

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2018-185744
(P2018-185744A)

(43) 公開日 平成30年11月22日(2018.11.22)

(51) Int.Cl.	F I	テーマコード (参考)
G06F 17/30 (2006.01)	G06F 17/30 210D	5L049
G06Q 30/02 (2012.01)	G06F 17/30 220Z	
	G06Q 30/02 300	
	G06F 17/30 170A	

審査請求 未請求 請求項の数 10 O L (全 39 頁)

(21) 出願番号 特願2017-88575 (P2017-88575)
(22) 出願日 平成29年4月27日 (2017. 4. 27)

(71) 出願人 504133110
国立大学法人電気通信大学
東京都調布市調布ヶ丘一丁目5番地1
(74) 代理人 100107766
弁理士 伊東 忠重
(74) 代理人 100070150
弁理士 伊東 忠彦
(72) 発明者 沼尾 雅之
東京都調布市調布ヶ丘一丁目5番地1 国立大学法人電気通信大学内
(72) 発明者 金兵 裕太
東京都調布市調布ヶ丘一丁目5番地1 国立大学法人電気通信大学内
Fターム(参考) 5L049 BB02

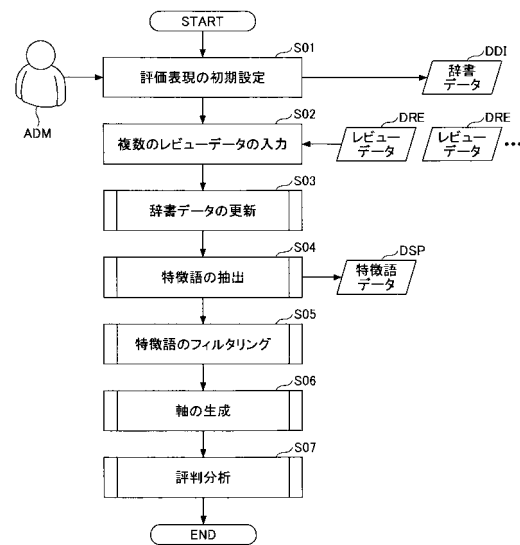
(54) 【発明の名称】 情報処理装置、情報処理システム、情報処理方法及びプログラム

(57) 【要約】

【課題】 情報処理装置が商品等の評判を示す軸を生成することを目的とする。

【解決手段】 情報処理装置は、所定の商品又はサービスに対する複数のレビューデータを入力する入力部と、前記複数のレビューデータに含まれ、前記商品又は前記サービスに対する意見、評判、感想又は評価を示す評価表現を抽出する評価表現抽出部と、前記評価表現に基づいて、前記複数のレビューデータに含まれ、前記商品又は前記サービスの特徴を示す特徴語を抽出する特徴語抽出部と、前記特徴語をクラスタに分け、前記クラスタごとに軸を生成するクラスタリング部とを含む。

【選択図】 図3



【特許請求の範囲】**【請求項 1】**

所定の商品又はサービスに対する複数のレビューデータを入力する入力部と、
前記複数のレビューデータに含まれ、前記商品又は前記サービスに対する意見、評判、感想又は評価を示す評価表現を抽出する評価表現抽出部と、
前記評価表現に基づいて、前記複数のレビューデータに含まれ、前記商品又は前記サービスの特徴を示す特徴語を抽出する特徴語抽出部と、
前記特徴語をクラスタに分け、前記クラスタごとに軸を生成するクラスタリング部とを含む情報処理装置。

【請求項 2】

前記評価表現には、
前記商品又は前記サービスに対して、肯定的な表現となる肯定語と、
前記商品又は前記サービスに対して、否定的な表現となる否定語と
が含まれる請求項 1 に記載の情報処理装置。

【請求項 3】

前記レビューデータから前記評価表現が抽出される抽出回数及び前記評価表現と係り受けの関係にある他の評価表現のスコアに基づいて、前記評価表現のスコアを計算する請求項 1 又は 2 に記載の情報処理装置。

【請求項 4】

前記スコアは、3 段階以上を表現できる値である請求項 3 に記載の情報処理装置。

【請求項 5】

前記クラスタリング部は、Gap 統計量に基づいて、前記クラスタの数を特定する請求項 1 乃至 4 のいずれか 1 項に記載の情報処理装置。

【請求項 6】

k - m e d o i d s 法に基づいて、複数の前記特徴語から、前記軸の軸名を決定する請求項 5 に記載の情報処理装置。

【請求項 7】

前記複数のレビューデータから前記軸ごとに、前記商品又は前記サービスの特性についての評判を示す評判値を計算する評判分析部を更に含む請求項 1 乃至 6 のいずれか 1 項に記載の情報処理装置。

【請求項 8】

1 台以上の情報処理装置を有する情報処理システムであって、
所定の商品又はサービスに対する複数のレビューデータを入力する入力部と、
前記複数のレビューデータに含まれ、前記商品又は前記サービスに対する意見、評判、感想又は評価を示す評価表現を抽出する評価表現抽出部と、
前記評価表現に基づいて、前記複数のレビューデータに含まれ、前記商品又は前記サービスの特徴を示す特徴語を抽出する特徴語抽出部と、
前記特徴語をクラスタに分け、前記クラスタごとに軸を生成するクラスタリング部とを含む情報処理システム。

【請求項 9】

情報処理装置が行う情報処理方法であって、
前記情報処理装置が、所定の商品又はサービスに対する複数のレビューデータを入力する入力手順と、
前記情報処理装置が、前記複数のレビューデータに含まれ、前記商品又は前記サービスに対する意見、評判、感想又は評価を示す評価表現を抽出する評価表現抽出手順と、
前記情報処理装置が、前記評価表現に基づいて、前記複数のレビューデータに含まれ、前記商品又は前記サービスの特徴を示す特徴語を抽出する特徴語抽出手順と、
前記情報処理装置が、前記特徴語をクラスタに分け、前記クラスタごとに軸を生成するクラスタリング手順と
を含む情報処理方法。

10

20

30

40

50

【請求項10】

コンピュータに情報処理方法を実行させるためのプログラムであって、
前記コンピュータが、所定の商品又はサービスに対する複数のレビューデータを入力する入力手順と、

前記コンピュータが、前記複数のレビューデータに含まれ、前記商品又は前記サービスに対する意見、評判、感想又は評価を示す評価表現を抽出する評価表現抽出手順と、

前記コンピュータが、前記評価表現に基づいて、前記複数のレビューデータに含まれ、前記商品又は前記サービスの特徴を示す特徴語を抽出する特徴語抽出手順と、

前記コンピュータが、前記特徴語をクラスタに分け、前記クラスタごとに軸を生成するクラスタリング手順と

を実行させるためのプログラム。

【発明の詳細な説明】**【技術分野】****【0001】**

本発明は、情報処理装置、情報処理システム、情報処理方法及びプログラムに関する。

【背景技術】**【0002】**

いわゆる口コミサイト、ブログ又はレビューサイト等において、商品又はサービス等（以下単に「商品等」という。）について、購入者等によるレビュー（review）が投稿される。そして、これらのレビューを解析する技術が知られている。

【0003】

例えば、非特許文献1に開示される技術は、オンラインショッピング等で使用される通販サイトで投稿される多数のレビューから、属性及び意見のペアを抽出する。具体的には、まず、機械学習によって、レビューから、属性を含む文節が抽出される。次に、機械学習によって、レビューから、意見を含む文節が抽出される。そして、抽出された属性及び意見の「係り受け」の関係に基づいて、属性及び意見のペアが抽出される。このようにして、効率良く商品等の評判を把握する方法が知られている。

【0004】

他にも、非特許文献2に開示される技術は、弱教師付きの手法によって、「係り受け」の関係と、相互情報量（Pointwise Mutual Information）とに基づいて、名詞又は名詞句のカテゴリ分類を行う。このようにして、ユーザがパラメータを設定しなくとも、高精度にレビューを分類することができる方法が知られている。

【先行技術文献】**【非特許文献】****【0005】**

【非特許文献1】"機械学習による商品レビューの属性 - 意見ペアの抽出", 中野裕介, 湯本高行, 新居学, 上浦尚武著, 情報処理学会研究報告 IPSJ SIG Technical Report, Vol. 2015-DBS-162 No. 14 2015/11/26

【非特許文献2】"新たな弱教師付き型分類手法 Bautext", ゲンファムタタオ, 岡部誠, 尾内理紀夫, 林貴宏, 西岡悠平, 竹中孝真, 森正弥, 情報処理学会論文誌 Vol. 52 No. 1 269-283 Jan. 2011

【発明の概要】**【発明が解決しようとする課題】****【0006】**

しかしながら、従来の技術では、複数のレビューを点数で評価するには、商品等の種類ごとに、商品等の特性を考慮して、管理者等があらかじめ軸を設定しなければならない問題がある。

【0007】

本発明は、上記問題点に鑑み提案されたものであり、その目的とするところは、情報処

10

20

30

40

50

理装置が商品等の評判を示す軸を生成できることにある。

【課題を解決するための手段】

【0008】

上記の課題を解決するため、本発明に係る一実施形態にあっては、
 情報処理装置は、
 所定の商品又はサービスに対する複数のレビューデータを入力する入力部と、
 前記複数のレビューデータに含まれ、前記商品又は前記サービスに対する意見、評判、
 感想又は評価を示す評価表現を抽出する評価表現抽出部と、
 前記評価表現に基づいて、前記複数のレビューデータに含まれ、前記商品又は前記サー
 ビスの特徴を示す特徴語を抽出する特徴語抽出部と、
 前記特徴語をクラスタに分け、前記クラスタごとに軸を生成するクラスタリング部と
 を含む。

10

【発明の効果】

【0009】

本発明にあっては、情報処理装置が商品等の評判を示す軸を生成できる。

【図面の簡単な説明】

【0010】

【図1】本発明の一実施形態に係る情報処理装置を使用したシステムの全体構成例を示す概略図である。

【図2】本発明の一実施形態に係る情報処理装置のハードウェア構成例を示すブロック図である。

20

【図3】本発明の一実施形態に係る情報処理装置による全体処理例を示すフローチャートである。

【図4】本発明の一実施形態に係る情報処理装置が表示する初期設定用の画面例を示す図である。

【図5】本発明の一実施形態に係る情報処理装置による辞書データの更新例を示すフローチャートである。

【図6】本発明の一実施形態に係る情報処理装置による「係り受け」解析の一例を示す図である。

【図7】本発明の一実施形態に係る情報処理装置による登録語のスコアの計算例を示す図である。

30

【図8】本発明の一実施形態に係る情報処理装置による登録語のフィルタリング例を示すフローチャートである。

【図9】本発明の一実施形態に係る情報処理装置による特徴語の抽出例を示すフローチャートである。

【図10】本発明の一実施形態に係る情報処理装置による特徴語のフィルタリング例を示すフローチャートである。

【図11】本発明の一実施形態に係る情報処理装置による第5閾値の計算例を示す図である。

【図12】本発明の一実施形態に係る情報処理装置による軸の生成例を示すフローチャートである。

40

【図13】本発明の一実施形態に係る特徴語と評価表現の「係り受け」を示す行列の一例を示す図である。

【図14】本発明の一実施形態に係る情報処理装置によるTF-IDF計算の計算結果例を示す図である。

【図15】本発明の一実施形態に係る情報処理装置によるk-medoids法による代表点の特定例を示す図である。

【図16】本発明の一実施形態に係る情報処理装置による軸の生成例を示す図である。

【図17】本発明の一実施形態に係る情報処理装置による評判分析例を示すフローチャートである。

50

- 【図 18】本発明の一実施形態に係る情報処理装置による評判分析例を示す図である。
- 【図 19】本発明の一実施形態に係る情報処理装置が表示するレーダーチャートの一例を示す図である。
- 【図 20】本発明の一実施形態に係る情報処理装置が全体処理の対象としたレビューの条件を示す表である。
- 【図 21】本発明の一実施形態に係る情報処理装置が全体処理の対象とした商品及びそれぞれのレビュー数を示す表である。
- 【図 22】本発明の一実施形態に係る情報処理装置が全体処理に用いた設定値を示す表である。
- 【図 23】本発明の一実施形態に係る情報処理装置が全体処理において生成した辞書データを示す表である。 10
- 【図 24】本発明の一実施形態に係る情報処理装置が全体処理において「掃除機」のレビューデータから抽出した評価表現を示す表である。
- 【図 25】本発明の一実施形態に係る情報処理装置が全体処理において「スーツ」のレビューデータから抽出した評価表現を示す表である。
- 【図 26】本発明の一実施形態に係る情報処理装置が全体処理において「和菓子」のレビューデータから抽出した評価表現を示す表である。
- 【図 27】本発明の一実施形態に係る情報処理装置による全体処理において出現した特徴語の分布を示す図である。
- 【図 28】本発明の一実施形態に係る情報処理装置が全体処理において抽出した特徴語を示す表である。 20
- 【図 29】本発明の一実施形態に係る情報処理装置が全体処理において「掃除機」のレビューデータから生成した軸を示す表である。
- 【図 30】本発明の一実施形態に係る情報処理装置が全体処理において「スーツ」のレビューデータから生成した軸を示す表である。
- 【図 31】本発明の一実施形態に係る情報処理装置が全体処理において「和菓子」のレビューデータから生成した軸を示す表である。
- 【図 32】本発明の一実施形態に係る情報処理装置が全体処理において「掃除機」のレビューデータから抽出及び分配した準特徴語を示す表である。
- 【図 33】本発明の一実施形態に係る情報処理装置が全体処理において「スーツ」のレビューデータから抽出及び分配した準特徴語を示す表である。 30
- 【図 34】本発明の一実施形態に係る情報処理装置が全体処理において「和菓子」のレビューデータから抽出及び分配した準特徴語を示す表である。
- 【図 35】本発明の一実施形態に係る情報処理装置が全体処理において「掃除機」のレビューデータから生成した軸及び代表評価表現を示す表である。
- 【図 36】本発明の一実施形態に係る情報処理装置が全体処理において「スーツ」のレビューデータから生成した軸及び代表評価表現を示す表である。
- 【図 37】本発明の一実施形態に係る情報処理装置が全体処理において「和菓子」のレビューデータから生成した軸及び代表評価表現を示す表である。
- 【図 38】本発明の一実施形態に係る情報処理装置が全体処理において「掃除機」のレビューデータから生成したレーダーチャートを示す図である。 40
- 【図 39】本発明の一実施形態に係る情報処理装置が全体処理において「スーツ」のレビューデータから生成したレーダーチャートを示す図である。
- 【図 40】本発明の一実施形態に係る情報処理装置が全体処理において「和菓子」のレビューデータから生成したレーダーチャートを示す図である。
- 【図 41】本発明の一実施形態に係る情報処理装置の機能構成例を示す機能ブロック図である。
- 【発明を実施するための形態】
- 【0011】
- 以下、本発明の好適な実施形態を例に説明する。 50

【0012】

<全体構成例>

図1は、本発明の一実施形態に係る情報処理装置を使用したシステムの全体構成例を示す概略図である。例えば、情報処理装置の例であるサーバSERは、図示するように、インターネット等のネットワークNTWに接続される。そして、ネットワークNTWを介してサーバSERにアクセスすると、様々なユーザが、商品等についてレビューを投稿できるように、管理者ADMは、サーバSERを設定する。すなわち、管理者ADMは、いわゆるレビューサイトをサーバSER上に開設し、ネットワークNTW上で開放する。このようにすると、例えば、ユーザUS1は、端末PC1に操作を入力して、レビューRE1を投稿できる。同様に、ユーザUS2は、端末PC2に操作を入力して、レビューRE2

