

# センサデータとテキスト運用記録を活用した宇宙機運用支援

## Operations Support for Artificial Satellites with Sensor Data and Logs

武石直也 \*<sup>1</sup> 矢入健久 \*<sup>1</sup>

Naoya Takeishi Takehisa Yairi

\*<sup>1</sup> 東京大学大学院工学系研究科

School of Engineering, the University of Tokyo

For secure operation of a spacecraft, it is of great importance to survey states of the spacecraft constantly, which is often too costly for human operators. One of promising solutions is a machine learning technique that learns models of the data and survey new data automatically. Since information on spacecraft' states is included not only in sensor data but also in text data that are written by designers and operators of the spacecraft, the model we learn has to be adequate for the both types of data. We proposed a model that consisted of mixture of factor analyzers (MFA) and latent Dirichlet allocation (LDA), and confirmed the validity of the model using actual telemetry data.

### 1. 背景

人工衛星や深宇宙探査機などの宇宙機は、打ち上げ後に修理することが極めて難しい。そのため、安全な運用のためには宇宙機の状態を監視することが欠かせない。ところが、宇宙機は複雑かつ大規模な機械/電気電子システムであるから、宇宙機の運用から得られるデータを全て運用者が監視するのはコストが高い。そこで、機械学習によって宇宙機運用から得られるデータのモデルを学習して、データ可視化や自動的な状態監視などを行いたい。本稿の目的は、宇宙機から得られるデータの性質を考慮したモデルを検討し、実際の人工衛星から得られたデータに適用して実利用をふまえた利用可能性の考察を行うことである。

宇宙機運用から得られるデータとして、一般的には図1に示す4種類のものが考えられる。1つめは、宇宙機に装備された各種センサから得られるセンサデータである。これは多くの場合、刻々と変化する内外の物理量を捉える連続値として得られる。2つめは、宇宙機が処理上取得している内部状態の記録である。これは（センサデータとは異なり）離散的な記号列として表すことができる。3つめは、地上局から宇宙機へ送られる運用コマンドであり、内部状態の変化と密接に関係していると考えられる。そして4つめは運用上発生した文書類で、運用計画、議事録、運用結果の解析などを指す。このように、センサデータや内部状態記録といった観測値だけでなく、運用コマンドや運用記録文書といったテキストも運用上発生している点には注意が必要である \*<sup>1</sup>。

本稿では、図1に示すようなデータを統一的な枠組みで扱うことのできるモデルを提案する。ただし、このとき考慮すべき宇宙機データの性質として次のような点がある。

#### センサデータ取得（保存）頻度が低い場合がある

宇宙機との通信可能時間や宇宙機のメモリ量は限られており、取得したセンサデータや内部状態記録の全てを保存・送信することは難しい。そのため、センサデータや内部状態記録の表す物理量や状態が本来時系列的なものであっても、その情報を十分に捉えられないことがある。これは、小型の人工衛星において特に顕著である。

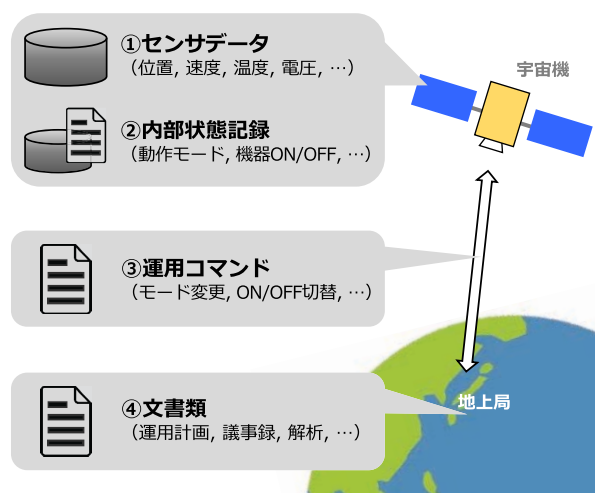


図1: 宇宙機運用上得られるデータのイメージ

#### センサデータと文書類の対応が1対1ではない

当然ながら、センサデータの取得間隔と文書類の生成間隔は大きく異なる。このような性質の異なるデータを扱うことのできるモデルである必要がある。

これらの性質をふまえたモデルとして、本稿では混合因子分析 (mixture of factor analyzers, MFA) [Ghahramani 96] と潜在ディリクレ配分 (latent Dirichlet allocation, LDA) [Blei 03] を組み合わせたモデルを用いる。これまでの研究 [矢入 11] で、MFAのようにクラスタリングと次元削減を行うことのできるモデルが宇宙機センサデータの性質（多モード性による非線形性など）に適切であることがわかっている。また、LDAもMFAと同様に混合モデルの一種であり、両者の組み合わせは相性が良い。

