(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7207568号
(P7207568)

(45)発行日　令和5年1月18日(2023.1.18)

(24)登録日　令和5年1月10日(2023.1.10)

(51)国際特許分類

Ｇ０６Ｎ　20/00　(2019.01)	Ｇ０６Ｎ	20/00	
Ｇ０６Ｔ　7/00　(2017.01)	Ｇ０６Ｔ	7/00	３５０Ｂ

請求項の数　7　（全29頁）

(21)出願番号　特願2021-555729(P2021-555729)
(86)(22)出願日　令和1年11月14日(2019.11.14)
(86)国際出願番号　PCT/JP2019/044770
(87)国際公開番号　WO2021/095212
(87)国際公開日　令和3年5月20日(2021.5.20)
審査請求日　令和4年1月18日(2022.1.18)

(73)特許権者　000005223
富士通株式会社
神奈川県川崎市中原区上小田中４丁目１番１号
(74)代理人　100104190
弁理士　酒井 昭徳
(72)発明者　山田 萌
神奈川県川崎市中原区上小田中４丁目１番１号 富士通株式会社内
審査官　藤原 敬利

最終頁に続く

(54)【発明の名称】　出力方法、出力プログラム、および出力装置

(57)【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。

【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。

【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。
【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。
【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】
【文献】Ｖａｓｗａｎｉ，Ａｓｈｉｓｈ，ｅｔ ａｌ．“Ａｔｔｅｎｔｉｏｎ ｉｓ ａｌｌ ｙｏｕ ｎｅｅｄ．”Ａｄｖａｎｃｅｓ ｉｎ ｎｅｕｒａｌ ｉｎｆｏｒｍａｔ

ｉｏｎ 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 Ｎｅｔｗｏｒｋ５００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、ＣＡＮ５００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、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）．
【００１４】
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。
【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３を参照することができる。
【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を基に補正した、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。
【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、画像に関するモー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を基に補正した、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。

【００２０】
非特許文献５：Ｌｕ，Ｊｉａｓｅｎ，ｅｔ ａｌ．“ｖｉｌｂｅｒｔ：Ｐｒｅｔｒａｉｎｉｎｇ ｔａｓｋ－ａｇｎｏｓｔｉｃ ｖｉｓｉｏｌｉｎｇｕｉｓｔｉｃ ｒｅｐｒｅｓｅｎｔａｔｉｏｎｓ ｆｏｒ ｖｉｓｉｏｎ－ａｎｄ－ｌａｎｇｕａｇｅ ｔａｓｋｓ．”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。
【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。
【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。
【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。
【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。
【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。
【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。
【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）
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 Ｎｅｔｗｏｒｋ）、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－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に示した変換モデル１１０を用いて形成される。

【００４０】

出力装置１００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。

【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－Ａｔｔｅｎｔｉｏｎ 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－Ａｔｔｅｎｔｉｏｎ 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－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。

【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。

【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。

【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。

【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。

【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／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。

【００５８】

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。

【００５９】
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。
【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。
【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。
【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。
【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。
【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、第一のモーダルの情報に基づくベクトルを、第二のモー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

ー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。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。
【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。
【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。
【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。
【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。

【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。

【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。

【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。

【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】

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／Ｆ３０３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。

【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。

【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 Ｎｅｔｗｏｒｋ５００の具体例について説明する。

【００８８】

図５は、Ｃｏ－Ａｔｔｅｎｔｉｏｎ Ｎｅｔｗｏｒｋ５００の具体例を示す説明図である。以下の説明では、Ｃｏ－Ａｔｔｅｎｔｉｏｎ Ｎｅｔｗｏｒｋ５００を「ＣＡＮ５００」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。

【００８９】

図５に示すように、ＣＡＮ５００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、ＣＡＮ５００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。

【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。

【００９２】

また、画像ＳＡ層５０２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ＳＡ層５０２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５０５に出力する。画像ＳＡ層５０２を実現するＳＡ層の具体例については、例えば、図６を用いて後述する。

