

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2021年9月16日(16.09.2021)



(10) 国際公開番号

WO 2021/181569 A1

(51) 国際特許分類:

G06F 16/383 (2019.01) G06F 40/216 (2020.01)
G06N 20/00 (2019.01) G06F 40/30 (2020.01)

(21) 国際出願番号: PCT/JP2020/010579

(22) 国際出願日: 2020年3月11日(11.03.2020)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(71) 出願人: 日本電信電話株式会社 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目5番1号 Tokyo (JP).

(72) 発明者: 大杉 康仁 (OSUGI, Yasuhito); 〒1808585 東京都武蔵野市緑町3丁目9-11 NTT

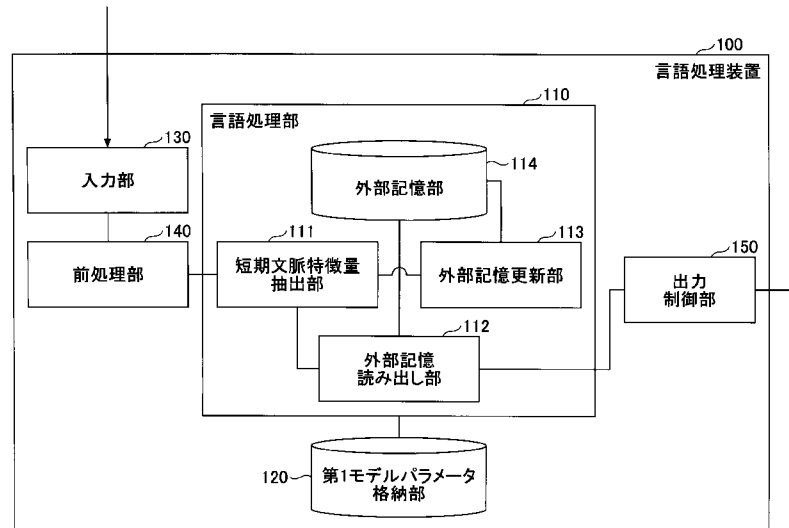
知的財産センタ内 Tokyo (JP). 斉藤 いつみ (SAITO, Itsumi); 〒1808585 東京都武蔵野市緑町3丁目9-11 NTT 知的財産センタ内 Tokyo (JP). 西田 京介 (NISHIDA, Kyosuke); 〒1808585 東京都武蔵野市緑町3丁目9-11 NTT 知的財産センタ内 Tokyo (JP). 浅野 久子 (ASANO, Hisako); 〒1808585 東京都武蔵野市緑町3丁目9-11 NTT 知的財産センタ内 Tokyo (JP). 富田 準二 (TOMITA, Junji); 〒1808585 東京都武蔵野市緑町3丁目9-11 NTT 知的財産センタ内 Tokyo (JP).

(74) 代理人: 伊東 忠重, 外 (ITO, Tadashige et al.); 〒1000005 東京都千代田区丸の内二丁目1

(54) Title: LANGUAGE PROCESSING DEVICE, TRAINING DEVICE, LANGUAGE PROCESSING METHOD, TRAINING METHOD, AND PROGRAM

(54) 発明の名称: 言語処理装置、学習装置、言語処理方法、学習方法、及びプログラム

[図1]



100 Language processing device

110 Language processing unit

111 Short-term context feature amount extraction unit

112 External storage reading unit

113 External storage update unit

114 External storage unit

120 First model parameter storage unit

130 Input unit

140 Preprocessing unit

150 Output control unit

(57) Abstract: This language processing device comprises: a preprocessing unit which divides an input text into a plurality of short texts; a language processing unit which calculates, for each of the plurality of short texts, a first feature amount and a second feature amount by using a trained model; and an external storage unit for storing a third feature amount for one or more short texts, wherein the language processing unit uses the trained model and calculates the second feature amount for a short text by using the first feature amount of the short text and the third feature amount stored in the external storage unit.



番 1 号 丸の内 M Y P L A Z A (明治安
田生命ビル) 1 6 階 Tokyo (JP).

- (81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.
- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

- 一 国際調査報告(条約第21条(3))

(57) 要約: 言語処理装置において、入力されたテキストを複数の短テキストに分割する前処理部と、前記複数の短テキストのそれぞれに対し、学習済みモデルを用いて第1の特徴量及び第2の特徴量を算出する言語処理部と、1以上の短テキストについての第3の特徴量を格納するための外部記憶部と、を備え、前記言語処理部は、前記学習済みモデルを用いて、ある短テキストに対する第2の特徴量を、当該短テキストの第1の特徴量と、前記外部記憶部に格納された第3の特徴量とを用いて算出する。

明 細 書

発明の名称：

言語処理装置、学習装置、言語処理方法、学習方法、及びプログラム
技術分野

[0001] 本発明は、言語理解モデルに関連するものである。

背景技術

[0002] 言語理解モデルに関する研究が近年盛んに行われている。言語理解モデルとは、トークンの分散表現を得るニューラルネットワークモデルの一つである。言語理解モデルでは、単一のトークンをモデルに入力するのではなく、トークンが使用されているテキスト全てをモデルに入力するため、テキスト内の他のトークンとの意味的な関係性を反映した分散表現を得ることができ

[0003] 上記のような言語理解モデルとして、例えば、非特許文献1に開示されている言語理解モデルがある。

先行技術文献

非特許文献

[0004] 非特許文献1：BERT、<https://arxiv.org/abs/1810.04805>、インターネット、
2020年2月26日検索

発明の概要

発明が解決しようとする課題

[0005] しかし、非特許文献1に開示されている言語理解モデルでは、長いテキスト（長いトークン系列）を上手く扱うことができないという課題がある。なお、長いテキストとは、所定の長さ（例：非特許文献1の言語理解モデルで適切に扱える512トークン）よりも長いテキストである。

[0006] 本発明は上記の点に鑑みてなされたものであり、長いテキストが入力された場合でも、テキスト内のトークン間の関係性を反映した特徴量を適切に抽出することを可能とする技術を提供することを目的とする。

課題を解決するための手段

[0007] 開示の技術によれば、入力されたテキストを複数の短テキストに分割する前処理部と、

前記複数の短テキストのそれぞれに対し、学習済みモデルを用いて第1の特徴量及び第2の特徴量を算出する言語処理部と、

1以上の短テキストについての第3の特徴量を格納するための外部記憶部と、を備え、

前記言語処理部は、

前記学習済みモデルを用いて、ある短テキストに対する第2の特徴量を、当該短テキストの第1の特徴量と、前記外部記憶部に格納された第3の特徴量とを用いて算出する

言語処理装置が提供される。

発明の効果

[0008] 開示の技術によれば、長いテキストが入力された場合でも、テキスト内のトークン間の関係性を反映した特徴量を適切に抽出することが可能となる。

[0009] 開示の技術によれば、データの分類を精度良く行う技術が提供される。

図面の簡単な説明

[0010] [図1]実施例1における言語処理装置100の構成図である。

[図2]実施例1における言語処理装置100の処理手順を示すフローチャートである。

[図3]外部記憶読み出し部112の構成と処理を説明するための図である。

[図4]外部記憶更新部113の構成と処理を説明するための図である。

[図5]実施例2における言語処理装置100の構成図である。

[図6]実施例2における言語処理装置100の処理手順を示すフローチャートである。

[図7]言語処理装置100のハードウェア構成の例を示す図である。

発明を実施するための形態

- [0011] 以下、図面を参照して本発明の実施の形態（本実施の形態）を説明する。
以下で説明する実施の形態は一例に過ぎず、本発明が適用される実施の形態は、以下の実施の形態に限られるわけではない。
- [0012] なお、本実施の形態において、「テキスト」とは文字の羅列であり、「テキスト」を「文」と呼んでもよい。また、「トークン」は、テキスト中の単語などの分散表現の単位を表す。例えば、非特許文献1では、単語を更に細かい単位サブワードへと分割しているため、非特許文献1でのトークンは当該サブワードとなる。
- [0013] 非特許文献1に開示されている言語理解モデルにおいては、Transformerのattention機構とposition encodingが重要な要素となっている。attention機構では、あるトークンとその他のトークンがどの程度関連しているかを表す重みを計算し、それに基づいてトークンの分散表現を計算する。position encodingでは、あるトークンがテキスト内のどの位置にあるかを表す特徴量を算出する。
- [0014] しかし、前述したように、非特許文献1に開示されている従来の言語理解モデルでは、長いテキストを上手く扱うことができない。その理由は2点あり、下記のとおりである。
- [0015] 1点目の理由は、position encodingが予め決められた数しか学習されていないということである。非特許文献1のposition encodingは512個学習されており、テキスト内の512トークンまでの位置を扱うことができる。従って、もしもテキストが512トークンよりも長ければ、513番目以降のトークンはそれ以前のトークンと同時に扱うことができない。
- [0016] 2点目の理由は、attention機構における計算コストが大きいということである。すなわち、attention機構では、入力テキスト内のトークン毎に全トークンとの間の関連性のスコアを計算するため、トークン系列が長くなればなるほど、スコア計算にかかるコストが増えてしまい、

計算機上で計算できなくなってしまう。

[0017] 以上の2つの理由から、非特許文献1に開示されている従来の言語理解モデルでは、長いトークン系列で構成されるテキストを上手く扱うことができない。本実施形態では、この課題を解決する言語処理装置100について説明する。

[0018] 以下、学習済の言語理解モデルを備える言語処理装置100が、入力テキストから文脈特徴量集合を得るための構成及び処理を実施例1として説明し、言語理解モデルの学習のための構成及び処理を実施例2として説明する。

