

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7635366号
(P7635366)

(45)発行日 令和7年2月25日(2025.2.25)

(24)登録日 令和7年2月14日(2025.2.14)

(51)国際特許分類	F I	
G 0 6 F 16/28 (2019.01)	G 0 6 F 16/28	
G 0 6 F 16/90 (2019.01)	G 0 6 F 16/90	1 0 0
G 0 6 F 16/36 (2019.01)	G 0 6 F 16/36	

請求項の数 11 (全27頁)

(21)出願番号	特願2023-517266(P2023-517266)	(73)特許権者	522380893
(86)(22)出願日	令和3年12月21日(2021.12.21)		北京沃▲東▼天▲駿▼信息技▲術▼有限 公司
(65)公表番号	特表2023-542665(P2023-542665 A)		Beijing Wodong Tian jun Information Tec hnology Co., Ltd.
(43)公表日	令和5年10月11日(2023.10.11)		中華人民共和国 1 0 0 1 7 6 北京市北 京▲経▼▲済▼技▲術▼▲開▼▲発▼区 科▲創▼十一街 1 8 号院 2 号楼 4 ▲層▼ A 4 0 2 室
(86)国際出願番号	PCT/CN2021/140074		Room A 4 0 2 , 4 / f , No . 2
(87)国際公開番号	WO2022/156468		Building , No . 1 8 Kec huang 1 1 t h Street , E
(87)国際公開日	令和4年7月28日(2022.7.28)		conomic and Technol ogical Development
審査請求日	令和5年3月14日(2023.3.14)		最終頁に続く
(31)優先権主張番号	202110080903.5		
(32)優先日	令和3年1月21日(2021.1.21)		
(33)優先権主張国・地域又は機関	中国(CN)		

(54)【発明の名称】 モデルデータ処理方法、装置、電子機器、コンピュータ可読媒体およびコンピュータプログラム

(57)【特許請求の範囲】

【請求項 1】

コンピュータによって実行されるモデルデータ処理方法であって、
処理対象となるデータセット内の少なくとも2種の異なるモーダルのデータを取得する
ステップと、
前記少なくとも2種の異なるモーダルのデータに対して特徴抽出を行った後にコンカッ
トおよび/または重ね合わせにより特徴系列を取得するステップと、
前記特徴系列に対してモデルマッピング処理を行い、自己回帰モデルに適合するマルチ
モーダルの入力データを取得するステップと、
前記マルチモーダルの入力データを前記自己回帰モデルに入力して、前記自己回帰モデ
ルから出力されたシングルモーダルの結果を取得するステップと、を含み、
前記少なくとも2種の異なるモーダルのデータは、処理対象テキストデータと知識グラ
フデータとを含み、
前記少なくとも2種の異なるモーダルのデータに対して特徴抽出を行った後にコンカッ
トおよび/または重ね合わせにより特徴系列を取得するステップは、
前記知識グラフデータに対応するターナリテキストデータを採集するステップと、
前記ターナリテキストデータおよび前記処理対象テキストデータに対してそれぞれ単語
分割処理を行い、全ての単語分割結果をコンカットして特徴系列を得るステップと、
を含むモデルデータ処理方法。

【請求項 2】

前記モデルデータ処理方法は、

前記特徴系列においてコンカットされた全ての単語分割結果のうちの各単語分割結果の総位置データを確定するステップをさらに含み、

前記特徴系列に対してモデルマッピング処理を行い、自己回帰モデルに適合するマルチモーダルの入力データを取得するステップは、

前記特徴系列に対して自己回帰モデルに適合するベクトルマッピングを行い、前記特徴系列に対応するデータ特徴ベクトルを得るステップと、

前記総位置データに対して前記自己回帰モデルに適合するベクトルマッピングを行い、総位置ベクトルを得るステップと、

前記データ特徴ベクトルと前記総位置ベクトルとを加算して、前記自己回帰モデルに適合するマルチモーダルの入力データを得るステップと、

を含む請求項 1 に記載のモデルデータ処理方法。

【請求項 3】

前記モデルデータ処理方法は、

前記特徴系列においてコンカットされた全ての単語分割結果のうちの各単語分割結果に対して分類処理を行って総分類データを得るステップをさらに含み、

前記特徴系列に対してモデルマッピング処理を行い、自己回帰モデルに適合するマルチモーダルの入力データを取得するステップは、

前記特徴系列に対して自己回帰モデルに適合するベクトルマッピングを行い、前記特徴系列に対応するデータ特徴ベクトルを得るステップと、

前記総分類データに対して前記自己回帰モデルに適合するベクトルマッピングを行い、総分類ベクトルを得るステップと、

前記データ特徴ベクトルと前記総分類ベクトルとを加算して、前記自己回帰モデルに適合するマルチモーダルの入力データを得るステップと、

を含む請求項 1 に記載のモデルデータ処理方法。

【請求項 4】

前記少なくとも 2 種の異なるモーダルのデータは、画像データをさらに含み、

前記少なくとも 2 種の異なるモーダルのデータに対して特徴抽出を行った後にコンカットおよび / または重ね合わせにより特徴系列を取得するステップは、

前記画像データに対して特徴抽出を行って画像特徴を得るステップさらに含み、得られた特徴系列に前記画像特徴がさらに重ね合わせされており、

前記特徴系列に対してモデルマッピング処理を行い、自己回帰モデルに適合するマルチモーダルの入力データを取得するステップは、

前記特徴系列中のコンカットされた全ての単語分割結果に対して自己回帰モデルに適合するベクトルマッピングを行い、コンカットされた全ての単語分割結果に対応するデータ特徴ベクトルを得るステップと、

前記特徴系列中の画像特徴に対して前記自己回帰モデルの入力層と同じ次元のベクトルマッピングを行い、画像特徴ベクトルを得るステップと、

前記データ特徴ベクトルと前記画像特徴ベクトルとを加算して、前記自己回帰モデルに適合するマルチモーダルの入力データを得るステップと、

を含む請求項 1 に記載のモデルデータ処理方法。

【請求項 5】

前記モデルデータ処理方法は、

前記特徴系列においてコンカットされた全ての単語分割結果のうちの各単語分割結果の総位置データを確定するステップと、

前記特徴系列においてコンカットされた全ての単語分割結果のうちの各単語分割結果に対して分類処理を行って総分類データを得るステップとをさらに含み、

前記特徴系列に対してモデルマッピング処理を行い、自己回帰モデルに適合するマルチモーダルの入力データを取得するステップは、

前記総位置データに対して前記自己回帰モデルに適合するベクトルマッピングを行い、

10

20

30

40

50

総位置ベクトルを得るステップと、

前記総分類データに対して前記自己回帰モデルに適合するベクトルマッピングを行い、総分類ベクトルを得るステップと、をさらに含み、

得られた、前記自己回帰モデルに適合するマルチモーダルの入力データは、前記総位置ベクトルと前記総分類ベクトルとをさらに含む

請求項 4 に記載のモデルデータ処理方法。

【請求項 6】

前記少なくとも 2 種の異なるモーダルのデータは、処理対象テキストデータと画像データとを含み、

前記少なくとも 2 種の異なるモーダルのデータに対して特徴抽出を行った後にコンカットおよび / または重ね合わせにより特徴系列を取得するステップは、

前記処理対象テキストデータに対して単語分割処理を行い、全ての単語分割結果をコンカットして分割語系列を得るステップと、

前記画像データに対して特徴抽出を行って画像特徴を得るステップと、

前記分割語系列と前記画像特徴とを重ね合わせして特徴系列を得るステップと、を含み、

前記特徴系列に対してモデルマッピング処理を行い、自己回帰モデルに適合するマルチモーダルの入力データを得るステップは、

前記特徴系列中の分割語系列に対して自己回帰モデルに適合するベクトルマッピングを行い、前記分割語系列のデータ特徴ベクトルを得るステップと、

前記特徴系列中の画像特徴に対して前記自己回帰モデルの入力層と同じ次元のベクトルマッピングを行い、画像特徴ベクトルを得るステップと、

前記データ特徴ベクトルと前記画像特徴ベクトルとを加算して、前記自己回帰モデルに適合するマルチモーダルの入力データを得るステップと、

を含む請求項 1 に記載のモデルデータ処理方法。

【請求項 7】

前記モデルデータ処理方法は、

前記分割語系列において各語の語位置データを確定するステップと、

前記分割語系列における各語を分類処理して、語分類データを得るステップとをさらに含み、

前記特徴系列に対してモデルマッピング処理を行い、自己回帰モデルに適合するマルチモーダルの入力データを取得するステップは、

前記語位置データに対して前記自己回帰モデルに適合するベクトルマッピングを行い、語位置ベクトルを得るステップと、

前記語分類データに対して前記自己回帰モデルに適合するベクトルマッピングを行い、語分類ベクトルを得るステップと、をさらに含み、

得られた、前記自己回帰モデルに適合するマルチモーダルの入力データは、前記語位置ベクトルと前記語分類ベクトルとをさらに含む

請求項 6 に記載のモデルデータ処理方法。

【請求項 8】

モデルデータ処理装置であって、

処理対象となるデータセット内の少なくとも 2 種の異なるモーダルのデータを取得するように構成される獲得ユニットと、

前記少なくとも 2 種の異なるモーダルのデータに対して特徴抽出を行った後にコンカットおよび / または重ね合わせにより特徴系列を取得するように構成される抽出ユニットと、

前記特徴系列に対してモデルマッピング処理を行い、自己回帰モデルに適合するマルチモーダルの入力データを取得するように構成される取得ユニットと、

前記マルチモーダルの入力データを前記自己回帰モデルに入力して、前記自己回帰モデルから出力されたシングルモーダルの結果を取得するように構成される処理ユニットと、を備え、

前記少なくとも 2 種の異なるモーダルのデータは、処理対象テキストデータと知識グラ

10

20

30

40

50

フデータとを含み、

前記抽出ユニットは、

前記知識グラフデータに対応するターナリテキストデータを採集し、

前記ターナリテキストデータおよび前記処理対象テキストデータに対してそれぞれ単語分割処理を行い、全ての単語分割結果をコンカットして特徴系列を得るようにさらに構成されるモデルデータ処理装置。

【請求項 9】

電子機器であって、

1 つまたは複数のプロセッサと、1 つまたは複数のプログラムが格納されている記憶装置と、を備える電子機器であって、

前記 1 つまたは複数のプログラムが前記 1 つまたは複数のプロセッサによって実行されると、前記 1 つまたは複数のプロセッサに請求項 1 ～ 7 のいずれか一項に記載のモデルデータ処理方法を実現させる、電子機器。

【請求項 10】

コンピュータプログラムが格納されているコンピュータ可読媒体であって、

当該コンピュータプログラムがプロセッサによって実行されると、請求項 1 ～ 7 のいずれか 1 項に記載のモデルデータ処理方法を実現する、コンピュータ可読媒体。

【請求項 11】

プロセッサによって実行されると請求項 1 ～ 7 のいずれか 1 項に記載のモデルデータ処理方法が実現されるコンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

