

混合確率主成分分析と隠れセミマルコフモデルによる宇宙機データ監視

Data Monitoring of Spacecraft Using Mixture Probabilistic Principal Component Analysis and Hidden Semi-Markov Models

田川 貴章^{*1} 矢入 健久^{*2} 高田 昇^{*3} 山口 由仁^{*4}
 Takaaki Tagawa Takehisa Yairi Noboru Takada Yukihito Yamaguchi

^{*1} 東京大学大学院工学系研究科
 School of Engineering, University of Tokyo

^{*2} 東京大学先端科学技術研究センター
 RCAST, University of Tokyo

^{*3} 宇宙航空研究開発機構
 Japan Aerospace Exploration Agency

^{*4} TIS 株式会社
 TIS Corporation

Recently there are some researches for anomaly detection of spacecraft assuming that a data distribution of spacecraft is constrained to some modes. Based on this assumption, we learn the distribution as a mixture of local models and efficiently detect an anomaly by monitoring the local data distribution of the each mode. In this paper, because a system of a spacecraft has a periodic mode transition pattern, we propose the new anomaly detection system which can detect not only an anomaly inside the mode but also a mode transition pattern error by combining two methods, one is Mixture Probabilistic Principal Component Analysis to learn a mixture of local linear models and the other is Hidden Semi-Markov Models to learn mode transition patterns between the modes. Then the experiments were conducted to demonstrate the effectiveness of our new approach.

1. はじめに

最近の宇宙開発において、各々のミッションにかかるコストは膨大でありミスは許されないが、宇宙機の故障をゼロにすることは出来ていないのが現状である。そこでそうした異常を早期発見し対策を講じることが大変重要になっている。しかし、それら宇宙機のデータは数百系列にも及んでいるため人の手でそれらを処理するには限界があるというのが現状となっている。

近年の研究においては、機械学習やデータマイニングといった技術により宇宙機のデータから知識を獲得し、モデルを学習させることで異常を自動で検知するシステムが特に注目されてきている。このシステムでは、従来主に研究されてきた専門家の事前知識をベースに演繹的な推論を行う異常検知システムの欠点を補うことができる。

特に本研究では宇宙機のデータ分布はいくつかのモードに拘束されるという仮定(例えば日照、日陰による急激な温度変化など)をおき、それに基づいた異常検知手法を提案する。従来研究[乾 08, 吉木 09]においては、次元削減法の1つである Mixture Probabilistic Principal Component Analysis を用いた手法により対象の宇宙機をモードごとの局所線形な部分空間の混合モデルとしてシステムのモデル化を行い、現在のモード内におけるテストデータとの比較によって異常検知を行っている。この時、対象のシステムは一定の周期性をもったモード遷移を行うが、この手法においてはそうしたモード遷移の異常までは検知することが出来ず False Positive, False Negative が発生する原因となっている。そうした問題に対して従来研究[吉木 09]ではクラス間遷移確率及び文字列検索を用いた手法を提案しているがまだ大きな改善には至っていないのが現状となっている。

そこで本研究においては、そのような従来研究では検知する事が出来なかった再構成誤差およびモード遷移異常に基づく

False Positive 及び False Negative に同時対応した異常検知手法の提案を目的とした研究を行った。具体的には Mixture Probabilistic Principal Component Analysis によるモード内のミクロな異常データの検知に加え、Hidden Semi-Markov Models を用いたモード遷移学習によるモード遷移異常検知をも同時に行う自動異常検知システムの提案を行う。

本論文の構成としては、第 2 章で本研究において用いた 2 つのモデル学習法についての説明を行い、第 3 章ではそれらの学習法を元にした提案する自動異常検知システムについて記述する。続いて第 4 章にて宇宙機データに対して提案手法を適用した実験結果について示し、その有用性について考察する。最後に第 5 章において結論と展望について述べる。

2. モデル学習法

本研究において用いたモデル学習手法として、次元削減手法の 1 つである Mixture Probabilistic Principal Component Analysis および確率過程として知られている Hidden Semi-Markov Models について述べる。

