Vol.2019-MPS-123 No.38
Vol.2019-BIO-58 No.38
2019/6/19

少数異常データを有効活用する部分空間法による異常検知手法

江渕 文人[1,3,a)]　長谷川 隆徳[2,3,b)]　村川 正宏[3,1,c)]

概要：本稿では部分空間法において, 少数の異常データを活用して正常部分空間を生成することにより, 異常検知能力を向上させる手法を提案する. 従来の異常検知手法では, 正常データのみで学習を行うが, 現実の問題では異常データもわずかに取得できる. そこで提案手法では, 部分空間法の従来の目的関数に加えて異常データの平均射影長の最小化を導入する. このような定式化により, 異常データの分布を考慮した正常部分空間を生成できるため, 異常検知能力の向上が期待できる. また, 提案手法は平均射影長を用いて異常データの情報を与えているため, 異常データ数が極端に少ない場合においても安定した識別が期待できる. MNIST データセットを使い提案手法の有効性を評価した結果, 提案手法はわずかな異常データの利用ができる条件下において既存の正常モデルや識別モデルに比べて高い異常検知性能を示し, 学習データに含まれない未知の異常パターンに対しても頑健に異常検知が行えることを明らかにした.

1. はじめに

パターン認識 [1] における識別手法のひとつに部分空間法 (SM : Subspace Method)[2][3] が挙げられる. 部分空間法は汎化能力が高く, カーネル法 [4] の適用により非線形問題への拡張も容易であることから様々な分野, 特に画像認識 [5] において広く用いられている. 部分空間法では, あるクラスの教師データを部分空間に射影したときにその射影長が最大となるような低次元部分空間を生成する. クラスの特徴をよく表現するような部分空間をクラスごとに生成し, それらの部分空間に未知のデータを射影したときにその射影長が最大となるクラスに未知のデータを分類する. このとき部分空間は対象となるクラスの教師データのみを用いることで生成できる. このため, 部分空間法はクラス追加型学習や一つの代表クラスしか持たないような問題にも利用できる.

一方で近年, センサーデータの収集コストが低下したことにより, 機械学習を用いた異常検知 [6][7] の研究が盛んに行われており, 機械の故障検知などに応用されている [8][9]. 現実で得られるセンサーデータからは, 膨大な正常データを得ることができる一方で異常データはほとんど観測されない. そのため, 機械学習を用いた多くの異常検知手法では, 正常データのみを用いて正常状態を学習し, その正常状態からの逸脱度を基に異常を検知する [10][11][12]. この問題設定は一つの代表クラスしか持たないような問題設定であるため, 部分空間法が利用できる [13]. ここで重要なのは, 実用的な課題では, 正常データだけでなくわずかではあるが異常データも得られることである. 従って, このような貴重な少数の異常データの情報を有効活用することができれば, 正常データのみで学習した部分空間モデルよりも異常検知能力を向上できると考えられる.

そこで, 本稿では少数異常データを有効活用する部分空間法による異常検知手法と非線形問題への拡張方法を提案する. 従来の部分空間法の目的関数は正常データの平均射影長最大化であるが, 提案手法では通常の部分空間法の目的関数に加えて, 異常データの平均射影長最小化を導入する. 提案手法の最適化問題を解くことで, 入力次元数の大きさの固有値問題に帰着し, この固有値問題を解くことで正常クラス部分空間の基底ベクトルを得る. この基底ベクトルが張る部分空間に未知のデータを射影したときの射影長に基づき正常, または, 異常を判定する. このような定式化により, 異常データの分布を考慮した正常クラス部分空間を生成できるため, 異常検知能力の向上が期待できる. 平均射影長により異常データの情報を与えているため, 異常

[1] 筑波大学 大学院システム情報工学研究科
Graduate School of Systems and Information Engineering, Tsukuba University
[2] 早稲田大学 基幹理工学研究科
School of Fundamental Science and Engineering, Waseda University
[3] 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology (AIST)
[a)] f.ebuchi@aist.go.jp
[b)] takanori-hasegawa@aist.go.jp
[c)] m.murakawa@aist.go.jp

ⓒ 2019 Information Processing Society of Japan

データが極端に少ない場合においても異常データの情報を有効活用でき, 安定した異常検知が期待できる.

