

(19) 日本国特許庁(JP)

(12) 特許公報(B1)

(11) 特許番号

特許第5901001号
(P5901001)

(45) 発行日 平成28年4月6日(2016.4.6)

(24) 登録日 平成28年3月18日(2016.3.18)

(51) Int.Cl.	F I
G 1 0 L 15/06 (2013.01)	G 1 0 L 15/06 3 0 0 C

請求項の数 20 (全 24 頁)

(21) 出願番号	特願2015-555551 (P2015-555551)	(73) 特許権者	514187420
(86) (22) 出願日	平成25年10月25日(2013.10.25)		テンセント・テクノロジー・(シェンジェン)・カンパニー・リミテッド
(86) 国際出願番号	PCT/CN2013/085948		中華人民共和国・グアンドン・518000・シェンジェン・フーティアン・ジェン
(87) 国際公開番号	W02014/117548		シン・ロード・エスイージー・パーク・イースト・ブロック・2・ルーム・403
(87) 国際公開日	平成26年8月7日(2014.8.7)	(74) 代理人	100146835
審査請求日	平成27年9月24日(2015.9.24)		弁理士 佐伯 義文
(31) 優先権主張番号	201310040085.1	(74) 代理人	100140534
(32) 優先日	平成25年2月1日(2013.2.1)		弁理士 木内 敬二
(33) 優先権主張国	中国 (CN)		
早期審査対象出願			

最終頁に続く

(54) 【発明の名称】 音響言語モデルトレーニングのための方法およびデバイス

(57) 【特許請求の範囲】

【請求項1】

音響言語モデルをトレーニングする方法であって、

1つまたは複数のプロセッサとメモリとを有するデバイスにおいて、

単語クラスラベルを含まない初期単語分割データを得るために、単語クラスラベルを含まない初期言語モデルを使用して、トレーニングコーパス内のトレーニングサンプルに対して単語分割を行うステップと、

単語クラスラベルを含む第1の単語分割データを得るために、単語クラスラベルを含まない前記初期単語分割データに対して単語クラス置換を実行するステップと、

単語クラスラベルを含む第1の言語モデルをトレーニングするために、単語クラスラベルを含む前記第1の単語分割データを使用するステップと、

単語クラスラベルを含む第2の単語分割データを得るように、前記トレーニングコーパス内の前記トレーニングサンプルに対して単語分割を行うために、単語クラスラベルを含む前記第1の言語モデルを使用するステップと、

1つまたは複数の所定の基準を満たす前記第2の単語分割データに従って、前記音響言語モデルをトレーニングするために、単語クラスラベルを含む前記第2の単語分割データを使用するステップと

を含む、方法。

【請求項2】

単語クラスラベルを含む第1の単語分割データを得るために、単語クラスラベルを含ま

10

20

ない前記初期単語分割データに対して単語クラス置換を実行するステップが、

分類用語集内で、単語クラスラベルを含まない前記初期単語分割データ内の1つまたは複数のそれぞれの単語に関するそれぞれの単語クラスラベルを識別するステップと、

単語クラスラベルを含む前記第1の単語分割データを得るために、単語クラスラベルを含まない前記初期単語分割データ内の前記1つまたは複数のそれぞれの単語を、識別されたそれぞれの単語クラスラベルで置換するステップと

を含む、請求項1に記載の方法。

【請求項3】

単語クラスラベルを含む第2の単語分割データを得るように、前記トレーニングコーパス内の前記トレーニングサンプルに対して単語分割を行うために、単語クラスラベルを含む前記第1の言語モデルを使用するステップが、

10

分類用語集内で、前記トレーニングコーパス内の前記トレーニングサンプル内の1つまたは複数のそれぞれの単語に関するそれぞれの単語クラスラベルを識別するステップと、

単語クラスラベルを含む新たなトレーニングサンプルを得るために、前記トレーニングサンプル内の前記1つまたは複数のそれぞれの単語を、識別されたそれぞれの単語クラスラベルで置換するステップと、

単語クラスラベルを含む前記第2の単語分割データを得るために、単語クラスラベルを含む前記第1の言語モデルを使用し、前記新たなトレーニングサンプルに対して単語分割を行うステップと

を含む、請求項1に記載の方法。

20

【請求項4】

単語クラスラベルを含む前記第2の単語分割データを得た後、

前記第1および第2の単語分割データ内の対応するトレーニングサンプルの分割結果を比較するステップと、

前記第1の単語分割データが前記第2の単語分割データと一致するという判断に従って、前記音響言語モデルの前記トレーニングにおいて使用するために前記第2の単語分割データを承認するステップと

をさらに含む、請求項3に記載の方法。

【請求項5】

単語クラスラベルを含む前記第2の単語分割データを得た後、

30

前記第1の単語分割データが前記第2の単語分割データと一致しないという判断に従って、前記第2の単語分割データを使用して前記第1の言語モデルを再トレーニングするステップをさらに含む、請求項4に記載の方法。

【請求項6】

前記第1の言語モデルが再トレーニングされた後、改訂された第2の単語分割データを得るために、単語クラスラベルを含む前記第1の言語モデルを使用して、第2のトレーニングサンプルに対する前記単語分割を繰り返すステップと、

前記改訂された第2の単語分割データが前記第2の単語分割データと一致するという判断に従って、前記音響言語モデルの前記トレーニングにおいて使用するために前記改訂された第2の単語分割データを承認するステップと

40

をさらに含む、請求項5に記載の方法。

【請求項7】

前記第1の単語分割データが前記第2の単語分割データと一致するという判断が、前記第1の単語分割データにおけるそれぞれの単語クラスラベル置換が、前記第2の単語分割データ内のそれぞれの単語クラスラベル置換と同一であるという判断をさらに含む、請求項4に記載の方法。

【請求項8】

音響言語モデルをトレーニングするためのシステムであって、

1つまたは複数のプロセッサと、

命令が記憶されているメモリとを含み、前記命令が、前記1つまたは複数のプロセッサ

50

によって実行されると、前記プロセッサに、

単語クラスラベルを含まない初期単語分割データを得るために、単語クラスラベルを含まない初期言語モデルを使用して、トレーニングコーパス内のトレーニングサンプルに対して単語分割を行う動作と、

単語クラスラベルを含む第1の単語分割データを得るために、単語クラスラベルを含まない前記初期単語分割データに対して単語クラス置換を実行する動作と、

単語クラスラベルを含む第1の言語モデルをトレーニングするために、単語クラスラベルを含む前記第1の単語分割データを使用する動作と、

単語クラスラベルを含む第2の単語分割データを得るように、前記トレーニングコーパス内の前記トレーニングサンプルに対して単語分割を行うために、単語クラスラベルを含む前記第1の言語モデルを使用する動作と、

1つまたは複数の所定の基準を満たす前記第2の単語分割データに従って、前記音響言語モデルをトレーニングするために、単語クラスラベルを含む前記第2の単語分割データを使用する動作と

を実行させる、システム。

【請求項9】

単語クラスラベルを含む第1の単語分割データを得るために、単語クラスラベルを含まない前記初期単語分割データに対して単語クラス置換を実行する動作が、

分類用語集内で、単語クラスラベルを含まない前記初期単語分割データ内の1つまたは複数のそれぞれの単語に関するそれぞれの単語クラスラベルを識別する動作と、

単語クラスラベルを含む前記第1の単語分割データを得るために、単語クラスラベルを含まない前記初期単語分割データ内の前記1つまたは複数のそれぞれの単語を、識別されたそれぞれの単語クラスラベルで置換する動作と

を含む、請求項8に記載のシステム。

【請求項10】

単語クラスラベルを含む第2の単語分割データを得るように、前記トレーニングコーパス内の前記トレーニングサンプルに対して単語分割を行うために、単語クラスラベルを含む前記第1の言語モデルを使用する動作が、

分類用語集内で、前記トレーニングコーパス内の前記トレーニングサンプル内の1つまたは複数のそれぞれの単語に関するそれぞれの単語クラスラベルを識別する動作と、

単語クラスラベルを含む新たなトレーニングサンプルを得るために、前記トレーニングサンプル内の前記1つまたは複数のそれぞれの単語を、識別されたそれぞれの単語クラスラベルで置換する動作と、

単語クラスラベルを含む前記第2の単語分割データを得るために、単語クラスラベルを含む前記第1の言語モデルを使用し、前記新たなトレーニングサンプルに対して単語分割を行う動作と

を含む、請求項8に記載のシステム。

【請求項11】

前記動作が、

単語クラスラベルを含む前記第2の単語分割データを得た後、

前記第1および第2の単語分割データ内の対応するトレーニングサンプルの分割結果を比較する動作と、

前記第1の単語分割データが前記第2の単語分割データと一致するという判断に従って、前記音響言語モデルの前記トレーニングで使用するために前記第2の単語分割データを承認する動作と

をさらに含む、請求項10に記載のシステム。

【請求項12】

前記動作が、

単語クラスラベルを含む前記第2の単語分割データを得た後、

前記第1の単語分割データが前記第2の単語分割データと一致しないという判断に従っ

て、前記第2の単語分割データを使用して前記第1の言語モデルを再トレーニングする動作をさらに含む、請求項11に記載のシステム。

【請求項13】

前記動作が、

前記第1の言語モデルが再トレーニングされた後、改訂された第2の単語分割データを得るために、単語クラスラベルを含む前記第1の言語モデルを使用して、第2のトレーニングサンプルに対する前記単語分割を繰り返す動作と、

前記改訂された第2の単語分割データが前記第2の単語分割データと一致するという判断に従って、前記音響言語モデルの前記トレーニングで使用するために前記改訂された第2の単語分割データを承認する動作と

10

をさらに含む、請求項12に記載のシステム。

【請求項14】

前記第1の単語分割データが前記第2の単語分割データと一致するという判断が、前記第1の単語分割データにおけるそれぞれの単語クラスラベル置換が、前記第2の単語分割データ内のそれぞれの単語クラスラベル置換と同一であるという判断をさらに含む、請求項11に記載のシステム。

