

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第5817491号
(P5817491)

(45) 発行日 平成27年11月18日 (2015.11.18)

(24) 登録日 平成27年10月9日 (2015.10.9)

(51) Int. Cl.	F I
G 0 6 F 17/30 (2006.01)	G 0 6 F 17/30 2 2 0 Z
G 0 6 Q 30/02 (2012.01)	G 0 6 F 17/30 1 7 0 A
	G 0 6 F 17/30 3 8 0 E
	G 0 6 Q 30/02 1 5 0

請求項の数 6 (全 26 頁)

(21) 出願番号	特願2011-271245 (P2011-271245)	(73) 特許権者	000005496
(22) 出願日	平成23年12月12日 (2011.12.12)		富士ゼロックス株式会社
(65) 公開番号	特開2012-168925 (P2012-168925A)		東京都港区赤坂九丁目7番3号
(43) 公開日	平成24年9月6日 (2012.9.6)	(74) 代理人	100079049
審査請求日	平成26年11月21日 (2014.11.21)		弁理士 中島 淳
(31) 優先権主張番号	13/025,960	(74) 代理人	100084995
(32) 優先日	平成23年2月11日 (2011.2.11)		弁理士 加藤 和詳
(33) 優先権主張国	米国 (US)	(72) 発明者	フランシーン チェン
			アメリカ合衆国94304 カリフォルニア州 パロアルト ポータードライブ 3174 エフエックス パロアルトラボラトリーインク内

最終頁に続く

(54) 【発明の名称】 商品検索装置およびプログラム

(57) 【特許請求の範囲】

【請求項 1】

複数の商品に関する情報であって、少なくともそれぞれの商品に対するユーザーレビューを含む情報を分析して、個々の商品の用途と特性とを抽出する抽出手段と、
抽出した用途と特性とを関連付けて記憶する記憶手段と、
ユーザーによる用途の選択を入力する入力手段と、
選択された用途と関連付けて記憶された特性に基づいて、該選択された用途との関連の深さに応じて商品のランク付けを行うランク付け手段と、
上位にランク付けされた商品の一覧を出力する出力手段と、
を備える商品検索装置。

【請求項 2】

前記抽出手段は、ユーザーレビュー中の文を名詞列を類似するものをまとめるようにクラスタリングした上で、特徴を表す語句として予め定めた語句を含まないクラスターを除外することで、類似する商品の特徴をグループ化し、このグループ化した商品の特徴を商品の特性として用いることを特徴とする、

請求項 1 記載の商品検索装置。

【請求項 3】

前記抽出手段は、ユーザーレビュー中の文を主観的な文と客観的な文とに分類し、ユーザーレビュー中の文のうちの主観的な文のみに対して商品の特徴を抽出する処理を行うことを特徴とする、

請求項 2 記載の商品検索装置。

【請求項 4】

前記抽出手段は、ユーザーレビューに対して、用途として一般的に用いられる語句が特定のパターンで現れる句を抽出し、その句に含まれる名詞列から前記商品の特性を除外する処理を行って商品の用途として抽出することを特徴とする、

請求項 2 記載の商品検索装置。

【請求項 5】

前記記憶手段は、用途各々に関連付けて複数の特性をそれぞれ重み付けして記憶し、

前記ランク付け手段は、選択された用途に関連付けられた複数の特性を、それぞれの重み付けに応じて評価した上で商品のランク付けを行うことを特徴とする、

請求項 1 に記載の商品検索装置。

【請求項 6】

複数の商品に関する情報であって、少なくともそれぞれの商品に対するユーザーレビューを含む情報を分析して、個々の商品の用途と特性とを抽出する抽出ステップと、

抽出した用途と特性とを関連付けてデータベースに記憶する記憶ステップと、

ユーザーによる用途の選択を入力する入力ステップと、

選択された用途と関連付けて前記データベースに記憶された特性に基づいて、該選択された用途との関連の深さに応じて商品のランク付けを行うランク付けステップと、

上位にランク付けされた商品の一覧を出力する出力ステップと、

をコンピュータに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】

商品を検索してユーザーに提示する装置及びプログラムに関する。

【背景技術】

【0002】

非特許文献 1 には、ウェブ上に存在する製品に対するレビューから、その製品の特徴やその製品に対するユーザーの評価を抽出する技術が記載されている。

【0003】

非特許文献 2 には、複数のユーザーの意見を統合して、一つの評価値を計算する技術が記載されている。

【0004】

非特許文献 3 及び 4 には、ウェブ上のテキストから教師なし学習によってデータ集合を抽出する方法が記載されている。

【0005】

非特許文献 5 には、自己相互情報量 (Point-wise Mutual Information, PMI) とサポートベクターマシン (SVM) を用いたレビューの教師なし分類法が記載されている。

【0006】

非特許文献 6 には、文書中の全ての単語およびその全ての依存関係を用いて、2 つの単語間の類度を算出する技術が記載されている。

【0007】

非特許文献 7 には、文を主観的か否かで分類し、主観的と分類された文のみを用いて評価情報分析を行う例が記載されている。

【0008】

非特許文献 8 には、文書の検索用インデックスとして、検索用クエリーのセット、すなわちクエリーとその検索結果の文書のセットを用いる文書検索方法が記載されている。

【先行技術文献】

【非特許文献】

【0009】

10

20

30

40

50

【非特許文献1】ポペスキュー他(Popescu et al.)、「レビューに基づいた製品の特徴および意見の抽出(Extracting Product Features and Opinions from Reviews)」、人間の言語に関する技術会議予稿集(Proceedings of Human Language Technology Conference)、2005年、p.339-346

【非特許文献2】フー他(Hu et al.)、「カスタマーレビューからの意見特徴の検索(Mining Opinion Features In Customer Reviews)」、人口知能に関する国内会議予稿集(Proceedings of the National Conference on Artificial Intelligence)、2004年、p.755-760

10

【非特許文献3】イエイツ他(Yates et al.)、「ウェブ上のオブジェクトと関係性の教師なし分解(Unsupervised Resolution of Objects and Relations on the Web)」、コンピュータ言語学会北米支部会議予稿集(Proceedings of NAACL-HLT)、2007年、p.121-130

【非特許文献4】エツィオニ他(Etzioni et al.)、「ウェブからの教師なし固有表現抽出：実験的研究(Unsupervised Named-Entity Extraction From the Web: an Experimental Study)」、人工知能(Artificial Intelligence)165(1)、2005年、p.91-134

20

【非特許文献5】ターニー(Turney PD)、「賛成？反対？：レビューの教師なし分類への意味的オリエンテーションの適用(Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews)」、コンピュータ言語学会第40回年次会議予稿集(Proceedings of the 40th Annual Meeting of Association for Computational Linguistics)、2002年、p.417-424

【非特許文献6】リン(D. Lin.)、「類似語の自動取得とクラスタリング(Automatic retrieval and clustering of similar words)」、コンピュータ言語学会第17回国際会議予稿集(Proceedings of the 17th International Conference on Computational Linguistics)、1998年、p.768-774

30

【非特許文献7】パン及びリー(B. Pang and L. Lee)、「意見検索と評価情報分析(Opinion Mining and Sentiment Analysis)」、情報取得における基礎と傾向(Foundations and Trends in Information Retrieval)、vol.2、2008年、p.1-135

【非特許文献8】ピケنز他(Pickens et al.)、「フィードバックおよび拡張のための反転インデキシング(Reverted Indexing for Feedback and Expansion)」、情報および知識管理に関する第19回ACM会議予稿集(Proceedings of the 19th ACM international conference on Information and knowledge management)、2010年、p.1049-1058

40

【発明の概要】

【発明が解決しようとする課題】

【0010】

商品に関する技術的な知識が無くともその商品の用途から所望の商品を検索できるよう

50

にする商品検索装置を提供することを目的とする。

【課題を解決するための手段】

【0011】

本発明の第1の態様としては、複数の商品に関する情報であって、少なくともそれぞれの商品に対するユーザーレビューを含む情報を分析して、個々の商品の用途と特性とを抽出する抽出手段と、抽出した用途と特性とを関連付けて記憶する記憶手段と、ユーザーによる用途の選択を入力する入力手段と、選択された用途と関連付けて記憶された特性に基づいて、該選択された用途との関連の深さに応じて商品のランク付けを行うランク付け手段と、上位にランク付けされた商品の一覧を出力する出力手段と、を備える商品検索装置が提供される。

10

【0012】

本発明の第2の態様としては、第1の態様においてさらに、前記抽出手段は、ユーザーレビュー中の文を名詞列を類似するものをまとめるようにクラスタリングした上で、特徴を表す語句として予め定めた語句を含まないクラスターを除外することで、類似する商品の特徴をグループ化し、このグループ化した商品の特徴を商品の特性として用いてもよい。

【0013】

本発明の第3の態様としては、第2の態様においてさらに、前記抽出手段は、ユーザーレビュー中の文を主観的な文と客観的な文とに分類し、ユーザーレビュー中の文のうちの主観的な文のみに対して商品の特徴を抽出する処理を行ってもよい。

