

(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2017-167378

(P2017-167378A)

(43) 公開日 平成29年9月21日 (2017.9.21)

(51) Int.Cl.	F I	テーマコード (参考)
G 1 0 L 15/19 (2013.01)	G 1 0 L 15/19	5 B 0 9 1
G 1 0 L 15/16 (2006.01)	G 1 0 L 15/16	
G 0 6 F 17/27 (2006.01)	G 0 6 F 17/27 6 1 5	

審査請求 未請求 請求項の数 14 O L (全 18 頁)

(21) 出願番号	特願2016-53438 (P2016-53438)	(71) 出願人	000003078
(22) 出願日	平成28年3月17日 (2016.3.17)		株式会社東芝
			東京都港区芝浦一丁目1番1号
		(74) 代理人	110002147
			特許業務法人酒井国際特許事務所
		(72) 発明者	森下 皓文
			東京都港区芝浦一丁目1番1号 株式会社東芝内
		(72) 発明者	益子 貴史
			東京都港区芝浦一丁目1番1号 株式会社東芝内
		Fターム(参考)	5B091 CA02 CC02 EA01

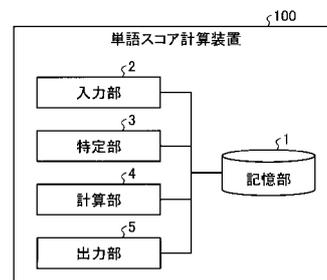
(54) 【発明の名称】 単語スコア計算装置、単語スコア計算方法及びプログラム

(57) 【要約】

【課題】 新語を含む単語列の単語スコアを計算する。

【解決手段】 実施形態の単語スコア計算装置は、入力部と計算部とを備える。入力部は、既に知られている既知語とまだ知られていない新語とを含む単語列の入力を受け付ける。計算部は、入力層に単語の語彙素性を入力する1以上の第1のユニット群を有するニューラルネットワークに対して、単語列に含まれる単語の語彙素性を入力することにより、単語列に含まれるターゲット単語の単語スコアを計算する。

【選択図】 図1



【特許請求の範囲】

【請求項 1】

既に知られている既知語とまだ知られていない新語とを含む単語列の入力を受け付ける入力部と、

入力層に単語の語彙素性を入力する 1 以上の第 1 のユニット群を有するニューラルネットワークに対して、前記単語列に含まれる単語の語彙素性を入力することにより、前記単語列に含まれるターゲット単語の単語スコアを計算する計算部と、

を備える単語スコア計算装置。

【請求項 2】

前記既知語と前記既知語の語彙素性とを関連付けて記憶する既知語辞書と、前記新語と前記新語の語彙素性とを関連付けて記憶する新語辞書と、を記憶する記憶部と、

前記単語が既知語である場合、前記既知語辞書により前記語彙素性を特定し、前記単語が新語である場合、前記新語辞書により前記語彙素性を特定する特定部と、

を更に備える請求項 1 に記載の単語スコア計算装置。

10

【請求項 3】

前記記憶部は、複数の単語を語彙として記憶し、

前記既知語は、前記語彙に含まれる単語であり、前記新語は前記語彙に含まれない単語である、

請求項 2 に記載の単語スコア計算装置。

【請求項 4】

前記単語の語彙素性は、前記単語の品詞である、

請求項 1 に記載の単語スコア計算装置。

20

【請求項 5】

前記ニューラルネットワークの入力層は、前記既知語を入力する 1 以上の第 2 のユニット群を更に有し、

前記計算部は、前記単語列に含まれる単語が前記新語である場合、前記新語の語彙素性を前記第 1 のユニット群に入力し、前記単語列に含まれる単語が前記既知語である場合、前記既知語を前記第 2 のユニット群に入力する、

請求項 1 に記載の単語スコア計算装置。

【請求項 6】

前記ニューラルネットワークの出力層は、前記既知語のスコアを示す 1 以上の第 3 のユニット群と、前記新語の語彙素性のスコアを示す 1 以上の第 4 のユニット群と、を有し、

前記ターゲット単語が前記既知語である場合、前記第 3 のユニット群のスコアを前記単語スコアとして計算し、前記ターゲット単語が前記新語である場合、前記第 4 のユニット群のスコアを前記単語スコアとして計算する、

請求項 5 に記載の単語スコア計算装置。

30

【請求項 7】

前記計算部は、前記ターゲット単語が前記新語である場合、前記新語と前記語彙素性との組み合わせ毎に定義された新語出現スコア、及び、前記第 4 のユニット群のスコアに基づいて前記単語スコアを計算する、

請求項 6 に記載の単語スコア計算装置。

40

【請求項 8】

前記計算部は、前記ターゲット単語が前記新語である場合、前記新語と前記語彙素性との組み合わせ毎に定義された新語出現スコア、及び、前記第 4 のユニット群のスコアの積により前記単語スコアを計算する、

請求項 7 に記載の単語スコア計算装置。

【請求項 9】

前記ニューラルネットワークの出力層は、前記既知語のスコアを示す 1 以上の第 3 のユニット群を有し、

前記計算部は、前記ターゲット単語が前記既知語である場合、前記第 3 のユニット群の

50

スコアを前記単語スコアとして計算し、前記ターゲット単語が前記新語である場合、前記第3のユニット群に含まれる前記既知語のスコアの全て又は一部のうち、前記既知語の語彙素性が、前記ターゲット単語の新語の語彙素性と一致する前記既知語のスコアに基づいて、前記新語の単語スコアを計算する、

請求項5に記載の単語スコア計算装置。

【請求項10】

前記計算部は、前記ターゲット単語が前記新語である場合、前記第3のユニット群に含まれる前記既知語のスコアの全て又は一部のうち、前記既知語の語彙素性が、前記ターゲット単語の新語の語彙素性と一致する前記既知語のスコアの和を、前記新語の単語スコアとして計算する、

10

請求項9に記載の単語スコア計算装置。

【請求項11】

前記ニューラルネットワークは、第1のニューラルネットワークと、第2のニューラルネットワークと、を含み、

前記第1のニューラルネットワークの入力層は、前記既知語を入力する1以上の第2のユニット群を更に有し、

前記第1のニューラルネットワークの出力層は、前記既知語のスコアを示す1以上の第3のユニット群を有し、

前記第2のニューラルネットワークの入力層は、前記第2のユニット群を更に有し、

前記第2のニューラルネットワークの出力層は、前記新語の語彙素性のスコアを示す第4のユニット群を有し、

20

前記計算部は、前記単語列に含まれる単語が前記新語である場合、前記新語の語彙素性を、前記第1のニューラルネットワークの前記第1のユニット群と、前記第2のニューラルネットワークの前記第1のユニット群と、に入力し、前記単語列に含まれる単語が前記既知語である場合、前記既知語を、前記第1のニューラルネットワークの前記第2のユニット群と、前記第2のニューラルネットワークの前記第2のユニット群に入力し、前記ターゲット単語が前記既知語である場合、前記第3のユニット群のスコアを前記単語スコアとして計算し、前記ターゲット単語が前記新語である場合、前記第4のユニット群のスコアを前記単語スコアとして計算する、

請求項1に記載の単語スコア計算装置。

30

【請求項12】

前記ニューラルネットワークの出力層は、前記単語の語彙素性のスコアを示す1以上の第3のユニット群を有し、

前記計算部は、前記単語列に含まれる単語の語彙素性を前記第1のユニット群に入力し、前記ターゲット単語の単語スコアを、前記第3のユニット群のスコアにより計算する、

請求項1に記載の単語スコア計算装置。

【請求項13】

既に知られている既知語とまだ知られていない新語とを含む単語列の入力を受け付けるステップと、

入力層に単語の語彙素性を入力する1以上の第1のユニット群を有するニューラルネットワークに対して、前記単語列に含まれる単語の語彙素性を入力することにより、前記単語列に含まれるターゲット単語の単語スコアを計算するステップと、

