

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7700862号
(P7700862)

(45)発行日 令和7年7月1日(2025.7.1)

(24)登録日 令和7年6月23日(2025.6.23)

(51)国際特許分類	F I			
G 0 6 F 40/279 (2020.01)	G 0 6 F	40/279		
G 0 6 F 40/216 (2020.01)	G 0 6 F	40/216		
G 0 6 F 16/90 (2019.01)	G 0 6 F	16/90	1 0 0	

請求項の数 6 (全12頁)

(21)出願番号	特願2023-543588(P2023-543588)	(73)特許権者	000004226 日本電信電話株式会社 東京都千代田区大手町一丁目5番1号
(86)(22)出願日	令和3年8月26日(2021.8.26)	(74)代理人	110004381 弁理士法人 I T O H
(86)国際出願番号	PCT/JP2021/031420	(74)代理人	100107766 弁理士 伊東 忠重
(87)国際公開番号	WO2023/026444	(74)代理人	100070150 弁理士 伊東 忠彦
(87)国際公開日	令和5年3月2日(2023.3.2)	(74)代理人	100124844 弁理士 石原 隆治
審査請求日	令和5年11月20日(2023.11.20)	(72)発明者	斉藤 いつみ 東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内
		(72)発明者	西田 京介

最終頁に続く

(54)【発明の名称】 要約学習支援装置、要約学習支援方法及びプログラム

(57)【特許請求の範囲】

【請求項1】

複数の文字列について、所定のモデルに基づいて第1の文書の要約の際に追加される入力パラメータとしての適切さを表すスコアを計算する計算部と、

前記スコアに基づいて、前記複数の文字列の中から一部の文字列群を、文書の要約を生成する要約生成モデルの学習データを構成する前記入力パラメータとして選択する選択部と、

を有し、

前記スコアは、文書の本文と前記文書の標題を構成する文字列群との対応関係を学習済みのモデルに対して前記第1の文書の要約である第2の文書を入力した場合に当該モデルが出力候補の文字列の中から出力対象の文字列を選択するために前記出力候補の文字列ごとに計算するスコアである、

10

ことを特徴とする要約学習支援装置。

【請求項2】

前記スコアは、前記第1の文書と前記第1の文書の要約である第2の文書とのうちのいずれか一方又は双方を含む第3の文書を構成する複数の文字列のそれぞれについての前記第3の文書における重要度である、

ことを特徴とする請求項1記載の要約学習支援装置。

【請求項3】

前記第1の文書及び前記文字列群と、前記第1の文書の要約である第2の文書とを含む

20

学習データを用いて、前記要約生成モデルを学習する学習部、
を有することを特徴とする請求項 1 又は 2 記載の要約学習支援装置。

【請求項 4】

前記学習部により学習された前記要約生成モデルに対して、或る文書と前記或る文書の要約に関する文字列とを入力して、前記或る文書の要約を生成する要約部、
を有することを特徴とする請求項 3 記載の要約学習支援装置。

【請求項 5】

複数の文字列について、所定のモデルに基づいて第 1 の文書の要約の際に追加される入力パラメータとしての適切さを表すスコアを計算する計算手順と、

前記スコアに基づいて、前記複数の文字列の中から一部の文字列群を、文書の要約を生成する要約生成モデルの学習データを構成する前記入力パラメータとして選択する選択手順と、

をコンピュータが実行し、

前記スコアは、文書の本文と前記文書の標題を構成する文字列群との対応関係を学習済みのモデルに対して前記第 1 の文書の要約である第 2 の文書を入力した場合に当該モデルが出力候補の文字列の中から出力対象の文字列を選択するために前記出力候補の文字列ごとに計算するスコアである、

ことを特徴とする要約学習支援方法。

【請求項 6】

請求項 1 乃至 4 いずれか一項記載の要約学習支援装置としてコンピュータを機能させることを特徴とするプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、要約学習支援装置、要約学習支援方法及びプログラムに関する。

【背景技術】

【0002】

ニューラルネットワークを用いて要約文を生成するモデルの学習データとして、要約対象のソーステキストと正しい要約結果である要約データとのペアが一般的である。

【0003】

一方で、ソーステキスト以外の入力パラメータ（以下、「クエリ」という。）が必要とされるモデルが有る（例えば、非特許文献 1）。斯かるモデルによれば、クエリに即した要約文を生成することができる。斯かるモデルは、ソーステキスト、クエリ及び要約データ等のパラメータの組が学習データとされる。

【0004】

他方において、要約文の生成方法には、抽出型と生成型とが有る。抽出型とは、ソーステキストに含まれている一部分がそのまま抽出される方法である。生成型とは、ソーステキストに含まれる単語等に基づいて、要約データが生成される方法である。以下、入力としてクエリを必要とし、生成型によって要約データを生成するモデルを「クエリ依存生成型モデル」という。

【先行技術文献】

【非特許文献】

【0005】

【文献】 Gonc , alo M. Correia , Andre F. T. Martins、 A Simple and Effective Approach to Automatic Post-Editing with Transfer Learning、 Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 3050-3056、 July 28 August 2, 2019.

