

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6909832号
(P6909832)

(45) 発行日 令和3年7月28日(2021.7.28)

(24) 登録日 令和3年7月7日(2021.7.7)

(51) Int.Cl.		F I	
G 1 O L 25/54	(2013.01)	G 1 O L 25/54	
G 1 O L 15/10	(2006.01)	G 1 O L 15/10	2 O O W
G 1 O L 15/18	(2013.01)	G 1 O L 15/18	3 O O G
G 1 O L 15/16	(2006.01)	G 1 O L 15/16	

請求項の数 30 (全 18 頁)

(21) 出願番号	特願2019-149241 (P2019-149241)	(73) 特許権者	514322098
(22) 出願日	令和1年8月16日(2019.8.16)		ベイジン バイドウ ネットコム サイエ ンス アンド テクノロジー カンパニー リミテッド
(65) 公開番号	特開2020-30408 (P2020-30408A)		中華人民共和国 ペキン 100085, ハイディアン ディストリクト, シャ ンディ テンス ストリート, 10番, バイドウ キャンパス 2階
(43) 公開日	令和2年2月27日(2020.2.27)		
審査請求日	令和1年8月16日(2019.8.16)	(74) 代理人	100107515
(31) 優先権主張番号	201810949733.8		弁理士 廣田 浩一
(32) 優先日	平成30年8月20日(2018.8.20)	(74) 代理人	100107733
(33) 優先権主張国・地域又は機関	中国 (CN)		弁理士 流 良広
		(74) 代理人	100115347
			弁理士 松田 奈緒子

最終頁に続く

(54) 【発明の名称】 オーディオにおける重要語句を認識するための方法、装置、機器及び媒体

(57) 【特許請求の範囲】

【請求項 1】

第1のトレーニングテキストにおける単語の特徴情報を標識する第1のトレーニングデータを取得するステップと、

第2のトレーニングテキストにおける、主人公、所定のシーン、時間又はそれらの組み合わせを含む重要語句を標識する第2のトレーニングデータを取得するステップと、

前記第1のトレーニングデータと前記第2のトレーニングデータとに基づいて重要語句認識モデルをトレーニングすることにより、オーディオデータにおける重要語句を認識するステップと、をコンピュータが行うことを含む、

ことを特徴とする重要語句認識モデルをトレーニングするための方法。

10

【請求項 2】

前記第1のトレーニングデータを取得するステップは、

前記第1のトレーニングテキストを取得するステップと、

前記第1のトレーニングテキストを前処理することにより前記第1のトレーニングデータを生成するステップと、を含む、

ことを特徴とする請求項1に記載の重要語句認識モデルをトレーニングするための方法

。

【請求項 3】

前記第1のトレーニングテキストを取得するステップは、

前記重要語句認識モデルをトレーニングするための第1のオーディオサンプルを取得す

20

るステップと、

音声認識技術で前記第 1 のオーディオサンプルを前記第 1 のトレーニングテキストに変換するステップと、を含む、

ことを特徴とする請求項 2 に記載の重要語句認識モデルをトレーニングするための方法。

【請求項 4】

前記第 1 のトレーニングテキストを前処理するステップは、

前記第 1 のトレーニングテキストを少なくとも一つの文に分割するステップと、

自然言語処理技術で前記少なくとも一つの文における前記単語の前記特徴情報を決定するステップと、を含む、

ことを特徴とする請求項 2 に記載の重要語句認識モデルをトレーニングするための方法。

10

【請求項 5】

前記特徴情報は、前記単語のテキストと、品詞と、意味と、文法情報とのうち少なくとも一つを含む、

ことを特徴とする請求項 4 に記載の重要語句認識モデルをトレーニングするための方法。

【請求項 6】

前記第 2 のトレーニングデータを取得するステップは、

前記第 2 のトレーニングテキストを取得するステップと、

前記第 2 のトレーニングテキストを前処理することにより、前記第 2 のトレーニングデータを生成するステップと、を含む、

ことを特徴とする請求項 1 に記載の重要語句認識モデルをトレーニングするための方法。

20

【請求項 7】

前記第 2 のトレーニングテキストを取得するステップは、

前記重要語句認識モデルをトレーニングするための第 2 のオーディオサンプルを取得するステップと、

音声認識技術で前記第 2 のオーディオサンプルを前記第 2 のトレーニングテキストに変換するステップと、を含む、

ことを特徴とする請求項 6 に記載の重要語句認識モデルをトレーニングするための方法。

30

【請求項 8】

前記第 2 のトレーニングテキストを前処理するステップは、

前記第 2 のトレーニングテキストを少なくとも一つの文に分割するステップと、

複数のタグで前記少なくとも一つの文における前記重要語句を標識するステップと、を含む、

ことを特徴とする請求項 6 に記載の重要語句認識モデルをトレーニングするための方法。

【請求項 9】

複数のタグで前記少なくとも一つの文における前記重要語句を標識するステップは、

第 1 のタグで前記重要語句の先頭文字を標識するステップと、

第 2 のタグで前記重要語句における前記先頭文字に続く後続文字を標識するステップと

、
第 3 のタグで前記少なくとも一つの文における前記重要語句に属しない文字を標識するステップと、を含む、

ことを特徴とする請求項 8 に記載の重要語句認識モデルをトレーニングするための方法。

40

【請求項 10】

認識すべきオーディオデータを取得するステップと、

50

トレーニングされた重要語句認識モデルにより前記オーディオデータにおける重要語句を認識するステップと、をコンピュータが行うことを含み、

前記重要語句認識モデルは、第1のトレーニングデータと第2のトレーニングデータとに基づいてトレーニングされ、

前記第1のトレーニングデータは、第1のトレーニングテキストにおける単語の特徴情報を標識し、

前記第2のトレーニングデータは、第2のトレーニングテキストにおける重要語句を標識し、

前記重要語句は、主人公、所定のシーン、時間又はそれらの組み合わせを含む、

ことを特徴とするオーディオにおける重要語句を認識するための方法。

10

【請求項11】

前記オーディオデータにおける重要語句を認識するステップは、

音声認識技術で前記オーディオデータを前記オーディオデータに対応したテキストに変換するステップと、

前記テキストを少なくとも一つの文に分割するステップと、

前記重要語句認識モデルで前記少なくとも一つの文における文字に対応する対応タグを決定するステップと、

前記対応タグに基づいて前記オーディオデータにおける重要語句を認識するステップと、を含む、

ことを特徴とする請求項10に記載のオーディオにおける重要語句を認識するための方法。

20

【請求項12】

前記対応タグは、

前記文字が前記重要語句の先頭文字であることを示す第1のタグと、

前記文字が前記重要語句における前記先頭文字に続く後続文字であることを示す第2のタグと、

前記文字が前記重要語句に属しないことを示す第3のタグと、のうち一つを含む、

ことを特徴とする請求項11に記載のオーディオにおける重要語句を認識するための方法。

30

【請求項13】

前記対応タグに基づいて前記オーディオデータにおける重要語句を認識するステップは、

前記第1のタグで標識された前記先頭文字と前記第2のタグで標識された前記後続文字との集合を前記重要語句として認識するステップを含む、

ことを特徴とする請求項12に記載のオーディオにおける重要語句を認識するための方法。

【請求項14】

第1のトレーニングテキストにおける単語の特徴情報を標識する第1のトレーニングデータを取得するように構成される第1の取得モジュールと、

第2のトレーニングテキストにおける、主人公、所定のシーン、時間又はそれらの組み合わせを含む重要語句を標識する第2のトレーニングデータを取得するように構成される第2の取得モジュールと、

