次元削減の再構成誤差を用いた異常検知手法の比較

Comparison of Anomaly Detection Methods
Using Dimensionality Reduction and Reconstruction Error

乾 稔[*1] 矢入 健久[*2] 河原 吉伸[*3] 町田 和雄[*2]
Minoru Inui Takehisa Yairi Yoshinobu Kawahara Kazuo Machida

[*1]東京大学大学院工学系研究科 [*2]東京大学先端科学技術研究センター
School of Engineering, University of Tokyo RCAST, University of Tokyo

[*3]東京工業大学大学院情報理工学研究科
Department of Computer Science, Tokyo Institute of Technology

Anomaly detection has been one of the most active areas in the datamining study. On the other hand, in the field of machine learning, a variety of non-linear dimensionality reduction (NLDR) methods have been developed recently. In this paper, we focus on the anomaly detection methodology based on the dimensionality reduction, and compared several modern NLDR algorithms with the classical methods such as linear PCA and k-means clustering on this framework. In the experiment, these NLDR methods were shown to be very suitable when the distribution of input data is complicated. Especially, Mixture Probabilistic PCA (MPPCA) showed an outstanding performance for the high-dimensional sensor data of an artificial satellite.

1. はじめに

生産プラントや航空機など，大規模で複雑な人工システムの稼働を監視し，異常やその兆候を検知することは，古くから工学における重要テーマの一つとなっている．従来，この分野では，最も基本的なリミットチェックをはじめとして，ルールベーズドあるいはモデルベーズドな異常検知・診断法など，いわゆる「専門家の知識に基づく」手法が主流であった．しかし時代の趨勢により，システム自体に対しては高機能化・高精度化や多様な運用条件への対応が求められる一方で開発コストや期間の削減が求められており，また，そもそも，各分野において対象システムに精通した専門家の絶対数が減少しているという事情も絡んで，従来からの専門知識に頼ったシステム監視の維持が困難になりつつある．他方，近年の計測・通信技術の発達により，システムの運用状況に関する様々なデータは豊富に得られるようになってきた．そのため，事前知識の代替，あるいは補助として，過去のデータからの学習によるデータ駆動型 (data-driven) のシステム健全性監視・異常検知技術の重要性が各分野で高まっている．

ところで，過去のデータから得られた知識によってシステムを監視すること自体は決して新しいことではない．まず，外れ値検知自体，統計学の古典的問題であるし，プラントの監視においてはケモメトリクスの一分野として統計的プロセス制御 (SPC) が発展している．また，制御工学の世界においてもデータからシステムモデルを獲得する方法論としてシステム同定理論がある． しかし，これらの従来技術は主に線形システムと比較的低次元な計測データを対象としているの対して，より非線形なシステムと高次元なデータへの対応の必要性が高まっている．近年，この分野に様々な機械学習・データマイニング技術が盛んに応用されるようになった背景にはこのような事情によるところが大きい．

一般に，機械学習・データマイニングにおける異常検知は，まず，正常時のデータを訓練データとして用いて何らかの「正常モデル」を学習した後，テストデータをそのモデルに対して評価することによって正常か異常かを判定することによって行われる．本研究では，次元削減による異常検知法，すなわち，訓練データを次元削減することによって得られる部分空間によって「正常モデル」を表すアプローチに着目する．この方法では，まず，正常なデータが存在するべき部分空間あるいはデータ多様体 (data manifold) を訓練データから求める．そして，テストデータに対しては，その正常部分空間からどれだけ乖離しているかを求め，それを「異常度」とみなす．言い換えれば，学習した部分空間によって入力テストデータを近似 (再構築) したときの誤差を評価する．特に，次元削減法として通常の線形主成分分析 (PCA) を用いたケースは，統計的プロセス制御や計算機視覚の分野などで古くから用いられている．また，データマイニングの分野では，近年，Kernel PCA を用いた手法について Hoffman の研究 [Hoffmann 07] があり，One-class SVM などとの比較が議論されている．

本研究の主な目的は，この次元削減・再構築誤差に基づく異常検知法について，以下の 3 つの疑問に答えることである．第一に，近年，様々な教師なし非線形次元削減法が提案されており，それらを異常検知に用いた場合にどのような結果が得られるかという疑問である．第二に，関連する従来手法としてクラスタリングを利用した異常検知法が知られているが，システムの本質的な状態を連続的とするか離散的とするかという違いが実際のシステム異常検知問題においてどのような差をもたらすか，という疑問である．最後に，従来から次元削減は高次元データの可視化に利用されてきたが，一旦次元を削減した後に再度元のデータ空間に復元した結果を提示することで，異常の原因を考察するのに有効な情報を提供できないかという興味である．

