F-030

説明可能な多変量時系列異常検知手法
A study of explainable anomaly detection for multivariate time series

中原 英里[1)]　塩田 哲哉[1)]　豊田 真智子[1)]
Eri Nakahara　Tetsuya Shioda　Machiko Toyoda

1　はじめに

製造業分野の生産性向上を目指す Industry4.0 という概念が 2011 年頃から提唱されている．工場の生産プロセスでは IoT 機器の導入が加速しており，大量のデータ収集を低コストで取得し，それを多様な用途で活用することが可能になった [1]．そして，多数の企業が工場の運用効率化や運用コスト削減を目的に，IoT 機器から取得したデータを用いて，設備の監視業務を効率化している．この監視業務効率化の一環として，取得したデータから通常とは異なる挙動を検出することで設備の異常を発見する異常検知技術の導入が進んでいる [2]．監視する IoT 機器や監視項目が少数である場合は，取得したデータに基づき人手設計のルールの設定することにより，設備監視や異常検知が行われることが多い．しかし，監視する IoT 機器や監視項目が増加すると異常検出のためのルールが複雑化するため，閾値やルールを人手で適切に設定することが困難となる．このように，多数の異常検出ルールを人手で適切に設定することには限界があるため，取得したデータに対して機械学習を用いた異常検知を行う事例が増加している [3]．

異常検知の文脈においては，一般に異常が発生する頻度は正常が発生する頻度に比べて極めて少なく，また収集したすべてのデータに異常または正常のラベルを付与することが難しいため，正常データのみを用いて教師なし学習でモデルを作成する場合が多い．そして，学習したモデルを用いて正常状態からの乖離を表す異常スコアを算出し，異常スコアに対して異常と判定するための閾値を設定する．近年ではニューラルネットワークを利用して非線形な特徴を学習し，モデルを作成することが増加している [4]．

このような機械学習による異常検知には課題が 2 つある．1 つ目の課題は，異常検知モデルでは異常スコアは正常状態との乖離度合いを計算しているに過ぎず，異常イベントを発生させた原因を明らかにすることが難しいということである．この課題により，異常発生時間前後のデータの可視化や異常原因特定技術を用いて原因を特定するといった追加分析に時間が掛かってしまい，結果として工場生産プロセスの稼働効率を低下させてしまう．2 つ目の課題は，追加分析時に用いる多くの異常原因特定技術は時系列性を考慮しておらず，異常と判定された時点前後の時間的変化を考慮した分析を行うことが困難であるということである．工場生産プロセスから取得されたデータは時系列データであることが多く，時系列データは前の時刻に依存して振舞いが変化する性質を持つため，異常スコアの上昇がどの特徴量およびどの時刻からの影響であるのかを明らかにすることができる異常原因技術が重要である．代表的な異常原因特定技術として再構成誤差や機械学習の説明手法が挙げられるが，これらの技術はデータに含まれる時系列性を考慮しておらず，異常と判定された時点前後の時間的変化を考慮した分析を行うことが困難である．

この 2 つの課題に対処すべく，本研究では時系列データを対象とする異常検知モデルから異常スコアを算出し，それに加えて異常スコア上昇に影響を与えた時刻を示す時間寄与度と，影響を与えた特徴量を示す特徴量寄与度を出力可能な Convolutional Neural Network(CNN)[5] ベースの手法を提案する．人工データを用いて評価実験を行った結果，提案手法は時系列データに対して高精度に異常を検知可能なモデルであることがわかった．そして，モデルから特徴量寄与度と時間寄与度を算出することで，異常原因の一部を特定することが可能になった．

2　関連研究

時系列データを対象とする異常検知の代表的な手法として，CNN や Recurrent Neural Network(RNN)[6][7] がある．これらの時系列異常検知手法は，正常状態の時系列データに対して時間窓を設定し，時間方向にずらしながら学習を行うことで，データの時系列性を考慮したモデルを作成する．そして，学習したモデルを用いて時系列予測を行い，予測値と実測値の差である予測誤差を異常スコアとして算出する．学習済の正常な時系列データと類似している場合は予測誤差が小さくなり，正常な時系列データと類似していない場合は予測誤差が大きくなるため，この性質を利用して時系列データから異常を検知することができる．しかし，これらの手法によって作成したモデルが異常と判定したときに，その判定を下した根拠や理由を示すことが難しく人間が理解できないという問題がある．また，時系列データはデータ点の前後に依存関係があり，前の時刻によってデータ点が示す観測値が変動する傾向を持つため，どの特徴量の変動が異常に加担したのかを明らかにするためには，データの可視化や異常原因の特定などの追加分析が必要になる．

追加分析を行う際に用いられる代表的な 2 つの異常原因特定技術について説明する．1 つ目は再構成誤差を用いる技術である．再構成誤差は入力層，中間層，出力層を持つモデルの入力層と出力層の差分によって，特徴量毎に算出される値である．再構成誤差は，Autoencoder[8] や主成分分析といった中間層でデータの圧縮表現を得る手法であれば算出可能である．モデルに入力されたデータ点が，学習済みである正常データの挙動と類似している場合は，出力層において入力されたデータ点が復元されることで各特徴量の再構成誤差が小さくなる．反対に，モデルに入力されたデータ点が正常データの挙動と類似していない場合は，出力層で入力されたデータ点の復元がうまくできず，再構成誤差が大き

