

【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 2 年 1 月 23 日 (2020.1.23)

【公表番号】特表 2019-502988 (P2019-502988A)

【公表日】平成 31 年 1 月 31 日 (2019.1.31)

【年通号数】公開・登録公報 2019-004

【出願番号】特願 2018-528803 (P2018-528803)

【国際特許分類】

G 0 6 N 3/02 (2006.01)

G 0 6 N 20/00 (2019.01)

【F I】

G 0 6 N 3/02

G 0 6 N 99/00 1 5 3

【手続補正書】

【提出日】令和 1 年 12 月 2 日 (2019.12.2)

【手続補正 1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項 1】

入力から取得される表現の生成のためのコンピュータシステムであって、

(i) ニューラルネットワークを含むエンコーダ及びニューラルネットワークを含むデコーダを含むオートエンコーダを有し、

(1) 前記エンコーダは、化合物に関する情報である入力を潜在変数として符号化するように構成され、

(2) 前記デコーダは、前記潜在変数に基づき情報を復号化し、ランダム変数を出力するように構成され、

前記システムは、前記入力と前記入力に関連する訓練ラベルとの双方を前記エンコーダに符号化させ、前記入力の再構成物を前記デコーダに生成させることによって訓練され、前記システムの訓練が再構成誤差によって制約される、
コンピュータシステム。

【請求項 2】

前記訓練ラベルが、所定の値を有する 1 つまたは複数のラベル要素を含む、請求項 1 に記載のコンピュータシステム。

【請求項 3】

前記システムが、1 つまたは複数のラベル要素を含むターゲットラベルを受け取り、前記 1 つまたは複数のラベル要素の各々についての規定値を満たす化合物指紋を生成するように構成される、請求項 1 又は 2 に記載のコンピュータシステム。

【請求項 4】

前記訓練ラベルが前記ターゲットラベルを含まない、請求項 3 に記載のコンピュータシステム。

【請求項 5】

前記訓練が、前記エンコーダと前記デコーダとの間の情報フロー全体をさらに制約する、請求項 1 乃至 4 何れか一項に記載のコンピュータシステム。

【請求項 6】

前記エンコーダが、平均のベクトルおよび標準偏差のベクトルのペアを含む出力を提供

するように構成される、請求項 1 乃至 5 何れか一項に記載のコンピュータシステム。

【請求項 7】

サンプリングモジュールを更に含み、

前記サンプリングモジュールが、前記エンコーダの出力を受け取り、前記エンコーダの前記出力に基づいて前記潜在変数を定義し、1つまたは複数の潜在的表現を生成するように構成され、前記潜在変数が確率分布によってモデル化される、請求項 6 に記載のコンピュータシステム。

【請求項 8】

前記オートエンコーダが生成モデルを含む、請求項 1 乃至 7 何れか一項に記載のコンピュータシステム。

【請求項 9】

前記訓練ラベルが、バイオアッセイ結果、毒性、交差反応性、薬物動態、薬力学、バイオアベイラビリティ、および溶解性からなるグループから選択される1つまたは複数のラベル要素を含む、請求項 2 に記載のコンピュータシステム。

【請求項 10】

入力から取得される表現の生成のためのコンピュータによって実現される訓練方法であって、前記訓練方法が生成モデルを訓練するステップを含み、

前記訓練が、

(i) 化合物に関する情報である入力と関連する訓練ラベルとの双方を前記生成モデルに入力すること、

ならびに

(i i) 前記入力の再構成物を生成すること

を含み、

前記生成モデルが、ニューラルネットワークを含むエンコーダ及びニューラルネットワークを含むデコーダを含むオートエンコーダを有し、

(1) 前記エンコーダは、前記入力及び前記訓練ラベルを潜在変数として符号化するように構成され、

(2) 前記デコーダは、ランダム変数として前記潜在変数に基づき情報を復号化するように構成され、

前記訓練が再構成誤差によって制約される、

訓練方法。

【請求項 11】

薬物予測のためのコンピュータシステムであって、前記システムが

(i) 1つ以上のニューラルネットワークを含む生成モデルを含む機械学習モデル

を含み、

前記生成モデルが、化合物に関する情報である入力データと、1つ以上のラベル要素を含む関連する訓練ラベルとを含む訓練データセットによって訓練され、

前記生成モデルは、前記訓練ラベルを入力としてエンコーダに入力し、前記エンコーダ以外から供給された前記訓練ラベルを前記生成モデルの訓練のためのデコーダに入力することによって訓練される、

システム。

【請求項 12】

前記ラベル要素が、バイオアッセイ結果、毒性、交差反応性、薬物動態、薬力学、バイオアベイラビリティ、および溶解性からなるグループから選択される1つまたは複数の要素を含む、請求項 11 に記載のシステム。

【請求項 13】

前記生成モデルがオートエンコーダを含む、請求項 11 又は 12 に記載のシステム。

【請求項 14】

前記生成モデルが、確率的エンコーダ、確率的デコーダ、およびサンプリングモジュールを含む確率的オートエンコーダまたは変分オートエンコーダを含む、請求項 11 乃至 1

3 何れか一項に記載のシステム。

【請求項 15】

前記確率的エンコーダが、平均のベクトルおよび標準偏差のベクトルのペアを含む出力を提供するように構成される、請求項 14 に記載のシステム。

【請求項 16】

前記サンプリングモジュールが、前記確率的エンコーダの出力を受け取り、前記確率的エンコーダの前記出力に基づいて潜在変数を定義し、1つまたは複数の潜在的表現を生成するように構成され、前記潜在変数が確率分布である、請求項 14 に記載のシステム。