さらに, 上記手法は入力空間においては部分空間の基底ベクトルを得られるが, カーネル法への拡張を行う場合, 通常の部分空間法とは異なり, そのままでは解を得ることができない. これは, カーネル法を用いた非線形問題への拡張は, 一般的に RBF カーネル等が利用されるが, これにより定義される特徴ベクトルは無限次元であり, 通常の部分空間法では式変形することで有限次元の問題に帰着させて解くことが可能であるが, 上記手法ではそれが困難である. そこで, Xiong らにより提案されている標本特徴空間 [14] を導入することでこの問題を解決する. 標本特徴空間はカーネル行列の固有値問題を解くことで得られる特徴空間であり, 高次元特徴空間とカーネルの値が同じになる, すなわち, 高次元特徴空間と等価な有限次元の特徴空間である. このため, 標本特徴空間へ写像することにより, 有限次元の特徴ベクトルを得ることができ, 線形分離性の高い特徴空間でカーネル法に拡張した上記手法を解くことができる.

以下では, 2 章で提案手法に関連する部分空間法と標本特徴空間について述べ, 3 章で少数異常データを有効活用する部分空間法を提案する. 4 章で MNIST データセットを用いた比較実験を行い提案手法の有効性を示し, 5 章でまとめを述べる.

2. 関連研究

2.1 正常データのみを学習する部分空間法

部分空間法は対象のクラスの教師データの射影長の二乗和が最大となる部分空間を求める手法である. 異常検知において正常クラスを表す部分空間の基底ベクトルは以下の最適化問題を解くことで得られる.

$$\begin{aligned} \max \quad & L = \frac{1}{|S_+|} \sum_{i \in S_+} (\boldsymbol{x}_i^\top \boldsymbol{v})^2 \\ \text{s.t.} \quad & \boldsymbol{v}^\top \boldsymbol{v} = 1 \end{aligned} \tag{1}$$

ここで, S_+ は正常データの添字集合であり, $|S_+|$ はその大きさである. また, $\boldsymbol{x}_i$ は m 次元入力データ, $\boldsymbol{v}$ は m 次元ベクトルである. 式 (1) にラグランジュ定数 λ を導入すると, 以下の最適化問題となる.

$$\max \quad L = \frac{1}{|S_+|} \sum_{i \in S_+} (\boldsymbol{x}_i^\top \boldsymbol{v})^2 - \lambda(\boldsymbol{v}^\top \boldsymbol{v} - 1) \tag{2}$$

式 (2) の最適化問題の $\boldsymbol{v}$ に対する最適条件は次式の固有値問題となる.

$$\frac{1}{|S_+|} \sum_{i \in S_+} \boldsymbol{x}_i \boldsymbol{x}_i^\top \boldsymbol{v} = \lambda \boldsymbol{v} \tag{3}$$

次に, m 次元入力データ $\boldsymbol{x}$ を l 次元空間 $(m < l)$ に写像する写像関数 $\boldsymbol{\phi}(\boldsymbol{x})$ を考える. ここで, $\boldsymbol{\phi}(\boldsymbol{x})$ はカーネル関数 $K(\boldsymbol{x}, \boldsymbol{x}') = \boldsymbol{\phi}^\top(\boldsymbol{x})\boldsymbol{\phi}(\boldsymbol{x}')$ を満たす関数とする [4]. 式 (3) の $\boldsymbol{x}_i$ を $\boldsymbol{\phi}(\boldsymbol{x}_i)$ に置き換えると次式となる.

$$\frac{1}{|S_+|} \sum_{i \in S_+} \boldsymbol{\phi}(\boldsymbol{x}_i)\boldsymbol{\phi}^\top(\boldsymbol{x}_i)\boldsymbol{v} = \lambda \boldsymbol{v} \tag{4}$$

ここで, 高次元特徴空間の正常データを並べた行列 $G = (\boldsymbol{\phi}(\boldsymbol{x}_1)...\boldsymbol{\phi}(\boldsymbol{x}_{|S_+|}))$ を用いると, 次式のように書き換えることができる [16].

$$\begin{aligned} & \frac{1}{|S_+|} G G^\top \boldsymbol{v} = \lambda \boldsymbol{v} \\ \Longleftrightarrow \quad & \frac{1}{|S_+|} G^\top G G^\top \boldsymbol{v} = \lambda G^\top \boldsymbol{v} \\ \Longleftrightarrow \quad & \frac{1}{|S_+|} K_+ \boldsymbol{\alpha} = \lambda \boldsymbol{\alpha} \end{aligned} \tag{5}$$

ここで, $\boldsymbol{\alpha} = G^\top \boldsymbol{v}$, K_+ は正常データによるカーネル行列である. 式 (5) により得られた固有ベクトル $\boldsymbol{\alpha}$ より, $\boldsymbol{v}$ を求めると次式となる.