1）　日本電信電話株式会社
NTT ソフトウェアイノベーションセンタ
Nippon Telegraph and Telephone Corporation
NTT Software Innovation Center

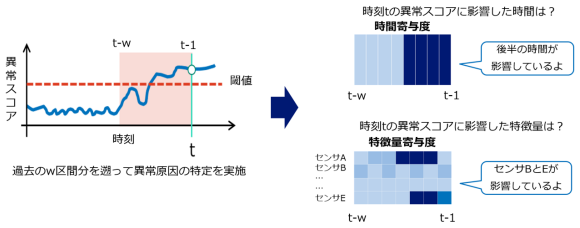

図 1: 提案する寄与度の出力方法

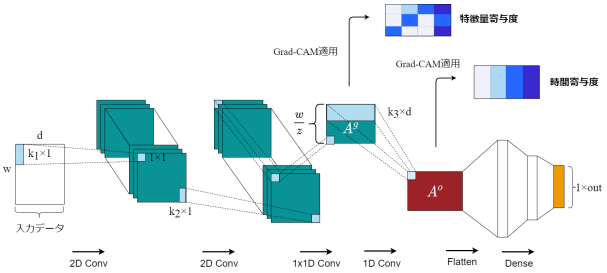

図 2: MTEX-CNN

くなる．このように算出した再構成誤差の可視化や統計量の算出を行うことにより，再構成誤差の値が大きい特徴量を異常原因と推定することが可能になる．

2 つ目は機械学習モデルの説明手法を用いて学習済モデルにより寄与度を算出する技術である．LIME[9] や SHAP[10] では，原因を特定したいデータ点を選択し，原因特定用の新たなモデルを作成することで各特徴量の寄与度を出力する．寄与度とは機械学習モデルの予測結果に影響を与えた度合いのことを示し，特徴量ごとに算出される．この手法を異常と判定されたデータ点に適用し寄与度を算出することで，異常原因特定を実施することが可能である．具体的には，寄与度が大きい特徴量ほど異常の原因であると判断することができる．これらの説明手法は主に教師あり学習を用いたモデルに適用する技術であるが，教師なし異常検知手法に対しても異常スコアと閾値に基づき，各データ点に正常異常のラベルを付けることで適用可能である．教師なし異常検知においても，各データ点をに異常または正常のラベルを付けることにより，説明手法を適用することができる．具体的には，異常判定したデータ点も対して寄与度を算出し，寄与度が大きい特徴量は異常原因であるという判断を行うことができる．

先述した通り，時系列データはデータ点の前後に対して依存関係があり，前の時刻によってデータ点が示す観測値が変動する傾向を持つため，どの特徴量の変動が異常の要因になっているかを明らかにすること難しい．異常原因特定技術として再構成誤差や機械学習の説明手法について述べたが，これらは各データ点を時間方向に独立に扱っており，データの時系列性を考慮しない寄与度が出力される．そのため，異常と判定された時点前後の時間的変化を考慮した分析を行うことが困難である．

3 提案手法

本研究では異常スコアに加えて，異常スコア上昇に影響を与えた時刻と特徴量を示す寄与度を算出し，異常原因特定を可能にする手法を提案する．図 1 に示すように，時刻 t から w 時間分遡って特徴量寄与度と時間寄与度を算出することで，時系列性を考慮した寄与度を算出する．具体的には，特徴量方向および特徴量方向の畳み込みによって得られた特徴マップそれぞれに対して，CNN から寄与度を算出する手法である Grad-CAM[11] を適用する．また，異常原因を強調する寄与度を算出するため，正常データに対して寄与度が示す値が低くなるように正則化を加える．本稿では，3.1 節に提案手法のベースとなった先行研究である MTEX-CNN[12] について説明を行い，3.2 節以降にて異常原因特定を行うための提案手法について詳細な説明を行う．

3.1 MTEX-CNN

MTEX-CNN は教師あり学習を用いて，時系列データを対象とした分類問題を解く手法である．また，分類結果に影響を与えた時刻と特徴量を示す寄与度を算出でき，それにより分類結果に対する判断根拠を示すことができる．MTEX-CNN のアーキテクチャを図 2 に示す．特徴量が d 次元，時間窓が w であるサンプルを入力として受け取り，2 段階の特徴抽出を行う．1 段階目は特徴量方向の特徴抽出で，1 回目の畳み込みでは $k_1 \times 1$ のサイズのフィルタを，2 回目の畳み込みでは $k_2 \times 1$ のサイズのフィルタを用いて，特徴量ごとに時間方向の 2 次元畳み込みを行い，$\frac{w}{z}$ に圧縮する．z は時間窓 w に対してどれくらい圧縮するかを定める値である．そして，行列を転置することにより特徴マップ A^g を得る．なお，畳み込みに使用するフィルタ k_1，k_2 は，$1 < k_1, k_2 < w$ でなければならないという制限がある．2 段階目は時間方向の特徴抽出で，1 段階目の特徴量方向の畳み込みによって得た特徴マップ A^g に対して，d 次元の全ての特徴量から特徴抽出するため $k_3 \times d$ のフィルタを用いて 1 次元畳み込みを行うことで，サンプル全体の特徴量方向の特徴抽出及び時間方向の特徴抽出を実施する．なお，この畳み込みに使用するフィルタ k_3 は，$1 < k_3 < z$ でなければならないという制限がある．2 段階目の時間方向の畳込みによって得られた特徴マップ A^o を全結合層に通しクラス c の予測確率 s_c を出力することで分類問題を解く．