次章では，次元削減・再構築による異常検知法の一般的なフレームワークと，今回比較を行った次元削減法に説明する．また，3 章では，2 つの異なる対象領域 – 手書き文字データと人工衛星データについて実験を行い，考察を行う．

連絡先: 矢入 健久，東京大学先端科学技術研究センター，東京都目黒区駒場 4-6-1
e-mail: yairi@space.rcast.u-tokyo.ac.jp

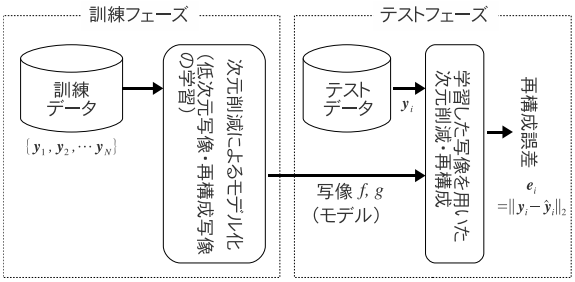

図 1: 次元削減・再構成による異常検知フレームワーク

2. 次元削減・再構築による異常検知

2.1 異常検知のフレームワーク

次元削減と再構築による異常検知は，一般的に以下に示す訓練フェーズとテストフェーズによって実現される．

2.1.1 訓練フェーズ

訓練フェーズでは，システム正常時に得られた訓練データ $\{\boldsymbol{y}_i\}(i = 1, \ldots, N)$ に対してある次元削減アルゴリズムを適用し，低次元空間への写像 f:$\boldsymbol{y}_i \in R^D \rightarrow \boldsymbol{z}_i \in R^d$ $(D > d)$ と，低次元空間から元のデータ空間への復元写像 g:$\boldsymbol{z}_i \in R^d \rightarrow \hat{\boldsymbol{y}}_i \in R^D$ を学習する．

2.1.2 テストフェーズ

各テストデータ $\boldsymbol{y}_i$ に対して，訓練フェーズで得られた 2 つの写像 f,g を順に適用し，再構成データ $\hat{\boldsymbol{y}}_i = g(f(\boldsymbol{y}_i))$ を得る．そして，次式で表される再構成誤差 e_i を計算し，あらかじめ (訓練フェーズにおいて) 設定した閾値を超えた場合に異常が発生したと判定する．

$$e_i = \|\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i\|_2 \tag{1}$$

ここで $\|\cdot\|_2$ は ユークリッドノルムである．

図 1 は，次元削減・再構成による異常検知フレームワークを図示したものである．

2.2 比較する次元削減アルゴリズム

以下では，今回比較した次元削減アルゴリズム 5 種類の概要を述べる．

2.2.1 Principal Component Analysis

最も基本的な線形の次元削減法として，主成分分析 (PCA) を用いる．

2.2.2 Kernel PCA

Kernel PCA (KPCA)[Sholkopf 98] は，線形 PCA をカーネル法によって非線形の PCA ととして拡張したものである．本論文では，カーネル関数としてガウスカーネルを用いた．KPCA では再構成誤差は特徴空間上で非明示的に算出される [Hoffmann 07] ため，再構成されたデータ $\hat{\boldsymbol{y}}_i$ を明示的に求めるためには，別途，数値的最適化手法などを利用する必要がある [Kwok 03].

2.2.3 Mixture Probabilistic PCA

Mixture Probabilistic PCA (MPPCA) は，非線形な分布を局所線形 PCA の混合モデルによって近似する次元削減手法である [Tipping 99]．テストデータの再構成および評価は，各コンポーネント (局所線形 PCA モデル) のうち，最も再構成誤差が小さいものに対して行う．本研究では，コンポーネント数はデータの分布や [Vlassis 02] を参考に決定した．また EM algorithm の初期値依存性回避には，事前に k-means 法により PPCA の中心の算出を行っている．k-means 法も初期値による依存があるため，初期値を変えた 100 回の試行から二乗誤差を最小にするクラスタリング結果を利用した．