$$\boldsymbol{v}_i = \frac{1}{\sqrt{\lambda_i}} \sum_{j=1}^{|S_+|} \alpha_{ij} \boldsymbol{\phi}(\boldsymbol{x}_j) \quad (\boldsymbol{x}_j \in S_+) \tag{6}$$

式 (5), (6) より, 正常データ数と同数の固有ベクトルを得るが, 冗長な表現が含まれるため, 次式で定義する寄与率を用いることで, 正常クラス部分空間に必要な固有ベクトルを選択する.

$$\frac{\sum_{i=1}^{p} \lambda_i}{\sum_{i=1}^{|S_+|} \lambda_i} > \eta \tag{7}$$

ここで, η は累積寄与率と呼ばれ, 予め人為的に定めるパラメータである. 式 (7) を満たす最小の p を求め, 対応する固有ベクトルが張る部分空間を生成する. 未知のデータ $\boldsymbol{x}$ に対して, 正常クラス部分空間への射影長 $D_+(\boldsymbol{x})$ を次式で定義する.

$$D_+(\boldsymbol{x}) = \sum_{i=1}^{p} \left(\frac{\boldsymbol{v}_i^\top \boldsymbol{\phi}(\boldsymbol{x})}{||\boldsymbol{v}_i|| ||\boldsymbol{\phi}(\boldsymbol{x})||} \right)^2 \tag{8}$$

これを基に正常と異常を判定できる.

2.2 標本特徴空間

一般に高次元特徴空間のベクトル $\boldsymbol{\phi}(\boldsymbol{x})$ は RBF カーネル等を利用するため無限次元になりうる. 標本特徴空間 [14] は高次元特徴空間のデータ分布を解析するために提案された有限次元の空間である. データ数を M とする. カーネル行列は半正定値対称行列であるため, $p = \mathrm{rank}(K)$ とすると, 固有値 λ と固有ベクトル $\boldsymbol{v}$ を並べた行列 $V = (\boldsymbol{v}_1 \cdots \boldsymbol{v}_p \cdots \boldsymbol{v}_M)$ を用いて次式のように表される.

$$K = V \begin{pmatrix} \lambda_1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ 0 & \cdots & \lambda_p & \cdots & 0 \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix} V^\top \tag{9}$$

ⓒ 2019 Information Processing Society of Japan

ここで, 固有ベクトルの定義より, $V^\top V = VV^\top = I_{M\times M}$ である. 正の固有値を対角成分に持つ行列 $\Lambda = \mathrm{diag}(\lambda_1 \cdots \lambda_p)$ と, それらの固有値に対応する固有ベクトルを並べた行列 $U = (\boldsymbol{v}_1 \cdots \boldsymbol{v}_p)$ を用いると, 式 (9) は以下のように書き換えられる.

$$K = U\Lambda U^\top \tag{10}$$

このとき, $U^\top U = I_{p\times p}$ であるが, $UU^\top \neq I_{M\times M}$ である. ここで, $U = (\boldsymbol{v}_1 \cdots \boldsymbol{v}_p) = (\boldsymbol{u}_1 \cdots \boldsymbol{u}_M)^\top$ を満たす $\boldsymbol{u}_i$ を考えると, 教師データ $\boldsymbol{x}_i$ に対するカーネルは次式となる.

$$(K(\boldsymbol{x}_1, \boldsymbol{x}_i), \cdots, K(\boldsymbol{x}_M, \boldsymbol{x}_i))^\top = U\Lambda\boldsymbol{u}_i \tag{11}$$

このとき, m 次元入力ベクトル $\boldsymbol{x}_i$ を p 次元標本特徴空間へ写像する写像関数は次式で定義される.

$$\begin{aligned}\boldsymbol{h}(\boldsymbol{x}_i) &= \Lambda^{-\frac{1}{2}}U^\top(K(\boldsymbol{x}_1, \boldsymbol{x}_i), \cdots, K(\boldsymbol{x}_M, \boldsymbol{x}_i))^\top \\ &= \Lambda^{-\frac{1}{2}}U^\top U\Lambda\boldsymbol{u}_i \\ &= \Lambda^{\frac{1}{2}}\boldsymbol{u}_i \end{aligned} \tag{12}$$

したがって, 標本特徴空間で与えられるカーネル値 $K_e(\boldsymbol{x}_i, \boldsymbol{x}_j)$, すなわち, 標本特徴空間上のベクトルの内積値は, 次のようになる.

