

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3767516号

(P3767516)

(45) 発行日 平成18年4月19日(2006.4.19)

(24) 登録日 平成18年2月10日(2006.2.10)

(51) Int. Cl.

F I

G06F 17/27 (2006.01)

G06F 17/27 Z

G06F 17/30 (2006.01)

G06F 17/30 170A

G06Q 10/00 (2006.01)

G06F 17/30 220Z

G06F 17/60 170A

請求項の数 11 (全 10 頁)

(21) 出願番号 特願2002-134334 (P2002-134334)
 (22) 出願日 平成14年5月9日(2002.5.9)
 (65) 公開番号 特開2003-330922 (P2003-330922A)
 (43) 公開日 平成15年11月21日(2003.11.21)
 審査請求日 平成15年8月12日(2003.8.12)

(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
 (74) 代理人 100064621
 弁理士 山川 政樹
 (72) 発明者 森永 聡
 東京都港区芝五丁目7番1号 日本電気株
 式会社内
 (72) 発明者 山西 健司
 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 和田 財太

最終頁に続く

(54) 【発明の名称】 典型文分析方法、装置および記録媒体

(57) 【特許請求の範囲】

【請求項1】

各種情報を蓄積する蓄積部と、この蓄積部の情報に対して演算処理を行う処理部とを有する典型文分析装置により、所定の分析対象に関する記述を示す複数のテキストを要素とする分析対象集合について、前記各テキストからその分析対象集合を代表する度合い进行分析する典型文分析方法であって、

前記蓄積部に蓄積されている、前記分析対象集合と、前記分析対象と比較する各比較対象に関する記述を示す複数のテキストをそれぞれ要素とする複数の比較対象集合とから、前記処理部により、前記分析対象集合に含まれる各テキストがその分析対象集合を代表する度合いとして、前記分析対象集合に含まれる各テキストごとに、前記分析対象集合に対する当該テキストのベイズ事後確率を計算する事後確率計算ステップを備える

10

ことを特徴とする典型文分析方法。

【請求項2】

請求項1記載の典型文分析方法において、

前記処理部により、前記分析対象集合に含まれる各テキストのうち、ベイズ事後確率が高い順に1つ以上のテキストを前記分析対象集合の典型文として抽出するテキスト抽出ステップをさらに備えることを特徴とする典型文分析方法。

【請求項3】

請求項1または2記載の典型文分析方法において、

前記事後確率計算ステップでベイズ事後確率を計算する際、

20

前記分析対象集合および前記各比較対象集合に含まれる各テキストから、当該テキストを分解して得られる各单位表現の当該集合における出現回数を計数するとともに、前記分析対象集合および前記各比較対象集合の各集合に含まれるテキスト数をそれぞれの集合ごとに計数し、これら計数結果に基づいて前記分析対象集合に含まれる各テキストごとのベイズ事後確率を計算することを特徴とする典型文分析方法。

【請求項 4】

請求項 3 記載の典型文分析方法において、

前記事後確率計算ステップで前記計数結果からベイズ事後確率を計算する際、

前記各单位表現の出現回数に基づき個々の単位表現の前記各集合における出現確率を推定するとともに、前記各集合のテキスト数に基づき前記各集合の出現確率を推定し、これら推定結果を用いて前記分析対象集合に含まれる各テキストごとのベイズ事後確率を計算することを特徴とする典型文分析方法。

10

【請求項 5】

請求項 1 ～ 4 のいずれか記載の典型文分析方法において、

前記比較対象集合群の各集合を空集合とすることを特徴とする典型文分析方法。

【請求項 6】

各種情報を蓄積する蓄積部と、この蓄積部の情報に対して演算処理を行う処理部とを有し、所定の分析対象に関する記述を示す複数のテキストを要素とする分析対象集合について、前記各テキストからその分析対象集合を代表する度合い进行分析する典型文分析装置であって、

20

前記蓄積部に蓄積されている、前記分析対象集合と、前記分析対象と比較する各比較対象に関する記述を示す複数のテキストをそれぞれ要素とする複数の比較対象集合とから、前記処理部により、前記分析対象集合に含まれる各テキストがその分析対象集合を代表する度合いとして、前記分析対象集合に含まれる各テキストごとに、前記分析対象集合に対する当該テキストのベイズ事後確率を計算する事後確率計算手段を備えることを特徴とする典型文分析装置。

