敵対的生成ネットワークを用いた集団型異常検知

丸　千尋[1]　小林一郎[2]

Generative Adversarial Network (GAN) は，現実世界の高次元分布をモデル化することができ，異常検知にも適用され始めている．しかし，GANを用いた異常検知の既存研究は，特定時点の観測値を扱うモデルであるため，観測値自体は正常であるが，その観測値のふるまいが変化する集団型異常を検知することはできない．そこで本研究では，時系列データに存在する集団型異常をGANを用いて検知することを目的とする．集団型異常の検知に向け，提案するGANモデルのEncoderとGeneratorにsequence to sequenceのEncoder側とDecoder側を，DiscriminatorにRecurrent Neural Networkと全結合Neural Networkをそれぞれ採用した．さらに，人工データと自然データの二種類のデータセットを用いて評価実験を行い，提案モデルの有効性を検証した．

1　はじめに

あらゆるものがインターネットと繋がる，Internet of Thingsの出現により，機械や設備等に取り付けられた各種センサーから膨大な時系列データを容易に収集することが可能になっている．同時に，これらのデータ活用の一つとして，大量のデータをリアルタイムに監視することによって，平常時と異なる状況の発生やその予兆を検知可能な異常検知が盛んに行われている [1]．例えば，クレジットカードの不正利用の検出，病気の診断の援助，サイバーセキュリティの侵入検知，及び安全性が重視されるシステムの障害検知等，様々なアプリケーションで広く使用されている．

現在の異常検知においては，教師あり異常検知と教師なし異常検知の2つの手法が存在する．教師あり異常検知は，入力データにあらかじめ付けられた正常／異常の正解のラベルに基づき，異常判定モデルを学習する手法である．一方，教師なし異常検知は，入力データに正解ラベルを付けず，正常データのみから成る入力データを用いて異常判定モデルを学習する手法である．これら2つの手法のうち，教師あり異常検知では，複数の問題が存在する．まず1つ目は，機械や設備等に異常が発生することは稀であるため，異常データを大量に収集することが難しいことである．2つ目は，ラベル付けは人手で行われるため，ラベル付けがされているデータが少ないことである．したがって，近年，教師なし異常検知が盛んに研究されている．この手法を用いれば，正常なデータのみを用いてあらかじめ異常判定モデルを学習しておくことで，モデルから異なる異常な観測値が与えられたとき，未知の異常を含めた異常を検知することが可能となる．

近年，現実世界の高次元分布をモデル化することができる，Generative Adversarial Network (GAN) [2]が提案されており，教師なし異常検知にGANが適用され始めている [3–5]．例えば，Efficient GAN [4]は，上記のGANを用いて，正常データのモデルをあらかじめ学習しておき，このモデルに従わない観測値が与えられたときに異常と判定する．しかし，Efficient GANは，集団型異常を検知することができない．Efficient GANは特定時点の観測値を扱うモデルであるため，各時点の観測値の単変量もしくは多変量を考慮した際に，他の観測値から値が大きく異なる，点異常／文脈依存型異常を検知することはできる．一方で，観測値自体は正常であるが，その観測値のふるまいが変化する集団型異常については，Efficient GANは複数の観測値を扱うことができないため，検知することができない．また，多次元の集団型異常検知のために，LSTMやAutoEncoderといった，Neural Network (NN) を用いた手法が提案されているが，実運用のためには精度が低いというのが現状である．そこで，本研究は，GANモデルを用いて時系列データに潜む集団型異常を検知することを目的とする．

集団型異常を検知するため，Efficient GANのEncoder，Generator及びDiscriminatorをそれぞれ複数の観測値を扱えるネットワークに拡張した，Multivariate Anomaly detection with Recurrent Units-GAN (MARU-GAN) を提案する．具体的には，Encoderにsequence to sequence (seq2seq) [6] のEncoder側を，Generatorにseq2seqのDecoder側を，そしてDiscriminatorにRecurrent Neural Network (RNN) と全結合NNを採用することで，時系列データに対応することが可能になる．このMARU-GANに対して，正常な時系列データから成る，SWaTデータセット [7] とWADIデータセット [8]の一部の正常な観測値を他の正常な観測値と入れ替えて生成した新たなデータセットを用いて評価を行った．その結果，集団型異常を検知するためには複数の観測値を扱うネットワークを利用する必要があること，我々のMARU-GANは複数の観測値を扱うネットワークを採用した既存手法と比較して，高い精度で集団型異常を検知できることが明らかになった．さらに，自然データであるてんかん患者の皮質脳波の信号に対してMARU-GANを適用し，自然データに対してもMARU-GANが有効であることを示した．

2　関連研究

異常検知とは，予測されるふるまいに適合しないデータ内のパターンを見つける技術である．異常検知における異常は，点異常，文脈依存型異常，集団型異常の3種類に分類でき，点異常は最も単純な種類の異常で，異常検知のための研究の多くはこの異常を対象にしている [9]．確率分布に基づく異常検知 [10–12]，クラスタリングに基づく異常検知 [13–16]，最近傍法に基づく異常検知 [17, 18]，分類に基づく手法 [19, 20]は点異常を検知するために提案されている．文脈依存型異常の検知には，点異常を検知するための手法を拡張したものを使うことができる．例えば，Hayesら [21]は，クラスタリングアルゴリズムを用いた手法を提案している．まず，クラスタリングアルゴリズムを用いて，文脈

[1] 学生会員　お茶の水女子大学
maru.chihiro@is.ocha.ac.jp
[2] 正会員　お茶の水女子大学
koba@is.ocha.ac.jp

依存型属性に基づいて各観測値を分類する．そして，複数の観測値から成る各クラスタごとに，分類器が文脈依存型異常であるかを判定する．したがって，点異常と文脈依存型異常は，個々の観測値を用いることで異常を検知することができるため，似た手法を使うことができる．

一方，集団型異常は，個々の観測値自体は異常ではないが，複数の観測値が集まったときの挙動が異常であるため，異なる戦略を取らなくてはならない．集団型異常の検知の例として，Keoghら [22]とLinら [23]は，スライディングウィンドウを用いて，与えられた時系列データから部分時系列を抽出し，最も近い部分時系列との距離を異常度とする手法を提案した．クラスタリングや最近傍法に基づく集団型異常検知のための手法は多く存在するが，実世界における機械や設備での運用を考慮すると，多次元のデータに対応する必要がある．

