

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第5714702号
(P5714702)

(45) 発行日 平成27年5月7日 (2015.5.7)

(24) 登録日 平成27年3月20日 (2015.3.20)

(51) Int. Cl.

F I

G 0 6 Q 30/06 (2012.01)

G 0 6 Q 30/06 1 2 6 E

請求項の数 23 (全 22 頁)

(21) 出願番号	特願2013-512600 (P2013-512600)	(73) 特許権者	510330264
(86) (22) 出願日	平成23年5月25日 (2011.5.25)		アリババ・グループ・ホールディング・リミテッド
(65) 公表番号	特表2013-543154 (P2013-543154A)		ALIBABA GROUP HOLDING LIMITED
(43) 公表日	平成25年11月28日 (2013.11.28)		英国領、ケイマン諸島、グランド・ケイマン、ジョージ・タウン、ワン・キャピタル・プレイス、フォース・フロア、ビー・オー、ボックス 847
(86) 国際出願番号	PCT/US2011/000932		
(87) 国際公開番号	W02011/149527	(74) 代理人	110000028
(87) 国際公開日	平成23年12月1日 (2011.12.1)		特許業務法人明成国際特許事務所
審査請求日	平成25年12月18日 (2013.12.18)	(74) 代理人	100102989
(31) 優先権主張番号	13/068, 976		弁理士 井上 佳知
(32) 優先日	平成23年5月24日 (2011.5.24)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	201010187445.7		
(32) 優先日	平成22年5月27日 (2010.5.27)		
(33) 優先権主張国	中国 (CN)		

最終頁に続く

(54) 【発明の名称】 商品情報の乱雑さの解析

(57) 【特許請求の範囲】

【請求項 1】

商品情報を解析する方法であって、

ユーザによって入力された商品情報を受信し、

前記商品情報から 1 または複数の特性属性に対応する値を取得することを少なくとも含む、前記商品情報の解析を実行し、1 または複数の特性属性に対応する前記値は、前記商品情報が乱雑であるか否かを判定するために用いられ、

1 または複数の特性属性に対応する前記取得された値に少なくとも部分的に基づいて、前記商品情報に関連する乱雑さの信頼水準を決定し、

前記商品情報に関連する前記乱雑さの信頼水準が、事前設定された閾値を超えるか否かを判定し、前記乱雑さの信頼水準が前記事前設定された閾値を超えた場合には、前記商品情報の公開を停止する旨の示唆を送信し、前記乱雑さの信頼水準が前記事前設定された閾値を超えていない場合には、前記商品情報の公開を停止する旨の示唆を送信しない、ことを備える、方法。

【請求項 2】

請求項 1 に記載の方法であって、前記商品情報は、電子商取引ウェブサイトに関連して受信される、方法。

【請求項 3】

請求項 1 に記載の方法であって、前記商品情報は、商品タイトル、商品説明情報、商品

10

20

紹介情報、商品レビュー、および、商品の製品仕様の内の 1 または複数を含む、方法。

【請求項 4】

請求項 1 に記載の方法であって、1 または複数の特性属性に対応する前記取得された値に少なくとも部分的に基づいて、前記商品情報に関連する乱雑さの信頼水準を決定することは、

1 または複数の特性属性に対応する前記取得された値を条件付き確率モデルに入力し、

1 または複数の特性属性に対応する前記取得された値と、前記条件付き確率モデルとを少なくとも用いて、前記商品情報が乱雑である可能性に関連する事後確率を計算することを含む、

前記乱雑さの信頼水準は、前記事後確率を含む、方法。

【請求項 5】

請求項 1 に記載の方法であって、前記 1 または複数の特性属性は、少なくとも 1 つの形態的特性属性を含む、方法。

【請求項 6】

請求項 5 に記載の方法であって、前記少なくとも 1 つの形態的特性属性は、

前記商品情報に含まれるコンマの数、前記商品情報の文の長さ、前記商品情報内の総単語数に対する反復単語の除去後の前記商品情報に含まれる単語数の比、前記商品情報における最頻出単語の出現回数、事前設定されたルールに基づいて前記商品情報がセグメントに分割された後に各セグメント内の特定の位置にある単語で構成されたセット内の総単語数に対する反復単語の除去後の単語数の比、事前設定されたルールに基づいて前記商品情報がセグメントに分割された後の各セグメント長の分散、の内の 1 または複数を含む、方法。

【請求項 7】

請求項 1 に記載の方法であって、前記 1 または複数の特性属性は、少なくとも 1 つの構文的特性属性を含む、方法。

【請求項 8】

請求項 7 に記載の方法であって、前記少なくとも 1 つの構文的特性属性は、

前記商品情報内の単語に対応する品詞の総数に対する反復品詞の除去後の前記商品情報に含まれる単語に対応する品詞の数の比、名詞の総単語数に対する反復単語を除去した後の前記商品情報内の名詞の単語数の比、最頻出品詞の出現回数、事前設定されたルールに基づいて前記商品情報がセグメントに分割された後に各セグメント内の特定の位置にある単語に対応する品詞で構成されたセット内の総品詞数に対する、反復品詞を除去した後の品詞数の比、の内の 1 または複数を含む、方法。

【請求項 9】

請求項 6 に記載の方法であって、さらに、事前設定されたルールに基づいて前記商品情報をセグメントに分割することを備え、前記ルールは、

前記商品情報内のコンマの位置に基づいて前記商品情報を分割して、前記商品情報に含まれる単語のサブセットを含む 1 または複数のセグメントを形成すること、

および / または、

前記商品情報における最頻出単語の位置に基づいて前記商品情報を分割して、1 または複数のセグメントを形成することを含む、方法。

【請求項 10】

請求項 8 に記載の方法であって、さらに、事前設定されたルールに基づいて前記商品情報をセグメントに分割することを備え、前記ルールは、

前記商品情報内のコンマの位置に基づいて前記商品情報を分割して、前記商品情報に含まれる単語のサブセットを含む 1 または複数のセグメントを形成すること、

および / または、

前記商品情報における最頻出単語の位置に基づいて前記商品情報を分割して、1 または複数のセグメントを形成することを含む、方法。

10

20

30

40

50

【請求項 1 1】

請求項 1 に記載の方法であって、前記乱雑さの信頼水準が前記事前設定された閾値を超えた場合に、前記商品情報が乱雑な商品情報を含むと判定することを備える、方法。

【請求項 1 2】

請求項 1 1 に記載の方法であって、前記乱雑さの信頼水準が前記事前設定された閾値を超えた場合に、

前記商品情報に関連する乱雑さを引き起こした可能性のある前記商品情報のキーワードを決定し、

前記ユーザによってアクセス可能なインターフェース要素を介して前記キーワードに関する示唆を提示すること、

をさらに備える、方法。

10

【請求項 1 3】

請求項 1 2 に記載の方法であって、さらに、前記インターフェース要素を介して前記商品情報への修正を入力することを前記ユーザに促すことを備える、方法。

【請求項 1 4】

商品情報を解析するためのシステムであって、

プロセッサであって、

ユーザによって入力された商品情報を受信し、

前記商品情報から 1 または複数の特性属性に対応する値を取得することを少なくとも含む、前記商品情報の解析を実行し、1 または複数の特性属性に対応する前記値は、前記商品情報が乱雑であるか否かを判定するために用いられ、

20

1 または複数の特性属性に対応する前記取得された値に少なくとも部分的に基づいて、前記商品情報に関連する乱雑さの信頼水準を決定し、

前記商品情報に関連する前記乱雑さの信頼水準が、事前設定された閾値を超えるか否かを判定し、前記乱雑さの信頼水準が前記事前設定された閾値を超えた場合には、前記商品情報の公開を停止する旨の示唆を送信し、前記乱雑さの信頼水準が前記事前設定された閾値を超えていない場合には、前記商品情報の公開を停止する旨の示唆を送信しないように構成されているプロセッサと、

前記プロセッサに接続され、前記プロセッサに命令を提供するよう構成されているメモリと、

30

を備える、システム。

【請求項 1 5】

請求項 1 4 に記載のシステムであって、前記商品情報は、電子商取引ウェブサイトに関連して受信される、システム。

【請求項 1 6】

請求項 1 4 に記載のシステムであって、前記商品情報は、商品タイトル、商品説明情報、商品紹介情報、商品レビュー、および、商品の製品仕様の内の 1 または複数を含む、システム。

【請求項 1 7】

請求項 1 4 に記載のシステムであって、1 または複数の特性属性に対応する前記取得された値に少なくとも部分的に基づいて、前記商品情報に関連する乱雑さの信頼水準を決定するように構成されている前記プロセッサは、

40

1 または複数の特性属性に対応する前記取得された値を条件付き確率モデルに入力し、

1 または複数の特性属性に対応する前記取得された値と、前記条件付き確率モデルとを少なくとも用いて、前記商品情報が乱雑である可能性に関連する事後確率を計算するように構成されているプロセッサを含み、

