(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
2020年9月3日(03.09.2020)

WIPO | PCT

(10) 国際公開番号
WO 2020/174826 A1

(51) 国際特許分類:
G06F 16/90 (2019.01)

(21) 国際出願番号： PCT/JP2019/049385

(22) 国際出願日： 2019年12月17日(17.12.2019)

(25) 国際出願の言語： 日本語

(26) 国際公開の言語： 日本語

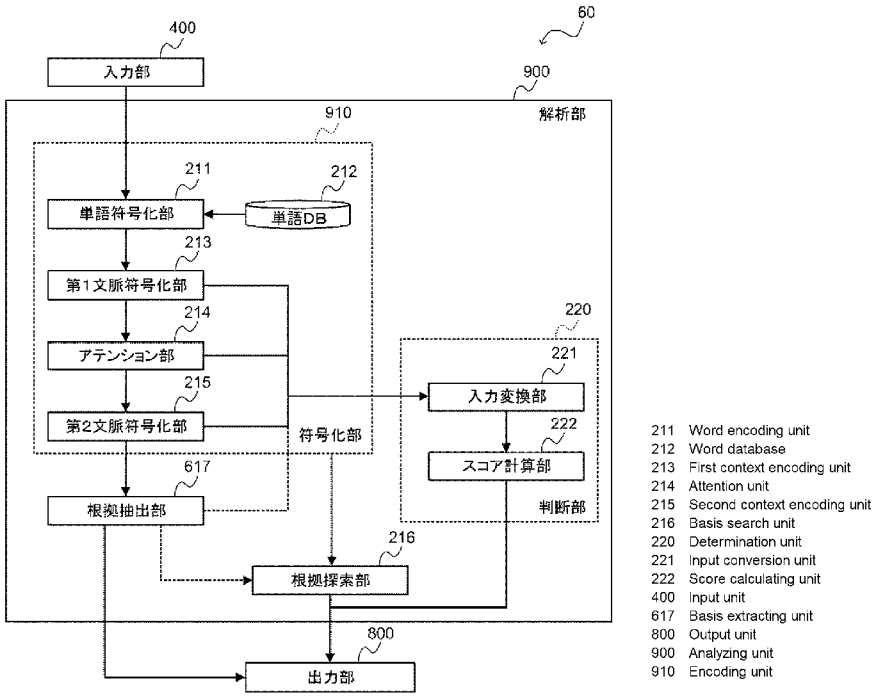
(30) 優先権データ：
特願 2019-032127 2019年2月25日(25.02.2019) JP

(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: 西田 光甫 (**NISHIDA, Kosuke**); 〒1808585 東京都武蔵野市緑町三丁目９番１１号 ＮＴＴ 知的財産センタ内 Tokyo (JP). 大塚 淳史(**OTSUKA, Atsushi**); 〒1808585 東京都武蔵野市緑町三丁目９番１１号 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三丁目９番１１号 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三丁目９番１１号 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三丁目９番１１号 ＮＴＴ 知的財産センタ内 Tokyo (JP). 斉藤 いつみ(**SAITO, Itsumi**);

(54) **Title:** ANSWER GENERATING DEVICE, ANSWER LEARNING DEVICE, ANSWER GENERATING METHOD, AND ANSWER GENERATING PROGRAM

(54) 発明の名称： 回答生成装置、回答学習装置、回答生成方法、及び回答生成プログラム



(57) **Abstract:** An encoding unit, on the basis of a passage and a question sentence that have been input, the passage being divided into a plurality of spans which are units of division of the passage, converts the input text into a vector representation series representing the meaning of the spans and the question sentence, using a previously learned encoding model for converting the input text into a vector representation series representing the meaning of the input text. A basis extracting unit, on the basis of the vector representation series, estimates a basis score of each of the spans using a

WO 2020/174826 A1

［続葉有］

〒1808585 東京都武蔵野市緑町三丁目９番１１号ＮＴＴ知的財産センタ内 Tokyo (JP).

(74) 代理人：特許業務法人太陽国際特許事務所 (TAIYO, NAKAJIMA & KATO)；〒1600022 東京都新宿区新宿４丁目３番１７号 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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：

— 国際調査報告（条約第21条(3)）

previously learned extracting model for calculating the basis score indicating the degree to which the spans are suitable as the basis for extracting the answer.

(57) 要約：符号化部が、入力された、文章の分割単位である複数のスパンに分割された文章及び質問文に基づいて、入力文を入力文の意味を表すベクトル表現系列に変換するための予め学習された符号化モデルを用いて、前記スパン及び前記質問文の意味を表すベクトル表現系列に変換する。根拠抽出部が、前記ベクトル表現系列に基づいて、前記スパンが前記回答を抽出する根拠に適している度合いを示す根拠スコアを算出するための予め学習された抽出モデルを用いて、前記スパンの各々の根拠スコアを推定する。

明　細　書

発明の名称 ：

回答生成装置、回答学習装置、回答生成方法、及び回答生成プログラム

技術分野

[0001] 本開示は、回答生成装置、回答学習装置、回答生成方法、及び回答生成プログラムに係り、特に、質問文に対して極性で回答するための回答生成装置、回答学習装置、回答生成方法、及び回答生成プログラムに関する。

背景技術

[0002] 近年、機械が文章を読み解いて質問に答える機械読解技術（例えば、ＢｉＤＡＦ（非特許文献１））が注目を集めている。機械読解の代表的なデータセットにはＳＱｕＡＤ（非特許文献２）が存在し、大規模な深層学習技術の適用が可能となっている。

[0003] ＳＱｕＡＤは１つの質問に対して１段落の文章が紐づき、文章に書いてある回答をそのまま抽出して回答とする抽出型のタスクのためのデータセットである。

先行技術文献

非特許文献

[0004] 非特許文献1：Minjoon Seo, Aniruddha Kembhavi, Ali Farhadi, Hananneh Hajishirzi, ”BI-DIRECTIONAL ATTENTION FLOW FOR MACHINE COMPREHENSION”, Published as a conference paper at ICLR, 2017.

非特許文献2：Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, Percy Liang, ”SQuAD: 100,000+ Questions for Machine Comprehension of Text”, Computer Science Department Stanford University, 2016.

発明の概要

発明が解決しようとする課題

[0005] しかし、抽出型タスクのための手法では、テキストに書いていない形式で答えを出力することができない、という問題があった。具体的には、Ｙｅｓ

又はＮｏ等の極性で回答することができる質問に対して、その極性（Ｙｅｓ又はＮｏ）で回答する、ということができない。このようなテキストに書いていない形式で答えを出力するためには、機械が文章の中から質問に関連する部分に注目するだけでなく、関連部分から質問に対する回答を判断する必要がある。

[0006] 本開示は上記の点に鑑みてなされたものであり、極性で回答することができる質問に対して、精度よく、極性で回答することができる回答生成装置、回答生成方法、及び回答生成プログラムを提供することを目的とする。

[0007] また、本開示は上記の点に鑑みてなされたものであり、極性で回答することができる質問に対して、精度よく、極性で回答するためのモデルを学習することができる回答学習装置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

[0008] 本開示の第１態様は、回答生成装置であって、入力された、文章の分割単位である複数のスパンに分割された文章及び質問文に基づいて、入力文を入力文の意味を表すベクトル表現系列に変換するための予め学習された符号化モデルを用いて、前記スパン及び前記質問文の意味を表すベクトル表現系列に変換する符号化部と、前記ベクトル表現系列に基づいて、前記回答を抽出する根拠に適しているスパンを抽出する予め学習された抽出モデルを用いて、前記回答を抽出する根拠に適しているスパンである根拠情報を推定する根拠抽出部と、前記ベクトル表現系列と、前記根拠情報とに基づいて、前記文章における前記回答の根拠となる範囲である根拠範囲を抽出する予め学習された探索モデルを用いて、前記根拠範囲を推定する根拠探索部と、前記ベクトル表現系列と、前記根拠情報とに基づいて、前記文章における前記質問文に対する回答の正解である回答種別を判断する予め学習された判断モデルを用いて、前記質問文に対する回答種別を判断する判断部と、を備えて構成される。

[0009] 本開示の第２態様は、回答生成装置であって、入力された、文章の分割単位である複数のスパンに分割された文章及び質問文に基づいて、入力文を入

力文の意味を表すベクトル表現系列に変換するための予め学習された符号化モデルを用いて、前記スパン及び前記質問文の意味を表すベクトル表現系列に変換する符号化部と、前記ベクトル表現系列に基づいて、前記スパンが前記回答を抽出する根拠に適している度合いを示す根拠スコアを算出するための予め学習された抽出モデルを用いて、前記スパンの各々の根拠スコアを推定する根拠抽出部と、を備え、前記符号化モデル及び前記抽出モデルは、前記符号化モデルを用いて変換された前記ベクトル表現系列に基づいて、前記文章における前記回答の根拠となる範囲である根拠範囲を抽出する探索モデル、及び前記ベクトル表現系列に基づいて、前記文章における前記質問文に対する回答の正解である回答種別を判断する判断モデルを更に含む複数のモデルのマルチタスク学習により予め学習されたものである。

[0010] 本開示の第３態様は、回答生成学習装置であって、文章の分割単位である複数のスパンに分割された文章と、質問文と、前記文章における前記質問文に対する回答の正解である回答種別と、前記文章における前記回答の根拠となる範囲である根拠範囲と、前記文章における前記回答の根拠となるスパンである根拠情報とを含む学習データの入力を受け付ける入力部と、複数のスパンに分割した前記文章と、前記質問文とを、入力文を入力文の意味を表すベクトル表現系列に変換するための符号化モデルを用いて、前記スパン及び前記質問文の意味を表すベクトル表現系列に変換する符号化部と、前記ベクトル表現系列に基づいて、前記根拠情報を抽出する抽出モデルを用いて、前記根拠情報を推定する根拠抽出部と、前記ベクトル表現系列と、前記学習データの前記根拠情報とに基づいて、前記根拠範囲を抽出する探索モデルを用いて、前記根拠範囲を推定する根拠探索部と、前記ベクトル表現系列と、前記学習データの前記根拠情報とに基づいて、前記質問文に対する回答種別を判断する判断モデルを用いて、前記質問文に対する回答種別を判断する判断部と、前記根拠抽出部により抽出された前記根拠情報が前記学習データの前記根拠情報と一致し、前記根拠探索部により推定された前記根拠範囲が前記学習データの前記根拠範囲と一致し、前記判断部により判断された前記回答

種別が前記学習データの前記回答種別と一致するように、前記符号化モデル、前記抽出モデル、前記探索モデル、及び前記判断モデルのパラメータを学習するパラメータ学習部と、を備えて構成される。

[0011] 本開示の第４態様は、回答生成方法であって、符号化部が、入力された、文章の分割単位である複数のスパンに分割された文章及び質問文に基づいて、入力文を入力文の意味を表すベクトル表現系列に変換するための予め学習された符号化モデルを用いて、前記スパン及び前記質問文の意味を表すベクトル表現系列に変換し、根拠抽出部が、前記ベクトル表現系列に基づいて、前記回答を抽出する根拠に適しているスパンを抽出する予め学習された抽出モデルを用いて、前記回答を抽出する根拠に適しているスパンである根拠情報を推定し、根拠探索部が、前記ベクトル表現系列と、前記根拠情報とに基づいて、前記文章における前記回答の根拠となる範囲である根拠範囲を抽出する予め学習された探索モデルを用いて、前記根拠範囲を推定し、判断部が、前記ベクトル表現系列と、前記根拠情報とに基づいて、前記文章における前記質問文に対する回答の正解である回答種別を判断する予め学習された判断モデルを用いて、前記質問文に対する回答種別を判断する。

[0012] 本開示の第５態様は、回答生成方法であって、符号化部が、入力された、文章の分割単位である複数のスパンに分割された文章及び質問文に基づいて、入力文を入力文の意味を表すベクトル表現系列に変換するための予め学習された符号化モデルを用いて、前記スパン及び前記質問文の意味を表すベクトル表現系列に変換し、根拠抽出部が、前記ベクトル表現系列に基づいて、前記スパンが前記回答を抽出する根拠に適している度合いを示す根拠スコアを算出するための予め学習された抽出モデルを用いて、前記スパンの各々の根拠スコアを推定する回答生成方法であって、前記符号化モデル及び前記抽出モデルは、前記符号化モデルを用いて変換された前記ベクトル表現系列に基づいて、前記文章における前記回答の根拠となる範囲である根拠範囲を抽出する探索モデル、及び前記ベクトル表現系列に基づいて、前記文章における前記質問文に対する回答の正解である回答種別を判断する判断モデルを更

に含む複数のモデルのマルチタスク学習により予め学習されたものである。

[0013] 本開示の第６態様は、回答生成プログラムであって、入力された、文章の分割単位である複数のスパンに分割された文章及び質問文に基づいて、入力文を入力文の意味を表すベクトル表現系列に変換するための予め学習された符号化モデルを用いて、前記スパン及び前記質問文の意味を表すベクトル表現系列に変換し、前記ベクトル表現系列に基づいて、前記回答を抽出する根拠に適しているスパンを抽出する予め学習された抽出モデルを用いて、前記回答を抽出する根拠に適しているスパンである根拠情報を推定し、前記ベクトル表現系列と、前記根拠情報とに基づいて、前記文章における前記回答の根拠となる範囲である根拠範囲を抽出する予め学習された探索モデルを用いて、前記根拠範囲を推定し、前記ベクトル表現系列と、前記根拠情報とに基づいて、前記文章における前記質問文に対する回答の正解である回答種別を判断する予め学習された判断モデルを用いて、前記質問文に対する回答種別を判断する処理をコンピュータに実行させるための回答生成プログラムである。

[0014] 本開示の第７態様は、回答生成プログラムであって、入力された、文章の分割単位である複数のスパンに分割された文章及び質問文に基づいて、入力文を入力文の意味を表すベクトル表現系列に変換するための予め学習された符号化モデルを用いて、前記スパン及び前記質問文の意味を表すベクトル表現系列に変換し、前記ベクトル表現系列に基づいて、前記スパンが前記回答を抽出する根拠に適している度合いを示す根拠スコアを算出するための予め学習された抽出モデルを用いて、前記スパンの各々の根拠スコアを推定する処理をコンピュータに実行させるための回答生成プログラムであって、前記符号化モデル及び前記抽出モデルは、前記符号化モデルを用いて変換された前記ベクトル表現系列に基づいて、前記文章における前記回答の根拠となる範囲である根拠範囲を抽出する探索モデル、及び前記ベクトル表現系列に基づいて、前記文章における前記質問文に対する回答の正解である回答種別を判断する判断モデルを更に含む複数のモデルのマルチタスク学習により予

め学習されたものである。

発明の効果

[0015] 本開示の回答生成装置、回答生成方法、及びプログラムによれば、極性で回答することができる質問に対して、精度よく、極性で回答することができる。

[0016] また、本開示の回答学習装置によれば、極性で回答することができる質問に対して、精度よく、極性で回答するためのモデルを学習することができる。

図面の簡単な説明

[0017] [図1]第１の実施形態の形態に係る回答学習装置の構成を示す機能ブロック図である。

[図2]第１の実施形態に係る回答学習装置の回答学習処理ルーチンを示すフローチャートである。

[図3]第１の実施形態に係る回答生成装置の構成を示す機能ブロック図である。

[図4]第１の実施形態に係る回答生成装置の回答生成処理ルーチンを示すフローチャートである。

[図5]第２の実施形態に係る回答学習装置の構成を示す機能ブロック図である。

[図6]第２の実施形態に係る回答学習装置の回答学習処理ルーチンを示すフローチャートである。

[図7]第２の実施形態に係る回答学習装置の根拠情報抽出処理ルーチンを示すフローチャートである。

[図8]第２の実施形態に係る回答生成装置の構成を示す機能ブロック図である。

[図9]第２の実施形態に係る回答生成装置の回答生成処理ルーチンを示すフローチャートである。

[図10]第２の実施形態に係る回答生成装置のベースラインモデルの例を示す

図である。

[図11]第２の実施形態に係る根拠抽出部の抽出モデルの構成例を示す図である。

[図12]第３の実施形態に係る回答学習装置の構成を示す機能ブロック図である。

[図13]第３の実施形態に係る回答生成装置の構成を示す機能ブロック図である。

発明を実施するための形態

[0018] 以下、本開示の実施形態について図面を用いて説明する。

[0019] ＜第１の実施形態に係る回答学習装置の概要＞

第１の実施形態は、入力された質問に対し、テキストに書いていない形式で答えを出力する新しいタスク設定として、「Ｙｅｓ又はＮｏ等の極性で回答することができる質問に対してＹｅｓ又はＮｏ等の極性で回答する」タスクを提案する。本実施形態では、回答の極性がＹｅｓ又はＮｏである場合を例に説明する。このＹｅｓ又はＮｏで回答するタスクは、既存研究の存在しない全く新しいタスクである。

[0020] 機械読解の代表的なデータセットには、ＳＱｕＡＤ（非特許文献２）の他にＭＳ－ＭＡＲＣＯ（参考文献１）が存在する。ＭＳ－ＭＡＲＣＯは１つの質問に１０近くの段落が紐づき、その段落群から人間が答えを生成したデータセットである。このような、質問に対して、文章に書かれていない形式で回答を出力するタスクを生成型タスクという。

［参考文献１］Tri Nguyen, Mir Rosenberg, Xia Song, Jianfeng Gao, Saurabh Tiwary, Rangan Majumder, Li Deng, ”MS MARCO: A Human Generated MAchine Reading COmprehension Dataset”, 2016.

[0021] 抽出型・生成型の２種類のタスクが存在する一方で、既存の機械読解技術の多くは抽出型のタスクを設定した技術が多い。

[0022] 生成型のタスクは、「テキストに書いていない形式で答えを出力する」という特性から、抽出型のタスクに比べて難しい課題となっている。

[0023] 生成型のタスクは、人間がゼロから生成した答えを正解とするデータセットを用いるので、機械も答えをゼロから作り出す必要がある。生成型のタスクの手法には、Ｓ－Ｎｅｔ（参考文献２）が存在する。

［参考文献２］Chuanqi Tan, Furu Weiz, Nan Yang, Bowen Du, Weifeng Lv, Ming Zhouz, ”S-NET: FROM ANSWER EXTRACTION TO ANSWER GENERATION FOR MACHINE READING COMPREHENSION”, 2017.

[0024] 一般的な質問応答において、Ｙｅｓ又はＮｏで回答すべき状況は多く現れる。しかし、参考文献２のような生成型の手法をこのような状況で適用した場合、回答としてＹｅｓ又はＮｏを生成する可能性を含むものの、その可能性は極めて低く、正しく応答をすることができない。

[0025] 本実施形態では、「Ｙｅｓ又はＮｏで回答することができる質問に対してＹｅｓ又はＮｏで回答する」タスクに特化した手法を提案するため、Ｙｅｓ又はＮｏで回答すべき状況で正しく応答することが可能である。そのため、機械によって質問応答可能な範囲を大きく広げることができる。

[0026] 本実施形態に係る回答学習装置は、単語系列である文章Ｐと質問文Ｑをベクトル系列に変換し、機械読解部が読解技術を用いて回答範囲スコア（s_d：s_e）に変換し、ベクトル系列と回答範囲のスコアから新しい技術である判断部を用いて判断スコアに変換し、回答範囲スコアと判断スコアを用いて学習する。

[0027] すなわち、Ｙｅｓ、Ｎｏの単なる２値判定（文章Ｐ全体を特徴量として、何も考えず機械学習で判定）を行うのではなく、機械読解技術によって、質問文Ｑに対する回答が書かれている場所を同定し、それを根拠としてＹｅｓかＮｏかの判定を行う。

[0028] この際、機械読解部と判定部とのニューラルネットワークは層を共有しているので、Ｙｅｓ／Ｎｏ判定に合わせた機械読解、読解に基づくＹｅｓ／Ｎｏ判定の両側面から学習することが可能となる。

[0029] ＜第１の実施形態に係る回答学習装置の構成＞

図１を参照して、第１の実施形態に係る回答学習装置１０の構成について

説明する。図１は、第１の実施形態に係る回答学習装置１０の構成を示すブロック図である。

[0030] 回答学習装置１０は、ＣＰＵと、ＲＡＭと、後述する回答学習処理ルーチンを実行するためのプログラムを記憶したＲＯＭとを備えたコンピュータで構成され、機能的には次に示すように構成されている。図１に示すように、本実施形態に係る回答学習装置１０は、入力部１００と、解析部２００と、パラメータ学習部３００とを備えて構成される。

[0031] 入力部１００は、文章Ｐと、質問文Ｑと、文章Ｐにおける当該質問文に対する回答の極性を示す正解Ｙと、文章Ｐにおける回答の根拠となる範囲の始端Ｄ及び終端Ｅとを含む複数の学習データの入力を受け付ける。

[0032] 具体的には、学習データは、テキストデータからなる文章Ｐ及び質問文Ｑと、回答がＹｅｓ／Ｎｏのいずれかであるかを示す正解Ｙと、文章Ｐにおける回答の根拠となる範囲（Ｄ：Ｅ）で構成される。ここで、Ｄ、Ｅは文章Ｐ中の単語の位置番号で表現され、Ｄは回答の根拠となる範囲の開始位置の単語の位置番号、Ｅは回答の根拠となる範囲の終了位置の単語の位置番号である。

[0033] テキストデータである文章Ｐ及び質問文Ｑは、既存のトークナイザによってトークン系列として表現されている。なお、トークンとして任意の単位を用いることができるが、本実施形態では、トークンを単語と表記する。

[0034] 単語系列で表現されている文章Ｐ及び質問文Ｑの長さを単語の数で定義し、文章Ｐの単語の数をL_P、質問文Ｑの単語の数をL_Qとする。

[0035] なお、複数の学習データをミニバッチとしてミニバッチ毎にまとめて処理してもよいし、学習データ毎に処理されてもよい。

[0036] そして、入力部１００は、受け付けた学習データのうち、文章Ｐと質問文Ｑとを、機械読解部２１０に、学習データをパラメータ学習部３００に渡す。

[0037] 解析部２００は、機械読解部２１０と、判断部２２０とを備えて構成される。

[0038] 機械読解部２１０は、文章Ｐ及び質問文Ｑに基づいて、文章Ｐにおける回答の根拠となる範囲Ｄ：Ｅを推定するための読解モデルを用いて、当該範囲の始端s_d及び終端s_eを推定する。

[0039] 具体的には、機械読解部２１０は、単語符号化部２１１と、単語データベース（ＤＢ）２１２と、第１文脈符号化部２１３と、アテンション部２１４と、第２文脈符号化部２１５と、根拠探索部２１６とを備えて構成される。

[0040] 単語符号化部２１１は、文章Ｐ及び質問文Ｑに基づいて、単語ベクトルの系列P_1及びQ_1を生成する。

[0041] 具体的には、単語符号化部２１１は、単語ＤＢ２１２から文章Ｐ及び質問文Ｑの各単語に対応するベクトルを抽出し、単語ベクトルの系列P_1及びQ_1を生成する。

[0042] 単語ＤＢ２１２に格納されるベクトルの次元をｄとすると、単語ベクトルの系列P_1は$L_P \times d$、単語ベクトルの系列Q_1は$L_Q \times d$の大きさの行列である。

[0043] そして、単語符号化部２１１は、生成した単語ベクトルの系列P_1及びQ_1を、第１文脈符号化部２１３に渡す。

[0044] 単語ＤＢ２１２は、複数の単語ベクトルが格納されている。単語ベクトルは、単語を表す所定次元の実数値ベクトルの集合である。

[0045] 具体的には、単語ＤＢ２１２は、予めニューラルネットワークにより学習された複数の単語ベクトル（ｗｏｒｄ　ｅｍｂｅｄｄｉｎｇ）を用いる。これには例えばｗｏｒｄ２ｖｅｃやＧｌｏＶｅのような、既存のものを用いてもよい。単語ベクトルには、既存の複数の単語ベクトルから抽出される単語ベクトルに、新しく学習した単語ベクトルを繋げることができる。なお、単語の文字情報を符号化する技術（参考文献３）等、任意の単語ｅｍｂｅｄｄｉｎｇ技術が使用可能である。単語ベクトルは、誤差逆伝播法によって計算できる勾配から学習することも可能である。

［参考文献３］Yoon Kim, Yacine Jernite, David Sontag, Alexander M. Rush, ”Character-Aware Neural Language Models”, arXiv:1508.06615, 201

6.