近年，多次元の集団型異常検知のために，LSTMやAutoEncoderといった，NNを用いた手法が提案されている．LSTMを用いた異常検知 [24–26] では，正常なデータのみから成る訓練データを用いて，入力されたd点の観測値から次のl点を予測するLSTMを学習する．そして，学習されたLSTMから予測された値と実際の値がどの程度異なるかによって，異常度を算出する．AutoEncoderを用いた異常検知 [27] は，正常なデータのみから成る訓練データを用いて，入力された部分時系列をそのまま復元するAutoEncoderを学習する．そして，学習されたAutoEncoderから復元された値と実際の値がどの程度異なるかによって，異常度を算出する．しかし，これらの手法は，実運用のためにはまだ精度が低いというのが現状である．

近年，NNの一つとして，現実世界の高次元分布をモデル化することができる，GANが提案されており，異常検知にも適用され始めている [3–5]．Efficient GAN [4]は，画像およびネットワーク侵入データセットにおいて，最先端の性能を達成したことを示し，異常検知にGANを用いることが有用であることを明らかにした．しかし，Efficient GANは，特定時点の観測値を扱うモデルであるため，観測値自体は正常であるが，その観測値のふるまいが変化した異常を検知することができない．MAD-GAN [5]は，時系列データの異常検知のためのGANモデルであるが，時系列データの中の他の観測値から大きく異なる観測値を検知しており，集団型異常を検知することを目的にしていない．そこで，我々は，GANを用いて時系列データに潜む集団型異常を検知することを目指す．

3 Efficient GANに基づいた異常検知

本章では，GANを異常検知に用いたEfficient GANと，それがもつ問題点について説明する．

3.1 Efficient GANの詳細

Efficient GANは，3つのネットワーク，Generator，Encoder及びDiscriminatorから構成される．このモデルの全体像を図1に示す．

Efficient GANは正常なデータのみから成るデータセットを用いて学習される．Generator (G) は，潜在空間のノイズをデータ空間にマッピングすることにより，実データ$\boldsymbol{x}$のデータ分布$p_{data(\boldsymbol{x})}(\boldsymbol{x})$を学習する．そして，潜在変数$\boldsymbol{z}$を与えたとき，$p_{data(\boldsymbol{x})}(\boldsymbol{x})$を用いて実データに近い正常データ$G(\boldsymbol{z})$を生成する．Encoder (E) は，Generatorの逆の機能を有し，データ$\boldsymbol{x}$を潜在空間にマッピングし$E(\boldsymbol{x})$を得る．Discriminator (D) は，データと潜在変数の組から成る入力データ $((\boldsymbol{x}, E(\boldsymbol{x}))$もしくは$(G(\boldsymbol{z}), \boldsymbol{z}))$ が与えられたときに，それが実データ（本物）であるのか，Generatorによって生成されたデータ (偽物) であるのかを識別し，与えられた入力データが本物である確率$P(y)$を出力する．

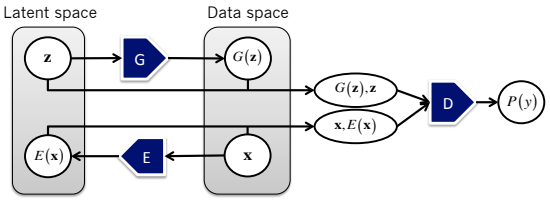

図1 Efficient GANの全体像

Generatorの最終的な目標は，Discriminatorに本物と識別されるような，実データに近い正常データを生成することである．Encoderの場合は，Discriminatorに実データ$\boldsymbol{x}$を偽物だと識別されるように，データを潜在空間にマッピングすることである．Discriminatorは，Generatorに騙されないよう，与えられた入力データが本物であるのか，偽物であるのかを正確に識別することを目標とする．上記の目標を達成するため，以下の目的関数(1)をGenerator，Encoder，及びDiscriminatorで共有し，Discriminatorに関しては目的関数を最大化，GeneratorとEncoderに関しては最小化するように交互に学習する．

$$V(D,E,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{data}(\boldsymbol{x})}[\log(D(\boldsymbol{x}, E(\boldsymbol{x})))] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z}), \boldsymbol{z}))] \quad (1)$$

ここで，$p_{data}(\boldsymbol{x})$は実データ分布，$p_{\boldsymbol{z}}(\boldsymbol{z})$は潜在変数$\boldsymbol{z}$の分布で，訓練データを用いて学習される．$G(\boldsymbol{z})$は潜在変数$\boldsymbol{z}$をGeneratorに与えたときに生成されるデータを表す．さらに，$D(\boldsymbol{x}, E(\boldsymbol{x}))$と$D(G(\boldsymbol{z}), \boldsymbol{z})$は，入力データ$(\boldsymbol{x}, E(\boldsymbol{x}))$もしくは$(G(\boldsymbol{z}), \boldsymbol{z})$がDiscriminatorに与えられたときにDiscriminatorによって出力される，入力データが本物である確率を表す．

正常データのみを用いて学習されたEfficient GANを用いて，未知の入力データ$\boldsymbol{x}$に対して，異常度$A(\boldsymbol{x})$ (式(2)) を算出する．$A(\boldsymbol{x})$は，再構築損失$L_G(\boldsymbol{x})$ (式(3)) と識別損失$L_D(\boldsymbol{x})$ (式(4)) の2つの項から構成される．αは係数である．$A(\boldsymbol{x})$の値が大きくなるほど，$\boldsymbol{x}$が異常であるということを意味する．

$$A(\boldsymbol{x}) = \alpha L_G(\boldsymbol{x}) + (1-\alpha) L_D(\boldsymbol{x}) \quad (2)$$

$$L_G(\boldsymbol{x}) = \|\boldsymbol{x} - G(E(\boldsymbol{x}))\|_1 \quad (3)$$

$$L_D(\boldsymbol{x}) = \sigma(D(\boldsymbol{x}, E(\boldsymbol{x})), 1) \quad (4)$$

再構築損失$L_G(\boldsymbol{x})$は，未知の入力データ$\boldsymbol{x}$と再構築されたデータ$G(E(\boldsymbol{x}))$の$L1$ノルムである．$G(E(\boldsymbol{x}))$は，