$$\begin{aligned}K_e(\boldsymbol{x}_i, \boldsymbol{x}_j) &= \boldsymbol{h}^\top(\boldsymbol{x}_i)\boldsymbol{h}(\boldsymbol{x}_j) \\ &= \boldsymbol{u}_i^\top\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}\boldsymbol{u}_j \\ &= \boldsymbol{u}_i^\top\Lambda\boldsymbol{u}_j \\ &= K(\boldsymbol{x}_i, \boldsymbol{x}_j) \end{aligned} \tag{13}$$

式 (13) より, 標本特徴空間と高次元特徴空間のカーネル値が一致するため, 二つの空間は等価であるといえる. この手法により得られる標本特徴空間は教師データ数と同数の次元の有限次元空間となる.

3. 少数異常データを有効活用する部分空間法

部分空間法では式 (3), (4) から分かるように正常データのみを用いて部分空間を生成しているため, 異常データを考慮した正常クラス部分空間を生成することにより異常検知能力が向上しうる. そこで, 目的関数に異常クラスのデータの平均射影長を最小化する条件を導入することにより実現する. 式 (1) に異常クラスのデータの平均射影長を最小化する目的を加えると次式となる.

$$\begin{aligned}&\max \quad L = \frac{1}{|S_+|}\sum_{i\in S_+}(\boldsymbol{x}_i^\top\boldsymbol{v})^2 - \frac{C}{|S_-|}\sum_{j\in S_-}(\boldsymbol{x}_j^\top\boldsymbol{v})^2 \\ &\text{s.t.} \quad \boldsymbol{v}^\top\boldsymbol{v} = 1\end{aligned} \tag{14}$$

ここで, C は第一項と第二項のトレードオフを決定する人為的パラメータ, S_- は異常データの添え字集合である. 特に $C = 0$ の場合は通常の部分空間法と等価である. ラグランジュ乗数 λ を導入すると, 次式の無制約最適化問題が得られる.

$$\begin{aligned}\max \quad L = &\frac{1}{|S_+|}\sum_{i\in S_+}(\boldsymbol{x}_i^\top\boldsymbol{v})^2 - \frac{C}{|S_-|}\sum_{j\in S_-}(\boldsymbol{x}_j^\top\boldsymbol{v})^2 \\ &- \lambda(\boldsymbol{v}^\top\boldsymbol{v} - 1)\end{aligned} \tag{15}$$

式 (15) の $\boldsymbol{v}$ に対する最適条件は次式の固有値問題で与えられる.

$$\left(\frac{1}{|S_+|}\sum_{i\in S_+}\boldsymbol{x}_i\boldsymbol{x}_i^\top - \frac{C}{|S_-|}\sum_{j\in S_-}\boldsymbol{x}_j\boldsymbol{x}_j^\top\right)\boldsymbol{v} = \lambda\boldsymbol{v} \tag{16}$$

式 (16) の固有値問題を解くことにより, 正常クラス部分空間が得られる. 異常クラスのデータの情報を平均射影長により与えているため, 異常データがごくわずかにしか得られていない場合でも有効に活用することができる.

本手法を高次元特徴空間に拡張する場合, 通常の部分空間法と異なり, 式 (5) のように式変形を行うことができないため, 解が得られない. そこで, 本手法に標本特徴空間を導入することでこの問題を解決する. 具体的には, 式 (16) の $\boldsymbol{x}$ を式 (12) により得られた標本特徴空間のベクトル $\boldsymbol{h}(\boldsymbol{x})$ に置き換える. すなわち, 式 (16) を以下のように書き換える.

$$\begin{aligned}&(A - B)\boldsymbol{v} = \lambda\boldsymbol{v} \\ &A = \frac{1}{|S_+|}\sum_{i\in S_+}\boldsymbol{h}(\boldsymbol{x}_i)\boldsymbol{h}^\top(\boldsymbol{x}_i) \\ &B = \frac{C}{|S_-|}\sum_{j\in S_-}\boldsymbol{h}(\boldsymbol{x}_j)\boldsymbol{h}^\top(\boldsymbol{x}_j)\end{aligned} \tag{17}$$

$\boldsymbol{h}(\boldsymbol{x})$ は有限次元の特徴ベクトルであるため, 式 (17) の固有値問題の解を得ることができる. 得られた固有値から式 (7) を基に正常クラス部分空間の基底ベクトルを選択する. 選択した基底ベクトル $\boldsymbol{v}$ を用いて, 射影長は次のようにあらわされる.

$$D_+(\boldsymbol{x}) = \sum_{i=1}^{p}\left(\frac{\boldsymbol{v}_i^\top\boldsymbol{h}(\boldsymbol{x})}{||\boldsymbol{v}_i||||\boldsymbol{h}(\boldsymbol{x})||}\right)^2 \tag{18}$$

未知のデータ $\boldsymbol{x}$ に対して式 (18) を求め, これを基に正常と異常を判定できる.

