工場設備センサデータのための次元圧縮に基づく故障予測アルゴリズム

後雄貴[1]　松原靖子[2]　東口慎吾[3]　木村輔[4]
櫻井保志[5]

近年の工場設備では，時系列センサデータを活用した設備保全が重要な課題となっている． IoTセンサから取得される時系列データは通常，膨大な量であり，これらのデータから効果的に異常兆候を検知し故障の発生を予測することは，運用効率を高め，故障による損害を防ぐ上で不可欠である．そこで本研究では，工場設備より取得された時系列センサデータから設備故障の兆候を捉え，故障発生を予測する手法を提案する．提案手法では，クラスタリングとマハラノビス距離を組み合わせた次元圧縮アルゴリズムにより多次元時系列データを低次元へ圧縮し，CubeMarkerを用いて異常の兆候を示す典型的な時系列パターンを抽出する．さらに，SplitCastを用いて，パターン変化に応じて予測モデルを適応的に切り替えることで，高精度な故障予測を実現する．オープンデータを用いた実験により，次元圧縮アルゴリズムが故障予測の精度の改善に寄与していることを示し，また，提案手法が工場設備の故障予測において高い有効性を示すことを確認した．

1　まえがき

IoTセンサ技術の急速な発展により，工場設備の稼働状況を示す時系列センサデータが大量に生成・蓄積されるようになっている [12,14–16]．工場設備において故障が発生すると，修理費や生産ラインの停止に伴う生産性の低下により多大な経済的な損失となるため，工場設備の時系列センサデータから将来の故障を予測する機械学習技術は重要な研究課題となっている．

工場設備の時系列センサデータから将来の故障を事前に予測するために， CubeMarker [6]やSplitCast [17]という手法を活用した研究がこれまでに行われてきた．具体的な応用例としては，エンジン主軸受の摩耗予測，真空装置の故障予測，パワーモジュールの寿命予測などが挙げられ，これらの研究は工場設備の故障の兆候を捉えて故障の発生を事前に予測することで，故障による生産ラインの停止や修理費による損失を軽減することを目的としている．上記の研究によって提案されたCubeMarkerやSplitCastを用いた故障予測手法は，特定のデータセットにおいて高い予測精度を発揮するものの，汎用性の観点で課題が存在している．具体的には，多くの工場設備の時系列センサデータは次のような特徴を持ち，そのことが正確な故障予測を困難なものにしている．

(1)高次元：

本研究で想定する工場設備で取得される時系列センサデータは，（設備，センサ，時間）の3つの属性から構成されるテンソル形式で表現される．

センサ数や設備数は数十から数百に及ぶ可能性があり，データは高次元テンソルを構成する．データが高次元テンソルであることは，CubeMarkerを用いたセグメンテーションの精度を低下させ，その結果故障予測の精度を低下させる． **(2)類似した時系列パターン：** 高次元なセンサデータの中には，類似した時系列パターンを示すセンサが存在することがある．

例えば，ある工場設備の温度と電流の強さを計測するセンサの値は，電流が強くなった際に温度が上昇するという相関関係から，類似した時系列パターンが存在する可能性がある．そうしたデータに対しては，人手で特徴選択を行い入力データの次元を減らすことが考えられるが，高次元のデータにおける特徴選択は人的・時間的コストが必要である．

本研究では，上記の技術的課題に対応するため，次元圧縮に基づく故障予測アルゴリズムを提案する．

提案手法は，まず初めにクラスタリングとマハラノビス距離を組み合わせた次元圧縮アルゴリズムにより，高次元テンソルである時系列センサデータから類似する時系列パターンをもつ特徴次元をまとめることで低次元テンソルへ圧縮する．その後，圧縮されたテンソルに対し，CubeMarkerを用いたセグメンテーションによる機械状態の推定やSplitCastを用いた将来予測を行う．圧縮されたテンソルは元のテンソルの中の主要なパターンが抽出されたものであるため，元のテンソルに対して直接セグメンテーションや予測モデルの学習を行うよりも，高精度なセグメンテーションと故障予測が可能となる．要約すると，本研究の貢献は以下の通りである．

本研究の貢献：

本研究では，工場設備のセンサから取得される多次元時系列データを対象とする次元圧縮に基づく新たな故障予測モデルを提案する．提案手法は次のような特徴を持つ．

- クラスタリングとマハラノビス距離に基づく次元削減アルゴリズムにより，入力された高次元テンソルを主要な時系列パターンのみが抽出された低次元テンソルに圧縮することで，高次元テンソルに対する高精度なセグメンテーションや故障予測を実現する
- これにより，データの次元圧縮を自動化し，従来は手動で選定されていた特徴量選択プロセスを効率化する
- レジームに応じて予測モデルを切り替えることで高精度に異常を予測する

2　関連研究

センサデータの解析に関する研究は，データベース，データマイニング，機械学習など多様な分野で活発に進められている．これまでに自己回帰モデルや線形動的システムを用いた時

[1] 学生会員　大阪大学産業科学研究所,大阪大学工学部電子情報工学科
yuki88@sanken.osaka-u.ac.jp
[2] 正会員　大阪大学産業科学研究所
yasuko@sanken.osaka-u.ac.jp
[3] 非会員　大阪大学産業科学研究所
shingo88@sanken.osaka-u.ac.jp
[4] 正会員　大阪大学産業科学研究所
tasuku@sanken.osaka-u.ac.jp
[5] 正会員　大阪大学産業科学研究所
yasushi@sanken.osaka-u.ac.jp

系列解析手法が多く提案されており，センサデータを活用した設備保全や故障予測への応用が行われてきた [9]．一例として，CubeCast [8]は大規模なテンソルストリームを解析し，時間と季節性の両観点から非線形な動的トレンドを捉える手法である．本手法は，センサデータの実測値予測において高い性能を示す一方で，設備の故障のようなラベル付きイベントデータの予測には対応していない．また，Deep Neural Network（DNN）を用いた設備故障予測手法も複数提案されている [10].