【請求項 7】

請求項 6 記載の典型文分析装置において、

前記処理部により、前記分析対象集合に含まれる各テキストのうち、ベイズ事後確率が高い順に 1 つ以上のテキストを前記分析対象集合の典型文として抽出するテキスト抽出手段を備えることを特徴とする典型文分析装置。

30

【請求項 8】

請求項 6 または 7 記載の典型文分析装置において、

前記処理部により、前記分析対象集合および前記各比較対象集合に含まれる各テキストから、当該テキストを分解して得られる各单位表現の当該集合における出現回数を計数する表現ヒストグラム計算手段と、

前記処理部により、前記分析対象集合および前記各比較対象集合の各集合に含まれるテキスト数をそれぞれの集合ごとに計数するカテゴリヒストグラム計算手段とをさらに備え、

前記事後確率計算手段は、これら計数結果に基づいて前記分析対象集合に含まれる各テキストごとのベイズ事後確率を計算することを特徴とする典型文分析装置。

40

【請求項 9】

請求項 8 記載の典型文分析装置において、

前記事後確率計算手段は、前記各单位表現の出現回数に基づき個々の単位表現の前記各集合における出現確率を推定するとともに、前記各集合のテキスト数に基づき前記各集合の出現確率を推定し、これら推定結果を用いて前記分析対象集合に含まれる各テキストごとのベイズ事後確率を計算することを特徴とする典型文分析装置。

【請求項 10】

請求項 6 ～ 9 のいずれか記載の典型文分析装置において、

前記比較対象集合群の各集合として空集合を用いることを特徴とする典型文分析装置。

50

【請求項 1 1】

各種情報を蓄積する蓄積部と、この蓄積部の情報に対して演算処理を行う処理部とを有する典型文分析装置により、所定の分析対象に関する記述を示す複数のテキストを要素とする分析対象集合について、前記各テキストからその分析対象集合を代表する度合いを分析する際に用いられ、前記典型文分析装置のコンピュータで、請求項 1～5 のいずれか記載の典型文分析方法を実行させるためのプログラムを記録した記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、典型文分析方法、装置および記録媒体に関し、特に特定の集合に属する各テキストから、その集合を代表する典型文を分析する典型文分析方法、装置および記録媒体に関するものである。

10

【0002】

【従来の技術】

ビジネス分野において、アンケート調査に基づくマーケティング分析は、商品やサービスに関する顧客の評価を取得する上で非常に重要視されている。特に、自由記述回答欄を含むアンケートでは、顧客の生の声を汲み取ることができ、アンケート以外にはweb上で自由に書き込まれた各種意見も同様である。

従来、このような自由記述欄などに記載された複数のテキスト（文字情報）からなる集合について、その集合を代表するテキストを抽出する場合、これらテキストを要約する技術が用いられていた（例えば、特開平10-134066号公報など参照）。この種の文書分析技術は、その集合に含まれる各テキストから1つの要約した文章を作成するものである。

20

【0003】

【発明が解決しようとする課題】

しかしながら、このような従来の文書分析技術では、分析対象となるテキスト集合を適切かつ詳細に分析できないという問題点があった。

通常、アンケート回答の分析では、「男性からの回答テキストの集合」などのように、趣味・嗜好の異なる複数の顧客層を含み類似性の低いテキスト集合についても分析対象となり、有効な代表テキストが必要とされる。しかし、従来の文書分析技術では、分析対象とする集合に含まれる各テキストが互いに類似していない場合、その集合を代表する適切な文章が得られない。

30

【0004】

また「男性からの回答における典型例」や「男性からの回答としては珍しい例」に相当するテキストの抽出など、各テキストがその集合を代表する度合いすなわち典型度が得られれば、より有益な情報が得られる。しかし、従来の文書分析技術では、元のテキスト集合から新たなテキストが生成されることから、その集合の要素である各テキストについて「その集合を代表するものとして、どれだけ相応しいか」を分析することができない。

本発明はこのような課題を解決するためのものであり、分析対象となるテキスト集合を適切かつ詳細に分析できる典型文分析方法、装置および記録媒体を提供することを目的としている。