< 関連出願の相互参照 >

本開示は 2021 年 1 月 21 日に出願され、出願番号が 202110080903.5 であり、発明の名称が「モデルデータ処理方法、装置、電子機器およびコンピュータ可読媒体」である中国特許出願の優先権を主張し、当該出願の全文は引用により本開示に組み込まれる。

【0002】

本開示は、コンピュータ技術分野に関し、具体的に、人工知能の技術分野、特にモデルデータ処理方法、装置、電子機器、コンピュータ可読媒体およびコンピュータプログラムに関する。

【背景技術】

【0003】

既存の対話システムのデータセットはシングルモーダルデータセットであることが多く、それに対応する対話生成システムの研究はデータセットの制約を受けており、研究されているモデルもシングルモーダルのものである。しかし、実際の応用場面では、対話システムにおいてクライアントから提出された問題は、テキスト情報および画像情報などからなるマルチモーダルのものであることが多い。これは、スマートフォンの広範な応用に伴い、クライアントが気軽に写真を撮ってカスタマーサービスに送信し、写真を通じて問題をより詳細に説明することを望んでいるが、従来の技術では、マルチモーダルデータセットと事前訓練されたシングルモーダル言語モデルとを組み合わせた例はない。

【発明の概要】

【0004】

本開示の実施形態は、モデルデータ処理方法、モデルデータ処理装置、電子機器、コンピュータ可読媒体、およびコンピュータプログラムを提供する。

【0005】

第 1 の態様において、本開示の実施形態は、処理対象となるデータセット内の少なくとも 2 種の異なるモーダルのデータを取得するステップと、少なくとも 2 種の異なるモーダルのデータに対して特徴抽出を行った後に、コンカット (concat) および / または

10

20

30

40

50

重ね合わせにより特徴系列を取得するステップと、特徴系列に対してモデルマッピング処理を行って、自己回帰モデルに適合するマルチモーダルの入力データを取得するステップと、マルチモーダルの入力データを自己回帰モデルに入力して、自己回帰モデルから出力されたシングルモーダルの結果を取得するステップと、を含むモデルデータ処理方法を提供する。

【0006】

いくつかの実施形態では、上記の少なくとも2種の異なるモーダルのデータは、処理対象テキストデータと知識グラフデータとを含み、上記の少なくとも2種の異なるモーダルのデータに対して特徴抽出を行った後にコンカットおよび/または重ね合わせにより特徴系列を取得するステップは、知識グラフデータに対応するターナリテキストデータ (T e r n a r y T e x t D a t a) を採集するステップと、ターナリテキストデータおよび処理対象テキストデータに対してそれぞれ単語分割処理を行い、全ての単語分割結果をコンカットして特徴系列を得るステップとを含む。

10

【0007】

いくつかの実施形態では、上記のモデルデータ処理方法は、特徴系列においてコンカットされた全ての単語分割結果のうちの各単語分割結果の総位置データを確定するステップをさらに含み、上記の特徴系列に対してモデルマッピング処理を行い、自己回帰モデルに適合するマルチモーダルの入力データを取得するステップは、特徴系列に対して自己回帰モデルに適合するベクトルマッピングを行い、特徴系列に対応するデータ特徴ベクトルを得るステップと、総位置データに対して自己回帰モデルに適合するベクトルマッピングを行い、総位置ベクトルを得るステップと、データ特徴ベクトルと総位置ベクトルとを加算して、自己回帰モデルに適合するマルチモーダルの入力データを得るステップと、を含む。

20

【0008】

いくつかの実施形態では、上記のモデルデータ処理方法は、特徴系列においてコンカットされた全ての単語分割結果のうちの各単語分割結果に対して分類処理を行って総分類データを得るステップをさらに含み、上記の特徴系列に対してモデルマッピング処理を行い、自己回帰モデルに適合するマルチモーダルの入力データを取得するステップは、特徴系列に対して自己回帰モデルに適合するベクトルマッピングを行い、特徴系列に対応するデータ特徴ベクトルを得るステップと、総分類データに対して自己回帰モデルに適合するベクトルマッピングを行い、総分類ベクトルを得るステップと、データ特徴ベクトルと総分類ベクトルとを加算して、自己回帰モデルに適合するマルチモーダルの入力データを得るステップと、を含む。

30

【0009】

いくつかの実施形態では、上記の少なくとも2種の異なるモーダルのデータは、画像データをさらに含み、少なくとも2種の異なるモーダルのデータに対して特徴抽出を行った後にコンカットおよび/または重ね合わせにより特徴系列を取得するステップは、画像データに対して特徴抽出を行って画像特徴を得るステップさらに含み、得られた特徴系列に画像特徴がさらに重ね合わせられ、上記の特徴系列に対してモデルマッピング処理を行い、自己回帰モデルに適合するマルチモーダルの入力データを取得するステップは、特徴系列中のコンカットされた全ての単語分割結果に対して自己回帰モデルに適合するベクトルマッピングを行い、コンカットされた全ての単語分割結果に対応するデータ特徴ベクトルを得るステップと、特徴系列中の画像特徴に対して自己回帰モデルの入力層と同じ次元のベクトルマッピングを行い、画像特徴ベクトルを得るステップと、データ特徴ベクトルと画像特徴ベクトルとを加算して、自己回帰モデルに適合するマルチモーダルの入力データを得るステップと、を含む。

40

【0010】

いくつかの実施形態では、上記のモデルデータ処理方法は、特徴系列においてコンカットされた全ての単語分割結果のうちの各単語分割結果の総位置データを確定するステップと、特徴系列においてコンカットされた全ての単語分割結果のうちの各単語分割結果に対

50

して分類処理を行って総分類データを得るステップとをさらに含み、上記の特徴系列に対してモデルマッピング処理を行い、自己回帰モデルに適合するマルチモーダルの入力データを取得するステップは、総位置データに対して自己回帰モデルに適合するベクトルマッピングを行い、総位置ベクトルを得るステップと、総分類データに対して自己回帰モデルに適合するベクトルマッピングを行い、総分類ベクトルを得るステップと、をさらに含み、得られた自己回帰モデルに適合するマルチモーダルの入力データは、総位置ベクトルと総分類ベクトルとをさらに含む。

【 0 0 1 1 】

いくつかの実施形態では、上記の少なくとも2種の異なるモーダルのデータは、処理対象テキストデータと画像データとを含み、上記の少なくとも2種の異なるモーダルのデータに対して特徴抽出を行った後にコンカットおよび／または重ね合わせにより特徴系列を取得するステップは、処理対象テキストデータに対して単語分割処理を行い、全ての単語分割結果をコンカットして分割語系列を得るステップと、画像データに対して特徴抽出を行って画像特徴を得るステップと、分割語系列と画像特徴とを重ね合わせして特徴系列を得るステップと、を含み、上記の特徴系列に対してモデルマッピング処理を行い、自己回帰モデルに適合するマルチモーダルの入力データを得るステップは、特徴系列中の分割語系列に対して自己回帰モデルに適合するベクトルマッピングを行い、分割語系列のデータ特徴ベクトルを得るステップと、特徴系列中の画像特徴に対して自己回帰モデルの入力層と同じ次元のベクトルマッピングを行い、画像特徴ベクトルを得るステップと、データ特徴ベクトルと画像特徴ベクトルとを加算して、自己回帰モデルに適合するマルチモーダル

10

20

【 0 0 1 2 】

いくつかの実施形態では、上記のモデルデータ処理方法は、分割語系列において各語の語位置データを確定するステップと、分割語系列の各語を分類処理して、語分類データを得るステップとをさらに含み、上記の特徴系列に対してモデルマッピング処理を行い、自己回帰モデルに適合するマルチモーダルの入力データを取得するステップは、語位置データに対して自己回帰モデルに適合するベクトルマッピングを行い、語位置ベクトルを得るステップと、語分類データに対して自己回帰モデルに適合するベクトルマッピングを行い、語分類ベクトルを得るステップと、をさらに含み、得られた自己回帰モデルに適合するマルチモーダルの入力データは、語位置ベクトルと語分類ベクトルとをさらに含む。

30

【 0 0 1 3 】

第2の態様において、本開示の実施形態は、モデルデータ処理装置であって、処理対象となるデータセット内の少なくとも2種の異なるモーダルのデータを取得するように構成される獲得ユニットと、少なくとも2種の異なるモーダルのデータに対して特徴抽出を行った後にコンカットおよび／または重ね合わせにより特徴系列を取得するように構成される抽出ユニットと、特徴系列に対してモデルマッピング処理を行い、自己回帰モデルに適合するマルチモーダルの入力データを取得するように構成される取得ユニットと、マルチモーダルの入力データを自己回帰モデルに入力して、自己回帰モデルから出力されたシングルモーダルの結果を取得するように構成される処理ユニットと、を備えるモデルデータ処理装置を提供する。

40

【 0 0 1 4 】

いくつかの実施形態では、上記の少なくとも2種の異なるモーダルのデータは、処理対象テキストデータと知識グラフデータとを含み、上記抽出ユニットは、知識グラフデータに対応するターナリテキストデータ (Ternary Text Data) を採集するように構成される採集モジュールと、ターナリテキストデータおよび処理対象テキストデータに対してそれぞれ単語分割処理を行うように構成される総単語分割モジュールと、全ての単語分割結果をコンカットして特徴系列を得るように構成される総コンカットモジュールとを備える。

【 0 0 1 5 】

いくつかの実施形態では、上記モデルデータ処理装置は、特徴系列においてコンカット

50

された全ての単語分割結果のうちの各単語分割結果の総位置データを確定するように構成される総位置特定ユニットをさらに備え、上記取得ユニットは、特徴系列に対して自己回帰モデルに適合するベクトルマッピングを行い、特徴系列に対応するデータ特徴ベクトルを得るように構成されるデータマッピングモジュールと、総位置データに対して自己回帰モデルに適合するベクトルマッピングを行い、総位置ベクトルを得るように構成される総位置マッピングモジュールと、データ特徴ベクトルと総位置ベクトルとを加算して、自己回帰モデルに適合するマルチモーダルの入力データを得るように構成されるデータ位置加算モジュールと、を備える。

【0016】

いくつかの実施形態では、上記モデルデータ処理装置は、特徴系列においてコンカットされた全ての単語分割結果のうちの各単語分割結果に対して分類処理を行って総分類データを得るように構成される総分類ユニットをさらに備え、上記取得ユニットは、特徴系列に対して自己回帰モデルに適合するベクトルマッピングを行い、特徴系列に対応するデータ特徴ベクトルを得るように構成されるデータマッピングモジュールと、総分類データに対して自己回帰モデルに適合するベクトルマッピングを行い、総分類ベクトルを得るように構成される総分類マッピングモジュールと、データ特徴ベクトルと総分類ベクトルとを加算して、自己回帰モデルに適合するマルチモーダルの入力データを得るように構成されるデータ分類加算モジュールと、を備える。