Encoderを使って$\boldsymbol{x}$を潜在変数$E(\boldsymbol{x})$にマッピングした後，$E(\boldsymbol{x})$をGeneratorに与えることで再構築されたデータである．EncoderとGeneratorは正常データを用いて学習されているため，$\boldsymbol{x}$が正常である場合，再構築された$G(E(\boldsymbol{x}))$は$\boldsymbol{x}$に似たデータとなるはずである．一方，$\boldsymbol{x}$が異常な場合は，EncoderとGeneratorが対応していないため，$G(E(\boldsymbol{x}))$は$\boldsymbol{x}$と大きく異なるデータとなる．よって，$L_G(\boldsymbol{x})$の値が大きくなる．

識別損失$L_D(\boldsymbol{x})$は，未知の入力データ$\boldsymbol{x}$と，それをEncoderを使ってマッピングした潜在変数$E(\boldsymbol{x})$の組をDiscriminatorが本物であると識別する確率と，クラス1の交差エントロピー損失σである．ここで，クラス1は入力データ$\boldsymbol{x}$が本物であることを意味する．交差エントロピーでは，$D(\boldsymbol{x}, E(\boldsymbol{x}))$の値が0に近くなる，すなわちDiscriminatorによって入力データ$\boldsymbol{x}$が偽物であると識別される程，$L_D(\boldsymbol{x})$の値が大きくなる．Discriminatorは正常データの識別を正確に行えるように学習されているため，異常データ$\boldsymbol{x}$が与えられると，Discriminatorは$\boldsymbol{x}$を偽物だと識別し，$D(\boldsymbol{x}, E(\boldsymbol{x}))$が0に近い値となる．その結果，$L_D(\boldsymbol{x})$の値が大きくなる．

3.2 異常検知における異常の種類

異常検知における異常は，3種類に分類することができる [9]．1つ目は，他の観測値から大きく異なる観測値を検知する，点異常である．例えば，気温の年間の推移において，気温が100度である観測値は点異常であると判定される． 2つ目は，多変量のデータにおいて，特定の状況において異常である観測値を検知する，文脈依存型異常である．例えば，6月の気温において，気温が0度である観測値は文脈依存型異常であると判定される．年間を通して気温が0度になる可能性はあるが，6月に0度になることは，通常起こり得ないことである．これは，時期と気温の2つの変数を考えたときに異常だと判定される．3つ目は，他のデータと比べ，ふるまいが異なる観測値の集まりを検知する，集団型異常である．これは，観測値自体は正常であるが，その観測値が複数集まった時のふるまいが変化したことを意味しており，時系列データに存在する異常である．例えば，人間の心電図のふるまいが変化したとき，その観測値の集まりは集団型異常であると判定される．

点異常，文脈依存型異常はそれぞれ，1点の観測値の1次元，K次元 (K >1) を扱えば検知することができる．一方で，集団型異常はN点の観測値のK次元 (K ≥1) を扱えば検知することができる．

3.3 Efficient GANの問題点

3.1節 のEfficient GANにおけるEncoder， Generator及びDiscriminatorは，観測値を個々に扱うネットワークであるため，特定時点の観測値の異常である点異常及び文脈依存型異常を検知することはできる．しかし，観測値自体は正常であるが，その観測値のふるまいに異常が存在するような，複数の観測値を扱うことによって異常を検知することが可能となる集団型異常を検知することはできない．そこで本研究では，GANモデルを用いて，時系列データに潜む集団型異常を検知することを目的とする．

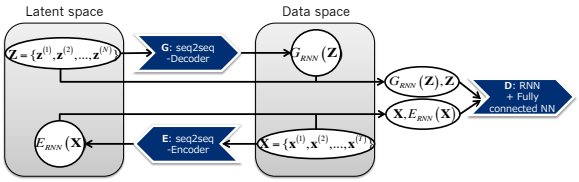

図2 MARU-GANの全体像

4 GANを用いた集団型異常検知

本稿では，長さTの部分時系列$\boldsymbol{X} = \{\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots, \boldsymbol{x}^{(T)}\}$を考える．時刻$t_i$の観測値$\boldsymbol{x}^{(i)} \in \mathbb{R}^M$は，$M$個の変数を持つ$M$次元ベクトルである．本稿では，このような長さ$T$の複数の部分時系列$\boldsymbol{X}$が時系列データを構成していると仮定する．部分時系列$\boldsymbol{X}$を扱うため，Efficient GAN の Encoder，Generator及びDiscriminatorをそれぞれ複数の観測値を扱えるネットワークに拡張した，Multivariate Anomaly detection with Recurrent Units-GAN (MARU-GAN) を提案する．MARU-GANの全体像を図2に示す．Efficient GANのEncoderにseq2seqのEncoder側，GeneratorにseqseqのDecoder側，DiscriminatorにRNNと全結合NNを採用する． MARU-GANのEncoder， Generator及びDiscriminatorについて，それぞれ詳細に説明する．

4.1 Encoder

Encoderには，seq2seqのEncoder側を採用する． Encoderは長さTの固定長の部分時系列$\boldsymbol{X} = \{\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots, \boldsymbol{x}^{(T)}\}$を入力部分時系列$\boldsymbol{X}$として受け取り，Encoderの最後の隠れ状態$(h_1^{(T)}, h_2^{(T)}, h_3^{(T)})$を$E_{RNN}(\boldsymbol{X})$として出力する．この$E_{RNN}(\boldsymbol{X})$には，入力部分時系列の特徴が圧縮されている．本稿では，隠れ層が多い程性能が良いことが明らかなため [6]，Encoderの隠れ層を3ユニットとして実装した． Encoderは式(5)の目的関数を最小化するように学習される．つまり，Dicsriminatorに入力部分時系列を偽物であると判定されるように学習される。

$$V(E) = \mathbb{E}_{\boldsymbol{X} \sim p_{data}(\boldsymbol{X})}[\log(D_{RNN}(\boldsymbol{X}, E_{RNN}(\boldsymbol{X})))] \quad (5)$$