MTEX-CNN の分類結果に対する寄与度の算出に，CNN から寄与度を算出する手法である Grad-CAM を用いる．最初に Grad-CAM による特徴量寄与度の算出方法について述べる．まず特徴量方向の 2 次元畳み込みによって得られた特徴マップ A^g からクラス c の予測確率 s_c に対する勾配を計算する．これは，すなわち $\frac{\partial s_c}{\partial A^g}$ と表せる．g はフィルタの数に依存する．次に，時間窓 w と次元数 d を用いて全体平均プーリングを求め，各特徴マップの重み p_g^c を計算する．これにより，p_g^c は $p_g^c = \frac{1}{w \times d} \sum_i \sum_j \frac{\partial s_c}{\partial A_{ij}^g}$ として表せる．最後に，畳み込み層から得た特徴マップと重みをかけ合わせた行列を，ReLU 関数等の活性化関数を用いて変換することで，特徴量寄与度を算出する．特徴量寄与度は，$M_{2D}^c = \mathrm{ReLU}\left(\sum_g p_g^c A^g\right)$ と表せる．

次に Grad-CAM による時間寄与度の算出方法について述べる．まず，時間方向の 1 次元畳み込みによって得られた特徴マップ A^o から，クラス c の予測確率 s_c に対する勾配を計算する．o はフィルタの数に依存する．次に，時間窓 w を用いて全体平均プーリングをかけ，各特徴マップの重み q_o^c を計算する．これにより，q_o^c は

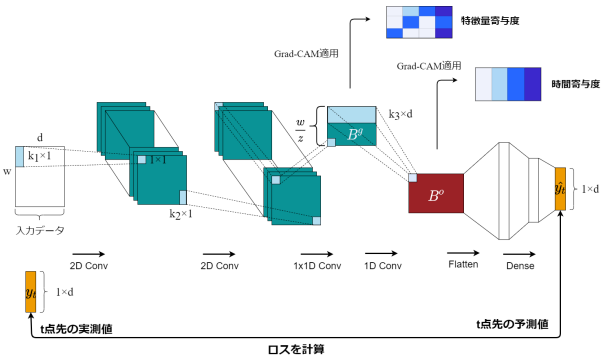

図 3: 提案手法

$q_o^c = \frac{1}{w}\sum_i \frac{\partial s_c}{\partial A_i^o}$ と表せる．最後に，畳み込み層から得た特徴マップと重みをかけ合わせた行列を，ReLU 関数等の活性化関数を用いて変換することで，時間寄与度を算出する．時間寄与度は，$M_{1D}^c = \mathrm{ReLU}\left(\sum_o q_o^c A^o\right)$ と表せる．

3.2 提案手法のアーキテクチャ

提案アーキテクチャは MTEX-CNN[12] をベースとし，異常スコアに加えて異常スコア上昇に影響を与えた時刻と特徴量を示す寄与度を算出する．MTEX-CNN は分類問題を解いているが，提案手法では回帰問題を解くことで時系列異常検知を行う．提案手法のアーキテクチャを図 3 に示す．以降，特徴量が d 次元，時間窓が w に区切ったデータ系列のことをサンプルとする．具体的には，特徴量が d 次元，時間窓が w である入力サンプルを使って特徴抽出を行い，入力サンプルの 1 時刻先の予測値 $\hat{y}_t$ を算出する回帰のアーキテクチャに変更する．学習時には正常な時系列データを用いて，実測値 y_t と予測値 $\hat{y}_t$ との誤差が最小化するようにモデルを学習する．そして予測時には，入力サンプルが学習済の正常データの特徴と類似している場合は誤差が小さくなり，入力サンプルが正常サンプルの特徴と類似していない場合は誤差が大きくなる性質を利用して，時系列データから異常を検出する．さらに式 (1) 示すように，提案手法では実測値 y_t と予測値 $\hat{y}_t$ との平均二乗誤差を計算することで異常スコア L_{ad} を算出する．

$$L_{ad} = \sum_{i=1}^{d} |y_t - \hat{y}_t|^2 \tag{1}$$

MTEX-CNN では，クラス予測確率 s_c を用いて Grad-CAM を適用し特徴量寄与度と時間寄与度を算出したが，提案手法では予測値 $\hat{y}_t$ を用いて Grad-CAM 適用することで特徴量寄与度 $B_{2D}^{\hat{y}_t}$ および時間寄与度 $B_{1D}^{\hat{y}_t}$ を算出する．

