

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7435749号
(P7435749)

(45)発行日 令和6年2月21日(2024.2.21)

(24)登録日 令和6年2月13日(2024.2.13)

(51)国際特許分類 F I
G 0 6 N 20/00 (2019.01) G 0 6 N 20/00

請求項の数 6 (全15頁)

(21)出願番号	特願2022-514299(P2022-514299)	(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
(86)(22)出願日	令和2年4月10日(2020.4.10)	(74)代理人	110002147 弁理士法人酒井国際特許事務所
(86)国際出願番号	PCT/JP2020/016212	(72)発明者	鎌田 裕一 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
(87)国際公開番号	WO2021/205669	(72)発明者	中川 章 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
(87)国際公開日	令和3年10月14日(2021.10.14)	(72)発明者	加藤 圭造 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
審査請求日	令和4年7月14日(2022.7.14)		最終頁に続く

(54)【発明の名称】 推定プログラム、推定方法および情報処理装置

(57)【特許請求の範囲】

【請求項1】

符号化部と復号化部とを有する訓練済みの変分オートエンコーダに入力データを入力し、前記入力に応じて前記訓練済みの変分オートエンコーダにより生成される潜在変数の確率分布を、前記符号化部から出力される標準偏差の大きさに基づいて、前記潜在変数から主成分座標への変換処理によって、第1の確率分布に変換し、

前記入力データに関する前記復号化部の出力誤差に基づき、前記第1の確率分布を第2の確率分布に変換し、

前記入力データの確率分布の推定値として、前記第2の確率分布を出力する、処理をコンピュータに実行させることを特徴とする推定プログラム。

10

【請求項2】

前記変換処理は、

前記符号化部から前記潜在変数の確率分布の分布パラメータである前記標準偏差と平均とを取得する処理と、

前記主成分座標と前記潜在変数との間のスケールの変化率を、前記標準偏差の大きさを用いて算出する処理と、

前記標準偏差と前記平均と前記変化率とを用いて前記潜在変数を前記主成分座標へ変換する処理と、

を含むことを特徴とする請求項1に記載の推定プログラム。

【請求項3】

20

前記第 2 の確率分布に変換する処理は、

前記第 1 の確率分布における主成分以外の確率分布を定数に設定し、入力データの前記スケールを設定する正規分布に従う前記出力誤差を用いて、前記第 1 の確率分布を第 2 の確率分布に変換する処理

を含むことを特徴とする請求項 2 に記載の推定プログラム。

【請求項 4】

前記入力データから生成された前記第 2 の確率分布に基づいて、特定の割合を占める確率が下位のデータを異常データ、または、前記第 2 の確率分布に基づいて、閾値以下の確率であるデータを異常データ、として検出する処理

を前記コンピュータにさらに実行させることを特徴とする請求項 1 に記載の推定プログラム。

10

【請求項 5】

符号化部と復号化部とを有する訓練済みの変分オートエンコーダに入力データを入力し、前記入力に応じて前記訓練済みの変分オートエンコーダにより生成される潜在変数の確率分布を、前記符号化部から出力される標準偏差の大きさに基づいて、前記潜在変数から主成分座標への変換処理によって、第 1 の確率分布に変換し、

前記入力データに関する前記復号化部の出力誤差に基づき、前記第 1 の確率分布を第 2 の確率分布に変換し、

前記入力データの確率分布の推定値として、前記第 2 の確率分布を出力する、

処理をコンピュータが実行することを特徴とする推定方法。

20

【請求項 6】

符号化部と復号化部とを有する訓練済みの変分オートエンコーダに入力データを入力し、前記入力に応じて前記訓練済みの変分オートエンコーダにより生成される潜在変数の確率分布を、前記符号化部から出力される標準偏差の大きさに基づいて、前記潜在変数から主成分座標への変換処理によって、第 1 の確率分布に変換し、

前記入力データに関する前記復号化部の出力誤差に基づき、前記第 1 の確率分布を第 2 の確率分布に変換し、

前記入力データの確率分布の推定値として、前記第 2 の確率分布を出力する、

処理を実行する制御部を含むことを特徴とする情報処理装置。

【発明の詳細な説明】

30

【技術分野】

【0001】

本発明は、確率分布の推定技術に関する。

【背景技術】

【0002】

データ解析において、オートエンコーダなどを用いて、複雑な多次元データから低次元の特徴量を抽出し、特徴量を用いてデータ解析を実行することが行われている。例えば、ベルトコンベアを流れる製品の画像の特徴量を抽出し、流れてくる製品から不良品を検出することなどが行われている。

【0003】

40

近年では、潜在変数を確率分布として学習する変分オートエンコーダ (VAE: Variational Autoencoder) を用いたデータ解析が利用されている。例えば、VAE は、符号化部と復号化部とを有し、潜在変数が入力される復号化部の出力を用いて計算される再構成誤差の期待値と、特徴量を入力する符号化部の出力を用いて計算される潜在変数の確率分布の正規化誤差とを最小化するように、符号化部と復号化部のパラメータの機械学習が実行される。このようにして学習 (訓練) された VAE に、複数の検出対象データを入力して異常データ (アノマリデータ) の検出が行われる。

【先行技術文献】

【非特許文献】

【0004】

50

【文献】Diederik P. Kingma, Max Welling, “Auto - Encoding Variational Bayes”, ICLR 2014.