前記乱雑さの信頼水準は、前記事後確率を含む、システム。

【請求項 1 8】

請求項 1 4 に記載のシステムであって、前記 1 または複数の特性属性は、少なくとも 1 つの形態的特性属性を含む、システム。

50

【請求項 19】

請求項 14 に記載のシステムであって、前記 1 または複数の特性属性は、少なくとも 1 つの構文的特性属性を含む、システム。

【請求項 20】

請求項 14 に記載のシステムであって、前記乱雑さの信頼水準が前記事前設定された閾値を超えた場合に、前記プロセッサは、前記商品情報が乱雑な商品情報を含むと判定するように構成されている、システム。

【請求項 21】

請求項 20 に記載のシステムであって、前記乱雑さの信頼水準が前記事前設定された閾値を超えた場合に、前記プロセッサは、さらに、

前記商品情報に関連する乱雑さを引き起こした可能性のある前記商品情報のキーワードを決定し、

前記ユーザによってアクセス可能なインターフェース要素を介して前記キーワードに関する示唆を提示するように構成されている、システム。

【請求項 22】

請求項 21 に記載のシステムであって、前記プロセッサは、さらに、前記インターフェース要素を介して前記商品情報への修正を入力することを前記ユーザに促すように構成されている、システム。

【請求項 23】

商品情報を解析するためのコンピュータプログラムであって、

ユーザによって入力された商品情報を受信するための機能と、

前記商品情報から 1 または複数の特性属性に対応する値を取得することを少なくとも含む、前記商品情報の解析を実行するための機能と、 1 または複数の特性属性に対応する前記値は、前記商品情報が乱雑であるか否かを判定するために用いられ、

1 または複数の特性属性に対応する前記取得された値に少なくとも部分的に基づいて、前記商品情報に関連する乱雑さの信頼水準を決定するための機能と、

前記商品情報に関連する前記乱雑さの信頼水準が、事前設定された閾値を超えるか否かを判定するための機能と、前記乱雑さの信頼水準が前記事前設定された閾値を超えた場合に、前記商品情報の公開を停止する旨の示唆を送信し、前記乱雑さの信頼水準が前記事前設定された閾値を超えていない場合に、前記商品情報の公開を停止する旨の示唆を送信しない機能と、

をコンピュータによって実現させる、コンピュータプログラム。

【発明の詳細な説明】**【技術分野】****【0001】****[他の出願の相互参照]**

本願は、すべての目的のために参照により本明細書に組み込まれる、2010年5月27日出願の発明の名称を「A METHOD AND DEVICE FOR PUBLISHING MERCHANDISE INFORMATION (商品情報を公開するための方法および装置)」とする中国特許出願第201010187445.7号に基づく優先権を主張する。当該出願は、

【0002】

本願は、オンラインウェブサイト技術に関し、特に、商品情報を公開する技術に関する。

【背景技術】**【0003】**

電子商取引の分野において、商品の説明情報(例えば、商品タイトル)は、その製品に関する重要情報を含む。例えば、図1の例に見られるように、表示された商品のタイトルは、「&New arrived&Fashion wind coat, ladies' coat, fashion coat, women's wind coat (Wh

10

20

30

40

50

olesale price + Dropship)」である。この例において、商品タイトルは、女性用ウィンドコートとしてユーザに商品を正確に提示できる。しかしながら、この商品タイトルは、冗長な情報を含んでおり、単語の使い方が「乱雑」である。例えば、「Fashion wind coat」、「fashion coat」、「ladies' coat」、および、「women's wind coat」という単語は、少なくとも部分的に意味が重複している。これらの意味の重複および単語利用の冗長性は、ウェブサイトでの商品情報の簡潔さ、そして、正確さまで損ないうる。さらに、例えば、ユーザによるウェブサイトでの商品情報の検索に回答して、ユーザに対して冗長および/または乱雑な商品情報を表示すると、検索処理の効率が低下しうる。

【図面の簡単な説明】

10

【0004】

以下の詳細な説明と添付の図面において、本発明の様々な実施形態を開示する。

【0005】

【図1】ウェブページでの商品情報の表示の一例を示す図。

【0006】

【図2】商品情報を解析するためのシステムの一実施形態を示す図。

【0007】

【図3】商品情報解析サーバの一実施形態を示す図。

【0008】

【図4】乱雑さ分類器の一実施形態を示す図。

20

【0009】

【図5】商品情報を解析するための処理の一実施形態を示すフローチャート。

【発明を実施するための形態】

【0010】

本発明は、処理、装置、システム、物質の組成、コンピュータ読み取り可能な格納媒体上に具現化されたコンピュータプログラム製品、および/または、プロセッサ（プロセッサに接続されたメモリに格納および/またはそのメモリによって提供される命令を実行するよう構成されたプロセッサ）を含め、様々な形態で実装されうる。本明細書では、これらの実装または本発明が取りうる任意の他の形態を、技術と呼ぶ。一般に、開示された処理の工程の順序は、本発明の範囲内で変更されてもよい。特に言及しない限り、タスクを実行するよう構成されるものとして記載されたプロセッサまたはメモリなどの構成要素は、ある時間にタスクを実行するよう一時的に構成された一般的な構成要素として、または、タスクを実行するよう製造された特定の構成要素として実装されてよい。本明細書では、「プロセッサ」という用語は、1または複数のデバイス、回路、および/または、コンピュータプログラム命令などのデータを処理するよう構成された処理コアを指すものとする。

30

【0011】

以下では、本発明の原理を示す図面を参照しつつ、本発明の1または複数の実施形態の詳細な説明を行う。本発明は、かかる実施形態に関連して説明されているが、どの実施形態にも限定されない。本発明の範囲は、特許請求の範囲によってのみ限定されるものであり、多くの代替物、変形物、および、等価物を含む。以下の説明では、本発明の完全な理解を提供するために、多くの具体的な詳細事項が記載されている。これらの詳細事項は、例示を目的としたものであり、本発明は、これらの具体的な詳細事項の一部または全てがなくとも特許請求の範囲に従って実施可能である。簡単のために、本発明に関連する技術分野で周知の技術事項については、本発明が必要以上にわかりにくくならないように、詳細には説明していない。

40

【0012】

商品情報の乱雑さの解析が開示されている。いくつかの実施形態において、ユーザによって入力された商品情報が受信される。いくつかの実施形態において、1または複数の特性属性に対応する値が商品情報から取得され、1または複数の特性属性に対応する値は、

50

商品情報が乱雑であるか否かを判定するために用いられる。いくつかの実施形態において、商品情報に関連する乱雑さの信頼水準が、1または複数の特性属性に対応する取得された値の最大エントロピ原理に少なくとも部分的に基づいて決定される。いくつかの実施形態において、最大エントロピ原理は、入力された商品情報に関連する特性属性の値の関数に基づいて乱雑さの信頼水準を決定する式である。いくつかの実施形態において、乱雑さの信頼水準が、事前設定された閾値を超えるか否かが判定される。事前設定された閾値を超える場合、商品情報の公開を停止する旨の示唆が送信される。事前設定された閾値を超えない場合、商品情報の公開を停止する旨の示唆は送信されない。いくつかの実施形態において、信頼水準が、事前設定された閾値を超える場合、商品情報は乱雑であるとみなされ、それに応じてイベントがトリガされる（例えば、商品情報の公開を停止する旨の示唆の送信）。

10

【0013】

いくつかの実施形態において、「乱雑さ」の概念は、同じ製品の「羅列」および異なる製品の「積み重ね」の概念によって説明することができる。本明細書で用いられているように、同じ製品の「羅列」とは、特定の製品の商品情報内に、互いに冗長であるかまたは実質的に同じ意味を表す単語が存在するという概念を指す。同じ製品の「羅列」の一例は、特定の製品の商品タイトル内で、多くの用語またはフレーズが互いに同義であるか、もしくは、特定のキーワードがタイトル内に数回出現することである（例えば、「coat（コート）」、「jacket（ジャケット）」、「outerwear（アウターウェア）」、「red（赤）」、および、2回目の「coat」を含む商品タイトル）。本明細書で用いられているように、異なる製品の「積み重ね」とは、商品情報内に、複数の異なる製品の商品名が含まれるという概念を指す。異なる製品の「積み重ね」の一例は、異なる製品を指す様々なキーワードを含む商品タイトルである（例えば、「mp3 player」、「mp4 player」、「ipod」、および、「walkman」というキーワードを含む商品タイトル）。本明細書で用いられているように、「乱雑さ」の程度とは、商品情報が「羅列」および/または「積み重ね」られている程度である。様々な実施形態において、乱雑である商品情報は、電子商取引ウェブサイトなどのウェブサイトで公開されることが望ましくない（例えば、閲覧者を誤解させる可能性がある不必要な情報を含みうるため）。

20

【0014】

いくつかの実施形態において、商品タイトルに加えて、商品情報は、1または複数の他のコンテンツを含みうる。例えば、商品説明情報、商品紹介情報、商品レビュー、商品の製品仕様。商品情報は、上記のものだけに限定されない。

30

【0015】

図2は、商品情報を解析するためのシステムの一実施形態を示す図である。システム200は、デバイス202、ネットワーク204、および、商品情報解析サーバ206を備える。ネットワーク204は、様々な高速データネットワークおよび/または遠隔通信ネットワークを含む。いくつかの実施形態において、デバイス202は、ネットワーク204を介して商品情報解析サーバ206と通信する。

