

【公報種別】特許法第17条の2の規定による補正の掲載

【部門区分】第6部門第3区分

【発行日】令和4年4月6日(2022.4.6)

【国際公開番号】WO2021/095211

【出願番号】特願2021-555728(P2021-555728)

【国際特許分類】

G 0 6 N 2 0 / 0 0 ( 2 0 1 9 . 0 1 )

【 F I 】

G 0 6 N 2 0 / 0 0

10

【手続補正書】

【提出日】令和4年1月18日(2022.1.18)

【手続補正1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項1】

20

第一のモーダルの情報に基づくベクトルと、前記第一のモーダルとは異なる第二のモーダルの情報に基づくベクトルとの相関に基づいて、前記第一のモーダルの情報に基づくベクトルを補正し、

前記第一のモーダルの情報に基づくベクトルと、前記第二のモーダルの情報に基づくベクトルとの相関に基づいて、前記第二のモーダルの情報に基づくベクトルを補正し、

補正後の前記第一のモーダルの情報に基づくベクトルから得た異なる種類の2つのベクトルの相関に基づいて、第一のベクトルを生成し、

補正後の前記第二のモーダルの情報に基づくベクトルから得た前記異なる種類の2つのベクトルの相関に基づいて、第二のベクトルを生成し、

所定のベクトルと、生成した前記第一のベクトルと、生成した前記第二のベクトルとを含む結合ベクトルから得た前記異なる種類の2つのベクトルの相関に基づいて、前記第一のベクトルと前記第二のベクトルとを集約した第三のベクトルを生成し、

30

生成した前記第三のベクトルを出力する、

処理をコンピュータが実行することを特徴とする出力方法。

【請求項2】

前記第一のモーダルの情報に基づくベクトルを補正する処理は、

前記第一のモーダルに関する第一のターゲットアテンション層を用いて、前記第一のモーダルの情報に基づくベクトルから得たベクトルと、前記第二のモーダルの情報に基づくベクトルから得たベクトルとの内積に基づいて、前記第一のモーダルの情報に基づくベクトルを補正し、

40

前記第二のモーダルの情報に基づくベクトルを補正する処理は、

前記第二のモーダルに関する第二のターゲットアテンション層を用いて、前記第一のモーダルの情報に基づくベクトルから得たベクトルと、前記第二のモーダルの情報に基づくベクトルから得たベクトルとの内積に基づいて、前記第二のモーダルの情報に基づくベクトルを補正し、

前記第一のベクトルを生成する処理は、

前記第一のモーダルに関する第一のセルフアテンション層を用いて、補正後の前記第一のモーダルの情報に基づくベクトルから得た前記異なる種類の2つのベクトルの内積に基づいて、補正後の前記第一のモーダルの情報に基づくベクトルをさらに補正し、前記第一のベクトルを生成し、

50

前記第二のベクトルを生成する処理は、

前記第二のモーダルに関する第二のセルフアテンション層を用いて、補正後の前記第二のモーダルの情報に基づくベクトルから得た前記異なる種類の2つのベクトルの内積に基づいて、補正後の前記第二のモーダルの情報に基づくベクトルをさらに補正し、前記第二のベクトルを生成し、

前記第三のベクトルを生成する処理は、

第三のセルフアテンション層を用いて、前記所定のベクトルと前記第一のベクトルと前記第二のベクトルとを結合した結合ベクトルから得た前記異なる種類の2つのベクトルの内積に基づいて、前記結合ベクトルを補正し、前記第三のベクトルを生成する、ことを特徴とする請求項1に記載の出力方法。

10

【請求項3】

生成した前記第三のベクトルに基づいて、前記第一のモーダルと前記第二のモーダルとに関する状況を判断して出力する、

処理を前記コンピュータが実行することを特徴とする請求項1または2に記載の出力方法。

【請求項4】

生成した前記第一のベクトルを、新たな前記第一のモーダルの情報に基づくベクトルに設定し、

生成した前記第二のベクトルを、新たな前記第二のモーダルの情報に基づくベクトルに設定し、

20

設定した前記第一のモーダルの情報に基づくベクトルと、設定した前記第二のモーダルの情報に基づくベクトルとの相関に基づいて、設定した前記第一のモーダルの情報に基づくベクトルを補正し、

設定した前記第一のモーダルの情報に基づくベクトルと、設定した前記第二のモーダルの情報に基づくベクトルとの相関に基づいて、設定した前記第二のモーダルの情報に基づくベクトルを補正し、

補正後の前記第一のモーダルの情報に基づくベクトルから得た前記異なる種類の2つのベクトルの相関に基づいて、前記第一のベクトルを生成し、

補正後の前記第二のモーダルの情報に基づくベクトルから得た前記異なる種類の2つのベクトルの相関に基づいて、前記第二のベクトルを生成する、

30

処理を前記コンピュータが1回以上繰り返し、

前記第三のベクトルを生成する処理は、

前記所定のベクトルと、生成した前記第一のベクトルと、生成した前記第二のベクトルとを含む結合ベクトルから得た前記異なる種類の2つのベクトルの相関に基づいて、前記第一のベクトルと前記第二のベクトルとを集約した第三のベクトルを生成する、ことを特徴とする請求項1～3のいずれか一つに記載の出力方法。

【請求項5】

前記第一のモーダルと前記第二のモーダルとの組は、画像に関するモーダルと文書に関するモーダルとの組、画像に関するモーダルと音声に関するモーダルとの組、第一の言語の文書に関するモーダルと第二の言語の文書に関するモーダルとの組のうちいずれかの組である、ことを特徴とする請求項1～4のいずれか一つに記載の出力方法。

40

【請求項6】

第一のモーダルの情報に基づくベクトルと、前記第一のモーダルとは異なる第二のモーダルの情報に基づくベクトルとの相関に基づいて、前記第一のモーダルの情報に基づくベクトルを補正し、

前記第一のモーダルの情報に基づくベクトルと、前記第二のモーダルの情報に基づくベクトルとの相関に基づいて、前記第二のモーダルの情報に基づくベクトルを補正し、

補正後の前記第一のモーダルの情報に基づくベクトルから得た異なる種類の2つのベクトルの相関に基づいて、第一のベクトルを生成し、

補正後の前記第二のモーダルの情報に基づくベクトルから得た前記異なる種類の2つのベ

50

クトルの相関に基づいて、第二のベクトルを生成し、  
 所定のベクトルと、生成した前記第一のベクトルと、生成した前記第二のベクトルとを含む結合ベクトルから得た前記異なる種類の2つのベクトルの相関に基づいて、前記第一のベクトルと前記第二のベクトルとを集約した第三のベクトルを生成し、  
 生成した前記第三のベクトルを出力する、  
 処理をコンピュータに実行させることを特徴とする出力プログラム。

【請求項7】

第一のモーダルの情報に基づくベクトルと、前記第一のモーダルとは異なる第二のモーダルの情報に基づくベクトルとの相関に基づいて、前記第一のモーダルの情報に基づくベクトルを補正し、

10

前記第一のモーダルの情報に基づくベクトルと、前記第二のモーダルの情報に基づくベクトルとの相関に基づいて、前記第二のモーダルの情報に基づくベクトルを補正し、

補正後の前記第一のモーダルの情報に基づくベクトルから得た異なる種類の2つのベクトルの相関に基づいて、第一のベクトルを生成し、

補正後の前記第二のモーダルの情報に基づくベクトルから得た前記異なる種類の2つのベクトルの相関に基づいて、第二のベクトルを生成し、

所定のベクトルと、生成した前記第一のベクトルと、生成した前記第二のベクトルとを含む結合ベクトルから得た前記異なる種類の2つのベクトルの相関に基づいて、前記第一のベクトルと前記第二のベクトルとを集約した第三のベクトルを生成し、

20

生成した前記第三のベクトルを出力する、

制御部を有することを特徴とする出力装置。

【手続補正2】

【補正対象書類名】明細書

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、出力方法、出力プログラム、および出力装置に関する。

30

【背景技術】

【0002】

従来、複数のモーダルの情報を用いて問題を解く技術がある。この技術は、例えば、文書翻訳や質疑応答、物体検出、状況判断などの問題を解く際に利用される。ここで、モーダルとは、情報の様式や種類を示す概念であり、具体例としては、画像、文書（テキスト）、音声などを挙げることができる。複数のモーダルを用いた機械学習はマルチモーダル学習と呼ばれる。

【0003】

先行技術としては、例えば、ViLBERT (Vision - and - Language Bidirectional Encoder Representations from Transformers) と呼ばれるものがある。具体的には、ViLBERTは、画像に関するモーダルの情報に基づくベクトルに基づいて補正した、文書に関するモーダルの情報に基づくベクトルと、文書に関するモーダルの情報に基づくベクトルに基づいて補正した、画像に関するモーダルの情報に基づくベクトルとを参照し、問題を解く技術である。

40

【先行技術文献】

【非特許文献】

【0004】

【非特許文献1】Lu, Jiasen, et al. "vilbert: Pretraining task-agnostic visiolinguistic re

50

presentations for vision - and - language tasks .” arXiv preprint arXiv : 1908 . 02265 ( 2019 ) .

【発明の概要】

【発明が解決しようとする課題】

【0005】

しかしながら、従来技術では、複数のモーダルの情報を用いて問題を解いた際の解の精度が悪い場合がある。例えば、ViLBERTにおいて、画像と文書とを基に状況を判断する問題を解くにあたり、補正した文書に関するモーダルの情報に基づくベクトルと、補正した画像に関するモーダルの情報に基づくベクトルとをそのまま参照するだけでは、問題を解いた際の解の精度が悪い。

10

【0006】

1つの側面では、本発明は、複数のモーダルの情報を用いて問題を解いた際の解の精度の向上を図ることを目的とする。

【課題を解決するための手段】

【0007】

1つの実施態様によれば、第一のモーダルの情報に基づくベクトルと、前記第一のモーダルとは異なる第二のモーダルの情報に基づくベクトルとの相関に基づいて、前記第一のモーダルの情報に基づくベクトルを補正し、前記第一のモーダルの情報に基づくベクトルと、前記第二のモーダルの情報に基づくベクトルとの相関に基づいて、前記第二のモーダルの情報に基づくベクトルを補正し、補正後の前記第一のモーダルの情報に基づくベクトルから得た異なる種類の2つのベクトルの相関に基づいて、第一のベクトルを生成し、補正後の前記第二のモーダルの情報に基づくベクトルから得た前記異なる種類の2つのベクトルの相関に基づいて、第二のベクトルを生成し、所定のベクトルと、生成した前記第一のベクトルと、生成した前記第二のベクトルとを含む結合ベクトルから得た前記異なる種類の2つのベクトルの相関に基づいて、前記第一のベクトルと前記第二のベクトルとを集約した第三のベクトルを生成し、生成した前記第三のベクトルを出力する出力方法、出力プログラム、および出力装置が提案される。

20

【発明の効果】

【0008】

一態様によれば、複数のモーダルの情報を用いて問題を解いた際の解の精度の向上を図ることが可能になる。

30

【図面の簡単な説明】

【0009】

【図1】図1は、実施の形態にかかる出力方法の一実施例を示す説明図である。

【図2】図2は、情報処理システム200の一例を示す説明図である。

【図3】図3は、出力装置100のハードウェア構成例を示すブロック図である。

【図4】図4は、出力装置100の機能的構成例を示すブロック図である。

【図5】図5は、Co - Attention Network 500の具体例を示す説明図である。

40

【図6】図6は、SA層600の具体例と、TA層610の具体例とを示す説明図である。

【図7】図7は、CAN 500を用いた動作の一例を示す説明図である。

【図8】図8は、出力装置100の利用例1を示す説明図(その1)である。

【図9】図9は、出力装置100の利用例1を示す説明図(その2)である。

【図10】図10は、出力装置100の利用例2を示す説明図(その1)である。

【図11】図11は、出力装置100の利用例2を示す説明図(その2)である。

【図12】図12は、学習処理手順の一例を示すフローチャートである。

【図13】図13は、推定処理手順の一例を示すフローチャートである。

【発明を実施するための形態】

50

## 【0010】

以下に、図面を参照して、本発明にかかる出力方法、出力プログラム、および出力装置の実施の形態を詳細に説明する。

## 【0011】

(実施の形態にかかる出力方法の一実施例)

図1は、実施の形態にかかる出力方法の一実施例を示す説明図である。出力装置100は、複数のモーダルの情報を用いて、問題の解決に有用な情報を得やすくすることにより、問題を解いた際の解の精度の向上を図るためのコンピュータである。

## 【0012】

従来、問題を解くための手法として、例えば、BERT(Bidirectional Encoder Representations from Transformers)と呼ばれるものがある。BERTは、具体的には、TransformerのEncoder部を積み重ねて形成される。BERTについては、例えば、下記非特許文献2や下記非特許文献3を参照することができる。ここで、BERTは、文書に関するモーダルの情報を用いて問題を解くような状況に適用することが想定されており、複数のモーダルの情報を用いて問題を解くような状況に適用することができない。

## 【0013】

非特許文献2 : Devlin, Jacob et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." NAACL-HLT (2019).