3.3 損失関数

異常原因を強調する寄与度を算出するため，提案手法では式 (2) に示す損失関数を用いてモデルを学習することで，異常に対して寄与度が高く，正常に対して寄与度が低くなるように学習を行う．

$$Loss = L_{ad} + L_{feature} + L_{time} \tag{2}$$

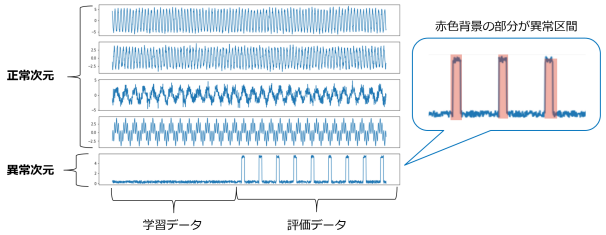

図 4: 人工データ

表 1: 人工データの内訳

	サンプル数	異常サンプル数 (異常の割合)
学習データ	2,619 件	0 件 (0%)
評価データ	1,079 件	350 件 (32%)

$$L_{feature} = \frac{1}{|B_{2D}^{\hat{y}_t}|}\sum_{i=1}^{d}\sum_{j=1}^{w}(1 - B_{i,j}^{\hat{y}_t}) \tag{3}$$

$$L_{time} = \frac{1}{|B_{1D}^{\hat{y}_t}|}\sum_{j=1}^{w}(1 - B_{j}^{\hat{y}_t}) \tag{4}$$

式 (2) の損失関数は，先行研究 [13] を参考に設計されている．先行研究 [13] は画像データに対する教師なし異常検知手法であり，正常画像の寄与度が小さく異常画像に含まれる異常領域の寄与度が大きくなることを目的とし，損失関数に正常サンプルの寄与度が一定になることを期待した正則化を加えている．式 (2) に示す損失関数は，先行研究 [13] を参考に，予測誤差に対するペナルティを表す L_{ad}，時間および特徴量の寄与度に対するペナルティを表す L_{time}，$L_{feature}$ から構成されており，予測誤差に対するペナルティを表す L_{ad} に，正常サンプルの寄与度が小さく異常サンプルに含まれる異常原因に対する寄与度が大きくなることを期待して，正常サンプルの寄与度が一定にする正則化を加えている．

4 実験と考察

提案手法の有効性を示すために，以下のように比較手法としてベースラインを定義し，異常を混入させた人工データを用いて有効性の評価を行った．

- ベースライン：提案モデルのアーキテクチャに式 (1) を損失関数として学習した手法
- 提案手法：提案モデルのアーキテクチャに式 (2) を損失関数として学習した手法

ベースラインと提案手法の違いは，損失関数に特徴量および時間寄与度に対する正則化が含むかどうかである．提案手法は異常原因に対して特徴量寄与度および時間寄与度が高くなるという改善が期待される一方，モデルが複雑になることにより，異常検知精度の劣化が懸念される．そこで，まずはベースラインを用いて提案したアーキテクチャが異常検知に活用できるか，および，特徴量寄与度と時間寄与度から異常原因を特定できるかについて評価を行った．その後，提案手法に対して同様の評価を行い，ベースラインと比較した時の異常検知精度の劣化，および，正則化によって異常原因が強調されることにより，特徴量寄与度と時間寄与度から異常原因特

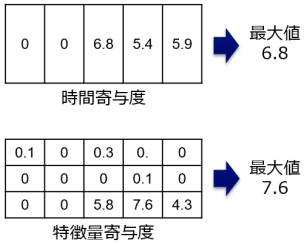

図 5: 各寄与度行列からの最大値の算出

定が容易になっているかを検証した．

4.1 使用した実験データ

図 4 に実験に使用した人工データを示す．図 4 の上から 1-4 次元目は正常次元とし，学習データと評価データともに三角関数を用いて同じ規則でデータを生成した．図 4 の上から 5 次元目は異常次元とし，学習データでは三角関数と一様分布を組み合わせてデータを生成した．一方，評価データでは，赤色の領域のように，著しく大きい値を定期的に加算することで異常状態を擬似的に生成し，生成した異常状態を異常区間として扱った．この人工データを用いて時間窓 w を 1 時刻ずつずらしながら切り出し，モデルに入力可能なデータ形式に変換して正解ラベルを付与した．データ件数と異常データ割合を表 1 に示す．時間窓で切り出したサンプルに 1 時刻以上異常区間が含まれていれば異常サンプルとし，そうでなければ正常サンプルとし，評価データに 3 割程度異常サンプルが入るように人工データを生成した．前処理として，学習データの最大値と最小値を用いて評価データを正規化した．

4.2 実験設定

時間窓 w は，異常区間の一部が入る長さである $w = 16$，$w = 20$ を候補として 5 試行ずつ予備実験を実施し，ベースライン，提案手法ともに異常検知精度が良かった w=20 を採用した．本稿では異常と判定する閾値について，学習データに対する異常スコアの最大値，99%tile 値，95%tile 値の 3 種類の中から，最も異常検知精度が高かった 95%tile 値を閾値として用いたときの結果について述べる．バッチサイズはベースライン，提案手法ともに 10 とした．学習エポック数は訓練ロスと検証ロスの推移より，ベースラインは 2000 エポック，提案手法は 4000 エポックとした．CNN に関するハイパーパラメータは，MTEX-CNN の設定に基づき設定した．具体的には，畳み込み層の数を 3，畳み込みのフィルタサイズ $k_1 \times 1$，$k_2 \times 1$，$k_3 \times 1$ は，順に 8×1，6×1，4×1 とし，フィルタ数は順に，64，128，128 とした．2 回目の 2 次元畳み込みにて，$\frac{w}{z} = \frac{20}{4} = 5$ に圧縮した．