[0019] (実施例1)

＜装置構成例＞

図1に示すように、実施例1の言語処理装置100は、言語処理部110、第1モデルパラメータ格納部120、入力部130、前処理部140、出力制御部150を備える。

[0020] 言語処理部110は、短文脈特徴量抽出部111、外部記憶読み出し部112、外部記憶更新部113、外部記憶部114を備える。言語処理部110による処理の詳細は後述するが、言語処理部110を構成する各部の概要は下記のとおりである。なお、外部記憶読み出し部112を特徴量算出部と呼んでもよい。また、言語処理装置100が備える外部記憶部114は、言語処理部110の外部に備えられてもよい。

[0021] 短文脈特徴量抽出部111は、入力されたテキストを分割して得られる短いトークン系列から特徴量を抽出する。外部記憶読み出し部112は、外部記憶部114に格納されている情報（外部記憶特徴量）を用いて中間特徴量を出力する。外部記憶更新部113は、外部記憶部114の情報を更新する。外部記憶部114には、長文脈の情報として、長文脈の中のキーワードやそれらの関係性を表す情報が格納されている。これらの情報は特徴量行列として行列の形で格納される。

[0022] 短文脈特徴量抽出部111、外部記憶読み出し部112、及び外部記憶更新部113は、それぞれ例えばニューラルネットワークのモデルとして実

装される。これら3つの機能部に外部記憶部114を加えた機能部である言語処理部110をメモリ付き言語理解モデルと称してもよい。第1モデルパラメータ格納部120には、当該メモリ付き言語理解モデルにおける学習済みのパラメータが格納されている。当該メモリ付き言語理解モデルに、当該学習済みのパラメータが設定されることで、言語処理部110は、実施例1の動作を実行することができる。

[0023] 入力部130は、装置外部から長期テキストを入力し、当該長期テキストを前処理部140に渡す。前処理部140は、入力された長期テキストを短期テキストの集合に変換し、短期テキストを1つずつ短文脈特徴量抽出部111に入力する。なお、実施例1（及び実施例2）における長期テキストを長いテキスト又は長テキストと言い換えてもよい。長いテキストとは、前述したとおり、所定の長さ（例：非特許文献1の言語理解モデルで適切に扱える512トークン）よりも長いテキストである。また、短期テキストを短いテキスト又は短テキストと言い換えてもよい。短いテキストは、テキストを分割することにより得られるテキストである。なお、入力部130から入力するテキストは、長いテキストに限られず、長いテキストよりも短いテキストであってもよい。

[0024] 出力制御部150は、外部記憶読み出し部112から、短期テキスト毎の中間特徴量を受信し、最後の短期テキストの中間特徴量を受信したら、中間特徴量を結合することで、入力された長期テキストの特徴量である長的文脈特徴量を出力する。

[0025] <装置の動作例>

以下、図2に示すフローチャートの手順に沿って、実施例1における言語処理装置100の動作例を説明する。実施例1（実施例2も同様）において、テキストは適切なトークナイザによって文字列からトークン系列に変換されており、テキストの長さはトークン系列の系列長（トークンの個数）を表すものとする。

[0026] <S101>

S 1 0 1において、入力部 1 3 0により長期テキストを入力する。長期テキストは入力部 1 3 0から前処理部 1 4 0に渡される。

[0027] <S 1 0 2>

S 1 0 2において、前処理部 1 4 0は、入力された長期テキストを、1以上の予め設定した長さ L^{seq} (L^{seq} は1以上の整数)の短期テキストへ分割し、短期テキスト集合 $S = \{s_1, s_2, \dots, s_N\}$ を得る。例えば、長さ5 1 2の長期テキストについて、 $L^{seq} = 3 2$ とすると、 $N = 1 6$ 、すなわち、1 6個の短期テキストが含まれた短期テキスト集合 S が生成される。

[0028] この集合 S の各要素(短期テキスト s_i)それぞれに対し、以下で説明するS 1 0 3～S 1 0 5の処理が行われる。

[0029] S 1 0 2において、より詳細には、前処理部 1 4 0は、パディングなどを使用する特殊なトークンを含めて、個々の短期テキストが長さ L^{seq} になるように分割する。

[0030] 例えば、短文脈特徴量抽出部 1 1 1として、非特許文献 1に開示されたモデルを使用する場合には、クラストークン([CLS])あるいはセパレータトークン([SEP])をトークン系列の先頭と末尾に追加するため、つまり、2トークンが追加されるため、実際には長期テキストを長さ「 $L^{seq} - 2$ 」の1以上のトークン系列に分割することになる。

[0031] <S 1 0 3>

S 1 0 3において、短期テキスト s_i が短文脈特徴量抽出部 1 1 1に入力され、短文脈特徴量抽出部 1 1 1は、短期テキスト s_i に対する短文脈特徴量 $h_i \in R^{d \times L^{seq}}$ を算出する。なお、記載の便宜上、 $R^{d \times L^{seq}}$ ($= d \times L^{seq}$ の実行列の集合)の右上の添え字「 $d \times L^{seq}$ 」を「 $d \times L^{seq}$ 」と記載している。ここで、 d は特徴量の次元数を表す。例えば、 $d = 7 6 8$ である。

[0032] 短文脈特徴量抽出部 1 1 1は、 s_i 内において各トークンと他の全てのトークンとの関係性を考慮して短文脈特徴量を算出する。短文脈特徴量抽出部 1 1 1は、特定のモデルに限定されないが、例えば、短文脈特徴量抽

出部 1 1 1 として、非特許文献 1 に開示されたニューラルネットワークモデル（BERT）を使用することができる。実施例 1（及び実施例 2）では、短文脈特徴量抽出部 1 1 1 として BERT を用いている。

[0033] BERT は、attention 機構を用いて、各トークンについて、当該トークンと他のトークンとの関係性を考慮し、それを反映した特徴量を出力することができる。参考文献（Transformer (<https://arxiv.org/abs/1706.03762>)) に開示されているように、attention 機構は以下の式（1）で表される。なお、下記の式（1）において、上記参考文献における d_k を d と記載している。

[0034] [数1]

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (1)$$

短文脈特徴量抽出部 1 1 1 は、 s_i から特徴量から Q 、 K 、 V を作成し、上記の式（1）で attention を算出する。式（1）において、 Q は Query の略であり、 K は Key の略であり、 V は Value の略である。短文脈特徴量抽出部 1 1 1（すなわち、BERT）において、トークンと他のトークンとの関係性を考慮する場合、上記式（1）における Q 、 K 、 V はそれぞれ各トークンの特徴量を線形変換した行列であり、 $Q, K, V \in \mathbb{R}^{d \times L_{seq}}$ となる。なお、本実施例では、線形変換して得られる Q 、 K 、 V の特徴量次元数を、 h_i の特徴量次元数 d と同一としているが、 Q 、 K 、 V の特徴量次元数が h_i の特徴量次元数 d と異なってもよい。

[0035] 上記式（1）における、

[0036]

[数2]

$$\text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)$$

のソフトマックス関数の計算は、トークンが他のトークンとどの程度関連しているかを表すスコア（確率）をトークンの特徴量間の内積（ QK^T ）に基づいて計算していることを示している。このスコアによって V を重み付け和したものが attention の出力、すなわち、他のトークンが当該トークンとどの程度関連しているかを表す特徴量となる。短文脈特徴量抽出部111は、この $\text{Attention}(Q, K, V)$ と当該トークンの特徴量を足し合わせることで、当該トークンと他のトークンとの関連性を反映した特徴量を得ている。

[0037] <S104>

S104において、S103で得られた短文脈特徴量 h_i と、外部記憶部114に格納されている外部記憶特徴量 $m \in R^{d \times M}$ とが外部記憶読み出し部112へ入力され、外部記憶読み出し部112は、入力情報から中間特徴量 $v_i \in R^{d \times L_{seq}}$ を算出し、出力する。本実施例では、 v_i 、 m の特徴量次元数は同じ d であるが、特徴量次元数が異なってもよい。

[0038] $m \in R^{d \times M}$ における M は外部記憶特徴量のスロット数を表す。外部記憶特徴量は、 $\{s_1, \dots, s_{i-1}\}$ から必要な情報を抽出・格納したベクトルである。どのように情報をベクトルとして抽出・格納するかについてはS105（更新の処理）で説明する。なお、 s_1 に関する処理を行う前に、外部記憶特徴量 m はランダムな数値で初期化するなど、予め適切な初期化を行っておく。

[0039] 外部記憶読み出し部112は、短文脈特徴量 h_i と外部記憶特徴量 m の各要素を比較し、外部記憶特徴量から必要な情報を抽出し、その抽出した情報と h_i の持つ情報とを足し合わせる。これにより、 $\{s_1, \dots, s_{i-1}\}$ の情報

を反映した s_i に関する中間特徴量を得ることができる。

[0040] すなわち、外部記憶読み出し部 112 は、2つの特徴量間 (h_i と m 間) のマッチングを行い必要な情報を抽出する。この処理を実行するニューラルネットワークモデルは、特定のモデルに限定されないが、例えば、前述した参考文献の *a t t e n t i o n* 機構 (式 (1)) を用いたモデルを使用することができる。本実施例では、当該 *a t t e n t i o n* 機構を用いたモデルを使用している。