4.2 Generator

Generatorには， seq2seqのDecoder側を採用する．Generatorは潜在変数$\boldsymbol{Z}$を受け取り，それをGeneratorの最初の隠れ状態$(\boldsymbol{s_1}^{(1)}, \boldsymbol{s_2}^{(1)}, \boldsymbol{s_3}^{(1)})$に設定する．そして，開始を表す入力を与えると，各時刻t_iで$\boldsymbol{x}^{(i)}$が出力される．全ての$\boldsymbol{x}^{(i)}$を結合した部分時系列$\boldsymbol{X} = \{\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots, \boldsymbol{x}^{(T)}\}$を$G_{RNN}(\boldsymbol{Z})$とする．本稿では，Generatorの隠れ層を3ユニットとして実装した．Generatorは式(6)の目的関数を最小化するように学習される．つまり，Dicsriminatorに入力部分時系列を本物であると判定されるように学習される．

$$V(G) = \mathbb{E}_{\boldsymbol{Z} \sim p_{\boldsymbol{Z}}(\boldsymbol{Z})}[\log(1 - D_{RNN}(G_{RNN}(\boldsymbol{Z}), \boldsymbol{Z}))] \quad (6)$$

4.2.1 Discriminator

Discriminatorには，RNNと全結合NNの2種類のNNを用いる． Discriminatorは，データ (部分時系列) と潜在変数を

結合したベクトルを入力として受け取る．RNNでは，部分時系列$G_{RNN}(\boldsymbol{Z})$もしくは，$\boldsymbol{X}$を入力として，最後の隠れ状態$\boldsymbol{h}_{\boldsymbol{dis}}^{(T)}$を得る．この$\boldsymbol{h}_{\boldsymbol{dis}}^{(T)}$には，入力された部分時系列の特徴が圧縮されている．そして，得られた$\boldsymbol{h}_{\boldsymbol{dis}}^{(T)}$と，潜在変数$\boldsymbol{Z}$もしくは，$E_{RNN}(\boldsymbol{X})$を結合し，全結合NNに入力する．全結合NNは，入力された部分時系列が本物である確率$P(y)$を出力し，これがDiscriminatorの出力となる． Discriminatorは式(7)の目的関数を最大化するように学習される．つまり，Dicsriminatorは入力部分時系列を正確に識別するように学習される．

$$V(D) = \mathbb{E}_{\boldsymbol{X}\sim p_{data}(\boldsymbol{X})}[\log(D_{RNN}(\boldsymbol{X}, E_{RNN}(\boldsymbol{X})))] + \mathbb{E}_{\boldsymbol{Z}\sim p_{\boldsymbol{Z}}(\boldsymbol{Z})}[\log(1 - D_{RNN}(G_{RNN}(\boldsymbol{Z}), \boldsymbol{Z}))] \quad (7)$$

4.2.2 異常度の算出

4節のEncoder，Generator及び Discriminatorを正常な部分時系列のみから構成される訓練データを用いて学習した後，各部分時系列$\boldsymbol{X}$の異常度$A_{RNN}(\boldsymbol{X})$を，式(2)を拡張した式(8)を用いて算出する．

$$A_{RNN}(\boldsymbol{X}) = \alpha L_{G_{RNN}}(\boldsymbol{X}) + (1-\alpha) L_{D_{RNN}}(\boldsymbol{X}) \quad (8)$$

$$L_{G_{RNN}}(\boldsymbol{X}) = \|\boldsymbol{X} - G_{RNN}(E_{RNN}(\boldsymbol{X}))\|_1 \quad (9)$$

$$L_{D_{RNN}}(\boldsymbol{X}) = \sigma(D_{RNN}(\boldsymbol{X}, E_{RNN}(\boldsymbol{X})), 1) \quad (10)$$

$L_{G_{RNN}}$ (式(9)) においては，入力部分時系列$\boldsymbol{X}$に含まれる観測値ごとに算出した値を集計して$\boldsymbol{X}$の$L_{G_{RNN}}$とする． $L_{D_{RNN}}(\boldsymbol{X})$ (式(10)) の$D_{RNN}(\boldsymbol{X}$fi$E_{RNN}(\boldsymbol{X}))$では，Discriminatorが部分時系列$\boldsymbol{X}$ごとに本物だと識別する確率を出力する．上記の$A_{RNN}(\boldsymbol{X})$を用いて，全ての未知の入力部分時系列$\boldsymbol{X}$に対して異常度を算出し，上位N%の部分時系列を異常と判定する．

MARU-GANによる異常検知のアルゴリズムをAlgorithm 1に示す．

5 人工データを用いた実験

4節で提案したMARU-GANが時系列データに潜む集団型異常を検知可能であるかを，SWaTデータセット [7]とWADIデータセット [8]を用いて本章で評価する．我々は，正常な観測値のみから構成されるSWaTデータセットとWADIデータセットに，それぞれ集団型異常を人工的に作成し混ぜることで，新たに生成されたデータセットを用いて実験を行った．比較対象として，GANを用いた特定時点の観測値の異常を検知する手法である，Efficient GAN，AutoEncoderを異常検知に用いた手法である，EncDec-AD，そしてLSTMを異常検知に用いた手法である，LSTM-ADを採用した．このうち，EncDec-ADとLSTM-ADは，集団型異常検知に向けた手法である．

5.1 実験データ

SWaTデータセット[*1]は，最新の水処理プラントを縮小したレプリカから収集された51次元のデータから構成される．WADIデータセット[*2]は，水分散システムから収集された123次元のデータから構成される．どちらのデータセットにおいても，データ

[*1] https://itrust.sutd.edu.sg/testbeds/secure-water-treatment-swat/
[*2] https://itrust.sutd.edu.sg/testbeds/water-distribution-wadi/