40

前記第1のトレーニングデータと前記第2のトレーニングデータとに基づいて重要語句認識モデルをトレーニングすることにより、オーディオデータにおける重要語句を認識するように構成されるモデルトレーニングモジュールと、を含む、

ことを特徴とする重要語句認識モデルをトレーニングするための装置。

【請求項15】

前記第1の取得モジュールは、

前記第1のトレーニングテキストを取得するように構成される第1の取得ユニットと、

前記第1のトレーニングテキストを前処理することにより前記第1のトレーニングデー

50

タを生成する第 1 の前処理ユニットと、を含む、

ことを特徴とする請求項 1 4 に記載の重要語句認識モデルをトレーニングするための装置。

【請求項 1 6】

前記第 1 の取得ユニットは、さらに、

前記重要語句認識モデルをトレーニングするための第 1 のオーディオサンプルを取得し、

音声認識技術で前記第 1 のオーディオサンプルを前記第 1 のトレーニングテキストに変換するように構成される、

ことを特徴とする請求項 1 5 に記載の重要語句認識モデルをトレーニングするための装置。 10

【請求項 1 7】

前記第 1 の前処理ユニットは、さらに、

前記第 1 のトレーニングテキストを少なくとも一つの文に分割し、

自然言語処理技術で前記少なくとも一つの文における前記単語の前記特徴情報を決定するように構成される、

ことを特徴とする請求項 1 5 に記載の重要語句認識モデルをトレーニングするための装置。

【請求項 1 8】

前記特徴情報は、前記単語のテキストと、品詞と、意味と、文法情報とのうち少なくとも一つを含む、 20

ことを特徴とする請求項 1 7 に記載の重要語句認識モデルをトレーニングするための装置。

【請求項 1 9】

前記第 2 の取得モジュールは、

前記第 2 のトレーニングテキストを取得するように構成される第 2 の取得ユニットと、

前記第 2 のトレーニングテキストを前処理することにより、前記第 2 のトレーニングデータを生成するように構成される第 2 の前処理ユニットと、を含む、

ことを特徴とする請求項 1 4 に記載の重要語句認識モデルをトレーニングするための装置。 30

【請求項 2 0】

前記第 2 の取得ユニットは、さらに、

前記重要語句認識モデルをトレーニングするための第 2 のオーディオサンプルを取得し、

音声認識技術で前記第 2 のオーディオサンプルを前記第 2 のトレーニングテキストに変換するように構成される、

ことを特徴とする請求項 1 9 に記載の重要語句認識モデルをトレーニングするための装置。

【請求項 2 1】

前記第 2 の前処理ユニットは、さらに、 40

前記第 2 のトレーニングテキストを少なくとも一つの文に分割し、

複数のタグで前記少なくとも一つの文における前記重要語句を標識するように構成される、

ことを特徴とする請求項 1 9 に記載の重要語句認識モデルをトレーニングするための装置。

【請求項 2 2】

前記第 2 の前処理ユニットは、さらに、

第 1 のタグで前記重要語句の先頭文字を標識し、

第 2 のタグで前記重要語句における前記先頭文字に続く後続文字を標識し、

第 3 のタグで前記少なくとも一つの文における前記重要語句に属しない文字を標識する 50

ように構成される、

ことを特徴とする請求項 2 1 に記載の重要語句認識モデルをトレーニングするための装置。

【請求項 2 3】

認識すべきオーディオデータを取得するように構成されるオーディオ取得モジュールと

、
トレーニングされた重要語句認識モデルにより前記オーディオデータにおける重要語句を認識するように構成される認識モジュールと、を含み、

前記重要語句認識モデルは、第 1 のトレーニングデータと第 2 のトレーニングデータとに基づいてトレーニングされ、

前記第 1 のトレーニングデータは、第 1 のトレーニングテキストにおける単語の特徴情報を標識し、

前記第 2 のトレーニングデータは、第 2 のトレーニングテキストにおける重要語句を標識し、

前記重要語句は、主人公、所定のシーン、時間又はそれらの組み合わせを含む、
ことを特徴とするオーディオにおける重要語句を認識するための装置。

【請求項 2 4】

前記認識モジュールは、

音声認識技術で前記オーディオデータを前記オーディオデータに対応したテキストに変換するように構成されるデータ変換ユニットと、

前記テキストを少なくとも一つの文に分割するように構成されるテキスト分割ユニットと、

前記重要語句認識モデルで前記少なくとも一つの文における文字に対応する対応タグを決定するように構成されるタグ決定ユニットと、

前記対応タグに基づいて前記オーディオデータにおける重要語句を認識するように構成される認識ユニットと、を含み、

ことを特徴とする請求項 2 3 に記載のオーディオにおける重要語句を認識するための装置。

【請求項 2 5】

前記対応タグは、

前記文字が前記重要語句の先頭文字であることを示す第 1 のタグと、

前記文字が前記重要語句における前記先頭文字に続く後続文字であることを示す第 2 のタグと、

前記文字が前記重要語句に属しないことを示す第 3 のタグと、のうち一つを含む、

ことを特徴とする請求項 2 4 に記載のオーディオにおける重要語句を認識するための装置。

【請求項 2 6】

前記認識ユニットは、さらに、

前記第 1 のタグで標識された前記先頭文字と前記第 2 のタグで標識された前記後続文字との集合を前記重要語句として認識するように構成される、

ことを特徴とする請求項 2 5 に記載のオーディオにおける重要語句を認識するための装置。

【請求項 2 7】

一つ又は複数のプロセッサと、

一つ又は複数のプログラムを記憶するための記憶装置と、を含み、

前記一つ又は複数のプログラムが前記一つ又は複数のプロセッサによって実行される場合に、前記一つ又は複数のプロセッサが請求項 1 ~ 9 のいずれかに記載の方法を実現する、