【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、上記VAEでは、単変数独立の正規分布を仮定しており、得られた潜在空間の確率分布が実空間の分布を反映する保証がない。このため、判定対象データを訓練済みのVAEに入力し、符号化部の出力を用いて入力データの確率分布を推定して、異常データを検出する場合、推定結果が保証できず、異常検出の精度が高くない。

【0006】

なお、潜在変数の情報エントロピーを最小化するRate - Distortion理論を応用したオートエンコーダを用いることも考えられる。このようなオートエンコーダを用いた場合、潜在空間の確率分布と実空間データの確率分布とほぼ同一になる。しかし、実空間の確率分布の形状が複雑である場合に、それと同一となるように設計する潜在空間の確率分布も、複数のパラメトリックな確率分布を混合させるなどの複雑な形状で表現する必要があるため、精度を高くするためにはコストが増大となり、現実的ではない。

【0007】

一つの側面では、入力データの異常検出の精度を向上することができる推定プログラム、推定方法および情報処理装置を提供することを目的とする。

【課題を解決するための手段】

【0008】

第1の案では、推定プログラムは、符号化部と復号化部とを有する訓練済みの変分オートエンコーダに入力データを入力し、前記入力に応じて前記訓練済みの変分オートエンコーダにより生成される潜在変数の確率分布を、前記符号化部から出力される標準偏差の大きさに基づき、第1の確率分布に変換する処理をコンピュータに実行させる。推定プログラムは、前記入力データに関する前記復号化部の出力誤差に基づき、前記第1の確率分布を第2の確率分布に変換する処理をコンピュータに実行させる。推定プログラムは、前記入力データの確率分布の推定値として、前記第2の確率分布を出力する処理をコンピュータに実行させる。

【発明の効果】

【0009】

一つの側面では、入力データの異常検出の精度を向上することができる。

【図面の簡単な説明】

【0010】

【図1】図1は、実施例1にかかる情報処理装置を説明する図である。

【図2】図2は、実施例1にかかる情報処理装置の機能構成を示す機能ブロック図である。

【図3】図3は、VAEの構成および機械学習を説明する図である。

【図4】図4は、入力データの確率分布の算出を説明する図である。

【図5】図5は、入力データと各変数の対応を説明する図である。

【図6】図6は、学習処理の流れを示すフローチャートである。

【図7】図7は、検出処理の流れを示すフローチャートである。

【図8】図8は、検証用に人工的に生成した入力データを説明する図である。

【図9】図9は、参考技術を用いた異常検出結果を説明する図である。

【図10】図10は、実施例1を用いた異常検出結果を説明する図である。

【図11】図11は、VAEの別例を説明する図である。

【図12】図12は、ハードウェア構成例を説明する図である。

【発明を実施するための形態】

【0011】

以下に、本発明にかかる推定プログラム、推定方法および情報処理装置の実施例を図面に基づいて詳細に説明する。なお、この実施例によりこの発明が限定されるものではない

10

20

30

40

50

。また、各実施例は、矛盾のない範囲内で適宜組み合わせることができる。

【0012】

図1は、実施例1にかかる情報処理装置10を説明する図である。図1に示す情報処理装置10は、VAEを用いて生成したモデルに実空間の入力データを入力し、VAEの符号化部によって推定した潜在変数の事後分布パラメータを用いて、事前確率を入力データの生成確率とほぼ同一となるよう補正する。このようにして、情報処理装置10は、VAEにより抽出された潜在空間の確率分布から入力データの確率分布を推定し、入力データの異常検出の精度向上を図るコンピュータ装置である。

【0013】

具体的には、情報処理装置10は、訓練データを用いて、符号化部と復号化部を有するVAEの機械学習を実行し、訓練済みのVAEを適用したモデルを生成する。そして、情報処理装置10は、訓練データと同じドメインの入力データをモデルの符号化部に入力して、モデルの復号化部から復元された入力データを取得する。

10

【0014】

ここで、情報処理装置10は、VAEの符号化部から出力される潜在変数の確率分布を、符号化部の出力の標準偏差の大きさに基づいて第1の確率分布に変換する。さらに、情報処理装置10は、VAEの復号化部の出力誤差に基づいて第1の確率分布を第2の確率分布に変換し、入力データの確率分布の推定値として、第2の確率分布を出力する。

【0015】

このようにして、情報処理装置10は、生成された第2の確率分布に基づいて、特定の割合を占める確率が下位のデータを異常データ(アノマリデータ)として検出する。また、情報処理装置10は、第2の確率分布に基づいて、複数の入力データのうち閾値以下の確率であるデータを異常データとして検出しても良い。

20

【0016】

図2は、実施例1にかかる情報処理装置10の機能構成を示す機能ブロック図である。図2に示すように、情報処理装置10は、通信部11、表示部12、記憶部13、制御部20を有する。

【0017】

通信部11は、他の装置の間の通信を制御する。例えば、通信部11は、管理者の端末から機械学習の開始指示や各種データを受信し、機械学習の結果や異常検出の結果などを管理者の端末に送信する。

30

【0018】

記憶部13は、各種データや制御部20が実行するプログラムなどを記憶する。例えば、記憶部13は、訓練データ14、入力データ15、モデル16などを記憶する。

【0019】

訓練データ14は、VAEの機械学習に利用される訓練データであり、同じドメインに属するデータである。例えば、ベルトコンベアーを流れる製品から不良品を検出するモデルを生成する場合、訓練データ14としては、製品の画像データなどが該当する。