2.1 Mixture Probabilistic Principal Component Analysis

Mixture Probabilistic Principal Component Analysis (MPPCA) [Tipping 99] は Probabilistic Principal Component Analysis (PPCA) [Bishop 06] を混合した確率的な主成分分析である。この手法においては、非線形なデータ分布を局所線形な混合モデルとして扱う。ここで主成分分析(PCA)とはデータの線形な相関構造を獲得するための線形次元削減法であり、PCA による次元削減の写像 W_d はデータの共分散行列の固有値分解によって求まる固有ベクトルを固有値の大きいほうから d 個採用したものである。また PPCA は PCA を確率的な解釈を持って行なったもので、以上において、各コンポーネント(PPCA)のうち以下

連絡先: 田川 貴章, 東京大学大学院工学系研究科, 東京都目黒区駒場 4-6-1 東京大学先端科学技術研究センター 14 号館, E-mail: t_tagawa@space.rcast.u-tokyo.ac.jp

の式で表される再構成誤差が最も小さいものに対して再構成データの評価を行う(図.1).

$$\hat{y}_t = W_d^{(t)} W_d^{(t)T} y_t \quad (1)$$

ここで, $t=1,2,\dots,N$ であり N はデータ数を示す. 今, 低次元空間におけるデータは $x_t = f(y_t) = W_d^{(t)T} y_t$ で表され, 再構成データは $\hat{y}_t = g(y_t) = W_d^{(t)} x_t$ で表される. この時の混合モデルの尤度は

$$L = \sum_{t=1}^N \left\{ \sum_{i=1}^K \pi_i p(y_t | i) \right\} \quad (2)$$

で表される. ここで, π_i は PCCA の混合率, K はコンポーネント数を示している. 今, コンポーネント数 K はデータの分布や greedy EM algorithm [Vlassis 02] を参考にして決定した. また EM algorithm の初期値依存性回避には, 事前に k-means 法により PCCA 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為, 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.

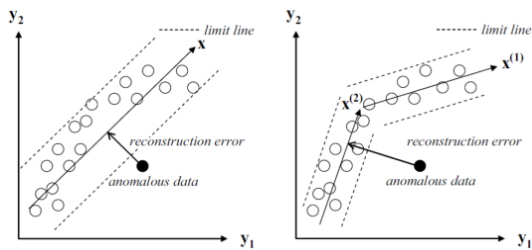


図. 1 左:PCA による再構成誤差, 右:MPPCA による再構成誤差

2.2 Hidden Semi-Markov Models

Hidden Semi-Markov Models (HSMMs) [Barbu 08] とはセミマルコフ過程におけるモデル化の手法である. ここで, セミマルコフ過程とは以下の 2 つの条件を満たすものを言う.

- マルコフ性を持つ
- どの状態においても任意の滞在時間分布を設定することが出来る.

これより HSMMs は, Hidden Markov Models (HMMs) において, そのマルコフ過程における任意の状態遷移においてその遷移をする時間の分布を設定することが出来るようにしたものと捉えられる. 今, HSMMs において観測過程 Y , 状態変数過程 Z の集合を $(Z, Y) = (Z_n, Y_n)_{n \in N}$ と置くとそのグラフィカルモデルは図.2 のように表される.

次に具体的なモデル化手法について述べる. HSMMs のモデル化においては, 次に示すセミマルコフ核 q_{ij} を訓練データから学習する.

$$q_{ij} = p_{ij} f_{ij} r_i \quad (3)$$

ここで,

- p_{ij} : あるモード i からモード j へ遷移する確率.
- f_{ij} : あるモード i からモード j へ遷移する時のモード i への滞在確率.
- r_i : あるモード i における観測データの出力確率.

であり, これら 3 つの確率分布を訓練データから学習する. 今回学習の際においては, p_{ij} は遷移パターンからヒストグラム型の確率を, f_{ij} および r_i はそれぞれの遷移パターンにおける滞在時間分布および出力分布を正規分布として学習した.

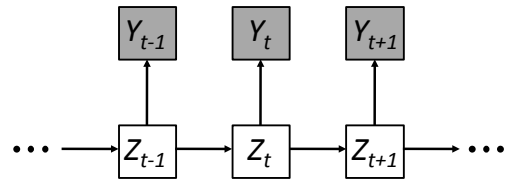


図. 2 HSMMs

3. MPPCA と HSMMs を用いたデータ監視

この章では, 前章で述べた MPPCA へ HSMMs を適用した新しい異常検知手法の枠組みについて述べ, さらにその際想定される 2 種類の異常について述べる.