[0041] 図 3 は、外部記憶読み出し部 112 に相当するモデルの構成 (及び処理内容) を示す図である。

[0042] 図 3 に示すように、当該モデルは、線形変換部 1、線形変換部 2、線形変換部 3、*A t t e n t i o n* 機構 4 (式 (1))、及び加算部 5 を有する。線形変換部 1 は、短期文脈特徴量 h_i を線形変換して Q を出力し、線形変換部 2、3 はそれぞれ m を線形変換して K 、 V を出力する。

[0043] Q 、 K 、 V が *A t t e n t i o n* 機構 4 (式 (1)) に入力され、*A t t e n t i o n* 機構 4 (式 (1)) は、 $u_i = A t t e n t i o n (Q, K, V)$ を出力する。

[0044] 上記のとおり、 h_i に基づいて Q (*Q u e r y*) が得られ、 m に基づいて K (*K e y*)、 V (*V a l u e*) が得られるため、

[0045] [数3]

$$\text{softmax} \left(\frac{QK^T}{\sqrt{d}} \right)$$

は短いテキスト (短期テキスト) の中の各トークンが外部記憶特徴量の各スロットとどの程度関連しているかを表す確率に相当し、その確率で外部記憶特徴量を重み付けて合計したものが u_i の各ベクトルである。すなわち、 u_i には短いテキストの各トークンについて関連した外部記憶特徴量の情報が格

納されている。図3に示すように、加算部5が、 u_i と h_i を加算することで外部記憶特徴量における長的文脈の情報を反映した中間特徴量 v_i を得ることができる。

[0046] <S105>

S105において、S103で得られた短文脈特徴量 h_i と、外部記憶特徴量 m とが外部記憶更新部113へ入力され、外部記憶更新部113は、これらの入力に基づいて新しい外部記憶特徴量 $m^{\wedge}$ を算出して外部記憶部114に出力し、 m を $m^{\wedge}$ で更新する。なお、記載の便宜上、本明細書では、 m の上に乗るハット(^)を、" $m^{\wedge}$ "のように m の後に記載している。

[0047] 外部記憶更新部113は、短文脈特徴量 h_i と外部記憶特徴量 m の各要素を比較し、 h_i の情報の中で保存すべき情報を抽出し、 m に上書きすることで情報の更新を行う。

[0048] すなわち、外部記憶更新部113は、2つの特徴量間(h_i と m 間)のマッチングを行い必要な情報を抽出する。この処理を実行するニューラルネットワークモデルは、特定のモデルに限定されないが、例えば、前述した参考文献のattention機構(式(1))を用いたモデルを使用することができる。本実施例では、当該attention機構を用いたモデルを使用している。

[0049] 図4は、外部記憶更新部113に相当するモデルの構成(及び処理内容)を示す図である。

[0050] 図4に示すように、当該モデルは、線形変換部11、線形変換部12、線形変換部13、Attention機構14(式(1))、及び加算部15を有する。線形変換部11は、 m を線形変換して Q を出力し、線形変換部12、13はそれぞれ短文脈特徴量 h_i を線形変換して K 、 V を出力する。

[0051] Q 、 K 、 V がAttention機構14(式(1))に入力され、Attention機構14(式(1))は、 $r = \text{Attention}(Q, K, V)$ を得る。

[0052] 上記のとおり、 m に基づいて Q が得られ、 h_i に基づいて K 、 V が得られる

ため、

[0053] [数4]

$$\text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)$$

は外部記憶特徴量の各スロットが短期テキストの各トークンとどの程度関連しているかを表す確率に相当し、その確率で短期テキストのトークンの特徴量を重み付けて合計したものが r の各ベクトルである。すなわち、 r には外部記憶特徴量の各スロットについて関連した短期テキスト内のトークンの情報が格納されている。図4に示すように、加算部15が、 r と m を加算する。これにより、 s_i から必要な情報 r を抜き出し、今まで抽出した情報 m と加算して特徴量 $m^{\wedge}$ が得られる。すなわち、 $\{s_1, \dots, s_i\}$ から必要な情報を抽出・格納した新しい外部記憶の特徴量 $m^{\wedge}$ を得ることができる。

[0054] <S106、S107>

S106において、出力制御部150は、外部記憶読み出し部112から受信した中間特徴量 v_i が、最後の短期テキストに対する中間特徴量であるかを判断し、最後でなければ次の短期テキストに対するS103からの処理を行うよう制御する。

[0055] 中間特徴量 v_i が、最後の短期テキストに対する中間特徴量である場合、すなわち、S103～S105が $S = \{s_1, s_2, \dots, s_N\}$ の全てに対して行われた場合、出力制御部150は、得られた中間特徴量の集合 $\{v_1, \dots, v_N\}$ における各 v_i を系列長方向に結合することで、長的文脈特徴量 V を得て、それを出力する。

[0056] 例えば、長さ512の長期テキストについて $L_{seq} = 32$ として、S103～S107を実行すると、 $\{v_1, \dots, v_{16}\}$ が得られる。 $d = 768$ とすれば、 v_i は768次元の列ベクトルが32個並んだ 768×32 の行列である

。これを列方向に結合した768×512の行列を、入力された長期テキストに対する長的文脈特徴量Vとする。

[0057] (実施例2)

次に、実施例2を説明する。実施例2では、言語処理部110、すなわち、メモリ付き言語理解モデルのモデルパラメータを学習するための構成及び処理内容について説明する。

[0058] メモリ付き言語理解モデルの学習方法は特定の方法に限定されないが、本実施例では、一例として、マスクされたトークンを予測するタスク（例：非特許文献1の3.1節Task#1 Masked LM）を通じてモデルパラメータを学習する方法について説明する。

[0059] <装置構成例>

図5に示すように、実施例2の言語処理装置100は、言語処理部110、第1モデルパラメータ格納部120、入力部130、前処理部140、第2モデルパラメータ格納部160、トークン予測部170、更新部180を備える。言語処理部110は、短文脈特徴量抽出部111、外部記憶読み出し部112、外部記憶更新部113、外部記憶部114を備える。言語処理装置100が備える外部記憶部114は、言語処理部110の外部に備えられてもよい。

[0060] すなわち、実施例2の言語処理装置100は、実施例1の言語処理装置100と比較して、出力制御部150が除かれ、第2モデルパラメータ格納部160、トークン予測部170、更新部180が追加されたものである。追加されたもの以外の構成と動作は基本的に実施例1と同じである。

[0061] なお、実施例1の言語処理装置100に、第2モデルパラメータ格納部160、トークン予測部170、更新部180を追加した言語処理装置100を用いることで、モデルパラメータの学習と、実施例1で説明した長的文脈特徴量の取得を1つの言語処理装置100で行うことができる。また、実施例2の言語処理装置100と実施例1の言語処理装置100は別装置であってもよく、その場合、実施例2の言語処理装置100の学習処理で得られた

モデルパラメータを実施例１の言語処理装置１００の第１モデルパラメータ格納部１２０に格納することで、実施例１の言語処理装置１００において、長的文脈特徴量の取得を行うことができる。また、実施例２の言語処理装置１００を学習装置と呼んでもよい。

[0062] トークン予測部１７０は、 v_i を用いてトークンを予測する。実施例２のトークン予測部１７０は、ニューラルネットワークのモデルとして実装される。更新部１８０は、トークンの正解とトークンの予測結果とに基づいて、短文脈特徴量抽出部１１１、外部記憶読み出し部１１２、及び外部記憶更新部１１３のモデルパラメータと、トークン予測部１７０のモデルパラメータを更新する。トークン予測部１７０のモデルパラメータは、第２モデルパラメータ格納部１６０に格納されている。

[0063] また、実施例２では、Ｗｅｂ上に公開されている長いテキストを収集し、図５に示すテキスト集合データベース２００に格納しておく。テキスト集合データベース２００から長期テキストが読み出される。例えば、ある文書の１段落の文章（文と呼んでもよい）を１つの長期テキストとして扱うことができる。

[0064] ＜装置の動作例＞

以下、図６に示すフローチャートの手順に沿って、実施例２における言語処理装置１００の動作例を説明する。短文脈特徴量抽出部１１１、外部記憶読み出し部１１２、及び外部記憶更新部１１３のモデルパラメータ、及びトークン予測部１７０のモデルパラメータは、任意の適当な値で初期化されているとする。

[0065] ＜Ｓ２０１＞

Ｓ２０１において、入力部１３０はテキスト集合データベースから長期テキストを読み出し、入力する。長期テキストは入力部１３０から前処理部１４０に渡される。

[0066] ＜Ｓ２０２＞

Ｓ１０２において、前処理部１４０は、入力された長期テキストを、１以

上の予め設定した長さ L_{seq} (L_{seq} は 1 以上の整数) の短期テキストへ分割し、短期テキスト集合 $S = \{s_1, s_2, \dots, s_N\}$ を得る。

[0067] S 2 0 2 で得た集合 S の各要素 (短期テキスト s_i) それぞれに対し、以下の処理を行う。

[0068] < S 2 0 3 >

前処理部 1 4 0 は、 s_i 中のトークンの内、いくつかのトークンを選択し、選択したトークンを、マスクトークン ([MASK]) やランダムに選んだ別のトークンに置換する、あるいは、選択したトークンそのままのトークンを維持し、マスクされた短期テキスト $s_i^{\wedge}$ を得る。ここで、置換や維持の条件は非特許文献 1 における条件と同じでよい。このとき置換もしくは維持の対象に選ばれたトークンが、トークン予測部 1 7 0 での予測対象となる。

