

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2018-49355
(P2018-49355A)

(43) 公開日 平成30年3月29日(2018.3.29)

(51) Int.Cl.
G06N 99/00 (2010.01)

F I
G06N 99/00 150

テーマコード (参考)

審査請求 未請求 請求項の数 10 O L (全 16 頁)

(21) 出願番号 特願2016-183085 (P2016-183085)
(22) 出願日 平成28年9月20日 (2016.9.20)

(71) 出願人 000003078
株式会社東芝
東京都港区芝浦一丁目1番1号
(71) 出願人 301063496
東芝デジタルソリューションズ株式会社
神奈川県川崎市幸区堀川町72番地34
(74) 代理人 110001634
特許業務法人 志賀国際特許事務所
(72) 発明者 伊藤 秀将
東京都港区芝浦一丁目1番1号 株式会社東芝内
(72) 発明者 森本 孝司
東京都港区芝浦一丁目1番1号 株式会社東芝内

最終頁に続く

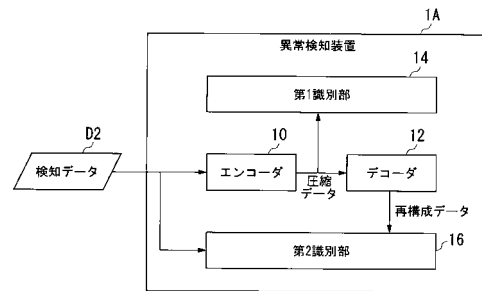
(54) 【発明の名称】 異常検知装置、学習装置、異常検知方法、学習方法、異常検知プログラム、および学習プログラム

(57) 【要約】

【課題】データの異常検知を高精度で行うことができる異常検知装置、学習装置、異常検知方法、学習方法、異常検知プログラム、および学習プログラムを提供することである。

【解決手段】実施形態の異常検知装置は、エンコーダと、第1識別部と、デコーダと、第2識別部とを持つ。前記エンコーダは、正常データに適合する圧縮パラメータを用いて、入力データを圧縮する。前記第1識別部は、前記エンコーダによって圧縮された圧縮データの事前分布上の対数確率密度と、圧縮データ分布および事前分布の対数密度比とに基づいて、前記入力データの第1異常を検知する。前記デコーダは、前記正常データに適合する復号パラメータを用いて、前記エンコーダによって圧縮された圧縮データを復号する。前記第2識別部は、前記デコーダにより復号された再構成データと、前記入力データとの差異を算出し、前記差異に基づいて、前記入力データの第2異常を検知する。

【選択図】 図3



【特許請求の範囲】**【請求項 1】**

正常データに適合する圧縮パラメータを用いて、入力データを圧縮するエンコーダと、前記エンコーダによって圧縮された圧縮データの事前分布上の対数確率密度と、圧縮データ分布および事前分布の対数密度比とに基づいて、前記入力データの第 1 異常を検知する第 1 識別部と、

前記正常データに適合する復号パラメータを用いて、前記エンコーダによって圧縮された圧縮データを復号するデコーダと、

前記デコーダにより復号された再構成データと、前記入力データとの差異を算出し、前記差異に基づいて、前記入力データの第 2 異常を検知する第 2 識別部と

を備える異常検知装置。

10

【請求項 2】

前記第 1 識別部は、前記エンコーダによって圧縮された圧縮データの間層データと、前記事前分布の間層データとを比較して前記第 1 異常を検知し、

前記第 2 識別部は、前記デコーダによって復号された再構成データの間層データと、前記入力データの間層データとを比較して前記第 2 異常を検知する、

請求項 1 に記載の異常検知装置。

【請求項 3】

前記第 1 識別部は、前記エンコーダによって圧縮された圧縮データの前記事前分布上の対数確率密度と、圧縮データ分布および事前分布の対数密度比との和が第 1 閾値以下である場合、前記入力データが前記第 1 異常であることを検知し、

20

前記第 2 識別部は、前記デコーダによって復号された再構成データと、前記入力データとの差異が第 2 閾値以上である場合、前記入力データが前記第 2 異常であることを検知する、

請求項 1 または 2 に記載の異常検知装置。

【請求項 4】

学習データを圧縮するエンコーダと、

入力されたデータの分布が、前記エンコーダによって圧縮された圧縮データの分布と、事前分布とのいずれの分布であることを識別して第 1 識別結果を出力する第 1 識別部と、

前記エンコーダによって圧縮された圧縮データを復号するデコーダと、

30

入力されたデータが、前記デコーダにより復号された再構成データと、前記学習データとのいずれのデータであることを識別して第 2 識別結果を出力する第 2 識別部と

を備え、

前記エンコーダは、前記第 1 識別結果および前記第 2 識別結果に基づいて、前記学習データの圧縮に用いる圧縮パラメータを調整し、

前記デコーダは、前記第 2 識別結果に基づいて、前記圧縮データの復号に用いる復号パラメータを調整する、学習装置。

【請求項 5】

前記エンコーダは、前記圧縮データの分布と前記事前分布との差異が低減するように、前記圧縮パラメータを調整し、

40

前記デコーダは、前記再構成データと前記学習データとの差異が低減するように、前記復号パラメータを調整する、

請求項 4 に記載の学習装置。

【請求項 6】

前記第 1 識別部は、前記エンコーダによって圧縮された圧縮データの間層データと、前記事前分布の間層データとを比較して前記識別を行い、

前記第 2 識別部は、前記デコーダによって復号された再構成データの間層データと、前記学習データの間層データとを比較して前記識別を行う、

請求項 4 または 5 に記載の学習装置。

【請求項 7】

50

正常データに適合する圧縮パラメータを用いて入力データを圧縮し、
 前記入力データを圧縮することにより生成された圧縮データの事前分布上の対数確率密度と、圧縮データ分布および事前分布の対数密度比とに基づいて、前記入力データの第1異常を検知し、
 前記正常データに適合する復号パラメータを用いて、前記入力データを圧縮することにより生成された圧縮データを復号し、
 前記圧縮データを復号することにより生成された再構成データと、前記入力データとの差異を算出し、前記差異に基づいて、前記入力データの第2異常を検知する
 異常検知方法。