10

【0013】

図示するように、レビューは、所定の商品等についての意見、評判、感想又は評価を含む文である。具体的には、図示する例では、ユーザUS1は、「商品A」という製品名の商品を購入し、使用后、「商品A」を使用した感想等を他人に公開するため、レビューRE1をサーバSERに投稿する。

【0014】

したがって、レビューRE1には、「商品A」の良かった点又は悪かった点等が表現された文が、主にユーザUS1による体験に基づいて、記載される。このようなレビューが、サーバSERには、レビューデータDREとなって蓄積される。例えば、レビューデータDREは、テキストデータ等である。なお、レビューデータDREには、ユーザ名又は書き込み日時等のデータが含まれてもよい。

20

【0015】

また、各レビューは、レビューの対象となった商品等と対応付けされる。具体的には、商品等には、各商品等を識別できるID(Identification)等があらかじめ付与される。そして、各レビューは、レビューデータDREに、IDと一緒に記憶される。したがって、各レビューがどの商品等を対象にした文であるかは、レビューデータDREに含まれるIDを参照すると、特定することができる。

【0016】

以下、図示するようにサーバSERに蓄積された複数のレビューデータに基づいて、所定の商品について、評判分析を行う場合を例に説明する。

30

【0017】

<情報処理装置のハードウェア構成例>

図2は、本発明の一実施形態に係る情報処理装置のハードウェア構成例を示すブロック図である。図示するように、サーバSERは、例えば、CPU(Central Processing Unit)H01と、記憶装置H02と、入力装置H03と、出力装置H04と、ネットワークインタフェースH05とを有するハードウェア構成である。これらのハードウェア資源は、バス(bus)等によって相互に接続される。また、サーバSER以外の情報処理装置も、例えば、図示するようなハードウェア構成である。以下、各情報処理装置がすべて図示するハードウェア構成である例で説明するが、情報処理装置のハードウェア構成は、図示するハードウェア構成に限られない。

40

【0018】

CPUH01は、情報処理装置が行う処理を実行するための演算を行う演算装置及び情報処理装置が有するハードウェア資源を制御する制御装置の例である。

【0019】

記憶装置H02は、例えば、メモリ等の主記憶装置である。なお、記憶装置H02は、更にハードディスク等の補助記憶装置を有してもよい。

【0020】

入力装置H03は、ユーザによる操作又は外部装置からのコマンド等を入力する装置である。例えば、入力装置H03は、キーボード及びコネクタ等である。

50

【0021】

出力装置H04は、ユーザ又は外部装置に対して処理結果等を入力する装置である。例えば、出力装置H04は、ディスプレイ及びコネクタ等である。

【0022】

ネットワークインタフェースH05は、ネットワークを介して、有線、無線又はこれらの組み合わせによって、外部装置とデータを入出力するインタフェースである。例えば、ネットワークインタフェースH05は、コネクタ及び処理IC(Integrated Circuit)等である。

【0023】

<全体処理例>

図3は、本発明の一実施形態に係る情報処理装置による全体処理例を示すフローチャートである。例えば、図1に示すように、レビューが投稿され、複数のレビューデータが集まった後、サーバSERは、図示するような全体処理を行って、所定の商品又はサービスについての評判を分析する。

【0024】

<評価表現の初期設定例>

ステップS01では、サーバSERは、評価表現を初期設定する。

【0025】

評価表現は、商品等に対する意見、評判、感想又は評価等を示す言葉である。例えば、「良い」又は「満足」等は、商品等に対して肯定的な意見等を示す評価表現(以下「肯定語」という。)の例である。一方で、「つまらない」又は「不満」等は、商品等に対して否定的な意見等を示す評価表現(以下「否定語」という。)の例である。また、評価表現は、肯定語と、否定語とに分けて記憶されるのが望ましい。さらに、各評価表現には、値(以下「スコア」という。)が登録される。

【0026】

例えば、初期設定は、管理者ADM等の操作に基づいて行われる。具体的には、初期設定は、例えば、以下のような画面で行われる。

【0027】

図4は、本発明の一実施形態に係る情報処理装置が表示する初期設定用の画面例を示す図である。この例では、評価表現は、「Positive」の欄に肯定語が設定され、「Negative」の欄に否定語が設定される例である。なお、設定された評価表現は、辞書データDDIに記憶される。

【0028】

図示する例は、肯定語及び否定語に1語ずつ設定する例である。具体的には、図示する例では、「満足」という言葉が、肯定語に設定される例である。一方で、図示する例では、「不満」という言葉が、否定語に設定される例である。このように初期設定されると、辞書データDDIに各言葉が登録される。そのため、以後、サーバSERは、「満足」という言葉を肯定語と認識し、さらに、「不満」という言葉を否定語と認識できる。

【0029】

なお、初期設定は、図示するような設定に限られず、例えば、複数の言葉を設定してもよい。

【0030】

<複数のレビューデータの入力例>

ステップS02では、サーバSERは、複数のレビューデータDREを入力する。すなわち、図1に示すように、サーバSERには、各商品等についてのレビューが、レビューデータDREとなってあらかじめ蓄積されている。そこで、サーバSERは、複数のレビューデータDREを読み出し、各レビューを示す文を読み込む。

【0031】

<辞書データの更新例>

ステップS03では、サーバSERは、辞書データを更新する。例えば、ステップS0

10

20

30

40

50

3 は、以下のような処理である。

【0032】

図5は、本発明の一実施形態に係る情報処理装置による辞書データの更新例を示すフローチャートである。

【0033】

以下、図示するような辞書データDDI1があらかじめステップS01によって設定される例で説明する。具体的には、辞書データDDI1には、肯定語に、「満足」という言葉が1語登録され、かつ、スコアが「+0.90」と設定される例である。さらに、辞書データDDI1には、否定語に、「不満」という言葉が1語登録され、かつ、スコアが「-0.95」と設定される例である。

10

【0034】

以下、スコアが正の値であると、肯定語であり、一方で、スコアが負の値であると、否定語であるとする。また、以下の例では、スコアは、「+1.00」乃至「-1.00」の範囲であるとする。なお、スコアは、すべての肯定語を「+1」とし、かつ、すべての否定語を「-1」とする、いわゆる2値でないのが望ましい。すなわち、スコアは、この例のように、3段階以上を表現できる値（以下「連続値」という。）であるのが望ましい。

【0035】

< レビューデータから1文を抽出する例 >

ステップS0301では、サーバSERは、レビューデータDREから1文を抽出する。以下、図示するように、「商品Aは、値段が安くて満足です。」という1文がレビューデータDREから抽出された場合を例に説明する。

20

【0036】

< 文中から評価表現を検索する例 >

ステップS0302では、サーバSERは、文中から評価表現を検索する。すなわち、サーバSERは、ステップS0301によって抽出された文に、辞書データDDI1に登録されている評価表現と同じ言葉が含まれているかを検索する。この例では、ステップS0301によって抽出された文には、辞書データDDI1に肯定語として登録されている「満足」という言葉が含まれている。そのため、図示するように、サーバSERは、「満足」という評価表現を抽出した検索結果を出力する。

30

【0037】

< 検索された評価表現と「係り受け」の関係にある言葉の抽出例 >

ステップS0303では、サーバSERは、検索された評価表現と「係り受け」の関係にある言葉を抽出する。すなわち、サーバSERは、ステップS0301によって抽出された文において、ステップS0302で検索される評価表現と「係り受け」の関係にある言葉を抽出する。例えば、以下のような「係り受け」解析によって、「係り受け」の関係にある言葉が抽出される。

【0038】

図6は、本発明の一実施形態に係る情報処理装置による「係り受け」解析の一例を示す図である。例えば、「商品Aは、値段が安くて満足です。」の文が解析される対象である例で説明する。

40

【0039】

まず、「係り受け」解析では、図6(A)に示すように、サーバSERは、レビューデータDREから抽出された文を文節で分解する。次に、図6(B)に示すように、サーバSERは、構文を解析して、各文節に「係り受け」の関係があるか否かを判断する。この例では、図6(B)に示すように、「値段が」の文節と、「安くて」の文節とが「係り受け」の関係であると判断され、同様に、「安くて」の文節と、「満足です。」の文節とが「係り受け」の関係であると判断される。したがって、この例では、「安い」という言葉が、「満足」という評価表現に対して順接関係となる。このように、サーバSERは、肯定語と順接関係となる言葉を肯定語と推定し、抽出する。

50

【 0 0 4 0 】

以下、ステップ S 0 3 0 3 によって抽出され、辞書データ D D I に登録される候補となる言葉を「登録語」という。つまり、この例では、「安い」という言葉が、登録語となる。

【 0 0 4 1 】

< 抽出された回数のカウント例 >

ステップ S 0 3 0 4 では、サーバ S E R は、抽出された回数をカウントする。すなわち、サーバ S E R は、登録語が文から抽出されると、「1回」とカウントし、抽出回数 D C N を記憶する。つまり、他の文でも、評価表現と「係り受け」の関係となる場合が多い登録語は、抽出回数 D C N が大きい値となる。

10

【 0 0 4 2 】

以上のようなステップ S 0 3 0 1 乃至ステップ S 0 3 0 4 が所定回数繰り返し行われる。なお、所定回数となるループ回数及びループ条件等は、あらかじめ設定される値である。

【 0 0 4 3 】

< 登録語のスコアの計算例 >

ステップ S 0 3 0 5 では、サーバ S E R は、登録語のスコアを計算する。例えば、スコアは、下記 (1) 式等によって計算される。

【 0 0 4 4 】

【 数 1 】

20

$$score(x) = \frac{\sum extract(x)}{\sum |extract(x)|}$$

$score(x)$: 登録語 x のスコア

$extract(x)$: 登録語 x の抽出値

(1)

30

例えば、登録語のそれぞれの抽出値 (上記 (1) 式における「 $extract(x)$ 」である。) が「+ 0 . 8」、「+ 0 . 9」及び「+ 1 . 0」であるとする、上記 (1) 式は、分子が「(+ 0 . 8) + (+ 0 . 9) + (+ 1 . 0)」となり、分母では、絶対値の総和である「(+ 0 . 8) + (+ 0 . 9) + (+ 1 . 0)」となる。したがって、この例では、スコアは、上記 (1) 式に基づいて、「1 . 0」と計算される。

【 0 0 4 5 】

40

また、上記 (1) 式における登録語の抽出値 (上記 (1) 式では、「 $extract(x)$ 」である。) は、ステップ S 0 3 0 3 によって抽出されるごとに、下記 (2) 式によって算出される値である。

【 0 0 4 6 】

【数 2】

$$\text{extract}(e, x) = \text{value}_e \cdot \text{conjunction}(e, x) \cdot \text{reverse}(e) \cdot \text{revers}(x)$$

value_e : 評価表現 e のスコア

(辞書データに登録されているスコア)

$\text{conjunction}(e, x)$

$$= \begin{cases} +1: \text{評価表現 } e \text{ と、登録語 } x \text{ とが順接関係である。} \\ -1: \text{評価表現 } e \text{ と、登録語 } x \text{ とが逆接関係である。} \end{cases}$$

$\text{reverse}(A)$

$$= \begin{cases} +1: A \text{ の評価極性が文中で反転していない。} \\ -1: A \text{ の評価極性が文中で反転している。} \end{cases}$$

(2)

10

20

したがって、ステップ S 0 3 0 3 によって抽出されると、上記 (2) 式に基づいて、サーバ S E R は、抽出値を計算する。上記 (2) 式に示すように、抽出値は、「係り受け」の関係にある評価表現 (上記 (2) 式では、「 e 」である。) のスコア等に基づいて定まる。

【0047】

また、上記 (2) 式において、「評価極性の反転」は、「・・・ない」等の言葉 (以下「反転子」という。) が文中において評価表現に続けて使われることにより、意味が反転することをいう。具体的には、「満足」という言葉は、肯定語であり、反転子がなければ、肯定的な意味を示す。一方で、「満足できない」等のように、反転が続くと、この文の全体の意味は、「満足」を反転させた否定的な意味を示す。

30

【0048】

さらに、反転子は、複数回用いられる場合がある。例えば、反転子が複数回用いられた文は、「満足できなくはない」等である。このような場合には、サーバ S E R は、反転子が用いられた回数を数える。奇数回反転子が用いられると、評価極性は、反転する。一方で、偶数回反転子が用いられると、評価極性は、反転しない。このように、反転子が用いられた回数に基づいて、上記 (2) 式では、「 $\text{reverse}(A)$ 」の関数に対して、「 -1 」又は「 $+1$ 」の値が返される。

40

【0049】

同様に、逆接の接続詞等が文に用いられると、反転子が使われるのと同様に、評価極性が反転する場合がある。そこで、上記 (2) 式では、評価表現と、登録語との関係が、順接関係にあるか逆接関係にあるかを判断する。順接関係にあれば、評価極性は、反転しない。一方で、逆接関係にあると、評価極性は、反転する。このように、順接関係にあるか逆接関係にあるかに基づいて、上記 (2) 式では、「 $\text{conjunction}(e, x)$ 」の関数に対して、「 -1 」又は「 $+1$ 」の値が返される。

【0050】

以上のようなステップ S 0 3 0 5 が行われると、例えば、以下のような結果となる。

【0051】

50

図7は、本発明の一実施形態に係る情報処理装置による登録語のスコアの計算例を示す図である。以下、サーバSERが図5に示す辞書データDDI1を使用してスコアを計算する例で説明する。また、以下、図7(A)に示すような文を例に説明する。

【0052】

図7(A)に示す文では、辞書データDDI1に登録されている「満足」という評価表現が検索される(ステップS0302)。そして、「係り受け」解析によって、「甘い」という登録語が抽出される(ステップS0303)。

【0053】

この例では、登録語と「係り受け」の関係にある「満足」という評価表現のスコアが「+0.90」であるため、上記(2)式における「value」は、「+0.90」となる。そして、登録語は、評価表現と順接関係であり、文中に反転子がないため、抽出値は、「+0.90」となる。さらに、「甘い」という登録語が抽出されるのが初めてであれば、抽出回数は、「1」であるため、「甘い」のスコアは、この時点では、「+0.90」と計算される。

10

【0054】

次に、図7(B)に示す文が抽出されたとする(ステップS0301)。図7(B)に示す文では、辞書データDDI1に登録されている「不満」という評価表現が検索される(ステップS0302)。そして、図7(A)と同様に、「係り受け」解析によって、「甘い」という登録語が抽出される(ステップS0303)。

【0055】

この例では、登録語と「係り受け」の関係にある「不満」という評価表現のスコアが「-0.95」であるため、上記(2)式における「value」は、「-0.95」となる。そして、登録語は、評価表現と順接関係であり、文中に反転子がないため、抽出値は、「-0.95」となる。

20

【0056】

なお、登録語は、以前の処理によって既に登録されている場合がある。このような場合には、スコアは、更新されてもよい。例えば、スコアは、下記(3)式で計算された値に更新される。

【0057】

【数3】

30

$$value'_x = score(x) \times \alpha + value_x \times (1 - \alpha)$$

$value'_x$: (更新後の)登録語 x のスコア

$score(x)$: 登録語 x の抽出値

$value_x$: (更新前の)登録語 x のスコア

40

(3)

