

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7408574号
(P7408574)

(45)発行日 令和6年1月5日(2024.1.5)

(24)登録日 令和5年12月22日(2023.12.22)

(51)国際特許分類

FI

G 0 6 F 40/216(2020.01) G 0 6 F 40/216

G 0 6 N 20/00 (2019.01) G 0 6 N 20/00

G 0 6 F 40/44 (2020.01) G 0 6 F 40/44

請求項の数 15 (全26頁)

(21)出願番号	特願2020-564520(P2020-564520)	(73)特許権者	506332063
(86)(22)出願日	令和1年5月14日(2019.5.14)		セールスフォース インコーポレイテッド
(65)公表番号	特表2021-524623(P2021-524623 A)		アメリカ合衆国 カリフォルニア州 9 4
(43)公表日	令和3年9月13日(2021.9.13)		1 0 5, サン フランシスコ, ミッショ
(86)国際出願番号	PCT/US2019/032207	(74)代理人	ン ストリート 4 1 5, サード フロアー
(87)国際公開番号	WO2019/222206		100107766
(87)国際公開日	令和1年11月21日(2019.11.21)	(74)代理人	弁理士 伊東 忠重
審査請求日	令和4年5月11日(2022.5.11)	(74)代理人	100070150
(31)優先権主張番号	62/673,606	(74)代理人	弁理士 伊東 忠彦
(32)優先日	平成30年5月18日(2018.5.18)	(74)代理人	100135079
(33)優先権主張国・地域又は機関	米国(US)	(72)発明者	弁理士 宮崎 修
(31)優先権主張番号	16/006,691		マッキャン, ブライアン
(32)優先日	平成30年6月12日(2018.6.12)		アメリカ合衆国 カリフォルニア州 9 4
	最終頁に続く		1 0 5, サンフランシスコ, ザ ランド
			マーク アット ワン マーケット ストリ
			最終頁に続く

(54)【発明の名称】 質問応答としてのマルチタスク学習

(57)【特許請求の範囲】

【請求項1】

自然言語処理のためのシステムであって、
コンテキストからの第1のワードと質問からの第2のワードを並列して符号化するための多層符号化器と、
符号化された前記コンテキストおよび符号化された前記質問を復号するための多層復号器と、
前記多層復号器からの出力に基づいて、前記コンテキストからの前記第1のワード、前記質問からの前記第2のワード、および語彙の第3のワードに対する分布を生成するためのポインタ生成器と、
スイッチであって、
前記コンテキストからの前記第1のワードに対する前記分布、前記質問からの前記第2のワードに対する前記分布、および前記語彙の前記第3のワードに対する前記分布の重みを生成し、
前記コンテキストからの前記第1のワードに対する前記分布、前記質問からの前記第2のワードに対する前記分布、および前記語彙の前記第3のワードに対する前記分布の前記重みに基づいて、複合分布を生成し、
前記複合分布を使用して、答えに含めるワードを選択するためのスイッチと、を含む、システム。

【請求項2】

前記コンテキストおよび前記質問は、質問応答、機械翻訳、文書要約、データベースクエリ生成、感情分析、自然言語推論、意味役割ラベル付け、関係抽出、目標指向対話、および代名詞解決から選択される自然言語処理タスクタイプに対応する、請求項 1 に記載のシステム。

【請求項 3】

前記多層符号化器は、

前記コンテキスト内の前記第 1 のワードと前記質問内の前記第 2 のワードとの間のコアテンションを決定するためのコアテンションネットワークと、

前記コアテンションネットワークからの出力を圧縮するための並列双方向長短期記憶と、を含む、請求項 1 または 2 に記載のシステム。

10

【請求項 4】

前記多層符号化器は、

前記コンテキストにわたるアテンションと前記質問にわたるアテンションを並列して生成するための並列セルフアテンション符号化器と、

生成された前記アテンションに基づいて、前記コンテキストおよび前記質問の最終符号化を並列して生成するための並列双方向長短期記憶と、を含む、請求項 1 ～ 3 のいずれか一項に記載のシステム。

【請求項 5】

前記多層符号化器は、

前記コンテキストからの前記第 1 のワードと前記質問からの前記第 2 のワードを並列して符号化するための並列符号化層と、

20

前記コンテキストからの前記第 1 のワードと前記質問からの前記第 2 のワードの前記符号化を並列して投影するための並列線形ネットワークと、

前記符号化の前記投影をさらに符号化するための双方向長短期記憶と、を含む、請求項 1 ～ 4 のいずれか一項に記載のシステム。

【請求項 6】

前記多層復号器は、

前記答えの中間バージョンを符号化し埋め込むための符号化および埋め込み層と、

前記答えの符号化および埋め込まれた前記中間バージョンと前記コンテキストの最終符号化との間のアテンションを生成するためのセルフアテンション復号器と、

30

前記セルフアテンション復号器の出力から中間復号器状態を生成するための長短期記憶と、

前記コンテキストの最終符号化、前記質問の最終符号化、および前記中間復号器状態に基づいて、コンテキストおよび質問復号器状態を生成するためのコンテキストおよび質問アテンションネットワークと、を含む、請求項 1 ～ 5 のいずれか一項に記載のシステム。

【請求項 7】

前記システムは、前記システムが処理するように設計されたタスクタイプの完全セットに対して訓練される前に、タスクタイプのサブセットに対して訓練される、請求項 1 ～ 6 のいずれか一項に記載のシステム。

【請求項 8】

40

前記タスクタイプの前記サブセットは、カリキュラム戦略に従って選択される、請求項 7 に記載のシステム。

【請求項 9】

前記タスクタイプの前記サブセットは、反カリキュラム戦略に従って選択される、請求項 7 に記載のシステム。

【請求項 10】

コンピュータデバイスに関連付けられた 1 つ以上のプロセッサによって実行される自然言語処理のための方法であって、

多層符号化器を使用して、コンテキストからの第 1 のワードと質問からの第 2 のワードを並列して符号化することと、

50

多層復号器を使用して、符号化された前記コンテキストおよび符号化された前記質問を復号することと、

前記多層復号器からの出力に基づいて、前記コンテキストからの前記第1のワード、前記質問からの前記第2のワード、および語彙の第3のワードに対する分布を生成することと、

前記コンテキストからの前記第1のワードに対する前記分布、前記質問からの前記第2のワードに対する前記分布、および前記語彙の前記第3のワードに対する前記分布の重みを生成することと、

前記コンテキストからの前記第1のワードに対する前記分布、前記質問からの前記第2のワードに対する前記分布、および前記語彙の前記第3のワードに対する前記分布の前記重みに基づいて、複合分布を生成することと、

前記複合分布を使用して、答えに含めるワードを選択することと、を含む、方法。

【請求項11】

前記コンテキスト内の前記第1のワードと前記質問内の前記第2のワードとの間のコアテンションを決定することをさらに含む、請求項10に記載の方法。

【請求項12】

前記コンテキストにわたるアテンションと前記質問にわたるアテンションを並列して生成することと、

前記アテンションに基づいて、前記コンテキストおよび前記質問の最終符号化を並列して生成することと、をさらに含む、請求項10または11に記載の方法。

【請求項13】

前記コンテキストからの前記第1のワードと前記質問からの前記第2のワードを並列して符号化することと、

前記コンテキストからの前記第1のワードと前記質問からの前記第2のワードの前記符号化を並列して投影することと、

前記符号化の前記投影をさらに符号化することと、をさらに含む、請求項10～12のいずれか一項に記載の方法。

【請求項14】

前記答えの中間バージョンを符号化し埋め込むことと、

前記答えの符号化および埋め込まれた前記中間バージョンと前記コンテキストの最終符号化との間のアテンションを生成することと、

生成された前記出力から中間復号器状態を生成することと、

前記コンテキストの最終符号化、前記質問の最終符号化、および前記中間復号器状態に基づいて、コンテキストおよび質問復号器状態を生成することと、をさらに含む、請求項10～13のいずれか一項に記載の方法。

【請求項15】

コンピュータデバイスに関連付けられた1つ以上のプロセッサによって実行されるときに、前記1つ以上のプロセッサが方法を行うように適合された実行可能コードを含む非一時的な機械可読媒体であって、前記方法は、

多層符号化器を使用して、コンテキストからの第1のワードと質問からの第2のワードを並列して符号化することと、

多層復号器を使用して、符号化された前記コンテキストおよび符号化された前記質問を復号することと、

前記多層復号器からの出力に基づいて、前記コンテキストからの前記第1のワード、前記質問からの前記第2のワード、および語彙の第3のワードに対する分布を生成することと、

前記コンテキストからの前記第1のワードに対する前記分布、前記質問からの前記第2のワードに対する前記分布、および前記語彙の前記第3のワードに対する前記分布の重みを生成することと、

前記コンテキストからの前記第1のワードに対する前記分布、前記質問からの前記第2

10

20

30

40

50

のワードに対する前記分布、および前記語彙の前記第3のワードに対する前記分布の前記重みに基づいて、複合分布を生成することと、

前記複合分布を使用して、答えに含めるワードを選択することと、を含む、非一時的な機械可読媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本出願は、2017年5月18日に出願された「Multitask Learning As Question Answering」と題する米国特許仮出願第62/673,606号、および2018年6月12日に出願された「Multitask Learning As Question Answering」と題する米国特許出願第16/006,691号の優先権を主張し、それらの全体が参照により本明細書に組み込まれる。

【0002】

本開示は、一般に、自然言語処理に関し、より具体的には、自然言語コンテキストに関する自然言語質問に答えることに関する。

【背景技術】

【0003】

自然言語処理と、自然言語のサンプルの内容に関する自然言語質問に答えるシステムの能力は、自然言語形式で提供される情報に関するコンテキスト特有の推論についてテストするためのベンチマークである。これは複雑な作業となる可能性がある。なぜなら、質問され得る自然言語質問には多くの異なるタイプがあり、その答えには異なるタイプの推論および／または異なるタイプの分析が必要になることがあるためである。