### 2. 使用するモデル

この節では、使用するモデルの概要と推論方法について述べる。

連絡先: takeishi@space.rcast.u-tokyo.ac.jp

\*<sup>1</sup> これは、運用者が係るようなシステムでは一般的にいえることであろう

## 2.1 生成モデル

使用するモデルのグラフィカルモデルを図2に示す。図中および以降の表記において、 $\mathbf{y}$  は連続的な刻々の観測値を、 $\mathbf{t}$  は離散的な刻々の観測値を、 $\boldsymbol{\eta}$  は文書中の単語を表す。すなわち、センサデータは  $\mathbf{y}$  に割り当て、文書類は  $\boldsymbol{\eta}$  に割り当てることができる。一方、内部状態記録は定期的を取得されるためセンサデータに対応するように  $\mathbf{t}$  に割り当てることができるし、変化が少ない場合は文書中の単語として見なして  $\boldsymbol{\eta}$  に割り当てることができる。

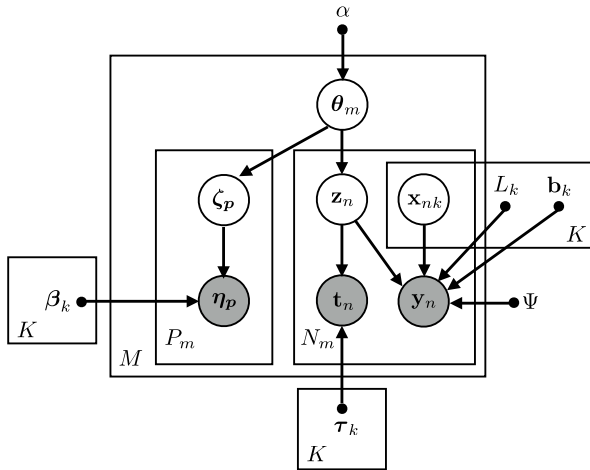


図 2: 使用するモデルのグラフィカルモデル

このモデルでは、データ全体を  $M$  個のまとまりとして考える。また、 $m$  番目のまとまりには  $N_m$  個の観測値からなるセンサデータ（および内部状態記録）が対応すると同時に、 $P_m$  個の単語からなる文書が対応する。例えば、1 年分のデータを 1 日ごとのまとまりとして考えた場合は  $M = 365$  となる。データのまとまりを考えることによって、刻々の観測値列と文書とを結びつけることが可能になる。

生成モデルの詳細は次のとおりである。まず、まとまりごとの潜在変数として  $\boldsymbol{\theta}$  をおく。

$$\boldsymbol{\theta}_m \sim \text{Dir}(\boldsymbol{\alpha}) \quad (\text{for } m = 1, \dots, M) \quad (1)$$

また、まとまり内の各観測値と各単語に対応する潜在変数  $\mathbf{z}$  と  $\boldsymbol{\zeta}$  をおく。

$$\mathbf{z}_n | \boldsymbol{\theta}_m \sim \text{Cat}(\boldsymbol{\theta}_m) \quad (\text{for } n = 1, \dots, N_m) \quad (2)$$

$$\boldsymbol{\zeta}_p | \boldsymbol{\theta}_m \sim \text{Cat}(\boldsymbol{\theta}_m) \quad (\text{for } p = 1, \dots, P_m) \quad (3)$$