【００９３】

また、文書ＴＡ層５０３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ＴＡ層５０３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ＴＡ層５０３は、補正後の文書に関する特徴量ベクトルＬを、文書ＳＡ層５０４に出力する。文書ＴＡ層５０３を実現するＴＡ層の具体例については、例えば、図６を用いて後述する。

【００９４】

また、文書ＳＡ層５０４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ＳＡ層５０４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ＳＡ層５０４を実現するＳＡ層の具体例については、例えば、図６を用いて後述する。

【００９５】

また、結合層５０５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５０５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ＳＡ層５０６に出力する。

【００９６】

また、統合ＳＡ層５０６は、結合ベクトルＣの入力を受け付ける。統合ＳＡ層５０６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_1～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１００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１００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。

【００９７】

ここでは、説明の簡略化のため、画像ＴＡ層５０１と、画像ＳＡ層５０２と、文書ＴＡ層５０３と、文書ＳＡ層５０４とのグループ５１０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ＴＡ層５０１と、画像ＳＡ層５０２と、文書ＴＡ層５０３と、文書ＳＡ層５０４とのグループ５１０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１００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。

【００９８】

ここでは、ＣＡＮ５００が、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とを有する場合について説明したが、これに限らない。例えば、ＣＡＮ５００が、結合層５０５と、統合ＳＡ層５０６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１００は、例えば、問題を解くにあたり、画像ＳＡ層５０２の出力と、文書ＳＡ層５０４の出力とを利用する。

【００９９】

次に、図６の説明に移行し、ＣＡＮ５００を形成する画像ＳＡ層５０２と文書ＳＡ層５

０４と統合ＳＡ層５０６とを実現するＳＡ層６００の具体例と、ＣＡＮ５００を形成する文書ＴＡ層５０３などを実現するＴＡ層６１０の具体例とについて説明する。ＣＡＮ５００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。

【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。

【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。

【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。

【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＆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。

【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、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】

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。

【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。

【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。

【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、Ｃ

ＡＮ５００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。

【０１２５】

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１３４０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１３２０，１３３０や識別器１３４０を学習する。

【０１２６】

これにより、出力装置１００は、画像１３００と字幕となる文書１３１０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１１２０，１１３０や識別器１１４０を更新することができる。次に、図１４の説明に移行する。

【０１２７】

図１４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１３２０，１３３０や識別器１３４０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１４００と、画像に対応する説明文となる文書１４１０とを取得する。画像１４００は、例えば、ももを切るシーンを写す。

【０１２８】

出力装置１００は、画像１４００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１４１０を変換器１３３０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１３４０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１４００と説明文となる文書１４１０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。

【０１２９】

（学習処理手順）

次に、図１５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／Ｆ３０３とによって実現される。

【０１３０】

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１５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Ｓ１５０１）。

【０１３１】

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Ｓ１５０２）。ここで、出力装置１００は、具体的には、図１４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。

【０１３２】

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Ｓ１５０３）。

【０１３３】

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Ｓ１５０４）。

【０１３４】

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５

０４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Ｓ１５０５）。

【０１３５】

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Ｓ１５０６）。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Ｓ１５０７）。

【０１３６】

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Ｓ１５０８）。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Ｓ１５０９）。

【０１３７】

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。

【０１３８】

ここで、出力装置１００は、図１５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Ｓ１５０２，Ｓ１５０３の処理と、ステップＳ１５０４，Ｓ１５０５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Ｓ１５０２～Ｓ１５０５の処理を繰り返し実行してもよい。

【０１３９】

（推定処理手順）

次に、図１６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／Ｆ３０３とによって実現される。

【０１４０】

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１６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Ｓ１６０１）。

【０１４１】

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Ｓ１６０２）。ここで、出力装置１００は、具体的には、図１４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。

【０１４２】

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Ｓ１６０３）。

【０１４３】

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Ｓ１６０４）。

【０１４４】

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Ｓ１６０５）。

【０１４５】

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Ｓ１６０６）。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Ｓ１６０７）。