【0016】

デバイス202はラップトップとして図示されているが、デバイス202の例としては、デスクトップコンピュータ、スマートフォン、携帯デバイス、または、タブレットデバイスが挙げられる。デバイス202は、ウェブブラウザ（例えば、マイクロソフトインターネットエクスプローラまたはグーグルクローム）を実行することができる。例えば、ユーザは、ウェブブラウザを介して電子商取引ウェブサイト（例えば、www.alibaba.com）にアクセスするためにデバイス202を用いることができる。ウェブサイトは、ウェブサイト上で製品を宣伝したいユーザがウェブインターフェースを介して情報を送信できるように対話型インターフェースを備えうる。

40

【0017】

商品情報解析サーバ206は、ユーザが送信した情報（例えば、商品情報）を受信し、情報が乱雑か否かを判定する。いくつかの実施形態において、商品情報解析サーバ206

50

は、商品情報に関連する信頼水準を決定する。いくつかの実施形態において、信頼水準が、事前設定された閾値に達するかまたは超える場合、商品情報は乱雑であるとみなされる。しかし、信頼水準が、事前設定された閾値に達しないかまたは超えない場合、商品情報は乱雑でないとみなされる。いくつかの実施形態において、商品情報が乱雑であると見なされた場合、情報解析サーバ206は、(例えば、関連ウェブサイトでの)商品情報の公開を停止する、および/または、関連する示唆をユーザに対して表示する。いくつかの実施形態において、商品情報が乱雑であると判定された場合、ウェブサイト情報解析サーバ206は、商品情報の修正をユーザに促す。

【0018】

図3は、商品情報解析サーバの一実施形態を示す図である。いくつかの実施形態において、図2の商品情報解析サーバ206は、少なくとも部分的に図3の例を用いて実装できる。図3に示すように、商品情報解析サーバ206は、通信要素10、解析要素11、第1の解析要素12、および、第2の解析要素13を備える。様々な実施形態において、商品情報解析サーバ206は、ウェブサイト(例えば、電子商取引ウェブサイトなど)をサポートするサーバと関連して(例えば、組み合わせて、構成要素として、または、通信するように)実装される。

【0019】

上述の要素は、1または複数の汎用プロセッサ上で実行されるソフトウェアコンポーネントとして、特定の機能を実行するよう設計されたプログラム可能論理デバイスおよび/または特定用途向け集積回路などのハードウェアとして、もしくは、それらの組み合わせとして実装することができる。いくつかの実施形態において、要素は、コンピュータデバイス(パーソナルコンピュータ、サーバ、ネットワーク装置など)に本発明の実施形態に記載された方法を実行させるための複数の命令など、不揮発性記憶媒体(光学ディスク、フラッシュ記憶装置、携帯用ハードディスクなど)に格納することができるソフトウェア製品の形態で具現化されてよい。要素は、単一のデバイス上に実装されてもよいし、複数のデバイスにわたって分散されてもよい。要素の機能は、互いに統合されてもよいし、複数のサブ要素にさらに分割されてもよい。

【0020】

通信要素10は、ユーザによって入力された商品情報を受信する。いくつかの実施形態において、通信要素10は、ユーザが情報を閲覧するおよび/または相互作用することを可能にする対話型インターフェース(例えば、電子商取引ウェブサイトのウェブページにあるもの)をサポートする。

【0021】

解析要素11は、商品情報を解析し、商品情報の特性属性値を取得する。いくつかの実施形態において、特性属性は、商品情報に含まれる単語の乱雑さを決定するために用いられる。

【0022】

計算要素12は、特性属性の値および最大エントロピ原理に基づいて、商品情報が乱雑な情報である信頼水準を計算する。乱雑さの信頼水準とは、どれだけの可能性で商品情報が乱雑な情報であるのかを意味する。

【0023】

いくつかの実施形態では、図3の例に示すように、計算要素12は、さらに、第1の計算サブ要素120および第2の計算サブ要素121を備える。

【0024】

第1の計算サブ要素120は、最大エントロピ原理に基づいた条件付き確率モデルのための入力情報として特性属性の値を取得するために用いられる。

【0025】

第2の計算サブ要素121は、条件付き確率モデルを用いて、商品情報が乱雑な情報である事後確率を入力情報で計算し、事後確率を、商品情報が乱雑な情報である信頼水準とするよう構成される。いくつかの実施形態において、ランダムな事象の事後確率は、関連

10

20

30

40

50

する証拠を考慮した後にランダムな事象に割り当てられる条件付き確率であると説明できる。

【 0 0 2 6 】

実行要素 1 3 は、信頼水準が、事前設定された閾値に達した、または、超えたと判定された時に、商品情報の公開を停止するよう構成される。

【 0 0 2 7 】

いくつかの実施形態において、任意選択的に、戦略要素 1 4 が商品情報解析サーバ 2 0 6 に備えられる。戦略要素 1 4 は、商品情報が乱雑であると判定された場合（例えば、関連する信頼水準が、事前設定された閾値に達したかまたは超えた場合）に、商品情報に含まれる単語の乱雑さを引き起こしていると思われる少なくとも 1 つのキーワードを決定する。いくつかの実施形態において、1 つのかかるキーワードは、商品情報内に最も頻繁に現れる単語である。いくつかの実施形態において、戦略要素 1 4 は、特定したキーワードを通信要素 1 0 を介してユーザに送信し、最初に送信された商品情報を修正するようユーザに促す。いくつかの実施形態において、戦略要素 1 4 は、さらに、商品情報のための任意選択的な修正オプションを含む。

【 0 0 2 8 】

いくつかの実施形態において、商品情報解析サーバ 2 0 6 は、機械学習に基づいた乱雑さ特定方法を採用するよう構成される。商品情報解析サーバ 2 0 6 は、ユーザが公開に向けて（例えば、電子商取引ウェブサイトで製品の提供に関連するウェブページに）送信する商品情報をテストするために、乱雑さ特定方法を用いる。公開に向けてユーザが送信した商品情報が乱雑さを含むと見なされた場合（例えば、商品情報に含まれる単語の乱雑さの信頼水準が、事前設定された閾値に達するかまたは超えると判定された場合）、商品情報の公開は停止される。いくつかの実施形態において、商品情報の公開が停止された場合、このイベントの示唆がユーザに（例えば、通信要素 1 0 によってサポートされるディスプレイを介して）送信される。

【 0 0 2 9 】

いくつかの実施形態において、信頼水準は、最大エントロピ原理に基づいた条件付き確率モデルを用いて計算される。ユーザが送信した商品情報に含まれる 1 または複数の単語の信頼水準を計算するために用いられる式の一例は、以下の通りである。

【 数 1 】

$$p(y|x) = \frac{1}{Z(x)} \exp \left(\sum_j \lambda_j f_j(x, y) \right) \quad \text{式 1}$$

【 0 0 3 0 】

ここで、 $y \in \{\text{タイトルが乱雑である, タイトルが乱雑でない}\}$ は、 y が 2 つの値「タイトルが乱雑である」および「タイトルが乱雑でない」を持ちうることを示す。どちらの値（「タイトルが乱雑である」または「タイトルが乱雑でない」）を y に割り当てるかについての決定は、事前設定されたパラメータに基づく。例えば、 y の値が「タイトルが乱雑である」である場合、計算された $p(y|x)$ は、タイトルが乱雑な情報を含む事後確率（すなわち、信頼水準）であり、 x は商品情報の特性属性である。いくつかの実施形態において、各特性属性に関連した y の値は、その特性属性の値に従う。 f_j は、最大エントロピモデルに基づく各特性属性の特性値である。 λ_j は、現行の商品情報の特性属性 j に対応する加重である。いくつかの実施形態において、 λ_j は事前に設定されうる（例えば、経験的な値に基づいて）。 $z(x)$ は、正規化因子であり、事前に設定することもできる（例えば、経験的な値に基づいて）。

【 0 0 3 1 】

いくつかの実施形態において、商品情報解析によって用いられる機械学習モデルは、条件付き確率モデルを確立するための線形回帰モデルであってよい。いくつかの実施形態において、商品情報解析によって用いられる機械学習モデルは、サポートベクターマシンモデルであってもよく、そのモデルは条件付き確率モデルではないが、それによって計算さ

10

20

30

40

50

れた比は信頼水準として用いることができる。

【 0 0 3 2 】

いくつかの実施形態において、上述の式 1 などの式を用いることによって、商品情報乱雑さ分類器が構築される。商品情報乱雑さ分類器の入力は商品情報を含み、分類器の出力は分類結果を含む。いくつかの実施形態において、分類結果の出力は、信頼水準の値であり、信頼水準の値が、事前設定された閾値より大きい場合、入力された商品情報は乱雑であると見なされると判定されるが、信頼水準が、事前設定された閾値より小さい場合、入力された商品情報は乱雑でないと判定される。