ことを特徴とする電子機器。

【請求項 2 8】

10

20

30

40

50

一つ又は複数のプロセッサと、
一つ又は複数のプログラムを記憶するための記憶装置と、を含み、
前記一つ又は複数のプログラムが前記一つ又は複数のプロセッサによって実行される場合に、前記一つ又は複数のプロセッサが請求項 10 ~ 13 のいずれかに記載の方法を実現する、
ことを特徴とする電子機器。

【請求項 29】

コンピュータプログラムが記憶されているコンピュータ読み取り可能な記憶媒体であって、
前記コンピュータプログラムがプロセッサによって実行される場合に、請求項 1 ~ 9 のいずれかに記載の方法が実現される、
ことを特徴とするコンピュータ読み取り可能な記憶媒体。

10

【請求項 30】

コンピュータプログラムが記憶されているコンピュータ読み取り可能な記憶媒体であって、
前記コンピュータプログラムがプロセッサによって実行される場合に、請求項 10 ~ 13 のいずれかに記載の方法が実現される、
ことを特徴とするコンピュータ読み取り可能な記憶媒体。

【発明の詳細な説明】

【技術分野】

20

【0001】

本開示の実施例は、データ処理分野に関し、具体的に、オーディオにおける重要語句を認識するための方法、装置、機器及びコンピュータ読み取り可能な記憶媒体に関する。

【背景技術】

【0002】

コンピュータ技術の発展とスマート端末の普及に伴い、スマートフォン又はスマートスピーカボックス等の端末機器を用いて種々なトーキングショー、例えばトーキングフィクション、童話、トレーニングコース等を聴取する機会が多くなっている。トーキングリソースのプロバイダは、通常、収録されたオーディオデータをサーバにアップロードして簡単なタイトルとイントロダクションを付けるにすぎず、このため、ユーザがこれらのリソースを検索する場合に、システムは、通常、タイトル及びイントロダクションによってしか検索できない。オーディオデータのコンテンツを全面、迅速に理解できないため、このような検索効果は、ユーザの要求を満足できない。例えば、ユーザからの照会命令におけるキーワードが、オーディオデータのタイトル又はイントロダクションに含まれない場合に、該オーディオデータは、検索されない。

30

【発明の概要】

【0003】

本開示の例示的な実施例によれば、オーディオにおける重要語句を認識するための態様を提供する。

【0004】

40

本開示の第 1 態様では、重要語句認識モデルをトレーニングするための方法を提供する。該方法は、第 1 のトレーニングテキストにおける単語の特徴情報を標識する第 1 のトレーニングデータを取得するステップを含む。該方法は、第 2 のトレーニングテキストにおける重要語句を標識する第 2 のトレーニングデータを取得するステップをさらに含む。該方法は、第 1 のトレーニングデータと第 2 のトレーニングデータとに基づいて重要語句認識モデルをトレーニングすることにより、オーディオデータにおける重要語句を認識するステップをさらに含む。

【0005】

本開示の第 2 態様では、オーディオにおける重要語句を認識するための方法を提供する。該方法は、認識すべきオーディオデータを取得するステップを含む。該方法は、トレー

50

ニングされた重要語句認識モデルによりオーディオデータにおける重要語句を認識するステップをさらに含む。該重要語句認識モデルは、第1のトレーニングデータと第2のトレーニングデータとに基づいてトレーニングされ、ここで、第1のトレーニングデータは、第1のトレーニングテキストにおける単語の特徴情報を標識し、第2のトレーニングデータは、第2のトレーニングテキストにおける重要語句を標識する。

【0006】

本開示の第3態様では、重要語句認識モデルをトレーニングするための装置を提供する。該装置は、第1のトレーニングテキストにおける単語の特徴情報を標識する第1のトレーニングデータを取得するように構成される第1の取得モジュールと、第2のトレーニングテキストにおける重要語句を標識する第2のトレーニングデータを取得するように構成される第2の取得モジュールと、第1のトレーニングデータと第2のトレーニングデータとに基づいて重要語句認識モデルをトレーニングすることにより、オーディオデータにおける重要語句を認識するモデルトレーニングモジュールとを含む。

10

【0007】

本開示の第4態様では、オーディオにおける重要語句を認識するための装置を提供する。該装置は、認識すべきオーディオデータを取得するように構成されるオーディオ取得モジュールと、トレーニングされた重要語句認識モデルによりオーディオデータにおける重要語句を認識するが、重要語句認識モデルは、第1のトレーニングデータと第2のトレーニングデータとに基づいてトレーニングされ、第1のトレーニングデータは、第1のトレーニングテキストにおける単語の特徴情報を標識し、第2のトレーニングデータは、第2

20

【0008】

本開示の第5態様では、一つ又は複数のプロセッサ及び記憶装置を含む電子機器を提供する。記憶装置は、一つ又は複数のプログラムを記憶するためのものである。一つ又は複数のプログラムが一つ又は複数のプロセッサによって実行される場合に、一つ又は複数のプロセッサは、本開示の第1態様に係る方法を実行させる。

【0009】

本開示の第6態様では、一つ又は複数のプロセッサ及び記憶装置を含む電子機器を提供する。記憶装置は、一つ又は複数のプログラムを記憶するためのものである。一つ又は複数のプログラムが一つ又は複数のプロセッサによって実行される場合に、一つ又は複数のプロセッサは、本開示の第2態様に係る方法を実行させる。

30

【0010】

本開示の第7態様では、コンピュータプログラムが記憶されているコンピュータ読み取り可能な媒体を提供し、該コンピュータプログラムがプロセッサによって実行される場合、本開示の第1態様に係る方法を実現する。

【0011】

本開示の第8態様では、コンピュータプログラムが記憶されているコンピュータ読み取り可能な媒体を提供し、該コンピュータプログラムがプロセッサによって実行される場合、本開示の第2態様に係る方法を実現する。

40

【0012】

なお、発明内容部分に説明する内容は、本開示の実施例の肝心又は重要特徴を制限することを意図しておらず、本開示の範囲を制限することを意図していない。本開示の他の特徴は、以下の説明により容易に理解され得る。

【図面の簡単な説明】

【0013】

本発明の上述及び/又は付加的特徴と利点は、下記の添付図面を参照した実施形態に対する説明により、明らかになり、理解されることが容易になる。

【図1】本開示の実施例を実現できる例示的なシステムを示す概略図である。

【図2】本開示の実施例に係る重要語句認識モデルをトレーニングするための例示的な方

50

法を示すフローチャートである。

【図3】本開示の実施例に係るオーディオにおける重要語句を認識するための例示的な方法を示すフローチャートである。

【図4】本開示の実施例に係る重要語句認識モデルをトレーニングするための装置を示す模式ブロック図である。

【図5】本開示の実施例に係るオーディオにおける重要語句を認識するための装置を示す模式ブロック図である。

【図6】本開示の一部の実施例を実現できる計算機器を示すブロック図である。

【発明を実施するための形態】

【0014】

以下に、図面を参照しながら本開示の実施例をさらに詳しく説明する。図面には、本開示の一部の実施例を示したが、本開示は、種々な形式により実現されることができ、ここで説明した実施例に限定されると理解されてはならない。逆に、本開示を明らかで、完全に理解するために、これらの実施例を提供する。なお、本開示の図面及び実施例は、例示的な作用を奏し、本開示の保護範囲を限定するものと理解されてはならない。