【0020】

入力データ15は、生成されたモデルに入力される各データであり、異常か否かの判定対象となるデータである。上記例で説明すると、製品の画像データを訓練データ14に用いてVAEの機械学習が実行された場合、入力データ15としては、ベルトコンベアーに流れる製品の画像などが該当する。

40

【0021】

モデル16は、制御部20によって生成されるモデルである。具体的には、モデル16は、訓練データ14を用いた機械学習により訓練されたVAEを適用したモデルである。

【0022】

制御部20は、情報処理装置10全体を司る処理部であり、学習部21と検出部22を有する。学習部21は、訓練データ14を用いてVAEの機械学習を実行して、モデル16を生成する処理部である。この学習部21は、後述する図3に示した機械学習により訓

50

練された V A E を適用したモデル 1 6 を生成して記憶する 1 3 に格納する。

【 0 0 2 3 】

図 3 は、V A E の構成および機械学習を説明する図である。図 3 に示すように、V A E は、符号化部 2 1 a ($f(x)$)、雑音生成部 2 1 b、復号化部 2 1 c ($g(z)$)、推定部 2 1 d (R)、最適化部 2 1 e (θ) を有する。

【 0 0 2 4 】

ここで、V A E の機械学習について説明する。符号化部 2 1 a は、ドメイン D に属する訓練データ x が入力されると、訓練データ x の特徴を圧縮して N 次元の正規分布の平均 $\mu(x)$ と標準偏差 $\sigma(x)$ とを出力する。そして、雑音生成部 2 1 b は、平均 0、標準偏差 I に従う N 次元のノイズ z を生成する。

10

【 0 0 2 5 】

雑音生成部 2 1 b により生成されたノイズ z に標準偏差 $\sigma(x)$ を乗じた値を平均 $\mu(x)$ に混入させることによって、標準偏差 $\sigma(x)$ と平均 $\mu(x)$ とに従う正規分布から、サンプリングにより復号化部 2 1 c に入力する潜在変数 z が決定される。そして、復号化部 2 1 c は、訓練データ x の特徴ベクトルに該当する潜在変数 z を用いて、訓練データ x を復号した再構成データを生成する。

【 0 0 2 6 】

その後、推定部 2 1 d は、符号化部 2 1 a から出力される平均 $\mu(x)$ と標準偏差 $\sigma(x)$ を用いて、訓練データ x から計算した潜在変数 z の確率分布と潜在変数 z の事前確率分布との誤差である正規化誤差 R を推定する。そして、最適化部 2 1 e は、推定部 2 1 d により推定される正規化誤差 R を最小化するとともに、訓練データ x と再構成データとの誤差である再構成誤差を最小化するように、符号化部 2 1 a の各パラメータと復号化部 2 1 c の各パラメータの調整 (機械学習) を実行する。

20

【 0 0 2 7 】

図 2 に戻り、検出部 2 2 は、モデル 1 6 を用いて、入力データから異常データを検出する。具体的には、検出部 2 2 は、V A E によってドメインの訓練データを学習した符号化部 2 1 a を用いて、ドメインの入力データを潜在変数の確率分布のパラメータに変換し、変換したパラメータを利用して入力データの生成確率を計算する。すなわち、検出部 2 2 は、ドメインにおける各入力データの確率分布を、訓練済みの V A E で抽出した潜在変数の確率分布から推定し、推定結果を用いて異常データを検出する。

30

【 0 0 2 8 】

まず、検出部 2 2 は、V A E で同定した潜在変数の確率分布から、入力データ (x) の確率分布 $p(X)$ を計算する。図 4 は、入力データの確率分布の算出を説明する図である。図 4 に示すように、訓練時も検出時も、入力データ (x) は、一旦、明示されていない主成分座標 y に変換された後、スケールが適宜変更されて潜在変数 z に変換されると仮定することができる。

【 0 0 2 9 】

具体的には、入力データ (x) に対して、正規直交変換のようなカルーネン・レーベ展開 (KLT : Karhunen - Loeve Expansion) および主成分分析 (PCA : Principal Component Analysis) を実行することにより、分散が大きい主成分の確率分布 ($p(y_1)$) から分散が小さい主成分の確率分布 ($p(y_m)$) が生成される。その後、各分散に対応する確率分布を正規分布の潜在変数に変換する際に、分散が等しくなるようにスケール調整が行われると仮定する。例えば、分散が大きい主成分の確率分布 ($p(y_1)$) から変換される潜在変数の確率分布 ($p(z_1)$) も、分散が小さい主成分の確率分布 ($p(y_m)$) から変換される潜在変数の確率分布 ($p(z_m)$) も、同じ分散となるようにスケールが調整される。

40

【 0 0 3 0 】

このような仮定に従うと、図 4 の (a) に示すように、潜在変数の確率分布 $p(z)$ を逆変換することで、主成分の確率分布 $p(y)$ を生成することができる。そして、図 4 の (b) に示すように、この主成分の確率分布 $p(y)$ をさらに逆変換することで、入力デ

50

ータ (x) の確率分布 p (X) を生成することができる。

【 0 0 3 1 】

さらに、検出部 2 2 は、入力データから推定した潜在変数での確率密度を、VAEの符号化部 2 1 a から出力される標準偏差により補正する。仮定した主成分の確率分布 (p (y)) から潜在変数 z の確率分布 (p (z)) へ変換するときスケールが変化する場合、変換後の確率密度はスケールに比例して変化する。