【 0 0 3 3 】

図 4 は、乱雑さ分類器の一実施形態を示す図である。図 4 の例に示すように、商品情報 4 0 2 は、乱雑さ分類器 4 0 4 に入力され、乱雑さ分類器 4 0 4 は、以下の 2 つの可能な分類結果の一方を出力する。クラス 1、信頼水準 1、または、クラス 2、信頼水準 2。いくつかの実施形態において、図 4 の出力領域に示すように、「タイトルが乱雑である」という分類結果をクラス 1 と呼び、「タイトルが乱雑でない」という分類結果をクラス 2 と呼ぶことができる。

【 0 0 3 4 】

いくつかの実施形態において、機械学習ベースの乱雑さ特定方法が用いられる場合、商品情報から取得された特性属性は、形態的特性属性および／または構文的特性属性に分けられる。これら 2 つの種類の特性属性（形態的または構文的）については、解析された商品情報の商品タイトルの例において以下で説明する。以下の例では、商品情報（例えば、商品タイトル）は、最初に形態的特性属性について、2 番目に構文的特性属性について解析されるが、いくつかの実施形態において、商品情報は、形態的特性属性の前に構文的特性属性について解析されてもよいし、同時に解析されてもよい。

【 0 0 3 5 】

最初に、形態的特性属性が商品タイトルから取得される。形態的特性属性に対応する値の例は、以下に挙げる値の内の 1 または複数を含みうるが、それらに限定されない。

【 0 0 3 6 】

1 . 商品タイトルに含まれるコンマの数

【 0 0 3 7 】

商品タイトルに含まれるコンマの数は、商品タイトルに含まれる単語が乱雑である（結果として、商品タイトルが乱雑である）可能性を、ある程度反映しうると考えられる。一般に、商品タイトル内のコンマの数が多いほど、商品タイトルに含まれる単語が乱雑である可能性が高くなる。

【 0 0 3 8 】

例えば、「#24 Baseball Jersey, Baseball Jerseys, Jerseys, Sports Jerseys, Sport Jersey, Jersey, 24# Baseball Jersey」という商品タイトルには、6 つのコンマがある。

【 0 0 3 9 】

2 . 商品タイトルの文の長さ（例えば、単語数 + コンマ数）

【 0 0 4 0 】

一般に、乱雑な商品タイトルは、より冗長な情報を含んでいるため、商品タイトルの文長が長いほど、商品タイトルの単語が乱雑である可能性が高くなる。

【 0 0 4 1 】

例えば、商品タイトル「100% Original Asus P6T7 WS SuperComputer Motherboard, ASUS Motherboard, Computer Motherboard, Computer Mainboard, Motherboard」の文長は、18 である。

【 0 0 4 2 】

3 . 商品タイトル内の総単語数に対する、反復単語を除去した後に商品タイトルに含まれる単語数の比

【 0 0 4 3 】

一般に、語幹抽出を受けた商品タイトルについて、商品タイトル内の総単語数に対する、反復単語を除去した後の単語数の比が小さくなるほど、タイトルが乱雑である可能性が高くなる。「語幹抽出」の意味するところは、英単語からの接尾辞の除去および語幹の保持である。語幹抽出の一例は、複数形に関連するすべての接尾辞の除去である（例えば、「laptops」から「s」を除去）。しかしながら、商品タイトルが中国語である時、「語幹抽出」工程は省略される。

【 0 0 4 4 】

例えば、商品タイトル「100% Original Asus P6T7 WS SuperComputer Motherboard, AS US Motherboard, Computer Motherboard, Computer Mainboard, Motherboard」が、接尾辞「e r」の除去を含む語幹抽出を受けた後、対応する単語列は、「100% Origin Asus P 6T7 WS SuperComput Motherboard ASUS Motherboard Comput Motherboard Comput Mainboard Motherboard」（14単語）になる。反復する単語が除去された後、文は「100% Origin Asus P6T7 WS SuperComput Motherboard Comput Mainboard」（9単語）となる。したがって、この例では、反復単語を除去した後の商品タイトル内の単語数の総単語数に対する比は、9 / 14である。

【 0 0 4 5 】

4．商品タイトル内の最頻出単語の出現回数

【 0 0 4 6 】

一般に、或る単語が商品タイトル内に出現する頻度が高くなるほど、商品タイトルが乱雑になる可能性が高くなる。いくつかの実施形態では、最頻出単語が、商品情報の乱雑さの主な原因になる単語であると見なされる。

【 0 0 4 7 】

例えば、商品タイトル「09 branded handbag, designer handbag, new style handbag, fashion handbag, ladies handbag, elegant handbag」が語幹抽出を受けた後、最も出現頻度の高い単語は、「handbag」であり、6回出現している。この例において、この商品タイトルは、単語「handbag」に関して乱雑であると決定される。

【 0 0 4 8 】

5．事前設定されたルールに基づいて商品タイトルがセグメント（セグメントとは、元の商品タイトルのすべての単語／句のサブセットを指す）に分割された後に各セグメント内の特定の位置にある単語で構成されたセット内の総単語数に対する、反復する単語を除去した後の単語数の比

【 0 0 4 9 】

一般に、上述の事前設定されたルールは、以下を含むがそれらに限定されない。商品タイトル中のコンマの位置に基づいて商品タイトルをセグメントに分割する、および／または、商品タイトルにおける最頻出単語の位置に基づいて商品タイトルをセグメントに分割する。上述の2つの方法は、単に例示であって、商品タイトルを分割する他の方法を除外するものではない。

【 0 0 5 0 】

a) セグメント化の形態としてコンマに基づく分割の一例を用いると、商品タイトルが、タイトルに含まれるコンマの位置に基づいてセグメントに分割された後、各セグメント内の最後の単語／句（例えば、分割が行われた商品タイトル内の位置の直前の単語／句）が、セットの要素として指定される。かかるセットにおいて、（反復単語を含む）セット内の総単語数に対する、セットから反復単語を除去した後の単語数の比が小さいほど、商品タイトルに含まれる単語が乱雑である可能性が高くなる。

【 0 0 5 1 】

例えば、商品タイトル「Paypal-Fashion sunglasses, ED sunglasses, CA sunglasses, Brand name sunglasses, designer sunglasses」について、単語が語幹抽出を受けて、タイトルがコンマに基づいて分割された後、結果として得られるセグメントのセットは、{ “Paypal-Fashion sunglass”, “ED sunglass”, “CA sunglass”, “Brand nam sunglass”, “design sunglass” } であり、各セグメントの最後の単語のセットは、{ “sunglass

10

20

30

40

50

”, “sunglass”, “sunglass”, “sunglass”, “sunglass” } である。反復する単語を除去した後、セットに残る唯一の単語は、{ “sunglass” } である。したがって、各セグメント内の最後の単語で構成された単語のセットにおいて、セット内の総単語数に対する、反復単語を除去した後の単語数の比は、 $1 / 5$ である。

【 0 0 5 2 】

b) セグメント化の形態としてコンマに基づく分割の別の例を用いると、商品タイトルが、タイトルに含まれるコンマの位置に基づいて特定の数のセグメントに分割された後、各セグメントの最後の2つの単語 / 句 (例えば、分割が行われた商品タイトル内の位置の直前の最後の2つの単語 / 句) が、セットの要素として指定される。(反復単語を含む) セット内のバイグラム (各セグメント内の最後の2つの単語からなる単語) の総数に対する、反復単語を除去した後のバイグラムの数の比が小さいほど、商品タイトルに含まれる単語が乱雑である可能性が高くなる。

10

【 0 0 5 3 】

例えば、商品タイトル「Degree name card holder, business card holder, name card case, business card case, card holder credit card holder」が語幹抽出とコンマに基づく分割を受けた後、結果として得られるセグメントのセットは、{ “Degree name card hold”, “business card hold”, “name card case”, “business card case”, “card holder”, “credit card holder” } である。各セグメントの最後の2つの単語 / 句からなるセットは、{ “card holder”, “card holder”, “card case”, “card case”, “card holder”, “card holder” } である。反復単語を除去した後のセットは、{ “card holder”, “card case” } である。したがって、反復単語を除去した後のバイグラムの全バイグラムに対する比は、 $1 / 3$ である。

20

【 0 0 5 4 】

c) 最頻出単語に基づいて商品タイトルをセグメントに分割する一例を用いると、商品タイトルが、タイトル内に含まれる最頻出単語に基づいてセグメントに分割された後、各セグメントの最後の単語 / 句が、セットの要素として指定される。一般に、(反復単語を含む) セット内の総単語数に対する、反復単語を除去した後の単語数の比が小さいほど、タイトルに含まれる単語が乱雑である可能性が高くなる。

【 0 0 5 5 】

例えば、商品タイトルが、「New style Brand tshirt Polo tshirt Fashion tshirt mens Top quality tshirt Paypal」であるとする。商品タイトルは、語幹抽出された後、「New styl Brand tshirt Polo tshirt Fashion tshirt men Top qualiti tshirt Payp」となり、最頻出単語は「tshirt」である。文は、「tshirt」を区切り記号として用いて分割される。したがって、結果として得られるセグメントセットは、{ “New styl Brand tshirt”, “Polo tshirt”, “Fashion tshirt”, “men Top qualiti tshirt”, “Payp” } である。各セグメントの最後の単語が要素に指定されたセットは、{ “tshirt”, “tshirt”, “tshirt”, “tshirt”, “Payp” } である。反復単語を除去した後のセットは、{ “Payp” } のみを含む。したがって、各セグメント内の最後の単語で構成されたセットにおいて、セット内の (反復単語を含む) 総単語数に対する、反復単語を除去した後の単語数の比は、 $1 / 5$ である。