40

を含む単語スコア計算方法。

【請求項14】

コンピュータを、

既に知られている既知語とまだ知られていない新語とを含む単語列の入力を受け付ける入力部と、

入力層に単語の語彙素性を入力する1以上の第1のユニット群を有するニューラルネットワークに対して、前記単語列に含まれる単語の語彙素性を入力することにより、前記単語列に含まれるターゲット単語の単語スコアを計算する計算部、

50

として動作させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明の実施形態は単語スコア計算装置、単語スコア計算方法及びプログラムに関する。

【背景技術】

【0002】

単語列が与えられた際に、その単語列中の特定の位置の単語の単語スコアを計算する技術が従来から知られている。近年では、単語スコアの計算をニューラルネットワークによって行う技術が開発されている。

10

【0003】

単語スコア計算技術は音声認識等の認識システムに 응용されている。例えば音声認識システムでは、ユーザが発話した際に、音声認識結果の単語列として複数の候補が生じうる。音声認識システムは、例えばこの候補単語列それぞれに対して、単語列中の各単語の単語スコアを単語スコア計算技術によって計算し足し合わせることで、その候補単語列の言語スコアを計算することができる。音声認識システムは、例えば候補単語列の言語スコアと、その候補単語列の音響スコアと、から候補単語列間の優劣をつけることにより、音声認識結果を決定することができる。

20

【先行技術文献】

【特許文献】

【0004】

【特許文献1】特許第4245530号公報

【特許文献2】特許第4543294号公報

【非特許文献】

【0005】

【非特許文献1】“Neural probabilistic language models” Yoshua Ben-gio, Holger Schwenk, Jean-Sebastien Senecal, Frederic Morin, Jean-Luc Gauvain

30

【非特許文献2】“Recurrent neural network based language model” Tomas Mikolov, Martin Karafiat, Lukas Burget, Jan Cer-nocky, Sanjeev Khudanpur

【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、従来の単語スコア計算装置は語彙と呼ばれる単語のリストを持っており、語彙に含まれない新語を含む単語列が与えられた場合には単語スコアの計算ができなかった。

40

【課題を解決するための手段】

【0007】

実施形態の単語スコア計算装置は、入力部と計算部とを備える。入力部は、既に知られている既知語とまだ知られていない新語とを含む単語列の入力を受け付ける。計算部は、入力層に単語の語彙素性を入力する1以上の第1のユニット群を有するニューラルネットワークに対して、前記単語列に含まれる単語の語彙素性を入力することにより、前記単語列に含まれるターゲット単語の単語スコアを計算する。

【図面の簡単な説明】

【0008】

【図1】第1実施形態の単語スコア計算装置の機能構成の例を示す図。

50

- 【図 2】単語スコアの計算例について説明するための図。
 【図 3】第 1 実施形態のニューラルネットワークの例を示す図。
 【図 4】第 1 実施形態の単語スコア計算方法の例を示すフローチャート。
 【図 5】第 1 実施形態の変形例の新語出現スコアの例を示す図。
 【図 6】第 1 実施形態の変形例の単語スコア計算方法の例を示すフローチャート。
 【図 7】第 2 実施形態のニューラルネットワークの例を示す図。
 【図 8】第 2 実施形態の単語スコア計算方法の例を示すフローチャート。
 【図 9】第 3 実施形態の第 1 のニューラルネットワークの例を示す図。
 【図 10】第 3 実施形態の第 2 のニューラルネットワークの例を示す図。
 【図 11】第 3 実施形態の単語スコア計算方法の例を示すフローチャート。
 【図 12】第 4 実施形態のニューラルネットワークの例を示す図。
 【図 13】第 4 実施形態の単語スコア計算方法の例を示すフローチャート。
 【図 14】第 1 ~ 第 4 実施形態の単語スコア計算装置のハードウェア構成の例を示す図。
 【発明を実施するための形態】

10

【0009】

以下に添付図面を参照して、単語スコア計算装置、単語スコア計算方法及びプログラムの実施形態を詳細に説明する。

【0010】

(第 1 実施形態)

はじめに、第 1 実施形態の単語スコア計算装置の機能構成の例について説明する。

20

【0011】

[単語スコア計算装置の機能構成]

【0012】

図 1 は第 1 実施形態の単語スコア計算装置 100 の機能構成の例を示す図である。第 1 実施形態の単語スコア計算装置 100 は、記憶部 1、入力部 2、特定部 3、計算部 4 及び出力部 5 を備える。

【0013】

記憶部 1 は情報を記憶する。記憶部 1 は、例えば単語リスト、既知語辞書、新語辞書、及び、後述のニューラルネットワーク(図 3 参照)等を記憶する。

【0014】

単語リストは、単語のリストである。以下、単語リストを語彙という。また語彙に含まれる単語を既知語という。また語彙に含まれない単語を新語という。

30

【0015】

既知語辞書は、既に知られている既知語と、当該既知語の語彙素性と、を関連付けて記憶する。既知語は、例えば国語辞書等に既に登録されている一般的によく知られている単語である。語彙素性は、単語をある基準で幾つかの集合に分類した際の集合の種類である。語彙素性は、例えば単語の品詞である。また例えば語彙素性は、単語の類似度により分類された集合の名称等である。なお単語は、自然言語に限らず、記号等を含む。

【0016】

新語辞書は、新語と、当該新語の語彙素性とを関連付けて記憶する。単語スコア計算装置 100 は、新語辞書により、未知語を新語として扱うことができる。新語辞書には、例えばユーザにより随時、新しい未知語が新語として追加される。

40

【0017】

以下、第 1 実施形態の説明では、単語の語彙素性が品詞である場合を例にして説明する。

【0018】

入力部 2 は、既知語と新語とを含む単語列の入力を受け付ける。

【0019】

特定部 3 は、単語列に含まれる新語の語彙素性(第 1 実施形態の説明では品詞)を特定する。具体的には、はじめに特定部 3 は、単語列を形態素解析することにより、当該単語

50

列を複数の単語に分割する。次に、特定部 3 は、単語が新語辞書に登録されている場合、当該新語辞書により、当該単語の品詞を特定する。

【 0 0 2 0 】

< 単語スコアの計算処理 >

計算部 4 は、単語列に含まれるターゲット単語の単語スコアを計算する。ターゲット単語は、単語スコアの計算の対象となる単語である。はじめに、簡単な例により単語スコアの例について説明する。ターゲット単語の単語スコアは、ターゲット単語以外の単語が、単語列中に与えられた条件での条件付き確率により計算される。

【 0 0 2 1 】

図 2 は単語スコアの計算例について説明するための図である。図 2 の例では、単語列 1 0 1 ~ 1 0 3 に含まれるターゲット単語 w 3 を計算する場合を示す。単語列 1 0 1 のターゲット単語 w 3 は「雨」である。単語列 1 0 2 のターゲット単語 w 3 は「カブトムシ」である。単語列 1 0 3 のターゲット単語 w 3 は「食べる」である。

10

【 0 0 2 2 】

ターゲット単語以外の単語は、単語 w 1 「明日」、及び、単語 w 2 「は」である。計算部 4 は、単語列 1 0 1 の単語スコア P 1、単語列 1 0 2 の単語スコア P 2、及び、単語列 1 0 3 の単語スコア P 3 を、下記式 (1) ~ (3) により計算する。