【0017】

いくつかの実施形態では、上記の少なくとも2種の異なるモーダルのデータは、画像データをさらに含み、上記抽出ユニットはさらに画像データに対して特徴抽出を行って画像特徴を得るように構成される。得られた特徴系列に画像特徴がさらに重ね合わせられる。上記取得ユニットは、特徴系列中のコンカットされた全ての単語分割結果に対して自己回帰モデルに適合するベクトルマッピングを行い、コンカットされた全ての単語分割結果に対応するデータ特徴ベクトルを得るように構成されるデータマッピングモジュールと、特徴系列中の画像特徴に対して自己回帰モデルの入力層と同じ次元のベクトルマッピングを行い、画像特徴ベクトルを得るように構成される画像マッピングモジュールと、データ特徴ベクトルと画像特徴ベクトルとを加算して、自己回帰モデルに適合するマルチモーダルの入力データを得るように構成されるデータ画像加算モジュールと、を備える。

【0018】

いくつかの実施形態では、上記のモデルデータ処理装置は、特徴系列においてコンカットされた全ての単語分割結果のうちの各単語分割結果の総位置データを確定するように構成される総位置特定ユニットと、特徴系列においてコンカットされた全ての単語分割結果のうちの各単語分割結果に対して分類処理を行って総分類データを得るように構成される総分類ユニットとをさらに備え、上記取得ユニットは、総位置データに対して自己回帰モデルに適合するベクトルマッピングを行い、総位置ベクトルを得るように構成される総位置マッピングモジュールと、総分類データに対して自己回帰モデルに適合するベクトルマッピングを行い、総分類ベクトルを得るように構成される総分類マッピングモジュールと、をさらに備え、得られた自己回帰モデルに適合するマルチモーダルの入力データは、総位置ベクトルと総分類ベクトルとをさらに含む。

【0019】

いくつかの実施形態では、上記少なくとも2種の異なるモーダルのデータは、処理対象テキストデータと画像データとを含み、上記抽出ユニットは、処理対象テキストデータに対して単語分割処理を行うように構成される単語分割待ちモジュールと、全ての単語分割結果をコンカットして分割語系列を得るように構成されるコンカット待ちモジュールと、画像データに対して特徴抽出を行って画像特徴を得るように構成される抽出モジュールと、分割語系列と画像特徴とを重ね合わせして特徴系列を得るように構成される重ね合わせモジュールと、を備え、上記取得ユニットは、特徴系列中の分割語系列に対して自己回帰モデルに適合するベクトルマッピングを行い、分割語系列のデータ特徴ベクトルを得るように構成される単語分割結果マッピングモジュールと、特徴系列中の画像特徴に対して自

10

20

30

40

50

己回帰モデルの入力層と同じ次元のベクトルマッピングを行い、画像特徴ベクトルを得るように構成される画像マッピングモジュールと、データ特徴ベクトルと画像特徴ベクトルとを加算して、自己回帰モデルに適合するマルチモーダルの入力データを得るように構成されるデータ・画像加算モジュールと、を備える。

【0020】

いくつかの実施形態では、上記モデルデータ処理装置は、分割語系列において各語の語位置データを確定するように構成される語位置特定ユニットと、分割語系列の各語を分類処理して、語分類データを得るように構成される語分類ユニットとをさらに備え、上記取得ユニットは、語位置データに対して自己回帰モデルに適合するベクトルマッピングを行い、語位置ベクトルを得るように構成される語位置マッピングモジュールと、語分類データに対して自己回帰モデルに適合するベクトルマッピングを行い、語分類ベクトルを得るように構成される語分類マッピングモジュールと、をさらに備え、得られた自己回帰モデルに適合するマルチモーダルの入力データは、語位置ベクトルと語分類ベクトルとをさらに含む。

10

【0021】

第3の態様において、本開示の実施形態は、1つまたは複数のプロセッサと、1つまたは複数のプログラムが格納されている記憶装置と、を備える電子機器であって、1つまたは複数のプログラムが1つまたは複数のプロセッサによって実行されると、1つまたは複数のプロセッサに第1の態様のいずれかの実施形態に記載の方法を実現させる、電子機器を提供する。

20

【0022】

第4の態様において、本開示の実施形態は、コンピュータプログラムが格納されているコンピュータ可読媒体であって、当該コンピュータプログラムがプロセッサによって実行されると、第1の態様のいずれかの実施形態に記載の方法を実現する、コンピュータ可読媒体を提供する。

【0023】

第5の態様、プロセッサによって実行されると、第1の態様のいずれかの実施形態に記載の方法が実行されるコンピュータプログラムを提供する。

【0024】

本開示の実施形態が提供するモデルデータ処理方法および装置は、まず処理対象となるデータセットの少なくとも2種の異なるモーダルのデータを取得する。次に、少なくとも2種の異なるモーダルのデータを特徴抽出した後、コンカットおよび/または重ね合わせにより特徴系列を得る。次に、特徴系列に対してモデルマッピング処理を行って、自己回帰モデルに適合するマルチモーダルの入力データを得る。最後に、マルチモーダルの入力データを自己回帰モデルに入力して、自己回帰モデルから出力されるシングルモーダルの結果を得る。このように、処理対象となるデータセットの少なくとも2種のモーダルデータを処理して得られたマルチモーダルの入力データを自己回帰モデルに入力することにより、自己回帰モデルがシングルモーダルデータ処理だけでなく、マルチモーダルデータ処理もでき、モデル全体が出力するシングルモーダルの結果がより高精度になり、モデルデータの処理能力が向上する。

30

40

【0025】

本開示の他の特徴、目的および利点は、以下の図面を参照してなされる非限定的な実施形態に係る詳細な説明を読むことにより、より明らかになる。

【図面の簡単な説明】

【0026】

【図1】本開示の一実施形態を適用可能な例示的なシステムアーキテクチャを示す図である。

【図2】本開示に係るモデルデータ処理方法の一実施形態を示すフローチャートである。

【図3】本開示に係るモデルデータ処理方法のもう一つの実施形態を示すフローチャートである。

50

【図４】本開示の実際の場面におけるデータ構造図である。

【図５】本開示に係るモデルデータ処理装置の実施形態を示す構造概略図である。

【図６】本開示の実施形態の実現に適する電子機器の構造概略図である。

【発明を実施するための形態】

【００２７】

以下、図面および実施形態を参照しながら本開示をより詳細に説明する。ここで述べている具体的な実施形態は関連発明を説明するためのものにすぎず、当該発明を限定するものではないことを理解すべきである。なお、説明の便宜上、図面には発明に関連する部分のみが示されている。

【００２８】

なお、本開示の実施形態および実施形態における特徴は、矛盾を生じない限り、相互に組み合わせることができる。以下、図面および実施形態を参照しながら本開示を詳細に説明する。

【００２９】

図１は、本開示に係るモデルデータ処理方法が適用可能な例示的なシステムアーキテクチャ１００を示している。

【００３０】

図１に示すように、システムアーキテクチャ１００は、端末装置１０１、１０２、１０３、ネットワーク１０４、およびサーバ１０５を含んでもよい。ネットワーク１０４は、端末装置１０１、１０２、１０３とサーバ１０５との間で通信リンクを提供するための媒体として使用される。ネットワーク１０４は、通常に無線通信リンクなどの様々なタイプの接続を含んでもよい。

【００３１】

メッセージを送受信するために、端末装置１０１、１０２、１０３はネットワーク１０４を介してサーバ１０５と情報のやり取りをする。端末装置１０１、１０２、１０３には、例えば、インスタントコミュニケーションツール、メールボックスクライアント等のような様々な通信クライアントアプリケーションをインストールすることができる。

【００３２】

端末装置１０１、１０２、１０３は、ハードウェアであってもソフトウェアであってもよい。端末装置１０１、１０２、１０３がハードウェアである場合には、サーバ１０５と通信可能な、通信および制御機能を有するユーザデバイスであってもよい。端末装置１０１、１０２、１０３がソフトウェアである場合、上記ユーザデバイスにインストールされてもよい。端末装置１０１、１０２、１０３は、複数のソフトウェアまたはソフトウェアモジュール（例えば、分散サービスを提供するためのソフトウェアまたはソフトウェアモジュール）として実装されてもよく、または単一のソフトウェア若しくはソフトウェアモジュールとして実装されてもよい。ここでは特に限定しない。

【００３３】

サーバ１０５は、様々なサービスを提供するサーバ、例えば、端末装置１０１、１０２、１０３上のクライアントＱ＆Ａシステムにおける処理対象となるデータセットをサポートするモデルデータ処理のバックエンドサーバであってもよい。バックエンドサーバは、ネットワーク上の自己回帰モデルに入力されようとする関連情報を解析処理し、処理結果（例えば、自己回帰モデルから出力されるシングルモーダルの結果）を端末装置にフィードバックすることができる。

【００３４】

なお、サーバは、ハードウェアであってもよく、ソフトウェアであってもよい。サーバがハードウェアである場合、複数のサーバから構成される分散サーバクラスターとしても、単一のサーバとしても実装され得る。サーバがソフトウェアである場合、複数のソフトウェアまたはソフトウェアモジュール（例えば、分散サービスを提供するためのソフトウェアまたはソフトウェアモジュール）として実現されてもよく、または単一のソフトウェアまたはソフトウェアモジュールとして実現されてもよい。ここでは特に限定しない。

10

20

30

40

50

【 0 0 3 5 】

なお、本開示の実施形態によって提供されるモデルデータ処理方法は、通常、サーバ 1 0 5 によって実行される。

【 0 0 3 6 】

図 1 における端末装置、ネットワークおよびサーバの数は例示的なものに過ぎないことを理解すべきである。実装のニーズに応じて、端末装置、ネットワークおよびサーバの数を任意に加減してもよい。

【 0 0 3 7 】

図 2 に示すように、本開示によるモデルデータ処理方法の一実施形態のフロー 2 0 0 を示し、このモデルデータ処理方法は以下のステップを含む。

10

【 0 0 3 8 】

ステップ 2 0 1 では、処理対象となるデータセット内の少なくとも 2 種の異なるモーダルのデータを取得する。

【 0 0 3 9 】

この実施形態では、処理対象となるデータセットは、複数種の異なるモーダルのデータを含み、複数種の異なるモーダルのデータは、複数の異なるソースから由来するシングルモーダルのデータ、または異なる表現形式のシングルモーダルのデータであってもよい。実際には、複数の異なるソースから由来するシングルモーダルのデータは、異なるセンサによって採集された同一物の記述データ（例えば、ビデオ検索時に採集された同一時刻における同一動物の音声データおよび画像データ）を指し、ここでの音声データおよび画像データは、2 種の異なるシングルモーダルのデータである。異なる表現形式のシングルモーダルデータとは、データを通じてデータの表示内容（例えば、タスク型対話コミュニケーションプロセス、ユーザから対話ロボットに送信された画像データ、テキストデータなど）を知ることができるものを指す。

20

【 0 0 4 0 】

この実施形態では、処理対象となるデータセットは、自己回帰モデルの構築または訓練を実現するために予め設定されたデータセットを含んでもよいし、処理対象となるデータセットは、自己回帰モデルの訓練が完了した後に自己回帰モデルが処理する必要のある全てのシングルモーダルデータセットを含んでもよい。