30

40

【 0 0 5 6 】

いくつかの実施形態において、上の a)、b)、および、c) に記載したセグメント分割方法の内の1または複数、ならびに、それらに対応する比の計算方法が用いられる。計算結果の精度を高めるために、セグメント分割方法 a)、b)、および、c) の組み合わせを実施してもよい。

【 0 0 5 7 】

6 . 商品タイトルが、事前設定されたルールに基づいてセグメントに分割された後の各セグメントの分散

【 0 0 5 8 】

コンマに基づく分割の別の例を用いると、商品タイトルがコンマの位置に基づいてセグ

50

メントに分割された後、各セグメントは、セグメント長（すなわち、セグメントが含む単語数）に関連付けられる。一般に、商品タイトルから得られたこれらのセグメントのセットについて、セット内でのセグメント長の分散が小さいほど、商品タイトルに含まれる単語が乱雑である可能性が高くなる。

【 0 0 5 9 】

例えば、商品タイトル「Paypal-Fashion sunglasses, ED sunglasses, CA sunglasses, Brand name sunglasses, designer sunglasses」が語幹抽出とコンマに基づく分割を受けた後、結果として得られるセグメントセットは、{ “Paypal-Fashion sunglass”, “ED sunglass”, “CA sunglass”, “Brand nam sunglass”, “design sunglass” } である。セグメントに対応する長さのセットは、{ 2, 2, 2, 3, 2 } であり、セグメント長の分散は 0.2 である。

10

【 0 0 6 0 】

第 2 に、商品タイトルの構文的特性属性が、商品情報から取得される。この処理は、最初に、商品タイトルの品詞タグ付け、すなわち、商品タイトルに含まれる各単語を、対応する品詞（名詞、動詞、形容詞、または、副詞など）でタグ付けすることを含む。品詞のカテゴリは比較的少数である（例えば、ペン・ツリーバンクは、36 の品詞を定義している）。したがって、品詞特性に基づく特徴の方が語彙特性に基づく特徴よりも一般化しやすいため、この技術的スキームの適用範囲を広く解釈できる。いくつかの実施形態において、一般化のレベルをさらに高めるために、品詞の上位カテゴリが定義される。いくつかの実施形態において、品詞の上位カテゴリは、以下のカテゴリで品詞を規定する。名詞（N）、動詞（V）、形容詞（JJ）、副詞（ADV）、前置詞（TO）、数詞（DT）。上記の構文的特性属性の説明に関連して、構文的特性属性に対応する値の例は、以下の内の 1 または複数を含みうるが、それらに限定されない。

20

【 0 0 6 1 】

1. 商品タイトルの単語の総品詞数に対する、反復品詞の除去後に商品タイトルに含まれる単語の品詞数の比

【 0 0 6 2 】

一般に、商品タイトルの単語の総品詞数に対する、反復品詞の除去後に商品タイトルに含まれる単語の品詞数の比が小さいほど、商品タイトルに含まれる単語が乱雑である可能性が高くなる。

30

【 0 0 6 3 】

例えば、商品タイトルが「100% Original Asus P6T7 WS SuperComputer Motherboard, ASUS Motherboard, Computer Motherboard, Computer Mainboard, Motherboard」であると仮定すると、対応する品詞は、「DT JJ N DT N N N, N N, N N, N N, N」になる。反復する品詞を除去すると、品詞のセットは、{ “DT”, “JJ”, “N” } となる。したがって、商品タイトル内の単語の総品詞数に対する、反復品詞を除去した後の品詞数の比は、3 / 14 である。

【 0 0 6 4 】

2. 名詞の総単語数に対する、反復単語を除去した後の商品タイトル内の名詞の単語数の比

40

【 0 0 6 5 】

電子商取引の分野において、商品タイトル内の名詞は、より重要な商品情報を記述するため、より豊富な情報を含む傾向がある。一般に、商品名（例えば、製品名）は名詞である。したがって、一般に、名詞の総数に対する、商品タイトルから反復単語を除去した後の名詞の数の比が小さいほど、商品タイトルに含まれる単語が乱雑である可能性が高くなる。

【 0 0 6 6 】

例えば、商品タイトル「100% Original Asus P6T7 WS SuperComputer Motherboard, ASUS Motherboard, Computer Motherboard, Computer Mainboard, Motherboard」において、名詞は、「Asus WS SuperComputer Motherboard ASUS Motherboard Computer Motherboard

50

d Computer Mainboard Motherboard」であり、反復する単語を除去した後の名詞のセットは、{ “Asus”, “WS”, “SuperComputer”, “Motherboard”, “Mainboard” } である。したがって、商品タイトル内の総名詞数に対する、反復単語の除去後の名詞数の比は、5 / 11 である。

【0067】

3. 最頻出品詞の出現回数

【0068】

句読点のない乱雑な商品タイトルの特定を改善するために、いくつかの実施形態では、1つの品詞が連続して（すなわち、バイグラムとして）出現する頻度が考慮される。一般に、品詞の連続の頻度が高いほど、商品タイトルに含まれる単語が乱雑である可能性が高くなる。

10

【0069】

例えば、商品タイトル「Power Amplifier Audio Amplifier Professional Power Amplifier Karaoke Amplifier Pa Pro Amplifier」について、対応する品詞列は「JJ N JJ N JJ N N N N N N N」であり、そこから抽出されたバイグラム品詞のセットは、{ “JJ N”, “N JJ”, “JJ N”, “N JJ”, “JJ N”, “N N”, “N N”, “N N”, “N N”, “N N”, “N N” } であり、最も頻繁に（7回）出現するバイグラム配列は、“N N” である。

【0070】

4. 事前設定されたルールに基づいて商品情報がセグメント（例えば、商品情報の単語 / 句のサブセット）に分割された後に各セグメント内の1または複数の特定の位置にある単語に対応する品詞を含むセット内の総品詞数に対する、反復単語を除去した後の品詞数の比

20

【0071】

いくつかの実施形態において、事前設定されたルールに基づいた商品情報のセグメントへの分割は、以下を含むがそれらに限定されない。商品情報（例えば、商品タイトルなど）を商品タイトル内のコンマの位置に基づいてセグメントに分割すること、および / または、商品タイトルにおける最頻出単語の位置に基づいて商品タイトルを分割すること。

【0072】

一般に、商品タイトルがセグメントに分割された後、各セグメント内の最後の2つの単語（バイグラム）に対応する品詞が、セットの要素として指定される。このセットにおいて、セット内のバイグラム品詞の総数に対する、反復品詞を除去した後のバイグラム品詞の数の比が小さいほど、商品タイトルに含まれる単語が乱雑である可能性が高くなる。

30

【0073】

例えば、商品タイトルが「100% Original Asus P6T7 WS SuperComputer Motherboard, ASUS Motherboard, Computer Motherboard, Computer Mainboard, Motherboard」であると仮定すると、各セグメント内の最後の2つの単語の品詞からなるセットは、{ “N N”, “N N”, “N N”, “N” } である。（最後のセグメントが1単語しか含まないため、そのバイグラム品詞配列は、“N” である）。反復単語の除去後、セットは、{ “N N”, “N” } となる。したがって、反復品詞を除去した後のバイグラム品詞と、セット内のバイグラム品詞の総数との比は、2 / 4 である。

40

【0074】

図5は、商品情報を解析するための処理の一実施形態を示すフローチャートである。いくつかの実施形態では、処理500は、少なくとも部分的にシステム200を用いることによって実施される。

【0075】

工程502：ユーザによって入力された商品情報が受信される。

【0076】

いくつかの実施形態において、商品情報は、電子商取引ウェブサイトでユーザ（例えば、アカウントを持つ個人）によって入力される。いくつかの実施形態において、1または複数のユーザは、電子商取引ウェブサイトのウェブページで製品を宣伝することによって電

50

子商取引ウェブサイトで製品を販売することができる。例えば、各ユーザは、電子商取引ウェブサイトに1または複数のウェブページを持つことができ、そこで、自らが提供する1または複数の製品を宣伝する。ユーザは、それらの製品に関する商品情報を入力して送信することも可能であり、かかる情報は適切なウェブサイトで公開されうる。例えば、ユーザは、電子商取引ウェブサイトのユーザインターフェースウェブページでユーザが販売している製品の内の1または複数のための商品情報を送信できる。

【0077】

工程504：商品情報から1または複数の特性属性に対応する値を取得することを少なくとも含む、商品情報の解析が実行される。1または複数の特性属性に対応する取得された値は、商品情報が乱雑であるか否かを判定するために用いられる。