## 【0014】

非特許文献3 : Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems. 2017.

## 【0015】

また、問題を解くための手法として、例えば、VideoBERTと呼ばれるものがある。VideoBERTは、具体的には、BERTを、文書に関するモーダルの情報と、画像に関するモーダルの情報とを用いて問題を解くような状況に適用可能に拡張したものである。VideoBERTについては、例えば、下記非特許文献4を参照することができる。ここで、VideoBERTは、問題を解くにあたり、文書に関するモーダルの情報と、画像に関するモーダルの情報とを明示的に区別せずに扱うため、問題を解いた際の解の精度が悪い場合がある。

## 【0016】

非特許文献4 : Sun, Chen, et al. "Videobert: A joint model for video and language representation learning." arXiv preprint arXiv:1904.01766 (2019).

## 【0017】

また、問題を解くための手法として、例えば、MCAN(Modular Co-Attention Network)と呼ばれるものがある。MCANは、文書に関するモーダルの情報と、文書に関するモーダルの情報で補正した画像に関するモーダルの情報とを参照し、問題を解くものである。MCANについては、例えば、下記非特許文献5を参照することができる。ここで、MCANは、問題を解くにあたり、文書に関するモーダルの情報を、画像に関するモーダルの情報で補正せず、そのまま参照するため、問題を解いた際の解の精度が悪い場合がある。

## 【0018】

非特許文献5 : Yu, Zhou, et al. "Deep Modular Co-Attention Networks for Visual Question Answering." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2019.

rence on Computer Vision and Pattern Recognition. 2019.

【0019】

また、上述した通り、問題を解くための手法として、例えば、ViLBERTと呼ばれるものがある。しかしながら、ViLBERTは、画像に関するモダルの情報で補正した文書に関するモダルの情報と、文書に関するモダルの情報で補正した、画像に関するモダルの情報とをそのまま参照するだけであるため、問題を解いた際の解の精度が悪い場合がある。

【0020】

そこで、本実施の形態では、複数のモダルの情報が集約された集約ベクトルを生成することにより、複数のモダルの情報を用いて問題を解くような状況に適用可能であり、問題を解いた際の解の精度を向上可能にすることができる出力方法について説明する。

【0021】

図1において、出力装置100は、第一のモダルの情報に基づくベクトルと、第二のモダルの情報に基づくベクトルとを取得する。モダルの情報は、情報の様式を意味する。第一のモダルの情報と、第二のモダルの情報は、それぞれ異なるモダルである。第一のモダルの情報は、例えば、画像に関するモダルである。第一のモダルの情報は、例えば、第一のモダルに従って表現された、画像である。第二のモダルの情報は、例えば、文書に関するモダルである。第二のモダルの情報は、例えば、第二のモダルに従って表現された、文書である。

【0022】

第一のモダルの情報に基づくベクトルは、例えば、第一のモダルに従って表現された、第一のモダルの情報に基づいて生成されたベクトルである。第一のモダルの情報に基づくベクトルは、具体的には、画像に基づいて生成されたベクトルである。第二のモダルの情報に基づくベクトルは、例えば、第二のモダルに従って表現された、第二のモダルの情報に基づいて生成されたベクトルである。第二のモダルの情報に基づくベクトルは、具体的には、文書に基づいて生成されたベクトルである。

【0023】

(1-1) 出力装置100は、第一のモダルの情報に基づくベクトルと、第二のモダルの情報に基づくベクトルとの相関に基づいて、第一のモダルの情報に基づくベクトルを補正する。出力装置100は、例えば、第一の補正モデル111を用いて、第一のモダルの情報に基づくベクトルを補正する。第一の補正モデル111は、例えば、第一のモダルに関するターゲットアテンション層である。

【0024】

(1-2) 出力装置100は、第一のモダルの情報に基づくベクトルと、第二のモダルの情報に基づくベクトルとの相関に基づいて、第二のモダルの情報に基づくベクトルを補正する。出力装置100は、例えば、第二の補正モデル112を用いて、第二のモダルの情報に基づくベクトルを補正する。第二の補正モデル112は、例えば、第二のモダルに関するターゲットアテンション層である。

【0025】

(1-3) 出力装置100は、補正後の第一のモダルの情報に基づくベクトルから得た異なる種類の2つのベクトルの相関に基づいて、第一のベクトルを生成する。異なる種類の2つのベクトルは、例えば、クエリとなるベクトルと、キーとなるベクトルとである。出力装置100は、例えば、第一の生成モデル121を用いて、第一のベクトルを生成する。第一の生成モデル121は、例えば、第一のモダルに関するセルフアテンション層である。

【0026】

(1-4) 出力装置100は、補正後の第二のモダルの情報に基づくベクトルから得た異なる種類の2つのベクトルの相関に基づいて、第二のベクトルを生成する。異なる種類の2つのベクトルは、例えば、クエリとなるベクトルと、キーとなるベクトルとである。

10

20

30

40

50

出力装置 100 は、例えば、第二の生成モデル 122 を用いて、第二のベクトルを生成する。第二の生成モデル 122 は、例えば、第二のモーダルに関するセルフアテンション層である。

【0027】

(1-5) 出力装置 100 は、所定のベクトルと、生成した第一のベクトルと、生成した第二のベクトルとを含む結合ベクトルを生成する。所定のベクトルは、例えば、予めユーザによって設定される。所定のベクトルは、第一のベクトルと、第二のベクトルとを集約するための集約用ベクトルである。所定のベクトルは、例えば、要素がランダムに設定されたベクトルである。所定のベクトルは、例えば、要素が、ユーザによって設定された既定値のベクトルである。結合ベクトルは、例えば、所定のベクトルと、第一のベクトルと、第二のベクトルとを順に結合することにより得られる。

10

【0028】

そして、出力装置 100 は、結合ベクトルから得た異なる種類の 2 つのベクトルの相関に基づいて、第三のベクトルを生成する。異なる種類の 2 つのベクトルは、例えば、クエリとなるベクトルと、キーとなるベクトルとである。第三のベクトルは、第一のベクトルと第二のベクトルとを集約したベクトルである。出力装置 100 は、第三の生成モデル 130 を用いて、第三のベクトルを生成する。第三の生成モデル 130 は、例えば、セルフアテンション層である。

【0029】

これによれば、出力装置 100 は、第一のベクトルと第二のベクトルとに基づき、キーとなるベクトルに含まれる部分と、所定のベクトルに基づき、クエリとなるベクトルに含まれる部分との相関に基づいて、所定のベクトルを補正することができる。出力装置 100 は、例えば、当該相関に基づいて、第一のベクトルと第二のベクトルとに基づき、バリュエーションとなるベクトルの部分により、所定のベクトルを補正することができる。このため、出力装置 100 は、所定のベクトルに対し、第一のベクトルと第二のベクトルとが集約されるような処理を行うことができ、第三のベクトルを得ることができる。

20

【0030】

(1-6) 出力装置 100 は、生成した第三のベクトルを出力する。出力形式は、例えば、ディスプレイへの表示、プリンタへの印刷出力、他のコンピュータへの送信、または、記憶領域への記憶などである。これにより、出力装置 100 は、第一のベクトルと第二のベクトルとが集約され、第一のモーダルの情報に基づくベクトルと第二のモーダルの情報に基づくベクトルとのうち問題の解決に有用な情報が反映される傾向がある第三のベクトルを生成し、利用可能にすることができる。出力装置 100 は、例えば、実世界の画像や文書の特徴のうち、問題の解決に有用な特徴を、コンピュータ上で精度よく表現した第三のベクトルを利用可能にすることができる。

30

【0031】

出力装置 100 は、例えば、第三のベクトルを利用し、第一の補正モデル 111 と、第二の補正モデル 112 と、第一の生成モデル 121 と、第二の生成モデル 122 と、第三の生成モデル 130 となどを更新することができる。このため、出力装置 100 は、第一のモーダルの情報に基づくベクトルと第二のモーダルの情報に基づくベクトルとのうち問題の解決に有用な情報が、第三のベクトルに反映されやすくすることができる。結果として、出力装置 100 は、以降の、問題を解いた際の解の精度の向上を図ることができる。

40

【0032】

出力装置 100 は、例えば、問題を解くにあたり、第一のモーダルの情報に基づくベクトルと第二のモーダルの情報に基づくベクトルとのうち問題の解決に有用な情報が反映される傾向がある第三のベクトルを利用することができ、問題を解いた際の解の精度を向上させることができる。出力装置 100 は、具体的には、画像と文書とを基に、対象の状況を判断する問題を解くにあたり、正確に対象の状況を判断することができる。対象の状況を判断する問題は、例えば、対象の状況が、ポジティブな状況であるか、またはネガティブな状況であるかを判断する問題である。

50

## 【 0 0 3 3 】

( 情報処理システム 2 0 0 の一例 )

次に、図 2 を用いて、図 1 に示した出力装置 1 0 0 を適用した、情報処理システム 2 0 0 の一例について説明する。

## 【 0 0 3 4 】

図 2 は、情報処理システム 2 0 0 の一例を示す説明図である。図 2 において、情報処理システム 2 0 0 は、出力装置 1 0 0 と、クライアント装置 2 0 1 と、端末装置 2 0 2 とを含む。

## 【 0 0 3 5 】

情報処理システム 2 0 0 において、出力装置 1 0 0 とクライアント装置 2 0 1 とは、有線または無線のネットワーク 2 1 0 を介して接続される。ネットワーク 2 1 0 は、例えば、LAN ( Local Area Network )、WAN ( Wide Area Network )、インターネットなどである。また、情報処理システム 2 0 0 において、出力装置 1 0 0 と端末装置 2 0 2 とは、有線または無線のネットワーク 2 1 0 を介して接続される。

10

## 【 0 0 3 6 】

出力装置 1 0 0 は、第一のモーダルの情報に基づくベクトルと、第二のモーダルの情報に基づくベクトルとに基づいて、第三のベクトルを生成する Co - Attention Network を有する。第一のモーダルは、例えば、画像に関するモーダルである。第二のモーダルは、例えば、文書に関するモーダルである。Co - Attention Network は、例えば、図 1 に示した、第一の補正モデル 1 1 1 と、第二の補正モデル 1 1 2 と、第一の生成モデル 1 2 1 と、第二の生成モデル 1 2 2 と、第三の生成モデル 1 3 0 との全体に対応する。

20

## 【 0 0 3 7 】

出力装置 1 0 0 は、教師データに基づいて、Co - Attention Network を更新する。教師データは、例えば、標本となる第一のモーダルの情報に基づくベクトルを生成する元となる第一のモーダルの情報と、標本となる第二のモーダルの情報に基づくベクトルを生成する元となる第二のモーダルの情報と、正解データとを対応付けた対応情報である。教師データは、例えば、出力装置 1 0 0 のユーザにより出力装置 1 0 0 に入力される。正解データは、例えば、第三のベクトルに基づいて、問題を解いた場合についての正解を示す。例えば、第一のモーダルが、画像に関するモーダルであれば、第一のモーダルの情報に基づくベクトルを生成する元となる第一のモーダルの情報は、画像である。例えば、第二のモーダルが、文書に関するモーダルであれば、第二のモーダルの情報に基づくベクトルを生成する元となる第二のモーダルの情報は、文書である。

30

## 【 0 0 3 8 】

出力装置 1 0 0 は、例えば、教師データのうち第一のモーダルの情報となる画像から、第一のモーダルの情報に基づくベクトルを生成することにより取得し、教師データのうち第二のモーダルの情報となる文書から、第二のモーダルの情報に基づくベクトルを生成することにより取得する。そして、出力装置 1 0 0 は、取得した第一のモーダルの情報に基づくベクトルと、取得した第二のモーダルの情報に基づくベクトルと、教師データの正解データとに基づいて、誤差逆伝搬などにより、Co - Attention Network を更新する。出力装置 1 0 0 は、誤差逆伝搬以外の学習方法により、Co - Attention Network を更新してもよい。

40

## 【 0 0 3 9 】

出力装置 1 0 0 は、第一のモーダルの情報に基づくベクトルと、第二のモーダルの情報に基づくベクトルとを取得する。そして、出力装置 1 0 0 は、Co - Attention Network を用いて、取得した第一のモーダルの情報に基づくベクトルと、取得した第二のモーダルの情報に基づくベクトルとに基づいて、第三のベクトルを生成し、生成した第三のベクトルに基づいて、問題を解く。その後、出力装置 1 0 0 は、問題を解いた結果を、クライアント装置 2 0 1 に送信する。

50

## 【 0 0 4 0 】

出力装置 1 0 0 は、例えば、出力装置 1 0 0 のユーザにより出力装置 1 0 0 に入力された第一のモーダルの情報に基づくベクトルを取得する。また、出力装置 1 0 0 は、第一のモーダルの情報に基づくベクトルを、クライアント装置 2 0 1 または端末装置 2 0 2 から受信することにより取得してもよい。また、出力装置 1 0 0 は、例えば、第一のモーダルの情報に基づくベクトルを生成する元となる第一のモーダルの情報を、クライアント装置 2 0 1 または端末装置 2 0 2 から受信することにより取得してもよい。例えば、第一のモーダルが、画像に関するモーダルであれば、第一のモーダルの情報に基づくベクトルを生成する元となる第一のモーダルの情報は、画像である。