【発明の概要】

【発明が解決しようとする課題】

【0006】

10

20

30

40

50

ソーステキストと要約データとのペアで構成される学習データは多数存在するが、クエリ依存生成型モデルを学習するため、ソーステキスト以外の追加の入力パラメータを含む学習データは、不十分である。

【0007】

本発明は、上記の点に鑑みてなされたものであって、追加の入力パラメータが必要とされる要約の学習を効率化することを目的とする。

【課題を解決するための手段】

【0008】

そこで上記課題を解決するため、要約学習支援装置は、複数の文字列について、所定のモデルに基づいて第1の文書の要約の際に追加される入力パラメータとしての適切さを表すスコアを計算する計算部と、前記スコアに基づいて、前記複数の文字列の中から一部の文字列群を、文書の要約を生成する要約生成モデルの学習データを構成する前記入力パラメータとして選択する選択部と、を有し、前記スコアは、文書の本文と前記文書の標題を構成する文字列群との対応関係を学習済みのモデルに対して前記第1の文書の要約である第2の文書を入力した場合に当該モデルが出力候補の文字列の中から出力対象の文字列を選択するために前記出力候補の文字列ごとに計算するスコアである。

10

【発明の効果】

【0009】

追加の入力パラメータが必要とされる要約の学習を効率化することができる。

【図面の簡単な説明】

20

【0010】

【図1】第1の実施の形態における要約生成装置10のハードウェア構成例を示す図である。

【図2】第1の実施の形態における要約生成装置10の機能構成例を示す図である。

【図3】第1の実施の形態におけるクエリ有り学習データ生成部11の構成例を示す図である。

【図4】第1の実施の形態におけるクエリ有り学習データの生成処理の処理手順の一例を説明するためのフローチャートである。

【図5】第2の実施の形態におけるクエリ有り学習データ生成部11の構成例を示す図である。

30

【図6】第2の実施の形態におけるクエリ有り学習データの生成処理の処理手順の一例を説明するためのフローチャートである。

【図7】第3の実施の形態における要約生成モデルの学習及び要約の生成を説明するための図である。

【図8】第4の実施の形態における要約生成モデルの学習及び要約の生成を説明するための図である。

【発明を実施するための形態】

【0011】

以下、図面に基づいて本発明の実施の形態を説明する。図1は、第1の実施の形態における要約生成装置10のハードウェア構成例を示す図である。図1の要約生成装置10は、それぞれバスBで相互に接続されているドライブ装置100、補助記憶装置102、メモリ装置103、プロセッサ104、及びインタフェース装置105等を有する。

40

【0012】

要約生成装置10での処理を実現するプログラムは、CD-ROM等の記録媒体101によって提供される。プログラムを記憶した記録媒体101がドライブ装置100にセットされると、プログラムが記録媒体101からドライブ装置100を介して補助記憶装置102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101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102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。

50

【 0 0 1 3 】

メモリ装置 1 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 2 からプログラムを読み出して格納する。プロセッサ 1 0 4 は、CPU 若しくは GPU (Graphics Processing Unit)、又は CPU 及び GPU であり、メモリ装置 1 0 3 に格納されたプログラムに従って要約生成装置 1 0 に係る機能を実行する。インタフェース装置 1 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。

【 0 0 1 4 】

図 2 は、第 1 の実施の形態における要約生成装置 1 0 の機能構成例を示す図である。図 2 において、要約生成装置 1 0 は、クエリ有り学習データ生成部 1 1、要約学習部 1 2 及び要約部 1 3 を有する。これら各部は、要約生成装置 1 0 にインストールされた 1 以上のプログラムが、プロセッサ 1 0 4 に実行させる処理により実現される。