【 0 0 3 2 】

図 5 は、入力データと各変数の対応を説明する図である。図 5 は、図 4 に示した変換過程の 1 つを選択して表示したものであり、入力データから潜在変数 z の確率分布 (p (z)) への変換を図示している。図 5 に示すように、入力データ (x) から特徴量 (平均 μ (x) 、標準偏差 σ (x)) が生成されるが、変換過程において、このうちの平均 μ (x) の潜在空間における確率分布 (p (μ (x))) は正規分布上に対応付けられる。

10

【 0 0 3 3 】

一方で、標準偏差 σ (x) はデータに混入されたノイズの大きさを表していることから、スケールが変化するとノイズを表す標準偏差 σ (x) も変化する。つまり、入力データに混入されたノイズは既知 (一定の分布) であるとすれば、スケールの変化率は、VAEの符号化部 2 1 a から出力される標準偏差 σ (x) の大きさに現れる。そこで、検出部 2 2 は、最適化式の正規化項の係数 β を用いて、変換スケールを式 (1) のように定義する。これらの結果、入力データ (x) の確率分布 p (X) は、式 (2) で表すことができる。つまり、確率分布 p (X) は、サンプリング確率の一例であり、各入力データ (x) が従う生成確率の分布である。

20

【 0 0 3 4 】

【 数 1 】

$$\left. \frac{dz_j}{dy_j} \right|_{z_j=\mu_j(x)} = \sqrt{2/\beta} \sigma_j(x) \quad \dots \text{式 (1)}$$

(β:最適化式の正規化項の係数)

30

【 数 2 】

$$p(X) = \left| \frac{\partial y}{\partial x} \right| \prod_{j=1}^m p(y_j) = \left| \frac{\partial y}{\partial x} \right| \prod_{j=1}^m \left(\frac{dz_j}{dy_j} p(z_j) \right)_{z_j=\mu_j(x)}$$
$$= (2/\beta)^{\frac{m}{2}} (2\pi)^{-\frac{n-m}{2}} |G_x|^{\frac{1}{2}} \prod_{j=1}^n \left(\sigma_j(x) \frac{p(\mu_j(x))}{N(0,1)} \right) \quad \dots \text{式 (2)}$$

A B

40

【 0 0 3 5 】

なお、式 (2) では、m 個の主成分を用いた式を例示しているが、標準偏差 σ (x) の - 2 乗の平均値 「 σ (x) ⁻² のバー」は、各主成分でのデータ分散 (固有値) と対応するので、どの主成分の分散が大きいのかの順番を特定することができる。したがって、所定次元に圧縮するときには、圧縮効果の高い主成分を選択することができる。

【 0 0 3 6 】

ここで、式 (2) における項目 A は、主成分以外での確率であり、一定値 (平均 μ (x) = 0 、標準偏差 σ (x) = 1) となり、項目 B は、主成分での確率に対応する。つまり、図 4 で示した変換の仮定において、例えば p (y_m) のように分散が小さい主成分の確率分

50

布などは、主成分が分散しないので定数として扱う。

【0037】

したがって、式(3)に示すように、入力データ(x)の確率分布 $p(X)$ はスケールの比例として表現できる。なお、式(3)における項目Cは、再構成誤差のスケールであり、再構成誤差に正規分布を仮定した場合に式(4)で定義できる。つまり、入力データ(x)の確率分布 $p(X)$ は、式(3)の項目Dで定義でき、潜在変数の(x)によって潜在変数の確率分布を入力データの生成確率を反映するように補正することができる。

【0038】

【数3】

$$p(X) \propto \underbrace{|G_x|}_{C} \underbrace{\frac{1}{N_n(0, I_n)} \prod_{j=1}^n \sigma_j(x)}_{D} \quad \dots \text{式(3)}$$

10

【数4】

$$|G_x| = \left(\frac{1}{2\sigma^2} \right)^m \quad \dots \text{式(4)}$$

20

【0039】

ここで、上述した入力データの確率分布の算出処理を詳細に説明する。具体的には、VAEと同様に、m次元のドメインの入力データxに対する潜在変数zの確率分布 $q(z|x)$ をm次元ガウス分布 $N(\mu(x), \sigma(x))$ とし、そのパラメータ $\mu(x), \sigma(x)$ を符号化部21(f(x))で同定する。そして、同定した分布からサンプリングした式(5)に示す潜在変数zを用いて、式(6)に示すように再構成データを復号化部21c(g(z))で推定する。また、符号化部21aと復号化部21cの各パラメータは、式(7)を最小化する機械学習により最適化される。

【0040】

【数5】

$$z = \mu(x) + \sigma(x) \odot \epsilon, \epsilon \sim N(0, I) \quad \dots \text{式(5)}$$

30

【数6】

$$\text{再構成データ } \hat{x} \rightarrow \hat{x} = g_\theta(z) \quad \dots \text{式(6)}$$

【数7】

$$\beta \cdot R + E \sim \beta \cdot D_{KL}(q_\phi(z|x) \| p(z)) + E_{z \sim q_\phi(z|x)}[-\log p_\theta(x|z)] \quad \dots \text{式(7)}$$

40

【0041】

その後、検出部22は、獲得したzについて、式(8)を満たすzをn個(n次元)、または、学習した入力データについて符号化部21aの出力の標準偏差 $\sigma(x)$ の-2乗の平均値を計算して大きい順にn個を抽出する。そして、検出部22は、ドメインの入力データ(x)を符号化部21a(f(x))で分布パラメータ $(\mu(x), \sigma(x))$ に変換し、式(9)により入力データ(x)の生成確率 $p(x)$ を推定する。つまり、生成確率 $p(x)$ は、各々の入力データ(x)に対する生成確率であり、各入力データ(x)が

