

【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】WO2021/095212
 【出願番号】特願2021-555729(P2021-555729)

【国際特許分類】
 G 0 6 N 2 0 / 0 0 (2 0 1 9 . 0 1)
 G 0 6 T 7 / 0 0 (2 0 1 7 . 0 1)

【 F I 】

G 0 6 N 2 0 / 0 0
 G 0 6 T 7 / 0 0 3 5 0 B

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】

20

【特許請求の範囲】

【請求項1】

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、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補正ベクトルを生成し、
 生成した前記補正ベクトルを、前記第一のモーダルの情報に基づくベクトルに結合し、
 所定のルールに従って、結合後の前記第一のモーダルの情報に基づくベクトルを圧縮し、
 圧縮後の前記第一のモーダルの情報に基づくベクトルに対して、正規化処理を実施し、
 前記正規化処理により得た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。

30

【請求項2】

前記生成する処理は、
 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、前記補正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。

【請求項3】

前記正規化処理を実施する処理は、
 前記第一のモーダルの情報に基づくベクトルと、前記補正ベクトルとの和を正規化し、
 当該正規化により得たベクトルと、圧縮後の前記第一のモーダルの情報に基づくベクトルとの和を正規化する、ことを特徴とする請求項1または2に記載の出力方法。

40

【請求項4】

前記正規化処理を実施する処理は、
 結合後の前記第一のモーダルの情報に基づくベクトルと、圧縮後の前記第一のモーダルの情報に基づくベクトルとの和を正規化する、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。

50

【請求項 6】

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、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補正ベクトルを生成し、

生成した前記補正ベクトルを、前記第一のモーダルの情報に基づくベクトルに結合し、所定のルールに従って、結合後の前記第一のモーダルの情報に基づくベクトルを圧縮し、圧縮後の前記第一のモーダルの情報に基づくベクトルに対して、正規化処理を実施し、前記正規化処理により得た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、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する補正ベクトルを生成し、

生成した前記補正ベクトルを、前記第一のモーダルの情報に基づくベクトルに結合し、所定のルールに従って、結合後の前記第一のモーダルの情報に基づくベクトルを圧縮し、圧縮後の前記第一のモーダルの情報に基づくベクトルに対して、正規化処理を実施し、前記正規化処理により得た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。

【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、Attentionにより情報を変換するTransformerと呼ばれるものがある。Attentionは、具体的には、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から得たバリューの重み付け和を算出し、第一のモーダルの情報に基づくベクトルに加算する。

【先行技術文献】

【非特許文献】

【0004】

【非特許文献 1】Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems. 2017.

【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

10

20

30

40

50

が悪い場合がある。例えば、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、Attentionにより、画像に関するモーダルの情報に基づくベクトルに、文書に関するモーダルの情報に基づくベクトルから得たバリューの重み付け和を、単純に加算すると、問題の解決に有用な情報が失われやすい。このため、問題を解いた際の解の精度が悪くなりやすい。

【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、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する補正ベクトルを生成し、生成した前記補正ベクトルを、前記第一のモーダルの情報に基づくベクトルに結合し、所定のルールに従って、結合後の前記第一のモーダルの情報に基づくベクトルを圧縮し、圧縮後の前記第一のモーダルの情報に基づくベクトルに対して、正規化処理を実施し、前記正規化処理により得た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。

【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。

【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。

【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、画像TA層501の具体例を示す説明図である。

【図8】図8は、画像TA層501の別の具体例を示す説明図である。

【図9】図9は、画像TA層501と文書TA層503との比較例を示す説明図である。

【図10】図10は、CAN 500を用いた動作の一例を示す説明図である。

【図11】図11は、出力装置100の利用例1を示す説明図(その1)である。

【図12】図12は、出力装置100の利用例1を示す説明図(その2)である。

【図13】図13は、出力装置100の利用例2を示す説明図(その1)である。

【図14】図14は、出力装置100の利用例2を示す説明図(その2)である。

【図15】図15は、学習処理手順の一例を示すフローチャートである。

【図16】図16は、推定処理手順の一例を示すフローチャートである。

【図17】図17は、アテンション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】

【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、

10

20

30

40

50

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、Attentionにより情報を変換するTransformerを利用した、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を参照することができる。

【0013】

非特許文献2 : Devlin, Jacob et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." NAACL-HLT (2019). 10

【0014】

ここで、BERTは、文書に関するモ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。

【0015】

これに対し、例えば、VideoBERTと呼ばれる手法がある。VideoBERTは、具体的には、BERTを、文書に関するモダルの情報と、画像に関するモ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3を参照することができる。 20

【0016】

非特許文献3 : Sun, Chen, et al. "Videobert: A joint model for video and language representation learning." arXiv preprint arXiv:1904.01766 (2019).

【0017】

また、例えば、MCAN(Modular Co-Attention Network)と呼ばれる手法がある。MCANは、文書に関するモダルの情報に基づくベクトルと、文書に関するモダルの情報に基づくベクトルを基に補正した、画像に関するモダルの情報に基づくベクトルとを参照し、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4を参照することができる。 30

【0018】

非特許文献4 : Yu, Zhou, et al. "Deep Modular Co-Attention Networks for Visual Question Answering." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.