例えば，Jalayerら [7]は，高速フーリエ変換（Fast Fourier transform，FFT）および連続ウェーブレット変換（Continuous wavelet transform，CWT）を用いてデータの統計的特徴を抽出し，それを長・短期記憶（Long short-term memory，LSTM）モデルに入力することで回転機械の故障予測を実現した．さらに，Pandarakoneら [13]は，電流データに対してFFTによる特徴抽出を行い，

畳み込みニューラルネットワーク（Convolutional neural network，CNN）を活用して誘導電動機のベアリング故障を予測する手法を提案している．加えて，電流や振動などの時系列データを活用した真空ポンプ（TMP）の故障予測に関する研究も行われている [11]． Ainapureら [3]は深層学習に基づく稼働状態の識別手法を提案し，ドメイン適応技術を用いて機械状態の診断を可能にした．さらに，Alessandroら [5]は，振動データとポンプの平均温度を用いてマハラノビス距離を算出し，設備の正常・異常状態を判定した．しかし，これらの研究は特定の設備（例：玉軸受型，ハイブリッド型のTMP）に焦点を当てており，他の構造や異なる運転条件における異常検知については十分に検討されていない．また，Bancheng [4]らは，磁気浮上型のTMPに対して，電流の損失や温度上昇を推定するモデルを提案しているが，故障予測には対応していないことが課題として挙げられる．

これらの既存研究を総括すると，特定の機器やドメインに特化した手法が多く，広範な設備環境に適応可能な汎用的な手法には未だ課題が残されている．また，マハラノビス距離や次元圧縮を活用したデータの要約や，異常検知の精度向上を目指す試みは十分には探求されていない．本研究では，これらの課題に対処するため，多次元時系列データの効率的な次元圧縮と解析を統合した新たな故障予測手法を提案する．

3 問題定義

本研究の目的は，時系列センサデータを用いて設備の異常を高精度に検知および予測することである．本章では，この問題を具体的に定義する．

データは，設備数，センサ数，データ点の3つの要素から構成される多次元時系列データであり，$\mathcal{X} \in \mathbb{R}^{w \times d \times n}$の3階テンソルで表される．

ここで，各変数の意味は以下の通りである：

- w：対象とする設備の数
- d：センサの種類（次元数）
- n：観測時刻の総数

テンソル $\mathcal{X}$ の要素 $x_{ij}(t)$ は，時刻 t における i 番目の設備に搭載された j 番目のセンサで観測された値を示す．異常状態に関する情報は，ラベルデータ $\mathbf{Y} \in \{0,1\}^{w \times n}$ で表され，要素 $y_i(t) = 1$ の場合は設備 i において時刻 t に異常が発生したことを示し，$y_i(t) = 0$ の場合は正常状態を示す．

本論文では，多次元時系列テンソル $\mathcal{X}$ に対してクラスタリングとマハラノビス距離を組み合わせた次元圧縮を行う．次に，次元圧縮後のテンソル $\mathcal{X}'$ を m 個のセグメント集合 $S = \{s_1, \ldots, s_m\}$ に分割してその特徴をとらえる． s_i は i 番目のセグメントの開始点 t_s，終了点 t_e，設備番号で構成され（つまり，$s_i = \{t_s, t_e, \text{facilityID}\}$），各セグメントは重複がないものとする．そして，発見したセグメント集合を類似セグメントのグループに分類する．本論文において，これらのグループをレジームと呼ぶ．

[定義1]（レジーム） セグメントのグループをレジームと定義する．ここで，r を最適なセグメントグループの個数とし，各セグメントs_iは，r個のセグメントグループのいずれかに属するものとする．それぞれのレジームは統計モデル$\theta_i (i = 1, ..., r)$によって表現される．

さらに，各セグメントが所属するレジームを表現するために，新たにセグメントメンバーシップを定義する．

[定義2]（セグメントメンバーシップ） $F = \{f_1, \ldots, f_m\}$ を m 個の整数列とし，f_i を i 番目のセグメントが所属するレジームの番号とする．

提案手法は，時系列テンソルをm個のセグメント，r個のレジームに分割することで，故障に関する重要な特徴を抽出する．r個の独立したレジームに対するパラメータは $\Theta = \{\theta_1, \ldots, \theta_r, \Delta_{r \times r}\}$ と表現される．ここで，$\Delta_{r \times r}$はレジーム遷移行列であり，次のように定義される．

[定義3]（レジーム遷移行列） $\Delta_{r \times r}$をr個のレジーム群の遷移行列と呼ぶ．ここで，要素$\delta_{ij} \in \Delta_{r \times r}$ は，i番目のレジームからj番目のレジームへの遷移確率を表す．

これにより，多次元時系列テンソル $\mathcal{X}'$ を m 個のセグメントと r 個のレジームで $C = \{m, r, S, \Theta, F\}$ として表現することができる．

最後に，本研究で取り組む問題を以下のように設定する．

問題： 与えられた幅l_cのセンサデータ $\mathcal{X}(t - l_c : t)$ に基づき，時刻 t における設備の状態を評価し，l_f ステップ先のラベル（異常または正常）$Y(t + l_f)$ を予測する．