【請求項8】

学習データを圧縮し、
 入力されたデータの分布が、前記学習データを圧縮することによって生成された圧縮データの分布と、事前分布とのいずれの分布であるかを識別して第1識別結果を出力し、
 前記学習データを圧縮することによって生成された圧縮データを復号し、
 入力されたデータが、前記圧縮データを復号することにより生成された再構成データと、前記学習データとのいずれのデータであるかを識別して第2識別結果を出力し、
 前記第1識別結果および前記第2識別結果に基づいて、前記学習データの圧縮に用いる圧縮パラメータを調整し、
 前記第2識別結果に基づいて、前記圧縮データの復号に用いる復号パラメータを調整する、
 学習方法。

【請求項9】

コンピュータに、
 正常データに適合する圧縮パラメータを用いて入力データを圧縮させ、
 前記入力データを圧縮することにより生成された圧縮データの事前分布上の対数確率密度と、圧縮データ分布および事前分布の対数密度比とに基づいて、前記入力データの第1異常を検知させ、
 前記正常データに適合する復号パラメータを用いて、前記入力データを圧縮することにより生成された圧縮データを復号させ、
 前記圧縮データを復号することにより生成された再構成データと、前記入力データとの差異を算出し、前記差異に基づいて、前記入力データの第2異常を検知させる
 異常検知プログラム。

【請求項10】

コンピュータに、
 学習データを圧縮させ、
 入力されたデータの分布が、前記学習データを圧縮することによって生成された圧縮データの分布と、事前分布とのいずれの分布であるかを識別させて第1識別結果を出力させ、
 前記学習データを圧縮することによって生成された圧縮データを復号させ、
 入力されたデータが、前記圧縮データを復号することにより生成された再構成データと、前記学習データとのいずれのデータであるかを識別させて第2識別結果を出力させ、
 前記第1識別結果および前記第2識別結果に基づいて、前記学習データの圧縮に用いる圧縮パラメータを調整させ、
 前記第2識別結果に基づいて、前記圧縮データの復号に用いる復号パラメータを調整させる、
 学習プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明の実施形態は、異常検知装置、学習装置、異常検知方法、学習方法、異常検知プ

10

20

30

40

50

プログラム、および学習プログラムに関する。

【背景技術】

【0002】

異常検知とは、データに存在する大多数の傾向をモデル化し、本来あるべきではないデータを発見する技術である。機器の故障検知や予測、ネットワークの不正検知など、様々な分野に応用されており、IoT (Internet of Things) の発展により、さらなる拡大が見込まれている。データの異常を検知するための方法として、オートエンコーダを用いた異常検知方法が知られている。この異常検知方法では、正常データ間における関係性やパターンを利用してデータをできるだけ損失無く圧縮および再構成するモデルを用いて異常の検知を行う。このモデルを用いて正常データを処理した場合、データ損失(再構成誤差)が少ない圧縮および再構成処理が行われるが、異常データを処理した場合、データ損失が大きくなる。この異常検知方法では、このデータ損失の大きさに基づいてデータの異常が検知される。

10

【0003】

上記のオートエンコーダは、高次元のデータ中の高密度の低次元多様体をモデル化している。すなわち、圧縮処理においてデータは多様体へ射影され、再構成処理において多様体が高次元空間へ射影される。例えば、図7に示すように、二次元で表されている複数のデータ(プロット)が、圧縮処理により一次元多様体A上に射影される。また、一次元多様体A上に射影されたデータは、再構成処理により二次元空間に射影される。このため、オートエンコーダが検知する異常(再構成誤差)は、データから多様体へのユークリッド距離(例えば、図7に示す距離D1)に基づいて判定される。例えば、図6に示すデータ番号1の自動車の画像データ(正常データ)をオートエンコーダで処理した場合、元の画像データと、再構成データとの間の再構成誤差が小さいため、正常と判定される。一方、データ番号2の自動車の一部が欠損した画像データ(異常データ)をオートエンコーダで処理した場合、欠損部分が補完された再構成データが生成されるため、元の画像データと、再構成データとの間の再構成誤差が大きくなり異常と判定される。

20

【0004】

しかしながら、上記のオートエンコーダによる処理では、多様体上のデータ密度を考慮していないため、データの異常を検知できない場合がある。例えば、図6に示すデータ番号3の高さ方向に圧縮された自動車の画像データ(異常データ)は、データ密度が低い範囲に位置するデータ(例えば、図7に示すデータB1)である。この画像データをオートエンコーダで処理した場合、元の画像データと、再構成データとの間の再構成誤差は大きくないため正常と判定されてしまう。また、同様に、図6に示すデータ番号4のドアノブが欠損した自動車の画像データ(異常データ)も正常と判定されてしまう。

30

【0005】

また、上記のオートエンコーダによる処理では、再構成誤差を得るために二乗誤差最小化を用いる場合、各データの多様体からの離れ方は一様に正規分布を仮定しているため、正確な再構成誤差を得ることができない場合がある。

【0006】

その他、学習データを用いた画像処理技術として、自由度の高い事前分布を定義可能な敵対的ネットワーク(Generative Adversarial Network: GAN)の研究が進められている。このGANにおいては、事前分布からデータを生成するデコーダと、データが真のデータであるか、生成されたデータであるかを識別する識別器とが用いられる。このGANはデータ生成モデルであり、事前分布からデータを生成する機能を持つが、データを事前分布に逆変換する機能を持たないため、異常検知には利用できない。

40

【先行技術文献】

【特許文献】

【0007】

【特許文献1】米国特許第8352216号明細書

50

【特許文献2】米国特許出願公開第2011/0313726号明細書

【特許文献3】米国特許出願公開第2007/0192863号明細書

【特許文献4】米国特許第7917335号明細書

【非特許文献】

【0008】

【非特許文献1】Jinwon Anら, "Variational Autoencoder based Anomaly Detection using Reconstruction Probability", 2015年12月27日,[online],[2016年9月1日検索], インターネット<URL: <http://dm.snu.ac.kr/static/docs/TR/SNUDM-TR-2015-03.pdf>>