10

【0078】

いくつかの実施形態において、特性属性は、形態的特性属性および/または構文的特性属性を含む。

【0079】

いくつかの実施形態において、形態的特性属性の例は、以下の内の任意の1または複数を含む：商品情報に含まれるコンマの数；商品情報の文の長さ；商品情報内の総単語数に対する、反復単語の除去後の商品情報に含まれる単語数の比；商品情報における最頻出単語の出現回数；事前設定されたルールに基づいて商品情報がセグメントに分割された後に各セグメント内の特定の位置にある単語で構成されたセット内の総単語数に対する、反復単語を除去した後の単語数の比；事前設定されたルールに基づいて商品情報がセグメントに分割された後の各セグメントの分散。

20

【0080】

いくつかの実施形態において、構文的特性属性の例は、以下の内の任意の1または複数を含む。商品情報内の単語に対応する品詞の総数に対する反復品詞の除去後に商品情報に含まれる単語に対応する品詞の数の比、名詞の総単語数に対する反復品詞を除去した後の商品情報内の名詞の単語数の比、最頻出品詞の出現回数、事前設定されたルールに基づいて商品情報がセグメントに分割された後に各セグメント内の特定の位置にある単語に対応する品詞からなるセット内の総品詞数に対する、反復品詞を除去した後の品詞数の比。

【0081】

工程506：商品情報に関連する乱雑さの信頼水準が、1または複数の特性属性に対応する取得された値の最大エントロピー原理に少なくとも部分的に基づいて決定される。

30

【0082】

いくつかの実施形態において、取得された1または複数の特性属性の最大エントロピー原理に少なくとも部分的に基づいて、商品情報に関連する乱雑さの信頼水準を決定する工程は、取得された特性属性の値を最大エントロピー原理に基づいた条件付き確率モデル

【数2】

$$p(y|x) = \frac{1}{Z(x)} \exp \left(\sum_j \lambda_j f_j(x, y) \right)$$

の入力情報とし、次いで、条件付き確率モデルを用いて、商品タイトルが乱雑な情報である事後確率 $p(y|x)$ を、与えられた入力情報について計算することを含む。事後確率 $p(y|x)$ は、商品情報に関連する信頼水準と見なされる。

40

【0083】

工程508：商品情報に関連する信頼水準が、事前設定された閾値を超えるか否かが判定される。信頼水準が、事前設定された閾値を超えると判定された場合、商品情報の公開を停止する旨の示唆が送信され、信頼水準が、事前設定された閾値を超えないと判定された場合、商品情報の公開を停止する旨の示唆は送信されない。

【0084】

いくつかの実施形態において、信頼水準の閾値は、システム200のオペレータによって事前に設定される。いくつかの実施形態において、信頼水準が閾値を超えた場合、商品

50

情報は乱雑であると見なされ、信頼水準が閾値を超えない場合、商品情報は乱雑でないと見なされる。信頼水準が、事前設定された閾値を超えると判定されると、商品情報の（例えば、関連するウェブページでの）公開が停止され、いくつかの実施形態では、商品情報の乱雑さの原因となっているキーワードを決定するために解析が行われる。いくつかの実施形態では、キーワードが、商品情報内での最頻出単語である場合に、商品情報の乱雑さの主な理由と見なされる。いくつかの実施形態において、商品情報の乱雑さの主な理由と見なされたキーワードは、（例えば、ユーザインターフェースウェブページで表示することにより）ユーザに返される。次に、ユーザは、このキーワードに関して商品情報を修正するよう促される。例えば、ユーザは、より少ない単語を含む情報および／またはキーワードの反復がより少ない情報など、新たな商品情報を送信できる。いくつかの実施形態において、ユーザは、商品情報の自動修正案を提示されてもよく、ユーザは、公開に向けて送信するために１つを選択するか、または、公開に向けて送信すべき新たな商品情報を作成する際にそれらを参照することができる。

10

【 0 0 8 5 】

処理 5 0 0 は、以下の実験データの例を用いてさらに説明されうる。

【 0 0 8 6 】

いくつかの実施形態において、各特性属性の値は、0 から 1 の間の値に正規化され、次いで、正規化された値は、後の計算処理を単純化するために整数に写像される。例えば、値「6」は、0.3（すなわち、6 / 20、20 は正規化パラメータであり、正規化されたデータの値に基づきうる）に正規化され、整数「3」に写像される。一例において、正規化された値と整数との間の写像関係は、以下の通りである：0 -> 0, (0, 0.05] -> 1, (0.05, 0.15] -> 2, (0.15, 0.3] -> 3, (0.3, 0.5] -> 4, (0.5, 1] -> 5

20

【 0 0 8 7 】

したがって、例えば、商品タイトルが「#24 Baseball Jersey, Baseball Jerseys, Jerseys, Sports Jerseys, Sport Jersey, Jersey, 24# Baseball Jersey」である場合、商品タイトルの解析結果に基づいて取得された特性属性は、以下の値であり、上述のように、式 1 で用いられる。

【 0 0 8 8 】

商品タイトルに含まれるコンマの数は 6 であり、正規化によって 0.3 に変換され、次いで、写像によって 3 に変換される。これは、 $f_1(x, y)$ に対応しており、ここで、 f_1 の仮説値は、0.0653117 であり、 $f_1(x, y)$ の値は、以下の通りである。

30

【 数 3 】

$$f_1(x, y) = \begin{cases} 1 & \text{if } x = \text{コンマ特性IDが3, そして } y = \text{タイトルが乱雑} \\ 0 & \text{その他} \end{cases}$$

【 0 0 8 9 】

商品タイトルの文章の長さは 20 であり、正規化によって 0.20 に変換され、次いで、写像によって整数 2 に変換される。これは、 $f_2(x, y)$ に対応する。 f_2 の仮説値は、0.853789 であり、 $f_2(x, y)$ の値は、以下の通りである。

40

【 数 4 】

$$f_2(x, y) = \begin{cases} 1 & \text{if } x = \text{文長特性IDが2, そして } y = \text{タイトルが乱雑} \\ 0 & \text{その他} \end{cases}$$

【 0 0 9 0 】

商品タイトル内の単語の総数に対する、反復単語を除去した後の商品タイトルに含まれる単語数の比は、4 / 14 であり、正規化によって 0.28 に変換され、次いで、写像によって整数 3 に変換される。これは、 $f_3(x, y)$ に対応する。 f_3 の値は、- 0 .

50

1 7 7 9 4 1 であり、 $f_3(x, y)$ の値は、以下のように仮定される。

【数 5】

$$f_3(x, y) = \begin{cases} 1 & \text{if } x = \text{単語繰返し比特性IDが5, そして } y = \text{タイトルが乱雑} \\ 0 & \text{その他} \end{cases}$$

【0 0 9 1】

商品タイトル内の最頻出単語の出現回数は7であり、正規化によって0.35に変換され、次いで、写像によって3に変換される。これは、 $f_4(x, y)$ に対応する。 f_4 の仮説値は、0.457743であり、 $f_4(x, y)$ の値は、以下の通りである。

10

【数 6】

$$f_4(x, y) = \begin{cases} 1 & \text{if } x = \text{最頻出単語数特性IDが3, そして } y = \text{タイトルが乱雑} \\ 0 & \text{その他} \end{cases}$$

(事前設定されたルールに基づいて商品タイトルがセグメントに分割された後に) 各セグメント内の特定の位置にある単語で構成されたセット内の総単語数に対する、反復単語を除去した後の単語数の比。これは、3つの状況に分けられる。

【0 0 9 2】

タイトルに含まれるコンマの位置に従って商品タイトルが特定の数のセグメントに分割された後に各セグメント内の最後の単語で構成されたセット内の総単語数に対する、反復単語の除去後の単語数の比は、 $1/7$ であり、正規化によって0.14に変換され、次いで、写像によって整数2に変換される。これは、 $f_5(x, y)$ に対応する。 f_5 の仮説値は、1.7743であり、 $f_5(x, y)$ の値は、以下の通りである。

20

【数 7】

$$f_5(x, y) = \begin{cases} 1 & \text{if } x = \text{特性IDが2, そして } y = \text{タイトルが乱雑} \\ 0 & \text{その他} \end{cases}$$

【0 0 9 3】

(タイトルに含まれるコンマの位置に基づいて商品タイトルがセグメントに分割された後に) 各セグメント内の最後の2つの単語で構成されたセット内の総単語数に対する、反復単語の除去後の単語数の比は、 $3/7$ であり、正規化によって0.42に変換され、次いで、写像によって整数4に変換される。これは、 $f_6(x, y)$ に対応する。

30

【0 0 9 4】

f_6 の仮説値は、-0.24332であり、 $f_6(x, y)$ の値は、以下の通りである。

【数 8】

$$f_6(x, y) = \begin{cases} 1 & \text{if } x = \text{特性IDが3, そして } y = \text{タイトルが乱雑} \\ 0 & \text{その他} \end{cases}$$

【0 0 9 5】

(タイトルに含まれる最頻出単語に基づいて商品タイトルがセグメントに分割された後に) 各セグメント内の最後の単語で構成されたセット内の総単語数に対する、反復単語の除去後の単語数の比は、 $2/7$ であり、正規化によって0.29に変換され、次いで、写像によって整数3に変換される。これは、 $f_7(x, y)$ に対応する。 f_7 の仮説値は、0.410227であり、 $f_7(x, y)$ の値は、以下の通りである。