[0046] 第１文脈符号化部２１３は、単語符号化部２１１により生成された単語ベクトルの系列P_1及びQ_1を、ニューラルネットワークを用いてベクトルの系列P_2及びQ_2にそれぞれ変換する。

[0047] 具体的には、第１文脈符号化部２１３は、単語ベクトルの系列P_1及びQ_1を、ＲＮＮによってベクトルの系列P_2及びQ_2にする。ＲＮＮの構造には、ＬＳＴＭ等の既存技術を用いることができる。

[0048] 本実施形態では、第１文脈符号化部２１３は、ベクトルの系列を順方向に処理するＲＮＮと、逆方向に処理するＲＮＮとの２種類のＲＮＮを組み合わせた双方向ＲＮＮを用いる。双方向ＲＮＮの出力するベクトルの次元をd_1とすると、第１文脈符号化部２１３が変換するベクトルの系列P_2は$L_P \times d_1$、ベクトルの系列Q_2は$L_Q \times d_1$の大きさの行列となる。

[0049] そして、第１文脈符号化部２１３は、変換したベクトルの系列P_2及びQ_2を、アテンション部２１４に、ベクトルの系列Q_2を、入力変換部２２１にそれぞれ渡す。

[0050] アテンション部２１４は、ニューラルネットワークを用いて、ベクトルの系列P_2及びQ_2に基づいて、文章Ｐ及び質問文Ｑのアテンションを表すベクトルの系列である読解行列Ｂを生成する。

[0051] 具体的には、アテンション部２１４は、まず、ベクトルの系列P_2及びQ_2から、アテンション行列

$$A \in \mathbb{R}^{L_P \times L_Q}$$

を計算する。アテンション行列Ａは、例えば、下記式（１）を用いることができる。

[0052] [数1]

$$A_{ij} = \left[P_{2,i:}, Q_{2,j:}, P_{2,i:} \circ Q_{2,j:}\right] w_S \quad \cdots (1)$$

[0053] 上記式（１）において、行列の添え字で成分を表し、”：”は全体を表す。例えば、$A_{i:}$は、アテンション行列Ａの第ｉ行全体を表す。また、上記式

（１）において、”〇”は要素積であり、”，”はベクトル・行列を縦方向に結合する演算子である。ｗ$_S$は、モデルの学習可能なパラメータであり、

$$w_S \in \mathbb{R}^{3d_1}$$

である。

[0054] アテンション部２１４は、アテンション行列Ａを基に、文章Ｐから質問文Ｑ方向へのアテンションベクトル

$$\tilde{Q}$$

、質問文Ｑから文章Ｐ方向へのアテンションベクトル

$$\tilde{p}$$

を計算する。

[0055] ここで、アテンションベクトル

$$\tilde{Q}$$

は、下記式（２）で表すことができる。

[0056] [数2]

$$\alpha_i = \mathrm{softmax}(A_{i:}) \in \mathbb{R}^{L_Q}$$

$$\tilde{Q}_{i:} = \sum_j \alpha_{i,j} Q_{2,j:} \in \mathbb{R}^{d_1} \quad \cdots (2)$$

[0057] ｓｏｆｔｍａｘは、ソフトマックス関数であり、

$$\tilde{Q} \in \mathbb{R}^{L_P \times d_1}$$

である。

[0058] また、アテンションベクトル

$$\tilde{p}$$

は、下記式（３）で表すことができる。

[0059]

[数3]

$$\beta = \mathrm{softmax}_i\left(\max_j A_{ij}\right) \in \mathbb{R}^{L_P}$$

$$\tilde{p} = \sum_i \beta_i P_{2,i} \in \mathbb{R}^{d_1} \qquad \cdots(3)$$

[0060] ここで、

$$\max_j A_{ij}$$

は、L_P次元のベクトルであり、そのｉ番目の要素（$1 \leqq i \leqq L_P$）は、アテンション行列Ａのｉ番目のベクトルの最大値（ｊ方向のｍａｘ値）である。

$\mathrm{softmax}_i$は、ｉ方向にｓｏｆｔｍａｘを用いるという意味である。

[0061] βは、アテンション行列Ａにｍａｘ関数を用いることにより、長さがL_Pのベクトルとなり、式（３）において、βの各成分を重みとしてP_2の各行の重みの和を取ることにより、

$$\tilde{p}$$

は長さd_1のベクトルとなる。

[0062] また、

$$\tilde{p}^T$$

をL_P回繰り返し計算して縦に並べた行列が、

$$\tilde{P} \in \mathbb{R}^{L_P \times d_1}$$

となる。

[0063] アテンション部２１４は、ベクトルの系列P_2、アテンションベクトル

$$\tilde{p}$$

、及びアテンションベクトル

$$\tilde{Q}$$

に基づいて、アテンションの結果を表現する長さL_Pの読解行列Bを求める。例えば、読解行列

$$B = \left[P_2, \hat{Q},\ P_2 \circ \hat{Q}, P_2 \circ \hat{P}\right] \in \mathbb{R}^{L_P \times 4d_1}$$

である。ただし、”，”は、ベクトル・行列を横に結合する演算子である。

[0064] そして、アテンション部２１４は、読解行列Bを、入力変換部２２１及び第２文脈符号化部２１５に渡す。

[0065] 第２文脈符号化部２１５は、アテンション部２１４により生成された読解行列Bを、ニューラルネットワークを用いてベクトルの系列である読解行列Mに変換する。

[0066] 具体的には、第２文脈符号化部２１５は、読解行列Bを、ＲＮＮによって読解行列Mにする。ＲＮＮの構造には、第１文脈符号化部２１３と同様に、ＬＳＴＭ等の既存技術を用いることができる。

[0067] 第２文脈符号化部２１５のＲＮＮが出力する単語ベクトルの次元をd_2とすると、読解行列

$$M \in \mathbb{R}^{L_P \times d_2}$$

となる。

[0068] そして、第２文脈符号化部２１５は、変換した読解行列Mを、入力変換部２２１及び根拠探索部２１６に渡す。

[0069] 根拠探索部２１６は、読解行列Mに基づいて、文章Pにおける回答の根拠となる範囲Ｄ：Ｅを推定するための読解モデルを用いて、当該範囲の始端s_d及び終端s_eを推定する。

[0070] 具体的には、根拠探索部２１６は、回答の根拠となる範囲の始端s_dを推定するための始端用ＲＮＮ及び終端s_eを推定するための終端用ＲＮＮの２つのニューラルネットワークによって構成される。

[0071] 根拠探索部２１６は、まず、読解行列Mを、始端用ＲＮＮに入力してベクトルの系列M_1を得る。

[0072] 根拠探索部２１６は、回答の根拠となる範囲の始端s_dを、下記式（４）を

用いて求める。

[0073] [数4]

$$s_d = \mathrm{softmax}([B, M_1]\, w_1) \in \mathbb{R}^{L_P} \quad \cdots (4)$$

[0074] ここで、始端s_dは、回答の根拠となる範囲の始端に関するスコアであり、ベクトルで表される。すなわち、ベクトルの各次元に対応する単語が回答範囲の始端になる確率（スコア）を表す。

[0075] 同様に、読解行列Ｍを、終端用ＲＮＮに入力して単語ベクトルM_2を得る。

[0076] 根拠探索部２１６は、回答の根拠となる範囲の終端s_eを、下記式（５）を用いて求める。

[0077] [数5]

$$s_e = \mathrm{softmax}([B, M_2]\, w_2) \in \mathbb{R}^{L_P} \quad \cdots (5)$$

[0078] ここで、終端s_eは、回答の根拠となる範囲の終端に関するスコアであり、ベクトルで表される。すなわち、ベクトルの各次元に対応する単語が回答範囲の終端になる確率（スコア）を表す。

[0079] 推定した始端s_d及び終端s_eをまとめて回答範囲スコアと呼ぶ。なお、上記式（４）及び式（５）において、w_1及びw_2は、式（４）及び式（５）で表される読解モデルのパラメータであり、パラメータ学習部３００により更新される。

[0080] そして、根拠探索部２１６は、推定した回答範囲スコアを、入力変換部２２１及びパラメータ学習部３００に渡す。

[0081] 判断部２２０は、機械読解部２１０の処理によって得られる情報に基づいて、質問文Ｑに対する回答の極性が正か否かを判断する判断モデルを用いて、質問文Ｑに対する回答の極性を判断する。

[0082] 具体的には、判断部２２０は、入力変換部２２１と、スコア計算部２２２とを備えて構成される。

[0083] 入力変換部２２１は、機械読解部２１０により文章Ｐを符号化した結果と、機械読解部２１０により質問文Ｑを符号化した結果とに基づいて、ベクト

ルの系列P_3及びQ_3を生成する。

[0084] 具体的には、入力変換部２２１は、まず、機械読解部２１０の処理によって得られる情報の入力を受け付ける。

[0085] 入力を受け付ける情報は、４種類に分類することができる。すなわち、（１）文章Ｐの符号化結果であり、かつ、質問文Ｑを考慮した長さL_Pのベクトルの系列（例えば、読解行列Ｂ又はＭ）、（２）質問文Ｑの符号化結果である長さL_Qのベクトル系列（例えば、ベクトルの系列Q_2）、（３）回答範囲に関する情報である長さL_Pのベクトル（例えば、推定した始端s_dと終端s_e）、（４）文章Ｐと質問文Ｑとの意味的マッチング結果である大きさ$L_P \times L_Q$の行列（例えば、アテンション行列Ａ）の４種類を受け付ける。

[0086] ここで、受け付ける情報は、必ずしも４種類全てを受け付ける必要はなく、最低限の構成として（１）の１種類（読解行列Ｂ又はＭ）があれば本実施形態の目的を達することができる。（２）、（３）及び（４）は、いずれかのみ、あるいは複数を追加として受け付けても良い。本実施形態では、単純な形式として（１）読解行列Ｂ、及び（２）ベクトルの系列Q_2を受け付ける場合を例に説明する。

[0087] 入力変換部２２１は、受け付けた読解行列Ｂ及びベクトルの系列Q_2に基づいて、長さL_Pのベクトルの系列

$$P_3 \in \mathbb{R}^{L_P \times d_3}$$

、長さL_Qのベクトルの系列

$$Q_3 \in \mathbb{R}^{L_Q \times d_3}$$

を計算する。

[0088] ベクトルの系列P_3及びQ_3の計算方法として、任意のニューラルネットワークを用いることができる。例えば、下記式（６）及び式（７）を用いることができる。

[0089]

[数6]

$$P_3 = \mathrm{RNN}(B) \quad \cdots(6)$$

$$Q_3 = Q_2 \quad \cdots(7)$$

[0090] なお、d_3の次元数は、任意に設定することができる。式（６）及び式（７）を用いた場合、Q_2との次元を合わせるため、$d_3 = d_2$であり、式（６）におけるＲＮＮの出力の次元も$d_3 = d_2$となる。

[0091] そして、入力変換部２２１は、生成したベクトルの系列P_3及びQ_3を、スコア計算部２２２に渡す。

[0092] スコア計算部２２２は、質問文Ｑに対する回答の極性が正か否かを判断する判断モデルを用いて、質問文Ｑに対する回答の極性を判断する。

[0093] 具体的には、スコア計算部２２２は、ベクトルの系列P_3及びQ_3に基づいて、任意の文ペア分類タスクのフレームワークを用いて、質問文Ｑに対する回答がＹｅｓかＮｏかに分類するために用いる判断スコアｋ（０から１の実数）を求める。

[0094] 例えば、文ペア分類タスクの１つである含意認識の代表的なモデルであるＥＳＩＭ（参考文献４）のｄｅｃｏｄｅｒ　ＬＳＴＭ後のフレームワークを分類問題に用いることができる。

［参考文献４］Qian Chen, Xiaodan Zhu, Zhenhua Ling, Si Wei, Hui Jiang, Diana Inkpen, ”Enhanced LSTM for Natural Language Inference”, arXiv:1609.06038, 2017.

[0095] この場合、ベクトルの系列P_3及びQ_3をａｖｅｒａｇｅ　ｐｏｏｌｉｎｇ（列方向の平均を取る操作）、又はｍａｘ　ｐｏｏｌｉｎｇ（列方向の最大値を取る操作）をして、ベクトル

$$P_a, Q_a, P_m, Q_m \in \mathbb{R}^{d_3}$$

を得る。

[0096] 得られたベクトルP_a、Q_a、P_m及びQ_mを結合して、$4d_3$次元のベクトルＪを得る。ベクトルＪを多層パーセプトロンによって、実数（１次元のベク

トル）にし、シグモイド変換をして判断スコアｋを得る。

[0097] なお、Ｙｅｓ／Ｎｏの分類問題でなく、Ｙｅｓ、Ｎｏ、又は不明の３つに分類するように構成してもよい。この場合、ベクトルＪを多層パーセプトロンによって３次元のベクトルに変換した後で、ソフトマックス変換したものを判断スコアｋとしてもよい。

[0098] そして、スコア計算部２２２は、判断スコアｋを、パラメータ学習部３００に渡す。

[0099] パラメータ学習部３００は、学習データに含まれる正解Ｙと、判断部２２０により判断された結果とが一致し、学習データに含まれる始端Ｄ及び終端Ｅと、機械読解部２１０により推定された始端s_d及び終端s_eとが一致するように、読解モデル及び判断モデルのパラメータを学習する。

[0100] 具体的には、パラメータ学習部３００は、機械読解部２１０で用いる読解モデルについての目的関数L_Cと、判断部２２０で用いる判断モデルについての目的関数L_Jの線形和を、最適化問題の目的関数とする（下記式（８））。

[0101] [数7]

$$L_C + \lambda L_J \quad \cdots(8)$$

[0102] ここで、λはモデルのパラメータであり、学習器によって学習可能である。λの値を事前に指定する場合、１や１／２等、学習が進むように適当な値を定める。

[0103] 目的関数L_Cは、任意の機械読解技術の目的関数を用いることができる。例えば、非特許文献１では、下記式（９）で表されるクロスエントロピー関数を提案している。

[0104] [数8]

$$L_C = \log s_{d,D} + \log s_{e,E} \quad \cdots(9)$$

[0105] 上記式（９）において、Ｄ及びＥは、それぞれ真の始端Ｄ及び終端Ｅの位置を表し、$s_{d,D}$は、ベクトルs_dにおけるＤ番目の要素の値を、$s_{e,E}$は、ベクトルs_eにおけるＥ番目の要素の値を表す。

[0106] 目的関数Ｌ」も任意の目的関数を用いることができる。例えば、クロスエントロピー関数を用いた場合、下記式（１０）となる。

[0107] [数9]

$$L_J = \log k_Y \quad \cdots(10)$$

[0108] 上記式（１０）において、Ｙは、真の回答の極性を示す正解Ｙであり、正解ＹがＹｅｓである場合、スコアk_{Yes}＝ｋ、正解ＹがＮｏである場合、スコアk_{No}＝１－ｋである。つまり、正解ＹがＹｅｓの場合Ｌ」＝ｌｏｇ（ｋ）、正解ＹがＮｏの場合Ｌ」＝ｌｏｇ（１－ｋ）となる。

[0109] そして、パラメータ学習部３００は、上記式（８）で表される目的関数の勾配を、誤差逆伝播勾配法を用いて計算し、任意の最適化手法を用いてパラメータを更新する。

[0110] ＜第１の実施形態に係る回答学習装置の作用＞

図２は、第１の実施形態に係る回答学習処理ルーチンを示すフローチャートである。また、以下では本実施形態に係る回答学習装置が、ミニバッチを用いて学習する場合について説明するが、一般的なニューラルネットワークの学習方法を用いてもよい。なお、簡便のため、ミニバッチのサイズを１とする。

[0111] 入力部１００に複数の学習データが入力されると、回答学習装置１０において、図２に示す回答学習処理ルーチンが実行される。

[0112] まず、ステップＳ１００において、入力部１００は、文章Ｐと、質問文Ｑと、文章Ｐにおける当該質問文に対する回答の極性を示す正解Ｙと、文章Ｐにおける回答の根拠となる範囲の始端Ｄ及び終端Ｅとを含む複数の学習データの入力を受け付ける。

[0113] ステップＳ１１０において、入力部１００は、ステップＳ１００により受け付けた学習データを、ミニバッチに分割する。ミニバッチとは、複数の学習データをランダムに分割した学習データε個の集合である。εは１以上の自然数である。

[0114] ステップＳ１２０において、単語符号化部２１１は、１番目のミニバッチ

を選択する。

[0115] ステップＳ１３０において、単語符号化部２１１は、選択されているミニバッチに含まれる文章Ｐ及び質問文Ｑに基づいて、単語ベクトルの系列$Ｐ_1$及び$Ｑ_1$を生成する。

[0116] ステップＳ１４０において、第１文脈符号化部２１３は、上記ステップＳ１３０により生成された単語ベクトルの系列$Ｐ_1$及び$Ｑ_1$を、ニューラルネットワークを用いてベクトルの系列$Ｐ_2$及び$Ｑ_2$にそれぞれ変換する。

[0117] ステップＳ１５０において、アテンション部２１４は、ニューラルネットワークを用いて、ベクトルの系列$Ｐ_2$及び$Ｑ_2$に基づいて、文章Ｐ及び質問文Ｑのアテンションを表す読解行列Ｂを生成する。

[0118] ステップＳ１６０において、第２文脈符号化部２１５は、上記ステップＳ１５０により生成された読解行列Ｂを、ニューラルネットワークを用いて読解行列Ｍに変換する。

[0119] ステップＳ１７０において、根拠探索部２１６は、読解行列Ｍに基づいて、文章Ｐにおける回答の根拠となる範囲Ｄ：Ｅを推定するための読解モデルを用いて、当該範囲の始端$ｓ_d$及び終端$ｓ_e$を推定する。

[0120] ステップＳ１８０において、入力変換部２２１は、機械読解部２１０により文章Ｐを符号化した結果と、機械読解部２１０により質問文Ｑを符号化した結果とに基づいて、ベクトルの系列$Ｐ_3$及び$Ｑ_3$を生成する。

[0121] ステップＳ１９０において、スコア計算部２２２は、ベクトルの系列$Ｐ_3$及び$Ｑ_3$に基づいて、質問文Ｑに対する回答の極性が正か否かを判断する判断モデルを用いて、質問文Ｑに対する回答の極性を判断する。

[0122] ステップＳ２００において、パラメータ学習部３００は、学習データに含まれる正解Ｙと、判断部２２０により判断された結果とが一致し、学習データに含まれる始端Ｄ及び終端Ｅと、機械読解部２１０により推定された始端$ｓ_d$及び終端$ｓ_e$とが一致するように、読解モデル及び判断モデルのパラメータを更新する。

[0123] ステップＳ２１０において、パラメータ学習部３００は、全てのミニバッ

チについて処理を行ったか否かを判定する。

[0124] 全てのミニバッチについて処理を行っていない場合の場合（ステップＳ２１０のＮＯ）、ステップＳ２２０において、次のミニバッチを選択し、ステップＳ１３０に戻る

[0125] 一方、全てのミニバッチについて処理を行っている場合の場合（ステップＳ２１０のＹＥＳ）、ステップＳ２３０において、パラメータ学習部３００は、学習が収束したか否かについての収束判定を行う。

[0126] 学習が収束していない場合（ステップＳ２３０のＮＯ）、ステップＳ１１０に戻り、再度ステップＳ１１０～ステップＳ２３０までの処理を行う。

[0127] 一方、学習が収束している場合（ステップＳ２３０のＹＥＳ）、ステップＳ２４０において、パラメータ学習部３００は、学習したパラメータを、メモリ（図示省略）に格納する。

[0128] なお、ミニバッチのサイズを２以上とする場合、上記ステップＳ１２０の後に、１番目の文章Ｐ及び質問Ｑを選択するステップと、上記ステップＳ２１０の前に、ミニバッチ内の全ての文章Ｐ及び質問Ｑについて処理を行ったか否かを判定し、当該判定結果が否定的な場合に、次の文章Ｐ及び質問Ｑを選択して上記ステップＳ１３０に戻り、当該判定が肯定的な場合に上記ステップＳ２１０に進むステップとを追加する構成とすればよい。

[0129] 以上説明したように、本実施形態に係る回答学習装置によれば、文章と、質問文と、当該文章における当該質問文に対する回答の極性を示す正解と、当該文章における回答の根拠となる範囲の始端及び終端とを含む学習データの入力を受け付け、当該文章及び当該質問文に基づいて、当該範囲を推定するための読解モデルを用いて、当該範囲の始端及び終端を推定する処理によって得られる情報に基づいて、質問文に対する回答の極性が正か否かを判断する判断モデルを用いて、当該質問文に対する回答の極性を判断し、学習データに含まれる正解と、判断された結果とが一致し、学習データに含まれる始端及び終端と、推定された始端及び終端とが一致するように、読解モデル及び判断モデルのパラメータを学習することにより、極性で回答することが