[0069] < S 2 0 4、S 2 0 5、S 2 0 6 >

実施例 1 の S 1 0 3、S 1 0 4、S 1 0 5 と同じ処理により、短期テキスト $s_i^{\wedge}$ に対する中間特徴量 v_i が得られ、外部記憶特徴量 m が更新される。

[0070] < S 2 0 7 >

外部記憶読み出し部 1 1 2 は、中間特徴量 v_i をトークン予測部 1 7 0 へ入力し、トークン予測部 1 7 0 が予測トークンを出力する。

[0071] 実施例 2 では、トークン予測部 1 7 0 は、 v_i の t 番目のトークンに関する特徴量 $v_i^{(t)} \in \mathbb{R}^d$ を基に t 番目のトークンを予め決められた語彙の中から予測する機構である。 t 番目のトークンは、置換や維持の対象となったトークンに相当する。当該機構により、例えば、1 層の Feed Forward Network を用いて、 $v_i^{(t)}$ を次元数が語彙サイズ d' である特徴量 $y^{(t)} \in \mathbb{R}^{d'}$ へと変換し $y^{(t)}$ の要素の値が最大となるインデックスを用いて、語彙からトークンを予測することができる。

[0072] 例えば、 $d' = 32000$ とし、 t 番目のトークンが 32000 個の語彙セット (リスト) のどの語彙かを予測するとする。32000 次元のベクトルである $y^{(t)}$ の要素について、3000 番目の要素が最大値となる場合、語彙リストの 3000 番目のトークンが求めるトークンとなる。

[0073] <S 2 0 8>

S 2 0 8において、マスクされた短期テキストと予測トークンが更新部 1 8 0に入力され、更新部 1 8 0は、教師あり学習で第 1 モデルパラメータ格納部 1 2 0におけるモデルパラメータと第 2 モデルパラメータ格納部 1 6 0におけるモデルパラメータを更新する。

[0074] <S 2 0 9>

S 2 0 9において、トークン予測部 1 7 0は、外部記憶読み出し部 1 1 2から受信した中間特徴量 v_i が、最後の短期テキストに対する中間特徴量であるか否かを判断し、最後でなければ次の短期テキストに対する S 2 0 3からの処理を行うよう制御する。

[0075] 中間特徴量 v_i が、最後の短期テキストに対する中間特徴量である場合、すなわち、S 2 0 3～S 2 0 8が $S = \{s_1, s_2, \dots, s_N\}$ の全てに対して行われた場合、処理を終了する。

[0076] (ハードウェア構成例)

本実施の形態における言語処理装置 1 0 0は、例えば、コンピュータに、本実施の形態で説明する処理内容を記述したプログラムを実行させることにより実現可能である。なお、この「コンピュータ」は、物理マシンであってもよいし、クラウド上の仮想マシンであってもよい。仮想マシンを使用する場合、ここで説明する「ハードウェア」は仮想的なハードウェアである。

[0077] 上記プログラムは、コンピュータが読み取り可能な記録媒体（可搬メモリ等）に記録して、保存したり、配布したりすることが可能である。また、上記プログラムをインターネットや電子メール等、ネットワークを通して提供することも可能である。

[0078] 図 7 は、上記コンピュータのハードウェア構成例を示す図である。図 7 のコンピュータは、それぞれバス B Sで相互に接続されているドライブ装置 1 0 0 0、補助記憶装置 1 0 0 2、メモリ装置 1 0 0 3、CPU 1 0 0 4、インタフェース装置 1 0 0 5、表示装置 1 0 0 6、及び入力装置 1 0 0 7等を有する。なお、当該コンピュータは、CPU 1 0 0 4の代わりに、又はCP

U 1 0 0 4 と共に G P U (Graphics Processing Unit) を有してもよい。

[0079] 当該コンピュータでの処理を実現するプログラムは、例えば、C D - R O M 又はメモリカード等の記録媒体 1 0 0 1 によって提供される。プログラムを記憶した記録媒体 1 0 0 1 がドライブ装置 1 0 0 0 にセットされると、プログラムが記録媒体 1 0 0 1 からドライブ装置 1 0 0 0 を介して補助記憶装置 1 0 0 2 にインストールされる。但し、プログラムのインストールは必ずしも記録媒体 1 0 0 1 より行う必要はなく、ネットワークを介して他のコンピュータよりダウンロードするようにしてもよい。補助記憶装置 1 0 0 2 は、インストールされたプログラムを格納すると共に、必要なファイルやデータ等を格納する。

[0080] メモリ装置 1 0 0 3 は、プログラムの起動指示があった場合に、補助記憶装置 1 0 0 2 からプログラムを読み出して格納する。C P U 1 0 0 4 (又は G P U、又は C P U 1 0 0 4 と G P U) は、メモリ装置 1 0 0 3 に格納されたプログラムに従って、当該装置に係る機能を実現する。インタフェース装置 1 0 0 5 は、ネットワークに接続するためのインタフェースとして用いられる。表示装置 1 0 0 6 はプログラムによる G U I (G r a p h i c a l U s e r I n t e r f a c e) 等を表示する。入力装置 1 0 0 7 はキーボード及びマウス、ボタン、又はタッチパネル等で構成され、様々な操作指示を入力させるために用いられる。

[0081] (実施の形態の効果等)

以上、説明したように、本実施の形態では、長期テキストを分割して得られる短期テキストの情報を順次外部記憶部 1 1 4 に書き込み、新たな短期テキストの特徴量を算出する際に、外部記憶部 1 1 4 に格納されているこれまでに書き込んだテキストの情報(長い文脈の情報)を利用することとしたので、長いテキストを一貫して扱うことができる。

[0082] すなわち、本実施の形態では、短期的な情報の処理と長期的な情報の処理を分けることで、a t t e n t i o n 機構にかかる計算コストを抑えることができる。また、外部記憶部 1 1 4 に長期的な情報を格納できるため、長い

テキストを系列長の制限なく扱うことができる。

(実施の形態のまとめ)

本明細書には、少なくとも下記の各項に記載した言語処理装置、学習装置、言語処理方法、学習方法、及びプログラムが記載されている。

(第1項)

入力されたテキストを複数の短テキストに分割する前処理部と、
前記複数の短テキストのそれぞれに対し、学習済みモデルを用いて第1の特徴量及び第2の特徴量を算出する言語処理部と、

1以上の短テキストについての第3の特徴量を格納するための外部記憶部と、を備え、

前記言語処理部は、

前記学習済みモデルを用いて、ある短テキストに対する第2の特徴量を、当該短テキストの第1の特徴量と、前記外部記憶部に格納された第3の特徴量とを用いて算出する

言語処理装置。

(第2項)

前記言語処理部は、前記学習済みモデルを用いて、

短テキストの第2の特徴量の算出を行う度に、当該短テキストについての、短テキスト内の各トークンと前記外部記憶部に格納された情報との間の関連性を反映させた特徴量を用いて前記外部記憶部に格納された第3の特徴量を更新する

第1項に記載の言語処理装置。

(第3項)

入力されたテキストから分割して得られた複数の短テキストにおけるある短テキストについて、当該短テキストに含まれる全トークンのうちの一部のトークンを別のトークンに変換する、又は、変換せずに維持する前処理部と、

前記一部のトークンが変換又は維持された前記短テキストに対し、モデル

を用いて第1の特徴量及び第2の特徴量を算出する言語処理部と、

前記一部のトークンが変換又は維持された1以上の前記短テキストについての第3の特徴量を格納するための外部記憶部と、

前記第2の特徴量を用いて、前記一部のトークンを予測するトークン予測部と、

前記一部のトークンと、前記トークン予測部による予測結果とに基づいて、前記言語処理部を構成する前記モデルのモデルパラメータを更新する更新部と、を備え、

前記言語処理部は、前記モデルを用いて、

前記一部のトークンが変換又は維持された前記短テキストに対する第2の特徴量を、当該短テキストの第1の特徴量と、前記外部記憶部に格納された第3の特徴量とを用いて算出し、

前記前処理部、前記言語処理部、前記トークン予測部、及び前記更新部の処理を、前記複数の短テキストのそれぞれに対して実行する

学習装置。

(第4項)

言語処理装置が実行する言語処理方法であって、

入力されたテキストを複数の短テキストに分割するステップと、

前記複数の短テキストのそれぞれに対し、学習済みモデルを用いて第1の特徴量及び第2の特徴量を算出する言語処理ステップと、を備え、

前記言語処理装置は、1以上の短テキストについての第3の特徴量を格納するための外部記憶部を備えており、

前記言語処理ステップにおいて、前記学習済みモデルを用いて

ある短テキストに対する第2の特徴量を、当該短テキストの第1の特徴量と、前記外部記憶部に格納された第3の特徴量とを用いて算出する

言語処理方法。

(第5項)

モデルを備える学習装置が実行する学習方法であって、

入力されたテキストから分割して得られた複数の短テキストにおけるある短テキストについて、当該短テキストに含まれる全トークンのうちの一部のトークンを別のトークンに変換する、又は、変換せずに維持する前処理ステップと、

前記一部のトークンが変換又は維持された前記短テキストに対し、前記モデルを用いて第1の特徴量及び第2の特徴量を算出する言語処理ステップと、

前記第2の特徴量を用いて、前記一部のトークンを予測するトークン予測ステップと、

前記一部のトークンと、前記トークン予測ステップによる予測結果とに基づいて、前記モデルのモデルパラメータを更新する更新ステップと、を備え、

前記学習装置は、前記一部のトークンが変換又は維持された1以上の前記短テキストについての第3の特徴量を格納するための外部記憶部を備えており、

前記言語処理ステップにおいて、前記モデルを用いて、

前記一部のトークンが変換又は維持された前記短テキストに対する第2の特徴量を、当該短テキストの第1の特徴量と、前記外部記憶部に格納された第3の特徴量とを用いて算出し、

