

(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2020-126472
(P2020-126472A)

(43) 公開日 令和2年8月20日(2020.8.20)

(51) Int. Cl.	F I	テーマコード (参考)
G06F 16/00 (2019.01)	G06F 17/30 340A	
G06F 16/30 (2019.01)	G06F 17/30 170A	
	G06F 17/30 210A	

審査請求 未請求 請求項の数 9 O L (全 16 頁)

(21) 出願番号 特願2019-18849 (P2019-18849)
(22) 出願日 平成31年2月5日(2019.2.5)

(71) 出願人 504133110
国立大学法人電気通信大学
東京都調布市調布ケ丘一丁目5番地1
(74) 代理人 100083806
弁理士 三好 秀和
(74) 代理人 100101247
弁理士 高橋 俊一
(72) 発明者 坂本 真樹
東京都調布市調布ケ丘一丁目5番地1 国立大学法人電気通信大学内
(72) 発明者 多田 佳歩
東京都調布市調布ケ丘一丁目5番地1 国立大学法人電気通信大学内

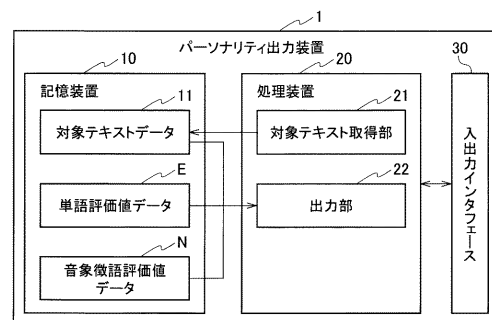
(54) 【発明の名称】 パーソナリティ出力装置、パーソナリティ出力方法、パーソナリティ出力プログラム、単語評価値生成装置、単語評価値生成方法および単語評価値生成プログラム

(57) 【要約】

【課題】テキストデータから、直感的に評価可能なパーソナリティを出力する。

【解決手段】パーソナリティ出力装置1は、音象徴語と、音象徴語の印象評価値を対応づけた音象徴語評価値データNと、単語と、単語の印象評価値を対応づけた単語評価値データEとを記憶する記憶装置10と、テキストデータに含まれる各単語から特定されるユーザのパーソナリティの印象評価値と音象徴語評価値データNの印象評価値との類似度に基づいて、音象徴語を出力する出力部22を備える。

【選択図】 図1



【特許請求の範囲】**【請求項 1】**

ユーザが入力したテキストデータから、前記ユーザのパーソナリティを出力するパーソナリティ出力装置であって、

音象徴語と、前記音象徴語の印象評価値を対応づけた音象徴語評価値データと、
単語と、前記単語の印象評価値を対応づけた単語評価値データとを記憶する記憶装置と

、
テキストデータに含まれる各単語から特定されるユーザのパーソナリティの印象評価値と前記音象徴語評価値データの印象評価値との類似度に基づいて、音象徴語を出力する出力部

を備えることを特徴とするパーソナリティ出力装置。

【請求項 2】

前記出力部は、所定期間毎に、前記所定期間に入力されたテキストデータについて音象徴語を出力する

ことを特徴とする請求項 1 に記載のパーソナリティ出力装置。

【請求項 3】

前記単語評価値データは、

テキストデータを入力した際のユーザのパーソナリティを表す音象徴語と、前記テキストデータに含まれる単語のうち前記音象徴語に対応するプリミティブワードを対応づけたプリミティブワードデータを参照して、

前記プリミティブワードに対応する前記音象徴語の印象評価値を、前記プリミティブワードの印象評価値として設定し、

プリミティブワード以外の未知語について、前記プリミティブワードの印象評価値と、前記プリミティブワードと前記未知語との類似度の乗算から、前記未知語の評価値を算出し、

前記プリミティブワードと前記プリミティブワードの印象評価値を対応づけるとともに、前記未知語と前記未知語の印象評価値を対応づけて、生成される

ことを特徴とする請求項 1 または 2 に記載のパーソナリティ出力装置。

【請求項 4】

ユーザが入力したテキストデータから、前記ユーザのパーソナリティを出力するパーソナリティ出力方法であって、

コンピュータが、音象徴語と、前記音象徴語の印象評価値を対応づけた音象徴語評価値データと、単語と、前記単語の印象評価値を対応づけた単語評価値データとを、記憶装置に記憶するステップと、

前記コンピュータが、テキストデータに含まれる各単語から特定されるユーザのパーソナリティの印象評価値と前記音象徴語評価値データの印象評価値との類似度に基づいて、音象徴語を出力するステップ

を備えることを特徴とするパーソナリティ出力方法。

【請求項 5】

コンピュータに、請求項 1 ないし請求項 3 のいずれか 1 項に記載のパーソナリティ出力装置として機能させるためのパーソナリティ出力プログラム。

【請求項 6】

単語と、前記単語の印象評価値を対応づけた単語評価値データを生成する単語評価値生成装置であって、

テキストデータを入力した際のユーザのパーソナリティを表す音象徴語と、前記テキストデータに含まれる単語のうち前記音象徴語に対応するプリミティブワードを対応づけたプリミティブワードデータを記憶する記憶装置と、

前記プリミティブワードに対応する前記音象徴語の印象評価値を、前記プリミティブワードの印象評価値として設定するプリミティブワード評価値設定部と、

プリミティブワード以外の未知語について、前記プリミティブワードの印象評価値と、

10

20

30

40

50

前記プリミティブワードと前記未知語との類似度の乗算から、前記未知語の評価値を算出する未知語評価値算出部と、

前記プリミティブワードと前記プリミティブワードの印象評価値を対応づけるとともに、前記未知語と前記未知語の印象評価値を対応づけて、単語評価値データを生成する単語評価値データ生成部

を備えることを特徴とする単語評価値生成装置。

【請求項 7】

前記未知語評価値算出部は、プリミティブワードを含むテキストデータと未知語を含むテキストデータの比較に基づいて、前記プリミティブワードと前記未知語との類似度を算出する