## 【 0 0 4 1 】

出力装置 1 0 0 は、例えば、出力装置 1 0 0 のユーザにより出力装置 1 0 0 に入力された第二のモーダルの情報に基づくベクトルを取得する。また、出力装置 1 0 0 は、第二のモーダルの情報に基づくベクトルを、クライアント装置 2 0 1 または端末装置 2 0 2 から受信することにより取得してもよい。また、出力装置 1 0 0 は、例えば、第二のモーダルの情報に基づくベクトルを生成する元となる第二のモーダルの情報を、クライアント装置 2 0 1 または端末装置 2 0 2 から受信することにより取得してもよい。例えば、第二のモーダルが、文書に関するモーダルであれば、第二のモーダルの情報に基づくベクトルを生成する元となる第二のモーダルの情報は、文書である。

## 【 0 0 4 2 】

そして、出力装置 1 0 0 は、Co - Attention Network を用いて、取得した第一のモーダルの情報に基づくベクトルと、取得した第二のモーダルの情報に基づくベクトルとに基づいて、第三のベクトルを生成し、生成した第三のベクトルに基づいて、問題を解く。その後、出力装置 1 0 0 は、問題を解いた結果を、クライアント装置 2 0 1 に送信する。出力装置 1 0 0 は、例えば、サーバや PC ( Personal Computer ) などである。

## 【 0 0 4 3 】

クライアント装置 2 0 1 は、出力装置 1 0 0 と通信可能なコンピュータである。クライアント装置 2 0 1 は、例えば、第一のモーダルの情報に基づくベクトルを、出力装置 1 0 0 に送信してもよい。また、クライアント装置 2 0 1 は、例えば、第一のモーダルの情報に基づくベクトルを生成する元となる第一のモーダルの情報を、出力装置 1 0 0 に送信してもよい。クライアント装置 2 0 1 は、例えば、第二のモーダルの情報に基づくベクトルを、出力装置 1 0 0 に送信してもよい。また、クライアント装置 2 0 1 は、例えば、第二のモーダルの情報に基づくベクトルを生成する元となる第二のモーダルの情報を、出力装置 1 0 0 に送信してもよい。

## 【 0 0 4 4 】

クライアント装置 2 0 1 は、出力装置 1 0 0 が問題を解いた結果を受信して出力する。出力形式は、例えば、ディスプレイへの表示、プリンタへの印刷出力、他のコンピュータへの送信、または、記憶領域への記憶などである。クライアント装置 2 0 1 は、例えば、PC、タブレット端末、またはスマートフォンなどである。

## 【 0 0 4 5 】

端末装置 2 0 2 は、出力装置 1 0 0 と通信可能なコンピュータである。端末装置 2 0 2 は、例えば、第一のモーダルの情報に基づくベクトルを、出力装置 1 0 0 に送信してもよい。また、端末装置 2 0 2 は、例えば、第一のモーダルの情報に基づくベクトルを生成する元となる第一のモーダルの情報を、出力装置 1 0 0 に送信してもよい。端末装置 2 0 2 は、例えば、第二のモーダルの情報に基づくベクトルを、出力装置 1 0 0 に送信してもよい。また、端末装置 2 0 2 は、例えば、第二のモーダルの情報に基づくベクトルを生成する元となる第二のモーダルの情報を、出力装置 1 0 0 に送信してもよい。端末装置 2 0 2 は、例えば、PC、タブレット端末、スマートフォン、電子機器、IoT 機器、またはセンサ装置などである。端末装置 2 0 2 は、具体的には、監視カメラであってもよい。

## 【 0 0 4 6 】

10

20

30

40

50

ここでは、出力装置100が、Co-Attention Networkを更新し、かつ、Co-Attention Networkを用いて、問題を解く場合について説明したが、これに限らない。例えば、他のコンピュータが、Co-Attention Networkを更新し、出力装置100が、他のコンピュータから受信したCo-Attention Networkを用いて、問題を解く場合があってもよい。また、例えば、出力装置100が、Co-Attention Networkを更新し、他のコンピュータに提供し、他のコンピュータで、Co-Attention Networkを用いて、問題を解く場合があってもよい。

**【0047】**

ここでは、教師データが、標本となる第一のモーダルの情報に基づくベクトルを生成する元となる第一のモーダルの情報と、標本となる第二のモーダルの情報に基づくベクトルを生成する元となる第二のモーダルの情報と、正解データとを対応付けた対応情報である場合について説明したが、これに限らない。例えば、教師データが、標本となる第一のモーダルの情報に基づくベクトルと、標本となる第二のモーダルの情報に基づくベクトルと、正解データとを対応付けた対応情報である場合があってもよい。

10

**【0048】**

ここでは、出力装置100が、クライアント装置201や端末装置202とは異なる装置である場合について説明したが、これに限らない。例えば、出力装置100が、クライアント装置201と一体である場合があってもよい。また、例えば、出力装置100が、端末装置202と一体である場合があってもよい。

20

**【0049】**

ここでは、出力装置100が、ソフトウェア的に、Co-Attention Networkを実現する場合について説明したが、これに限らない。例えば、出力装置100が、Co-Attention Networkを、電子回路的に実現する場合があってもよい。

**【0050】**

(情報処理システム200の適用例1)

適用例1において、出力装置100は、画像と、画像についての質問文となる文書とを記憶する。質問文は、例えば、「画像内で何を切っているか」である。そして、出力装置100は、画像と文書とに基づいて、質問文に対する回答文を推定する問題を解く。出力装置100は、例えば、画像と文書とに基づいて、画像内で何を切っているかの質問文に対する回答文を推定し、クライアント装置201に送信する。

30

**【0051】**

(情報処理システム200の適用例2)

適用例2において、端末装置202は、監視カメラであり、対象を撮像した画像を、出力装置100に送信する。対象は、具体的には、試着室の外観である。また、出力装置100は、対象についての説明文となる文書を記憶している。説明文は、具体的には、人間が試着室を利用中は、試着室のカーテンが閉まっている傾向があることの説明文である。そして、出力装置100は、画像と文書とに基づいて、危険度を判断する問題を解く。危険度は、例えば、試着室に避難が未完了の人間が残っている可能性の高さを示す指標値である。出力装置100は、例えば、災害時に、試着室に避難が未完了の人間が残っている可能性の高さを示す危険度を判断する。

40

**【0052】**

(情報処理システム200の適用例3)

適用例3において、出力装置100は、動画を形成する画像と、画像についての説明文となる文書を記憶している。動画は、例えば、料理の様子を写した動画である。説明文は、具体的には、料理の手順についての説明文である。そして、出力装置100は、画像と文書とに基づいて、危険度を判断する問題を解く。危険度は、例えば、料理中の危険性の高さを示す指標値である。出力装置100は、例えば、料理中の危険性の高さを示す危険度を判断する。

50

## 【0053】

(出力装置100のハードウェア構成例)

次に、図3を用いて、出力装置100のハードウェア構成例について説明する。

## 【0054】

図3は、出力装置100のハードウェア構成例を示すブロック図である。図3において、出力装置100は、CPU(Central Processing Unit)301と、メモリ302と、ネットワークI/F(Interface)303と、記録媒体I/F304と、記録媒体305とを有する。また、各構成部は、バス300によってそれぞれ接続される。

## 【0055】

ここで、CPU301は、出力装置100の全体の制御を司る。メモリ302は、例えば、ROM(Read Only Memory)、RAM(Random Access Memory)およびフラッシュROMなどを有する。具体的には、例えば、フラッシュROMやROMが各種プログラムを記憶し、RAMがCPU301のワークエリアとして使用される。メモリ302に記憶されるプログラムは、CPU301にロードされることで、コーディングされている処理をCPU301に実行させる。

## 【0056】

ネットワークI/F303は、通信回線を通じてネットワーク210に接続され、ネットワーク210を介して他のコンピュータに接続される。そして、ネットワークI/F303は、ネットワーク210と内部のインターフェースを司り、他のコンピュータからのデータの入出力を制御する。ネットワークI/F303は、例えば、モデムやLANアダプタなどである。

## 【0057】

記録媒体I/F304は、CPU301の制御に従って記録媒体305に対するデータのリード/ライトを制御する。記録媒体I/F304は、例えば、ディスクドライブ、SSD(Solid State Drive)、USB(Universal Serial Bus)ポートなどである。記録媒体305は、記録媒体I/F304の制御で書き込まれたデータを記憶する不揮発メモリである。記録媒体305は、例えば、ディスク、半導体メモリ、USBメモリなどである。記録媒体305は、出力装置100から着脱可能であってもよい。

## 【0058】

出力装置100は、上述した構成部のほか、例えば、キーボード、マウス、ディスプレイ、プリンタ、スキャナ、マイク、スピーカーなどを有してもよい。また、出力装置100は、記録媒体I/F304や記録媒体305を複数有していてもよい。また、出力装置100は、記録媒体I/F304や記録媒体305を有していなくてもよい。

## 【0059】

(クライアント装置201のハードウェア構成例)

クライアント装置201のハードウェア構成例は、具体的には、図3に示した出力装置100のハードウェア構成例と同様であるため、説明を省略する。

## 【0060】

(端末装置202のハードウェア構成例)

端末装置202のハードウェア構成例は、具体的には、図3に示した出力装置100のハードウェア構成例と同様であるため、説明を省略する。

## 【0061】

(出力装置100の機能的構成例)

次に、図4を用いて、出力装置100の機能的構成例について説明する。

## 【0062】

図4は、出力装置100の機能的構成例を示すブロック図である。出力装置100は、記憶部400と、取得部401と、第一の補正部402と、第一の生成部403と、第二の補正部404と、第二の生成部405と、第三の生成部406と、解析部407と、出力

10

20

30

40

50

部 4 0 8 とを含む。

【 0 0 6 3 】

記憶部 4 0 0 は、例えば、図 3 に示したメモリ 3 0 2 や記録媒体 3 0 5 などの記憶領域によって実現される。以下では、記憶部 4 0 0 が、出力装置 1 0 0 に含まれる場合について説明するが、これに限らない。例えば、記憶部 4 0 0 が、出力装置 1 0 0 とは異なる装置に含まれ、記憶部 4 0 0 の記憶内容が出力装置 1 0 0 から参照可能である場合があってもよい。

【 0 0 6 4 】

取得部 4 0 1 ~ 出力部 4 0 8 は、制御部の一例として機能する。取得部 4 0 1 ~ 出力部 4 0 8 は、具体的には、例えば、図 3 に示したメモリ 3 0 2 や記録媒体 3 0 5 などの記憶領域に記憶されたプログラムを CPU 3 0 1 に実行させることにより、または、ネットワーク I / F 3 0 3 により、その機能を実現する。各機能部の処理結果は、例えば、図 3 に示したメモリ 3 0 2 や記録媒体 3 0 5 などの記憶領域に記憶される。

10

【 0 0 6 5 】

記憶部 4 0 0 は、各機能部の処理において参照され、または更新される各種情報を記憶する。記憶部 4 0 0 は、Co - Attention Network を記憶する。Co - Attention Network は、第一のモーダルの情報に基づくベクトルと、第二のモーダルの情報に基づくベクトルとに基づいて、第三のベクトルを生成するモデルである。Co - Attention Network は、後述する第一のターゲットアテンション層と、第二のターゲットアテンション層と、第一のセルフアテンション層と、第二のセルフアテンション層と、第三のセルフアテンション層との全体に対応する。

20

【 0 0 6 6 】

第一のターゲットアテンション層は、例えば、第一のモーダルに関する。第一のターゲットアテンション層は、第一のモーダルの情報に基づくベクトルを補正するモデルである。第一のセルフアテンション層は、例えば、第一のモーダルに関する。第一のセルフアテンション層は、補正後の第一のモーダルの情報に基づくベクトルをさらに補正し、第一のベクトルを生成するモデルである。第二のターゲットアテンション層は、例えば、第二のモーダルに関する。第二のターゲットアテンション層は、第二のモーダルの情報に基づくベクトルを補正するモデルである。第二のセルフアテンション層は、例えば、第二のモーダルに関する。第二のセルフアテンション層は、補正後の第二のモーダルの情報に基づくベクトルをさらに補正し、第二のベクトルを生成するモデルである。第三のセルフアテンション層は、第一のベクトルと、第二のベクトルとに基づいて、第三のベクトルを生成するモデルである。

30

【 0 0 6 7 】

例えば、第一のモーダルは、画像に関するモーダルであり、第二のモーダルは、文書に関するモーダルである。例えば、第一のモーダルは、画像に関するモーダルであり、第二のモーダルは、音声に関するモーダルである。例えば、第一のモーダルは、第一の言語の文書に関するモーダルであり、第二のモーダルは、第二の言語の文書に関するモーダルである。Co - Attention Network は、解析部 4 0 7 によって更新され、または解析部 4 0 7 によって問題を解く際に利用される。