【0004】

従って、異なるタイプの自然言語質問に同時に答えることができるように統一されたシステムおよび方法を有することが有利であろう。

【図面の簡単な説明】

【0005】

【図1】いくつかの実施形態による質問応答として特徴付けられる自然言語処理タスクタイプの例の簡略図である。

【図2】いくつかの実施形態による計算デバイスの簡略図である。

【図3】いくつかの実施形態による符号化器の簡略図である。

【図4】いくつかの実施形態によるアテンションネットワークの簡略図である。

【図5】いくつかの実施形態によるアテンションベースのトランスフォーマネットワークのための層の簡略図である。

【図6】いくつかの実施形態による復号器の簡略図である。

【図7】いくつかの実施形態によるマルチタスク学習方法の簡略図である。

【図8】いくつかの実施形態による訓練セットの要約の簡略図である。

【図9】いくつかの実施形態によるシングルタスクおよびマルチタスク学習からの結果の簡略図である。

【図10】いくつかの実施形態による異なる訓練戦略のための結果の簡略図である。

【図11】いくつかの実施形態による答えのためのワード選択のためのソースの簡略図である。

【図12A】いくつかの実施形態による異なる訓練アプローチからの結果の簡略図である。

【図12B】いくつかの実施形態による異なる訓練アプローチからの結果の簡略図である。

【0006】

図において、同じ称呼有する要素は、同じまたは類似の機能を有する。

【発明を実施するための形態】

【0007】

自然言語情報の内容に関するコンテキスト特有の推論を含むコンテキスト特有の推論は、機械インテリジェンスおよび学習アプリケーションにおいて重要な問題である。コンテ

10

20

30

40

50

キスト特有の推論は、自然言語テキストの解釈に使用するための貴重な情報を提供することがあり、自然言語テキストの内容に関する質問への答え、言語翻訳、意味内容分析など、異なるタスクを含むことができる。しかしながら、これらの異なるタイプの自然言語処理タスクの各々は、異なるタイプの分析および／または異なるタイプの期待される答えを伴うことが多い。

【0008】

自然言語処理におけるマルチタスク学習は、タスクタイプが類似しているときに進展した。しかしながら、言語翻訳、質問応答および分類のような異なるタイプのタスクに取り組むときに、パラメータ共有は、ワードベクトルまたはパラメータのサブセットに制限されることが多い。最終的なアーキテクチャは、典型的には、各タスクタイプに対して高度に最適化され、設計されて、タスクタイプを横断的に一般化する能力を制限している。

【0009】

しかしながら、これらのタスクタイプの多くは、シングルタイプのタスクとしてフレーム化されるときに、同じアーキテクチャおよびモデルによって処理され得る。例えば、すべてではないが多くの自然言語処理タスクを、質問応答タスクとして処理することが可能である。例えば、質問応答、機械翻訳、文書要約、データベースクエリ生成、感情分析、自然言語推論、意味役割ラベル付け、関係抽出、目標指向対話、および代名詞解決のタスクタイプは、質問応答タスクとしてフレーム化されてもよい。図1は、いくつかの実施形態による質問応答として特徴付けられる自然言語処理タスクタイプの例の簡略図である。図1の各例は、質問、コンテキスト、および正解を備えた3つの列形式で示される。実施例105は、質問応答タスクの例であり、ここで、質問が、コンテキストの内容に関して尋ねられている。実施例110は、英語からドイツ語への機械翻訳タスクの例であり、ここで、質問が「what is the translation . . .」の形式で提示されている。実施例115は、文書要約タスクの例であり、ここで、質問が「what is the summary?」として提示されている。実施例120は、自然言語推論タスクの例であり、ここで、仮説と、その仮説がコンテキストに伴意するか、矛盾するか、または中立であるかについての問いとを使用して提示されている。実施例125は、感情分析タスクの例であり、ここで、コンテキストが肯定的か否定的かについて質問が提示されている。実施例130は、意味役割ラベル付けタスクの例であり、ここで、どのエンティティがコンテキストにおいて指示された役割を演じるかについて質問が提示されている。実施例135は、関係抽出タスクの例であり、ここで、コンテキストにおいて言及されたエンティティのうちの1つの関係に関して質問が提示されている。実施例140は、目標指向対話タスクの例であり、ここで、コンテキストから学習され得る知識に関して質問が提示されている。実施例145は、データベースクエリ生成タスクの例であり、ここで、質問が、データベースクエリ言語（例えば、SQL）への変換を要求している。実施例150は、代名詞解決タスクの例であり、ここで、質問が、コンテキスト中に現れる代名詞に関する質問に答えることに指向されている。

【0010】

図2は、いくつかの実施形態による計算デバイス200の簡略図である。図2に示すように、計算デバイス200は、メモリ220に結合されたプロセッサ210を含む。計算デバイス200の動作は、プロセッサ210によって制御される。また、計算デバイス200は、1つのプロセッサ210のみで示されているが、プロセッサ210は、計算デバイス200内の1つ以上の中央処理ユニット、マルチコアプロセッサ、マイクロプロセッサ、マイクロコントローラ、デジタル信号プロセッサ、フィールドプログラマブルゲートアレイ（FPGA）、特定用途向け集積回路（ASIC）、グラフィックス処理ユニット（GPU）などを代表するものであってもよいと理解される。計算デバイス200は、スタンドアロン型のサブシステムとして、計算デバイスに追加されたボードとして、および／または仮想マシンとして実装されてもよい。

【0011】

メモリ220は、計算デバイス200および／または計算デバイス200の動作中に使

10

20

30

40

50

用される１つ以上のデータ構造によって実行されるソフトウェアを記憶するために使用されてもよい。メモリ２２０は、１つ以上のタイプの機械可読媒体を含んでもよい。機械可読媒体のいくつかの一般的な形態は、フロッピーディスク、フレキシブルディスク、ハードディスク、磁気テープ、他の任意の磁気媒体、ＣＤ－ＲＯＭ、他の任意の光学媒体、パンチカード、紙テープ、穴のパターンを有する他の任意の物理媒体、ＲＡＭ、ＰＲＯＭ、ＥＰＲＯＭ、ＦＬＡＳＨ－ＥＰＲＯＭ、他の任意のメモリチップまたはカートリッジ、および／またはプロセッサまたはコンピュータが読むように適合される他の任意の媒体を含んでもよい。

【００１２】

プロセッサ２１０および／またはメモリ２２０は、任意の適切な物理的配置で配置されてもよい。いくつかの実施形態において、プロセッサ２１０および／またはメモリ２２０は、同じボード、同じパッケージ（例えば、システム・イン・パッケージ）、同じチップ（例えば、システム・オン・チップ）などの上に実装されてもよい。いくつかの実施形態において、プロセッサ２１０および／またはメモリ２２０は、分散、仮想化、および／またはコンテナ化された計算リソースを含んでもよい。そのような実施形態と一致して、プロセッサ２１０および／またはメモリ２２０は、１つ以上のデータセンタおよび／またはクラウド計算施設に位置してもよい。

【００１３】

図示のように、メモリ２２０は、本明細書でさらに説明される質問応答システムおよびモデルを実装および／またはエミュレートするために、および／または本明細書でさらに説明される任意の方法を実装するために使用され得る質問応答モジュール２３０を含む。いくつかの例において、質問応答モジュール２３０は、自然言語コンテキストに関する自然言語質問に答えるために使用されてもよい。いくつかの例において、質問応答モジュール２３０はまた、自然言語コンテキストに関する自然言語質問に答えるために使用される質問応答システムまたはモデルの反復訓練および／または評価を処理してもよい。いくつかの例において、メモリ２２０は、１つ以上のプロセッサ（例えば、プロセッサ２１０）によって実行されるときに、１つ以上のプロセッサが本明細書にさらに詳細に説明される計数方法を行うようにし得る実行可能コードを含む、非一時的、有形、機械可読媒体を含んでもよい。いくつかの例において、質問応答モジュール２３０は、ハードウェア、ソフトウェア、および／またはハードウェアとソフトウェアの組み合わせを使用して実装されてもよい。図示のように、計算デバイス２００は、自然言語コンテキスト２４０および自然言語コンテキスト２４０に関する自然言語質問２５０を受信し、これらは質問応答モジュール２３０に提供され、質問応答モジュール２３０は、次に、自然言語コンテキスト２４０の内容に基づいて、自然言語質問２５０に対する自然言語応答２６０を生成する。

【００１４】

図３は、いくつかの実施形態による符号化器３００の簡略図である。符号化器３００は、自然言語コンテキスト c および自然言語質問 q を受信し、これらは、各々、英語、フランス語、ドイツ語、スペイン語などのような自然言語で順序付けられたワードのシーケンスからなる。コンテキスト c および質問 q の両方は、それぞれの符号化層３１０および３１５を用いて、行列形式に符号化され、ここで、行列の i 番目の行は、式１に従って示されるように、シーケンス中の i 番目のトークンまたはワードに対する d_{emb} 次元埋め込みに対応し、ここで、 l は、コンテキスト中のワードまたはトークンの数に対応し、 m は、質問中のワードまたはトークンの数に対応する。いくつかの例において、各ワードの符号化は、 $Glove$ 符号化に基づき、ここで、各ワードが

[外１]

$\mathbb{R}^{300}$

の要素として符号化される。いくつかの例において、各ワードの符号化は、文字 $n-gram$ 符号化に基づき、ここで、各ワードが

〔外 2〕

 $\mathbb{R}^{100}$

の要素として符号化される。いくつかの例において、各ワードの符号化はG l o V e符号化と文字のn - g r a m符号化の連結に基づく。いくつかの例において、ワードに対してG l o V eおよび／または文字n - g r a m符号化がないとき（例えば、ワードが英語でないとき）に、ランダム符号化は、G l o V e符号化と同じ平均および標準偏差（例えば、平均ゼロおよび標準偏差0.4）を有する正規分布から選択され、そのランダム符号化がそれぞれのワードの各出現に対して一貫して使用される。いくつかの例において、ワードに対してG l o V eおよび／または文字n - g r a m符号化が存在しないとき（例えば、ワードが英語でないとき）に、ゼロの値が使用される。いくつかの例において、訓練中に、正解は、式1にも示されるように、同様に符号化され、ここで、nは、正解におけるワードまたはトークンの数に対応する。

〔0015〕

〔数1〕

$$\begin{aligned} C &\in \mathbb{R}^{l \times d_{\text{emb}}} \\ Q &\in \mathbb{R}^{m \times d_{\text{emb}}} \\ A &\in \mathbb{R}^{n \times d_{\text{emb}}} \end{aligned} \quad \text{式1}$$