ことを特徴とする請求項 6 に記載の単語評価値生成装置。

【請求項 8】

単語と、前記単語の印象評価値を対応づけた単語評価値データを生成する単語評価値生成方法であって、

コンピュータが、テキストデータを入力した際のユーザのパーソナリティを表す音象徴語と、前記テキストデータに含まれる単語のうち前記音象徴語に対応するプリミティブワードを対応づけたプリミティブワードデータを、記憶装置に記憶するステップと、

前記コンピュータが、前記プリミティブワードに対応する前記音象徴語の印象評価値を、前記プリミティブワードの印象評価値として設定するステップと、

前記コンピュータが、プリミティブワード以外の未知語について、前記プリミティブワードの印象評価値と、前記プリミティブワードと前記未知語との類似度の乗算から、前記未知語の評価値を算出するステップと、

前記コンピュータが、前記プリミティブワードと前記プリミティブワードの印象評価値を対応づけるとともに、前記未知語と前記未知語の印象評価値を対応づけて、単語評価値データを生成するステップ

を備えることを特徴とする単語評価値生成方法。

【請求項 9】

コンピュータに、請求項 6 または請求項 7 に記載の単語評価値生成装置として機能させるための単語評価値生成プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ユーザが入力したテキストデータから、ユーザのパーソナリティを出力するパーソナリティ出力装置、パーソナリティ出力方法およびパーソナリティ出力プログラム、パーソナリティ出力装置に用いられる単語評価値データを生成する単語評価値生成装置、単語評価値生成方法および単語評価値生成プログラムに関する。

【背景技術】

【0002】

近年、日々の所感を投稿するマイクロブログ等が普及している。またマイクロブログの投稿内容から、投稿者のパーソナリティを推定する技術もある。

【0003】

またオノマトペ等の音象徴語を用いて、パーソナリティを評価する方法がある（特許文献 1）。特許文献 1 は、音象徴語に含まれる音韻の要素に対して、人のパーソナリティに関する複数の評価尺度に基づいて音象徴語を評価する。

【先行技術文献】

【特許文献】

【0004】

【特許文献1】特開2018-22331号公報

【発明の概要】

【発明が解決しようとする課題】

【0005】

しかしながら、投稿者のパーソナリティの推定結果は、ダイアグラムでの表示、文章での表示などが多く、パーソナリティを直感的に把握することが困難な場合がある。また特許文献1に記載の方法では、マイクロブログ等のテキストデータからパーソナリティを推定するものではない。

【0006】

従って本発明の目的は、テキストデータから、直感的に把握可能なパーソナリティを出力するパーソナリティ出力装置パーソナリティ出力方法およびパーソナリティ出力プログラム、パーソナリティ出力装置に用いられる単語評価値データを生成する単語評価値生成装置、単語評価値生成方法および単語評価値生成プログラムを提供することである。

10

【課題を解決するための手段】

【0007】

上記課題を解決するために、本発明の第1の特徴は、ユーザが入力したテキストデータから、パーソナリティ出力装置ユーザのパーソナリティを出力するパーソナリティ出力装置に関する。本発明の第1の特徴に係る音象徴語と、パーソナリティ出力装置音象徴語の印象評価値を対応づけた音象徴語評価値データと、単語と、パーソナリティ出力装置単語の印象評価値を対応づけた単語評価値データとを記憶する記憶装置と、テキストデータに含まれる各単語から特定されるユーザのパーソナリティの印象評価値とパーソナリティ出力装置音象徴語評価値データの印象評価値との類似度に基づいて、音象徴語を出力する出力部を備える。

20

【0008】

出力部は、所定期間毎に、パーソナリティ出力装置所定期間に入力されたテキストデータについて音象徴語を出力しても良い。

【0009】

単語評価値データは、テキストデータを入力した際のユーザのパーソナリティを表す音象徴語と、パーソナリティ出力装置テキストデータに含まれる単語のうちパーソナリティ出力装置音象徴語に対応するプリミティブワードを対応づけたプリミティブワードデータを参照して、プリミティブワードに対応するパーソナリティ出力装置音象徴語の印象評価値を、パーソナリティ出力装置プリミティブワードの印象評価値として設定し、プリミティブワード以外の未知語について、パーソナリティ出力装置プリミティブワードの印象評価値と、パーソナリティ出力装置プリミティブワードとパーソナリティ出力装置未知語との類似度の乗算から、パーソナリティ出力装置未知語の評価値を算出し、プリミティブワードとパーソナリティ出力装置プリミティブワードの印象評価値を対応づけるとともに、パーソナリティ出力装置未知語とパーソナリティ出力装置未知語の印象評価値を対応づけて、生成されても良い。

30

【0010】

本発明の第2の特徴は、ユーザが入力したテキストデータから、パーソナリティ出力装置ユーザのパーソナリティを出力するパーソナリティ出力方法に関する。本発明の第2の特徴に係るパーソナリティ出力方法は、コンピュータが、音象徴語と、パーソナリティ出力装置音象徴語の印象評価値を対応づけた音象徴語評価値データと、単語と、パーソナリティ出力装置単語の印象評価値を対応づけた単語評価値データとを、記憶装置に記憶するステップと、コンピュータが、テキストデータに含まれる各単語から特定されるユーザのパーソナリティの印象評価値とパーソナリティ出力装置音象徴語評価値データの印象評価値との類似度に基づいて、音象徴語を出力するステップを備える。