【非特許文献2】日経Robotics, 2016年6月号, p4-5

10

【発明の概要】

【発明が解決しようとする課題】

【0009】

本発明が解決しようとする課題は、データの異常検知を高精度で行うことができる異常検知装置、学習装置、異常検知方法、学習方法、異常検知プログラム、および学習プログラムを提供することである。

【課題を解決するための手段】

【0010】

実施形態の異常検知装置は、エンコーダと、第1識別部と、デコーダと、第2識別部とを持つ。前記エンコーダは、正常データに適合する圧縮パラメータを用いて、入力データを圧縮する。前記第1識別部は、前記エンコーダによって圧縮された圧縮データの事前分布上の対数確率密度と、圧縮データ分布および事前分布の対数密度比との和に基づいて、前記入力データの第1異常を検知する。前記デコーダは、前記正常データに適合する復号パラメータを用いて、前記エンコーダによって圧縮された圧縮データを復号する。前記第2識別部は、前記デコーダにより復号された再構成データと、前記入力データとの差異を算出し、前記差異に基づいて、前記入力データの第2異常を検知する。

20

【図面の簡単な説明】

【0011】

【図1】実施形態の学習装置および異常検知装置の処理の概要を示すフロー図。

【図2】実施形態の学習装置の一例を示す機能ブロック図。

30

【図3】実施形態の異常検知装置の一例を示す機能ブロック図。

【図4】実施形態の学習装置の処理の流れの一例を示すフローチャート。

【図5】実施形態の異常検知装置の処理の流れの一例を示すフローチャート。

【図6】オートエンコーダおよび実施形態の異常検知装置による検知結果の一例を示す図。

【図7】オートエンコーダによるデータの圧縮および再構成処理を説明する図。

【発明を実施するための形態】

【0012】

以下、実施形態の異常検知装置、学習装置、異常検知方法、学習方法、異常検知プログラム、および学習プログラムを、図面を参照して説明する。

40

【0013】

図1は、実施形態の学習装置および異常検知装置の処理の概要を示すフロー図である。学習装置1は、データの異常検知を行うための前準備として、正常データの圧縮処理および復号処理（再構成処理）を行うことで、正常データにおけるデータ間の関係性やパターンを把握し、再構成誤差の小さい圧縮および復号処理を行うための手順を学習する。例えば、学習装置1は、学習データD1の圧縮および復号処理を行い、学習パラメータD1に適合する学習パラメータP1を算出する。

【0014】

学習データD1は、例えば、各種センサによって測定されたセンサデータ、各種機器の動作ログデータ、各種数値データ、各種カテゴリカルデータなど、任意のデータを含む。

50

学習データ D 1 は、これらの種類のデータの正常データを含んでいる。学習データ D 1 は、少量の異常データを含んでいてもよい。以下、学習装置 1 による上記の学習処理が行われる段階を「学習段階」と称する。

【0015】

異常検知装置 1 A は、学習装置 1 によって算出された学習パラメータ P 1 を利用して、検知データ（入力データ）D 2 の異常検知を行い、検知結果 R 1 を出力する。この異常検知装置 1 A は、正常データに対しては再構成誤差の小さい圧縮および復号処理を行うことができるが、正常データと異なる傾向のデータ（異常データ）に対して圧縮および復号処理を行った場合にはその再構成誤差が大きくなる。この再構成誤差とは、正常データと、正常データを圧縮した後で復号することで生成される再構成データとの差異を示す。この再構成誤差を把握することで、異常データを検知することができる。以下、この異常検知装置 1 A による上記の異常検知処理を行う段階を「異常検知段階」と称する。尚、学習装置 1 および異常検知装置 1 A は、その処理段階に応じて名称を使い分けしたものである。このため、学習装置 1 および異常検知装置 1 A は同一の装置であってよい。

10

【0016】

図 2 は、実施形態の学習装置 1 の一例を示す機能ブロック図である。学習装置 1 は、例えば、エンコーダ 1 0 と、デコーダ 1 2 と、第 1 識別部 1 4 と、第 2 識別部 1 6 とを備える。

【0017】

エンコーダ 1 0 は、学習データ D 1 を圧縮し、圧縮データを第 1 識別部 1 4 およびデコーダ 1 2 に出力する。また、エンコーダ 1 0 は、第 1 識別部 1 4 から入力された識別結果に基づいて、学習データ D 1 に適合するように圧縮処理に利用するパラメータ（圧縮パラメータ）の調整を行う。例えば、エンコーダ 1 0 は、エンコーダ 1 0 により生成された圧縮データの分布と、事前分布との差異が低減するように、圧縮パラメータを調整する。事前分布とは、学習段階前に定義される確率分布であり、正規分布、多項分布、一様分布など、任意の分布が定義される。すなわち、エンコーダ 1 0 は、第 1 識別部 1 4 が、エンコーダ 1 0 から入力された圧縮データの分布を事前分布であると判定するように、圧縮パラメータを調整する。

20

【0018】

デコーダ 1 2 は、エンコーダ 1 0 から入力された圧縮データを復号して再構成データを生成し、再構成データを第 2 識別部 1 6 に出力する。また、デコーダ 1 2 は、第 2 識別部 1 6 から入力された識別結果に基づいて、学習データ D 1 に適合するように復号処理に利用するパラメータ（復号パラメータ）の調整を行う。例えば、デコーダ 1 2 は、デコーダ 1 2 により生成された再構成データと学習データ D 1 との差異が低減するように、復号パラメータを調整する。すなわち、デコーダ 1 2 は、第 2 識別部 1 6 が、デコーダ 1 2 から入力された再構成データが学習データ D 1 であると判定するように、復号パラメータを調整する。