4.3 評価基準

異常検知の精度評価を行うため，評価指標として適合率，再現率，F 値，正解率，ROC-AUC を用いた．ベースライン，提案手法ともにそれぞれ 5 試行実施した．各評価指標は 5 試行の平均を取り精度評価を行った．先行研究 [14] より ROC-AUC が 0.8 以上であれば提案手法が有効であると判断した．

寄与度の評価について，可視化による主観評価とヒストグラムによる客観評価を実施した．可視化による主観

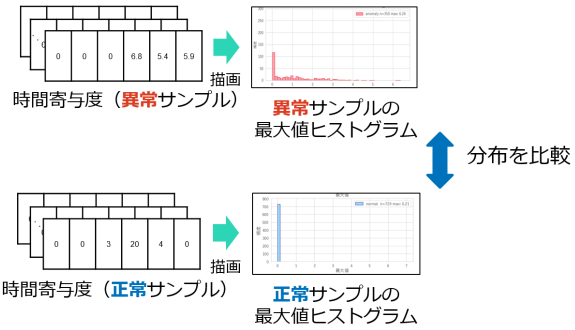

図 6: 寄与度の評価方法

評価では，各寄与度の可視化から異常原因が特定できるかを確認することを目的とし，(a) 入力した異常または正常サンプルの時系列データ，(b) 時間寄与度のヒートマップ，(c) 特徴量寄与度のヒートマップを描画した．本稿では，ベースライン，提案手法ともに同じサンプルに対して寄与度の描画を実施し，主観評価を行った．これらの可視化から，異常サンプルに対して寄与度が高くなり，正常サンプルに対して寄与度が低くなるか主観評価を行った．入力した異常または正常サンプルの時系列データのラインプロットでは，異常次元のみ赤の太い実線で，正常次元は黒の破線で描画した．特徴量寄与度と時間寄与度のヒートマップは，各寄与度行列の要素の最小値を 0，最大値を 1 に正規化し，寄与度が高くなるほど色が濃く，寄与度が低くなるほど色が薄くなるように描画した．特徴量寄与度は，異常次元である 5 次元目の特徴量の異常区間と同じ時刻に高くなることが望ましく，時間寄与度は，異常区間と同じ時刻に高くなることが望ましい。ヒストグラムによる客観評価では，期待通り異常サンプルに対して寄与度が高く，正常サンプルに対して寄与度が低くなっているか確認することを目的とした．まず，5 施行中最も ROC-AUC が高いモデルから，特徴量寄与度と時間寄与度を算出した．次に，図 5 のように特徴量寄与度および時間寄与度の各行列の要素から最大値を算出した．そして，図 6 のように算出した最大値を用いて，異常サンプルと正常サンプルごとにヒストグラムを描画した．このような方法でヒストグラムを描画することで，各寄与度行列が示す最大値の分布がわかり，異常サンプルに対して寄与度が高く，正常サンプルに対して寄与度が低くなっているかを評価することができる．異常サンプルは高い寄与度が得られるべきであるため，ヒストグラムは裾の重い分布となることが望ましく，正常サンプルは低い寄与度が得られるべきであるため，取りうる値がすべて 0 の一様分布になるのが望ましい．このような前提のもとで異常サンプルと正常サンプルに対する特徴量寄与度と時間寄与度を客観的に評価した．図 6 では時間寄与度を例に図示しているが，特徴量重要度も同様に実施した．

4.4 実験結果と考察

4.4.1 ベースラインの実験結果と考察

表 2 にベースラインの異常検知精度の評価結果を示す．ROC-AUC が 0.89 となり異常検知に十分活用できることがわかった．また，正解率は 0.89 であり正常およ

表 2: ベースラインの異常検知精度

	適合率	再現率	F 値	正解率	ROC-AUC
正常	0.87	0.99	0.92	0.89	0.89
異常	0.95	0.68	0.92		

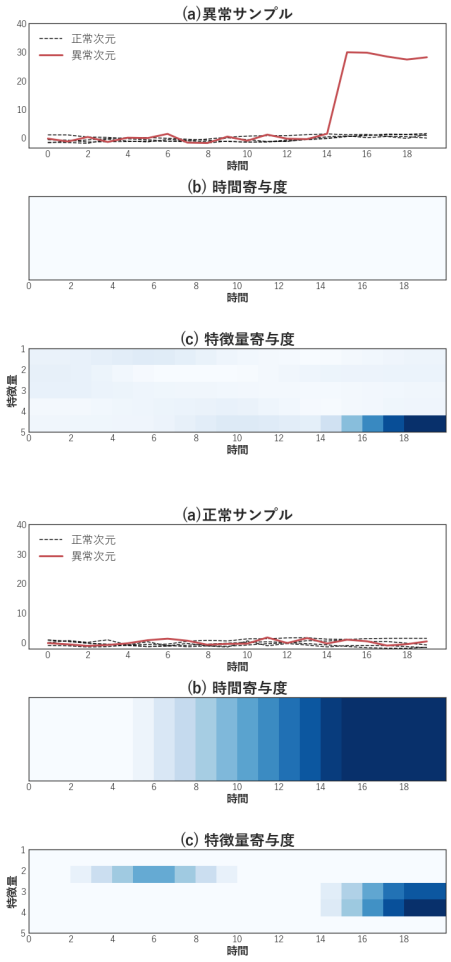

図 7: ベースラインの各寄与度の可視化

び異常に対する F 値は 0.92 と精度が向上した．異常に対する適合率は 0.95 と誤検知が少ない一方で，異常に対する再現率が 0.68 と低く，ベースラインは異常の見逃しが多い傾向がある手法であることがわかった．しかし、今回の評価では，1 時刻でも異常区間が含まれていれば異常サンプルであると設定しており，このことが再現率の低下に影響している可能性があると考えられる．