40

【0005】

【課題を解決するための手段】

このような目的を達成するために、本発明にかかる典型文分析方法は、各種情報を蓄積する蓄積部と、この蓄積部の情報に対して演算処理を行う処理部とを有する典型文分析装置により、所定の分析対象に関する記述を示す複数のテキストを要素とする分析対象集合について、各テキストからその分析対象集合を代表する度合いを分析する典型文分析方法であって、蓄積部に蓄積されている、分析対象集合と、分析対象と比較する各比較対象に関する記述を示す複数のテキストをそれぞれ要素とする複数の比較対象集合とから、処理部により、分析対象集合に含まれる各テキストがその分析対象集合を代表する度合いとし

50

て、分析対象集合に含まれる各テキストごとに、分析対象集合に対する当該テキストのベイズ事後確率を計算する事後確率計算ステップを備えている。

【0006】

これに加えて、処理部により、分析対象集合に含まれる各テキストのうち、ベイズ事後確率が高い順に1つ以上のテキストを分析対象集合の典型文として抽出するテキスト抽出ステップをさらに備えてもよい。

【0007】

事後確率計算ステップでベイズ事後確率を計算する際、分析対象集合および各比較対象集合に含まれる各テキストから、当該テキストを分解して得られる各单位表現の当該集合における出現回数を計数するとともに、分析対象集合および各比較対象集合の各集合に含まれるテキスト数をそれぞれの集合ごとに計数し、これら計数結果に基づいて分析対象集合に含まれる各テキストごとのベイズ事後確率を計算するようにしてもよい。

10

【0008】

事後確率計算ステップで計数結果からベイズ事後確率を計算する際、各单位表現の出現回数に基づき個々の単位表現の各集合における出現確率を推定するとともに、各集合のテキスト数に基づき各集合の出現確率を推定し、これら推定結果を用いて分析対象集合に含まれる各テキストごとのベイズ事後確率を計算するようにしてもよい。

比較対象集合群の各集合を空集合としてもよい。

【0009】

また、本発明にかかる典型文分析装置は、各種情報を蓄積する蓄積部と、この蓄積部の情報に対して演算処理を行う処理部とを有し、所定の分析対象に関する記述を示す複数のテキストを要素とする分析対象集合について、各テキストからその分析対象集合を代表する度合いを分析する典型文分析装置であって、蓄積部に蓄積されている、分析対象集合と、分析対象と比較する各比較対象に関する記述を示す複数のテキストをそれぞれ要素とする複数の比較対象集合とから、処理部により、分析対象集合に含まれる各テキストがその分析対象集合を代表する度合いとして、分析対象集合に含まれる各テキストごとに、分析対象集合に対する当該テキストのベイズ事後確率を計算する事後確率計算手段を備えるものである。

20

【0010】

これに加えて、処理部により、分析対象集合に含まれる各テキストのうち、ベイズ事後確率が高い順に1つ以上のテキストを分析対象集合の典型文として抽出するテキスト抽出手段を設けてもよい。

30

【0011】

また、処理部により、分析対象集合および各比較対象集合に含まれる各テキストから、当該テキストを分解して得られる各单位表現の当該集合における出現回数を計数する表現ヒストグラム計算手段と、処理部により、分析対象集合および各比較対象集合の各集合に含まれるテキスト数をそれぞれの集合ごとに計数するカテゴリヒストグラム計算手段とをさらに備え、事後確率計算手段で、これら計数結果に基づいて分析対象集合に含まれる各テキストごとのベイズ事後確率を計算するようにしてもよい。

【0012】

40

さらに、事後確率計算手段で、各单位表現の出現回数に基づき個々の単位表現の各集合における出現確率を推定するとともに、各集合のテキスト数に基づき各集合の出現確率を推定し、これら推定結果を用いて分析対象集合に含まれる各テキストごとのベイズ事後確率を計算するようにしてもよい。

また、比較対象集合群の各集合として空集合を用いるようにしてもよい。

【0013】

また、本発明にかかる記録媒体は、各種情報を蓄積する蓄積部と、この蓄積部の情報に対して演算処理を行う処理部とを有する典型文分析装置により、所定の分析対象に関する記述を示す複数のテキストを要素とする分析対象集合について、各テキストからその分析対象集合を代表する典型文を分析する際に用いられ、典型文分析装置のコンピュータで、