50

、式(2)の確率分布 $p(x)$ に従うと定義できる結果、各入力データの生成確率 $p(x)$ は、式(9)で定義できる。したがって、1つの入力データに対して、1つの生成確率 $p(x)$ が計算され、それらが複数集まることで確率分布 $p(x)$ が構成される。

【0042】

【数8】

$$\{z_j \in z | D_{KL}(q_\varphi(z_j|x) || p(z_j)) \neq 0\} \quad \dots \text{式(8)}$$

【数9】

$$p(x) = \prod_{j=1}^n \left(\sigma_j(x) \frac{p(\mu_j(x))}{N(0,1)} \right) \quad \dots \text{式(9)}$$

10

【0043】

そして、検出部22は、複数の入力データ15に対して、式(9)を用いて計算した生成確率のデータから、例えば全体の10%のように一定割合の下位データを異常データとして検出する。

【0044】

次に、学習処理の流れを説明する。ここでは、訓練データ14を訓練データ x として説明する。図6は、学習処理の流れを示すフローチャートである。図6に示すように、学習部21は、訓練データ x を符号化部21aに入力し、符号化部21aにより訓練データ x を符号化し、潜在変数 z の分布パラメータ $(\mu(x), \sigma(x))$ を取得する(S101)。

20

【0045】

続いて、学習部21は、所定数の潜在変数 z をサンプリングして N 次元のデータを生成する(S102)。そして、学習部21は、 N 次元のデータを復号化部21c $(g(z))$ に入力して訓練データ x を復号化したデータを取得する(S103)。

【0046】

その後、学習部21は、推定部21dにより推定される正規化誤差 R と、訓練データ x と再構成データとの誤差である再構成誤差 E とを用いた学習コストを算出し(S104)、学習コストが小さくなるように、符号化部21aと復号化部21cの各パラメータ $(\mu(x), \sigma(x))$ を更新する(S105)。

30

【0047】

その後、機械学習が収束していない場合(S106: No)、次の訓練データ14についてS101以降が繰り返される。一方、学習部21は、機械学習が収束した場合(S106: Yes)、機械学習が完了したVAEを適用したモデル16を生成する。なお、学習部21は、学習回数が閾値以上となった場合や復元誤差が閾値以下となった場合に、機械学習が収束したと判定することができる。

【0048】

次に、検出処理の流れを説明する。ここでは、入力データ15を入力データ (x) として説明する。図7は、検出処理の流れを示すフローチャートである。図7に示すように、検出部22は、入力データ (x) を読み込み(S201)、入力データ (x) を訓練済みのVAEの符号化部21aに入力して符号化し、潜在変数 z の分布パラメータ $(\mu(x), \sigma(x))$ を取得する(S201)。

40

【0049】

続いて、検出部22は、取得した分布パラメータ $(\mu(x), \sigma(x))$ をもとに、式(9)を用いて入力データ (x) の生成確率 $p(x)$ を算出する(S202)。ここで、検出部22は、未処理の入力データ (x) が存在する場合(S204: No)、次の入力データ (x) についてS201以降を繰り返す。

50

【 0 0 5 0 】

そして、検出部 2 2 は、全ての入力データ (x) について処理が完了すると (S 2 0 4 : Yes)、生成確率 $p(x)$ が低い順に、一定割合の入力データ (x) を異常データと検出する (S 2 0 5)。

【 0 0 5 1 】

上述したように、情報処理装置 1 0 は、V A E の符号化部 2 1 a から出力される標準偏差を用いて、入力データの確率分布を推定することができる。したがって、情報処理装置 1 0 は、潜在空間の確率分布が実空間の分布を反映することを保証でき、V A E の符号化部 2 1 a の出力を用いた高精度の異常検出を実行することができる。また、情報処理装置 1 0 は、異常検出 (アノマリ検出) など、実ドメインでの確率分布を扱うタスクにおいて、実ドメインのデータが潜在的に単変数のガウス分布で表せる場合に、複雑な分布を用いることなく、入力データの異常検出を実行することができ、計算コストを低減することができる。入力データは、例えば、画像データや音声データであっても良い。

10

【 0 0 5 2 】

ここで、人工的に生成した入力データを用いた検証を説明する。ここでは、式 (8) により入力データの生成確率 $p(x)$ を推定する実施例 1 と、V A E の潜在空間をそのまま用いて入力データの生成確率 ($p(\mu(x))$) を推定する参考技術とのそれぞれの検証結果を説明する。図 8 は、検証用に人工的に生成した入力データを説明する図であり、図 9 は、参考技術を用いた異常検出結果を説明する図であり、図 1 0 は、実施例 1 を用いた異常検出結果を説明する図である。

20

【 0 0 5 3 】

図 8 に示すように、確率密度関数 (PDF : Probability Density Function) にしたかった 3 つのデータを用いて入力データを生成する。例えば、PDF が「 0 から 1 」の範囲となる $p(S_1)$ に属する変数と、PDF が「 (ルート 2) / 4 」となる $p(S_2)$ に属する変数と、PDF が「 (ルート 3) / 6 から 0 」となる $p(S_3)$ に属する変数とを乗算した 3 次元の入力データを生成する。