なお、上記(3)式において、「 α 」は、あらかじめ設定される値であって、「 $0 < \alpha < 1$ 」の範囲で、設定される値である。

【0058】

<登録語のフィルタリング例>

ステップS0306では、サーバSERは、登録語をフィルタリングするのが望ましい。例えば、ステップS0306では、以下のような処理が行われる。

50

【 0 0 5 9 】

図 8 は、本発明の一実施形態に係る情報処理装置による登録語のフィルタリング例を示すフローチャートである。

【 0 0 6 0 】

< 新しい登録語があるか否かの判断例 >

ステップ S 3 6 0 1 では、サーバ S E R は、新しい登録語があるか否かを判断する。すなわち、図 5 に示す処理によって、以前には辞書データ D D I に登録されていなかった登録語が抽出されると、サーバ S E R は、新しい登録語があると判断する。

【 0 0 6 1 】

次に、新しい登録語があるとサーバ S E R が判断すると（ステップ S 3 6 0 1 で Y E S ）、サーバ S E R は、ステップ S 3 6 0 2 に進む。一方で、新しい登録語がないとサーバ S E R が判断すると（ステップ S 3 6 0 1 で N O ）、サーバ S E R は、登録語をフィルタリングする処理を終了する。

10

【 0 0 6 2 】

< 登録語の抽出回数の読み出し例 >

ステップ S 3 6 0 2 では、サーバ S E R は、登録語の抽出回数を読み出す。ステップ S 0 3 0 4 では、ステップ S 0 3 0 3 で抽出されるごとに、抽出回数 D C N がカウントされる。そこで、ステップ S 3 6 0 2 では、サーバ S E R は、ステップ S 0 3 0 4 でカウントされた抽出回数 D C N を読み出す。

【 0 0 6 3 】

< 登録語のスコアの読み出し例 >

ステップ S 3 6 0 3 では、サーバ S E R は、登録語のスコアを読み出す。ステップ S 0 3 0 5 では、ステップ S 0 3 0 3 で抽出されるごとに、スコアが計算される。そこで、ステップ S 3 6 0 3 では、サーバ S E R は、ステップ S 0 3 0 5 で計算されたスコアを読み出す。

20

【 0 0 6 4 】

< 抽出回数が第 1 閾値より大きい値であり、かつ、スコアが第 2 閾値より大きい値であるか否かの判断例 >

ステップ S 3 6 0 4 では、サーバ S E R は、抽出回数が第 1 閾値より大きい値であり、かつ、スコアが第 2 閾値より大きい値であるか否かを判断する。

30

【 0 0 6 5 】

まず、サーバ S E R は、ステップ S 3 6 0 2 で読み出された抽出回数と、第 1 閾値とを比較する。なお、第 1 閾値は、あらかじめ設定される値である。例えば、第 1 閾値に「 1 0 」が設定されると、抽出回数が「 1 0 」回より大きい値である、すなわち、出現する回数が多い登録語が、抽出回数が第 1 閾値より大きい値であると判断される。

【 0 0 6 6 】

次に、サーバ S E R は、ステップ S 3 6 0 3 で読み出されたスコアと、第 2 閾値とを比較する。なお、第 2 閾値は、あらかじめ設定される値である。例えば、第 2 閾値に「 0 . 8 5 」が設定されると、スコアが「 0 . 8 5 」より大きい値である登録語が、スコアが第 2 閾値より大きい値であると判断される。

40

【 0 0 6 7 】

そして、抽出回数が第 1 閾値より大きい値であり、かつ、スコアが第 2 閾値より大きい値であると（ステップ S 3 6 0 4 で Y E S ）、サーバ S E R は、登録語をフィルタリングする処理を終了する。一方で、抽出回数が第 1 閾値より大きい値でない、又は、スコアが第 2 閾値より大きい値でないと（ステップ S 3 6 0 4 で N O ）、サーバ S E R は、ステップ S 3 6 0 5 に進む。

【 0 0 6 8 】

< 登録語の削除例 >

ステップ S 3 6 0 5 では、サーバ S E R は、登録語を削除する。すなわち、ステップ S 3 6 0 4 によって、スコアが小さい値である、又は、抽出回数が小さい値である登録語は

50

、辞書データDDIから削除される。

【0069】

辞書データDDIには、人間による判断からしてみると、適切でない評価表現、いわゆるノイズとなる言葉が登録される場合が多い。

【0070】

ノイズは、商品等に対する評価等を示す表現ではなく、肯定的でも否定的でもない言葉である。具体的には、レビューにおいて、「使いやすい」、「うるさい」及び「高い」等の表現は、商品等の特性を肯定又は否定して評価する表現であるため、評価表現に適した表現である。一方で、商品等の特性を肯定又は否定する以外の言葉が、ノイズとなる。

【0071】

そこで、辞書データに登録される候補となる登録語には、例えば、図8に示すようなフィルタリング処理が行われるのが望ましい。このようにすると、辞書データに登録されるノイズが少なくなり、サーバSERは、精度良く評判を分析できる。

【0072】

<登録語及びスコアを辞書データに追加し、辞書データを更新する例>

ステップS0307では、サーバSERは、登録語及びスコアを辞書データに追加し、辞書データを更新する。例えば、図5に示す例において、ステップS0303によって、「安い」という登録語が、抽出され、ステップS0305によって、「+0.85」という値のスコアが計算されたとする。なお、「安い」という登録語は、ステップS0306によるフィルタリングで削除対象としない(ステップS3604でNO)とされたとする。

【0073】

このような場合には、図示するように、サーバSERは、辞書データDDI1に「安い」の登録語と、スコアを追加する。このようにすると、辞書データDDI1は、更新され、図示する辞書データDDI2となる。なお、スコアが正の値であるため、「安い」の登録語は、肯定語に登録される例である。

【0074】

<繰り返すか否かの判断例>

ステップS0308では、サーバSERは、図5に示す処理を繰り返すか否かを判断する。なお、繰り返し回数は、例えば、あらかじめサーバSERに設定されるとする。したがって、所定の繰り返し回数まで、図5に示す処理を繰り返し行った場合には、サーバSERは、繰り返さないと判断する。一方で、所定の繰り返し回数まで、図5に示す処理を繰り返し行っていない場合には、サーバSERは、繰り返すと判断する。

【0075】

次に、繰り返すとサーバSERが判断すると(ステップS0308でYES)、サーバSERは、ステップS0301に進む。一方で、繰り返さないサーバSERが判断すると(ステップS0308でNO)、サーバSERは、辞書データを更新する処理を終了する。

【0076】

<特徴語の抽出例>

ステップS04では、サーバSERは、特徴語を抽出する。

【0077】

なお、特徴語は、商品等の特徴を示す言葉である。また、特徴語となる言葉は、評判分析における軸に付けられる軸名の候補となる言葉である。そのため、特徴語は、多くのレビューで評価される項目であるのが望ましい。すなわち、特徴語は、商品等の特性等を示し、多くのユーザが商品等において関心のある項目を示す言葉であるのが望ましい。特徴語は、このような性質の言葉であるため、商品等又は商品等のカテゴリごとに異なる言葉となる。例えば、特徴語は、以下のような処理によって抽出される。

【0078】

図9は、本発明の一実施形態に係る情報処理装置による特徴語の抽出例を示すフローチ

10

20

30

40

50

ャートである。

【0079】

< レビューデータから1文を抽出する例 >

ステップS0401では、サーバSERは、レビューデータDREから1文を抽出する。例えば、サーバSERは、ステップS0301と同様の処理によってレビューデータから1文を抽出する。

【0080】

< 評価表現と「係り受け」の関係にある特徴語の抽出例 >

ステップS0402では、サーバSERは、評価表現と「係り受け」の関係にある特徴語を抽出する。まず、ステップS0302のように、サーバSERは、辞書データDDIに基づいて、評価表現を文中から検索する。そして、サーバSERは、検索した評価表現と「係り受け」の関係にある言葉があると、特徴語として抽出し、特徴語データDSPに記憶する。なお、特徴語データDSPには、ノイズが含まれる場合があるため、特徴語データDSPは、後段で、特徴語のフィルタリング処理(ステップS05)が行われるのが望ましい。

10

【0081】

< 特徴語の重要度の計算例 >

ステップS0403では、サーバSERは、特徴語の重要度を計算する。例えば、サーバSERは、下記(4)式によって重要度を計算する。

【0082】

20

【数4】

$$\begin{aligned} weight_y & \\ &= \sqrt{extract_num(y)} \\ &\quad \times rate_particle(y) \times rate_EE(y) \end{aligned}$$

$weight_y$: 特徴語 y の重要度

30

$extract_num(y)$: 特徴語 y の抽出回数

$rate_particle(y)$: 特徴語 y が特定の助詞を伴う確率

$rate_EE(y)$: 特徴語 y が評価表現と

「係り受け」の関係にある確率

(4)

40

上記(4)式が示すように、重要度は、文中において、特徴語が抽出される回数(上記(4)式における「 $extract_num(y)$ 」)を考慮して計算される値である。すなわち、「係り受け」の関係にあつて、多く登場する言葉ほど、特徴語である可能性が高い。したがって、重要度を計算する上で、上記(4)式のように、抽出回数が考慮されるのが望ましい。

【0083】

さらに、上記(4)式が示すように、重要度は、文中において、特徴語が特定の助詞を

50

伴う確率（上記（４）式における「rate_particle(y)」である。）が高いと、値が高くなる。

【 0 0 8 4 】

なお、特定の助詞は、係助詞及び格助詞等である。具体的には、特定の助詞は、「が」、「も」及び「は」等である。このような特定の助詞を伴う言葉は、主語、述語又は目的語である可能性が高い。

【 0 0 8 5 】

そして、主語、述語又は目的語となる言葉は、レビューでは、評価の対象となる場合が多い言葉である。このように、抽出された特徴語の前後にある言葉の品詞が考慮されて、重要度が計算されると、主語、述語又は目的語となる言葉が特徴語に抽出されやすくなる。すなわち、特徴語となる言葉は、主語、述語又は目的語となる可能性が高いため、重要度は、上記（４）式のように、特定の助詞を伴う確率等を考慮して計算されるのが望ましい。

10

【 0 0 8 6 】

また、上記（４）式が示すように、重要度は、文中において、特徴語が評価表現と「係り受け」の関係である確率（上記（４）式における「rate_EE(y)」である。）が高いと、値が高くなる。評価表現は、形容詞等の名詞を修飾する言葉である場合が多い。したがって、評価表現と「係り受け」の関係にある言葉は、評価表現によって、何らかの評価がされている可能性が高い言葉である。上記（４）式のように、評価表現と「係り受け」の関係となる確率が考慮されて、重要度が計算されると、評価表現と「係り受け」の関係があるように使われる言葉が特徴語に抽出されやすくなる。すなわち、特徴語となる言葉は、評価表現によって評価される可能性が高いため、重要度は、上記（４）式のように、評価表現と「係り受け」の関係となる確率等を考慮して計算されるのが望ましい。

20

【 0 0 8 7 】

なお、重要度は、計算され、特徴語と関連付けして、特徴語データDSPに記憶されるとする。

【 0 0 8 8 】

< 特徴語が抽出された回数のカウント例 >

ステップS0404では、サーバSERは、特徴語が抽出された回数をカウントする。すなわち、サーバSERは、特徴語が文から抽出されると、「１回」とカウントし、抽出回数DC1を記憶する。つまり、他の文でも、評価表現と「係り受け」の関係となる場合が多い特徴語は、抽出回数DC1が大きい値となる。

30

【 0 0 8 9 】

< 特徴語が格助詞を伴うか否かの判断例 >

ステップS0405では、サーバSERは、特徴語が格助詞を伴うか否かを判断する。すなわち、サーバSERは、抽出された特徴語の前後に格助詞となる言葉があるか否かを判断する。

【 0 0 9 0 】

次に、特徴語が格助詞を伴うとサーバSERが判断すると（ステップS0405でYES）、サーバSERは、ステップS0406に進む。一方で、特徴語が格助詞を伴わないとサーバSERが判断すると（ステップS0405でNO）、サーバSERは、ステップS0407に進む。

40

【 0 0 9 1 】

< 特徴語が格助詞を伴って抽出された回数のカウント例 >

ステップS0406では、サーバSERは、特徴語が格助詞を伴って抽出された回数（以下「格助詞回数DC2」という。）をカウントする。すなわち、ステップS0405において特徴語が格助詞を伴うと判断されると（ステップS0405でYES）、サーバSERは、「１回」とカウントし、格助詞回数DC2を記憶する。

【 0 0 9 2 】

< 特徴語が係助詞を伴うか否かの判断例 >

50

ステップS0407では、サーバSERは、特徴語が係助詞を伴うか否かを判断する。すなわち、サーバSERは、抽出された特徴語の前後に係助詞となる言葉があるか否かを判断する。

【0093】

次に、特徴語が係助詞を伴うとサーバSERが判断すると(ステップS0407でYES)、サーバSERは、ステップS0408に進む。一方で、特徴語が係助詞を伴わないとサーバSERが判断すると(ステップS0407でNO)、サーバSERは、特徴語の抽出処理を終了する。

【0094】

<特徴語が係助詞を伴って抽出された回数のカウント例>

ステップS0408では、サーバSERは、特徴語が係助詞を伴って抽出された回数(以下「係助詞回数DC3」という。)をカウントする。すなわち、ステップS0407において特徴語が係助詞を伴うと判断されると(ステップS0407でYES)、サーバSERは、「1回」とカウントし、係助詞回数DC3を記憶する。

【0095】

<特徴語のフィルタリング例>

ステップS05では、サーバSERは、特徴語をフィルタリングする。例えば、特徴語は、以下のような処理によってフィルタリングされる。

【0096】

図10は、本発明の一実施形態に係る情報処理装置による特徴語のフィルタリング例を示すフローチャートである。

【0097】

<新しく特徴語が抽出されたか否かの判断例>

ステップS0501では、サーバSERは、新しく特徴語が抽出されたか否かを判断する。例えば、特徴語データDSPに新たに特徴語が追加されると、サーバSERは、新しく特徴語が抽出されたと判断する。

【0098】

次に、新しく特徴語が抽出されたとサーバSERが判断すると(ステップS0501でYES)、サーバSERは、ステップS0502に進む。一方で、新しく特徴語が抽出されていないとサーバSERが判断すると(ステップS0501でNO)、サーバSERは、特徴語のフィルタリング処理を終了する。

【0099】

<特徴語が格助詞を伴って抽出された確率及び特徴語が係助詞を伴って抽出された確率の計算例>

ステップS0502では、サーバSERは、特徴語が格助詞を伴って抽出された確率及び特徴語が係助詞を伴って抽出された確率を計算する。

【0100】

まず、特徴語が格助詞を伴って抽出された確率(以下「格助詞確率」という。)は、抽出回数DC1に対する格助詞回数DC2の割合を計算すると求まる値である。具体的には、サーバSERは、「格助詞回数DC2÷抽出回数DC1」を計算すると、格助詞確率を求めることができる。

【0101】

同様に、特徴語が係助詞を伴って抽出された確率(以下「係助詞確率」という。)は、抽出回数DC1に対する係助詞回数DC3の割合を計算すると求まる値である。具体的には、サーバSERは、「係助詞回数DC3÷抽出回数DC1」を計算すると、係助詞確率を求めることができる。

【0102】

<格助詞確率が第3閾値より大きい値であり、かつ、係助詞確率が第4閾値より大きい値であるか否かの判断例>

ステップS0503では、サーバSERは、格助詞確率が第3閾値より大きい値であり

10

20

30

40

50

、かつ、係助詞確率が第4閾値より大きい値であるか否かを判断する。

【0103】

まず、サーバSERは、ステップS0502で計算された格助詞確率と、第3閾値とを比較する。なお、第3閾値は、あらかじめ設定される値である。例えば、第3閾値には、「0.1」以上の値が設定される。このような第3閾値であると、「10パーセント」より高い確率で格助詞を伴う特徴語が、格助詞確率が第3閾値より大きい値であると判断される。