【0019】

また、例えば、ViLBERT(Vision-and-Language Bidirectional Encoder Representations from Transformers)と呼ばれる手法がある。ViLBERTは、画像に関するモダルの情報に基づくベクトルを基に補正した、文書に関するモダルの情報に基づくベクトルと、文書に関するモダルの情報に基づくベクトルを基に補正した、画像に関するモ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

【0020】

非特許文献5 : Lu, Jiasen, et al. "vilbert: Pre-training task-agnostic visiolinguistic representations for vision-and-language t 50

asks.” arXiv preprint arXiv:1908.02265 (2019).

【0021】

しかしながら、上述したVideoBERT、MCAN、およびViLBERTなどの手法でも、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。具体的には、いずれの手法でも、Attentionにより、画像に関するモーダルの情報に基づくベクトルに、文書に関するモーダルの情報に基づくベクトルから得たバリューの重み付け和を、単純に加算するため、問題の解決に有用な情報が失われやすいという性質が存在する。このため、いずれの手法でも、問題を解いた際の解の精度が悪くなりやすい。また、VideoBERTでは、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。

10

【0022】

そこで、本実施の形態では、問題を解くにあたり有用なベクトルを生成可能にすることにより、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可能でありつつ、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。

【0023】

図1において、出力装置100は、例えば、Attentionを実現する変換モデル110を有する。変換モデルは、生成モデル101と、結合モデル102と、圧縮モデル103と、正規化モデル104とを含む。

20

【0024】

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を取得する。モーダルは、情報の様式を意味する。第一のモーダルと、第二のモーダルとは、それぞれ異なるモーダルである。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。

【0025】

第一のモーダルの情報に基づくベクトルは、例えば、第一のモーダルに従って表現されたベクトルである。第一のモーダルの情報に基づくベクトルは、例えば、第一のモーダルの情報に基づいて生成される。第一のモーダルの情報は、例えば、画像である。第一のモーダルの情報に基づくベクトルは、例えば、画像に基づいて生成されたベクトルである。

30

【0026】

第二のモーダルの情報に基づくベクトルは、例えば、第二のモーダルに従って表現されたベクトルである。第二のモーダルの情報に基づくベクトルは、例えば、第二のモーダルの情報に基づいて生成される。第二のモーダルの情報は、例えば、文書である。第二のモーダルの情報に基づくベクトルは、例えば、文書に基づいて生成されたベクトルである。

【0027】

(1-1)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補正ベクトルを生成する。出力装置100は、例えば、生成モデル101を用いて、第一のモーダルの情報に基づくベクトルを補正する補正ベクトルを生成する。

40

【0028】

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。

【0029】

(1-2)出力装置100は、生成した補正ベクトルを、第一のモーダルの情報に基づくベクトルに結合する。出力装置100は、例えば、結合モデル102を用いて、生成した

50

補正ベクトルを、第一のモーダルの情報に基づくベクトルに結合する。

【0030】

(1-3) 出力装置100は、所定のルールに従って、結合後の第一のモーダルの情報に基づくベクトルを圧縮する。出力装置100は、例えば、圧縮モデル103を用いて、結合後の第一のモーダルの情報に基づくベクトルを圧縮する。圧縮は、次元数を低減しない変換を含む。

【0031】

(1-4) 出力装置100は、圧縮後の第一のモーダルの情報に基づくベクトルに対して、正規化処理を実施する。出力装置100は、例えば、正規化モデル104を用いて、正規化処理を実施する。正規化処理を実施する具体例については、例えば、図7を用いて後述する。

10

【0032】

(1-5) 出力装置100は、正規化処理により得た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ベクトルを生成し、利用可能にすることができる。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。

【0033】

ここで、例えば、第一のモーダルが画像に関し、第二のモーダルが文書に関する場合、第二のモーダルは、第一のモーダルの上位の階層であるという特徴を有していると考えることができる。具体的には、「りんご(単語)」は、複数の「りんご(画像)」を包含する概念である。

20

【0034】

出力装置100は、この特徴を利用し、画像に関する第一のモーダルの情報に基づくベクトルに、文書に関する第二のモーダルの情報に基づくベクトルに基づく補正ベクトルを結合した上で、圧縮することができる。このため、出力装置100は、圧縮後のベクトルにおいて、画像と文書とのうち問題の解決に有用な情報が失われ辛く、反映され易くすることができる。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、コンピュータ上で効果的に表現した圧縮後のベクトルを利用可能にすることができる。結果として、出力装置100は、複数のモーダルの情報を用いて問題を解くにあたり、有用なベクトルを得ることができ、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

【0035】

ここでは、第一のモーダルと、第二のモーダルとが、それぞれ異なるモーダルである場合について説明したが、これに限らない。例えば、第一のモーダルと、第二のモーダルとが同一のモーダルである場合があってもよい。

【0036】

(情報処理システム200の一例)

40

次に、図2を用いて、図1に示した出力装置100を適用した、情報処理システム200の一例について説明する。

【0037】

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2において、情報処理システム200は、出力装置100と、クライアント装置201と、端末装置202とを含む。

【0038】

情報処理システム200において、出力装置100とクライアント装置201とは、有線または無線のネットワーク210を介して接続される。ネットワーク210は、例えば、LAN(Local Area Network)、WAN(Wide Area Ne

50

t w o r k)、インターネットなどである。また、情報処理システム 200 において、出力装置 100 と端末装置 202 とは、有線または無線のネットワーク 210 を介して接続される。

【0039】

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統合した統合ベクトルを生成する Co - A t t e n t i o n 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 - A t t e n t i o n Network は、例えば、図 1 に示した変換モデル 110 を用いて形成される。

10

【0040】

出力装置 100 は、教師データに基づいて、Co - A t t e n t i o n 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 100 のユーザにより出力装置 100 に入力される。正解データは、例えば、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は、文書である。