【請求項15】

音響言語モデルをトレーニングするための非一時的なコンピュータ可読記憶媒体であって、

1つまたは複数のプロセッサによって実行されると、前記プロセッサに、

20

単語クラスラベルを含まない初期単語分割データを得るために、単語クラスラベルを含まない初期言語モデルを使用して、トレーニングコーパス内のトレーニングサンプルに対して単語分割を行う動作と、

単語クラスラベルを含む第1の単語分割データを得るために、単語クラスラベルを含まない前記初期単語分割データに対して単語クラス置換を実行する動作と、

単語クラスラベルを含む第1の言語モデルをトレーニングするために、単語クラスラベルを含む前記第1の単語分割データを使用する動作と、

単語クラスラベルを含む第2の単語分割データを得るように、前記トレーニングコーパス内の前記トレーニングサンプルに対して単語分割を行うために、単語クラスラベルを含む前記第1の言語モデルを使用する動作と、

30

1つまたは複数の所定の基準を満たす前記第2の単語分割データに従って、前記音響言語モデルをトレーニングするために、単語クラスラベルを含む前記第2の単語分割データを使用する動作と

を実行させる命令が記憶された、コンピュータ可読記憶媒体。

【請求項16】

単語クラスラベルを含む第1の単語分割データを得るために、単語クラスラベルを含まない前記初期単語分割データに対して単語クラス置換を実行する動作が、

分類用語集内で、単語クラスラベルを含まない前記初期単語分割データ内の1つまたは複数のそれぞれの単語に関するそれぞれの単語クラスラベルを識別する動作と、

単語クラスラベルを含む前記第1の単語分割データを得るために、単語クラスラベルを含まない前記初期単語分割データ内の前記1つまたは複数のそれぞれの単語を、識別されたそれぞれの単語クラスラベルで置換する動作と

40

を含む、請求項15に記載のコンピュータ可読記憶媒体。

【請求項17】

単語クラスラベルを含む第2の単語分割データを得るように、前記トレーニングコーパス内の前記トレーニングサンプルに対して単語分割を行うために、単語クラスラベルを含む前記第1の言語モデルを使用する動作が、

分類用語集内で、前記トレーニングコーパス内の前記トレーニングサンプル内の1つまたは複数のそれぞれの単語に関するそれぞれの単語クラスラベルを識別する動作と、

単語クラスラベルを含む新たなトレーニングサンプルを得るために、前記トレーニング

50

サンプル内の前記1つまたは複数のそれぞれの単語を、識別されたそれぞれの単語クラスラベルで置換する動作と、

単語クラスラベルを含む前記第2の単語分割データを得るために、単語クラスラベルを含む前記第1の言語モデルを使用して、前記新たなトレーニングサンプルに対して単語分割を行う動作と

を含む、請求項15に記載のコンピュータ可読記憶媒体。

【請求項18】

前記動作が、

単語クラスラベルを含む前記第2の単語分割データを得た後、

前記第1および第2の単語分割データ内の対応するトレーニングサンプルの分割結果を比較する動作と、

前記第1の単語分割データが前記第2の単語分割データと一致するという判断に従って、前記音響言語モデルの前記トレーニングで使用するために前記第2の単語分割データを承認する動作と

をさらに含む、請求項17に記載のコンピュータ可読記憶媒体。

【請求項19】

前記動作が、

単語クラスラベルを含む前記第2の単語分割データを得た後、

前記第1の単語分割データが前記第2の単語分割データと一致しないという判断に従って、前記第2の単語分割データを使用して前記第1の言語モデルを再トレーニングする動作をさらに含む、請求項18に記載のコンピュータ可読記憶媒体。

【請求項20】

前記動作が、

前記第1の言語モデルが再トレーニングされた後、改訂された第2の単語分割データを得るために、単語クラスラベルを含む前記第1の言語モデルを使用して、第2のトレーニングサンプルに対する前記単語分割を繰り返す動作と、

前記改訂された第2の単語分割データが前記第2の単語分割データと一致するという判断に従って、前記音響言語モデルの前記トレーニングで使用するために前記改訂された第2の単語分割データを承認する動作と

をさらに含む、請求項19に記載のコンピュータ可読記憶媒体。

【発明の詳細な説明】

【技術分野】

【0001】

関連出願

本出願は、その全体が参照により本明細書に組み込まれている、2013年2月1日に出願した、「Method and Device for Training Acoustic Language Models」と題する中国特許出願第CN201310040085.1号の優先権を主張するものである。

【0002】

本開示は、自然言語処理技術の分野に関し、特に、音響言語モデルをトレーニングするための方法およびデバイスに関する。

【背景技術】

【0003】

自然言語は、人間のためのコミュニケーション方法の最も便利で、効果的で、迅速な形態であるので、コンピュータアプリケーションがますます一般的になるにつれて、コンピュータとの直接の自然言語コミュニケーションに対するユーザの期待が増加している。音声認識技術は、コンピュータベースの認識および理解プロセスによって、人間の音声信号を対応するテキストに変更するための技術である。言語モデルは、音声認識の精度を改善する重要な役割を演じる。

【0004】

ハードウェアの性能およびソフトウェアアルゴリズムの制限のために、現在の音声認識

10

20

30

40

50

システムは、言語モデルのサイズに厳しい制限を課す。対応して、言語モデルのサイズは、言語モデルによってカバーされる語彙サイズの増加に伴って、指数関数的に増大する。これらの2つの理由のために、音声認識システムの利用可能な語彙サイズは、無制限に増大され得ない。現在の技術の状態の下では、音声認識システムの用語集容量の上限は、10万をわずかに超える。用語集外の単語に関して、音声認識システムの認識精度は、大幅に低下することになる。

【0005】

さらに、通常の音声環境では、低い使用頻度の数百万の単語が存在する。例えば、これらは、短時間の間だけ関連性のある単語(例えば、テレビ番組または映画の名前)、特定の地理的領域にのみ関連性のある単語(例えば、地元のレストランの名前)、または、特定の専門分野にだけ現れる単語(例えば、専門用語または業界用語)、などであり得る。これらおよび他の理由のために、各単語が非常に低い統計的有意性を有する大量の低頻度語が存在するという現象が存在する。

10

【0006】

したがって、言語モデルのサイズを大幅に増加させることなく、または、その計算精度を損なうことなく、言語モデルの語彙範囲を拡大する方法の課題を解決するための緊急の必要性が存在する。

【発明の概要】

【発明が解決しようとする課題】

【0007】

本開示は、言語モデルの語彙範囲を拡大し、それによって、音声認識システムの認識精度を改善する、音声認識のための音響言語モデルをトレーニングするための方法およびデバイスを提案する。

20

【課題を解決するための手段】

【0008】

一態様では、音響言語モデルをトレーニングする方法は、1つまたは複数のプロセッサとメモリとを有するデバイスで、単語クラスラベルを含まない初期単語分割データを得るために、単語クラスラベルを含まない初期言語モデルを使用して、トレーニングコーパス内のトレーニングサンプルに対して単語分割を行うステップと、単語クラスラベルを含む第1の単語分割データを得るために、単語クラスラベルを含まない初期単語分割データに対して単語クラス置換を実行するステップと、単語クラスラベルを含む第1の言語モデルをトレーニングするために、単語クラスラベルを含む第1の単語分割データを使用するステップと、単語クラスラベルを含む第2の単語分割データを得るように、トレーニングコーパス内のトレーニングサンプルに対して単語分割を行うために、単語クラスラベルを含む第1の言語モデルを使用するステップと、1つまたは複数の所定の基準を満たす第2の単語分割データに従って、音響言語モデルをトレーニングするために、単語クラスラベルを含む第2の単語分割データを使用するステップとを含む。

30

【0009】

いくつかの実施形態では、音響言語モデルをトレーニングするためのシステムは、1つまたは複数のプロセッサと、命令が記憶されているメモリとを含み、命令は、1つまたは複数のプロセッサによって実行されると、プロセッサに上記の方法を実行させる。

40

【0010】

上記の技術的解決策によれば、いくつかの実施形態は、トレーニングコーパス内の言語サンプルに対して単語分割を実行するために、単語クラスラベルを含まない初期言語モデルを利用し、それによって、単語クラスラベルを含む初期単語分割データを得るステップと、初期単語分割データ内の少なくともいくつかの単語に対して単語クラス置換を実行し、それによって、単語クラスラベルを含む第1の単語分割データを得るステップと、第1の単語分割データを使用して、単語クラスラベルを含む第1の言語モデルをトレーニングし、トレーニングコーパス内のトレーニングサンプルに単語分割を実行するために、第1の言語モデルを利用し、それによって、単語クラスラベルを含む第2の単語分割データを得

50

るステップと、音響言語モデルをトレーニングするために、単語クラスラベルを含む第2の単語分割データを利用するステップとを含む。したがって、上記の解決策の実施形態は、単語クラスラベルを含む単語分割データに基づいて、音響言語モデルのトレーニングを実現する。単語クラスラベルは、トレーニング言語サンプル内のそのクラスに属するすべての単語(例えば、エンティティ名)を置換するために使用され得、それによって、言語モデル内の総語彙数を減少させる。所与の単語クラスの特定の単語(例えば、特定のエンティティ名)に関連付けられた確率を計算するとき、言語モデル内のその所与の単語クラスに関連付けられたパラメータが、計算に使用される。本解決策は、言語モデルの語彙範囲を拡大することができ、それによって、言語モデルの認識精度を改善する。加えて、本解決策は、音声認識システムの制限された用語集容量による、用語集外の単語に関する不十分な認識結果の問題に対処する。

10

【0011】

本発明の上述した特徴および利点、ならびにその追加の特徴および利点は、図面と併せたときに好ましい実施形態の詳細な説明の結果として、以下でより明確に理解されるであろう。

