

小型衛星向け知的テレメトリデータ監視システムの開発

Development of intelligent telemetry monitoring system for small satellites

小田 哲生^{*1} 塩井 宏亮^{*2} 谷本 啓^{*3} 武石 直也^{*4} 桑原 絢一^{*5} 矢入 健久^{*6} 酒匂 信匡^{*7}
 Tetsuo Oda Hiroaki Shioi Kei Tanimoto Naoya Takeishi Junichi Kuwabara Takehisa Yairi Nobutada Sako

^{*1*}^{*2*}^{*3*}^{*4*}^{*5} 東京大学大学院工学系研究科
 School of Engineering, University of Tokyo

^{*6} 東京大学先端科学技術研究センター
 Rcast, University of Tokyo

^{*7} キヤノン電子株式会社
 Canon Electric Inc.

Development of monitoring system detecting anomaly for satellites is of a great interest in space engineering. A number of researches on anomaly detection have been conducted, which utilize a variety of intelligent information technologies. We proposed a telemetry monitoring system integrating machine learning algorithms for detecting anomalies. We have implemented two anomaly detection algorithms — SLDS and MPPCA. This system is expected to help us to detect anomaly and deal with the problem on earlier phase and expands the possibility of satellite operation, however, we have been confronted with various difficulties on the application of this system. In this paper, we propose one of the solutions for the difficulties such as anomaly detection using telemetry data with missing values and outliers.

1. はじめに

宇宙開発ではミッションにかかるコストが膨大であるため、故障をゼロにするための取り組みがなされているが、宇宙機の故障を完全になくすことはできていない。そのため、異常を早期に発見し、対策をとることが重要になる。宇宙機の運用では、各サブシステムの状態を把握するためにテレメトリデータの解析を行う。しかし、テレメトリデータは多変数時系列データであり、変数間の相関も複雑なため、異常状態やその兆候を発見するのは容易ではない。

近年、機械学習・データマイニングの技術を応用して宇宙機の異常診断を行う研究がなされてきた[矢入 11]。この研究では、宇宙機のデータからモデルを学習し、それを基に異常を検知する。この技術を衛星の実運用で実証するために、我々はパターン認識及び機械学習を用いた異常診断による小型衛星のための知的データ監視システムの開発をしている。

小型衛星向けにテレメトリデータ監視システムの独自開発を行った理由は 2 点ある。1 点目は、小型衛星のための安価かつ高度な異常検知技術をもつ知的監視システムへのニーズが高いことである。大型の衛星ではテレメトリデータを監視できる地上システムを備えている施設もあるが、小型衛星においてはコストや人員の関係で十分な地上システムを備えられていないところも多く、テレメトリデータを網羅的に監視しきれない現状がある。

2 点目は、小型衛星は大型衛星に比べて新規機能を導入しやすいことである。知的監視システムの実用化には、数多くの衛星運用で実証されることが必要不可欠であるため、本システムを採用する衛星を増やすことで、システムの更なる向上を試みようとしている。

本システムでは、宇宙機のデータ分布が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、その仮定に基づいた異常検知手法を採用している。テレメトリデータは非線形性を持つが、非線形なモデルよりも局所線形なモデルの方が、異常検知能力が高いことが[矢入 11]の結果より示されている。本システムではモデル学習法として、主に確率的な局所線形モデルである Switching Linear Dynamical System(SLDS) [Murphy 98] と Mixture Probabilistic Principal Component Analysis(MPPCA) [Tipping 99]を採用した。これらのモデルを用いてモデル学習を行い、そのモデルとテストデータとのずれを評価することにより異常検知を行う。

異常診断システムの実用化を図るにあたり、2 つの面で工夫が必要となった。1 つは機械学習のアルゴリズム自体の改良、もう 1 つはデータの前処理である。これらについては改善を図ってきたが、学習データの選択方法を工夫する必要も出てきた。今までは同運用中のデータのみにおいて異常検知手法を適用させてきたが、試験中のデータなど異なる運用中のデータにも適用させる必要性が出てきたからである。この問題は異常検知を行う時点以前の全てのデータを用いて異常診断をすることで解決できるが、全てのデータをまとめてモデル化し異常診断をすることは計算コストの面から容易ではない。我々は前処理により、全てのデータの中から有用な学習データを効率的に作り出すことを考えている。本論文ではこの解決法についても言及する。