Algorithm 1 MARU-GAN-based anomaly detection

Require: K,$\boldsymbol{Z}$,$\boldsymbol{X}$,α,N

for K epochs **do**
  Training:
  Generate time-series subsequences from the latent variables
  $\boldsymbol{Z}$, <START> $\Rightarrow G_{RNN}(\boldsymbol{Z})$
  Map from time-series subsequences to the latent space
  $\boldsymbol{X} \Rightarrow E_{RNN}(\boldsymbol{X})$
  Discriminate time-series subsequences
  $D_{RNN}(G_{RNN}(\boldsymbol{Z}), \boldsymbol{Z})$
  $D_{RNN}(\boldsymbol{X}, E_{RNN}(\boldsymbol{X}))$
  Update the parameters by maximizing $V(D)$
  $V(D) = \mathbb{E}_{\boldsymbol{X}\sim p_{data}(\boldsymbol{X})}[\log(D_{RNN}(\boldsymbol{X}, E_{RNN}(\boldsymbol{X})))] + \mathbb{E}_{\boldsymbol{Z}\sim p_{\boldsymbol{Z}}(\boldsymbol{Z})}[\log(1 - D_{RNN}(G_{RNN}(\boldsymbol{Z}), \boldsymbol{Z}))]$ (equation (7))
  Update the parameters by minimizing $V(G)$
  $V(G) = \mathbb{E}_{\boldsymbol{Z}\sim p_{\boldsymbol{Z}}(\boldsymbol{Z})}[\log(1 - D_{RNN}(G_{RNN}(\boldsymbol{Z}), \boldsymbol{Z}))]$ (equation (6))
  Update the parameters by minimizing $V(E)$
  $V(E) = \mathbb{E}_{\boldsymbol{X}\sim p_{data}(\boldsymbol{X})}[\log(D_{RNN}(\boldsymbol{X}, E_{RNN}(\boldsymbol{X})))]$ (equation (5))
  Save the parameters of Encoder，Generator，and Discriminator in the current epoch
  Validating:
  Compute anomaly scores $A_{RNN}(\boldsymbol{X})$ using validation dataset
  $L_{G_{RNN}}(\boldsymbol{X}) = \|\boldsymbol{X} - G_{RNN}(E_{RNN}(\boldsymbol{X}))\|_1$ (equation (9))
  $L_{D_{RNN}}(\boldsymbol{X}) = \sigma(D_{RNN}(\boldsymbol{X}, E_{RNN}(\boldsymbol{X})), 1)$ (equation (10))
  $A_{RNN}(\boldsymbol{X}) = \alpha L_{G_{RNN}}(\boldsymbol{X}) + (1-\alpha) L_{D_{RNN}}(\boldsymbol{X})$ (equation (8))
  Define N% of the time-series subsequences with the highest $A_{RNN}(\boldsymbol{X})$ as anomalous
  Compute F1-value using the anomaly detection results
end for
Testing:
Restore the model with the highest F1-value in K epochs
Compute anomaly scores $A_{RNN}(\boldsymbol{X})$ using test dataset based on the above model
$L_{G_{RNN}}(\boldsymbol{X}) = \|\boldsymbol{X} - G_{RNN}(E_{RNN}(\boldsymbol{X}))\|_1$ (equation (9))
$L_{D_{RNN}}(\boldsymbol{X}) = \sigma(D_{RNN}(\boldsymbol{X}, E_{RNN}(\boldsymbol{X})), 1)$ (equation (10))
$A_{RNN}(\boldsymbol{X}) = \alpha L_{G_{RNN}}(\boldsymbol{X}) + (1-\alpha) L_{D_{RNN}}(\boldsymbol{X})$ (equation (8))
Define N% of the time-series subsequences with the highest $A_{RNN}(\boldsymbol{X})$ as anomalous
Compute F1-value，accuracy，and false positive rate using the anomaly detection results

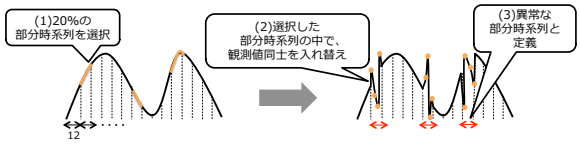

図3　疑似集団型異常時系列データの生成方法

は1秒ごとに測定される．本研究では，475,200 点の正常な観測値から成るSWaTデータセットと，1,048,560点の正常な観測値から成るWADIデータセットと用いて各手法の評価を行った．

5.2　集団型異常の生成

5.1節のSWaTデータセットとWADIデータセットは，どちらも正常な観測値のみから構成されたデータセットである．そこで，集団型異常を人工的に生成し，各データセットにそれぞれ混ぜることで，新しい実験データセットを生成する．

事前処理として，475,200点の正常なSWaTデータ，1,048,560点の正常なWADIデータを訓練／検証／テストデータとしてそれぞれ8:1:1に分割し，部分時系列の長さがT=12のデータセットを生成する．その結果，訓練／検証／テストデータは，SWaTデータセットにおいては，31,680部分時系列，3,960部分時系列，3,960部分時系列，WADIデータセットにおいては，69,904部分時系列，8,738部分時系列，8,738時系列となった．訓練データは，モデルを訓練するために使われる．検証データは，各epochの学習後の評価に使われ，最終的にF値が最も高いepochのモデルを選択するために使われる．テストデータは，検証データを用いて選択されたepochのモデルに対して，異常検知の精度を評価するために使われる．これらのデータのうち，各異常検知のモデルは，正常なデータのみを用いて学習されるため，訓練データはこのまま利用できる．一方，検証・テストデータは評価に使われるため，集団型異常を含む必要がある．

集団型異常とは，観測値自体は正常だが，その観測値のふるまいが変化した種類の異常である．したがって，正常な観測値と他の正常な観測値を入れ替えることで対応した．具体的な集団型異常の生成方法を次に示す．生成方法の概要を図3に示す．

(1) 各データ (検証／テスト) の中で，20%の部分時系列をランダムに選択
(2) 選択した部分時系列の中で，観測値同士をランダムに入れ替え
(3) 選択した20%の部分時系列を異常と定義

(2)で選択された20%の各部分時系列に含まれるT点の観測値を，他の選択された部分時系列の観測値とランダムに入れ替えることで，観測値自体は正常な値であるが，その観測値のふるまいに異常が存在する集団型異常を生成することが可能となる．上記で選択された20%を異常と定義し，MARU-GANの異常検知の結果と比較する．

5.3　比較手法

MARU-GANの比較手法として，Efficient GANと，集団型異常検知に向けた手法であるEncDec-AD，LSTM-ADを採用した．各異常検知の手法について説明する．これらの比較手法は，観測値$\boldsymbol{x}^{(i)}$ごとに異常判定を行うため，5.2節で選択された20%の部分時系列に含まれるT点の観測値全てを異常と定義し，観測値ごとに評価を実施した．本稿では，$T = 12$としたため，検証・テストデータにおける異常と定義された観測値は，SWaTデータセットにおいては，3,960部分時系列×20%×12点=9,504点，WADIデータセットにおいては，8,738部分時系列×20%×12点=104,856点である．

5.3.1　Efficient GAN

3.1節で説明したEfficient GANを用いる．Efficient GANは，観測値$\boldsymbol{x}^{(i)}$ごとに異常度を算出するため，異常度が高い上位20%の観測値を異常と判定する．