【０１４６】
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Ｓ１６０８）。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Ｓ１６０９）。
【０１４７】
次に、出力装置１００は、推定した状況を出力する（ステップＳ１６１０）。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。
【０１４８】
ここで、出力装置１００は、図１６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Ｓ１６０２，Ｓ１６０３の処理と、ステップＳ１６０４，Ｓ１６０５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Ｓ１６０２～Ｓ１６０５の処理を繰り返し実行してもよい。
【０１４９】
（アテンション処理手順）
次に、図１７を用いて、画像ＴＡ層により、出力装置１００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／Ｆ３０３とによって実現される。
【０１５０】
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。図１７において、出力装置１００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Ｓ１７０１）。
【０１５１】
次に、出力装置１００は、取得した画像の特徴量ベクトルからベクトルＱｕｅｒｙを生成する（ステップＳ１７０２）。そして、出力装置１００は、取得した文書の特徴量ベクトルからベクトルｋｅｙとベクトルＶａｌｕｅを生成する（ステップＳ１７０３）。
【０１５２】
次に、出力装置１００は、生成したベクトルＱｕｅｒｙと、生成したベクトルｋｅｙとの内積を算出する（ステップＳ１７０４）。そして、出力装置１００は、内積のｓｏｆｔｍａｘによりベクトルＡｔｔを生成する（ステップＳ１７０５）。
【０１５３】
次に、出力装置１００は、ベクトルＡｔｔとベクトルＶａｌｕｅとの内積によりベクトルＲを生成する（ステップＳ１７０６）。そして、出力装置１００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Ｓ１７０７）。
【０１５４】
次に、出力装置１００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Ｓ１７０８）。そして、出力装置１００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Ｓ１７０９）。
【０１５５】
次に、出力装置１００は、取得した正規化後のベクトルを出力する（ステップＳ１７１０）。そして、出力装置１００は、アテンション処理を終了する。これにより、出力装置１００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。
【０１５６】
ここで、出力装置１００は、図１７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Ｓ１７０２の処理と、ステップＳ１７０３の処理との順序は入れ替え可能である。

【０１５７】
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。
【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。
【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。
【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。
【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。

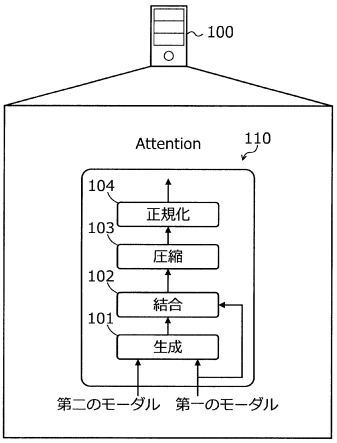
【０１６５】
なお、本実施の形態で説明した出力方法は、予め用意されたプログラムをＰＣやワークステーションなどのコンピュータで実行することにより実現することができる。本実施の形態で説明した出力プログラムは、コンピュータで読み取り可能な記録媒体に記録され、コンピュータによって記録媒体から読み出されることによって実行される。記録媒体は、ハードディスク、フレキシブルディスク、ＣＤ（Ｃｏｍｐａｃｔ　Ｄｉｓｃ）－ＲＯＭ、ＭＯ、ＤＶＤ（Ｄｉｇｉｔａｌ　Ｖｅｒｓａｔｉｌｅ　Ｄｉｓｃ）などである。また、本実施の形態で説明した出力プログラムは、インターネットなどのネットワークを介して配布してもよい。
【０１６６】
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。
【０１７４】
（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】
【０１７５】
１００　出力装置
１０１　生成モデル
１０２　結合モデル
１０３　圧縮モデル
１０４　正規化モデル
１１０　変換モデル
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　生成部
４０３　結合部
４０４　変換部
４０５　正規化部
４０６　出力部
５００　ＣＡＮ
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
６０１，６１１，７０１，８０１　ＭＨＡ層
６０２，６０４，６１２，６１４，７０２，７０５，８０４　Ａ＆Ｎ層
６０３，６１３，７０４，８０３　ＦＦ層
６１０　ＴＡ層
６２０　Ａｔｔｅｎｔｉｏｎ層

