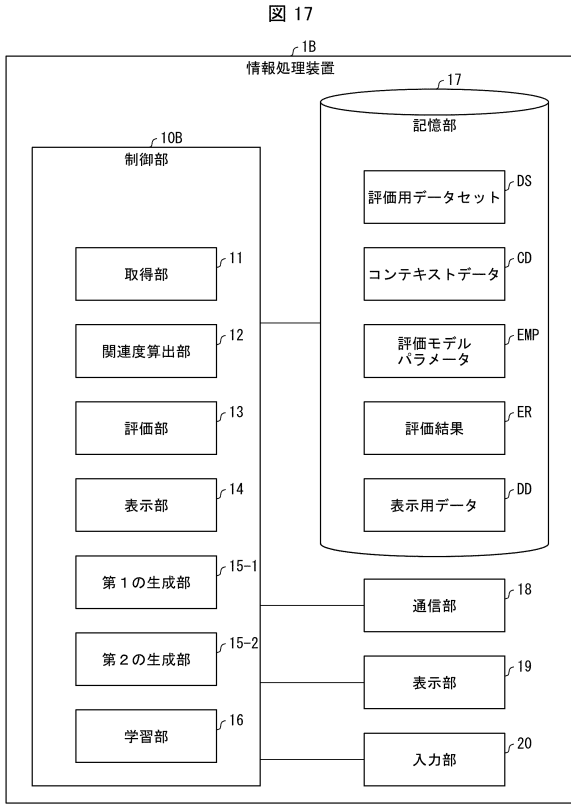
20

30

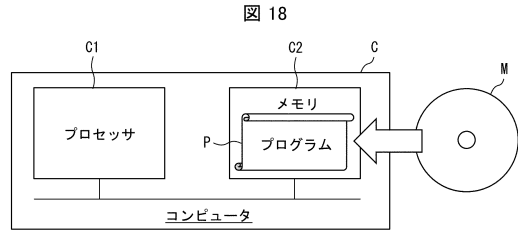
40

50

【 図 1 7 】



【 図 1 8 】



10

20

30

40

50

---

フロントページの続き

- 株式会社内  
(72)発明者 小山田 昌史  
東京都港区芝五丁目7番1号 日本電気株式会社内
- 審査官 塩田 徳彦
- (56)参考文献 特開2014-056573(JP,A)  
国際公開第2018/173251(WO,A1)  
特開2019-128634(JP,A)  
国際公開第2016/027318(WO,A1)
- (58)調査した分野 (Int.Cl., DB名)  
G06Q 10/00-99/00