4. 計算機実験

提案手法の有効性を検証するために MNIST データセットを用いた比較実験を行った. 本章では, 従来手法の入力空間における部分空間法を **SM**, カーネル部分空間法を **KSM** とし, 入力空間における提案手法 (標本特徴空間を用いない) を **ISM**, 標本特徴空間を導入した手法を **KISM** と記す. さらに, ニューラルネットワークを用いた異常検知でよく利用されるオートエンコーダの再構成誤差による異常検知モデルと比較を行う [11][12]. オートエンコーダは正常データ

Vol.2019-MPS-123 No.38
Vol.2019-BIO-58 No.38
2019/6/19

のみで学習される正常モデルである．更に，異常データを利用した識別モデルの多層パーセプトロンとも比較を行う．

実験で用いたオートエンコーダの構造は (784-100-64-64-100-784) で L2 ノルムによる誤差で評価した．実験で用いた多層パーセプトロンの構造は (784-100-100-1) で正常ならば 0, 異常ならば 1 を出力する．オートエンコーダ, 多層パーセプトロンはどちらも活性化関数には ReLU を使用し，Adam(学習率 0.001, $\beta 1 = 0.9, \beta 2 = 0.999$) により 100epoch 学習して最適化した．本章ではオートエンコーダを **AE**, 多層パーセプトロンを **MLP** と記す．

SM, **KSM**, **ISM**, **KISM** では, 正常クラス部分空間の基底ベクトルを選択するための累積寄与率 η を $\eta = 0.99$ とした．また, **KSM** と **KISM** では, RBF カーネル ($K(\boldsymbol{x}, \boldsymbol{x}') = \exp(-\gamma||\boldsymbol{x} - \boldsymbol{x}'||)$) を使用し, $\gamma = 0.005$ とした. **ISM**, **KISM** では, トレードオフパラメータとして C を設定する必要があるが, 本実験では $C = 0.05$ とした.

4.1 MNIST データセットと実験条件

MNIST データセット [15] はオープンソースのデータセットのひとつであり, 教師データ数が 60000, テストデータ数が 10000 の 0-9 の手書き数字画像である．提案手法の有効性を評価するために, 次の 3 つの実験を行った．各実験において, 学習に用いることができる正常データ数に対する異常データの比率を 0.1%から 10%まで変化させて学習して, 全 10 クラスのテストデータに対し ROC で評価し, AUC 値を算出する．異常データはランダムサンプリングとするが, 異常データのサンプリング結果により識別率が変化するため, 1 つの実験条件につき 10 回試行を行いその平均 AUC により比較を行う．

実験 1: 線形分離性の高いデータ分布に対する提案手法の有効性を評価する．正常データは 10 クラスのうち 1 個のクラスのデータを用い, 異常データは残りの 9 個のクラスのデータを用いモデルを構築する．学習データのうち，正常クラスを 10 通り変化させたときの各 AUC の平均値により, **SM**, **AE**, **MLP**, **ISM** を比較する．

実験 2: 線形分離性の低いデータ分布に対する提案手法の有効性を評価する．正常クラスは，クラス ‘8’ と残りの 9 個のクラスのうち 1 個のクラスをあわせたものを, 異常クラスは，残りの 8 個のクラスを用いモデルを構築する．正常クラスを 9 通り変化させたときの各 AUC の平均値により, **SM**, **AE**, **MLP**, **ISM**，**KISM** を比較する．

実験 3: テストデータに未知の異常データが含まれる場合に対して提案手法の有効性を評価する．正常クラスには，クラス ‘7’ とクラス ‘8’ をあわせたものを用い，異常クラスには，残りの 8 個のうち 1 個のクラスを用いモデルを構築する．異常クラスを変えたときの各 AUC の平均値により, **SM**, **AE**, **MLP**, **ISM**，**KISM** を比較する. 学習データには正常と異常データ合わせて 3 クラスのデータのみを用いるため，7 個のクラスは学習データに含めない．そのため，本実験では学習データに含まれないデータ未知の異常データに対していかに頑健に異常検知を行えるかを評価する．