〔0016〕

独立表現 いくつかの例において、符号化されたコンテキストおよび質問に対する独立表現C_{proj}およびQ_{proj}は、それぞれ、式2に従ってそれぞれの線形ネットワーク320および325を使用して生成される。いくつかの例において、それぞれの線形ネットワーク320および325は、符号化されたコンテキストおよび質問の次元をdに減らしてもよい。いくつかの例において、次元数dは200である。式2に示されるように、線形ネットワーク320および325の両方は、同じ重み行列W₁および同じバイアスb₁を使用するため、独立表現C_{proj}およびQ_{proj}は、d次元空間に一貫して投影される。いくつかの例において、バイアスb₁は任意選択であり、省略されてもよい。いくつかの例において、重み行列W₁およびバイアスb₁は、符号化器300の残りの部分と共に訓練される。

〔0017〕

〔数2〕

$$\begin{aligned} CW_1 + b_1 &= C_{\text{proj}} \in \mathbb{R}^{l \times d} \\ QW_1 + b_1 &= Q_{\text{proj}} \in \mathbb{R}^{m \times d} \end{aligned} \quad \text{式2}$$

〔0018〕

コンテキストC_{proj}および質問Q_{proj}のための投影表現は、式3に従ってC_{ind}およびQ_{ind}を形成するために、共有された1層の双方向長短期記憶ネットワーク（BiLSTM: bidirectional Long Short-Term Memory Network）330に渡される。

〔0019〕

〔数3〕

$$\begin{aligned} \text{BiLSTM}(C_{\text{proj}}) &= C_{\text{ind}} \in \mathbb{R}^{l \times d} \\ \text{BiLSTM}(Q_{\text{proj}}) &= Q_{\text{ind}} \in \mathbb{R}^{m \times d} \end{aligned} \quad \text{式3}$$

〔0020〕

B i L S T M 3 3 0は、式4に従って、
 [外3]

$$h_i^{\rightarrow}$$

および

[外4]

$$h_i^{\leftarrow}$$

10

の連結として h_i としての各時間ステップ*i*において出力を生成し、ここで、 x は、**b i L S T M 3 3 0**への入力であり、**L S T M**は、長短期記憶ネットワークに対応する。いくつかの例において、**B i L S T M 3 3 0**は、入力に対して0.2のドロップアウトを使用する。

[0021]

【数4】

$$\begin{aligned} h_i^{\rightarrow} &= \text{LSTM}(x_i, h_{i-1}^{\rightarrow}) \\ h_i^{\leftarrow} &= \text{LSTM}(x_i, h_{i+1}^{\leftarrow}) \end{aligned} \quad \text{式4}$$

20

[0022]

アライメント **b i L S T M 3 3 0**の出力 C_{ind} および Q_{ind} は、コアテンションネットワーク (**Coattention network**) 340を使用して、コアテンド表示 (**coattended representation**) C_{proj} 、 Q_{proj} 、 C_{sum} 、 Q_{sum} 、 C_{coa} および Q_{coa} を生成するために使用される。コアテンションネットワーク340は、まず、コンテキストおよび質問のシーケンスの符号化表現を整列する。いくつかの例において、質問またはコンテキストのシーケンスからのいくつかのトークンは、質問とコンテキストシーケンスの他のものにおけるいかなるトークンともうまく整合しないことがあるため、別々に訓練されたダミー埋め込みが C_{ind} および Q_{ind} の各々に追加され、これらは、

30

[外5]

$$\mathbb{R}^{(l+1) \times d}$$

および

[外6]

$$\mathbb{R}^{(m+1) \times d}$$

にある。

40

[0023]

コアテンションネットワーク340は、次に、式5に従って、質問およびコンテキストに対する符号化トークンの各々の間の正規化されたドット積類似性スコアを生成し、ここで、 $\text{softmax}(X)$ は、行列 X の各列を正規化し、合計が1になるようにエントリを有する列ごとのソフトマックスを示す。

[0024]

【数5】

$$\begin{aligned} \text{softmax}(C_{ind} Q_{ind}^T) &= S_{cq} \in \mathbb{R}^{(1 \times 1) \times (m+1)} \\ \text{softmax}(Q_{ind} C_{ind}^T) &= S_{qc} \in \mathbb{R}^{(m+1) \times (1+1)} \end{aligned} \quad \text{式5}$$

50

【 0 0 2 5 】

デュアルコアテンション 次いで、コアテンションネットワーク 3 4 0 は、正規化されたドット積類似性スコア $S_{c q}$ および $S_{q c}$ を重みとして使用し、式 6 に従って、コンテキストおよび質問のシーケンスから、コンテキストおよび質問のシーケンスの他方において各トークンに関連する情報の要約を決定する。

【 0 0 2 6 】

【数 6】

$$\begin{aligned} S_{c q}^T C_{ind} &= C_{sum} \in \mathbb{R}^{(m+1) \times d} \\ S_{q c}^T Q_{ind} &= Q_{sum} \in \mathbb{R}^{(l+1) \times d} \end{aligned} \quad \text{式 6}$$

10

【 0 0 2 7 】

次いで、コアテンションネットワーク 3 4 0 は、式 7 に従って、正規化されたドット積類似性スコア $S_{c q}$ および $S_{q c}$ の使用と共にコアテンド表示 C_{sum} および Q_{sum} を使用して、アライメントから得られた情報を元のシーケンスに戻すように伝達する。

【 0 0 2 8 】

【数 7】

$$\begin{aligned} S_{q c}^T C_{sum} &= C_{coa} \in \mathbb{R}^{(l+1) \times d} \\ S_{c q}^T Q_{sum} &= Q_{coa} \in \mathbb{R}^{(m+1) \times d} \end{aligned} \quad \text{式 7}$$

20

【 0 0 2 9 】

コアテンド表示 C_{coa} と Q_{coa} の第 1 の列は、以前に追加されたダミー埋め込みに対応する。この情報は必要とされないため、コアテンションネットワーク 3 4 0 は、行列のその列をドロップして、

【外 7】

$$C_{coa} \in \mathbb{R}^{l \times d}$$

および

30

【外 8】

$$Q_{coa} \in \mathbb{R}^{m \times d}$$

を得る。

【 0 0 3 0 】

圧縮 2 つの $BiLSTM_{350}$ および 355 が、2 つの別々のグループにあるコアテンションネットワーク 3 4 0 からの出力を圧縮するために使用される。コアテンションネットワーク 3 4 0 からの出力は、2 つのグループに連結され、各グループは、式 8 および式 9 に従って、それぞれ、 $BiLSTM_{350}$ および 355 によって処理される。いくつかの例において、 $BiLSTM_{350}$ および 355 は、入力に対して 0.2 のドロップアウトを使用する。

40

【 0 0 3 1 】

【数 8】

$$BiLSTM_{comC}([C_{proj}; C_{ind}; Q_{sum}; C_{coa}]) = C_{com} \in \mathbb{R}^{l \times d} \quad \text{式 8}$$

【 0 0 3 2 】

【数 9】

50

$$\text{BiLSTM}_{\text{comQ}}([Q_{\text{proj}}; Q_{\text{ind}}; C_{\text{sum}}; Q_{\text{coa}}]) = Q_{\text{com}} \in \mathbb{R}^{m \times d} \quad \text{式 9}$$

【 0 0 3 3 】

セルフアテンション b i L S T M 3 5 0 および 3 5 5 の出力 C_{com} および Q_{com} は、次に、それぞれの多層セルフアテンションベースのトランスフォーマの符号化器部分に渡される。より具体的には、b i L S T M 3 5 0 の出力 C_{com} はセルフアテンション符号化器 3 6 0 に渡され、セルフアテンション符号化器 3 6 0 の出力 C_{self1} はセルフアテンション符号化器 3 7 0 に渡されて、出力 C_{self2} を生成する。並列に、b i L S T M 3 5 5 の出力 Q_{com} はセルフアテンション符号化器 3 6 5 に渡され、セルフアテンション符号化器 3 6 5 の出力 Q_{self1} はセルフアテンション符号化器 3 7 5 に渡されて、出力 C_{self2} を生成する。いくつかの例において、セルフアテンション符号化器 3 6 0 および 3 7 0 は、コンテキスト内の長距離依存性を捕捉し、セルフアテンション符号化器 3 6 5 および 3 7 5 は、質問内の長距離依存性を捕捉する。また、符号化器 3 0 0 は、コンテキストおよび質問の各々について、2つのセルフアテンション符号化器とともに示されているが、符号化器 3 0 0 は、コンテキストおよび質問の一方または両方について、1つのセルフアテンション符号化器のみまたは3つ以上のセルフアテンション符号化器を含んでもよい。各セルフアテンション符号化器 3 6 0、3 6 5、3 7 0、および 3 7 5 は、位置ごとに完全に接続されたフィードフォワードネットワークに続いて、図 4 および図 5 に関して以下にさらに詳細に説明されるように、レジデュアルコネクション (r e s i d u a l c o n n e c t i o n) および層正規化と共に、マルチヘッドセルフアテンション機構を含む。

【 0 0 3 4 】

図 4 は、いくつかの実施形態によるアテンションネットワーク 4 0 0 の簡略図である。図 4 に示されるように、アテンションネットワーク 4 0 0 は、クエリ

【 外 9 】

$$q \in \mathbb{R}^{d_q}$$

、キー

【 外 1 0 】

$$k \in \mathbb{R}^{d_k}$$

、および値

【 外 1 1 】

$$v \in \mathbb{R}^{d_v}$$

を受信する。q、k、およびvの各々は、式 1 0 に従って、それぞれの重み W^{Q410} 、 W^{K420} 、および W^{V430} に従う。重み W^{Q410} 、 W^{K420} 、および W^{V430} は、逆伝播を使用して訓練中に変更される。

【 0 0 3 5 】

【 数 1 0 】

$$Q = qW^Q \in \mathbb{R}^{d_q}$$

$$K = kW^K \in \mathbb{R}^{d_k} \quad \text{式 1 0}$$

$$V = vW^V \in \mathbb{R}^{d_v}$$

【 0 0 3 6 】

得られたQベクトル、Kベクトル、およびVベクトルは、QとKのドット積を生成するアテンション伝達関数440を通過し、次に、式11に従ってVに適用される。

【0037】

【数11】

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \in \mathbb{R}^{d_v} \quad \text{式11}$$