40

【 0 0 6 8 】

記憶部 4 0 0 は、例えば、Co - Attention Network のパラメータを記憶する。記憶部 4 0 0 は、具体的には、第一のターゲットアテンション層と、第二のターゲットアテンション層と、第一のセルフアテンション層と、第二のセルフアテンション層と、第三のセルフアテンション層とのパラメータを記憶する。

【 0 0 6 9 】

記憶部 4 0 0 は、教師データを記憶してもよい。教師データは、例えば、標本となる第一のモーダルの情報に基づくベクトルを生成する元となる第一のモーダルの情報と、標本となる第二のモーダルの情報に基づくベクトルを生成する元となる第二のモーダルの情報と、正解データとを対応付けた対応情報である。教師データは、例えば、ユーザにより入力

50

される。正解データは、例えば、第三のベクトルに基づいて、問題を解いた場合についての正解を示す。

【0070】

例えば、第一のモーダルが、画像に関するモーダルであれば、第一のモーダルの情報に基づくベクトルを生成する元となる第一のモーダルの情報は、画像である。例えば、第二のモーダルが、文書に関するモーダルであれば、第二のモーダルの情報に基づくベクトルを生成する元となる第二のモーダルの情報は、文書である。教師データは、標本となる第一のモーダルの情報に基づくベクトルと、標本となる第二のモーダルの情報に基づくベクトルと、正解データとを対応付けた対応情報であってもよい。

【0071】

取得部401は、各機能部の処理に用いられる各種情報を取得する。取得部401は、取得した各種情報を、記憶部400に記憶し、または、各機能部へ出力する。また、取得部401は、記憶部400に記憶しておいた各種情報を、各機能部へ出力してもよい。取得部401は、例えば、ユーザの操作入力に基づき、各種情報を取得する。取得部401は、例えば、出力装置100とは異なる装置から、各種情報を受信してもよい。

【0072】

取得部401は、第一のモーダルの情報に基づくベクトルと、第二のモーダルの情報に基づくベクトルとを取得する。取得部401は、Co-Attention Networkを更新する際に、教師データを取得し、教師データに基づいて、第一のモーダルの情報に基づくベクトルと、第二のモーダルの情報に基づくベクトルとを取得する。

【0073】

取得部401は、例えば、ユーザによる教師データの入力を受け付け、教師データの中から、第一のモーダルの情報に基づくベクトルを生成する元となる第一のモーダルの情報と、第二のモーダルの情報に基づくベクトルを生成する元となる第二のモーダルの情報とを取得する。そして、取得部401は、取得した各種情報に基づいて、第一のモーダルの情報に基づくベクトルと、第二のモーダルの情報に基づくベクトルとを生成する。

【0074】

取得部401は、具体的には、教師データに含まれる画像を取得し、第一のモーダルの情報に基づくベクトルとして、取得した画像に関する特徴量ベクトルを生成する。画像に関する特徴量ベクトルは、例えば、画像に写る物体ごとの特徴量ベクトルを並べたものである。また、取得部401は、具体的には、教師データに含まれる文書を取得し、第二のモーダルの情報に基づくベクトルとして、取得した文書に関する特徴量ベクトルを生成する。文書に関する特徴量ベクトルは、例えば、文書に含まれる単語ごとの特徴量ベクトルを並べたものである。

【0075】

取得部401は、例えば、教師データを、クライアント装置201または端末装置202から受信し、受信した教師データの中から、第一のモーダルの情報に基づくベクトルを生成する元となる第一のモーダルの情報と、第二のモーダルの情報に基づくベクトルを生成する元となる第二のモーダルの情報とを取得してもよい。そして、取得部401は、取得した情報に基づいて、第一のモーダルの情報に基づくベクトルと、第二のモーダルの情報に基づくベクトルとを生成する。

【0076】

取得部401は、具体的には、教師データに含まれる画像を取得し、第一のモーダルの情報に基づくベクトルとして、取得した画像に関する特徴量ベクトルを生成する。画像に関する特徴量ベクトルは、例えば、画像に写る物体ごとの特徴量ベクトルを並べたものである。また、取得部401は、具体的には、教師データに含まれる文書を取得し、第二のモーダルの情報に基づくベクトルとして、取得した文書に関する特徴量ベクトルを生成する。文書に関する特徴量ベクトルは、例えば、文書に含まれる単語ごとの特徴量ベクトルを並べたものである。

【0077】

10

20

30

40

50

取得部 401 は、例えば、ユーザによる教師データの入力を受け付け、教師データの中から、第一のモーダルの情報に基づくベクトルと、第二のモーダルの情報に基づくベクトルとを、そのまま取得してもよい。取得部 401 は、例えば、教師データを、クライアント装置 201 または端末装置 202 から受信し、受信した教師データの中から、第一のモーダルの情報に基づくベクトルと、第二のモーダルの情報に基づくベクトルとを、そのまま取得してもよい。

【0078】

取得部 401 は、Co-Attention Network を利用し、問題を解く際に、第一のモーダルの情報に基づくベクトルと、第二のモーダルの情報に基づくベクトルとを取得する。取得部 401 は、例えば、ユーザによる、第一のモーダルの情報に基づくベクトルを生成する元となる第一のモーダルの情報と、第二のモーダルの情報に基づくベクトルを生成する元となる第二のモーダルの情報との入力を受け付ける。そして、取得部 401 は、入力された各種情報に基づいて、第一のモーダルの情報に基づくベクトルと、第二のモーダルの情報に基づくベクトルとを生成する。

10

【0079】

取得部 401 は、具体的には、画像を取得し、第一のモーダルの情報に基づくベクトルとして、取得した画像に関する特徴量ベクトルを生成する。画像に関する特徴量ベクトルは、例えば、画像に写る物体ごとの特徴量ベクトルを並べたものである。また、取得部 401 は、具体的には、文書を取得し、第二のモーダルの情報に基づくベクトルとして、取得した文書に関する特徴量ベクトルを生成する。文書に関する特徴量ベクトルは、例えば、文書に含まれる単語ごとの特徴量ベクトルを並べたものである。

20

【0080】

取得部 401 は、例えば、第一のモーダルの情報に基づくベクトルを生成する元となる第一のモーダルの情報と、第二のモーダルの情報に基づくベクトルを生成する元となる第二のモーダルの情報とを、クライアント装置 201 または端末装置 202 から受信してもよい。そして、取得部 401 は、取得した各種情報に基づいて、第一のモーダルの情報に基づくベクトルと、第二のモーダルの情報に基づくベクトルとを生成する。

【0081】

取得部 401 は、具体的には、画像を取得し、第一のモーダルの情報に基づくベクトルとして、取得した画像に関する特徴量ベクトルを生成する。画像に関する特徴量ベクトルは、例えば、画像に写る物体ごとの特徴量ベクトルを並べたものである。取得部 401 は、具体的には、文書を取得し、第二のモーダルの情報に基づくベクトルとして、取得した文書に関する特徴量ベクトルを生成する。文書に関する特徴量ベクトルは、例えば、文書に含まれる単語ごとの特徴量ベクトルを並べたものである。

30

【0082】

取得部 401 は、例えば、ユーザによる、第一のモーダルの情報に基づくベクトルと、第二のモーダルの情報に基づくベクトルとの入力を受け付けてもよい。取得部 401 は、例えば、第一のモーダルの情報に基づくベクトルと、第二のモーダルの情報に基づくベクトルとを、クライアント装置 201 または端末装置 202 から受信してもよい。

【0083】

取得部 401 は、いずれかの機能部の処理を開始する開始トリガーを受け付けてもよい。開始トリガーは、例えば、ユーザによる所定の操作入力があったことである。開始トリガーは、例えば、他のコンピュータから、所定の情報を受信したことであってもよい。開始トリガーは、例えば、いずれかの機能部が所定の情報を出力したことであってもよい。取得部 401 は、例えば、第一のモーダルの情報に基づくベクトルと、第二のモーダルの情報に基づくベクトルとを取得したことを、各機能部の処理を開始する開始トリガーとして受け付ける。

40

【0084】

第一の補正部 402 は、第一のモーダルの情報に基づくベクトルと、第二のモーダルの情報に基づくベクトルとの相関に基づいて、第一のモーダルの情報に基づくベクトルを補正

50

する。相関は、例えば、第一のモーダルの情報に基づくベクトルから得たベクトルと、第二のモーダルの情報に基づくベクトルから得たベクトルとの類似度によって表現される。第一のモーダルの情報に基づくベクトルから得たベクトルは、例えば、クエリである。第二のモーダルの情報に基づくベクトルから得たベクトルは、例えば、キーである。類似度は、例えば、内積によって表現される。類似度は、例えば、差分の二乗和などによって表現されてもよい。

【0085】

第一の補正部402は、例えば、第一のターゲットアテンション層を用いて、第一のモーダルの情報に基づくベクトルから得たベクトルと、第二のモーダルの情報に基づくベクトルから得たベクトルとの内積に基づいて、第一のモーダルの情報に基づくベクトルを補正する。

10

【0086】

第一の補正部402は、具体的には、第一のターゲットアテンション層を用いて、第一のモーダルの情報に基づくベクトルから得たクエリと、第二のモーダルの情報に基づくベクトルから得たキーとの内積に基づいて、第一のモーダルの情報に基づくベクトルを補正する。ここで、第一のモーダルの情報に基づくベクトルを補正する一例は、例えば、図5を用いて後述する動作例に示す。これにより、第一の補正部402は、第二のモーダルの情報に基づくベクトルのうち、第一のモーダルの情報に基づくベクトルと相対的に関連深い成分ほど、第一のモーダルの情報に基づくベクトルに強く反映されるように、第一のモーダルの情報に基づくベクトルを補正することができる。

20

【0087】

第一の生成部403は、補正後の第一のモーダルの情報に基づくベクトルから得た異なる種類の2つのベクトルの相関に基づいて、第一のベクトルを生成する。相関は、例えば、補正後の第一のモーダルの情報に基づくベクトルから得た異なる種類の2つのベクトルの類似度によって表現される。異なる種類の2つのベクトルは、例えば、クエリとキーとである。類似度は、例えば、内積によって表現される。類似度は、例えば、差分の二乗和などによって表現されてもよい。

【0088】

第一の生成部403は、例えば、第一のセルフアテンション層を用いて、補正後の第一のモーダルの情報に基づくベクトルから得た異なる種類の2つのベクトルの内積に基づいて、補正後の第一のモーダルの情報に基づくベクトルをさらに補正し、第一のベクトルを生成する。

30

【0089】

第一の生成部403は、具体的には、第一のセルフアテンション層を用いて、補正後の第一のモーダルの情報に基づくベクトルから得たクエリとキーとの内積に基づいて、補正後の第一のモーダルの情報に基づくベクトルをさらに補正し、第一のベクトルを生成する。ここで、第一のベクトルを生成する一例は、例えば、図5を用いて後述する動作例に示す。これにより、第一の生成部403は、補正後の第一のモーダルの情報に基づくベクトルのうち、より有用な成分ほど大きくなるように、補正後の第一のモーダルの情報に基づくベクトルをさらに補正することができる。

40

【0090】

第二の補正部404は、第一のモーダルの情報に基づくベクトルと、第二のモーダルの情報に基づくベクトルとの相関に基づいて、第二のモーダルの情報に基づくベクトルを補正する。相関は、例えば、第一のモーダルの情報に基づくベクトルから得たベクトルと、第二のモーダルの情報に基づくベクトルから得たベクトルとの類似度によって表現される。第一のモーダルの情報に基づくベクトルから得たベクトルは、例えば、キーである。第二のモーダルの情報に基づくベクトルから得たベクトルは、例えば、クエリである。類似度は、例えば、内積によって表現される。類似度は、例えば、差分の二乗和などによって表現されてもよい。

【0091】

50

第二の補正部 404 は、例えば、第二のターゲットアテンション層を用いて、第一のモーダルの情報に基づくベクトルから得たベクトルと、第二のモーダルの情報に基づくベクトルから得たベクトルとの内積に基づいて、第二のモーダルの情報に基づくベクトルを補正する。

【0092】

第二の補正部 404 は、具体的には、第二のターゲットアテンション層を用いて、第一のモーダルの情報に基づくベクトルから得たキーと、第二のモーダルの情報に基づくベクトルから得たクエリとの内積に基づいて、第二のモーダルの情報に基づくベクトルを補正する。ここで、第二のモーダルの情報に基づくベクトルを補正する一例は、例えば、図 5 を用いて後述する動作例に示す。これにより、第二の補正部 404 は、第一のモーダルの情報に基づくベクトルのうち、第二のモーダルの情報に基づくベクトルと相対的に関連深い成分ほど、第二のモーダルの情報に基づくベクトルに強く反映されるように、第二のモーダルの情報に基づくベクトルを補正することができる。