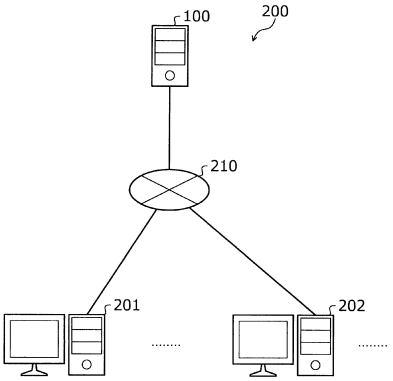
６２１，６２５　ＭａｔＭｕｌ層
６２２　Ｓｃａｌｅ層
６２３　Ｍａｓｋ層
６２４　ＳｏｆｔＭａｘ層
７０３，８０２　Ｃｏｎ層
１０００，１１１０，１２１０，１３１０，１４１０　文書
１００１　トークン集合
１００２，１０１２　特徴量ベクトル
１０１０，１１００，１２００，１３００，１４００　画像
１０１１　集合
１０３０　危険度推定器
１１２０，１１３０，１３２０，１３３０　変換器
１１４０，１３４０　識別器
１２２０　回答生成器

【図面】

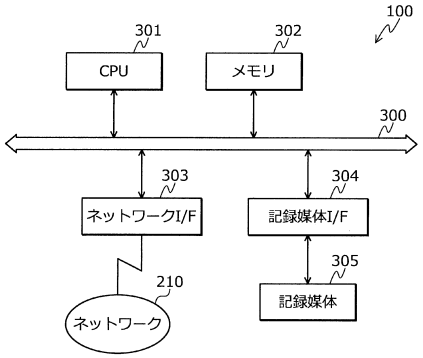
【図１】



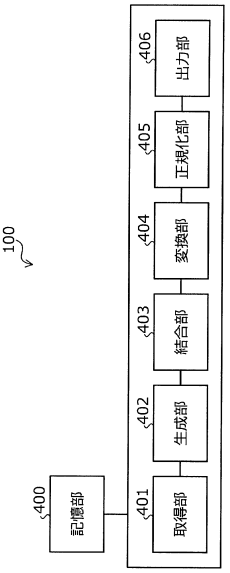
【図２】



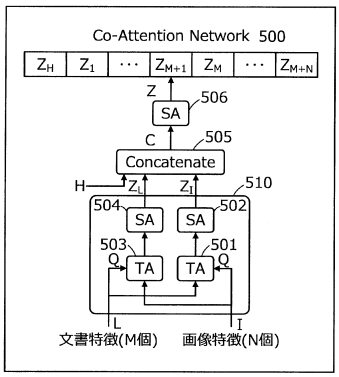
【図３】


100
301
CPU
302
メモリ
300
303
ネットワークI/F
304
記録媒体I/F
305
記録媒体
210
ネットワーク

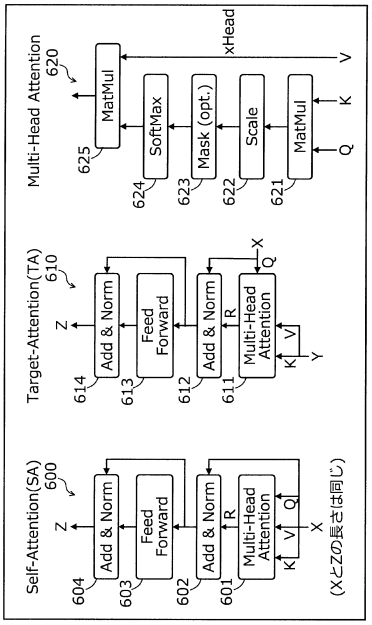
【図４】


100
400
記憶部
401
取得部
402
生成部
403
結合部
404
変換部
405
正規化部
406
出力部

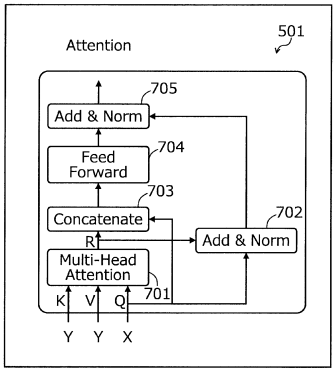
【図５】


Co-Attention Network 500
Z
506
SA
C
505
Concatenate
H
510
504
SA
502
SA
503
TA
501
TA
文書特徴(M個)
画像特徴(N個)