【0015】

本開示の実施例の説明において、「含む」の用語及びその類似用語とは、開放的に含まれることをいい、即ち、「含むがこれらに限定されない」ことを指す。なお、「基づく」は、「少なくとも部分に基づく」ことをいう。なお、「一実施例」又は「当該実施例」は、「少なくとも一つの実施例」をいう。「第1」、「第2」等は、異なる又は同一の対象を指す。以下、その他の明確及び暗黙的な定義があることもある。

【0016】

上述のとおり、コンピュータ技術の発展とスマート端末の普及に伴い、スマートフォン又はスマートスピーカボックス等の端末機器を用いて種々なトーキングショー、例えばトーキングフィクション、童話、トレーニングコース等を聴取するケースが多くなっている。トーキングリソースのプロバイダは、通常、収録されたオーディオデータをサーバにアップロードして簡単なタイトルとイントロダクションを付けるにすぎず、このため、ユーザがこれらのリソースを検索する場合に、システムは、通常、タイトル及びイントロダクションによってしか検索できない。オーディオデータのコンテンツを全面、迅速に理解するために、このような検索効果は、ユーザの要求を満足できない。この場合に、如何にオーディオファイルのコンテンツを快速に理解することにより、そのキーワード又は重要語句を正確、効率的に認識するのは、早急な解決の待たれるチャレンジである。いくつかの従来態様では、人気オーディオに対して手動標記の方式でそのキーワード又は重要語句、例えば人気童話での主人公等を標記する。しかしながら、この態様は、効率が低く、大規模で適用することができない。さらに、従来態様では、テキストに対してキーワード又は重要語句を検索する態様がある。しかしながら、これらの態様は、オーディオデータに対して正確、効率的に検索することができない。

【0017】

本開示の実施例によれば、オーディオにおける重要語句を認識するための態様を提供する。該態様は、自然言語処理されたトレーニングサンプルと手動序列標記されたトレーニングサンプルとの両方により、重要語句認識モデルをトレーニングする。該態様では、認識すべきオーディオデータを対応したテキストに変換するとともに、トレーニングされた重要語句認識モデルによりテキストにおけるキーワード又は重要語句を認識する。このように、該態様では、オーディオデータのコンテンツを正確、効率的に理解するとともに、その重要情報を抽出することができる。そのため、該態様では、種々なトーキング用途又はオーディオ検索システムに広く適用することができる。

【0018】

以下、図面を参照しながら本開示の実施例を具体的に説明する。本発明において、「モデル」の用語は、トレーニングデータから対応した入力と出力との関連を学習することにより、トレーニング終了後、所定の入力に対して、対応した出力を生成することができる

10

20

30

40

50

。なお、「モデル」は「ニューラルネット」、「学習モデル」又は「学習ネットワーク」とも呼ばれる。「重要語句」の用語とは、一のコンテンツに出現する一つ又は複数のキーワードをいう。「重要語句」は、ユーザの意図によって決定され、トーキングコンテンツでの主人公、所定のシーン、時間又はそれらの組み合わせが挙げられるが、これらに限定されない。

【 0 0 1 9 】

図1は、本開示の実施例を実現できる例示的なシステム100を示す概略図である。システム100は、全体的にモデルトレーニングサブシステム110とモデル応用サブシステム120とを含むことができる。なお、システム100の構造と機能への説明は、例示的なものに過ぎず、本開示の範囲に対するいかなる限定を示唆することない。本開示の実施例は、さらに構造及び/又は機能の異なる環境に適用することができる。

10

【 0 0 2 0 】

モデルトレーニングサブシステム110において、モデルトレーニング装置111は、第1のトレーニングデータ101と第2のトレーニングデータ102を取得することにより、重要語句認識モデル103をトレーニングすることができる。例えば、第1のトレーニングデータ101は、第1のトレーニングテキストにおける単語の特徴情報を標識することができる。第2のトレーニングデータ102は、第2のトレーニングテキストにおける重要語句を標識することができる。

【 0 0 2 1 】

一部の実施例では、第1のトレーニングテキスト又は第2のトレーニングテキストは、任意のテキスト又はオーディオデータから変換して得られたテキストであってもよい。例えば、第1のトレーニングテキストは、重要語句認識モデル103をトレーニングする第1のオーディオサンプルから変換して得ることができる。第2のトレーニングテキストは、重要語句認識モデル103をトレーニングする第2のオーディオサンプルから変換して得ることができる。一部の実施例では、第1のオーディオサンプルと第2のオーディオサンプルは、同一又は異なるオーディオデータであってもよい。即ち、第1のトレーニングテキストと第2のトレーニングテキストは、同一又は異なってもよい。

20

【 0 0 2 2 】

一部の実施例では、第1のトレーニングテキストを前処理することにより、第1のトレーニングデータ101を生成することができる。例えば、第1のトレーニングテキストを文の区切り(例えば、種々な句読点等)により複数の文に分割した後、その各文に対して自然言語処理を行うことにより、文における各単語の特徴情報を取得することができる。例えば、単語の特徴情報は、該単語のテキスト、品詞、意味及び/又は文法情報等を含むことができる。特徴情報における品詞は、例えば該単語が名詞、動詞、形容詞、又は副詞等のいずれかに属することを示すことができる。特徴情報における意味は、例えば、該単語の具体的な意味、例えば人物、動作、環境等を示すことができる。特徴情報における文法情報は、該単語が文における主語、述語、目的語、限定語、副詞又は補語等のいずれかに属することを示すことができる。

30

【 0 0 2 3 】

一部の実施例では、第1のトレーニングテキストの取得と第1のトレーニングデータ101の生成の一部又は全体は、モデルトレーニング装置111により実行されてもよい。即ち、モデルトレーニング装置111は、任意のテキストを第1のトレーニングテキストとするか、又は音声認識技術で第1のオーディオサンプルを第1のトレーニングテキストに変換することができる。付加的又は選択的に、モデルトレーニング装置111は、第1のトレーニングテキストに基づいて第1のトレーニングデータ101を自己生成することができる。

40

【 0 0 2 4 】

選択的に、他の一部の実施例において、第1のトレーニングテキストの取得と第1のトレーニングデータ101の生成の一部又は全体は、他の装置により実行されることができ。この場合に、モデルトレーニング装置111は、例えば当該他の装置から、生成され