10

【 0 0 1 5 】

クエリ有り学習データ生成部 1 1 は、入力として与えられるクエリ無し学習データ群に含まれる各クエリ無し学習データに基づいて、クエリ有り学習データを生成する。1 つのクエリ無し学習データに対して 1 つのクエリ有り学習データが生成される。したがって、複数のクエリ無し学習データの集合であるクエリ無し学習データ群に対して、複数のクエリ有り学習データの集合であるクエリ有り学習データ群が生成される。クエリ無し学習データ及びクエリ有り学習データは、いずれも、文書の要約を生成するニューラルネットワーク等のモデル(以下、「要約生成モデル」という。)の学習データとして用いられるデータをいう。クエリ無し学習データは、クエリを構成要素として含まない点において、クエリ有り学習データと異なる。クエリとは、要約に関する追加情報として、要約対象の文書と共に要約生成モデルに入力されるテキスト(文字列)をいう。例えば、要約の焦点がクエリとされてもよい。

20

【 0 0 1 6 】

クエリ無し学習データは、{ソーステキスト, 要約テキスト}の 2 つのテキストデータの組によって構成される学習データである。ソーステキストとは、要約対象の文書のテキストデータをいう。要約テキストとは、ソーステキストが要約された結果の正解を示すテキストデータをいう。

【 0 0 1 7 】

一方、クエリ有り学習データは、{ソーステキスト, クエリ, 要約テキスト}の 3 つのテキストデータの組によって構成される学習データである。

30

【 0 0 1 8 】

要約学習部 1 2 は、クエリ有り学習データを用いて要約生成モデルの学習を行う。

【 0 0 1 9 】

要約部 1 3 は、要約対象のソーステキスト及び当該ソーステキストに対するクエリ等の入力を受け付けると、当該ソーステキスト及び当該クエリを学習済みの要約生成モデルへ入力することで、当該ソーステキストに対する当該クエリに応じた要約を要約生成モデルに生成させる。

【 0 0 2 0 】

クエリ有り学習データ生成部 1 1 について更に詳しく説明する。図 3 は、第 1 の実施の形態におけるクエリ有り学習データ生成部 1 1 の構成例を示す図である。図 3 において、クエリ有り学習データ生成部 1 1 は、重要度計算部 1 1 1、クエリ選択部 1 1 2 及びクエリ追加部 1 1 3 を有する。これら各部の機能については、図 4 を用いて詳細に説明する。

40

【 0 0 2 1 】

図 4 は、第 1 の実施の形態におけるクエリ有り学習データの生成処理の処理手順の一例を説明するためのフローチャートである。

【 0 0 2 2 】

ステップ S 1 0 1 において、重要度計算部 1 1 1 は、クエリ無し学習データ群に含まれるクエリ無し学習データ(ソーステキスト及び要約テキストの組)ごとに、クエリの候補となる文字列の抽出元とする文書(以下、「抽出元文書」という。)を生成する。したが

50

って、N個のクエリ無し学習データからN個の抽出元文書が生成される。

【0023】

例えば、重要度計算部111は、クエリ無し学習データの以下の(a)~(d)のいずれかを、当該クエリ無し学習データに基づく抽出元文書として生成する。

(a) ソーステキストと要約テキストを結合した文書(ソーステキスト及び要約テキストの双方を含む文書)

(b) 要約テキストのみ

(c) ソーステキストのみ

(d) (a)~(c)のいずれかと、その他の付属情報テキスト(例えば、ソーステキストのタイトルなど)を結合した文書

10

続いて、重要度計算部111は、所定のモデルに基づいて、各抽出元文書を構成する所定単位の各文字列(例えば、単語)について、文書の要約の際に用いるクエリ(追加される入力パラメータ)としての適切さを表すスコアの一例として、これらの抽出元文書群における重要度を算出する(S102)。例えば、重要度計算部111は、所定のモデルとして、TF-IDFの計算モデルを用いる。この場合、重要度計算部111は、各単語のTF-IDFを重要度として算出する。文書群に含まれる各単語のTF-IDFの算出は、公知の方法を用いて行うことができる。なお、本実施の形態において、入力パラメータにおける「パラメータ」とは、例えば、ニューラルネットワーク等のモデルの学習用パラメータとは明確に区別される。入力パラメータは、モデルに対する入力として与えられるデータであるのに対し、学習用パラメータは、モデルの学習に応じて値が変化するデータ

20

【0024】

続いて、クエリ選択部112は、抽出元文書ごとに、当該抽出元文書を構成する所定単位の文字列(単語)の中から重要度の降順にK個の文字列を、当該抽出元文書に対応するクエリ無し学習データに対応するクエリとして選択する(S103)。なお、Kの値(K>=0)は、抽出元文書ごとにランダムに選択されてもよいし、全ての抽出元文書に対して同じであってもよい。また、各抽出元文書からのクエリの選択に際し、クエリ選択部112は、当該抽出元文書の要約テキストに含まれる単語のみをクエリとして選択するようにしてもよい。そうすることで、要約生成モデルについて、指定されたクエリを要約中に含むような学習をしやすくすることができる。