4 提案手法

本研究では，多次元時系列データを対象として，設備の異常検知および予測を行うための解析手法を提案する．提案手法は，以下の3つのアルゴリズムで構成されている．

- **入力テンソルの次元圧縮**：クラスタリングとマハラノビス距離を用いたアルゴリズムにより高次元データを低次元に圧縮する．これにより，高次元データから特徴的な時系列パターンを効果的に抽出することが可能になる．
- **時系列データの分割とレジーム検出**: 圧縮後のデータをCubeMarkerを用いて複数のセグメントに分割し，類似

セグメントをグループ化することでレジーム（時系列パターン）を検出する．

- **レジームごとの予測モデルの構築**: SplitCastを利用し，各レジームに特化した故障予測モデルを構築し，予測を行う．

以下では，各アルゴリズムの詳細について説明する．

4.1 マハラノビス距離を用いた次元圧縮

まず初めにクラスタリングを行うことで，類似した時系列パターンを持つセンサのクラスタを検出する．その後，それぞれのクラスタに対して，マハラノビス距離に基づく次元削減を行うことで，入力テンソル$\mathcal{X} \in \mathbb{R}^{w\times d\times n}$ を $\mathcal{X}' \in \mathbb{R}^{w\times d'\times n}$ $(d' < d)$ に変換する．本節では，クラスタリングのアルゴリズムと，マハラノビス距離を用いた次元削減のアルゴリズムについて詳細に記述する．

4.1.1 DTWとK-meansによるクラスタリング

本研究では，動的時間伸縮法（Dynamic time warping，DTW）を用いて時系列データ間の非線形な類似性を計算し，その結果に基づいてK-meansクラスタリングを実施する． DTWは，時系列データ間の時間的ずれを考慮しつつ類似性を評価するため，高次元データを適切にクラスタ化することができる．これにより，類似した特徴を示すセンサのグループを自動的に発見することが可能となる．

4.1.2 マハラノビス距離による次元削減

マハラノビス距離D_tは，次式で定義される．

$$D_t = (\mathbf{x}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \tag{1}$$

ここで，$\mathbf{x}_i$ は観測ベクトル，$\boldsymbol{\mu}$ はデータ全体の平均ベクトル，$\boldsymbol{\Sigma}$ は共分散行列を表す．

4.1.1節で得られた各クラスタについて，式(1)を用いて次元圧縮を行う．まず，各クラスタ内のデータ点を用いて観測ベクトルの共分散行列 $\boldsymbol{\Sigma}$ を計算し，得られた共分散行列に対して固有値分解を行う．次に，固有値の大きさに基づいて重要な固有ベクトルを選択し，これらを用いて低次元空間への射影を行う．このプロセスにより，多次元時系列テンソル $\mathcal{X} \in \mathbb{R}^{w\times d\times n}$ は $\mathcal{X}' \in \mathbb{R}^{w\times d'\times n}$ $(d' < d)$ に圧縮される．

ここで各クラスタは類似した時系列パターンを持つセンサのクラスタを表しており，それぞれのクラスタ内のデータを次元圧縮するため，データの非線形な構造や相関性を保持しつつ，解析に適した低次元表現を得ることができる．また，他の次元圧縮手法として主成分分析（Principal component analysis，PCA）が挙げられるが，PCAはデータの分散を最大化する主成分方向に基づいて次元を削減するため，次元間の非線形な関係や複雑な相関性を十分に捉えることが難しい．一方で，マハラノビス距離を用いた次元圧縮は，データの相関性を直接反映した次元削減が可能であり，各クラスタの特性をより適切に表現できる．

4.2 CubeMarkerによるセグメント分割とレジーム検出

CubeMarkerは多次元時系列テンソルをセグメンテーションし，部分時系列の集合に要約する手法である．ここで，CubeMarkerによって分割された個々の部分時系列をセグメントと呼び，また各セグメントが属する固有の時系列パターンをレジームと呼ぶ． CubeMarkerは時系列データ $\mathcal{X}(0:t)$ を入力として受け取り，そのセグメンテーション結果 $C = \{m, r, S, \mathcal{F}, \Theta\}$ を出力する．各変数の詳細はそれぞれ以下のとおりである．

- m：セグメントの総数
- r：レジームの総数
- S：セグメントの集合 $S = \{s_1, \ldots, s_m\}$． s_i は i 番目のセグメントを示し，（設備番号，セグメントの開始点，セグメントの終了点）の3つ組で表される．
- $\mathcal{F}$：セグメントのラベルの集合 $\mathcal{F} = \{f_1, \ldots, f_m\}$． f_i はセグメント s_i が所属するレジームのラベルを示す．
- Θ：レジームのモデルの集合およびレジーム間の遷移行列．$\Theta = \{\theta_1, \ldots, \theta_r, \Delta_{r\times r}\}$について，$\theta_i$は$i$番目のレジームのモデルを示し，$\Delta_{r\times r}$はレジーム間の遷移行列を示す．

各レジーム θ_i は隠れマルコフモデル（Hidden Markov model，HMM）を用いて表現される． HMMは隠れ状態を持つマルコフ過程を示した確率モデルであり，音声処理や言語処理などの分野で広く活用されている．また，CubeMarkerでは1つのレジーム θ_i をHMMにおける1つの隠れ状態とみなした，階層的なHMMを用いてレジーム間の遷移をモデリングする．すなわち，それぞれがHMMである各レジーム θ_i が下階層のHMMであり，上階層のHMMはそれらの隠れ状態 $\theta_1, \ldots, \theta_r$ を持つ．ここで，上階層のHMMにおける状態遷移確率が $\Delta_{r\times r}$ である．