40

【数 9】

$$f_7(x, y) = \begin{cases} 1 & \text{if } x = \text{特性IDが4, そして } y = \text{タイトルが乱雑} \\ 0 & \text{その他} \end{cases}$$

【0 0 9 6】

50

商品タイトルが、事前設定されたルールに基づいてセグメントに分割された後、各セグメントの分散は 0.28 であり、2 に写像される。これは、 $f_8(x, y)$ に対応する。 f_8 の仮説値は、-0.188554 であり、 $f_8(x, y)$ の値は、以下の通りである。

【数 10】

$$f_8(x, y) = \begin{cases} 1 & \text{if } x = \text{特性IDが2, そして } y = \text{タイトルが乱雑} \\ 0 & \text{その他} \end{cases}$$

【0097】

商品タイトル内の単語に対応する品詞の総数に対する、反復品詞の除去後に商品タイトルに含まれる単語に対応する品詞の数の比は、2 / 14 であり、正規化によって 0.14 に変換され、次いで、写像によって整数 2 に変換される。これは、 $f_9(x, y)$ に対応する。 f_9 の仮説値は、-0.0397724 であり、 $f_9(x, y)$ の値は、以下の通りである。

10

【数 11】

$$f_9(x, y) = \begin{cases} 1 & \text{if } x = \text{特性IDが2, そして } y = \text{タイトルが乱雑} \\ 0 & \text{その他} \end{cases}$$

【0098】

名詞である単語の総数に対する、反復品詞の除去後の商品タイトル内の名詞の単語数の比は、3 / 15 であり、正規化によって 0.2 に変換され、次いで、写像によって整数 2 に変換される。これは、 $f_{10}(x, y)$ に対応する。 f_{10} の仮説値は、0.305969 であり、 $f_{10}(x, y)$ の値は、以下の通りである。

20

【数 12】

$$f_{10}(x, y) = \begin{cases} 1 & \text{if } x = \text{特性IDが4, そして } y = \text{タイトルが乱雑} \\ 0 & \text{その他} \end{cases}$$

【0099】

最頻出品詞の出現回数は 12 であり、正規化によって 0.6 に変換され、次いで、写像によって整数 6 に変換される。これは、 $f_{11}(x, y)$ に対応する。 f_{11} の仮説値は、0.105729 であり、 $f_{11}(x, y)$ の値は、以下の通りである。

30

【数 13】

$$f_{11}(x, y) = \begin{cases} 1 & \text{if } x = \text{特性IDが24, そして } y = \text{タイトルが乱雑} \\ 0 & \text{その他} \end{cases}$$

【0100】

(商品情報がセグメントに分割された後に) 各セグメント内の特定の位置にある品詞で構成されたセット内の品詞の総数に対する、反復品詞の除去後の品詞数の比は、2 / 7 であり、正規化によって 0.28 に変換され、次いで、写像によって整数 3 に変換される。これは、 $f_{12}(x, y)$ に対応する。 f_{12} の仮説値は、-0.174333 であり、 $f_{12}(x, y)$ の値は、以下の通りである。

40

【数 14】

$$f_{12}(x, y) = \begin{cases} 1 & \text{if } x = \text{特性IDが2, そして } y = \text{タイトルが乱雑} \\ 0 & \text{その他} \end{cases}$$

【0101】

式 1 に与えられた入力情報として上述の特性属性に基づく、事後確率 $p(y | x)$ は 0.989271 であり、仮説閾値は 0.7 である。信頼水準として機能する事後確率は、閾値を超えている。したがって、ユーザによって入力された商品タイトルに含まれる単語は乱雑であり、公開を停止すべきであると判定される。特性属性の利用に関する上述の

50

説明は、一例にすぎず、特性属性の任意のサブセットを用いて、商品情報の信頼水準（例えば、事後確率）を計算することができる。

【 0 1 0 2 】

当業者は、本願の精神および範囲から逸脱することなく、開示された実施形態を変形および変更することができる。したがって、本願のこれらの変形例および変更例が、特許請求の範囲および等価の技術の範囲内にある場合、本願は、これらの変形例および変更例をも網羅するものである。

【 0 1 0 3 】

上述の実施形態は、理解しやすいようにいくぶん詳しく説明されているが、本発明は、提供された詳細事項に限定されるものではない。本発明を実施する多くの代替方法が存在する。開示された実施形態は、例示であり、限定を意図するものではない。

適用例 1：商品情報を解析する方法であって、ユーザによって入力された商品情報を受信し、前記商品情報から 1 または複数の特性属性に対応する値を取得することを少なくとも含む、前記商品情報の解析を実行し、1 または複数の特性属性に対応する前記値は、前記商品情報が乱雑であるか否かを判定するために用いられ、1 または複数の特性属性に対応する前記取得された値に少なくとも部分的に基づいて、前記商品情報に関連する乱雑さの信頼水準を決定し、前記商品情報に関連する前記乱雑さの信頼水準が、事前設定された閾値を超えるか否かを判定し、前記乱雑さの信頼水準が前記事前設定された閾値を超えた場合には、前記商品情報の公開を停止する旨の示唆を送信し、前記乱雑さの信頼水準が前記事前設定された閾値を超えていない場合には、前記商品情報の公開を停止する旨の示唆を送信しない、こと、を備える、方法。

適用例 2：適用例 1 に記載の方法であって、前記商品情報は、電子商取引ウェブサイトに関連して受信される、方法。

適用例 3：適用例 1 に記載の方法であって、前記商品情報は、商品タイトル、商品説明情報、商品紹介情報、商品レビュー、および、商品の製品仕様の内の 1 または複数を含む、方法。

適用例 4：適用例 1 に記載の方法であって、1 または複数の特性属性に対応する前記取得された値に少なくとも部分的に基づいて、前記商品情報に関連する乱雑さの信頼水準を決定することは、1 または複数の特性属性に対応する前記取得された値を条件付き確率モデルに入力し、1 または複数の特性属性に対応する前記取得された値と、前記条件付き確率モデルとを少なくとも用いて、前記商品情報が乱雑である可能性に関連する事後確率を計算すること、を含み、前記乱雑さの信頼水準は、前記事後確率を含む、方法。

適用例 5：適用例 1 に記載の方法であって、前記 1 または複数の特性属性は、少なくとも 1 つの形態的特性属性を含む、方法。

適用例 6：適用例 5 に記載の方法であって、前記少なくとも 1 つの形態的特性属性は、前記商品情報に含まれるコンマの数、前記商品情報の文の長さ、前記商品情報内の総単語数に対する反復単語の除去後の前記商品情報に含まれる単語数の比、前記商品情報における最頻出単語の出現回数、事前設定されたルールに基づいて前記商品情報がセグメントに分割された後に各セグメント内の特定の位置にある単語で構成されたセット内の総単語数に対する反復単語の除去後の単語数の比、事前設定されたルールに基づいて前記商品情報がセグメントに分割された後の各セグメントの分散、の内の 1 または複数を含む、方法。

適用例 7：適用例 1 に記載の方法であって、前記 1 または複数の特性属性は、少なくとも 1 つの構文的特性属性を含む、方法。

適用例 8：適用例 7 に記載の方法であって、前記少なくとも 1 つの構文的特性属性は、前記商品情報内の単語に対応する品詞の総数に対する反復品詞の除去後の前記商品情報に含まれる単語に対応する品詞の数の比、名詞の総単語数に対する反復単語を除去した後の前記商品情報内の名詞の単語数の比、最頻出品詞の出現回数、事前設定されたルールに基づいて前記商品情報がセグメントに分割された後に各セグメント内の特定の位置にある単語に対応する品詞で構成されたセット内の総品詞数に対する、反復品詞を除去した後の品詞数の比、の内の 1 または複数を含む、方法。

10

20

30

40

50

適用例 9：適用例 6 に記載の方法であって、さらに、事前設定されたルールに基づいて前記商品情報をセグメントに分割することを備え、前記ルールは、前記商品情報内のコンマの位置に基づいて前記商品情報を分割して、前記商品情報に含まれる単語のサブセットを含む 1 または複数のセグメントを形成すること、および / または、前記商品情報における最頻出単語の位置に基づいて前記商品情報を分割して、1 または複数のセグメントを形成することを含む、方法。

適用例 10：適用例 8 に記載の方法であって、さらに、事前設定されたルールに基づいて前記商品情報をセグメントに分割することを備え、前記ルールは、前記商品情報内のコンマの位置に基づいて前記商品情報を分割して、前記商品情報に含まれる単語のサブセットを含む 1 または複数のセグメントを形成すること、および / または、前記商品情報における最頻出単語の位置に基づいて前記商品情報を分割して、1 または複数のセグメントを形成することを含む、方法。

10

適用例 11：適用例 1 に記載の方法であって、前記乱雑さの信頼水準が前記事前設定された閾値を超えた場合に、前記商品情報が乱雑な商品情報を含むと判定することを備える、方法。

適用例 12：適用例 11 に記載の方法であって、前記乱雑さの信頼水準が前記事前設定された閾値を超えた場合に、前記商品情報に関連する乱雑さを引き起こした可能性のある前記商品情報のキーワードを決定し、前記ユーザによってアクセス可能なインターフェース要素を介して前記キーワードに関する示唆を提示すること、をさらに備える、方法。