【0104】

次に、サーバSERは、ステップS0502で計算された係助詞確率と、第4閾値とを比較する。なお、第4閾値は、あらかじめ設定される値である。例えば、第4閾値には、「0.1」以上の値が設定される。このような第4閾値であると、「10パーセント」より高い確率で係助詞を伴う特徴語が、係助詞確率が第4閾値より大きい値であると判断される。

10

【0105】

そして、格助詞確率が第3閾値より大きい値であり、かつ、係助詞確率が第4閾値より大きい値であると（ステップS0503でYES）、サーバSERは、ステップS0505に進む。一方で、格助詞確率が第3閾値より大きい値でない、又は、係助詞確率が第4閾値より大きい値でないと（ステップS0503でNO）、サーバSERは、ステップS0504に進む。

【0106】

<特徴語の削除例>

ステップS0504では、サーバSERは、特徴語を削除する。すなわち、ステップS0503によって、格助詞確率が小さい値である、又は、係助詞確率が小さい値である特徴語は、特徴語データから削除される。

20

【0107】

<特徴語の重要度の読み出し例>

ステップS0505では、サーバSERは、特徴語の重要度を読み出す。すなわち、サーバSERは、ステップS0403で計算された重要度を読み出す。

【0108】

<第5閾値の計算例>

ステップS0506では、サーバSERは、第5閾値を計算して求める。例えば、第5閾値は、以下のように計算される。

30

【0109】

図11は、本発明の一実施形態に係る情報処理装置による第5閾値の計算例を示す図である。なお、図では、横軸が、特徴語データに記憶される特徴語を示し、縦軸が、各特徴語の重要度を示す。したがって、図では、横軸において、左から右に向かって、重要度が高い順に特徴語が並べられた状態である。

【0110】

レビューでは、頻繁に評価される言葉は、重要度が高くなる場合が多い。そのため、特徴語及び特徴語の候補となる言葉の集合における重要度の分布は、図示するように、ロングテールで示せる。そして、この分布において、重要度が高い言葉が、特徴語である可能性が高い。そこで、サーバSERは、図示する分布において、全体のうち、上位「」パーセントを占める言葉の特徴語であると推定し、それ以外の言葉をノイズと推定する。なお、「」は、あらかじめ設定される値である。

40

【0111】

具体的には、まず、サーバSERは、特徴語データに記憶される特徴語を重要度が高い順に並べ、上位から「」パーセントとなる特徴語を特定する。そして、サーバSERは、特定された特徴語の重要度を第5閾値とする。すなわち、サーバSERは、いわゆる累積密度が「」パーセントとなる重要度を特定し、第5閾値とする。このようにすると、サーバSERは、上位「」パーセントに含まれる重要度が高い言葉であるか否かを第5

50

閾値と重要度を比較することによって判断できる。

【0112】

なお、「」パーセントに入らない言葉（図では、「」より右側の部分に該当する。）から、「準特徴語」が抽出されてもよい。「準特徴語」の詳細は、後述する。

【0113】

<重要度が第5閾値より大きい値か否かの判断例>

ステップS0507では、サーバSERは、重要度が第5閾値より大きい値か否かを判断する。すなわち、サーバSERは、第5閾値に基づいて、特徴語が図11に示す分布において、上位に位置する言葉か否かを判断する。

【0114】

そして、重要度が第5閾値より大きい値であると（ステップS0507でYES）、サーバSERは、特徴語のフィルタリング処理を終了する。一方で、重要度が第5閾値より大きい値でないと（ステップS0507でNO）、サーバSERは、ステップS0508に進む。

【0115】

<特徴語の削除例>

ステップS0508では、サーバSERは、特徴語を削除する。すなわち、ステップS0507によって、重要度が小さい値である特徴語は、特徴語データから削除される。

【0116】

なお、格助詞確率及び係助詞確率に基づく判断（ステップS0502及びステップS0503等）と、重要度に基づく判断（ステップS0507等）とは、実行される順序が逆でもよい。ただし、図示するような順序とすると、サーバSERは、精度良く特徴語をフィルタリングできる。

【0117】

<軸の生成例>

ステップS06では、サーバSERは、軸を生成する。軸は、商品等の評判を示す上で、商品等が有する機能等の評価指標を示す。なお、軸は、「評価軸」等と呼ばれる場合もある。例えば、軸は、以下のような処理によって生成される。

【0118】

図12は、本発明の一実施形態に係る情報処理装置による軸の生成例を示すフローチャートである。

【0119】

<特徴語と評価表現の「係り受け」を示す行列の生成例>

ステップS0601では、サーバSERは、特徴語と評価表現の「係り受け」を示す行列を生成する。例えば、行列は、以下のように生成される。

【0120】

図13は、本発明の一実施形態に係る特徴語と評価表現の「係り受け」を示す行列の一例を示す図である。以下、図示するように、縦方向jに、特徴語を並べ、一方で、横方向iに、評価表現を並べた行列を例に説明する。

【0121】

この例では、縦方向jには、特徴語データDSPに記憶される特徴語が並べられる。そして、各特徴語と「係り受け」の関係があった評価表現が抽出され、横方向iに抽出された評価表現が並べられる。次に、行列の各要素には、「係り受け」の関係にあった回数が入力される。

【0122】

具体的には、図示する例では、「値段」という特徴語は、レビューにおいて、「安い」及び「満足」という評価表現と「係り受け」の関係にある言葉である。そして、この例では、「値段」と、「安い」という言葉は、「10」回「係り受け」の関係にあったので、「10」の値が行列の要素に入力される。同様に、この例では、「値段」と、「満足」という言葉は、「3」回「係り受け」の関係にあったので、「3」の値が行列の要素に入力

10

20

30

40

50

される。

【 0 1 2 3 】

<TF - IDF (Term Frequency - Inverse Document Frequency) 計算例>

ステップ S 0 6 0 2 では、サーバ S E R は、TF - IDF 計算を行う。TF - IDF 計算は、言葉が出現する頻度 (TF) と、多くの文に出現する頻度 (逆文書頻度、IDF) とを指標にする方法であって、各文書における言葉がどれだけ重要であることを示す尺度を計算する。具体的には、サーバ S E R は、下記 (5) 式によって、TF - IDF 計算を行う。

【 0 1 2 4 】

10

【 数 5 】

$$tfidf_{i,j} = tf_{i,j} \cdot idf_i$$

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

20

$$idf_i = \log \frac{F}{F_i}$$

$tfidf_{i,j}$: TF - IDF 値

$tf_{i,j}$: TF 値

idf_i : IDF 値

30

$n_{i,j}$: 評価表現 e_i と特徴語 f_j とが

「係り受け」の関係にあった回数

F : 特徴語の総数

F_i : 評価表現 e_i と共起する特徴語の数

(5)

40

例えば、上記 (5) 式に基づいて、図 1 3 に示す行列を計算すると、計算結果は、以下のようなになる。

【 0 1 2 5 】

図 1 4 は、本発明の一実施形態に係る情報処理装置による TF - IDF 計算の計算結果例を示す図である。まず、図示する例では、特徴語は、4 種類、すなわち、縦方向 j に 4 つの特徴語が抽出されるため、上記 (5) 式における「 F : 特徴語の総数」は、「4」となる。

50

【0126】

以下、「安い」の評価表現を例に説明する（上記（5）式における評価表現 e_i とする）。この例では、「安い」という評価表現は、「値段」及び「価格」という2つの特徴語と共起する。したがって、上記（5）式における「 F_i ：評価表現 e_i と共起する特徴語の数」は、「2」となる。

【0127】

次に、「 $n_{i,j}$ ：評価表現 e_i と特徴語 f_j とが「係り受け」の関係にあった回数」は、図13に示す「安い」の列から特定される。具体的には、「安い」と、「値段」との「 $n_{i,j}$ 」は、「10」となる。同様に、「安い」と、「価格」との「 $n_{i,j}$ 」は、「8」となる。さらに、「安い」と、「音」との「 $n_{i,j}$ 」及び「安い」と、「吸引力」との「 $n_{i,j}$ 」は、「0」となる。

10

【0128】

また、上記（5）式における「 $tf_{i,j}$ ：TF値」の計算での分母である「 $\sum_k n_k$ 」は、評価表現ごとの「係り受け」の関係にあった回数を合計した値となる。具体的には、図13に示す例では、「値段」は、「 $10 + 0 + 0 + 3 = 13$ 」となる。次に、「価格」は、「 $8 + 0 + 0 + 5 = 13$ 」となる。また、「音」は、「 $0 + 10 + 0 + 0 = 10$ 」となる。さらに、「吸引力」は、「 $0 + 0 + 9 + 5 = 14$ 」となる。

【0129】

また、図14に示す例では、TF-IDF値は、特徴語ごとに、TF-IDFを合計した値がほぼ「1」となるように正規化した値である。具体的には、図14では、「値段」のTF-IDF値は、合計すると、「 $0.88 + 0 + 0 + 0.11 = 0.99$ 」である。

20

【0130】

このように、TF-IDF計算を行うと、特徴語と「係り受け」の関係があつて、出現する回数の多い評価表現は、数値が高くなりやすい（TF値）。ただし、どのような特徴語に対しても評価する評価表現は、数値が低くなりやすい（IDF値）。

【0131】

この例では、「満足」という評価表現は、どのような特徴語を評価するにも出現することが多い評価表現の例である。具体的には、「満足」は、「値段」、「価格」及び「吸引力」の3つの特徴語に対して、「係り受け」の関係がある。このように、汎用性が高い評価表現は、TF-IDF計算では、値が小さくなる。一方で、「静か」という評価表現は、「音」という特徴語としか「係り受け」の関係がないため、TF-IDF値は、「1.0」のように、値が大きくなる。

30

【0132】

<特徴語のクラスタリング例>

ステップS0603では、サーバSERは、特徴語をクラスタリングする。例えば、サーバSERは、k-medoids法及びGap統計量計算によって、特徴語をクラスタリングする。そして、クラスタリングされた各クラスタが、評判分析による軸となる。

【0133】

まず、クラスタ数 k は、Gap統計量を計算して決定する。具体的には、サーバSERは、例えば、クラスタ数 k を「 $k = 1, 2, 3 \dots$ 」の1つずつ増加させていく。そして、サーバSERは、それぞれのクラスタ数 k ごとに、Gap統計量を計算する。

40

【0134】

Gap統計量は、各クラスタにおけるデータの密集度を示す値である。すなわち、Gap統計量が高い値であるほど、クラスタ内では、データは、凝集している状態である。そして、サーバSERは、Gap統計量が高くなるクラスタ数 k を特定する。なお、複数のクラスタ数 k が特定される場合には、サーバSERは、最も小さい値のクラスタ数 k を選択する。

【0135】

Gap統計量によって、クラスタ数 k を特定すると、サーバSERは、X-means

50

法等と比較して、少ない計算コストによって、クラスタ数 k を特定できる。また、Gap 統計量によって、クラスタ数 k を特定すると、サーバSERは、X - means 法等と比較して、精度良くクラスタ数 k を特定できる。さらに、Gap 統計量によって、クラスタ数 k を特定すると、サーバSERは、X - means 法等と比較して、安定してクラスタ数 k を特定できる。

【0136】

次に、サーバSERは、 k - medoids 法によって、代表点を特定する。例えば、 k - medoids 法によって、以下のように代表点が特定される。

【0137】

図15は、本発明の一実施形態に係る情報処理装置による k - medoids 法による代表点の特定例を示す図である。まず、図示するように、特徴語が、クラスタCLUにクラスタリングされるとする。したがって、「 」で図示するように、それぞれのクラスタCLU内には、特徴語が点在する。

10

【0138】

k - medoids 法は、教師なし学習法の一例である。 k - medoids 法では、特徴語となる各データは、他のデータとの距離が計算され、距離の総和が最も小さい値となるデータが特定される。例えば、図示するように、クラスタCLUごとに、1つずつ代表点PREが特定される。このように、 k - medoids 法が行われると、サーバSERは、クラスタCLUごとに、中心となる特徴語を代表点PREとして特定することができる。

20

【0139】

また、 k - medoids 法であると、 k - means 法等と比較して、サーバSERは、いわゆる外れ値の影響を小さくできる。 k - means 法では、平均値が計算されるため、データがない点等が選ばれる可能性が高い。一方で、 k - medoids 法では、クラスタCLU内にあるデータから、代表点PREが特定されるため、特徴語の集合にない言葉が選ばれるようなことが防げる。

【0140】

さらに、サーバSERは、 k - medoids 法を用いると、同じような意味の言葉を1つの軸に絞ることができる。具体的には、「値段」及び「価格」等は、どちらも金銭についての言葉である。このような言葉は、別の言葉であっても、どちらも同じような意味で使われることが多い言葉である。そこで、サーバSERは、 k - medoids 法を用いると、同じような意味を示す特徴語のクラスタから、1つの特徴語を選び、言葉の揺らぎを吸収することができる。

30

【0141】

以上のように、 k - medoids 法及びGap 統計量計算を用いると、サーバSERは、例えば、以下のように、特徴語をクラスタリングして軸を生成することができる。以下、「掃除機」の商品に対するレビューで、図示するような特徴語がステップS04で抽出された場合を例に説明する。

【0142】

図16は、本発明の一実施形態に係る情報処理装置による軸の生成例を示す図である。

40

【0143】

図示する例では、図12に示す処理を行うと、サーバSERは、3つのクラスタに特徴語をクラスタリングできる。すなわち、この例では、サーバSERは、「軸1」、「軸2」及び「軸3」の3つの軸を生成することができる。なお、これらの軸とは別に、さらに、総合的な評価を示す「総合評価軸」を加えて、4つの軸としてもよい。以下、「総合評価軸」がある場合を例に説明する。

【0144】

そして、図示するように、 k - medoids 法を用いると、各軸において、代表点PREとなる特徴語が特定される。そして、代表点PREに特定された特徴語が、それぞれの軸名となる。このように、サーバSERは、軸を生成することができる。

50

【0145】

このように、k - m e d o i d s 法及び G a p 統計量計算を用いると、サーバ S E R は、多すぎず、少なすぎない軸の数に最適化し、かつ、それぞれの軸に、それぞれの軸名を決定することができる。

【0146】

なお、軸の生成では、サーバ S E R は、以下のような計算をして軸を削除してもよい。まず、サーバ S E R は、すべての特徴語の重要度を総和した値（以下「全体値」という。）を計算する。次に、サーバ S E R は、軸ごとの特徴語の重要度を総和した値（以下「軸値」という。）を計算する。そして、全体値に対する軸値が、所定の値以下であると、サーバ S E R は、計算対象とした軸を削除する。なお、所定の値は、あらかじめ設定される

10

【0147】

クラスタリングでは、同じような意味を示す特徴語の数が少ない場合、すなわち、データ数がとても少ないクラスタが発生する場合がある。このような場合には、サーバ S E R は、データ数がとても少ないクラスタをノイズとし、削除する。

【0148】

重要な商品等の特性、すなわち、軸については、頻繁にレビューされる場合が多い。したがって、重要な軸であると、軸にクラスタリングされた特徴語は、重要度が高くなる場合が多い。したがって、重要度が低い軸を削除すると、サーバ S E R は、重要な軸に絞って評判の分析結果を示すことができる。