【0038】

次に、追加および正規化モジュール450を使用して、クエリqをアテンション伝達関数からの出力と組み合わせて、アテンションネットワーク400による学習速度を改善するレジデュアルコネクションを提供する。加算および正規化モジュール450は、式12を実装し、ここで、 μ および σ は、それぞれ、入力ベクトルの平均および標準偏差であり、 g_i は、層の正規化をスケールリングするための利得パラメータである。加算および正規化モジュール450からの出力は、アテンションネットワーク400の出力である。

【0039】

【数12】

$$\text{LayerNorm}(\text{Attention}(Q,K,V)+q) \\ \text{LayerNorm}(a_i) = g_i \frac{a_i - \mu}{\sigma} \quad \text{式12}$$

【0040】

アテンションネットワーク400は、しばしば2つの変形形態で使用される。第1の変形形態は、マルチヘッドアテンションネットワークであり、ここで、アテンションネットワーク400と一致する複数のアテンションネットワークが並列に実装され、マルチヘッドアテンションネットワークの「ヘッド」の各々は、それ自身の重み W^Q 410、 W^K 420、および W^V 430を有し、これらは、異なる値に初期化され、異なる符号化を学習するように訓練される。次に、ヘッドの各々からの出力は、一緒に連結され、マルチヘッドアテンションネットワークの出力を形成する。第2の変形形態は、マルチヘッドアテンションネットワークであるセルフアテンションネットワークであり、ここで、q、k、およびvの入力はアテンションネットワークの各ヘッドに対して同じである。

【0041】

セルフアテンションベースの層は、さらに、2017年6月12日に提出された、Vaswaniらによる“Attention is All You Need,” arXiv preprint arXiv:1706.03762に説明されており、その全体が参照により本明細書に組み込まれる。

【0042】

図5は、いくつかの実施形態によるアテンションベースのトランスフォーマネットワークのための層500の簡略図である。図5に示されるように、層500は、符号化器510および復号器520を含む。

【0043】

符号化器510は、層入力（例えば、符号化スタック内の第1の層のための入力ネットワークから、または符号化スタックの他のすべての層のための次の最下位の層の層出力から）を受信し、マルチヘッドアテンションネットワーク511の3つの全ての入力にそれを提供する。したがって、マルチヘッドアテンションネットワーク511は、セルフアテンションネットワークとして構成されている。マルチヘッドアテンションネットワーク511の各ヘッドは、アテンションネットワーク400と一致する。いくつかの例において、マルチヘッドアテンションネットワーク511は、3つのヘッドを含むが、2つ、または3つより多い数などの他の数のヘッドが可能である。いくつかの例において、各アテン

10

20

30

40

50

ションネットワークは、寸法が200で、隠れたサイズが128である。マルチヘッドアテンションネットワーク511の出力は、フィードフォワードネットワーク512に提供され、フィードフォワードネットワーク512の入力および出力の両方が、符号化器510のための層出力を生成する加算および正規化モジュール513に提供される。いくつかの例において、フィードフォワードネットワーク512は、正規化線形ユニット（ReLU）活性化を備えた2層パーセプトロンネットワークであり、これは、式13を実装し、ここで、 γ はフィードフォワードネットワーク512への入力であり、 M_i および b_i は、パーセプトロンネットワーク内の各層の重みおよびバイアスである。いくつかの例において、加算および正規化モジュール513は、加算および正規化モジュール450と実質的に類似している。

【0044】

【数13】

$$FF(\gamma) = \max(0, \gamma M_1 + b_1) M_2 + b_2 \quad \text{式13}$$

【0045】

復号器520は、層入力を（例えば、復号スタック内の第1の層のための入力ネットワークから、または復号スタックの他のすべての層のための次の最下位の層の層出力から）受信し、マルチヘッドアテンションネットワーク521の3つの全ての入力にそれを提供する。したがって、マルチヘッドアテンションネットワーク521は、セルフアテンションネットワークとして構成されている。マルチヘッドアテンションネットワーク521の各ヘッドは、アテンションネットワーク400と一致する。いくつかの例において、マルチヘッドアテンションネットワーク521は、3つのヘッドを含むが、2つ、または3つより多い数などの他の数のヘッドが可能である。マルチヘッドアテンションネットワーク521の出力は、別のマルチヘッドアテンションネットワーク522への入力 q として提供され、マルチヘッドアテンションネットワーク522の入力 k および v は、符号化器からの出力で提供される。マルチヘッドアテンションネットワーク521の各ヘッドは、アテンションネットワーク400と一致する。いくつかの例において、マルチヘッドアテンションネットワーク522は、3つのヘッドを含むが、2つ、または3つより多い数などの他の数のヘッドが可能である。いくつかの例において、各アテンションネットワークは、寸法が200で、隠れたサイズが128である。マルチヘッドアテンションネットワーク522の出力は、フィードフォワードネットワーク523に提供され、フィードフォワードネットワーク523の入力および出力の両方が、符号化器510のための層出力を生成する加算および正規化モジュール524に提供される。いくつかの例において、フィードフォワードネットワーク523および追加および正規化モジュール524は、それぞれフィードフォワードネットワーク512および追加および正規化モジュール513と実質的に類似している。

【0046】

図3に戻って参照すると、セルフアテンション符号化器360、365、370、および／または375の各々は、符号化器510と一致し、式14に従ってセルフアテンションベースの符号化を生成し、ここで、 $SAEncode(X)$ は、符号化器510によって行われる符号化に対応し、 X は、マルチヘッドアテンションネットワーク511によって受信される入力に対応する。

【0047】

【数14】

$$\begin{aligned} SAEncode(C_{com}) &= C_{self1} \in \mathbb{R}^{1 \times d} \\ SAEncode(C_{self1}) &= C_{self2} \in \mathbb{R}^{1 \times d} \\ SAEncode(Q_{com}) &= Q_{self1} \in \mathbb{R}^{m \times d} \\ SAEncode(Q_{self1}) &= Q_{self2} \in \mathbb{R}^{m \times d} \end{aligned} \quad \text{式14}$$

10

20

30

40

50

【0048】

最終符号化コンテキストのための最後の符号化が、次に、式15に従ってBiLSTM380を使用して、コンテキスト情報を経時的に統合することによって、セルフアテンション符号化器370からの出力 C_{self2} から生成される。同様に、質問のための最後の符号化が、次に、式15に従ってBiLSTM385を使用して、質問情報を経時的に統合することによって、セルフアテンション符号化器375からの出力 Q_{self2} 出力から生成される。いくつかの例において、BiLSTM380および385は、入力に対して0.2のドロップアウトを使用する。

【0049】

【数15】

$$\begin{aligned} \text{BiLSTM}_{\text{finC}}(C_{\text{self2}}) &= C_{\text{fin}} \in \mathbb{R}^{1 \times d} \\ \text{BiLSTM}_{\text{finQ}}(Q_{\text{self2}}) &= Q_{\text{fin}} \in \mathbb{R}^{m \times d} \end{aligned} \quad \text{式15}$$

10

【0050】

図6は、いくつかの実施形態による復号器の簡略図である。復号器600は、反復アプローチを使用して、一度に単一のワードまたはトークンを生成する。図6は、答えの生成においてどの反復が参照されているかを示す下付き文字を使用する時間的／位置的關係を表す。（例えば、 X_t は t 回目の反復からの X の値を参照し、 X_{t-1} は $(t-1)$ 回目の反復、つまり t 回目の反復の前の反復からの X の値を参照する）。式1に関して上述したように、復号器600が訓練中に使用されるときに、正解は、コンテキストおよび質問に対して使用されるのと同じ符号化を使用して、最初に符号化される。

20

【0051】

答え表現埋め込みおよび符号化器610は、部分的に形成された答えを、式16に従ってコンテキストおよび質問に対する埋め込みと同様の d 次元空間に投影するために使用される。

【0052】

【数16】

$$AW_2 + b_2 = A_{\text{proj}} \in \mathbb{R}^{n \times d} \quad \text{式16}$$

30

【0053】

答え内のワードまたはトークンは、再帰および畳み込みの両方を欠くため、位置符号化PEは、式17に従って A_{proj} に加算される。

【0054】

【数17】

$$\text{PE}[t, k] = \begin{cases} \sin(t/10000^{k/2d}) & k \text{は偶数} \\ \cos(t/10000^{(k-1)/2d}) & k \text{は奇数} \end{cases} \quad \text{式17}$$

$$A_{\text{proj}} + \text{PE} = A_{\text{ppr}} \in \mathbb{R}^{n \times d}$$

40

【0055】

マルチヘッド復号器アテンション埋め込みおよび符号化器610の出力は、次に、セルフアテンション復号器620に渡され、その出力は、次に、式18に従ってセルフアテンション復号器630に渡されてASELFを生成し、ここで、 $\text{SADecode}(A, B)$ は、復号器520によって行われる復号に対応し、 A は、マルチヘッドアテンションネットワーク521によって受信される層入力に対応し、 B は、マルチヘッドアテンションネットワーク522によって受信される符号化入力に対応する。セルフアテンションが使用されて、復号器600が、答えに追加される次のワードまたはトークンに備えるために

50

、コンテキストに対する前の出力およびアテンションを認識するようにする。しかしながら、訓練中、および復号器が段階的に動作するため、埋め込みおよび符号化された正解の適切なエントリは、式 18 を適用するときに復号器 600 が将来の時間ステップを見ることを防止するために、大きな負の数に設定される。

【0056】

【数 18】

$$\begin{aligned} \text{SADecode}(A_{\text{ppr}}, C_{\text{fin}}) &= A_1 \in \mathbb{R}^{l \times d} \\ \text{SADecode}(A_1, C_{\text{fin}}) &= A_{\text{SELF}} \in \mathbb{R}^{l \times d} \end{aligned} \quad \text{式 18}$$

10

【0057】

中間復号器状態 LSTM 640 は、式 19 に従って、前の答えワードまたはトークン $A_{\text{SELF}_{t-1}}$ 、前の反復からのコンテキスト復号器状態