【 0 0 5 4 】

そして、複数の入力データに対して、参考技術を用いて入力データの生成確率を推定した結果を図 9 であり、実施例 1 を用いて入力データの生成確率を推定した結果を図 1 0 である。図 9 の横軸は生成確率 $p(x)$ であり、縦軸は推定確率 ($p(\mu(x))$) である。図 1 0 の横軸は式 (9) で算出される生成確率 $p(x)$ であり、縦軸は式 (3) で算出される推定確率である。図 9 に示すように、参考技術では、入力データの生成確率が分散してしまい、異常と検出する範囲を判定することが難しく、異常検出の精度は高くない。一方で、図 1 0 に示すように、実施例 1 では、入力データの生成確率が線形となり、生成確率が低い一定割合を正確に特定して異常と検出することができるので、異常検出の精度が向上している。

30

【 0 0 5 5 】

上記実施例で用いた数値、データ、次元数等は、あくまで一例であり、任意に変更することができる。また、V A E を機械学習する装置と、異常検知を実行する装置とを別々の装置で実現することもできる。

40

【 0 0 5 6 】

また、V A E は、図 3 に示した構成以外にも、例えば Rate - Distortion 理論を応用した V A E の構成などを採用することもできる。図 1 1 は、V A E の別例を説明する図である。図 1 1 に示すように、Rate - Distortion 理論を応用した V A E は、符号化部 2 1 a、雑音生成部 2 1 b、復号化部 2 1 c - 1、復号化部 2 1 c - 2、推定部 2 1 d、最適化部 2 1 e を有する。

【 0 0 5 7 】

符号化部 2 1 a は、ドメイン D に属する訓練データ x が入力されると、訓練データ x の特徴を圧縮して N 次元の正規分布の平均 $\mu(x)$ と標準偏差 $\sigma(x)$ を出力する。復号化部 2 1 c - 1 は、符号化部 2 1 a から出力される平均 $\mu(x)$ を用いて入力データを復号した

50

再構成データを生成する。復号化部 2 1 c - 2 は、雑音生成部 2 1 b が生成したノイズが標準偏差 (x) に混入された後、ノイズを含む標準偏差 (x) と平均 $\mu(x)$ とを用いて、入力データを復号した再構成データを生成する。

【 0 0 5 8 】

その後、推定部 2 1 d は、符号化部 2 1 a から出力される平均 $\mu(x)$ と標準偏差 (x) を用いて、訓練データ (x) の確率分布と潜在変数 z の確率分布との正規化誤差 R を推定する。そして、最適化部 2 1 e は、推定部 2 1 d により推定される正規化誤差 R を最小化するとともに、復号化部 2 1 c - 1 が生成した再構成データと入力データとの誤差である再構成誤差 $D 1$ および復号化部 2 1 c - 1 が生成した再構成データと復号化部 2 1 c - 2 が生成した再構成データとの誤差である再構成誤差 $D 2$ を最小化するように、符号化部 2 1 a と各復号化部の各パラメータの調整（機械学習）を実行する。

10

【 0 0 5 9 】

なお、機械学習が完了した後は、実施例 1 と同様の手法により、訓練済みの $V A E$ を用いて、入力データの確率分布を推定することができる。

【 0 0 6 0 】

上記文書中や図面中で示した処理手順、制御手順、具体的名称、各種のデータやパラメータを含む情報については、特記する場合を除いて任意に変更することができる。

【 0 0 6 1 】

また、図示した各装置の各構成要素は機能概念的なものであり、必ずしも物理的に図示の如く構成されていることを要しない。すなわち、各装置の分散や統合の具体的な形態は図示のものに限られない。つまり、その全部または一部を、各種の負荷や使用状況などに応じて、任意の単位で機能的または物理的に分散・統合して構成することができる。

20

【 0 0 6 2 】

さらに、各装置にて行なわれる各処理機能は、その全部または任意の一部が、 $C P U$ および当該 $C P U$ にて解析実行されるプログラムにて実現され、あるいは、ワイヤードロジックによるハードウェアとして実現され得る。

【 0 0 6 3 】

次に、情報処理装置 1 0 のハードウェア構成例を説明する。図 1 2 は、ハードウェア構成例を説明する図である。図 1 2 に示すように、情報処理装置 1 0 は、通信装置 1 0 a、表示装置 1 0 b、 $H D D$ (Hard Disk Drive) 1 0 c、メモリ 1 0 d、プロセッサ 1 0 e を有する。また、図 1 2 に示した各部は、バス等で相互に接続される。

30

【 0 0 6 4 】

通信装置 1 0 a は、ネットワークインタフェースカードなどであり、他のサーバとの通信を行う。表示装置 1 0 b は、学習結果、検出結果などを表示する装置であり、例えばタッチパネルやディスプレイなどである。 $H D D$ 1 0 c は、図 2 に示した機能を動作させるプログラムや $D B$ を記憶する。

【 0 0 6 5 】

プロセッサ 1 0 e は、図 2 に示した各処理部と同様の処理を実行するプログラムを $H D D$ 1 0 c 等から読み出してメモリ 1 0 d に展開することで、図 2 等で説明した各機能を実行するプロセスを動作させる。例えば、このプロセスは、情報処理装置 1 0 が有する各処理部と同様の機能を実行する。具体的には、プロセッサ 1 0 e は、学習部 2 1 と検出部 2 2 等と同様の機能を有するプログラムを $H D D$ 1 0 c 等から読み出す。そして、プロセッサ 1 0 e は、学習部 2 1 と検出部 2 2 等と同様の処理を実行するプロセスを実行する。