20

【0041】

出力装置 100 は、例えば、第一のモーダルの情報となる教師データの画像から、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、第二のモーダルの情報となる教師データの文書から、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 100 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 - A t t e n t i o n Network を更新する。出力装置 100 は、誤差逆伝搬以外の学習方法により、Co - A t t e n t i o n Network を更新してもよい。

【0042】

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を取得する。そして、出力装置 100 は、Co - A t t e n t i o n 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、統合ベクトルを生成し、生成した統合ベクトルに基づいて、問題を解く。その後、出力装置 100 は、問題を解いた結果を、クライアント装置 201 に送信する。

30

【0043】

出力装置 100 は、例えば、出力装置 100 のユーザにより出力装置 100 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 100 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 201 または端末装置 202 から受信することにより取得してもよい。また、出力装置 100 は、例えば、第一のモーダルの情報を、クライアント装置 201 または端末装置 202 から受信し、受信した第一のモーダルの情報から、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てもよい。

40

【0044】

出力装置 100 は、例えば、出力装置 100 のユーザにより出力装置 100 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 100 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 201 または端末装置 202 から受信することにより取得してもよい。また、出力装置 100 は、例えば、第二のモーダルの情報を、クライアント装置 201 または端末装置 202 から受信し、受信した第二のモー

50

ダルの情報から、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してもよい。

【0045】

そして、出力装置100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、統合ベクトルを生成し、生成した統合ベクトルに基づいて、問題を解く。その後、出力装置100は、問題を解いた結果を、クライアント装置201に送信する。出力装置100は、例えば、サーバやPC(Personal Computer)などである。

【0046】

クライアント装置201は、出力装置100と通信可能なコンピュータである。クライアント装置201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100に送信してもよい。また、クライアント装置201は、例えば、第一のモーダルの情報を、出力装置100に送信してもよい。クライアント装置201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100に送信してもよい。また、クライアント装置201は、例えば、第二のモーダルの情報を、出力装置100に送信してもよい。

【0047】

クライアント装置201は、出力装置100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201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。

【0048】

端末装置202は、出力装置100と通信可能なコンピュータである。端末装置202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100に送信してもよい。また、端末装置202は、例えば、第一のモーダルの情報を、出力装置100に送信してもよい。端末装置202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100に送信してもよい。また、端末装置202は、例えば、第二のモーダルの情報を、出力装置100に送信してもよい。端末装置202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202は、具体的には、監視カメラであってもよい。

【0049】

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。

【0050】

ここでは、教師データが、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、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。

【0051】

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

10

20

30

40

50

末装置 202 と一体である場合があってもよい。

【0052】

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。

【0053】

(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。

10

【0054】

(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。

20

【0055】

(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。

30

【0056】

(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。

【0057】

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/F 304 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。

40

【0058】

ここで、CPU 301 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 CPU 301 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、CPU 301 にロードされることで、コーディングされている処理を CPU 301 に実行させる。

【0059】

ネットワーク I/F 303 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/F 30

50

3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。

【0060】

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。

10

【0061】

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。

【0062】

(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。

20

【0063】

(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。

【0064】

(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。

【0065】

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、生成部402と、結合部403と、変換部404と、正規化部405と、出力部406とを含む。

30

【0066】

記憶部400は、例えば、図3に示したメモリ302や記録媒体305などの記憶領域によって実現される。以下では、記憶部400が、出力装置100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400が、出力装置100とは異なる装置に含まれ、記憶部400の記憶内容が出力装置100から参照可能である場合があってもよい。

【0067】

取得部401～出力部406は、制御部の一例として機能する。取得部401～出力部406は、具体的には、例えば、図3に示したメモリ302や記録媒体305などの記憶領域に記憶されたプログラムをCPU301に実行させることにより、または、ネットワークI/F303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302や記録媒体305などの記憶領域に記憶される。

40

【0068】

記憶部400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400は、Attentionを実現し、第一のモダルの情報に基づくベクトルを、第二のモダルの情報に基づくベクトルに基づいて補正し、補正後の第一のモダルの情報に基づくベクトルを出力する変換モデルを記憶する。

【0069】

例えば、第一のモダルは、画像に関するモダルであり、第二のモダルは、文書に関

50

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。例えば、第一のモーダルは、第二のモーダルと同一であってもよい。

【0070】

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、各機能部に出力する。また、取得部401は、記憶部400に記憶しておいた各種情報を、各機能部に出力してもよい。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。

10

【0071】

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、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。

【0072】

取得部401は、具体的には、第一のモーダルの情報として、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、第二のモーダルの情報として、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

【0073】

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。

30

【0074】

取得部401は、具体的には、第一のモーダルの情報として、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、第二のモーダルの情報として、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、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ることにより、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を取得してもよい。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から受信することにより取得してもよい。

40

【0076】

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

50

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。

【0077】

生成部 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する補正ベクトルを生成する。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。

10

【0078】

生成部 402 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内積に基づいて、補正ベクトルを生成する。生成部 402 は、具体的には、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補正ベクトルを生成する。

【0079】

生成部 402 は、より具体的には、画像に関するモーダルの情報に基づくベクトルから得たクエリと、文書に関する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、画像に関するモーダルの情報に基づくベクトルを補正する補正ベクトルを生成する。ここで、補正ベクトルを生成する一例は、例えば、図 7 を用いて後述する動作例に示す。これにより、生成部 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可能な補正ベクトルを生成することができる。