【 0 0 2 3 】

$$P 1 (w 3 = \text{雨} \mid w 1 = \text{明日}, w 2 = \text{は}) = 0 . 3 \quad \dots (1)$$

$$P 2 (w 3 = \text{カブトムシ} \mid w 1 = \text{明日}, w 2 = \text{は}) = 0 . 1 \quad \dots (2)$$

$$P 3 (w 3 = \text{食べる} \mid w 1 = \text{明日}, w 2 = \text{は}) = 0 . 2 \quad \dots (3)$$

20

【 0 0 2 4 】

式 (1) ~ (3) の例では、単語列 1 0 1 の単語スコアが最も高い。そのため、例えば音声認識システムの認識結果として、単語列 1 0 1 ~ 1 0 3 が得られた場合には、言語的な観点からは、認識結果は単語列 1 0 1 である可能性が最も高いことがわかる。

【 0 0 2 5 】

具体的には、計算部 4 は、ニューラルネットワーク 2 0 0 を用いて上述の単語スコアを計算する。そして出力部 5 は、計算部 4 により計算された単語スコアを出力する。

【 0 0 2 6 】

図 3 は第 1 実施形態のニューラルネットワーク 2 0 0 の例を示す図である。図 3 の例は、ターゲット単語の 3 つ前までの単語に基づいて、当該ターゲット単語の単語スコアを計算するフィードフォワード型ニューラルネットワークの場合を示す。なおターゲット単語の単語スコアの計算に使用される単語の数は任意でよい。例えばターゲット単語の単語スコアは、ターゲット単語の 5 つ前までの単語に基づいて計算されてもよい。

30

【 0 0 2 7 】

第 1 実施形態のニューラルネットワーク 2 0 0 は、入力層 1 0、中間層 2 0 a ~ 2 0 c、中間層 3 0 及び出力層 4 0 を有する。以下、中間層 2 0 a ~ 2 0 c を区別しない場合、単に中間層 2 0 という。なお図 3 の例では、ニューラルネットワーク 2 0 0 は、中間層 2 0 及び中間装置 3 0 の 2 層構成になっているが、中間層の層数は任意でよい。

【 0 0 2 8 】

入力層 1 0 は、既知語ユニット群 1 1 a ~ 1 1 c、及び、品詞ユニット群 1 2 a ~ 1 2 c を有する。以下、既知語ユニット群 1 1 a ~ 1 1 c を区別しない場合、単に既知語ユニット群 1 1 という。同様に、品詞ユニット群 1 2 a ~ 1 2 c を区別しない場合、単に品詞ユニット群 1 2 という。既知語ユニット群 1 1 は、1 以上の既知語ユニットを含む。同様に、品詞ユニット群 1 2 は、1 以上の品詞ユニットを含む。

40

【 0 0 2 9 】

はじめに、計算部 4 は、ターゲット単語の 3 つ前までの単語に基づいて、入力層 1 0 に情報の入力を行う。

【 0 0 3 0 】

具体的には、計算部 4 は、ターゲット単語の 1 つ前の単語が既知語である場合、当該既

50

知語を既知語ユニット群 1 1 a に入力する。一方、計算部 4 は、ターゲット単語の 1 つ前の単語が新語である場合、当該新語の品詞を品詞ユニット群 1 2 a に入力する。

【 0 0 3 1 】

また計算部 4 は、ターゲット単語の 2 つ前の単語が既知語である場合、当該既知語を既知語ユニット群 1 1 b に入力する。一方、計算部 4 は、ターゲット単語の 2 つ前の単語が新語である場合、当該新語の品詞を品詞ユニット群 1 2 b に入力する。

【 0 0 3 2 】

また計算部 4 は、ターゲット単語の 3 つ前の単語が既知語である場合、当該既知語を既知語ユニット群 1 1 c に入力する。一方、計算部 4 は、ターゲット単語の 3 つ前の単語が新語である場合、当該新語の品詞を品詞ユニット群 1 2 c に入力する。

10

【 0 0 3 3 】

なお計算部 4 が、既知語ユニット群 1 1 に既知語を入力する方法は任意でよい。計算部 4 は、例えば各々の既知語に対して、予め定められた当該既知語を識別するベクトルを既知語ユニット群 1 1 に入力する。既知語を識別するベクトルは、既知語辞書に登録されているそれぞれの既知語に対応する。なお既知語を識別するベクトルは任意に決めてよい。既知語を識別するベクトルは、例えば (0 , 0 , 0 , . . . , 0 , 1 , 0 , . . . , 0 , 0 , 0) 等の 1 つの該当成分が 1 であり、他の成分は全て 0 であるベクトルである。この場合、ベクトルの各成分が、既知語ユニット群 1 1 に含まれる各ユニットに対応する。1 が入力されたユニットは、活性化されたユニットと呼ばれる。

【 0 0 3 4 】

同様に、計算部 4 が、品詞ユニット群 1 2 に既知語の品詞を入力する方法は任意でよい。計算部 4 は、例えば各々の品詞に対して、予め定められた当該品詞を識別するベクトルを品詞ユニット群 1 2 に入力する。

20

【 0 0 3 5 】

以下、入力層 1 0 の既知語ユニット群 1 1 に入力されたベクトルと、品詞ユニット群 1 2 に入力されたベクトルと、を縦に連結することにより得られる列ベクトルを、単語ベクトル v^k ($k = 1, 2, 3$) という。

【 0 0 3 6 】

次に、計算部 4 は、入力層 1 0 の単語ベクトル v^k ($k = 1, 2, 3$) に、下記式 (4) による線形演算を行うことにより得られたベクトル w^k ($k = 1, 2, 3$) を、中間層 2 0 に入力する。

30

【 0 0 3 7 】

$$w^k = W v^k \quad \cdot \cdot \cdot (4)$$

【 0 0 3 8 】

ここで W は、単語ベクトル v^k の結合重みを示すパラメータを表す行列である。

【 0 0 3 9 】

次に、計算部 4 は、中間層 2 0 のベクトル w^k ($k = 1, 2, 3$) に、下記式 (5) による演算を行うことにより得られたベクトル h を、中間層 3 0 に入力する。

【 0 0 4 0 】

$$h_i = \tanh [(S w + u)_i] \quad \cdot \cdot \cdot (5)$$

40

【 0 0 4 1 】

ここで i は、ベクトル h の i 番目の成分を示す。また w は、中間層 2 0 のベクトル w^k ($k = 1, 2, 3$) を縦に連結することにより得られる列ベクトルを示す。また S はベクトル w の結合重みを示すパラメータを表す行列である。また u は中間層 3 0 の閾値パラメータを示す。また \tanh はハイパボリックタンジェント関数を示す。

【 0 0 4 2 】

すなわちベクトル h は、ベクトル w に線形演算を行った後に、ハイパボリックタンジェント関数を適用することにより得られる。

【 0 0 4 3 】

次に、計算部 4 は、中間層 3 0 のベクトル h に、下記式 (6) による線形演算を行うこ

50

とにより得られたベクトル a に、下記式 (7) によるソフトマックス演算を行う。そして計算部 4 は、ベクトル y を出力層 40 に入力する。

【0044】

$$a = Th + r \dots (6)$$

$$y_i = \exp(a_i) / \sum_j \exp(a_j) \dots (7)$$

【0045】

ここで T はベクトル h の結合重みを示すパラメータを表す行列である。また r は出力層 40 の閾値パラメータを示す。また i は、ベクトル y 及びベクトル a の i 番目の成分を示す。また j は、ベクトル a の j 番目の成分を示す。また \exp は指数関数を示す。