【 0 0 4 1 】

30

自己回帰モデルは、時系列を統計的に扱う方法の一つであり、同一変数の前の各期の表現状況を用いて、その変数の現期の表現状況を予測し、それらが線形関係を有すると仮定する。例えば、自己回帰言語モデルでは、前述の内容に基づいて次にたどる可能性のある単語を予測したり、逆に後述の内容に基づいて前の単語を予測したりする。

【 0 0 4 2 】

実際の応用場面では、例えば、タスク型対話コミュニケーション中に、質問に効果的に答えるために、いくつかの知識ベースの補助は必要となる。例えば、あるショッピングモールのオンラインカスタマーサービスシステムでは、顧客は常に商品のいくつかの属性について質問する必要がある。知識グラフ情報による対話生成は、コンテンツがより正確であり、含まれる情報量がより十分であり、実際の応用場面においてクライアントのニーズを満たすことを可能にする。上記の場面では、少なくとも 2 種の異なるモーダルのデータは、処理対象テキストデータ、画像データ、知識グラフデータのいずれか 2 種または複数種を含んでもよい。

40

【 0 0 4 3 】

本実施形態では、モデルデータ処理方法の実行主体（例えば、図 1 に示すサーバ 1 0 5）は、例えば、ユーザ端末（図 1 に示す端末装置 1 0 1、1 0 2、1 0 3）から処理対象となるデータセットをリアルタイムに採集し、処理対象となるデータセットから少なくとも 2 種の異なるモーダルのデータを抽出するなど、様々な手段で処理対象となるデータセットを取得することができる。あるいは、処理対象となるデータセットをローカルメモリから取得し、処理対象となるデータセットから少なくとも 2 種の異なるモーダルのデータ

50

を抽出する。

【 0 0 4 4 】

ステップ 2 0 2 では、少なくとも 2 種の異なるモーダルのデータに対して特徴抽出を行った後にコンカットおよび / または重ね合わせにより特徴系列を取得する。

【 0 0 4 5 】

本実施形態では、得られた少なくとも 2 種の異なるモーダルのデータがそれぞれの基準を有しているため、データ変換後に統一された形式を有する各シングルモーダルのデータに対して、特徴抽出後に抽出された特徴をコンカットしてコンカットされた系列を得ることができ、例えば、知識ベース中の知識グラフデータは、ヘッダ実体、関係実体、テイル実体のトリプレットとして存在し、データ処理過程において、対話に関わるトリプレットのうちヘッダ実体、関係実体、テイル実体に対応する文字内容を抽出してターナリテキストデータを得ることができ、ターナリテキストデータとテキストデータとが統一された形式を有することにより、抽出された処理対象テキストデータとターナリテキストデータとをコンカットすることができる。

10

【 0 0 4 6 】

したがって、データ変換後に統一データ形式を持たない各シングルモーダルデータについては、特徴抽出後に抽出された特徴を重ね合わせすることができるが、本実施形態での重ね合わせはただ情報の重ね合わせ方式のみであり、両者は別々の特徴ベクトル空間に独立して存在する。例えば、画像データと処理対象テキストデータとを統一データ形式に変換することはできないが、両者は同一物を表しており、テキストデータ形式に比べて画像データ形式の方がよりイメージ性が高く、より豊かに同一物を表現することができる。

20

【 0 0 4 7 】

本実施形態のいくつかのオプション的な実施形態では、少なくとも 2 種の異なるモーダルのデータは、処理対象テキストデータと知識グラフデータとを含む。上記の少なくとも 2 種の異なるモーダルのデータに対して特徴抽出を行った後にコンカットおよび / または重ね合わせにより特徴系列を取得するステップは、知識グラフデータに対応するターナリテキストデータ (T e r n a r y T e x t D a t a) を採集するステップと、ターナリテキストデータおよび処理対象テキストデータに対してそれぞれ単語分割処理を行い、全ての単語分割結果をコンカットして特徴系列を得るステップとを含む。

【 0 0 4 8 】

具体的には、図 4 に示す実施形態において、ターナリテキストデータおよび処理対象テキストデータの単語分割処理の過程を詳細に説明する。

30

【 0 0 4 9 】

このオプション的な実施形態では、少なくとも 2 種の異なるモーダルのデータが処理対象テキストデータと知識グラフデータとを含む場合に、両者から抽出した特徴をコンカットすることにより、データ情報を効率的に豊富化することができる。

【 0 0 5 0 】

あるいは、上記のオプション的な実施形態では、少なくとも 2 種の異なるモーダルのデータは、画像データを含んでもよい。少なくとも 2 種の異なるモーダルのデータに対して特徴抽出を行った後にコンカットおよび / または重ね合わせにより特徴系列を取得するステップは、画像データを特徴抽出して画像特徴を得ることを含み、得られた特徴系列にはさらに画像特徴が重ね合わされる。

40

【 0 0 5 1 】

このオプション的な実施形態では、全ての単語分割結果をコンカットした上で画像特徴を重ね合わせして特徴系列を得ることができ、知識グラフやテキスト以外の形式でデータ情報を豊富化することができる。

【 0 0 5 2 】

本実施形態のいくつかのオプション的な実施形態において、上記少なくとも 2 種の異なるモーダルのデータは、処理対象テキストデータおよび画像データを含み、上記の少なくとも 2 種の異なるモーダルのデータに対して特徴抽出を行った後にコンカットおよび / ま

50

たは重ね合わせにより特徴系列を取得するステップは、
処理対象テキストデータに単語分割処理を施し、得られた全ての単語分割結果をコンカットして分割語系列を得るステップと、画像データに特徴抽出を施して画像特徴を得るステップと、分割語系列と画像特徴を重ね合わせして特徴系列を得るステップとを含む。

【0053】

本実施形態では、分割語系列と画像特徴とは、2つの異なる次元空間における異なるデータであるため、両者の重ね合わせは、両者を重ね合わせする1種の方式のみであり、この重ね合わせ方式によりその特徴系列を得た後、分割語系列と画像特徴とを迅速に分離することができる。

【0054】

このオプション的な実施形態では、少なくとも2種の異なるモダルのデータが処理対象テキストデータと画像データとを含む場合に、両者から抽出された特徴を重ね合わせると、データ情報を効果的に豊富化することができる。

【0055】

ステップ203では、特徴系列に対してモデルマッピング処理を行い、自己回帰モデルに適合するマルチモダルの入力データを得る。

【0056】

本実施形態では、特徴系列をモデルマッピング処理することは、特徴系列を自己回帰モデルの受信できるデータに変換するためであり、すなわち、変換後のデータの次元が自己回帰モデルの入力層の次元と同じになることで、変換後のデータを自己回帰モデルに入力することができ、また、変換後のデータを自己回帰モデルの訓練中のパラメータの変化に応じて調整することもできる。

【0057】

本実施形態では、少なくとも2種の異なるモダルのデータの内容が異なり、得られた特徴系列の内容が異なるため、マルチモダルの入力データの内容も異なる。

【0058】

本実施形態のいくつかのオプション的な実施形態では、少なくとも2種の異なるモダルのデータが、処理対象テキストデータおよび画像データを含む場合、上記の特徴系列に対してモデルマッピング処理を行い、自己回帰モデルに適合するマルチモダルの入力データを得るステップは、

特徴系列中の分割語系列に対して自己回帰モデルに適合するベクトルマッピングを行い、分割語系列のデータ特徴ベクトルを得るステップと、特徴系列中の画像特徴に対して自己回帰モデルの入力層と同じ次元のベクトルマッピングを行い、画像特徴ベクトルを得るステップと、データ特徴ベクトルおよび画像特徴ベクトルを加算し、自己回帰モデルに適合するマルチモダルの入力データを得るステップとを含む。

【0059】

本実施形態において、異なるデータに対して自己回帰モデルに適合するベクトルマッピングを行うステップとは、自己回帰モデルの入力層が変換後の異なるベクトルを受け取ることができるように、異なるデータを自己回帰モデルの入力層と同じ次元のベクトルに変換することを指し、当然ながら、異なる形式のデータを自己回帰モデルの入力層と同じ次元のベクトルに変換する手段が異なってもよい。

【0060】

このオプション的な実施形態では、少なくとも2種の異なるモダルのデータが、処理対象テキストデータと画像データとを含む場合、処理対象テキストデータおよび画像特徴のそれぞれにベクトルマッピングを行うことで、マルチモダルの入力データの内容の多様性を豊かにした。

【0061】

あるいは、少なくとも2種の異なるモダルのデータが、処理対象テキストデータと知識グラフデータとを含む場合には、上記の特徴系列に対してモデルマッピング処理を行って、自己回帰モデルに適合するマルチモダルの入力データを得るステップは、

10

20

30

40

50

特徴系列に対して自己回帰モデルに適合するベクトルマッピングを行い、特徴系列に対応したデータ特徴ベクトルを得るステップを含み、このデータ特徴ベクトルは、得られた自己回帰モデルに適合するマルチモーダルの入力データである。なお、このオプション的な実施形態において、上記特徴系列はすべての単語分割結果をコンカットしたものである。

【0062】

本実施形態のいくつかのオプション的な実施形態では、少なくとも2種の異なるモーダルのデータが、処理対象テキストデータ、知識グラフデータ、および画像データを含む場合、上記の特徴系列に対してモデルマッピング処理を行い、自己回帰モデルに適合するマルチモーダルの入力データを得るステップは、

特徴系列中のコンカットされた全ての単語分割結果に対して、自己回帰モデルに適合するベクトルマッピングを行い、コンカットされた全ての単語分割結果に対応するデータ特徴ベクトルを得るステップと、特徴系列中の画像特徴に対して自己回帰モデルの入力層と同じ次元のベクトルマッピングを行い、画像特徴ベクトルを得るステップと、データ特徴ベクトルと画像特徴ベクトルとを加算して、自己回帰モデルに適合するマルチモーダルの入力データを得るステップとを含む。

10

【0063】

このオプション的な実施形態では、少なくとも2種の異なるモーダルのデータが、処理対象テキストデータ、知識グラフデータおよび画像データを含む場合、処理対象テキストデータおよび知識グラフデータに対して単語分割処理を行って複数の単語分割結果を得、すべての単語分割結果をコンカットして、コンカットされた単語分割結果を得、コンカットされた単語分割結果および画像特徴のそれぞれに対してベクトルマッピングを行うことにより、複数の異なるモーダルのデータに基づいてマルチモーダルの入力データの内容を豊かにする。

20

【0064】

さらに、特徴系列中の特徴のそれぞれが位置およびタイプの特徴を有するので、上記の得られた特徴系列に対して位置アノテーションおよびタイプアノテーションを付与してもよい。

【0065】

本実施形態のいくつかのオプション的な実施形態では、少なくとも2種の異なるモーダルのデータが処理対象テキストデータおよび知識グラフデータを含む場合、上記の方法は、特徴系列においてコンカットされた全ての単語分割結果のうち、各単語分割結果の総位置データを確定するステップをさらに含む。