20

【0080】

結合部 403 は、生成した補正ベクトルを、第一のモーダルの情報に基づくベクトルに結合する。結合部 403 は、例えば、補正ベクトルを、第一のモーダルの情報に基づくベクトルに加算せず、第一のモーダルの前後いずれかに結合する。これにより、結合部 403 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失われ辛く、反映され易いように、第一のモーダルの情報に基づくベクトルを加工することができる。

30

【0081】

変換部 404 は、所定のルールに従って、結合後の第一のモーダルの情報に基づくベクトルを圧縮する。所定のルールは、例えば、学習により自動で設定される。変換部 404 は、例えば、多層ニューラルネットワークを用いて、結合後の第一のモーダルの情報に基づくベクトルを圧縮する。これにより、変換部 404 は、結合後の第一のモーダルの情報に基づくベクトルの次元数を、扱いやすい次元数に変換することができる。

40

【0082】

正規化部 405 は、圧縮後の第一のモーダルの情報に基づくベクトルに対して、正規化処理を実施する。正規化部 405 は、例えば、第一のモーダルの情報に基づくベクトルと、補正ベクトルとの和を正規化し、当該正規化により得たベクトルと、圧縮後の第一のモーダルの情報に基づくベクトルとの和を正規化する。これにより、正規化部 405 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効率よく反映された、問題の解決に有用なベクトルを得ることができる。

【0083】

50

正規化部 405 は、例えば、結合後の第一のモーダルの情報に基づくベクトルと、圧縮後の第一のモーダルの情報に基づくベクトルとの和を正規化する。これにより、正規化部 405 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効率よく反映された、問題の解決に有用なベクトルを得ることができる。

【0084】

出力部 406 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 406 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。

10

【0085】

出力部 406 は、正規化処理により得たベクトルを出力する。これにより、出力部 406 は、正規化処理により得たベクトルを利用し、Attention を実現することができる。そして、出力部 406 は、Attention により、Co-Attention Network を実現可能にすることができる。

【0086】

出力部 406 は、例えば、Attention により、問題の解決に有用に、正規化処理により得られたベクトルを出力することができる。このため、出力部 406 は、Co-Attention Network を、問題の解決に有用になるように学習可能にすることができる。また、出力部 406 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。

20

【0087】

(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。

【0088】

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

【0089】

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。

【0090】

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_T 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

【0091】

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 7 および図 8 を用いて後述する。

50

【 0 0 9 2 】

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_I を生成し、結合層 5 0 5 に出力する。画像 S A 層 5 0 2 を実現する S A 層の具体例については、例えば、図 6 を用いて後述する。

【 0 0 9 3 】

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 を実現する T A 層の具体例については、例えば、図 6 を用いて後述する。

10

【 0 0 9 4 】

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_L を生成して出力する。文書 S A 層 5 0 4 を実現する S A 層の具体例については、例えば、図 6 を用いて後述する。

【 0 0 9 5 】

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_I と、特徴量ベクトル Z_L 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_I と、特徴量ベクトル Z_L 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 0 9 6 】

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_T を生成して出力する。特徴量ベクトル Z_T は、集約ベクトル Z_H と、文書に関する統合特徴量ベクトル Z_{Y1} ~ Z_M と、画像に関する統合特徴量ベクトル Z_{M+1} ~ Z_{M+N} 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_H を含む特徴量ベクトル Z_T 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 0 9 7 】

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 0 9 8 】

ここでは、CAN 5 0 0 が、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、結合層 5 0 5 と、統合 S A 層 5 0 6 とを有する場合について説明したが、これに限らない。例えば、CAN 5 0 0 が、結合層 5 0 5 と、統合 S A 層 5 0 6 とを有していない場合があってもよい。この場合、出力装置 1 0 0 は、例えば、問題を解くにあたり、画像 S A 層 5 0 2 の出力と、文書 S A 層 5 0 4 の出力とを利用する。

【 0 0 9 9 】

次に、図 6 の説明に移行し、CAN 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 となどを実現する S A 層 6 0 0 の具体例と、CAN 5 0 0 を形成する文書 T A 層 5 0 3 などを実現する T A 層 6 1 0 の具体例とについて説明する。CAN 5

50

00を形成する画像TA層501の具体例については、図7を用いて後述する。

【0100】

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。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。

【0101】

図6に示すように、SA層600は、MHA層601と、A&N層602と、FF層603と、A&N層604とを有する。MHA層601は、入力ベクトルXから得たクエリQとキーKとバリュー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602に出力する。MHA層601は、具体的には、入力ベクトルXを、Head個のベクトルに分割して処理する。Headは、1以上の自然数である。

10

【0102】

A&N層602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、FF層603とA&N層604とに出力する。FF層603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604に出力する。A&N層604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。

【0103】

また、TA層610は、MHA層611と、A&N層612と、FF層613と、A&N層614とを有する。MHA層611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とバリュー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612に出力する。A&N層612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、FF層613とA&N層614とに出力する。FF層613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614に出力する。A&N層614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

【0104】

上述したMHA層601やMHA層611は、より具体的には、Headの個数分のAttention層620により形成される。Attention層620は、MatMul層621と、Scale層622と、Mask層623と、SoftMax層624と、MatMul層625とを有する。

30

【0105】

MatMul層621は、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、Score全体を定数aで除算し、更新する。Mask層623は、更新後のScoreをマスク処理してもよい。SoftMax層624は、更新後のScoreを、正規化し、Attに設定する。MatMul層625は、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。次に、図7および図8を用いて、CAN500を形成する画像TA層501の具体例について説明する。