CubeMarkerでは次の手順で時系列データをセグメンテーションする：

1. 初期化：階層的HMMモデル Θ を初期化する．
2. 階層的Viterbiアルゴリズムを適用：時系列データ $\mathcal{X}(0:t) = \{x_0, \ldots, x_t\}$ が与えられたとき，次式により最大尤度 $P(x_t \mid \Theta)$ を動的に求める．

$$P(x_i \mid \Theta) = \max_{1\le j\le r} \left\{ \max_{1\le k\le k_j} \{p_{j,k}(i)\} \right\} \tag{2}$$

$$p_{j;k}(i) = \max \begin{cases} \alpha \cdot \pi_{j;k} \cdot b_{j;k}(x_i), \\ \beta \cdot b_{j;k}(x_i) \end{cases} \tag{3}$$

ただし, $\alpha = \max_l \delta_{li} \cdot \max_v p_{l;v}(t-1)$,
$\beta = \delta_{ij} \cdot \max_w p_{j;w}(i-1) \cdot a_{j;w;k}$.

ここで，$p_{j;k}(i)$ は x_i がレジーム j の隠れ状態 k に属する確率である．δ_{ij} は $\Delta_{r\times r}$ の i 行 j 列目の要素であり，レジーム i からレジーム j へ遷移する確率である．$\pi_{j;k}$ はレジーム j における隠れ状態 k の初期確率，$b_{j;k}(x_i)$ はレジーム j の隠れ状態 k が値 x_i を出力する出力確率，$a_{j;wk}$ はレジーム j において隠れ状態 w から k へ遷移する遷移確率である．

3. Baum-Welchアルゴリズムを適用：セグメンテーション結果を用いて各レジーム θ_i のパラメータ $\{\pi, A, B\}$ を推定し，遷移確率 $\Delta_{r\times r}$ を更新する．
4. 繰り返し：上記の2, 3の手順を，スコアが収束するまで繰り返す．CubeMarkerでは，モデル選択指標として最小記述長（Minimum Description Length，MDL）に基づいたスコアを

定義しており，このスコアが小さいほど良いモデルと判定する．

最終的に，スコアが収束した時点でモデルとセグメンテーション結果が出力される．以上の工程により，時系列データ $\mathcal{X}(0:t)$ をセグメンテーションし，C を得る．得られたセグメンテーション結果は，4.3節において説明されるアルゴリズムの入力となる．

4.3 SplitCastによる時系列予測モデルの構築

本節では，故障予測を行うアルゴリズムであるSplitCastについて記述する． SplitCastは，4.2節で記述したアルゴリズム（すなわち，CubeMarker）の出力Cを入力として受け取り，それぞれのレジーム$\{\theta_1,\ldots,\theta_r\}$に対して，それぞれ1つの時系列予測モデル $\{\mathcal{M}_1,\ldots,\mathcal{M}_r\}$ を学習する．通常の時系列予測手法では，時系列全体 $\mathcal{X}(0:t)$ に対して1つの時系列予測モデルを学習させるが，SplitCastでは各モデルが個々の時系列パターンを集中的に学習できるため，より高精度に時系列予測を行うことができる．具体的には，SplitCastは，CubeMarkerによるセグメンテーション結果 $S,\mathcal{F}$ を入力として受け取り，r 個の時系列予測モデル $\{\mathcal{M}_1,\ldots,\mathcal{M}_r\}$ を出力する．また，推論時にはこれらのモデルを用いて，センサデータおよびセグメンテーション結果に基づいて将来の故障ラベルを予測する． SplitCastによる学習と推論の工程を以下に示す．

■モデルの学習 SplitCastでは，CubeMarkerで得られたそれぞれのレジームに対して固有の故障予測モデルを学習する．そのため，学習においてはまずはじめに，CubeMarkerにより得られたセグメント$S=\{s_1,\ldots,s_m\}$を，それぞれのレジームに属するセグメントの集合 $S^{(1)}_{train},S^{(2)}_{train},\ldots,S^{(r)}_{train}$ に分割する．この分割は次のように定義される．

$$S^{(i)}_{\text{train}}=\{\,s_j\in\mathcal{S}\mid f(s_j)=i\,\},\quad i=1,2,\ldots,r. \tag{4}$$

次に，すべてのレジーム$i\in\{1,\ldots,r\}$に対して，$S^{(i)}_{train}$を学習データとして故障予測モデル$\mathcal{M}_i$を学習する．故障予測には任意のモデルを用いることが可能であり，本研究では，GRU，LSTM，RNN，Transformer，MLP，ロジスティック回帰（LogisticRegression），サポートベクターマシン（SVM），線形サポートベクターマシン（LinearSVM），ランダムフォレスト（RandomForest），エクストラツリー分類木（ExtraTrees）を採用し，それぞれのモデルに対して評価実験を行った．予測に用いるモデルが任意であることは，時系列データの多様な特性に応じて適切なモデルを選択し，柔軟かつ高精度な予測を行うことを可能にする．