40

【0011】

本発明の第3の特徴は、コンピュータに、本発明の第1の特徴に記載のパーソナリティ出力装置として機能させるためのパーソナリティ出力プログラムに関する。

50

【0012】

本発明の第4の特徴は、単語と、パーソナリティ出力装置単語の印象評価値を対応づけた単語評価値データを生成する単語評価値生成装置に関する。本発明の第4の特徴に係る単語評価値生成装置は、テキストデータを入力した際のユーザのパーソナリティを表す音象徴語と、パーソナリティ出力装置テキストデータに含まれる単語のうちパーソナリティ出力装置音象徴語に対応するプリミティブワードを対応づけたプリミティブワードデータを記憶する記憶装置と、プリミティブワードに対応するパーソナリティ出力装置音象徴語の印象評価値を、パーソナリティ出力装置プリミティブワードの印象評価値として設定するプリミティブワード評価値設定部と、プリミティブワード以外の未知語について、パーソナリティ出力装置プリミティブワードの印象評価値と、パーソナリティ出力装置プリミティブワードとパーソナリティ出力装置未知語との類似度の乗算から、パーソナリティ出力装置未知語の評価値を算出する未知語評価値算出部と、プリミティブワードとパーソナリティ出力装置プリミティブワードの印象評価値を対応づけるとともに、パーソナリティ出力装置未知語とパーソナリティ出力装置未知語の印象評価値を対応づけて、単語評価値データを生成する単語評価値データ生成部を備える。

10

【0013】

未知語評価値算出部は、プリミティブワードを含むテキストデータと未知語を含むテキストデータの比較に基づいて、パーソナリティ出力装置プリミティブワードとパーソナリティ出力装置未知語との類似度を算出して良い。

【0014】

本発明の第5の特徴は、単語と、パーソナリティ出力装置単語の印象評価値を対応づけた単語評価値データを生成する単語評価値生成方法に関する。本発明の第5の特徴に係る単語評価値生成方法は、コンピュータが、テキストデータを入力した際のユーザのパーソナリティを表す音象徴語と、パーソナリティ出力装置テキストデータに含まれる単語のうちパーソナリティ出力装置音象徴語に対応するプリミティブワードを対応づけたプリミティブワードデータを、記憶装置に記憶するステップと、コンピュータが、パーソナリティ出力装置プリミティブワードに対応するパーソナリティ出力装置音象徴語の印象評価値を、パーソナリティ出力装置プリミティブワードの印象評価値として設定するステップと、コンピュータが、プリミティブワード以外の未知語について、パーソナリティ出力装置プリミティブワードの印象評価値と、パーソナリティ出力装置プリミティブワードとパーソナリティ出力装置未知語との類似度の乗算から、パーソナリティ出力装置未知語の評価値を算出するステップと、コンピュータが、パーソナリティ出力装置プリミティブワードとパーソナリティ出力装置未知語とパーソナリティ出力装置未知語の印象評価値を対応づけるとともに、パーソナリティ出力装置未知語とパーソナリティ出力装置未知語の印象評価値を対応づけて、単語評価値データを生成するステップを備える。

20

30

【0015】

本発明の第6の特徴は、コンピュータに、本発明の第4の特徴に記載の単語評価値生成装置として機能させるための単語評価値生成プログラムに関する。

【発明の効果】

【0016】

本発明によれば、テキストデータから、直感的に把握可能なパーソナリティを出力するパーソナリティ出力装置パーソナリティ出力方法およびパーソナリティ出力プログラム、パーソナリティ出力装置に用いられる単語評価値データを生成する単語評価値生成装置、単語評価値生成方法および単語評価値生成プログラムを提供することができる。

40

【図面の簡単な説明】

【0017】

【図1】本発明の実施の形態に係るパーソナリティ出力装置のハードウェア構成と機能ブロックを説明する図である。

【図2】本発明の実施の形態に係るパーソナリティ出力装置が出力する結果表示画面の一例である。

50

【図 3】単語評価値データのデータ構造の一例を説明する図である。

【図 4】印象評価値を説明する図である。

【図 5】音象徴語評価値データのデータ構造の一例を説明する図である。

【図 6】本発明の実施の形態に係るパーソナリティ出力装置の出力部の処理を説明するフローチャートである。

【図 7】本発明の実施の形態に係る単語評価値生成装置のハードウェア構成と機能ブロックを説明する図である。

【図 8】プリミティブワードデータのデータ構造とデータの一例を説明する図である。

【図 9】類似度データのデータ構造の一例を説明する図である。

【図 10】本発明の実施の形態に係る単語評価値生成装置の未知語評価値算出部において類似度を算出する処理を説明するフローチャートである。

10

【図 11】類似度を算出する処理で参照される単語文章行列のデータ構造とデータの一例を説明する図である。

【発明を実施するための形態】

【0018】

次に、図面を参照して、本発明の実施の形態を説明する。以下の図面の記載において、同一または類似の部分には同一または類似の符号を付している。

【0019】

(パーソナリティ出力装置)

図 1 に示すパーソナリティ出力装置 1 は、ミニブログ、マイクロブログ等のユーザが入力したテキストデータから、ユーザのパーソナリティを出力する。パーソナリティ出力装置 1 は、そのテキストデータを入力した際のユーザの感情の詳細を分析し、オノマトペに代表される音象徴語で表現する。本発明の実施の形態において音象徴語は、語音そのものが、ある特定のイメージを喚起する語であって、「おろおろ」、「ほんわか」等のオノマトペに代表される。パーソナリティ出力装置 1 は、ダイアグラムまたは文章等の煩雑な表現ではなく、1 単語の音象徴語でユーザの感情を表現することにより、簡潔かつわかりやすく、ユーザの心情を伝えることができる。

20

【0020】

パーソナリティ出力装置 1 は、例えば図 2 に示す結果表示画面 P 1 を出力する。マイクロブログのユーザアカウント等のテキストデータを特定するためのテキスト特定情報が入力され、入力ボタン B 1 が選択されると、パーソナリティ出力装置 1 は、テキスト特定情報から、パーソナリティ出力対象のテキストデータを取得する。テキスト特定情報には、例えば、ログイン ID など、マイクロブログの管理サイトにおいて、対象となるユーザの投稿を特定する情報が設定される。