【0106】

図7は、画像TA層501の具体例を示す説明図である。図7において、画像TA層501は、MHA層701と、A&N層702と、Con層703と、FF層704と、A&N層705とを含む。MHA層701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とバリュー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702およびCon層703に出力する。A&N層702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、A&N層705に出力する。

40

【0107】

Con層703は、入力ベクトルXと補正ベクトルRとを結合し、結合ベクトルをFF層704に出力する。FF層704は、結合ベクトルを圧縮し、圧縮後のベクトルを、A&

50

N層705に出力する。A&N層705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、正規化で得た出力ベクトルを出力する。次に、図8を用いて、画像TA層501の別の具体例について説明する。

【0108】

図8は、画像TA層501の別の具体例を示す説明図である。図8において、画像TA層501は、MHA層801と、Con層802と、FF層803と、A&N層804とを含む。MHA層801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とバリュー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、Con層802に出力する。

【0109】

Con層802は、入力ベクトルXと補正ベクトルRとを結合し、結合ベクトルをFF層803およびA&N層804に出力する。FF層803は、結合ベクトルを圧縮し、圧縮後のベクトルを、A&N層804に出力する。A&N層804は、結合ベクトルと、圧縮後のベクトルとを加算した上で正規化し、正規化で得た出力ベクトルを出力する。次に、図9を用いて、画像TA層501と文書TA層503との比較例について説明する。

【0110】

図9は、画像TA層501と文書TA層503との比較例を示す説明図である。図9に示すように、画像TA層501と、文書TA層503とは、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。しかしながら、画像TA層501と、文書TA層503とは、それぞれ、異なる手法で、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とを扱うことになる。

【0111】

例えば、画像TA層501は、画像に関する特徴量ベクトルIに、ベクトル Z_{I1} を結合することにより、新たな特徴量ベクトル Z_{I2} を生成する。一方で、文書TA層503は、文書に関する特徴量ベクトルLに、ベクトル Z_{L1} を加算することにより、新たな特徴量ベクトル Z_{L2} を生成する。これにより、出力装置100は、それぞれ性質が異なる、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とに対し、異なる扱い方をすることができる。

【0112】

そして、出力装置100は、画像TA層501において、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とのうち、問題の解決に有用な情報が失われ辛くすることができる。結果として、出力装置100は、複数のモダルの情報を用いて問題を解くにあたり有用なベクトルを得ることができ、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。

【0113】

ここでは、画像TA層501を、図7および図8に示す具体例のように形成する場合について説明したが、これに限らない。例えば、画像SA層502と、文書TA層503と、文書SA層504と、統合SA層506との少なくともいずれかを、図7および図8に示す具体例と同様に形成する場合があってもよい。次に、図10を用いて、出力装置100による、CAN500を用いた動作の一例について説明する。

【0114】

図10は、CAN500を用いた動作の一例を示す説明図である。図10において、出力装置100は、文書1000を取得し、画像1010を取得する。出力装置100は、文書1000をトークン化し、トークン集合1001をベクトル化し、文書1000に関する特徴量ベクトル1002を生成し、CAN500に入力する。また、出力装置100は、画像1010から物体を検出し、物体ごとの部分画像の集合1011をベクトル化し、画像1010に関する特徴量ベクトル1012を生成し、CAN500に入力する。

【0115】

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1030に入力する。出力装置100は

10

20

30

40

50

、危険度推定器 1030 から推定結果 No 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 1030 に危険であるか否かを推定させることができ、危険であるか否かを精度よく推定可能にすることができる。危険度推定器 1030 は、例えば、銃を持った人物が写っている画像 1010 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 No = 危険ではないと推定することができる。

【0116】

(出力装置 100 の利用例)

次に、図 11 ~ 図 14 を用いて、出力装置 100 の利用例について説明する。

【0117】

10

図 11 および図 12 は、出力装置 100 の利用例 1 を示す説明図である。図 11 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 1100 と、画像 1100 に対応する字幕となる文書 1110 とを取得する。画像 1100 は、例えば、りんごを切るシーンを写す。

【0118】

出力装置 100 は、画像 1100 を変換器 1120 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 の単語 apple をマスクした上で、変換器 1130 により特徴量ベクトルに変換し、CAN500 に入力する。

【0119】

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 1140 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 1120, 1130 や識別器 1140 を学習してもよい。

20

【0120】

これにより、出力装置 100 は、画像 1100 と字幕となる文書 1110 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 1120, 1130 や識別器 1140 を更新することができる。次に、図 12 の説明に移行する。

【0121】

図 12 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 1120, 1130 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 1200 と、画像 1200 に対応する質問文となる文書 1210 とを取得する。画像 1200 は、例えば、りんごを切るシーンを写す。

30

【0122】

出力装置 100 は、画像 1200 を変換器 1120 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 1210 を変換器 1130 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 1220 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 1200 と質問文となる文書 1210 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。

40

【0123】

図 13 および図 14 は、出力装置 100 の利用例 2 を示す説明図である。図 13 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 1300 と、画像 1300 に対応する字幕となる文書 1310 とを取得する。画像 1300 は、例えば、りんごを切るシーンを写す。

【0124】

出力装置 100 は、画像 1300 を変換器 1320 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 1310 の単語 apple をマスクした上で、変換器 1330 により特徴量ベクトルに変換し、CAN500 に入力する。