3.1 提案する異常検知手法

今, MPPCA においては宇宙機のデータ分布がいくつかのモードを持つという仮定のもと, データを局所線形な部分空間の混合モデルとして扱い, テストデータの再構成・評価においては最も再構成誤差が小さくなるコンポーネント(PCCA)との再構成誤差を計算することにより異常検知を行っている. この時, 時系列データにおいてそれらは周期性を持つモード遷移を行う. しかしこの手法においてはモード遷移パターンを考慮しないため異常検知が難しく False Negative 及び False Positive の問題が起きてしまう. そこで本手法では HSMMs を MPPCA に適用し拡張することで再構成誤差だけではなく, モード滞在時間, 遷移確率も同時に学習することができ, モード遷移パターンを考慮したモデルを得ることを提案する.

具体的には, あるモードに滞在する時に得られる入力テストデータに対して, 全ての次モード遷移パターンにおけるセミマルコフ核を各々計算し, もっとも確率が高くなる時のモードを推定モードとして採用するというものである. ただし, 出力確率 r_i としては各モードにおける PCCA の再構成誤差を確率的に解釈したものを利用した.

以上において, HSMMs により得られるセミマルコフ核の確率分布より, 入力テストデータのエン트로ピを計算することで異常度を計算する.

3.2 想定される 2 種類の異常

以上で述べた異常検知システムにおいては次の 2 種類の異常が想定される(図.3).

1. 再構成誤差による異常.
2. モード遷移パターンによる異常.

そして今回提案する手法においては, 1.の異常に対しては MPPCA を, また 2.の異常に対しては HSMMs を用いた異常検知を想定している.

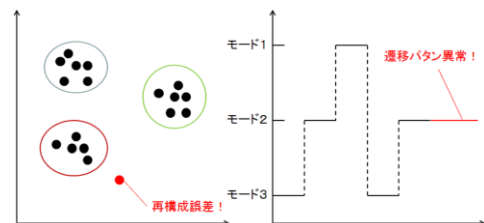


図. 3 左:再構成誤差による異常, 右:モード遷移パターンによる異常

4. 宇宙機データへの適用

本研究にて提案したモード遷移異常検知手法を, 実際の宇宙機テレメトリデータに適用した際の実験結果について示す.

4.1 モード内誤差異常検知

図.4 に示すような 900step 周辺から異常が発生している電源系の宇宙機テレメトリデータに対し、本研究で提案する手法を適用した。このケースにおいては、モード遷移パターンは正常だが、モード内において異常が発生している。その際の MPPCA による再構成誤差およびセミマルコフ核分布に対するエントロピの結果を図.5 に示す。

図.5 を見ると MPPCA による再構成誤差、セミマルコフ核分布に対するエントロピどちらにおいても 900step 付近で大きく値が上昇しており、異常を検知できていることが分かる。このことから、セミマルコフ核分布に対するエントロピは MPPCA による再構成誤差と比較してほぼ同等性能の検知能力を有していることが確認できる。

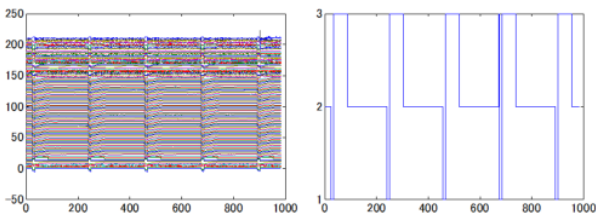


図. 4 左:宇宙機データ(モード内異常), 右:MPPCA による再構成誤差

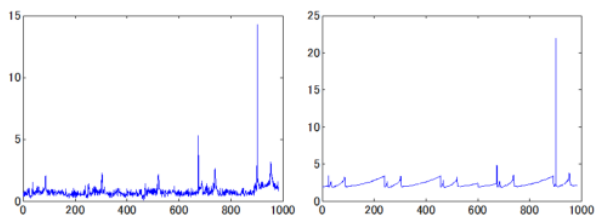


図. 5 左:MPPCA による再構成誤差, 右:セミマルコフ核分布に対するエントロピ

4.2 モード遷移異常検知

図.6 に示すような、今回用いたテレメトリデータに対して意図的に 600step からモードを一定としたデータを作成し、本手法を適用した。このケースにおいては、各モード内におけるデータ分布は正常だが、モード遷移パターンが異常となっている。その際の MPPCA による再構成誤差およびセミマルコフ核分布に対するエントロピの結果を図.7 に示す。