図 7 にベースラインの特徴量寄与度と時間寄与度の可視化の一例を示す．異常サンプルの (b) 時間寄与度は，異常区間がデータに含まれているにも関わらず寄与度が低くなっており，異常サンプルと対応した寄与度を示せていない．一方，(c) 特徴量寄与度では，異常区間が混入した時間に異常次元である 5 次元目の寄与度が高くなっており，異常サンプルと対応した寄与度を示せている．また，正常サンプルの (b) 時間寄与度と (c) 特徴量寄与度は，異常区間が入っていない正常サンプルであるにも関わらず，ともに高い寄与度を示している．

図 8，図 9 にベースラインの時間寄与度及び特徴量寄与度のヒストグラムを示す．x 軸は各寄与度行列の要素

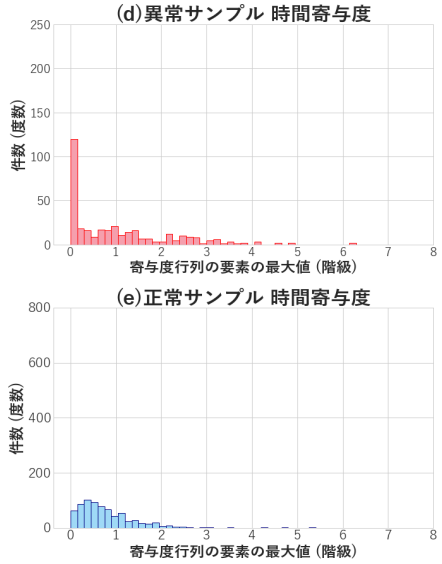

図 8: ベースライン時間寄与度のヒストグラム

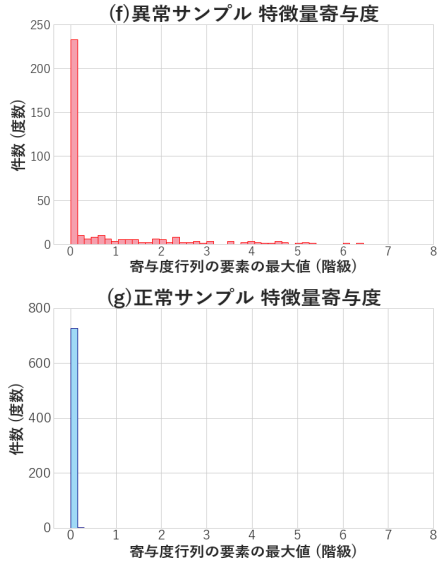

図 9: ベースラインの特徴量寄与度のヒストグラム

の最大値 (階級)，y 軸は件数 (度数) を示す．上図が異常サンプル，下図が正常サンプルのヒストグラムである．x 軸の取りうる値の範囲は上下同じであるが，y 軸の取りうる値の範囲は異常サンプルが正常サンプルに比べて少ないため，最大値が上下で異なっていることに注意していただきたい．

まず，時間寄与度のヒストグラムに関して述べる．図 8 の (d) 異常サンプルの時間寄与度では，ビンが x 軸の 1 から 7 の間で立っており，裾の重い分布となった．しかし，x 軸の 0 付近に分布のピークが立っており，異常サンプルに対する時間寄与度の 350 件中 100 件程度が，異常サンプルに対して低い寄与度を示していたことがわかった．一方，図 8 の (e) 正常サンプルの時間寄与度では，大部分が x 軸の 0 付近の値を示していたが，ビンが x 軸の 1 から 3 の間に複数立っていた．この結果より，一部の時間寄与度は正常サンプルに対して高い寄与度を