50

【 0 1 2 5 】

出力装置 1 0 0 は、CAN 5 0 0 により生成された特徴量ベクトルを、識別器 1 3 4 0 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 1 0 0 は、算出した誤差に基づいて、誤差逆伝搬により CAN 5 0 0 を学習する。また、出力装置 1 0 0 は、誤差逆伝搬により、変換器 1 3 2 0 , 1 3 3 0 や識別器 1 3 4 0 を学習する。

【 0 1 2 6 】

これにより、出力装置 1 0 0 は、画像 1 3 0 0 と字幕となる文書 1 3 1 0 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、CAN 5 0 0、および変換器 1 1 2 0 , 1 1 3 0 や識別器 1 1 4 0 を更新することができる。次に、図 1 4 の説明に移行する。

10

【 0 1 2 7 】

図 1 4 において、出力装置 1 0 0 は、試験フェーズを実施し、学習した変換器 1 3 2 0 , 1 3 3 0 や識別器 1 3 4 0 と、学習した CAN 5 0 0 とを用いて、危険度を予測して出力する。出力装置 1 0 0 は、例えば、何らかのシーンを写した画像 1 4 0 0 と、画像に対応する説明文となる文書 1 4 1 0 とを取得する。画像 1 4 0 0 は、例えば、ももを切るシーンを写す。

【 0 1 2 8 】

出力装置 1 0 0 は、画像 1 4 0 0 を変換器 1 3 2 0 により特徴量ベクトルに変換し、CAN 5 0 0 に入力する。また、出力装置 1 0 0 は、文書 1 4 1 0 を変換器 1 3 3 0 により特徴量ベクトルに変換し、CAN 5 0 0 に入力する。出力装置 1 0 0 は、CAN 5 0 0 により生成された特徴量ベクトルを、識別器 1 3 4 0 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 1 0 0 は、画像 1 4 0 0 と説明文となる文書 1 4 1 0 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。

20

【 0 1 2 9 】

(学習処理手順)

次に、図 1 5 を用いて、出力装置 1 0 0 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 CPU 3 0 1 と、メモリ 3 0 2 や記録媒体 3 0 5 などの記憶領域と、ネットワーク I / F 3 0 3 とによって実現される。

【 0 1 3 0 】

図 1 5 は、学習処理手順の一例を示すフローチャートである。図 1 5 において、出力装置 1 0 0 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 S 1 5 0 1) 。

30

【 0 1 3 1 】

次に、出力装置 1 0 0 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トに基づいて、画像 T A 層 5 0 1 を用いて、画像の特徴量ベクトルを補正する (ステップ S 1 5 0 2) 。ここで、出力装置 1 0 0 は、具体的には、図 1 4 に後述するアテンション処理を実行することにより、画像の特徴量ベクトルを補正する。

【 0 1 3 2 】

そして、出力装置 1 0 0 は、補正後の画像の特徴量ベクトルに基づいて、画像 S A 層 5 0 2 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 S 1 5 0 3) 。

40

【 0 1 3 3 】

次に、出力装置 1 0 0 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トに基づいて、文書 T A 層 5 0 3 を用いて、文書の特徴量ベクトルを補正する (ステップ S 1 5 0 4) 。

【 0 1 3 4 】

そして、出力装置 1 0 0 は、補正後の文書の特徴量ベクトルに基づいて、文書 S A 層 5 0 4 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 S 1 5 0 5) 。

50

【 0 1 3 5 】

次に、出力装置 1 0 0 は、集約用ベクトルを初期化する（ステップ S 1 5 0 6）。そして、出力装置 1 0 0 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 S 1 5 0 7）。

【 0 1 3 6 】

次に、出力装置 1 0 0 は、結合ベクトルに基づいて、統合 S A 層 5 0 6 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 S 1 5 0 8）。そして、出力装置 1 0 0 は、集約ベクトルに基づいて、C A N 5 0 0 を学習する（ステップ S 1 5 0 9）。

【 0 1 3 7 】

その後、出力装置 1 0 0 は、学習処理を終了する。これにより、出力装置 1 0 0 は、C A N 5 0 0 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、C A N 5 0 0 のパラメータを更新することができる。 10

【 0 1 3 8 】

ここで、出力装置 1 0 0 は、図 1 5 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 S 1 5 0 2 , S 1 5 0 3 の処理と、ステップ S 1 5 0 4 , S 1 5 0 5 の処理との順序は入れ替え可能である。また、出力装置 1 0 0 は、ステップ S 1 5 0 2 ~ S 1 5 0 5 の処理を繰り返し実行してもよい。

【 0 1 3 9 】

（推定処理手順）

次に、図 1 6 を用いて、出力装置 1 0 0 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 C P U 3 0 1 と、メモリ 3 0 2 や記録媒体 3 0 5 などの記憶領域と、ネットワーク I / F 3 0 3 とによって実現される。 20

【 0 1 4 0 】

図 1 6 は、推定処理手順の一例を示すフローチャートである。図 1 6 において、出力装置 1 0 0 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 S 1 6 0 1）。

【 0 1 4 1 】

次に、出力装置 1 0 0 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トに基づいて、画像 T A 層 5 0 1 を用いて、画像の特徴量ベクトルを補正する（ステップ S 1 6 0 2）。ここで、出力装置 1 0 0 は、具体的には、図 1 4 に後述するアテンション処理を実行することにより、画像の特徴量ベクトルを補正する。 30

【 0 1 4 2 】

そして、出力装置 1 0 0 は、補正後の画像の特徴量ベクトルに基づいて、画像 S A 層 5 0 2 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 S 1 6 0 3）。