10

【0093】

第二の生成部 405 は、補正後の第二のモーダルの情報に基づくベクトルから得た異なる種類の 2 つのベクトルの相関に基づいて、第二のベクトルを生成する。相関は、例えば、補正後の第二のモーダルの情報に基づくベクトルから得た異なる種類の 2 つのベクトルの類似度によって表現される。異なる種類の 2 つのベクトルは、例えば、クエリとキーとである。類似度は、例えば、内積によって表現される。類似度は、例えば、差分の二乗和などによって表現されてもよい。

20

【0094】

第二の生成部 405 は、例えば、第二のセルフアテンション層を用いて、補正後の第二のモーダルの情報に基づくベクトルから得た異なる種類の 2 つのベクトルの内積に基づいて、補正後の第二のモーダルの情報に基づくベクトルをさらに補正し、第二のベクトルを生成する。

【0095】

第二の生成部 405 は、具体的には、第二のセルフアテンション層を用いて、補正後の第二のモーダルの情報に基づくベクトルから得たクエリとキーとの内積に基づいて、補正後の第二のモーダルの情報に基づくベクトルをさらに補正し、第二のベクトルを生成する。ここで、第二のベクトルを生成する一例は、例えば、図 5 を用いて後述する動作例に示す。これにより、第二の生成部 405 は、補正後の第二のモーダルの情報に基づくベクトルのうち、より有用な成分ほど大きくなるように、補正後の第二のモーダルの情報に基づくベクトルをさらに補正することができる。

30

【0096】

ここで、出力装置 100 は、第一の補正部 402 ~ 第二の生成部 405 の動作を、1 回以上繰り返してもよい。出力装置 100 は、例えば、第一の補正部 402 ~ 第二の生成部 405 の動作を繰り返す際には、生成した第一のベクトルを、新たな第一のモーダルの情報に基づくベクトルに設定し、生成した第二のベクトルを、新たな第二のモーダルの情報に基づくベクトルに設定する。これにより、出力装置 100 は、問題を解いた際の解の精度をさらに向上可能にすることができる。出力装置 100 は、例えば、問題を解いた際の解の精度を向上させる観点で、第三のベクトルをより有用な状態で生成可能にすることができる。

40

【0097】

第三の生成部 406 は、結合ベクトルを生成する。結合ベクトルは、所定のベクトルと、生成した第一のベクトルと、生成した第二のベクトルとを含む。第三の生成部 406 は、例えば、所定のベクトルと第一のベクトルと第二のベクトルとを結合した結合ベクトルを生成する。第三の生成部 406 は、例えば、第一の補正部 402 ~ 第二の生成部 405 の動作を繰り返した後であれば、所定のベクトルと最後に生成した第一のベクトルと最後に生成した第二のベクトルとを結合した結合ベクトルを生成する。

【0098】

50

第三の生成部 406 は、結合ベクトルから得た異なる種類の 2 つのベクトルの相関に基づいて、第一のベクトルと第二のベクトルとを集約した第三のベクトルを生成する。相関は、例えば、結合ベクトルから得た異なる種類の 2 つのベクトルの類似度によって表現される。異なる種類の 2 つのベクトルは、例えば、クエリとキーとである。類似度は、例えば、内積によって表現される。類似度は、例えば、差分の二乗和などによって表現されてもよい。

【0099】

第三の生成部 406 は、例えば、第三のセルフアテンション層を用いて、結合ベクトルから得た異なる種類の 2 つのベクトルの内積に基づいて、結合ベクトルを補正し、第三のベクトルを生成する。第三のベクトルは、例えば、補正後の結合ベクトルのうち、所定のベクトルに対応する位置に含まれる部分的なベクトルである。

10

【0100】

第三の生成部 406 は、具体的には、第三のセルフアテンション層を用いて、結合ベクトルから得たクエリとキーとの内積に基づいて、結合ベクトルを補正することにより、第三のベクトルを含む補正後の結合ベクトルを生成する。ここで、第三のベクトルを生成する一例は、例えば、図 5 を用いて後述する動作例に示す。これにより、第三の生成部 406 は、問題を解いた際の解の精度を向上させる観点で有用な第三のベクトルを生成し、参照可能にすることができる。

【0101】

解析部 407 は、生成した第三のベクトルに基づいて、Co-Attention Network を更新する。解析部 407 は、例えば、第三のベクトルに基づいて、第一のターゲットアテンション層と、第二のターゲットアテンション層と、第一のセルフアテンション層と、第二のセルフアテンション層と、第三のセルフアテンション層とを更新する。更新は、例えば、誤差逆伝搬によって実施される。

20

【0102】

解析部 407 は、具体的には、生成した第三のベクトルを用いて、試験的に問題を解き、正解データと比較する。問題の一例は、例えば、第一のモーダルと第二のモーダルとに関する状況が、ポジティブな状況であるか、または、ネガティブな状況であるかを判断する問題である。問題の一例は、具体的には、画像が示唆する状況が、人間に危害が及び得る状況であるか、または、人間に危害が及ばない状況であるかを判断する問題である。

30

【0103】

そして、解析部 407 は、比較した結果に基づいて、第一のターゲットアテンション層と、第二のターゲットアテンション層と、第一のセルフアテンション層と、第二のセルフアテンション層と、第三のセルフアテンション層とを更新する。これにより、解析部 407 は、第三のベクトルをより有用な状態で生成可能に、各種アテンション層を更新し、問題を解いた際の解の精度を向上可能にすることができる。

【0104】

解析部 407 は、生成した第三のベクトルを用いて、実際の問題を解く。問題の一例は、例えば、第一のモーダルと第二のモーダルとに関する状況が、ポジティブな状況であるか、または、ネガティブな状況であるかを判断する問題である。問題の一例は、具体的には、画像が示唆する状況が、人間に危害が及び得る状況であるか、または、人間に危害が及ばない状況であるかを判断する問題である。これにより、解析部 407 は、問題を解いた際の解の精度を向上させることができる。

40

【0105】

出力部 408 は、いずれかの機能部の処理結果を出力する。出力形式は、例えば、ディスプレイへの表示、プリンタへの印刷出力、ネットワーク I/F 303 による外部装置への送信、または、メモリ 302 や記録媒体 305 などの記憶領域への記憶である。これにより、出力部 408 は、各機能部の処理結果をユーザに通知可能にし、出力装置 100 の利便性の向上を図ることができる。

【0106】

50

出力部 408 は、例えば、更新済みの Co - Attention Network を出力する。出力部 408 は、具体的には、更新済みの第一のターゲットアテンション層と、第二のターゲットアテンション層と、第一のセルフアテンション層と、第二のセルフアテンション層と、第三のセルフアテンション層とを出力する。これにより、出力部 408 は、更新済みの Co - Attention Network を参照可能にすることができる。このため、出力部 408 は、例えば、他のコンピュータで、更新済みの Co - Attention Network を用いて、問題を解いた際の解の精度を向上可能にすることができる。

【0107】

出力部 408 は、例えば、生成した第三のベクトルを出力する。これにより、出力部 408 は、第三のベクトルを参照可能にし、Co - Attention Network を更新可能にしたり、または、問題を解いた際の解の精度を向上可能にすることができる。

10

【0108】

出力部 408 は、例えば、第三のベクトルを、実際の問題を解いた結果に対応付けて出力する。出力部 408 は、具体的には、第三のベクトルを、判断した状況に対応付けて出力する。これにより、出力部 408 は、問題を解いた結果を、ユーザなどに参照可能にすることができる。

【0109】

出力部 408 は、例えば、第三のベクトルを出力せずに、実際の問題を解いた結果を出力してもよい。出力部 408 は、具体的には、第三のベクトルを出力せずに、判断した状況

20

【0110】

(出力装置 100 の動作例)

次に、図 5 ~ 図 7 を用いて、出力装置 100 の動作例について説明する。まず、図 5 を用いて、出力装置 100 によって用いられる Co - Attention Network 500 の具体例について説明する。

【0111】

図 5 は、Co - Attention Network 500 の具体例を示す説明図である。以下の説明では、Co - Attention Network 500 を「CAN 500」と表記する場合がある。また、ターゲットアテンションを「TA」と表記する場合がある。また、セルフアテンションを「SA」と表記する場合がある。

30

【0112】

図 5 に示すように、CAN 500 は、画像 TA 層 501 と、画像 SA 層 502 と、文書 TA 層 503 と、文書 SA 層 504 と、結合層 505 と、統合 SA 層 506 とを有する。

【0113】

図 5 において、CAN 500 は、文書に関する特徴量ベクトル L と画像に関する特徴量ベクトル I とが入力されたことに応じて、ベクトル Z<sub>T</sub> を出力する。文書に関する特徴量ベクトル L は、例えば、文書に関する M 個の特徴量ベクトルを並べたものである。M 個の特徴量ベクトルは、例えば、文書に含まれる M 個の単語を示す特徴量ベクトルである。画像に関する特徴量ベクトル I は、例えば、画像に関する N 個の特徴量ベクトルを並べたものである。N 個の特徴量ベクトルは、例えば、画像に写った N 個の物体を示す特徴量ベクトルである。

40

【0114】

具体的には、画像 TA 層 501 は、画像に関する特徴量ベクトル I と、文書に関する特徴量ベクトル L との入力を受け付ける。画像 TA 層 501 は、画像に関する特徴量ベクトル I から得たクエリと、文書に関する特徴量ベクトル L から得たキーおよびバリューとに基づいて、画像に関する特徴量ベクトル I を補正する。画像 TA 層 501 は、補正後の画像に関する特徴量ベクトル I を、画像 SA 層 502 に出力する。画像 TA 層 501 の具体例については、例えば、図 6 を用いて後述する。

50

## 【 0 1 1 5 】

また、画像 S A 層 5 0 2 は、補正後の画像に関する特徴量ベクトル I の入力を受け付ける。画像 S A 層 5 0 2 は、補正後の画像に関する特徴量ベクトル I から得たクエリ、キーおよびバリューに基づいて、補正後の画像に関する特徴量ベクトル I をさらに補正し、新たな特徴量ベクトル Z<sub>I</sub> を生成し、結合層 5 0 5 に出力する。画像 S A 層 5 0 2 の具体例については、例えば、図 6 を用いて後述する。

## 【 0 1 1 6 】

また、文書 T A 層 5 0 3 は、文書に関する特徴量ベクトル L と、画像に関する特徴量ベクトル I との入力を受け付ける。文書 T A 層 5 0 3 は、文書に関する特徴量ベクトル L から得たクエリと、画像に関する特徴量ベクトル I から得たキーおよびバリューとに基づいて、文書に関する特徴量ベクトル L を補正する。文書 T A 層 5 0 3 は、補正後の文書に関する特徴量ベクトル L を、文書 S A 層 5 0 4 に出力する。文書 T A 層 5 0 3 の具体例については、例えば、図 6 を用いて後述する。

10

## 【 0 1 1 7 】

また、文書 S A 層 5 0 4 は、補正後の文書に関する特徴量ベクトル L の入力を受け付ける。文書 S A 層 5 0 4 は、補正後の文書に関する特徴量ベクトル L から得たクエリ、キーおよびバリューに基づいて、補正後の文書に関する特徴量ベクトル L をさらに補正し、新たな特徴量ベクトル Z<sub>L</sub> を生成して出力する。文書 S A 層 5 0 4 の具体例については、例えば、図 6 を用いて後述する。

## 【 0 1 1 8 】

また、結合層 5 0 5 は、集約用ベクトル H と、特徴量ベクトル Z<sub>I</sub> と、特徴量ベクトル Z<sub>L</sub> との入力を受け付ける。結合層 5 0 5 は、集約用ベクトル H と、特徴量ベクトル Z<sub>I</sub> と、特徴量ベクトル Z<sub>L</sub> とを結合し、結合ベクトル C を生成し、統合 S A 層 5 0 6 に出力する。

20

## 【 0 1 1 9 】

また、統合 S A 層 5 0 6 は、結合ベクトル C の入力を受け付ける。統合 S A 層 5 0 6 は、結合ベクトル C から得たクエリ、キーおよびバリューに基づいて、結合ベクトル C を補正し、特徴量ベクトル Z<sub>T</sub> を生成して出力する。特徴量ベクトル Z<sub>T</sub> は、集約ベクトル Z<sub>H</sub> と、文書に関する統合特徴量ベクトル Z<sub>Y1</sub> ~ Z<sub>M</sub> と、画像に関する統合特徴量ベクトル Z<sub>M+1</sub> ~ Z<sub>M+N</sub> とを含む。これにより、出力装置 1 0 0 は、問題を解いた際の解の精度を向上させる観点で有用な集約ベクトル Z<sub>H</sub> を含む特徴量ベクトル Z<sub>T</sub> を生成し、参照可能にすることができる。このため、出力装置 1 0 0 は、問題を解いた際の解の精度を向上可能にすることができる。