できる質問に対して、精度よく、極性で回答するためのモデルを学習することができる。

[0130] ＜第１の実施形態に係る回答生成装置の構成＞

図３を参照して、第１の実施形態に係る回答生成装置２０の構成について説明する。図３は、第１の実施形態に係る回答生成装置２０の構成を示すブロック図である。なお、上述の回答学習装置１０と同様の構成については、同一の符号を付して詳細な説明は省略する。

[0131] 回答生成装置２０は、ＣＰＵと、ＲＡＭと、後述する回答生成処理ルーチンを実行するためのプログラムを記憶したＲＯＭとを備えたコンピュータで構成され、機能的には次に示すように構成されている。図３に示すように、本実施形態に係る回答生成装置２０は、入力部４００と、解析部２００と、出力部５００とを備えて構成される。なお、解析部２００は、回答学習装置１０により学習されたパラメータを用いる。

[0132] 入力部４００は、文章Ｐと、質問文Ｑとの入力を受け付ける。

[0133] そして、入力部４００は、受け付けた文章Ｐ及び質問文Ｑを、機械読解部２１０に渡す。

[0134] 出力部５００は、機械読解部２１０の根拠探索部２１６により得られた回答範囲スコアを回答の根拠とし、判断部２２０のスコア計算部２２２により得られた判断スコアｋを回答として出力する。

[0135] ここで、出力部５００は、判断スコアｋのＹｅｓのスコア、Ｎｏのスコアのうち、スコアが大きい判断結果を回答として出力する、閾値を超えたスコアの判断結果だけを出力するといった任意の出力形式を選択することができる。

[0136] また、出力部５００は、回答範囲スコアについても同様に、任意の出力形式を選択することができる。回答範囲スコアには始端s_dと終端s_eとが含まれるので、出力の計算方法として様々な手法を用いることが考えられる。例えば、非特許文献１のように、始端s_dが終端s_eよりも前になる制約下で、始端s_dと終端s_eとの積が最大になる範囲の単語列を出力する、といった手

法を用いることができる。

[0137] ＜第１の実施形態に係る回答生成装置の作用＞

図４は、第１の実施形態に係る回答生成処理ルーチンを示すフローチャートである。なお、第１の実施形態に係る回答学習処理ルーチンと同様の処理については、同一の符号を付して詳細な説明は省略する。

[0138] 入力部４００に文章Ｐと、質問文Ｑとが入力されると、回答生成装置２０において、図４に示す回答生成処理ルーチンが実行される。

[0139] ステップＳ３００において、入力部４００は、文章Ｐと、質問文Ｑとの入力を受け付ける。

[0140] ステップＳ４００において、出力部５００は、上記ステップＳ１７０により得られた回答範囲スコアを所定の方法により回答の根拠とし、上記ステップＳ１９０により得られた判断スコアｋを所定の方法により回答として生成する。

[0141] ステップＳ４３０において、出力部５００は、上記ステップＳ４００により得られた全ての回答の根拠及び回答を出力する。

[0142] 以上説明したように、本実施形態に係る回答生成装置によれば、入力された文章及び質問文に基づいて、当該文章における当該質問文に対する回答の根拠となる範囲を推定するための読解モデルを用いて、当該範囲の始端及び終端を推定する処理によって得られる情報に基づいて、質問文に対する回答の極性が正か否かを判断するための予め学習された判断モデルを用いて、当該質問文に対する回答の極性を判断することにより、極性で回答することができる質問に対して、精度よく、極性で回答することができる。

[0143] ＜第２の実施形態に係る回答学習装置の概要＞

人間が自然言語を理解して回答する場合は、自身のもつ経験、常識、及び世界知識を踏まえて、理解した質問に対して回答を推論することができる。例えば、人間が文章を読んでその文章に対する質問に回答をする場合には、文章からだけでなく、自分のもつ経験等から回答を見つけている。しかし、ＡＩの場合は質問の対象となっている文章に含まれている情報だけから回答

を推論する必要がある。

[0144] 特にＹｅｓ／Ｎｏで答えるべき質問は、質問に答えるために必要な知識が一か所に記載されているとは限らない。例えば、必要な知識が文章内の複数箇所に書いてある場合や世界知識から補わなければならない場合がある。しかし、文章内の複数箇所にある記述や世界知識を組み合わせて考えるためには、テキストの長期の依存関係を理解する必要がある。そのため、Ｙｅｓ／Ｎｏの質問応答を精度よく行うことは難しい。

[0145] そこで、第２の実施形態では、「Ｙｅｓ又はＮｏで回答することができる質問に対してＹｅｓ又はＮｏで回答する」タスクを精度よく行うために、必要な知識が文章内の複数箇所に書いてある質問や必要な知識を世界知識から補わなければならない質問に注目する。本実施形態では、第１の実施形態と同様に、回答の極性がＹｅｓ又はＮｏである場合を例に説明する。

[0146] 文章内の複数箇所にある記述を組み合わせて考える質問応答は、ニューラルネットワークが苦手とする長期の依存関係の理解を要求するため、難しい質問応答である。本実施形態では、回答に必要な文だけを根拠文として抽出することにより、位置が離れた根拠文同士のマッチングを可能にし、長期の依存関係を理解することを実現する。

[0147] この根拠文の抽出によって、ユーザーはＹｅｓ／Ｎｏの回答だけでなくその根拠となる文を過不足なく確認することが可能となり、解釈性を向上することもできる。

[0148] また、必要な知識を世界知識から補う必要がある質問応答に対しては、必要な知識が書いてあるテキストをＷｅｂでの検索等によって得て、質問対象の文章に繋げた新しい文章に対して質問応答を行うことで実現する。通常、単純に文章を繋げるだけでは、元の文章中の回答に必要な部分と新しく繋げたテキストが離れた箇所にあるためマッチングを取ることが難しい。しかし、本実施形態においては根拠文としてそれらを抽出することによって、根拠文が離れた箇所にある場合であってもマッチングが可能となる。

[0149] ＜第２の実施形態に係る回答学習装置の構成＞

図５を参照して、第２の実施形態に係る回答学習装置３０の構成について説明する。図５は、第２の実施形態に係る回答学習装置３０の構成を示すブロック図である。なお、上述の第１の実施形態に係る回答学習装置１０と同様の構成については、同一の符号を付して詳細な説明は省略する。

[0150] 回答学習装置３０は、ＣＰＵと、ＲＡＭと、後述する回答学習処理ルーチンを実行するためのプログラムを記憶したＲＯＭとを備えたコンピュータで構成され、機能的には次に示すように構成されている。図５に示すように、本実施形態に係る回答学習装置３０は、入力部１００と、解析部６００と、パラメータ学習部７００とを備えて構成される。

[0151] 解析部６００は、機械読解部６１０と、判断部２２０とを備えて構成される。機械読解部６１０は、文章Ｐ及び質問文Ｑに基づいて、文章Ｐにおける回答の根拠となる範囲Ｄ：Ｅを推定するための読解モデルを用いて、当該範囲の始端s_d及び終端s_eを推定する。

[0152] 具体的には、機械読解部２１０は、単語符号化部２１１と、単語データベース（ＤＢ）２１２と、第１文脈符号化部２１３と、アテンション部２１４と、第２文脈符号化部２１５と、根拠抽出部６１７と、根拠探索部２１６とを備えて構成される。

[0153] 根拠抽出部６１７は、機械読解部６１０の処理によって得られる情報に基づいて、質問文に対する回答の根拠となる情報である根拠情報を抽出する抽出モデルを用いて、質問文Ｑに対する回答の根拠情報を抽出する。

[0154] 具体的には、根拠抽出部６１７は、まず、第２文脈符号化部２１５により変換された読解行列Ｍ（変換前の読解行列Ｂでもよい）を入力とし、ニューラルネットワークを用いて文章Ｐの各文の意味を表すベクトルの系列Ｈを抽出する。根拠抽出部６１７は、例えば、ニューラルネットワークとして、Ｕｎｄｉｒｅｃｔｉｏｎａｌ－ＲＮＮを用いることができる。

[0155] 次に、根拠抽出部６１７は、根拠文を１つ抽出する操作を１時刻と定義し、状態z_tを抽出モデルのＲＮＮによって生成する。すなわち、根拠抽出部６１７は、時刻ｔ－１に抽出された根拠文に対応するベクトルの系列Ｈの要素

$h_{s_{t-1}}$

を抽出モデルのＲＮＮに入力することにより、状態z_tを生成する。ただし、s_{t-1}は時刻ｔ－１に抽出された根拠文の添字である。また、時刻ｔまでに抽出された文s_tの集合をS_tとする。

[0156] 根拠抽出部６１７は、状態z_tと、質問文の各単語に対するベクトルy_jからなるベクトルの系列Ｙ’に基づいて、抽出モデルにより、時刻ｔにおける重要性を考慮した質問文ベクトルであるｇｌｉｍｐｓｅベクトルe_t（下記式（１３））を、質問文Ｑに対するｇｌｉｍｐｓｅ操作（参考文献５）を行うことで生成する。このように、抽出モデルでは質問文Ｑに対するｇｌｉｍｐｓｅ操作を行うことで、根拠文の抽出結果が質問全体に対応する内容を包含することができる。

［参考文献５］Ｏ．Ｖｉｎｙａｌｓ，Ｓ．Ｂｅｎｇｉｏ ａｎｄ Ｍ．Ｋｕｄｌｕｒ，“Ｏｒｄｅｒ ｍａｔｔｅｒｓ：Ｓｅｑｕｅｎｃｅ ｔｏ ｓｅｑｕｅｎｃｅ ｆｏｒ ｓｅｔｓ”，ＩＣＬＲ（２０１６）．

[0157] [数10]

$$a_j^t = v_g^T \tanh(W_{g1} y_j + W_{g2} z_t) \quad \cdots (11)$$

$$\varphi^t = \mathrm{softmax}(a^t) \in \mathbb{R}^n \quad \cdots (12)$$

$$e_t = \sum_j \varphi_j^t W_{g1} y_j \in \mathbb{R}^d \quad \cdots (13)$$

[0158] 抽出モデルのＲＮＮの初期値はベクトルの系列Ｈをａｆｆｉｎｅ変換したベクトル系列をｍａｘｐｏｏｌｉｎｇしたベクトルとする。

[0159] 根拠抽出部６１７は、状態z_tと、ｇｌｉｍｐｓｅベクトルe_tと、ベクトルの系列Ｈとに基づいて、抽出モデルにより、時刻ｔにおいて下記式（１４）で表される確率分布に従って第δ文を選び、文$s_t = \delta$を、時刻ｔに抽出された根拠文とする。

[0160] [数11]

$$P(\delta; S_{t-1}) = \mathrm{softmax}_\delta(u_\delta^t) \qquad \cdots (14)$$

$$u_j^t = \begin{cases} v_p^T \tanh(W_{p1} h_j + W_{p2} e_t + W_{p3} z_t) & (j \notin S_{t-1}) \\ -\infty & (\mathrm{otherwise}) \end{cases}$$

[0161] そして、根拠抽出部６１７は、抽出した文s_tの集合S_tを根拠情報として、根拠探索部２１６及びパラメータ学習部７００に渡す。

[0162] 根拠探索部２１６は、上記第１の実施形態と同様に、文章Ｐにおける回答の根拠となる範囲Ｄ：Ｅを推定するための読解モデルを用いて、当該範囲の始端s_d及び終端s_eを推定する。

[0163] 具体的には、上記第１の実施形態で説明した読解行列

$$\mathrm{M} \in \mathbb{R}^{L_p \times d_2}$$

を利用すると共に、文の意味を表すベクトルの系列Ｈを利用する。

[0164] また、ベクトルの系列

$$\mathrm{H} \in \mathbb{R}^{m \times d}$$

は文レベルのベクトル系列であるため、以下のように、単語レベルのベクトル系列に変換する。

[0165] まず、単語のインデックスをｉ（ｉ＝１，…，L_p）、文のインデックスをｊ（ｊ＝１，…，ｍ）とする。関数ｗｏｒｄ＿ｔｏ＿ｓｅｎｔを、単語ｉが文ｊに含まれるときｗｏｒｄ＿ｔｏ＿ｓｅｎｔ（ｉ）＝ｊで定める。新しい単語レベルのベクトル系列Ｈ’を、$h_i' = h_{__word_to_sent(i)}$と定義することで、

$$\mathrm{H}' \in \mathbb{R}^{L_p \times d}$$

を定める。

[0166] そして、ＭとＨ’を縦に連結した行列を

$$M' \in \mathbb{R}^{L_p \times (d_2+d)}$$

として、このM’を後段の処理におけるMの代わりに用いてもよい。なお、上記では、読解行列Mを利用する場合を例に説明したが、読解行列Bなど別の行列を利用してもよい。また、上記の根拠探索部２１６の説明で用いたｉやｊといった変数は、ここの説明に限るものである。

[0167] パラメータ学習部７００は、学習データに含まれる正解Ｙと、判断部２２０により判断された結果とが一致し、学習データに含まれる始端Ｄ及び終端Ｅと、機械読解部６１０により推定された始端s_d及び終端s_eとが一致し、学習データに含まれる文章Ｐにおける正解の根拠情報と、根拠抽出部６１７により抽出された根拠情報とが一致するように、読解モデル、判断モデル及び抽出モデルのパラメータを学習する。

[0168] 具体的には、文章Ｐと、質問Ｑと、ＹＥＳ、ＮＯ、及び抽出型の何れかである正解となる回答Ｙ^、回答Ｙの正解となる根拠範囲である始端Ｄ^及び終端Ｅ^と、回答Ｙの正解となる根拠情報である根拠段落の集合S_t^とを１セットとし、学習データが、複数セット含むものとする。また、パラメータ学習部７００は、機械読解部６１０で用いる読解モデルについての目的関数L_Cと、判断部２２０で用いる判断モデルについての目的関数L_Jと、根拠抽出部６１７で用いる抽出モデルについての目的関数L_sとの線形和を、最適化問題の目的関数とする（下記式（１５））。

[0169] [数12]

$$\lambda_1 L_C + \lambda_2 L_J + \lambda_3 L_S \quad \cdots (15)$$

[0170] ここで、λ_1、λ_2、λ_3はハイパーパラメータであり、１／３等の学習が進むように適当な値を定める。また、サンプルによって持つ教師データが異なる場合も、持たないデータに関する項のλを０とすることで一律に扱うことができる。例えば、根拠探索部２１６の出力に対応するデータがないサンプルに対しては、$\lambda_1 = 0$とする。

[0171] 目的関数L_C及びL_Jについては、第１の実施形態と同様である。目的関数

Ｌｓは、ｃｏｖｅｒａｇｅ正則化（参考文献６）を行った目的関数である。例えば、目的関数Ｌｓは下記式（１６）のような目的関数を用いることができる。

［参考文献６］Ａ．　Ｓｅｅ，　Ｐ．　Ｊ．　Ｌｉｕ　ａｎｄ　Ｃ．　Ｄ．　Ｍａｎｎｉｎｇ，　“Ｇｅｔ　ｔｏ　ｔｈｅ　ｐｏｉｎｔ：　ｕｍｍａｒｉｚａｔｉｏｎ　ｗｉｔｈ　ｐｏｉｎｔｅｒ－ｇｅｎｅｒａｔｏｒ　ｎｅｔｗｏｒｋｓ”，　ＡＣＬ，　２０１７，　ｐｐ．１０７３－１０８３．

[0172] [数13]

$$L_S = -\sum_{t=1}^{T} \log P(\hat{s}_t; S_{t-1}) + \sum_i \min(c_i^t, \varphi_i^t) \quad \cdots (16)$$

[0173] 上記式（１６）において、

$\hat{s}_t$

は、正解の根拠情報として与えられた根拠文の集合S_tの中で時刻ｔの抽出確率Ｐ（δ；S_{t-1}）が最小の文ｓとし、c^tは、ｃｏｖｅｒａｇｅベクトルであり、

$c^t = \sum_{r=1}^{t-1} \alpha^r$

である。Ｔは終了時刻である。すなわち、ｔ＝Ｔが学習の終了条件となる。このｃｏｖｅｒａｇｅにより、抽出結果を質問全体に対応する内容を包含させることが可能となる。ただし、抽出の終了条件を学習するために、抽出終了ベクトル

h_{EOE}

を学習可能なパラメータとする。文の意味を表すベクトルの系列Ｈに抽出終了ベクトル

h_{EOE}

を加え、文章Ｐの文数ｍを実際の文数＋１とする。Ｔも真の根拠文の数＋１

とし、学習時は時刻Ｔ－１までに全ての根拠文を出力した後、時刻Ｔに抽出終了ベクトル

$$h_{EOE}$$

を抽出するように学習を行う。テスト時は、抽出終了ベクトルを出力した時点で抽出を終了する。

[0174] そして、パラメータ学習部７００は、上記式（１６）で表される目的関数の勾配を、誤差逆伝播勾配法を用いて計算し、任意の最適化手法を用いて各パラメータを更新する。

[0175] ＜第２の実施形態に係る回答学習装置の作用＞

図６は、第２の実施形態に係る回答学習処理ルーチンを示すフローチャートである。また、以下では本実施形態に係る回答学習装置が、ミニバッチを用いて学習する場合について説明するが、一般的なニューラルネットワークの学習方法を用いてもよい。なお、簡便のため、ミニバッチのサイズを１とする。なお、上述の第１の実施形態に係る回答学習処理ルーチンと同様の構成については、同一の符号を付して詳細な説明は省略する。

[0176] ステップＳ５５５において、根拠抽出部６１７は、根拠情報抽出処理を実行する。

[0177] ステップＳ６００において、パラメータ学習部７００は、学習データに含まれる正解Ｙと、判断部２２０により判断された結果とが一致し、学習データに含まれる始端Ｄ及び終端Ｅと、機械読解部２１０により推定された始端s_d及び終端s_eとが一致し、学習データに含まれる文章Ｐにおける回答の根拠情報と、根拠抽出部６１７により抽出された根拠情報とが一致するように、読解モデル、判断モデル及び抽出モデルのパラメータを学習する。

[0178] 図７は、第２の実施形態に係る回答学習装置における根拠情報抽出処理ルーチンを示すフローチャートである。根拠抽出部６１７は、根拠情報抽出処理により、機械読解部６１０の処理によって得られる情報に基づいて、質問文に対する回答の根拠となる情報である根拠情報を抽出する抽出モデルを用

いて、質問文Ｑに対する回答の根拠情報を抽出する。

[0179] ステップＳ５００において、根拠抽出部６１７は、ｔ＝１とする。

[0180] ステップＳ５１０において、根拠抽出部６１７は、根拠文を１つ抽出する操作を１時刻と定義し、時刻ｔにおける状態z_tを抽出モデルのＲＮＮによって生成する。

[0181] ステップＳ５２０において、根拠抽出部６１７は、時刻ｔにおける重要性を考慮した質問文ベクトルであるｇｌｉｍｐｓｅベクトルe_tを、質問文Ｑに対してｇｌｉｍｐｓｅ操作を行うことにより生成する。

[0182] ステップＳ５３０において、根拠抽出部６１７は、時刻ｔにおいて上記式（１４）で表される確率分布に従って第δ文を選び、文$s_t=\delta$とする。

[0183] ステップＳ５４０において、根拠抽出部６１７は、終了条件を満たしているか否かを判定する。

[0184] 終了条件を満たしていない場合（上記ステップＳ５４０のＮＯ）、根拠抽出部６１７は、ステップＳ５５０においてｔに１を加算し、ステップＳ５１０に戻る。一方、終了条件を満たしている場合（上記ステップＳ５４０のＹＥＳ）、根拠抽出部６１７は、リターンする。

[0185] 以上説明したように、本実施形態に係る回答学習装置によれば、機械読解部の処理によって得られる情報に基づいて、質問文に対する回答の根拠となる情報である根拠情報を抽出する抽出モデルを用いて、質問文に対する回答の根拠情報を抽出し、学習データに含まれる文章における回答の根拠情報と、根拠抽出部により抽出された根拠情報とが一致するように、抽出モデルのパラメータを学習することにより、極性で回答することができる質問に対して、更に精度よく、極性で回答するためのモデルを学習することができる。

[0186] また、質問文に対する回答の根拠となる情報である根拠情報を抽出する抽出モデルと、根拠（回答）範囲を抽出する読解モデルと、回答を判断する判断モデルとのマルチタスク学習を行うことにより、根拠文が離れた個所にある場合でも、質問文とのマッチングを取ることを可能にする。ここで、マルチタスク学習とは、複数の異なるタスクを解く複数のモデルについて、当該

複数のモデルの一部分を共有した状態で学習することをいう。本実施形態では、各符号化部のモデル（厳密には各符号化部のモデルのパラメータ）を共有し、共有した各符号化部による符号化結果を共通の入力とした抽出モデル、読解モデル、及び判断モデルの出力が正解となる情報に近付くよう、抽出モデル、読解モデル、判断モデル、及び符号化モデルのマルチタスク学習を行う。

[0187] ＜第２の実施形態に係る回答生成装置の構成＞

図８を参照して、第２の実施形態に係る回答生成装置４０の構成について説明する。図８は、第２の実施形態に係る回答生成装置４０の構成を示すブロック図である。なお、上述の回答学習装置３０と同様の構成については、同一の符号を付して詳細な説明は省略する。回答生成装置４０は、ＣＰＵと、ＲＡＭと、後述する回答生成処理ルーチンを実行するためのプログラムを記憶したＲＯＭとを備えたコンピュータで構成され、機能的には次に示すように構成されている。図８に示すように、第２の実施形態に係る回答生成装置４０は、入力部４００と、解析部６００と、出力部８００とを備えて構成される。

[0188] 出力部８００は、判断部２２０により判断された回答の極性と、根拠抽出部６１７により抽出された根拠情報とを回答として出力する。

[0189] ＜第２の実施形態に係る回答生成装置の作用＞

図９は、第２の実施形態に係る回答生成処理ルーチンを示すフローチャートである。なお、第１の実施形態に係る回答生成処理ルーチン及び第２の実施形態に係る回答学習処理ルーチンと同様の処理については、同一の符号を付して詳細な説明は省略する。

[0190] ステップＳ７００において、出力部８００は、上記ステップＳ４００により得られた全ての回答の根拠及び回答、及び上記ステップＳ５５５により得られた根拠情報を出力する。