20

【0014】

本発明の第4の態様としては、第2の態様においてさらに、前記抽出手段は、ユーザーレビューに対して、用途として一般的に用いられる語句が特定のパターンで現れる句を抽出し、その句に含まれる名詞列から前記商品の特性を除外する処理を行って商品の用途として抽出してもよい。

【0015】

本発明の第5の態様としては、第1の態様においてさらに、前記記憶手段は、用途各々に関連付けて複数の特性をそれぞれ重み付けして記憶し、前記ランク付け手段は、選択された用途に関連付けられた複数の特性を、それぞれの重み付けに応じて評価した上で商品のランク付けを行ってもよい。

30

【0016】

本発明の第6の態様としては、複数の商品に関する情報であって、少なくともそれぞれの商品に対するユーザーレビューを含む情報を分析して、個々の商品の用途と特性とを抽出する抽出ステップと、抽出した用途と特性とを関連付けてデータベースに記憶する記憶ステップと、ユーザーによる用途の選択を入力する入力ステップと、選択された用途と関連付けて前記データベースに記憶された特性に基づいて、該選択された用途との関連の深さに応じて商品のランク付けを行うランク付けステップと、上位にランク付けされた商品の一覧を出力する出力ステップと、をコンピュータに実行させるためのプログラムが提供される。

【発明の効果】

40

【0017】

本発明の第1及び第6の態様によれば、商品に関する技術的な知識が無くとも所望の商品を検索することができる。

【0018】

本発明の第2の態様によれば、類似する商品の特徴を別々に扱うことによるランク付けの不具合を回避できる。

【0019】

本発明の第3の態様によれば、主観的な文及び客観的な文の両者に基づいて商品特徴を抽出する場合に比べて、精度の高い商品特徴の抽出が可能となる。

【0020】

50

本発明の第４の態様によれば、商品の用途をユーザーレビューから自動的に抽出することが可能となる。

【００２１】

本発明の第５の態様によれば、商品の用途に対応する特性ごとの重みに応じた商品のランク付けが可能となる。

【図面の簡単な説明】

【００２２】

【図１】未加工の商品情報データを分析してユーザーインターフェースを生成するシステムの一例を示すブロック図である。

【図２】頻出名詞列のクラスタリングおよびフィルタリングを行う方法の一例を示すフローチャートである。

【図３】ベータ二項分布モデルを用いて、商品の特性に関する意見を抽出する方法の一例を示す図である。

【図４】ユーザーレビューから要約文を選択する方法の一例を示すフローチャートである。

【図５】商品の用途を特定する方法の一例を示すフローチャートである。

【図６】用途に関する質問に対する回答を促すグラフィカルユーザーインターフェース（GUI）の一例を示す図である。

【図７】関連する商品のリストと対応する特性を表示しユーザーの操作を受け付けるGUIの一例を示す図である。

【図８】上位にランク付けされた商品、および、選択された商品の詳細な情報（例：仕様、用途、レビュー例）を表示するGUIの一例を示す図である。

【図９】特性の重みを操作するために用いるユーザーインターフェースの一例を示す図である。

【図１０】並行座標を用いて商品の特性値を表示する比較用インターフェースの一例を示す図である。

【図１１】用途を基礎にしたユーザーインターフェースを使用する方法の一例を示すフローチャートである。

【図１２】本発明が実装されるコンピュータシステムの一例を示すブロック図である。

【発明を実施するための形態】

【００２３】

商品の一覧、検索または比較を支援するための商用システムは既に数多くあるが、これらのインターフェースは一般に、商品を主体としたもの、すなわち商品情報を一覧表示するようなものとなっている。一方で、商品購入の意思決定に用いる商品情報は、公式なものおよび非公式なものを含め、これまでに無いほどの量がインターネット上にあふれている。

【００２４】

ここで、公式な商品情報とは、その商品を提供する者（生産者や販売者）がその商品に関して公式に発表した情報であり、その商品の特性、仕様、設定および価格に関する情報を含むものである。また、ここでの非公式な商品情報とは、一般消費者がその商品について作成した情報（消費者商品情報）のことを意味しており、いわゆるユーザーレビューを含むものである。また、このユーザーレビューは、商品に対する評価や意見をその商品のユーザーが表明するべく作成した文章等の情報であり、その商品に関して公式商品情報では触れられていない追加的な情報、またはその商品の機能や、有用性や、特定用途に対する適合性等に関する意見が含まれる。また、ユーザーが行うその商品に対する評価（商品全般に関するもの、および特定の特性に関するもの）もこれに含まれる。このような評価によって、潜在的な顧客は、自分が重要だと考える特性について既存顧客がどのように評価しているかを知るのである。

【００２５】

以下、商品情報を分析を行い、商品の用途を基礎にした商品検索を行うシステム及び方

10

20

30

40

50

法の例を説明する。ユーザーの要求に合致した上位レベルの商品情報に主眼を置いた、カスタマイズ可能なユーザーインターフェースを提供するべく、あらゆる種類の商品情報、例えば仕様、属性、ユーザーレビューが、商品の特性および用途を決定するべく利用される。

【 0 0 2 6 】

ここでは、仕様、属性および商品の特徴（商品特徴）を含む、その商品の性質を現す情報を「特性」と総称する（以下、商品の特性の意味で商品特性とも言う）。また、ユーザーがその商品で何を行うのか、または、その商品を使用しているときにどんな種類の活動をしているのかを表す用語として「用途」を用いる。

【 0 0 2 7 】

まず、本実施例のシステムの概要を説明する。

【 0 0 2 8 】

ユーザーには、商品の用途を選択するためのグラフィカルユーザーインターフェース（GUI）が、個々の特性の重み、すなわちユーザーにとっての重要性、を調整する領域とともに提供される。この特性に対する重み付けは、ユーザーによる用途の選択に伴って自動的に行われるようにしても良い。これは、特定の用途に対して、関連する特性およびその個々の重みを予め決定してこれを記憶しておくことによって可能となる。

【 0 0 2 9 】

その後、これらのユーザーからの情報（用途や特性の重み）に基づいて商品のランク付けを行い、それぞれの商品の関連情報を表示する。ここで、ユーザーが特性の重みを調整すると、その重み付けに応じて再度ランク付けが行われる。また、ユーザーによって手動で、またはシステムによって自動で選択された複数の商品を、比較用に表示することもできる。

【 0 0 3 0 】

本システム及びインターフェース（UI）は、用途を中心としたものである。したがって、ユーザーは最初に、その商品の想定される使用状況に関する質問に答えることになる。これを受けてシステムは、ユーザーの要求に合致した種類の商品をGUIに表示し、ユーザーが興味を示した種類の用途に関連する上位レベルの商品特性を併せて表示する。

【 0 0 3 1 】

このUIを操作していくうちに、ユーザーは上位レベルの商品特性と具体的な商品特徴との間の関係に気付くことになる。このように、下位レベルの商品の詳細ではなく上位レベルのユーザーの目的に重きをおくという意味で、本システムの手法は、従来の商品検索のちょうど逆の発想のものである。ここでは、生の（未加工の）商品情報、すなわちユーザーレビューや商品仕様等から、システムで使用する上位レベルの特性を抽出するべく、半自動の手法を採用する。

【 0 0 3 2 】

この方法は大まかに言えば、（１）商品特徴を特定および分類し、（２）商品に対するレビューからその商品特徴に関する意見を分析して要約し、（３）特定された特徴から商品の用途を特定するものである。

【 0 0 3 3 】

システムは、仕様データ、属性データおよびユーザーレビューを前処理して、それぞれの商品の商品特性および候補となる用途を抽出する。この抽出されたデータをデータベースに格納し、GUIアプリケーションによってアクセスできるようにする。このGUIによって、特性の重み付けをおこなうための一連の質問をユーザーに投げかける。ここで決定された重みは、商品のランク付けに用いられることになる。このような処理の後に、ユーザーに対して、商品の詳細情報にアクセスしたり、商品の比較をしたり、重みを変更したりすることを許可する。上位レベルの簡単な質問による特性の重み付けと、重み付けに対するより詳細な直接操作のこの組み合わせは、ユーザーに対して複雑な操作や高度な技術知識を求めることができないような状況（たとえば店頭での操作端末による商品検討）、および一般的な状況（たとえばウェブブラウザによる商品検討）の両者に対してこのGUI

10

20

30

40

50

I が利用可能である。

【 0 0 3 4 】

この GUI は、商品の技術仕様だけに着目した検索ではなく、その商品の日常の使用形態に着目した検索が出来るように、様々な種類の商品情報を統合する。このような検索を可能とするために、例えば、ユーザーの意見（レビュー）から商品特徴を抽出し、これを上位レベルの用途と結びつける処理を行うのである。また、上位レベルの特徴（ハイキングや結婚式でカメラが使われるか等）を用いて、上位レベルの商品の用途を下位レベルの特徴及び仕様（価格、画素数等）と統合する。技術的な特徴が特定されると、実際のユーザーによって報告された用途にこの特徴を当てはめるのである。