表 3: 提案手法の異常検知精度

	適合率	再現率	F 値	正解率	ROC-AUC
正常	0.94	0.98	0.96	0.95	0.95
異常	0.96	0.87	0.96		

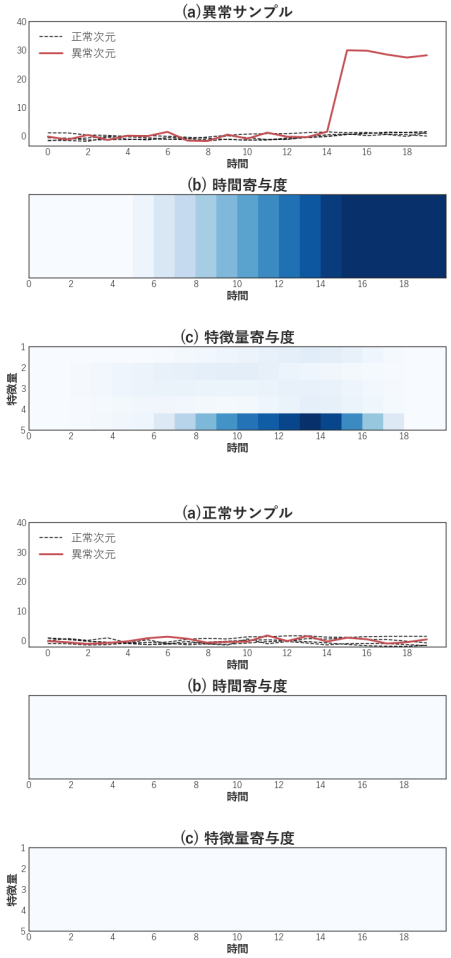

図 10: 提案手法の各寄与度の可視化

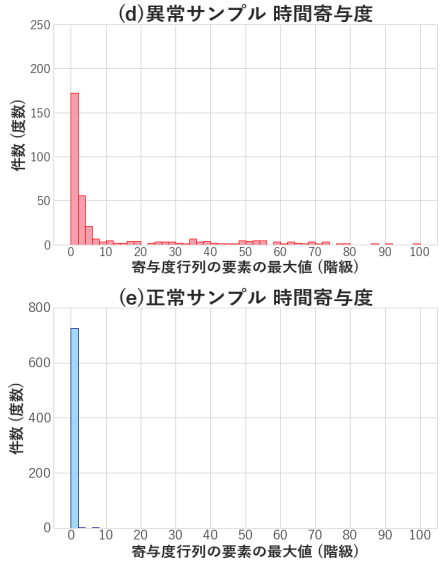

図 11: 提案手法の時間寄与度のヒストグラム

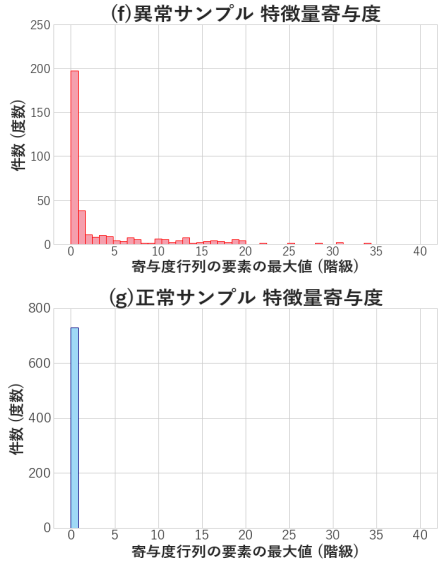

図 12: 提案手法の特徴量寄与度のヒストグラム

示していたことがわかった．次に，特徴量寄与度のヒストグラムに関して述べる．図 9 の (f) 異常サンプルの特徴量寄与度では，ビンが x 軸の 1 から 7 の間で立っており，裾の重い分布となった．しかし，時間寄与度と同様に，x 軸の 0 付近に分布のピークが立っており，異常サンプルに対する特徴量寄与度の 350 件中 230 件程度が，異常サンプルに対して低い寄与度を示していたことがわかった．一方，図 9 の (g) 正常サンプルの特徴量寄与度は，x 軸の 0 付近に分布のピークが立っており期待通りの挙動となった．これらの結果より，時間寄与度，特徴量寄与度ともに異常に対して高い寄与度が出力されることを期待していたが，高い割合でそうはなっていないことがわかった．また，正常サンプルに対して低い寄与度が算出されることを期待していたが，時間寄与度はそうはなっておらず，正常および異常原因の分離ができていないことがわかった．

4.4.2 提案手法の実験結果と考察

表 3 に提案手法の異常検知精度の評価結果を示す．提案手法の ROC-AUC は 0.95 であり，提案手法のほうがベースラインと比較して高い精度を示し，異常検知に十分活用できることがわかった．正解率は 0.96 であり，正常および異常に対する F 値は 0.92 とベースラインと比較して精度が向上していた．異常に対する適合率は 0.96 であり，ベースラインと同様に誤検知が少ない傾向であった．正則化を加えることで，異常に対する再現率が 0.87 とベースラインと比較して精度が向上し，異常の見逃しが多い傾向を改善することができた．これらの結果から，正則化による精度劣化の影響は見られなかった．また，ベースラインの異常検知精度の考察でも述べたとおり，時間窓に含まれる異常区間の割合を考慮した評価を行うといった，さらなる検証が必要であると考える．

図 10 に提案手法の特徴量寄与度と時間寄与度の可視

化の一例を示す．異常サンプルの (b) 時間寄与度は，異常区間と同じ時刻で寄与度が高くなっている．また，(c) 特徴量寄与度では，異常区間と同じ時刻に異常次元である 5 次元目の寄与度が高くなっており，入力サンプルと対応した寄与度を示せている．正常サンプルの (b) 時間寄与度，(c) 特徴量寄与度ともに，寄与度が低くなっている．このような結果から，特徴量寄与度，時間寄与度ともに，正常サンプルに対しては低い寄与度を示し，異常サンプルに対して高い寄与度を示していることが確認できた．