【0046】

出力層 40 は、既知語スコアユニット群 41 及び品詞スコアユニット群 42 を有する。既知語スコアユニット群 41 は、1 以上の既知語スコアユニットを含む。既知語スコアユニットの各々は、一の既知語のスコアを示す。品詞スコアユニット群 42 は、1 以上の品詞スコアユニットを含む。品詞スコアユニットの各々は、一の品詞のスコアを示す。品詞は、例えば動詞、名詞及び形容詞等である。

【0047】

計算部 4 は、上述の構造のニューラルネットワーク 200 を用いることにより、既知語と品詞とを含む単語列が入力層 10 に与えられた場合でも、ターゲット単語の単語スコアを計算することができる。

【0048】

具体的には、ターゲット単語 s_i に対応する出力層 40 の値 y_i が、ターゲット単語 s_i の単語スコア $P(s_i | v^1, v^2, v^3)$ を表す。計算部 4 は、ターゲット単語 s_i が新語の場合でも、出力層 40 に品詞スコアユニット群 42 が存在するので、品詞のスコアにより当該新語の単語スコアを計算することができる。

【0049】

なお上述の図 3 のニューラルネットワーク 200 は、フィードフォワード型ニューラルネットワークであるが、他の構造のニューラルネットワークでもよい。ニューラルネットワーク 200 は、例えばリカレント型ニューラルネットワークでもよい。

【0050】

[単語スコア計算方法]

次に第 1 実施形態の単語スコア計算方法の例について説明する。

【0051】

図 4 は第 1 実施形態の単語スコア計算方法の例を示すフローチャートである。はじめに、入力部 2 が、既知語と新語とを単語として含む単語列の入力を受け付ける (ステップ S1)。次に、特定部 3 が、単語列に含まれる新語の品詞を特定する (ステップ S2)。

【0052】

次に、計算部 4 が、入力層 10 の既知語ユニット群 11 に既知語を入力する (ステップ S3)。具体的には、計算部 4 は、ターゲット単語の 1 つ前の単語が既知語である場合、当該既知語を既知語ユニット群 11 a に入力する。また、計算部 4 は、ターゲット単語の 2 つ前の単語が既知語である場合、当該既知語を既知語ユニット群 11 b に入力する。また、計算部 4 は、ターゲット単語の 3 つ前の単語が既知語である場合、当該既知語を既知語ユニット群 11 c に入力する。

【0053】

次に、計算部 4 が、入力層 10 の品詞ユニット群 12 に新語の品詞を入力する (ステップ S4)。具体的には、計算部 4 は、ターゲット単語の 1 つ前の単語が新語である場合、当該新語の品詞を品詞ユニット群 12 a に入力する。また、計算部 4 は、ターゲット単語の 2 つ前の単語が新語である場合、当該新語の品詞を品詞ユニット群 12 b に入力する。また、計算部 4 は、ターゲット単語の 3 つ前の単語が新語である場合、当該新語の品詞を品詞ユニット群 12 c に入力する。

【0054】

10

20

30

40

50

次に、計算部 4 が、ターゲット単語が新語であるか否かを判定する（ステップ S 5）。ターゲット単語が新語である場合（ステップ S 5、Yes）、計算部 4 は、出力層 4 0 の品詞スコアユニット群 4 2 の品詞スコアを、ターゲット単語の単語スコアとして計算する（ステップ S 6）。

【0055】

ターゲット単語が新語でない場合（ステップ S 5、No）、計算部 4 は、出力層 4 0 の既知語スコアユニット群 4 1 の既知語スコアを、ターゲット単語の単語スコアとして計算する（ステップ S 7）。

【0056】

以上説明したように、第 1 実施形態の単語スコア計算装置 1 0 0 では、入力部 2 が、既に知られている既知語とまだ知られていない新語とを含む単語列の入力を受け付ける。そして計算部 4 が、ニューラルネットワーク 2 0 0（図 3 参照）を使用して、単語列に含まれる計算対象の単語であるターゲット単語の単語スコアを計算する。

【0057】

具体的には、計算部 4 は、単語列に含まれる単語が新語である場合、新語の品詞を品詞ユニット群 1 2 に入力し、単語列に含まれる単語が既知語である場合、既知語を既知語ユニット群 1 1 に入力する。そして計算部 4 は、ターゲット単語が既知語である場合、既知語スコアユニット群 4 1 のスコアを単語スコアとして計算し、ターゲット単語が新語である場合、品詞スコアユニット群 4 2 のスコアを単語スコアとして計算する。

【0058】

これにより第 1 実施形態の単語スコア計算装置 1 0 0 によれば、単語スコア計算装置 1 0 0 に語彙に含まれない新語を含む単語列が、単語スコア計算装置 1 0 0 に与えられた場合でも、当該単語列の単語スコアを計算することができる。

【0059】

なお特定部 3 は、上述の新語辞書を使用せずに、他の方法で新語の品詞を特定してもよい。特定部 3 は、例えば品詞クラス推定用文例辞書（特許文献 1 参照）を使用してもよい。具体的には、特定部 3 は、品詞クラス推定用文例辞書中に含まれる文例と、新語とのマッチングを取ることににより、当該新語の品詞を特定してもよい。

【0060】

また、特定部 3 は、例えば単語末尾の音韻に基づいて品詞を特定してもよい。

【0061】

（第 1 実施形態の変形例）

次に第 1 実施形態の変形例について説明する。第 1 実施形態の変形例の説明では、第 1 実施形態と同様の説明については省略し、第 1 実施形態と異なる箇所について説明する。

【0062】

第 1 実施形態の変形例では、記憶部 1 が新語出現スコアを更に記憶し、計算部 4 が、当該新語出現スコアを更に使用して単語スコアを計算する。第 1 実施形態の変形例の説明では、語彙素性が品詞である場合を例にして説明する。

【0063】

図 5 は第 1 実施形態の変形例の新語出現スコアの例を示す図である。新語出現スコアは、新語と品詞との組み合わせ毎に定義される。新語出現スコアは、新語の出現し易さを表す。新語 1 1 1 が動詞で出現した場合の新語出現スコアは、0.05 である。また、新語 1 1 2 が動詞で出現した場合の新語出現スコアは、0.01 である。

【0064】

第 1 実施形態の説明では、計算部 4 は、新語の単語スコアを、品詞スコアユニット群 4 2 のスコアにより計算した。しかしながら第 1 実施形態の単語スコア計算装置 1 0 0 では、例えば新語 1 1 1 及び新語 1 1 2 の単語スコアは、どちらも動詞なので同じになる。第 1 実施形態の変形例では、計算部 4 が、同じ品詞の新語についても、単語スコアに差がでるようにするために、新語出現スコアを更に使用する。

【0065】

10

20

30

40

50

図5の例では、新語111の新語出現スコアは、新語112の新語出現スコアよりも高い。計算部4は、ターゲット単語が新語である場合、新語出現スコア、及び、品詞スコアユニット群42のスコアに基づいて単語スコアを計算する。計算部4は、例えば新語出現スコア、及び、品詞スコアユニット群42の品詞スコアの積により単語スコアを計算する。

【0066】

[単語スコア計算方法]

次に第1実施形態の変形例の単語スコア計算方法について説明する。

【0067】

図6は第1実施形態の変形例の単語スコア計算方法の例を示すフローチャートである。ステップS21～ステップS25及びステップS27の説明は、第1実施形態のステップS1～ステップS5及びステップS7の説明(図4参照)と同じなので省略する。