【図面の簡単な説明】

【0012】

【図1A】いくつかの実施形態による音響言語モデルトレーニング方法のフロー図である。

【図1B】いくつかの実施形態による音響言語モデルトレーニング方法のフローチャート図である。

20

【図1C】いくつかの実施形態による音響言語モデルトレーニング方法のフローチャート図である。

【図2】いくつかの実施形態による音響言語モデルトレーニングの全体の概要処理フローである。

【図3】いくつかの実施形態による音響言語モデルトレーニングのための装置のブロック図である。

【図4】いくつかの実施形態による音響言語モデルトレーニングのための装置のブロック図である。

【発明を実施するための形態】

30

【0013】

図面のいくつかの図を通して、同じ参照番号は、対応する部分を指す。

【0014】

ここで、実施形態を詳細に参照し、これらの実施形態の例は、添付図面に示される。以下の詳細な説明では、本明細書で提示される主題の完全な理解を提供するために、多数の特定の詳細が記載される。しかし、主題は、これらの特定の詳細なしに実施され得ることが、当業者には明らかであろう。他の例では、周知の方法、手順、構成要素、および回路は、実施形態の態様を不必要に曖昧にしないように、詳細には説明されていない。

【0015】

本開示の目的、技術的スキーム、および利点をより明確に理解するために、本開示は、添付図面と組み合わせて以下に詳細に説明される。

40

【0016】

いくつかの実施形態では、語彙の以下の特性の知識が考慮される。

【0017】

(1)多数の低頻度単語は、人、本、映画、場所、および他の名前付きエンティティの名前のような、エンティティ名である。

【0018】

(2)同じ単語クラスのエンティティ名のコンテキスト情報は、類似性を示し、例えば、同じ単語クラスの異なるエンティティ名が、同じテキストの文脈(例えば、文書中の周囲のテキスト)中に配置された場合、文章の表意文字の能力(ideographic ability)は、ほと

50

んど影響を受けない。例えば、「I like watching the cartoon 'Crayon Shin-Chan'」という文章、および「I like watching the cartoon 'Chibi Maruko Chan'」という文章は、かなり類似した表意文字の能力を有する。

【 0 0 1 9 】

上記の特性に従って、本開示の実施形態は、単語クラスによる単語分割に基づいて音響言語モデルをトレーニングするための戦略を提案する。

【 0 0 2 0 】

本明細書で説明するように、単語クラスは、それらのテキストのコンテキストの統計的特徴(例えば、すべての所与のテキストのコンテキスト中の単語の出現確率)に従って同様である単語(例えば、エンティティ名)を含むセットを指す。

10

【 0 0 2 1 】

本明細書で説明するように、単語分割は、連続的な言語サンプル(例えば、テキスト文字列)を一連の明確な意味の単位(例えば、単語)に分割するプロセスを指す。例えば、中国語では、漢字またはピンインを含むテキスト文字列は、単語間に自然の区切り文字を含まず、テキスト文字列内の意味の単位の間での分割は、明らかではない。したがって、テキスト文字列の意味を解釈するために、文字列は、各々がそれぞれの単語を表す一連のチャンクに分割される。

【 0 0 2 2 】

本明細書で説明するように、単語クラスラベリングまたは単語クラス置換は、言語サンプル(例えば、トレーニング言語サンプル)内の単語(例えば、エンティティ名)をその対応する単語クラスラベルで代用または置換するプロセスを指す。例えば、「I like watching the cartoon 'Crayon Shin-Chan'」という文字列は、単語クラス置換によって「I like watching the cartoon <cartoon name>」という別の文字列に変換され得、「Crayon Shin-Chan」という単語は、「<cartoon name>」というそのクラスラベルで置換される。単語クラス置換は、文章の複数の単語に対して実行され得る。単語クラス置換の結果は、より一般的な文章であり、単語クラス置換によって同じ一般的な文章をもたらすトレーニングサンプルに関連付けられた統計値は、一般的な文章に関連付けられた一般的な統計値に組み合わされ得る。一般的な文章が、言語モデルのトレーニングで使用されるとき、もたらされる言語モデルは、単語クラスラベルを含む言語モデルである。単語クラスラベルを含む一般的な文章を使用するトレーニングは、特定の低頻度単語(例えば、それらの対応するクラスラベルで置き換えられている単語)を含むトレーニングサンプルに関するデータスパースネス(data sparseness)の問題に対処する。

20

30

【 0 0 2 3 】

本明細書で説明するように、単語クラスラベルに基づく単語分割は、単語クラスラベルが候補分割結果の確率経路で表される、言語サンプルを分割する方法を指し、単語クラスが最高の経路に現れる場合、特定の単語(例えば、エンティティ名)の単語クラスラベルは、分割結果の対応する位置に現れる。例えば、複数の可能な分割経路の中で最も高い確率を有する経路が「[word i][word j][<word class p>][word k]」である場合、分割結果は、「[word i][word j][<word class p>][word k]」である。この分割結果は、「[word i][word j][string x][word k]」というフォームのすべての文章に対して有効であり、ここで[string x]は、<word class p>という単語クラス内の単語である。

40

【 0 0 2 4 】

本明細書で説明するように、言語モデルは、所与の過去の単語列に対して、例えばnグラム統計および平滑化アルゴリズムに基づいて、所与の過去の単語列に続く単語の出現確率を与える確率モデルを指す。

【 0 0 2 5 】

言語モデルは、しばしば、音声認識、機械翻訳、品詞タグ付け、構文解析、情報検索、などのような多くの自然言語処理の局面で使用される。単語および文章の長さは、任意であり得るので、(例えば、言語モデルトレーニング中のデータスパースネスの問題のため)トレーニングされた言語モデルによって十分にカバーされない、まれなまたは隠れた文字

50

列が存在する可能性がある。これは、コーパス内の文字列の確率の推定を困難にし、ほぼ平滑なNグラムモデルを使用する理由の1つでもある。

【0026】

本開示のいくつかの実施形態によれば、単語クラスラベルは、トレーニングコーパスのトレーニングサンプル(例えば、テキストサンプル)内の対応する単語クラスのすべての単語(例えば、エンティティ名)を置換するために使用される。したがって、同じ単語クラスの多数の単語(例えば、エンティティ名)が、言語モデル内で同じ単語クラスラベルとして表されるので、言語モデルの語彙サイズは、言語モデルの語彙範囲を犠牲にすることなく、低減され得る。同じ単語クラスの単語(例えば、エンティティ名)に関する確率を計算するとき、言語モデル内の単語クラスラベルのパラメータが使用される。トレーニングテキスト内に現れなかった単語(例えば、エンティティ名)に関して、この種類の単語(例えば、エンティティ名)を表す単語クラスラベルが言語モデル内に存在する限り、これらの隠れた単語(例えば、隠れたエンティティ名)に関連付けられた確率分布を近似するために、単語クラスラベルに関連付けられた確率分布は、考慮され得る。実際には、単語クラスラベルを含む言語モデルによって、(例えば、出現確率が計算され得る単語の総数によって測定される)言語モデルの語彙範囲は、言語モデルに実際に含まれる語彙サイズおよび単語クラス量よりもはるかに大きくすることができる。そのように、音声認識システムの制限された用語集容量によって引き起こされる、用語集の範囲を超える語彙に関する不十分な認識性能の問題は、対処され得る。

10

【0027】

図1は、いくつかの実施形態による音響言語モデルトレーニング方法のフローチャートである。

20

【0028】

図1に示すように、方法は、以下を含む。

【0029】

ステップ101:単語クラスラベルを含まない単語分割データを得るために、単語クラスラベルを含まない言語モデルを使用することによって、トレーニングコーパス内の言語サンプル(例えば、テキスト文字列)に対する単語分割を行う。

【0030】

ここで、単語クラスラベルを含まない言語モデルを使用することによってトレーニングコーパスの単語分割を行う前に、いくつかの好ましい実施形態では、トレーニングコーパスの正規化処理が実行される。正規化処理は、オプションで、(例えば、大文字と小文字との間の)大文字小文字の変換、(例えば、Unicodeと他の符号化との間の)変換の符号化、時間フォーマットの統一、などを含む。

30

【0031】

具体的には、様々なタイプのトークナイザ(例えば、純粋な単語を含み、単語クラスラベルを含まない言語サンプルのみでトレーニングされた従来の言語モデル)は、単語クラスラベルなしの単語分割結果を得るように、トレーニングコーパスに対する単語分割を行うために、初期言語モデルとして、すなわち、単語クラスラベルを含まない言語モデルとして使用され得る。

40

【0032】

いくつかの実施形態では、単語クラスラベルなしの言語モデルは、確率統計理論に基づく数学的モデルであり得、中国語の文章に関連付けられた確率分布を計算するためのものであり得る。言語モデルは、正しい文章(例えば、正しい単語の分割を有する文章)を出力する確率を、誤った文章を出力する確率よりも高くするように構成される。例えば、中国語の文章

【0033】

【数1】

「说明此处汉语语句的概率」

50

【 0 0 3 4 】

に関して、統計言語モデルでは、中国語の文章は、

【 0 0 3 5 】

【数 2】

「/説明/ /此处/ /汉语/ /语句/ /的/ /概率/」

【 0 0 3 6 】

のようなくつかの分離した単語に分解または分割され得る。

【 0 0 3 7 】

m (m は、自然数)の単語を含む中国語の文章、例えば、

10

【 0 0 3 8 】

【数 3】

$$S = w_1 w_2 \dots w_m \bar{S} = \bar{w}_1 \bar{w}_2 \dots \bar{w}_m$$

【 0 0 3 9 】

に関して、ベイズ理論によれば、この中国語の文章の確率(この文章が正しい出力である確率)は、1つよりも多くの単語を含む条件付き確率の算術積に分解され得、すなわち、

【 0 0 4 0 】

【数 4】

$$P(S) = P(w_1 w_2 \dots w_m) = \prod_{i=1}^m p(w_i / w_1 w_2 \dots w_{i-1})$$

20

【 0 0 4 1 】

となる。

【 0 0 4 2 】

上記の式では、 w_i は、中国語の文章 S 中に含まれる i 番目の単語であり、 $p(w_i / w_1 w_2 \dots w_{i-1})$ は、文章 $w_1 w_2 \dots w_m$ 中に現れる単語 w_i の条件付き確率である。