20

【0149】

< 準特徴語の分配例 >

軸の生成後、準特徴語が分配されるのが望ましい。

【0150】

ステップ S 0 6 0 4 では、サーバ S E R は、準特徴語を分配する。まず、準特徴語が分配される場合には、ステップ S 0 5 0 8 では、サーバ S E R は、対象となる特徴語を削除せず、「準特徴語」として記憶する。そして、サーバ S E R は、軸ごとに、各準特徴語の T F - I D F 値を計算する。このように、T F - I D F 値によって、各準特徴語と、軸との距離が定まる。次に、サーバ S E R は、T F - I D F 値のうち、最も高い値（以下「1番目の値」という。）と、1番目の値の次に高い値（以下「2番目の値」という。）とを

30

【0151】

レビューでは、全体では、出現する頻度が低い、特定の話題又は特性では、よく使われる言葉がある。このような言葉は、全体としては、出現する頻度が低いため、重要度等が低い値になる可能性が高い。しかし、このような言葉は、ある特性を評価する上では、重要な言葉である場合が多い。そこで、ステップ S 0 5 0 7 で重要度が低いと判断されても、このような言葉を削除せず、準特徴語とする。

【0152】

そして、（1番目の値 ÷ 2番目の値）は、所定の軸に対してだけ、T F - I D F 値が高いか否かを判定するための値である。つまり、ある特性に対するレビューでよく出現する言葉であれば、1番目の値は、高い値となるが、他の特性に対するレビューでは、あまり出現しないため、2番目の値等は、低い値となる場合が多い。そこで、（1番目の値 ÷ 2番目の値）が高い値であれば、準特徴語は、ある特性を評価する上では、重要な言葉である可能性が高い。

40

【0153】

したがって、サーバ S E R は、（1番目の値 ÷ 2番目の値）が高い値であれば、1番目の値となる軸に、準特徴語を分配する。このようにすると、サーバ S E R は、特定の話題又は特性に偏って使われる言葉も特徴語として使用することができる。

【0154】

50

< 評判分析例 >

ステップ S 0 7 では、サーバ S E R は、評判分析を行う。例えば、評判分析は、以下のように行われる。以下、図 1 6 に示すように、軸が生成され（なお、以下の説明では、「総合評価」軸もあるとする。）、軸ごとに、評判を示す値（以下「評価値」という。）が計算されて、評判分析が行われるとする。

【 0 1 5 5 】

図 1 7 は、本発明の一実施形態に係る情報処理装置による評判分析例を示すフローチャートである。

【 0 1 5 6 】

< 商品又はサービスの選択例 >

ステップ S 0 7 0 1 では、ユーザは、商品等を選択する。一方で、レビューデータには、各レビューが、どの商品等についてのレビューであるかを識別できる ID が付されている。そのため、以下の処理で処理対象となるレビューデータは、すべてのレビューデータから、ステップ S 0 7 0 1 で選択された商品等と同一の商品等についてのレビューデータに絞られる。

【 0 1 5 7 】

< レビューデータから 1 文を抽出する例 >

ステップ S 0 7 0 2 では、サーバ S E R は、レビューデータ D R E から 1 文を抽出する。例えば、サーバ S E R は、ステップ S 0 3 0 1 と同様の処理によってレビューデータから 1 文を抽出する。

【 0 1 5 8 】

< 「係り受け」解析例 >

ステップ S 0 7 0 3 では、サーバ S E R は、「係り受け」解析を行う。「係り受け」解析が行われると、サーバ S E R は、各文節における「係り受け」の関係と、主節となる文節を把握できる。

【 0 1 5 9 】

< 評価表現があるか否かの判断例 >

ステップ S 0 7 0 4 では、サーバ S E R は、評価表現があるか否かを判断する。具体的には、ステップ S 0 7 0 2 によって抽出された文のいずれかに、評価表現となる言葉が含まれると、サーバ S E R は、評価表現があると判断する。

【 0 1 6 0 】

次に、評価表現があるとサーバ S E R が判断すると（ステップ S 0 7 0 4 で Y E S ）、サーバ S E R は、ステップ S 0 7 0 5 に進む。一方で、評価表現がないとサーバ S E R が判断すると（ステップ S 0 7 0 4 で N O ）、サーバ S E R は、ステップ S 0 7 0 2 又はステップ S 0 7 1 2 に進む。

【 0 1 6 1 】

< 評価表現が主節であるか否かの判断例 >

ステップ S 0 7 0 5 では、サーバ S E R は、評価表現が主節であるか否かを判断する。すなわち、ステップ S 0 7 0 3 による解析等の構文解析が行われると、サーバ S E R は、対象となる評価表現が主文の述語であるか否かを判断できる。

【 0 1 6 2 】

次に、評価表現が主節であるとサーバ S E R が判断すると（ステップ S 0 7 0 5 で Y E S ）、サーバ S E R は、ステップ S 0 7 0 6 に進む。一方で、評価表現が主節でないとサーバ S E R が判断すると（ステップ S 0 7 0 5 で N O ）、サーバ S E R は、ステップ S 0 7 0 7 に進む。

【 0 1 6 3 】

< 「総合評価」軸の評価値に評価表現のスコアを加算する例 >

ステップ S 0 7 0 6 では、サーバ S E R は、「総合評価」軸の評価値に評価表現のスコアを加算する。すなわち、評価表現が主節である場合には、評価表現は、商品等を総合的に評価した結果を示す表現である場合が多い。そこで、サーバ S E R は、評価表現が主節

10

20

30

40

50

である場合には、「総合評価」軸の評価値に、評価表現のスコアを反映させる。

【0164】

<特徴語と「係り受け」の関係があるか否かの判断例>

ステップS0707では、サーバSERは、評価表現が特徴語と「係り受け」の関係にあるか否かを判断する。

【0165】

次に、評価表現が特徴語と「係り受け」の関係にあるとサーバSERが判断すると(ステップS0707でYES)、サーバSERは、ステップS0708に進む。一方で、評価表現が特徴語と「係り受け」の関係にないとサーバSERが判断すると(ステップS0707でNO)、サーバSERは、ステップS0709に進む。

10

【0166】

<「係り受け」の関係がある特徴語の軸の選択例>

ステップS0708では、サーバSERは、「係り受け」の関係がある特徴語の軸を選択する。すなわち、サーバSERは、「係り受け」の関係がある特徴語がクラスタリングされた軸を特定する。具体的には、図16に示す例では、「値段」が「係り受け」の関係がある特徴語とであると、サーバSERは、「軸1」を選択する。

【0167】

<代表評価表現であるか否かの判断例>

ステップS0709では、サーバSERは、評価表現が代表評価表現であるか否かを判断する。なお、代表評価表現は、例えば、軸の生成後、軸ごとに決定される。

20

【0168】

評価表現には、特徴語がなくとも特定の特性を評価できる表現がある。例えば、「安い」という評価表現は、「値段」等の特徴語が伴わなくても、金銭についての評価を示して使用される場合が多い。したがって、このような言葉は、特徴語と「係り受け」の関係がなくとも、特定の軸について評価を示すため、評価表現として評価値に反映されることが望ましい言葉である。以下、このような言葉を「代表評価表現」という。

【0169】

代表評価表現は、評価表現とは別に決定される。例えば、代表評価表現は、以下のように決定される。まず、ステップS0603でk-medoids法が行われると、軸ごとに、代表点となる特徴語が特定される。そして、代表点となる特徴語と、各評価表現との間には、TF-IDF値がある。

30

【0170】

次に、TF-IDF値が所定の閾値(以下「代表評価表現判定値」という。)以上である評価表現があると、サーバSERは、対象となる評価表現を代表評価表現に決定する。なお、代表評価表現判定値は、あらかじめ設定される。例えば、図14に示す例では、「静か」という評価表現は、TF-IDF値が高いため、代表評価表現と決定される。同様に、「安い」という評価表現は、TF-IDF値が高いため、代表評価表現と決定される。また、「強い」という評価表現も、TF-IDF値が高いため、代表評価表現と決定される。

【0171】

ステップS0709では、サーバSERは、決定された代表評価表現と同一の言葉であると、評価表現が代表評価表現であると判断する。

40

【0172】

次に、評価表現が代表評価表現であるとサーバSERが判断すると(ステップS0709でYES)、サーバSERは、ステップS0710に進む。一方で、評価表現が代表評価表現でないとサーバSERが判断すると(ステップS0709でNO)、サーバSERは、ステップS0702又はステップS0712に進む。

【0173】

<代表評価表現に対応する軸の選択例>

ステップS0710では、サーバSERは、代表評価表現に対応する軸を選択する。す

50

なわち、サーバSERは、代表評価表現が決定された軸を特定する。具体的には、上記に示す例では、「静か」が「軸2」の代表評価表現と決定されているため、サーバSERは、「静か」という代表評価表現がある場合には、「軸2」を選択する。

【0174】

< 選択された軸の評価値に評価表現のスコアを加算する例 >

ステップS0711では、サーバSERは、選択された軸の評価値に評価表現のスコアを加算する。すなわち、ステップS0708又はステップS0710によって、軸があらかじめ選択される。そして、サーバSERは、あらかじめ選択される軸の評価値に評価表現のスコアを反映させる。

【0175】

以上のようなステップS0702乃至ステップS0711が所定回数繰り返し行われる。そして、所定回数繰り返した後、サーバSERは、ステップS0712に進む。一方で、繰り返しの場合には、ステップS0704、ステップS0709又はステップS0711の後、サーバSERは、ステップS0702に進む。なお、所定回数となるループ回数及びループ条件等は、あらかじめ設定される値である。

【0176】

< レーダーチャートへのプロット例 >

ステップS0712では、サーバSERは、レーダーチャートへの評価値をプロットする。例えば、サーバSERは、レーダーチャートによって軸及び評判の解析結果をユーザに表示する。

【0177】

以上のような処理が具体的には以下のように行われる。

【0178】

図18は、本発明の一実施形態に係る情報処理装置による評判分析例を示す図である。以下、まず、ステップS03によって、図示するような辞書データDDI3が生成される例であるとする。さらに、ステップS07が図示するような生成結果AXIであって、それぞれの軸に、図示するように、特徴語がクラスタリングされ、かつ、代表評価表現が決定された例であるとする。

【0179】

また、「掃除機」の商品について、評判分析を行うとする(ステップS0701)。次に、評判分析において、レビューデータから、「少しうるさいけど吸引力は強いし満足です。」という文(以下「対象文SEN」という。)が抽出された例とする(ステップS0702)。

【0180】

続いて、「係り受け」解析が行われると(ステップS0703)、対象文SENは、図示するような解析結果RESとなる。具体的には、サーバSERは、対象文SENにおいて、「少し」と、「うるさい」という言葉は、「係り受け」の関係があると解析する。同様に、「吸引力」と、「強い」という言葉は、「係り受け」の関係があると解析する。また、「満足」という言葉は、対象文SENでは、最後の文節に含まれるため、主節にあると解析される。なお、「思う」又は「考える」等の言葉が最後の文節に含まれ場合等では、最後から1つ前の文節が、主節であるとされてもよい。

【0181】

次に、解析結果RESに基づいて、ステップS0704乃至ステップS0711の処理が繰り返し行われる。まず、辞書データDDI3に基づいて、「満足」という言葉は、評価表現であると判断される(ステップS0704でYES)。続いて、解析結果RESに基づいて、「満足」という言葉は、主節であると判断される(ステップS0705でYES)。したがって、「総合評価」軸が選択され、「総合評価」軸の評価値に、辞書データDDI3が示す「満足」のスコアである「+0.95」が加算される(ステップS0706)。

【0182】

10

20

30

40

50

次に、辞書データDDI3に基づいて、「強い」という言葉は、評価表現であると判断される(ステップS0704でYES)。続いて、解析結果RESに基づいて、「満足」という言葉は、「軸2」にクラスタリングされている「吸引力」という特徴語と「係り受け」の関係があると判断される(ステップS0707でYES)。また、「吸引力」という特徴語が「軸2」にクラスタリングされているため、「軸2」が選択される(ステップS0708)。したがって、「軸2」が選択され、「軸2」の評価値に、辞書データDDI3が示す「強い」のスコアである「+0.90」が加算される(ステップS0711)。

【0183】

次に、辞書データDDI3に基づいて、「うるさい」という言葉は、評価表現であると判断される(ステップS0704でYES)。続いて、解析結果RES及び生成結果AXIに基づいて、「うるさい」という言葉は、「軸1」に決定されている代表評価表現であると判断される(ステップS0709でYES)。また、「うるさい」という評価表現が「軸1」の代表評価表現に決定されているため、「軸1」が選択される(ステップS0710)。したがって、「軸1」が選択され、「軸1」の評価値に、辞書データDDI3が示す「うるさい」のスコアである「-0.92」が加算される(ステップS0711)。

10

【0184】

以上のように評判分析が行われると、サーバSERは、図示するような評判分析結果ANSを生成できる。なお、評判分析結果ANSでは、1つの対象文SENのみの評判分析の結果であるため、「レビュー数」は、「1」である。以降、同様に、他の文についても評判分析を行うと、評判分析結果ANSが示す「評価値」が、ステップS0707又はステップS0711によって変化する。以上のような評判分析結果ANSが得られると、サーバSERは、例えば、以下のようなレーダーチャート等によって軸及び評判の解析結果をユーザに表示できる。

20

【0185】

図19は、本発明の一実施形態に係る情報処理装置が表示するレーダーチャートの一例を示す図である。まず、図示するように、図18に示す例では、生成結果AXIに基づいて、サーバSERは、「軸1」乃至「軸3」となる軸AX1、軸AX2及び軸AX3をレーダーチャート上に表示する。さらに、サーバSERは、「総合評価」軸となる軸AXTをレーダーチャート上に表示する。

30

【0186】

次に、サーバSERは、評判分析結果ANSに基づいて、軸AX1上に「軸1」の評価値をプロットする(ステップS0712)。同様に、サーバSERは、評判分析結果ANSに基づいて、軸AX2上に「軸2」の評価値をプロットする(ステップS0712)。さらに、サーバSERは、評判分析結果ANSに基づいて、軸AX3上に「軸3」の評価値をプロットする(ステップS0712)。さらにまた、サーバSERは、評判分析結果ANSに基づいて、軸AXT上に「総合評価」軸の評価値をプロットする(ステップS0712)。

【0187】

以上のようなレーダーチャートが表示されると、ユーザは、サーバSERが生成した軸を見ることができる。また、評価値に基づいて、商品等の様々な特性が評価された結果が各軸上に表示されるため、ユーザは、商品等が有する機能等が評判の良い機能であるかを把握することができる。

40

【0188】

< 実行結果例 >

以下、全体処理を実行した評価実験結果の例を示す。

【0189】

なお、評価実験は、以下のようなレビューを対象とした。

【0190】

図20は、本発明の一実施形態に係る情報処理装置が全体処理の対象としたレビューの

50

条件を示す表である。図示するような条件下において、投稿されたレビューのレビューデータを全体処理の対象として評価した結果を以下に説明する。まず、この例では、商品及びそれぞれのレビュー数は、以下のようであった。

【0191】

図21は、本発明の一実施形態に係る情報処理装置が全体処理の対象とした商品及びそれぞれのレビュー数を示す表である。図示するように、以下に示す評価実験結果は、「掃除機」、「スーツ」及び「和菓子」の3カテゴリの商品についてのレビューがあった例である。

【0192】

また、評価実験結果は、以下のような設定によって行われた結果である。

【0193】

図22は、本発明の一実施形態に係る情報処理装置が全体処理に用いた設定値を示す表である。図示するような設定値が、全体処理が行われる前に、あらかじめ情報処理装置にされて全体処理が実行された結果が以下の通りである。

