

【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】令和 3 年 7 月 29 日 (2021.7.29)

【公表番号】特表 2021-510858 (P2021-510858A)

【公表日】令和 3 年 4 月 30 日 (2021.4.30)

【年通号数】公開・登録公報 2021-020

【出願番号】特願 2019-569437 (P2019-569437)

【国際特許分類】

G 0 6 F 16/33 (2019.01)

G 0 6 F 16/383 (2019.01)

【F I】

G 0 6 F 16/33

G 0 6 F 16/383

【手続補正書】

【提出日】令和 3 年 6 月 16 日 (2021.6.16)

【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】

第 1 のテキストと第 2 のテキストとの間の関連性を算出する方法であって、
前記第 1 のテキスト及び前記第 2 のテキストをベクトルにそれぞれマッピングすること
と、

前記ベクトルに基づいて前記第 1 のテキストと前記第 2 のテキストとの間の類似部分及
び相違部分を判断することと、

前記類似部分及び前記相違部分の両方を用いて前記第 1 のテキストと前記第 2 のテキス
トとの間の関連性を算出することを含む、方法。

【請求項 2】

前記類似部分及び前記相違部分を判断することは、

前記第 1 のテキストの類似部分及び相違部分を判断することと、

前記第 2 のテキストの類似部分及び相違部分を判断することを含む、請求項 1 に記載
の方法。

【請求項 3】

前記第 1 のテキスト及び前記第 2 のテキストをベクトルにそれぞれマッピングした後、
前記ベクトルに対して次元削減を行って、低次元密ベクトルを用いて前記第 1 のテキス
ト及び前記第 2 のテキストを表すことをさらに含む、請求項 1 に記載の方法。

【請求項 4】

次元削減は、Word 2 Vec、Sentence 2 Vec 及び Doc 2 Vec からな
る群より選択される 1 つ以上の方法を用いて行われる、請求項 3 に記載の方法。

【請求項 5】

前記類似部分及び相違部分を判断することは、前記第 1 のテキスト及び前記第 2 のテキ
ストに基づいてセマンティックコンテンツ分析を行うことを含む、請求項 1 から 4 のいづ
れか 1 項に記載の方法。

【請求項 6】

前記セマンティックコンテンツ分析は、前記第 1 のテキストと前記第 2 のテキストとの
間でセマンティックマッチング処理を行って前記第 1 のテキストと前記第 2 のテキストと

の間のセマンティックオーバーレイを取得することを含む、請求項 5 に記載の方法。

【請求項 7】

前記セマンティックマッチング処理を行うことは、前記第 2 のテキストの単語のベクトルを用いて前記第 1 のテキストにおける単語のベクトルを再構成することと、

前記再構成の結果に基づいてセマンティックオーバーレイを算出することを含む、請求項 6 に記載の方法。

【請求項 8】

前記類似部分及び相違部分を判断することは、前記第 1 のテキスト及び前記第 2 のテキストに対してセマンティック分解を行うことを含む、請求項 1 から 7 のいずれか 1 項に記載の方法。

【請求項 9】

前記セマンティックオーバーレイは、式 (1) に基づいて算出され、

【数 1】

$$\min(\left\|S_i - \sum_{j=1}^m \alpha_{i,j} \times T_j\right\|^2 + \lambda \sum_{j=1}^m \alpha_{i,j}^2) \quad (1)$$

ここで、 S_i は前記第 1 のテキストの $d \times 1$ 列ベクトルであり、 T_j は前記第 2 のテキストの $d \times 1$ 列ベクトルであり、 $\alpha_{i,j}$ はセマンティックオーバーレイを算出するためのパラメータであり、 $\lambda > 0$ であり、 λ は正の実数である、請求項 7 に記載の方法。

【請求項 10】

前記第 1 のテキストの前記類似部分は、

【数 2】

$$S'_i = \sum_{j=1}^m A_{i,j} \times T_j$$

に基づいて算出され、

ここで、 S'_i は前記第 1 のテキストの前記類似部分を表し、 $A_{i,j}$ は $\alpha_{i,j}$ の行列であり、

$S_i - S'_i$ は前記第 1 のテキストの前記相違部分を表す、請求項 9 に記載の方法。

【請求項 11】

前記セマンティックオーバーレイは、式 (2) に基づいて算出され、

【数 3】

$$\min(\left\|T_i - \sum_{j=1}^m \beta_{i,j} \times S_j\right\|^2 + \lambda \sum_{j=1}^m \beta_{i,j}^2) \quad (2)$$

ここで、 T_i は前記第 2 のテキストの $d \times 1$ 列ベクトルであり、 S_j は前記第 1 のテキストの $d \times 1$ 列ベクトルであり、 $\beta_{i,j}$ はセマンティックオーバーレイを算出するためのパラメータであり、 $\lambda > 0$ であり、 λ は正の実数である、請求項 7 に記載の方法。