【 0 0 3 5 】

ここで例として説明するシステム及び方法では、直接（重みを操作する GUI 部品によって）、間接（上位レベルの用途から推測された重みによって）を問わず対話形式で商品のランク付けを行うインターフェースと、データの抽出処理とを組み合わせる。

【 0 0 3 6 】

次に、図を用いて、より詳しい説明を行う。

【 0 0 3 7 】

図 1 に示すように、システム 1 0 0 は大きく分けて 3 つの要素を含んでいる。すなわち、（ 1 ）前処理のステップの大半を行う抽出ユニット 1 0 2、（ 2 ）生データ及び抽出したデータを格納するデータベース 1 0 4、（ 3 ）リアルタイムのユーザーインターフェースユニット 1 0 6 である。この前処理では、全てのデータをデータベース 1 0 4 に格納する。格納後のデータに対しては、ユーザーインターフェースユニット 1 0 6 がこれにアクセスし、ユーザーに表示するグラフィカルユーザーインターフェース（ GUI ）を生成する。ここには図示していないが、ユーザーインターフェースユニット 1 0 6 は、ディスプレイ、マウス、タッチスクリーンといったユーザーとのやり取りを行う装置に接続される。

【 0 0 3 8 】

また、図 1 から分かるように、このシステムは用途に基づいて商品情報を生成するが、その方法はまず、抽出ユニットが仕様、属性およびユーザーレビューから、生の商品情報データを取得する（ S 1 2 2 ）。次に、この生の商品情報データを分析して、特性と用途を抽出する（ S 1 2 4 ）。次に、この特性を対応する用途に対応付ける（ S 1 2 6 ）。この対応付けの処理は、詳しくは後述するが、抽出ユニット 1 0 2 とは独立して、手動または半自動で行う。そして、抽出されたデータを、データベース 1 0 4 に格納し、ユーザーインターフェース 1 0 6 からアクセスできるようにする（ S 1 2 8 ）。このユーザーインターフェース 1 0 6 は、格納されたデータを取り出して、個々の特性と用途との間の関係に基づいて、その特性の重みを提供する（ S 1 1 0 ）。その後、用途がシステムまたはユーザーによって選択される（ S 1 1 2 ）。次に、選択された用途に対応する特性の重みに基づいて商品をランク付けし（ S 1 1 4 ）、ランク付けした商品をランクに応じてディスプレイ上に表示する（ S 1 1 8 ）。

【 0 0 3 9 】

ここで、各種特性の重みの直接操作や選択した用途の変更をユーザーに対して許可して（ S 1 1 6 ）、この結果としての更新された商品リストを表示する（ S 1 1 8 ）ようにしてもよい。

【 0 0 4 0 】

以下に、インターフェースユニット 1 0 6 が、以下のようなデジタルカメラの商品情報を利用する例を示す。もちろんこれに限らず、種々の商品情報データも利用可能である。

【 0 0 4 1 】

（ 1 ）仕様

最大ズーム倍率、最大画素数、重量といった標準的な商品仕様のことである。

【 0 0 4 2 】

（ 2 ）ユーザーレビューから得られる商品特徴

商品の特徴は、一般ユーザーが作成した文章によるレビューから作成される。これは、例えば、日常的な使用に耐えられる耐久性があるか、顔検出機能はうまく機能するか、それともただの邪魔な機能が等、標準的な仕様から得られる情報以上のものを与えてくれるものである。

【 0 0 4 3 】

この特徴は、後述するように、表現のバリエーションを吸収すべくグループ分けしてもよい。また、これも後述するように、それぞれの特徴に対する意見を表示するようにしても良い。

【 0 0 4 4 】

(3) ユーザーレビューにおいて評価された商品の属性

商品属性は、一般にはフォーマットの定まっていない文章から抽出されるが、ここでは、その属性に対する評価が明示的に選択されているものがその対象である。一方で商品特徴は、文脈から（例えば形容詞を抽出する等して）推論しなければならないものであり、この点で両者は異なる。

【 0 0 4 5 】

(4) レビューから抽出された用途

用途としては、例えば以下のようなものが挙げられる。(a) その商品をしようする際に行っている活動の種類（写真を撮る際に何を行っているか。泳いでいる、山に登っている等）。(b) どのようにその商品を用いるのか（何の写真を撮るのか）。(c) その商品が何のために使われるのか（撮った写真を何に使うのか）。

【 0 0 4 6 】

ここで、(b) は、多くの商品においてユーザーに具体例を選択させることによって特定可能である。カメラの例で言えば、ユーザーが選択した写真のサンプルから、そのユーザーが撮影する写真の種類を特定することができる。同様に、オフィスソフトウェアの場合には、ユーザーに作成したファイルの種類を選択させることで、特定可能である。用途は、仕様、特徴および属性を含む一つ以上の特性と関連付けられる。例えば、ハイキングという用途は、サイズや重量といった仕様、耐久性のような特徴、そして製造品質のような属性を含む特性と関連付けられる。

【 0 0 4 7 】

I . データ抽出および分析

商品に対するレビュー、その商品の仕様および属性といった生の商品情報から、特性と用途を抽出するためのデータ分析について、以下に説明する。

【 0 0 4 8 】

生の商品情報は、例えば、アマゾン（登録商標）のようなインターネット上のウェブサイトに掲載されているレビューデータ等から入手可能である。

【 0 0 4 9 】

I - 1 . 信頼できる商品特徴の抽出

ここにおける商品の「特徴」とは、ユーザーレビューに明確に記載された、その商品の部品および性質である。商品特徴の候補を特定するためには、ウェブを対象にした傾向分析による高精度のデータ抽出法を用いることができる。例えば、非特許文献 3 や非特許文献 4 に記載されたような、教師なし学習法によってデータ集合を抽出する方法である。そして、これらのデータ集合抽出法は、非特許文献 1 に記載されているような、商品特徴抽出法と組み合わせられる。具体的には、パターンを使って名詞句の特徴候補を特定する。そして、機械学習等の統計処理によって信頼できる商品特徴を特定するのである。

【 0 0 5 0 】

ここで、商品特徴の抽出プロセスを一例を以下に示す。なお、付加的な自然言語処理ステップ (5) は、商品レビューとして頻繁に利用されるであろう小規模スケールのデータを補うためのものである。

【 0 0 5 1 】

1) 商品特徴になり得るか否かの観点で、適当な単語と不適当な単語の例を手動でリス

10

20

30

40

50

ト化する。例えば、レンズ、ズーム、画質等は、カメラの商品特徴の候補として適当であるが、娘、クリスマス、休日等は不適当である。

【0052】

2) 適当と判断された単語を用いて、ユーザーレビューから商品特徴の種となる単語を見つけ、その単語を含んで前後にずらした4単語までのパターンを抽出する。例えば、lens (レンズ) という種が "The lens scratches easily (レンズが傷つき易い)" という文中に見つかった場合、次のようなパターンが抽出される。なお、NP は名詞句を表している。

- ・The NP scratches easily .
- ・The NP
- ・NP scratches easily .
- ・NP scratches

10

【0053】

3) 抽出したパターンの確からしさを計算する。そのパターンにおいて、肯定的な例の出現頻度が否定的な例の出現頻度に対して大きければ大きいほど、確かであると判断する。

【0054】

4) ユーザーレビュー全てに対して、確からしさの上位500のパターンとマッチする単語列(連続する複数の単語)を抽出し、名詞列に対応する部品または性質を商品特徴の候補として抽出する。名詞列は、"Stanford Log-linear Part-Of-Speech Tagger" (nlp.stanford.edu/software/tagger.shtml) 等の品詞分析器を用いて特定することができる。

20

【0055】

5) ウェブベースの自己相互情報量(Point-wise Mutual Information, PMI)とサポートベクターマシン(SVM)を用いて信頼できる商品特徴を選択する。これには、例えば非特許文献5に記載されている手法が適用できる。個々の商品特徴の候補に対して、SVMに渡される特徴量ベクトルの要素は、ウェブを利用したPMI統計量であり、これは、識別句として "<商品> features <候補>" および "<商品> has <候補>" を用いて求められる。具体的には、"camera has lens" や "camera features optical zoom" 等である。

30

【0056】

I-2. 用語の類似度とグループ化

名詞句の類似度を計算するためには、例えば、非特許文献6に記載の方法を用いることができる。一方で、本システムは単語の集まりである句を取り扱うが、その種類は必然的に単語の種類に比べてはるかに多いので、コーパス中の他の全ての単語および句を考慮せずとも句同士の類似度を計算できるようにすることが、計算コストの観点から重要である。そこで、後述するように、本実施例においては、非特許文献6を簡略化した方法を用いる。

40

【0057】

また、分布が類似する(類似した文脈で出現する)単語のグループを後処理する方法として、例えば以下のような2つの方法がある。(1) 2つの単語が近接して出現するとして検索された数と、2つの句("from X to Y"、"either X or Y")中にその2つの単語が出現する回数との比を計算する(非特許文献6に記載)。(2) 2カ国語間の対訳辞書を使う。2カ国語のコーパスの使用も可能である。