■モデルによる予測（推論） 学習した予測モデル $\{\mathcal{M}_1,\ldots,\mathcal{M}_r\}$ を用いて，それぞれの時刻における将来の異常ラベル（故障または正常）を予測する．具体的には，各時点 t において，ウィンドウ幅 l_c のデータ$\mathcal{X}(t-l_c:t)$ を受け取り，l_f 時点先のラベル $\boldsymbol{Y}(t+l_f)$ を予測する．このとき，ウィンドウ内のデータがどのセグメントに属しているのかを S から取得し，そのセグメントに割り当てられたレジームに対応する学習済みモデルを用いて，将来の異常ラベルを予測する．なお，ウィンドウがセグメントの切り替わりを含む場合，ウィンドウ内で最も多く出現するセグメントを算出し，そのセグメントに対応する学習済みモデルを適用して故障予測を行う．すなわち，予測値$\hat{Y}(t+l_f)$ は以下のように算出される．

$$\hat{Y}(t+l_f)=\mathcal{M}(\mathcal{X}(t-l_c:t)) \tag{5}$$

5 実験と議論

本研究では提案手法の有効性を検証するため，オープンデータを用いた実験を行った．

5.1 データセット

本研究では2種類のオープンデータセットを用いて実験を行った．

Microsoft Azure Predictive Maintenance（MAPM）：MAPMデータセット [1]は，実際に稼働していた機械の運転状態，メンテナンス記録，故障履歴など，機械の状態に関する情報を含んでおり，2015年1月1日から2016年1月1日の間，1時間ごとに記録したデータである．本研究では，テレメトリ時系列データ，エラーログ，故障記録を利用した．テレメトリ時系列データには，電圧，回転，圧力，振動の1時間ごとの平均値が記録されており，機械の動作状態の変化を定量的に把握することが可能である．エラーログには，機械の運転中に発生したエラーが時系列で記録されており，故障の前兆を検出するための重要な指標となる．メンテナンス記録は，予防的な部品交換（プロアクティブメンテナンス）と故障後の修理（リアクティブメンテナンス）の両方を含み，各機械のメンテナンス履歴を詳細に確認することができる．故障記録には，コンポーネントの故障に起因する交換事例の詳細が記録されており，故障パターンの分析に不可欠なデータを提供する．機械のメタデータには，モデルタイプと使用年数が含まれる．これらは機械固有の特性や経年劣化の傾向を理解する上で重要な情報であり，予測モデルの精度向上に寄与する要素となる．

Skoltech Anomaly Benchmark（SKAB）：SKABデータセット [2]は，異常検知アルゴリズムを評価するために設計されたベンチマークデータセットであり，産業用IoTシステムから得られた多次元時系列データで構成されている．本研究で用いたデータセットは，2020年3月1日に収集された流体漏れおよび流体添加のシミュレーションに関するものであり，約1秒ごとに記録された時系列データで構成されている．各データポイントには，データベースに書き込まれた日時を示すdatetimeが含まれ，加速度計の測定値として振動加速度（g単位）を示すAccelerometer1RMSおよびAccelerometer2RMS，電動モーターの電流値を示すCurrent（アンペア），水ポンプ後の循環ループ内の圧力を示すPressure（バール），エンジン本体の温度を示すTemperature（摂氏），循環ループ内の流体の温度を示すThermocouple（摂氏），電動モーターの電圧を示すVoltage（ボルト），および循環ループ内の流体循環流量を示すRateRMS（リットル/分）が含まれる．さらに，異常値であるかを示すanomaly（0または1）と，データの変化点を示すchangepoint（0または1）がラベリングされている．

5.2 実験設定

実験では，特徴量間のスケールの違いによる影響を抑えるため，データの各次元について平均を0，分散を1となるようにを正規化した．また各オープンデータセットにおける実験設定はそれぞれ以下のとおりである．

MAPM: 本研究では，特にテレメトリデータを主軸としながら，エラーログと故障記録を補完的に用い，故障予測を行う．$w = 7$個の機械における，$d = 4$個の項目，$n = 1,000$点のデータ点について実験を行った．MAPMデータセットでは7件の故障事例について予測する．テレメトリ時系列データに含まれる電圧，回転，圧力および振動の4つの成分を4.1.1節および4.1.2節の方法で，2次元に圧縮した．したがって，$\mathcal{X} \in \mathbb{R}^{7\times4\times1000}$を$\mathcal{X}' \in \mathbb{R}^{7\times2\times1000}$に変換した．提案手法に対して，$batch\ size = 32$，$epoch = 20$，$learning\ rate = 0.02$として実験を行った．最適化アルゴリズムにはNadamを使用した．

SKAB: 本研究において，SKABデータセットでは特徴量として，Accelerometer1RMS，Accelerometer2RMS，Current，Pressure，Temperature，Thermocouple，VoltageおよびVolume Flow Rate RMSの8つを用いた．$w = 3$個の機械における，$d = 8$個の特徴量，$n = 740$点のデータ点について実験を行った．SKABデータセットでは3件の故障事例について予測する．8つの特徴量を4.1.1節および4.1.2節の方法で，2次元に圧縮した．したがって，$\mathcal{X} \in \mathbb{R}^{3\times8\times740}$を$\mathcal{X}' \in \mathbb{R}^{3\times2\times740}$に変換した．提案手法に対して，$batch\ size = 32$，$epoch = 20$，$learning\ rate = 0.02$として実験を行った．最適化アルゴリズムにはNadamを使用した．