本論文の構成としては、第 2 章で開発中のシステムの概要と本システムで採用した異常検知手法の説明を行い、第 3 章で現在までに実装した実用化のための工夫について述べる。続いて第 4 章で実際の衛星の運用試験中のデータに適用した実験結果を示し、第 5 章で現在直面している課題に対する解決法を提案する。最後に第 6 章において結論を述べる。

2. 異常診断システムについて

2.1 システムの概要

本システムにおけるテレメトリデータの解析処理は基礎的な手法と機械学習アルゴリズムに大別される。基礎的な手法としては、以下の手法を採用した。

- 統計値の表示

連絡先: 小田 哲生, 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4号館, E-mail: oda@space.rcast.u-tokyo.ac.jp

- 固有値解析
- FFT 解析

統計値は最大値, 最小値, 平均値, 中央値, 標準偏差, 四分位点を表示する。ここで, 固有値解析の結果に基づき, 機械学習アルゴリズムにおける次元削減の次元数を決定する。FFT 解析では, データのスペクトルを評価する。これら基本的な統計・信号処理機能は, それら自体で高度な異常検知を実現できるわけではないが, 運用者がテレメトリデータの特徴を把握するうえで有用である。

機械学習アルゴリズムとしては以下の手法を採用した。

- Principal Component Analysis (PCA)
- Switching Linear Dynamical System (SLDS)
- Vector Quantized PCA (VQPCA)
- Mixture Probabilistic PCA (MPPCA)



図. 1 メインメニュー: MATLAB GUIにより実装

PCA ではデータのもつ潜在的な構造を可視化し, VQPCA では再構成誤差を, SLDS, MPPCA では対数尤度を評価する。これらの複数の手法を統合し, 異常診断を行っている。

2.2 採用した手法の概要

ここでは採用した異常検知手法のなかでも特に有用と思われる SLDS と MPPCA について, それらの表すモデルの概要を述べる。どちらも確率的な局所線形モデルであるが, MPPCA に比べ, SLDS は時系列遷移まで考えているため, SLDS の方がモデルの説明能力は高い。しかし, テレメトリデータは外れ値を含み, さらに, サンプリング周期が異なるデータを含むこともあるなど理想的なデータではない場合も多くある。このようにデータが理想的でない場合, SLDS は適用しづらい。そこで, 本システムでは時系列遷移を考えないモデルである MPPCA も同時に採用している。

(1) SLDS

SLDS は Linear Dynamical System(LDS)[Bishop 06]にモード遷移性を追加し拡張したものであり, 以下の方程式に従う状態空間モデルである。

$$\mathbf{x}_t = \mathbf{A}(s_t)\mathbf{x}_{t-1} + \mathbf{w}(s_t) \quad (1)$$

$$\mathbf{y}_t = \mathbf{C}(s_t)\mathbf{x}_t + \mathbf{v}(s_t) \quad (2)$$

$$\mathbf{w}(s_t) \sim \mathcal{N}(\mathbf{w} | \mathbf{0}, \mathbf{Q}(s_t)) \quad (3)$$

$$\mathbf{v}(s_t) \sim \mathcal{N}(\mathbf{v} | \mathbf{0}, \mathbf{R}(s_t)) \quad (4)$$

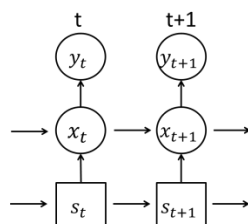


図. 2 SLDS のグラフィカルモデル

このモデルにおいて, パラメータ $\theta = \{\mathbf{A}, \mathbf{C}, \mathbf{Q}, \mathbf{R}, \mathbf{Z}\}$ を EM アルゴリズムにより推定する。ここで \mathbf{Z} はモード遷移を表す行列

$$\mathbf{Z}(i, j) = \Pr(s_{t+1} = j | s_t = i) \quad (5)$$

を表す。(ただし, $i, j = 1, \dots, k$ (k はモード数))