50

上記各典型文分析方法を実行させるためのプログラムを記録したものである。

【0014】

【発明の実施の形態】

次に、本発明の実施の形態について図面を参照して説明する。

図1は本発明の一実施の形態にかかる典型文分析装置の構成を示すブロック図である。

この典型文分析装置100は、全体としてコンピュータからなり、蓄積部1、処理部2、分析対象テキスト入力手段3、比較対象テキスト集合群入力手段4、記憶部5、操作入力部6および画面表示部7が設けられている。

【0015】

分析対象テキスト入力手段3は、アンケート調査やwebから収集した分析対象、例えば
10 所定商品の評価に関して記述したテキスト（文字情報）の集合を入力する手段である。比較対象テキスト集合群入力手段4は、同じくアンケート調査やwebから収集した分析対象、例えば所定商品と同じ商品種別に属する他の商品の評価に関して記述したテキスト（文字情報）の集合を入力する手段である。これら分析対象テキスト入力手段3や比較対象テキスト集合群入力手段4は、テキストデータを操作入力するためのキーボードやweb上のテキストデータをダウンロードするアプリケーションなどの一般的な情報入力手段により構成される。

【0016】

蓄積部1は、分析対象テキスト入力手段3から入力されたテキストデータをカテゴリ10
20 というテキスト集合として蓄積するとともに、比較対象テキスト集合群入力手段4から入力されたテキストデータを各比較対象ごとにカテゴリ11～1nというテキスト集合として蓄積する手段である。この蓄積部1は、ハードディスクやメモリなどの情報記憶装置から構成される。

図2に各カテゴリの構成例を示す。例えばカテゴリ10は「商品Aへの意見」に関して記述されたテキストの集合であり、テキスト集合の要素として「商品Aは使いにくい。遅い。」や「商品Aは大きくなってしまった。小さい方がよい。」などのテキストが含まれている。

【0017】

操作入力部6は、オペレータが処理部2での分析処理を指示／制御するための操作を入力するための手段であり、キーボードやマウスから構成される。
30

画面表示部7は、処理部2での処理過程や分析結果を画面表示するための手段であり、CRT装置やLCD装置から構成される。

【0018】

記憶部5は、記録媒体9から予め読み込まれ処理部2で実行されるプログラム51や、処理部2での分析処理に用いる各種情報を記憶するハードディスクやメモリなどの情報記憶装置から構成される。

処理部2は、CPUなどのマイクロプロセッサおよびその周辺回路と、記憶部5のプログラム51とが協働することにより各種機能手段を構成する機能部である。

この機能手段としては、表現ヒストグラム計算手段21、カテゴリヒストグラム計算手段22、事後確率計算手段23およびテキスト抽出手段24とが設けられている。
40

【0019】

表現ヒストグラム計算手段21では、各カテゴリ10～1nごとに、当該カテゴリに含まれる各テキストで使用されている各表現の出現回数を計数する。この表現とは、各テキストを分解して得られる単語や文節などの単位表現であり、例えば自立語など特定の品詞に属する単語や、時制（現在形、過去形、未来形あるいは進行形など）の違いを無視して同一分類とした文節などで区分されたものである。

図3に表現の区分例を示す。ここでは自立語という品詞に基づきテキストから表現を抽出した例が示されており、「商品Aは使いにくい。遅い。」というテキスト30から、「商品A」、「使いにくい」、「遅い」という3つの表現31～33が抽出されている。このようなテキスト中にどのような表現が存在しているかを調べるには、形態素解析に関する
50

公知技術を応用できる。このようにして各テキストから抽出された表現が、同一表現ごとに計数される。

【0020】

カテゴリヒストグラム計算手段22では、各カテゴリ10～1nごとに、当該カテゴリに含まれているテキストすなわち要素の数を計数する。

事後確率計算手段23では、表現ヒストグラム計算手段21およびカテゴリヒストグラム計算手段22で得られた計数結果に基づき、分析対象のテキスト集合すなわちカテゴリ10の各テキストごとに、事後確率を計算する。

【0021】

一般に、確率モデルの考え方では、与えられたデータxは、ある確率変数Xの実現値と見なされる。特にこの確率変数の確率密度関数が有限次元のパラメータtを持つ固定された関数形 $f(x|t)$ を持つと仮定すると、その確率密度関数族 $F = \{f(x|t); t \in T\}$ をパラメトリック確率モデルという。