30

## 【 0 1 2 0 】

ここでは、説明の簡略化のため、画像 T A 層 5 0 1 と、画像 S A 層 5 0 2 と、文書 T A 層 5 0 3 と、文書 S A 層 5 0 4 とのグループ 5 1 0 が、1 段である場合について説明したが、これに限らない。例えば、画像 T A 層 5 0 1 と、画像 S A 層 5 0 2 と、文書 T A 層 5 0 3 と、文書 S A 層 5 0 4 とのグループ 5 1 0 が、複数段存在する場合があってもよい。これによれば、出力装置 1 0 0 は、問題を解いた際の解の精度のさらなる向上を図ることができる。

40

## 【 0 1 2 1 】

次に、図 6 の説明に移行し、C A N 5 0 0 を形成する画像 S A 層 5 0 2 と文書 S A 層 5 0 4 と統合 S A 層 5 0 6 となどのような S A 層 6 0 0 の具体例について説明する。また、C A N 5 0 0 を形成する画像 T A 層 5 0 1 と文書 T A 層 5 0 3 となどのような T A 層 6 1 0 の具体例について説明する。

## 【 0 1 2 2 】

図 6 は、S A 層 6 0 0 の具体例と、T A 層 6 1 0 の具体例とを示す説明図である。以下の説明では、Multi-Head Attention を「MHA」と表記する場合がある。また、Add & Norm を「A & N」と表記する場合がある。また、Feed Forward を「FF」と表記する場合がある。

50

## 【 0 1 2 3 】

図 6 に示すように、S A 層 6 0 0 は、M H A 層 6 0 1 と、A & N 層 6 0 2 と、F F 層 6 0 3 と、A & N 層 6 0 4 とを有する。M H A 層 6 0 1 は、入力ベクトル X から得たクエリ Q とキー K とバリュウ V とに基づいて、入力ベクトル X を補正する補正ベクトル R を生成し、A & N 層 6 0 2 に出力する。M H A 層 6 0 1 は、具体的には、入力ベクトル X を、H e a d 個のベクトルに分割して処理する。H e a d は、1 以上の自然数である。

## 【 0 1 2 4 】

A & N 層 6 0 2 は、入力ベクトル X と補正ベクトル R とを加算した上で正規化し、正規化後のベクトルを、F F 層 6 0 3 と A & N 層 6 0 4 とに出力する。F F 層 6 0 3 は、正規化後のベクトルを圧縮し、圧縮後のベクトルを、A & N 層 6 0 4 に出力する。A & N 層 6 0 4 は、正規化後のベクトルと、圧縮後のベクトルとを加算した上で正規化し、出力ベクトル Z を生成して出力する。

10

## 【 0 1 2 5 】

また、T A 層 6 1 0 は、M H A 層 6 1 1 と、A & N 層 6 1 2 と、F F 層 6 1 3 と、A & N 層 6 1 4 とを有する。M H A 層 6 1 1 は、入力ベクトル X から得たクエリ Q と、入力ベクトル Y から得たキー K とバリュウ V とに基づいて、入力ベクトル X を補正する補正ベクトル R を生成し、A & N 層 6 1 2 に出力する。A & N 層 6 1 2 は、入力ベクトル X と補正ベクトル R とを加算した上で正規化し、正規化後のベクトルを、F F 層 6 1 3 と A & N 層 6 1 4 とに出力する。F F 層 6 1 3 は、正規化後のベクトルを圧縮し、圧縮後のベクトルを、A & N 層 6 1 4 に出力する。A & N 層 6 1 4 は、正規化後のベクトルと、圧縮後のベクトルとを加算した上で正規化し、出力ベクトル Z を生成して出力する。

20

## 【 0 1 2 6 】

上述した M H A 層 6 0 1 や M H A 層 6 1 1 は、より具体的には、H e a d 個の A t t e n t i o n 層 6 2 0 により形成される。A t t e n t i o n 層 6 2 0 は、M a t M u l 層 6 2 1 と、S c a l e 層 6 2 2 と、M a s k 層 6 2 3 と、S o f t M a x 層 6 2 4 と、M a t M u l 層 6 2 5 とを有する。

## 【 0 1 2 7 】

M a t M u l 層 6 2 1 は、クエリ Q とキー K との内積を算出し、S c o r e に設定する。S c a l e 層 6 2 2 は、S c o r e 全体を定数 a で除算し、更新する。M a s k 層 6 2 3 は、更新後の S c o r e をマスク処理してもよい。S o f t M a x 層 6 2 4 は、更新後の S c o r e を、正規化し、A t t に設定する。M a t M u l 層 6 2 5 は、A t t とバリュウ V との内積を算出し、補正ベクトル R に設定する。

30

## 【 0 1 2 8 】

ここで、S A 層 6 0 0 の計算例について説明する。具体的には、S A 層 6 0 0 の計算例の一つとして、S A 層 6 0 0 で画像 S A 層 5 0 2 を実現する場合における計算例を示す。また、説明の簡略化のため、H e a d = 1 であるとする。

## 【 0 1 2 9 】

ここで、入力ベクトル X は、下記式 ( 1 ) により表現される画像に関する特徴量ベクトル X であるとする。x<sub>1</sub>, x<sub>2</sub>, x<sub>3</sub> は、d 次元のベクトルである。x<sub>1</sub>, x<sub>2</sub>, x<sub>3</sub> は、それぞれ、画像に写った物体に対応する。

40

## 【 0 1 3 0 】

## 【 数 1 】

$$X: \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \dots (1)$$

## 【 0 1 3 1 】

クエリ Q は、下記式 ( 2 ) により算出される。W<sub>Q</sub> は、変換行列であり、学習により設定される。キー K は、下記式 ( 3 ) により算出される。W<sub>K</sub> は、変換行列であり、学習によ

50

り設定される。バリューVは、下記式(4)により算出される。W<sub>V</sub>は、変換行列であり、学習により設定される。クエリQと、キーKと、バリューVとは、入力ベクトルXと同じ次元である。

【0132】

【数2】

$$Q = W_Q X = \begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix} \dots (2)$$

10

【0133】

【数3】

$$K = W_K X = \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} \dots (3)$$

【0134】

【数4】

$$V = W_V X = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \dots (4)$$

20

【0135】

MatMul層621は、下記式(5)に示すように、クエリQとキーKとの内積を算出し、Scoreに設定する。Scale層622は、下記式(6)に示すように、Score全体を定数aで除算し、更新する。ここでは、Mask層623は、マスク処理を省略する。SoftMax層624は、下記式(7)に示すように、更新後のScoreを、正規化し、Attに設定する。MatMul層625は、下記式(8)に示すように、AttとバリューVとの内積を算出し、補正ベクトルRに設定する。

30

【0136】

【数5】

$$Score = Q \cdot K^T = \begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix} \begin{bmatrix} k_1^T & k_2^T & k_3^T \end{bmatrix} \dots (5)$$

【0137】

【数6】

$$Score = \frac{Score}{a} \dots (6)$$

40

【0138】

【数7】

$$Att = Softmax(Score) \dots (7)$$

【0139】

50

【数 8】

$$R = Att \cdot V = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} \dots (8)$$

【0140】

MHA層601は、上述したように、補正ベクトルRを生成する。A&N層602は、下記式(9)および下記式(10)に示すように、入力ベクトルXと補正ベクトルRとを加算した上で正規化し、入力ベクトルXを更新する。 $\mu$ は、下記式(11)により定義される。 $\sigma$ は、下記式(12)により定義される。FF層603は、下記式(13)に示すように、更新後の入力ベクトルXを変換し、変換ベクトルX'を設定する。fは、活性化関数である。A&N層604は、更新後の入力ベクトルXと、設定した変換ベクトルX'とを加算した上で正規化し、出力ベクトルZを生成する。

10

【0141】

【数 9】

$$X = X + R \dots (9)$$

【0142】

20

【数 10】

$$X = \frac{X - \mu}{\sigma} \dots (10)$$

【0143】

【数 11】

$$\mu = \frac{1}{d} \sum_{i=1}^d x_i \dots (11)$$

30

【0144】

【数 12】

$$\sigma = \frac{1}{d} \sum_{i=1}^d (x_i - \mu)^2 \dots (12)$$

【0145】

【数 13】

$$X' = W_1 f(W_2 X) \dots (13)$$

40

【0146】

次に、TA層610の計算例について説明する。具体的には、TA層610の計算例の一つとして、TA層610で画像TA層501を実現する場合における計算例を示す。また、説明の簡略化のため、Head = 1であるとする。

【0147】

ここで、入力ベクトルXは、上記式(1)により表現される画像に関する特徴量ベクトルXであるとする。 $x_1, x_2, x_3$ は、d次元のベクトルである。 $x_1, x_2, x_3$ は、それぞれ、画像に写った物体に対応する。入力ベクトルYは、下記式(14)により表現さ

50

れる文書に関する特徴量ベクトル $Y$ であるとする。 $y_1, y_2, y_3$ は、 $d$ 次元のベクトルである。 $y_1, y_2, y_3$ は、それぞれ、文書に含まれる単語に対応する。

【 0 1 4 8 】

【数 1 4】

$$Y: \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \dots (14)$$

【 0 1 4 9 】

10

クエリ $Q$ は、下記式(15)により算出される。 $W_Q$ は、変換行列であり、学習により設定される。キー $K$ は、下記式(16)により算出される。 $W_K$ は、変換行列であり、学習により設定される。バリュー $V$ は、下記式(17)により算出される。 $W_V$ は、変換行列であり、学習により設定される。クエリ $Q$ は、入力ベクトル $X$ と同じ次元である。キー $K$ と、バリュー $V$ とは、入力ベクトル $Y$ と同じ次元である。

【 0 1 5 0 】

【数 1 5】

$$Q = W_Q X = \begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix} \dots (15)$$

20

【 0 1 5 1 】

【数 1 6】

$$K = W_K Y = \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} \dots (16)$$

【 0 1 5 2 】

30

【数 1 7】

$$V = W_V Y = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \dots (17)$$

【 0 1 5 3 】

MatMul層621は、上記式(5)に示すように、クエリ $Q$ とキー $K$ との内積を算出し、Scoreに設定する。Scale層622は、上記式(6)に示すように、Score全体を定数 $a$ で除算し、更新する。ここでは、Mask層623は、マスク処理を省略する。SoftMax層624は、上記式(7)に示すように、更新後のScoreを、正規化し、Attに設定する。MatMul層625は、上記式(8)に示すように、Attとバリュー $V$ との内積を算出し、補正ベクトル $R$ に設定する。

40

【 0 1 5 4 】

MHA層601は、上述したように、補正ベクトル $R$ を生成する。A&N層602は、上記式(9)および上記式(10)に示すように、入力ベクトル $X$ と補正ベクトル $R$ とを加算した上で正規化し、入力ベクトル $X$ を更新する。FF層603は、上記式(13)に示すように、更新後の入力ベクトル $X$ を変換し、変換ベクトル $X'$ を設定する。A&N層604は、更新後の入力ベクトル $X$ と、設定した変換ベクトル $X'$ とを加算した上で正規化し、出力ベクトル $Z$ を生成する。次に、図7を用いて、出力装置100による、CAN5

50

00を用いた動作の一例について説明する。

【0155】

図7は、CAN500を用いた動作の一例を示す説明図である。出力装置100は、文書700を取得し、画像710を取得する。出力装置100は、文書700をトークン化し、トークン集合701をベクトル化し、文書700に関する特徴量ベクトル702を生成し、CAN500に入力する。また、出力装置100は、画像710から物体を検出し、物体ごとの部分画像の集合711をベクトル化し、画像710に関する特徴量ベクトル712を生成し、CAN500に入力する。

【0156】

出力装置100は、CAN500から、特徴量ベクトル $Z_T$ を取得し、特徴量ベクトル $Z_T$ に含まれる集約ベクトル $Z_H$ を、危険度推定器720に入力する。出力装置100は、危険度推定器720から推定結果 $N_o$ を取得する。これにより、出力装置100は、画像と文書との特徴が反映された集約ベクトル $Z_H$ を用いて、危険度推定器720に推定させることができ、精度よく推定可能にすることができる。危険度推定器720は、例えば、銃を持った人物が写っている画像710があるが、ミュージアムの展示物であることを示す文書もあるため、推定結果 $N_o$  = 危険ではないと推定することができる。

10

【0157】

(出力装置100の利用例)

次に、図8～図11を用いて、出力装置100の利用例について説明する。

【0158】

図8および図9は、出力装置100の利用例1を示す説明図である。図8において、出力装置100は、学習フェーズを実施し、CAN500を学習する。出力装置100は、例えば、何らかのシーンを写した画像800と、画像800に対応する字幕となる文書810とを取得する。画像800は、例えば、りんごを切るシーンを写す。出力装置100は、画像800を変換器820により特徴量ベクトルに変換し、CAN500に入力する。また、出力装置100は、文書810の単語appleをマスクした上で、変換器830により特徴量ベクトルに変換し、CAN500に入力する。