10

【0068】

ターゲット単語が新語である場合(ステップS25、Yes)、計算部4は、出力層40の品詞スコアユニット群42の品詞スコアと、当該新語の新語出現スコアとの積により、ターゲット単語の単語スコアを計算する(ステップS26)。

【0069】

以上説明したように、第1実施形態の変形例の単語スコア計算装置100によれば、新語間の出現のし易さの差が更に考慮された、より高精度な単語スコアを計算することができる。

20

【0070】

(第2実施形態)

次に第2実施形態について説明する。第2実施形態の説明では、第1実施形態と同様の説明については省略し、第1実施形態と異なる箇所について説明する。

【0071】

第2実施形態の単語スコア計算装置100の機能構成の説明は、第1実施形態の単語スコア計算装置100の機能構成の例の説明(図1参照)と同じなので省略する。第2実施形態の説明では、語彙素性が品詞である場合を例にして説明する。

【0072】

第2実施形態では、ニューラルネットワーク200の出力層40が、品詞スコアユニット群42を有さない点が、第1実施形態の場合と異なる。

30

【0073】

図7は第2実施形態のニューラルネットワーク200の例を示す図である。第2実施形態のニューラルネットワーク200は、入力層10、中間層20a～20c、中間層30及び出力層40を有する。第2実施形態の入力層10、中間層20a～20c及び中間層30の説明は、第1実施形態の場合と同じなので省略する。

【0074】

出力層40は、既知語スコアユニット群41を有する。既知語スコアユニット群41は、既知語のスコアを示す。既知語スコアユニット群41は、1以上の既知語スコアユニットを含む。既知語スコアユニットの各々は、一の既知語のスコアを示す。

40

【0075】

計算部4は、ターゲット単語が新語である場合、既知語スコアユニット群41に含まれる既知語のスコアの全て又は一部のうち、既知語の品詞が、ターゲット単語の新語の品詞と一致する既知語のスコアに基づいて、新語の単語スコアを計算する。計算部4は、例えばターゲット単語が新語である場合、既知語スコアユニット群41に含まれる既知語のスコアの全て又は一部のうち、既知語の品詞が、ターゲット単語の新語の品詞と一致する既知語のスコアの和を、新語の単語スコアとして計算する。

【0076】

図7の例は、ベクトル y の成分 y_2 、 y_i 、 y_j 及び y_k が示す既知語のスコアに対応する既知語の品詞が、ターゲット単語である新語の品詞と一致する場合を示す。この場合

50

、計算部 4 は、ターゲット単語が新語である場合、 y_2 、 y_i 、 y_j 及び y_k が示す既知語のスコアの和を、新語の単語スコアとして計算する。

【0077】

[単語スコア計算方法]

次に第 2 実施形態の単語スコア計算方法について説明する。

【0078】

図 8 は第 2 実施形態の単語スコア計算方法の例を示すフローチャートである。はじめに、入力部 2 が、既知語と新語とを単語として含む単語列の入力を受け付ける（ステップ S 4 1）。次に、特定部 3 が、単語列に含まれる単語の品詞を特定する（ステップ S 4 2）。

10

【0079】

ステップ S 4 3 及びステップ S 4 4 の説明は、第 1 実施形態のステップ S 3 及びステップ S 4 の説明（図 4 参照）と同じなので省略する。

【0080】

次に、計算部 4 が、ターゲット単語が新語であるか否かを判定する（ステップ S 4 5）。ターゲット単語が新語である場合（ステップ S 4 5、Yes）、特定部 3 が、ターゲット単語の品詞を特定する（ステップ S 4 6）。次に、計算部 4 が、既知語スコアユニット群 4 1 に含まれる既知語のスコアの全て又は一部のうち、既知語の品詞が、ステップ S 4 6 の処理により特定された品詞と一致する既知語のスコアの和を、新語の品詞スコアとして計算する（ステップ S 4 7）。次に、計算部 4 が、ステップ S 4 7 の処理により計算された品詞スコアを、ターゲット単語の単語スコアとする（ステップ S 4 8）。

20

【0081】

ターゲット単語が新語でない場合（ステップ S 4 5、No）、計算部 4 は、出力層 4 0 の既知語スコアユニット群 4 1 の既知語スコアを、ターゲット単語の単語スコアとして計算する（ステップ S 4 9）。

【0082】

以上説明したように、第 2 実施形態の単語スコア計算装置 100 では、出力層 4 0 に品詞スコアユニット群 4 2 がないので、第 1 実施形態に比べて、ニューラルネットワーク 200 のパラメータを減らすことができる。すなわち第 2 実施形態の単語スコア計算装置 100 によれば、記憶部 1 に記憶されるニューラルネットワーク 200 の記憶容量を削減することができる。

30

【0083】

(第 3 実施形態)

次に第 3 実施形態について説明する。第 3 実施形態の説明では、第 1 実施形態と同様の説明については省略し、第 1 実施形態と異なる箇所について説明する。

【0084】

第 3 実施形態の単語スコア計算装置 100 では、計算部 4 が、構造的に 2 つに分離された第 1 のニューラルネットワーク及び第 2 のニューラルネットワークを含むニューラルネットワーク 200 を使用する点が、第 1 実施形態の場合と異なる。第 3 実施形態の説明では、語彙素性が品詞である場合を例にして説明する。

40

【0085】

図 9 は第 3 実施形態の第 1 のニューラルネットワーク 200 a の例を示す図である。第 3 実施形態の第 1 のニューラルネットワーク 200 a は、入力層 10、中間層 20 a ~ 20 c、中間層 30 及び出力層 40 を有する。第 3 実施形態の第 1 のニューラルネットワーク 200 a の入力層 10、中間層 20 a ~ 20 c 及び中間層 30 の説明は、第 1 実施形態のニューラルネットワーク 200 と同じなので省略する。

【0086】

出力層 40 は、既知語スコアユニット群 4 1 を有する。既知語スコアユニット群 4 1 は、1 以上の既知語スコアユニットを含む。既知語スコアユニットの各々は、一の既知語のスコアを示す。

50

【 0 0 8 7 】

図 1 0 は第 3 実施形態の第 2 のニューラルネットワーク 2 0 0 b の例を示す図である。第 3 実施形態の第 2 のニューラルネットワーク 2 0 0 b は、入力層 1 0、中間層 2 0 a ~ 2 0 c、中間層 3 0 及び出力層 4 0 を有する。第 3 実施形態の第 2 のニューラルネットワーク 2 0 0 b の入力層 1 0、中間層 2 0 a ~ 2 0 c 及び中間層 3 0 の説明は、第 1 実施形態のニューラルネットワーク 2 0 0 と同じなので省略する。

【 0 0 8 8 】

出力層 4 0 は、品詞スコアユニット群 4 2 を有する。品詞スコアユニット群 4 2 は、1 以上の品詞スコアユニットを含む。品詞スコアユニットの各々は、一の品詞のスコアを示す。品詞は、例えば動詞、名詞及び形容詞等である。

10

【 0 0 8 9 】

[単語スコア計算方法]

次に第 3 実施形態の単語スコア計算方法について説明する。

【 0 0 9 0 】

図 1 1 は第 3 実施形態の単語スコア計算方法の例を示すフローチャートである。ステップ S 6 1 ~ ステップ S 6 4 の説明は、第 1 実施形態のステップ S 1 ~ ステップ S 4 の説明 (図 4 参照) と同じなので省略する。