30

【0066】

上記の特徴系列に対してモデルマッピング処理を行い、自己回帰モデルに適合するマルチモーダルの入力データを取得するステップは、特徴系列に対して自己回帰モデルに適合するベクトルマッピングを行い、特徴系列に対応するデータ特徴ベクトルを得るステップと、総位置データに対して自己回帰モデルに適合するベクトルマッピングを行い、総位置ベクトルを得るステップと、データ特徴ベクトルと総位置ベクトルとを加算して、自己回帰モデルに適合するマルチモーダルの入力データを得るステップと、を含む。

【0067】

このオプション的な実施形態では、コンカットされた全ての単語分割結果における各単語分割結果の位置の順序に基づいて、各単語分割結果に位置データ（例えばコード）を割り当てることができる。総位置データは、全ての単語分割結果のうちの各単語分割結果の位置データを含む集合であり、位置データに基づいて、コンカットされた全ての単語分割結果における各単語分割結果の順序値または表現の情報を容易かつ迅速に特定することができる。

40

【0068】

このオプション的な実施形態では、単語分割結果の総位置データを確定することで、特徴系列中の各単語分割結果を効率的に位置特定することができ、マルチモーダルの入力データの位置特徴を豊かにした。

50

【 0 0 6 9 】

本実施形態のいくつかのオプション的な実施形態では、少なくとも2種の異なるモダルのデータが処理対象テキストデータおよび知識グラフデータを含む場合、上記の方法は、特徴系列においてコンカットされた全ての単語分割結果のうちの各単語分割結果に対して分類処理を行って、総分類データを得るステップをさらに含む。上記の特徴系列に対してモデルマッピング処理を行い、自己回帰モデルに適合するマルチモダルの入力データを取得するステップは、特徴系列に対して自己回帰モデルに適合するベクトルマッピングを行い、特徴系列に対応するデータ特徴ベクトルを得るステップと、総分類データに対して自己回帰モデルに適合するベクトルマッピングを行い、総分類ベクトルを得るステップと、データ特徴ベクトルと総分類ベクトルとを加算して、自己回帰モデルに適合するマルチモダルの入力データを取得するステップと、を含む。

10

【 0 0 7 0 】

このオプション的な実施形態では、コンカットされた全ての単語分割結果における各単語分割結果のタイプに基づいて、各単語分割結果に分類データ（例えば、コード）を割り当てることができる。総分類データは、全ての単語分割結果のうちの各単語分割結果の分類データを含む集合であり、分類データに基づいて、コンカットされた全ての単語分割結果における各単語分割結果のタイプを容易かつ迅速に特定することができる。

【 0 0 7 1 】

このオプション的な実施形態では、単語分割結果の総分類データを確定することで、特徴系列中の各単語分割結果の所属するタイプを効率的に特定することができ、マルチモダルの入力データのタイプ特徴を豊かにした。

20

【 0 0 7 2 】

あるいは、少なくとも2種の異なるモダルのデータが処理対象テキストデータおよび知識グラフデータを含む場合、上記の方法は、特徴系列におけるコンカットされた全ての単語分割結果のうちの各単語分割結果の総位置データを確定するステップと、特徴系列においてコンカットされた全ての単語分割結果のうちの各単語分割結果に対して分類処理を行って、総分類データを得るステップをさらに含む。上記の特徴系列に対してモデルマッピング処理を行い、自己回帰モデルに適合するマルチモダルの入力データを取得するステップは、特徴系列に対して自己回帰モデルに適合するベクトルマッピングを行い、特徴系列に対応するデータ特徴ベクトルを得るステップと、総位置データに対して自己回帰モデルに適合するベクトルマッピングを行い、総位置ベクトルを得るステップと、総分類データに対して自己回帰モデルに適合するベクトルマッピングを行い、総分類ベクトルを得るステップと、データ特徴ベクトルと、総位置ベクトルと、総分類ベクトルとを加算して、自己回帰モデルに適合するマルチモダルの入力データを取得するステップと、を含む。

30

【 0 0 7 3 】

本実施形態のいくつかのオプション的な実施形態において、少なくとも2種の異なるモダルのデータが処理対象テキストデータおよび画像データを含む場合、上記の方法は、分割語系列の各語の語位置データを確定するステップと、分割語系列の各語を分類処理して、語分類データを得るステップとをさらに含み、特徴系列に対してモデルマッピング処理を行い、自己回帰モデルに適合するマルチモダルの入力データを取得するステップは、特徴系列中の分割語系列に対して自己回帰モデルに適合するベクトルマッピングを行い、分割語系列のデータ特徴ベクトルを得るステップと、特徴系列中の画像特徴に対して自己回帰モデルの入力層と同じ次元のベクトルマッピングを行い、画像特徴ベクトルを得るステップと、語位置データに対して自己回帰モデルに適合するベクトルマッピングを行い、語位置ベクトルを得るステップと、語分類データに対して自己回帰モデルに適合するベクトルマッピングを行い、語分類ベクトルを得るステップと、データ特徴ベクトル、語位置ベクトル、語分類ベクトルおよび画像特徴ベクトルを加算し、自己回帰モデルに適合するマルチモダルの入力データを取得するステップとを含む。

40

【 0 0 7 4 】

このオプション的な実施形態では、分割語系列中の各語の語位置データおよび語分類デ

50

ータを確定することにより、分割語系列中の各語の所属するタイプおよび各語の位置を効率的に確定することができ、マルチモーダルの入力データの位置特徴とタイプ特徴を豊かにした。

【0075】

ステップ204では、マルチモーダルの入力データを自己回帰モデルに入力して、自己回帰モデルから出力されたシングルモーダルの結果を取得する。

【0076】

本実施形態では、自己回帰モデルは、出力データがシングルモーダルデータである自己回帰モデル、例えばUniLM(Unified Language Model Pre-training for Natural Language Understanding and Generation, 自然言語理解・生成モデル)、GPT(Generative Pre-Training, 事前訓練生成モデル)等のモデルとしてもよい。自己回帰モデルは、新たに構築されたモデル(例えば、入力側がマルチモーダルデータ、出力側がシングルモーダルデータ)であってもよい。

【0077】

あるいは、自己回帰モデルは、事前訓練済みの自己回帰モデルであってもよく、当該自己回帰モデルの入出力はいずれもシングルモーダルのデータであり、例えば、当該自己回帰モデルは言語モデルであり、その入出力はいずれもテキストデータであり、その言語モデルに対して適用される処理対象となるデータセットは、対話記録または文章の章もしくは段落などのデータであってもよい。

【0078】

上記マルチモーダルの入力データを自己回帰モデルに入力した後、この自己回帰モデルをモデル訓練することにより、訓練済みの自己回帰モデルが得られる。

【0079】

Q & Aタイプの自己回帰モデルに対しては、訓練済みの自己回帰モデルにユーザの質問(画像、知識ベースなどのマルチモーダルデータを有する)を入力し、ユーザの質問に対する回答を得ることができる。

【0080】

実際の応用場面では、自己回帰モデルとしてUniLMまたはDialogPTモデルなどのTransformer自己回帰モデル構造を採用してもよく、本開示の方法により、自己回帰モデルは、従来のテキストデータを処理できるだけでなく、画像、知識グラフ、テキストなどの情報を含むマルチモーダルデータを処理することもでき、自己回帰モデル全体として、マルチモーダルを有する処理対象となるデータセットに対して比較的良好的な処理効果を得ることができる。

【0081】

本開示の実施形態が提供するモデルデータ処理方法は、まず処理対象となるデータセット内の少なくとも2種の異なるモーダルのデータを取得する。次に、少なくとも2種の異なるモーダルのデータを特徴抽出した後、コンカットおよび/または重ね合わせにより特徴系列を得る。次に、特徴系列に対してモデルマッピング処理を行って、自己回帰モデルに適合するマルチモーダルの入力データを得る。最後に、マルチモーダルの入力データを自己回帰モデルに入力して、自己回帰モデルから出力されるシングルモーダルの結果を得る。このように、処理対象となるデータセット内の少なくとも2種のモーダルデータを処理して得られたマルチモーダルの入力データを自己回帰モデルに入力することにより、自己回帰モデルがシングルモーダルデータ処理だけでなく、マルチモーダルデータ処理もでき、モデル全体が出力するシングルモーダルの結果がより高精度になり、モデルデータの処理能力が向上した。

【0082】

本開示のもう一つの実施形態では、少なくとも2種の異なるモーダルのデータは、処理対象テキストデータ、知識グラフデータ、および画像データを含み、図3に示すように、本開示によるモデルデータ処理方法のもう一つの実施形態のフロー300を示しており、

このモデルデータ処理方法は以下のステップを含む。

【 0 0 8 3 】

ステップ 3 0 1 では、処理対象となるデータセット内の処理対象テキストデータ、知識グラフデータおよび画像データを取得する。

【 0 0 8 4 】

図 4 に示すように、処理対象となるデータセットは、あるデパートのカスタマーサービスシステムなどの対話システムから切り出されて処理された対話データであってもよく、この対話データは、テキストデータ、画像データを含み、予め作成された知識グラフから対話内容の知識グラフデータを得ることもできる。なお、処理対象テキストデータに対応するテキストは、少なくとも 1 つの文からなり、各文は記号により分割されていてもよい。

10

【 0 0 8 5 】

図 4 において、処理対象テキストデータに対応する処理すべきテキストには、1 つの質問文と 1 つの回答文が含まれており、質問文は：スケートボードスニーカーはありますか？このうち、質問文の中のスケートボードスニーカーは、ユーザが送付した一枚の画像に表示される。回答文は、このスタイルのスケートボードスニーカーですか。このうち、回答文の中のスケートボードスニーカーは、実行主体が送付した別の画像の中に表示される。

【 0 0 8 6 】

ステップ 3 0 2 では、知識グラフデータに対応するターナリテキストデータを採集する。

【 0 0 8 7 】

本実施形態では、知識グラフデータは、ヘッダ実体、関係実体、テイル実体のトリプレットの形式で格納されており、知識グラフデータを抽出した後に、知識グラフデータのヘッダ実体、関係実体、テイル実体を容易に得ることができる。

20

【 0 0 8 8 】

図 4 において、知識グラフを検索することにより、得られた知識グラフデータは、スケートボードスニーカー—材質—キャンバス生地、スケートボードスニーカー—スリッポン／バンド／シューレースタイプ—シューレース、スケートボードスニーカー—アッパー—ローカット。ターナリテキストデータに対応するターナリテキストは、それぞれスケートボードスニーカー—材質—キャンバス生地、スケートボードスニーカー—スリッポン／バンド／シューレースタイプ—シューレース、スケートボードスニーカー—アッパー—ローカットである。

30

【 0 0 8 9 】

ステップ 3 0 3 では、ターナリテキストデータおよび処理対象テキストデータに対してそれぞれ単語分割処理を行う。

【 0 0 9 0 】

本実施形態では、従来の単語分割ツールにより、処理対象テキストデータに対応する処理対象テキストと、ターナリテキストデータに対応するターナリテキストとをそれぞれ単語分割することができる。