EM アルゴリズムを用いた SLDS の学習法については文献 [Murphy 98] を参照されたい。正常状態のデータでモデルを学習し, テストデータにおける尤度の変化を評価することで, 尤度低下時に異常と判断している。

(2) MPPCA

テレメトリのデータは高次元のデータだが, 各系列間には一定の相関があるため, 本質的にはより低次元の空間で表すことができる。今までにも次元削減手法を用いた異常診断研究がなされてきた [矢入 11, 田川 11]。その 1 つが MPPCA である。MPPCA は Probabilistic Principal Component Analysis (PPCA) [Bishop 06] を混合した確率的成分分析である。

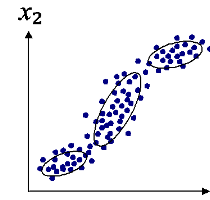


図. 3 MPPCA の確率密度分布モデル

SLDS と同様, 正常状態のデータでモデルを学習し, テストデータにおける尤度の変化を評価することで, 尤度低下時に異常と判断している。

混合モデルの尤度は

$$\mathcal{L} = \sum_{t=1}^N \left\{ \sum_{i=1}^K \pi_i \Pr(\mathbf{y}_t | i) \right\} \quad (6)$$

で表される。

ここで, π_i は PPCA の混合率, K は PPCA の混合数, N はデータ数, \mathbf{y}_t は時刻 t における観測データを示す。

また, PPCA の各コンポーネントのうち帰属度 (Responsibility) が最も高いものに対して再構成データを評価する。再構成データは以下の式で表される。

$$\hat{\mathbf{y}}_t = \mathbf{W}_d^{(t)T} \mathbf{W}_d^{(t)} \mathbf{y}_t \quad (7)$$

ここで, $\hat{\mathbf{y}}_t$ は再構成データ, $\mathbf{W}_d^{(t)}$ は次元削減の射影行列を表す。このようにして擬似的に再構成誤差を求め, 系列ごとの再構成誤差を表示することで, 異常度の高い系列を調べることができる。

3. 実用化のための工夫

3.1 前処理における工夫

2.2 節でも述べたように実際の衛星運用では理想的なデータが得られないことがよくある。そこで, 我々は実データに対し, 外れ値除去・正規化の 2 段階の前処理を行い, より安定的に異常検知を行うことができるようにしている。

(1) 外れ値除去

本稿では, 「外れ値」とは, センサにおける異常値でありシステムにおける異常値とは区別されるべきものとして扱う。宇宙機の性質上, 突発的にノイズが加わったり, 値が欠損し続けたりし, 近傍の時刻の値との差が大きくなることが多い。外れ値を含む

データを用いて学習を行うと正常時のモデルを正しく学習できない。そこで、本システムでは、外れ値を検知・除去・補正する前処理として、ロバスト統計量と呼ばれるパーセンタイルを用いた外れ値検知を採用した。

(2) 正規化

テレメトリデータは電圧、温度、回転数など様々な物理量を表す変数群によって構成されているため、系列ごとに値のオーダーは異なる。系列ごとにオーダーの異なるデータをそのままモデル化すると、値の大きい系列の寄与が相対的に大きくなってしまふという問題がある。そこで、現在は学習データにおいて各系列の平均を 0、分散を 1 にするような正規化変数を求め、その正規化変数を用いてテストデータも正規化している。

しかし、この単純な正規化法には 2 つの問題がある。1 つ目は、値の変動の少ない変数の寄与が過度に大きくなることである。微小なノイズが拡大されることにより、異常診断結果が不安定になってしまう。2 つ目は、学習データ(学習期間)ごとに正規化変数が異なることである。全てのデータから有用な学習データを作り出す際に、全データにおいて正規化の変数が異なるとデータを扱いづらいという問題がある。

これらの問題の改善を図るためには、全区間で共通かつ適切な正規化パラメータが必要である。我々は設計情報を利用し、設計上限値・下限値がそれぞれ+1,-1 になるように正規化を行った。また、設計情報が得られない系列においては、十分に長いデータにおける 1 パーセンタイル点・99 パーセンタイル点をそれぞれ+1,-1 にするような正規化を行った。