また、データxに基づきパラメータtの値を推測することを推定という。例えば、 $f(x|t)$ をtの関数（尤度関数）と見なし、これを最大にするtを推定値とする最尤推定法などが一般的である。

【0022】

一方、ベイズ統計（Bayesian Statistics）では、パラメータtも確率変数と見なし、その確率密度関数 $g(t)$ すなわち事前確率（ベイズ事前確率）を考える。そしてこの事前確率と尤度関数とからベイズの定理によって、tの事後確率（ベイズ事後確率） $g(t|x)$ が求められる。事後確率 $g(t|x)$ に関するtの平均をtの推定値とする方法をベイズ推定法といい、 $g(t|x)$ を最大にするtを推定する方法を、最大事後確率推定法（MAP推定法）という。

【0023】

事後確率計算手段23では、カテゴリ10の各テキストごとの事後確率を以下のようにして計算する。

今、カテゴリ10を c_0 、カテゴリ10に含まれる任意のテキストをS、そのテキストSで出現する表現を $w_1 \sim w_m$ とすると、そのテキストSに対する事後確率 $p(c_0|S)$ は、数1で求められる。

【0024】

【数1】

$$p(c_0|S) = \frac{p(c_0) \cdot \prod_{j=1}^m p(w_j|c_0)}{\sum_{i=0}^n \left(p(c_i) \cdot \prod_{j=1}^m p(w_j|c_i) \right)}$$

【0025】

ここで、カテゴリヒストグラム計算手段22での計数結果から、すべてのカテゴリに含まれるテキストの総数をN、カテゴリ c_0 に含まれるテキストの数を N_0 、カテゴリ c_i に含まれるテキストの数を N_i とすると、数1で用いられる、各カテゴリの出現確率を示す事前確率 $p(c_0)$ 、 $p(c_i)$ は、ラプラス型の推定量から数2により推定される。但し、nは比較対象テキスト集合の数を示す。

【0026】

【数2】

10

20

30

40

$$p(c_0) = \frac{N_0 + \frac{1}{2}}{N + \frac{n+1}{2}}, \quad p(c_i) = \frac{N_i + \frac{1}{2}}{N + \frac{n+1}{2}}$$

【0027】

また、表現ヒストグラム計算手段21での計数結果から、すべてのカテゴリに含まれる各テキストで出現した表現wの総数をK、カテゴリ c_0 に含まれる各テキストで出現した表現 w_j の出現回数を K_{0wj} 、カテゴリ c_i に含まれる各テキストで出現した表現 w_j の出現回数を K_{ijw} とすると、数1で用いられる、カテゴリ c_0 における表現 w_j の出現確率 $p(w_j | c_0)$ 、およびカテゴリ c_i における表現 w_j の出現確率 $p(w_j | c_i)$ は、ラプラス型の推定量から数3により推定される。但し、 K_0 は、カテゴリ C_0 における表現の総出現回数、 K_i はカテゴリ C_i における表現の総出現回数を示す。

10

【0028】

【数3】

$$p(w_j | c_0) = \frac{K_{0wj} + \frac{1}{2}}{K_0 + \frac{K}{2}}, \quad p(w_j | c_i) = \frac{K_{ijw} + \frac{1}{2}}{K_i + \frac{K}{2}}$$

20

【0029】

テキスト抽出手段24では、このようにして事後確率計算手段23により求められたカテゴリ10の各テキストに関する事後確率に基づき、事後確率の高い順あるいは低い順にカテゴリ10からテキストを抽出し、画面表示部7へ表示出力する。

【0030】

次に、図4を参照して、本実施の形態にかかる典型文分析装置の動作について説明する。図4は典型文分析装置の動作を示すフローチャートである。

操作入力部6によるオペレータの処理開始操作に応じて、まず、分析対象テキスト入力手段3では、分析対象となるテキスト集合を取り込んで蓄積部1へカテゴリ10として格納する（ステップ200）。これと前後して、比較対象テキスト集合群入力手段4では、比較対象となるテキスト集合群を取り込んで蓄積部1へカテゴリ11～1nとして格納する（ステップ201）。