20

【0159】

出力装置100は、CAN500により生成された特徴量ベクトルを、識別器840に入力し、マスクされた単語を予測した結果を取得し、マスクされた単語の正解「apple」との誤差を算出する。出力装置100は、算出した誤差に基づいて、誤差逆伝搬によりCAN500を学習する。さらに、出力装置100は、誤差逆伝搬により、変換器820、830や識別器840を学習してもよい。これにより、出力装置100は、画像800と字幕となる文書810の文脈とを考慮して単語を推定する観点で有用なように、CAN500、および変換器820、830や識別器840を更新することができる。次に、図9の説明に移行する。

30

【0160】

図9において、出力装置100は、試験フェーズを実施し、学習した変換器820、830と、学習したCAN500とを用いて、回答を生成して出力する。出力装置100は、例えば、何らかのシーンを写した画像900と、画像900に対応する質問文となる文書910とを取得する。画像900は、例えば、りんごを切るシーンを写す。

40

【0161】

出力装置100は、画像900を変換器820により特徴量ベクトルに変換し、CAN500に入力する。また、出力装置100は、文書910を変換器830により特徴量ベクトルに変換し、CAN500に入力する。出力装置100は、CAN500により生成された特徴量ベクトルを、回答生成器920に入力し、回答となる単語を取得して出力する。これにより、出力装置100は、画像900と質問文となる文書910の文脈とを考慮して、精度よく回答となる単語を推定することができる。

【0162】

図10および図11は、出力装置100の利用例2を示す説明図である。図10において

50

、出力装置 100 は、学習フェーズを実施し、CAN500 を学習する。出力装置 100 は、例えば、何らかのシーンを写した画像 1000 と、画像 1000 に対応する字幕となる文書 1010 とを取得する。画像 1000 は、例えば、りんごを切るシーンを写す。出力装置 100 は、画像 1000 を変換器 1020 により特徴量ベクトルに変換し、CAN500 に入力する。また、出力装置 100 は、文書 1010 の単語 apple をマスクした上で、変換器 1030 により特徴量ベクトルに変換し、CAN500 に入力する。

【0163】

出力装置 100 は、CAN500 により生成された特徴量ベクトルを、識別器 1040 に入力し、画像に写ったシーンの危険度を予測した結果を取得し、危険度の正解との誤差を算出する。出力装置 100 は、算出した誤差に基づいて、誤差逆伝搬により CAN500 を学習する。また、出力装置 100 は、誤差逆伝搬により、変換器 1020、1030 や識別器 1040 を学習する。これにより、出力装置 100 は、画像 1000 と字幕となる文書 1010 の文脈とを考慮して危険度を予測する観点で有用なように、CAN500、および変換器 1020、1030 や識別器 1040 を更新することができる。次に、図 11 の説明に移行する。

10

【0164】

図 11 において、出力装置 100 は、試験フェーズを実施し、学習した変換器 1020、1030 や識別器 1040 と、学習した CAN500 とを用いて、危険度を予測して出力する。出力装置 100 は、例えば、何らかのシーンを写した画像 1100 と、画像に対応する説明文となる文書 1110 とを取得する。画像 1100 は、例えば、ももを切るシーンを写す。

20

【0165】

出力装置 100 は、画像 1100 を変換器 1020 により特徴量ベクトルに変換し、CAN500 に入力する。また、出力装置 100 は、文書 1110 を変換器 1030 により特徴量ベクトルに変換し、CAN500 に入力する。出力装置 100 は、CAN500 により生成された特徴量ベクトルを、識別器 1040 に入力し、危険度を取得して出力する。これにより、出力装置 100 は、画像 1100 と説明文となる文書 1110 の文脈とを考慮して、精度よく危険度を予測することができる。

【0166】

(学習処理手順)

次に、図 12 を用いて、出力装置 100 が実行する、学習処理手順の一例について説明する。学習処理は、例えば、図 3 に示した CPU301 と、メモリ 302 や記録媒体 305 などの記憶領域と、ネットワーク I/F 303 とによって実現される。

30

【0167】

図 12 は、学習処理手順の一例を示すフローチャートである。図 12 において、出力装置 100 は、画像の特徴量ベクトルと、文書の特徴量ベクトルとを取得する(ステップ S1201)。

【0168】

次に、出力装置 100 は、取得した画像の特徴量ベクトルから生成したクエリと、取得した文書の特徴量ベクトルから生成したキーおよびバリューとに基づいて、画像 TA 層 501 を用いて、画像の特徴量ベクトルを補正する(ステップ S1202)。

40

【0169】

そして、出力装置 100 は、補正後の画像の特徴量ベクトルに基づいて、画像 SA 層 502 を用いて、補正後の画像の特徴量ベクトルをさらに補正し、新たに画像の特徴量ベクトルを生成する(ステップ S1203)。

【0170】

次に、出力装置 100 は、取得した文書の特徴量ベクトルから生成したクエリと、取得した画像の特徴量ベクトルから生成したキーおよびバリューとに基づいて、文書 TA 層 503 を用いて、文書の特徴量ベクトルを補正する(ステップ S1204)。

【0171】

50

そして、出力装置 100 は、補正後の文書の特徴量ベクトルに基づいて、文書 SA 層 504 を用いて、補正後の文書の特徴量ベクトルをさらに補正し、新たに文書の特徴量ベクトルを生成する（ステップ S1205）。

【0172】

次に、出力装置 100 は、集約用ベクトルを初期化する（ステップ S1206）。そして、出力装置 100 は、集約用ベクトルと、生成した画像の特徴量ベクトルと、生成した文書の特徴量ベクトルとを結合し、結合ベクトルを生成する（ステップ S1207）。

【0173】

次に、出力装置 100 は、結合ベクトルに基づいて、統合 SA 層 506 を用いて、結合ベクトルを補正し、集約ベクトルを生成する（ステップ S1208）。そして、出力装置 100 は、集約ベクトルに基づいて、CAN500 を学習する（ステップ S1209）。

その後、出力装置 100 は、学習処理を終了する。これにより、出力装置 100 は、CAN500 を用いて問題を解くにあたり、問題を解いた際の解の精度が向上するように、CAN500 のパラメータを更新することができる。

【0175】

ここで、出力装置 100 は、図 12 の一部ステップの処理の順序を入れ替えて実行してもよい。例えば、ステップ S1202, S1203 の処理と、ステップ S1204, S1205 の処理との順序は入れ替え可能である。また、出力装置 100 は、ステップ S1202 ~ S1205 の処理を繰り返し実行してもよい。

【0176】

（推定処理手順）

次に、図 13 を用いて、出力装置 100 が実行する、推定処理手順の一例について説明する。推定処理は、例えば、図 3 に示した CPU301 と、メモリ 302 や記録媒体 305 などの記憶領域と、ネットワーク I/F 303 とによって実現される。

【0177】

図 13 は、推定処理手順の一例を示すフローチャートである。図 13 において、出力装置 100 は、画像の特徴量ベクトルと、文書の特徴量ベクトルとを取得する（ステップ S1301）。

【0178】

次に、出力装置 100 は、取得した画像の特徴量ベクトルから生成したクエリと、取得した文書の特徴量ベクトルから生成したキーおよびバリューとに基づいて、画像 TA 層 501 を用いて、画像の特徴量ベクトルを補正する（ステップ S1302）。

【0179】

そして、出力装置 100 は、補正後の画像の特徴量ベクトルに基づいて、画像 SA 層 502 を用いて、補正後の画像の特徴量ベクトルをさらに補正し、新たに画像の特徴量ベクトルを生成する（ステップ S1303）。

【0180】

次に、出力装置 100 は、取得した文書の特徴量ベクトルから生成したクエリと、取得した画像の特徴量ベクトルから生成したキーおよびバリューとに基づいて、文書 TA 層 503 を用いて、文書の特徴量ベクトルを補正する（ステップ S1304）。

【0181】

そして、出力装置 100 は、補正後の文書の特徴量ベクトルに基づいて、文書 SA 層 504 を用いて、補正後の文書の特徴量ベクトルをさらに補正し、新たに文書の特徴量ベクトルを生成する（ステップ S1305）。

【0182】

次に、出力装置 100 は、集約用ベクトルを初期化する（ステップ S1306）。そして、出力装置 100 は、集約用ベクトルと、生成した画像の特徴量ベクトルと、生成した文書の特徴量ベクトルとを結合し、結合ベクトルを生成する（ステップ S1307）。

【0183】

10

20

30

40

50

次に、出力装置 100 は、結合ベクトルに基づいて、統合 SA 層 506 を用いて、結合ベクトルを補正し、集約ベクトルを生成する（ステップ S1308）。そして、出力装置 100 は、集約ベクトルに基づいて、識別モデルを用いて、状況を推定する（ステップ S1309）。

【0184】

次に、出力装置 100 は、推定した状況を出力する（ステップ S1310）。そして、出力装置 100 は、推定処理を終了する。これにより、出力装置 100 は、CAN500 を用いて、問題を解いた際の解の精度を向上させることができる。

【0185】

ここで、出力装置 100 は、図 13 の一部ステップの処理の順序を入れ替えて実行してもよい。例えば、ステップ S1302, S1303 の処理と、ステップ S1304, S1305 の処理との順序は入れ替え可能である。また、出力装置 100 は、ステップ S1302 ~ S1305 の処理を繰り返し実行してもよい。

【0186】

以上説明したように、出力装置 100 によれば、第一のモーダルの情報に基づくベクトルと、第二のモーダルの情報に基づくベクトルとの相関に基づいて、第一のモーダルの情報に基づくベクトルを補正することができる。出力装置 100 によれば、第一のモーダルの情報に基づくベクトルと、第二のモーダルの情報に基づくベクトルとの相関に基づいて、第二のモーダルの情報に基づくベクトルを補正することができる。出力装置 100 によれば、補正後の第一のモーダルの情報に基づくベクトルから得た異なる種類の 2 つのベクトルの相関に基づいて、第一のベクトルを生成することができる。出力装置 100 によれば、補正後の第二のモーダルの情報に基づくベクトルから得た異なる種類の 2 つのベクトルの相関に基づいて、第二のベクトルを生成することができる。出力装置 100 によれば、所定のベクトルと、生成した第一のベクトルと、生成した第二のベクトルとを含む結合ベクトルから得た異なる種類の 2 つのベクトルの相関に基づいて、第一のベクトルと第二のベクトルとを集約した第三のベクトルを生成することができる。出力装置 100 によれば、生成した第三のベクトルを出力することができる。これにより、出力装置 100 は、第一のベクトルと第二のベクトルとが集約され、第一のモーダルの情報に基づくベクトルと第二のモーダルの情報に基づくベクトルとのうち問題の解決に有用な情報が反映される傾向がある第三のベクトルを生成し、利用可能にすることができる。このため、出力装置 100 は、第三のベクトルを利用し、問題を解いた際の解の精度を向上可能にすることができる。

【0187】

出力装置 100 によれば、第一のターゲットアテンション層を用いて、第一のモーダルの情報に基づくベクトルから得たベクトルと、第二のモーダルの情報に基づくベクトルから得たベクトルとの内積に基づいて、第一のモーダルの情報に基づくベクトルを補正することができる。出力装置 100 によれば、第二のターゲットアテンション層を用いて、第一のモーダルの情報に基づくベクトルから得たベクトルと、第二のモーダルの情報に基づくベクトルから得たベクトルとの内積に基づいて、第二のモーダルの情報に基づくベクトルを補正することができる。出力装置 100 によれば、第一のセルフアテンション層を用いて、補正後の第一のモーダルの情報に基づくベクトルから得た異なる種類の 2 つのベクトルの内積に基づいて、補正後の第一のモーダルの情報に基づくベクトルをさらに補正し、第一のベクトルを生成することができる。出力装置 100 によれば、第二のセルフアテンション層を用いて、補正後の第二のモーダルの情報に基づくベクトルから得た異なる種類の 2 つのベクトルの内積に基づいて、補正後の第二のモーダルの情報に基づくベクトルをさらに補正し、第二のベクトルを生成することができる。出力装置 100 によれば、第三のセルフアテンション層を用いて、所定のベクトルと第一のベクトルと第二のベクトルとを結合した結合ベクトルから得た異なる種類の 2 つのベクトルの内積に基づいて、第三のベクトルを生成することができる。これにより、出力装置 100 は、各種アテンション層を用いて、第三のベクトルを生成する処理を、容易に実現することができる。

## 【 0 1 8 8 】

出力装置 1 0 0 によれば、生成した第三のベクトルに基づいて、第一のモーダルと第二のモーダルとに関する状況を判断して出力することができる。これにより、出力装置 1 0 0 は、状況を判別する問題を解く場合に適用可能にすることができ、問題を解いた結果を参照可能にすることができる。