[0191] ＜第２の実施形態に係る回答生成装置の実施例＞

次に、第２の実施形態に係る回答生成装置の実施例について説明する。本

実施例では、回答生成装置の各部の構成として、図１０に示した構成を用いる。具体的には、判断部２２０は、ＲＮＮと線形変換とを用いて構成され、Ｙｅｓ／Ｎｏ／抽出型の回答の何れかで答えるかを判断し、Ｙｅｓ／Ｎｏ／抽出型の回答の３値の何れかを出力とする。また、根拠探索部２１６は、ＲＮＮと線形変換との組を２つ用いて構成され、一方の組は回答の終点、他方の組は回答の始点を出力とする。根拠抽出部６１７は、ＲＮＮと抽出モデル６１７Ａとを用いて構成される。第２文脈符号化部２１５は、ＲＮＮとセルフアテンションとを用いて構成され、アテンション部２１４は、双方向アテンションにより構成される。

[0192] 第１文脈符号化部２１３は、２つのＲＮＮを用いて構成され、単語符号化部２１１は、単語埋め込みと文字埋め込みとの組を２つ用いて構成される。

[0193] また、抽出モデル６１７Ａの構成として、図１１に示す構成を用いている。この構成は、参考文献７に提案されている抽出型文章要約モデルをベースとしている。

［参考文献７］Ｙ．Ｃ．　Ｃｈｅｎ　ａｎｄ　Ｍ．　Ｂａｎｓａｌ，　“Ｆａｓｔ　ａｂｓｔｒａｃｔｉｖｅ　ｓｕｍｍａｒｉｚａｔｉｏｎ　ｗｉｔｈ　ｒｅｉｎｆｏｒｃｅ－ｓｅｌｅｃｔｅｄ　ｓｅｎｔｅｎｃｅ　ｒｅｗｒｉｔｉｎｇ”，　ＡＣＬ，　２０１８，　ｐｐ．６７５－６８６．

[0194] 参考文献７の手法は、要約元文章に注意しながら要約元文章中の文を抽出する手法であるが、本実施例では質問文Ｑに注意しながら文章Ｐ中の文を抽出する。抽出モデル６１７Ａでは、質問文Ｑに対するｇｌｉｍｐｓｅ操作を行うことで、抽出結果が質問全体に対応する内容を包含することを意図している。

[0195] ＜第２の実施形態に係る回答生成装置の実施例における実験結果＞

次に、第２の実施形態に係る回答生成装置の実施例における実験結果について説明する。

[0196] ＜＜実験設定＞＞

実験はＧＰＵに、”ＮＶＩＤＩＡ　Ｔｅｓｌａ　Ｐ１００（株式会社エル

ザジャパン製）”を４枚用いて行った。実装にはＰｙｔｏｒｃｈを用いた。Ｂｉ－ＲＮＮの出力の次元をｄ＝３００で統一した。ｄｒｏｐｏｕｔのｋｅｅｐ　ｒａｔｉｏは０．８とした。バッチサイズを７２、学習率を０．００１とした。上記以外の設定はベースラインモデルと同じ設定である。抽出モデル６１７ＡはＲＮＮにＧＲＵを用いた、ベクトルの初期化を正規分布で、行列の初期化をｘａｖｉｅｒ　ｎｏｒｍａｌ分布で行った。デコード時のｂｅａｍ　ｓｉｚｅを２とした。

[0197] また、ベースラインモデルとして、本実施例に係る回答生成装置の構成（図１０）のうち、抽出モデル６１７Ａをａｆｆｉｎｅ変換とｓｉｇｍｏｉｄ関数により各文の根拠スコアを得るモデルに変更したモデルを用いた。

[0198] 本実験では、回答タイプＴ・回答Ａ・根拠文Ｓの予測精度を評価した。ここで、回答タイプＴは、ＨｏｔｐｏｔＱＡのタスク設定における「Ｙｅｓ・Ｎｏ・抽出」の３ラベルから構成される。回答、根拠文抽出ともに完全一致（ＥＭ）と部分一致を評価した。部分一致の指標は適合率と再現率の調和平均（Ｆ１）である。回答は、回答タイプＴの一致で評価し、抽出の場合は回答Ａの一致でも評価する。根拠文抽出の部分一致については抽出された文のｉｄの真の根拠文ｉｄへの一致で測った。そのため、単語レベルでの部分一致は考慮されない。回答タイプに関して、「Ｙｅｓ・Ｎｏ」質問に限定したときの回答精度をＹＮと記した。また、回答と根拠の精度双方を考慮した指標としてｊｏｉｎｔ　ＥＭ及びｊｏｉｎｔ　Ｆ１（参考文献８）を用いる。

［参考文献８］Ｚ．　Ｙａｎｇ，　Ｐ．　Ｑｉ，　Ｓ．　Ｚｈａｎｇ，　Ｙ．　Ｂｅｎｇｉｏ，　Ｗ．　Ｗ．　Ｃｏｈｅｎ，　Ｒ．　Ｓａｌａｋｈｕｔｄｉｎｏｖ　ａｎｄ　Ｃ．　Ｄ．　Ｍａｎｎｉｎｇ，　“ＨｏｔｐｏｔＱＡ：　Ａ　ｄａｔａｓｅｔ　ｆｏｒ　ｄｉｖｅｒｓｅ，　ｅｘｐｌａｉｎａｂｌｅ　ｍｕｌｔｉ－ｈｏｐ　ｑｕｅｓｔｉｏｎ　ａｎｓｗｅｒｉｎｇ”，　ＥＭＮＬＰ，　２０１８，　ｐｐ．２３６９－２３８０．

[0199] 本実験では、ｄｉｓｔｒａｃｔｏｒ設定とｆｕｌｌｗｉｋｉ設定との場合について行う。ｄｉｓｔｒａｃｔｏｒ設定は、大量のテキストを質問に関連

する少量のテキストに絞ることが既存技術によって可能であるという仮定を置いた設定である。また、ｆｕｌｌｗｉｋｉ設定は、ＴＦ－ＩＤＦ類似度検索によって少量テキストへの絞り込みを行った設定である。

[0200] ＜＜実験結果＞＞

テストデータにおける実験結果は、ｄｉｓｔｒａｃｔｏｒ設定の結果を表１に、ｆｕｌｌｗｉｋｉ設定の結果を表２に示す。

[0201] [表1]

	回答		根拠		Joint	
	EM	F1	EM	F1	EM	F1
ベースライン	45.6	59.0	20.3	64.5	10.8	40.2
提案手法	**53.9**	**68.1**	**57.8**	**84.5**	**34.6**	**59.6**

[0202] [表2]

	回答		根拠		Joint	
	EM	F1	EM	F1	EM	F1
ベースライン	24.0	32.9	3.86	37.7	1.85	16.2
提案手法	**28.7**	**38.1**	**14.2**	**44.4**	**8.69**	**23.1**

[0203] ｄｉｓｔｒａｃｔｏｒ設定、ｆｕｌｌｗｉｋｉ設定共に、本実施例はベースラインモデルを大きく上回り、ｓｔａｔｅ－ｏｆ－ｔｈｅ－ａｒｔの精度を達成した。特に根拠文の完全一致はｄｉｓｔｒａｃｔｏｒ設定で３７．５ポイント（＋１８５％）、ｆｕｌｌｗｉｋｉ設定で１０．３ポイント（＋２６８％）と大きく向上している。そのため、本実施例は根拠文を過不足なく抽出することに秀でた手法であると言える。開発データでのｄｉｓｔｒａｃｔｏｒ設定における実験結果を表３に示す。

[0204]

[表3]

	回答			根拠		Joint	
	YN	EM	F1	EM	F1	EM	F1
ベースライン	57.8	52.7	67.3	34.3	78.0	19.9	54.4
提案手法	**63.4**	**53.7**	**68.7**	**58.8**	**84.7**	**35.4**	**60.6**
− glimpse	61.7	53.1	67.9	58.4	84.3	34.8	59.6

[0205] 開発データでのベースラインモデルは我々の追実験によって学習されたため、精度がテストデータでの数値と大きく異なっている。これはハイパーパラメータの違いに起因する。まず、本実施例はベースラインモデルを根拠文抽出のＥＭで２４．５ポイント上回っている。Ｆ１でも６．７ポイントの向上が見られる。さらに、回答でもＥＭで１．０ポイント、Ｆ１で１．４ポイントの上昇がある。特に「Ｙｅｓ・Ｎｏ」の判断精度については、５．６ポイント向上している。ベースラインモデルと本実施例とでは、抽出モデル６１７Ａ以外は全て同じモデルを用いている。それにも関わらず「Ｙｅｓ・Ｎｏ」の判断精度が向上していることは、抽出モデル６１７Ａとのマルチタスク学習が下層のＲＮＮを回答にも資する特徴量を獲得するように学習できると解釈できる。結果として、Ｊｏｉｎｔ指標でも精度が向上している。比較手法として、ｇｌｉｍｐｓｅ操作を用いずにＲＮＮによる文抽出だけを行う手法を実験したが、全ての指標で本実施例が上回ることを確認した。

[0206] 開発データでのｆｕｌｌｗｉｋｉ設定における実験結果を表４に示す。

[0207] [表4]

	回答			根拠		Joint	
	YN	EM	F1	EM	F1	EM	F1
ベースライン	59.2	28.5	38.2	7.87	**45.2**	4.77	22.8
提案手法	**62.2**	**29.0**	**38.8**	**14.4**	44.8	**8.43**	**23.3**
− glimpse	60.8	28.6	38.2	13.9	44.5	8.30	23.0

[0208] 本実施例はベースラインモデルを根拠のＥＭで６．５ポイント上回っているが、Ｆ１ではベースラインモデルを下回っている。回答ではＥＭで０．９ポイント、Ｆ１で０．８ポイントの上昇がある。特に「Ｙｅｓ・Ｎｏ」の判断精度については、３．０ポイント向上している。そのため、やはり抽出モデル６１７Ａによって下層のＲＮＮの学習が進んでいると解釈できる。結果として、Ｊｏｉｎｔ指標でも精度が向上している。また、本実施例がｇｌｉｍｐｓｅ操作を用いない手法と比較して全ての指標で上回っていることを確認した。

[0209] 以上の結果から、少量の関連テキストの中から特に必要な文を検索することについては、ｄｉｓｔｒａｃｔｏｒ設定では部分一致で８４．７％の精度を達成したこと、及び必要な文を使って「Ｙｅｓ・Ｎｏ」の判断精度を上げることについては、５．６ポイントの精度の向上が観察できた。

[0210] 以上説明したように、本実施形態に係る回答生成装置によれば、機械読解部の処理によって得られる情報に基づいて、質問文に対する回答の根拠となる情報である根拠情報を抽出する抽出モデルを用いて、前記質問文に対する回答の根拠情報を抽出し、判断された回答の極性と、抽出された根拠情報とを回答として出力することにより、極性で回答することができる質問に対して、更に精度よく、極性で回答することができる。

[0211] また、回答学習装置では、根拠抽出部、根拠探索部、及び判断部のそれぞれのタスクに関してマルチタスク学習を行っているため、根拠情報を参考にした、根拠となる範囲の推定や判断スコアの計算が実現される。

[0212] ＜第３の実施形態に係る回答学習装置の概要＞

上記第２の実施形態では、質問文に対する回答を出力するとき、回答の根拠として、根拠となる範囲と、文の集合である根拠情報を出力することで、回答結果をユーザーが解釈しやすい形にしていた。

[0213] この際、根拠情報に含まれるそれぞれの文が、回答を抽出する根拠としてどれぐらい適しているかを示す、根拠スコアを提示していなかった。

[0214] また、学習のためには、根拠文となる正解の文を定義した教師データが必

要であった。この教師データの作成は、人間が「回答のためにどの文が必要か」を考えながら作業を行う必要があるため、難しく高コストなタスクである。

[0215] そこで、本実施形態では、根拠抽出部において、文章を分割したスパンごとに、そのスパンが回答を抽出する根拠としてどれぐらい適しているかを示す、根拠スコアを算出する。

[0216] また、本実施形態では、根拠情報の単位であるスパンを、文ではなく段落とする。根拠情報の単位として文ではなく段落を用いることで、教師データの作成が容易となる。なぜなら、段落は文よりも粒度が大きく多くの情報を含むため、多くの教師データでは、回答を含む１段落をそのまま根拠段落として使用すればよい。そのため、人間の追加の作業が不要である。

[0217] なお、段落は、一般的に使われる段落と異なってもよく、例えば箇条書きの全体を段落としてみなすことも可能であるし、１つ１つのアイテムをそれぞれ別段落とみなすことも可能である。文章を分割する単位として、以降は「段落」を例として説明を行うが、上記の実施形態と同様に、文を単位として以降の処理を行うことも可能である。文章を所定単位で分割したものを「スパン」と呼称し、文や段落はこれに含まれるものとする。

[0218] また、回答種別とは、「ＹＥＳ」、「ＮＯ」、又は文章から回答を抽出する「抽出型」のいずれである。根拠範囲とは、文章の中の根拠を表す範囲の単語列であり、回答種別が抽出型である場合、この範囲に含まれる文字列を回答としても用いるため、回答範囲とも称する。また、根拠情報とは、根拠となるスパンの集合である。すなわち、根拠範囲（回答範囲）は単語レベルの根拠であり、根拠情報は、スパン（文や段落）レベルの根拠とも言える。

[0219] また、本実施形態では、回答生成装置が、根拠スコアを算出して可視化する点が、第２の実施形態と異なっている。

[0220] また、文章を構成する段落を１つの文章とみなし、段落毎に符号化部に入力し、符号化部が、単語符号化部２１１、第１文脈符号化部２１３、アテンション部２１４、及び第２文脈符号化部２１５に相当するモデルとして、参

考文献９に記載のＢＥＲＴ（Ｂｉｄｉｒｅｃｔｉｏｎａｌ Ｅｎｃｏｄｅｒ Ｒｅｐｒｅｓｅｎｔａｔｉｏｎｓ ｆｒｏｍ Ｔｒａｎｓｆｏｒｍｅｒｓ）などの事前学習済み言語モデルを利用する点が、第２の実施形態と異なっている。

[0221] ［参考文献９］ＢＥＲＴ： Ｐｒｅ－ｔｒａｉｎｉｎｇ ｏｆ Ｄｅｅｐ Ｂｉｄｉｒｅｃｔｉｏｎａｌ Ｔｒａｎｓｆｏｒｍｅｒｓ ｆｏｒ Ｌａｎｇｕａｇｅ Ｕｎｄｅｒｓｔａｎｄｉｎｇ，＜ｏｎｌｉｎｅ＞，＜ＵＲＬ： ｈｔｔｐｓ：／／ｗｗｗ．ａｃｌｗｅｂ．ｏｒｇ／ａｎｔｈｏｌｏｇｙ／Ｎ１９－１４２３／＞．

[0222] また、回答学習装置では、根拠情報の単位として段落を用い、学習データの正解の根拠情報として、段落単位で設定したものを用いる点が、第２の実施形態と異なっている。

[0223] なお、上記以外の点が、第２の実施形態と同様であるため、同一符号を付して説明を省略する。

[0224] ＜第３の実施形態に係る回答学習装置の構成＞

図１２を参照して、第３の実施形態に係る回答学習装置５０の構成について説明する。図１２は、第３の実施形態に係る回答学習装置５０の構成を示すブロック図である。なお、上述の第２の実施形態に係る回答学習装置３０と同様の構成については、同一の符号を付して詳細な説明は省略する。

[0225] 回答学習装置５０は、ＣＰＵと、ＲＡＭと、上記図６に示した回答学習処理ルーチンと同様の処理ルーチンを実行するためのプログラムを記憶したＲＯＭとを備えたコンピュータで構成され、機能的には次に示すように構成されている。図１２に示すように、本実施形態に係る回答学習装置５０は、入力部１００と、解析部９００と、パラメータ学習部７００とを備えて構成される。

[0226] 解析部９００は、符号化部９１０と、根拠抽出部６１７と、根拠探索部２１６と、判断部２２０とを備えて構成される。

[0227] 符号化部９１０は、単語符号化部２１１と、単語データベース（ＤＢ）２

１２と、第１文脈符号化部２１３と、アテンション部２１４と、第２文脈符号化部２１５と、を備えている。

[0228] 符号化部９１０で用いられる事前学習済み言語モデルであるＢＥＲＴモデルには入力可能なテキストの長さに制約があるため、文章を段落ごとに分割して入力する。

[0229] 具体的には、ＢＥＲＴモデルの入力を、質問文と１段落のテキストを連結して生成した１テキストとする。ＢＥＲＴモデルの出力は、１つの固定長ベクトルと、１つのベクトル系列である。固定長ベクトルは、質問と１段落全体の意味をあらわすd_4次元のベクトルである。ベクトル系列の長さ（トークン数）は、入力となる「質問文と１段落のテキストを連結して生成した１テキスト」のトークン数と同じである。ベクトル系列のうちの１ベクトルが、１トークンの意味を表すd_4次元であり、ベクトル系列のサイズは、d_4次元×トークン数となる。本実施形態では、トークンの単位を単語とするが、部分単語単位などを用いてもよい。

[0230] 符号化部９１０は、質問文と１段落のテキストを連結して生成した１テキストをＢＥＲＴモデルに入力して、１つの固定長ベクトルと、１つのベクトル系列を求める処理を、文章の各段落（段落数をｎとする）について行う。これにより、符号化部９１０では、ｎ個のd_4次元ベクトルである固定長ベクトルV_1，…，V_nとｎ個のベクトル系列U_1，…，U_nを得る。V_iは、文脈符号化結果であるｉ番目の段落を表す固定長ベクトル、U_iは質問文とｉ番目の段落の各単語の意味を表すベクトル系列である。なお、段落数ｎは、上記式（１２）中の「ｎ」とは別の変数である。ｎ個のd_4次元ベクトルである固定長ベクトルV_1，…，V_nとｎ個のベクトル系列U_1，…，U_nが、質問文とスパンの意味を表すベクトル表現系列の一例である。固定長ベクトルは単語や文章の意味をベクトルで表現したものであり、「ベクトル表現」とも表される。また、ベクトル系列は文章の意味を表現する際などに、ベクトル表現を並べたものであり、これは「ベクトル表現系列」とも表される。入力文が１単語からなる場合、当該入力文に対応する「ベクトル表現系列」は「ベクト

ル表現」になるため、「ベクトル表現系列」は、「ベクトル系列」をも含む概念である。

[0231] また、符号化部９１０は、質問文の各単語の意味を表すベクトル系列U_i'を、以下の２つの処理で作成する。

[0232] 第１処理では、符号化部９１０の出力するｎ個のベクトル系列U_iの各々について、当該ベクトル系列U_iから、質問文に相当する位置のベクトルだけを抽出して、ベクトル系列を抽出する。ここで、抽出されるベクトル系列は、質問文の単語数と長さが一致し、ベクトル系列に含まれる各ベクトルは質問文の１単語を表す。

[0233] 第２処理では、ｎ個のベクトル系列U_iの各々について抽出したベクトル系列の平均又は和をとり、１つのベクトル系列を求め、ベクトル系列U_i'とする。

[0234] 根拠抽出部６１７は、符号化部９１０の処理によって得られる情報に基づいて、質問文に対する回答の根拠となる情報である根拠情報を抽出する抽出モデルを用いて、質問文Ｑに対する回答の根拠情報を抽出する。

[0235] 具体的には、根拠抽出部６１７は、まず、符号化部９１０の処理によって得られたｎ個の固定長ベクトルV_iを入力とし、ニューラルネットワークを用いて文章Ｐの各段落の意味を表すベクトルの系列Ｈを抽出する。根拠抽出部６１７は、例えば、ニューラルネットワークとして、Ｕｎｄｉｒｅｃｔｉｏｎａｌ－ＲＮＮを用いることができる。若しくは、根拠抽出部６１７は、ニューラルネットワークを用いずに、ｎ個の固定長ベクトルV_iを並べてベクトルの系列Ｈを生成してもよい。

[0236] 次に、根拠抽出部６１７は、根拠段落を１つ抽出する操作を１時刻と定義し、状態z_tを抽出モデルのＲＮＮによって生成する。すなわち、根拠抽出部６１７は、時刻ｔ－１に抽出された根拠段落に対応するベクトルの系列Ｈの要素

$$h_{st-1}$$

を抽出モデルのＲＮＮに入力することにより、状態z_tを生成する。ただし、

s_{t-1}は時刻ｔ－１に抽出された根拠段落の添字である。また、時刻ｔまでに抽出された段落s_tの集合をS_tとする。

[0237] 根拠抽出部６１７は、状態z_tと、質問文の各単語の意味を表すベクトル系列U_i’に基づいて、抽出モデルにより、時刻ｔにおける重要性を考慮した質問文ベクトルであるｇｌｉｍｐｓｅベクトルe_t（上記式（１３））を、質問文Ｑに対するｇｌｉｍｐｓｅ操作（上記参考文献５）を行うことで生成する。なお、上記式（１３）におけるベクトルy_jに、ベクトル系列U_i’のｊ番目のベクトルを代入して計算する。また、上記式（１１）～式（１３）におけるv_g、W_{g1}、及びW_{g2}と、上記式（１４）におけるv_p、W_{p1}、W_{p2}、及びW_{p3}とは、抽出モデルのパラメータであり、パラメータ学習部７００により更新される。また、上記式（１４）におけるh_jは、ベクトルの系列Ｈのｊ番目のベクトルである。

[0238] このように、抽出モデルでは質問文Ｑに対するｇｌｉｍｐｓｅ操作を行うことで、根拠段落の抽出結果が質問全体に対応する内容を包含することができる。

[0239] 抽出モデルのＲＮＮの初期値はベクトルの系列Ｈをａｆｆｉｎｅ変換したベクトル系列をｍａｘｐｏｏｌｉｎｇしたベクトルとする。

[0240] 根拠抽出部６１７は、状態z_tと、ｇｌｉｍｐｓｅベクトルe_tと、ベクトルの系列Ｈとに基づいて、抽出モデルにより、時刻ｔにおいて上記式（１４）で表される確率分布に従って第δ段落を選び、段落$s_t=\delta$を、時刻ｔに抽出された根拠段落とする。ただし、上記式（１４）における値$u_j{}^t$は、ベクトル系列U_i’とは別の値である。