前記前処理ステップ、前記言語処理ステップ、前記トークン予測ステップ、及び前記更新ステップの処理を、前記複数の短テキストのそれぞれに対して実行する

学習方法。

(第6項)

コンピュータを、第1項又は第2項に記載の言語処理装置における各部として機能させるためのプログラム。

(第7項)

コンピュータを、第3項に記載の学習装置における各部として機能させる

ためのプログラム。

[0083] 以上、本実施の形態について説明したが、本発明はかかる特定の実施形態に限定されるものではなく、特許請求の範囲に記載された本発明の要旨の範囲内において、種々の変形・変更が可能である。

符号の説明

- [0084] 1 0 0 言語処理装置
- 1 1 0 言語処理部
- 1 1 1 短期文脈特徴量抽出部
- 1 1 2 外部記憶読み出し部
- 1 1 3 外部記憶更新部
- 1 1 4 外部記憶部
- 1 2 0 第1モデルパラメータ格納部
- 1 3 0 入力部
- 1 4 0 前処理部
- 1 5 0 出力制御部
- 1 6 0 第2モデルパラメータ格納部
- 1 7 0 トークン予測部
- 1 8 0 更新部
- 2 0 0 テキスト集合データベース
- 1 0 0 0 ドライブ装置
- 1 0 0 1 記録媒体
- 1 0 0 2 補助記憶装置
- 1 0 0 3 メモリ装置
- 1 0 0 4 C P U
- 1 0 0 5 インタフェース装置
- 1 0 0 6 表示装置
- 1 0 0 7 入力装置

請求の範囲

- [請求項1] 入力されたテキストを複数の短テキストに分割する前処理部と、
 前記複数の短テキストのそれぞれに対し、学習済みモデルを用いて
 第1の特徴量及び第2の特徴量を算出する言語処理部と、
 1以上の短テキストについての第3の特徴量を格納するための外部
 記憶部と、を備え、
 前記言語処理部は、
 前記学習済みモデルを用いて、ある短テキストに対する第2の特徴
 量を、当該短テキストの第1の特徴量と、前記外部記憶部に格納され
 た第3の特徴量とを用いて算出する
 言語処理装置。
- [請求項2] 前記言語処理部は、前記学習済みモデルを用いて、
 短テキストの第2の特徴量の算出を行う度に、当該短テキストにつ
 いての、短テキスト内の各トークンと前記外部記憶部に格納された情
 報との間の関連性を反映させた特徴量を用いて前記外部記憶部に格納
 された第3の特徴量を更新する
 請求項1に記載の言語処理装置。
- [請求項3] 入力されたテキストから分割して得られた複数の短テキストにおけ
 るある短テキストについて、当該短テキストに含まれる全トークンの
 うちの一部のトークンを別のトークンに変換する、又は、変換せずに
 維持する前処理部と、
 前記一部のトークンが変換又は維持された前記短テキストに対し、
 モデルを用いて第1の特徴量及び第2の特徴量を算出する言語処理部
 と、
 前記一部のトークンが変換又は維持された1以上の前記短テキスト
 についての第3の特徴量を格納するための外部記憶部と、
 前記第2の特徴量を用いて、前記一部のトークンを予測するトーク
 ン予測部と、

前記一部のトークンと、前記トークン予測部による予測結果とに基づいて、前記言語処理部を構成する前記モデルのモデルパラメータを更新する更新部と、を備え、

前記言語処理部は、前記モデルを用いて、

前記一部のトークンが変換又は維持された前記短テキストに対する第2の特徴量を、当該短テキストの第1の特徴量と、前記外部記憶部に格納された第3の特徴量とを用いて算出し、

前記前処理部、前記言語処理部、前記トークン予測部、及び前記更新部の処理を、前記複数の短テキストのそれぞれに対して実行する学習装置。

[請求項4]

言語処理装置が実行する言語処理方法であって、

入力されたテキストを複数の短テキストに分割するステップと、

前記複数の短テキストのそれぞれに対し、学習済みモデルを用いて第1の特徴量及び第2の特徴量を算出する言語処理ステップと、を備え、

前記言語処理装置は、1以上の短テキストについての第3の特徴量を格納するための外部記憶部を備えており、

前記言語処理ステップにおいて、前記学習済みモデルを用いて

ある短テキストに対する第2の特徴量を、当該短テキストの第1の特徴量と、前記外部記憶部に格納された第3の特徴量とを用いて算出する

言語処理方法。

[請求項5]

モデルを備える学習装置が実行する学習方法であって、

入力されたテキストから分割して得られた複数の短テキストにおけるある短テキストについて、当該短テキストに含まれる全トークンのうちの一部のトークンを別のトークンに変換する、又は、変換せずに維持する前処理ステップと、

前記一部のトークンが変換又は維持された前記短テキストに対し、

前記モデルを用いて第1の特徴量及び第2の特徴量を算出する言語処理ステップと、

前記第2の特徴量を用いて、前記一部のトークンを予測するトークン予測ステップと、

前記一部のトークンと、前記トークン予測ステップによる予測結果とに基づいて、前記モデルのモデルパラメータを更新する更新ステップと、を備え、

前記学習装置は、前記一部のトークンが変換又は維持された1以上の前記短テキストについての第3の特徴量を格納するための外部記憶部を備えており、

前記言語処理ステップにおいて、前記モデルを用いて、

前記一部のトークンが変換又は維持された前記短テキストに対する第2の特徴量を、当該短テキストの第1の特徴量と、前記外部記憶部に格納された第3の特徴量とを用いて算出し、

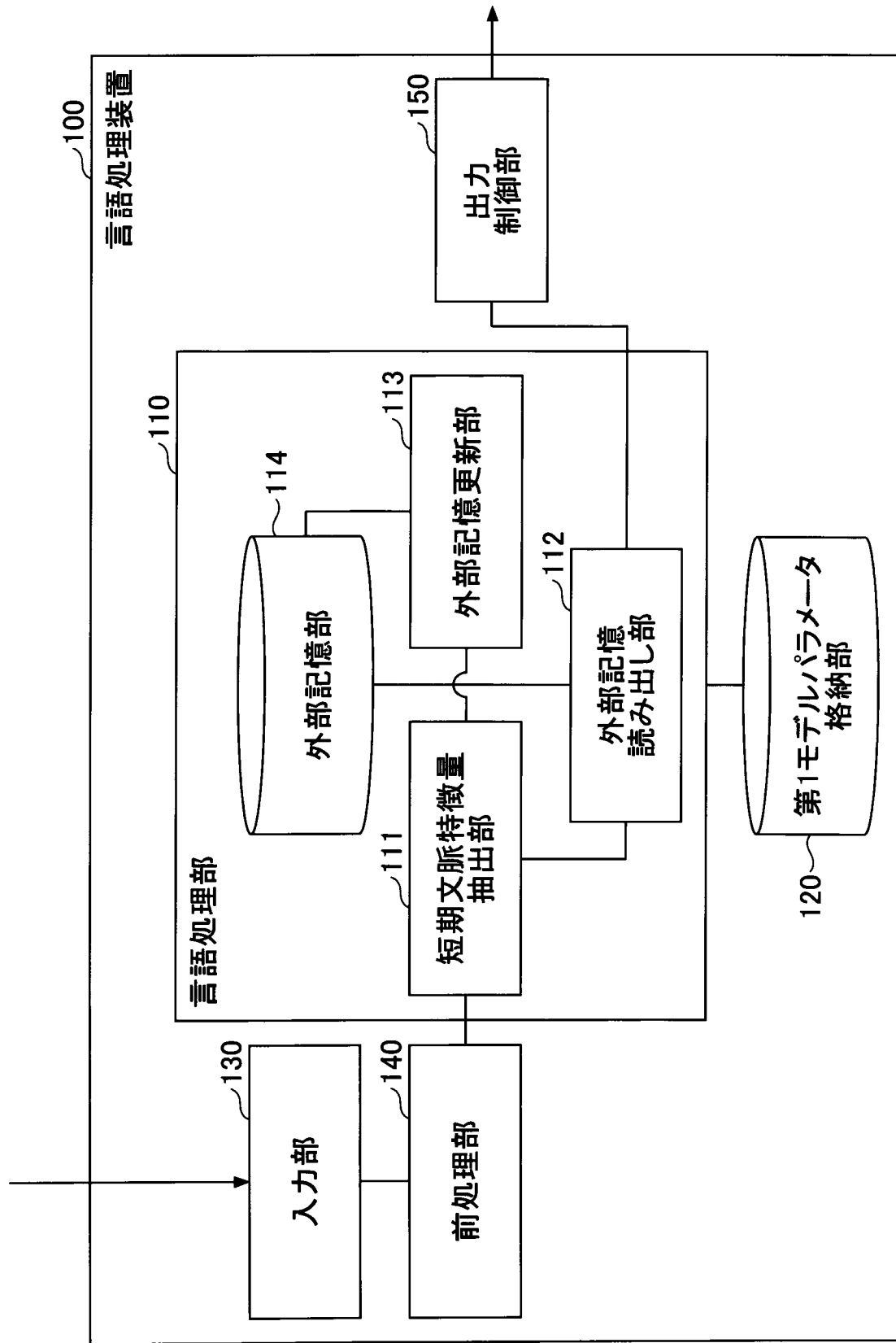
前記前処理ステップ、前記言語処理ステップ、前記トークン予測ステップ、及び前記更新ステップの処理を、前記複数の短テキストのそれぞれに対して実行する