40

【 0 0 6 6 】

このように、情報処理装置 1 0 は、プログラムを読み出して実行することで推定方法を実行する情報処理装置として動作する。また、情報処理装置 1 0 は、媒体読取装置によって記録媒体から上記プログラムを読み出し、読み出された上記プログラムを実行することで上記した実施例と同様の機能を実現することもできる。なお、この他の実施例というプログラムは、情報処理装置 1 0 によって実行されることに限定されるものではない。例えば、他のコンピュータまたはサーバがプログラムを実行する場合や、これらが協働してプ

50

プログラムを実行するような場合にも、本発明を同様に適用することができる。

【符号の説明】

【0067】

- 10 情報処理装置
- 11 通信部
- 12 表示部
- 13 記憶部
- 14 訓練データ
- 15 入力データ
- 16 モデル
- 20 制御部
- 21 学習部
- 22 検出部

10

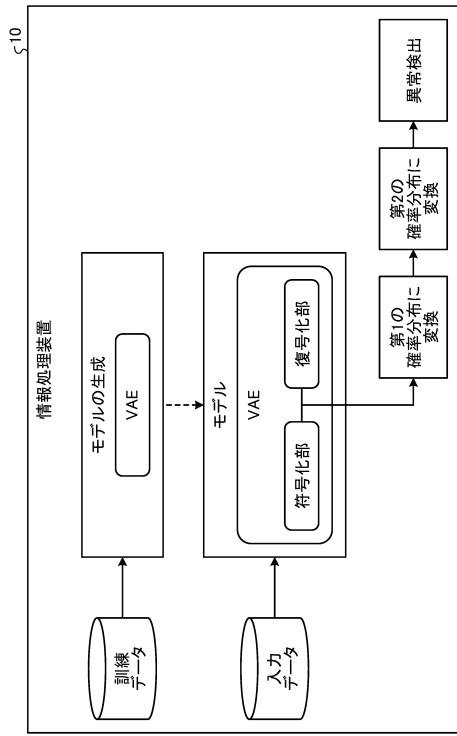
20

30

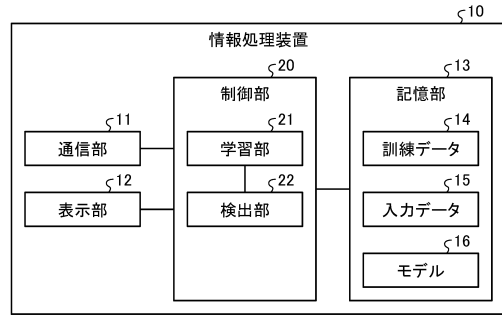
40

50

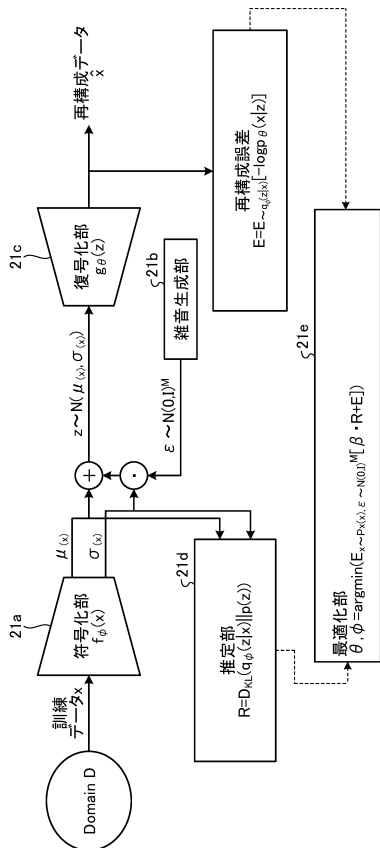
【図面】
【図 1】



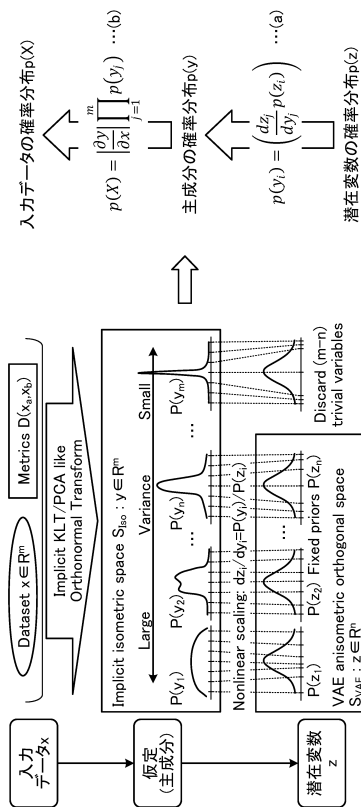
【図 2】



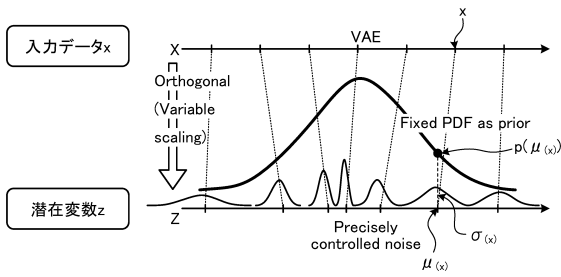
【図 3】



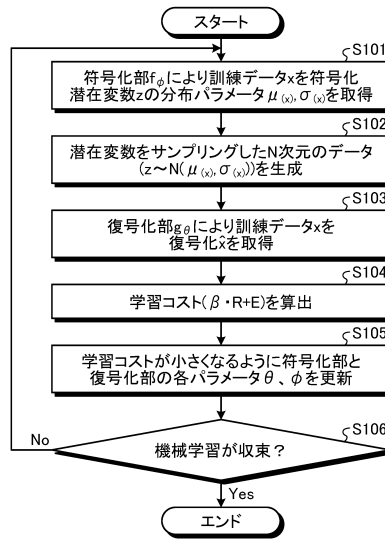
【図 4】



【 図 5 】



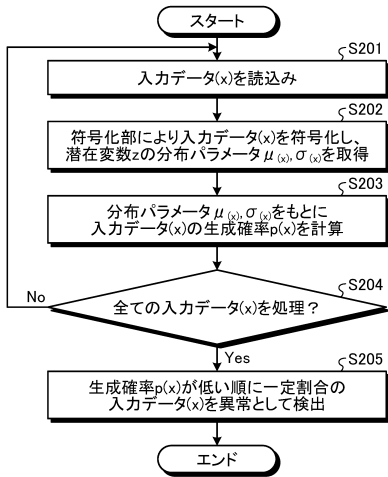
【 図 6 】



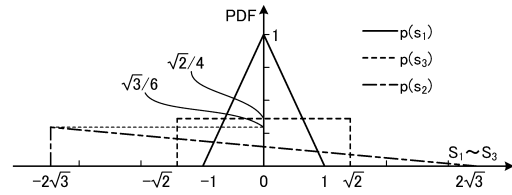