50

た第1のトレーニングデータ101を取得して、そのままモデルトレーニングに用いることができる。

【0025】

一部の実施例では、第2のトレーニングテキストを前処理することにより、第2のトレーニングデータ102を生成することができる。例えば、第2のトレーニングテキストを文の区切り（例えば、種々な句読点等）により複数の文に分割した後、複数のタグで各文における重要語句を標識することができる。一部の実施例では、例えば、B（Begin）、I（Inside）、O（Outside）タグで文における重要語句を手動に標識することができるが、ここでBタグ（本発明で、「第1のタグ」とも記す）は、重要語句の先頭文字を標記するためのものであり、Iタグ（本発明で、「第2のタグ」とも記す）は、重要語句における先頭文字以外の他の文字を標記するためのものであり、Oタグ（本発明で、「第3のタグ」とも記す）は、文における重要語句に属しない他の文字を標記するためのものである。

10

【0026】

例えば、標記される文が「白雪姫と七人の小人が森にいる」であり、且つ標記される重要語句が該文における主人公（即ち、「白雪姫」及び「七人の小人」）であると、「白」と「七」は、それぞれBタグで標記され、「雪姫」と「人の小人」とは、それぞれIタグで標記され、他の文字は、Oタグで標記される。「白雪姫と七人の小人が森にいる」を例とし、標記される重要語句は、該文におけるシーン（即ち、「森にいる」）であると、「森」は、Bタグで標記され、「にいる」は、Iタグで標記され、他の文字は、Oタグで標記される。

20

【0027】

以上、BIOタグを例として、第2のトレーニングテキストにおける重要語句に対して手動に標記することを説明している。しかしながら、なお、これは説明のための例示に過ぎず、本発明を限定するものではない。本開示の他の一部の実施例では、BIOタグ以外の他のタグで第2のトレーニングテキストにおける重要語句を標記することもでき、本開示の範囲は、この態様で限定されない。

【0028】

一部の実施例では、第2のトレーニングテキストの取得と第2のトレーニングデータ102の生成の一部又は全体は、モデルトレーニング装置111によって実行されてもよい。即ち、モデルトレーニング装置111は、任意のテキストを第2のトレーニングテキストとするか、又は音声認識技術で第2のオーディオサンプルを第2のトレーニングテキストに変換することができる。付加的又は選択的に、モデルトレーニング装置111は、第2のトレーニングテキストに基づいて第2のトレーニングデータ102を自己生成することができる。

30

【0029】

選択的に、他の一部の実施例において、第2のトレーニングテキストの取得と第2のトレーニングデータ102の生成の一部又は全体は、他の装置によって実行されてもよい。この場合に、モデルトレーニング装置111は、例えば当該他の装置から生成された第2のトレーニングデータ102を取得して、そのままモデルトレーニングに用いることができる。

40

【0030】

一部の実施例では、モデルトレーニング装置111は、第1のトレーニングデータ101と第2のトレーニングデータ102に基づいて重要語句認識モデル103をトレーニングすることができる。例えば、重要語句認識モデル103の例は、再帰ニューラルネットワーク（例えば、長期及び短期記憶（LSTM）モデル）、コンボリユーションアルニューラルネットワーク（CNN）モデル等を含むことができるが、これらに限定されない。トレーニングされた重要語句認識モデル103は、オーディオデータにおける重要語句を認識するためのものであってもよい。

【0031】

50

図1に示すように、モデル応用サブシステム120では、モデル応用装置121は、認識すべきオーディオデータ104を取得し、トレーニングされた重要語句認識モデル103によりオーディオデータ104における重要語句を認識することができる。オーディオデータ104は、任意の認識すべきオーディオファイル、オーディオデータフロー等を含むことができる。例えば、モデル応用装置121は、認識すべきオーディオデータフローをリアルタイムに取得し、又は所定の記憶装置から認識すべきオーディオファイルをオーディオデータ104として取得することができる。

【0032】

一部の実施例では、モデル応用装置121は、音声認識技術でオーディオデータ104をそれに対応したテキストに変換するとともに、得られたテキストを文の区切りにより一つ又は複数の文に分割することができる。その後、モデル応用装置121は、重要語句認識モデル103により各文における文字に対応する対応タグを決定し、対応タグに基づいて文における重要語句を認識することができる。

10

【0033】

一部の実施例では、例えば重要語句認識モデル103をトレーニングするための第2のトレーニングデータ102において、BIOタグで第2のトレーニングテキストにおける重要語句を標記することができる。この場合に、モデル応用装置121は、重要語句認識モデル103を認識し、各文における各文字のタグをBタグ、Iタグ及びOタグの一つに決定することができる。ここで、Bタグは、該文字が重要語句の先頭文字であることを示し、Iタグは、該文字が重要語句における先頭文字以外の文字であることを示し、Oタグは、該文字が重要語句に属しないことを示すことができる。さらに、モデル応用装置121は、Bタグで標識された文字とIタグで標識された文字の集合を重要語句として認識してもよい。選択的に、他の一部の実施例では、重要語句認識モデル103をトレーニングするための第2のトレーニングデータ102において、BIOタグ以外の複数の他のタグで第2のトレーニングテキストにおける重要語句を標記することもできる。この場合に、モデル応用装置121は、重要語句認識モデル103により各文における文字又は単語の対応タグを、例えば複数の他のタグの一つとして決定し、さらに決定されたタグで重要語句を認識することができる。

20

【0034】

図2は、本開示の実施例に係る重要語句認識モデルをトレーニングするための例示的な方法200を示すフローチャートである。例えば、方法200は、図1に示すようなモデルトレーニング装置111によって実行されることができる。以下、図1を組み合わせる方法200の各動作を詳しく説明する。なお、方法200は、不図示の付加動作及び/又は図示を省略してもよい動作を含むことができる。本開示の範囲は、この態様で限定されない。

30

【0035】

ブロック210で、モデルトレーニング装置111は、第1のトレーニングデータ101を取得する。例えば、第1のトレーニングデータ101は、第1のトレーニングテキストにおける単語の特徴情報を標識することができる。

【0036】

一部の実施例では、第1のトレーニングテキストは取得されることができ、また第1のトレーニングテキストを前処理することにより第1のトレーニングデータ101を生成することができる。例えば、任意のテキストを第1のトレーニングテキストとして取得してもよい。また、重要語句認識モデル103をトレーニングするための第1のオーディオサンプルを取得するとともに、音声認識技術で第1のオーディオサンプルを第1のトレーニングテキストに変換することができる。一部の実施例では、第1のトレーニングデータ101は、文の区切りにより第1のトレーニングテキストを少なくとも一つの文に分割した後、自然言語処理技術で少なくとも一つの文における単語の特徴情報を決定することにより生成されることができる。例えば、単語の特徴情報は、該単語のテキスト、品詞、意味と文法情報の少なくとも一つを含むことができる。