30

【0021】

パーソナリティ出力装置 1 は、取得したテキストデータにおいてユーザが用いた単語からユーザのパーソナリティを分析し、パーソナリティ表示部 D 1 に示すように、音象徴語でユーザのパーソナリティを出力する。図 2 に示す例では、1 ヶ月単位に「おろおろ」、「ほんわか」および「にこにこ」とパーソナリティを出力しており、パーソナリティの遷移も示す。処理数表示部 D 2 に示すように、各月においてパーソナリティを算出するために用いられたテキストデータ数(マイクロブログの投稿数)が表示されても良い。

40

【0022】

図 1 を参照して、本発明の実施の形態に係るパーソナリティ出力装置 1 を説明する。パーソナリティ出力装置 1 は、記憶装置 10、処理装置 20、入出力インタフェース 30 を備える一般的なコンピュータである。一般的なコンピュータがパーソナリティ出力プログラムを実行することにより、図 1 に示す機能を実現する。

【0023】

記憶装置 10 は、ROM (Read Only Memory)、RAM (Random access memory)、ハードディスク等であって、処理装置 20 が処理を実行するための入力データ、出力データおよび中間データなどの各種データを記憶する。処理装置 20 は、CPU (Central Proc

50

essing Unit) であって、記憶装置 10 に記憶されたデータを読み書きしたり、入出力インタフェース 30 とデータをやりとりしたりして、パーソナリティ出力装置 1 における処理を実行する。入出力インタフェース 30 は、マウス、キーボード等の入力装置、ディスプレイ、プリンタ等の出力装置、通信ネットワークに接続するための通信制御装置等と、処理装置 20 を接続する。入出力インタフェース 30 は、入力装置または通信ネットワークから入力された指示等を処理装置 20 に入力し、処理装置 20 の出力結果等を出力装置または通信ネットワークに出力する。

【0024】

記憶装置 10 は、パーソナリティ出力プログラムを記憶するとともに、対象テキストデータ 11、単語評価値データ E および音象徴語評価値データ N を記憶する。

10

【0025】

対象テキストデータ 11 は、パーソナリティ出力装置 1 がユーザのパーソナリティを出力する対象のテキストデータである。対象テキストデータ 11 は、特定の事実に対する解説など投稿するユーザによって差異が出にくい内容よりも、単一のユーザによる日々の所感など、ユーザの個性が出やすく、パーソナリティが現れやすい内容であることが好ましい。

【0026】

また対象テキストデータ 11 は、マイクロブログなどの 1 つの投稿または一文であっても良いし、数ヶ月などの長期に亘って入力されたテキストデータの集合であっても良い。パーソナリティ出力装置 1 は、数ヶ月間におけるパーソナリティを出力することも可能であるし、1ヶ月等の期間毎のパーソナリティを出力し、パーソナリティの遷移を出力することも可能である。

20

【0027】

単語評価値データ E は、図 3 に示すように、単語と、単語の印象評価値を対応づけたデータである。本発明の実施の形態において印象評価値は、図 4 に示すように「無口な - 話好きな」、「臆病な - 果敢な」などの性格を表す形容詞対毎の評価値を含む。音象徴語の印象評価値は、音韻の語中の位置によって、音象徴語の印象が異なる可能性がある。従って音象徴語は第 1 モーラと第 2 モーラに分解され、各モーラについて、子音行、濁音、拗音、小母音、母音、語尾の要素と反復の有無等を特定した結果に基づいて、音象徴語の印象評価値が算出される。単語評価値データ E に登録される単語は、単語評価値データ E を生成する際に最初に登録されたプリミティブワードと、プリミティブワードに対してさらに追加された未知語に区分されても良い。

30

【0028】

単語評価値データ E は、後述の単語評価値生成装置 6 によって生成される。単語評価値生成装置 6 は、単語評価値データ E を生成する際、テキストデータを入力した際のユーザのパーソナリティを表す音象徴語と、テキストデータに含まれる単語のうち音象徴語に対応するプリミティブワードを対応づけたプリミティブワードデータを参照して、プリミティブワードに対応する音象徴語の印象評価値を、プリミティブワードの印象評価値として設定する。次に単語評価値生成装置 6 は、プリミティブワード以外の未知語について、プリミティブワードの印象評価値と、プリミティブワードと未知語との類似度の乗算から、未知語の評価値を算出する。単語評価値生成装置 6 は、プリミティブワードとプリミティブワードの印象評価値を対応づけるとともに、未知語と未知語の印象評価値を対応づけて、単語評価値データ E を生成する。

40

【0029】

音象徴語評価値データ N は、図 5 に示すように、音象徴語と、音象徴語の印象評価値を対応づけたデータである。音象徴語の印象評価値は、図 4 に示すように形容詞対毎の評価値を含む。音象徴語評価値データ N の音象徴語と単語評価値データ E の単語の各印象評価値は、互いに対応する形式で表現される。

【0030】

処理装置 20 は、対象テキスト取得部 21 および出力部 22 を備える。

50

【 0 0 3 1 】

対象テキスト取得部 2 1 は、パーソナリティ出力装置 1 がパーソナリティを出力する元となるテキストデータを取得し、対象テキストデータ 1 1 として記憶装置 1 0 に記憶する。

【 0 0 3 2 】

出力部 2 2 は、対象テキストデータ 1 1 のテキストデータに含まれる各単語から特定されるユーザのパーソナリティの印象評価値と音象徴語評価値データ N の印象評価値との類似度に基づいて、音象徴語を出力する。出力部 2 2 は、対象テキストデータ 1 1 から、処理対象のテキストデータを抽出し、抽出したテキストデータに含まれる単語と、単語評価値データ E における単語の印象評価値とに基づいて、ユーザのパーソナリティを示す印象評価値を算出する。さらに出力部 2 2 は、音象徴語評価値データ N を参照して、算出した印象評価値に近い印象評価値を有する音象徴語を出力する。