[0241] そして、根拠抽出部６１７は、抽出した段落s_tの集合S_tを根拠情報として、根拠探索部２１６及びパラメータ学習部７００に渡す。

[0242] 根拠探索部２１６は、符号化部９１０の出力するｎ個のベクトル系列U_iのうち、学習データ中の正解となる根拠情報（段落）に対応するベクトル系列U_iに基づいて、文章Ｐにおける回答の根拠となる範囲Ｄ：Ｅを推定するための読解モデルを用いて、当該範囲の始端s_d及び終端s_eを推定する。

[0243] 具体的には、根拠探索部２１６は、回答の根拠となる範囲の始端s_dを推定するための始端用の深層学習モデル及び終端s_eを推定するための終端用の深層学習モデルの２つのニューラルネットワークによって構成される。この深層学習モデルは、例えば、線形変換層又は多層線形変換層などを含む。

[0244] 根拠探索部２１６は、まず、正解となる根拠情報（段落）に対応するベクトル系列U_iを、始端用の深層学習モデルに入力して、回答の根拠となる範囲である回答範囲の始端に関するスコアを、正解となる根拠情報（段落）中の各単語について計算する。

[0245] また、根拠探索部２１６は、正解となる根拠情報（段落）に対応するベクトル系列U_iを、終端用の深層学習モデルに入力して、回答の根拠となる範囲である回答範囲の終端に関するスコアを、正解となる根拠情報（段落）中の各単語について計算する。

[0246] 正解となる根拠情報（段落）中の各単語について計算した始端に関するスコア及び終端に関するスコアをまとめて回答範囲スコアと呼ぶ。なお、始端用の深層学習モデル及び終端用の深層学習モデルのパラメータは、パラメータ学習部７００により更新される。

[0247] そして、根拠探索部２１６は、計算した回答範囲スコアを、パラメータ学習部７００に渡す。

[0248] なお、根拠探索部２１６の内部の処理やニューラルネットワークの構成については、上記第２の実施形態の根拠探索部２１６と同様のものを用いてもよい。

[0249] 判断部２２０は、符号化部９１０の処理によって得られる情報に基づいて、質問文Ｑに対する回答の極性が正か否かを判断する判断モデルを用いて、質問文Ｑに対する回答の極性を判断する。

[0250] 具体的には、判断部２２０は、入力変換部２２１と、スコア計算部２２２とを備えて構成される。

[0251] 入力変換部２２１は、符号化部９１０が出力するｎ個の固定長ベクトルV_iのうち、学習データの正解の根拠情報（段落）に対応する固定長ベクトルV_i

を抽出し、抽出した固定長ベクトルV_iを、スコア計算部２２２に渡す。

[0252] スコア計算部２２２は、質問文Ｑに対する回答の極性が正か否かを判断する判断モデルを用いて、質問文Ｑに対する回答の極性を判断する。

[0253] 具体的には、スコア計算部２２２は、正解の根拠情報（段落）に対応する固定長ベクトルV_iに基づいて、線形変換層又は多層線形変換層などを含む深層学習モデルである判断モデルを用いて、Ｙｅｓ、Ｎｏ、及び抽出型の回答の何れかで答えるかの判断結果を表す判断スコアを計算する。ここで計算される判断スコアは、３次元のベクトルである。

[0254] なお、判断部２２０の内部の処理やニューラルネットワークの構成については、上記第１の実施形態の判断部２２０と同様のものを用いてもよい。例えば、Ｙｅｓ、Ｎｏ、又は抽出型の３つに分類するように構成して、上記第１の実施形態の判断部２２０と同様の処理を行うようにしてもよい。

[0255] そして、スコア計算部２２２は、判断スコアを、パラメータ学習部７００に渡す。

[0256] パラメータ学習部７００は、学習データに含まれる正解Ｙ＾と、判断部２２０により判断された結果とが一致し、学習データに含まれる始端Ｄ＾及び終端Ｅ＾と、根拠探索部２１６により推定された始端s_d及び終端s_eとが一致し、学習データに含まれる文章Ｐにおける正解の根拠情報と、根拠抽出部６１７により抽出された根拠情報とが一致するように、読解モデル、判断モデル及び抽出モデルのパラメータを学習する。

[0257] 具体的には、文章Ｐと、質問Ｑと、ＹＥＳ、ＮＯ、及び抽出型の何れかである正解となる回答Ｙ＾、回答Ｙの正解となる根拠範囲である始端Ｄ＾及び終端Ｅ＾と、回答Ｙの正解となる根拠情報である根拠段落の集合S_t＾とを１セットとし、学習データが、複数セット含むものとする。

[0258] ここで、正解となる回答Ｙ＾は、回答種別がＹＥＳ、ＮＯ、抽出型のいずれであるかを示す情報とする。正解となる根拠情報とは、回答種別を判断する、又は根拠範囲を抽出する際の根拠となるスパンとしての段落の集合であり、文章Ｐに含まれる全段落の部分集合である。回答種別が極性であるＹＥ

Ｓ又はＮＯである場合、根拠情報はＹＥＳ又はＮＯの判断の根拠となる段落からなる集合である。回答種別が抽出型の場合、根拠情報は、真の回答範囲を含む段落からなる集合である。

[0259] なお、回答学習装置５０の他の構成及び作用については、第２の実施形態と同様であるため、説明を省略する。

[0260] ＜第３の実施形態に係る回答生成装置の構成＞

図１３を参照して、本開示の第３の実施形態に係る回答生成装置６０の構成について説明する。図１３は、本開示の第３の実施形態に係る回答生成装置６０の構成を示すブロック図である。なお、上述の第３の実施形態に係る回答学習装置５０と同様の構成については、同一の符号を付して詳細な説明は省略する。

[0261] 回答生成装置６０は、ＣＰＵと、ＲＡＭと、上記図９に示した回答生成処理ルーチンと同様の処理ルーチンを実行するためのプログラムを記憶したＲＯＭとを備えたコンピュータで構成され、機能的には次に示すように構成されている。図１３に示すように、本実施形態に係る回答生成装置６０は、入力部４００と、解析部９００と、出力部８００とを備えて構成される。

[0262] 解析部９００は、符号化部９１０と、根拠抽出部６１７と、根拠探索部２１６と、判断部２２０とを備えて構成される。

[0263] 根拠抽出部６１７は、符号化部９１０の処理によって得られる情報に基づいて、質問文に対する回答の根拠となる情報である根拠情報を抽出する抽出モデルを用いて、質問文Ｑに対する回答の根拠情報を抽出すると共に、文章Ｐのスパンである各段落について根拠スコアを算出する。

[0264] 具体的には、根拠抽出部６１７は、根拠スコアを算出する際に、状態z_tと、ｇｌｉｍｐｓｅベクトルe_tと、ベクトルの系列Ｈとに基づいて、抽出モデルを用いて、文章Ｐの各段落について、当該段落が回答を抽出する根拠に適している度合いを示す根拠スコアを算出する。

[0265] 例えば、時刻ｔにおいて段落δについて上記（１４）式に従って値Ｐ（δ；$S_{(t-1)}$）を計算し、この値をスコアA_t（δ）とする。そして、段落δの

根拠スコアを、$max_t\ A_t(\delta)$又は$sum_t\ A_t(\delta)$で定義する。

[0266] なお、根拠スコアは、スパンが段落でなく文であった場合も、同様の処理で計算することができる。その場合、根拠スコアは、その文が回答を抽出する根拠に適している度合いを示すスコアを表す。

[0267] 根拠抽出部６１７は、抽出した根拠情報と、各段落についての根拠スコアとを、出力部８００に渡すと共に、抽出した根拠情報を、根拠探索部２１６及び入力変換部２２１に渡す。

[0268] 根拠探索部２１６は、符号化部９１０の出力するｎ個のベクトル系列U_iのうち、根拠抽出部６１７が根拠情報として抽出した段落に対応するベクトル系列U_iに基づいて、文章Ｐにおける回答の根拠となる範囲Ｄ：Ｅを推定するための読解モデルを用いて、当該範囲の始端s_d及び終端s_eを推定する。

[0269] 入力変換部２２１は、符号化部９１０が出力するｎ個の固定長ベクトルV_iのうち、根拠抽出部６１７が根拠情報として抽出した段落に対応する固定長ベクトルV_iを抽出し、抽出した固定長ベクトルV_iを、スコア計算部２２２に渡す。

[0270] スコア計算部２２２は、質問文Ｑに対する回答の極性が正か否かを判断する判断モデルを用いて、質問文Ｑに対する回答の極性を判断する。

[0271] 具体的には、スコア計算部２２２は、根拠抽出部６１７が根拠情報として抽出した段落に対応する固定長ベクトルV_iに基づいて、線形変換層又は多層線形変換層などを含む深層学習モデルである判断モデルを用いて、Ｙｅｓ、Ｎｏ、及び抽出型の回答の何れかで答えるかの判断結果を表す判断スコアを計算する。

[0272] 出力部８００は、根拠抽出部６１７によって抽出した段落の集合Ｓｔと、各段落の根拠スコアと、判断部２２０によって出力された、Ｙｅｓ／Ｎｏ／抽出型の３種の判断結果を示す判断スコアと、根拠探索部２１６によって得られた、各単語の回答始端スコア及び回答終端スコアを入力とする。

[0273] また、回答生成装置６０に入力する時点で、文章を構成する各段落にスコアが与えられている場合、そのスコアを検索スコアとして出力部８００に入

力することができる。例えばＷｉｋｉｐｅｄｉａ（Ｒ）全体から、当該段落を一般的な検索によって絞り込む際に、検索手法が出力するスコアを、検索スコアとして用いてもよい。

[0274] 出力部８００は、根拠抽出部６１７が根拠情報として抽出した段落の任意の単語列に対して、当該単語列の始端の回答始端スコア及び当該区間の終端の回答終端スコアの積又は和で、回答単語列スコアを計算する。回答単語列スコアは、その単語列が回答または回答の根拠であることを示すスコアである。

[0275] この回答単語列スコアの計算によって、回答種別の判断結果が「抽出型」の場合、文章中の任意の単語列、つまり抽出型の回答として考えうる全ての候補に対して、回答単語列スコアが与えられる。候補数は、根拠抽出部６１７が抽出した根拠段落の単語数をＬとすると、Ｌ個から２個を選ぶ重複組み合わせ$_{L}C_{2}$通りである。

[0276] また、出力部８００は、以下の式に従って、根拠抽出部６１７が抽出した根拠段落、回答種別の判断結果、及び回答単語列の全ての組み合わせに対して読解スコアを計算する。

[0277] 読解スコア＝（検索スコア×検索スコア重み）＋（根拠スコア×根拠スコア重み）＋（回答単語列スコア×回答単語列スコア重み）＋（判断スコア×判断スコア重み）

[0278] ただし、各スコア重みはユーザーが決定できる数値であり、例えばすべて１である。検索スコアの項は検索スコアが与えられている場合のみ用いる。

[0279] 出力部８００は、読解スコア、根拠段落、判断結果、及び回答単語列の組み合わせを、読解スコアが高い順に出力する。出力する個数は、事前に指定された個数や読解スコアの閾値で決定する。

[0280] なお、回答生成装置６０の他の構成及び作用については、第２の実施形態と同様であるため、説明を省略する。

[0281] 以上説明したように、第３の実施形態の回答生成装置によれば、根拠探索部及び判断部が、根拠抽出部が抽出した根拠情報に基づいて処理を行うため

、より明示的に根拠情報に基づいた処理が可能になり、精度良く回答することができる。また、段落毎に根拠スコアを算出して可視化することができ、ユーザーがより解釈しやすい形で回答を出力することができる。

[0282] また、根拠情報として根拠段落を抽出するため、回答学習装置に入力する教師データの作成が容易になるとと共に、精度よく回答の極性を推定できるモデルを学習することができる。

[0283] なお、本開示は、上述した実施の形態に限定されるものではなく、この発明の要旨を逸脱しない範囲内で様々な変形や応用が可能である。

[0284] 上述の実施形態では、機械読解部２１０により文章Ｐを符号化した結果と、機械読解部２１０により質問文Ｑを符号化した結果とに基づいて、ベクトルの系列$Ｐ_3$及び$Ｑ_3$を生成したが、機械読解部２１０により推定された回答の根拠となる範囲の始端$ｓ_d$及び終端$ｓ_e$の少なくとも一方、又は文章Ｐと質問文Ｑとの関係性を表すアテンション行列Ａを更に入力として、質問文Ｑに対する回答の極性が正か否かを判断する判断モデルを用いて、質問文Ｑに対する回答の極性を判断してもよい。

[0285] この場合、第２文脈符号化部２１５は、変換した読解行列Ｍを、根拠探索部２１６は、推定した回答範囲スコアを、それぞれ入力変換部２２１に渡す。

[0286] 例えば、入力変換部２２１は、ベクトルの系列$Ｐ_3$の計算方法として、下記式（１７）や、式（１８）を用いることができる。

[0287] [数14]

$$\mathrm{P}_3 = \mathrm{Linear}(B) + M \quad \cdots(17)$$

$$\mathrm{P}_3 = [B, s_d, s_e] \quad \cdots(18)$$

[0288] ただし，Ｌｉｎｅａｒ（）は線形変換を示す。

[0289] また、例えば、入力変換部２２１は、ベクトルの系列$Ｑ_3$の計算方法として、下記式（１９）を用いることができる。

[0290]

[数15]

$$\alpha_i = softmax(A_{i:}) \in \mathbb{R}^{L_Q}$$
$$\tilde{Q}_{i:} = \sum_j \alpha_{i,j} Q_{2,j:} \in \mathbb{R}^{d_1} \quad \cdots(19)$$

[0291] 同様の操作をアテンション行列A^T、ベクトルの系列Ｐに対して行い、得られたベクトル系列を、ベクトルの系列Q_3としてもよく、得られたベクトル系列にベクトルの系列Q_2を結合したものとしてもよい。

[0292] このようなバリエーションによって、入力変換部２２１で必要な変数が決定する。

[0293] また、タスク特有の問題に対処するため、スコア計算部２２２は、文ペア分類タスクの既存フレームワークに工夫を加えたものを用いることができる。

[0294] 例えば、上記ＥＳＩＭのフレームワークを用いた場合に、以下の工夫を用いることができる。

[0295] ＜＜工夫１＞＞

文章Ｐが文ではなく、文章であるので、文ペア分類タスクに比べて系列の長さL_Pが大きくなってしまう。この問題に対処するため、ｍａｘ　ｐｏｏｌｉｎｇ、ａｖｅｒａｇｅ　ｐｏｏｌｉｎｇをより長い系列向きの手法に置き換える。

[0296] 具体的には、ベクトルの系列Q_3を、ＬＳＴＭに入力したときのＬＳＴＭの出力の最終状態を使う手法や、ａｔｔｅｎｔｉｖｅ　ｐｏｏｌｉｎｇ（列方向の重み付き平均を取る操作であり、重みとしてベクトルの系列P_3の線形変換や推定した始端s_d、終端s_e等を用いる）に置き換えることができる。

[0297] ＜＜工夫２＞＞

文ペア分類タスクに比べて、上記実施形態の分類対象となるベクトルの系列P_3は、文章Ｐの情報だけでなく、質問文Ｑの情報も豊富に含んでいる傾向がある。そのため、スコア計算部２２２においてベクトルの系列Q_3を用いず

、ベクトルの系列P_3のみを用いてベクトルＪを求めてもよい。

[0298] この場合、入力変換部２２１が受け付ける情報は、（１）読解行列Ｂのみとすることができる。また、ベクトルの系列P_3への変換は、上記式（６）を用いる。このとき、Ｊの定義は、

$$J = [P_m; P_a] \in \mathbb{R}^{2d_3}$$

である。

[0299] また、回答学習装置１０は、入力された質問文Ｑが、「Ｙｅｓ又はＮｏで答えることができる質問」なのか否かを判定する質問判定部を更に備える構成としてもよい。

[0300] 質問判定部の判定方法については、ルールベースや、機械学習による判定など、従来手法を用いればよい。この場合、質問判定部の判定の結果、「Ｙｅｓ又はＮｏで答えることができる質問ではない」と判定された時には、判断部２２０からの出力（Ｙｅｓ／Ｎｏ）を行わない、すなわち、機械読解部２１０からの出力のみを行うように構成することもできる。

[0301] このように、質問判定部を備えることにより、判断部２２０の出力がＹｅｓ／Ｎｏの２値の場合、ＹｅｓかＮｏかで答えることが不適切な場合に、ＹｅｓかＮｏかで答えてしまう事を防ぐことができる。また、学習データからＹｅｓかＮｏかで答えることが不適切な質問を除外でき、より適切な学習を行うことができる。

[0302] また、判断部２２０の出力がＹｅｓ／Ｎｏ／不明の３値である場合、「不明」となった場合の意味合いが、より明確となる。質問判定部を備えていない場合、「不明」の意味は、「Ｙｅｓ又はＮｏで答えることが不適切な質問である」、又は「（文章Ｐに回答の根拠となる記載がない等の理由で）分からない」の２つが混在してしまうが、質問判定部による判定を行えば、「不明」の意味は後者に絞ることができる。

[0303] また、当該質問判定部は、回答生成装置２０に備えることもできる。回答生成装置２０は、質問判定部を備えることにより、判断部２２０の出力がＹｅｓ／Ｎｏの２値の場合、ＹｅｓかＮｏかで答えることが不適切な場合に、

ＹｅｓかＮｏかで答えてしまう事を防ぐことができる。

[0304] また、本実施形態は、回答が、Ｙｅｓ／Ｎｏの何れであるかを判断する判断モデルを用いる場合を例に説明したが、これに限定されるものではなく、判断モデルが、回答が、Ｙｅｓ／Ｎｏ／抽出型の回答の何れかであるかを判断し、抽出型の回答である場合に、出力部が、抽出型の回答として、根拠抽出部６１７により出力された根拠文、又は根拠探索部２１６により出力された回答の根拠の範囲を出力してもよい。

[0305] また、上述の実施形態では、回答の極性を、Ｙｅｓ又はＮｏである場合を例に説明したが、これに限定されるものではなく、回答の極性を、例えば、ＯＫ又はＮＧとしてもよい。

[0306] また、第３の実施形態の回答生成装置６０において、根拠抽出部６１７は、根拠情報を根拠探索部２１６へ渡さないようにしてもよい。この場合には、根拠探索部２１６が、根拠情報を用いずに、文章Ｐにおける回答の根拠となる範囲の始端及び終端を推定する。具体的には、根拠探索部２１６は、符号化部９１０の出力するｎ個のベクトル系列U_iをそのまま用いて、文章Ｐにおける回答の根拠となる範囲の始端及び終端を推定するようにすればよい。

[0307] また、第３の実施形態の回答生成装置６０において、根拠抽出部６１７から、根拠情報を判断部２２０の入力変換部２２１へ渡さないようにしてもよい。この場合には、入力変換部２２１は、符号化部９１０が出力するｎ個の固定長ベクトルV_iをそのままスコア計算部２２２に渡す。また、スコア計算部２２２は、ｎ個の固定長ベクトルV_iに基づいて、判断スコアを計算するようにすればよい。

[0308] 上記のように、根拠探索部２１６及び判断部２２０が、根拠抽出部６１７が抽出した根拠情報を用いない場合であっても、回答学習装置５０では、根拠抽出部６１７、根拠探索部２１６、判断部２２０のそれぞれのタスクに関してマルチタスク学習を行っているため、根拠情報を参考にした、根拠となる範囲の推定や判断スコアの計算が、実現される。

[0309] また、本願明細書中において、プログラムが予めインストールされている

実施形態として説明したが、当該プログラムを、コンピュータ読み取り可能な記録媒体に格納して提供することも可能である。

[0310] 以上の実施形態に関し、更に以下の付記を開示する。

[0311] （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、

前記プロセッサは、

入力された、文章の分割単位である複数のスパンに分割された文章及び質問文に基づいて、入力文を入力文の意味を表すベクトル表現系列に変換するための予め学習された符号化モデルを用いて、前記スパン及び前記質問文の意味を表すベクトル表現系列に変換し、

前記ベクトル表現系列に基づいて、前記回答を抽出する根拠に適しているスパンを抽出する予め学習された抽出モデルを用いて、前記回答を抽出する根拠に適しているスパンである根拠情報を推定し、

前記ベクトル表現系列と、前記根拠情報とに基づいて、前記文章における前記回答の根拠となる範囲である根拠範囲を抽出する予め学習された探索モデルを用いて、前記根拠範囲を推定し、

前記ベクトル表現系列と、前記根拠情報とに基づいて、前記文章における前記質問文に対する回答の正解である回答種別を判断する予め学習された判断モデルを用いて、前記質問文に対する回答種別を判断する、

回答生成装置。

[0312] （付記項２）

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、

前記プロセッサは、

入力された、文章の分割単位である複数のスパンに分割された文章及び質

問文に基づいて、入力文を入力文の意味を表すベクトル表現系列に変換するための予め学習された符号化モデルを用いて、前記スパン及び前記質問文の意味を表すベクトル表現系列に変換し、

前記ベクトル表現系列に基づいて、前記スパンが前記回答を抽出する根拠に適している度合いを示す根拠スコアを算出するための予め学習された抽出モデルを用いて、前記スパンの各々の根拠スコアを推定する、

回答生成装置であって、

前記符号化モデル及び前記抽出モデルは、

前記符号化モデルを用いて変換された前記ベクトル表現系列に基づいて、前記文章における前記回答の根拠となる範囲である根拠範囲を抽出する探索モデル、及び

前記ベクトル表現系列に基づいて、前記文章における前記質問文に対する回答の正解である回答種別を判断する判断モデル

を更に含む複数のモデルのマルチタスク学習により予め学習されたものである

回答生成装置。

（付記項３）

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、

前記プロセッサは、

文章の分割単位である複数のスパンに分割された文章と、質問文と、前記文章における前記質問文に対する回答の正解である回答種別と、前記文章における前記回答の根拠となる範囲である根拠範囲と、前記文章における前記回答の根拠となるスパンである根拠情報とを含む学習データの入力を受け付け、

複数のスパンに分割した前記文章と、前記質問文とを、入力文を入力文の意味を表すベクトル表現系列に変換するための符号化モデルを用いて、前記

スパン及び前記質問文の意味を表すベクトル表現系列に変換し、

前記ベクトル表現系列に基づいて、前記根拠情報を抽出する抽出モデルを用いて、前記根拠情報を推定し、

前記ベクトル表現系列と、前記学習データの前記根拠情報とに基づいて、前記根拠範囲を抽出する探索モデルを用いて、前記根拠範囲を推定し、

前記ベクトル表現系列と、前記学習データの前記根拠情報とに基づいて、前記質問文に対する回答種別を判断する判断モデルを用いて、前記質問文に対する回答種別を判断し、

前記抽出された前記根拠情報が前記学習データの前記根拠情報と一致し、前記推定された前記根拠範囲が前記学習データの前記根拠範囲と一致し、前記判断された前記回答種別が前記学習データの前記回答種別と一致するように、前記符号化モデル、前記抽出モデル、前記探索モデル、及び前記判断モデルのパラメータを学習する、

