

【公報種別】特許法第17条の2の規定による補正の掲載

【部門区分】第6部門第3区分

【発行日】平成26年7月3日(2014.7.3)

【公表番号】特表2013-545160(P2013-545160A)

【公表日】平成25年12月19日(2013.12.19)

【年通号数】公開・登録公報2013-068

【出願番号】特願2013-530136(P2013-530136)

【国際特許分類】

G 06 F 17/27 (2006.01)

G 06 F 17/30 (2006.01)

【F I】

G 06 F 17/27 Z

G 06 F 17/30 220Z

G 06 F 17/30 170A

【手続補正書】

【提出日】平成26年5月16日(2014.5.16)

【手続補正1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項1】

コンピュータによって実行されるターゲット単語認識方法であって、

候補単語集合および対応する特性計算データを取得し、前記候補単語集合はテキストデータを備え、前記特性計算データは前記候補単語集合に関連付けられており、

複数のテキストセグメントを生成するために前記特性計算データのセグメント化を実行し、

テキストデータ組合せ集合を生成するために前記複数のテキストセグメントを組み合わせ、

前記候補単語集合と前記テキストデータ組合せ集合との共通集合を決定し、前記共通集合は複数のテキストデータ組合せを含み、

前記複数のテキストデータ組合せに関して複数の指定特性値を決定し、

基準を決定し、前記基準の決定は、

訓練サンプル単語集合およびサンプル特性計算データを取得し、前記サンプル特性計算データは、前記複数のサンプル単語と、前記複数のサンプル単語の指定特性値とを含み、

前記複数のサンプル単語に基づいて、サンプルテキストデータ組合せ集合を取得し、

前記サンプルテキストデータ組合せ集合と前記訓練サンプル単語集合との共通集合内のサンプルテキストデータ組合せの複数の指定特性値を決定し、

前記基準の一部として、前記共通集合内のサンプルテキストデータ組合せの指定特性値のしきい値を設定すること、を含み、

少なくとも一部は前記複数の指定特性値に基づいて、かつ少なくとも1つの前記基準に従って、前記複数のテキストデータ組合せの中から、特性値が前記基準を満たすターゲット単語を認識こと、

を備えるターゲット単語認識方法。

【請求項2】

前記候補単語集合は、ウェブサイトへのユーザ入力クエリキーワードに基づく請求項1に記載の方法。

【請求項 3】

前記特性計算データは、ウェブサイトへのユーザ入力クエリキーワードに応答したサーチ結果の記述から取得される請求項1に記載の方法。

【請求項 4】

前記ターゲット単語の少なくともいくつかは、辞書に登録されていない単語であり、前記方法は、さらに、前記登録されていない単語を前記辞書に追加することを備える請求項1に記載の方法。

【請求項 5】

前記指定特性値は相互情報を含む請求項1に記載の方法。

【請求項 6】

前記指定特性値は対数尤度を含む請求項1に記載の方法。

【請求項 7】

前記指定特性値は文脈エントロピーを含む請求項1に記載の方法。

【請求項 8】

前記指定特性値は位置ベースの単語内出現確率を含む請求項1に記載の方法。

【請求項 9】

前記複数のサンプルセグメントを組み合わせは、前記複数のサンプルセグメントにn-gram言語モデルを適用することを含む、請求項1に記載の方法。

【請求項 10】

前記しきい値を設定は、

前記共通集合内の前記訓練サンプル単語がターゲット単語であるか否かを判断するために、前記しきい値を使用して前記共通集合内の訓練サンプル単語をソートし、

前記判断を既知の結果と比較し、

前記判断が前記既知の結果と一致しない場合に、前記しきい値を調節することを含む、請求項1に記載の方法。

【請求項 11】

テキストデータ組合せ集合を生成するために前記複数のテキストセグメントを組み合わせることは、

n-gram窓に基づくn-gramモデルを採用し、所定の順序に従って前記n-gram窓をシフトさせ、組合せ後のテキストデータ組合せを得るために前記窓に含まれる前記単語セグメントの単語セグメント組合せを実行することを含む請求項1に記載の方法。