さらに、混合因子分析の各観測値に対応する潜在変数を考える。

$$\mathbf{x}_{nk} \sim \mathcal{N}(\mathbf{0}, I) \quad (4)$$

これらの潜在変数が与えられたとき、各観測値と単語の尤度は次のとおり。

$$\mathbf{y}_n | \mathbf{z}_n = k, \mathbf{x}_{nk} \sim \mathcal{N}(L_k \mathbf{x}_{nk} + \mathbf{b}_k, \Psi) \quad (5)$$

$$\mathbf{t}_n | \mathbf{z}_n = k \sim \text{Cat}(\boldsymbol{\tau}_k) \quad (6)$$

$$\boldsymbol{\eta}_p | \boldsymbol{\zeta}_p = k \sim \text{Cat}(\boldsymbol{\beta}_k) \quad (7)$$

なお、ここでは簡単のためひとつの  $\mathbf{t}$  についてのみ記しているが、実際は複数の  $\mathbf{t}_1, \dots, \mathbf{t}_t$  を考えることができる。

## 2.2 推論と学習

推論と学習は EM アルゴリズムで行った。E ステップでは  $\boldsymbol{\theta}, \mathbf{z}, \boldsymbol{\zeta}$  の事後分布に関する変分推論と  $\mathbf{x}$  に関する事後分布の計算を行い、M ステップではパラメタ  $L_{1:K}, \mathbf{b}_{1:K}, \Psi, \boldsymbol{\tau}_{1:K}, \boldsymbol{\beta}_{1:K}, \boldsymbol{\alpha}$  について不完全対数尤度の下限を最大化する。紙面の都合上、計算式などは省略する。詳細は [Blei 03] や [Ghahramani 96] などを参照されたい。

## 3. 実験

この節では、第 2 節のモデルを実データに対して適用した結果を示す。使用データの詳細は 3.1 節に示すとおりである。また、この節で述べる実験全てにおいて、データのまとまりは 1 日ごととした。モデルの設定として、全ての実験で混合要素数  $K$  は 8 とし、 $\mathbf{x}$  の次元数は 15 とした。

### 3.1 使用データ

使用したデータは、宇宙航空研究開発機構の運用する小型実証衛星 4 型 (SDS-4) \*2 で取得されたテレメトリデータと運用記録である。使用したデータの構成イメージを図3に示す。本稿ではテレメトリデータのうち 92 種類のセンサデータと 356 種類の内部状態記録を用いた。なお、テレメトリは約 1 分などの間隔で刻々と取得されているが、必ずしも時系列性を捉えられるほど取得頻度が高いわけではない\*3。さらに、各日に行った軌道上実験の名称を運用記録として用いた。これは、軌道上実験の名称を単語（実験機器、実験設定など）に分解した短い文書である。

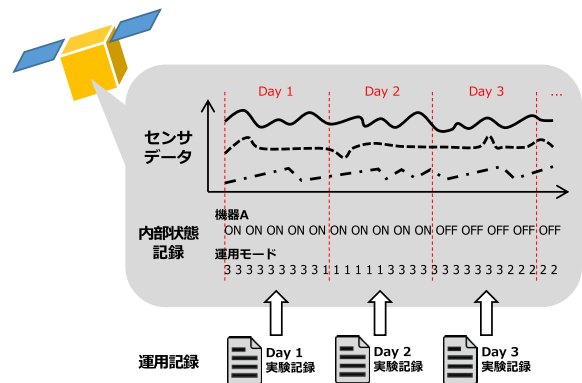


図 3: 使用したデータの構成イメージ

SDS-4 のテレメトリデータに関しては、これまで混合確率主成分分析 [Tipping 99] などを用いて異常検知を行ってきた [矢入 14, 西村 14]。本稿では、異常検知に限らず一般的なテレメトリデータの把握や、その他宇宙機の運用を支援するような利用法を意識して実験を行った。

### 3.2 可視化と異常検知

図4(a)に、SDS-4で2013-2014年に取得されたセンサデータ(92変数)を示す。このように多くのセンサデータを一度に図示しても人間の目で即座にわかることは少なく、センサデータから知見を得るためには詳細な解析が不可欠である。しかし、コストなどの点から詳細な解析が現実的でない場合がある。そこで、潜在変数を含むモデルを学習することでデー