30

【0019】

第 1 識別部 1 4 は、入力されたデータの分布が、エンコーダ 1 0 から入力された圧縮データの分布と、予め準備された事前分布とのいずれの分布であるのかを識別し、識別結果（第 1 識別結果）をエンコーダ 1 0 に出力する。例えば、第 1 識別部 1 4 は、ニューラルネットを使用してエンコーダ 1 0 から入力された圧縮データを処理して、最終層データおよび中間層データを生成する。最終層データとは、ニューラルネットの出力値に相当し、中間層データとは、最終層データおよび入力層データ以外の全ての層のデータに相当する。例えば、最終層の一つ手前の層のデータを、中間層データと称する。また、第 1 識別部 1 4 は、ニューラルネットを使用して事前分布からサンプリングされたデータを処理して、最終層データおよび中間層データを生成する。第 1 識別部 1 4 は、圧縮データから得られた中間層データと、事前分布からサンプリングされたデータから得られた中間層データとを比較して上記の識別処理を行う。

40

【0020】

50

第2識別部16は、入力されたデータが、デコーダ12から入力された再構成データと、上記の圧縮および復号処理が施されていない学習データD1とのいずれのデータであるかを識別し、識別結果(第2識別結果)をエンコーダ10およびデコーダ12に出力する。例えば、第2識別部16は、ニューラルネットを使用してデコーダ12から入力された再構成データを処理して、最終層データおよび中間層データを生成する。また、第2識別部16は、ニューラルネットを使用して学習データD1を処理して、最終層データおよび中間層データを生成する。第2識別部16は、再構成データから得られた中間層データと、学習データから得られた中間層データとを比較して上記の識別処理を行う。

【0021】

エンコーダ10およびデコーダ12の各々は、上記の第2識別結果に基づいて、再構成データと学習データD1との差異が低減するように、圧縮パラメータおよび復号パラメータを調整する。すなわち、エンコーダ10およびデコーダ12の各々は、第2識別部16がデコーダ12から入力された再構成データを学習データD1と判定するように、圧縮パラメータおよび復号パラメータを調整する。

10

【0022】

上記の第1識別部14による識別処理およびエンコーダ10によるパラメータの調整処理を複数の学習データに対して繰り返すことで、エンコーダ10は、正常なデータにおけるデータ間の関係性やパターンを把握し、事前分布に近い分布を持つ圧縮データを生成する圧縮処理を行うための手順を学習する。また、上記の第2識別部16による識別処理、並びにエンコーダ10およびデコーダ12によるパラメータの調整処理を複数の学習データに対して繰り返すことで、エンコーダ10およびデコーダ12は、正常なデータにおけるデータ間の関係性やパターンを把握し、再構成誤差の少ない圧縮および復号処理を行うための手順を学習する。

20

【0023】

図3は、実施形態の異常検知装置1Aの一例を示す機能ブロック図である。異常検知装置1Aは、例えば、上記の学習装置1と同様に、エンコーダ10と、デコーダ12と、第1識別部14と、第2識別部16とを備える。

【0024】

エンコーダ10は、検知データD2を圧縮し、圧縮データを第1識別部14およびデコーダ12に出力する。例えば、エンコーダ10は、検知データD2を事前分布にエンコードする。これにより、検知データD2の事前分布上での対数確率を計算することができる。検知データD2は、例えば、各種センサによって測定されたセンサデータ、各種機器の動作ログデータ、各種数値データ、各種カテゴリカルデータなど、任意のデータを含む。

30

【0025】

デコーダ12は、エンコーダ10から入力された圧縮データを復号して再構成データを生成し、再構成データを第2識別部16に出力する。

【0026】

第1識別部14は、エンコーダ10から入力された圧縮データの事前分布上の対数確率密度と、圧縮データ分布および事前分布の対数密度比との和が所定の閾値(第1閾値)以下である場合には異常(第1異常)と判定し、第1閾値よりも大きい場合には正常と判定する。圧縮データ分布および事前分布の対数密度比は、第1識別部14の最終層として出力される圧縮データ確率に対する事前分布確率の商の対数として得られる。すなわち、第1識別部14は、異常データの定義を正常データ分布から見たレアデータとする。データxの出現確率 $p(x)$ をモデル化し、レアデータの検知を行う。第1識別部14は、異常検知結果を表示部(図示しない)に表示してもよい。また、第1識別部14は、異常検知結果を外部の管理端末(図示しない)に出力してもよい。

40

【0027】

第2識別部16は、デコーダ12から入力された再構成データと、検知データD2との差異の程度を示す異常度を算出し、この異常度に基づいて、検知データD2の異常検知を行う。第2識別部16は、例えば、この異常度が所定の閾値(第2閾値)以上の場合には

50

異常（第2異常）と判定し、第2閾値未満の場合には正常と判定する。第2識別部16は、異常検知結果を、図示しない表示部に表示してもよい。また、第2識別部16は、異常検知結果を外部の管理端末に出力してもよい。

【0028】

上記の学習装置1および異常検知装置1Aの各機能部のうち一部または全部は、プロセッサがプログラム（ソフトウェア）を実行することにより実現されてよい。また、学習装置1および異常検知装置1Aの各機能部のうち一部または全部は、LSI（Large Scale Integration）やASIC（Application Specific Integrated Circuit）、FPGA（Field-Programmable Gate Array）等のハードウェアによって実現されてもよいし、ソフトウェアとハードウェアの組み合わせによって実現されてもよい。

10

【0029】

次に、実施形態の学習装置1の動作について説明する。図4は、実施形態の学習装置1の処理の流れの一例を示すフローチャートである。

【0030】

まず、エンコーダ10は、例えば、学習データベース（図示しない）に記憶されている複数の学習データD1（以下、「学習データ群」と称する）の中から少なくとも1つの学習データD1を抽出し、この学習データD1の圧縮を行い、圧縮データを第1識別部14およびデコーダ12に出力する（ステップS101）。尚、エンコーダ10は、学習データベースから上記の学習データD1をランダムに抽出してよい。