学習方法。

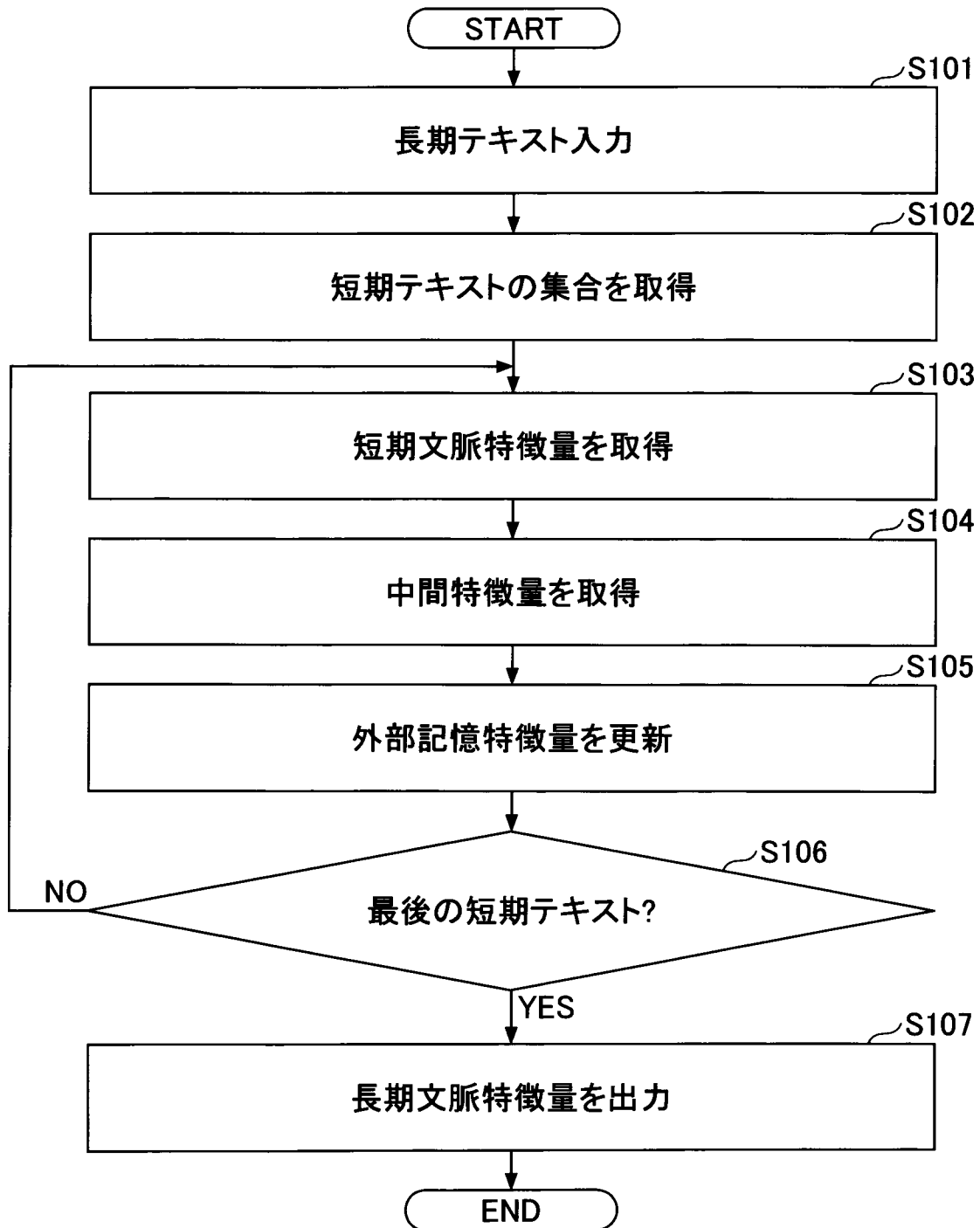
[請求項6] コンピュータを、請求項1又は2に記載の言語処理装置における各部として機能させるためのプログラム。

[請求項7] コンピュータを、請求項3に記載の学習装置における各部として機能させるためのプログラム。

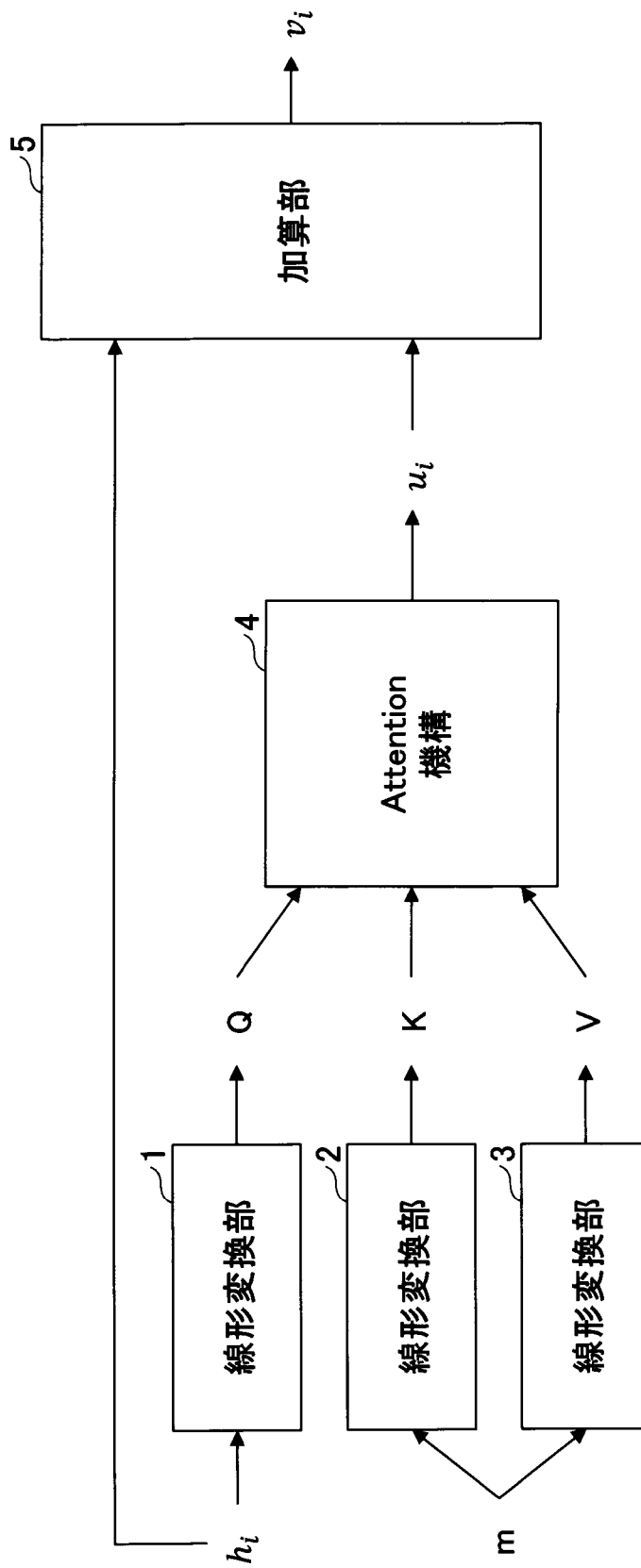
[図1]



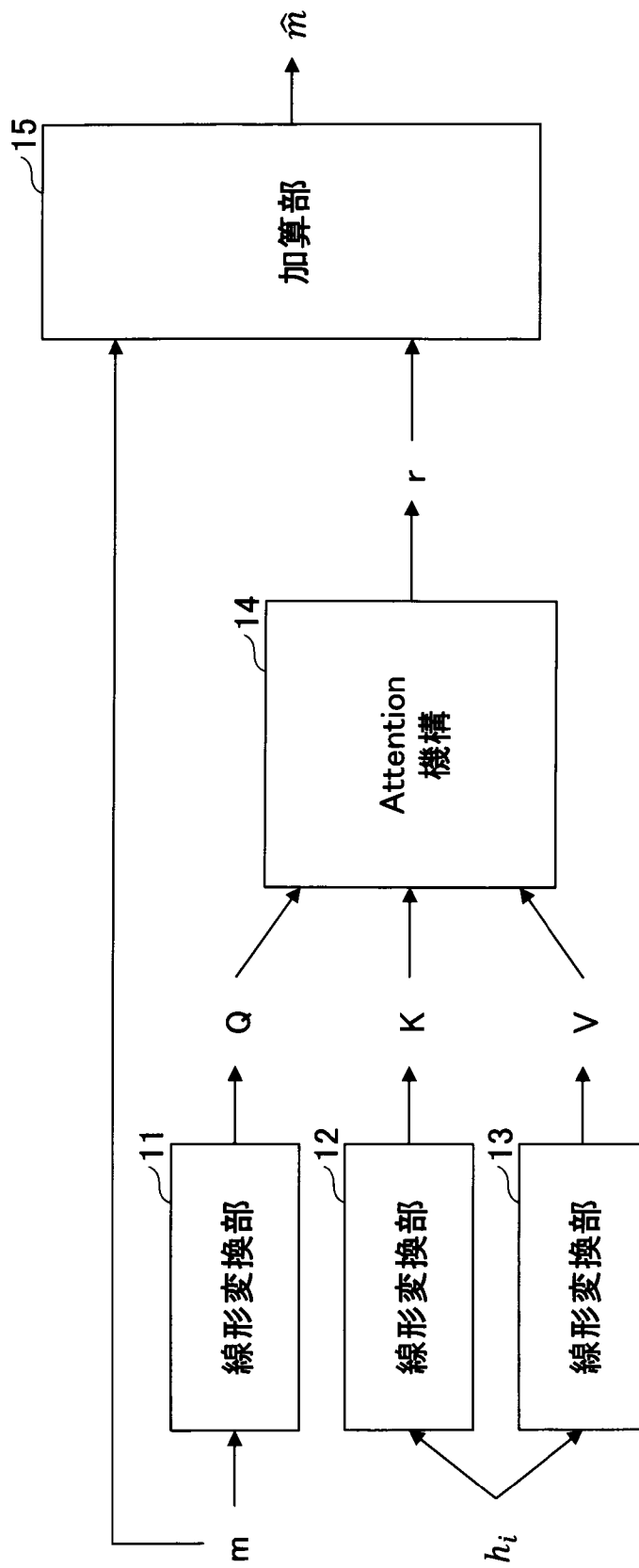
[図2]