回答学習装置。

（付記項４）

回答生成処理を実行するようにコンピュータによって実行可能なプログラムを記憶した非一時的記憶媒体であって、

前記回答生成処理は、

入力された、文章の分割単位である複数のスパンに分割された文章及び質問文に基づいて、入力文を入力文の意味を表すベクトル表現系列に変換するための予め学習された符号化モデルを用いて、前記スパン及び前記質問文の意味を表すベクトル表現系列に変換し、

前記ベクトル表現系列に基づいて、前記回答を抽出する根拠に適しているスパンを抽出する予め学習された抽出モデルを用いて、前記回答を抽出する根拠に適しているスパンである根拠情報を推定し、

前記ベクトル表現系列と、前記根拠情報とに基づいて、前記文章における前記回答の根拠となる範囲である根拠範囲を抽出する予め学習された探索モデルを用いて、前記根拠範囲を推定し、

前記ベクトル表現系列と、前記根拠情報とに基づいて、前記文章における前記質問文に対する回答の正解である回答種別を判断する予め学習された判断モデルを用いて、前記質問文に対する回答種別を判断する、

非一時的記憶媒体。

（付記項５）

回答生成処理を実行するようにコンピュータによって実行可能なプログラムを記憶した非一時的記憶媒体であって、

前記回答生成処理は、

入力された、文章の分割単位である複数のスパンに分割された文章及び質問文に基づいて、入力文を入力文の意味を表すベクトル表現系列に変換するための予め学習された符号化モデルを用いて、前記スパン及び前記質問文の意味を表すベクトル表現系列に変換し、

前記ベクトル表現系列に基づいて、前記スパンが前記回答を抽出する根拠に適している度合いを示す根拠スコアを算出するための予め学習された抽出モデルを用いて、前記スパンの各々の根拠スコアを推定し、

前記符号化モデル及び前記抽出モデルは、

前記符号化モデルを用いて変換された前記ベクトル表現系列に基づいて、前記文章における前記回答の根拠となる範囲である根拠範囲を抽出する探索モデル、及び

前記ベクトル表現系列に基づいて、前記文章における前記質問文に対する回答の正解である回答種別を判断する判断モデル

を更に含む複数のモデルのマルチタスク学習により予め学習されたものである、

非一時的記憶媒体。

（付記項６）

回答学習処理を実行するようにコンピュータによって実行可能なプログラムを記憶した非一時的記憶媒体であって、

前記回答学習処理は、

文章の分割単位である複数のスパンに分割された文章と、質問文と、前記文章における前記質問文に対する回答の正解である回答種別と、前記文章における前記回答の根拠となる範囲である根拠範囲と、前記文章における前記回答の根拠となるスパンである根拠情報とを含む学習データの入力を受け付け、

複数のスパンに分割した前記文章と、前記質問文とを、入力文を入力文の意味を表すベクトル表現系列に変換するための符号化モデルを用いて、前記スパン及び前記質問文の意味を表すベクトル表現系列に変換し、

前記ベクトル表現系列に基づいて、前記根拠情報を抽出する抽出モデルを用いて、前記根拠情報を推定し、

前記ベクトル表現系列と、前記学習データの前記根拠情報とに基づいて、前記根拠範囲を抽出する探索モデルを用いて、前記根拠範囲を推定し、

前記ベクトル表現系列と、前記学習データの前記根拠情報とに基づいて、前記質問文に対する回答種別を判断する判断モデルを用いて、前記質問文に対する回答種別を判断し、

前記抽出された前記根拠情報が前記学習データの前記根拠情報と一致し、前記推定された前記根拠範囲が前記学習データの前記根拠範囲と一致し、前記判断された前記回答種別が前記学習データの前記回答種別と一致するように、前記符号化モデル、前記抽出モデル、前記探索モデル、及び前記判断モデルのパラメータを学習する、

非一時的記憶媒体。

符号の説明

[0313] １０、３０、５０　　回答学習装置

２０、４０、６０　　回答生成装置

１００ 入力部

２００、６００、９００　　解析部

２１０、６１０　　機械読解部

２１１ 単語符号化部

２１３ 第１文脈符号化部

２１４ アテンション部

２１５ 第２文脈符号化部

２１６ 根拠探索部

２２０ 判断部

２２１ 入力変換部

２２２ スコア計算部

３００、７００ パラメータ学習部

４００ 入力部

５００、８００ 出力部

６１７ 根拠抽出部

９１０ 符号化部

請求の範囲

[請求項1] 入力された、文章の分割単位である複数のスパンに分割された文章及び質問文に基づいて、入力文を入力文の意味を表すベクトル表現系列に変換するための予め学習された符号化モデルを用いて、前記スパン及び前記質問文の意味を表すベクトル表現系列に変換する符号化部と、

前記ベクトル表現系列に基づいて、前記質問文に対する回答を抽出する根拠に適しているスパンを抽出する予め学習された抽出モデルを用いて、前記回答を抽出する根拠に適しているスパンである根拠情報を推定する根拠抽出部と、

前記ベクトル表現系列と、前記根拠情報とに基づいて、前記文章における前記回答の根拠となる範囲である根拠範囲を抽出する予め学習された探索モデルを用いて、前記根拠範囲を推定する根拠探索部と、

前記ベクトル表現系列と、前記根拠情報とに基づいて、前記文章における前記質問文に対する回答の正解である回答種別を判断する予め学習された判断モデルを用いて、前記質問文に対する回答種別を判断する判断部と、

を備えることを特徴とする回答生成装置。

[請求項2] 前記回答種別は、前記質問文に対する回答の極性、又は前記文章から回答を抽出することであり、

前記判断部が判断した前記回答種別が、前記文章から回答を抽出することである場合、前記根拠探索部により得られた前記根拠範囲に含まれる文字列を、前記質問文に対する回答として出力する出力部

を更に備えることを特徴とする請求項１記載の回答生成装置。

[請求項3] 入力された、文章の分割単位である複数のスパンに分割された文章及び質問文に基づいて、入力文を入力文の意味を表すベクトル表現系列に変換するための予め学習された符号化モデルを用いて、前記スパン及び前記質問文の意味を表すベクトル表現系列に変換する符号化部

と、

前記ベクトル表現系列に基づいて、前記スパンが前記質問文に対する回答を抽出する根拠に適している度合いを示す根拠スコアを算出するための予め学習された抽出モデルを用いて、前記スパンの各々の根拠スコアを推定する根拠抽出部と、

を備え、

前記符号化モデル及び前記抽出モデルは、

前記符号化モデルを用いて変換された前記ベクトル表現系列に基づいて、前記文章における前記回答の根拠となる範囲である根拠範囲を抽出する探索モデル、及び

前記ベクトル表現系列に基づいて、前記文章における前記質問文に対する回答の正解である回答種別を判断する判断モデル

を更に含む複数のモデルのマルチタスク学習により予め学習されたものである

ことを特徴とする回答生成装置。

[請求項4] 文章の分割単位である複数のスパンに分割された文章と、質問文と、前記文章における前記質問文に対する回答の正解である回答種別と、前記文章における前記回答の根拠となる範囲である根拠範囲と、前記文章における前記回答の根拠となるスパンである根拠情報とを含む学習データの入力を受け付ける入力部と、

複数のスパンに分割した前記文章と、前記質問文とを、入力文を入力文の意味を表すベクトル表現系列に変換するための符号化モデルを用いて、前記スパン及び前記質問文の意味を表すベクトル表現系列に変換する符号化部と、

前記ベクトル表現系列に基づいて、前記根拠情報を抽出する抽出モデルを用いて、前記根拠情報を推定する根拠抽出部と、

前記ベクトル表現系列と、前記学習データの前記根拠情報とに基づいて、前記根拠範囲を抽出する探索モデルを用いて、前記根拠範囲を

推定する根拠探索部と、

前記ベクトル表現系列と、前記学習データの前記根拠情報とに基づいて、前記質問文に対する回答種別を判断する判断モデルを用いて、前記質問文に対する回答種別を判断する判断部と、

前記根拠抽出部により抽出された前記根拠情報が前記学習データの前記根拠情報と一致し、前記根拠探索部により推定された前記根拠範囲が前記学習データの前記根拠範囲と一致し、前記判断部により判断された前記回答種別が前記学習データの前記回答種別と一致するように、前記符号化モデル、前記抽出モデル、前記探索モデル、及び前記判断モデルのパラメータを学習するパラメータ学習部と、

を備えることを特徴とする回答学習装置。

[請求項5] 符号化部が、入力された、文章の分割単位である複数のスパンに分割された文章及び質問文に基づいて、入力文を入力文の意味を表すベクトル表現系列に変換するための予め学習された符号化モデルを用いて、前記スパン及び前記質問文の意味を表すベクトル表現系列に変換し、

根拠抽出部が、前記ベクトル表現系列に基づいて、前記質問文に対する回答を抽出する根拠に適しているスパンを抽出する予め学習された抽出モデルを用いて、前記回答を抽出する根拠に適しているスパンである根拠情報を推定し、

根拠探索部が、前記ベクトル表現系列と、前記根拠情報とに基づいて、前記文章における前記回答の根拠となる範囲である根拠範囲を抽出する予め学習された探索モデルを用いて、前記根拠範囲を推定し、

判断部が、前記ベクトル表現系列と、前記根拠情報とに基づいて、前記文章における前記質問文に対する回答の正解である回答種別を判断する予め学習された判断モデルを用いて、前記質問文に対する回答種別を判断する

ことを特徴とする回答生成方法。

[請求項6] 符号化部が、入力された、文章の分割単位である複数のスパンに分割された文章及び質問文に基づいて、入力文を入力文の意味を表すベクトル表現系列に変換するための予め学習された符号化モデルを用いて、前記スパン及び前記質問文の意味を表すベクトル表現系列に変換し、

根拠抽出部が、前記ベクトル表現系列に基づいて、前記スパンが前記質問文に対する回答を抽出する根拠に適している度合いを示す根拠スコアを算出するための予め学習された抽出モデルを用いて、前記スパンの各々の根拠スコアを推定する

回答生成方法であって、

前記符号化モデル及び前記抽出モデルは、

前記符号化モデルを用いて変換された前記ベクトル表現系列に基づいて、前記文章における前記回答の根拠となる範囲である根拠範囲を抽出する探索モデル、及び

前記ベクトル表現系列に基づいて、前記文章における前記質問文に対する回答の正解である回答種別を判断する判断モデル

を更に含む複数のモデルのマルチタスク学習により予め学習されたものである

ことを特徴とする回答生成方法。

[請求項7] 入力された、文章の分割単位である複数のスパンに分割された文章及び質問文に基づいて、入力文を入力文の意味を表すベクトル表現系列に変換するための予め学習された符号化モデルを用いて、前記スパン及び前記質問文の意味を表すベクトル表現系列に変換し、

前記ベクトル表現系列に基づいて、前記質問文に対する回答を抽出する根拠に適しているスパンを抽出する予め学習された抽出モデルを用いて、前記回答を抽出する根拠に適しているスパンである根拠情報を推定し、

前記ベクトル表現系列と、前記根拠情報とに基づいて、前記文章に

おける前記回答の根拠となる範囲である根拠範囲を抽出する予め学習された探索モデルを用いて、前記根拠範囲を推定し、

前記ベクトル表現系列と、前記根拠情報とに基づいて、前記文章における前記質問文に対する回答の正解である回答種別を判断する予め学習された判断モデルを用いて、前記質問文に対する回答種別を判断する

処理をコンピュータに実行させるための回答生成プログラム。

[請求項8] 入力された、文章の分割単位である複数のスパンに分割された文章及び質問文に基づいて、入力文を入力文の意味を表すベクトル表現系列に変換するための予め学習された符号化モデルを用いて、前記スパン及び前記質問文の意味を表すベクトル表現系列に変換し、

前記ベクトル表現系列に基づいて、前記スパンが前記質問文に対する回答を抽出する根拠に適している度合いを示す根拠スコアを算出するための予め学習された抽出モデルを用いて、前記スパンの各々の根拠スコアを推定する

処理をコンピュータに実行させるための回答生成プログラムであって、

前記符号化モデル及び前記抽出モデルは、

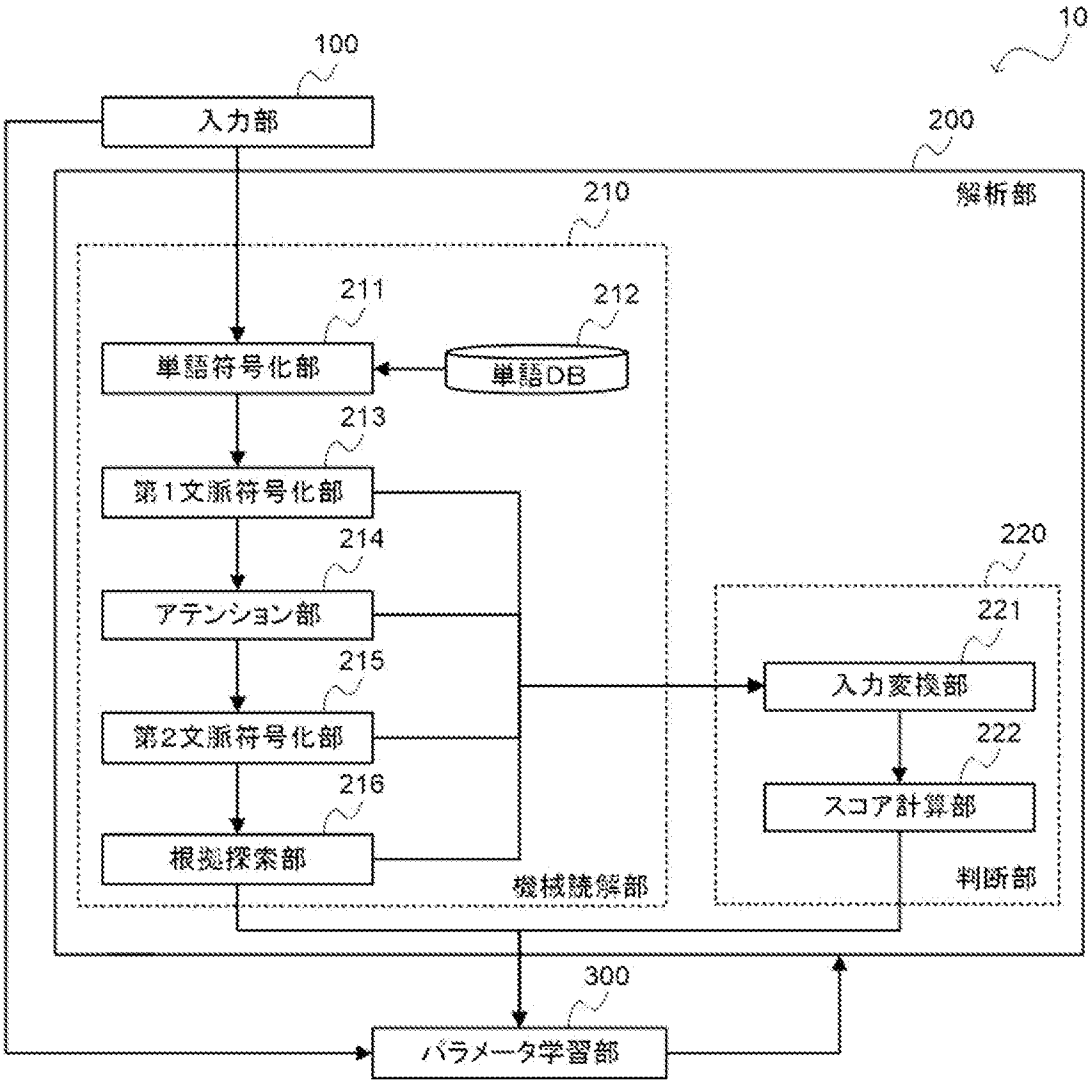
前記符号化モデルを用いて変換された前記ベクトル表現系列に基づいて、前記文章における前記回答の根拠となる範囲である根拠範囲を抽出する探索モデル、及び

前記ベクトル表現系列に基づいて、前記文章における前記質問文に対する回答の正解である回答種別を判断する判断モデル

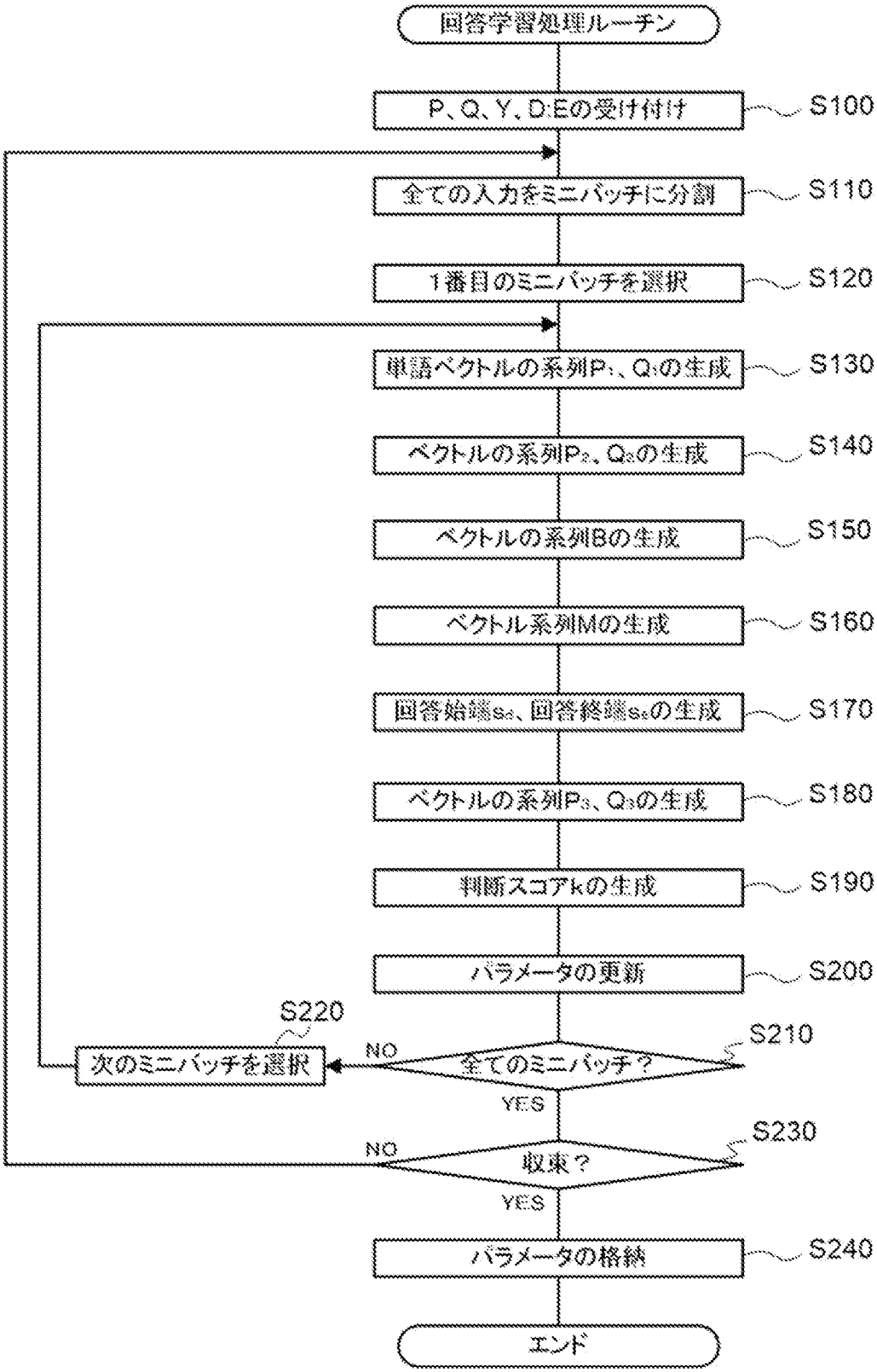
を更に含む複数のモデルのマルチタスク学習により予め学習されたものである

回答生成プログラム。

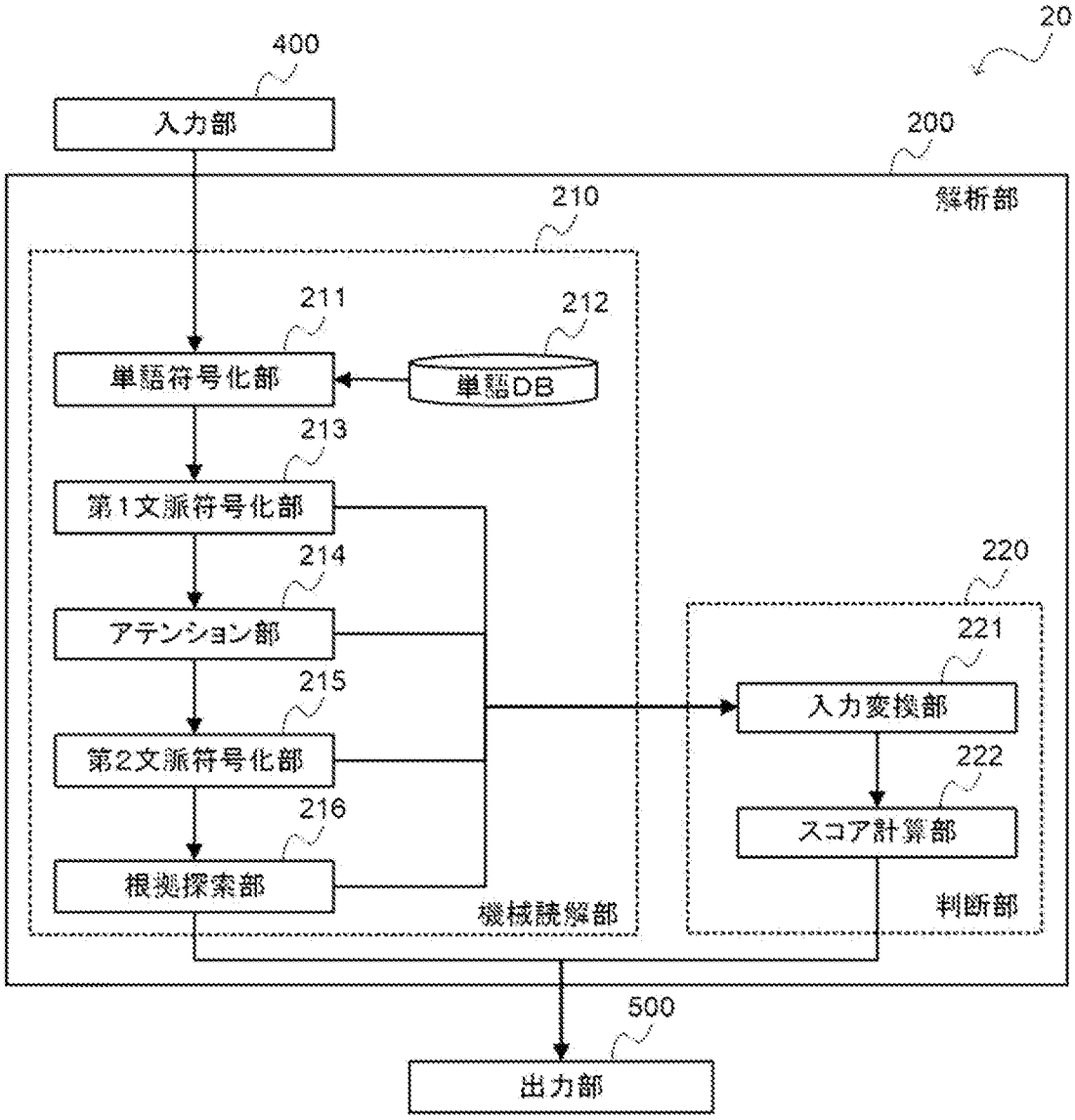
[図1]