【 0 0 9 1 】

具体的には、ターナリテキスト単語分割に対して、 $S_{11}, S_{12}, \dots, S_{1n}, P_{11}, P_{12}, \dots, P_{1n}, O_{11}, O_{12}, \dots, O_{1n}, S_{21}, S_{22}, \dots, S_{2n}, P_{21}, P_{22}, \dots, P_{2n}, O_{21}, O_{22}, \dots, O_{2n}, \dots, S_{n1}, S_{n2}, \dots, S_{nm}, P_{n1}, P_{n2}, \dots, P_{nm}, O_{n1}, O_{n2}, \dots, O_{nm}$ のような単語分割処理された知識表現を得ることができ、ここで S_{nm} は第 n ($n > 0, m > 0$) 個のトリプレット中のヘッダ実体での単語分割結果による m 番目のタグを示し、 P_{nm} は第 n 個のトリプレット中の関係実体での単語分割結果による m 番目のタグを示し、 O_{nm} は第 n 個のトリプレット中のテイル実体での単語分割結果による m 番目のタグを示す。

40

【 0 0 9 2 】

具体的には、本実施形態では、処理対象テキストデータに対応するコンテキスト中の各文を単語分割して、 $W_{n1}, W_{n2}, W_{n3}, \dots, W_{nm}$ を得る。ここで、 W_{nm} は、第 n ($n > 0, m > 0$) 番目の文における単語分割結果の後の m 番目のタグを示す。実際の単

50

語分割では、特別な意味の文字記号、例えばURL (Uniform Resource Location, ユニフォームリソースロケーション) リンク、表情記号文字表現などの特別な意味のテキストは、特別なタグで置き換えることができ、それによって文の意味を最大限に保つことができる。

【 0 0 9 3 】

ステップ 3 0 4 では、全ての単語分割結果をコンカットして、コンカットされた全ての単語分割結果を得る。

【 0 0 9 4 】

図 4 に示すように、特殊記号 [C L S] を文頭として、上記知識グラフデータのタグ (知識タグと略称する)、処理対象テキスト中の各文の単語分割結果タグを順次コンカットする。このうち、知識グラフデータと処理対象テキストとの間、および文と文の間には、異なる内容のタグの間隔を表すための特殊記号 [S E P] を付加する必要がある。

10

【 0 0 9 5 】

ステップ 3 0 5 では、画像データに対して特徴抽出を行って画像特徴を得る。

【 0 0 9 6 】

本実施形態では、処理対象となるデータセットに示された画像に対して、事前訓練済みの特徴採集モデル (例えば、R e s N e t 5 0) を用いて画像の特徴抽出を行うことができる。

【 0 0 9 7 】

ステップ 3 0 6 では、コンカットされた全ての単語分割結果と画像特徴とを重ね合わせして特徴系列を得る。

20

【 0 0 9 8 】

本実施形態では、コンカットされた全ての単語分割結果と画像特徴とを重ね合わせするようにしたのは、全体の特徴系列を後続きの処理に便利に適用するために、統一した全体的な特徴系列を取得するためである。

【 0 0 9 9 】

ステップ 3 0 7 では、特徴系列においてコンカットされた全ての単語分割結果のうちの各単語分割結果の総位置データを確定する。

【 0 1 0 0 】

本実施形態では、総位置データの確定は、特徴系列中の各単語分割結果に対して位置コードの割り当てを行い、図 4 に示すように、文頭に位置する [C L S] の位置データを 0 とし、すべての知識タグの位置データを 1 とし、その他の非知識タグを順に 2 から順に遞増し、図 4 のように、得られた総位置データはそれぞれ、0 1 1 1 1 1 1 1 1 2 3 4 5 6 7 8 9 1 0 1 1 1 2 である。

30

【 0 1 0 1 】

ステップ 3 0 8 では、特徴系列においてコンカットされた全ての単語分割結果のうちの各単語分割結果に対して分類処理を行って総分類データを得る。

【 0 1 0 2 】

本実施形態では、コンカットタグの異なるグループを区別するために、図 4 に示すように、特に、異なるタグを示すカテゴリをいくつか導入し、[C L S] カテゴリは、そのタグが文のヘッダタグであることを示し、[S E P] カテゴリは、そのタグが知識と文、または文と文の区切りであることを示し、[S U B] カテゴリは、そのタグが知識トリプレットに所属するヘッダ実体であることを示し、[P R E] カテゴリは、そのタグが知識トリプレットに所属する関係実体であることを示し、[O B J] カテゴリは、そのタグが知識トリプレットに所属するテイル実体であることを示す。[S P 1] カテゴリは、そのタグが対話中の役 1 で述べた内容に所属することを示し、[S P 2] カテゴリは、そのタグが対話中の役 2 で述べた内容に所属することを示す。

40

【 0 1 0 3 】

ステップ 3 0 9 では、特徴系列中のコンカットされた全ての単語分割結果に対して自己回帰モデルに適合するベクトルマッピングを行い、コンカットされた全ての単語分割結果

50

に対応するデータ特徴ベクトルを得る。

【0104】

本実施形態におけるベクトルマッピングとは、1つの特徴データを自己回帰モデルに適合する特徴ベクトルにマッピングして融合することを指す。

【0105】

本実施形態では、ステップ303で生成された知識タグ(S_{nm} 、 P_{nm} 、 O_{nm})と、テキストタグ W_{nm} と、ステップ304で生成された $[CLS]$ 、 $[SEP]$ との三者はセマンティック上で一致しているため、同一語彙リストと対応するベクトルマッピングを用いてデータ特徴ベクトルを得、得られたデータ特徴ベクトルの次元は、自己回帰モデルの入力層の次元と同じである。

10

【0106】

実際には、`tf.nn.embedding_lookup(params, ids)`関数によってデータ特徴ベクトルを得ることができ、この関数は主にテンソル内のインデックスに対応する要素を選択し、ここで`params`はテンソルでも配列でもよく、`ids`は対応するインデックス値でもあり、自己回帰モデルに対応するベクトルでもある。

【0107】

ステップ310では、特徴系列中の画像特徴に対して自己回帰モデルの入力層と同じ次元のベクトルマッピングを行い、画像特徴ベクトルを得る。

【0108】

本実施形態では、特徴系列中の画像特徴に対して、線形回帰モデル(`Linear`)を用いて画像特徴を処理することにより、自己回帰モデルの入力層と同じ次元である指定された次元に画像特徴をマッピングすることができる。

20

【0109】

なお、自己回帰モデルは、問答モデルなどの言語モデルであってもよく、自己回帰モデルの入力層の次元は、その中の`embedding`(埋め込み層)の次元によって定められ、`embedding`の次元は、スーパーパラメータに属し、例えば、300または512次元であってもよい。

【0110】

ステップ311では、総位置データに対して自己回帰モデルに適合するベクトルマッピングを行い、総位置ベクトルを得る。

30

【0111】

本実施形態において、総位置ベクトルは、従来の`Transformer`の三角関数を用いた位置符号化方法により求めることができ、それによって対応する総位置データを指定された次元の連続空間にマッピングする。

【0112】

ステップ312では、総分類データに対して自己回帰モデルに適合するベクトルマッピングを行い、総分類ベクトルを得る。

【0113】

本実施形態では、ステップ308で得られた7種の異なるカテゴリタグを、1つの語彙リストと対応するベクトルマッピングを用いて、7種の異なるカテゴリタグを指定された次元の連続空間にマッピングする。

40

【0114】

ステップ313では、データ特徴ベクトルと、画像特徴ベクトルと、総位置ベクトルと、総分類ベクトルとを加算して、自己回帰モデルに適合するマルチモーダルの入力データを得る。

【0115】

本実施形態では、上記の処理を経て、データ特徴ベクトル、画像特徴ベクトル、総位置ベクトル、総分類ベクトルがいずれも指定された次元にマッピングされているので、これら4種類のベクトルを加算することにより、自己回帰モデルの入力データを得ることができる。

50

【0116】

ステップ314では、マルチモーダルの入力データを自己回帰モデルに入力して、自己回帰モデルから出力されたシングルモーダルの結果を取得する。

【0117】

本実施形態において、自己回帰モデルは、Transformer（エンコーダとデコーダからなるモデル構造）に基づく自己回帰言語モデルであってもよい。図4に示すように、対話のコンテキスト、画像、知識情報をデータ処理してマルチモーダルの入力データを得て、自己回帰言語モデルに入力し、モデルを訓練することにより、ユーザの質問応答を生成する際に、マルチモーダル情報と知識情報を利用することができ、このように生成した回答がよりスムーズで正確となる。

10

【0118】

本実施形態では、知識グラフデータ、画像データおよびテキストデータの三者の情報を効率的に融合することにより、対話のコンテキスト情報を十分に表現でき、自己回帰構造に基づくモデルによりエンドツーエンドで訓練を行い、スムーズで情報量に富んだ対話応答を生成できる。

【0119】

本実施形態のモデルデータ処理方法は、少なくとも2種の異なるモーダルのデータが処理対象テキストデータ、知識グラフデータおよび画像データを含む場合に、画像データの特徴から抽出された画像特徴を、処理対象テキストデータと知識グラフデータとがコンカットされた全ての単語分割結果に重ね合わせして、特徴系列を得る。コンカットされた全ての単語分割結果の特徴系列の中でコンカットされた全ての単語分割結果のうちの各単語分割結果の総位置データを確定し、コンカットされた全ての単語分割結果の特徴系列の中でコンカットされた全ての単語分割結果のうちの各単語分割結果を分類処理して、総分類データを得る。総位置データ、総分類データ、コンカットされた全ての単語分割結果に対して自己回帰モデルに適合するベクトルマッピングを行い、それぞれデータ特徴ベクトル、画像特徴ベクトル、総位置ベクトル、総分類ベクトルを得る。画像データを自己回帰モデルの入力層と同じ次元のベクトルマッピングを行い、画像特徴ベクトルを得、データ特徴ベクトルと、画像特徴ベクトルと、総位置ベクトルと、総分類ベクトルとを加算して、自己回帰モデルに適合するマルチモーダルの入力データを得る。これにより、知識ベース、テキスト、画像などの異なるモーダルのデータをカバーすることができ、知識の異なるモーダルのデータを効果的に位置特定・分類することができ、得られたマルチモーダルの入力データの正確性が向上し、自己回帰モデルから出力されたシングルモーダルの結果がよりスムーズでかつ正確となる。

20

30

【0120】

更に図5を参照すると、上記の各図に示された方法の実施態様として、本開示は、モデルデータ処理装置の一実施形態を提供し、当該装置の実施形態は、図2に示された方法の実施形態に対応しており、当該装置は、具体的に様々な電子機器に適用することができる。