2.2.4 Gaussian Process Latent Variable Models

Gaussian Process Latent Variable Models (GPLVM) [Lawrence 04] は確率的な非線形次元削減手法である．Probabilistic PCA (PPCA) をカーネル法を用いて非線形へ拡張したものと見ることができる．あるテストデータ $\boldsymbol{y}_i$ に対してその再構成データ $\hat{\boldsymbol{y}}_i$ を得るために，訓練データによって推定されたモデルを用いて潜在変数を最尤法によって推定する．このとき再構成されたデータ $\hat{\boldsymbol{y}}_i$ はガウス過程 (GP) 回帰と同様に予測分布の期待値を採用する．

2.2.5 Laplacian Eigenmaps Latent Variable Model

Laplacian Eigenmaps Latent Variable Model (LELVM)[Carreira 07] は，非線形次元削減手法 Laplacian Eigenmaps (LE) [Belkin 03] を，確率的なモデルとして拡張した手法である．LELVM では，テストデータに対する次元削減・再構成は，カーネル回帰を用いて行うことができる．なお，今回の実験では，LE やカーネル回帰に必要なパラメータは試行錯誤的に選択した．

2.3 各次元削減法の計算コスト

現実問題への適用では計算コストが重要である，特に異常監視の場合，高速なデータ評価が求められる．表 1 に各次元削減手法の主な計算コストを学習時と評価時についてまとめた．評価時についてのみ述べると，線形手法は高速に新しいデータを評価することが出来るが，非線形手法は pre-image 問題の解決にかかる計算コストが大きい．ただし，LELVM ではカーネル回帰によって pre-image 問題を近似的に解くため，比較的高速なデータ評価が行える．

表 1: データの学習と評価に掛かる主な計算コスト

手法	学習時	評価時
PCA	固有値分解 $O(D^3)$	
KPCA	固有値分解 $O(N^3)$	pre-image 最適化
MPPCA	EM アルゴリズム	
GPLVM	パラメータ最適化	潜在変数最適化
LELVM	固有値分解 $O(N^3)$	
k-means	$O(iN)$	

N:サンプル数，D:系列数，i:繰り返し回数

3. 実験

各次元削減手法による異常検知性能を比較するため，今回は性質の異なる 2 つの分野 – 手書き数字画像認識と，人工衛星サブシステムのデータを用いて実験を行った．このうち，後者の人工衛星については，2 つの衛星で実際に発生した不具合事例に対して適用した．

いずれの実験においても，前章で述べた 5 つの次元削減手法と，比較のためにクラスタリング手法である k-means 法を用いて実験を行った．k-means 法を利用した異常検知方法では，学習データを k 個にクラスタリングし，テストデータと最も近いクラスタ中心とのユークリッド距離を再構成誤差 Er とした．

図 2: 次元削減後，再構成された画像 ($d = 20$)

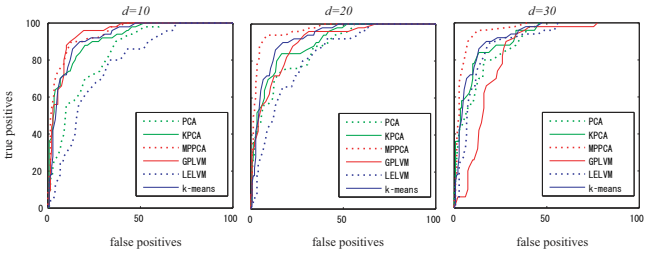

図 3: 手書き数字データにおける ROC 曲線

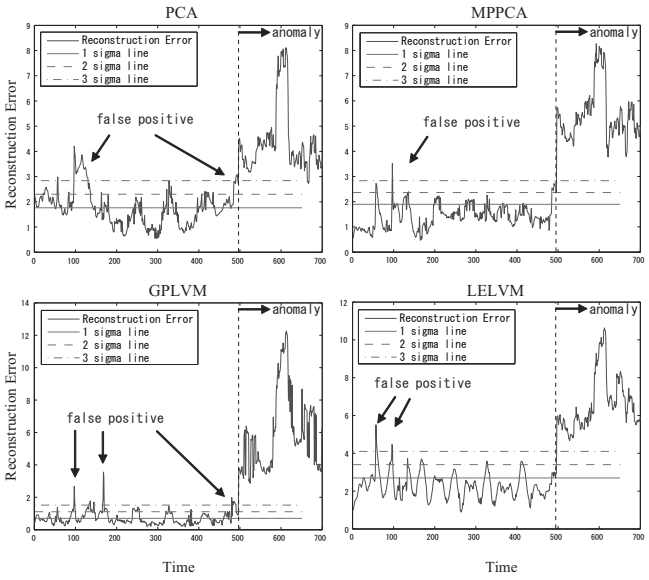

図 4: 再構成誤差と 1,2,3 σ によるリミットライン (衛星 A 異常データ)