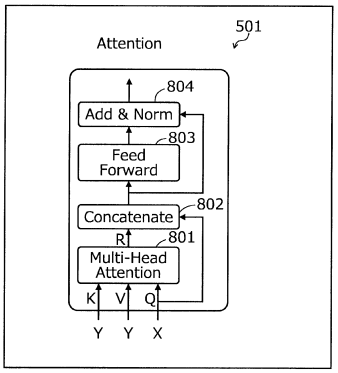
【図６】


Multi-Head Attention
620
xHead
MatMul
SoftMax
Mask (opt.)
Scale
MatMul
625
624
623
622
621
Target-Attention(TA)
610
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
614
613
612
611
Self-Attention(SA)
600
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
604
603
602
601
(XとZの長さは同じ)

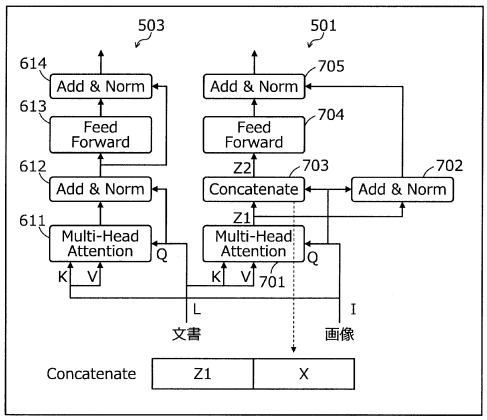
【図７】


Attention
501
705
Add & Norm
704
Feed Forward
703
Concatenate
702
Add & Norm
R
Multi-Head Attention
701
K V Q
Y Y X

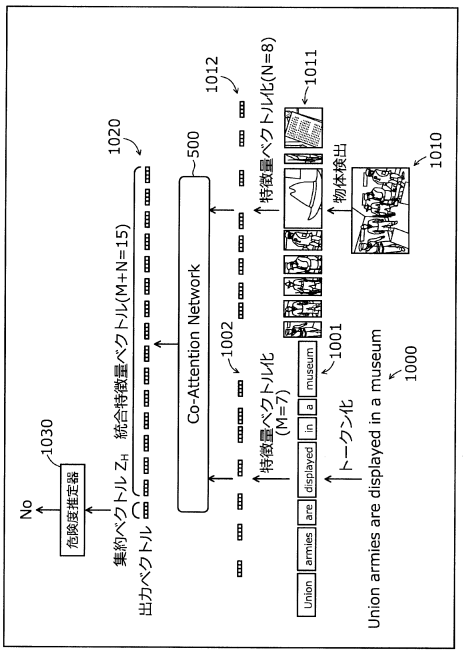
【図８】


Attention
501
804
Add & Norm
803
Feed Forward
802
Concatenate
R
801
Multi-Head Attention
K V Q
Y Y X

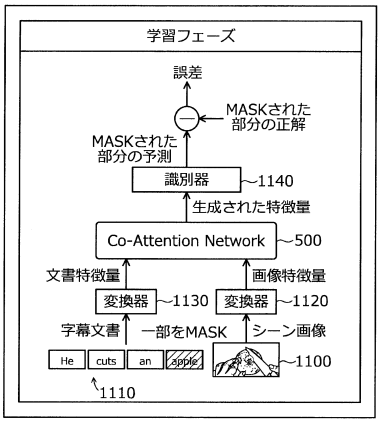
【図９】


503
501
614
Add & Norm
705
Add & Norm
613
Feed Forward
704
Feed Forward
612
Add & Norm
Z2
703
Concatenate
702
Add & Norm
611
Multi-Head Attention
Q
Z1
Multi-Head Attention
Q
K V
K V
701
L
文書
I
画像
Concatenate
Z1
X