【0031】

次に、デコーダ12は、エンコーダ10から入力された圧縮データを復号して再構成データを生成し、再構成データを第2識別部16に出力する（ステップS103）。

20

【0032】

次に、第1識別部14は、入力されたデータが、エンコーダ10から入力される圧縮データのベクトルと、予め準備された事前分布からサンプリングされたベクトルとのいずれのベクトルであるのかを識別する（ステップS105）。例えば、第1識別部14は、圧縮データをニューラルネットで処理して、最終層データおよび中間層データのベクトルを取得する。また、第1識別部14は、事前分布からサンプリングされたベクトルをニューラルネットで処理して、最終層データおよび中間層データのベクトルを取得する。次に、第1識別部14は、圧縮データの中間層データのベクトルの平均と、事前分布からサンプリングされたベクトルの中間層データのベクトルの平均との差分（例えば、差分の二乗）を算出する。この差分は以下の式（1）によって表される。

30

【0033】

【数1】

$$smse \leftarrow (\text{mean}(f_s) - \text{mean}(f'_s))^2 \quad \dots \text{式(1)}$$

上記式（1）において、 $\text{mean}(f_s)$ は圧縮データの中間層データのベクトルの平均を示し、 $\text{mean}(f'_s)$ は事前分布からサンプリングされたベクトルの中間層データのベクトルの平均を示し、 $smse$ がこれらの差分を二乗した値を示す。

40

【0034】

次に、第2識別部16は、入力されたデータが、デコーダ12から入力された再構成データと、圧縮および復号処理が施されていない学習データD1とのいずれのデータであるのかを識別する（ステップS107）。例えば、第2識別部16は、学習データD1をニューラルネットで処理して、最終層データおよび中間層データのベクトルを取得する。また、第2識別部16は、再構成データをニューラルネットで処理して、最終層データおよび中間層データのベクトルを取得する。次に、第2識別部16は、学習データD1の中間層データのベクトルと、再構成データの中間層データのベクトルとの差分（例えば、差分の二乗）を算出する。この差分は以下の式（2）によって表される。

【0035】

50

【数 2】

$$mse \leftarrow (f - f')^2 \quad \dots \text{式(2)}$$

上記式(2)において、 f は学習データ D 1 の中間層データのベクトルを示し、 f' は再構成データの中間層データのベクトルを示し、 mse がこれらの差分を二乗した値を示す。

【0036】

さらに、第 2 識別部 16 は、事前分布からサンプリングされたベクトルを圧縮することにより得られた圧縮データをニューラルネットで処理して、最終層データおよび中間層データのベクトルを取得する。次に、第 2 識別部 16 は、学習データ D 1 をニューラルネットで処理することにより得られた中間層データのベクトルの平均と、事前分布からサンプリングされたベクトルを圧縮することにより得られた圧縮データの中間層データのベクトルの平均との差分(例えば、差分の二乗)を算出する。この差分は以下の式(3)によって表される。

10

【0037】

【数 3】

$$fmmse \leftarrow (\text{mean}(f) - \text{mean}(f''))^2 \quad \dots \text{式(3)}$$

20

上記式(3)において、 $\text{mean}(f)$ は学習データ D 1 の中間層データのベクトルの平均を示し、 $\text{mean}(f'')$ は事前分布からサンプリングされたベクトルの圧縮データの中間層データのベクトルの平均を示し、 $fmmse$ がこれらの差分を二乗した値を示す。

【0038】

次に、エンコーダ 10 およびデコーダ 12 の各々は、上記の式(1)から(3)に基づいて定義された損失関数を用いて、圧縮パラメータおよび復号パラメータを調整する(ステップ S 109)。損失関数 $loss_{AE}$ は、以下の式(4)によって表される。およびは損失関数の相対的重要度を定義する重みである。

【0039】

30

【数 4】

$$loss_{AE} \leftarrow mse + \beta smse + \gamma fmmse \quad \dots \text{式(4)}$$

【0040】

エンコーダ 10 は、第 1 識別部 14 から入力された第 1 識別結果に基づいて、圧縮データの分布が、事前分布に近づくように、圧縮パラメータを調整する。すなわち、エンコーダ 10 は、第 1 識別部 14 がエンコーダ 10 から入力される圧縮データを事前分布からサンプリングされたベクトルであると判定するように、圧縮パラメータを調整する。また、エンコーダ 10 およびデコーダ 12 の各々は、第 2 識別部 16 から入力された第 2 識別結果に基づいて、デコーダ 12 によって生成される再構成データと、圧縮および復号処理が施されていない学習データ D 1 との差異が低減するように、圧縮パラメータおよび復号パラメータを調整する。すなわち、エンコーダ 10 およびデコーダ 12 の各々は、第 2 識別部 16 がデコーダ 12 から入力された再構成データを学習データ D 1 と判定するように、圧縮パラメータおよび復号パラメータを調整する。

40

【0041】

パラメータの調整処理においては、上記の損失関数 $loss_{AE}$ を低減させるように、例えば、以下の式(5)から(8)を用いて圧縮パラメータおよび復号パラメータが調整される。

【0042】

【数5】

$$W_{enc} \leftarrow W_{enc} - \alpha \frac{\partial loss_{AE}}{\partial W_{enc}} \quad \dots \text{式(5)}$$

上記式(5)において、 W_{enc} は、エンコーダ10のニューラルネットの持つ重みパラメータ(コネクシオンの強さ)であり、 α は係数(例えば、0以上1以下の係数)である。右辺の微分計算は、損失関数 $loss_{AE}$ を低減させる W_{enc} の勾配を求めている。この勾配方向に α 値をかけて、 W_{enc} を更新する。この W_{enc} の勾配計算および更新処理においては、確率的勾配降下法(Stochastic Gradient Descent: SGD)と呼ばれる学習処理を使用してもよいし、その他の学習アルゴリズムを使用してもよい。

10

【0043】

【数6】

$$W_{dec} \leftarrow W_{dec} - \alpha \frac{\partial loss_{AE}}{\partial W_{dec}} \quad \dots \text{式(6)}$$