【請求項 12】

1つまたは複数のプロセッサを備えるターゲット単語認識システムであって、

候補単語集合および対応する特性計算データを取得し、前記候補単語集合はテキストデータを備え、前記特性計算データは前記候補単語集合に関連付けられており、

複数のテキストセグメントを生成するために前記特性計算データのセグメント化を実行し、

テキストデータ組合せ集合を生成するために前記複数のテキストセグメントを組み合わせ、

前記候補単語集合と前記テキストデータ組合せ集合との共通集合を決定し、前記共通集合は複数のテキストデータ組合せを含み、

前記複数のテキストデータ組合せに関して複数の指定特性値を決定し、

基準を決定し、前記基準の決定は、

訓練サンプル単語集合およびサンプル特性計算データを取得し、前記サンプル特性計算データは、前記複数のサンプル単語と、前記複数のサンプル単語の指定特性値とを含み、

前記複数のサンプル単語に基づいて、サンプルテキストデータ組合せ集合を取得し、

前記サンプルテキストデータ組合せ集合と前記訓練サンプル単語集合との共通集合

内のサンプルテキストデータ組合せの複数の指定特性値を決定し、

前記基準の一部として、前記共通集合内のサンプルテキストデータ組合せの指定特性値のしきい値を設定すること、を含み、

少なくとも一部は前記複数の指定特性値に基づいて、かつ少なくとも1つの前記基準に従って、前記複数のテキストデータ組合せの中から、特性値が前記基準を満たすターゲット単語を認識するように構成されている1つまたは複数のプロセッサと、

前記1つまたは複数のプロセッサに結合され、前記1つまたは複数のプロセッサに命令を提供するように構成されている1つまたは複数のメモリを備えるターゲット単語認識システム。

【請求項13】

前記候補単語集合は、ウェブサイトへのユーザ入力クエリキーワードに基づく請求項1_2記載のシステム。

【請求項14】

前記特性計算データは、ウェブサイトへのユーザ入力クエリキーワードに応答したサーチ結果の記述から取得される請求項1_2に記載のシステム。

【請求項15】

前記複数のサンプルセグメントを組み合わせは、前記複数のサンプルセグメントにn-gram言語モデルを適用することを含む請求項1_2に記載のシステム。

【請求項16】

前記しきい値の設定は、

前記共通集合内の前記訓練サンプル単語がターゲット単語であるか否かを判断するために、前記しきい値を使用して前記共通集合内の訓練サンプル単語をソートし、

前記判断を既知の結果と比較し、

前記判断が前記既知の結果と一致しない場合に、前記しきい値を調節することを含む、請求項1_2に記載のシステム。

【請求項17】

テキストデータ組合せ集合を生成するために前記複数のテキストセグメントを組み合わせることは、

n-gram窓に基づくn-gramモデルを採用し、所定の順序に従って前記n-gram窓をシフトさせ、組合せ後のテキストデータ組合せを得るために前記窓に含まれる前記単語セグメントの単語セグメント組合せを実行することを含む、請求項1_2に記載のシステム。

【請求項18】

ターゲット単語認識用のコンピュータプログラムであって、

候補単語集合および対応する特性計算データを取得するための機能と、前記候補単語集合はテキストデータを備え、前記特性計算データは前記候補単語集合に関連付けられており、

複数のテキストセグメントを生成するために前記特性計算データのセグメント化を実行するための機能と、

テキストデータ組合せ集合を生成するために前記複数のテキストセグメントを組み合わせるための機能と、

前記候補単語集合と前記テキストデータ組合せ集合との共通集合を決定するための機能と、前記共通集合は複数のテキストデータ組合せを含み、

前記複数のテキストデータ組合せに関して複数の指定特性値を決定するための機能と、基準を決定するための機能であって、

訓練サンプル単語集合およびサンプル特性計算データを取得し、前記サンプル特性計算データは、前記複数のサンプル単語と、前記複数のサンプル単語の指定特性値とを含み、

前記複数のサンプル単語に基づいて、サンプルテキストデータ組合せ集合を取得し、前記サンプルテキストデータ組合せ集合と前記訓練サンプル単語集合との共通集合内

のサンプルテキストデータ組合せの複数の指定特性値を決定し、

前記基準の一部として、前記共通集合内のサンプルテキストデータ組合せの指定特性値のしきい値を設定すること、を含む、基準を決定するための機能と、

少なくとも一部は前記複数の指定特性値に基づいて、かつ少なくとも1つの前記基準に従って、前記複数のテキストデータ組合せの中から、特性値が前記基準を満たすターゲット単語を認識するための機能とを含むコンピュータプログラム。

【請求項19】

コンピュータによって実行されるターゲット単語認識方法であって、
候補単語集合および対応する特性計算データを取得し、前記候補単語集合はテキストデータを備え、特性計算データは前記候補単語集合に関連付けられており、

複数のテキストセグメントを生成するために前記特性計算データのセグメント化を実行し、

テキストデータ組合せ集合を生成するために前記複数のテキストセグメントを組み合わせ、

前記候補単語集合と前記テキストデータ組合せ集合との共通集合を決定し、前記共通集合は複数のテキストデータ組合せを含み、

前記複数のテキストデータ組合せに関して複数の指定特性値を決定し、
基準を決定し、前記基準の決定は、

訓練サンプル単語集合およびサンプル特性計算データを取得し、前記訓練サンプル単語集合は複数のサンプル単語を備え、ソートの結果は前記複数のサンプル単語がそれぞれターゲット単語であるか否かを示し、前記サンプル特性計算データは、前記複数のサンプル単語と、前記複数のサンプル単語の指定特性値とを含み、