\*2 <http://www.jaxa.jp/projects/sat/sds4/>

\*3 欠損部分には 0 次補間を施した

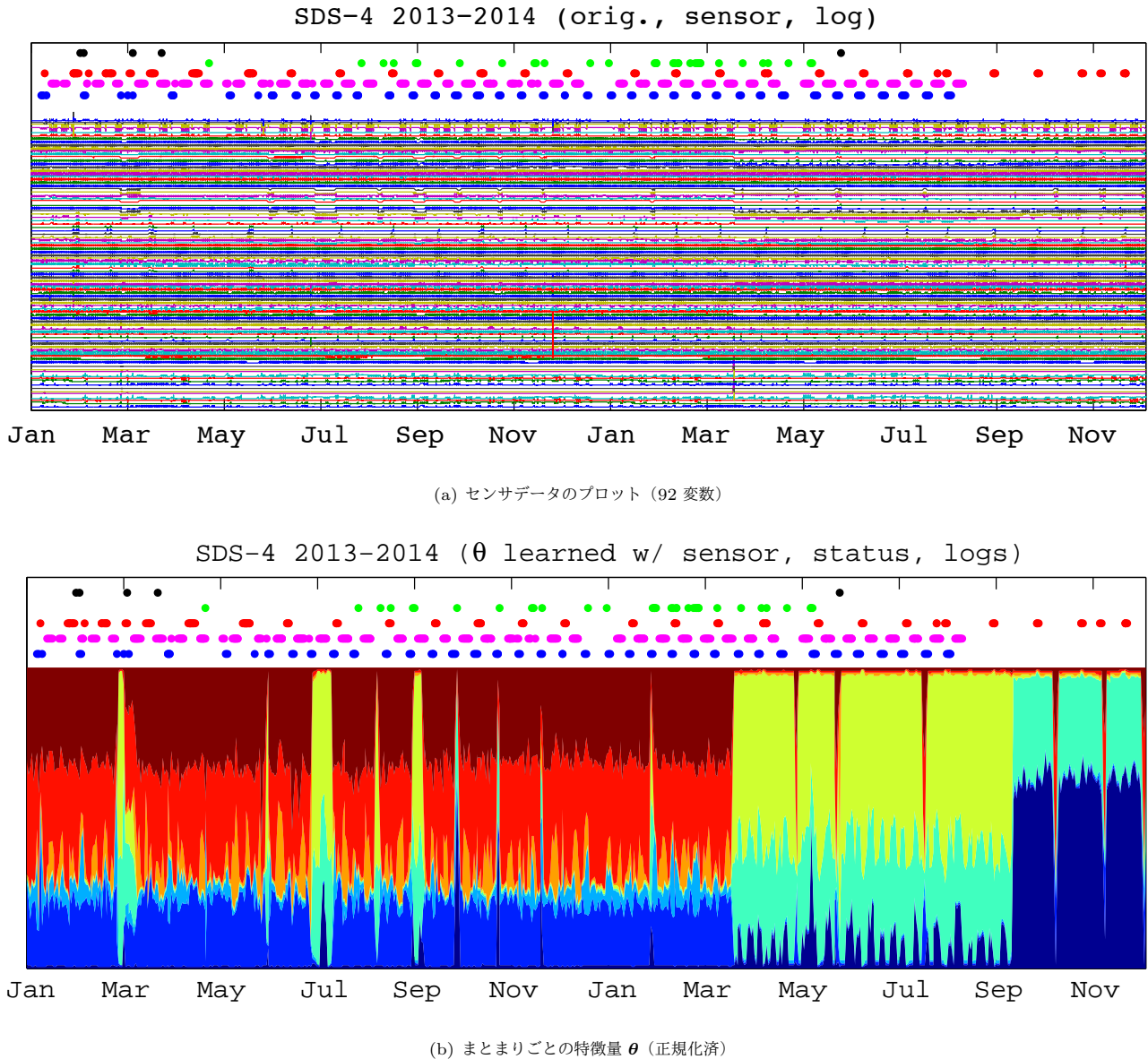


図 4: 2013-2014 年のセンサデータ元値とモデル適用結果

データの概観を把握できれば有用である。ここでは、SDS-4 のセンサデータ、内部状態記録、および運用記録に対して第 2 節で紹介したモデルを適用し、まとまり (1 日) ごとの特徴量  $\theta$  を正規化したものを図 4(b) に示した。なお、図 4(a) および図 4(b) の上部に示した点は、運用記録から実施している軌道上実験の種類を判別して色分けしプロットしたものである。