40

50

【 0 0 3 7 】

ブロック 2 2 0 で、モデルトレーニング装置 1 1 1 は、第 2 のトレーニングデータ 1 0 2 を取得する。例えば、第 2 のトレーニングデータ 1 0 2 は、第 2 のトレーニングテキストにおける重要語句を標識することができる。ここでいう「重要語句」は、ユーザの意図に基づいて決定されることができ、その例は、第 2 のトレーニングテキストにおける主人公、所定のシーン、時間又はその組み合わせ等を含むが、これらに限定されない。

【 0 0 3 8 】

一部の実施例では、第 2 のトレーニングテキストは、取得されることができ、また第 2 のトレーニングテキストを前処理することにより、第 2 のトレーニングデータ 1 0 2 を生成することができる。例えば、任意のテキストを第 2 のトレーニングテキストとして取得することができる。また、重要語句認識モデル 1 0 3 をトレーニングするための第 2 のオーディオサンプルを取得するとともに、音声認識技術で第 2 のオーディオサンプルを第 2 のトレーニングテキストに変換することができる。一部の実施例では、第 2 のトレーニングデータ 1 0 2 は、文の区切りにより第 2 のトレーニングテキストを少なくとも一つの文に分割した後、複数のタグで少なくとも一つの文における重要語句を分割することにより生成されることができ、例えば、第 1 のタグで重要語句の先頭文字を表示することができ、第 2 のタグで重要語句での先頭文字に続く後続文字を標識することができ、第 3 のタグで文における重要語句に属しない文字を標識することができる。

【 0 0 3 9 】

ブロック 2 3 0 で、モデルトレーニング装置 1 1 1 は、第 1 のトレーニングデータ 1 0 1 と第 2 のトレーニングデータ 1 0 2 に基づいて重要語句認識モデル 1 0 3 をトレーニングする。例えば、重要語句認識モデル 1 0 3 の例は、再帰ニューラルネットワーク（例えば、長期及び短期記憶（LSTM）モデル）、コンボリユーションアルニューラルネットワーク（CNN）モデル等を含むが、これらに限定されない。トレーニングされた重要語句認識モデル 1 0 3 は、オーディオデータにおける重要語句を認識するためのものであってもよい。

【 0 0 4 0 】

図 3 は、本開示の実施例に係るオーディオにおける重要語句を認識するための方法 3 0 0 を示すフローチャートである。例えば、方法 3 0 0 は、図 1 に示すようなモデル応用装置 1 2 1 によって実行されることができ、以下、図 1 を組み合わせて方法 3 0 0 の各動作を詳しく説明する。なお、方法 3 0 0 は、不図示の付加動作及び / 又は図示を省略してもよい動作を含むことができる。本開示の範囲は、この態様で限定されない。

【 0 0 4 1 】

ブロック 3 1 0 で、モデル応用装置 1 2 1 は、認識すべきオーディオデータ 1 0 4 を取得する。例えば、オーディオデータ 1 0 4 は、いかなる認識すべきオーディオフィール、オーディオデータフロー等を含むことができる。一部の実施例では、モデル応用装置 1 2 1 は、リアルタイムに認識すべきオーディオデータフローをオーディオデータ 1 0 4 として取得することができる。選択的に、モデル応用装置 1 2 1 は、所定記憶装置から認識すべきオーディオフィールをオーディオデータ 1 0 4 として取得することができる。

【 0 0 4 2 】

ブロック 3 2 0 で、モデル応用装置 1 2 1 は、トレーニングされた重要語句認識モデル 1 0 3 によりオーディオデータ 1 0 4 における重要語句を認識することができる。

【 0 0 4 3 】

一部の実施例では、モデル応用装置 1 2 1 は、音声認識技術でオーディオデータ 1 0 4 をそれに対応したテキストに変換するとともに、得られたテキストを文の区切りにより一つ又は複数の文に分割することができる。その後、モデル応用装置 1 2 1 は、重要語句認識モデル 1 0 3 により各文における文字に対応する対応タグを決定するとともに、対応タグに基づいて文における重要語句を認識することができる。

【 0 0 4 4 】

一部の実施例では、例えば、文字のタグは、該文字が重要語句の先頭文字であることを

10

20

30

40

50

示す第1のタグと、該文字が重要語句における先頭文字に続く後続文字であることを示す第2のタグと、該文字が重要語句に属しないことを示す第3のタグと、の一つを含むことができる。一部の実施例では、モデル応用装置121は、第1のタグで標識された先頭文字と第2のタグで標識された後続文字との集合を重要語句として認識することができる。

【0045】

以上の説明から分かるように、本開示の実施例は、オーディオにおける重要語句を認識するための態様を提出する。該態様では、自然言語処理されたトレーニングサンプルと手動序列標記されたトレーニングサンプルとの両方により、重要語句認識モデルをトレーニングする。該態様では、認識すべきオーディオデータを対応したテキストに変換し、またトレーニングされた重要語句認識モデルにより、テキストにおけるキーワード又は重要語句を認識することができる。このように、該態様は、オーディオデータのコンテンツを正確、効率的に理解するとともに、その重要情報を抽出することができる。そのため、該態様では、種々なトークング用途又はオーディオ検索システムに広く適用することができる。

10

【0046】

図4は、本開示の実施例に係る重要語句認識モデルをトレーニングするための装置400を示す模式ブロック図である。図4に示すように、装置400は、第1のトレーニングテキストにおける単語の特徴情報を標識する第1のトレーニングデータを取得するように構成される第1の取得モジュール410を含むことができる。装置400は、第2のトレーニングテキストにおける重要語句を標識する第2のトレーニングデータを取得するように構成される第2の取得モジュール420をさらに含むことができる。さらに、装置400は、第1のトレーニングデータと第2のトレーニングデータとに基づいて重要語句認識モデルをトレーニングすることにより、オーディオデータにおける重要語句を認識するように構成されるモデルトレーニングモジュール430をさらに含むことができる。

20

【0047】

一部の実施例では、第1の取得モジュール410は、第1のトレーニングテキストを取得するように構成される第1の取得ユニットと、第1のトレーニングテキストを前処理することにより第1のトレーニングデータを生成するように構成される第1の前処理ユニットとを含む。

【0048】

一部の実施例では、第1の取得ユニットは、さらに、重要語句認識モデルをトレーニングするための第1のオーディオサンプルを取得し、音声認識技術で第1のオーディオサンプルを第1のトレーニングテキストに変換するように構成される。