上記式(6)において、 W_{dec} はデコーダ12のニューラルネットの持つ重みパラメータ(コネクシオンの強さ)であり、 α は係数(例えば、0以上1以下の係数)である。右辺の微分計算は、損失関数 $loss_{AE}$ を低減させる W_{dec} の勾配を求めている。この勾配方向に α 値をかけて、 W_{dec} を更新する。

20

【0044】

【数7】

$$b_{enc} \leftarrow b_{enc} - \alpha \frac{\partial loss_{AE}}{\partial b_{enc}} \quad \dots \text{式(7)}$$

上記式(7)において、 b_{enc} はバイアスであり、 α は係数(例えば、0以上1以下の係数)である。右辺の微分計算は、損失関数 $loss_{AE}$ を低減させる b_{enc} の勾配を求めている。この勾配方向に α 値をかけて、 b_{enc} を更新する。

【0045】

【数8】

$$b_{dec} \leftarrow b_{dec} - \alpha \frac{\partial loss_{AE}}{\partial b_{dec}} \quad \dots \text{式(8)}$$

上記式(8)において、 b_{dec} はバイアスであり、 α は係数(例えば、0以上1以下の係数)である。右辺の微分計算は、損失関数 $loss_{AE}$ を低減させる b_{dec} の勾配を求めている。この勾配方向に α 値をかけて、 b_{dec} を更新する。

30

【0046】

さらに、上記のパラメータの調整処理の後、第1識別部14および第2識別部16の学習損失を低減させるように、第1識別部14および第2識別部16のパラメータが調整される。第1識別部14の学習損失 $loss_{smalldis}$ および第2識別部16の学習損失 $loss_{dis}$ は、例えば、以下の式(9)および(10)によって表される。

40

【0047】

【数9】

$$loss_{smalldis} \leftarrow -\log y_s[0] - \log y'_s[1] \quad \dots \text{式(9)}$$

上記式(9)において、 $-\log y_s[0]$ は、第1識別部14が、入力された圧縮データを圧縮データとして識別できるほど小さくなる項であり、 $-\log y'_s[1]$ は、事前分布からサンプリングされたベクトルをこのベクトルとして識別できるほど小さくなる項である。第1識別部14は、この学習損失 $loss_{smalldis}$ を最小化するようにパラメータの調整を行う。 $y_s[0]$ および $y'_s[1]$ はプログラムで扱う際のベ

50

クトルのインデックスあり、第1識別部14が出力する二次元ベクトル y_s の1次元目および2次元目に対応する。

【0048】

【数10】

$$loss_{dis} \leftarrow -\log y[0] - \log \hat{y}[1] \quad \dots \text{式(10)}$$

上記式(10)において、 $-\log y[0]$ は、第2識別部16が、入力された学習データD1を学習データD1として識別できるほど小さくなり項であり、 $-\log y'[1]$ は、再構成データを再構成データとして識別できるほど小さくなる項である。第2識別部16は、この学習損失 $loss_{dis}$ を最小化するようにパラメータの調整を行う。 $y[0]$ および $y'[1]$ はプログラムで扱う際のベクトルのインデックスあり、第2識別部16が出力する二次元ベクトル y の1次元目および2次元目に対応する。

10

【0049】

上記の第1識別部14および第2識別部16における調整処理においては、上記の学習損失 $loss_{smalldis}$ および学習損失 $loss_{dis}$ を低減させるように、例えば、以下の式(5)から(8)を用いてパラメータが調整される。

【0050】

【数11】

$$W_{smalldis} \leftarrow W_{smalldis} - \alpha \frac{\partial loss_{smalldis}}{\partial W_{smalldis}} \quad \dots \text{式(11)}$$

20

上記式(11)において、 $W_{smalldis}$ は、第1識別部14のニューラルネットの持つ重みパラメータ(コネクシオンの強さ)であり、 α は係数(例えば、0以上1以下の係数)である。右辺の微分計算は、学習損失 $loss_{smalldis}$ を低減させる $W_{smalldis}$ の勾配を求めている。この勾配方向に α 値をかけて、 $W_{smalldis}$ を更新する。

【0051】

【数12】

$$W_{dis} \leftarrow W_{dis} - \alpha \frac{\partial loss_{dis}}{\partial W_{dis}} \quad \dots \text{式(12)}$$

30

上記式(12)において、 W_{dis} は、第2識別部16のニューラルネットの持つ重みパラメータ(コネクシオンの強さ)であり、 α は係数(例えば、0以上1以下の係数)である。右辺の微分計算は、学習損失 $loss_{dis}$ を低減させる W_{dis} の勾配を求めている。この勾配方向に α 値をかけて、 W_{dis} を更新する。

【0052】

【数13】

$$b_{smalldis} \leftarrow b_{smalldis} - \alpha \frac{\partial loss_{smalldis}}{\partial b_{smalldis}} \quad \dots \text{式(13)}$$

40

上記式(13)において、 $b_{smalldis}$ はバイアスであり、 α は係数(例えば、0以上1以下の係数)である。右辺の微分計算は、学習損失 $loss_{smalldis}$ を低減させる $b_{smalldis}$ の勾配を求めている。この勾配方向に α 値をかけて、 $b_{smalldis}$ を更新する。

【0053】

【数14】

$$b_{dis} \leftarrow b_{dis} - \alpha \frac{\partial loss_{dis}}{\partial b_{dis}} \quad \dots \text{式(14)}$$

上記式(14)において、 b_{dis} はバイアスであり、 α は係数(例えば、0以上1以

50

下の係数)である。右辺の微分計算は、学習損失 l_{oss}_{dis} を低減させる b_{dis} の勾配を求めている。この勾配方向に η 値をかけて、 b_{dis} を更新する。

【0054】

次に、エンコーダ10は、学習データベースに記憶されている学習データD1の抽出が完了したか否かを判定する(ステップS111)。エンコーダ10は、学習データD1の抽出が完了していないと判定した場合、残りの学習データD1のうち少なくとも1つを抽出し、上述の圧縮および復号処理、並びにパラメータ調整処理を行う。一方、エンコーダ10は、学習データD1の抽出が完了したと判定した場合、上記の学習データ群に対する1回の学習処理サイクルが完了したと判断し、学習処理の回数を計数する。例えば、エンコーダ10は、内部に設けられたメモリ(図示しない)上に設定された処理回数計測用のパラメータをインクリメントすることで、上記の学習処理の回数を計数する。