5.3.2　EncDec-AD

EncDec-ADはAutoEncoderを用いた異常検知の手法である．まず，正常なデータのみから成る訓練データの一部を用いて，Encoderで入力部分時系列$\boldsymbol{X}$を低次元のベクトルに圧縮し，Decoderで圧縮されたベクトルから時系列$\boldsymbol{X'}$を復元するAutoEncoderを学習する．EncDec-ADは，入力部分時系列$\boldsymbol{X} = \{\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots, \boldsymbol{x}^{(T)}\}$とAutoEncoderによって復元される部分時系列$\boldsymbol{X'} = \{\boldsymbol{x}^{(1)'}, \boldsymbol{x}^{(2)'}, \cdots, \boldsymbol{x}^{(T)'}\}$に対して，目的関数$\sum_{i=1}^{T} \|\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(i)'}\|^2$を最小化するように学習される．そして学習で使用されなかった残りの訓練データを用いて，学習されたAutoEncoderで復元された時系列$\boldsymbol{X}^{(i)'}$と，入力部分時系列$\boldsymbol{X}^{(i)}$における各観測値$\boldsymbol{x}^{(i)}$のエラーベクトル$\boldsymbol{e}^{(i)} = |\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(i)'}|$を算出し，$\boldsymbol{e}^{(i)}$が従う正規分布$N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$の平均$\boldsymbol{\mu}$と標準偏差$\boldsymbol{\Sigma}$を求める．このパラメータを用い，テストデータの各観測値$\boldsymbol{x}^{(i)}$に対して，異常度$A^{(i)} = (\boldsymbol{e}^{(i)} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{e}^{(i)} - \boldsymbol{\mu})$を算出し，この値が高い上位20%の観測値を異常と判定する．

5.3.3　LSTM-AD

LSTM-ADは，LSTMを用いた異常検知の手法である．まず，正常なデータのみから成る訓練データの一部を用いて，d点からl点を予測するLSTMを学習する．LSTM-ADは，LSTMによって予測されるl点$\{\boldsymbol{x}^{(1)'}, \boldsymbol{x}^{(2)'}, \ldots, \boldsymbol{x}^{(l)'}\}$と実データ$\{\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(l)}\}$に対して，目的関数$\sum_{i=1}^{l} \|\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(i)'}\|^2$を最小化するように学習される．そして，学習で使用されなかった残りの訓練データを用いて，学習されたLSTMから予測された各観測値$\boldsymbol{x}^{(i)'}$と，実際の$\boldsymbol{x}^{(i)}$のエラーベクトル$\boldsymbol{e}^{(i)} = |\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(i)'}|$を算出し，$\boldsymbol{e}^{(i)}$が従う正規分布$N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$の平均$\boldsymbol{\mu}$と標準偏差$\boldsymbol{\Sigma}$を求める．このパラメータを用い，テストデータの各観測値$\boldsymbol{x}^{(i)}$に対して，異常度$A^{(i)} = (\boldsymbol{e}^{(i)} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{e}^{(i)} - \boldsymbol{\mu})$を算出し，この値が高い上位20%の観測値を異常と判定する．

5.4　実験設定

本稿の実験で用いたハイパーパラメータについて説明する．表1にMARU-GANのEncoder，Generator及びDiscriminatorの詳細，表2に実験設定を示す．比較のため，各表のハイパーパラメータは，先行研究に基づいて調節した．1000epochの各epochの学習後に検証データを用いてF値を算出し，F値が

表1　MARU-GANのハイパーパラメータ

	ユニット数	層数	ドロップアウト率
Encoder			
$E(\boldsymbol{X})$: RNN	100	3	0.0
Generator			
$G(\boldsymbol{Z})$: RNN	100	3	0.0
Discriminator			
$D(\boldsymbol{X})$: RNN	100	1	0.2
$D(\boldsymbol{X}$fi$\boldsymbol{Z})$: 全結合NN	1	1	0.0

表2　実験設定

データセット	SWaTデータセット, WADIデータセット
勾配法	Adam
ハイパーパラメータ	$\alpha = 1e-5$，$\beta_1 = 0.5$，$\beta_2 = 0.999$，$\varepsilon = 1e-8$
時系列の長さT	12
バッチサイズ	50
Epoch数	1000
潜在変数の次元	100

最も高いepochのモデルを使ってテストデータで評価を実施した．5.2節で異常と定義された部分時系列と，MARU-GANによって算出される異常度が高い20%の部分時系列を比較し，F値，Accuracy，False positive率を求める．

5.5　実験結果と考察

実験結果を表3と表4に示す．MARU-GANは全ての評価値において，他の既存手法よりも高い精度を達成することができた．

表3　実験結果（SWaTデータセット）

Method	F1-value	Accuracy	False positive rate
Efficient GAN [4]	0.18	0.67	0.21
EncDec-AD [27]	0.19	0.68	0.20
LSTM-AD [24]	0.55	0.82	0.11
MAD-GAN [5]	0.46	0.67	0.33
MARU-GAN	**0.62**	**0.85**	**0.09**

表4　実験結果（WADIデータセット）

Method	F1-value	Accuracy	False positive rate
Efficient GAN [4]	0.20	0.68	0.20
EncDec-AD [27]	0.19	0.68	0.20
LSTM-AD [24]	0.45	0.78	0.14
MAD-GAN [5]	0.32	0.59	0.38
MARU-GAN	**0.61**	**0.85**	**0.10**

Efficient GANのような特定時点の観測値を扱うモデルでは，観測値自体は正常であるが，その観測値のふるまいが変化した集団型異常を検知することが不可能であった．Efficient GANでは，F値が実験データに占める異常の割合20%とほぼ変わらず，つまり，異常をランダムに選択してしまっているのと区別がつかない．したがって，集団型異常を検知するためには，複数の観測値を扱うためのネットワークを利用する必要があることが明らかになった．