30

【0049】

一部の実施例では、第1の前処理ユニットは、さらに第1のトレーニングテキストを少なくとも一つの文に分割し、自然言語処理技術で少なくとも一つの文における単語の特徴情報を決定するように構成される。

【0050】

一部の実施例では、特徴情報は、該単語のテキストと、品詞と、意味と、文法情報とのうち少なくとも一つを含む。

40

【0051】

一部の実施例では、第2の取得モジュール420は、第2のトレーニングテキストを取得するように構成される第2の取得ユニットと、第2のトレーニングテキストを前処理することにより第2のトレーニングデータを生成するように構成される第2の前処理ユニットとを含む。

【0052】

一部の実施例では、第2の取得ユニットは、さらに、重要語句認識モデルをトレーニングするための第2のオーディオサンプルを取得し、音声認識技術で第2のオーディオサンプルを第2のトレーニングテキストに変換するように構成される。

【0053】

50

一部の実施例では、第2の前処理ユニットは、さらに、第2のトレーニングテキストを少なくとも一つの文に分割し、複数のタグで少なくとも一つの文における重要語句を標識するように構成される。

【0054】

一部の実施例では、第2の前処理ユニットは、さらに、第1のタグで重要語句の先頭文字を標識し、第2のタグで重要語句における先頭文字に続く後続文字を標識し、第3のタグで少なくとも一つの文における重要語句に属しない文字を標識するように構成される。

【0055】

図5は、本開示の実施例に係るオーディオにおける重要語句を認識するための装置500を示す模式ブロック図である。図5に示すように、装置500は、認識すべきオーディオデータを取得するように構成されるオーディオ取得モジュール510を含むことができる。装置500は、トレーニングされた重要語句認識モデルによりオーディオデータにおける重要語句を認識するように構成される認識モジュール520をさらに含むことができる。ここで、重要語句認識モデルは、第1のトレーニングデータと第2のトレーニングデータとに基づいてトレーニングされ、第1のトレーニングデータは、第1のトレーニングテキストにおける単語の特徴情報を標識し、第2のトレーニングデータは、第2のトレーニングテキストにおける重要語句を標識する。

【0056】

一部の実施例では、認識モジュール520は、音声認識技術でオーディオデータを該オーディオデータに対応したテキストに変換するように構成されるデータ変換ユニットと、該テキストを少なくとも一つの文に分割するように構成されるテキスト分割ユニットと、重要語句認識モデルで少なくとも一つの文における文字に対応する対応タグを決定するように構成されるタグ決定ユニットと、対応タグに基づいてオーディオデータにおける重要語句を認識するように構成される認識ユニットとを含む。

【0057】

一部の実施例では、対応タグは、該文字が重要語句の先頭文字であることを示す第1のタグと、該文字が重要語句における先頭文字に続く後続文字であることを示す第2のタグと、該文字が重要語句に属しないことを示す第3のタグと、のうち一つを含む。

【0058】

一部の実施例では、認識ユニットは、さらに、第1のタグで標識された先頭文字と第2のタグで標識された後続文字との集合を重要語句として認識するように構成される。

【0059】

図6は、本開示の実施例を実現するための例示的な機器600を示す模式ブロック図である。機器600は、図1に記載のようなモデルトレーニング装置111及び/又はモデル応用装置121を実現するためのものであってもよい。図に示すように、コンピュータシステム600は、ROM602に記憶されているコンピュータプログラム命令、又は記憶ユニット608からRAM603にローディングされたコンピュータプログラム命令に基づいて、各種の適切な動作と処理を実行できる中央処理装置(CPU)601を含む。RAM603には、システム600の操作に必要な各種のプログラムとデータがさらに記憶されている。CPU601と、ROM602と、RAM603とは、バス604を介して互いに接続されている。入力/出力(I/O)インタフェース605もバス604に接続される。

【0060】

機器600における複数の部材はI/Oインタフェース605に接続され、前記複数の部材は、例えばキーボード、マウス等の入力ユニット606と、例えば種々なディスプレイ、スピーカ等の出力ユニット607と、例えば磁気ディスク、光ディスク等の記憶ユニット608と、例えばネットワークカード、モデム、無線通信送受信機等の通信ユニット609と、を含む。通信ユニット609は、機器600が例えばインターネットのようなコンピュータネット及び/又は種々なキャリアネットワークによりその他の機器に情報/データを交換することを許可する。

10

20

30

40

50

【0061】

処理ユニット601は、上述の各方法及び処理、例えば方法200、及び/又は300を実行する。例えば、一部の実施例では、方法200、及び/又は300は、コンピュータソフトウェアプログラムとして実現されることができ、機器読取可能な媒体、例えば記憶ユニット608に有形的に含まれる。一部の実施例では、コンピュータプログラムの一部又は全ては、ROM602及び/又は通信ユニット609を経て、機器600にロード及び/又はインストールされる。コンピュータプログラムがRAM603にロードされCPU601によって実行される場合に、上述の方法200、及び/又は300の一つ又は複数のステップに実行される。或いは、その他の実施例では、CPU601は、その他の任意の適当な方式（例えば、ファームウェアを経る）により方法200、及び/又は300を実行されるように構成される。

10

【0062】

本開示で、上記した機能は、少なくとも部分的に一つ又は複数のハードウェア論理装置によって実行されることができ、例えば、非限定的に、使用可能なハードウェア論理装置としては、プログラム可能なゲートアレイ（FPGA）、特定用途向け集積回路（ASIC）、特定用途向け標準品（ASSP）、システムオンチップ（SOC）、複雑プログラム可能論理装置（CPLD）等が挙げられる。

【0063】

本開示の方法を実施するためのプログラムコードは、一つ又は複数のプログラミング言語の任意の組み合わせにより記述されることができ、これらのプログラムコードは、汎用コンピュータ、専門コンピュータ又はその他のプログラミングデータ処理装置のプロセッサ又は制御器に提供し、プログラムコードがプロセッサ又は制御器によって実行される場合に、フローチャート及び/又はブロック図に規定の機能/操作を実施させることができる。プログラムコードは、完全に機器で実行されてもよく、部分に機器で実行されてもよく、独立のパッケージとして部分に機器で実行されるとともに、部分にリモート機器で実行されてもよく、又は完全にリモート機器又はサーバで実行されてもよい。