【 0 0 9 1 】

計算部 4 は、ターゲット単語が新語であるか否かを判定する (ステップ S 6 5)。ターゲット単語が新語である場合 (ステップ S 6 5、Yes)、計算部 4 は、第 2 のニューラルネットワーク 2 0 0 b の品詞スコアユニット群 4 2 の品詞スコアを、ターゲット単語の単語スコアとして計算する (ステップ S 6 6)。

20

【 0 0 9 2 】

ターゲット単語が新語でない場合 (ステップ S 6 5、No)、計算部 4 は、第 1 のニューラルネットワーク 2 0 0 a の既知語スコアユニット群 4 1 の既知語スコアを、ターゲット単語の単語スコアとして計算する (ステップ S 6 7)。

【 0 0 9 3 】

以上説明したように、第 3 実施形態の単語スコア計算装置 1 0 0 によれば、既知語スコアを計算する第 1 のニューラルネットワーク 2 0 0 a の構造、及び、品詞スコアを計算する第 2 のニューラルネットワーク 2 0 0 b の構造を、それぞれのスコア計算に適した構造にすることができる。

30

【 0 0 9 4 】

(第 4 実施形態)

次に第 4 実施形態について説明する。第 4 実施形態の説明では、第 1 実施形態と同様の説明については省略し、第 1 実施形態と異なる箇所について説明する。

【 0 0 9 5 】

第 4 実施形態の単語スコア計算装置 1 0 0 では、ニューラルネットワーク 2 0 0 の入力層 1 0 に既知語ユニット群 1 1 がなく、ニューラルネットワーク 2 0 0 の出力層 4 0 に既知語スコアユニット群 4 1 がない点が、第 1 実施形態の場合と異なる。第 4 実施形態の説明では、語彙素性が品詞である場合を例にして説明する。

40

【 0 0 9 6 】

図 1 2 は第 4 実施形態のニューラルネットワーク 2 0 0 の例を示す図である。第 4 実施形態のニューラルネットワーク 2 0 0 は、入力層 1 0、中間層 2 0 a ~ 2 0 c、中間層 3 0 及び出力層 4 0 を有する。

【 0 0 9 7 】

入力層 1 0 は、品詞ユニット群 1 2 a ~ 1 2 c を有する。品詞ユニット群 1 2 は、1 以上の品詞ユニットを含む。

【 0 0 9 8 】

はじめに、計算部 4 は、ターゲット単語の 3 つ前までの単語に基づいて、入力層 1 0 に情報の入力を行う。具体的には、計算部 4 は、ターゲット単語の 1 つ前の単語の品詞を品

50

詞ユニット群 1 2 a に入力する。計算部 4 は、ターゲット単語の 2 つ前の単語の品詞を品詞ユニット群 1 2 b に入力する。計算部 4 は、ターゲット単語の 3 つ前の単語の品詞を品詞ユニット群 1 2 c に入力する。

【 0 0 9 9 】

第 4 実施形態の中間層 2 0 a ~ 2 0 c 及び中間層 3 0 の説明は、第 1 実施形態の中間層 2 0 a ~ 2 0 c 及び中間層 3 0 の説明と同じなので省略する。

【 0 1 0 0 】

出力層 4 0 は、品詞スコアユニット群 4 2 を有する。品詞スコアユニット群 4 2 は、1 以上の品詞スコアユニットを含む。品詞スコアユニットの各々は、一の品詞のスコアを示す。品詞は、例えば動詞、名詞及び形容詞等である。

10

【 0 1 0 1 】

[単語スコア計算方法]

次に第 4 実施形態の単語スコア計算方法について説明する。

【 0 1 0 2 】

図 1 3 は第 4 実施形態の単語スコア計算方法の例を示すフローチャートである。はじめに、入力部 2 が、既知語と新語とを単語として含む単語列の入力を受け付ける（ステップ S 8 1）。次に、特定部 3 が、単語列に含まれる単語の品詞を特定する（ステップ S 8 2）。

【 0 1 0 3 】

次に、計算部 4 が、入力層 1 0 の品詞ユニット群 1 2 に単語の品詞を入力する（ステップ S 8 3）。具体的には、計算部 4 は、ターゲット単語の 1 つ前の単語の品詞を品詞ユニット群 1 2 a に入力する。計算部 4 は、ターゲット単語の 2 つ前の単語の品詞を品詞ユニット群 1 2 b に入力する。計算部 4 は、ターゲット単語の 3 つ前の単語の品詞を品詞ユニット群 1 2 c に入力する。

20

【 0 1 0 4 】

次に、計算部 4 は、出力層 4 0 の品詞スコアユニット群 4 2 の品詞スコアを、ターゲット単語の単語スコアとして計算する（ステップ S 8 4）。

【 0 1 0 5 】

以上説明したように、第 4 実施形態の単語スコア計算装置 1 0 0 では、ニューラルネットワーク 2 0 0 の入力層 1 0 に既知語ユニット群 1 1 がなく、ニューラルネットワーク 2 0 0 の出力層 4 0 に既知語スコアユニット群 4 1 がない。そのため第 4 実施形態の単語スコア計算装置 1 0 0 によれば、第 1 実施形態に比べて、ニューラルネットワーク 2 0 0 のパラメータを減らすことができる。すなわち第 4 実施形態の単語スコア計算装置 1 0 0 によれば、記憶部 1 に記憶されるニューラルネットワーク 2 0 0 の記憶容量を削減することができる。

30

【 0 1 0 6 】

また、第 4 実施形態の単語スコア計算装置 1 0 0 では、入力層 1 0 に入力される情報が品詞であり、出力層 4 0 から出力される情報が品詞スコアであるので、ニューラルネットワーク 2 0 0 で扱われる情報を品詞に特化させることができる。そのため第 4 実施形態の単語スコア計算装置 1 0 0 によれば、ニューラルネットワーク 2 0 0 の統計学習を有利に行うことができる。

40

【 0 1 0 7 】

[単語スコア計算装置のハードウェア構成]

最後に、第 1 ~ 第 4 実施形態の単語スコア計算装置 1 0 0 のハードウェア構成の例について説明する。

【 0 1 0 8 】

図 1 4 は第 1 ~ 第 4 実施形態の単語スコア計算装置 1 0 0 のハードウェア構成の例を示す図である。第 1 ~ 第 4 実施形態の単語スコア計算装置 1 0 0 は、制御装置 3 0 1、主記憶装置 3 0 2、補助記憶装置 3 0 3、表示装置 3 0 4、入力装置 3 0 5 及び通信装置 3 0 6 を備える。制御装置 3 0 1、主記憶装置 3 0 2、補助記憶装置 3 0 3、表示装置 3 0 4

50

、入力装置 305 及び通信装置 306 は、バス 310 を介して接続されている。

【0109】

制御装置 301 は補助記憶装置 303 から主記憶装置 302 に読み出されたプログラムを実行する。主記憶装置 302 は ROM 及び RAM 等のメモリである。補助記憶装置 303 はメモリカード及び SSD (Solid State Drive) 等である。

【0110】

表示装置 304 は情報を表示する。表示装置 304 は、例えば液晶ディスプレイである。入力装置 305 は、情報の入力を受け付ける。入力装置 305 は、例えばキーボード及びマウス等である。なお表示装置 304 及び入力装置 305 は、表示機能と入力機能とを兼ねる液晶タッチパネル等でもよい。通信装置 306 は他の装置と情報を送受信する。

