

【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5月31日(2022.5.31)

【国際公開番号】WO2019/221985

【公表番号】特表2021-524099(P2021-524099A)

【公表日】令和3年9月9日(2021.9.9)

【出願番号】特願2020-564186(P2020-564186)

【国際特許分類】

G 06 N 20/00(2019.01)

10

G 06 T 7/00(2017.01)

【F I】

G 06 N 20/00

G 06 T 7/00 350 C

【手続補正書】

【提出日】令和4年5月23日(2022.5.23)

【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

20

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項1】

第1モダリティからの入力データおよび前記第1モダリティとは異なる第2モダリティからの入力データを含む複数のモダリティからの入力データを受信するように構成されたマルチモーダル統計モデルを訓練する方法であって、前記方法は

前記第1モダリティの訓練データおよび前記第2モダリティの訓練データを含む訓練データにアクセスする訓練データアクセス工程と、

前記マルチモーダル統計モデルを、前記訓練データを用いて訓練する、訓練工程であって、前記マルチモーダル統計モデルは、前記第1モダリティおよび前記第2モダリティの入力データをそれぞれ処理する第1エンコーダおよび第2エンコーダと、第1モダリティ埋め込みおよび第2モダリティ埋め込みと、共同モダリティ表現と、予測子とを含む複数の構成要素を含み、前記第1モダリティ埋め込みおよび前記第2モダリティ埋め込みは、前記共同モダリティ表現を第1モダリティの空間および第2モダリティの空間にそれぞれ投影するための複数のベクトルおよび複数の重みを含み、前記訓練工程は、

前記訓練データを使用して、前記複数の構成要素のパラメータの値を推定することを含む訓練工程と、

前記マルチモーダル統計モデルの前記複数の構成要素のパラメータの推定した前記値を記憶することにより、前記マルチモーダル統計モデルを指定する情報を少なくとも部分的に記憶する工程と、を備える方法。

【請求項2】

前記訓練工程は、前記第1モダリティ埋め込みおよび前記第2モダリティ埋め込みと前記共同モダリティ表現と予測子とのパラメータの値を推定する前に前記第1エンコーダおよび前記第2エンコーダのパラメータの値を推定する工程をさらに含む、請求項1に記載の方法。

【請求項3】

前記訓練工程は、前記第1モダリティおよび前記第2モダリティの第1デコーダおよび第2デコーダのパラメータの値を推定する工程をさらに含む、請求項2に記載の方法。

【請求項4】

50

前記訓練工程は、前記第1エンコーダおよび前記第2エンコーダのパラメータの値の推定を、前記共同モダリティ表現のパラメータの値の推定と共同でする工程をさらに含む、請求項1に記載の方法。

#### 【請求項5】

前記訓練工程は、前記第1モダリティの第1デコーダおよび前記第2モダリティの第2デコーダのパラメータの値を推定する工程をさらに含む、請求項4に記載の方法。

#### 【請求項6】

前記訓練工程は、

前記第1モダリティの前記訓練データにおける第1データ入力にアクセスする工程と、前記第1データ入力を前記第1エンコーダに提供して、第1特徴ベクトルを生成する工程と、

10

前記共同モダリティ表現、前記第1モダリティ埋め込み、および前記第1特徴ベクトルを使用して、第2特徴ベクトルを特定する、第2特徴ベクトル特定工程と、

前記第2特徴ベクトルを入力として第1デコーダに提供して、第1データ出力を生成する工程と、を含む請求項1に記載の方法。

#### 【請求項7】

前記第1データ出力を前記第1データ入力と比較する工程と、

前記比較の結果に基づき、前記共同モダリティ表現の1つ以上のパラメータの1つ以上の値を更新する工程と、をさらに含む請求項6に記載の方法。

20

#### 【請求項8】

前記訓練工程は、

前記第1モダリティ埋め込みおよび前記第2モダリティ埋め込みと前記共同モダリティ表現とのパラメータの値を推定することにより、少なくとも部分的に第1訓練段階を実行する、第1訓練段階実行工程および、

前記予測子のパラメータの値を推定することにより、少なくとも部分的に第2訓練段階を実行する、第2訓練段階実行工程、を含む請求項1に記載の方法。

30

#### 【請求項9】

前記第1エンコーダはd次元ベクトルを出力するように構成され、前記共同モダリティ表現はN個のm次元ベクトルを含み、前記第1モダリティ埋め込みはm\*dの重みを含む、請求項6に記載の方法。

30

#### 【請求項10】

前記第2特徴ベクトル特定工程は、

前記第1モダリティ埋め込みを使用することにより、前記共同モダリティ表現を前記第1モダリティの空間に投影して、N個のd次元ベクトルを取得する工程と、

前記共同モダリティ表現における前記N個のd次元ベクトルの中から、類似性メトリックに従って前記第1特徴ベクトルに最も類似する第3特徴ベクトルを特定する工程と、

前記第1特徴ベクトルを前記第3特徴ベクトルと集約することにより前記第2特徴ベクトルを生成する工程と、を含む請求項9に記載の方法。

40

#### 【請求項11】

前記第2特徴ベクトル特定工程は、

前記第1モダリティ埋め込みを使用することにより、前記共同モダリティ表現を前記第1モダリティの空間に投影して、N個のd次元ベクトルを取得する工程と、

前記N個のd次元ベクトルの少なくとも一部と前記第1特徴ベクトルとの間の類似性に従って、前記共同モダリティ表現における前記N個のd次元ベクトルの前記少なくとも一部の重みを算出する工程と、