3.2 学習アルゴリズムにおける工夫

(1) k-means 法の安定化

本システムでは、SLDS, MPPCA を用いたモデル学習において、モード分けの初期値を得る手段として k-means 法を用いている。これは最も単純なクラスタリング手法の 1 つであるが、クラスタ数が大きい場合、非最適な局所解に収束する可能性が高くなる。本システムでは主成分分析を用いた漸次的分割法[Ding 04]を採用することで、この点を改善した。

(2) EM アルゴリズムの安定化

SLDS を学習する際に EM アルゴリズムを用いる。しかし、EM アルゴリズムには、初期値依存性を持ち局所解に収束しやすいという欠点があるため、実装する際にはその安定性に注意する必要がある。

この問題を解決するために、確定的アニーリング EM アルゴリズム[上田 97]を用いた。この方法は確率密度関数に統計力学の温度パラメータに相当する変数を導入することで、より簡単な確率密度関数から学習する方法である。簡単な関数から段階的に学習させていくことで学習の精度を向上させることができる。

4. テレメトリデータへの適用例

本章では、東京大学中須賀研究室で開発中の小型衛星 Nano-Jasmine の地上運用試験中における電源系のデータへの本システムの適用例を示す。用いたデータは図 4 のような 22 次元のデータである。このデータは 6000step のデータであり、このうち 1~4000step を学習データ、4001~6000step をテストデータとして適用した。ただし、step 数が 142, 1980, 4509 のところで実時間上では途切れており、細切れのデータをつなげたものである。また、このデータに異常は含まれていない。4.1 節に SLDS における適用例を、4.2 節に MPPCA における適用例を示す。

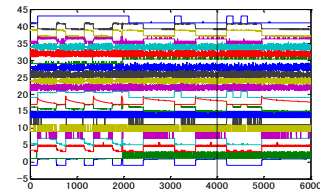


図 4 Nano-Jasmine 電源系データ

4.1 SLDS

SLDS における適用例を図 5 に示す。図 5 の左図は SLDS で学習したモード遷移図を、右図は SLDS による負の対数尤度を表す。図 4 と比較すると、モード遷移状態も確からしく、正しく学習が行われていることがわかる。

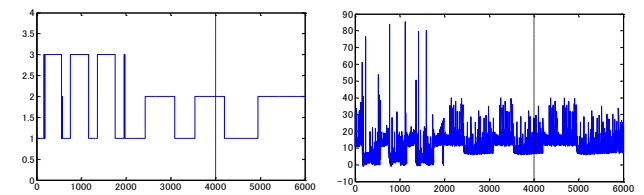


図 5 左:SLDS によるモード遷移図 右:SLDS による負の対数尤度

他の既存の衛星の異常データで同様に適用させた際には、異常発生時に尤度が急激に低下し異常を検知することができた。

4.2 MPPCA

同様に、MPPCA における適用例を図 6 に示す。左図は MPPCA で学習したモード遷移図を、右図は MPPCA による負の対数尤度を表す。SLDS と同様、モード遷移の学習結果は確からしいことがわかる。

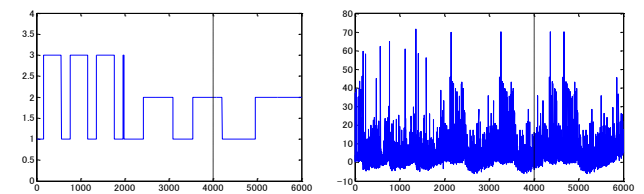


図 6 左:MPPCA によるモード遷移図 右:MPPCA による負の対数尤度

他の既存の衛星の異常データで同様に適用させた際には、異常発生時に尤度が急激に低下し異常を検知することができた。

加えて、MPPCA による系列ごとの再構成誤差を図 7 に示す。

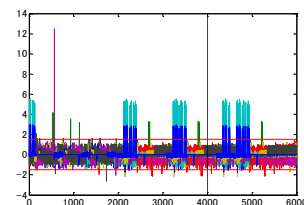


図 7 MPPCA による系列ごとの再構成誤差

擬似的に再構成誤差を評価することもできるため、特に系列ごとの再構成誤差を評価することで、系列ごとの異常度を評価することができる。

5. 今後の展望

本システムの実用化を目指すなかで、我々は様々な問題に直面した。テレメトリデータは、地上との通信中のデータはサンプリング周期が短く、欠損値がほとんどないが、通信中でないときのデータはサンプリング周期が長く、欠損値が非常に多いという特徴を持つ。加えて、試験中など異なる運用中のデータにも対応する必要がある。これらの課題を解決していかなければならない。