3.1 実験 1：手書き数字画像データの異常検知

最初の実験では，MNIST 白黒画像手書き数字データ*1 を用いて以下のような異常検知問題を扱った．まず，数字'0','1','2' についてそれぞれ 300 サンプルずつ，計 900 サンプルを抽出し，これを「正常」なデータとして訓練 (学習) を行った．一方，テストフェーズでは，'0','1','2' の「正常」データに加えて，'9' を「異常」データとみなして各 50 サンプルずつ，計 200 サンプルをテストデータとして検知結果を評価した．

文献 [Levina 04] による手法をこのデータに適用したところ，本質的な次元数は 20 と推定されたが，今回はその前後の次元数を含めた $d = 10, 20, 30$ で実験を行った．

図 2 は，テストデータについて各アルゴリズムで次元削減後再構成した結果を示したものである．いずれの手法でも,「正常」な'0','1','2' については入力画像とほぼ同じ出力が得られているのに対して,「異常」である'9' の画像については入力と大きく異なる出力が得られており，直感的に次元削減・再構成誤差による異常検知が機能していることが分かる．ただし，線形 PCA の結果 (最上段) では期待に反して'9' もある程度復元されてしまっている．これは，線形 PCA では「正常」クラス ('0','1','2') 間の違い (分散) を表現するような部分空間が求められ，その部分空間が結果として,「異常」クラス'9' のデータも比較的良く表現してしまったためであると考えられる．

図 3 は異なる次元数を用いたときの各手法の異常検知能力を評価したものである．ここで着目すべき第一の点は，クラスタリング (k-means 法) による手法の性能が比較的高いことである．これは，本実験では，正常データがクラスター状に (すなわち，不連続的に) 分布しており，かつ，各クラス内の分散が比較的小さく等方的であるためと考えられる．注目すべきもうひとつの点は，MPPCA がその k-means や他の次元削減手法よりも高い性能を示していることである．これは上で述べた不連続的な正常クラス間の分布と，連続的かつ局所的な各正常クラス内の分布をバランス良く表現しているためと考えられる．

*1 MNIST 手書き数字データは下記よりダウンロード可能．
http://yann.lecun.com/exdb/mnist/

3.2 実験 2：人工衛星データの異常検知

3.2.1 実験 2-A

ある人工衛星の電力サブシステムで発生した異常事例について実験を行った．この実験では関連する 14 変数のデータを用いている．訓練データは，異常が発生する以前のデータからサンプリングした 1400 点から成る．一方，テストデータは，異常発生時を含む前後の 700 点から成る．この事例における異常は，突発的に衛星の状態が変化する類のものであったが，14 変数いずれについても事前に専門家が設定したリミット値を超えていない．なお，各変数の値は，平均 0, 分散 1 となるように事前に規格化されている．

図 4 は，いくつかの次元削減アルゴリズムについて，異常発生前後の再構成誤差をプロットしたものである．結論から言えば，本ケースにおいては，いずれの次元削減手法 (および k-means 法によるクラスタリング) を用いた場合でも異常を検知することが出来た．ただし，線形 PCA と比べて他の非線形次元削減法や MPPCA を用いた場合の方が，正常時と異常時の差がより顕著に現れており，正常時における誤警報 (false alarm) の割合も低い．その一つの要因としては，一般に衛星データ (特に電力系) の分布は，衛星本体が日照・日陰のいずれの領域にあるかに影響を受けるため，単純な線形モデルでは正常データを十分に表現しきれないことが考えられる．

3.2.2 実験 2-B

次に，上で用いたのとは別の衛星の電力サブシステムに生じた異常について適用した．この実験では同サブシステムに属する変数のうち，値の変動が全く無いものを除いた全 106 変数を用いた．

図 5 は，各アルゴリズムを用いた場合の検知性能を ROC 曲線によって示したものである．注目すべき第 1 点は，線形 PCA だけでなく，[実験 1] では良好であった k-means 法の結果が著しく悪いことである．これは，この実験におけるデータが比較的高次元であり，かつ，正常時のデータがいくつかの不連続なモードごとに連続的かつ非等方的に分布しているためと考えられる．次に注目すべき点は，MPPCA を用いた場合が