【 0 1 4 3 】

次に、出力装置 1 0 0 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トに基づいて、文書 T A 層 5 0 3 を用いて、文書の特徴量ベクトルを補正する（ステップ S 1 6 0 4）。 40

【 0 1 4 4 】

そして、出力装置 1 0 0 は、補正後の文書の特徴量ベクトルに基づいて、文書 S A 層 5 0 4 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 S 1 6 0 5）。

【 0 1 4 5 】

次に、出力装置 1 0 0 は、集約用ベクトルを初期化する（ステップ S 1 6 0 6）。そして、出力装置 1 0 0 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 S 1 6 0 7）。

【 0 1 4 6 】

次に、出力装置 1 0 0 は、結合ベクトルに基づいて、統合 S A 層 5 0 6 を用いて、結合ベ 50

クトルを補正し、集約ベクトルを生成する（ステップ S 1 6 0 8）。そして、出力装置 1 0 0 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 S 1 6 0 9）。

【 0 1 4 7 】

次に、出力装置 1 0 0 は、推定した状況を入力する（ステップ S 1 6 1 0）。そして、出力装置 1 0 0 は、推定処理を終了する。これにより、出力装置 1 0 0 は、CAN 5 0 0 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。

【 0 1 4 8 】

ここで、出力装置 1 0 0 は、図 1 6 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 S 1 6 0 2, S 1 6 0 3 の処理と、ステップ S 1 6 0 4, S 1 6 0 5 の処理との順序は入れ替え可能である。また、出力装置 1 0 0 は、ステップ S 1 6 0 2 ~ S 1 6 0 5 の処理を繰り返し実行してもよい。

10

【 0 1 4 9 】

（アテンション処理手順）

次に、図 1 7 を用いて、画像 T A 層により、出力装置 1 0 0 が実行する、アテンション処理手順の一例について説明する。アテンション処理は、例えば、図 3 に示した CPU 3 0 1 と、メモリ 3 0 2 や記録媒体 3 0 5 などの記憶領域と、ネットワーク I / F 3 0 3 とによって実現される。

【 0 1 5 0 】

図 1 7 は、アテンション処理手順の一例を示すフローチャートである。図 1 7 において、出力装置 1 0 0 は、ベクトル X となる画像の特徴量ベクトルと、ベクトル Y となる文書の特徴量ベクトルとを取得する（ステップ S 1 7 0 1）。

20

【 0 1 5 1 】

次に、出力装置 1 0 0 は、取得した画像の特徴量ベクトルからベクトル Q u e r y を生成する（ステップ S 1 7 0 2）。そして、出力装置 1 0 0 は、取得した文書の特徴量ベクトルからベクトル k e y とベクトル V a l u e を生成する（ステップ S 1 7 0 3）。

【 0 1 5 2 】

次に、出力装置 1 0 0 は、生成したベクトル Q u e r y と、生成したベクトル k e y との内積を算出する（ステップ S 1 7 0 4）。そして、出力装置 1 0 0 は、内積の s o f t m a x によりベクトル A t t を生成する（ステップ S 1 7 0 5）。

30

【 0 1 5 3 】

次に、出力装置 1 0 0 は、ベクトル A t t とベクトル V a l u e との内積によりベクトル R を生成する（ステップ S 1 7 0 6）。そして、出力装置 1 0 0 は、ベクトル R とベクトル X とを結合したベクトル X ' を生成する（ステップ S 1 7 0 7）。

【 0 1 5 4 】

次に、出力装置 1 0 0 は、多層ニューラルネットワークにより、ベクトル X ' を、ベクトル X と同じ次元に圧縮し、ベクトル X " を生成する（ステップ S 1 7 0 8）。そして、出力装置 1 0 0 は、ベクトル R とベクトル X とを用いて、ベクトル X " を正規化し、正規化後のベクトルを取得する（ステップ S 1 7 0 9）。

【 0 1 5 5 】

次に、出力装置 1 0 0 は、取得した正規化後のベクトルを出力する（ステップ S 1 7 1 0）。そして、出力装置 1 0 0 は、アテンション処理を終了する。これにより、出力装置 1 0 0 は、画像と文書とのうち問題の解決に有用な情報が失われ辛いように、正規化後のベクトルを生成して取得することができる。

40

【 0 1 5 6 】

ここで、出力装置 1 0 0 は、図 1 7 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 S 1 7 0 2 の処理と、ステップ S 1 7 0 3 の処理との順序は入れ替え可能である。

【 0 1 5 7 】

以上説明したように、出力装置 1 0 0 によれば、第一のモーダルの情報に基づくベクトル

50

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補正ベクトルを生成することができる。出力装置 100 によれば、生成した補正ベクトルを、第一のモーダルの情報に基づくベクトルに結合することができる。出力装置 100 によれば、所定のルールに従って、結合後の第一のモーダルの情報に基づくベクトルを圧縮することができる。出力装置 100 によれば、圧縮後の第一のモーダルの情報に基づくベクトルに対して、正規化処理を実施することができる。出力装置 100 によれば、正規化処理により得た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を残して、問題を解くのに有用なベクトルを得ることができ、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

【0158】

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、補正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、アテンションを実現することができる。また、出力装置 100 は、問題を解くのに有用な補正ベクトルを得ることができる。

【0159】

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、補正ベクトルとの和を正規化し、当該正規化により得たベクトルと、圧縮後の第一のモーダルの情報に基づくベクトルとの和を正規化することができる。これにより、出力装置 100 は、正規化処理を実現することができる。