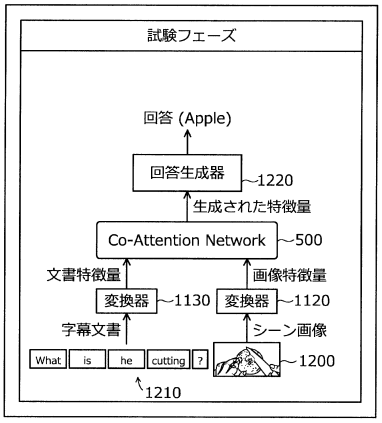
【図１０】


1012
1011
特徴量ベクトル化(N=8)
1020
500
物体検出
1010
Co-Attention Network
1002
特徴量ベクトル化
(M=7)
1001
トークン化
1000
No
1030
危険度推定器
集約ベクトル Z_H 統合特徴量ベクトル(M+N=15)
出力ベクトル
Union armies are displayed in a museum
Union armies are displayed in a museum

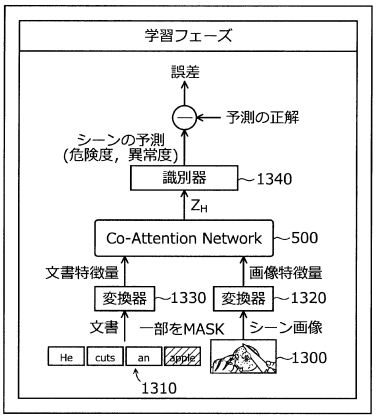
【図１１】



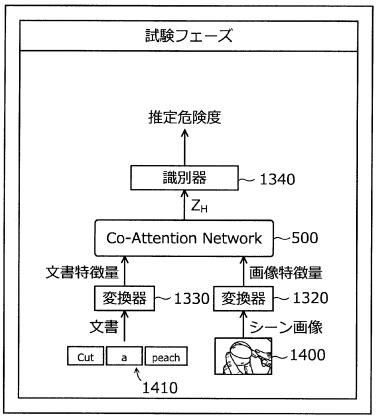
【図１２】



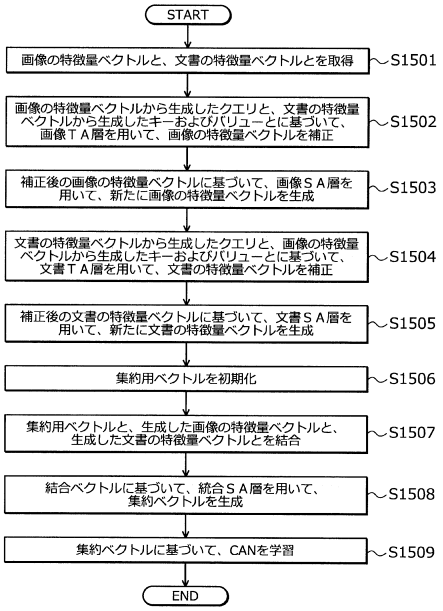
【図１３】



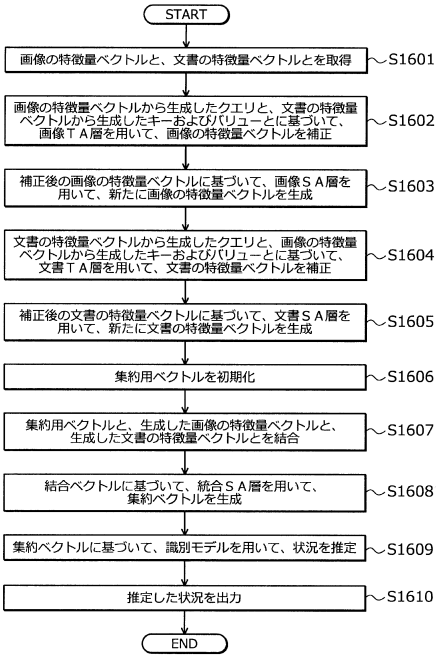
【図１４】



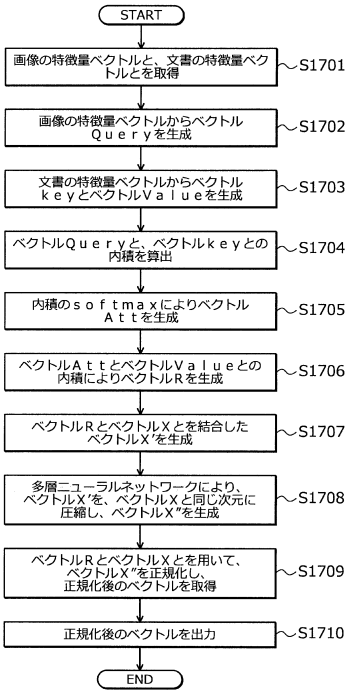
【図１５】


START
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
S1501
画像の特徴量ベクトルから生成したクエリと、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層を用いて、画像の特徴量ベクトルを補正
S1502
補正後の画像の特徴量ベクトルに基づいて、画像ＳＡ層を用いて、新たに画像の特徴量ベクトルを生成
S1503
文書の特徴量ベクトルから生成したクエリと、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層を用いて、文書の特徴量ベクトルを補正
S1504
補正後の文書の特徴量ベクトルに基づいて、文書ＳＡ層を用いて、新たに文書の特徴量ベクトルを生成
S1505
集約用ベクトルを初期化
S1506
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
S1507
結合ベクトルに基づいて、統合ＳＡ層を用いて、集約ベクトルを生成
S1508
集約ベクトルに基づいて、CANを学習
S1509
END

【図１６】


START
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
S1601
画像の特徴量ベクトルから生成したクエリと、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層を用いて、画像の特徴量ベクトルを補正
S1602
補正後の画像の特徴量ベクトルに基づいて、画像ＳＡ層を用いて、新たに画像の特徴量ベクトルを生成
S1603
文書の特徴量ベクトルから生成したクエリと、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層を用いて、文書の特徴量ベクトルを補正