30

【0025】

続いて、クエリ追加部113は、クエリ無し学習データごとに、当該クエリ無し学習データに基づく抽出元文書から選択されたK個の単語を、クエリとして当該クエリ無し学習データに追加することで、クエリ有り学習データを生成する(S104)。したがって、生成されるクエリ有り学習データは、クエリ無し学習データが含んでいたソーステキスト及び要約テキストと、当該クエリ無し学習データから抽出されたK個のクエリ(クエリ列)とを含むことになる。

【0026】

上述したように、第1の実施の形態によれば、クエリを含まない学習データから、疑似的なクエリを生成することができる。したがって、追加の入力パラメータが必要とされる要約の学習を効率化することができる。

40

【0027】

次に、第2の実施の形態について説明する。第2の実施の形態では第1の実施の形態と異なる点について説明する。第2の実施の形態において特に言及されない点については、第1の実施の形態と同様でもよい。

【0028】

第2の実施の形態では、クエリ有り学習データ生成部11の構成及びクエリ有り学習データ生成部11が実行する処理手順が第1の実施の形態と異なる。

【0029】

50

図5は、第2の実施の形態におけるクエリ有り学習データ生成部11の構成例を示す図である。図5中、図3と同一又は対応する部分には、同一符号を付している。第2の実施の形態において、クエリ有り学習データ生成部11は、重要度計算部111の代わりにクエリ生成モデル学習部114及びクエリ候補生成部115を有する。クエリ生成モデル学習部114は、クエリ無し学習データから1以上のクエリを生成するモデル（以下、「クエリ生成モデル」という。）を学習する。クエリ生成モデルは、例えば、ニューラルネットワーク等により構成される。クエリ生成モデル学習部114は、クエリ生成モデルの学習データの元となる学習用文書群を入力とする。学習用文書群は、複数の学習用文書の集合をいう。学習用文書とは、wikipedia等のようにインターネットにおいて公開されている百科事典や、新聞などのように、標題（見出し）と本文とを含むテキスト形式の文書データをいう。

10

【0030】

クエリ候補生成部115は、学習済みのクエリ生成モデルに基づいて、クエリの候補を生成（出力）する。

【0031】

図6は、第2の実施の形態におけるクエリ有り学習データの生成処理の処理手順の一例を説明するためのフローチャートである。

【0032】

ステップS201において、クエリ生成モデル学習部114は、学習用文書群に含まれる学習用文書ごとに、クエリ生成モデルの学習データを生成する。具体的には、クエリ生成モデル学習部114は、各学習用文書の標題を、所定単位の文字列（例えば、単語）に分解（分割）する。したがって、例えば、学習用文書ごとに、標題を構成する単語列（以下、単に「単語列」という。）が生成される。この際、クエリ生成モデル学習部114は、ストップワードを削除した単語列を生成するようにしてもよい。クエリ生成モデル学習部114は、学習用文書ごとに、当該学習用文書の本文（パラグラフテキスト）と、当該本文に対応する標題から生成された単語列との組を学習データとして生成する。

20

【0033】

続いて、クエリ生成モデル学習部114は、ステップS201において生成された学習データ群を用いて、クエリ生成モデルの学習を行う（S202）。具体的には、クエリ生成モデル学習部114は、各学習データの本文を入力とし、標題の単語列を出力とした場合における、当該本文と当該単語列との対応関係をクエリ生成モデルに学習させる。したがって、クエリ生成モデルは、或る文書の本文を入力すると、当該文書の標題に関連する単語列を出力するように学習される。なお、クエリ生成モデルは、例えば、公知のencoder-decoderモデルによって構成されてもよいし、公知の他の文生成モデルによって構成されてもよい。

30

【0034】

続いて、クエリ候補生成部115は、クエリ無し学習データごとに、当該クエリ無し学習データの要約テキストを学習済みのクエリ生成モデルに入力して、当該クエリ生成モデルが出力する文字列群（単語列）を、当該クエリ無し学習データに対応するクエリ候補列として生成する（S203）。