図.7 を見ると MPPCA による再構成誤差ではモード遷移異常をうまく検知出来ていないのに対して(False Positive), セミマルコフ核分布に対するエントロピを見るとモード遷移に異常が起きると値が大きく上昇しているのが見てとれる。このことから、セミマルコフ核分布に対するエントロピにてうまくモード遷移異常を検知できていることが確認できる。

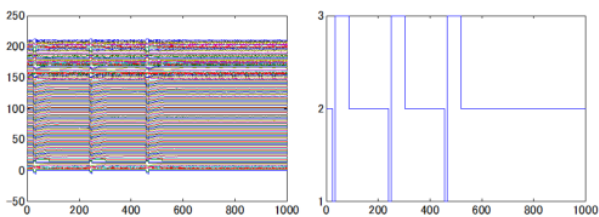


図. 6 左:宇宙機データ(モード遷移異常), 右: MPPCA により得られたモード遷移 (モード数 3)

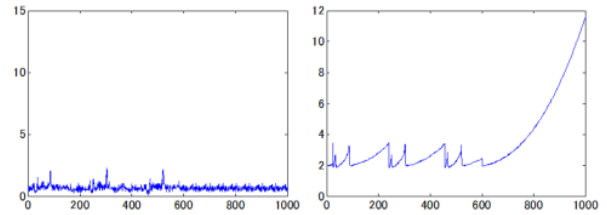


図. 7 左:MPPCA による再構成誤差, 右:セミマルコフ核分布に対するエントロピ

5. 結論

本研究における結論は以下の通りである。

- 従来研究では検知する事が出来なかった再構成誤差およびモード遷移異常に基づく False Positive 及び False Negative に同時対応した異常検知手法の提案を目的とした研究を行った。
- Mixture Probabilistic Component Analysis 及び Hidden Semi-Markov Models を組み合わせた新しい異常検知システムの提案をした。
- 実際の宇宙機データに適用する事でその検知性能の実証を行った。

次に今後の展望について述べる。

現状の問題点としては、まだ実データに対する適用例が少ないこと、HSMs の実装に関して改良の余地が残っていることが挙げられる。加えて、実際の利用においてモード遷移というものは実際の宇宙機を取り巻く環境の変化や、運用者が行うコマンド入力による状態変更の際に発生することが多い。このことにより、そうした実際のイベントとデータ挙動との相関を考慮したモード遷移監視システムの構築が必要であると考えられる。そこで今後は、運用者の提示する運用計画を事前知識として実際の宇宙機挙動のモード遷移を監視するシステムに関する研究を行っていく予定である。

参考文献

- [Bishop 06] Christopher M. Bishop: "Pattern Recognition and Machine Learning", Springer: Information Science and Statistics, 2006.
- [Tipping 99] M. E. Tipping and C. M. Bishop: "Mixtures of probabilistic principal component analyzers", Neural Computation, Vol. 11, No. 2, 1999.
- [Vlassis 02] N. Vlassis, A. Likas: "A greedy EM algorithm for Gaussian mixture learning", In Neural Processing Letters, Vol. 15, No. 1, 2002.
- [Barbu 08] Barbu, Vlad Stefan, Limnios and Nikolas: "Semi-Markov Chains and Hidden Semi-Markov Models toward Applications", Springer: Lecture Notes in Statistics, Vol.191, 2008.
- [吉木 09] 吉木明博: "次元削減と動的的特性推定による宇宙機テレメトリ監視法", 東京大学工学系研究科航空宇宙工学専攻修士論文, 2009.
- [乾 08] 乾稔: "次元削減による再構成誤差を用いた宇宙機異常検知手法に関する研究", 東京大学工学系研究科航空宇宙工学専攻修士論文, 2008.