4.2 実験 1

表 1 に各手法における全クラスの平均 AUC を示す. また, 図 1 にはその結果をグラフで示す. 図 1 を見てわかるように, **ISM** は異常データ率がどの値であっても正常データのみで学習する **SM**, **AE** よりも性能が高い. この結果より, 異常データの分布を考慮した正常クラス部分空間を生成することで, 異常検知能力が向上していることがわかる. また, 異常データも学習を行う **MLP** と比較すると, 異常データ率が 0.1%から 2%までの範囲では **ISM** が上回っており, 提案手法が少ない異常データを有効活用できていることが確認できる.

以上のことから, 提案手法は線形分離性が高い分布の問題において, 異常データ数が極端に少ない場合においても, それらの異常データを有効に活用して, 従来手法よりも高い異常検知能力を得られることが確認できた.

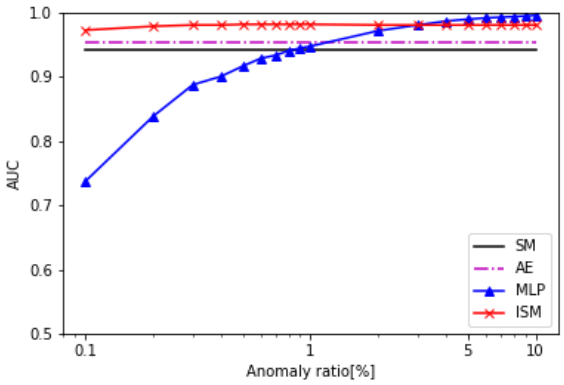

図 1　実験 1 の平均結果
Fig. 1　Result of average in EXPERIMENT 1

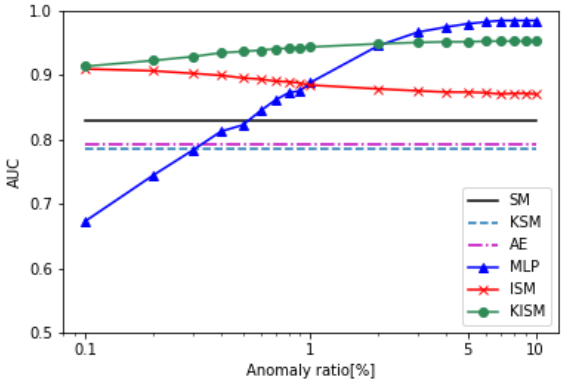

図 2　実験 2 の平均結果
Fig. 2　Result of average in EXPERIMENT 2

ⓒ 2019 Information Processing Society of Japan

表 1　実験 1 の結果
Table 1　Comparison of average AUC in EXPERIMENT 1

	0.0%	0.1%	0.2%	0.3%	0.4%	0.5%	0.6%	0.7%	0.8%	0.9%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
SM	0.941	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AE	0.952	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MLP	-	0.737	0.838	0.887	0.9	0.916	0.928	0.933	0.94	0.944	0.947	0.971	0.98	0.986	0.989	0.991	0.992	0.993	0.994	0.995
ISM	-	0.972	0.978	0.98	0.98	0.981	0.981	0.981	0.981	0.981	0.981	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98

表 2　実験 2 の結果
Table 2　Comparison of average AUC in EXPERIMENT 2

	0.0%	0.1%	0.2%	0.3%	0.4%	0.5%	0.6%	0.7%	0.8%	0.9%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
SM	0.829	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
KSM	0.786	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AE	0.792	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MLP	-	0.673	0.744	0.783	0.812	0.822	0.844	0.861	0.872	0.875	0.888	0.945	0.966	0.974	0.979	0.982	0.984	0.984	0.984	0.984
ISM	-	0.909	0.906	0.902	0.899	0.895	0.893	0.89	0.889	0.887	0.884	0.878	0.875	0.873	0.873	0.872	0.87	0.871	0.871	0.87
KISM	-	0.913	0.922	0.928	0.934	0.936	0.938	0.94	0.941	0.942	0.943	0.948	0.95	0.951	0.951	0.952	0.952	0.952	0.952	0.952

4.3　実験 2

表 2 に各手法における全クラスの平均 AUC を示す. また, 図 2 にはその結果をグラフで示す.