【0121】

図5に示すように、本開示の実施形態に係るモデルデータ処理装置500を提供し、当該装置500は、獲得ユニット501と、抽出ユニット502と、取得ユニット503と、処理ユニット504とを備える。獲得ユニット501は、処理対象となるデータセット内の少なくとも2種の異なるモーダルのデータを取得するように構成されてもよい。抽出ユニット502は、少なくとも2種の異なるモーダルのデータに対して特徴抽出を行った後にコンカットおよび/または重ね合わせにより特徴系列を取得するように構成されてもよい。取得ユニット503は、特徴系列に対してモデルマッピング処理を行い、自己回帰モデルに適合するマルチモーダルの入力データを得るように構成されてもよい。処理ユニット504は、マルチモーダルの入力データを自己回帰モデルに入力して、自己回帰モデルから出力されたシングルモーダルの結果を得るように構成されてもよい。

40

【0122】

本実施形態において、モデルデータ処理装置500における獲得ユニット501と、抽

50

出ユニット 5 0 2 と、取得ユニット 5 0 3 と、処理ユニット 5 0 4 の具体的な処理およびそれらによって奏される技術的效果は、それぞれ図 2 の対応する実施形態におけるステップ 2 0 1、ステップ 2 0 2、ステップ 2 0 3、およびステップ 2 0 4 を参照することができる。

【 0 1 2 3 】

いくつかの実施形態では、上記の少なくとも 2 種の異なるモダルのデータは、処理対象テキストデータと知識グラフデータとを含み、上記抽出ユニット 5 0 2 は、採集モジュール（図示せず）、総単語分割モジュール（図示せず）、および総コンカットモジュール（図示せず）を含む。このうち、採集モジュールは、知識グラフデータに対応するターナリテキストデータを採集するように構成されてもよい。総単語分割モジュールは、ターナリテキストデータおよび処理対象テキストデータのそれぞれに対して単語分割処理を行うように構成されてもよい。総コンカットモジュールは、全ての単語分割結果をコンカットして特徴系列を得るように構成されてもよい。

10

【 0 1 2 4 】

いくつかの実施形態では、上記装置 5 0 0 は、特徴系列においてコンカットされた全ての単語分割結果のうちの各単語分割結果の総位置データを確定するように構成され得る総位置特定ユニット（図示せず）をさらに含む。上記取得ユニット 5 0 3 は、データマッピングモジュール（図示せず）と、総位置マッピングモジュール（図示せず）と、データ・位置加算モジュール（図示せず）とを含む。このうち、データマッピングモジュールは、特徴系列に対して自己回帰モデルに適合するベクトルマッピングを行い、特徴系列に対応するデータ特徴ベクトルを得るように構成されてもよい。総位置マッピングモジュールは、総位置データに対して自己回帰モデルに適合するベクトルマッピングを行い、総位置ベクトルを得るように構成されてもよい。データ・位置加算モジュールは、データ特徴ベクトルと総位置ベクトルとを加算して、自己回帰モデルに適合するマルチモダルの入力データを得るように構成されてもよい。

20

【 0 1 2 5 】

いくつかの実施形態では、上記装置 5 0 0 は、特徴系列内のコンカットされた全ての単語分割結果のうちの各単語分割結果に対して分類処理を行い、総分類データを得るように構成され得る総分類ユニット（図示せず）をさらに備える。上記取得ユニット 5 0 3 は、データマッピングモジュール（図示せず）と、総分類マッピングモジュール（図示せず）と、データ・分類加算モジュール（図示せず）とを備える。このうち、データマッピングモジュールは、特徴系列に対して自己回帰モデルに適合するベクトルマッピングを行い、特徴系列に対応するデータ特徴ベクトルを得るように構成されてもよい。総分類マッピングモジュールは、総分類データに対して自己回帰モデルに適合するベクトルマッピングを行い、総分類ベクトルを得るように構成されてもよい。データ・分類加算モジュールは、データ特徴ベクトルと総分類ベクトルとを加算して、自己回帰モデルに適合するマルチモダルの入力データを得るように構成されてもよい。

30

【 0 1 2 6 】

いくつかの実施形態では、上記の少なくとも 2 種の異なるモダルのデータは、画像データをさらに含む。上記抽出ユニット 5 0 2 は、画像データに対して特徴抽出を行って画像特徴を得るように構成されてもよい。本実施形態では、得られた特徴系列にさらに画像特徴が重ね合わせされている。上記取得ユニット 5 0 3 は、データマッピングモジュール（図示せず）と、画像マッピングモジュール（図示せず）と、データ・画像加算モジュール（図示せず）とを備える。このうち、データマッピングモジュールは、特徴系列中のコンカットされた全ての単語分割結果に対して自己回帰モデルに適合するベクトルマッピングを行い、コンカットされた全ての単語分割結果に対応するデータ特徴ベクトルを得るように構成されてもよい。画像マッピングモジュールは、特徴系列中の画像特徴に対して、自己回帰モデルの入力層と同じ次元のベクトルマッピングを行い、画像特徴ベクトルを得るように構成されてもよい。データ・画像加算モジュールは、データ特徴ベクトルと画像特徴ベクトルとを加算して、自己回帰モデルに適合するマルチモダルの入力データを得

40

50

るように構成されてもよい。

【0127】

いくつか実施形態では、上記装置500は、総位置特定ユニット（図示せず）と、総分類ユニット（図示せず）とをさらに備える。総位置特定ユニットは、特徴系列においてコンカットされた全ての単語分割結果のうちの各単語分割結果の総位置データを特定するように構成されてもよい。総分類ユニットは、特徴系列においてコンカットされた全ての単語分割結果のうちの各単語分割結果に対して分類処理を行って総分類データを得るように構成されてもよい。上記取得ユニット503は、総位置マッピングモジュール（図示せず）と、総分類マッピングモジュール（図示せず）とをさらに備える。ただし、総位置マッピングモジュールは、総位置データに対して自己回帰モデルに適合するベクトルマッピングを行い、総位置ベクトルを得るように構成される。総分類マッピングモジュールは、総分類データに対して自己回帰モデルに適合するベクトルマッピングを行い、総分類ベクトルを得るように構成される。本実施形態では、得られた自己回帰モデルに適合するマルチモーダルの入力データは、総位置ベクトルと総分類ベクトルとをさらに含む。

10

【0128】

いくつかの実施形態では、上記の少なくとも2種の異なるモーダルのデータは、処理対象テキストデータおよび画像データを含む。上記抽出ユニット502は、単語分割待ちモジュール（図示せず）と、コンカット待ちモジュール（図示せず）と、抽出モジュール（図示せず）と、重ね合わせモジュール（図示せず）とを備える。このうち、単語分割待ちモジュールは、処理対象テキストデータに対して単語分割処理を行うように構成されてもよい。コンカット待ちモジュールは、全ての単語分割結果をコンカットして、分割語系列を得るように構成されてもよい。抽出モジュールは、画像データに対して特徴抽出を行って画像特徴を得るように構成されてもよい。重ね合わせモジュールは、分割語系列と画像特徴とを重ね合わせして特徴系列を得るように構成されてもよい。上記取得ユニット503は、単語分割結果マッピングモジュール（図示せず）と、画像マッピングモジュール（図示せず）と、語・画像加算モジュール（図示せず）とを備える。このうち、単語分割結果マッピングモジュールは、特徴系列中の分割語系列に対して自己回帰モデルに適合するベクトルマッピングを行い、分割語系列のデータ特徴ベクトルを得るように構成されてもよい。画像マッピングモジュールは、特徴系列中の画像特徴に対して、自己回帰モデルの入力層と同じ次元のベクトルマッピングを行い、画像特徴ベクトルを得るように構成されてもよい。語・画像加算モジュールは、データ特徴ベクトルと画像特徴ベクトルとを加算して、自己回帰モデルに適合するマルチモーダルの入力データを得るように構成されてもよい。

20

30

【0129】

いくつか実施形態では、上記装置500は、語位置特定ユニット（図示せず）と、語分類ユニット（図示せず）とをさらに備える。語位置特定ユニットは、分割語系列における各語の語位置データを特定するように構成される。語分類ユニットは、分割語系列における各語を分類処理して、語分類データを得るように構成される。上記取得ユニット503は、語位置マッピングモジュール（図示せず）と、語分類マッピングモジュール（図示せず）とをさらに備える。このうち、語位置マッピングモジュールは、語位置データに対して自己回帰モデルに適合するベクトルマッピングを行い、語位置ベクトルを得るように構成されてもよい。語分類マッピングモジュールは、語分類データに対して自己回帰モデルに適合するベクトルマッピングを行い、語分類ベクトルを得るように構成されてもよい。本実施形態では、得られた自己回帰モデルに適合するマルチモーダルの入力データには、語位置ベクトルと語分類ベクトルもさらに含まれる。

40

【0130】

以下、本開示の実施形態を実現するために適用される電子機器600を示す構造概略図である図6を参照する。

【0131】

図6に示すように、電子機器600は、読み出し専用メモリ（ROM）602に格納さ

50

れているプログラムまたは記憶装置 608 からランダムアクセスメモリ (RAM) 603 にロードされたプログラムによって様々な適当な動作および処理を実行可能な処理装置 (例えば、中央処理装置、グラフィックスプロセッサなど) 601 を含んでもよい。RAM 603 には、電子機器 600 の動作に必要な様々なプログラムおよびデータが更に格納されている。処理装置 601、ROM 602 および RAM 603 は、バス 604 を介して互いに接続されている。入/出力 (I/O) インターフェース 605 もバス 604 に接続されている。

【0132】

例えば、タッチスクリーン、タッチパッド、キーボード、マウスなどを含む入力装置 606、液晶ディスプレイ (LCD, Liquid Crystal Display)、スピーカ、振動子などを含む出力装置 607、例えば、磁気テープ、ハードディスクなどを含む記憶装置 608、および通信装置 609 が I/O インターフェース 605 に接続されてもよい。通信装置 609 により、電子機器 600 は、データを交換するために他のデバイスと無線または有線で通信可能になる。図 6 は、様々な装置を有する電子機器 600 を示しているが、図示された装置のすべてを実装または具備することが要求されないことを理解すべきである。オプション的に実行されるか、またはより多いまたはより少ない装置が実装されてもよい。図 6 に示す各ブロックは、1 つの装置を表すことができ、必要に応じて複数の装置を表すこともできる。

【0133】

特に、本開示の実施形態によれば、上述したフローチャートを参照しながら記載されたプロセスは、コンピュータのソフトウェアプログラムとして実装されてもよい。例えば、本開示の実施形態は、コンピュータ可読媒体に具現化されるコンピュータプログラムを備え、当該コンピュータプログラムは、フローチャートで示される方法を実行するためのプログラムコードを含む。このような実施形態では、該コンピュータプログラムは、通信装置 609 を介してネットワークからダウンロードされてインストールされることが可能であり、または記憶装置 608 または ROM 602 からインストールされ得る。当該コンピュータプログラムが処理装置 601 によって実行されると、本開示の実施形態の方法で限定された上記機能を実行する。