また実験には376GBのメモリ，NVIDIA Corporation GA102GL [RTX A6000]を2枚搭載したLinuxマシンを使用した．

5.3 比較手法

本研究では，2種類のオープンデータセット（MAPMおよびSKAB）を対象に，提案手法である次元圧縮を適用した場合と適用しない場合で故障予測の性能がどのように変化するかを検証した．具体的には，次元圧縮を行わず元のテンソルに対してセグメンテーション・故障予測を行った場合の精度と，次元圧縮を適用した圧縮テンソルに対してセグメンテーション・故障予測を行った場合の精度を比較した．故障予測のベース手法には，GRU，LSTM，RNN，Transformer，MLP，ロジスティック回帰（Logistic），サポートベクタマシン（SVM），線形サポートベクタマシン（LinearSVM），ランダムフォレスト分類木（RandomForest），および，エクストラツリー分類木（ExtraTrees）を用いた．

5.4 評価指標

提案手法の故障予測の精度を評価するために，評価指標としてF1スコアを用いた7分割交差検証を行った．F1スコアは，Precision（適合率）およびRecall（再現率）の調和平均であり，以下の式で計算される．

$$F_1 = \frac{2 \times \mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{6}$$

Precisionは，予測されたイベントの合計数とそのうち正解であったイベントの合計数の割合を示す．一方，Recallは全てのイベントの正解値の数と予測されたイベントの中で正解した合計数の割合を示す．これらの値は0から1の間をとり，1に近いほど予測精度が高いことを意味する．

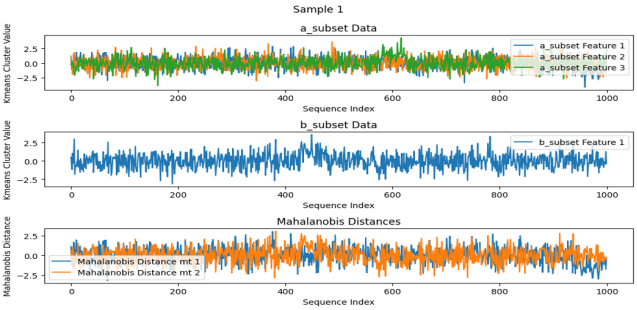

図1 MAPMデータセットをK-meansでクラスタ化しマハラノビス距離による圧縮を行った結果

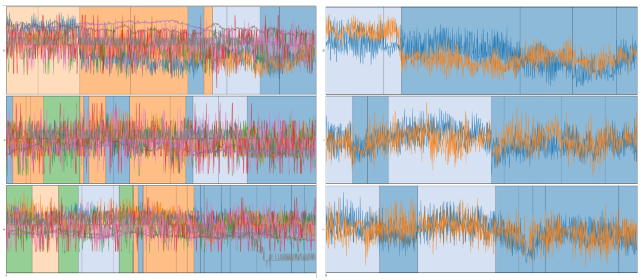

図2 SKABデータセットの次元圧縮前後のCubeMarkerによるセグメント分割の結果

5.5 実験結果

本節では，提案手法の有効性および故障予測における精度について検証した結果を示す．

まず，提案手法による次元圧縮の具体例を示す．図1は，4次元の特徴量（電圧，回転，圧力および振動）から構成されたMAPMデータセットの生データに対し，提案手法の次元圧縮により特徴量を2次元へ圧縮した結果を示している．具体的には，まずDTWおよびK-meansに基づいて，生データの4つの特徴量を2つのクラスタへ分類する．図1の上段は1つめのクラスタであるa_subset Dataを示し，中段は2つめのクラスタであるb_subset Dataを示す．クラスタリング結果から時系列の振る舞いが類似している回転，圧力および振動が1つのクラスタ（a_subset Data）へまとめられていることがわかる．そして，これらの各クラスタごとにマハラノビス距離を用いて次元圧縮を行う．図1の下段は，上段および中段のクラスタそれぞれを圧縮することで得られた新たな特徴量を示す．

次に，時系列パターンの抽出における提案手法による次元圧縮の有効性について示す．

図2は，SKABデータセットに対して CubeMarkerによる時系列パターン抽出した際の次元圧縮の適用時および未適用時におけるそれぞれのセグメント分割の結果を示している．ここで，抽出された時系列パターンの範囲は矩形によって表現され，同じ色の範囲は同一の時系列パターンであることを示す．図2の左列は，SKABデータセットの8次元の特徴量について次元圧縮せずにセ