図 2 を見てわかるように, **ISM**, **KISM** は異常データ率に依らず，正常データのみで学習する **SM**, **KSM**, **AE** よりも性能が高い. なお, **ISM** では, 異常データ率が増加するに従い AUC が低下しているが, これは, 学習時の異常データによる影響であると考えられ, パラメータ C を適切な値に設定することで影響を小さくすることができると考えられる.

一方で, 異常データも学習を行う **MLP** と比較すると, 異常データ率が 0.1%から 2%の範囲で **KISM** が **MLP** を上回っている. 3%を超えると **MLP** が **KISM** を上回るが, **KISM** も異常データ率が増加するに従い AUC が増加している. これは, 異常データが増加したことで標本特徴空間の生成において, より正確に高次元特徴空間を表現できたためであると考えられる.

以上のことから, 提案手法は線形分離性が低い分布の問題において, カーネル法により非線形問題へ拡張することの有効性が確認できた.

4.4　実験 3

表 3 に各手法における全クラスの平均 AUC を示す. また, 図 3 にはその結果をグラフで示す.

図 3 を見てわかるように, **ISM**, **KISM** は異常データ率に依らず，正常データのみで学習する **SM**, **KSM**, **AE** よりも性能が高い.

MLP と比較すると, **MLP** は識別モデルであるため, テストデータに未知の異常がある場合, それらを十分に検出することができないため性能が悪い一方, 提案手法は正常クラス部分空間を生成する手法であるため, テストデータに未知の異常データが多く含まれる場合においても安定して異常検知ができていることがわかる.

また提案手法では, 異常データ率が 0.1%から 10%までの AUC がほとんど変化していないことがわかる. 提案手法では異常データの情報を平均射影長により与えていることから, 同じクラスの異常データしか与えられなければほとんど同じ正常クラス部分空間が生成されるため, 異常データ率が変化しても AUC がほとんど変化しない.

表 4 に実験 2 のクラス 8,7 の結果と実験 3 における異常データ率が 1%のときの結果を比較する. 実験 2 と実験 3 では正常データは同じであるが, 異常データをどのクラスからサンプリングするかが異なる. 具体的には, 実験 2 では正常データのクラスを除く全てのクラスから異常データをサンプリングするが, 実験 3 では特定の 1 個のクラスのみから異常データをサンプリングする. 表 4 より, **ISM**, **KISM** どちらにおいても実験 2 の AUC の値が実験 3 の平均値を上回っていることから, 提案手法においては学習データ中の異常データ数よりも, 含まれる異常の種類が重要であることが示唆される.

既知の異常データを有効活用して安定的に正常クラス部分空間を生成できているため，提案手法は未知の異常データに対しても頑健であることが示された．また, モデル構築に用いることのできる異常データ数が極端に限られた場合においても, それらの異常を有効に活用して異常検知できていることが確認できた.

5.　まとめ

本稿では, 少数異常データを有効活用する部分空間法と非線形問題への拡張を提案した. 提案手法では正常クラス部分空間を生成する目的関数に, 異常データの平均射影長最小化の目的関数を導入することにより, 異常データの分布を考慮した正常クラス部分空間を生成する. また, 提案手法はカーネル法への拡張が式変形により得られないため, 標本特徴空間を導入することでこの問題を解決した. 提案手法の有効性を検証するために, MNIST データセットを用いた 3 つの実験を行った. 実験結果から, 提案手法は正常データのみで学習したモデルよりも異常検知能力が高く, また, 異常データ数が非常に少ない場合においてもそれらを有効

ⓒ 2019 Information Processing Society of Japan

表 3　実験 3 の結果

Table 3　Comparison of average AUC in EXPERIMENT 3

	0.0%	0.1%	0.2%	0.3%	0.4%	0.5%	0.6%	0.7%	0.8%	0.9%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
SM	0.792	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
KSM	0.832	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AE	0.786	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MLP	-	0.614	0.629	0.641	0.649	0.646	0.65	0.644	0.659	0.659	0.657	0.67	0.686	0.699	0.702	0.714	0.722	0.724	0.732	0.74
ISM	-	0.866	0.868	0.867	0.867	0.865	0.865	0.864	0.864	0.864	0.863	0.861	0.861	0.86	0.86	0.86	0.859	0.86	0.859	0.859
KISM	-	0.874	0.878	0.878	0.878	0.878	0.878	0.878	0.878	0.878	0.879	0.879	0.879	0.879	0.879	0.879	0.879	0.879	0.879	0.879

表 4　異常データ率 1%における実験 2 と実験 3 の比較

Table 4　Comparison of average AUC for 1% anomaly rate in EXPERIMENT 2 and 3

		実験 2	実験 3		
			最小値	最大値	平均値
1%	ISM	0.869	0.845	0.886	0.863
	KISM	0.923	0.849	0.923	0.879