【請求項 12】

前記第 2 のテキストの前記類似部分は、

【数 4】

$$T_{j'} = \sum_{j=1}^m A_{i,j} \times S_j$$

に基づいて算出され、

ここで、 $T_{j'}$ は前記第 2 のテキストの前記類似部分を表し、 $A_{i,j}$ は $\beta_{i,j}$ の行列であり、

$T_j - T_{j'}$ は、前記第 2 のテキストの前記相違部分を表す、請求項 1 1 に記載の方法。

【請求項 1 3】

前記第 1 のテキストと前記第 2 のテキストとの間の前記類似部分及び前記相違部分を入力として用いて、リカレントニューラルネットワークを用いて前記第 1 のテキストと前記第 2 のテキストとの間の関連性を算出する、請求項 1 から 1 2 のいずれか 1 項に記載の方法。

【請求項 1 4】

前記入力は、

前記第 1 のテキストの類似部分及び相違部分と、

前記第 2 のテキストの類似部分及び相違部分とを含む、請求項 1 3 に記載の方法。

【請求項 1 5】

(S_s , T_s , L) で表されるサンプルデータを用いて前記リカレントニューラルネットワークをトレーニングすることをさらに含み、ここで、 S_s は第 1 のサンプルテキストを表し、 T_s は第 2 のサンプルテキストを表し、 L はサンプルの関連性を表す、請求項 1 3 に記載の方法。

【請求項 1 6】

前記リカレントニューラルネットワークをトレーニングすることは、

第 1 の範囲にある第 1 の関連性に第 1 の粒度値を、第 2 の範囲にある第 2 の関連性に第 2 の粒度値を割り当てることをさらに含み、請求項 1 5 に記載の方法。

【請求項 1 7】

第 1 のテキストと第 2 のテキストとの間の関連性を算出する装置であって、

メモリと、

1 つ以上のプロセッサとを備え、

前記メモリと、前記 1 つ以上のプロセッサとは互いに接続され、

前記メモリは、

前記第 1 のテキスト及び前記第 2 のテキストをベクトルにそれぞれマッピングし、

前記ベクトルに基づいて前記第 1 のテキストと前記第 2 のテキストとの間の類似部分及び相違部分を判断し、

前記類似部分及び前記相違部分の両方を用いて前記第 1 のテキストと前記第 2 のテキストとの間の関連性を算出するように、前記 1 つ以上のプロセッサを制御するためのコンピュータ実行可能命令を記憶する、装置。

【請求項 1 8】

前記メモリは、

前記第 1 のテキストの類似部分及び相違部分を判断し、

前記第 2 のテキストの類似部分及び相違部分を判断するように、前記 1 つ以上のプロセッサを制御するためのコンピュータ実行可能命令を記憶する、請求項 1 7 に記載の装置。

【請求項 1 9】

前記メモリは、前記第 1 のテキスト及び前記第 2 のテキストをベクトルにそれぞれマッピングした後に、前記ベクトルに対して次元削減を行って低次元密ベクトルを用いて前記第 1 のテキスト及び前記第 2 のテキストを表すように、前記 1 つ以上のプロセッサを制御するためのコンピュータ実行可能命令をさらに記憶する、請求項 1 7 に記載の装置。

【請求項 20】

次元削減は、Word 2 Vec、Sentence 2 Vec 及び Doc 2 Vec からなる群から選択される 1 つ以上の方法を用いて行われる、請求項 19 に記載の装置。

【請求項 21】

前記メモリは、前記第 1 のテキスト及び前記第 2 のテキストに基づいてセマンティックコンテンツ分析を行って前記類似部分及び前記相違部分を判断するように、前記 1 つ以上のプロセッサを制御するためのコンピュータ実行可能命令をさらに記憶する、請求項 17 から 20 のいずれか 1 項に記載の装置。

【請求項 22】

前記メモリは、前記第 1 のテキストと前記第 2 のテキストとの間でセマンティックマッチング処理を行って前記第 1 のテキストと前記第 2 のテキストとの間のセマンティックオーバーレイを取得するように、前記 1 つ以上のプロセッサを制御するためのコンピュータ実行可能命令をさらに記憶する、請求項 21 に記載の装置。

【請求項 23】

前記メモリは、前記第 2 のテキストの単語のベクトルを用いて前記第 1 のテキストにおける単語のベクトルを再構成し、再構成の結果に基づいてセマンティックオーバーレイを算出するように、前記 1 つ以上のプロセッサを制御するためのコンピュータ実行可能命令をさらに記憶する、請求項 22 に記載の装置。

【請求項 24】

前記メモリは、前記第 1 のテキスト及び前記第 2 のテキストに対してセマンティック分解を行って前記類似部分及び前記相違部分を判断するように、前記 1 つ以上のプロセッサを制御するためのコンピュータ実行可能命令をさらに記憶する、請求項 17 から 23 のいずれか 1 項に記載の装置。