20

【0064】

本開示の説明において、機器読取可能な媒体は、有形な媒体であってもよく、命令実行システム、装置又は機器の使用、又は命令実行システム、装置又は機器との併用に提供されるプログラムを含み、又は記憶する。機器読取可能な媒体は、機器読取可能な信号媒体又は機器読取可能な記憶媒体であってもよい。機器読取可能な媒体は、電子、磁気、光学、電磁、赤外線、若しくは半導体のシステム、装置、若しくは機器、又は前述の任意の適切な組み合わせを含むことができるがこれらに限定されない。機器読取可能な記憶媒体のさらなる具体例として、1つ若しくは複数のワイヤーを有する電氣的接続、携帯型フロッピー（登録商標）ディスク、ハードディスク、ランダムアクセスメモリ（RAM）、読み取り専用メモリ（ROM）、消去可能プログラム可能ROM（EPROM又はフラッシュメモリ）、光ファイバー、携帯型コンパクトディスク読み取り専用メモリ（CD-ROM）、光学記憶装置、磁気記憶装置、又は前述の任意の組み合わせを含む。

30

【0065】

更に、特定の順番で各操作を説明したが、このような操作を、示される特定の順番又は順位で実行することが求められ、又は図示した操作の全てを実行して所望の結果を取得することが求められる。一定の環境において、複数の任務と並列処理が有利である可能性がある。同様に、以上の説明には、若干の具体的な実現詳細を含むが、本開示の範囲を限定するものと理解されてはならない。単独の実施例の前後に説明したある特徴は、一つの実現に組み合わせて実現されることができ、逆に、一つの実現の前後に説明した種々な特徴は、単独又は任意の適合のサブコンビネーションとして複数の実現に実現されることができ、

40

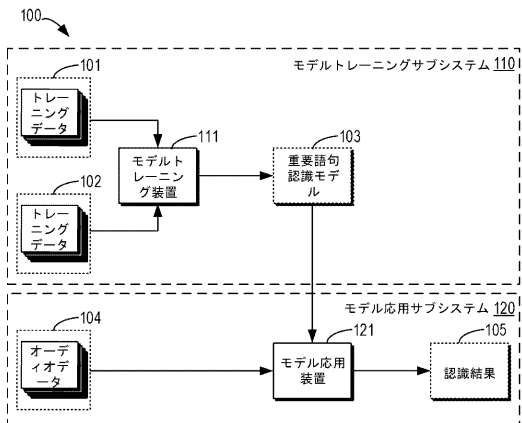
【0066】

構造特徴及び/又は方法論理動作を特定した言語により、本対象を説明したが、特許請求の範囲に限定される対象は、上記した特定の特徴又は動作に限らない。逆に、上記した

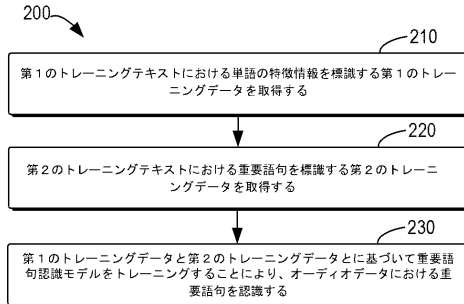
50

特定特徴と動作は、特許請求の範囲の例を実現するためのものに過ぎない。

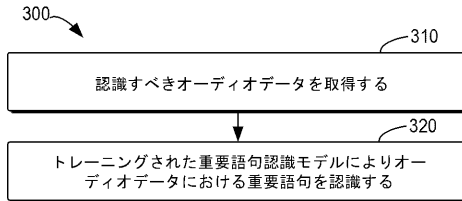
【図1】



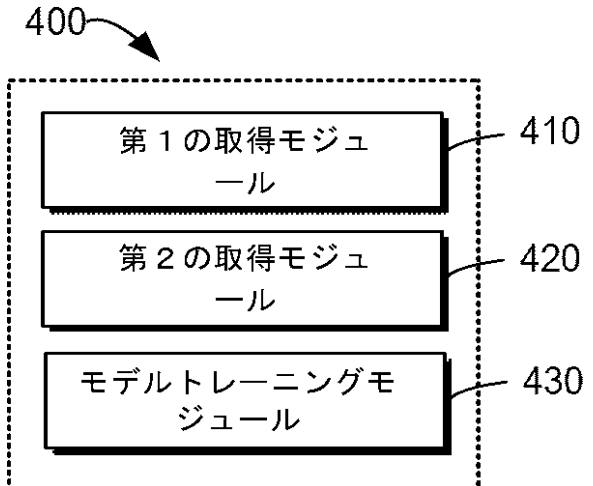
【図2】



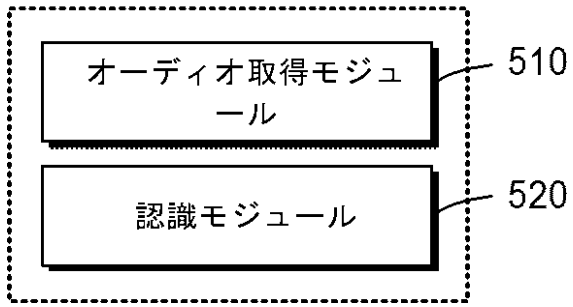
【図3】



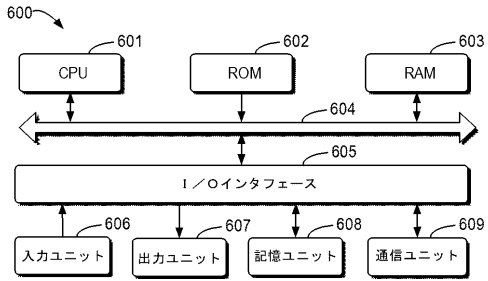
【図4】



【図5】
500



【図6】



フロントページの続き

- (72)発明者 ジーファ・ワン
中華人民共和国 100085 ベイジン ハイディエン ディストリクト シャンディ 10
イーエイチ ストリート ナンバー 10 バイドゥ キャンパス 2 / フロア
- (72)発明者 ティエンシン・ヤン
中華人民共和国 100085 ベイジン ハイディエン ディストリクト シャンディ 10
イーエイチ ストリート ナンバー 10 バイドゥ キャンパス 2 / フロア
- (72)発明者 ジーボン・ウー
中華人民共和国 100085 ベイジン ハイディエン ディストリクト シャンディ 10
イーエイチ ストリート ナンバー 10 バイドゥ キャンパス 2 / フロア
- (72)発明者 ピン・ボン
中華人民共和国 100085 ベイジン ハイディエン ディストリクト シャンディ 10
イーエイチ ストリート ナンバー 10 バイドゥ キャンパス 2 / フロア
- (72)発明者 チェンユェン・チャオ
中華人民共和国 100085 ベイジン ハイディエン ディストリクト シャンディ 10
イーエイチ ストリート ナンバー 10 バイドゥ キャンパス 2 / フロア

審査官 大野 弘

- (56)参考文献 特開2016-061954(JP, A)
特開2004-361873(JP, A)
特開2005-292958(JP, A)

(58)調査した分野(Int.Cl., DB名)

G10L 25/54
G10L 15/10
G10L 15/16
G10L 15/18