10
100
入力部
200
解析部
210
211
212
単語符号化部
単語DB
213
第1文脈符号化部
214
アテンション部
215
第2文脈符号化部
216
根拠探索部
機械読解部
220
221
入力変換部
222
スコア計算部
判断部
300
パラメータ学習部

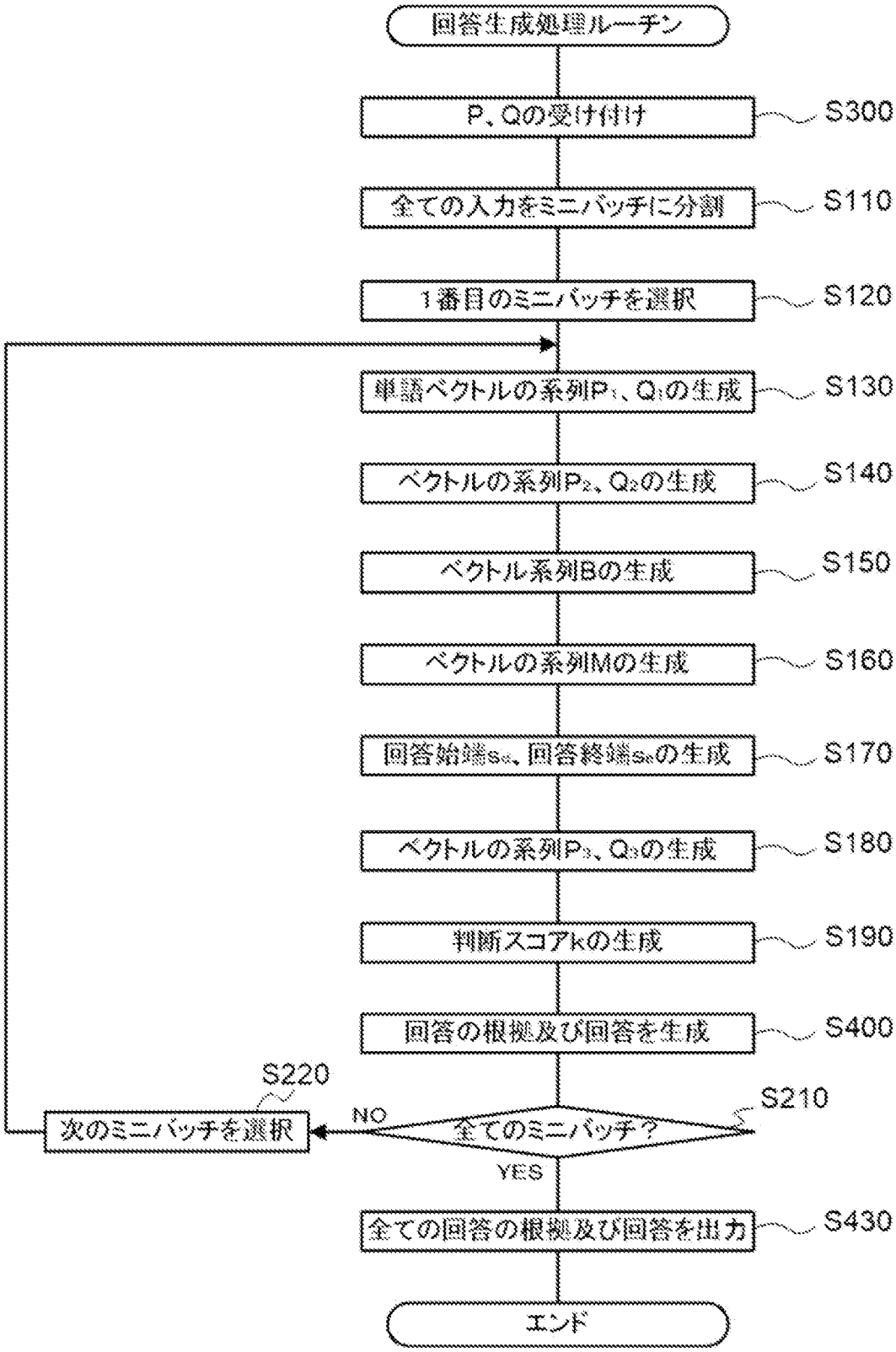
[図2]


回答学習処理ルーチン
P、Q、Y、D:Eの受け付け
S100
全ての入力をミニバッチに分割
S110
1番目のミニバッチを選択
S120
単語ベクトルの系列P1、Q1の生成
S130
ベクトルの系列P2、Q2の生成
S140
ベクトルの系列Bの生成
S150
ベクトル系列Mの生成
S160
回答始端sa、回答終端saの生成
S170
ベクトルの系列P3、Q3の生成
S180
判断スコアkの生成
S190
パラメータの更新
S200
S210
全てのミニバッチ？
NO
YES
S220
次のミニバッチを選択
S230
収束？
NO
YES
パラメータの格納
S240
エンド

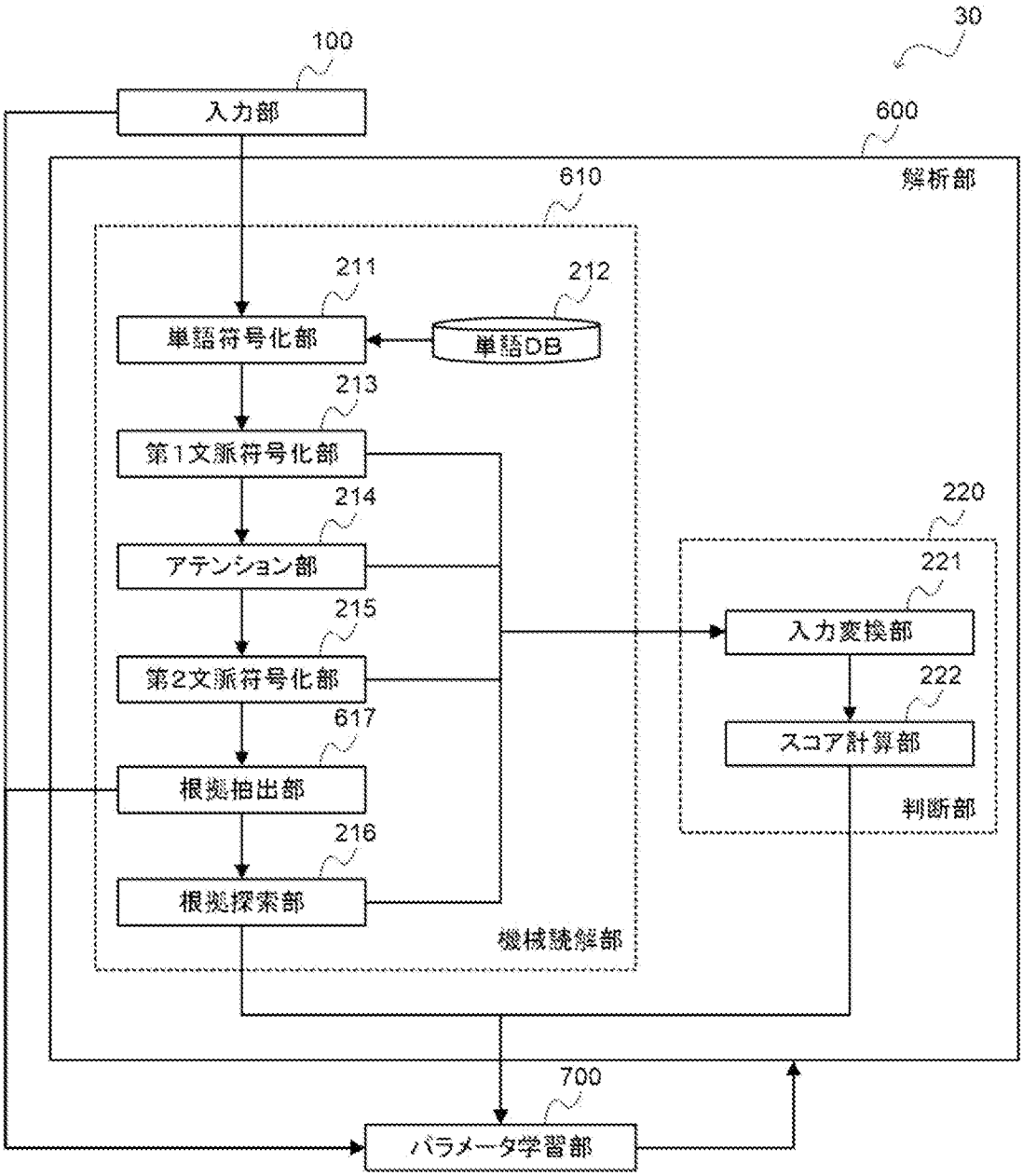
[図3]


20
400
入力部
200
解析部
210
211
212
単語符号化部
単語DB
213
第1文脈符号化部
214
アテンション部
215
第2文脈符号化部
216
根拠探索部
機械読解部
220
221
入力変換部
222
スコア計算部
判断部
500
出力部

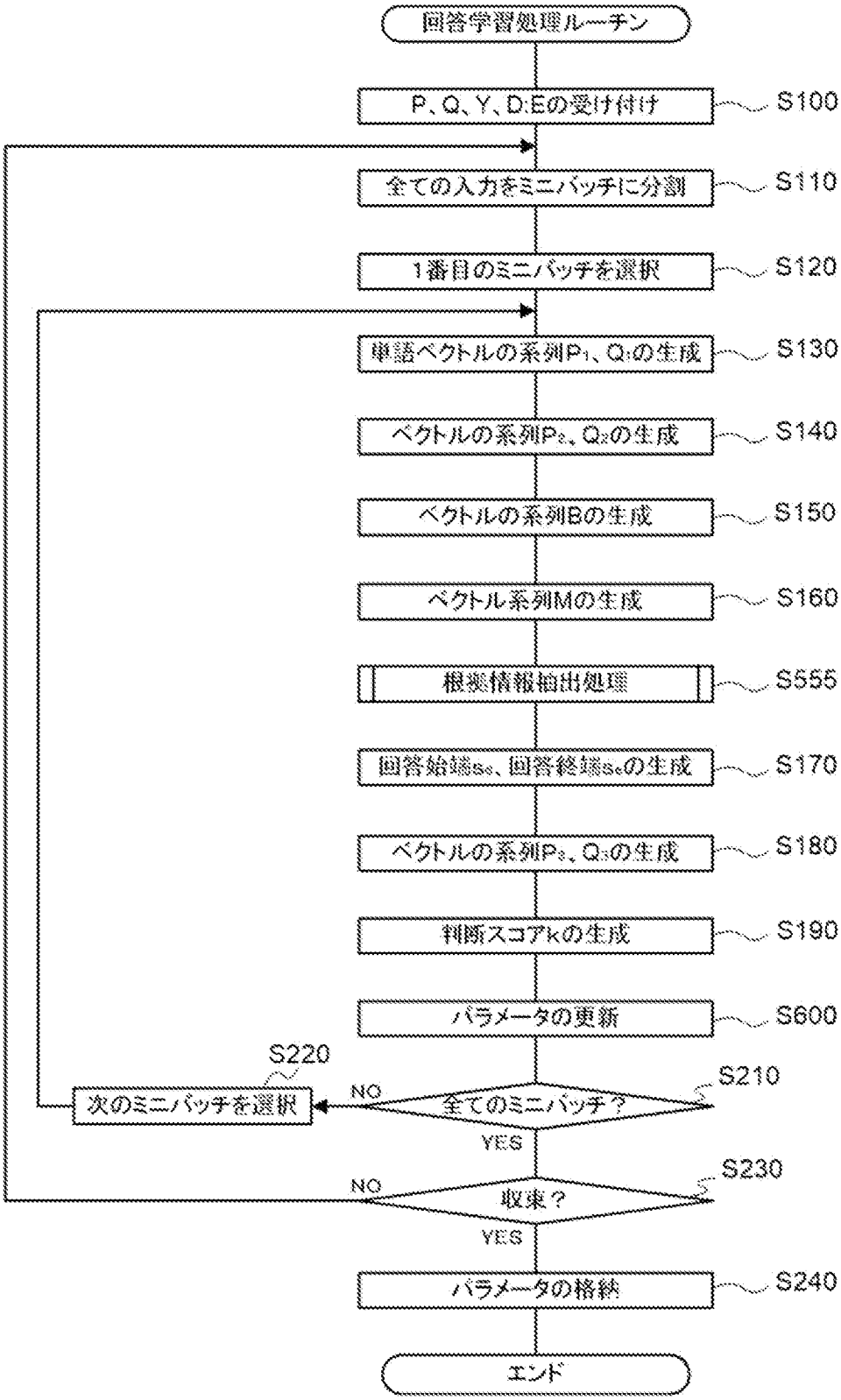
[図4]


回答生成処理ルーチン
P、Qの受け付け
S300
全ての入力をミニバッチに分割
S110
1番目のミニバッチを選択
S120
単語ベクトルの系列P1、Q1の生成
S130
ベクトルの系列P2、Q2の生成
S140
ベクトル系列Bの生成
S150
ベクトルの系列Mの生成
S160
回答始端sa、回答終端saの生成
S170
ベクトルの系列P3、Q3の生成
S180
判断スコアkの生成
S190
回答の根拠及び回答を生成
S400
S220
S210
次のミニバッチを選択
NO
全てのミニバッチ？
YES
全ての回答の根拠及び回答を出力
S430
エンド

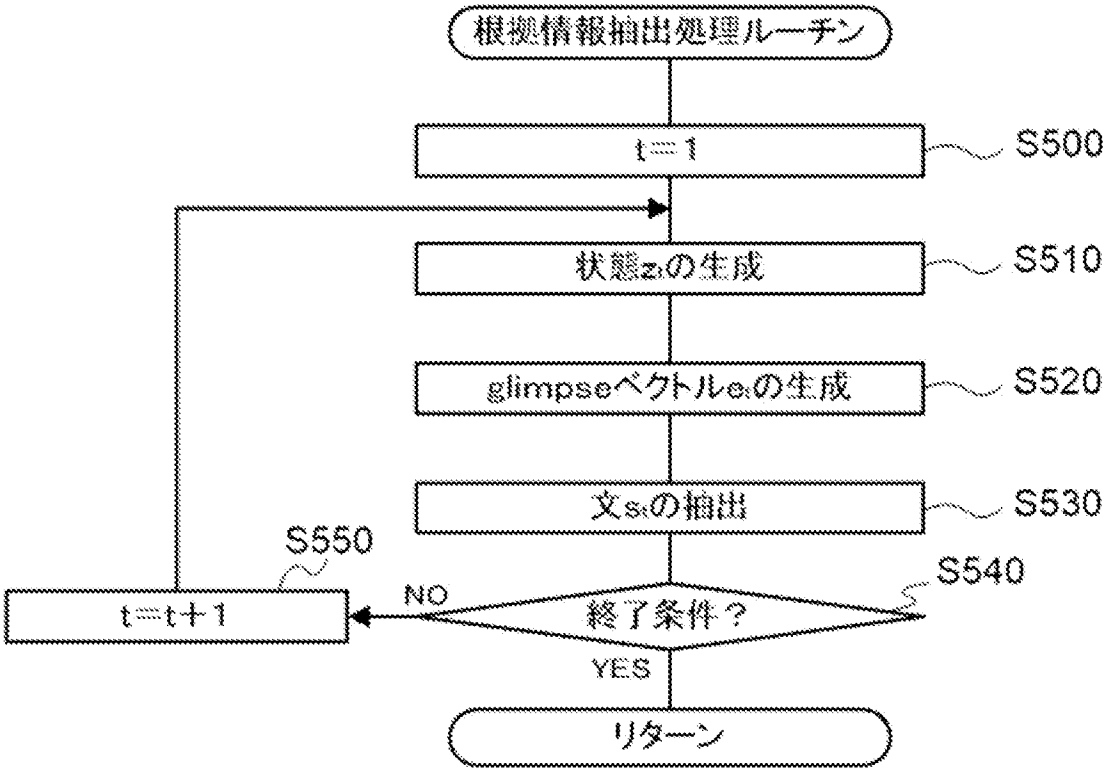
[図5]


30
100
入力部
600
解析部
610
211
212
単語符号化部
単語DB
213
第1文脈符号化部
214
アテンション部
215
第2文脈符号化部
617
根拠抽出部
216
根拠探索部
機械読解部
220
221
入力変換部
222
スコア計算部
判断部
700
パラメータ学習部

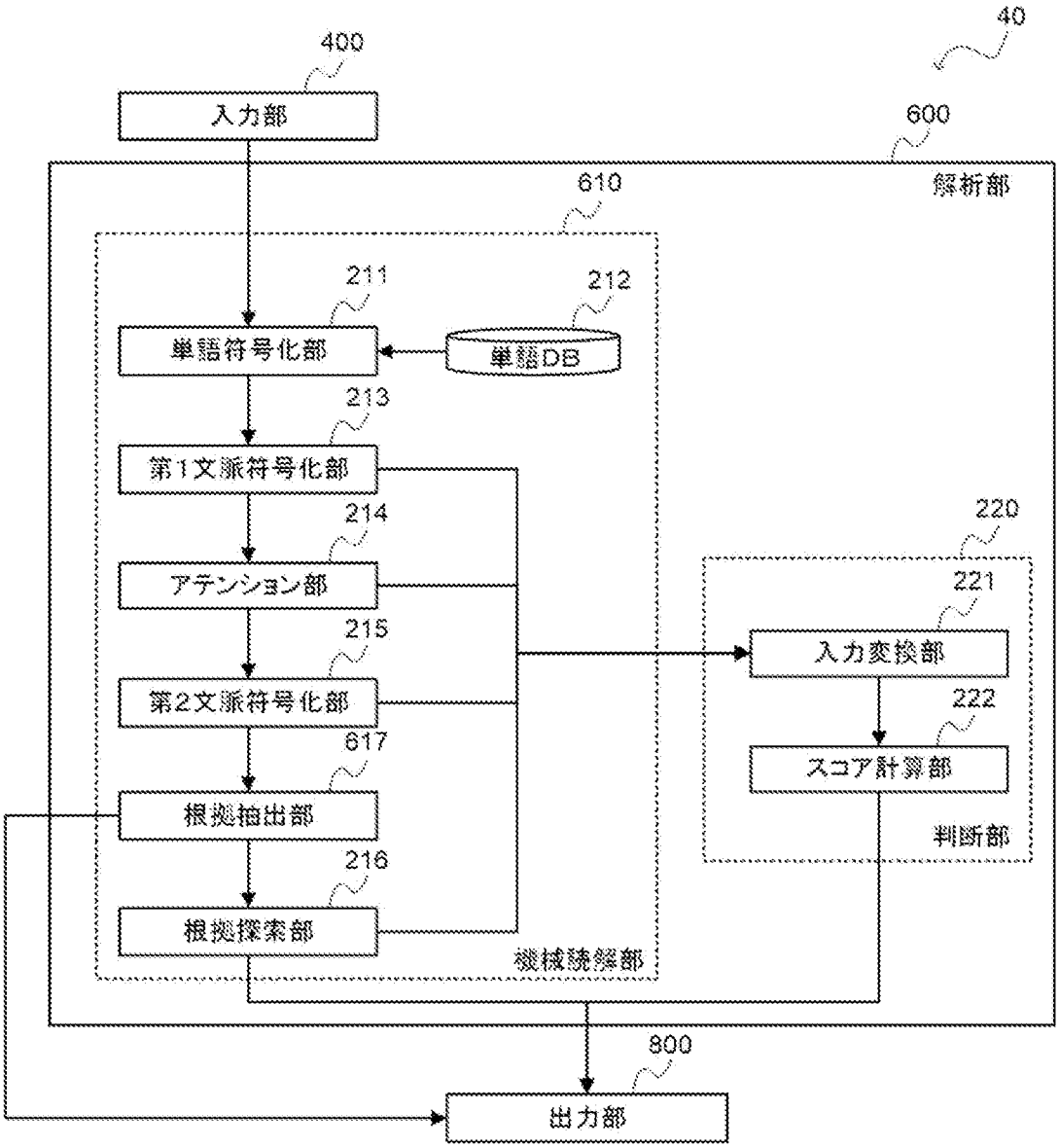
[図6]


回答学習処理ルーチン
P、Q、Y、D:Eの受け付け
S100
全ての入力をミニバッチに分割
S110
1番目のミニバッチを選択
S120
単語ベクトルの系列P1、Q1の生成
S130
ベクトルの系列P2、Q2の生成
S140
ベクトルの系列Bの生成
S150
ベクトル系列Mの生成
S160
根拠情報抽出処理
S555
回答始端s1、回答終端s2の生成
S170
ベクトルの系列P3、Q3の生成
S180
判断スコアkの生成
S190
パラメータの更新
S600
S220
次のミニバッチを選択
NO
全てのミニバッチ？
S210
YES
NO
収束？
S230
YES
パラメータの格納
S240
エンド