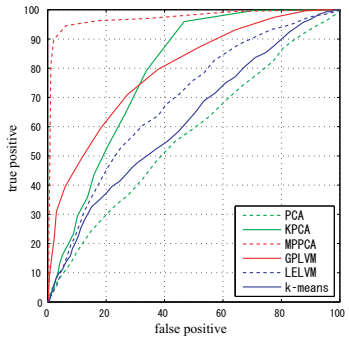

図 5: 衛星 B 異常データにおける ROC 曲線

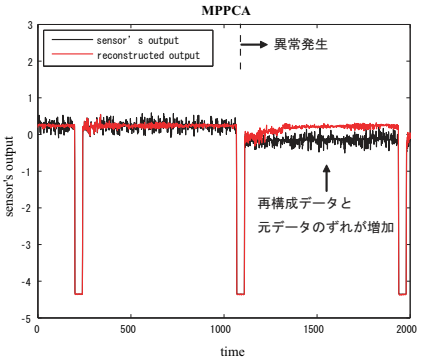

図 6: 衛星 B 異常データと再構成後のデータ

他の非線形次元削減法よりも顕著に良いことである．これは，不連続的な各モードの中ではデータの分布が局所的に線形になっているためと考えられる．一般に，人工衛星などの高度な人工システムは，いくつかの異なる運用モード間を不連続的に遷移する一方，各モード内では平衡点を中心に線形なフィードバック制御によって安定が保たれるように設計されている場合が多い．すなわち，局所線形モデルの混合によるモデル化が適しているという上記の結果は合理的であると言える．

最後に，図 6 は，この異常と深く関連する変数の元データと，MPPCA によって次元削減した後に再構成したデータを示している．異常発生後，再構成された値が元の値と大きく異なっており，この変数が異常に関係していることが想像される．紙面の都合によりここでは例示できないが，変数毎にこのような再構成誤差を見ることにより，どの系列が異常にどれだけ寄与しているかを分析することが可能である．

4. 結論

本研究では，次元削減による再構成誤差に基づく異常検知法のフレームワークに着目し，いくつかの異なる次元削減アルゴリズムおよび k-means クラスタリングアルゴリズムを用いた場合について実験的な比較を行った．

今回の実験と考察から得られる主な結論は，以下の 3 点である．(1) 正常時のデータが複雑な分布をしている場合，非線形な次元削減法が線形 PCA に比べて有意に高い検知能力を示す．(2) 複雑な状態遷移を行う人工衛星の高次元データに適用したケースでは，局所線形 PCA の混合モデルである MPPCA を用いた場合が極めて良好な結果を示した．一方，k-means によるクラスタリングはほとんど異常を検知することができなかった．(3) 次元削減した後に再構成されたデータと，元の入力データを比較することにより，異常原因等を解析する上で有用な情報が得られる．

今後の研究では，より多くの種類の異常検知問題において検証を行う予定である．

謝辞

本研究の実施にあたっては，宇宙航空研究開発機構・研究開発本部・通信・データ処理グループの高田昇氏より多大なご支援とご助言を頂いた．ここに感謝を述べる次第である．

参考文献

[Sholkopf 98] B. Schölkopf, A.J. Smola, K.R. Müller: Nonlinear component analysis as a kernel eigenvalue problem, *Neural Computation*, Vol. 10, No. 5, pp. 1299–1319 (1998).

[Kwok 03] James T. Kwok and Ivor W. Tsang: The pre-image problem in kernel methods, In *ICML*, pp. 408–415 (2003).

[Hoffmann 07] H. Hoffmann: Kernel PCA for novelty detection, *Pattern Recognition*, Vol. 40, No. 3, pp. 863–874 (2007).

[Tipping 99] M. E. Tipping and C. M. Bishop: Mixtures of probabilistic principal component analyzers, *Neural Computation*, Vol. 11, No. 2, pp. 443–482 (1999).

[Lawrence 04] N. D. Lawrence: Gaussian process latent variable models for visualisation of high dimensional data, In *Proc. NIPS* (2004).

[Carreira 07] M. Á. Carreira-Perpiñán and Z. Lu: The Laplacian Eigenmaps Latent Variable Model, In *AISTATS* (2007).

[Belkin 03] M. Belkin and P. Niyogi: Laplacian eigenmaps for dimensionality reduction and data representation, *Neural Computation*, Vol. 15, No. 6, 1373–1396 (2003).

[Levina 04] E. Levina and P. J. Bickel: Maximum likelihood estimation of intrinsic dimension, In *NIPS* (2004).

[Vlassis 02] N. Vlassis, A. Likas: A greedy EM algorithm for Gaussian mixture learning, In *Neural Processing Letters*, Vol. 15, No. 1, pp. 77–87 (2002).