【 0 0 4 3 】

上記の式から、条件付き確率 $p(w_i / w_1 w_2 \dots w_{i-1})$ のパラメータ空間は、変数 i の増加に伴って指数関数的に増大するようになることがわかる。変数 i が大きいとき、利用可能なトレーニングコーパスの制限されたサイズでは、確率値 $p(w_i / w_1 w_2 \dots w_{i-1})$ は、正確に推定され得ない。

30

【 0 0 4 4 】

トレーニングコーパスは、統計的方法を使用して大量のトレーニングテキストから特定のカテゴリに従って編成および形成されている、順序付けられたテキストサンプルを含むセットである。トレーニングコーパスは、バッチ内のコンピュータによって処理され得る。したがって、現在の実際的な言語モデルのモデリング方法では、条件付き確率 $p(w_i / w_1 w_2 \dots w_{i-1})$ は、様々な程度に単純化される。これらのモデリング方法から出現するのは、標準的なNグラム言語モデルのモデリング方法である。

【 0 0 4 5 】

40

標準的なNグラム言語モデルは、現在最も一般的に使用される統計的言語モデルの1つである。いくつかの実施形態では、標準的なNグラム言語モデルは、中国語の文章を、マルコフ性を満たすマルコフ列とみなす。具体的には、標準的なNグラム言語モデルは、統計的言語モデルにおける条件付き確率 $p(w_i / w_1 w_2 \dots w_{i-1})$ に関する以下の基本的な仮説をする。

【 0 0 4 6 】

(1)制限された履歴仮説:現在の入力文章中の単語の条件付き確率は、中国語の文章全体の代わりに、単語の前の $n-1$ の単語にのみ関連し、ここで、 n は、予め設定された自然数である。

【 0 0 4 7 】

50

(2)時間均一性仮説:現在の単語の条件付き確率は、中国語の文章中のその位置に関連しない。

【0048】

上記2つの仮説に基づいて、標準的なNグラム言語モデルの下での文章に関する確率計算の式は、次のように単純化され得る。

【0049】

【数5】

$$P(S) = \prod_{i=1}^m p(w_i / w_{i-n+1} w_{i-n+2} \dots w_{i-1})$$

10

【0050】

このように、上記の2つの仮説に基づいて、標準的な統計的言語モデルにおける条件付き確率 $p(w_i / w_1 w_2 \dots w_{i-1})$ は、標準的なNグラム言語モデルにおける条件付き確率 $p(w_i / w_{i-n+1} w_{i-n+2} \dots w_{i-1})$ に単純化される。確率計算の後者の式では、現在の単語に関連する過去の単語の数は、標準的な統計的言語モデルでの変数 $i-1$ ではなく、定数 $n-1$ に固定される。したがって、言語モデル全体に関するパラメータ空間のサイズは減少し、Nグラム確率の値は、現在利用可能なトレーニングコーパスに基づいて正確に推定され得る。その結果、標準的なNグラム言語モデルは、合理的に実施され得る。

【0051】

標準的なNグラム言語モデルでは、条件付き確率 $p(w_i / w_{i-n+1} w_{i-n+2} \dots w_{i-1})$ の値の推定は、最尤推定の方法を採用し、推定式は、以下の通りである。

20

【0052】

【数6】

$$p(w_i / w_{i-n+1} w_{i-n+2} \dots w_{i-1}) = \frac{c(w_{i-n+1} \dots w_{i-1} w_i)}{c(w_{i-n+1} \dots w_{i-2} w_{i-1})}$$

【0053】

上記の式では、 $c(w_{i-n+1} \dots w_{i-1} w_i)$ は、標準的なNグラム言語モデルのトレーニングコーパスに出現する単語列 $w_{i-n+1} \dots w_{i-1} w_i$ (文章中の単語のサブ列)の頻度を表す。

【0054】

30

いくつかの実施形態では、単語分割は、単語クラスラベルを含まない単語分割データを得るために、標準的なNグラム言語モデルを使用して行われる。

【0055】

ステップ102:単語クラスラベルを含む第1の単語分割データを得るために、単語クラスラベルを含まない単語分割データに対して単語クラス置換を実行する。

【0056】

ここで、分類用語集は、事前に確立され得、この分類用語集内の語彙は、単語クラスに従って分類される。いくつかの実施形態では、単語クラスラベルを含む単語分割データ(例えば、 $w_1 / w_2 / \dots / w_i / \dots / w_m$)内の識別された単語 w_i に関して、同一の単語(例えば、 w_i)は、予め確立された分類用語集から検索され得、検索された単語のそれぞれの単語クラスラベル(例えば、 $\langle \text{label}_x \rangle$)は、単語分割データ内の単語に関するそれぞれの単語クラスラベルとして設定される(例えば、変更された単語分割データ $w_1 / w_2 / \dots / \langle \text{label}_x \rangle / \dots / w_m$ をもたらす)。

40

【0057】

【数7】

例えば、いくつかの実施形態では、分類用語集は、単語クラスラベル<家具>(家具を意味する)を含むと仮定する。用語集では、ワードクラスラベル<家具>に対応する単語は、窗户(「窓」を意味する)、桌子(「テーブル」を意味する)、椅子(「椅子」を意味する)、大门(「ドア」を意味する)などを含む。さらに、これらの家具の単語のための共通の単語クラスラベルは、<家具>である。トレーニングコーパスは、中国語の文章「我打开窗户」(「私は、窓を開ける」を意味する)を含むとき、文章は、「/我/打开/窗户/」として分割される。単語分割データ「/我/打开/窗户/」内の分割された単語を分類用語集中の既知の単語と比較することによって、トレーニングコーパス内の分割された単語「窗户」は、家具単語クラスに属し、分類用語集内の<家具>の単語クラスラベルを有すると判断され得る。したがって、分類データ内の分割された単語「窗户」は、分類用語集内で発見された単語クラスラベル<家具>で置換され、単語クラスラベルを含まない分割データ(例えば、「/我/打开/窗户/」)は、単語クラスラベルを含む単語分割データ(例えば、「/我/打开/<家具>/」)に変換される。

10

【0058】

【数8】

同様に、トレーニングコーパスが中国語の文章「我打开大门」(「私は、ドアを開ける」を意味する)を含むとき、文章は、「/我/打开/大门/」として分割される。単語分割データ「/我/打开/大门/」内の分割された単語を、分類用語集内の既知の単語と比較することによって、トレーニングコーパス内の分割された単語「大门」も、家具単語クラスに属し、分類用語集内の<家具>の単語クラスラベルを有すると判断され得る。したがって、単語分割データ「/我/打开/大门/」内の分割された単語「大门」は、単語クラスラベル<家具>で置換され、単語クラスラベルを含む単語分割データ(例えば、「/我/打开/<家具>/」)をもたらす。

20

【0059】

【数9】

同様に、トレーニングコーパスが中国語の文章「我踢开桌子」(「私は、机を蹴飛ばす」を意味する)を含むとき、文章は、「/我/踢开/桌子」として分割される。単語分割データ「/我/踢开/桌子」内の分割された単語を、分類用語集内の既知の単語と比較することによって、トレーニングコーパス内の分割された単語「桌子」も、家具単語クラスに属し、分類用語集内の<家具>の単語クラスラベルを有すると判断され得る。したがって、分割データ「/我/踢开/桌子」内の分割された単語「桌子」は、単語クラスラベル<家具>で置換され、単語クラスラベルを含む単語分割データ(例えば、「/我/踢开/<家具>/」)をもたらす。

30

【0060】

いくつかの実施形態では、単語クラスラベルを含まない単語分割データ内の分割された単語と同じである既知の単語が識別され、予め確立された分類用語集から検索された後、方法は、以下を含む。

【0061】

分類用語集から検索された既知の単語の単語長が、一致する単語長の予め設定されたしきい値を超えるかどうかを判断し、そうである場合、検索された既知の単語の単語クラスラベルを、単語クラスラベルを含まない単語分割データ内の分割された単語の単語クラスラベルとして設定する。例えば、いくつかの実施形態では、予め設定されたしきい値単語長は、2文字である。いくつかの実施形態では、予め設定されたしきい値単語長は、1文字である。いくつかの実施形態では、予め設定されたしきい値単語長は、3文字である。分類用語集から検索された既知の単語の単語長が、一致する単語長の予め設定されたしきい値を超えない場合、検索された既知の単語の単語クラスラベルは、単語分割データ内の分割された単語を置換するために使用されない。いくつかの実施形態では、2つ以上の連続する分割された単語は、分類用語集内の十分な長さの一致する既知の単語を検索するとき、単一の分割された単語として結合され、一致が見つかった場合、検索された既知の単

40

50

語の単語クラスラベルは、単語分割データ内で、2つ以上の連続する分割された単語を全体として置換するために使用される。いくつかの実施形態では、検索された単語の単語長をチェックする代わりに、分割された単語の単語長は、分類用語集を検索するために使用される前にチェックされ、予め設定されたしきい値長よりも長い分割された単語のみが、分類用語集を検索するために使用される。

【 0 0 6 2 】

ステップ103:単語クラスラベルを含む言語モデルをトレーニングするために、単語クラスラベルを含む第1の単語分割データを使用し、単語クラスラベルを含む第2の単語分割データを得るように、トレーニングコーパスの単語分割を行うために、単語クラスラベルを含む言語モデルを使用する。

10

【 0 0 6 3 】

いくつかの実施形態では、単語クラスラベルを含む第1の単語分割データは、少なくとも1つの分割された単語がその対応する単語クラスラベルに置換されている分割されたトレーニングサンプル(例えば、中国語の文章)を含む。いくつかの実施形態では、第1の単語分割データ内のいくつかの分割されたトレーニングサンプルは、1つまたは複数の元の分割された単語の中に1つまたは複数の単語クラスラベルを含むことができる。いくつかの実施形態では、ステップ102での単語クラス置換は、いくつかの異なる段階で、1つまたは複数の所定の基準に従って実行される。いくつかの実施形態では、単語クラス置換プロセスの各段階では、特定の単語(例えば、特定の長さ要件を満たす単語、特定のしきい値を上回る分割スコアを有する単語、特定の単語クラスの単語、など)のみが、それらの対応する単語クラスラベルによって置換される。