40

【0035】

この際、クエリ生成モデルがencoder-decoderモデルであれば、クエリ生成モデルは、クエリ無し学習データの入力に応じ、単語列を構成する各単語を逐次的に出力する。単語の逐次的な出力において、クエリ生成モデルは、自らの語彙（クエリ生成モデルの出力候補の単語の集合）を構成するD個の各単語について、出力候補の中から出力対象を選択するためのスコアを計算し、スコアが最大である単語を出力する。第2の実施の形態では、当該スコアが、文書の要約の際に用いるクエリ（追加される入力パラメータ）としての適切さを表すスコアの一例に相当する。

【0036】

続いて、クエリ選択部112は、クエリ無し学習データごとに、当該クエリ無し学習デ

50

ータについてクエリ候補生成部 115 が生成したクエリ候補列の中からクエリとして利用する 1 以上の単語（クエリ列）を選択する（S204）。この際、クエリ選択部 112 は、クエリ候補列の全てをクエリ列として選択してもよいし、クエリ候補列の一部をクエリ列として選択してもよい。クエリ候補列の一部をクエリ列として選択する場合、クエリ選択部 112 は、クエリ候補列の先頭から K 番目までの単語をクエリ列として選択してもよい。すなわち、クエリ生成モデルが逐次的に行う単語の出力のうち、K 番目までの単語がクエリとして選択されてもよい。又は、ステップ S203 において、クエリ生成モデルからの逐次的な単語の出力回数が K 回に抑制されてもよい。この場合、クエリ候補列は、K 個の単語から構成されることになる。したがって、この場合、ステップ S204 では、クエリ候補列の全部がクエリ列として選択されればよい。

10

【0037】

続いて、クエリ追加部 113 は、クエリ無し学習データごとに、当該クエリ無し学習データについて選択されたクエリ列を当該クエリ無し学習データに追加することで、クエリ有り学習データを生成する（S205）。

【0038】

上述したように、第 2 の実施の形態によれば、第 1 の実施の形態と同様の効果を得ることができる。

【0039】

次に、第 3 の実施の形態として、クエリ有り学習データを用いた要約生成モデルの学習、及び学習済みの要約生成モデルを用いた要約の生成に関する第 1 の例について説明する。なお、第 3 の実施の形態は、第 1 の実施の形態及び第 2 の実施の形態のいずれに対しても適用可能である。

20

【0040】

図 7 は、第 3 の実施の形態における要約生成モデルの学習及び要約の生成を説明するための図である。図 7 において、要約部 13 は、内容選択部 131、エンコーダ 132 及びデコーダ 133 を含む。これら各部が要約生成モデルを構成する。

【0041】

要約生成モデルの学習時において、要約学習部 12 は、要約部 13 に対して、クエリ有り学習データ群に含まれる学習データ（ソーステキスト、クエリ列、要約テキスト）ごとに、当該学習データのソーステキスト及びクエリ列を入力する。

30

【0042】

内容選択部 131 は、当該ソーステキスト及び当該クエリ列を結合したテキスト（以下、「結合テキスト」という。）を構成する文字列（例えば、単語）ごとに、重要度を算出するモデル（例えば、ニューラルネットワーク）である。内容選択部 131 は、BERT や MASS などの事前学習済モデルを *fine tune* することで構成されてもよい。なお、BERT については、例えば、「<https://arxiv.org/abs/1810.04805>」等に詳しい。また、MASS については、例えば、「<https://arxiv.org/abs/1905.02450>」等に詳しい。

【0043】

内容選択部 131 は、結合テキストの中から重要度の降順に N 個の単語列（重要語列）を抽出し、当該重要語列、入力として与えられたソーステキスト及びクエリ列をエンコーダ 132 へ入力する。この際、内容選択部 131 は、クエリ列、重要語列及びソーステキストを「クエリ列 [SEP] 重要語列 [SEP] ソーステキスト」のように、[SEP] などの特殊トークンで結合する。なお、N の値は、クエリ等と共に内容選択部 131 に対する入力とされてもよい。

40

【0044】

エンコーダ 132 及びデコーダ 133 は、例えば、BERT 又は MASS 等の公知の *encoder-decoder* モデル（ニューラルネットワーク）である。

【0045】

エンコーダ 132 は、入力されたテキストを符号化する。デコーダ 133 は、符号化結

50

果に基づいて要約テキストを生成及び出力する。

【0046】

要約学習部12は、学習データに含まれる要約テキストと、デコーダ133が出力した要約テキストとの比較に基づいて、エンコーダ132及びデコーダ133の学習用パラメータを更新する。なお、当該比較及び学習用パラメータの更新は、公知技術に基づいて行われればよい。