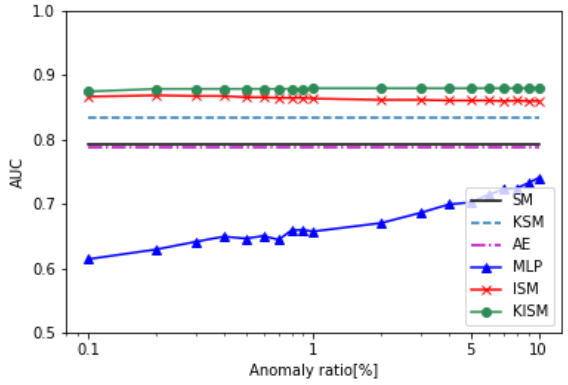

図 3　実験 3 の平均結果

Fig. 3　Result of average in EXPERIMENT 3

活用して安定した異常検知が可能であることが確認できた. さらに, 提案手法は正常クラス部分空間を生成しているため, テストデータに未知の種類の異常が含まれている場合においても, それらの異常を検出することができた. これらのことから, 提案手法は通常の部分空間法や AE, MLP と比較して, 少数の異常データを有効活用して安定して異常検知能力を向上できる手法であり, 実用上の応用範囲が広い.

参考文献

[1] C.M. Bishop:*Pattern Recognition and Machine Learning*, Springer, New York, NY, (2006).

[2] S. Watanabe and N. Pakvasa:*Subspace Methods of Pattern Recognition*, Proc. 1st International Joint Conference on Pattern Recognition, 283/328,(1973).

[3] E. Oja:*Subspace Methods of Pattern Recognition*, Research Studies Press, Letch-worch, UK, (1983).

[4] B. Schölkopf :*The kernel trick for distances*, Proc. Neural Information Processing Systems 13 (NIPS2000), 301/307 (2000)

[5] Belhumeur, Peter N., Joo P. Hespanha, and David J. Kriegman.:*Eigenfaces vs. fisherfaces, Recognition using class specific linear projection.*, IEEE Transactions on pattern analysis and machine intelligence 19.7 ,711/720, (1997).

[6] Denning, Dorothy E. :*An intrusion-detection model.*, IEEE Transactions on software engineering 2 (1987)

[7] Laskov, Pavel, et al. :*Learning intrusion detection: supervised or unsupervised?.*, International Conference on Image Analysis and Processing. Springer, Berlin, Heidelberg, (2005).

[8] T. Hasegawa, J. Ogata, M. Murakawa and T. Ogawa:*Tandem Connectionist Anomaly Detection : Use of Faulty Vibration Signals in Feature Representation Learning*, 2018 IEEE International Conference on Prognostics and Health Management(ICPHM), (2018)

[9] Ye, Jiaxing, et al.:*Statistical impact-echo analysis based on grassmann manifold learning: Its preliminary results for concrete condition assessment*, EWSHM-7th European workshop on structural health monitoring, (2014).

[10] Wang, Yanxin, Johnny Wong, and Andrew Miner.:*Anomaly intrusion detection using one class SVM.*, Proc. the Fifth Annual IEEE SMC Information Assurance Workshop, (2004).

[11] An, Jinwon, and Sungzoon Cho. :*Variational autoencoder based anomaly detection using reconstruction probability.*, Special Lecture on IE 2 (2015).

[12] Zhou, Chong, and Randy C. Paffenroth. :*Anomaly detection with robust deep autoencoders.* Proc. the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, (2017).

[13] Shyu, Mei-Ling, et al. :*A novel anomaly detection scheme based on principal component classifier*, Proc. ICDM Foundation and New Direction of Data Mining workshop, (2003).

[14] H. Xiong, M. N. S. Swamy and M. O. Ahmad :*Optimizing the kernel in the empirical feature space*, IEEE Trans. Neural Networks, vol. 16 no. 2, 460/474 (2005).

[15] Y. Lecun, L. Bottou, Y.Bengio, and P. Haffner:*Gradient-based learning applied to document recognition*, Proc. the IEEE, vol.86, no. 11, 2278/2324, (1998).

[16] Maeda, Eisaku, and Hiroshi Murase.:*Multi-category classification by kernel based nonlinear subspace method.*, Proc. 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Vol. 2. IEEE, (1999).

ⓒ 2019 Information Processing Society of Japan