【0194】

まず、辞書データは、ステップS03によって、以下のように生成された。

【0195】

図23は、本発明の一実施形態に係る情報処理装置が全体処理において生成した辞書データを示す表である。図示するような数の肯定語及び否定語が、レビューデータより抽出された。具体的には、以下のような評価表現が抽出された。

【0196】

図24は、本発明の一実施形態に係る情報処理装置が全体処理において「掃除機」のレビューデータから抽出した評価表現を示す表である。

【0197】

図25は、本発明の一実施形態に係る情報処理装置が全体処理において「スーツ」のレビューデータから抽出した評価表現を示す表である。

【0198】

図26は、本発明の一実施形態に係る情報処理装置が全体処理において「和菓子」のレビューデータから抽出した評価表現を示す表である。

【0199】

また、特徴語は、以下のように分布した。

【0200】

図27は、本発明の一実施形態に係る情報処理装置による全体処理において出現した特徴語の分布を示す図である。

【0201】

そして、特徴語は、以下のように抽出された。

【0202】

図28は、本発明の一実施形態に係る情報処理装置が全体処理において抽出した特徴語を示す表である。

【0203】

続いて、軸は、以下のような評価軸が生成された。

【0204】

図29は、本発明の一実施形態に係る情報処理装置が全体処理において「掃除機」のレビューデータから生成した軸を示す表である。なお、表における「除去フラグ」は、軸を削除する処理を行った結果である。すなわち、この評価実験では、図3に示すステップS06において、軸を生成する処理の後、重要度が低い軸を削除する処理を行っている。そして、削除対象になるほど、重要度が低いと判断されると、「除去フラグ」には、「T」が入力される。一方で、削除対象としない場合には、「除去フラグ」には、「F」が入力される。以下、同様に記載する。

【0205】

10

20

30

40

50

図30は、本発明の一実施形態に係る情報処理装置が全体処理において「スーツ」のレビューデータから生成した軸を示す表である。

【0206】

図31は、本発明の一実施形態に係る情報処理装置が全体処理において「和菓子」のレビューデータから生成した軸を示す表である。

【0207】

また、準特徴語は、以下のように抽出及び分配された。

【0208】

図32は、本発明の一実施形態に係る情報処理装置が全体処理において「掃除機」のレビューデータから抽出及び分配した準特徴語を示す表である。

10

【0209】

図33は、本発明の一実施形態に係る情報処理装置が全体処理において「スーツ」のレビューデータから抽出及び分配した準特徴語を示す表である。

【0210】

図34は、本発明の一実施形態に係る情報処理装置が全体処理において「和菓子」のレビューデータから抽出及び分配した準特徴語を示す表である。

【0211】

また、代表評価表現は、以下のように決定された。

【0212】

図35は、本発明の一実施形態に係る情報処理装置が全体処理において「掃除機」のレビューデータから生成した軸及び代表評価表現を示す表である。

20

【0213】

図36は、本発明の一実施形態に係る情報処理装置が全体処理において「スーツ」のレビューデータから生成した軸及び代表評価表現を示す表である。

【0214】

図37は、本発明の一実施形態に係る情報処理装置が全体処理において「和菓子」のレビューデータから生成した軸及び代表評価表現を示す表である。

【0215】

以上のような結果から、以下のようなレーダーチャートが生成された。

【0216】

30

図38は、本発明の一実施形態に係る情報処理装置が全体処理において「掃除機」のレビューデータから生成したレーダーチャートを示す図である。「掃除機」については、全体処理において、図35に示すように、「評価軸1」乃至「評価軸8」の8つの軸が生成された。したがって、図38に示すレーダーチャートでは、「モーター音」、「使い勝手」、「吸引力」、「形」、「価格」、「ヘッド」、「ごみ捨て」及び「使い心地」の8つの軸が生成され、各軸上に、それぞれの評判分析結果が表示された。なお、このレーダーチャートでは、総合評価の結果は、「4.07」及び「1.64」のように数値で示している。

【0217】

図39は、本発明の一実施形態に係る情報処理装置が全体処理において「スーツ」のレビューデータから生成したレーダーチャートを示す図である。「スーツ」については、全体処理において、図36に示すように、「評価軸1」乃至「評価軸7」の7つの軸が生成された。したがって、図39に示すレーダーチャートでは、「生地」、「質感」、「お値段」、「サイズ展開」、「シルエット」、「質」及び「丈」の7つの軸が生成され、各軸上に、それぞれの評判分析結果が表示された。なお、このレーダーチャートでは、総合評価の結果は、「4.09」及び「2.64」のように数値で示している。

40

【0218】

図40は、本発明の一実施形態に係る情報処理装置が全体処理において「和菓子」のレビューデータから生成したレーダーチャートを示す図である。「和菓子」については、全体処理において、図37に示すように、「評価軸1」乃至「評価軸7」の7つの軸が生成

50

された。したがって、図40に示すレーダーチャートでは、「餡」、「縁起」、「甘さ」、「見た目」、「お値段」、「いちご」及び「生地」の7つの軸が生成され、各軸上に、それぞれの評判分析結果が表示された。なお、このレーダーチャートでは、総合評価の結果は、「3.80」及び「3.29」のように数値で示している。

【0219】

図38乃至図40では、各カテゴリにおいて、2つの商品について、評判分析を行っている。また、評価値は、「1」乃至「5」の範囲となるように正規化した値である。そして、評価値は、値が大きいと、肯定的である。

【0220】

具体的には、図38に示すように、「掃除機」のカテゴリでは、「Product1」の商品が、「Product2」の商品より、「使い心地」、「モーター音」及び「ヘッド」といった特性で評判が良いことがわかる。

10

【0221】

以上のように、情報処理装置は、商品等の評判を示す軸を生成できる。

【0222】

<機能構成例>

図41は、本発明の一実施形態に係る情報処理装置の機能構成例を示す機能ブロック図である。図示するように、サーバSERは、入力部F01と、評価表現抽出部F02と、特徴語抽出部F03と、クラスタリング部F04と、評判分析部F05とを含む機能構成である。また、図示するように、サーバSERは、図示するように、評判分析部F05を更に含む機能構成であるのが望ましい。以下、図示する機能構成を例に説明する。

20

【0223】

入力部F01は、商品等に対する複数のレビューデータDREを入力する入力手順を行う。例えば、入力部F01は、入力装置H03(図2参照)又はネットワークインタフェースH05(図2参照)等によって実現される。

【0224】

評価表現抽出部F02は、複数のレビューデータDREに含まれ、商品等に対する意見、評判、感想又は評価を示す評価表現を抽出する評価表現抽出手順を行う。例えば、評価表現抽出部F02は、CPUH01(図2参照)等によって実現される。

【0225】

特徴語抽出部F03は、評価表現に基づいて、複数のレビューデータDREに含まれ、商品等の特徴を示す特徴語を抽出する特徴語抽出手順を行う。例えば、特徴語抽出部F03は、CPUH01(図2参照)等によって実現される。

30

【0226】

クラスタリング部F04は、特徴語抽出部F03が抽出する特徴語をクラスタに分け、クラスタごとに軸を生成するクラスタリング手順を行う。例えば、クラスタリング部F04は、CPUH01(図2参照)等によって実現される。

【0227】

評判分析部F05は、複数のレビューデータDREから、クラスタリング部F04が生成する軸ごとに、商品等の特性についての評判を示す評判値を計算する評判分析手順を行う。例えば、評判分析部F05は、CPUH01(図2参照)等によって実現される。

40

【0228】

まず、サーバSERは、入力部F01によって、複数のレビューを示す複数のレビューデータDREを入力する。次に、サーバSERは、評価表現抽出部F02によって、例えば、ステップS03のように、評価表現を抽出し、辞書データDDIを更新して、辞書データDDIを構築する。なお、サーバSERは、評価表現を例えば、ステップS0306のようにフィルタリングするのが望ましい。このようにすると、サーバSERは、複数のレビューデータDREから、商品等に対する評価等を示す言葉である評価表現を抽出できる。

【0229】

50

次に、サーバ S E R は、特徴語抽出部 F 0 3 によって、複数のレビューデータ D R E から、例えば、ステップ S 0 4 のように、特徴語を抽出し、特徴語データ D S P に記憶する。なお、サーバ S E R は、特徴語をステップ S 0 5 のようにフィルタリングするのが望ましい。さらに、サーバ S E R は、ステップ S 0 6 0 4 のように、準特徴語を分配するのが望ましい。このようにすると、サーバ S E R は、複数のレビューデータ D R E から、商品等の特性を示す言葉である特徴語を抽出できる。

【 0 2 3 0 】

続いて、サーバ S E R は、クラスタリング部 F 0 4 によって、ステップ S 0 6 0 3 のように、複数の特徴語をクラスタに分ける。このようにすると、サーバ S E R は、複数のレビューデータ D R E に含まれる特徴語を例えば、図 3 5 乃至図 3 7 のようにクラスタリング
10

【 0 2 3 1 】

そして、評判分析部 F 0 5 によって評判分析が例えば、ステップ S 0 7 のように行われると、サーバ S E R は、例えば、図 1 9 のように、レーダーチャートによって、商品等の評判をユーザに示すことができる。

【 0 2 3 2 】

< その他の実施形態 >

情報処理装置は、1台の装置でなく、複数の装置で構成されてもよい。すなわち、本発明に係る処理は、1台以上の情報処理装置を有する情報処理システムによって実行されてもよい。そして、情報処理システムは、各処理を並行、分散、冗長、仮想化又はこれらの組み合わせで実行してもよい。
20

【 0 2 3 3 】

なお、本発明に係る各処理の全部又は一部は、アセンブラ等の低水準言語又はオブジェクト指向言語等の高水準言語で記述され、コンピュータに情報処理方法を実行させるためのプログラムによって実現されてもよい。すなわち、プログラムは、情報処理装置又は1以上の情報処理装置を含む情報処理システム等のコンピュータに各処理を実行させるためのコンピュータプログラムである。

【 0 2 3 4 】

また、プログラムは、コンピュータが読み取り可能な記録媒体に記録されて頒布することができる。なお、記録媒体は、磁気テープ、フラッシュメモリ、光ディスク、光磁気ディスク又は磁気ディスク等のメディアである。さらに、プログラムは、電気通信回線を通じて頒布することができる。
30

【 0 2 3 5 】

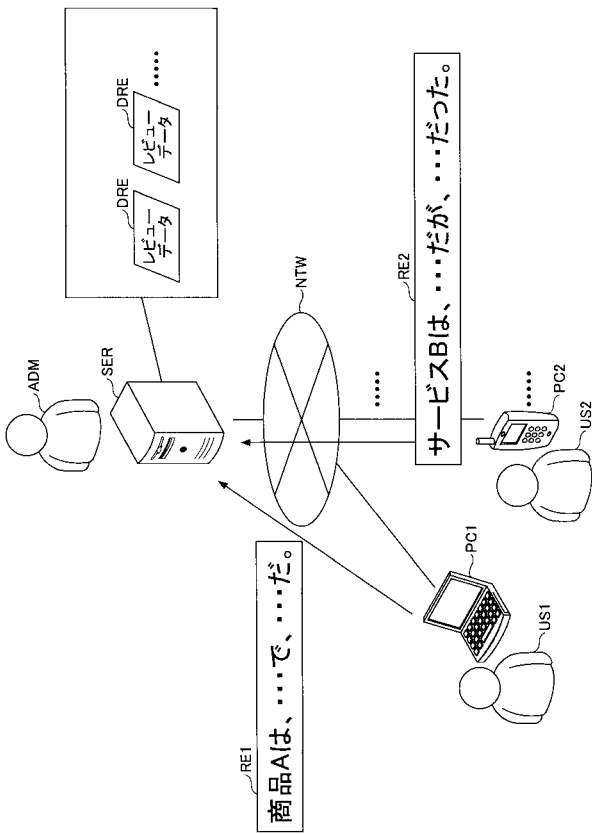
以上、本発明の好適な一実施形態を例に説明した。なお、上記の説明では、特定の具体例を示して本発明を説明したが、特許請求の範囲に定義された本発明の広範な趣旨及び範囲から逸脱することなく、上記に説明した具体例に、様々な修正及び変更が加えられてもよい。すなわち、具体例の詳細及び添付の図面によって説明した実施形態に、本発明が限定されるものと解釈してはならない。
40

【 符号の説明 】

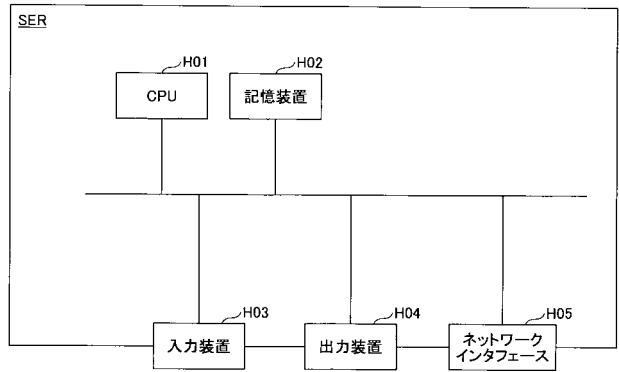
【 0 2 3 6 】

S E R サーバ
D R E レビューデータ
D D I、D D I 1、D D I 2、D D I 3 辞書データ
D S P 特徴語データ
C L U クラスタ