10

【 0 0 3 3 】

出力部 2 2 は、音象徴語を出力する際、その音象徴語を出力する際に参照したテキストデータの数（マイクロブログの投稿数）を、併せて出力しても良い。

【 0 0 3 4 】

出力部 2 2 は、一つの文またはマイクロブログへの一つの投稿など、データ量が比較的小さいテキストデータの単位で音象徴語を算出しても良いし、複数の文または複数の投稿など、データ量が比較的大きいテキストデータの単位で音象徴語を算出しても良い。例えば出力部 2 2 は、所定期間毎に、所定期間に入力されたテキストデータについて音象徴語を出力して、図 2 に示すように、音象徴語の遷移を出力しても良い。

20

【 0 0 3 5 】

図 6 を参照して、出力部 2 2 による出力処理を説明する。

【 0 0 3 6 】

まずステップ S 1 0 1 において出力部 2 2 は、処理対象のテキストデータに含まれる単語のうち、単語評価値データ E で定義される単語を抽出する。ステップ S 1 0 2 において出力部 2 2 は、単語評価値データ E から、ステップ S 1 0 1 で抽出した各単語の印象評価値を抽出し、各単語の印象評価値の平均値を算出する。ここで出力部 2 2 は、図 4 に示す形容詞対毎に各単語の評価値を特定し、各形容詞対の評価値の和を単語の数（のべ数）で割った値から、各単語の印象評価値の平均を算出する。ここで算出された各単語の印象評価値の平均は、ユーザが入力したテキストデータの各単語から特定される、ユーザのパーソナリティの印象評価値である。

30

【 0 0 3 7 】

ステップ S 1 0 3 において出力部 2 2 は、音象徴語評価値データ N を参照して、ステップ S 1 0 2 で算出した平均に近い音象徴語を出力する。出力部 2 2 は、音象徴語評価値データ N から、ステップ S 1 0 2 で算出された各単語の印象評価値の平均とコサイン類似度が高い印象評価値を持つ音象徴語を取得する。出力部 2 2 は、形容詞対毎に、音象徴語評価値データ N の評価値とステップ S 1 0 2 で算出した平均とを比較して、コサイン類似度が最も高い音象徴語を特定する。出力部 2 2 は、ステップ S 1 0 3 において、処理対象のテキストデータが表現するユーザのパーソナリティとして、特定された音象徴語を出力する。

40

【 0 0 3 8 】

ここで出力部 2 2 は、複数の投稿を統合した一つのテキストデータからユーザのパーソナリティの印象評価値を算出しても良いし、個々の投稿から印象評価値を算出して、その印象評価値からユーザの印象評価値を算出しても良い。また単語評価値データ E の単語が、プリミティブワードと未知語に区分される場合、出力部 2 2 は、単語評価値データ E のプリミティブワードと未知語をそれぞれ特定し、式 (1) によって、処理対象の印象評価値を算出しても良い。出力部 2 2 は、式 (1) において、プリミティブワードと未知語に異なる係数をつけて重み付けするなどにより、処理対象のテキストデータの印象評価値を算出しても良い。

50

【 0 0 3 9 】

【 数 1 】

$$v_t = \frac{\sum v_p \in P(t) v_p + \sum v_u \in U(t) v_u}{|P(t)| + |U(t)|} \quad \dots \text{式(1)}$$

$P(t)$: テキストデータ t に出現するプリミティブワードの集合
$U(t)$: テキストデータ t に出現する未知語の集合
v_p	: $P(t)$ の印象評価値
v_u	: 未知語の印象評価値
v_t	: テキストデータ t の印象評価値

【 0 0 4 0 】

各投稿についてユーザの印象評価値を算出したのち、複数の投稿についてのユーザの印象評価値を算出する場合、出力部 22 は、式 (2) によりユーザの印象評価値として算出する。

【 0 0 4 1 】

【 数 2 】

$$PV = \frac{\sum_{k=1}^n v_{t_k}}{n} \quad \dots \text{式(2)}$$

PV	: ユーザの印象評価値
t_k	: n 個のテキストデータのうちの k 番目のテキストデータ
v_{t_k}	: t_k のテキストデータに対する印象評価値

20

【 0 0 4 2 】

このように、パーソナリティ出力装置 1 は、ユーザが入力したテキストデータに含まれる単語に基づいて、そのユーザのパーソナリティを音象徴語で表すことができる。音象徴語は、1 単語の音象徴語でユーザの感情を表現することにより、簡潔かつ判りやすくユーザのパーソナリティを伝えることができる。

【 0 0 4 3 】

(単語評価値生成装置)

図 7 を参照して、本発明の実施の形態に係る単語評価値生成装置 6 を説明する。単語評価値生成装置 6 は、図 1 を参照して説明したパーソナリティ出力装置 1 で参照される単語評価値データ E を生成する。

【 0 0 4 4 】

単語評価値生成装置 6 は、記憶装置 60、処理装置 70 および入出力インタフェース 80 を備える一般的なコンピュータである。一般的なコンピュータが単語評価生成プログラムを実行することにより、図 7 に示す機能を実現する。記憶装置 60、処理装置 70 および入出力インタフェース 80 は、図 1 を参照して説明したパーソナリティ出力装置 1 の記憶装置 10、処理装置 20 および入出力インタフェース 30 と、それぞれ同様である。

【 0 0 4 5 】

記憶装置 60 は、単語評価値生成プログラムを記憶するとともに、プリミティブワードデータ 61、類似度データ 62、単語評価値データ E および音象徴語評価値データ N を記憶する。

【 0 0 4 6 】

30

40

50

プリミティブワードデータ61は、テキストデータを入力した際のユーザのパーソナリティを表す音象徴語と、テキストデータに含まれる単語のうち音象徴語に対応するプリミティブワードを対応づけたデータである。プリミティブワードデータ61は、図8に示すように、テキストデータと、そのテキストデータにおけるユーザのパーソナリティを示す音象徴語と、テキストデータに含まれる単語のうち、その音象徴語に関連する単語（プリミティブワード）とを対応づけたデータである。プリミティブワードデータ61について、一つのテキストデータについて、一つの音象徴語と、その音象徴語に関連する複数の単語が対応づけられる。