【0058】

しかしながら、本実施例では、先に求めた信頼できる特徴や予め定められた属性を用いる。例えば、AMAZON.COM(登録商標)のサイトにおける属性(attribute)は、カスタマーレビューのページの冒頭に表示される商品特徴であって、ユーザー

50

が個々の特徴に対して星1つから星5つまでの評価を行うものである。通常は10個以内の属性がリスト表示されるが、個々の商品によって異なる。とあるカメラの場合では、画質、大きさと重量、使い易さ等が挙げられている。構造化されていないテキストから評価を推論する場合と異なり、このような属性の平均評価は、生のデータから直接（テキスト分析による推論を経ずとも）抽出することができる。

【0059】

I-3. 商品特徴のグループ化

商品特徴の基本セットが特定された後には、クラスタリングによって商品特徴（信頼できる商品特徴を含む）の同義グループへのグループ化を行う。ここで、同義グループとは、同一の部品や性質に対してユーザーが言及する際の種々の表現を、まとめて一つのグループとしたものである。

10

【0060】

信頼できる特徴は直接クラスタリングできるが、頻出する名詞列（近接して出現する名詞の集合）をクラスタリングして、クラスター中の名詞列のフィルタリングに信頼できる特徴を用いる方が、よりよい結果が得られるであろう。

【0061】

この処理の概要を図2を用いて説明する。

【0062】

まず、抽出ユニットが、アマゾン（登録商標）等のウェブサイト上のユーザーレビューから、ユーザーレビューデータを取得する（S202）。次に、そのレビュー中の文を品詞分解して、名詞列を抽出する（S204）。次に、上述したような信頼できる特徴の抽出処理（S206）と並行して、名詞列のクラスタリングを行う。そのためにまず、頻出する名詞列の全てのペアに対する類似度を、それらの形容詞的修飾語句がいかに類似しているかという観点に基づいて計算する（S208）。すなわち、形容詞的修飾語句の類似性が高ければ、類似度も高くなるようにするのである。これは、非特許文献6に記載された方法を簡略化したものである。当該文献記載の方法は、文書中の全ての単語およびその全ての依存関係を用いて、2つの単語間の類度を算出している。しかしながら本実施例においては、単語同士ではなく名詞列同士の類似度を算出しており、また、全ての単語の依存関係ではなく、形容詞的修飾語句の依存関係のみを用いている。とりわけ、形容詞と名詞と間の関係として2つの種類のみを考慮している。すなわち、"brilliant sunset（素晴らしい夕日）"に見られるような直接修飾関係（名詞を限定的に修飾する関係）、および、"The block was yellow（ブロックは黄色い）"に見られるような動詞を通じた修飾関係（形容詞が主語を修飾する関係）の2つである。

20

30

【0063】

次に、名詞列が出現するレビューデータ中の個々の文について、対応する形容詞的修飾語句、および形容詞と名詞列の関係を、解析木を用いて抽出する。ここで、修飾関係が与えられたときに、句および形容詞が条件付きで独立だと仮定する。すると、名詞列N、形容詞Aおよび、名詞列と共起する形容詞との間の修飾関係Rは以下のように表せる。

【数1】

40

$$P(R)P(N|R)P(A|R)$$

また、Rで関係付けられたNとAの間の相互情報量 $I(N, R, A)$ は以下のように表せる。

【数2】

$$I(N, R, A) = \log \left(\frac{P(N, R, A)}{(P(R)P(N|R)P(A|R))} \right)$$

50

又は、

【数 3】

$$I(N, R, A) = \log \left(\frac{\|n, r, a\| \times \|*, r, *\|}{\|n, r, *\| \times \|*, r, a\|} \right)$$

r が関係で a が形容詞である場合に、T(w) を (r, a) の組と定義する。この場合、I(n, r, a) は正である。2つの名詞列 n1 および n2 の間の類似度は、以下の式で計算できる。

10

【数 4】

$$sim = \frac{\sum_{(r,a) \in T(\omega_1) \cap T(\omega_2)} (I(n_1, r, a) + I(n_2, r, a))}{\sum_{(r,a) \in T(\omega_1)} I(n_1, r, a) + \sum_{(r,a) \in T(\omega_2)} I(n_2, r, a)}$$

【0064】

頻出する（実施例では、100万強の文に対して50回を閾値とした）名詞列のペア全てにおいて計算された類似度を、クラスタリングに用いる（S210）。これには、種々のクラスタリングアルゴリズムが採用可能であるが、本実施例では、完全連結凝集法（complete-linkage agglomerative clustering）を用いて、名詞列をコンパクトに保つようにする。そして、別途設定した閾値を用いて階層木をクラスターに分割する。また、絞込みステップS214においては、まず、クラスターに対して、S206で信頼できるものとして特定された名詞列のみを残すフィルタリング処理を行う（S212）。

20

【0065】

以上のような処理の結果として自動生成されたクラスターで総合頻度が最も大きいもののリストの例は以下のようなものである。

camera, body;
photos, pics, pictures and shots;
battery life, photo quality, quality, picture quality, image quality;
zooms, zoom;
screen, lcd, view screen, lcd screen, lcd display, display;
lens, lenses;
image shot, picture;
bang, deal, value, job;
settings, setting;
battery, batteries;

30

40

【0066】

これから分かるように、商品としてカメラを用いた本実施例においては、最大規模のクラスターの大半は、カメラに関係の深い話題に関するものであるが、“bang（「バン」という音）”、“deal（取引）”、“value（価値）”、“job（仕事）”等、一見関係の無さそうなものも挙がっている。これらは、予め定めた属性（AMAZON.COM（登録商標）のサイトで評価対象となっているもの等）を含むクラスターのみを残すようにフィルタリングすることで、削除するようにしてもよい。フィルタリングの他の手法の例としては、「良い商品特徴」をウェブベースのPMIを用いてフィルタリングする方法がある。また、前述した信頼できる商品特徴も、「良い商品特徴」の例である

50

。

【 0 0 6 7 】

I - 4 . 意見検索

このようにして自動的に特定した商品特徴に対して、その極性、すなわち肯定的か否定的かを、意見検索を用いて評価する。これは、様々な粒度で行うことができる。たとえば、特徴として採用できそうな全ての意見を特定した上で、個々の極性を統合してスコアを算出しようとする場合には、より細かい粒度の特徴を取り扱う。この意見スコアと、特徴の重みと、他の属性に対するスコア（または評価）とを組み合わせることで商品のスコアを算出し、そのスコアに基づいて商品をランク付けする。

【 0 0 6 8 】

用途の抽出法については、後述する。

【 0 0 6 9 】

用途と特性の関連付けは、手動で行ってもよい。しかしながら、用途について触れているレビューから、属性値を特定したり、またその商品（カメラ）に関する活動をその用途と属性値とともに生成したりすることもできる。

【 0 0 7 0 】

個々の意見を1つのスコアに統合することは、評価情報分析において一般的に行われていることである。しかしながら、これまで知られている手法では、満足のいく評価値を得ることができない。なぜならば、これらは、真の評価情報に対する信頼できる評価を得るのに十分な数の意見があることを仮定または保証しているからである。しかしながら、本実施例のように、多数の商品（カメラ）それぞれについての個別の評価を取り扱う場合には、ある商品の特徴を表現するのに、せいぜい1つか2つの形容詞しか出現しない場合がある。したがって、既存の評価情報分析システムでは、このように非常に少ない数の情報から意見を評価することができない。

【 0 0 7 1 】

そこで、本実施例においては、商品特徴に関する意見を抽出するために、まず、ユーザーレビューの文を主観的なものか客観的なもののいずれかに分類する。次いで、意見を表す単語を特定し分類する。そして、その意見を表す単語の極性を統合して意見スコアを算出する。

【 0 0 7 2 】

主観的な文を特定するには、あらかじめ分類済みのコーパスを用いて学習させた $n - g r a m$ 分類器を用いることができる。また、商品特徴を修飾する単語を形容詞と定義してもよい。また、意見が肯定的か否定的かを特定するには、非特許文献5に記載されているウェブPMI法を用いることができる。

【 0 0 7 3 】

図3に示されるように、ベータ二項分布を用いて、 n 個の単語 $\{w_1, \dots, w_n\}$ から、商品特徴に対する意見スコア $S(306)$ の補間処理を行う。ここで、 w_i は、 $E\{+1, -1\}$ である。 $a+$ および $a-$ をパラメータとしてベータ分布から s が生成されるような生成的なモデルを用いることができる。また、 s を用いて、肯定的な極性を有する形容詞が出現する確率を決定する。すなわち、

$$P(p = +1) = s, \quad P(p = -1) = 1 - s$$

【 0 0 7 4 】

SVMによる分類が正確であるとは言い切れないので、他のレイヤーをこのモデルに追加する。分類された極性が、 $P(\text{分類が正しい}) = 0.8$ という二項分布によって生成されると仮定する。