20

【 0 0 6 4 】

いくつかの実施形態では、単語クラスラベルを含む第1の単語分割データは、それぞれのトレーニングサンプル内の各単語クラスラベルがトレーニングサンプル内の単語と同様に処理されるように、Nグラム言語モデルのためのトレーニングサンプルとして使用される。元のトレーニングコーパスのトレーニングサンプル内の多くの単語は、ここでは、それらの対応する単語クラスラベルによって置換されるので、トレーニングコーパス内の固有のトレーニングサンプルの数は減少し、置換された単語の使用頻度は、ここでは、それらのそれぞれの単語クラスラベルの使用頻度に寄与する。したがって、単語クラスラベルを含む第1の単語分割データでトレーニングされて得られる言語モデルは、ステップ101で使用された初期の言語モデルよりも小さく、よりよい精度および語彙範囲を有する。いくつかの実施形態では、言語モデルをトレーニングする様々な方法が使用され得る。

30

【 0 0 6 5 】

いくつかの実施形態では、言語モデルが、単語クラスラベルを含む第1の単語分割データを使用してトレーニングされた後、言語モデルは、単語クラスラベルを含む文章を分割するために使用され得る。いくつかの実施形態では、トレーニングコーパス内の少なくともいくつかの単語が、分類用語集内で見つかったそれらの対応する単語クラスラベルで置換されるように、単語クラス置換が、元のトレーニングコーパス内のトレーニングサンプルに対して実行される。いくつかの実施形態では、特定の単語を、ステップ102で使用されたそれらの対応する単語クラスラベルで置換するかどうかを判断するために使用される同じ基準は、ここで、どの単語がそれらの単語クラスラベルで置換されるべきかを判断するために使用される。いくつかの実施形態では、単語クラス置換は、このステップで単語クラスラベルを含まない元のトレーニングコーパスのトレーニングサンプルに対して実行される必要はなく、代わりに、第1の単語分割データ(すなわち、ステップ102での単語クラス置換プロセスによっていくつかの分割された単語がそれらの対応する単語クラスラベルで置換されている、分割されたトレーニングサンプル)は、単語クラスラベルを含む第2の単語分割データを得るために、テスト入力として、単語クラスラベルを含む言語モデルに提供される。

40

【 0 0 6 6 】

ステップ104:音響言語モデルを得るために、単語クラスラベルを含む第2の単語分割デ

50

ータを使用する。

【 0 0 6 7 】

ステップ103で得られた単語クラスラベルを含む言語モデルは、単語クラスラベルを含む第1の分割データの統計的要約であるので、単語クラスラベルを含む第2の単語分割データは、単語クラスラベルを含む第1の分割データと異なる可能性が高い。ステップ103で得られた言語モデルがますます正確になる場合、単語クラスラベルを含む第2の単語分割データは、単語クラスラベルを含む第1の単語分割データにますます近似する可能性が高い。下記で説明するように、単語クラスラベルを含む言語モデルを取得し、単語クラスラベルを含む新たな分割データを得るために、トレーニングコーパスに対して分割化を実行し、新たな言語モデルを得るために、単語クラスラベルを含む新たな分割データで、単語クラスラベルを含む言語モデルをトレーニングするプロセスは、所定の収束条件が満たされるまで、1回または複数回の反復のために繰り返され得る。いくつかの実施形態では、所定の収束条件は、新たな言語モデルを使用して得られた分割データが、新たな言語モデルをトレーニングするために使用される分割データと十分に類似していることである。いくつかの実施形態では、新たな言語モデルを使用して得られた分割データと、新たな言語モデルをトレーニングするために使用される分割データとの間の類似性を測定するための基準は、単語分割データの2つのセット内の単語クラスラベルが同じ位置を有するかどうかである。

10

【 0 0 6 8 】

いくつかの実施形態では、音響言語モデルが単語クラスラベルを含む第2の単語分割データに基づいて得られた後、音響言語モデルは、入力音声を受信することと、音響言語モデルの使用によって入力音声を目的とした音声認識処理を実行することを含む、入力を受信し、認識結果を提供する準備ができています。

20

【 0 0 6 9 】

いくつかの実施形態では、単語クラスラベルを含む第2の単語分割データが得られた後、方法は、さらに以下を含む。

【 0 0 7 0 】

単語クラスラベルを含む第1の単語分割データが、単語クラスラベルを含む第2の単語分割データと同じ単語クラスラベルの置換部分を有するかどうかを判断し、そうである場合、音響言語モデルを得るために、前記単語クラスラベルを含む第2の単語分割データを使用し、そうでない場合、単語クラスラベルを含む第2の単語分割データに対して単語クラス置換を実行する。

30

【 0 0 7 1 】

本開示のいくつかの好ましい実施形態では、プロセスフローは、反復的であり、すなわち、単語クラスラベルを含む言語モデルは、反復的な最適化プロセスを通過する。いくつかの実施形態では、プロセスは、以下を含む。

【 0 0 7 2 】

(1)第1に、トレーニングコーパスの正規化処理を実行する。

【 0 0 7 3 】

(2)次いで、トークナイザは、トレーニングコーパス内のトレーニングサンプルに単語分割を実行するために、初期言語モデル(例えば、単語クラスラベルを含まない言語モデル)を使用し、単語分割結果(例えば、単語クラスラベルを含まない単語分割データ)を得る。

40

【 0 0 7 4 】

(3)単語クラスラベルを含む単語分割結果を得るために、条件トリガールールのセットに基づいて、単語分割結果に対して単語クラス置換を行う。このとき、正解率を保証するために、条件トリガールールは、予め設定された確実性しきい値を満たすエンティティの名前の単語クラス置換を実行することができる。現在の反復で得られた単語クラスラベルを含む単語分割結果が、前の反復で得られた単語クラスラベルを含む単語分割結果と同じである場合、反復プロセスは、終了され得、後処理が開始され得る。

50

【 0 0 7 5 】

(4)言語モデルをトレーニングするために、単語クラスラベルを含む単語分割データを使用し、単語クラスラベルを含む言語モデルを得る。

【 0 0 7 6 】

(5)トークナイザは、ステップ(1)で得られたトレーニングコーパス内の単語の単語クラスラベルに従ってトレーニングコーパスの単語分割を実行し、単語クラスラベルを含む単語分割データを取得し、次の反復のステップ(3)を実行しに戻るために、単語クラスラベルを含む言語モデルを使用する。このとき、統計的言語モデルの使用は、トリガールールが満たさない単語クラス置換を作成する可能性がある一方で、かつ、単語分割結果の変化のため、トリガールールを満たすことができる元の位置は、もはや有効ではない可能性がある。そのように、ステップ(3)～(5)の反復プロセスを継続するかどうかの決定は、言語モデルが十分に正確になり、2つの反復の分割結果が収束するまで、反復プロセスの継続をもたらすことができる。

10

【 0 0 7 7 】

その後、音声認識に適した後処理が、単語クラスラベルを含む単語分割結果に対して実行され得、単語分割結果は、単語クラスラベルを含む音響言語モデルをトレーニングするために使用され得る。

【 0 0 7 8 】

図1B～図1Cは、いくつかの実施形態による音響言語モデルをトレーニングする上記方法のフローチャートである。

20

【 0 0 7 9 】

いくつかの実施形態では、1つまたは複数のプロセッサと、メモリとを有するデバイスで、単語クラスラベルを含まない初期単語分割データを得るために、単語クラスラベルを含まない初期言語モデルを使用して、トレーニングコーパス内のトレーニングサンプルに対して単語分割が行われる(112)。

【 0 0 8 0 】

次いで、単語クラスラベルを含む第1の単語分割データを得るために、単語クラスラベルを含まない初期単語分割データに対して単語クラス置換が実行される(114)。

【 0 0 8 1 】

いくつかの実施形態では、単語クラスラベルを含む第1の単語分割データを得るために、単語クラスラベルを含まない初期単語分割データに対して単語クラス置換を実行するとき、単語クラスラベルを含まない初期単語分割データ内の1つまたは複数のそれぞれの単語に関するそれぞれの単語クラスラベルは、分類用語集内で識別される(116)。単語クラスラベルを含まない初期単語分割データ内の1つまたは複数のそれぞれの単語は、次いで、単語クラスラベルを含む第1の単語分割データを得るために、識別されたそれぞれの単語クラスラベルで置換される(118)。

30

【 0 0 8 2 】

いくつかの実施形態では、単語クラスラベルを含む第1の単語分割データは、単語クラスラベルを含む第1の言語モデルをトレーニングするために使用される(120)。

【 0 0 8 3 】

いくつかの実施形態では、トレーニングの後、単語クラスラベルを含む第1の言語モデルは、単語クラスラベルを含む第2の単語分割データを得るために、トレーニングコーパス内のトレーニングサンプルに対して単語分割を行うために使用される(122)。

40

【 0 0 8 4 】

いくつかの実施形態では、単語クラスラベルを含む第2の単語分割データを得るように、トレーニングコーパス内のトレーニングサンプルに対して単語分割を行うために、単語クラスラベルを含む第1の言語モデルを使用するとき、トレーニングコーパス内のトレーニングサンプル内の1つまたは複数のそれぞれの単語に関するそれぞれの単語クラスラベルが、分類用語集内で識別され(124)、単語クラスラベルを含む新たなトレーニングサンプルを得るために、トレーニングサンプル内の1つまたは複数のそれぞれの単語は、識

50

別されたそれぞれの単語クラスラベルで置換され(126)、単語クラスラベルを含む第2の単語分割データを得るために、新たなトレーニングサンプルに対する単語分割が、単語クラスラベルを含む第1の言語モデルを使用して行われる(128)。

【0085】

いくつかの実施形態では、1つまたは複数の所定の条件を満たす第2の単語分割データに従って、単語クラスラベルを含む第2の単語分割データは、音響言語モデルをトレーニングするために使用される(130)。