## 【 0 1 8 9 】

出力装置 1 0 0 によれば、生成した第一のベクトルを、新たな第一のモーダルの情報に基づくベクトルに設定することができる。出力装置 1 0 0 によれば、生成した第二のベクトルを、新たな第二のモーダルの情報に基づくベクトルに設定することができる。出力装置 1 0 0 によれば、設定した第一のモーダルの情報に基づくベクトルを補正し、設定した第二のモーダルの情報に基づくベクトルを補正し、第一のベクトルを生成し、第二のベクトルを生成する、処理を 1 回以上繰り返すことができる。出力装置 1 0 0 によれば、所定のベクトルと、生成した第一のベクトルと、生成した第二のベクトルとを含む結合ベクトルから得た異なる種類の 2 つのベクトルの相関に基づいて、第一のベクトルと第二のベクトルとを集約した第三のベクトルを生成することができる。これにより、出力装置 1 0 0 は、各種ベクトルを多段階に補正し、問題を解いた際の解の精度をさらに向上可能にすることができる。

10

## 【 0 1 9 0 】

出力装置 1 0 0 によれば、第一のモーダルとして、画像に関するモーダルを採用することができる。出力装置 1 0 0 によれば、第二のモーダルとして、文書に関するモーダルを採用することができる。これにより、出力装置 1 0 0 は、画像と文書とに基づいて問題を解く場合に適用可能にすることができる。

20

## 【 0 1 9 1 】

出力装置 1 0 0 によれば、第一のモーダルとして、画像に関するモーダルを採用することができる。出力装置 1 0 0 によれば、第二のモーダルとして、音声に関するモーダルを採用することができる。これにより、出力装置 1 0 0 は、画像と音声とに基づいて問題を解く場合に適用可能にすることができる。

## 【 0 1 9 2 】

出力装置 1 0 0 によれば、第一のモーダルとして、第一の言語の文書に関するモーダルを採用することができる。出力装置 1 0 0 によれば、第二のモーダルとして、第二の言語の文書に関するモーダルを採用することができる。これにより、出力装置 1 0 0 は、異なる言語の 2 つの文書に基づいて問題を解く場合に適用可能にすることができる。

30

## 【 0 1 9 3 】

出力装置 1 0 0 によれば、生成した第三のベクトルに基づいて、ポジティブな状況、または、ネガティブな状況を判断して出力することができる。これにより、出力装置 1 0 0 は、ポジティブな状況、または、ネガティブな状況を判別する問題を解く場合に適用可能にすることができ、問題を解いた結果を参照可能にすることができる。

## 【 0 1 9 4 】

出力装置 1 0 0 によれば、生成した第三のベクトルに基づいて、第一のターゲットアテンション層と、第二のターゲットアテンション層と、第一のセルフアテンション層と、第二のセルフアテンション層と、第三のセルフアテンション層とを更新することができる。これにより、出力装置 1 0 0 は、第三のベクトルをより有用な状態で生成可能に、各種アテンション層を更新し、問題を解いた際の解の精度を向上可能にすることができる。

40

## 【 0 1 9 5 】

なお、本実施の形態で説明した出力方法は、予め用意されたプログラムを P C やワークステーションなどのコンピュータで実行することにより実現することができる。本実施の形態で説明した出力プログラムは、コンピュータで読み取り可能な記録媒体に記録され、コンピュータによって記録媒体から読み出されることによって実行される。記録媒体は、ハードディスク、フレキシブルディスク、C D ( C o m p a c t D i s c ) - R O M、M O、D V D ( D i g i t a l V e r s a t i l e D i s c ) などである。また、本実

50

施の形態で説明した出力プログラムは、インターネットなどのネットワークを介して配布してもよい。

【0196】

上述した実施の形態に関し、さらに以下の付記を開示する。

【0197】

(付記1) 第一のモーダルの情報に基づくベクトルと、前記第一のモーダルとは異なる第二のモーダルの情報に基づくベクトルとの相関に基づいて、前記第一のモーダルの情報に基づくベクトルを補正し、

前記第一のモーダルの情報に基づくベクトルと、前記第二のモーダルの情報に基づくベクトルとの相関に基づいて、前記第二のモーダルの情報に基づくベクトルを補正し、

補正後の前記第一のモーダルの情報に基づくベクトルから得た異なる種類の2つのベクトルの相関に基づいて、第一のベクトルを生成し、

補正後の前記第二のモーダルの情報に基づくベクトルから得た前記異なる種類の2つのベクトルの相関に基づいて、第二のベクトルを生成し、

所定のベクトルと、生成した前記第一のベクトルと、生成した前記第二のベクトルとを含む結合ベクトルから得た前記異なる種類の2つのベクトルの相関に基づいて、前記第一のベクトルと前記第二のベクトルとを集約した第三のベクトルを生成し、

生成した前記第三のベクトルを出力する、

処理をコンピュータが実行することを特徴とする出力方法。

【0198】

(付記2) 前記第一のモーダルの情報に基づくベクトルを補正する処理は、

前記第一のモーダルに関する第一のターゲットアテンション層を用いて、前記第一のモーダルの情報に基づくベクトルから得たベクトルと、前記第二のモーダルの情報に基づくベクトルから得たベクトルとの内積に基づいて、前記第一のモーダルの情報に基づくベクトルを補正し、

前記第二のモーダルの情報に基づくベクトルを補正する処理は、

前記第二のモーダルに関する第二のターゲットアテンション層を用いて、前記第一のモーダルの情報に基づくベクトルから得たベクトルと、前記第二のモーダルの情報に基づくベクトルから得たベクトルとの内積に基づいて、前記第二のモーダルの情報に基づくベクトルを補正し、

前記第一のベクトルを生成する処理は、

前記第一のモーダルに関する第一のセルフアテンション層を用いて、補正後の前記第一のモーダルの情報に基づくベクトルから得た前記異なる種類の2つのベクトルの内積に基づいて、補正後の前記第一のモーダルの情報に基づくベクトルをさらに補正し、前記第一のベクトルを生成し、

前記第二のベクトルを生成する処理は、

前記第二のモーダルに関する第二のセルフアテンション層を用いて、補正後の前記第二のモーダルの情報に基づくベクトルから得た前記異なる種類の2つのベクトルの内積に基づいて、補正後の前記第二のモーダルの情報に基づくベクトルをさらに補正し、前記第二のベクトルを生成し、

前記第三のベクトルを生成する処理は、

第三のセルフアテンション層を用いて、前記所定のベクトルと前記第一のベクトルと前記第二のベクトルとを結合した結合ベクトルから得た前記異なる種類の2つのベクトルの内積に基づいて、前記結合ベクトルを補正し、前記第三のベクトルを生成する、ことを特徴とする付記1に記載の出力方法。

【0199】

(付記3) 生成した前記第三のベクトルに基づいて、前記第一のモーダルと前記第二のモーダルとに関する状況を判断して出力する、

処理を前記コンピュータが実行することを特徴とする付記1または2に記載の出力方法。

【0200】

10

20

30

40

50

(付記4) 生成した前記第一のベクトルを、新たな前記第一のモーダルの情報に基づくベクトルに設定し、  
 生成した前記第二のベクトルを、新たな前記第二のモーダルの情報に基づくベクトルに設定し、  
 設定した前記第一のモーダルの情報に基づくベクトルと、設定した前記第二のモーダルの情報に基づくベクトルとの相関に基づいて、設定した前記第一のモーダルの情報に基づくベクトルを補正し、  
 設定した前記第一のモーダルの情報に基づくベクトルと、設定した前記第二のモーダルの情報に基づくベクトルとの相関に基づいて、設定した前記第二のモーダルの情報に基づくベクトルを補正し、  
 補正後の前記第一のモーダルの情報に基づくベクトルから得た前記異なる種類の2つのベクトルの相関に基づいて、前記第一のベクトルを生成し、  
 補正後の前記第二のモーダルの情報に基づくベクトルから得た前記異なる種類の2つのベクトルの相関に基づいて、前記第二のベクトルを生成する、  
 処理を前記コンピュータが1回以上繰り返し、  
 前記第三のベクトルを生成する処理は、  
 前記所定のベクトルと、生成した前記第一のベクトルと、生成した前記第二のベクトルとを含む結合ベクトルから得た前記異なる種類の2つのベクトルの相関に基づいて、前記第一のベクトルと前記第二のベクトルとを集約した第三のベクトルを生成する、ことを特徴とする付記1~3のいずれか一つに記載の出力方法。

10

20

## 【0201】

(付記5) 前記第一のモーダルと前記第二のモーダルとの組は、画像に関するモーダルと文書に関するモーダルとの組、画像に関するモーダルと音声に関するモーダルとの組、第一の言語の文書に関するモーダルと第二の言語の文書に関するモーダルとの組のうちいずれかの組である、ことを特徴とする付記1~4のいずれか一つに記載の出力方法。

## 【0202】

(付記6) 前記状況は、ポジティブな状況、または、ネガティブな状況である、ことを特徴とする付記3に記載の出力方法。

## 【0203】

(付記7) 生成した前記第三のベクトルに基づいて、前記第一のターゲットアテンション層と、前記第二のターゲットアテンション層と、前記第一のセルフアテンション層と、前記第二のセルフアテンション層と、前記第三のセルフアテンション層とを更新する、処理を前記コンピュータが実行することを特徴とする付記2に記載の出力方法。

30

## 【0204】

(付記8) 第一のモーダルの情報に基づくベクトルと、前記第一のモーダルとは異なる第二のモーダルの情報に基づくベクトルとの相関に基づいて、前記第一のモーダルの情報に基づくベクトルを補正し、  
 前記第一のモーダルの情報に基づくベクトルと、前記第二のモーダルの情報に基づくベクトルとの相関に基づいて、前記第二のモーダルの情報に基づくベクトルを補正し、  
 補正後の前記第一のモーダルの情報に基づくベクトルから得た異なる種類の2つのベクトルの相関に基づいて、第一のベクトルを生成し、  
 補正後の前記第二のモーダルの情報に基づくベクトルから得た前記異なる種類の2つのベクトルの相関に基づいて、第二のベクトルを生成し、  
 所定のベクトルと、生成した前記第一のベクトルと、生成した前記第二のベクトルとを含む結合ベクトルから得た前記異なる種類の2つのベクトルの相関に基づいて、前記第一のベクトルと前記第二のベクトルとを集約した第三のベクトルを生成し、  
 生成した前記第三のベクトルを出力する、  
 処理をコンピュータに実行させることを特徴とする出力プログラム。

40

## 【0205】

(付記9) 第一のモーダルの情報に基づくベクトルと、前記第一のモーダルとは異なる第

50

二のモーダルの情報に基づくベクトルとの相関に基づいて、前記第一のモーダルの情報に基づくベクトルを補正し、  
 前記第一のモーダルの情報に基づくベクトルと、前記第二のモーダルの情報に基づくベクトルとの相関に基づいて、前記第二のモーダルの情報に基づくベクトルを補正し、  
 補正後の前記第一のモーダルの情報に基づくベクトルから得た異なる種類の2つのベクトルの相関に基づいて、第一のベクトルを生成し、  
 補正後の前記第二のモーダルの情報に基づくベクトルから得た前記異なる種類の2つのベクトルの相関に基づいて、第二のベクトルを生成し、  
 所定のベクトルと、生成した前記第一のベクトルと、生成した前記第二のベクトルとを含む結合ベクトルから得た前記異なる種類の2つのベクトルの相関に基づいて、前記第一のベクトルと前記第二のベクトルとを集約した第三のベクトルを生成し、  
 生成した前記第三のベクトルを出力する、  
 制御部を有することを特徴とする出力装置。

10

## 【符号の説明】

## 【0206】

- 100 出力装置
- 111, 112 補正モデル
- 121, 122, 130 生成モデル
- 200 情報処理システム
- 201 クライアント装置
- 202 端末装置
- 210 ネットワーク
- 300 バス
- 301 CPU
- 302 メモリ
- 303 ネットワークI/F
- 304 記録媒体I/F
- 305 記録媒体
- 400 記憶部
- 401 取得部
- 402 第一の補正部
- 403 第一の生成部
- 404 第二の補正部
- 405 第二の生成部
- 406 第三の生成部
- 407 解析部
- 408 出力部
- 500 Co-Attention Network
- 501 画像TA層
- 502 画像SA層
- 503 文書TA層
- 504 文書SA層
- 505 結合層
- 506 統合SA層
- 510 グループ
- 600 SA層
- 601, 611 MHA層
- 602, 604, 612, 614 A&N層
- 603, 613 FF層
- 610 TA層

20

30

40

50

6 2 0 A t t e n t i o n 層  
6 2 1 , 6 2 5 M a t M u l 層  
6 2 2 S c a l e 層  
6 2 3 M a s k 層  
6 2 4 S o f t M a x 層  
7 0 0 , 8 1 0 , 9 1 0 , 1 0 1 0 , 1 1 1 0 文書  
7 0 1 トークン集合  
7 0 2 , 7 1 2 特徴量ベクトル  
7 1 0 , 8 0 0 , 9 0 0 , 1 0 0 0 , 1 1 0 0 画像  
7 1 1 集合  
7 2 0 危険度推定器  
8 2 0 , 8 3 0 , 1 0 2 0 , 1 0 3 0 変換器  
8 4 0 , 1 0 4 0 識別器  
9 2 0 回答生成器

10

20

30

40

50