【外 12】

$\tilde{c}_{t-1}$

、および前の反復からの中間状態 h_{t-1} を使用して、現在の反復の中間状態 h_t を生成する。いくつかの例において、LSTM 640 は、入力に対して 0.2 のドロップアウトを使用する。

【0058】

【数 19】

$$\text{LSTM}([A_{\text{SELF}_{t-1}}, \tilde{c}_{t-1}], h_{t-1}) = h_t \in \mathbb{R}^d \quad \text{式 19}$$

20

【0059】

コンテキストおよび質問アテンション LSTM 640 からの中間状態 h_t は、コンテキストおよび質問アテンションネットワーク 650 に提供される、すなわち、式 20 に従って、最初に、コンテキスト C_{fin} の最終符号化および質問 Q_{fin} の最終符号化に対して、それぞれ、アテンション重み α_{c_t} および α_{q_t} を生成し、ここで、 W_2 および W_3 は、訓練可能な重み行列である。いくつかの例において、コンテキストおよび質問アテンションネットワーク 650 は、式 20 内に訓練可能なバイアス b_2 および b_3 を任意選択で含めてもよい。アテンション重みにより、復号器 600 が各反復 t に関連する符号化情報に焦点を合わせることを可能にする。

30

【0060】

【数 20】

$$\begin{aligned} \text{softmax}(C_{\text{fin}}(W_2 h_t)) &= \alpha_{c_t} \in \mathbb{R}^1 \\ \text{softmax}(Q_{\text{fin}}(W_3 h_t)) &= \alpha_{q_t} \in \mathbb{R}^m \end{aligned} \quad \text{式 20}$$

【0061】

再帰コンテキスト状態コンテキストおよび質問アテンションネットワーク 650 は、次に、アテンション重み α_{c_t} および α_{q_t} によってそれぞれ重み付けされたコンテキスト C_{fin} の最終符号化および質問 Q_{fin} の最終符号化を、式 21 に従って隠れ状態 h_t と組み合わせ、コンテキスト表現

40

【外 13】

$\tilde{c}_t$

および質問表現

【外 14】

50

$\tilde{Q}_t$

を生成し、ここで、 W_4 および W_5 は、訓練可能な重みであり、 $\tanh$ は、双曲線正接伝達関数である。いくつかの例において、コンテキストおよび質問アテンションネットワーク650は、式21内に訓練可能なバイアス b_4 および b_5 を任意選択で含めてもよい。いくつかの例において、他の非線形伝達関数が式21と共に使用され得る。

【0062】

【数21】

$$\begin{aligned}\tanh(W_4 C_{fin}^T \alpha_{C_t}; h_t) &= \tilde{C}_t \in \mathbb{R}^d \\ \tanh(W_5 Q_{fin}^T \alpha_{Q_t}; h_t) &= \tilde{Q}_t \in \mathbb{R}^d\end{aligned}\quad \text{式21}$$

10

【0063】

マルチポインタ生成器 ポインタ生成器660は、コンテキスト内のワードまたはトークンおよび質問内のワードまたはトークンに対する分布だけではなく、生成語彙からのワードまたはトークン v に対する分布も生成するために使用される。分布は、コンテキスト、質問、および語彙からのワードまたはトークンの各々が、現在の反復 t において復号器600によって選択される次のワードまたはトークンである可能性を割り当てる。分布は、式22に従って生成され、ここで、 W^V は、訓練可能な重み行列である。いくつかの例において、ポインタ生成器660は、式22内の生成語彙に対する分布 p_v を生成するために、式22内に訓練可能なバイアス b_v を任意選択で含めてもよい。

20

【0064】

【数22】

$$\begin{aligned}\sum_{i: c_i = W_t} (\alpha_{C_t})_i &= p_c(W_t) \in \mathbb{R}^n \\ \sum_{i: q_i = W_t} (\alpha_{Q_t})_i &= p_q(W_t) \in \mathbb{R}^m \\ \text{softmax}(W_v \tilde{C}_t) &= p_v(W_t) \in \mathbb{R}^v\end{aligned}\quad \text{式22}$$

30

【0065】

次いで、ポインタ生成器660は、各部分 p_c 、 p_q 、 p_v が

【外15】

$\mathbb{R}^{l+m \times v}$

にあるように、各分布における欠落エントリを0に設定することによって、コンテキスト、質問、および生成語彙におけるトークンの結合に対して分布 p_c 、 p_q 、および p_v を拡張する。

40

【0066】

スイッチ670が、次に、式23に従って、現在の反復に対する次のワードまたはトークンを決定する際のコンテキストおよび質問の重要性を決定するために使用し、ここで、 σ は、 $\log \text{sig}$ 、 tansig などのシグモイド伝達関数である。

【0067】

【数23】

$$\begin{aligned}\sigma(W_{pv}[\tilde{C}_t; h_t; \text{ASELF}_{t-1}]) &= \gamma \in [0,1] \\ \sigma(W_{cq}[\tilde{Q}_t; h_t; \text{ASELF}_{t-1}]) &= \lambda \in [0,1]\end{aligned}\quad \text{式23}$$

50

【 0 0 6 8 】

スイッチ 6 7 0 は、式 2 4 に従って合成分布を生成し、現在の反復に対する次のワードまたはトークンが、最大の可能性を有する合成分布からのワードまたはトークンに基づいて選択される。選択されたワードまたはトークンが、次に、答えに追加され、埋め込みおよび符号化器 6 1 0 にフィードバックされ、ここで、復号器 6 0 0 の別の反復が開始される。復号器 6 0 0 は、最大の数、ワードまたはトークンが選択されるまで、および／または答えの終わりを示すセンチネルワードまたはトークンが選択されるまで、反復し続ける。

【 0 0 6 9 】

【 数 2 4 】

$$\gamma p_v(w_t) + (1 - \gamma)[\lambda_{p_r}(w_t) + (1 - \lambda)p_q(w_t)] = p(w_t) \in \mathbb{R}^{1+m+v} \quad \text{式 2 4}$$

10

【 0 0 7 0 】

いくつかの実施形態によれば、訓練中に、符号化器 3 0 0 および復号器 6 0 0 は、式 2 5 に従って、各時間ステップにわたってトークンレベルの負の対数尤度損失関数 $\mathcal{L}$ を使用して、逆伝播によって訓練され、ここで、 a_t は、答え内の t 番目のワードまたはトークンに対応する。

【 0 0 7 1 】

【 数 2 5 】

$$\mathcal{L} = - \sum_t^T \log p(a_t) \quad \text{式 2 5}$$

20

【 0 0 7 2 】

図 3 ～ 図 6 のネットワークは、複数のタスク（例えば、質問応答、機械翻訳、文書要約、データベースクエリ生成、感情分析、自然言語推論、意味役割ラベル付け、関係抽出、目標指向対話、代名詞解決）に使用され、各タスクタイプを横断する様々な層およびネットワークに対するそのパラメータを共有するため、注意深く訓練されない場合、破滅的な忘却の影響を受けやすいかもしれない。これに対処するために、いくつかの実施形態では、図 3 ～ 図 6 のネットワークは、修正ジョイント戦略に従って訓練されてもよく、ここで、図 3 ～ 図 6 のネットワークは、訓練サンプルが提示される順序を用いて訓練され、図 3 ～ 図 6 のネットワークを、タスクタイプの各々のバランスのとれた混合に対して同時に訓練する。すなわち、図 3 ～ 6 のネットワークに訓練サンプルが提示される順序は、異なるタスクタイプから連続訓練サンプルまたは連続する小グループ（例えば、2 ～ 1 0 程度）の訓練サンプルを選択する。いくつかの例において、ジョイント戦略は、訓練の各反復に伴うタスクタイプの異なる 1 つから訓練サンプル（コンテキスト c 、質問 q 、および正解 a ）を選択することを含む。ジョイント戦略の目標は、別のタスクタイプに対してあるタスクタイプに重点を置きすぎることなく、タスクタイプの各々に対して同時に訓練することである。

30

【 0 0 7 3 】

図 7 は、いくつかの実施形態によるマルチタスク学習方法の簡略図である。方法 7 0 0 の 7 1 0 ～ 7 8 0 のうちの 1 つ以上は、少なくとも部分的に、1 つ以上のプロセッサによって動作されるときに、1 つ以上のプロセッサがプロセス 7 1 0 ～ 7 8 0 のうちの 1 つ以上を行うようにし得る、非一時的な有形の機械可読媒体に記憶された実行可能コードの形態で実装されてもよい。いくつかの実施形態において、方法 7 0 0 は、図 3 ～ 図 6 のネットワークを訓練するためのハイブリッド訓練戦略として使用され得るが、方法 7 0 0 は、図 3 ～ 図 6 のネットワーク以外の他のマルチタスクシステムを訓練するためにも使用され得る。いくつかの例において、方法 7 0 0 は、カリキュラム (curriculum) および／または反カリキュラム (anti-curriculum) 戦略と一致してもよい。いくつかの実施形態において、方法 7 0 0 によって訓練されるタスクタイプは、質問応

40

50

答、機械翻訳、文書要約、データベースクエリ生成、感情分析、自然言語推論、意味役割ラベル付け、関係抽出、目標指向対話、代名詞解決などの様々な自然言語処理タスクタイプのいずれかを含んでもよい。

【0074】

プロセス710では、訓練サンプルは、第1の訓練戦略に従って選択される。いくつかの実施形態では、第1の訓練戦略は、訓練サンプルが、ネットワークが訓練されているタスクタイプのサブセットから選択されるジョイント訓練戦略である。いくつかの例において、タスクタイプのサブセットは、タスクタイプの完全セットから選択されてもよい。いくつかの例において、タスクタイプの完全セットは、質問応答、機械翻訳、文書要約、データベース質問生成、感情分析、自然言語推論、意味的役割ラベリング、関係抽出、目標指向対話、代名詞解決などから選択された1つ以上のタスクタイプを含んでもよい。

10

【0075】

いくつかの実施形態において、タスクタイプのサブセットは、カリキュラム戦略に従って選択されてもよく、ここで、訓練サンプルは、訓練が比較的少数の訓練反復で収束するタスクタイプから選択される。いくつかの例において、カリキュラム訓練のためのタスクタイプのサブセットは、データベースクエリ生成、感情分析、意味役割ラベル付け、関係抽出、目標指向対話、代名詞解決などから選択された1つ以上のタスクタイプを含んでもよい。