10

【0055】

次に、エンコーダ10は、上記の学習処理の回数が、所定の処理回数に到達したか否かを判定する(ステップS113)。エンコーダ10は、上記の学習処理の回数が、所定の処理回数未満であると判定した場合、上記の学習データ群に対する学習処理を再度行う。このように、同一の学習データ群に対して学習処理を繰り返すことで、パラメータ調整の精度を向上させることができる。一方、エンコーダ10は、上記の学習処理の回数が、所定の処理回数以上であると判定した場合、本フローチャートの処理を終了する。

【0056】

尚、上記の学習装置1における学習処理においては、多層構造のニューラルネットワーク(Deep Neural Network: DNN)、畳み込みニューラルネットワーク(Convolutional Neural Network: CNN)、再帰型ニューラルネットワーク(Recurrent Neural Network: RNN)を採用してもよい。

20

【0057】

次に、実施形態の異常検知装置1Aの動作について説明する。図5は、実施形態の異常検知装置1Aの処理の流れの一例を示すフローチャートである。図6は、実施形態の異常検知装置1Aによる検知結果の一例を示す図である。

【0058】

まず、エンコーダ10は、検知データD2を圧縮し、圧縮データを第1識別部14およびデコーダ12に出力する(ステップS201)。

30

【0059】

次に、第1識別部14は、エンコーダ10から入力された圧縮データの異常検知を行う(ステップS203)。例えば、第1識別部14は、エンコーダ10から入力された圧縮データの事前分布上の対数確率密度と、圧縮データ分布および事前分布の対数密度比との和が所定の閾値(第1閾値)以下である場合には異常(第1異常)と判定し、第1閾値よりも大きい場合には正常と判定する。圧縮データ分布および事前分布の対数密度比は、第1識別部14の最終層として出力される圧縮データ確率に対する事前分布確率の商の対数として得られる。

【0060】

40

この第1識別部14による異常検知処理においては、多様体上の密度を考慮し、多様体上の密度が低い範囲に位置するデータを異常と判定し、多様体上の密度が高い範囲に位置するデータを正常と判定する。例えば、エンコーダ10が、図6に示すデータ番号1の自動車の画像を示すデータ(正常データ)の圧縮データを第1識別部14に出力した場合、この圧縮データは多様体上の密度が高い範囲に位置するデータであるため、第1識別部14は、この圧縮データを正常と判定する。一方、図6に示すデータ番号3の自動車の高さ方向に圧縮された画像を示すデータ(異常データ)は、多様体上の密度が低い範囲に位置するデータである。エンコーダ10が、この画像データの圧縮データを第1識別部14に出力した場合、この圧縮データは多様体上の密度が第1閾値よりも低いため、第1識別部14はこの圧縮データを異常と判定する。

50

【0061】

次に、デコーダ12は、エンコーダ10から入力された圧縮データを復号して再構成データを生成し、再構成データを第2識別部16に出力する(ステップS205)。

【0062】

次に、第2識別部16は、デコーダ12から入力された再構成データと、検知データD2との差異の程度を示す異常度を算出し、この異常度に基づいて、データの異常検知を行う(ステップS207)。第2識別部16は、例えば、この異常度が所定の閾値(第2閾値)以上の場合には異常(第2異常)と判定し、第2閾値未満の場合には正常と判定する。

【0063】

例えば、エンコーダ10が、検知データD2として図6に示すデータ番号1の自動車の画像を示すデータ(正常データ)を圧縮し、デコーダ12が圧縮データを復号して第2識別部16に出力した場合、この再構成データは検知データD1と同様なデータであるため、第2識別部16は、この検知データD2を正常と判定する。一方、エンコーダ10が、検知データD2として図6に示すデータ番号2の自動車の一部が欠損した画像を示すデータ(異常データ)や、データ番号4の自動車のドアノブが無い画像を示すデータ(異常データ)のデータを圧縮し、デコーダ12が圧縮データを復号して第2識別部16に出力した場合、この再構成データと検知データD1との差異が大きい(異常度が第2閾値よりも高い)データであるため、第2識別部16は、この検知データD2を異常と判定する。以上により、本フローチャートの処理を終了する。

【0064】

以上説明した少なくとも一つの実施形態によれば、正常データに適合する圧縮パラメータを用いて、入力データを圧縮するエンコーダと、前記エンコーダによって圧縮された圧縮データの事前分布上の対数確率密度と、圧縮データ分布および事前分布の対数密度比とに基づいて、前記入力データの第1異常を検知する第1識別部と、前記正常データに適合する復号パラメータを用いて、前記エンコーダによって圧縮された圧縮データを復号するデコーダと、前記デコーダにより復号された再構成データと、前記入力データとの差異を算出し、前記差異に基づいて、前記入力データの第2異常を検知する第2識別部とを備えることにより、データの異常検知を高精度で行うことができる。

【0065】

また、上述した実施形態における学習装置1および異常検知装置1Aの一部の機能をコンピュータで実現するようにしてもよい。その場合、この機能を実現するためのプログラムをコンピュータ読み取り可能な情報記録媒体に記録する。そして、上述したプログラムを記録した情報記録媒体に記録されたプログラムをコンピュータシステムに読み込ませ、実行することによって実現してもよい。なお、ここでいう「コンピュータシステム」とは、オペレーティングシステムや周辺機器等のハードウェアを含むものとする。また、「コンピュータ読み取り可能な情報記録媒体」とは、可搬媒体や記憶装置等のことをいう。可搬媒体は、フレキシブルディスク、光磁気ディスク、ROM、CD-ROM等である。また、記憶装置は、コンピュータシステムに内蔵されるハードディスク等である。