【0047】

本発明の実施の形態においてプリミティブワードデータ61は、テキストデータを入力したユーザによって、音象徴語とその音象徴語に関連する単語が入力されることにより生成される。この際、ユーザは、音象徴語評価値データNに登録される音象徴語から、テキストデータにおけるユーザのパーソナリティを示す音象徴語が選択されても良い。

【0048】

プリミティブワードデータ61に登録されるテキストデータは、ユーザのパーソナリティを表すものが好ましい。事実のみを記載しているなど、ユーザのパーソナリティが現れていないテキストデータは、プリミティブワードデータ61に登録されないのが好ましい。またプリミティブワードデータ61に設定される単語は、ユーザのパーソナリティに関連するネットスラング、顔文字等であっても良い。

【0049】

ユーザは、テキストデータに対する音象徴語を選択した後、その音象徴語に関連する単語を入力して、プリミティブワードデータを生成するのが好ましい。テキストデータから、単語を入力した後に音象徴語を入力する場合、単語の持つ意味から音象徴語を選択する可能性を排除できず、テキストデータから想起される音象徴語と単語の関係に影響が生じる場合があるからである。

【0050】

類似度データ62は、後述の未知語評価値算出部72が出力するデータであって、図9に示すように、プリミティブワードと未知語との類似度を含む。類似度データ62は、プリミティブワードの印象評価値から、未知語の印象評価値を算出するために用いられる。プリミティブワードは、テキストデータに対してユーザが選択した音象徴語を表す単語として、ユーザがテキストデータから選択した単語である。一方未知語は、マイクロブログで用いられるテキストデータ等に含まれる単語であって、プリミティブワード以外の単語である。未知語は、後述の未知語評価値算出部72によって抽出される。

【0051】

単語評価値データEは、図1等を参照して説明したように、単語と、単語の印象評価値を対応づけたデータである。

【0052】

音象徴語評価値データNは、図5を参照して説明したように、音象徴語と、音象徴語の印象評価値を対応づけたデータである。

【0053】

処理装置70は、プリミティブワード評価値設定部71、未知語評価値算出部72および単語評価値データ生成部73を備える。

【0054】

プリミティブワード評価値設定部71は、プリミティブワードに対応する音象徴語の印象評価値を、プリミティブワードの印象評価値として設定する。プリミティブワードとその印象評価値は、単語評価値データEに設定される。

【0055】

プリミティブワード評価値設定部71は、プリミティブワードデータ61に含まれる各プリミティブワードについて、このプリミティブワードに対応づけられた音象徴語の印象評価値を、このプリミティブワードの印象評価値として設定する。音象徴語の印象評価値

10

20

30

40

50

は、音象徴語評価値データNから取得される。図8に示す例において、プリミティブワード「花火」の印象評価値は、音象徴語「うきうき」の印象評価値となる。

【0056】

プリミティブワード評価値設定部71は、プリミティブワードデータ61におけるプリミティブワードと音象評価値の組み合わせのうち、単語評価値データEに設定する組み合わせを取捨選択しても良い。

【0057】

例えばプリミティブワードデータ61においてユーザが設定した単語のうち、マイクロブログにおける高頻出の単語については除外して、単語評価値データEに設定されても良い。マイクロブログにおける高頻出の単語は、例えば、「する」、「です」などの語尾に多用する単語、「だけど」、「だから」など接続詞として多用する単語などである。プリミティブワード評価値設定部71は、プリミティブワードデータ61において設定された各単語の頻出回数を算出して、所定の頻出回数以上出現する単語については、単語評価値データEに設定されないようにしても良い。また他の方法として、マイクロブログに出現する単語のうち、名詞のみを単語評価値データEに設定するなど、所定のルールで取捨選択しても良い。

10

【0058】

またプリミティブワードに複数の音象徴語が対応づけられる場合、プリミティブワード評価値設定部71は、複数の音象徴語のうち単一の音象徴語の印象評価値を、そのプリミティブワードの印象評価値として設定する。プリミティブワード評価値設定部71は、複数の音象徴語のうち、ランダムに単一の音象徴語を選択しても良いし、一つのプリミティブワードに対して最初に登場した音象徴語など所定のロジックで選択しても良い。

20

【0059】

また一つの音象徴語に複数のプリミティブワードが対応づけられる場合、プリミティブワード評価値設定部71は、一つの音象徴語に所定数のプリミティブワードが対応づけられるように、プリミティブワードを選定しても良い。

【0060】

未知語評価値算出部72は、プリミティブワード以外の未知語について、プリミティブワードの印象評価値と、プリミティブワードと未知語との類似度の乗算から、未知語の評価値を算出する。

30

【0061】

プリミティブワード評価値設定部71により、プリミティブワードについて印象評価値が設定されたが、多種多様な単語が用いられるマイクロブログから、プリミティブワードに基づいてユーザのパーソナリティを判断することが困難な場合がある。そこで、プリミティブワード以外の未知語についても、印象評価値を設定し、より多くの単語に基づいてユーザのパーソナリティを判断できることが好ましい。そこで未知語評価値算出部72は、テキストデータ等からプリミティブワード以外の未知語を抽出し、未知語とプリミティブワードとの類似度から、未知語の印象評価値を算出する。

【0062】

未知語評価値算出部72は、マイクロブログから抽出した多数のテキストデータに対して形態素解析を行い、所定条件によって、未知語を抽出する。ここで参照される多数のテキストデータは、プリミティブワードデータ61において設定されたテキストデータ以外のテキストデータであることが好ましい。未知語を抽出する所定条件は例えば、プリミティブワード以外の単語であって、複数のテキストデータにおいて出現する語である。

40

【0063】

未知語評価値算出部72は、プリミティブワードと未知語との類似度を算出し、式(3)によりプリミティブワードの印象評価値と類似度を乗算して、未知語の印象評価値を算出する。