20

【0160】

出力装置 100 によれば、結合後の第一のモーダルの情報に基づくベクトルと、圧縮後の第一のモーダルの情報に基づくベクトルとの和を正規化することができる。これにより、出力装置 100 は、正規化処理を実現することができる。

【0161】

出力装置 100 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 100 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 100 は、ターゲットアテンション層を実現することができる。また、出力装置 100 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。

30

【0162】

出力装置 100 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 100 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 100 は、ターゲットアテンション層を実現することができる。また、出力装置 100 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。

【0163】

出力装置 100 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 100 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 100 は、ターゲットアテンション層を実現することができる。また、出力装置 100 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。

40

【0164】

出力装置 100 によれば、第一のモーダルと、第二のモーダルとに、同一のモーダルを採用することができる。これにより、出力装置 100 は、セルフアテンション層を実現することができる。また、出力装置 100 は、同一のモーダルの異なる情報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。

【0165】

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 PC やワークス

50

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、CD (Compact Disc) - ROM、MO、DVD (Digital Versatile Disc) などである。また、本実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。

【0166】

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。

【0167】

(付記1) 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、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補正ベクトルを生成し、
生成した前記補正ベクトルを、前記第一のモーダルの情報に基づくベクトルに結合し、
所定のルールに従って、結合後の前記第一のモーダルの情報に基づくベクトルを圧縮し、
圧縮後の前記第一のモーダルの情報に基づくベクトルに対して、正規化処理を実施し、
前記正規化処理により得た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。

【0168】

(付記2) 前記生成する処理は、
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、前記補正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。

【0169】

(付記3) 前記正規化処理を実施する処理は、
前記第一のモーダルの情報に基づくベクトルと、前記補正ベクトルとの和を正規化し、
当該正規化により得たベクトルと、圧縮後の前記第一のモーダルの情報に基づくベクトルとの和を正規化する、
ことを特徴とする付記1または2に記載の出力方法。

【0170】

(付記4) 前記正規化処理を実施する処理は、
結合後の前記第一のモーダルの情報に基づくベクトルと、圧縮後の前記第一のモーダルの情報に基づくベクトルとの和を正規化する、
ことを特徴とする付記1または2に記載の出力方法。

【0171】

(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。

【0172】

(付記6) 前記第一のモーダルは、前記第二のモーダルと同一である、
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。

【0173】

(付記7) 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、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補正ベクトルを生成し、
生成した前記補正ベクトルを、前記第一のモーダルの情報に基づくベクトルに結合し、
所定のルールに従って、結合後の前記第一のモーダルの情報に基づくベクトルを圧縮し、
圧縮後の前記第一のモーダルの情報に基づくベクトルに対して、正規化処理を実施し、
前記正規化処理により得た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。

10

20

30

40

50

【 0 1 7 4 】

(付記 8) 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、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補正ベクトルを生成し、

生成した前記補正ベクトルを、前記第一のモーダルの情報に基づくベクトルに結合し、所定のルールに従って、結合後の前記第一のモーダルの情報に基づくベクトルを圧縮し、圧縮後の前記第一のモーダルの情報に基づくベクトルに対して、正規化処理を実施し、前記正規化処理により得た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。

【 符号の説明 】

10

【 0 1 7 5 】

1 0 0	出力装置	
1 0 1	生成モデル	
1 0 2	結合モデル	
1 0 3	圧縮モデル	
1 0 4	正規化モデル	
1 1 0	変換モデル	
2 0 0	情報処理システム	
2 0 1	クライアント装置	
2 0 2	端末装置	20
2 1 0	ネットワーク	
3 0 0	バス	
3 0 1	C P U	
3 0 2	メモリ	
3 0 3	ネットワーク I / F	
3 0 4	記録媒体 I / F	
3 0 5	記録媒体	
4 0 0	記憶部	
4 0 1	取得部	
4 0 2	生成部	30
4 0 3	結合部	
4 0 4	変換部	
4 0 5	正規化部	
4 0 6	出力部	
5 0 0	C A N	
5 0 1	画像 T A 層	
5 0 2	画像 S A 層	
5 0 3	文書 T A 層	
5 0 4	文書 S A 層	
5 0 5	結合層	40
5 0 6	統合 S A 層	
5 1 0	グループ	
6 0 0	S A 層	
6 0 1 , 6 1 1 , 7 0 1 , 8 0 1	M H A 層	
6 0 2 , 6 0 4 , 6 1 2 , 6 1 4 , 7 0 2 , 7 0 5 , 8 0 4	A & N 層	
6 0 3 , 6 1 3 , 7 0 4 , 8 0 3	F F 層	
6 1 0	T A 層	
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	50

6 2 3	M a s k 層	
6 2 4	S o f t M a x 層	
7 0 3 , 8 0 2	C o n 層	
1 0 0 0 , 1 1 1 0 , 1 2 1 0 , 1 3 1 0 , 1 4 1 0	文書	
1 0 0 1	トークン集合	
1 0 0 2 , 1 0 1 2	特徴量ベクトル	
1 0 1 0 , 1 1 0 0 , 1 2 0 0 , 1 3 0 0 , 1 4 0 0	画像	
1 0 1 1	集合	
1 0 3 0	危険度推定器	
1 1 2 0 , 1 1 3 0 , 1 3 2 0 , 1 3 3 0	変換器	10
1 1 4 0 , 1 3 4 0	識別器	
1 2 2 0	回答生成器	

20

30

40

50