前記第1特徴ベクトルを、算出された前記重みによって重み付けされた前記N個のd次元ベクトルの前記少なくとも一部の加重和と集約することにより前記第2特徴ベクトルを生成する工程と、を含む請求項9に記載の方法。

#### 【請求項12】

前記マルチモーダル統計モデルは、第1タスク埋め込みおよび第2タスク埋め込みをさら

50

に備え、前記訓練工程は、前記第1タスク埋め込みおよび前記第2タスク埋め込みのパラメータの値の推定を、前記予測子のパラメータの値の推定と共同でする工程をさらに含む、請求項1に記載の方法。

【請求項13】

前記第1エンコーダはニューラルネットワークを含む、請求項1に記載の方法。

【請求項14】

前記ニューラルネットワークは畳み込みニューラルネットワークである、請求項13に記載の方法。

【請求項15】

前記ニューラルネットワークは回帰型ニューラルネットワークである、請求項13に記載の方法。 10

【請求項16】

前記第1モダリティの前記訓練データは画像を含む、請求項1に記載の方法。

【請求項17】

前記第2モダリティの訓練データはテキストを含む、請求項16に記載の方法。

【請求項18】

1つ以上のコンピュータハードウェアプロセッサと、

1つ以上の非一時的なコンピュータ可読記憶媒体と、を備えるシステムであって、前記非一時的なコンピュータ可読記憶媒体は、前記1つ以上のコンピュータハードウェアプロセッサによって実行された場合、前記1つ以上のコンピュータハードウェアプロセッサに、第1モダリティからの入力データおよび前記第1モダリティとは異なる第2モダリティからの入力データを含む複数のモダリティからの入力データを受信するように構成されたマルチモーダル統計モデルを訓練する方法、を実行させるプロセッサ実行可能な命令を記憶し、前記方法は 20

前記第1モダリティの訓練データおよび前記第2モダリティの訓練データを含む訓練データにアクセスする工程と、

前記マルチモーダル統計モデルを2段階で訓練する、訓練工程であって、前記マルチモーダル統計モデルは、前記第1モダリティおよび前記第2モダリティの入力データをそれぞれ処理する第1エンコーダおよび第2エンコーダと、第1モダリティ埋め込みおよび第2モダリティ埋め込みと、共同モダリティ表現と、予測子とを含む複数の構成要素を含み、前記第1モダリティ埋め込みおよび前記第2モダリティ埋め込みは、前記共同モダリティ表現を第1モダリティの空間および第2モダリティの空間にそれぞれ投影するための複数のベクトルおよび複数の重みを含み、前記訓練工程は、 30

前記訓練データを使用して、前記複数の構成要素のパラメータの値を推定することを含む訓練工程と、

前記マルチモーダル統計モデルの前記複数の構成要素のパラメータの推定した前記値を記憶することにより、前記マルチモーダル統計モデルを指定する情報を少なくとも部分的に記憶する工程と、を含むシステム。

【請求項19】

前記訓練工程は、 40

前記第1モダリティ埋め込みおよび前記第2モダリティ埋め込みと前記共同モダリティ表現とのパラメータの値を推定することにより、少なくとも部分的に第1訓練段階を実行する、第1訓練段階実行工程および、

前記予測子のパラメータの値を推定することにより、少なくとも部分的に第2訓練段階を実行する、第2訓練段階実行工程、を含む請求項18に記載のシステム。

【請求項20】

1つ以上の非一時的なコンピュータ可読記憶媒体であって、前記非一時的なコンピュータ可読記憶媒体は、1つ以上のコンピュータハードウェアプロセッサによって実行された場合、前記1つ以上のコンピュータハードウェアプロセッサに、第1モダリティからの入力データおよび前記第1モダリティとは異なる第2モダリティからの入力データを含む複数 50

のモダリティからの入力データを受信するように構成されたマルチモーダル統計モデルを訓練する方法、を実行させるプロセッサ実行可能な命令を記憶し、前記方法は前記第1モダリティの訓練データおよび前記第2モダリティの訓練データを含む訓練データにアクセスする工程と、

前記マルチモーダル統計モデルを2段階で訓練する、訓練工程であって、前記マルチモーダル統計モデルは、前記第1モダリティおよび前記第2モダリティの入力データをそれぞれ処理する第1エンコーダおよび第2エンコーダと、第1モダリティ埋め込みおよび第2モダリティ埋め込みと、共同モダリティ表現と、予測子とを含む複数の構成要素を含み、前記第1モダリティ埋め込みおよび前記第2モダリティ埋め込みは、前記共同モダリティ表現を第1モダリティの空間および第2モダリティの空間にそれぞれ投影するための複数のベクトルおよび複数の重みを含み、前記訓練工程は、

前記訓練データを使用して、前記複数の構成要素のパラメータの値を推定することを含む訓練工程と、

前記マルチモーダル統計モデルの前記複数の構成要素のパラメータの推定した前記値を記憶することにより、前記マルチモーダル統計モデルを指定する情報を少なくとも部分的に記憶する工程と、を含む非一時的なコンピュータ可読記憶媒体。

10

20

30

40

50