10

20

【 図 7 】



【 図 8 】

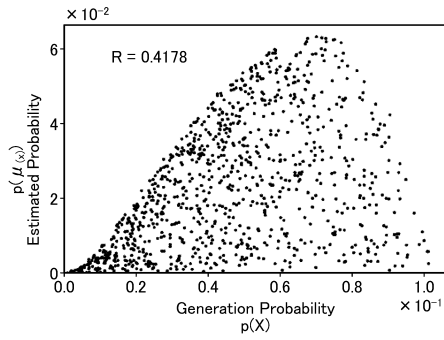


30

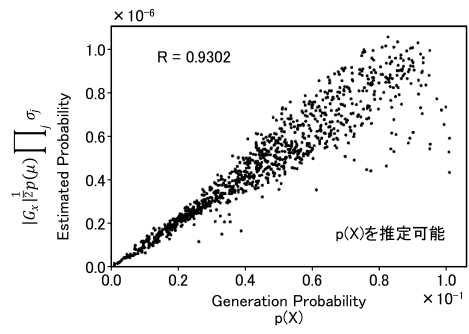
40

50

【図 9】



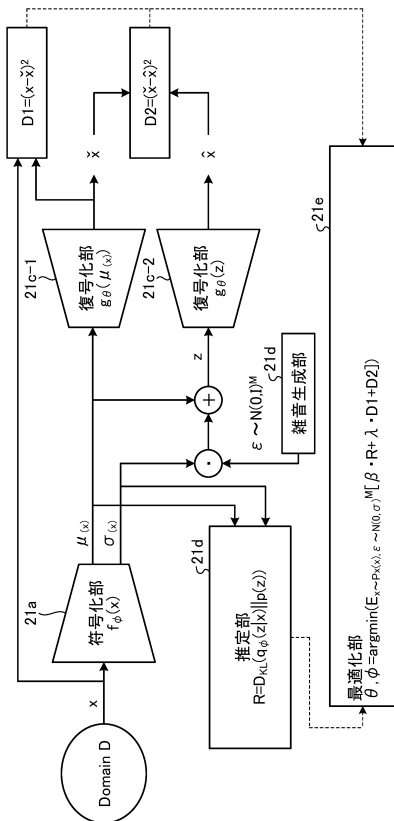
【図 10】



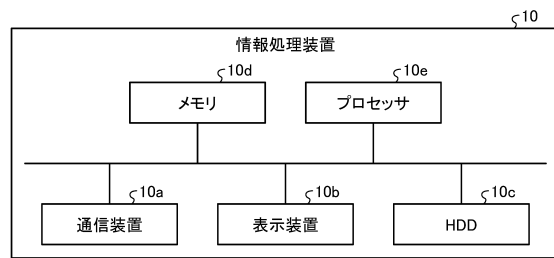
10

20

【図 11】



【図 12】



30

40

50

フロントページの続き

審査官 児玉 崇晶

- (56)参考文献 Keizo Kato et al. , Rate-Distortion Optimization Guided Autoencoder for Isometric Embedding in Euclidean Latent Space [online] , arXiv , [検索日 2023.08.31] , 2020年03月03日 , Internet: URL:<https://arxiv.org/pdf/1910.04329v2.pdf>
- 上原雅俊 ほか , 生成モデルを利用したX線写真に対する異常検知の試行と考察 [online] , 人工知能学会第二種研究会資料 , 2016年11月09日 , 2016巻 AIMED-002号 , pp.10-1 ~ 10.5 , [検索日 2023.08.31] Internet: https://www.jstage.jst.go.jp/article/jsaisigtwo/2016/AIMED-002/2016_10/_article/-char/ja
- (58)調査した分野 (Int.Cl. , D B 名)
G 0 6 N 2 0 / 0 0