【 0 0 7 5 】

最後に、 $a+ = a- = 1$ と設定する。これは、肯定的および否定的な形容詞が同程度出現するという前提を表している。このモデルを、ギブスサンプリング (Gibbs sampling) を用いて、商品特徴内に出現する形容詞の極性と適合させた。また、 $S(306)$ を最終的な評価情報スコアとして用いる。要するにこれは、少数の形容詞しか

10

20

30

40

50

いときには、商品特徴の質として極端な評価が行われないということである。

【0076】

ここで、主観的な文を特定した後に、その主観的な文だけに基づいて評価を行ったときに、意見評価の正確さが改善される。主観的な文を特定することは、たとえば、あらかじめ分類が行われているデータベース（たとえば、引用文献7で言及されているもの）によって学習させたn-gram分類器を、ウェブ上のレビューに含まれる文に対して使用することで実現することができる。

【0077】

商品特徴を修飾する形容詞を意見とみなして、主観的な文から意見を抽出および分類する。商品特徴が出現するレビューデータ中の主観的な文から、形容詞的修飾語句、副詞および名詞句主語（動詞を介して）によってその特徴に関連付けられている形容詞が抽出される。もしも否定語句が形容詞を修飾している場合には、その形容詞については否定的とマークする。

【0078】

例えば、形容詞に対して、代表的な肯定的評価を表す単語excellentおよびfantastic、否定的な評価を表す単語terribleおよびawfulを用いたウェブベースのPMIからなる特徴ベクトルを計算する。PMIを計算する際に用いる個数(counts)は、ウェブの検索エンジンで各単語を検索した結果のヒット数を用いることができる。また、www.cs.pitt.edu/mpqa/から得られるような主観性評価のための用語集を用いてSVMを学習させ、これをもって特徴ベクトル进行分类する。以上のような方法を用いて意見の極性の分類を行った際の正確さを評価した結果を以下の表に示す。

【表1】

意見の極性分類の正確さ

極性	精度	再現率	F1値
肯定的	0.84	0.69	0.76
否定的	0.75	0.81	0.78

【0079】

I-5. 要約文の選択

本システムは、一例として、選択されたカメラの特徴に関する意見を代表する少数の文（要約文の代表例）を自動的に選択して、そのカメラに関する意見の雰囲気を利用者に伝えるようにする。この方法では、図4に示すように、特定の商品（例えばカメラ）に関し、その商品のレビューの中で信頼できる商品特徴のクラスターを取得する(S402)。そして、それらに対してスコアを求め数値化する(S404)。この商品特徴に対するスコアは、(1)その商品特徴に関する意見を表している文の実数（重複カウントを排除した数）(2)特徴を現す語句と「カメラ」のPMIスコア、の2つに基づいて計算される。すなわち、PMIスコアが大きい商品特徴ほど、また多くの文に出現する商品特徴ほど、高いスコアが割り当てられる。この2つ方法で算出した数値の統合法には様々なものが考えられるが、ここでは単純に両者を掛け合わせることにする。

【0080】

その後、商品特徴をスコアに基づいてソートする(S406)。

【0081】

次に、このソート（順位付け）した商品特徴に関し、その特徴を含むレビューの文のスコアを算出し、文ごとに順位付けを行う。そして、対象となる商品特徴ごとに、上位N個の代表的な文（ここでは、2個の文）を選択するが、これを、予め設定しておいた文の最大数に達するまで、または予め設定しておいた商品特徴の数に達するまで繰り返す(S408)。

【0082】

次に、商品特徴のクラスターの一部または全部を代表する文を選択する。要約文の最大

数は、ユーザーが指定してもよいし、例えば文のスコアが予め設定した値を超えているものを選択するなど、システムが自動的に決定しても良い。

【0083】

次に、ある商品特徴について、特定の極性を有し且つその商品特徴と関連する文のそれぞれに対して、スコアの計算を行う。ここで、商品特徴が<形容詞><名詞列>のパターンに含まれる文だけが、このスコア計算の対象となる。このスコアは、出現数の多い商品特徴ほど、出現数の多い形容詞と名詞句のセットほど、そして形容詞と名詞列の間のPMI値が高いほど、大きい値となる。

【0084】

I - 6 . 商品の用途の抽出

「商品の用途」は、カメラに関して言えば、以下のいずれかを表す用語であると定義できる。すなわち、(1)何の写真を撮影するか、(2)写真撮影時に撮影者は何をしているか、(3)撮影した写真で何をするか、である。

【0085】

これら3種類の用途は、相互に関連することが多い。例えば、"birthday party (誕生パーティー)"、"結婚式"、"running of the bulls (牛追い祭り)"、"ballroom dancing (社交ダンス)"、および、"Garden of the Gods in Colorado Springs (コロラドスプリングスの神々の庭園)"は、いずれも何の写真を撮影するかを表している。しかしながら、これらは同時に、写真撮影時に撮影者が何をしているかをも表し得る。したがって、異なる種類の用途を互いに排他的になるように自動的に分類することはできない。

【0086】

図5は、カメラの用途を特定する方法の一例を示すフローチャートである。カメラの用途は、用途を示すのに用いられる一般的な表現のパターンを検索することで特定する。このために、本実施例では、"picture (写真)"と関連する名詞列を検索に用いている。すなわち、{picture, pictures, photo, photos, pic, pics}のいずれかの写真関連用語を、<写真関連用語><前置詞句>のパターンとして含む前置詞句を検索するのである。この前置詞句は、レビューデータから抽出する(S502)。このパターンにマッチする前置詞句に対して、信頼できる商品特徴("lens"や"shutter")および数値を含んだ名詞列を削除するフィルタリング処理を行って当該前置詞句中の名詞列を抽出する(S504)。なお、この前置詞句が複合句(2以上の名詞)を含んでいる場合、この名詞列中の名詞を別々に抽出するようにする(S504)。例えば"picture of people and pets (人とペットの写真)"の場合、"people"と"pets"とに分離して抽出する。ここで、無視すべき用語(例えば"anything")のリストを別途用意し、これに載っている名詞列を抽出対象から外すようにしても良い。

【0087】

次に、抽出され残った名詞列グループ化する(S506)。この際には、同じ名詞で終わる名詞列を同一のグループにグループ化する。例えば、"zoo (動物園)"、"Washington Zoo (ワシントン動物園)"および"San Diego Zoo (サンディエゴ動物園)"を全て"zoo"という一つのグループにする。次に、このグループを、出現頻度の順にソートする(S508)。こうしてソートされた名詞列が、カメラの用途として特定される(S510)。また、この生成したリストをチェックのためにユーザーに提示してもよい。

【0088】

以上のような方法で特定しソート(順位付け)した用途の例の上位25個を、それぞれの用途を用いた前置詞句の最頻出例3つと同時に表2として示す。

【表 2】

カメラの用途の自動検出例
(用途上位 25 個、用途ごとの最頻出句 3 個)

light	in low light	in bright light	in good light
people	of people	with people	of two people
conditions	in low light conditions	in all conditions	under most conditions
time	at a time	at one time	at the same time
kids	of the kids	of kids	of kids and pets
family	of family	of the family	of family and friends
friends	of friends	of family and friends	with friends
computer	on the computer	on computer	on a computer
price	for the price	at a great price	at a reasonable price
flowers	of flowers	of the flowers	of flowers and birds
day	during the day	on a sunny day	from day
dark	in the dark	in dark	in complete dark
color	with great color	with good color	in color
cameras	with both cameras	from both cameras	with other cameras
tv	on TV	on the TV	on a TV
set	on a set	on one set	with one set
succession	in rapid succession	in quick succession	in succession
row	in a row		
children	of children	of small children	of the children
moon	of the moon	of the full moon	of the Moon
items	of items	of small items	of the same items
birds	of birds	of birds and wildlife	of flowers and birds
room	in a room	in a dark room	in a darker room
water	under water	in the water	under the water

10

20

30

【 0 0 8 9 】

また、自動的に検出した「写真撮影時に撮影者は何をしているか」の例を最頻出の句 2 つと共に、表 3 として示す。

【表 3】

自動抽出した用途の例「写真撮影時に撮影者は何をしているか」

holidays	the holidays	the EID Holidays
snorkeling	snorkeling	snorkeling and scuba diving
hiking	hiking	camping and hiking
skiing	skiing	
diving	diving	snorkeling and scuba diving
game	a basketball game	a double header or a full football game
kayaking	kayaking	kayaking or horseback riding or hiking
dinner	dinner	a special occasion dinner
distances	all different lights and distances	
meetings	public meetings	
snowfall	a heavy snowfall	
header	a double header or a full football game	
sightseeing	sightseeing	
camping	camping and hiking	
stingrays	chasing fish or stingrays	
riding	kayaking or horseback riding or hiking	
christmas	christmas	
fish	chasing fish or stingrays	