【0076】

いくつかの実施形態において、タスクタイプのサブセットは、反カリキュラム戦略に従って選択されてもよく、ここで、訓練サンプルは、より学習困難である、より長い答えシーケンスを有する、および／または異なるタイプの復号を含むと特徴付けられるタスクタイプから選択される。いくつかの例において、反カリキュラム訓練のためのタスクタイプのサブセットは、質問応答、要約、機械翻訳、および／または自然言語推論から選択された1つ以上のタスクタイプを含んでもよい。

20

【0077】

いくつかの例において、選択された訓練サンプルは、コンテキストおよび質問にそれぞれ対応する自然言語コンテキストおよび自然言語質問、ならびに正解自然言語答えを含む。

【0078】

プロセス720では、選択された訓練サンプルがシステムに提示される。いくつかの例において、システムは、符号化器300および復号器600を含む。訓練サンプルがシステムに適用されるときに、現在訓練されているパラメータ（例えば、重みおよびバイアス）に従って、訓練サンプルはシステムの種々の層を通してフィードフォワードされ、答えが生成される。いくつかの例において、答えは自然言語句である。

30

【0079】

プロセス730では、システムは誤差に基づいて調整される。プロセス720中にシステムによって生成された答えが、選択された訓練サンプルについての正解と比較され、選択された訓練サンプルについての誤差が決定される。誤差は、次に、層の様々なパラメータ（例えば、重みおよびバイアス）を更新するために、逆伝播を使用してシステムにフィードバックされてもよい。いくつかの例において、逆伝播は、確率勾配降下（SGD）訓練アルゴリズム、適応モーメント推定（ADAM）訓練アルゴリズムなどを使用して行われてもよい。いくつかの例において、ADAMのメタパラメータは

40

【外16】

$$(\beta_1, \beta_2, \epsilon) = (0.9, 0.98, 10^{-9})$$

に設定される。いくつかの例において、学習速度は、 $k^{-1/2}$ で減衰される前の最初の800回の反復で、0から 2.5×10^{-3} まで直線的に増加され、ここで、 k は反復カウントである。いくつかの例において、逆伝播に使用される勾配は、1.0にクリップされてもよい。

50

【0080】

プロセス740では、タスクタイプのサブセットに基づく訓練から、タスクタイプの完全セットに基づく訓練に切り替えるかどうか決定される。いくつかの例において、所定の数の訓練サンプルが提示された後に、タスクタイプの完全なセットに切り替える決定が発生する。いくつかの例において、所定の数の訓練サンプルは300, 000であってもよいが、他の所定の数の訓練サンプルが可能である。いくつかの例において、タスクタイプのサブセット内のタスクタイプの各々に対する訓練サンプルの各々が、所定の回数だけ選択された後に、タスクタイプの完全セットに切り替える決定が発生する。いくつかの例において、所定の回数は5回であってもよいが、3回、4回、および／または6回以上のような任意の他の回数も使用されてもよい。いくつかの例において、1つ以上の他のファクタが使用されて、タスクタイプの完全セットを使用して訓練にいつ切り替えるかに関する決定をなしてもよい。いくつかの例において、1つまたは他のファクタは、訓練サンプルを通した各パスを伴うタスクタイプのサブセット内のタスクタイプの各々に対するパフォーマンスメトリックの変化を監視することと、各パス後のパフォーマンスメトリックの各々における改善が閾値量よりも小さく改善されたときに切り替えをなすことと、を含んでもよい。タスクタイプの完全セットに対する訓練に切り替えないと決定されたときに、方法700は、プロセス710に戻り、ここで、訓練サンプルは、タスクタイプのサブセットから選択され続ける。タスクタイプの完全セットから訓練に切り替えることが決定されるときに、訓練サンプルの選択は、プロセス750で開始するタスクタイプの完全セットを使用して発生する。

10

20

【0081】

プロセス750では、訓練サンプルは、タスクタイプの完全なセットから選択される。いくつかの例において、タスクタイプの完全セットは、質問応答、機械翻訳、文書要約、データベースクエリ生成、感情分析、自然言語推論、意味役割ラベル付け、関係抽出、目標指向対話、代名詞解決などから選択される1つ以上のタスクを含む。

【0082】

プロセス760では、選択された訓練サンプルは、プロセス720と実質的に同じプロセスを使用してシステムに提示される。

【0083】

プロセス770では、システムは、プロセス730と実質的に同じプロセスを使用して誤差に基づいて調整される。

30

【0084】

プロセス780では、訓練が完了しているかどうか決定される。いくつかの例において、所定の数の訓練サンプルが提示された後に訓練が完了する。いくつかの例において、訓練は、完全セットのタスクタイプ内のタスクタイプの各々に対する訓練サンプルが所定の回数システムに提示された後に完了する。いくつかの例において、所定の回数は8回であってもよいが、2〜7回および／または9回以上のような任意の他の回数も使用されてもよい。いくつかの例において、1つ以上の他のファクタが、訓練が完了したときに関する決定をなすために使用されてもよい。いくつかの例において、1つまたは他のファクタは、訓練サンプルを通した各パスを伴うタスクタイプの各々に対するパフォーマンスメトリックの変化を監視することと、各パス後のパフォーマンスメトリックの各々における改善が閾値量よりも小さく改善されたときになにもないことと、を含んでもよい。訓練が完了していないと決定されたときに、方法700は、プロセス740に戻り、ここで、訓練サンプルは、タスクタイプの完全なセットから選択され続ける。訓練が完了したと決定されたときに、方法700は終了し、訓練されたシステムは、訓練されたタスクのいずれに対しても使用され得る。

40

【0085】

訓練が完了した後、訓練されたシステムは、プロセス720および／または760と実質的に同様のプロセスを使用して、タスクタイプのいずれに対しても使用されてもよく、ここで、コンテキストcおよび質問qは、システムに提示され、方法700に従って訓練

50

されたパラメータ（例えば、重みおよびバイアス）に従って、システムの様々な層を通じてフィードフォワードされてもよい。生成された答えは、提示されたコンテキスト c および質問 q への答えに対応する。

【0086】

図8は、いくつかの実施形態による訓練セットの要約の簡略図である。より具体的には、図8は、各タスクタイプ、使用したデータセット、データセット中の訓練サンプル数、データセット中の開発／検証サンプル中のサンプル数、データセット中の試験サンプル数、および使用したパフォーマンス測定基準を示す。

【0087】

質問応答 (Question Answering) タスクタイプに対する訓練サンプルは、Wikipedia 記事の段落サンプルに関連する質問に基づく訓練サンプルを含む、Stanford Question Answering Dataset (SQuAD) に基づく。質問応答タスクタイプに使用されるパフォーマンス測定基準は、正規化 F1 (nF1) スコアである。

10

【0088】

機械翻訳 (Machine Translation) タスクタイプに対する訓練サンプルは、TEDトークから転記された文対を含む、英語からドイツ語への音声言語翻訳についての国際ワークショップ (IWSLT EN→DE) 訓練セットに基づく。機械翻訳タスクタイプに使用されるパフォーマンス測定基準は、BLEUスコアである。

【0089】

20

要約 (Summarization) タスクタイプに対する訓練サンプルは、CNN日報 (CNN/DM) 訓練セットから選択される。要約タスクタイプに使用されるパフォーマンス測定基準は、ROUGEスコアである。

【0090】

自然言語推論 (Natural Language Inference) タスクタイプに対する訓練サンプルは、Multi-genre Natural Language Inference Corpus (MNLI) 訓練セットから選択される。自然言語推論タスクタイプに使用されるパフォーマンス測定基準は、正規化 F1 (nF1) スコアである。

【0091】

30

感情分析 (Sentiment Analysis) タスクタイプに対する訓練サンプルは、Stanford Sentiment Treebank (SST) に基づく。SSTは、映画レビューと彼らの感情に基づいた訓練サンプルを含む。感情分類タスクタイプに使用されるパフォーマンス測定基準は、正規化 F1 (nF1) スコアである。

【0092】

意味役割ラベル付け (Semantic Role Labeling) タスクタイプに対する訓練サンプルは、Question-Answer Semantic Role Labeling (QA-SRL) 訓練セットから選択される。意味役割ラベル付けタスクタイプに使用されるパフォーマンス測定基準は、正規化 F1 (nF1) スコアである。

【0093】

40

関係抽出 (Relationship Extraction) タスクタイプに対する訓練サンプルは、Question-Answer Zero-shot Relation Extraction (QA-ZRE) 訓練セットから選択される。関係抽出タスクタイプに使用されるパフォーマンス測定基準は、F1スコアです。

【0094】

目標指向対話 (Goal-Oriented Dialog) タスクタイプに対する訓練サンプルは、Wizard of Oz (WOZ) 訓練セットから選択される。目標指向対話タスクタイプに使用されるパフォーマンス測定基準は、完全一致 (EM) スコアである。

【0095】

50

データベースクエリ生成 (Database Query Generation) タスクタイプに対する訓練サンプルは、WikiSQL 訓練セットから選択される。データベースクエリ生成タスクタイプに使用されるパフォーマンス測定基準は、完全一致 (EM) スコアである。

【0096】

代名詞解決 (Pronoun Resolution) タスクタイプに対する訓練サンプルは、Modified Winograd Schema Challenge (MWSC) 訓練セットから選択される。代名詞解決タスクタイプに使用されるパフォーマンス測定基準は、正規化 F1 (nF1) スコアである。

【0097】

図9は、いくつかの実施形態によるシングルタスクおよびマルチタスク学習からの結果の簡略図である。より具体的には、図8は、質問応答、機械翻訳、文書要約、データベースクエリ生成、感情分析、自然言語推論、意味役割ラベル付け、関係抽出、目標指向対話、代名詞解決を含む様々なタスクタイプに対する、符号化器300および復号器600などの訓練システムの結果を示す。

【0098】

訓練の結果は、複合Decathlon Score (decaScore) に従ってさらに評価される。decaScoreは、タスク固有のパフォーマンス測定基準の加算的な組み合わせを含む。個々のパフォーマンス測定基準は各々0~100の間にあるため、10タイプのタスクにわたる複合パフォーマンススコアのdecaScoreは0~1000の間にある。

