

弱教師あり学習による人工衛星テレメトリデータの状態監視手法

二木 浩司[†]
 東京大学工学部航空宇宙工学科

リュー ジュンフイ[‡] 矢入 健久[‡] 福島 裕介[‡]
 東京大学先端科学技術研究センター

1 研究の目的

人工衛星の状態を監視し、異常を検知するためにデータ駆動の機械学習アルゴリズムが用いられている [1]。これに関して、衛星テレメトリのうち正常データをいくつかのモードにクラスタ分けすることは一つの状態監視法であるが、完全に教師なしのアルゴリズムを用いた場合にはそのクラスタリングの出力は専門家の知見に必ずしも一致しない。

本報告では、教師なし学習により行われた衛星状態のクラスタリングに対して、運用者が知見に基づき一部修正を加えられるという弱教師あり学習アルゴリズムによる状態監視法の結果を示す。また、修正されたクラスタリングによって衛星の状態遷移パターンを把握できることを示す。

2 制約付き K-means 法によるクラスタリング

k-means 法は最も一般的なクラスタリング手法の一つである。上に書いた理由により、教師なしで得られたクラスタリングを運用者が自由に統合・分割をすることで解釈性を高めることが重要である。このような弱教師あり学習の思想に基づく k-means 法として、制約付き k-means 法 [2] がある。これは入力データセットに対して、選択したデータ同士が同じクラスターに所属しなければならない/同じクラスターに所属してはならないという条件 (must-link, cannot-link) を加えて k-means 法を実行するものである。

この制約付き k-means 法を用いて実際の人工衛星テレメトリデータのクラスタリング実験を行った。用意したデータは宇宙航空研究開発機構 (JAXA) より提供された静止衛星のバス系データ (電気系および熱系) 計 176 次元の変数の変動を 1 年間記録したものである。外れ値処理を適切に行ったのち、176 次元の変数を主成分分析 (PCA) によって 20 次元に削減し、クラスター数 k を 3 とした k-means 法によってクラスタリングした。このとき、制約条件をつけた場合とつけなかった場合について、以下のようなクラスタリングの違いが見られた。

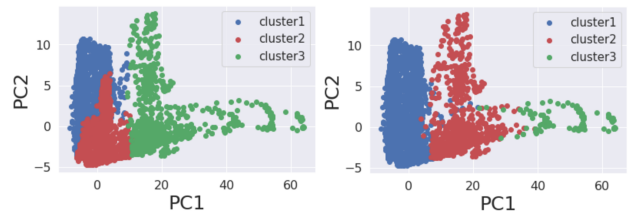


図1 第1, 第2主成分空間におけるクラスタリングの様子。制約条件なしの場合が左側, ありの場合が右側。

制約条件なしの場合のクラスタリングのうち、クラスター1・2に相当する部分が統合されて制約条件ありの場合のクラスター1となっている。また制約条件なしの場合のクラスター3が分割されて制約条件ありの場合のクラスター2・3となっていることが分かる。このようなクラスターの整理は、運用者にとっての解釈性を高めることを目指したものである。

制約条件なしの場合に得られる所属クラスターの3か月間の遷移を図2に示す。なおこのグラフではクラスターの切り替わりの様子を強調するために切り替わるタイミングで黒色の線が入っているが、これはクラスター遷移中にクラスター番号が整数でない値をとるということを意味しない。比較のために入力データの第1主成分の値の変動も合わせて示す。

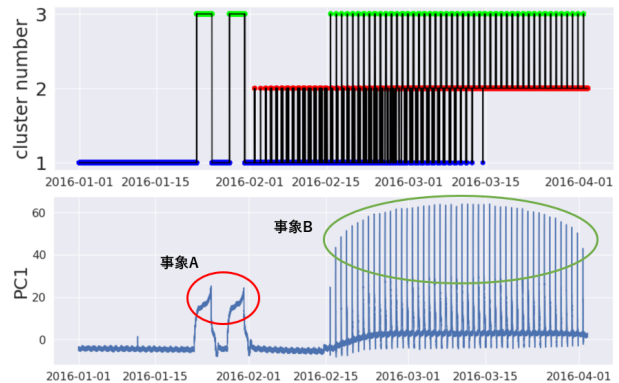


図2 制約条件なしのクラスタリングによる3か月間の所属クラスターの遷移。下図は第1主成分の変動。

第1主成分の変動を観察すると、図に事象A, Bと示した特異な挙動が観察できる。ゆえに、その発生期間中は他の期間 (正常期間) とはそれぞれ異なるクラスターに所属するように分類されることが望ましい。しかしながら、上のクラスター遷移はそのような要請に合致しない。

このような要請に基づき、クラスタリングに制約条件を加

A Condition Monitoring Method for Artificial Satellites based on Semi-Supervised Learning

[†]Hiroshi Futatsugi, Department of Aeronautics and Astronautics, the University of Tokyo

[‡]Chun Fui Liew, Takehisa Yairi, Yusuke Fukushima, Research Center for Advanced Science and Technology, the University of Tokyo