40

50

【 0 0 9 0 】

I - 7 . 用途と特性との関連付け

データ抽出の最終ステップは、特性をそれぞれの用途と関連付けることである。この関連付けは、例えば、それぞれの用途に適したカメラ（予め選択しておく）に対して、最も頻繁に出現した特性を、その用途と関連付けるようにすることで実現できる。また、これに関しては、手動すなわち人が手作業で行うのも精度と手間の観点から現実的な方法である。

【 0 0 9 1 】

I - 8 . ランキング

特性と用途とが抽出され及び関連付けられた後は、商品がユーザーインターフェースにランキングの結果に応じて表示される。例えば、ユーザーの指定した重み付けに応じてランキングを行うアルゴリズムによって、商品の順位付けを行い、その順序に応じて商品を表示する。

【 0 0 9 2 】

図7のGUIスクリーン700中の目盛りセクター702が、現在の重み、すなわち、仕様、特徴、属性の重要度を示している。このセクター702をユーザーが操作していずれかの重み付けを変更した場合、その変更された重みを正規化した上で、仕様、特徴、属性に反映させる。また、ユーザーの活動や興味から、間接的に重み付けを行うようにすることもできる。例えば、非特許文献8に記載されている「反転インデックス」による検索法、すなわちクエリーと結果のセットをインデックスとして保持しておく手法を用いれば、特性および用途が、検索結果として取得した商品のセットと関連付けられ、インデックスとして保存される。そして、ユーザーによって商品のセットが選択されると、このインデックスを用いて最も関係の深い特性および用途が検索される。そして、仕様または属性に割り当てられた関連性スコアが、その重みとして割り当てられ、再度最も関連の深い商品が検索される。

【 0 0 9 3 】

II . インターフェース

II - 1 . 詳細画面

商品の決定は、数値範囲の設定やリストからの選択のように簡単にはいかない。そこで、ユーザーが商品及びその特性をより詳しく検討するためのGUIを提供する。そのために、図8におけるGUI画面800は、カメラ表示部802を含んでいる。また、このカメラ表示部802は、個々のカメラの仕様804だけでなく、特定の特性に関するレビューの代表例806を表示している。この代表例は、上述のように要約文の代表例として自動的に抽出されたものであり、レビューの重要な点を要約している。重要な点として、この要約文の代表例806は、実際のレビューと関連付けられており、ユーザーの指示に応じてそれを表示する。このように、GUIは商品の特性の抜粋から詳細なレビューへのリンクを提供するのである。ユーザーがレビューの代表例の隣に表示されているウィジェットをクリックすると、レビューの詳細表示から元の表示に戻す。なお、このウィジェットは、その代表例と関連付けられている特性を直接選択するためのものである。

【 0 0 9 4 】

上述のように、商品の順位は、ユーザーが指定した仕様と特徴の重み付けに依存する。このインターフェースは、ユーザーが直接的及び間接的にこの重みを調節できるようになっている。

【 0 0 9 5 】

図6のGUI画面600では、何の写真を撮影するかという意味での用途602を選択することで、重みを間接的に指定するインターフェースの例を示している。より詳しい質問を取り扱えるように、用途を手動で整理してグループ化してもよい。ここでは一例として、用途を3つの用途にグループ化する例を示す。すなわち、（1）撮影時にユーザーが何をしているか（例：ハイキング）、（2）何の写真を撮影するか（例：山の風景）、（3）写真を何に使うか（例：スクラップブックに貼り付ける）の3つである。

【 0 0 9 6 】

上述のように用途が特性に割り当てられるので、用途の選択は間接的に重みを調整することになる。

【 0 0 9 7 】

ユーザーは、図 7 の G U I 画面 7 0 0 において、個々の特性 7 0 6 についてその重みのレベルを選択することによって、直接的に重み付けを調整することができる。ここで、レベルが小さい（最小値 0）の特性は重要でない特性であり、レベルが大きい（最大値 7）の特性は重要な特性であるという意味である。

【 0 0 9 8 】

特性の重みを操作するという検索法は、通常の実索における操作と趣を異にする。というのも、大半の実索用インターフェースは、ファセット（検索に用いる側面、観念または属性のこと）を選択したり、対象となる数値範囲を設定したりするようになっているからである。ファセットや数値範囲ではなく重みに注目するのは、重みが詳細な技術的知識を要求しないからである。すなわち、ユーザーがカメラの最大画素数をどの程度気に掛けているということを重みによって指定するが、これは最大画素数を具体的に指定する場合と異なり、その特徴に関して技術動向（どの程度の画素数が普通なのか等）を知らずとも行うことができる。

10

【 0 0 9 9 】

ここで、重みの指定には、様々な G U I を用いることができる。その例を図 9 に示す。重みを指定するのに最も単純なインタラクタ（操作用の G U I 部品）は、線形スライダー図 9（A）であろう。図 9（B）には、2つの指標の間のトレードオフとして重みを指定する二分スライダーの例を示す。

20

【 0 1 0 0 】

図 7 の G U I としては、最も単純な種類の重み付け調整用インターフェースを示しているが、これに限らず様々な種類のインターフェースを用いることができる。

【 0 1 0 1 】

（1）連続 - 増加（図 9（C））：曲線状のインタラクタがカテゴリー（この場合は色）ごとの重みを表している。この場合、色は波長に対応して連続的な値を取り得るので、カテゴリーそのものが連続的である。また、波長の上限が決まっているわけではないので、カテゴリーの範囲を予め制限することは出来ない（データ分析の結果増加し得る）。特定の色に対応する線上の点をドラッグしてその重みを変化させる場合、その周辺も曲線状に変化させて連続性を保つようにする。この際に、ある色の重みを増加させる場合には、他の色の重みを減らすようにして、インタラクタの曲線が囲む面積（重みの積分値）が一定になるようにする。

30

【 0 1 0 2 】

（2）離散 - 増加（図 9（D））：直線状のインタラクタが、離散的なカテゴリーごとの重みを表している。ここでは、例としてカーオーディオのスピーカーの数をカテゴリーとし、その数ごとの重みを取り扱っている。スピーカーの数の上限が決まっていなくて、カテゴリーの数に制限をつけることは出来ない（データ分析の結果増加し得る）。ここで、インタラクタを操作して、あるカテゴリーの重みを大きくすると、すなわち直線の長さを長くすると、他のカテゴリーの直線の長さを短くし、全ての直線の長さの和が一定になるようにする。

40

【 0 1 0 3 】

（3）連続 - カテゴリー（図 9（E））：レーダーチャート状のインタラクタが、連続的な値を取るカテゴリーの重みを表している。この場合は、賃貸住宅の実索用に、ある町の中における好ましい地域（重みが大きい）を方角で表している。方角は、決まった範囲内に収まるので、（1）とは異なり、予めカテゴリーの範囲を制限することができる（データ分析の結果増えることは無い）。一方で、（1）と同様に、インタラクタに囲まれる図形の面積を一定にするようにする。

【 0 1 0 4 】

50

(4) 離散 - カテゴリー (図9 (F)) : 直線状のインタラクタが、離散的なカテゴリそれぞれの重みを表している。この場合は、ノートPCの検索用に、どのようなアプリケーションに適しているものを望むか (重みが大きい) を示している。カテゴリーの数は、予め決めたアプリケーションの分類の数によって一意に定まるので、(2)とは異なり、カテゴリーの数を制限できる (データ分析の結果増えることは無い)。一方で、(2)と同様に、直線の長さの和を一定にするようにする。

【0105】

II-2. 比較画面

重み付けの調整によって整列された商品のリストを生成することになるが、この重み付けのプロセスは静的なものではない。すなわち、ユーザーは、重みを調整しながらそれがランキングにどのように影響を与えるかを検討しながら、商品を選ぶのである。そうすると、その過程で見つけた好みの商品が、重み調整後にはランキング上位から外れて見えなくなってしまうかもしれない。したがって、重み調整をしながら見つけた商品を保持しておいて比較できるようにすることが肝要である。このために、図10に示すような平行座標インターフェース1000を用いることができる。これは、概要表示、拡大表示及びフィルタリング、ならびに詳細表示の各機能を統合したものである。一般的な平行座標表示とは異なり、データポイント1002が数えるほどしかない。各カメラに対応する重みを表す線1004がクリックされると、そのカメラの詳細情報を表示するようにする。ディスプレイボックス1006は、画面の右側にあり、評価、QRコード (登録商標) および商品特性に対する意見のスコアを表示している。

【0106】

図11は、用途を基礎としたユーザーインターフェースを使用する方法の例を示している。ユーザーが希望する用途を入力する (S1102) と、システムはGUIにその用途に応じた検討対象となる商品のリストを表示する。その後ユーザーが、商品特性の重みを操作する (S1104) と、その変更後の重み付けで商品の再ランキングが行われるが、これによって、それまでは表示されていなかった商品画面に表示されることになる。ユーザーが商品を選択する (S1106) と、その商品の詳細な情報を表示する。ユーザーはこのときに、比較表示を行うように指示することもできる (S1108)。この場合、システムは、上述の平行座標インターフェースを表示する。また、ユーザーは選択した商品を比較用のコレクションに加えることもできる (S1110)。