【0099】

図9のシングルタスク訓練列は、Seeらの“Get to the Point: Summarization with Pointer-generator Networks,” arXiv preprint arXiv:1704.04368, 2017のポインタ生成器sequence-to-sequence model (S2S)、Vaswaniらの“Attention is All You Need,” arXiv preprint arXiv:1706.03762, 2017のS2Sの下位層がセルフアテンション符号化器および符号器層に置換されたS2Sのバージョン (w/SAtt)、コンテキストおよび質問の両方の表現を別々にビルドし、その2つがどのように相互作用するかを明示的にモデル化するコアテンションメカニズムで強化されたS2Sのバージョン (+CAtt)、および質問ポインタが追加された+CAttのバリエーション (+QPtt) を使用して分離して訓練されたときに、それぞれのパフォーマンス測定基準に対するタスクタイプの各々についてのベースライン結果を示す。

【0100】

図9のマルチタスク訓練カラムは、さらに、訓練を通してタスクタイプの各々から訓練サンプルが選択されるジョイント訓練戦略を使用したモデルの各々の訓練の結果を示す。符号化器300および復号器600の訓練のさらなる結果 (+ACurr) は、方法700の訓練の反カリキュラムバージョンを使用し、ここで、タスクタイプの初期サブセットは、質問応答 (SQuAD) 訓練セットからの訓練サンプルのみを含む。図8が示すように、符号化器300および復号器600を使用する方法700の反カリキュラム訓練は、これらのモデルの各々について、より良い全体的な複合結果 (デカスコアによって測定されるように) をもたらす。

【0101】

図10は、いくつかの実施形態に従った、異なる訓練戦略に関する結果の簡略図である。より具体的には、図10は、完全ジョイント (Fully Joint) 戦略の使用、タスクタイプのサブセットがデータベースクエリ生成、感情分析、意味役割ラベル付け、関係抽出、目標指向対話、および代名詞解決に限定されるカリキュラムベースのアプローチ (Curriculum) を採用する方法700の使用、タスクタイプのサブセットが質問応答に限定されたもの (SQuAD)、質問応答、機械翻訳、および要約に限定され

10

20

30

40

50

たもの (+IWSLT+CNN/DM)、質問応答、機械翻訳、要約、および自然言語推論に限定されたもの (+MNLI) である反カリキュラムベースのアプローチ (anti-curriculum) を採用する方法 700 の使用での訓練符号化器 300 と復号器 600 との差異を示す。図に示すように、カリキュラムベースの方法 700 のバリエーションは、完全ジョイント戦略ほどには機能しなかった。しかし、初期サブセットにおける質問応答訓練サンプルのみを使用した反カリキュラムアプローチ (SQUD) を有する方法 700 の使用は、完全ジョイント戦略に対して全体的な改善を示し、タスクタイプの各々に対する改善またはわずかな劣化を示した。タスクタイプのサブセットが、他の作業タイプの訓練サンプルを含むように拡張されたときに、パフォーマンスの全体的な改善は発生しなかった。

【0102】

図 11 は、いくつかの実施形態による答えのためのワード選択のためのソースの簡略図である。より具体的には、図 11 は、タスクタイプの各々について、コンテキスト、質問、または生成語彙のうち、どれがスイッチ 670 によって選択され、式 23 および式 24 で使用される γ 値および入値に基づいて、答えのために選択されるワードまたはトークンのソースとなる可能性がより高いかを示す。図示されるように、感情分析 (SST)、自然言語推論 (MNLI)、および代名詞解決 (MWSC) タスクタイプは、質問が答えの分類オプションを含む傾向があるため、質問からのワードまたはトークンを好む。機械翻訳 (IWSLT) および目標指向対話 (WOZ) タスクタイプは、コンテキストも質問も答えに望ましいワードを含まない傾向があるため、語彙からのワードまたはトークンを好む。

【0103】

図 12A は、いくつかの実施形態による、新しい機械翻訳タスクタイプのための異なる訓練アプローチからの結果の簡略図である。より具体的には、図 12A は、IWSLT 英語からチェコへの訓練サンプルのみに対してランダムスタートから訓練された符号化器 300 と復号器 600 (曲線 1210) と比較して、方法 700 に従って予備訓練された符号化器 300 と復号器 600 を使用する IWSLT 英語からチェコへの (IWSLT EN→CS) 訓練セットからの訓練サンプルを使用する新しい機械翻訳タスクタイプに対する訓練の結果を示す。

【0104】

図 12B は、いくつかの実施形態による新しい分類タスクのための異なる訓練アプローチからの結果の簡略図である。より具体的には、図 12B は、NER 訓練サンプルのみに対してランダムスタートから訓練された符号化器 300 と復号器 600 (曲線 1240) と比較して、方法 700 に従って予備訓練された符号化器 300 と復号器 600 を使用する固有表現抽出 (NER) タスクタイプ訓練サンプルに対する訓練の結果を示す。

【0105】

図示されるように、方法 700 (曲線 1210 および 1230) を使用する予備訓練は、シングルタスク訓練 (曲線 1220 および 1240) よりも良好なパフォーマンス結果を提示し、方法 700 のマルチタスク訓練を使用して、自然言語処理タスクの予備訓練を行い、追加的に、符号化器 300 および復号器 600 に基づく既に訓練されたネットワークを新たなタスクに対して適合させる利点を提示している。

【0106】

計算デバイス 200 のような計算デバイスのいくつかの例は、1つ以上のプロセッサ (例えば、プロセッサ 210) によって動作されるとき、1つ以上のプロセッサが方法 700 のプロセスを行う、および／または図 3～6 の構造をエミュレートし得る実行可能コードを含む、非一時的な有形の機械可読媒体を含んでもよい。方法 700 のプロセスおよび図 3～6 のエミュレーションを含み得る機械可読媒体のいくつかの一般的な形態は、例えば、フロッピーディスク、フレキシブルディスク、ハードディスク、磁気テープ、他の任意の磁気媒体、CD-ROM、他の任意の光学媒体、パンチカード、紙テープ、穴のパターンを有する他の任意の物理媒体、RAM、PROM、EPROM、FLASH-EPROM

OM、他の任意のメモリチップまたはカートリッジ、および／またはプロセッサまたはコンピュータが読むように適合される任意の他の媒体を含んでもよい。

【0107】

発明の態様、実施形態、実装、または用途を例示するこの説明および添付の図面は、限定的なものとして解釈されるべきではない。様々な機械的、組成的、構造的、電気的、および動作上の変更は、この説明および特許請求の範囲の精神および範囲から逸脱することなくなされてもよい。いくつかの例において、本開示の実施形態を不明瞭にしないために、周知の回路、構造、または技術が詳細に示されていないか、または説明されていない。2つ以上の図の同様の数字は、同じまたは類似の要素を表す。

【0108】

この説明では、本開示と矛盾しないいくつかの実施形態を説明する特定の詳細が記載される。実施形態の完全な理解を提供するために、多数の具体的な詳細が記載される。しかしながら、いくつかの実施形態が、これらの特定の詳細の一部または全部なしに実施され得ることは、当業者には明らかであろう。本明細書に開示される特定の実施形態は、例示的であるが、限定的ではないことを意味する。当業者は、本明細書に具体的に説明されていないが、本開示の範囲および精神内にある他の要素を認識してもよい。追加的に、不必要な繰り返しを避けるために、1つの実施形態に関連して示され説明された1つ以上の特徴は、具体的に述べられていない限り、または1つ以上の特徴が実施形態を非機能的にする場合を除いて、他の実施形態に組み込まれてもよい。

【0109】

例示的な実施形態が示され説明されたが、広範囲の修正、変更および置換が、前述の開示において考えられ、いくつかの例において、実施形態のいくつかの特徴は、他の特徴の対応する使用なしに使用され得る。当業者であれば、多くのバリエーション、代替案、および修正を認識するであろう。したがって、本発明の範囲は、以下の特許請求の範囲によってのみ限定されるべきであり、特許請求の範囲は、本明細書に開示された実施形態の範囲と一致する方式で広く解釈されることが適切である。

10

20

30

40

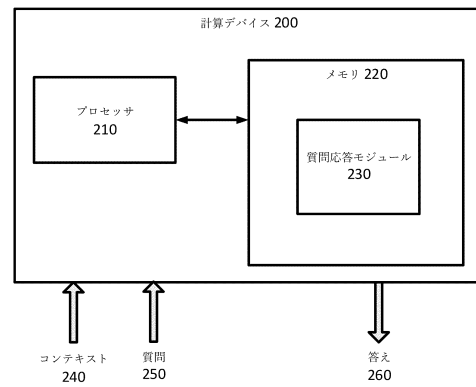
50

【図面】

【図 1】

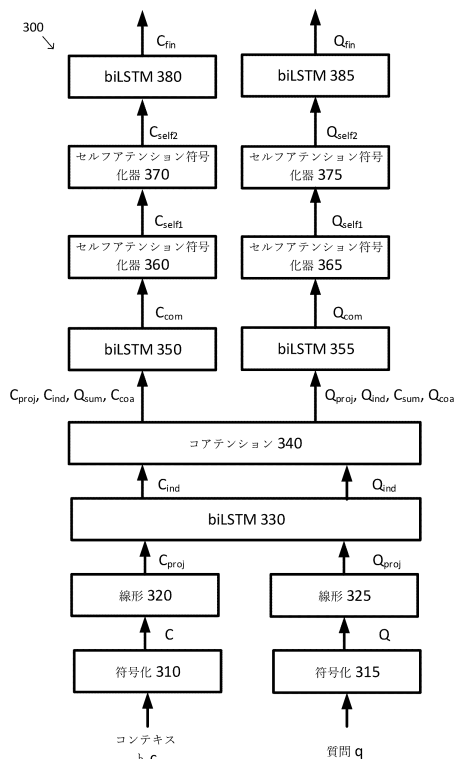
105	Question To whom did the Virgin Mary appear?	Context ...the Virgin Mary reputedly appeared to Saint Bernadette Soubirous...	Answer Saint Bernadette Soubirous
110	Question What is the translation from English to German?	Context Most of the planet is ocean water.	Answer Der Großteil der Erde ist Meerwasser
115	Question What is the summary?	Context Harry Potter star Daniel Radcliffe gains access to a reported £320 million fortune...	Answer Harry Potter star Daniel Radcliffe gets £320M fortune...
120	Question Hypothesis: Product and geography are what make cream skimming work. Entailment, neutral, or contradiction?	Context Premise: Conceptually cream skimming has two basic dimensions – product and geography. A stirring, funny and finally transporting re-imagining of Beauty and the Beast and 1930s horror film.	Answer Entailment
125	Question Is this sentence positive or negative?	Context A stirring, funny and finally transporting re-imagining of Beauty and the Beast and 1930s horror film.	Answer positive
130	Question What rows?	Context There are four boat clubs that row on the River Dee:...	Answer four boat clubs
135	Question What was the title that Timur held?	Context Both Timur and Demir are popular male names in Turkey today.	Answer unanswerable
140	Question What information have you gathered about the user?	Context Are there any Eritrean restaurants in town?	Answer food: Eritrean
145	Question What is the translation from English to SQL?	Context The table has column names... Tell me what the notes are for South Australia	Answer SELECT notes from table WHERE 'Current Slogan' = 'South Australia'
150	Question Who had given help? Susan or Joan?	Context Joan made sure to thank Susan for all the help she had given.	Answer Susan