【0134】

なお、本開示の実施形態に記載されたコンピュータ可読媒体は、コンピュータ可読信号媒体またはコンピュータ可読記憶媒体、またはこれらの任意の組み合わせであってもよい。コンピュータ可読記憶媒体は、例えば、電氣的、磁氣的、光学的、電磁氣的、赤外線、または半導体のシステム、装置もしくはデバイス、またはこれらの任意の組み合わせであってもよいが、これらに限定されない。コンピュータ可読記憶媒体のより具体的な例としては、1 本または複数本の導線により電氣的に接続された、ポータブルコンピュータディスク、ハードディスク、ランダムアクセスメモリ (RAM)、読取り専用メモリ (ROM)、消去可能プログラマブル読取り専用メモリ (EPROM もしくはフラッシュメモリ)、ポータブルコンパクトディスク読取り専用メモリ (CD-ROM)、光メモリ、磁気メモリ、またはこれらの任意の適切な組み合わせを含むことができるが、これらに限定されない。本開示の実施形態において、コンピュータ可読記憶媒体は、指令実行システム、装置もしくはデバイスによって使用可能な、またはそれらに組み込まれて使用可能なプログラムを包含または格納する任意の有形の媒体であってもよい。本開示の実施形態において、コンピュータ可読信号媒体は、ベースバンドにおける、または搬送波の一部として伝搬されるデータ信号を含んでもよく、その中にコンピュータ可読プログラムコードが担持されている。かかる伝搬されたデータ信号は、様々な形態をとることができ、電磁信号、光信号、またはこれらの任意の適切な組み合わせを含むが、これらに限定されない。コンピュータ可読信号媒体は、更にコンピュータ可読記憶媒体以外の任意のコンピュータ可読媒体であってもよい。当該コンピュータ可読信号媒体は、指令実行システム、装置もしくはデバイスによって使用されるか、またはそれらに組み込まれて使用されるプログラムを、送信、伝搬または伝送することができる。コンピュータ可読媒体に含まれるプログラムコ

10

20

30

40

50

ードは任意の適切な媒体で伝送することができ、当該任意の適切な媒体とは、電線、光ケーブル、RF (Radio Frequency, 無線周波数) など、またはこれらの任意の適切な組み合わせを含むが、これらに限定されない。

【0135】

上記コンピュータ可読媒体は、上記サーバに含まれるものであってもよく、当該サーバに実装されずに別体として存在するものであってもよい。上記コンピュータ可読媒体は1つまたは複数のプログラムを担持しており、上記1つまたは複数のプログラムが当該サーバによって実行されると、当該サーバに、処理対象となるデータセット内の少なくとも2種の異なるモダルのデータを取得するステップと、少なくとも2種の異なるモダルのデータに対して特徴抽出を行った後に、コンカットおよび/または重ね合わせにより特徴

10

【0136】

本開示の実施形態の動作を実行するためのコンピュータプログラムコードは、種以上のプログラミング言語、またはそれらの組み合わせで作成されることができ、プログラミング言語は、Java、Smalltalk、C++などのオブジェクト指向プログラミング言語と、「C」言語または同様のプログラミング言語などの従来の手続き型プログラミング言語とを含む。プログラムコードは、完全にユーザのコンピュータで実行されることも、部分的にユーザのコンピュータで実行されることも、単独のソフトウェアパッケージとして実行されることも、部分的にユーザのコンピュータで実行されながら部分的にリモートコンピュータで実行されることも、または完全にリモートコンピュータもしくはサーバで実行されることも可能である。リモートコンピュータの場合、リモートコンピュータは、ローカルエリアネットワーク(LAN)またはワイドエリアネットワーク(WAN)を含む任意の種類のネットワークを介してユーザコンピュータに接続することができ、または(例えば、インターネットサービスプロバイダによるインターネットサービスを介して)外部コンピュータに接続することができる。

20

【0137】

図面のうちのフローチャートおよびブロック図は、本開示の様々な実施形態に係るシステム、方法およびコンピュータプログラムによって実現できるアーキテクチャ、機能および動作を図示している。これについては、フローチャートまたはブロック図における各ブロックは、モジュール、プログラムセグメント、またはコードの一部を表すことができる。当該モジュール、プログラムセグメント、またはコードの一部には、所定のロジック機能を実現するための1つまたは複数の実行可能な指令が含まれている。なお、いくつかのオプション的な実施形態においては、ブロックに示されている機能は図面に示されているものとは異なる順番で実行することも可能である。例えば、連続して示された2つのブロックは、実際には係る機能に応答して、ほぼ並行して実行されてもよく、時には逆の順序で実行されてもよい。さらに注意すべきなのは、ブロック図および/またはフローチャートにおけるすべてのブロック、ならびにブロック図および/またはフローチャートにおけるブロックの組み合わせは、所定の機能または動作を実行する専用のハードウェアベースのシステムで実装されてもよく、または専用のハードウェアとコンピュータ指令との組み合わせで実装されてもよい。

30

40

【0138】

本開示の実施形態に記載されたユニットは、ソフトウェアで実装されてもよく、ハードウェアで実装されてもよい。記載されたユニットは、プロセッサに設置されてもよく、例えば、「獲得ユニットと、抽出ユニットと、取得ユニットと、処理ユニットとを備えるプロセッサ」と記載されてもよい。ここで、これらのユニットの名称は、ある場合において当該ユニットその自体を限定するものではない。例えば、獲得ユニットは、「処理対象となるデータセット内の少なくとも2種の異なるモダルのデータを取得するように構成さ

50

れる」ユニットとして記載されてもよい。

【 0 1 3 9 】

以上の記載は、本開示の好ましい実施形態、および適用される技術的原理に関する説明に過ぎない。当業者であれば、本開示に係る発明の範囲が、上述した技術的特徴の特定の組み合わせからなる技術案に限定されるものではなく、上述した本開示の趣旨を逸脱しない範囲で、上述した技術的特徴またはそれらの均等の特徴の任意の組み合わせからなる他の技術案も含むべきであることを理解すべきである。例えば、上記の特徴と、本開示の実施形態に開示された類似の機能を持っている技術的特徴（これらに限定されていない）と互いに置き換えてなる技術案が挙げられる。

10

20

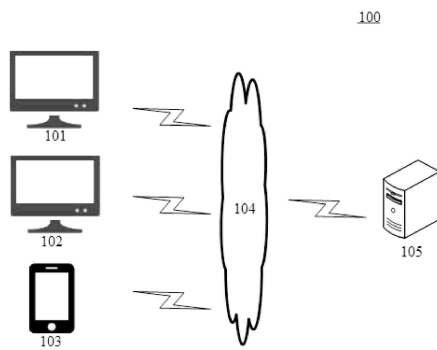
30

40

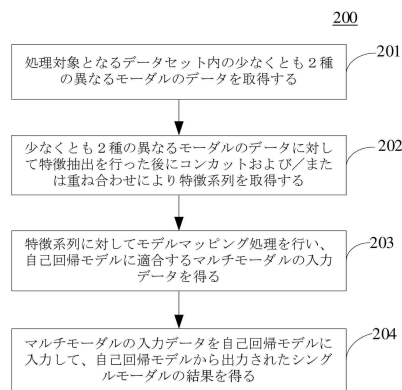
50

【図面】

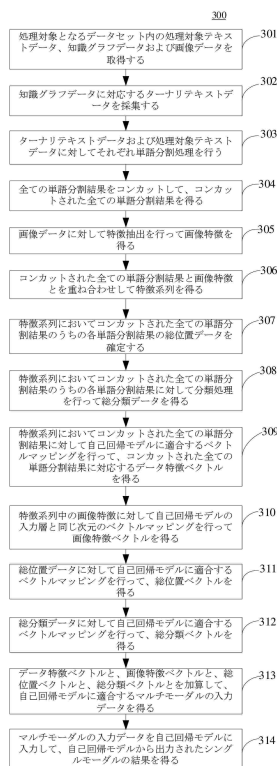
【 図 1 】



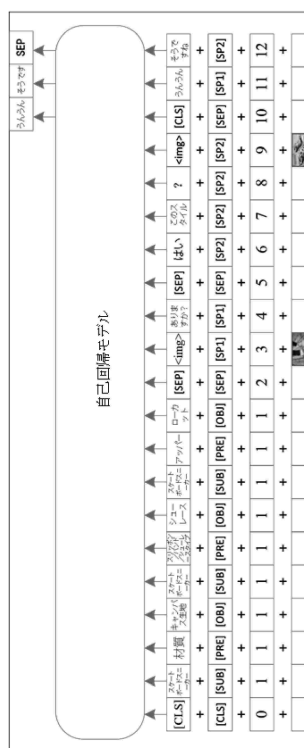
【 図 2 】



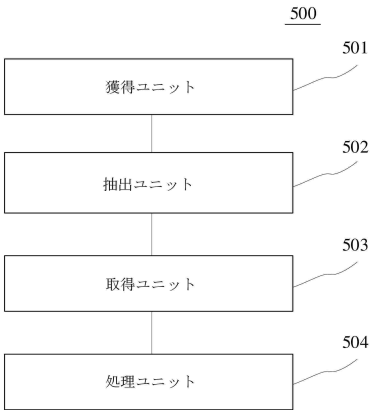
【圖 3】



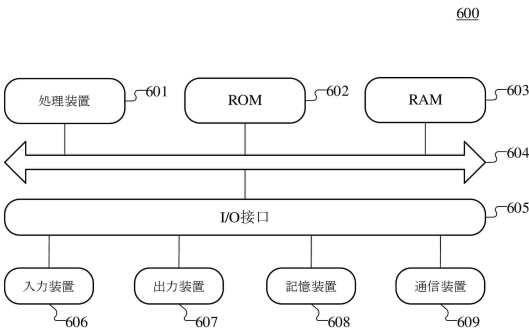
【図 4】



【図 5】



【図 6】



10

20

30

40

50

フロントページの続き

Zone, Beijing 100176, China

(73)特許権者 517241916

北京京東世紀貿易有限公司

BEIJING JINGDONG CENTURY TRADING CO., LTD.

中華人民共和国北京市北京▲経▼▲済▼技▲術▼▲開▼▲発▼区科▲創▼十一街18号C座2▲層▼201室

Room 201, 2/F, Block C, No.18, Kechuang 11th Street, Beijing Economic and Technological Development Zone, Beijing 100176, China

(74)代理人 100108453

弁理士 村山 靖彦

(74)代理人 100110364

弁理士 実広 信哉

(74)代理人 100133400

弁理士 阿部 達彦

(72)発明者 ▲趙▼楠

中華人民共和国100176北京市▲経▼▲済▼技▲術▼▲開▼▲発▼区科▲創▼十一街18号院2号楼4▲層▼A402室

(72)発明者 ▲吳▼友政

中華人民共和国100176北京市▲経▼▲済▼技▲術▼▲開▼▲発▼区科▲創▼十一街18号院2号楼4▲層▼A402室

審査官 早川 学

(56)参考文献 中国特許出願公開第111461203(CN, A)

特開2019-219827(JP, A)

米国特許第10467274(US, B1)

(58)調査した分野 (Int.Cl., DB名)

G06F 16/00 - 16/958

G06F 40/00 - 40/58

G06Q 10/00 - 99/00