【請求項 25】

前記セマンティックオーバーレイは、式 (1) に基づいて算出され、

【数 5】

$$\min \left(\left\| S_i - \sum_{j=1}^m \alpha_{i,j} \times T_j \right\|^2 + \lambda \sum_{j=1}^m \alpha_{i,j}^2 \right) \quad (1)$$

ここで、 S_i は前記第 1 のテキストの $d \times 1$ 列ベクトルであり、 T_j は前記第 2 のテキストの $d \times 1$ 列ベクトルであり、 $\alpha_{i,j}$ はセマンティックオーバーレイを算出するためのパラメータであり、 $\lambda > 0$ であり、 λ は正の実数である、請求項 23 に記載の装置。

【請求項 26】

前記第 1 のテキストの前記類似部分は、

【数 6】

$$S'_i = \sum_{j=1}^m A_{i,j} \times T_j$$

に基づいて算出され、

ここで、 S'_i は前記第 1 のテキストの前記類似部分を表し、 $A_{i,j}$ は $\alpha_{i,j}$ の行列であり、

$S_i - S'_i$ は前記第 1 のテキストの相違部分を表す、請求項 24 に記載の装置。

【請求項 27】

前記セマンティックオーバーレイは、式 (2) に基づいて算出され、

【数 7】

$$\min(\left\|T_i - \sum_{j=1}^m \beta_{ij} \times S_j\right\|^2 + \lambda \sum_{j=1}^m \beta_{ij}^2) \quad (2)$$

ここで、 T_i は前記第 2 のテキストの $d \times 1$ 列ベクトルであり、 S_j は前記第 1 のテキストの $d \times 1$ 列ベクトルであり、 $\beta_{i,j}$ はセマンティックオーバーレイを算出するためのパラメータであり、 $\lambda > 0$ であり、 λ は正の実数である、請求項 23 に記載の装置。

【請求項 28】

前記第 2 のテキストの前記類似部分は、

【数 8】

$$T_j' = \sum_{j=1}^m A_{i,j} \times S_j$$

に基づいて算出され、

ここで、 T_j' は前記第 2 のテキストの前記類似部分を表し、 $A_{i,j}$ は $\beta_{i,j}$ の行列であり、

$T_j - T_j'$ は、前記第 2 のテキストの前記相違部分を表す、請求項 27 に記載の装置。

【請求項 29】

前記第 1 のテキストと前記第 2 のテキストとの間の前記類似部分及び前記相違部分を入力として用いて、前記第 1 のテキストと前記第 2 のテキストとの間の関連性を算出するリカレントニューラルネットワークをさらに備える、請求項 17 から 28 のいずれか 1 項に記載の装置。

【請求項 30】

前記入力は、

前記第 1 のテキストの類似部分及び相違部分と、

前記第 2 のテキストの類似部分及び相違部分とを含む、請求項 29 に記載の装置。

【請求項 31】

前記メモリは、 (S_s, T_s, L) で表されるサンプルデータを用いて前記リカレントニューラルネットワークをトレーニングするように、前記 1 つ以上のプロセッサを制御するためのコンピュータ実行可能命令をさらに記憶し、 S_s は第 1 のサンプルテキストを表し、 T_s は第 2 のサンプルテキストを表し、 L はサンプルの関連性を表す、請求項 29 に記載の装置。

【請求項 32】

前記メモリは、第 1 の範囲にある第 1 の関連性に第 1 の粒度値を、第 2 の範囲にある第 2 の関連性に第 2 の粒度値を割り当てるように、前記 1 つ以上のプロセッサを制御するためのコンピュータ実行可能命令をさらに記憶する、請求項 31 に記載の装置。

【請求項 33】

請求項 17 から 32 のいずれか 1 項に記載の関連性を算出する装置を備える、データクエリ装置。

【請求項 34】

請求項 1 から 16 のいずれか 1 項に記載の方法により第 1 のテキストと複数の潜在的なターゲットテキストから選択される第 2 のテキストとの間の関連性を算出することと、

前記複数の潜在的なターゲットテキストと前記第 1 のテキストとの間の関連性の算出に基づいて、前記複数の潜在的なターゲットテキストをランク付けすることと、

前記複数の潜在的なターゲットテキストのランク付けの結果に基づいて、前記複数の潜

在的なターゲットテキストからターゲットテキストを選択することを含む、データクエリ方法。

【請求項 35】

コンピュータ可読命令を含むコンピュータプログラムであって、

前記コンピュータ可読命令は、プロセッサによって実行可能であり、前記プロセッサに

、

第1のテキスト及び第2のテキストをベクトルにそれぞれマッピングさせ、

前記ベクトルに基づいて前記第1のテキストと前記第2のテキストとの間の類似部分及び相違部分を判断させ、

前記類似部分及び前記相違部分の両方を用いて第1のテキストと第2のテキストとの間の関連性を算出させる、コンピュータプログラム。