欠損値・サンプリング周期の問題はテレメトリデータを通信中のデータとそれ以外のデータに分類し、それぞれでモデル化を行うことで、解決可能であると考えられる。テストデータが通信中のデータの場合は、通信中のデータで学習したモデルを用いて、通信中でないときのデータの場合は、通信中でないときのモデルを用いて異常診断を行えばよい。

異なる運用への対応としては、テストデータにしか含まれないモードが学習データにも含めれば、正しく異常診断をすることができる。現時点以前の全てのデータを学習データとしてモデル化すれば、正しく異常診断をすることが可能である。しかし、この方法は計算コストの面から非現実的である。前処理により、現時点以前の全てのデータの中から有用な学習データを効率的に作り出す必要がある。

我々は欠損値・サンプリング周期の問題の解決法と同様の方法により、異なる運用への対応もできると考えている。つまり、通信中とそれ以外で分類するのと同様に、現時点以前の全てのデータをより詳細なモード(通信中かつ日照かつ定常運用中のモードなど)に分類する。そして、各モードのもつ物理的意味を全区間で統一したうえで、全区間のデータをモードごとにモデル化し、これらのモデルをシステム上で保持しておく。学習データに含まれていないモードがテストデータに含まれている場合には、そのモードに対応するモデルを学習データに含めることで異常検知を行う。新しいデータが取得された際に、データをモードごとに分類し、それぞれで更新を行うことで正確なモデル学習ができると考えている。

加えて、全区間のデータをモードごとに分類する際に、モード分けを運用者の知識と合わせることも、実用化するうえでは重要になる。運用者の知識との乖離が生じるとシステムによる異常検知が信頼されなくなる可能性が考えられるからである。そこで、教師なし学習によるモード分けを運用者によるフィードバックにより修正する必要もあると思われる。

この方策をとることで、宇宙機の異常検知をより正確に行うことができると考えている。

6. まとめ

本研究において、我々は小型衛星の運用支援を目的とした知的テレメトリデータ監視システムを開発中である。現在のところ SLDS, MPPCA が有効だと考え、システムの中心的アルゴリズムにしている。また、数値データだけでなく、“ON”、“OFF”の2値をとるようなステータス変数も含めて異常診断を行う仕組みを導入するなど、異常検知手法の拡張を図っている。しかし、実際のテレメトリデータに適用するには、実データ特有の問題が多く存在する。本稿では、それらの問題を具体的に挙げて整理し、解決策を示した。

参考文献

[Bishop 06] Christopher M. Bishop: “Pattern Recognition and Machine Learning”, Springer: Information Science and Statistics, 2006

[Ding 04] Chris Ding and Xiaofeng He: Principal Component Analysis and Effective K-means Clustering, Proceeding of SIAM International Conference on Data Mining, 2004

[Murphy 98] Kevin P. Murphy: “Switching Kalman Filters” Technical report, Dept. of Computer Science, University of California, Berkeley, 1998

[田川 11] 田川貴章, 矢入健久, 高田昇, 山口由仁: 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, 人工知能学会, 2011

[Tipping 99] M. E. Tipping and C. M. Bishop: “Mixtures of probabilistic principal component analyzers”, Neural Computation, Vol. 11, No. 2, 1999.

[上田 97] 上田修功, 中野良平: 確定的アンニーリング EM アルゴリズム, 電子情報通信学会論文誌, D-II, Vol. J80-D-II, No.1, pp267-276, 1997

[矢入 11] 矢入健久, 乾稔, 河原吉伸, 高田昇: 次元削減とクラスタリングによる宇宙機テレメトリ監視法, 日本航空宇宙学会論文集, Vol. 59, No.691, pp.197-205, 2011