【図 2】

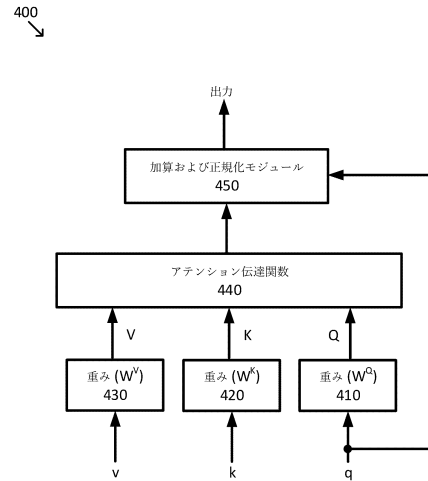


10

【図 3】



【図 4】



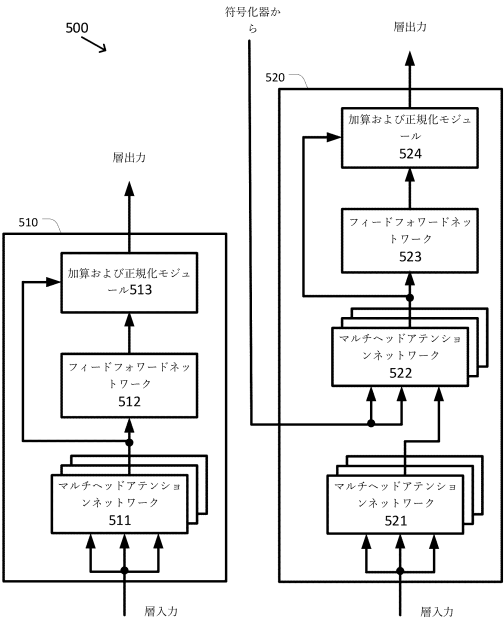
20

30

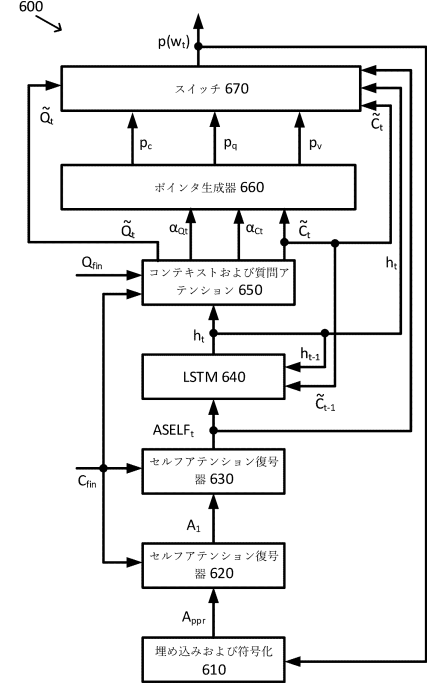
40

50

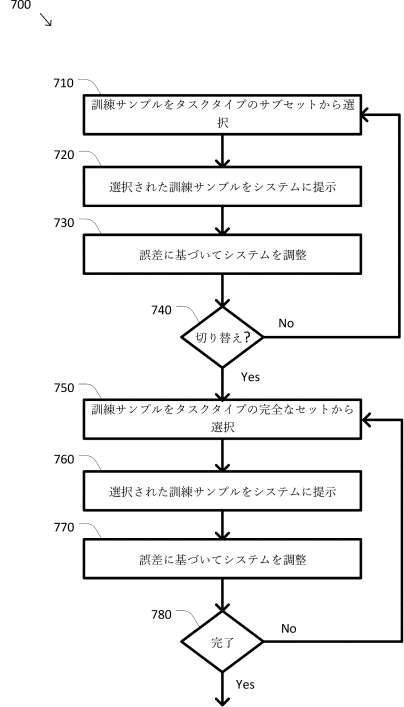
【図 5】



【図 6】



【図 7】



【図 8】

タスク	データセット	# Train	# Dev	# Test	Metric
Question Answering	SQuAD	87599	10570	9616	nF1
Machine Translation	IWSLT	196884	993	1305	BLEU
Summarization	CNN/DM	287227	13368	11490	ROUGE
Natural Language Inference	MNLI	392702	20000	20000	nF1
Sentiment Analysis	SST	6920	872	1821	nF1
Semantic Role Labeling	QA-SRL	6414	2183	2201	nF1
Zero-Shot Relation Extraction	QA-ZRE	840000	600	12000	F1
Goal-Oriented Dialogue	WOZ	2536	830	1646	EM
Database Query Generation	WikiSQL	56355	8421	15878	EM
Pronoun Resolution	MWSC	80	82	100	nF1

10

20

30

40

50

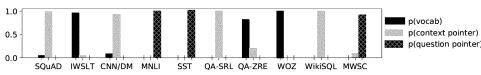
【図 9】

データセット	シングルタスク訓練				マルチタスク訓練				
	S2S	w/SAIt	+CAIt	+QPtr	S2S	w/SAIt	+CAIt	+QPtr	+ACurr
SQuAD	48.2	68.2	74.6	75.5	47.5	66.8	71.8	70.8	74.3
IWSLT	25.0	23.3	26.0	25.5	14.2	13.6	9.0	16.1	13.7
CNN/DM	19.0	20.0	25.1	24.0	25.7	14.0	15.7	23.9	24.6
MNLI	67.5	68.5	34.7	72.8	60.9	69.0	70.4	70.5	69.2
SST	86.4	86.8	86.2	88.1	85.9	84.7	86.5	86.2	86.4
QA-SRL	63.5	67.8	74.8	75.2	68.7	75.1	76.1	75.8	77.6
QA-ZRE	20.0	19.9	16.6	15.6	28.5	31.7	28.5	28.0	34.7
WOZ	85.3	86.0	86.5	84.4	84.0	82.8	75.1	80.6	84.1
WikiSQL	60.0	72.4	72.3	72.6	45.8	64.8	62.9	62.0	58.7
MWSC	43.9	46.3	40.4	52.4	52.4	43.9	37.8	48.8	48.4
decaScore	-	-	-	-	473.6	546.4	533.8	562.7	571.7

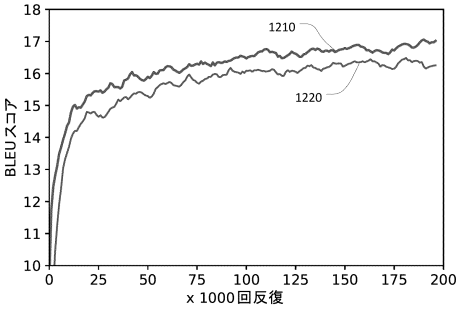
【図 10】

データセット	Fully Joint	Curriculum	Anti-Curriculum		
			SQuAD	+IWSLT+CNN/DM	+MNLI
SQuAD	70.8	43.4	74.3	74.5	74.6
IWSLT	16.1	4.3	13.7	18.7	19.0
CNN/DM	23.9	21.3	24.6	20.8	21.6
MNLI	70.5	58.9	69.2	69.6	72.7
SST	86.2	84.5	86.4	83.6	86.8
QA-SRL	75.8	70.6	77.6	77.5	75.1
QA-ZRE	28.0	24.6	34.7	30.1	37.7
WOZ	80.6	81.9	84.1	81.7	85.6
WikiSQL	62.0	68.6	58.7	54.8	42.6
MWSC	48.8	41.5	48.4	34.9	41.5
decaScore	562.7	499.6	571.7	546.2	557.2

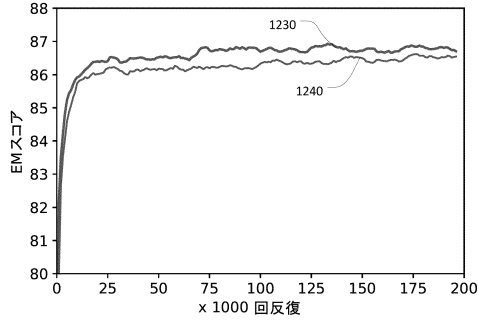
【図 11】



【図 12 A】



【図 12 B】



10

20

30

40

50

フロントページの続き

(33)優先権主張国・地域又は機関

米国(US)

ート, スイート 300

(72)発明者 ケスカー, ニティーシュ シリス

アメリカ合衆国 カリフォルニア州 94105, サンフランシスコ, ザ ランドマーク アット ワ
ン マーケット ストリート, スイート 300

(72)発明者 ション, ツァイミン

アメリカ合衆国 カリフォルニア州 94105, サンフランシスコ, ザ ランドマーク アット ワ
ン マーケット ストリート, スイート 300

(72)発明者 ソシエール, リチャード

アメリカ合衆国 カリフォルニア州 94105, サンフランシスコ, ザ ランドマーク アット ワ
ン マーケット ストリート, スイート 300

審査官 木村 大吾

(56)参考文献 西田 京介、他4名、情報検索とのマルチタスク学習による大規模機械読解、言語処理学会
第24回年次大会 発表論文集 [online], 日本、言語処理学会, 2018年03月05
日, pp.963~966

(58)調査した分野 (Int.Cl., DB名)

G06F 40/00-40/58

G06F 16/00-16/958

G06N 20/00