図 11 と図 12 に，提案手法の時間寄与度および特徴量のヒストグラムを示す．まず，時間寄与度のヒストグラムに関して述べる．図 11 の (d) 異常サンプルの時間寄与度では，ビンが x 軸の 1 から 100 の間で立っており，ベースラインと比較してより裾の重い分布となった．しかし，ベースラインと同様に x 軸の 0 付近に分布のピークが立っており，異常サンプルに対する時間寄与度の 350 件中 170 件程度が，異常サンプルに対して低い寄与度を示していたことがわかった．一方，図 12 の (g) 正常サンプルの特徴量寄与度は，x 軸の 0 付近に分布のピークが立っており，正則化によって期待通りの挙動となった．

次に，特徴量寄与度のヒストグラムに関して述べる．図 9 の (f) 異常サンプルの特徴量寄与度では，ビンが x 軸の 1 から 40 の間で立っており，時間寄与度と同様にベースラインと比較してより裾の重い分布となっていた．しかし，時間寄与度と同様に x 軸の 0 付近に分布のピークが立っており，異常サンプルに対する特徴量寄与度の 350 件中 200 件程度が，異常サンプルに対して低い寄与度を示していたことがわかった．一方，図 9 の (g) 正常サンプルの特徴量寄与度は，ベースラインと同様に x 軸の 0 付近に分布のピークが立っており期待通りの挙動となっていた．これらの結果から，正則化によって時間寄与度および特徴量寄与度の取りうる値が，ベースラインと比較して大きくなったが，異常サンプルに対して低い寄与度を示す傾向は改善しなかった．一方，正常サンプルの時間寄与度，特徴量寄与度ともに，期待通り低い寄与度が算出されるように改善されたものの，異常サンプルに対して低い寄与度を示す傾向は改善されず正常および異常原因の分離ができていないことがわかった．

5 まとめ

本研究では，時系列データを対象とする異常検知モデルから異常スコアを算出し，それに加えて異常スコア上昇に影響を与えた時刻を示す時間寄与度と，影響を与えた特徴量を示す特徴量寄与度を出力可能な MTEX-CNN ベースの手法を提案した．人工データを用いて評価実験を行った結果，提案手法は時系列データに対して高精度に異常を検知できることを確認した．そして，モデルから特徴量寄与度と時間寄与度を算出することで，一部は異常原因を特定することが可能であった．しかしながら，半数以上が異常サンプルに対しても低い寄与度を示し，異常原因を特定することができなかったため，提案手法の改善が必要であることがわかった．

今後の課題としては，モデルのアーキテクチャから見直し，異常サンプルに対して高い寄与度が出力できるような仕組みを検討したい．また，異常発見時にどのようにユーザに寄与度を提示することが異常原因特定に有効であるのかを検討するとともに，ユーザテストでその有効性を検証したい．

参考文献

[1] 総務省，情報通信白書平成 30 年度版，2018.

[2] 総務省，経済産業省，文部科学省，ものづくり白書 2019 年度版，2019.

[3] Mohammad Braei and Sebastian Wagner. Anomaly Detection in Univariate Time-series: A Survey on the State-of-the-Art. arXiv preprint arXiv:2004.00433, 2020.

[4] Chalapathy, Raghavendra, and Sanjay Chawla. Deep learning for anomaly detection: A survey. arXiv preprint arXiv:1901.03407, 2019.

[5] Hansheng Ren, Bixiong Xu, Yujing Wang, Chao Yi, Congrui Huang, Xiaoyu Kou, Tony Xing, Mao Yang, Jie Tong, and Qi Zhang. Time-Series Anomaly Detection Service at Microsoft. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD), 2019.

[6] Anvardh Nanduri and Lance Sherry. Anomaly detection in aircraft data using Recurrent Neural Networks(RNN). In Proceedings of the 2015 Integrated Communications Navigation and Surveillance (ICNS), 2016.

[7] Kyle Hundman, Valentino Constantinou, Christopher Laporte, Ian Colwell, and Tom Soderstrom. Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD), 2018.

[8] Geoffrey E. Hinton, and Ruslan Salakhutdinov. Reducing the dimensionality of data with neural networks. Science, vol. 313, no. 5786, pp. 504-507, 2006.

[9] Marco Tulio Ribeiro, Sameer Singh and Carlos Guestrin. "Why Should I Trust You?":Explaining the Predictions of Any Classifier. In Proceedings of the 23rd ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD), 2017.

[10] Lundberg Scott and Su-In Lee. A Unified Approach to Interpreting Model Predictions. In Proceedings of the 31st Neural Information Processing Systems (NIPS), 2017.

[11] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. In Proceedings of the 18th IEEE International Conference on Computer Vision (ICCV), 2017.

[12] Roy Assaf, Ioana Giurgiu, Frank Bagehorn and Anika Schumann. MTEX-CNN: Multivariate Time Series EXplanations for Predictions with Convolutional Neural Networks. In Proceedings of the 19th IEEE International Conference on Data Mining (ICDM), 2019.

[13] Shashanka Venkataramanan, Kuan-Chuan Peng, Rajat Vikram Singh and Abhijit Mahalanobis. Attention Guided Anomaly Localization in Images. In Proceedings of the 16th European Conference on Computer Vision (ECCV), 2020.

[14] Jayawant N. Mandrekar. Receiver operating characteristic curve in diagnostic test assessment. Journal of thoracic oncology, vol. 5, no. 9, pp. 1315-1316, 2010.