30

【0031】

続いて、処理部2の表現ヒストグラム計算手段21では、各カテゴリ10～1nごとに、当該カテゴリに含まれる各テキストで使用されている各表現の出現回数を計数する（ステップ202）。これと前後して、カテゴリヒストグラム計算手段22では、各カテゴリ10～1nごとに、当該カテゴリに含まれているテキストすなわち要素の数を計数する（ステップ203）。

40

そして、事後確率計算手段23では、表現ヒストグラム計算手段21での計数結果に基づく各表現の出現確率 $p(w_j | c_0)$ 、 $p(w_j | c_i)$ の推定（ステップ204）、およびカテゴリヒストグラム計算手段22での計数結果に基づく事前確率 $p(c_0)$ 、 $p(c_i)$ の推定（ステップ205）を前後して実行し、これら推定結果を用いて、カテゴリ10の各テキストSに関する事後確率 $p(c_0 | S)$ を求める（ステップ206）。

【0032】

その後、必要に応じてテキスト抽出手段24では、カテゴリ10に含まれる各テキストのうち、それぞれの事後確率の高い順あるいは低い順にテキストを抽出して画面表示部7へ表示出力し（ステップ207）、一連の処理を終了する。

【0033】

50

分析対象となるカテゴリ 10 の各テキストごとに求めた事後確率は、「当該テキストと同じ表現を持つテキストが与えられた場合、そのテキストがカテゴリ 10 に属する確率」と解釈することができる。

したがって、分析対象の各テキストごとに事後確率を求め、その事後確率を各テキストが分析対象を代表する度合いすなわち典型度として用いることにより、従来のように各テキスト相互間で類似性がない場合でも適切に分析できるとともに、従来のように新たなテキストを生成するものではなく個々のテキストについて個別に典型度が得られることから、より詳細な分析結果が得られる。

【0034】

なお、以上の説明では、事後確率そのものを典型度として用いる場合について説明したが、事後確率の関数のうち計算処理がより簡単なものを典型度として用いても良い。例えば、事後確率の逆数から 1 を引いたものは、事後確率そのものに比べて計算処理が少なく済む。

また、上記事後確率やその関数で定義した典型度を用いて計算した 1 表現当たりの典型度などを、改めて修正典型度と定義しても良い。この場合、テキスト S の修正典型度は修正前の典型度をテキスト S における表現数でわり算したものになる。

【0035】

図 5 に本実施の形態にかかる典型文分析装置 100 による分析結果出力例を示す。ここでは、図 4 に示したように、商品 A に関する記述を示すテキストを分析対象であるカテゴリ 10 とし、他の商品 B, C, … に関する記述を示すテキストを比較対象であるカテゴリ 11 ~ 1n とし、分析を行い、典型度が高い順に 10 個のテキストを典型文として抽出した。なお、典型度の値としては、事後確率の逆数から 1 を引いたものの自然対数値を、当該テキストの表現数で割ったものが用いられている。

【0036】

なお、以上の説明では、比較対象テキスト集合群すなわちカテゴリ 11 ~ 1n を用いる場合を例として説明したが、こゝでテキスト集合が空集合であっても、上記と同様に問題なく各テキストの典型度を求めることができる。比較対象として空集合しか入力しない場合、比較対象テキスト集合群入力手段 4 は不要となる。

【0037】

【発明の効果】

以上説明したように、本発明は、所定の分析対象に関する複数のテキストを要素とする分析対象集合と、分析対象と比較する各比較対象に関する複数のテキストをそれぞれ要素とする複数の比較対象集合からなる比較対象群とから、分析対象集合に含まれる各テキストがその分析対象集合を代表する度合いとして、分析対象集合に含まれる各テキストごとに、分析対象集合に対する当該テキストのベイズ事後確率を計算するようにしたので、従来のように各テキスト相互間で類似性がない場合でも適切に分析できるとともに、従来のように新たなテキストを生成するものではなく個々のテキストについて個別に典型度が得られることから、より詳細な分析結果が得られる。

【図面の簡単な説明】

【図 1】 本発明の一実施の形態にかかる典型文分析装置の構成を示すブロック図である。

【図 2】 各カテゴリの構成例である。

【図 3】 表現の区分例である。

【図 4】 典型文分析装置の動作を示すフローチャートである。

【図 5】 分析結果出力例である。

【符号の説明】

1…蓄積部、10~1n カテゴリ、2…処理部、21…表現ヒストグラム計算手段、22…カテゴリヒストグラム計算手段、23…事後確率計算手段、24…テキスト抽出手段、3…分析対象テキスト入力手段、4…比較対象テキスト集合群入力手段、5…記憶部、6…操作入力部、7…画面表示部、9…記録媒体。