EncDec-ADとLSTM-ADは，複数の観測値を扱うモデルであるにも関わらず，集団型異常を検知することができなかった．

EncDec-ADは，異常データの異常度が正常データの異常度と変わらないため，異常を検知することができなかった．EncDec-ADは，正常データのみから成る訓練データを用いて，入力部分時系列$\boldsymbol{X} = \{\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots, \boldsymbol{x}^{(T)}\}$とAutoEncoderによって復元される部分時系列$\boldsymbol{X'} = \{\boldsymbol{x}^{(1)'}, \boldsymbol{x}^{(2)'}, \cdots, \boldsymbol{x}^{(T)'}\}$に対して，目的関数$\sum_{i=1}^{T} \|\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(i)'}\|^2$を最小化するように学習される．目的関数を最適化すると，正常データのみを使ってAutoEncoderを学習しているにも関わらず，訓練データ (正常データ) に存在しない異常データに対しても，学習したAutoEncoderを使ってほぼ完全に復元することが可能であった．AutoEncoderは，Encoder側で入力時系列の特徴を圧縮したベクトルを使ってDecoder側で復元を行っている．よって，答えが明らかな状態で復元が行われているため，異常なデータに対してもほぼ完全に復元できてしまうと考えられる．そのため，異常度を算出する際に用いるエラーベクトル$\boldsymbol{e}^{(i)} = |\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(i)'}|$の値が，正常データであっても，異常データであっても，変わらないため，異常を検知することができなかった．一方，MARU-GANの異常度$A_{RNN}(\boldsymbol{X}) = \alpha L_{G_{RNN}}(\boldsymbol{X}) + (1-\alpha) L_{D_{RNN}}(\boldsymbol{X})$においても，EncDec-ADと同様に，再構築誤差$L_{G_{RNN}}(\boldsymbol{X}) = \|\boldsymbol{X} - G_{RNN}(E_{RNN}(\boldsymbol{X}))\|_1$を用いて異常の検出を行っている．しかし，MARU-GANの目的関数 (式(1)) はEncDec-ADとは異なっており，正常な部分時系列を生成するように学習される．つまり，完全な復元を目的にしていない．学習後，EncDec-ADは異常な部分時系列も正常な部分時系列と判定されるように再構築されるが，MARU-GANの場合，それを避けることができる．その結果が再構築誤差$L_{G_{RNN}}$で得られるため，MARU-GANは集団型異常を検知することが可能となる．

LSTM-ADは，異常な部分時系列の前半の観測値を異常と識別することができなかった．これは，LSTMに入力される観測値が少ない前半である程，予測のために必要な手がかりが少ないため，上手く予測ができなかったことが原因と考えられる．また，複数の観測値d点間に異常データのようなランダム性が存在する場合，その後の観測値l点を上手く予測することができなかった．その結果，頻繁にl点が異常だと判定されてしまった．

以上より，集団型異常を検知するためには，複数の観測値を扱うネットワークを利用する必要があること，提案モデルであるMARU-GANは，複数の観測値を扱うネットワークを採用した既存手法と比較して，高い精度で集団型異常を検知できることが明らかになった．

6　自然データを用いた実験

本章では自然データに対してMARU-GANを適用し，その有効性を検証する．自然データとして，てんかん患者の皮質脳波を用いる．これは，患者の脳表に留意された直径1-3mmの皿状の電極から計測された脳信号である．実験においては，72種類の電

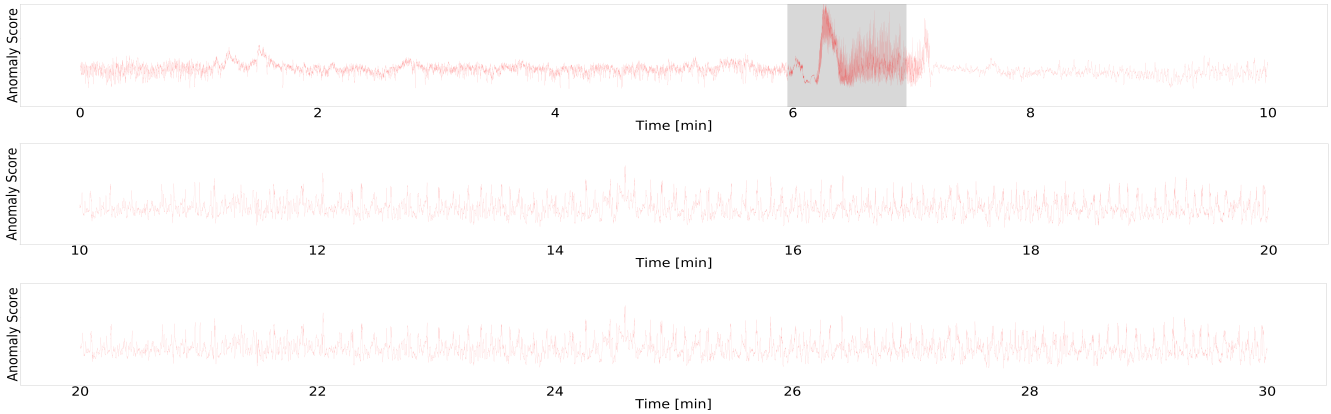

図4　異常度の推移（皮質脳波データセット）

極から取得された72次元のデータを用いる．皮質脳波は，電極直下の脳活動を計測するため空間精度が高く，通常の脳波と比べて多くの脳信号を得ることができる．

先行研究 [28] を参考に部分時系列の長さを40とし，訓練／検証／テストデータを生成した．その結果，訓練／検証／テストデータは，それぞれ16,500部分時系列，36,000部分時系列，36,000部分時系列となった．検証データ，テストデータには，てんかん発作が含まれており，発作が発生した時間が医師によって記録されている．本実験では，検証データを各epochの学習後の評価に利用し，発作が発生している期間における異常度の和が最も大きいモデルを選択した．そして，この選択されたモデルに対して，テストデータを用いて異常検知の評価を行なった．

異常度の推移を図4に示す。横軸は経過時間，縦軸が異常度である。図中の色が付いている区間は，医師によって記録された発作が発生している期間を示す．

結果より，てんかん発作の検知を正しく行えた上に，医師によって認識されていなかった，大きな発作の後にわずかな異常が続いているという事実を捉えることができた．これは，MARU-GANが多次元の時系列データを扱うため，他の電極と相関があるような小さい異常を見るけることが可能になったことを示している。我々は，MARU-GANが自然データに対しても有効であることを検証することができた．