【0047】

学習が終了すると、要約部13は、クエリ列及び入力テキストを入力とし、要約テキストを出力とする学習済みの要約生成モデルとして機能する。

【0048】

なお、図7の要約部13は、国際公開第2021/064907号に開示された技術を用いて構成されてもよい。

【0049】

次に、第4の実施の形態として、クエリ有り学習データを用いた要約生成モデルの学習、及び学習済みの要約生成モデルを用いた要約の生成に関する第2の例について説明する。なお、第4の実施の形態は、第1の実施の形態及び第2の実施の形態のいずれに対しても適用可能である。

【0050】

図8は、第4の実施の形態における要約生成モデルの学習及び要約の生成を説明するための図である。図8中、図7と同一又は対応する部分には同一符号を付している。図8において、要約部13は、エンコーダ132及びデコーダ133を含む。これら各部が要約生成モデルを構成する。すなわち、第4の実施の形態の要約生成モデルは、内容選択部131を有さない。

【0051】

要約生成モデルの学習時において、要約学習部12は、要約部13に対して、クエリ有り学習データ群に含まれる学習データ(ソーステキスト、クエリ列、要約テキスト)ごとに、当該学習データのソーステキスト及びクエリ列を入力する。この際、要約学習部12は、クエリ列及びソーステキストを「クエリ列[SEP]ソーステキスト」のように、[SEP]などの特殊トークンで結合する。

【0052】

エンコーダ132は、入力されたテキストを符号化する。デコーダ133は、符号化結果に基づいて要約テキストを生成及び出力する。

【0053】

要約学習部12は、学習データに含まれる要約テキストと、デコーダ133が出力した要約テキストとの比較に基づいて、エンコーダ132及びデコーダ133の学習用パラメータを更新する。なお、当該比較及び学習用パラメータの更新は、公知技術に基づいて行われればよい。

【0054】

学習が終了すると、要約部13は、クエリ列及び入力テキストを入力とし、要約テキストを出力とする学習済みの要約生成モデルとして機能する。

【0055】

なお、第4及び第5の実施の形態において、要約部13は、encoder-decoderモデル以外の文生成モデルに基づいて構成されてもよい。

【0056】

以上の実施形態に関し、更に以下の付記を開示する。

【0057】

(付記項1)

メモリと、

前記メモリに接続された少なくとも1つのプロセッサと、

を含み、

10

20

30

40

50

前記プロセッサは、

複数の文字列について、所定のモデルに基づいて第1の文書の要約の際に追加される入力パラメータとしての適切さを表すスコアを計算し、

前記スコアに基づいて、前記複数の文字列の中から一部の文字列群を、文書の要約を生成する要約生成モデルの学習データを構成する前記入力パラメータとして選択する、ことを特徴とする要約学習支援装置。

【0058】

(付記項2)

複数の文字列について、所定のモデルに基づいて第1の文書の要約の際に追加される入力パラメータとしての適切さを表すスコアを計算し、

前記スコアに基づいて、前記複数の文字列の中から一部の文字列群を、文書の要約を生成する要約生成モデルの学習データを構成する前記入力パラメータとして選択する、処理をコンピュータに実行させるプログラムを記録した記録媒体。

【0059】

なお、上記各実施の形態において、要約生成装置10は、要約学習支援装置の一例である。重要度計算部111又はクエリ候補生成部115(クエリ生成モデル)は、計算部の一例である。クエリ選択部112は、選択部の一例である。要約学習部12は、学習部の一例である。

【0060】

以上、本発明の実施の形態について詳述したが、本発明は斯かる特定の実施形態に限定されるものではなく、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】

【0061】

10 要約生成装置
 11 クエリ有り学習データ生成部
 12 要約学習部
 13 要約部
 100 ドライブ装置
 101 記録媒体
 102 補助記憶装置
 103 メモリ装置
 104 プロセッサ
 105 インタフェース装置
 111 重要度計算部
 112 クエリ選択部
 113 クエリ追加部
 114 クエリ生成モデル学習部
 115 クエリ候補生成部
 131 内容選択部
 132 エンコーダ
 133 デコーダ
 B バス