表1　次元圧縮アルゴリズムを適用した場合（+DR）および適用しない場合（−DR）における故障予測の精度比較（F_1値）

Model		GRU		LSTM		RNN		Transformer		MLP		LogisticRegression		SVM		LinearSVM		RandomForest		ExtraTrees	
Dataset	Window	−DR	+DR	−DR	+DR	−DR	+DR	−DR	+DR	−DR	+DR	−DR	+DR	−DR	+DR	−DR	+DR	−DR	+DR	−DR	+DR
MAPM	16	0.2993	**0.3039**	**0.7118**	0.3039	0.0860	**0.3375**	0.2705	**0.3604**	0.0802	**0.5873**	0.3240	**0.3897**	0.0737	**0.8219**	0.0791	**0.3819**	0.3294	**0.8127**	0.7200	**0.7965**
	32	0.0991	**0.1784**	**0.2995**	0.1688	0.0548	**0.1737**	0.1054	**0.1500**	0.0608	**0.2186**	0.1416	**0.2367**	0.0430	**0.6130**	0.0542	**0.2298**	0.1434	**0.5937**	0.4917	**0.5802**
	48	0.0484	**0.1274**	**0.2170**	0.1209	0.0306	**0.1147**	0.0189	**0.0992**	0.0317	**0.1259**	0.0771	**0.1434**	0.0180	**0.3687**	0.0288	**0.1424**	0.0787	**0.3791**	0.1470	**0.3839**
	64	0.0473	**0.0743**	0.0592	**0.0778**	0.0411	**0.0771**	0.0226	**0.0613**	0.0226	**0.0911**	0.0441	**0.0836**	0.0109	**0.1871**	0.0225	**0.0896**	0.0378	**0.1132**	**0.1153**	0.0951
	128	**0.0968**	0.0468	**0.1387**	0.0489	0.0096	**0.0625**	0.0248	**0.0466**	0.0248	**0.0521**	0.0190	**0.0496**	0.0190	**0.2015**	0.0164	**0.0632**	**0.0180**	0.0150	**0.0436**	0.0150
SKAB	32	0.2340	**0.4616**	**0.7897**	0.4930	**0.7168**	0.4546	**0.6639**	0.5371	**0.6688**	0.5575	**0.7099**	0.3303	**0.5831**	0.4130	**0.7134**	0.3738	**0.5942**	0.4526	**0.8524**	0.4297
	48	0.2620	**0.5729**	**0.7753**	0.5857	**0.7207**	0.5586	**0.7076**	0.5602	**0.6632**	0.5916	**0.6671**	0.4536	0.4533	**0.5972**	**0.6527**	0.4907	**0.5407**	0.5329	**0.5145**	0.4300
	64	0.4171	**0.6781**	0.5834	**0.6682**	0.5180	**0.6738**	0.5766	**0.6408**	0.5430	**0.6610**	0.5175	**0.6288**	0.2571	**0.6300**	0.5081	**0.6194**	0.4131	**0.6039**	0.4221	**0.5748**
	128	0.0000	**0.8376**	0.2526	**0.8190**	0.2416	**0.8045**	0.2526	**0.7956**	0.2416	**0.8650**	0.1562	**0.7704**	0.0000	**0.6280**	0.2325	**0.7486**	0.2114	**0.6823**	0.2080	**0.7486**

グメント分割した結果を示す．一方で，図2の右列は，提案手法によって8次元の特徴量を2次元へ圧縮した後にセグメント分割した結果を示す．また図の左列および右列において，各段（上段，中段および下段）の時系列データはもともと同一の時系列データであるため，同じ段の左右の列を比較することで次元圧縮の有無による時系列パターン抽出の変化を比較できる．図から，次元圧縮を行わない場合に比べて，次元圧縮を適用した場合にはセグメント数およびレジーム数が大幅に削減されていることが確認できる．次元圧縮後のセグメント分割では，データの時間的な変化をより明確に捉えつつも，不要なセグメントの分割が抑制されている．次元圧縮がデータの特徴を簡潔に表現することに寄与し，解析精度の向上と処理負荷の削減の両立を可能にしていることを示している．

続いて，提案手法の次元圧縮による故障予測の精度について検証した結果を示す．表1はMAPMデータセットおよびSKABデータセットに対する次元圧縮アルゴリズム適用の有無における予測精度（F1値）の比較結果を示している．この表において，DRは次元圧縮（**D**imensionality **R**eduction）を示し，`DRは次元圧縮を適用しない場合の結果，+DRは次元圧縮を適用した場合の結果をそれぞれ表している．

また太字で示された数値は，次元圧縮の適用および非適用について各条件の精度を比較した際にF1値が最大となった実験条件を示す．

まずMAPMデータセットでは，予測モデルとしてLSTMを用いた場合を除いて，すべての手法において次元圧縮アルゴリズムが精度の向上に寄与している．いくつかの手法では，ウィンドウサイズが大きくなった場合に精度が低下していることから，長期の故障予測において次元圧縮アルゴリズムが逆効果となっていることが読み取れる．しかし，長期の故障予測においては次元圧縮を用いなかった場合でも精度が高いわけではないことから，予測モデルそのものの限界であると推察される．ウィンドウサイズが小さい場合には，本研究で提案された次元圧縮アルゴリズムが故障予測の精度の改善に寄与していることがわかる．SKABデータセットでは，ウィンドウサイズが大きい場合には精度が大幅に改善しており，次元圧縮アルゴリズムの有効性が示されている．しかし，予測モデルとしてGRUを用いた場合に次元圧縮アルゴリズムが精度向上に寄与している一方で，その他のすべての手法に対して，ウィンドウサイズが小さい場合に精度が低下している．次元圧縮アルゴリズムの故障予測への寄与は，ウィンドウサイズの観点において，MAPMデータセットとSKABデータセットにおいて真逆の傾向を示しており，これはデータセットの特性によるものであると考えられるが，その原因の究明は今後の課題である．

5.6　今後の課題

本研究の結果から，モデルごとの性能差や次元圧縮の効果がデータセットごとに異なることが確認された．一方で，一部の条件においては次元圧縮が予測精度を低下させる場合もみられた．例えば，MAPMデータセットにおいては，LSTMモデルについて，次元圧縮による予測精度が向上する傾向がみられなかった．また，5.5節で記述したように，ウィンドウサイズごとの予測精度の改善の観点からは，データセットにより傾向の違いがみられた．このため，今後の研究では，次元圧縮が与える効果を深く理解するために，各データセットの特徴量分布や情報量を詳細に解析することが求められる．また，次元圧縮アルゴリズムの選択やパラメータ調整を行い，モデルごとに最適な設定を特定する必要がある．さらに，刻々と生成されるセンサデータに対するリアルタイム処理が可能なアルゴリズムへと提案手法を拡張することで，実応用が可能なアルゴリズムとなることが期待される．