【0086】

いくつかの実施形態では、単語クラスラベルを含む第2の単語分割データを得た後、第1および第2の単語分割データ内の対応するトレーニングサンプルの分割結果は、比較される(132)。第1の単語分割データが第2の単語分割データと一致するという判断に従って、第2の単語分割データは、音響言語モデルのトレーニングで使用するために承認される(134)。いくつかの実施形態では、単語クラスラベルを含む第2の単語分割データを得た後、第1の単語分割データが第2の単語分割データと一致しないという判断に従って、第1の言語モデルは、例えば、第2の単語分割データを使用して再トレーニングされる(136)。

【0087】

いくつかの実施形態では、第1の言語モデルが再トレーニングされた後、改訂された第2の単語分割データを得るために、第2のトレーニングサンプルに対する単語分割が、単語クラスラベルを含む第1の言語モデルを使用して繰り返される。いくつかの実施形態では、改定された第2の単語分割データが第2の単語分割データと一致するという判断に従って、音響言語モデルのトレーニングで使用するために、改訂された第2の単語分割データを承認する。

【0088】

いくつかの実施形態では、単語クラスラベルを含む第2の単語分割データを得た後、第1の単語分割データが第2の単語分割データと一致しないという判断に従って、単語クラスラベルを含む第1の単語分割データは、以前とは異なる単語クラス置換で改訂される。第1の言語モデルは、改訂された第1の言語モデルを得るために、改訂された第1の単語分割データを使用して再トレーニングされる。改訂された第1の言語モデルは、改訂された第2の単語分割データを得るために、トレーニングコーパスに対する分割を実行するために使用される。いくつかの実施形態では、改定された第2の単語分割データが改定された第1の単語分割データと一致するという判断に従って、改定された第2の単語分割データは、音響言語モデルのトレーニングで使用するために承認される。

【0089】

いくつかの実施形態では、第1の単語分割データが第2の単語分割データと一致すると判断することは、さらに、第1の単語分割データ内のそれぞれの単語クラスラベル置換が、第2の単語分割データ内のそれぞれの単語クラスラベル置換と同一であるという判断を含む。

【0090】

図2は、本開示の実施形態に基づく音響言語モデルトレーニング方法の全体的な概略図である。図1A~図1Cに示すように、いくつかの実施形態では、図2に示すプロセスは、(1) トレーニングコーパスの前処理、それに続く、(2)(例えば、単語クラスラベルを含む言語モデルを使用する)単語クラスに基づく単語分割、それに続く、(3)単語クラス置換(例えば、追加のまたは改訂された単語クラス置換)、それに続く、(4)新たなトレーニングデータ(例えば、追加のまたは改訂された単語クラス置換で変更された、前の言語モデルから得られた単語分割データ)を使用する改善された言語モデルのトレーニングを含み、(5)(2)~(4)のプロセスは、各追加の反復で言語モデルを改善するために、複数の反復を繰り返すことができ、所定の基準のセット(例えば、現在の言語モデルをトレーニングするために使用された単語分割データ、および、現在の言語モデルから得られた単語分割データ)が互いに一致すると、次いで、(2)~(4)のサイクルは、中断され得、(6)言語モデルから最後に得られた分割データは、後処理され、(7)音声認識で使用される言語モデル(音響言

10

20

30

40

50

語モデル)のトレーニングで使用される。

【0091】

図2に示すように、前述の特定の分析によれば、いくつかの実施形態では、単語分割段階に関して、最適化は、パターン単語、辞書単語にマッチングするだけでなく、完全な分割のプロセス中に、単語クラスリソース(例えば、分類用語集)に基づいて、一致した単語に対する単語クラスラベルの置換も実行する。単語分割段階に関して、隠れマルコフモデルの確率経路に基づく最適化選択のプロセスの間、言語モデルにおける単語クラスパラメータは、単語クラスの確率を計算するために使用され得る。最適経路が単語クラスラベルを含む場合、単語クラスラベルは、単語分割結果として直接出力されることになる。リソースに基づく単語クラス置換の段階では、単語クラス置換のデータは、自動認識に由来するリストにない単語の代わりに、信頼できるリソースデータでなければならない。それは、語彙単語クラスの置換方法が、トリガールールおよび統計的言語モデルを含む2つの方法の組み合わせであることを示す。

10

【0092】

本開示の実施形態では、トリガールールは、補償された部分に対する単語クラス置換のみを実行し、単語クラスに基づく言語モデルの反復を通して、単語クラス置換の範囲の割合は、徐々に改善され得る。本開示の実施形態では、分割リソースとして機能する語彙の分類は、単語分割のマッチング処理に関与し、単語クラスの確率パラメータを使用することによって、最適な確率経路の計算に関わり、勝った単語クラス語彙がラベルの分類の形態で単語分割結果に示されることを可能にする。

20

【0093】

前述の具体的な分析に基づいて、本開示の実施形態は、一種の音響言語モデルトレーニングデバイスも提唱する。

【0094】

図3は、いくつかの実施形態による音響言語モデルトレーニング方法のデバイス構成図である。

【0095】

デバイスは、単語分割ユニット301と、単語クラス置換ユニット302と、言語モデルトレーニングユニット303と、音響言語モデル取得ユニット304とを含む。

【0096】

単語分割ユニット301は、単語クラスラベルなしの単語分割データを取得するために、トレーニングコーパスに対して単語分割を実行するために、単語クラスラベルなしの言語モデルを使用する。

30

【0097】

単語クラス置換ユニット302は、単語クラスラベルを有する第1の単語分割データを得るために、単語クラスラベルなしの単語分割データに対して単語クラス置換を実行するために使用される。

【0098】

言語モデルトレーニングユニット303は、単語クラスラベルを有する第1の単語分割データを使用して、単語クラスラベルを含む言語モデルをトレーニングし、単語クラスラベルを含む第2の単語分割データを得るように、トレーニングコーパスの単語分割を行うために、単語クラスラベルを含む言語モデルを使用するために使用される。

40

【0099】

音響言語モデル取得ユニット304は、音響言語モデルを得るために、単語クラスラベルを含む第2の単語分割データを使用する。

【0100】

いくつかの実施形態では、単語クラス置換ユニット302は、予め確立された分類用語集から、単語クラスラベルなしの単語分割データ内の分割された単語と同一の既知の単語を取得し、取得された既知の単語のそれぞれの単語クラスラベルを、単語クラスラベルなしの単語分割データ内の分割された単語のそれぞれの単語クラスラベルとして設定するため

50

に使用される。

【0101】

いくつかの実施形態では、言語モデルトレーニングユニット303は、さらに、単語クラスラベルを有する第2の単語分割データが得られた後、単語クラスラベルを有する第1の単語分割データが、単語クラスラベルを有する第2の単語分割データと単語クラスラベルの同じ置換位置を有するかどうかを判断するために使用される。そうである場合、音響言語モデル取得ユニットは、有効にされ、そうでない場合、単語クラスラベルを有する第2の単語分割データに対して単語クラス置換を実行するために、単語クラス置換ユニットが有効にされる。

【0102】

いくつかの実施形態では、単語クラス置換ユニット302は、さらに、単語クラスラベルなしの単語分割データと同じ語彙が、予め確立された分類用語集から取得された後、分類用語集から取得された語彙の長さが、一致する単語長の予め設定されたしきい値を超えるかどうかを判断するために使用される。そうである場合、取得された語彙の単語クラスラベルは、単語クラスラベルなしの単語分割データ内の取得された単語の単語クラスラベルとして設定される。

【0103】

一実施形態では、単語クラス置換ユニット302は、さらに、単語クラスラベルなしの単語分割データ内の分割された単語と同一の既知の単語が、予め確立された分類用語集から取得された後、分類用語集から取得された語彙の出現頻度が、単語頻度の予め設定されたしきい値を超えるかどうかを判断するために使用される。そうである場合、取得された既知の単語の単語クラスラベルは、単語クラスラベルなしの単語分割データ内の分割された単語の単語クラスラベルとして設定される。

【0104】

好ましくは、装置は、さらに、音声認識ユニット305を含む。

【0105】

音声認識ユニット305は、入力音声を受信し、音響言語モデルを使用することによって、入力音声を目的とした音声認識処理を実行するために使用される。

【0106】

好ましくは、このデバイスは、さらに、正規化ユニット306を含む。

【0107】

正規化ユニット306は、単語クラスラベルなしの言語モデルを使用することによってトレーニングコーパスに対する単語分割を行う前に、トレーニングコーパスに対する正規化処理を行うために使用される。

【0108】

図3に示すデバイスを、様々なネットワークのハードウェアエンティティに組み込むことは、許容できる。例えば、音響言語モデルトレーニングデバイスは、フィーチャーフォン、スマートフォン、パームトップコンピュータ、パーソナルコンピュータ(PC)、タブレットコンピュータ、または携帯情報端末(PDA)、などのようなデバイスに組み込まれ得る。

【0109】

実際には、本開示の実施形態で述べた音響言語モデルトレーニングデバイスを具体的に実施するための様々な形態が存在する。例えば、特定の仕様に従うアプリケーションインターフェースを介して、音響言語モデルトレーニングデバイスは、ブラウザにインストールされたプラグインとして記述され得、同様に、ユーザ自身によってダウンロードするために使用されるアプリケーションとしてパッケージ化され得る。プラグインとして記述される場合、ocx、dll、cab、などを含む様々なプラグインフォーマットで実施され得る。Flashプラグイン、RealPlayerプラグイン、MMSプラグイン、Mistakeプラグイン、ActiveXプラグイン、などを含む特定の技術を介して、本開示の実施形態で述べた音響言語モデルトレーニングデバイスを実現することも許容され得る。

10

20

30

40

50

【0110】

命令または命令セットの記憶方法によって、本開示の実施形態で述べた音響言語モデルトレーニング方法は、様々な記憶媒体に記憶され得る。これらの記憶媒体は、限定はしないが、フロッピー(登録商標)ディスク、CD、DVD、ハードディスク、NANDフラッシュ、USBフラッシュディスク、CFカード、SDカード、MMCカード、SMカード、メモリースティック(Memory Stick)、xDカード、などを含む。