【0107】

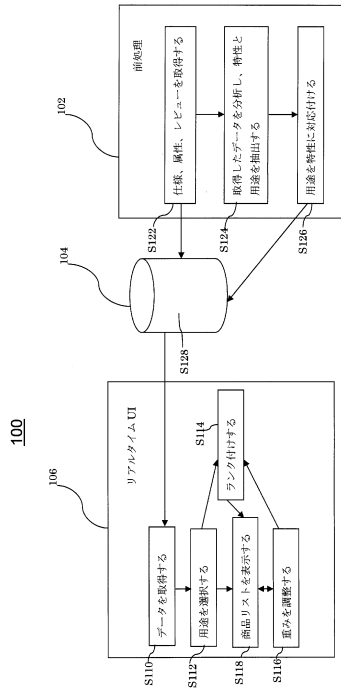
III. コンピュータ実装の例

図12は、これまで説明してきたシステム / 方法を実装するコンピュータ / サーバシステム1200の例を示すものである。システム1200は、コンピュータ / サーバプラットフォーム1201を含み、このプラットフォーム1201は、プロセッサ1202およびメモリ1203を含む。ここで、機械読取可能な記録媒体は、ディスクや半導体メモリ等、任意の有体の記録媒体を意味する。

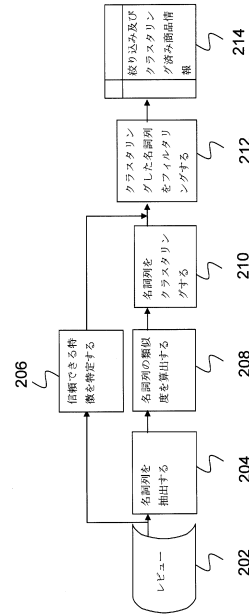
【0108】

さらに、コンピュータプラットフォーム1201は、キーボード、マウス、タッチデバイスまたは音声入力装置等、複数の入力装置1204からの入力を受ける。コンピュータプラットフォーム1201は、また、ポータブルハードディスク、光学ドライブ (CDやDVD) 等のリムーバブルストレージデバイス1205を接続してもよい。また、インターネットまたは他のローカルネットワーク上のネットワーク資源1206に接続してもよい。このコンピュータは、ディスプレイ1208にデータ又は他の情報をユーザーに向けて出力する。ディスプレイが入力装置1204を兼ねても良い。