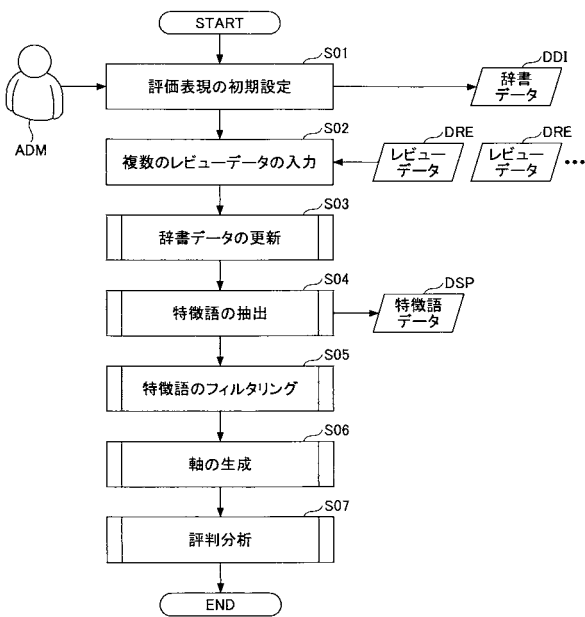
【 図 1 】



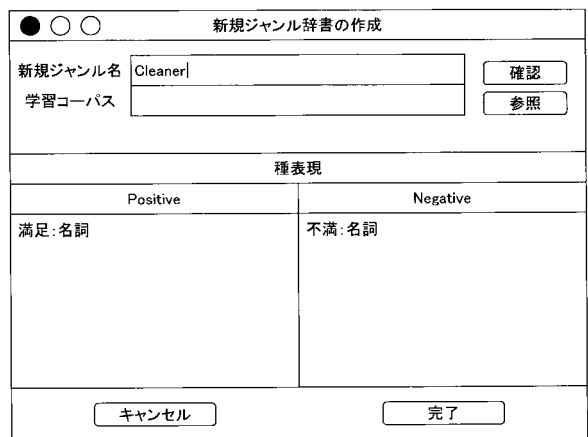
【 図 2 】



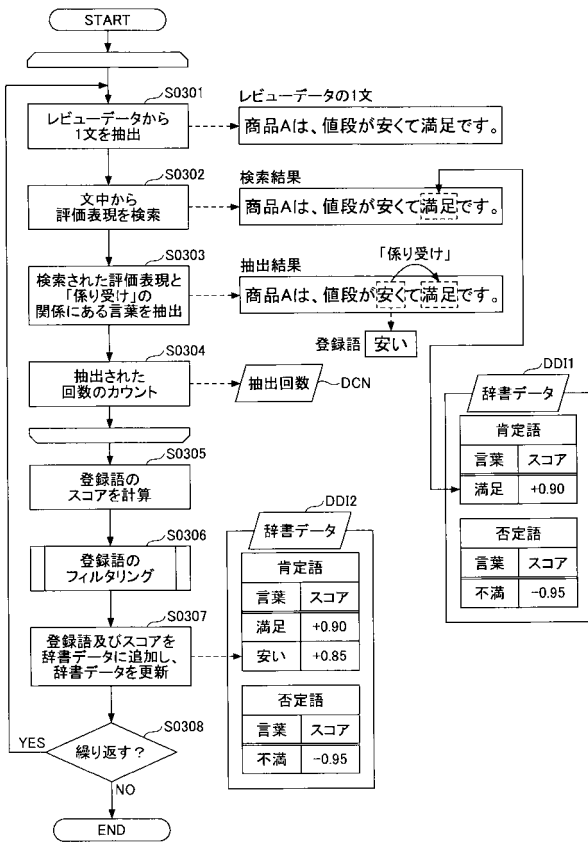
【 図 3 】



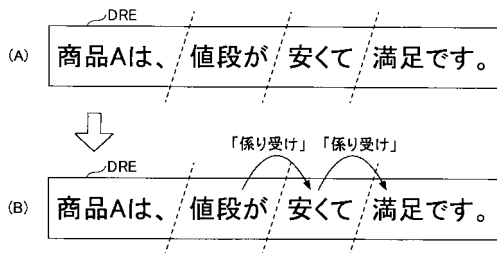
【 図 4 】



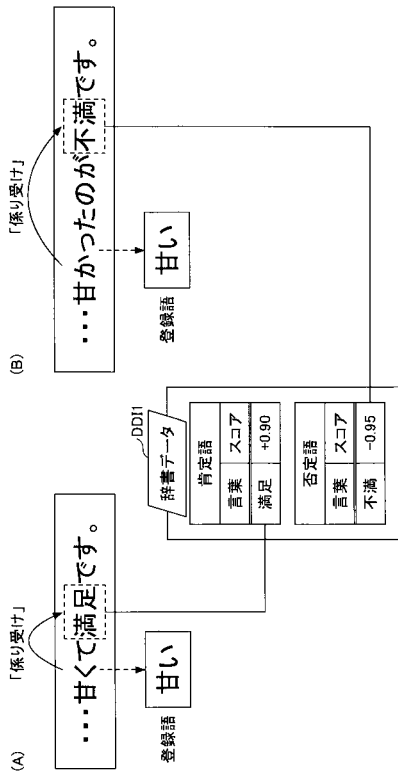
【図5】



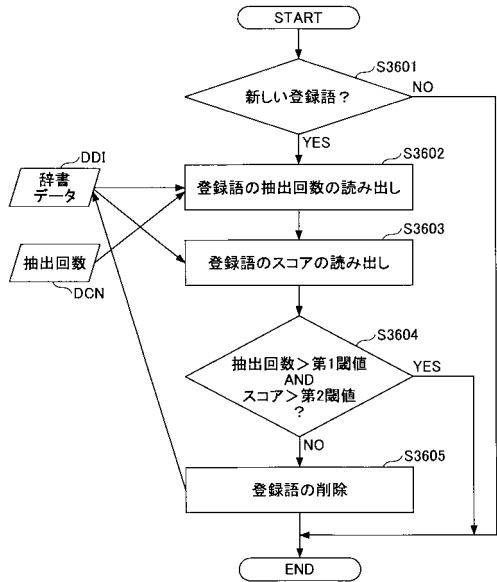
【図6】



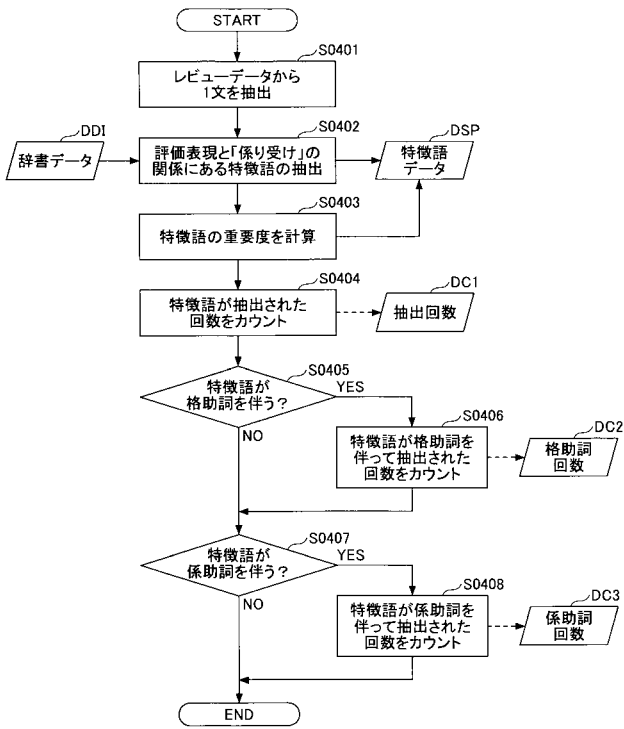
【図7】



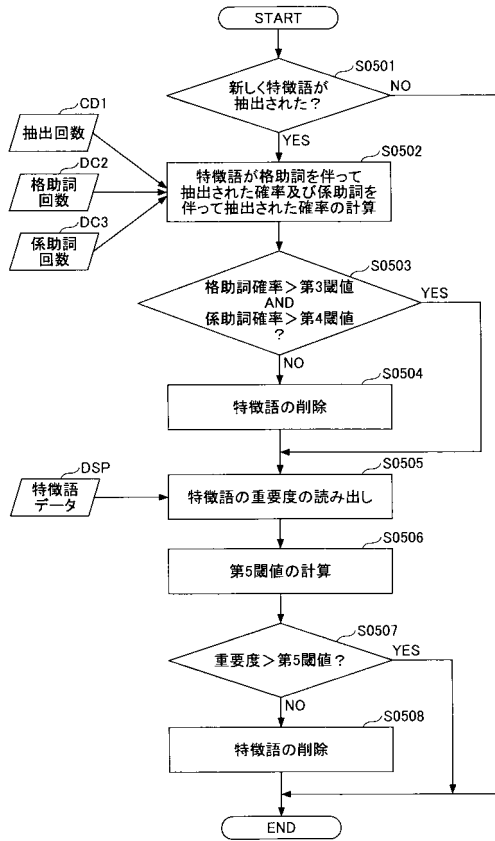
【図8】



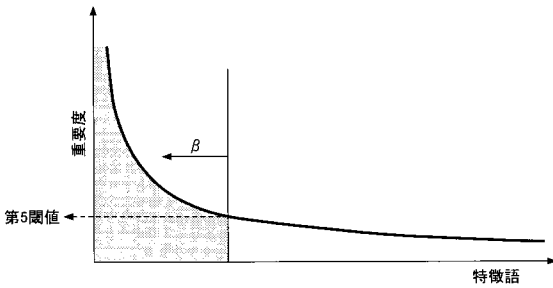
【 図 9 】



【 図 10 】



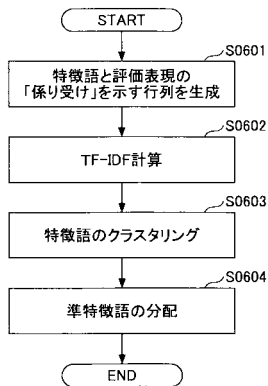
【 図 11 】



【 図 13 】

		評価表現			
		安い	静か	強い	満足
特徴語	値段	10	0	0	3
	価格	8	0	0	5
	音	0	10	0	0
	吸引力	0	0	9	5

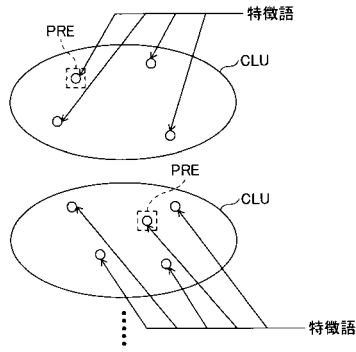
【 図 12 】



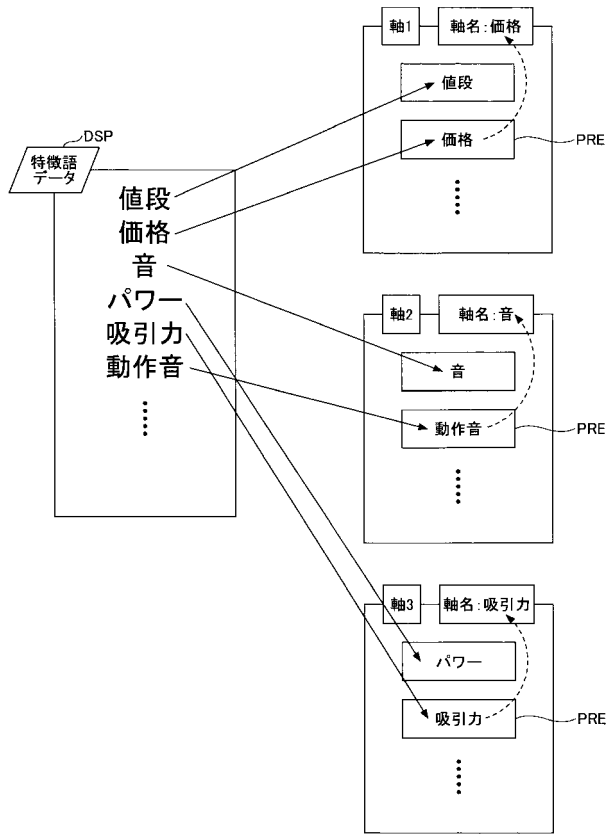
【 図 14 】

		評価表現			
		安い	静か	強い	満足
特徴語	値段	0.88	0	0	0.11
	価格	0.79	0	0	0.21
	音	0	1.0	0	0
	吸引力	0	0.0	0.90	0.10

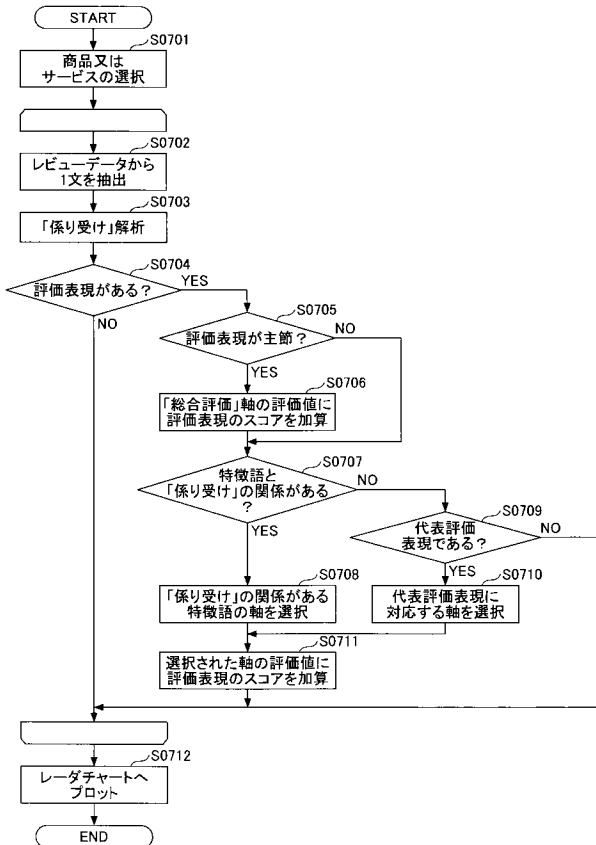
【 図 1 5 】



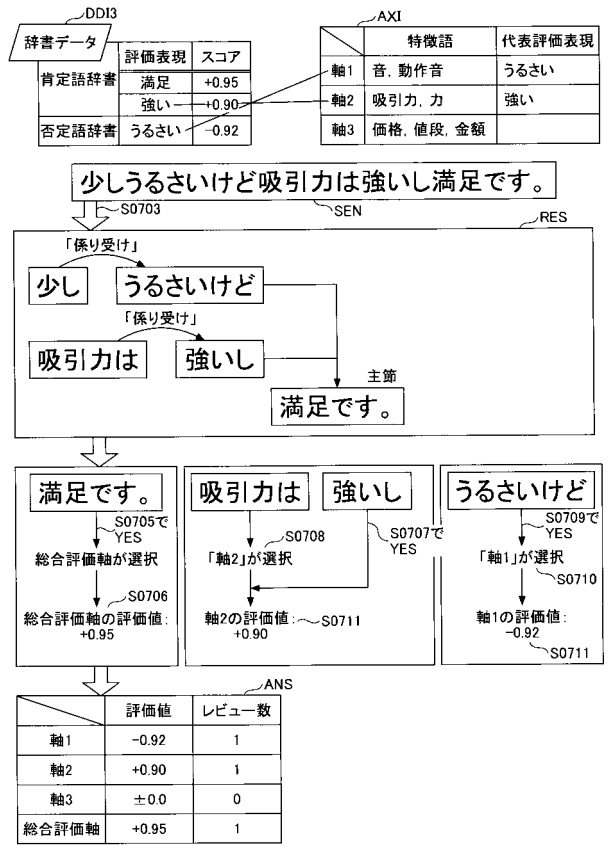
【 図 1 6 】



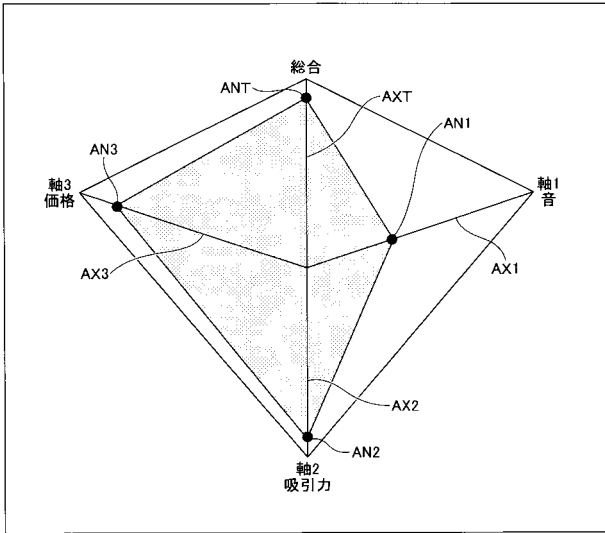
【 図 1 7 】



【 図 1 8 】



【図 19】



【図 20】

分類	カラム	内容についての補足
投稿者情報	投稿者	「user1」のようにマスクしたユーザ名
	年齢	「10歳代」等
	性別	[0:男 1:女 2:不明]
商品情報	商品コード	商品の主キー
	商品名	—
	店舗名	—
	商品URL	—
	商品ジャンルID	商品ジャンルの主キー
レビュー情報	商品価格	—
	購入フラグ	購入者かどうか
	内容	「実用品・普段使い」等の文字列
	目的	「自分用」等の文字列
	頻度	「はじめて」等の文字列
	評価ポイント	0から5までの6段階評価
	レビュータイトル	—
レビュー内容	—	
レビュー登録日時	—	—