10

20

30

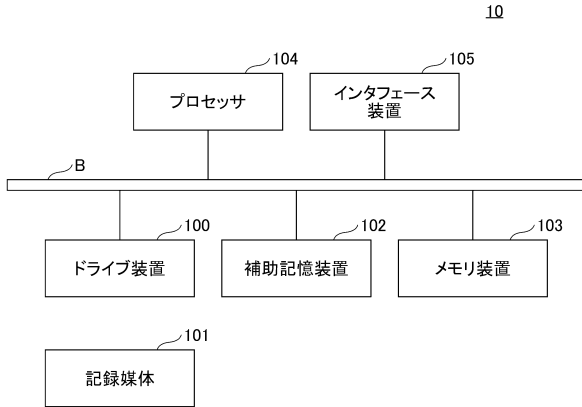
40

50

【図面】

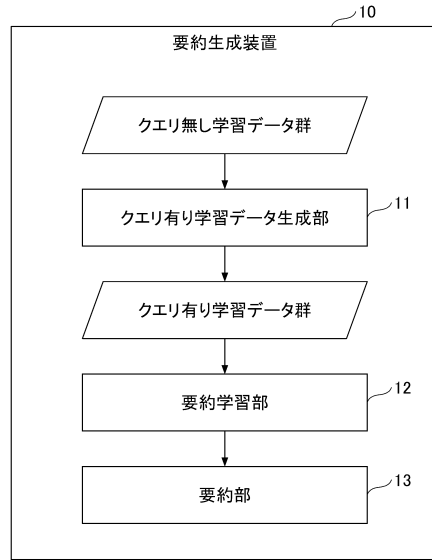
【図 1】

第1の実施の形態における要約生成装置10のハードウェア構成例を示す図



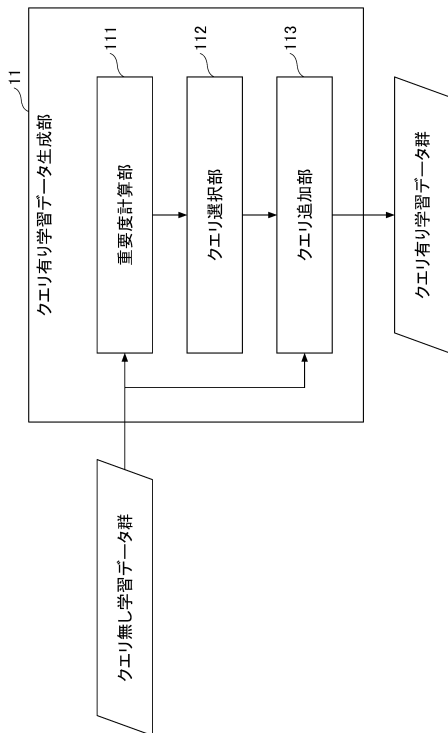
【図 2】

第1の実施の形態における要約生成装置10の機能構成例を示す図



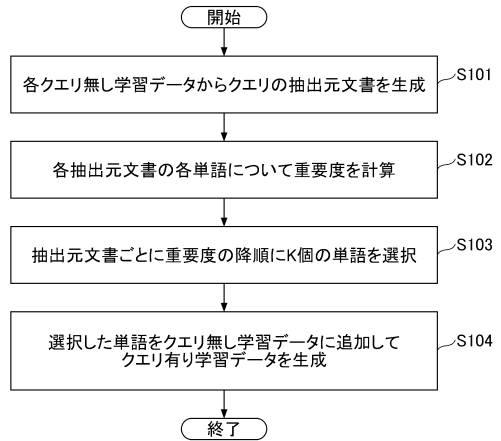
【図 3】

第1の実施の形態におけるクエリ有り学習データ生成部11の構成例を示す図



【図 4】

第1の実施の形態におけるクエリ有り学習データの生成処理の処理手順の一例を説明するためのフローチャート



10

20

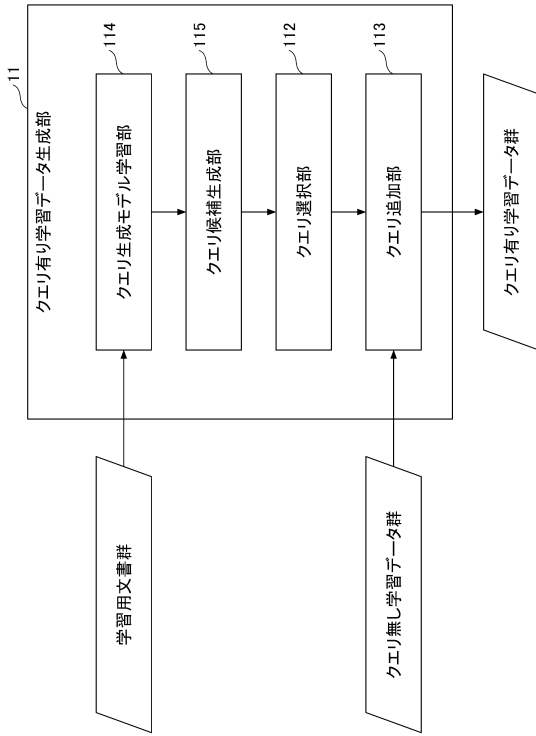
30

40

50

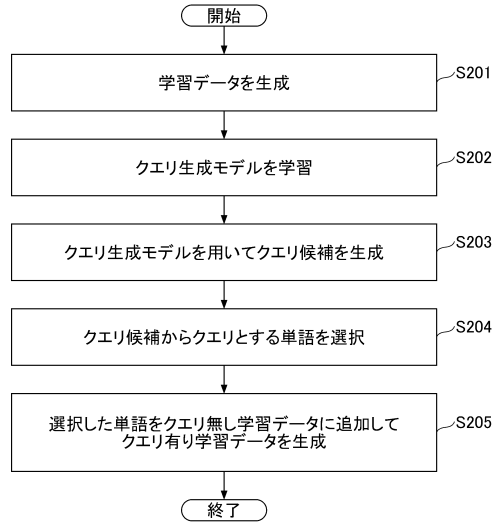
【図5】

第2の実施の形態におけるクエリ有り学習データ生成部11の構成例を示す図



【図6】

第2の実施の形態におけるクエリ有り学習データの生成処理の処理手順の一例を説明するためのフローチャート

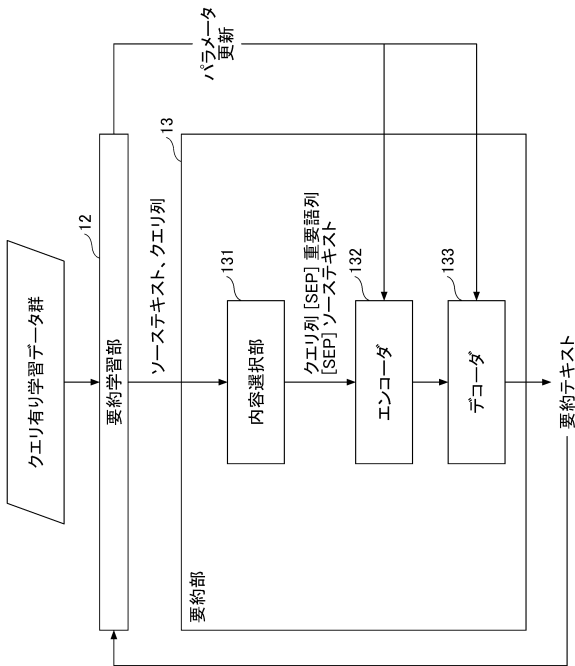