10

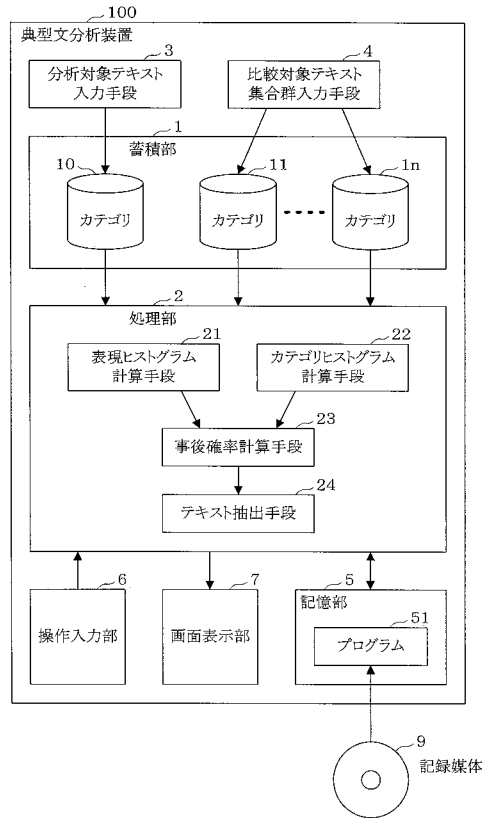
20

30

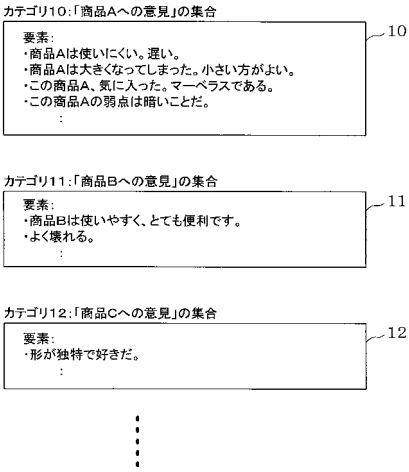
40

50

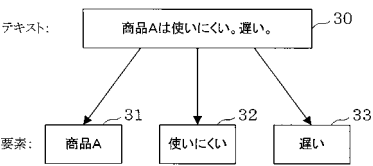
【図 1】



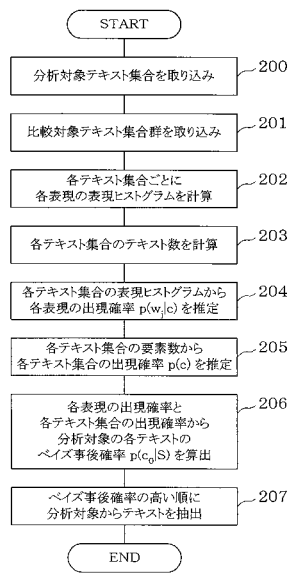
【図 2】



【図 3】



【図 4】



【図 5】

順位	典型文	典型度
1	テキストだけなら、モノクロの商品A330も良い。	1.798008
2	商品A役に立つようにがんばってみます。	1.665427
3	商品Aいいですよ。	1.645893
4	夢さまが惚れ込んでいる、商品A II あたりいかかで	1.644226
5	入力は商品Aですか、いいですね。	1.605591
6	商品Aの方が、ぐぐぐ、つよいのはわかっていても、どーも。	1.605239
7	商品Aは最高です。	1.588651
8	商品Aあたり良いかな？	1.523441
9	また、外出先で電子メールのやりとりをしたいのなら、商品Aがお薦めです。	1.523019
10	面倒なので商品Aで世間とつながる。	1.520433

フロントページの続き

(56)参考文献 H.Li and K.Yamanishi, Mining from Open Answers in Questionare Data, Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Minin, 2001年, p. 443-p. 449

森永聡・山西健司, テキストマイニングによる自由記述アンケート分析, 計測と制御, 2002年 5月10日, Vol.41, No. 5, p. 354-p. 357

(58)調査した分野(Int.Cl., DB名)

G06F 17/21-17/30