7　おわりに

本論文では，時系列データに含まれる集団型異常を検知する手法を提案した．集団型異常を検知するため，Efficient GANのEncoder，Generator及び Discriminatorをそれぞれ複数の観測値を扱えるよう，Encoderにseq2seqのEncoder側を，Generatorにseq2seqのDecoder側を，そしてDiscriminatorにRNNと全結合NNをそれぞれ採用した．この提案モデルに対して，正常な時系列データから成る，SWaTデータセットとWADIデータセットの一部の正常な観測値を他の正常な観測値と入れ替えて生成した新たなデータセットを用いて評価を行った．その結果，集団型異常を検知するためには複数の観測値を扱うネットワークを利用する必要があること，提案モデルは複数の観測値を扱うネットワークを採用した既存手法と比較して，高い精度で集団型異常を検知できることが明らかになった．また，自然データである，てんかん患者の皮質脳波の信号に対して提案モデルを適用したところ，てんかん発作の検知を正しく行えた．さらに，医師が認識していなかった，大きな発作の後にわずかな異常が続いているという事実も捉えることができた．このことより，我々は，自然データに対しても提案モデルの有効性を示すことができた．

今後は，提案モデルを異常検知の精度を高めるようなモデルに拡張し，より多くの種類の自然データを用いて提案モデルの有効性を検証していきたい．

謝辞

本研究を進めるにあたり，貴重なデータを提供していただいた大阪大学脳神経外科高等共創研究院の柳澤琢史教授に感謝申し上げます．

参考文献

[1] 山西健司, データマイニングによる異常検知, 共立出版, 2009.
[2] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. ”Generative adversarial nets,” In Advances in Neural Information Processing Systems, pp. 2672–2680, 2014.
[3] T. Schlegl, P. Seebock, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, ”Unsupervised anomaly detection with generative adversarial networks to guide marker discovery,” In International Conference on Information Processing in Medical Imaging, pp. 146–157, 2017.
[4] H. Zenati, C. S. Foo, B. Lecouat, G. Manek, and V. R. Chandrasekhar, ”Efficient GAN-based anomaly detection,” arXiv:1802.06222, 2018.
[5] D. Li, D. Chen, L. Shi, B. Jin, J. Goh, and S. K. Ng, "MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks," arXiv:1901.04997, 2019.
[6] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," In Advances in Neural Information Processing Systems, pp. 3104–3112, 2014.
[7] A.P. Mathur, and N. O. Tippenhauer, ”SWaT: A water treatment testbed for research and training on ICS security,” In Cyber-physical Systems for Smart Water Networks, pp. 31–36, 2016.

[8] C. M. Ahmed, V. R. Palleti, and A. P. Mathur, "WADI: A water distribution testbed for research in the design of secure cyber physical systems," In Proceedings of the 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks, pp. 25–28, 2017.
[9] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," ACM Computing Surveys, Vol. 41, No. 3, 2007.
[10] F. E. Grubbs, "Procedures for detecting outlying observations in samples," Technometrics, Vol. 11, No. 1, pp. 1–21, 1969.
[11] J. Laurikkala, M. Juhola, E. Kentala, N. Lavrac, S. Miksch, and B. Kavsek, "Informal identification of outliers in medical data," In Fifth International Workshop on Intelligent Data Analysis in Medicine and Pharmacology, Vol. 1, pp. 20–24, 2000.
[12] E. Eskin, "Anomaly detection over noisy data using learned probability distributions," In Proceedings of the International Conference on Machine Learning, 2000.
[13] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," In Proceedings of Second International Conference on Knowledge Discovery and Data Mining, Vol. 96, No. 34, pp. 226–231, 1996.
[14] S. Guha, R. Rastogi, and K. Shim, "ROCK: A robust clustering algorithm for categorical attributes," Information Systems, Vol. 25, No. 5, pp. 345–366, 2000.
[15] R. Smith, A. Bivens, M. Embrechts, C. Palagiri, and B. Szymanski, "Clustering approaches for anomaly based intrusion detection," In Proceedings of Intelligent Engineering Systems through Artificial Neural Networks, pp. 579–584, 2002.
[16] L. Ertoz, M. Steinbach, and V. Kumar, "Finding clusters of different sizes, shapes, and densities in noisy, high dimensional data," In Proceedings of the Society for Industrial and Applied Mathematics International Conference on Data Mining, pp. 47–58. 2003.
[17] S. Ramaswamy, R. Rastogi, and K. Shim, "Efficient algorithms for mining outliers from large data sets," In ACM Sigmod Record, Vol. 29, No. 2, pp. 427–438, 2000.
[18] E. M. Knorr, R. T. Ng, and V. Tucakov, "Distance-based outliers: algorithms and applications," the VLDB Journal—the International Journal on Very Large Data Bases, Vol. 8, No. 3-4, pp. 237–253, 2000.
[19] O. Taylor, and D. Addison, "Novelty detection using neural network technology," In Proceedings of International Congress on Condition Monitoring and Diagnostic Engineering Management, pp. 731–743, 2000.
[20] B. Scholkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," Neural Computation, Vol. 13, No. 7, pp. 1443–1471, 2001.
[21] M. A. Hayes, and M. A. M. Capretz, "Contextual anomaly detection in big sensor data," In Proceedings of the 3rd Int. Congress on Big Data (IEEE BigData 2014), 2014.
[22] E. Keogh, J. Lin, and A. Fu, "Hot sax: Efficiently finding the most unusual time series subseuquence, " In Proceeding of the Fifth IEEE international Conference on Data Mining, pp. 226–233, 2006.
[23] J. Lin, E. Keogh, A. Fu, and H. V. Herle, "Approximations to magic: Finding unusualmedical time series, " In Proceedings of the 18th IEEE Symposium on Computer-Based Medical-Systems, pp. 329–334, 2005.
[24] P. Malhotra, L. Vig, G, Shroff, and P. Agarwal, "Long short term memory networks for anomaly detection in time series," European Symposium on Artificial Neural Networks, pp. 89–94, 2015.
[25] S. Chauhan, and L. Vig, "Anomaly detection in ECG time signals via deep long short-term memory networks," In Data Science and Advanced Analytics, pp. 1–7, 2015.
[26] J. Goh, S. Adepu, M. Tan, and Z. S. Lee, "Anomaly detection in cyber physical systems using recurrent neural network," High Assurance Systems Engineering, 2017 IEEE 18th International Symposium, pp. 140–145, 2017.
[27] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff, "LSTM-based encoder-decoder for multi-sensor anomaly detection," arXiv:1607.00148, 2016.
[28] J. Aoe, R. Fukuma, T. Yanagisawa, T. Harada, M. Tanaka, M. Kobayashi, and H. Kishima, "Automatic diagnosis of neurological diseases using MEG signals with a deep neural network," Scientific reports, Vol. 9, No. 1, 2019.