【請求項 17】

前記確率的デコーダが、潜在的表現を復号し、指紋要素の値にわたって確率変数を生成するように構成される、請求項 14 に記載のシステム。

【請求項 18】

前記確率的エンコーダおよび前記確率的デコーダが同時に訓練される、請求項 14 に記載のシステム。

【請求項 19】

前記生成モデルが予測子をさらに含む、請求項 11 乃至 18 何れか一項に記載のシステム。

【請求項 20】

前記予測子が、前記訓練ラベルの少なくともサブセットについて1つまたは複数のラベル要素の値を予測するように構成される、請求項 19 に記載のシステム。

【請求項 21】

前記機械学習モデルが、前記訓練データセットにないシステム生成化合物指紋を含む出力を提供するように構成される、請求項 11 乃至 20 何れか一項に記載のシステム。

【請求項 22】

薬物予測のためのコンピュータにより実現される方法であって、

(i) 化合物に関する情報である入力データと、1つ以上のラベル要素を含む関連する訓練ラベルとを含む訓練データセットによって機械学習モデルの生成モデルを訓練するステップ

を含み、

前記生成モデルは、1つ以上のニューラルネットワークを含み、

前記生成モデルは、前記訓練ラベルを入力としてエンコーダに入力し、前記エンコーダ以外から供給された前記訓練ラベルを前記生成モデルの訓練のためのデコーダに入力することによって訓練される、方法。

【請求項 23】

前記訓練ラベルが、バイオアッセイ結果、毒性、交差反応性、薬物動態、薬力学、バイオアベイラビリティ、および溶解性からなるグループから選択される1つまたは複数の要素を含む、請求項 22 に記載の方法。

【請求項 24】

前記生成モデルがオートエンコーダを含む、請求項 22 又は 23 に記載の方法。

【請求項 25】

前記生成モデルが、確率的エンコーダと確率的デコーダとサンプリングモジュールとを含む確率的オートエンコーダを含む、請求項 22 乃至 24 何れか一項に記載の方法。

【請求項 26】

前記訓練データセット内の化合物指紋ごとに平均のベクトルおよび標準偏差のベクトルのペアを含む出力を前記エンコーダから提供するステップをさらに含む、請求項 22 乃至 25 何れか一項に記載の方法。

【請求項 27】

前記確率的エンコーダおよび前記確率的デコーダが同時に訓練される、請求項 25 に記載の方法。

【請求項 28】

前記訓練が、

前記確率的エンコーダを、潜在変数を定義する平均のベクトルおよび標準偏差のベクトルとして化合物指紋を符号化するように訓練することと、

前記潜在変数から潜在的表現を引き出すことと、

前記確率的デコーダを、前記化合物指紋の確率的再構成物として前記潜在的表現を復号するように訓練することを含む、請求項 2 7 に記載の方法。

【請求項 2 9】

前記訓練が、逆伝搬を使用して前記確率的エンコーダについての変分下限を最適化することを含む、請求項 2 5 に記載の方法。

【請求項 3 0】

前記生成モデルが予測子モジュールを含む、請求項 2 2 乃至 2 9 何れか一項に記載の方法。

【請求項 3 1】

前記訓練データセット内の 1 つまたは複数の化合物指紋に関連付けられたラベル要素についての 1 つまたは複数の値を予測するステップをさらに含む、請求項 3 0 に記載の方法。

【請求項 3 2】

前記訓練データセット内に表されていない化合物についての同定情報を含む出力を前記生成モデルから生成するステップをさらに含む、請求項 2 2 乃至 3 1 何れか一項に記載の方法。

【請求項 3 3】

化合物表現の生成のためのコンピュータシステムであって、

ニューラルネットワークとして実現されるデコーダであって、前記化合物表現に対応するランダム変数を生成するため、潜在表現及びラベルを受け取るように構成される生成モデルであるデコーダ

を含み、

ニューラルネットワークとして実現されるエンコーダに、化合物に関する情報及びラベルを符号化させ、潜在表現を提供するための潜在変数を生成し、前記デコーダに前記潜在表現及び前記ラベルを復号化させ、前記化合物に関する情報の再構成物に対応するランダム変数を生成することによって、前記デコーダは訓練され、前記デコーダの訓練は再構成誤差によって制約される、

コンピュータシステム。

【請求項 3 4】

前記システムが、1 つまたは複数のラベル要素について選択された値を満たす化合物指紋を生成するように構成される、請求項 3 3 に記載のコンピュータシステム。

【請求項 3 5】

前記ラベルが、前記 1 つまたは複数のラベル要素についての前記選択された値を含まない、請求項 3 4 に記載のコンピュータシステム。

【請求項 3 6】

訓練データセット内の 1 つまたは複数の化合物指紋に関連付けられた 1 つまたは複数のラベル要素についての値を予測するように構成された予測子をさらに含む、請求項 3 4 に記載のコンピュータシステム。

【請求項 3 7】

前記ラベル要素が、バイオアッセイ、毒性、交差反応性、薬物動態、薬力学、バイオベイヤビリティ、および溶解性からなるグループから選択される 1 つまたは複数の要素を含む、請求項 3 4 乃至 3 6 何れか一項に記載のコンピュータシステム。