【図 21】

ジャンル名	レビュー数
掃除機	78504
スーツ	93475
和菓子	78847

【図 22】

		設定した値
評価表現	第1閾値	10
	第2閾値	0.85
	α	0.8
特徴語	第3閾値	0.1
	第4閾値	0.1
	β	50[%]
軸	軸を削除するか否かを判断する所定の値	0.05
	準特徴語用判断値	0.5
	代表評価表現判定値	0.1

【図 24】

	評価表現
肯定表現	ありがたい, いい, おしゃれ, お洒落, カッコいい, かつこよい, かわいい, きれい, すばらしい, ぼつちり, びたつり, よい, イイ, オシャレ, カッコイイ, キレイ, コンパクト, シンプル, スムーズ, スリム, ハンディ, パワフル, ラッキー, 丁寧, 不要, 便利, 優秀, 十分, 可愛い, 可能, 大丈夫, 好き, 嬉しい, 安い, 安価, 安心, 小さい, 小型, 強い, 快適, 手ごろ, 手軽, 手頃, 抜群, 早い, 易い, 有り難い, 楽, 楽しい, 楽ちん, 気持ちいい, 満足, 簡単, 素敵, 素晴らしい, 綺麗, 良い, 薄い, 見やすい, 赤い, 軽い, 迅速, 透明, 速い, 静か, 面白い, 高性能
否定表現	うるさい, ごつい, しんどい, つらい, だかい, 不便, 不明, 不満, 億劫, 固い, 大きい, 大変, 太い, 残念, 無理, 煩い, 痛い, 辛い, 遅い, 邪魔, 重い, 重たい, 難しい, 高い

【図 23】

ジャンル名	肯定語	否定語
掃除機	67	24
スーツ	48	11
和菓子	51	14

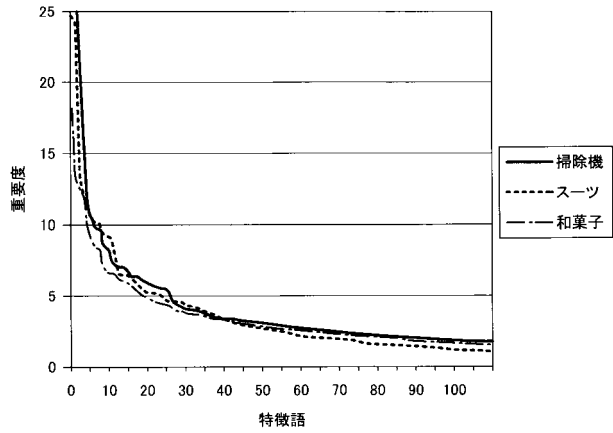
【図 25】

	評価表現
肯定表現	ありがたい, いい, うれしい, おしゃれ, お洒落, かつこよい, かつこよい, かわいい, きれい, すばらしい, はやい, ぼつちり, やすい, やわらかい, よい, イイ, オシャレ, カッコイイ, キレイ, 丁寧, 丈夫, 便利, 充分, 可愛い, 大丈夫, 嬉しい, 安い, 安心, 快適, 手ごろ, 手頃, 早い, 暖かい, 有難い, 柔らかい, 楽, 気持ちいい, 洗い, 満足, 素敵, 素晴らしい, 良い, 親切, 豊富, 軽い, 迅速, 速い, 駄目
否定表現	うすい, 不安, 不満, 地味, 残念, 派手, 短い, 薄い, 遅い, 重い, 雑

【図 2 6】

評価表現	
肯定表現	うまい、うれしい、おいしい、かりかり、かわいい、きれい、すごい、たまらない、ぴったり、みずみずしい、めずらしい、もったいない、やわらかい、よい、カリカリ、ヘルシー、丁寧、丁度、便利、優しい、冷たい、寒い、可愛い、大きい、大好き、安心、幸せ、懐かしい、手ごろ、手頃、早い、有名、有難い、柔らかい、楽しい、満足、濃い、珍しい、甘酸っぱい、程よい、綺麗、美味しい、美味しい、良い、親切、軟らかい、軽い、迅速、面白い、香ばしい
否定表現	くどい、しつこい、不安、不満、割高、大変、小さい、小さめ、微妙、普通、残念、物足りない、遅い、高い

【図 2 7】



【図 2 8】

商品ジャンル	特徴語
掃除機	音、使い勝手、吸引力、見た目、値段、動き、色、本体、評判、お手入れ、静音性音、お掃除、サイズ、容量、価格、ごみ捨て、パワー、滑り、組み立て、お値段、ヘッド、空気、コストパフォーマンス、コード、性能、充電時間、乾き、コードレス、形、取り扱い、動作音、力、腰、難点、レビュー評価、使い心地、部屋、紙パック、モーター音、調子、部分、階段掃除
スーツ	生地、着心地、肌触り、値段、サイズ、シルエット、質、品質、コストパフォーマンス、価格、評判、形、色、物、使い勝手、センス、見た目、質感、丈、ライン、お値段、種類、乾き、仕立て、袖、着丈、サイズ展開、アイロンかけ
和菓子	味、評判、甘さ、見た目、値段、あんこ、バランス、大きさ、餡、相性、サイズ、お値段、皮、食感、送料、香り、お餅、口当たり、風味、おもち、価格、ろう、形、あん、量、大福、パッケージ、クリーム、縁起、使い勝手、抹茶、外、色、単価、生地、甘み、箱、笹団子、部分、彩り、ここの、スイートポテト、塩加減、ウケ、腹持ち、商品、いちご、餅、栗、イチゴ、苺、抹茶餡、甘さ加減、種類、のど越し

【図 2 9】

	特徴語群	重要度合計	除去フラグ
評価軸1	モーター音、音、静音性音、動作音	66.4	F
評価軸2	使い勝手、評判、調子	38.7	F
評価軸3	吸引力、パワー、力	32.3	F
評価軸4	形、色、見た目	29.7	F
評価軸5	価格、値段、お値段	23.7	F
評価軸6	動き	10.4	T
評価軸7	ヘッド、サイズ、本体、部分、難点、容量、紙パック	40.0	F
評価軸8	ごみ捨て、お手入れ、組み立て、取り扱い	24.9	F
評価軸9	使い心地、滑り、性能、コードレス、お掃除、階段掃除	30.2	F
評価軸10	部屋、空気	9.09	T
評価軸11	レビュー評価、コストパフォーマンス	9.50	T
評価軸12	コード	5.49	T
評価軸13	乾き、充電時間	9.43	T
評価軸14	腰	3.83	T

【 図 3 0 】

	特徴語群	重要度合計	除去フラグ
評価軸1	生地	24.7	F
評価軸2	質感、肌触り、着心地	43.2	F
評価軸3	お値段、価格、値段	26.4	F
評価軸4	サイズ展開、種類、サイズ	20.8	F
評価軸5	シルエット、ライン、形、色	29.7	F
評価軸6	質、品質、見た目、物、コストパフォーマンス、センス	48.3	F
評価軸7	評判	8.97	T
評価軸8	使い勝手	6.41	T
評価軸9	丈、袖、着丈	14.7	F
評価軸10	乾き	5.12	T
評価軸11	仕立て	4.95	T
評価軸12	アイロンがけ	4.55	T

【 図 3 1 】

	特徴語群	重要度合計	除去フラグ
評価軸1	餡、あんこ、あん、大福、クリーム、商品、風味、笹団子、スイートポテト、食感、抹茶餡、味、大きさ、ここの、量、相性、ろう、香り、口当たり、種類、抹茶、のど越し	117	F
評価軸2	縁起、使い勝手、腹持ち、ウケ、評判、バランス	35.3	F
評価軸3	甘さ、甘み、甘さ加減、塩加減	21.6	F
評価軸4	見た目、形、箱、パッケージ、色、彩り	30.7	F
評価軸5	お値段、値段、価格、送料、単価	30.5	F
評価軸6	いちご、イチゴ、苺、栗、サイズ	18.0	F
評価軸7	生地、餅、お餅、おもち、部分、皮、外	29.9	F

【 図 3 2 】

	代表点	特徴語群
評価軸1	モーター音	おと、排気音、騒音、若干音、作動音
評価軸2	使い勝手	
評価軸3	吸引力	吸塵力、吸収力、威力、馬力
評価軸4	形	ピンク、姿
評価軸5	価格	金額、その他値段、料金、店、電気代
評価軸6	ヘッド	ヘッド部分
評価軸7	ごみ捨て	ごみ処理、取り出し、取り付け、使い方、ゴミ、ゴミ処理、後片付け
評価軸8	使い心地	

【 図 3 3 】

	代表点	特徴語群
評価軸1	生地	布地、記事、生地自体、厚み、秋冬物、パープル、たより生地
評価軸2	質感	
評価軸3	お値段	金額
評価軸4	サイズ展開	品数、バリエーション、品ぞろえ、カラーバリエーション
評価軸5	シルエット	状態
評価軸6	質	
評価軸7	丈	手、袖丈、腕、桁丈、股下、長さ

【 図 3 4 】

	代表点	特徴語群
評価軸1	餡	
評価軸2	縁起	口コミ、口溶け
評価軸3	甘さ	
評価軸4	見た目	包装紙、入れ物、巾着
評価軸5	お値段	金額
評価軸6	いちご	
評価軸7	生地	

【 図 3 5 】

	特徴語群	代表評価表現
評価軸1	モーター音、音、おと、排気音、静音性音、動作音、騒音、若干音、作動音	うるさい、大きい、静か
評価軸2	使い勝手、評判、調子	良い、よい、いい
評価軸3	吸引力、吸塵力、吸収力、威力、馬力、パワー、力	強い、十分
評価軸4	形、色、見た目、ピンク、姿	かわいい、可愛い
評価軸5	価格、値段、お値段、金額、その他値段、料金、店、電気代	安い、手頃、手ごろ
評価軸6	ヘッド、ヘッド部分、サイズ、本体、部分、難点、容量、紙パック	小さい、軽い
評価軸7	ごみ捨て、お手入れ、ごみ処理、取り出し、取り付け、組み立て、使い方、ゴミ、ゴミ処理、後片付け、取扱い	簡単、楽
評価軸8	使い心地、滑り、性能、コードレス、お掃除、階段掃除	快適、満足

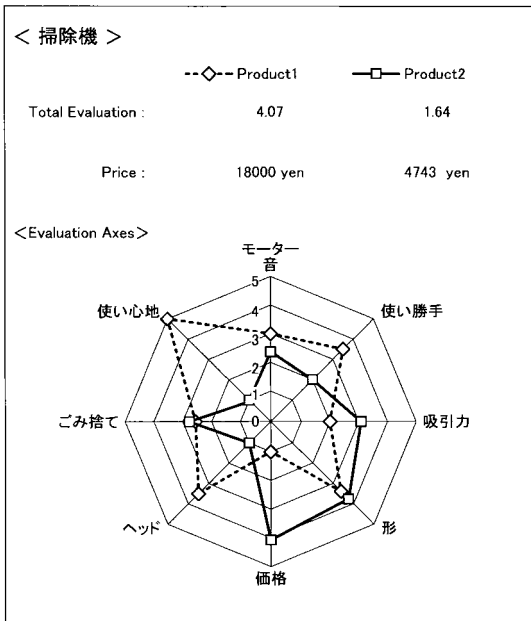
【図 3 6】

	特徴語群	代表評価表現
評価軸1	生地, 布地, 記事, 生地自体, 厚み, 秋冬物, パープル, たより生地	薄い
評価軸2	質感, 肌触り, 着心地	柔らかい, やわらかい
評価軸3	お値段, 価格, 値段, 金額	安い, 手頃
評価軸4	サイズ展開, 品数, バリエーション, 種類, 品ぞろえ, カラーバリエーション, サイズ	豊富
評価軸5	シルエット, ライン, 形, 状態, 色	きれいな, キレイ
評価軸6	質, 品質, 見た目, 物, コストパフォーマンス, センス	満足, 駄目
評価軸7	丈, 袖, 手, 袖丈, 腕, 着丈, 桁丈, 股下, 長さ	短い

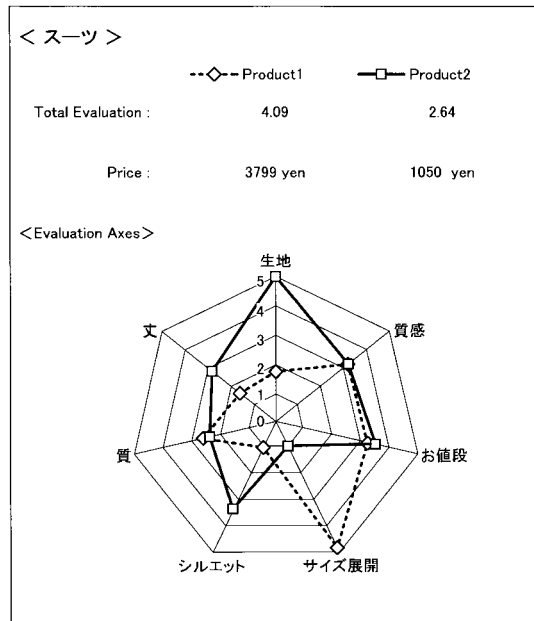
【図 3 7】

	特徴語群	代表評価表現
評価軸1	餡, あんこ, あん, 大福, クリーム, 商品, 風味, 笹団子, スイートポテト, 食感, 抹茶餡, 味, 大きさ, ここの, 量, 相性, ろう, 香り, 口当たり, 種類, 抹茶, のど越し	おいしい, 美味しい
評価軸2	縁起, 使い勝手, 腹持ち, ウケ, 口コミ, 口溶け, 評判, バランス	良い, よい
評価軸3	甘さ, 甘み, 甘さ加減, 塩加減	程よい, くだい, しつこい, 優しい
評価軸4	見た目, 包装紙, 形, 箱, パッケージ, 入れ物, 巾着, 色, 彩り	可愛い, かわいい, 綺麗, きれい
評価軸5	お値段, 値段, 価格, 金額, 送料, 単価	安い, 高い, 手頃, 手ごろ
評価軸6	いちご, イチゴ, 苺, 栗, サイズ	大きい, みずみずしい, 甘酸っぱい
評価軸7	生地, 餅, お餅, おもち, 部分, 皮, 外	やわらかい, 柔らかい, カリカリ

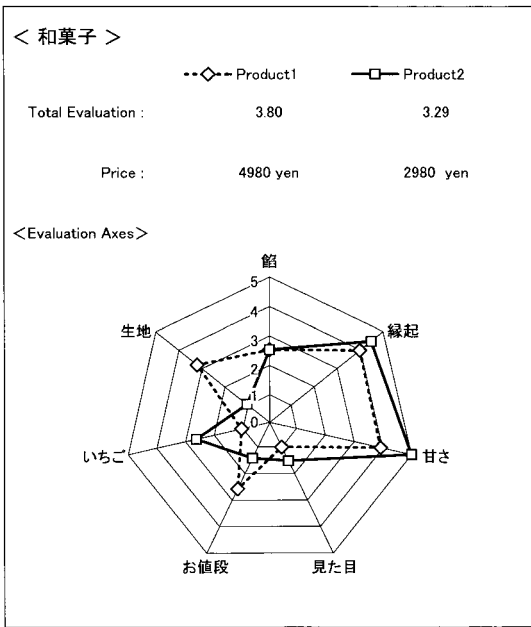
【図 3 8】



【図 3 9】



【図 40】



【図 41】