【0111】

加えて、本開示の実施形態で述べた音響言語モデルトレーニング方法は、また、NANDフラッシュに基づく記憶媒体、例えば、USBフラッシュディスク、CFカード、SDカード、SDHCカード、MMCカード、SMカード、メモリースティック、xDカード、などに適用され得る。

10

【0112】

要約すると、本開示のいくつかの実施形態は、単語クラスラベルなしの単語分割データを得るために、単語クラスラベルなしの言語モデルを使用することによって、トレーニングコーパスに対して単語分割を行うステップと、単語クラスラベルを有する第1の単語分割データを得るために、単語クラスラベルなしの単語分割データに対する単語クラス置換を実行するステップと、単語クラスラベルを有する第1の単語分割データを使用して、単語クラスラベルを有する言語モデルをトレーニングし、単語クラスラベルを有する第2の単語分割データを得るように、トレーニングコーパスの単語分割を行うために、単語クラスラベルを有する言語モデルを使用するステップと、音響言語モデルを得るために、単語クラスラベルを有する第2の単語分割データを使用するステップとを含む。したがって、単語クラスの単語分割に基づく音響言語モデルトレーニングは、本開示の実施形態の適用後に、果たされ得る。本開示の実施形態は、言語モデルトレーニングコーパス内の同じクラスのエンティティ名のすべてを置換するために、単語クラスラベルを使用することができ、その結果、言語モデル内の語彙の数を減少させる。加えて、同じ単語クラスのエンティティ名に関連付けられた確率を計算するとき、言語モデル内の単語クラスラベルに対応する同じパラメータが使用され得る。本開示の実施形態は、言語モデルの語彙範囲を拡大し、したがって、音声認識システムの認識精度を改善する。

20

【0113】

加えて、本開示の実施形態は、音声認識システムの制限された用語集容量によって引き起こされる語彙以外の単語に関する不十分な認識結果の問題に対処する。

30

【0114】

図4は、いくつかの実施形態による、本開示を実施するシステム400のブロック図である。いくつかの実施形態では、システム400は、上記で説明した、トレーニング、分割、および認識処理に関与する複数の処理装置のうちの1つである。いくつかの実施形態では、システム400は、上記で説明した機能およびプロセスを実行するための複数の処理ユニットを有する単一の機械を表す。

【0115】

図4に示すように、システム400は、1つまたは複数の処理ユニット(または「プロセッサ」)402と、メモリ404と、入力/出力(I/O)インターフェース406と、ネットワーク通信インターフェース408とを含む。これらの構成要素は、1つまたは複数の通信バスまたは信号ライン410を介して互いに通信する。いくつかの実施形態では、メモリ404、または、メモリ404のコンピュータ可読記憶媒体は、オペレーティングシステム412、I/Oモジュール414、通信モジュール416、および動作制御モジュール418のすべてまたはサブセットを含む、プログラム、モジュール、命令、およびデータ構造を記憶する。1つまたは複数のプロセッサ402は、メモリ404に結合され、これらのプログラム、モジュール、および命令を実行するように動作可能であり、データ構造から読み出す/データ構造に書き込む。

40

【0116】

いくつかの実施形態では、処理ユニット402は、シングルコアまたはマルチコアマイクロプロセッサのような、1つまたは複数のプロセッサ、またはマイクロプロセッサを含む。いくつかの実施形態では、処理ユニット402は、1つまたは複数の汎用プロセッサを含む

50

。いくつかの実施形態では、処理ユニット402は、1つまたは複数の専用プロセッサを含む。いくつかの実施形態では、処理ユニット402は、1つまたは複数のパーソナルコンピュータ、モバイルデバイス、ハンドヘルドコンピュータ、タブレットコンピュータ、または、1つもしくは複数の処理ユニットを含み、様々なオペレーティングシステム上で動作する多様なハードウェアプラットフォームのうちの1つを含む。

【0117】

いくつかの実施形態では、メモリ404は、DRAM、SRAM、DDR RAM、または他のランダムアクセス固体メモリデバイスのような高速ランダムアクセスメモリを含む。いくつかの実施形態では、メモリ404は、1つまたは複数の磁気ディスク記憶デバイス、光ディスク記憶デバイス、フラッシュメモリデバイス、または他の不揮発性固体記憶デバイスのような不揮発性メモリを含む。いくつかの実施形態では、メモリ404は、処理ユニット402から遠隔に配置された1つまたは複数の記憶デバイスを含む。メモリ404、または、代替的には、メモリ404内の不揮発性メモリデバイスは、不揮発性コンピュータ可読記憶媒体を備える。

10

【0118】

いくつかの実施形態では、I/Oインターフェース406は、ディスプレイ、キーボード、タッチスクリーン、スピーカ、およびマイクロホンのような入力/出力デバイスを、システム400のI/Oモジュール414に結合する。I/Oインターフェース406は、I/Oモジュール414と協働して、ユーザ入力(例えば、音声入力、キーボード入力、タッチ入力、など)を受信し、それらを適宜処理する。I/Oインターフェース406およびユーザインターフェースモジュール414は、また、システム400で実施される様々なプログラム命令に従って、出力(例えば、音、画像、テキスト、など)をユーザに提示する。

20

【0119】

いくつかの実施形態では、ネットワーク通信インターフェース408は、ワイヤード通信ポートならびに/またはワイヤレス送信および受信回路網を含む。ワイヤード通信ポートは、1つまたは複数のワイヤードインターフェース、例えば、イーサネット(登録商標)、ユニバーサルシリアルバス(USB)、FIREWIRE、などを介して、通信信号を送信および受信する。ワイヤレス回路網は、RF信号および/または光信号を、通信ネットワークおよび他の通信デバイスとの間で受信および送信する。ワイヤレス通信は、GSM(登録商標)、EDGE、CDMA、TDMA、Bluetooth(登録商標)、Wi-Fi、VoIP、Wi-MAX、または他の適切な通信プロトコルのような、複数の通信規格、プロトコル、および技術のいずれかを使用することができる。ネットワーク通信インターフェース408は、システム400と、インターネット、イントラネット、ならびに/または、セルラ電話ネットワーク、ワイヤレスローカルエリアネットワーク(LAN)、および/もしくはメトロポリタンエリアネットワーク(MAN)のようなワイヤレスネットワークのようなネットワーク、ならびに他のデバイスとの間の通信を可能にする。通信モジュール416は、ネットワーク通信インターフェース408を介する、システム400と他のデバイス(例えば、並列トレーニングおよび/または復号化プロセスに關与する他のデバイス)との間の通信を容易にする。

30

【0120】

いくつかの実施形態では、オペレーティングシステム402(例えば、Darwin、RTXC、LINUX(登録商標)、UNIX(登録商標)、OS X、ウィンドウズ(登録商標)、または、VxWorksのような組み込みオペレーティングシステム)は、一般的なシステムタスク(例えば、メモリ管理、記憶デバイス制御、電力管理、など)を制御および管理するための様々なソフトウェア構成要素および/またはドライバを含み、様々なハードウェア、ファームウェア、およびソフトウェア構成要素間の通信を容易にする。

40

【0121】

図4に示すように、システム400は、メモリ404に動作制御モジュール418を記憶する。いくつかの実施形態では、動作制御モジュール418は、さらに、以下のサブモジュール、すなわち、前処理モジュール420、言語モデルトレーニングモジュール422、単語クラス置換モジュール424、単語分割モジュール426、後処理モジュール428、および音声認識モジュール430、またはそのサブセットもしくはスーパーセットを含む。加えて、これらのモジ

50

ジュールの各々は、動作制御モジュール418の1つまたは複数の以下のデータ構造およびデータ源、すなわち、トレーニングコーパス432、分類用語集434、単語クラスラベルなしの初期言語モデル436、および、単語クラスラベルを有する言語モデル438、または、そのサブセットもしくはスーパーセットへのアクセスを有する。いくつかの実施形態では、動作制御モジュール418は、オプションで、本明細書に記載の他の関連する機能を提供するために、1つまたは複数の他のモジュールを含む。動作制御モジュール418のサブモジュールおよびデータ構造の構造、機能、および相互作用の詳細は、図1A~図3および付随する説明に関連して提供される。

【0122】

特定の実施形態を上記で説明したが、本発明をこれらの特定の実施形態に限定することは、意図されないことが理解されるであろう。それどころか、本発明は、添付の特許請求の範囲の要旨および範囲内にある代替態様、修正態様、および等価態様を含む。多数の特定の詳細が、本明細書で提示される主題の完全な理解を提供するために記載されている。しかし、主題は、これらの特定の詳細なしに実施され得ることが、当業者には明らかであろう。他の例では、周知の方法、手順、構成要素、および回路は、実施形態の態様を不必要に曖昧にしないように、詳細には説明されていない。

【符号の説明】

【0123】

- | | | |
|-----|-------------------|----|
| 301 | 単語分割ユニット | |
| 302 | 単語クラス置換ユニット | 20 |
| 303 | 言語モデルトレーニングユニット | |
| 304 | 音響言語モデル取得ユニット | |
| 305 | 音声認識ユニット | |
| 306 | 正規化ユニット | |
| 400 | システム | |
| 402 | プロセッサ | |
| 404 | メモリ | |
| 406 | I/Oインターフェース | |
| 408 | ネットワーク通信インターフェース | |
| 410 | 通信バス、信号ライン | 30 |
| 412 | オペレーティングシステム | |
| 414 | I/Oモジュール | |
| 416 | 通信モジュール | |
| 418 | 動作制御モジュール | |
| 420 | 前処理モジュール | |
| 422 | 言語モデルトレーニングモジュール | |
| 424 | 単語クラス置換モジュール | |
| 426 | 単語分割モジュール | |
| 428 | 後処理モジュール | |
| 430 | 音声認識モジュール | 40 |
| 432 | トレーニングコーパス | |
| 434 | 分類用語集 | |
| 436 | 初期言語モデル | |
| 438 | 単語クラスラベルを有する言語モデル | |