適用例 13：適用例 12 に記載の方法であって、さらに、前記インターフェース要素を介して前記商品情報への修正を入力することを前記ユーザに促すことを備える、方法。

20

適用例 14：商品情報を解析するためのシステムであって、プロセッサであって、ユーザによって入力された商品情報を受信し、前記商品情報から 1 または複数の特性属性に対応する値を取得することを少なくとも含む、前記商品情報の解析を実行し、1 または複数の特性属性に対応する前記値は、前記商品情報が乱雑であるか否かを判定するために用いられ、1 または複数の特性属性に対応する前記取得された値に少なくとも部分的に基づいて、前記商品情報に関連する乱雑さの信頼水準を決定し、前記商品情報に関連する前記乱雑さの信頼水準が、事前設定された閾値を超えるか否かを判定し、前記乱雑さの信頼水準が前記事前設定された閾値を超えた場合には、前記商品情報の公開を停止する旨の示唆を送信し、前記乱雑さの信頼水準が前記事前設定された閾値を超えていない場合には、前記商品情報の公開を停止する旨の示唆を送信しないように構成されているプロセッサと、前記プロセッサに接続され、前記プロセッサに命令を提供するように構成されているメモリと、を備える、システム。

30

適用例 15：適用例 14 に記載のシステムであって、前記商品情報は、電子商取引ウェブサイトに関連して受信される、システム。

適用例 16：適用例 14 に記載のシステムであって、前記商品情報は、商品タイトル、商品説明情報、商品紹介情報、商品レビュー、および、商品の製品仕様の内の 1 または複数を含む、システム。

適用例 17：適用例 14 に記載のシステムであって、1 または複数の特性属性に対応する前記取得された値に少なくとも部分的に基づいて、前記商品情報に関連する乱雑さの信頼水準を決定するように構成されている前記プロセッサは、1 または複数の特性属性に対応する前記取得された値を条件付き確率モデルに入力し、1 または複数の特性属性に対応する前記取得された値と、前記条件付き確率モデルとを少なくとも用いて、前記商品情報が乱雑である可能性に関連する事後確率を計算するように構成されているプロセッサを含み、前記乱雑さの信頼水準は、前記事後確率を含む、システム。

40

適用例 18：適用例 14 に記載のシステムであって、前記 1 または複数の特性属性は、少なくとも 1 つの形態的特性属性を含む、システム。

適用例 19：適用例 14 に記載のシステムであって、前記 1 または複数の特性属性は、少なくとも 1 つの構文的特性属性を含む、システム。

適用例 20：適用例 14 に記載のシステムであって、前記乱雑さの信頼水準が前記事前

50

設定された閾値を超えた場合に、前記プロセッサは、前記商品情報が乱雑な商品情報を含むと判定するように構成されている、システム。

適用例 2 1：適用例 2 0 に記載のシステムであって、前記乱雑さの信頼水準が前記事前設定された閾値を超えた場合に、前記プロセッサは、さらに、前記商品情報に関連する乱雑さを引き起こした可能性のある前記商品情報のキーワードを決定し、前記ユーザによってアクセス可能なインターフェース要素を介して前記キーワードに関する示唆を提示するように構成されている、システム。

適用例 2 2：適用例 2 1 に記載のシステムであって、前記プロセッサは、さらに、前記インターフェース要素を介して前記商品情報への修正を入力することを前記ユーザに促すように構成されている、システム。

適用例 2 3：商品情報を解析するためのコンピュータプログラム製品であって、前記コンピュータプログラム製品は、コンピュータ読み取り可能な記憶媒体内に具現化され、ユーザによって入力された商品情報を受信するためのコンピュータ命令と、前記商品情報から 1 または複数の特性属性に対応する値を取得することを少なくとも含む、前記商品情報の解析を実行するためのコンピュータ命令と、1 または複数の特性属性に対応する前記値は、前記商品情報が乱雑であるか否かを判定するために用いられ、1 または複数の特性属性に対応する前記取得された値に少なくとも部分的に基づいて、前記商品情報に関連する乱雑さの信頼水準を決定するためのコンピュータ命令と、前記商品情報に関連する前記乱雑さの信頼水準が、事前設定された閾値を超えるか否かを判定するためのコンピュータ命令と、前記乱雑さの信頼水準が前記事前設定された閾値を超えた場合に、前記商品情報の公開を停止する旨の示唆を送信し、前記乱雑さの信頼水準が前記事前設定された閾値を超えていない場合に、前記商品情報の公開を停止する旨の示唆を送信しないコンピュータ命令と、を備える、コンピュータプログラム製品。

10

20

【図 1】



&New arrived&Fashion wind coat,ladies' coat,fashion coat, women's wind coat(Wholesa price +Do dropship) 4 Similar

Super Hot, "Wind coat", new arrived ladies' winter coat, Nice and fashion style, top quality, Wholesale price !...

Min. Order: 5 Pieces FOB Price: US \$35-45

FIG. 1

【図 2】

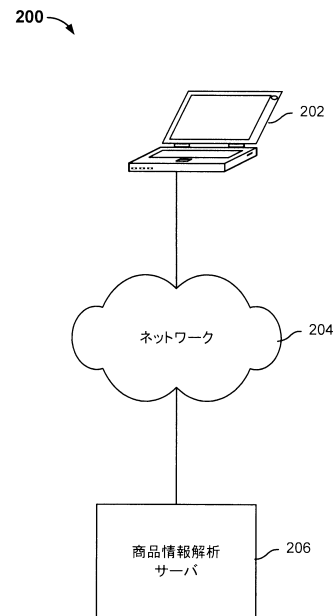


FIG. 2

【図 3】

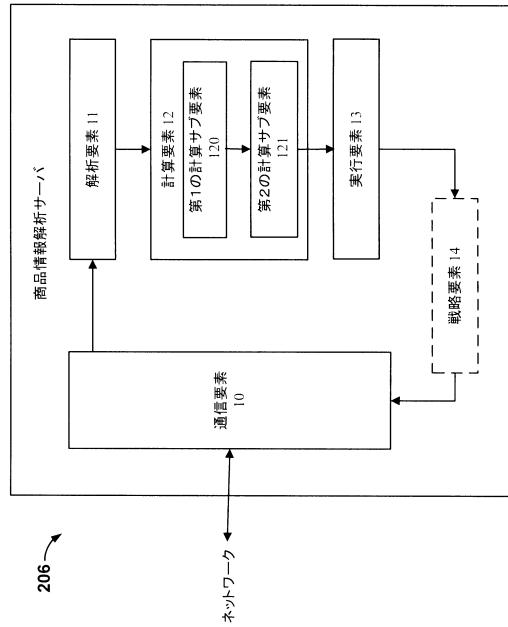


FIG. 3

【図 4】

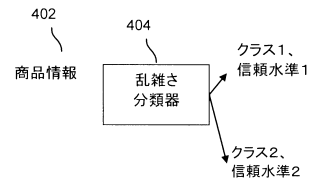


FIG. 4

【図 5】

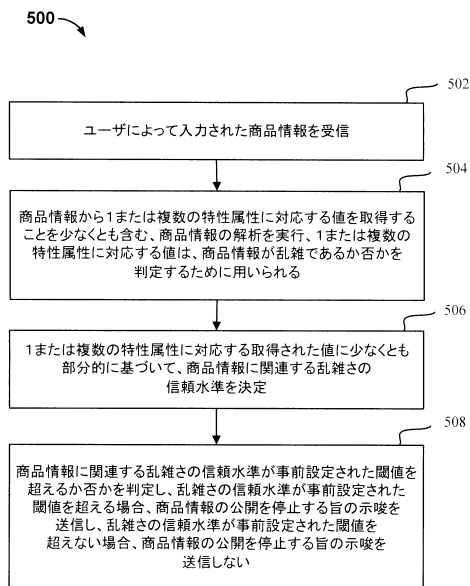


FIG. 5

フロントページの続き

(72)発明者 リン・フォン

中華人民共和国 ハンチョウ, ワーナー・ロード, ザ・ウエスト・レイク・インターナショナル・
プラザ・オブ・エス アンド ティー, ビルディング エー, 10階, ナンバー391, アリババ
・グループ・リーガル・デパートメント内

(72)発明者 チャン・ショウソーン

中華人民共和国 ハンチョウ, ワーナー・ロード, ザ・ウエスト・レイク・インターナショナル・
プラザ・オブ・エス アンド ティー, ビルディング エー, 10階, ナンバー391, アリババ
・グループ・リーガル・デパートメント内

(72)発明者 チャン・チン

中華人民共和国 ハンチョウ, ワーナー・ロード, ザ・ウエスト・レイク・インターナショナル・
プラザ・オブ・エス アンド ティー, ビルディング エー, 10階, ナンバー391, アリババ
・グループ・リーガル・デパートメント内

審査官 野崎 大進

(56)参考文献 米国特許出願公開第2009/0083096 (US, A1)

特開2008-097230 (JP, A)

特開平07-021201 (JP, A)

特開平01-231164 (JP, A)

(58)調査した分野(Int.Cl., DB名)

G06Q 10/00 50/34