えた分類結果が図1の右側の様子であった。この場合の所属クラスタの遷移を図3に示す。制約条件を加えるにあたって、1年間のデータをすべて入力すると計算コストが大きくなりすぎたため、データを50タイムステップおき(26分40秒おきに相当する)に間引き、そのうち1%程度にあたる223個の制約条件(must-link 211個, cannot-link 12個)を与えた。

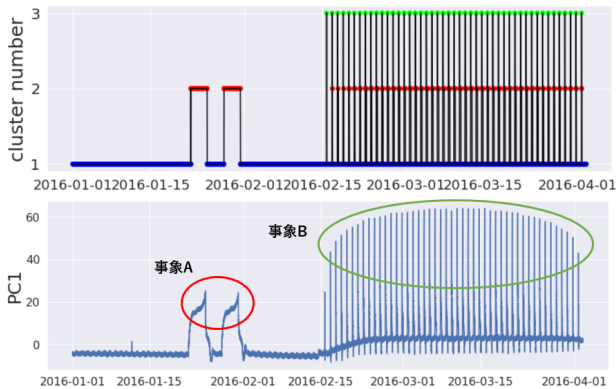


図3 制約条件がある場合の3か月間の所属クラスタ遷移。事象A, Bの所属クラスタが分離されている。

制約条件を加えたことで要求したようなクラスタリングを行うことができている。事象Aはクラスタ2に所属するように分類されている。事象Bの期間中はクラスタ3,2,1,...という遷移パターンを持ち、クラスタ3はこの期間に特徴的なものになっている。

3 クラスタ遷移に対するマルコフモデルの適用

次に、上で得られた事象B発生期間中のクラスタ遷移に対し、通常と異なる状態遷移が混入している場合に発見できるかを確認する。事象B発生期間中はクラスタ3,2,1,...という遷移が典型的なパターンのように見て取れる。仮にこの中に全く異なるパターンの遷移が混入していたならば、それはシステムの異常であると判断できる。このようなパターンをマルコフモデルを用いて抽出することを試みる。3つのクラスタが潜在状態かつそのまま観測量であるとしてモデル化する。

モデルの有効性を確認するため、期間中に通常と異なる遷移パターンを含むデータを人工的に作成した。すなわち、3月4日の未明から3月6日の夕方にかけての期間中は、図4のように実際の所属クラスタよりも番号が1だけ大きなクラスタ(もとの所属クラスタが3の時はクラスタ1)に所属しているとした。

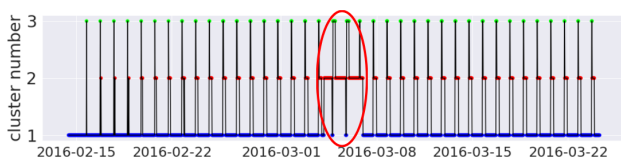


図4 通常と異なる状態遷移を含めたクラスタ遷移。

このデータに対して、上の確率モデルに基づいて計算した負の対数尤度、すなわち異常度の値は以下のようになった。通常と異なる状態遷移の期間で値が他の部分よりも顕著に増大していることを確認した。

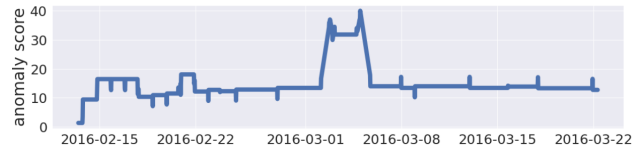


図5 通常と異なる遷移を含むデータの異常度の変化。

この手法により、弱教師あり学習によるクラスタリングと時系列解析を組み合わせた異常検知が可能である。

4 総括

本報告では、代表的な弱教師あり学習のアルゴリズムである制約付きk-means法を人工衛星テレメトリデータに適用した。事象A, 事象Bの存在を運用者の知見として入力した結果、アルゴリズムはそれを考慮したクラスタリングを提示した。そのクラスタリングの遷移の様子をマルコフモデルによって解析し、元データに人為的に挿入した通常とは異なる状態遷移を検知可能であることを示した。

本報告により、弱教師あり学習に基づく異常検知・状態監視の必要性が改めて確認された。人間と機械が協働するというコンセプトを今後も発展させ、たとえば衛星開発時に専門家が想定したルール・ベースのモデルと、テレメトリデータからそのモデルを修正するというデータ・ベースのシステムを相互に組み合わせるなどのアイデアがあり得る。

謝辞

本研究開発の一部は、JST START (JPMJST1814)の支援により行われた。また、宇宙航空研究開発機構(JAXA)からデータ提供と利用許諾を得ている。

参考文献

- [1] Yairi, T. et al. (2017). A Data-Driven Health Monitoring Method for Satellite Data Based on Probabilistic Clustering and Dimensionality Reduction. *IEEE Trans. Aerosp. Electron. Syst.*, 53(3), 1384-1401.
- [2] Wagstaff, K., Cardie, C., Rogers, S., & Schroedl, S. (2001). Constrained K-means Clustering with Background Knowledge. *Proc. of the 18th Intl. Conf. on Machine Learning*, 577-584.