【0064】

50

【数 3】

$$v_u = \frac{\sum_{w_i \in P(w_u, \theta)} s(w_u, w_i) v_i}{|P(w_u, \theta)|} \quad \dots \text{式(3)}$$

P	: プリミティブワード $w_1, w_2, \dots, w_i, \dots$ の集合
v_i	: $w_i \in P$ の印象評価値
w_u	: 未知語
v_u	: 未知語の印象評価値
$s(w_u, w_i)$: 未知語 w_u とプリミティブワード w_i とのコサイン類似度
$P(w_u, \theta)$: $s(w_u, w_i)$ が 0.4 以上のプリミティブワードの集合

【0065】

式(3)は、複数のプリミティブワードのそれぞれについて、プリミティブワードの印象評価値とそのプリミティブワードと未知語 w_u との類似度を乗算した値を算出し、その値を加算して正規化することにより、未知語 w_u の印象評価値を算出することを示す。また式(3)で用いるプリミティブワードを、未知語 w_u と所定値以上の類似度を有するプリミティブワードに限定することにより、印象評価値の処理負担を軽減する。

20

【0066】

本発明の実施の形態において未知語評価値算出部 72 は、プリミティブワードを含むテキストデータと未知語を含むテキストデータの比較に基づいて、プリミティブワードと未知語との類似度を算出する。未知語評価値算出部 72 は、プリミティブワードが含まれるテキストデータと、未知語を含むテキストデータが類似する場合、プリミティブワードと未知語との類似度が高くなり、類似しない場合、プリミティブワードと未知語との類似度が低くなるように、類似度を算出する。

【0067】

図 10 および図 11 を参照して、プリミティブワードと未知語の類似度を算出する処理を説明する。

30

【0068】

未知語評価値算出部 72 は、類似度を算出する際、ステップ S201 において、単語文章行列を生成する。単語文章行列は、図 11 に示すように、プリミティブワードと未知語を行に設定し、列に複数のテキストデータを設定する。単語文章行列の列に設定されるテキストデータは、図 1 で説明した対象テキストデータ 11 と同様に、特定の事実に対する解説など投稿するユーザによって差異が出にくい内容よりも、単一のユーザによる日々の所感など、ユーザの個性が出やすく、パーソナリティが現れやすい内容であることが好ましい。単語文章行列の列に設定されるテキストデータは、プリミティブワードデータ 61 に含まれるテキストデータが含まれても良いし、含まれなくても良い。

【0069】

40

行列の各値には、テキストデータにおいて含まれる未知語またはプリミティブワードの数が設定される。図 11 に示す例において、テキストデータ 1 において、未知語 u_2 が 1 回、プリミティブワード w_1 が 1 回等の各単語が、それぞれ出現することを示す。テキストデータ 2 において、未知語 u_1 が 2 回、プリミティブワード w_2 が 1 回等の各単語が、それぞれ出現することを示す。テキストデータ 3 において、未知語 u_2 が 3 回、プリミティブワード w_1 が 2 回、プリミティブワード w_2 が 2 回等の各単語が、それぞれ出現することを示す。

【0070】

次に未知語評価値算出部 72 は、未知語とプリミティブワードの各組み合わせについて、ステップ S202 の処理を行う。未知語評価値算出部 72 は、処理対象の未知語の行ベ

50

クトルと処理対象のプリミティブワードの行ベクトルとの、コサイン類似度を算出する。例えば未知語 u_1 とプリミティブワード w_1 のコサイン類似度を算出する際、図 11 に示す行列の未知語 u_1 の行ベクトル $(0, 2, 0, \dots)$ と、プリミティブワード w_1 の行ベクトル $(1, 0, 2, \dots)$ のコサイン類似度を算出する。ここで未知語評価値算出部 72 は、図 11 に示した単語文章行列を、 $tf-idf$ 法を用いて重み付けしたり、特異値分解により次元を圧縮した近似行列に変換したりして、コサイン類似度を算出しても良い。

【0071】

未知語とプリミティブワードの各組み合わせについて、ステップ S202 の処理を行うと、未知語評価値算出部 72 は、ステップ S203 においてステップ S202 で算出した類似度を含む類似度データ 62 を出力する。

10

【0072】

未知語評価値算出部 72 は、図 10 および図 11 で示す処理により算出された類似度を用いて、式 (3) により、未知語の印象評価値を算出することができる。

【0073】

単語評価値データ生成部 73 は、単語と、その単語の印象評価値を対応づけて単語評価値データ E を生成する。単語評価値データ E は、プリミティブワードとプリミティブワードの印象評価値を対応づけるとともに、未知語と未知語の印象評価値を対応づける。単語評価値データ生成部 73 は、プリミティブワード評価値設定部 71 が設定したプリミティブワードの印象評価値と、未知語評価値算出部 72 が算出した未知語の印象評価値をマージして、単語評価値データ E を生成する。

20

【0074】

このように本発明の実施の形態に係る単語評価値生成装置 6 は、印象評価値が付与されていない未知語についても印象評価値を付与し、多くの単語を単語評価値データ E に設定することができる。これによりパーソナリティ出力装置 1 は、ユーザが入力したテキストデータから、ユーザのパーソナリティを示す音象徴語を適切に出力することができる。

【0075】

また本発明の実施の形態に係る単語評価値生成装置 6 は、ネットスラング、顔文字等の未知語についても、印象評価値を算出することができる。マイクロブログは、ネットスラング等の通信ネットワーク固有の用語で表現されたり、砕けた表現が多用されたりすると

30

【0076】

(その他の実施の形態)

上記のように、本発明の実施の形態によって記載したが、この開示の一部をなす論述および図面はこの発明を限定するものであると理解すべきではない。この開示から当業者には様々な代替実施の形態、実施例および運用技術が明らかとなる。