S1604
補正後の文書の特徴量ベクトルに基づいて、文書ＳＡ層を用いて、新たに文書の特徴量ベクトルを生成
S1605
集約用ベクトルを初期化
S1606
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
S1607
結合ベクトルに基づいて、統合ＳＡ層を用いて、集約ベクトルを生成
S1608
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
S1609
推定した状況を出力
S1610
END

【図１７】


START
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
S1701
画像の特徴量ベクトルからベクトルＱｕｅｒｙを生成
S1702
文書の特徴量ベクトルからベクトルｋｅｙとベクトルＶａｌｕｅを生成
S1703
ベクトルＱｕｅｒｙと、ベクトルｋｅｙとの内積を算出
S1704
内積のｓｏｆｔｍａｘによりベクトルＡｔｔを生成
S1705
ベクトルＡｔｔとベクトルＶａｌｕｅとの内積によりベクトルＲを生成
S1706
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
S1707
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
S1708
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
S1709
正規化後のベクトルを出力
S1710
END

フロントページの続き

(56)参考文献 LU, Jiasen, et al.，ViLBERT: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks，arXiv.org [online]，2019年08月06日，pp.1-11，[検索日 2019.12.13], インターネット：〈URL：https://arxiv.org/pdf/1908.02265v1.pdf〉
NGUYEN, Duy-Kien, et al.，Improved fusion of visual and language representations by dense symmetric co-attention for visual qu，[online]，2018年，pp.6087-6096，http://openaccess.thecvf.com/content_cvpr_2018/html/Nguyen_Improved_Fusion_of_CVPR_2018_paper.html，[検索日 2019.12.13], インターネット：〈URL：http://openaccess.thecvf.com/content_cvpr_2018/html/Nguye

(58)調査した分野 (Int.Cl.，ＤＢ名)

Ｇ０６Ｎ　３／００　－９９／００
Ｇ０６Ｔ　７／００　－　７／９０
Ｇ０６Ｖ　１０／００　－２０／９０
Ｇ０６Ｖ　３０／４１８
Ｇ０６Ｖ　４０／１６　、４０／２０