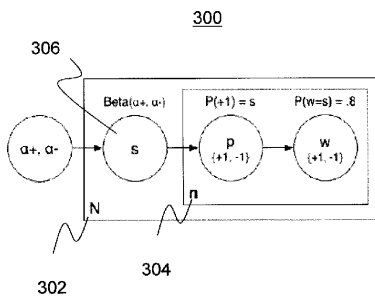
【図 1】



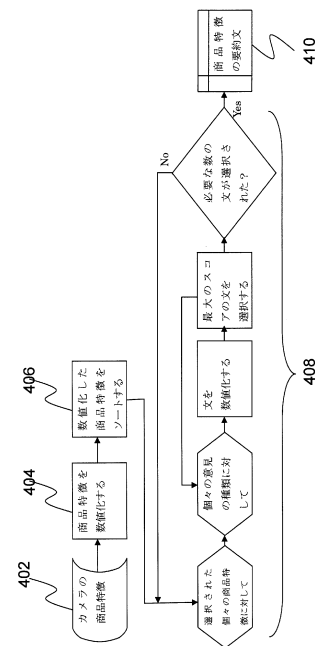
【図 2】



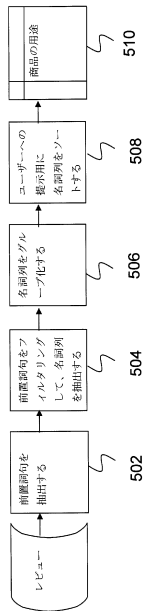
【図 3】



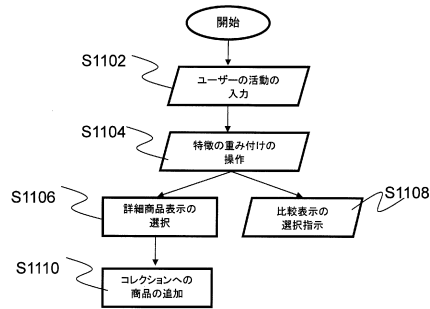
【図 4】



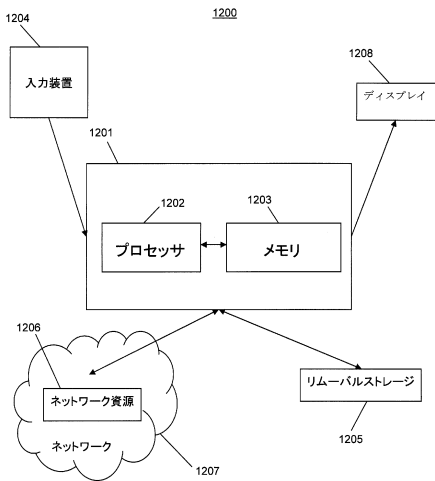
【図 5】



【図 1 1】



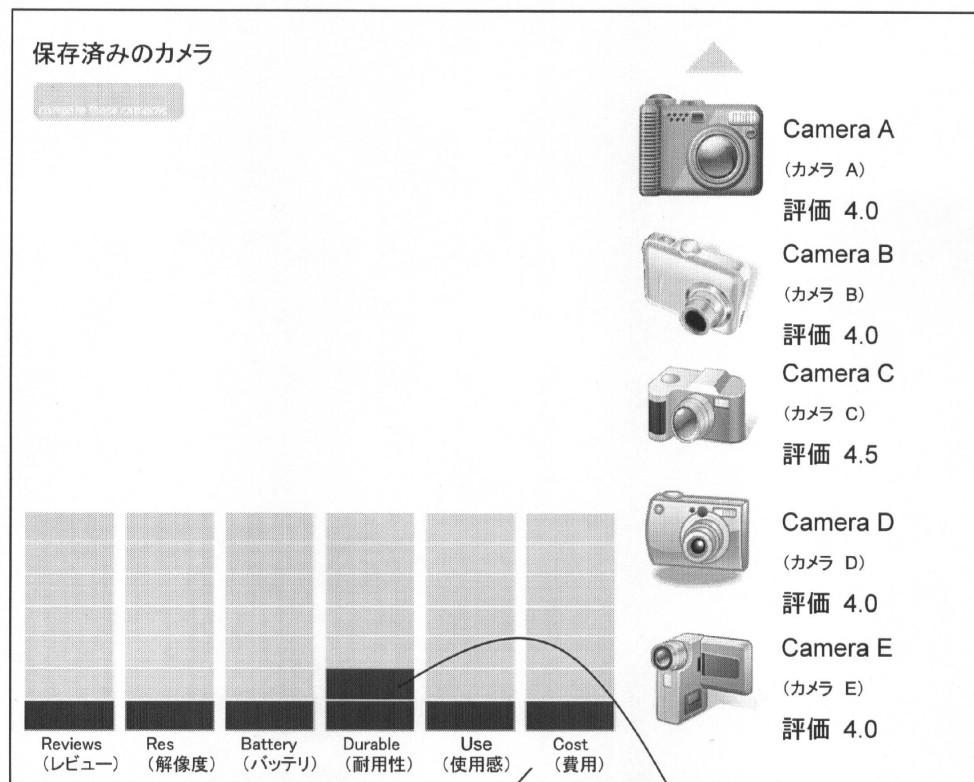
【図 1 2】



【図 6】



【図 7】

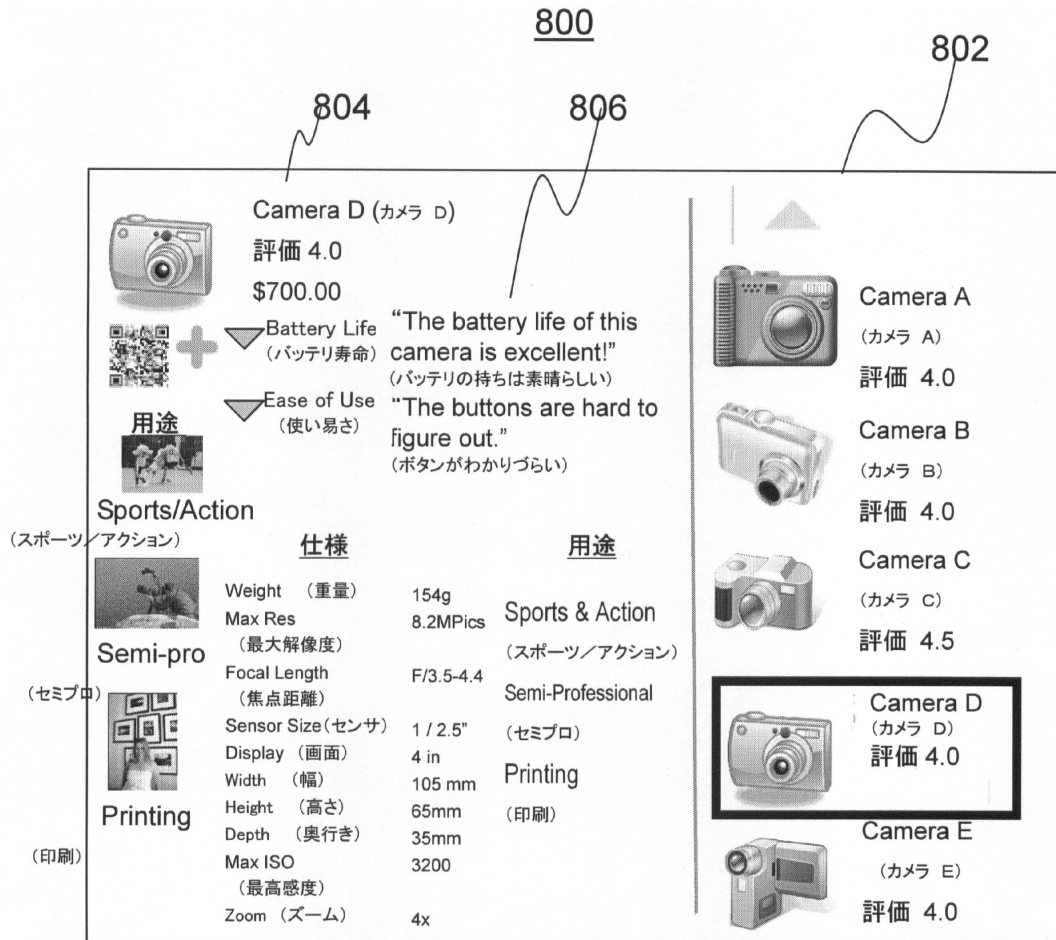
700

702

706

704

【図 8】



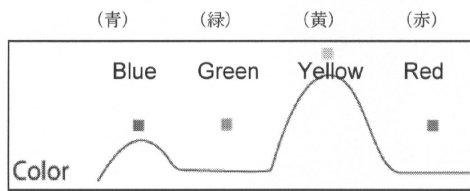
【図9】



(A)

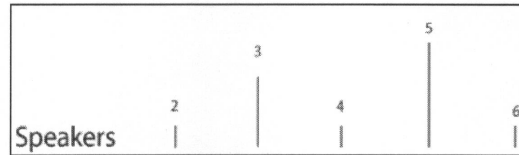


(B)



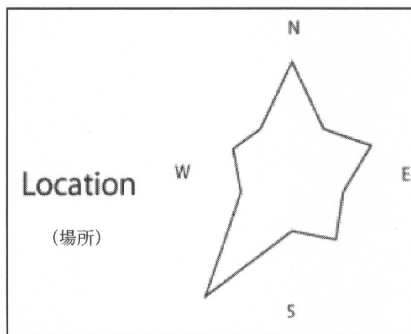
(色)

(C)

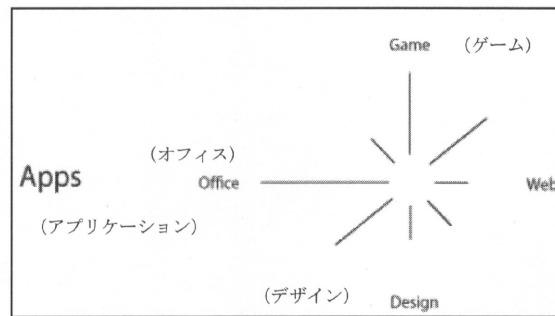


(スピーカ)

(D)

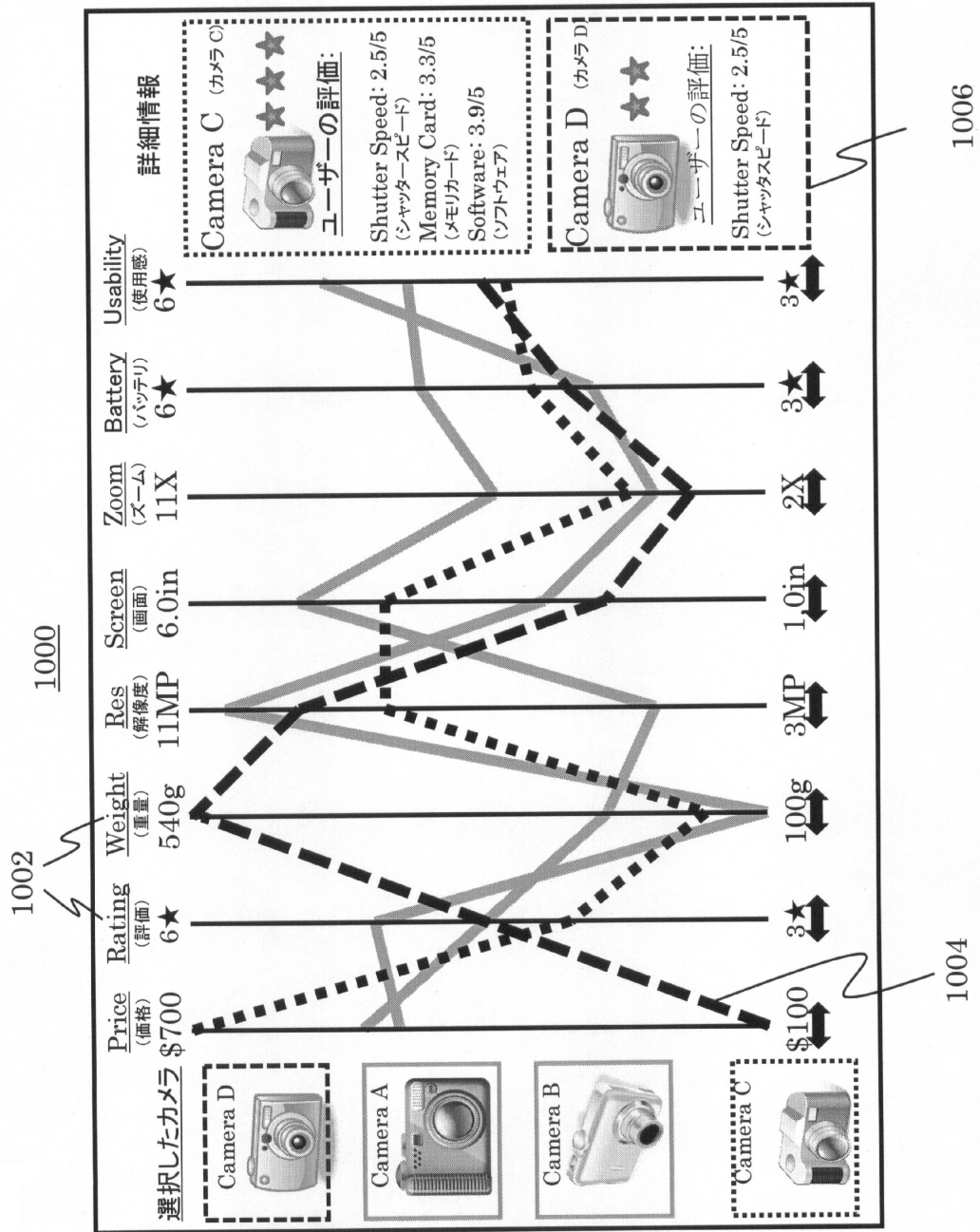


(E)



(F)

【図10】



フロントページの続き

- (72)発明者 スコット カーター
アメリカ合衆国 9 4 3 0 4 カリフォルニア州 パロアルト ポータードライブ 3 1 7 4 エフ
エックス パロアルトラボラトリーインク内
- (72)発明者 アディティ シュリクマー
アメリカ合衆国 9 4 3 0 4 カリフォルニア州 パロアルト ポータードライブ 3 1 7 4 エフ
エックス パロアルトラボラトリーインク内
- (72)発明者 ジェレミー ピケンズ
アメリカ合衆国 9 4 3 0 4 カリフォルニア州 パロアルト ポータードライブ 3 1 7 4 エフ
エックス パロアルトラボラトリーインク内

審査官 齊藤 貴孝

- (56)参考文献 特開 2 0 0 9 - 1 4 5 9 8 1 (J P , A)
国際公開第 2 0 0 9 / 1 5 5 3 7 5 (W O , A 1)
特開 2 0 0 4 - 0 3 8 7 2 9 (J P , A)
特開平 0 4 - 2 7 3 5 2 1 (J P , A)
米国特許第 0 7 7 6 1 3 4 5 (U S , B 1)
島松 千春、外 3 名、クチコミ情報からの関連商品マイニング、電子情報通信学会 第 1 9 回デ
ータ工学ワークショップ論文集、日本、電子情報通信学会データ工学研究専門委員会、2 0 0 9
年 6 月 2 5 日、p . 1 - 4

- (58)調査した分野(Int.Cl. , D B 名)
G 0 6 F 1 7 / 3 0
G 0 6 Q 3 0 / 0 2