10

20

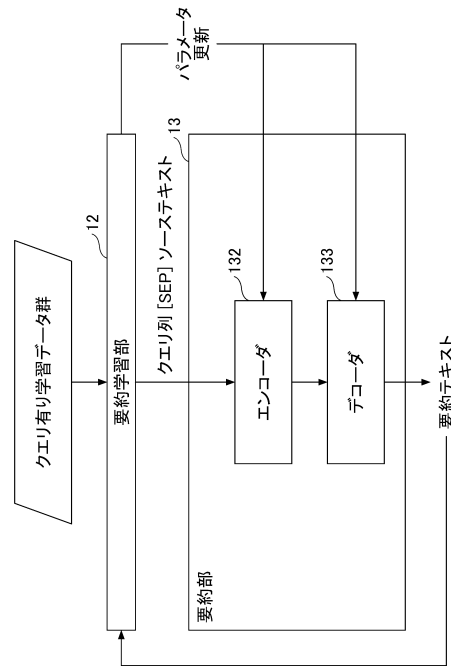
【図7】

第3の実施の形態における要約生成モデルの学習及び要約の生成を説明するための図



【図8】

第4の実施の形態における要約生成モデルの学習及び要約の生成を説明するための図



30

40

50

フロントページの続き

東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内

(72)発明者 吉田 仙

東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内

審査官 成瀬 博之

(56)参考文献 国際公開第2021/124489(WO, A1)

齊藤いつみ 他5名, クエリ・出力長を考慮可能な文書要約モデル, 言語処理学会第25回年次大会 発表論文集, 2019年03月04日, 497-500頁

廣嶋伸章 他2名, Webページのヘッドライン生成のための統計的要約, 自然言語処理, 言語処理学会, 2005年11月10日, 第12巻第6号, 113-128頁, ISSN 1340-7619

(58)調査した分野 (Int.Cl., DB名)

G 0 6 F 4 0 / 2 0 - 4 0 / 5 8

G 0 6 F 1 6 / 0 0 - 1 6 / 9 5 8