10

【0111】

第 1 ~ 第 4 実施形態の単語スコア計算装置 100 で実行されるプログラムは、インストール可能な形式又は実行可能な形式のファイルで CD-ROM、メモリカード、CD-R 及び DVD (Digital Versatile Disk) 等のコンピュータで読み取り可能な記憶媒体に記憶されてコンピュータ・プログラム・プロダクトとして提供される。

【0112】

また第 1 ~ 第 4 実施形態の単語スコア計算装置 100 で実行されるプログラムを、インターネット等のネットワークに接続されたコンピュータ上に格納し、ネットワーク経由でダウンロードさせることにより提供するように構成してもよい。また第 1 ~ 第 4 実施形態の単語スコア計算装置 100 が実行するプログラムを、ダウンロードさせずにインターネット等のネットワーク経由で提供するように構成してもよい。

20

【0113】

また第 1 ~ 第 4 実施形態の単語スコア計算装置 100 で実行されるプログラムを、ROM 等に予め組み込んで提供するように構成してもよい。

【0114】

第 1 ~ 第 4 実施形態の単語スコア計算装置 100 で実行されるプログラムは、上述の第 1 ~ 第 4 実施形態の単語スコア計算装置 100 の機能構成のうち、プログラムにより実現可能な機能を含むモジュール構成となっている。

【0115】

プログラムにより実現される機能は、制御装置 301 が補助記憶装置 303 等の記憶媒体からプログラムを読み出して実行することにより、プログラムにより実現される機能が主記憶装置 302 にロードされる。すなわちプログラムにより実現される機能は、主記憶装置 302 上に生成される。

30

【0116】

なお第 1 ~ 第 4 実施形態の単語スコア計算装置 100 の機能の一部又は全部を、IC (Integrated Circuit) 等のハードウェアにより実現してもよい。

【0117】

本発明のいくつかの実施形態を説明したが、これらの実施形態は、例として提示したものであり、発明の範囲を限定することは意図していない。これら新規な実施形態は、その他の様々な形態で実施されることが可能であり、発明の要旨を逸脱しない範囲で、種々の省略、置き換え、変更を行うことができる。これら実施形態やその変形は、発明の範囲や要旨に含まれるとともに、特許請求の範囲に記載された発明とその均等の範囲に含まれる。

40

【符号の説明】

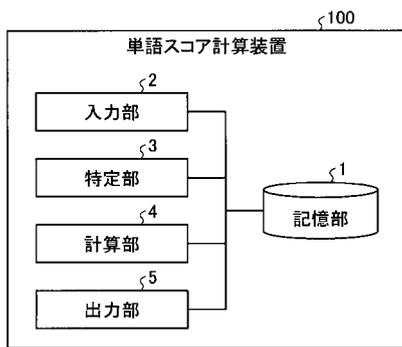
【0118】

- 1 記憶部
- 2 入力部
- 3 特定部
- 4 計算部

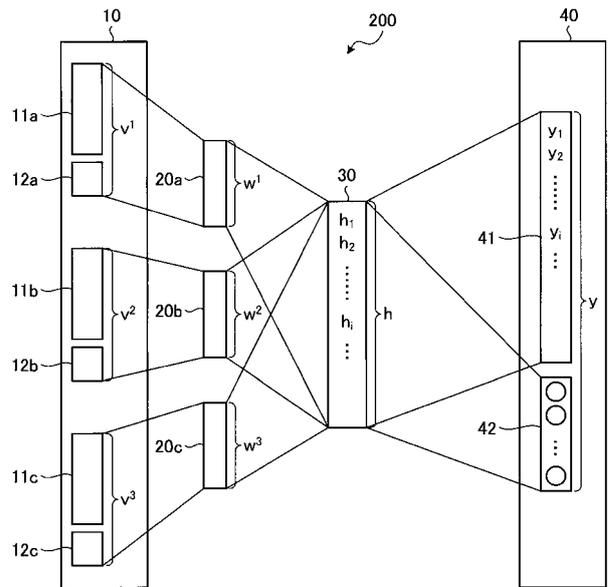
50

- 5 出力部
- 100 単語スコア計算装置
- 200 ニューラルネットワーク
- 301 制御装置
- 302 主記憶装置
- 303 補助記憶装置
- 304 表示装置
- 305 入力装置
- 306 通信装置