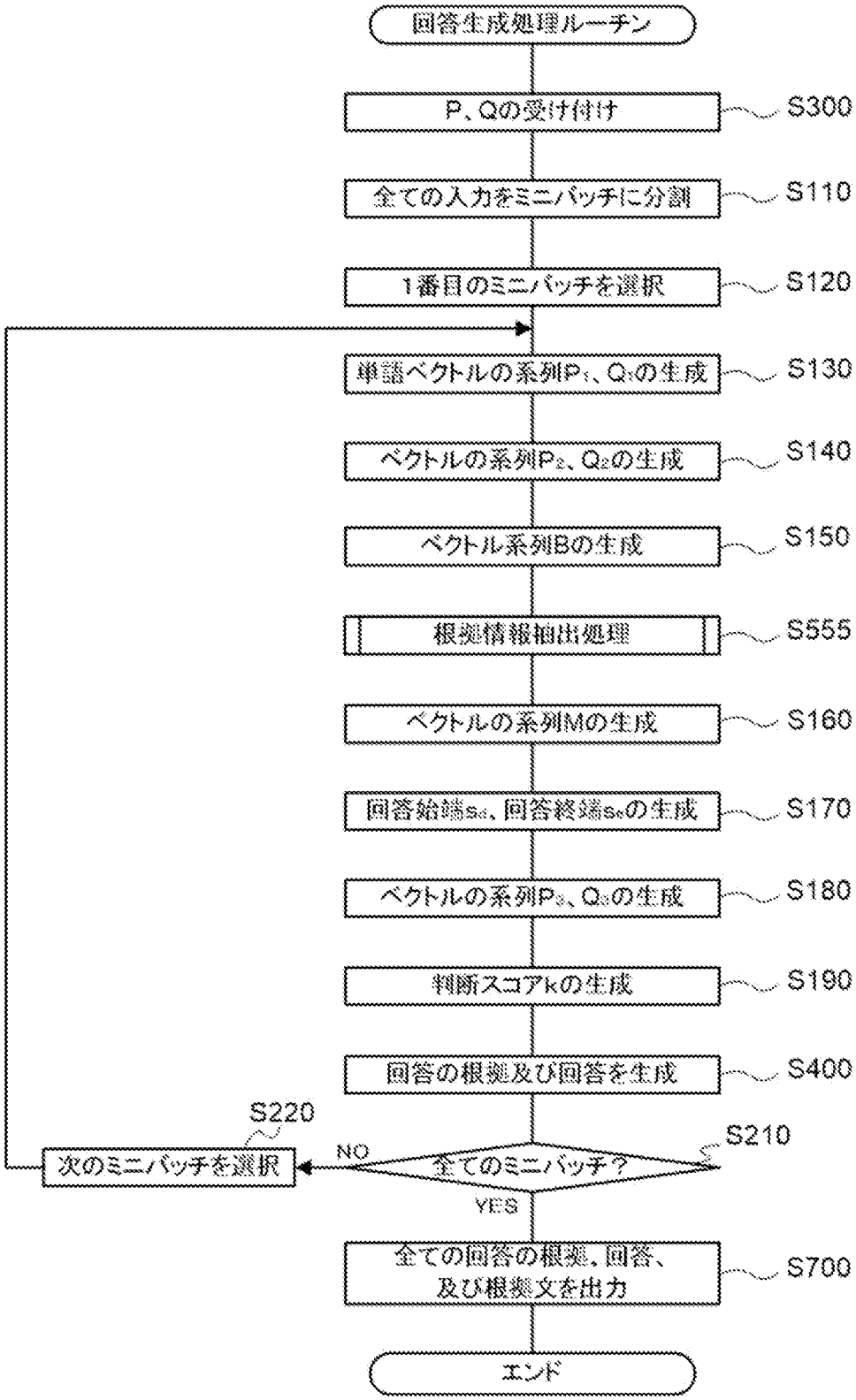
[図7]


根拠情報抽出処理ルーチン
t=1
S500
状態z_tの生成
S510
glimpseベクトルe_tの生成
S520
文s_tの抽出
S530
S540
終了条件？
NO
YES
t=t+1
S550
リターン

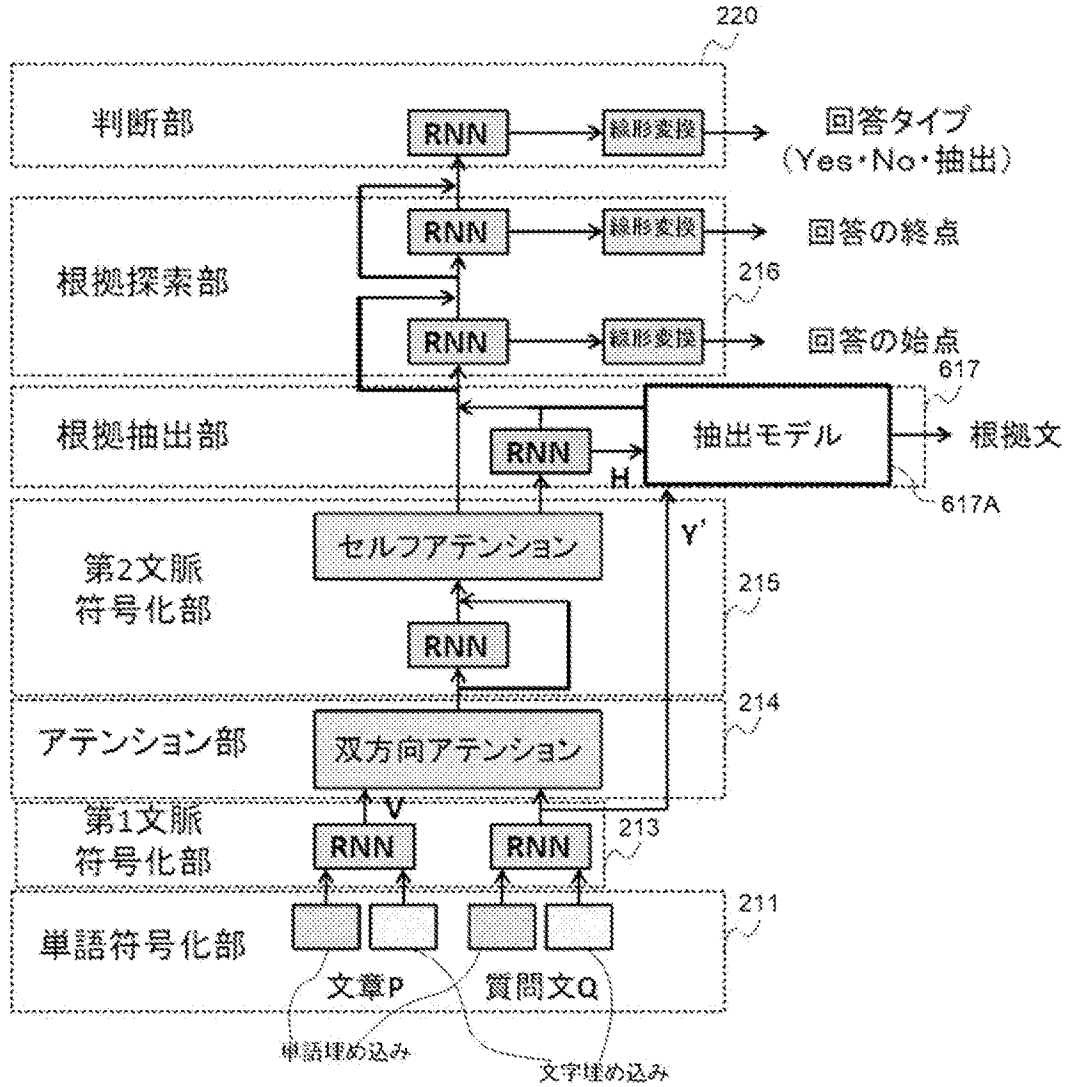
[図8]


40
400
入力部
600
解析部
610
211
212
単語符号化部
単語DB
213
第1文脈符号化部
214
アテンション部
215
第2文脈符号化部
617
根拠抽出部
216
根拠探索部
機械読解部
220
221
入力変換部
222
スコア計算部
判断部
800
出力部

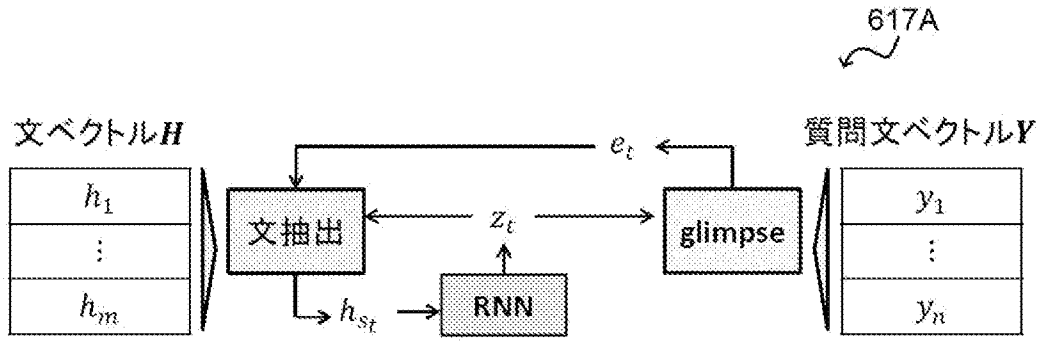
[図9]


回答生成処理ルーチン
P、Qの受け付け
S300
全ての入力をミニバッチに分割
S110
1番目のミニバッチを選択
S120
単語ベクトルの系列P1、Q1の生成
S130
ベクトルの系列P2、Q2の生成
S140
ベクトル系列Bの生成
S150
根拠情報抽出処理
S555
ベクトルの系列Mの生成
S160
回答始端sd、回答終端seの生成
S170
ベクトルの系列P3、Q3の生成
S180
判断スコアkの生成
S190
回答の根拠及び回答を生成
S400
S220
次のミニバッチを選択
NO
全てのミニバッチ？
S210
YES
全ての回答の根拠、回答、
及び根拠文を出力
S700
エンド

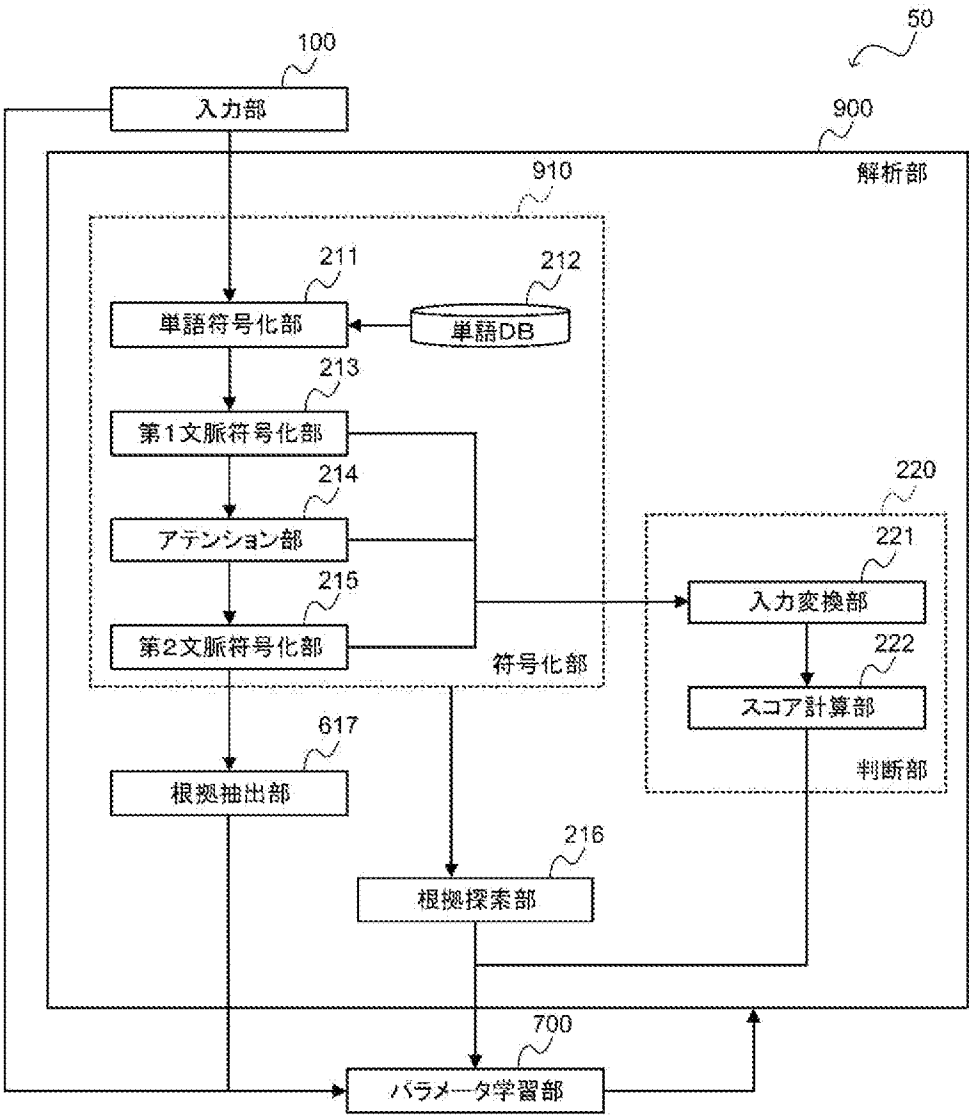
[図10]


220
判断部
RNN
線形変換
回答タイプ
（Yes・No・抽出）
根拠探索部
RNN
線形変換
回答の終点
216
RNN
線形変換
回答の始点
617
根拠抽出部
RNN
H
抽出モデル
根拠文
617A
Y'
第2文脈
符号化部
セルフアテンション
215
RNN
214
アテンション部
双方向アテンション
V
第1文脈
符号化部
RNN
RNN
213
211
単語符号化部
文章P
質問文Q
単語埋め込み
文字埋め込み

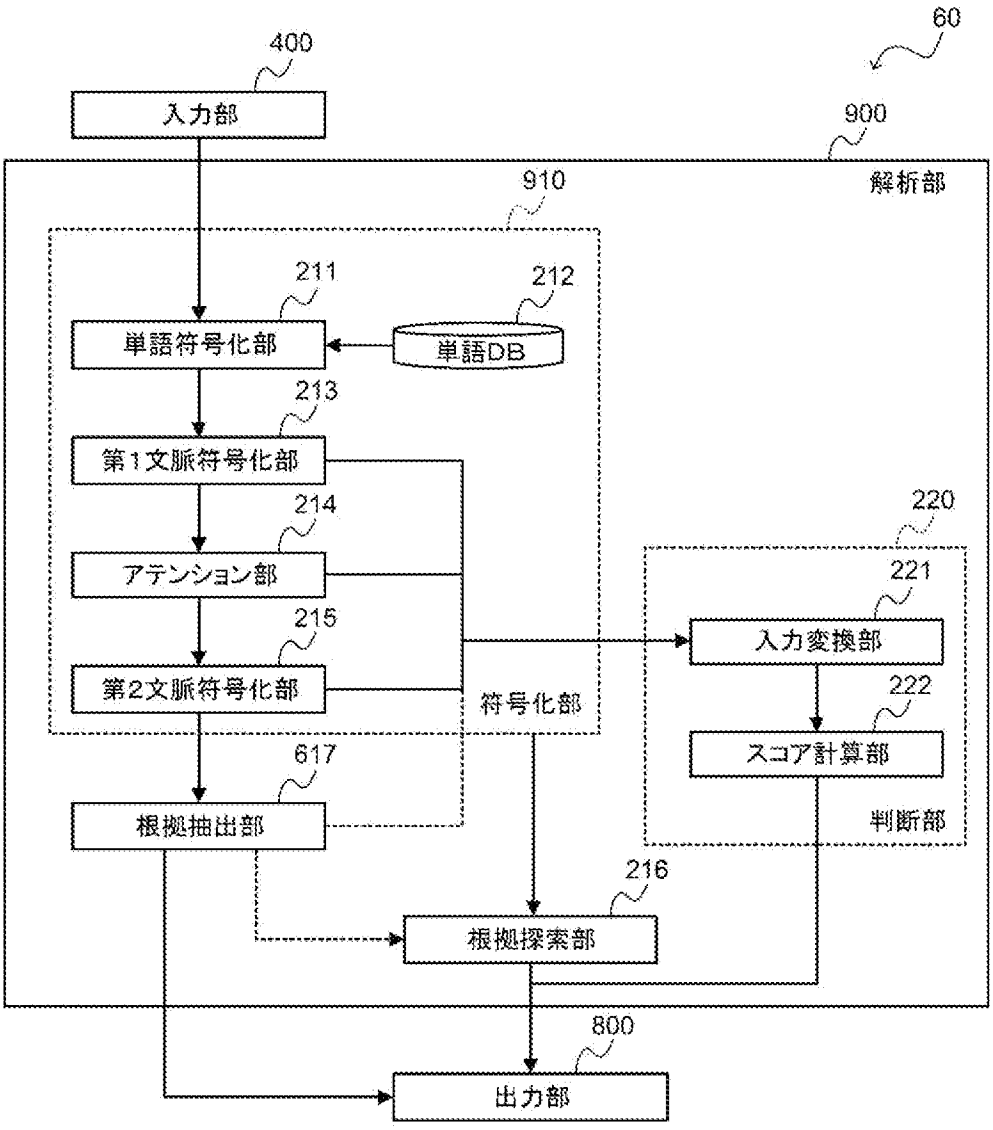
[図11]


617A
文ベクトルH
質問文ベクトルY
e_t
h_1
h_m
文抽出
z_t
glimpse
y_1
y_n
h_{st}
RNN

[図12]


50
100
入力部
900
解析部
910
211
212
単語符号化部
単語DB
213
第1文脈符号化部
214
アテンション部
215
第2文脈符号化部
符号化部
220
221
入力変換部
222
スコア計算部
判断部
617
根拠抽出部
216
根拠探索部
700
パラメータ学習部

[図13]


60
400
入力部
900
解析部
910
211
212
単語符号化部
単語DB
213
第1文脈符号化部
214
アテンション部
215
第2文脈符号化部
符号化部
220
221
入力変換部
222
スコア計算部
判断部
617
根拠抽出部
216
根拠探索部
800
出力部

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2019/049385

A. CLASSIFICATION OF SUBJECT MATTER
Int. Cl. G06F 16/90(2019.01)i
FI: G06F16/90 100

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. G06F16/90

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.
A	WO 2019/012908 A1 (NATIONAL INSTITUTE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY) 17 January 2019, entire text, all drawings	1-8
A	JP 2015-011426 A (NATIONAL INSTITUTE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY) 19 January 2015, entire text, all drawings	1-8
A	JP 2013-254420 A (NIPPON TELEGRAPH AND TELEPHONE CORP.) 19 December 2013, entire text, all drawings	1-8
A	JP 2008-282366 A (NIPPON TELEGRAPH AND TELEPHONE CORP.) 20 November 2008, entire text, all drawings	1-8

☒ 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	Date of mailing of the international search report
17.01.2020	28.01.2020
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/JP2019/049385

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2018-060273 A (NTT RESONANT INC.) 12 April 2018, entire text, all drawings	1-8
A	前田英作、外5名，質問応答システム：SAIQA 一何でも答える物知り博士, NTT R&D, 10 February 2003, vol. 52, no. 2, pp. 122-133, in particular, p. 123, left column to p. 130, right column, chapters 2-3, non-official translation (MAEDA, Eisaku et al., Question and Answer System: SAIQA - Experts that can Answer Anything	1-8

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/JP2019/049385

Patent Documents referred to in the Report	Publication Date	Patent Family	Publication Date
WO 2019/012908 A1	17.01.2019	JP 2019-020893 A entire text, all drawings	
JP 2015-011426 A	19.01.2015	US 2016/0155058 A1 entire text, all drawings WO 2014/208213 A1 entire text, all drawings EP 3016002 A1 entire text, all drawings CN 105393248 A entire text, all drawings KR 10-2016-0026892 A entire text, all drawings	
JP 2013-254420 A	19.12.2013	(Family: none)	
JP 2008-282366 A	20.11.2008	(Family: none)	
JP 2018-060273 A	12.04.2018	(Family: none)	

国際調査報告

国際出願番号
PCT/JP2019/049385

A. 発明の属する分野の分類（国際特許分類（ＩＰＣ））

G06F 16/90(2019.01)i
FI: G06F16/90 100

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））

G06F16/90

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

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
A	WO 2019/012908 A1 (国立研究開発法人情報通信研究機構) 17.01.2019 (2019-01-17) 全文,全図	1-8
A	JP 2015-011426 A (独立行政法人情報通信研究機構) 19.01.2015 (2015-01-19) 全文,全図	1-8
A	JP 2013-254420 A (日本電信電話株式会社) 19.12.2013 (2013-12-19) 全文,全図	1-8
A	JP 2008-282366 A (日本電信電話株式会社) 20.11.2008 (2008-11-20) 全文,全図	1-8
A	JP 2018-060273 A (エヌ・ティ・ティ　レゾナント株式会社) 12.04.2018 (2018-04-12) 全文,全図	1-8
A	前田 英作、外５名，質問応答システム：ＳＡＩＱＡ－何でも答える物知り博士，ＮＴＴ Ｒ＆Ｄ，2003.02.10，第５２巻，第２号，ｐ．１２２－１３３ 特に、ｐ．１２３左欄～ｐ．１３０左欄 第２～３章	1-8

☐ 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。

＊ 引用文献のカテゴリー
“A” 特に関連のある文献ではなく、一般的技術水準を示すもの
“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
“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）
“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
“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
“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
“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
“&” 同一パテントファミリー文献

国際調査を完了した日	国際調査報告の発送日
17.01.2020	28.01.2020

名称及びあて先	権限のある職員（特許庁審査官）
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	齊藤　貴孝 5N 4774 電話番号 03-3581-1101 内線 3586

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/JP2019/049385

引用文献	公表日	パテントファミリー文献	公表日
WO 2019/012908 A1	17.01.2019	JP 2019-020893 A 全文,全図	
JP 2015-011426 A	19.01.2015	US 2016/0155058 A1 全文,全図 WO 2014/208213 A1 全文,全図 EP 3016002 A1 全文,全図 CN 105393248 A 全文,全図 KR 10-2016-0026892 A 全文,全図	
JP 2013-254420 A	19.12.2013	（ファミリーなし）	
JP 2008-282366 A	20.11.2008	（ファミリーなし）	
JP 2018-060273 A	12.04.2018	（ファミリーなし）	