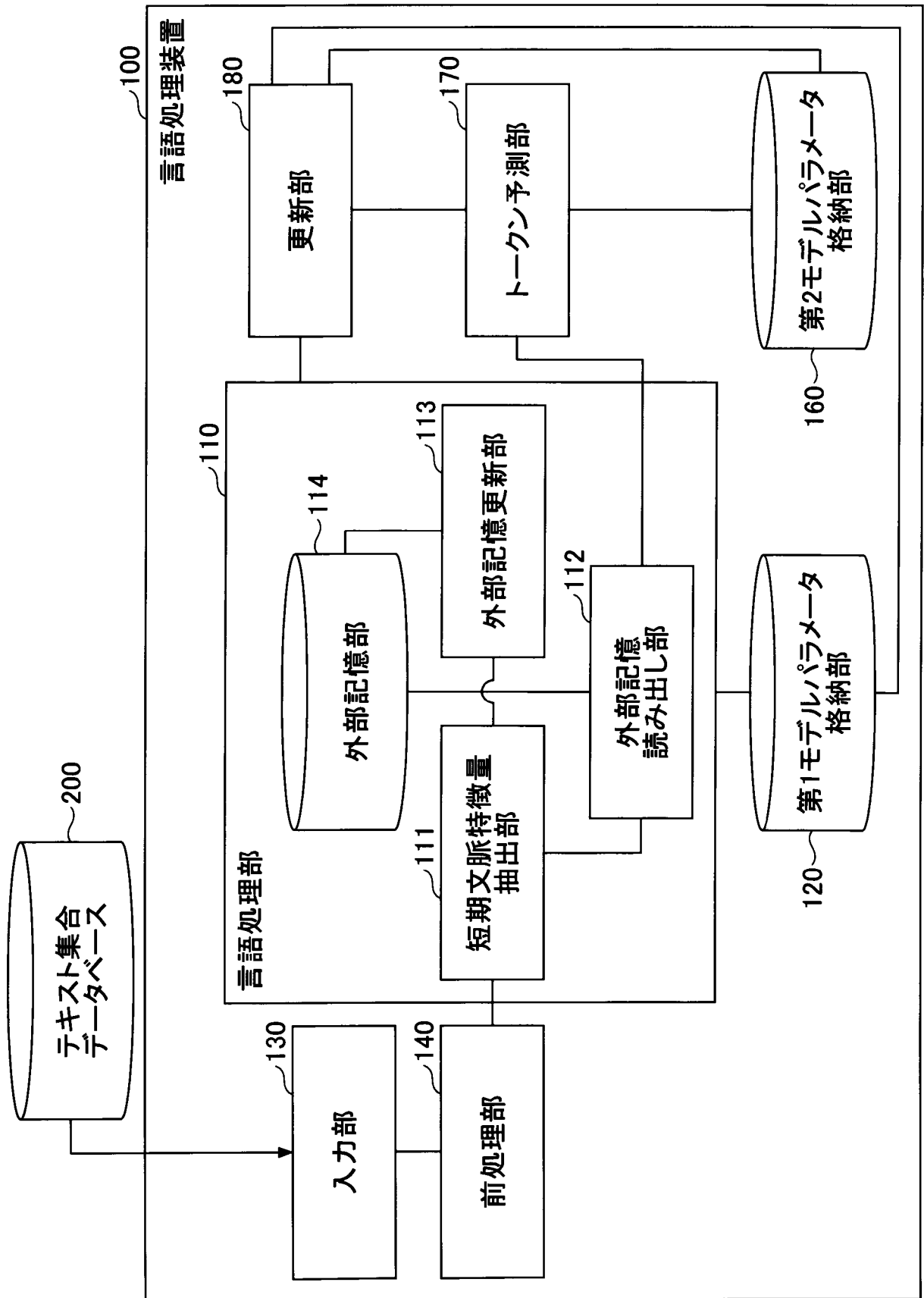
[図3]



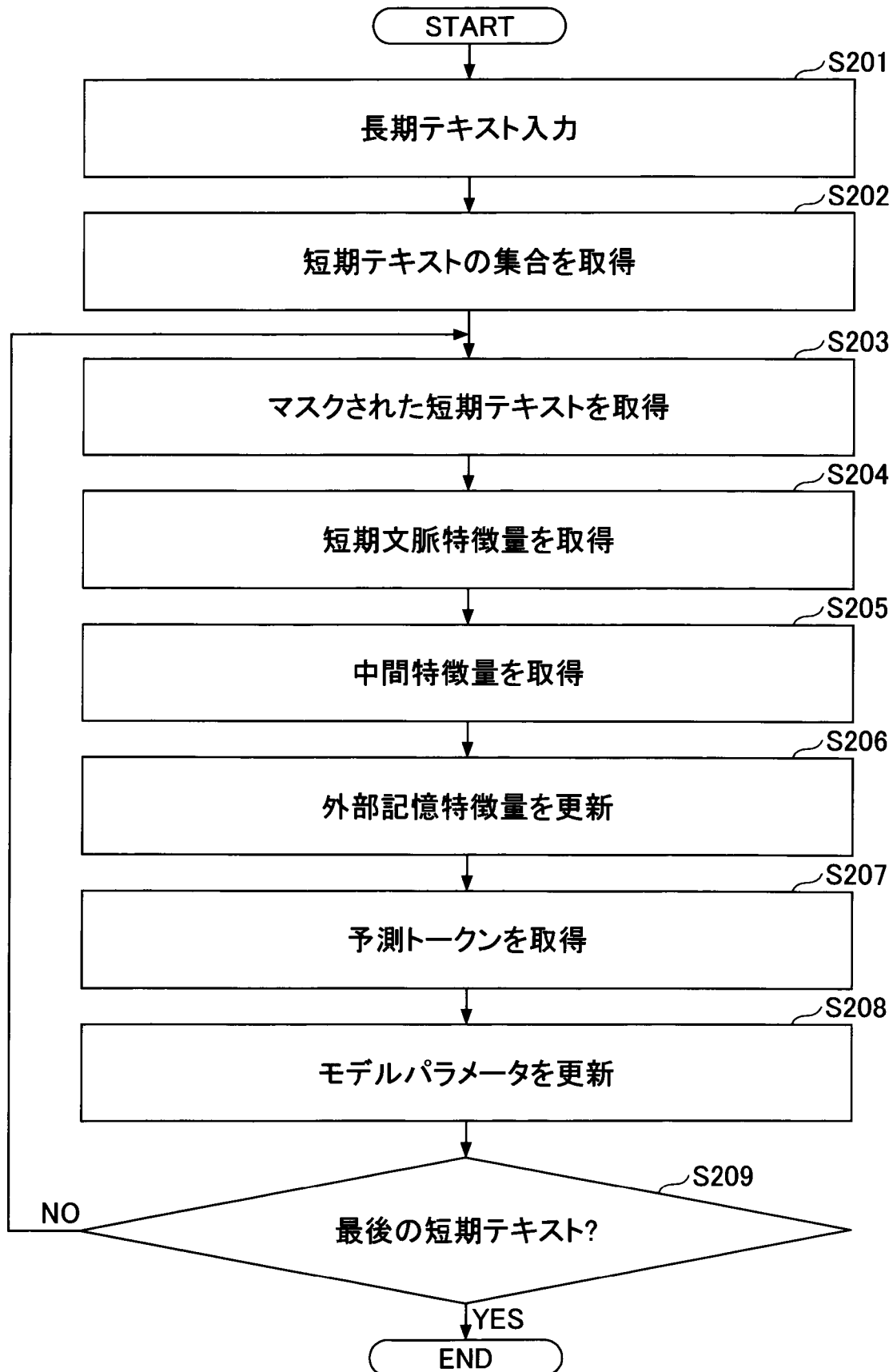
[図4]