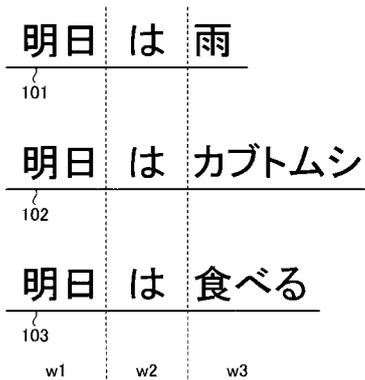
【図1】



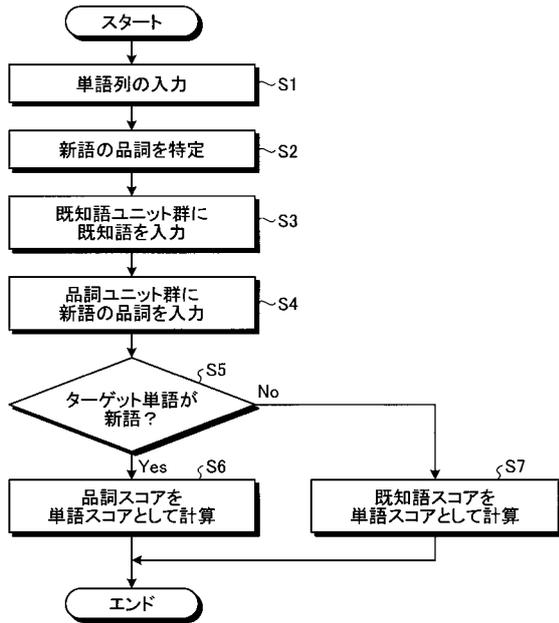
【図3】



【図2】



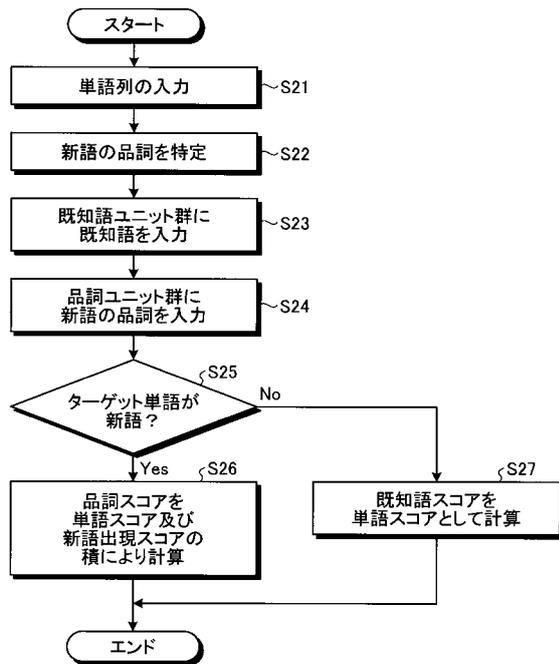
【図4】



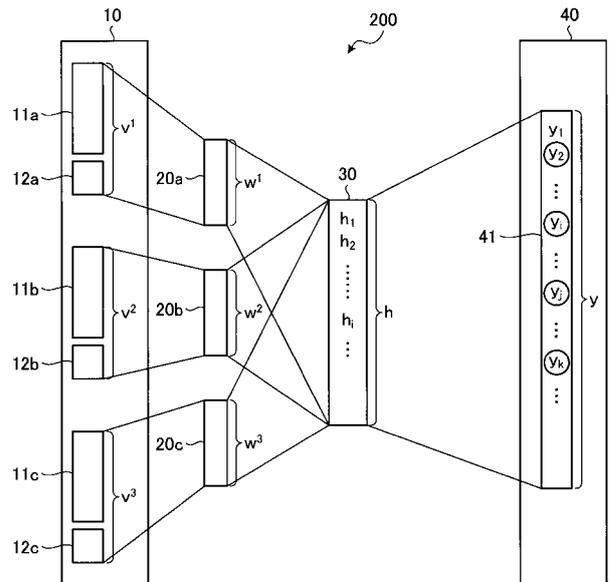
【図5】

新語	品詞	新語出現スコア
111 <u>クグ</u> る	動詞	0.05
112 <u>バミ</u> る	動詞	0.01
⋮	⋮	⋮

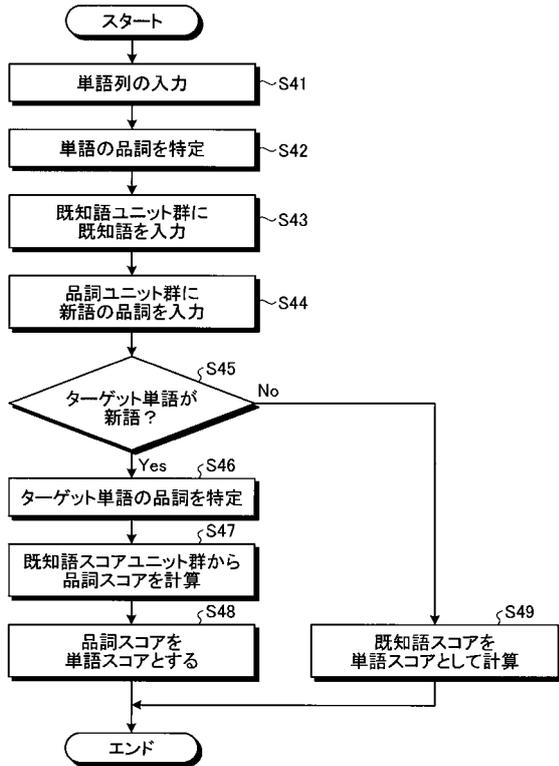
【図6】



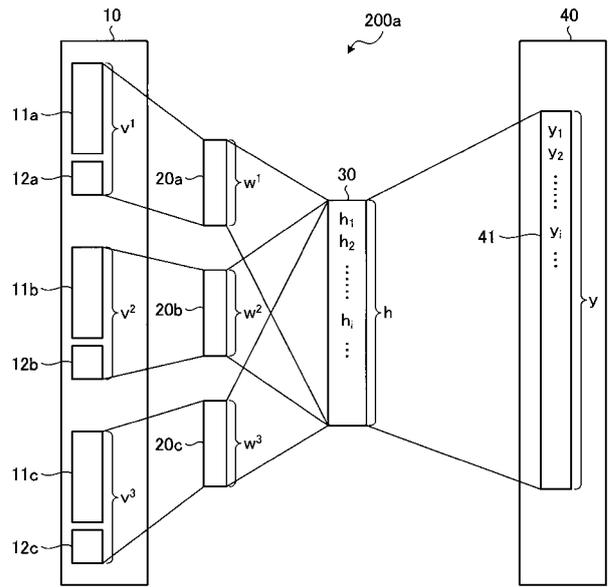
【図7】



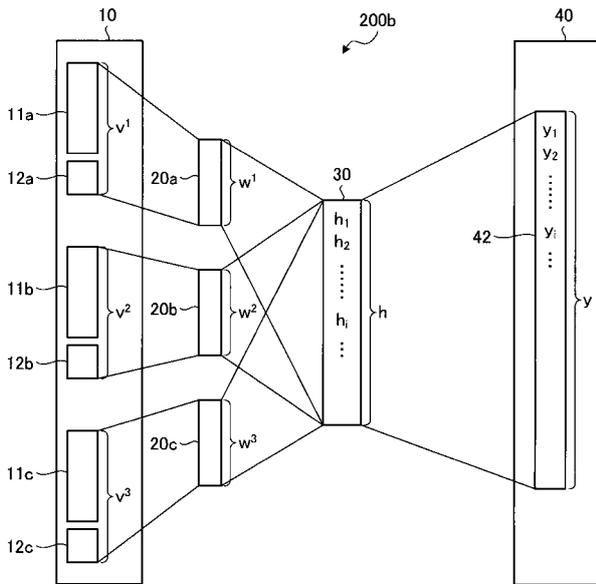
【図8】



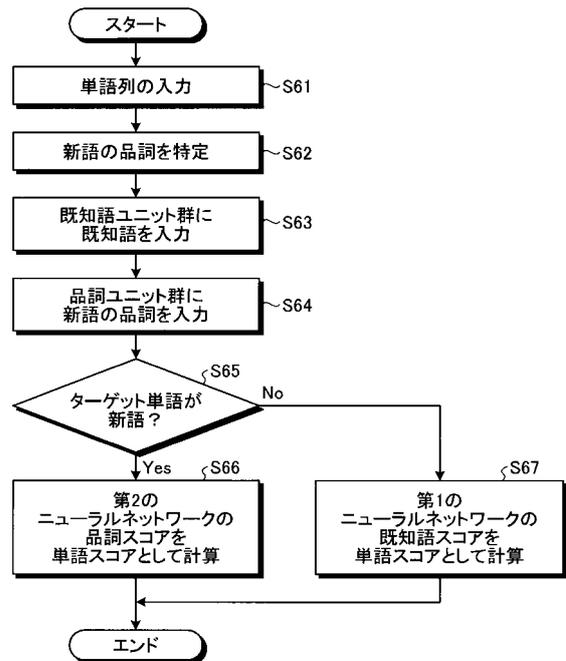
【図9】



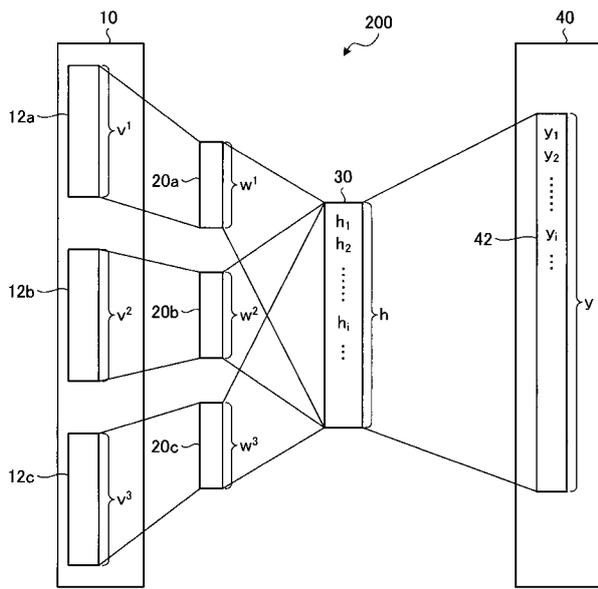
【図10】



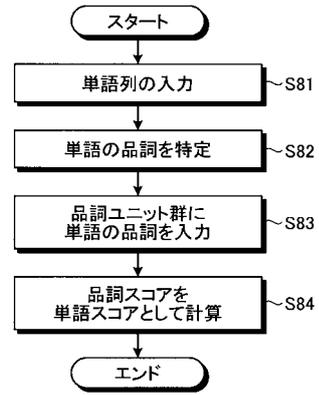
【図11】



【図12】



【図13】



【図14】