【0077】

例えば、本発明の実施の形態に記載したパーソナリティ出力装置および単語評価値生成装置は、図 1 および図 7 に示すように、それぞれ一つのハードウェア上に構成されても良いし、その機能や処理数に応じて複数のハードウェア上に構成されても良い。また、パーソナリティ出力装置および単語評価値生成装置の各機能は、一つのハードウェア上に構成されても良い。

40

【0078】

また本発明の実施の形態に示した処理順序は一例であって、この順序に限定されるものではない。

【0079】

本発明はここでは記載していない様々な実施の形態等を含むことは勿論である。従って、本発明の技術的範囲は上記の説明から妥当な特許請求の範囲に係る発明特定事項によってのみ定められるものである。

50

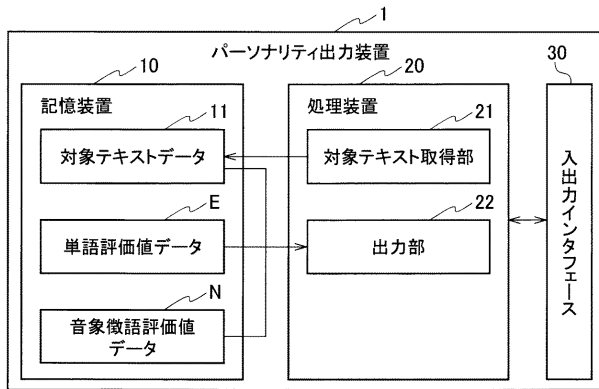
【符号の説明】

【0080】

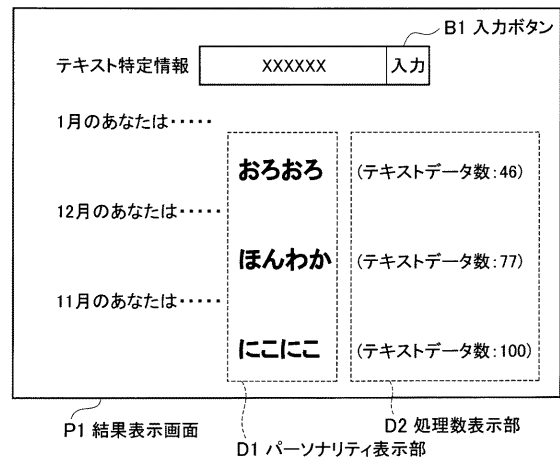
- 1 パーソナリティ出力装置
- 6 単語評価値生成装置
- 10、60 記憶装置
- 11 対象テキストデータ
- 20、70 処理装置
- 21 対象テキスト取得部
- 22 出力部
- 61 プリミティブワードデータ
- 62 類似度データ
- 71 プリミティブワード評価値設定部
- 72 未知語評価値算出部
- 73 単語評価値データ生成部
- E 単語評価値データ
- N 音象徴語評価値データ

10

【図1】



【図2】



【図3】

E 単語評価値データ

単語	印象評価値
...	...
...	...

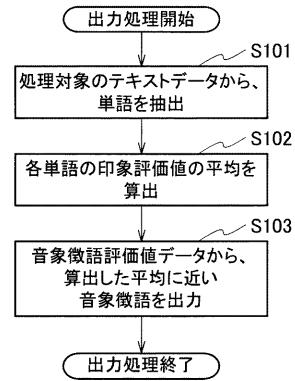
40

【 図 4 】

形容詞対	評価値
無口な-話好きな	...
臆病な-果敢な	...
迎合的な-断言的な	...
消極的な-意欲的な	...
受動的な-能動的な	...
...	...

印象評価値

【 図 6 】

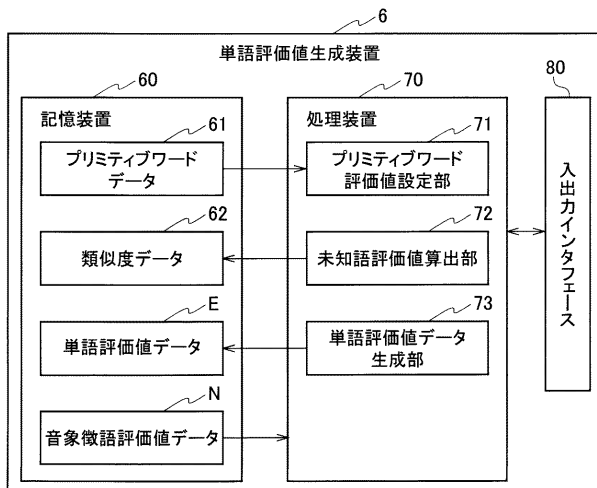


【 図 5 】

N 音象徴語評価値データ

音象徴語	印象評価値
...	...
...	...

【 図 7 】



【 図 8 】

61 プリミティブワードデータ

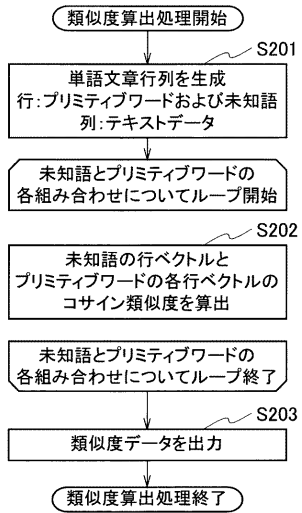
テキストデータ	音象徴語	音象徴語に関連する単語
今日は雨が降らなそうだから、 花火を友達と見に行くぞい	うきうき	花火
明日からまた学校~~ ずっと家に引きこもりたいよ	めそめそ	引きこもって
窓の外を眺めてたら 小学生が遊んでていいな~と思った	ぼんやり	いいな、眺めて
...

【 図 9 】

62 類似度データ

	プリミティブ ワードw1	プリミティブ ワードw2	プリミティブ ワードw3	...
未知語u1
未知語u2
未知語u3
...

【図 10】



【図 11】

	テキストデータ1	テキストデータ2	テキストデータ3	...
未知語u1	0	2	0	...
未知語u2	1	0	3	...
...
プリミティブワードw1	1	0	2	...
プリミティブワードw2	0	1	2	...
...