【0066】

さらに「コンピュータ読み取り可能な情報記録媒体」とは、通信回線を介してプログラムを送信する場合の通信線のように、短時間の間、動的にプログラムを保持するものである。通信回線は、インターネット等のネットワークや電話回線等である。また、「コンピュータ読み取り可能な情報記録媒体」は、サーバやクライアントとなるコンピュータシステム内部の揮発性メモリであってもよい。揮発性メモリは、一定時間プログラムを保持しているものである。また上記プログラムは、前述した機能の一部を実現するためのものであってもよい。また上記プログラムは、さらに前述した機能をコンピュータシステムに既に記録されているプログラムとの組み合わせで実現できるものであってもよい。

【0067】

本発明のいくつかの実施形態を説明したが、これらの実施形態は、例として提示したも

10

20

30

40

50

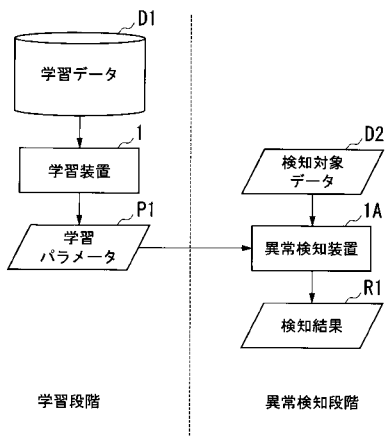
のであり、発明の範囲を限定することは意図していない。これら実施形態は、その他の様々な形態で実施されることが可能であり、発明の要旨を逸脱しない範囲で、種々の省略、置き換え、変更を行うことができる。これら実施形態やその変形は、発明の範囲や要旨に含まれると同様に、特許請求の範囲に記載された発明とその均等の範囲に含まれるものである。

【符号の説明】

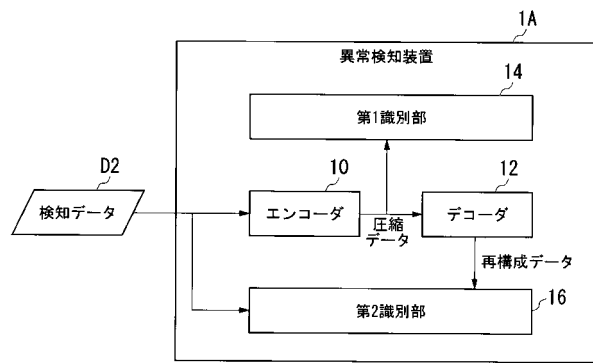
【0068】

1 ... 学習装置、1A ... 異常検知装置、10 ... エンコーダ、12 ... デコーダ、14 ... 第1識別部、16 ... 第2識別部

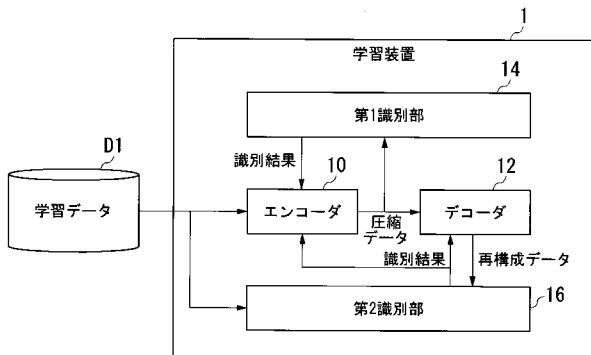
【図1】



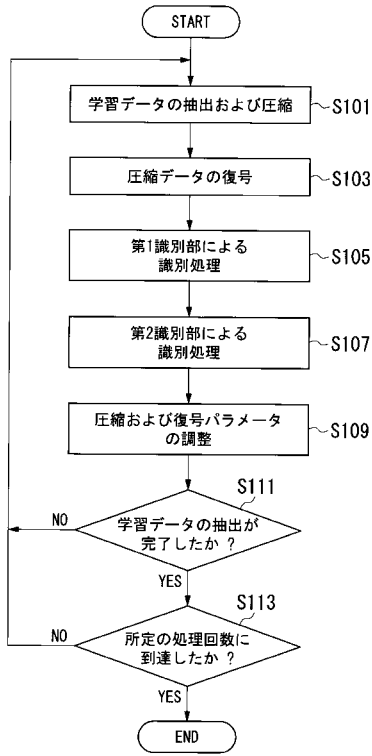
【図3】



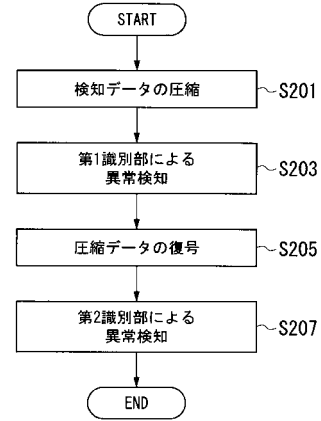
【図2】



【 図 4 】



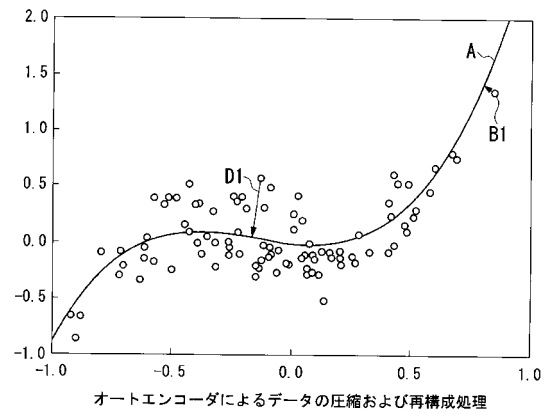
【 図 5 】



【 図 6 】

データ番号	従来のオートエンコーダ		実施形態の異常検知装置	
	再構成データ	検知結果	再構成データ	検知結果
1		正常		正常
2		異常		異常
3		正常		異常
4		正常		異常

【 図 7 】



フロントページの続き

- (72)発明者 高橋 信太郎
東京都港区芝浦一丁目1番1号 株式会社東芝内
- (72)発明者 加藤 利幸
東京都港区芝浦一丁目1番1号 株式会社東芝内