6　むすび

本研究では，工場設備の時系列センサデータに対する故障予測アルゴリズムを提案した．本手法は，クラスタリングとマハラノビス距離に基づく次元圧縮とCubeMarkerによるセグメンテーションを組み合わせることで，データの冗長性を低減しつつ，各レジームに特化した特徴抽出を実現した．さらに，SplitCastを活用し，レジームごとに個別の予測モデルを構築することで，複雑な多次元時系列テンソルにおける予測精度を向上させた．MAPMデータセットを用いた実験では，多くの予測モデルにおいて，提案された次元圧縮アルゴリズムが故障予測の精度向上に寄与することが示された．また，SKABデータセットを用いた実験では，長期間の故障予測において，次元圧縮アルゴリズムが貢献することが示された．一方，短期の故障予測においては，提案された次元圧縮アルゴリズムは故障予測の精度向上に寄与しないことが示されており，その原因をデータセットの特徴などから解析し，手法を改良することが今後の課題である．

謝辞 本研究の一部は JST CREST JPMJCR23M3, JST START JPMJST2553, JST CREST JPMJCR20C6, JST K Program JPMJKP25Y6, JST COI-NEXT JPMJPF2009, JST COI-NEXT JP-

MJPF2115 の助成を受けたものです.

参考文献

[1] https://www.kaggle.com/datasets/arnabbiswas1/microsoft-azure-predictive-maintenance/data.
[2] https://www.kaggle.com/datasets/yuriykatser/skoltech-anomaly-benchmark-skab.
[3] Abhijeet Ainapure, Xiang Li, Jaskaran Singh, Qibo Yang, and Jay Lee. Deep learning-based cross-machine health identification method for vacuum pumps with domain adaptation. *Procedia Manufacturing*, 48:1088–1093, 2020.
[4] Han Bangcheng, He Zan, Zhang Xu, Liu Xu, Wen Tong, and Zheng Shiqiang. Loss estimation, thermal analysis, and measurement of a large-scale turbomolecular pump with active magnetic bearings. *IET Electric Power Applications*, 14(7):1283–1290, 2020.
[5] Alessandro Paolo Daga, Luigi Garibaldi, and Luca Bonmassar. Turbomolecular high-vacuum pump bearings diagnostics using temperature and vibration measurements. In *2021 IEEE International Workshop on Metrology for Industry 4.0 & IoT (MetroInd4. 0&IoT)*, pages 264–269. IEEE, 2021.
[6] Takato Honda, Yasuko Matsubara, Ryo Neyama, Mutsumi Abe, and Yasushi Sakurai. Multi-aspect mining of complex sensor sequences. In *2019 IEEE International Conference on Data Mining (ICDM)*, pages 299–308. IEEE, 2019.
[7] Masoud Jalayer, Carlotta Orsenigo, and Carlo Vercellis. Fault detection and diagnosis for rotating machinery: A model based on convolutional lstm, fast fourier and continuous wavelet transforms. *Computers in Industry*, 125:103378, 2021.
[8] Koki Kawabata, Yasuko Matsubara, Takato Honda, and Yasushi Sakurai. Non-linear mining of social activities in tensor streams. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2093–2102, 2020.
[9] Lei Li, James McCann, Nancy S Pollard, and Christos Faloutsos. Dynammo: Mining and summarization of coevolving sequences with missing values. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 507–516, 2009.
[10] Ahlam Mallak and Madjid Fathi. Sensor and component fault detection and diagnosis for hydraulic machinery integrating lstm autoencoder detector and diagnostic classifiers. *Sensors*, 21(2):433, 2021.
[11] Alysson B Barbosa Moreira and Fabrice Thouverez. Dynamic modelling and vibration control of a turbomolecular pump with magnetic bearings in the presence of blade flexibility. In *Rotating Machinery, Vibro-Acoustics & Laser Vibrometry, Volume 7: Proceedings of the 36th IMAC, A Conference and Exposition on Structural Dynamics 2018*, pages 101–110. Springer, 2019.
[12] Qing Ni, JC Ji, and Ke Feng. Data-driven prognostic scheme for bearings based on a novel health indicator and gated recurrent unit network. *IEEE Transactions on Industrial Informatics*, 19(2):1301–1311, 2022.
[13] Shrinathan Esakimuthu Pandarakone, Makoto Masuko, Yukio Mizuno, and Hisahide Nakamura. Deep neural network based bearing fault diagnosis of induction motor using fast fourier transform analysis. In *2018 IEEE energy conversion congress and exposition (ECCE)*, pages 3214–3221. IEEE, 2018.
[14] Akhand Rai and Sanjay H Upadhyay. An integrated approach to bearing prognostics based on eemd-multi feature extraction, gaussian mixture models and jensen-rényi divergence. *Applied Soft Computing*, 71:36–50, 2018.
[15] Qingqing Zhai and Zhi-Sheng Ye. Rul prediction of deteriorating products using an adaptive wiener process model. *IEEE Transactions on Industrial Informatics*, 13(6):2911–2921, 2017.
[16] Xiaodong Zhang, Roger Xu, Chiman Kwan, Steven Y Liang, Qiulin Xie, and Leonard Haynes. An integrated approach to bearing fault diagnostics and prognostics. In *Proceedings of the 2005, American Control Conference, 2005.*, pages 2750–2755. IEEE, 2005.
[17] 本田崇人, 松原靖子, 川畑光希, 櫻井保志, et al. 大規模時系列テンソルによる多角的イベント予測. 情報処理学会論文誌データベース *(TOD)*, 13(1):8–19, 2020.