【要約】

音響言語モデルをトレーニングする方法およびデバイスは、単語クラスラベルを含まない初期単語分割データを得るために、単語クラスラベルを含まない初期言語モデルを使用して、トレーニングコーパス内のトレーニングサンプルに対して単語分割を行うステップと、単語クラスラベルを含む第1の単語分割データを得るために、単語クラスラベルを含まない初期単語分割データに対して単語クラス置換を実行するステップと、単語クラスラ

10

20

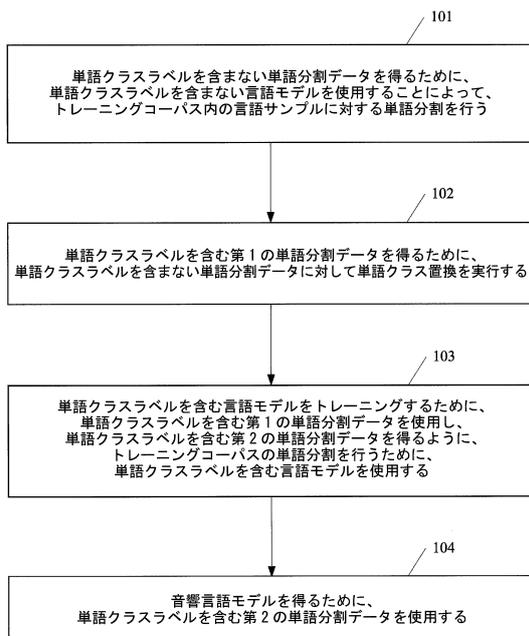
30

40

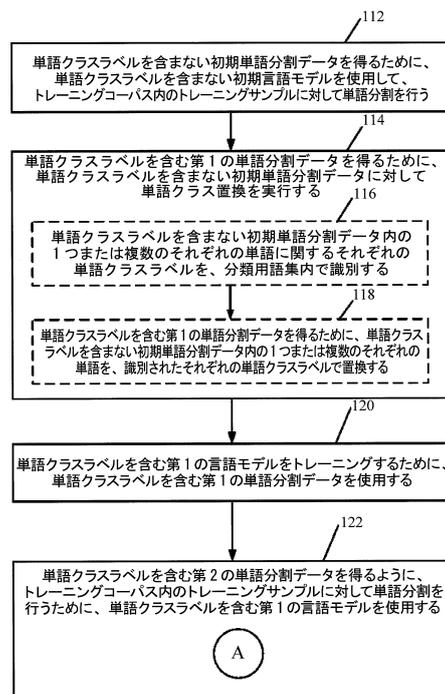
50

ベルを含む第1の言語モデルをトレーニングするために、単語クラスラベルを含む第1の単語分割データを使用するステップと、単語クラスラベルを含む第2の単語分割データを得るように、トレーニングコーパス内のトレーニングサンプルに対して単語分割を行うために、単語クラスラベルを含む第1の言語モデルを使用するステップと、1つまたは複数の所定の基準を満たす第2の単語分割データに従って、音響言語モデルをトレーニングするために、単語クラスラベルを含む第2の単語分割データを使用するステップとを含む。

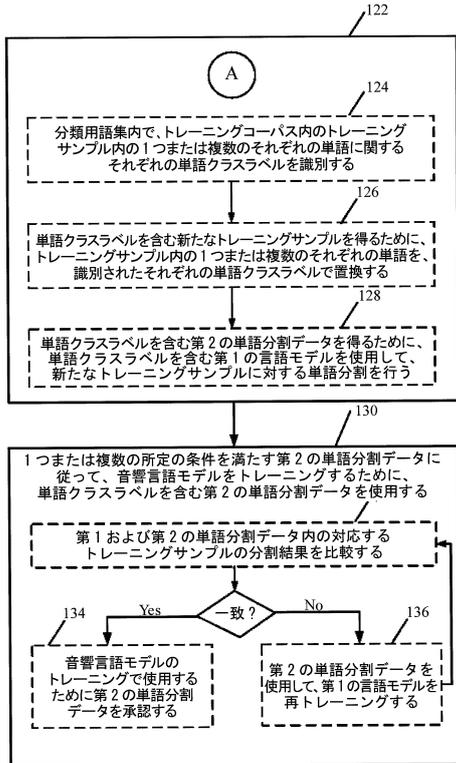
【図 1 A】



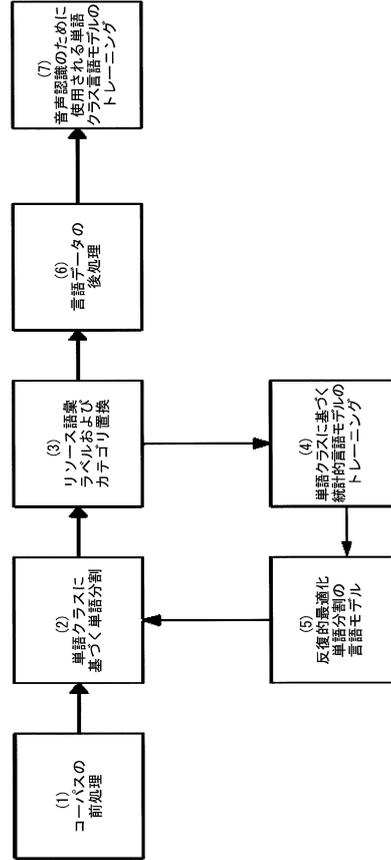
【図 1 B】



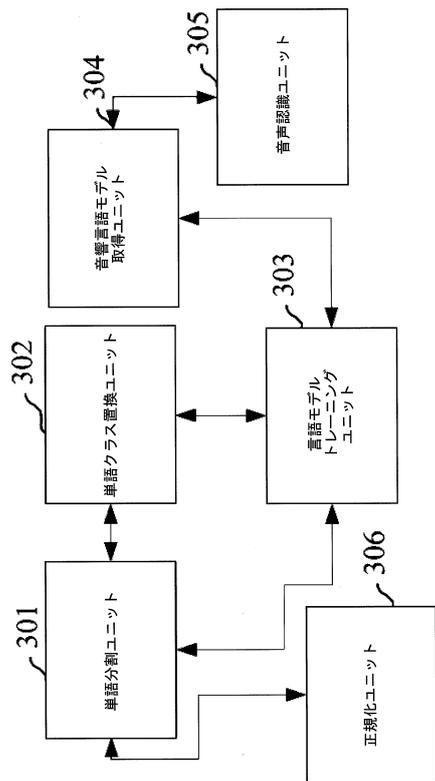
【図1C】



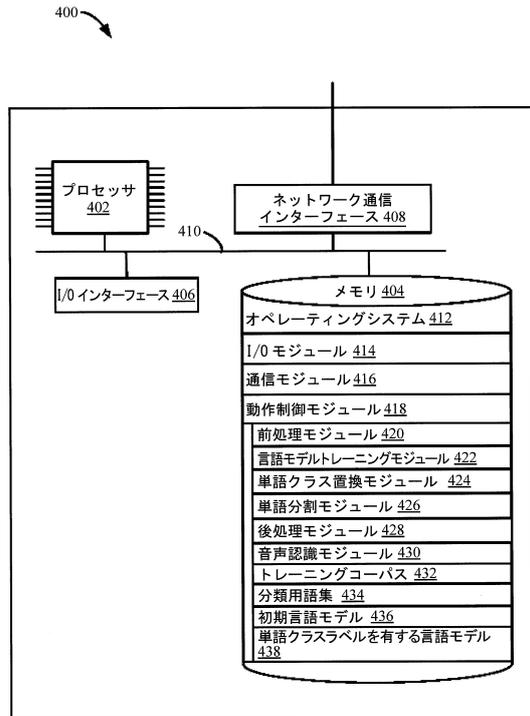
【図2】



【図3】



【図4】



フロントページの続き

- (72)発明者 ドゥリン・ルウ
中華人民共和国・グアンドン・518044・シェンジェン・フーティアン・ディストリクト・ジ
エンシン・ロード・エスイージー・パーク・イースト・ブロック・2・ルーム・403
- (72)発明者 ルウ・リ
中華人民共和国・グアンドン・518044・シェンジェン・フーティアン・ディストリクト・ジ
エンシン・ロード・エスイージー・パーク・イースト・ブロック・2・ルーム・403
- (72)発明者 フェン・ラオ
中華人民共和国・グアンドン・518044・シェンジェン・フーティアン・ディストリクト・ジ
エンシン・ロード・エスイージー・パーク・イースト・ブロック・2・ルーム・403
- (72)発明者 ボ・チェン
中華人民共和国・グアンドン・518044・シェンジェン・フーティアン・ディストリクト・ジ
エンシン・ロード・エスイージー・パーク・イースト・ブロック・2・ルーム・403
- (72)発明者 リ・ルウ
中華人民共和国・グアンドン・518044・シェンジェン・フーティアン・ディストリクト・ジ
エンシン・ロード・エスイージー・パーク・イースト・ブロック・2・ルーム・403
- (72)発明者 シャン・ジャン
中華人民共和国・グアンドン・518044・シェンジェン・フーティアン・ディストリクト・ジ
エンシン・ロード・エスイージー・パーク・イースト・ブロック・2・ルーム・403
- (72)発明者 エリュウ・ワン
中華人民共和国・グアンドン・518044・シェンジェン・フーティアン・ディストリクト・ジ
エンシン・ロード・エスイージー・パーク・イースト・ブロック・2・ルーム・403
- (72)発明者 シュアイ・ユエ
中華人民共和国・グアンドン・518044・シェンジェン・フーティアン・ディストリクト・ジ
エンシン・ロード・エスイージー・パーク・イースト・ブロック・2・ルーム・403

審査官 安田 勇太

- (56)参考文献 国際公開第2004/034378(WO, A1)
国際公開第2007/138875(WO, A1)

- (58)調査した分野(Int.Cl., DB名)
G10L 15/06