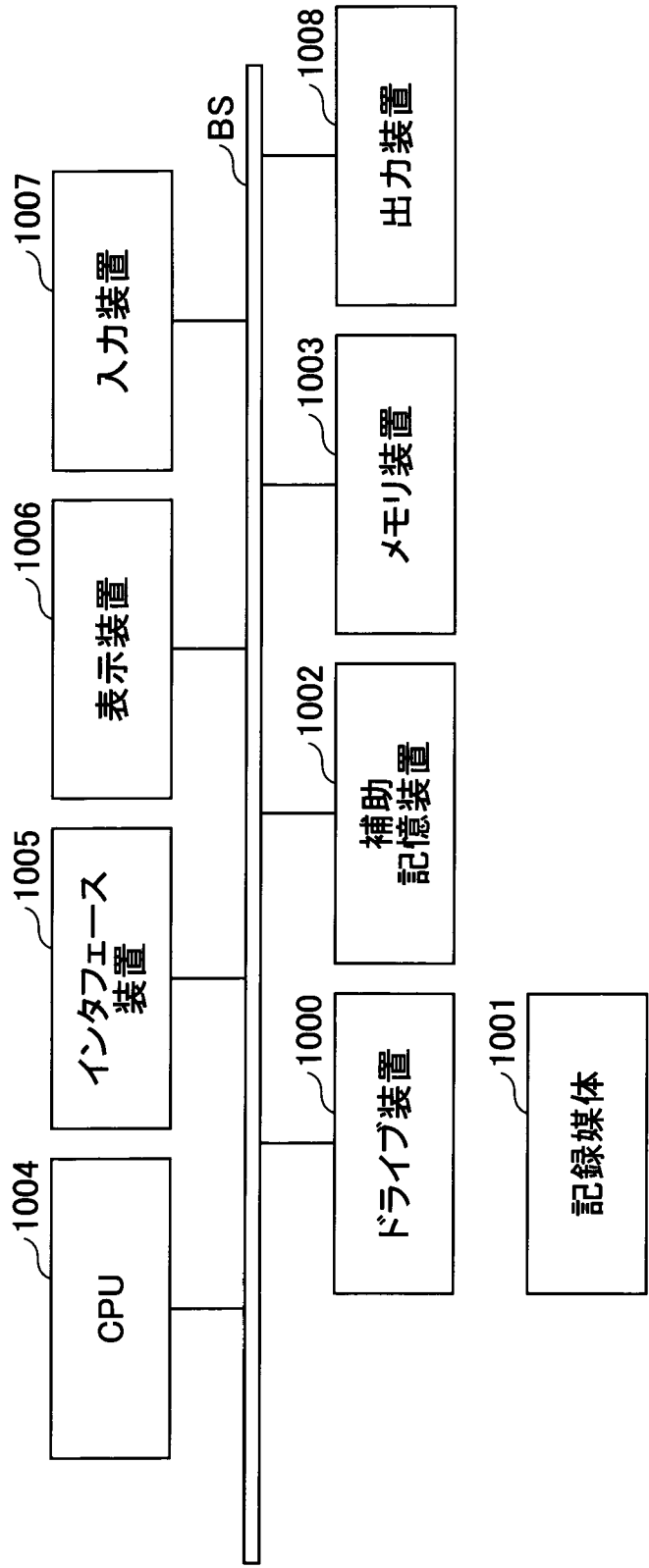
[図5]



[図6]



[図7]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2020/010579

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl. G06F16/383 (2019.01) i, G06N20/00 (2019.01) i, G06F40/216 (2020.01) i, G06F40/30 (2020.01) i

FI: G06F40/30, G06F40/216, G06F16/383, G06N20/00 160

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int. Cl. G06F16/383, G06N20/00, G06F40/216, G06F40/30

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan 1922-1996

Published unexamined utility model applications of Japan 1971-2020

Registered utility model specifications of Japan 1996-2020

Published registered utility model applications of Japan 1994-2020

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y A	US 2012/0150532 A1 (AT & T INTELLECTUAL PROPERTY, L.P.) 14 June 2012, paragraphs [0009], [0010], [0042], [0061], paragraphs [0009], [0010], [0042], [0061]	1, 3-7 2
Y A	US 2020/0073947 A1 (MMT SRL, TRENTO (IT)) 05 March 2020, paragraphs [0002], [0041], paragraphs [0002], [0041]	1, 3-7 2
Y A	US 2015/0186361 A1 (KABUSHIKI KAISHA TOSHIBA) 02 July 2015, paragraphs [0003], [0004], paragraphs [0003], [0004]	1, 3-7 2



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search
17.09.2020

Date of mailing of the international search report
29.09.2020

Name and mailing address of the ISA/
Japan Patent Office
3-4-3, Kasumigaseki, Chiyoda-ku,
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2020/010579

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y A	DEVLIN, Jacob et.al., BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, [online], [retrieved on 17 September 2020], Internet: 24 May 2019, pp. 1-16, URL: https://arxiv.org/pdf/1810.04805.pdf , in particular, p. 4, left column, line 38 to right column, line 29, entire text	3, 5, 7 1, 2, 4, 6
A	田中裕隆外 4 名. BERT を利用した文書の特徴ベクトルの作成. 情報処理学会研究報告自然言語処理 (NL). 2019-NL-243 [online], 27 November 2019, pp. 1-6, entire text, (TANAKA, Hirotaka et al. Construction of document feature vectors using BERT. IPSJ SIG Technical Report (NL).)	1-7

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/JP2020/010579

Patent Documents referred to in the Report	Publication Date	Patent Family	Publication Date
US 2012/0150532 A1	14.06.2012	(Family: none)	
US 2020/0073947 A1	05.03.2020	(Family: none)	
US 2015/0186361 A1	02.07.2015	CN 104750687 A	

A. 発明の属する分野の分類（国際特許分類（IPC）） G06F 16/383(2019.01)i; G06N 20/00(2019.01)i; G06F 40/216(2020.01)i; G06F 40/30(2020.01)i FI: G06F40/30; G06F40/216; G06F16/383; G06N20/00 160														
B. 調査を行った分野														
調査を行った最小限資料（国際特許分類（IPC）） G06F16/383; G06N20/00; G06F40/216; G06F40/30														
最小限資料以外の資料で調査を行った分野に含まれるもの <table border="0"> <tr> <td>日本国実用新案公報</td> <td>1922-1996年</td> </tr> <tr> <td>日本国公開実用新案公報</td> <td>1971-2020年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996-2020年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994-2020年</td> </tr> </table>			日本国実用新案公報	1922-1996年	日本国公開実用新案公報	1971-2020年	日本国実用新案登録公報	1996-2020年	日本国登録実用新案公報	1994-2020年				
日本国実用新案公報	1922-1996年													
日本国公開実用新案公報	1971-2020年													
日本国実用新案登録公報	1996-2020年													
日本国登録実用新案公報	1994-2020年													
国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）														
C. 関連すると認められる文献														
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号												
Y	US 2012/0150532 A1 (AT&T INTELLECTUAL PROPERTY, L.P.) 14.06.2012 (2012-06-14)	1,3-7												
A	段落0009-0010, 0042, 0061	2												
Y	US 2020/0073947 A1 (MMT SRL, TRENTO(IT)) 05.03.2020 (2020-03-05)	1,3-7												
A	段落0002, 0041	2												
Y	US 2015/0186361 A1 (KABUSHIKI KAISHA TOSHIBA) 02.07.2015 (2015-07-02)	1,3-7												
A	段落0003-0004	2												
Y	DEVLIN Jacob et.al., BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, [オンライン], [検索日 2020.09.17], インターネット: , 2019.05.24, pp.1-16, URL: https://arxiv.org/pdf/1810.04805.pdf	3,5,7												
A	特に、第4頁左欄第38行-同頁右欄第29行 全文	1,2,4,6												
☒ C欄の続きにも文献が列挙されている。 ☒ パテントファミリーに関する別紙を参照。														
<table border="0"> <tr> <td>* 引用文献のカテゴリー</td> <td>"T" 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの</td> </tr> <tr> <td>"A" 特に関連のある文献ではなく、一般的な技術水準を示すもの</td> <td>"X" 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの</td> </tr> <tr> <td>"E" 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの</td> <td>"Y" 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの</td> </tr> <tr> <td>"L" 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）</td> <td>"&" 同一パテントファミリー文献</td> </tr> <tr> <td>"O" 口頭による開示、使用、展示等に言及する文献</td> <td></td> </tr> <tr> <td>"P" 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献</td> <td></td> </tr> </table>			* 引用文献のカテゴリー	"T" 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの	"A" 特に関連のある文献ではなく、一般的な技術水準を示すもの	"X" 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの	"E" 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの	"Y" 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの	"L" 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）	"&" 同一パテントファミリー文献	"O" 口頭による開示、使用、展示等に言及する文献		"P" 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献	
* 引用文献のカテゴリー	"T" 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの													
"A" 特に関連のある文献ではなく、一般的な技術水準を示すもの	"X" 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの													
"E" 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの	"Y" 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの													
"L" 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）	"&" 同一パテントファミリー文献													
"O" 口頭による開示、使用、展示等に言及する文献														
"P" 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献														
国際調査を完了した日 17.09.2020	国際調査報告の発送日 29.09.2020													
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号	権限のある職員（特許庁審査官） 長 由紀子 5N 4233 電話番号 03-3581-1101 内線 3586													

C. 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	田中 裕隆 外4名, BERTを利用した文書の特徴ベクトルの作成, 情報処理学会 研究報告 自然言語処理 (NL) 2019-NL-243 [online], 2019.11.27, pp.1-6 全文	1-7

国際調査報告
パテントファミリーに関する情報

国際出願番号

PCT/JP2020/010579

引用文献			公表日	パテントファミリー文献			公表日
US	2012/0150532	A1	14.06.2012	(ファミリーなし)			
US	2020/0073947	A1	05.03.2020	(ファミリーなし)			
US	2015/0186361	A1	02.07.2015	CN	104750687	A	