まとまりごとの特徴量  $\theta$  を図示することで、データの挙動を概観することができる。特に、図 4(b) 右部の 2014 年 3 月末あたりから挙動の大きな変化が見られるが、これは 3 月末に発生した衛星姿勢異常の影響によるものと考えられる。

### 3.3 類似運用の検索

ここでは、ある日のセンサデータ等が過去の運用のうちどの日と似ているかを検索するというタスクに取り組む。このような類似運用の検索を自動的に行うことができれば、検索結果を運用者の判断の補助とすることができるなどの点で有用で

ある。

今回は、類似運用の検索を次のような手順で試した。まず、訓練期間中のセンサデータおよび運用記録に対して第 2 節のモデルを学習した<sup>\*4</sup>。次に学習したモデルを用いて、テスト期間 (検索キューとなる期間) 中のセンサデータに関して潜在変数を推論した。テスト時には運用記録を用いていない点に留意されたい。類似運用検索には、日ごとの潜在変数  $\theta$  を用いた。具体的には、キューの  $\theta_{\text{queue}}$  と最も類似した  $\theta$  を訓練期間中から検索した。また、検索性能の評価のために運用記録を用いた。すなわち、検索結果 (キューと最も類似した日) の運用記録と正解の運用記録とを比較し、両者の記録に同じ軌道上実験が含まれていれば正解とした。

2013 年 1-6 月を訓練期間、同年 7-12 月をテスト期間として類似運用検索の実験を行ったところ、正解率は 66.3 [%] で

<sup>\*4</sup> よりチャレンジングな問題とするため、内部状態記録は用いていない

あった。図 4(a) および図 4(b) の上部のプロットに示すように軌道上実験は主に 4 種類あるから、十分高い正解率といえる。

## 4. まとめ

本稿では、宇宙機運用から得られるセンサデータ、内部状態記録、運用記録文書などの様々なデータを扱うためのモデルを提案し、実際の人工衛星のデータを用いてその有効性を確かめた。例えば、学習した潜在変数を用いたデータの可視化によって、極めて明白に異常な状態を検出することができた。これは既知の(かつ明白な)事象であったが、このようにデータ要約によって発見できる事象もありうる。本稿で提案したモデルを実際のデータにさらに適用して効果を検証していくほか、時系列性を考慮するようモデルを改良することなどが今後の課題である。

## 謝辞

宇宙航空研究開発機構・宇宙実証研究共同センターの中島祐太様、西村尚樹様には SDS-4 テレメトリデータの提供をいただくとともに、データ解析についての議論や示唆をいただきました。深く感謝いたします。

## 参考文献

- [Blei 03] Blei, D. M., Ng, A. Y., and Jordan, M. I.: Latent Dirichlet Allocation, *Journal of Machine Learning Research*, Vol. 3, pp. 993–1022 (2003)
- [Ghahramani 96] Ghahramani, Z. and Hinton, G. E.: The EM Algorithm for Mixtures of Factor Analyzers, Technical report, University of Toronto (1996)
- [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)
- [Tipping 99] Tipping, M. E. and Bishop, C. M.: Mixtures of Probabilistic Principal Component Analysers, *Neural Computation*, Vol. 11, No. 2, pp. 443–482 (1999)
- [矢入 14] 矢入健久, 武石直也, 櫻田麻由, 西村尚樹, 中島佑太, 関川知里, 井上浩一, 高田昇: SDS-4 運用における学習型テレメトリ監視システムの構築 (1) – アルゴリズムと実装, 第 58 回宇宙科学技術連合講演会 (2014)
- [西村 14] 西村尚樹, 中島佑太, 関川知里, 井上浩一, 高田昇, 矢入健久, 武石直也, 櫻田麻由: SDS-4 運用における学習型テレメトリ監視システムの構築 (2) – 運用システムによる検証と評価, 第 58 回宇宙科学技術連合講演会 (2014)