最小粒度の複数のサンプルセグメントを得るために前記複数のサンプル単語のセグメント化を実行し、

サンプルテキストデータ組合せ集合を得るために前記複数のサンプルセグメントを組み合わせ、

前記サンプルテキストデータ組合せ集合と前記訓練サンプル単語集合との共通集合を決定し、

前記共通集合内のサンプルテキストデータ組合せの複数の指定特性値を決定し、
前記基準の一部として、前記共通集合内のサンプルテキストデータ組合せの指定特性値のしきい値を設定することを含み、

少なくとも一部は前記複数の指定特性値に基づいて、かつ少なくとも1つの前記基準に従って、前記複数のテキストデータ組合せの中から、特性値が前記基準を満たすターゲット単語を認識こと、

を備えるターゲット単語認識方法。

【請求項20】

1つまたは複数のプロセッサを備えるターゲット単語認識システムであって、
候補単語集合および対応する特性計算データを取得し、前記候補単語集合はテキストデータを備え、特性計算データは前記候補単語集合に関連付けられており、

複数のテキストセグメントを生成するために前記特性計算データのセグメント化を実行し、

テキストデータ組合せ集合を生成するために前記複数のテキストセグメントを組み合わせ、

前記候補単語集合と前記テキストデータ組合せ集合との共通集合を決定し、前記共通集合は複数のテキストデータ組合せを含み、

前記複数のテキストデータ組合せに関して複数の指定特性値を決定し、
基準を決定し、前記基準の決定は、

訓練サンプル単語集合およびサンプル特性計算データを取得し、前記訓練サンプル単語集合は複数のサンプル単語を備え、ソートの結果は前記複数のサンプル単語がそれぞれターゲット単語であるか否かを示し、前記サンプル特性計算データは、前記複数のサン

プル単語と、前記複数のサンプル単語の指定特性値とを含み、

最小粒度の複数のサンプルセグメントを得るために前記複数のサンプル単語のセグメント化を実行し、

サンプルテキストデータ組合せ集合を得るために前記複数のサンプルセグメントを組み合わせ、

前記サンプルテキストデータ組合せ集合と前記訓練サンプル単語集合との共通集合を決定し、

前記共通集合内のサンプルテキストデータ組合せの複数の指定特性値を決定し、

前記基準の一部として、前記共通集合内のサンプルテキストデータ組合せの指定特性値のしきい値を設定することを含み、

少なくとも一部は前記複数の指定特性値に基づいて、かつ少なくとも1つの前記基準に従って、前記複数のテキストデータ組合せの中から、特性値が前記基準を満たすターゲット単語を認識するように構成されている1つまたは複数のプロセッサと、

前記1つまたは複数のプロセッサに結合され、前記1つまたは複数のプロセッサに命令を提供するように構成されている1つまたは複数のメモリを備えるターゲット単語認識システム。

【請求項21】

ターゲット単語認識用のコンピュータプログラムであって、

候補単語集合および対応する特性計算データを取得するための機能と、前記候補単語集合はテキストデータを備え、特性計算データは前記候補単語集合に関連付けられており、

複数のテキストセグメントを生成するために前記特性計算データのセグメント化を実行するための機能と、

テキストデータ組合せ集合を生成するために前記複数のテキストセグメントを組み合わせるための機能と、

前記候補単語集合と前記テキストデータ組合せ集合との共通集合を決定するための機能と、前記共通集合は複数のテキストデータ組合せを含み、

前記複数のテキストデータ組合せに関して複数の指定特性値を決定するための機能と、基準を決定するための機能であって、

訓練サンプル単語集合およびサンプル特性計算データを取得し、前記訓練サンプル単語集合は複数のサンプル単語を備え、ソートの結果は前記複数のサンプル単語がそれぞれターゲット単語であるか否かを示し、前記サンプル特性計算データは、前記複数のサンプル単語と、前記複数のサンプル単語の指定特性値とを含み、

最小粒度の複数のサンプルセグメントを得るために前記複数のサンプル単語のセグメント化を実行し、

サンプルテキストデータ組合せ集合を得るために前記複数のサンプルセグメントを組み合わせ、

前記サンプルテキストデータ組合せ集合と前記訓練サンプル単語集合との共通集合を決定し、

前記共通集合内のサンプルテキストデータ組合せの複数の指定特性値を決定し、

前記基準の一部として、前記共通集合内のサンプルテキストデータ組合せの指定特性値のしきい値を設定することを含む、基準を決定するための機能と、

少なくとも一部は前記複数の指定特性値に基づいて、かつ少なくとも1つの前記基準に従って、前記複数のテキストデータ組合せの中から、特性値が前記基準を満たすターゲット単語を認識するための機能とを含むコンピュータプログラム。