

## トピックモデルによる探査衛星の成果測定に関する一考察

日大生産工 ○高宗大起 日大生産工 水上祐治, 宇宙航空研究開発機構 大畠昭子

## 1 はじめに

宇宙開発プロジェクト等の大規模プロジェクトでは、莫大な予算とリソースを必要とするため、その成果が社会発展に貢献することが求められ、国会等で議論されている[1]。宇宙航空研究開発機構は、観測衛星、探査衛星から得られた分析データに関して、それら情報の利用促進を目指してオープンデータ化して公開している[2]。

本稿は、それら情報の学術界での活用度合いについて、テキストマイニング分析のトピックモデルを用いて分析を施して、その傾向について考察するものである。

## 2 先行研究と本稿

先行研究として中村ら[1]が挙げられる。中村らはX線探査衛星SUZAKUに関する論文の要旨に、テキストマイニング分析のコレスポンディング分析を施して、年代別の研究トレンドを簡易的に抽出する手法を提示した。図1に中村らの分析結果を示す。この研究により、SUZAKUの研究トレンドが、13年間の間にXIS、NGC、AGN、SNRに移行したことが示された。一方、中村らの手法では、各分析の詳細まで把握することができない制約があった。

本稿では、中村らのSUZAKU研究トレンドにおいて、前期(2007年、2008年)がXIS、後期(2017年、2018年)がSNR研究であることに着目、これらグループの詳細把握を目指して、Takei[3]

のトピックモデルによる分析フレームワークを適用した論文要旨分析とその考察を行うものとした。

## 3 研究方法

論文情報は、Clarivate Analytics社の論文データベースWeb of Scienceを用いて、X線探査衛星SUZAKUの論文収集を行った。

分析では、統計ソフトRを用いて論文要旨にトピックモデル分析を施した。トピックモデルとはテキストデータ中で扱われている話題を分析・推定する統計的手法であり、含まれる単語の出現頻度や、特定の語彙の結びつきから、文章の内容や分野を推定するものである[3]。

## 4 分析結果

図2の(a)は前期、(b)は後期のトピックモデル分析の結果である。分析には名詞、固有名詞を使用し、各トピックに特徴的な15単語を抽出した。前期は、各トピックにおいて、diskとhole、spectrometerとgrb、hessとgamma-ray、plasmaとngcである。後期は、各トピックにおいて、Nustarとhole、snrとplasma、halo、とChandraである。

## 5 考察

分析の結果、前期では、4つの話題が抽出され、その3つにXISがあり、先行研究が示した研究の特徴を示すことができた。トピック1は、hole、diskからブラックホールの研究、variability、energy、fluxからエネルギーの変動性と流れの研究であった。トピック2はgrb(ガンマ線バースト)やspectrometerから分光計を用いたガンマ線の研究であった。トピック3はhessやgamma-ray、supernovaからhess(ガンマ線天体物理学を研究する天文台)を用いた放射線や超新星の研究であった。トピック4は、plasma、ngc、temperature、fluxからプラズマの銀河温度、cyclotronからプラズマ中

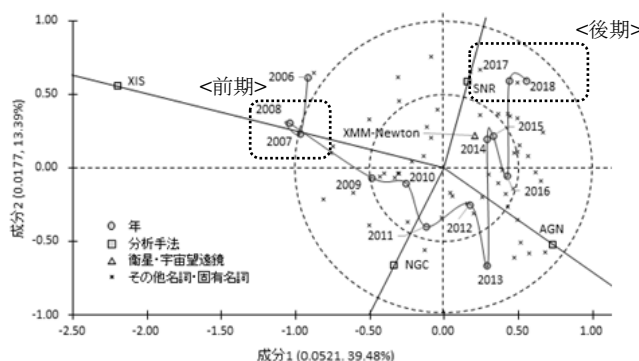
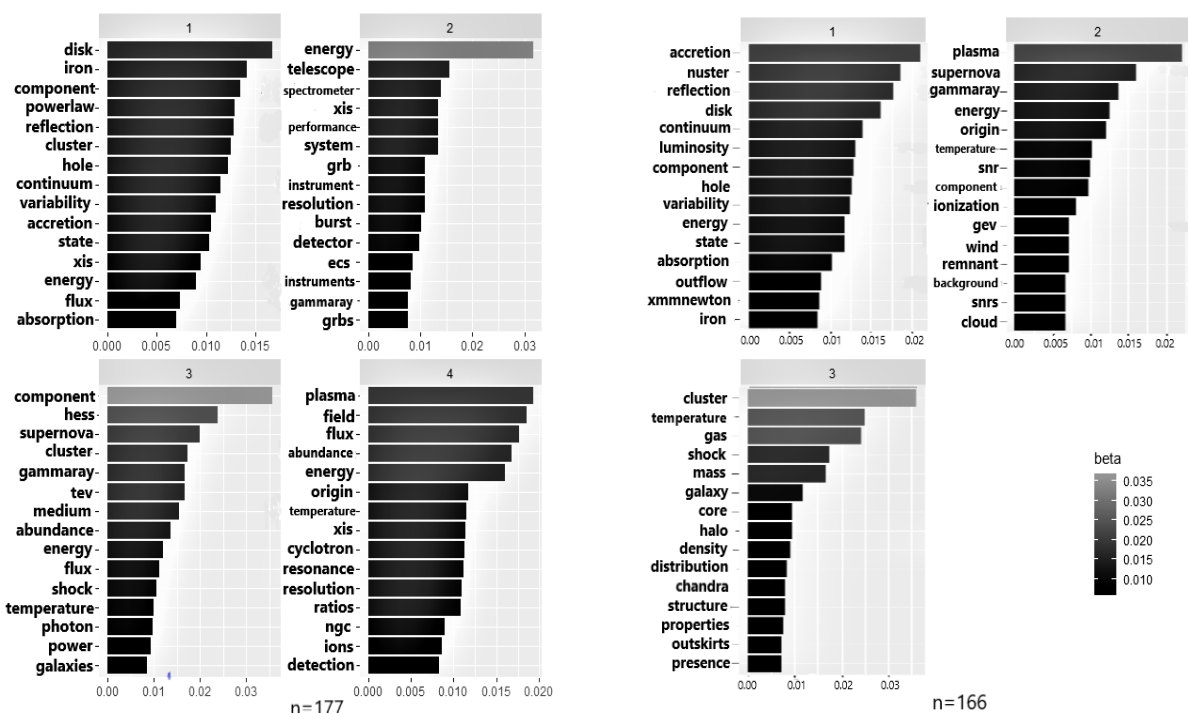


図1 中村らのコレスポンディング分析

## A Study on Measuring Results of Exploration Satellites Using Topic Models

Taiki Takamune Yuji Mizukami and Akiko Ohata



(a) 前期 (2007 年、2008 年) の分析

(b) 後期 (2017 年、2018 年) の分析

図 2 トピックモデル分析の結果

の電磁サイクロトロン波の研究であった。後期では、3つの話題が抽出され、その1つに snr 関連のトピックが検出された。トピック 1 は nuster, xmmnewton, hole からブラックホールの研究、nuster (NASA 宇宙望遠鏡)、xmmnewton (ESA の X 線観測衛星) の観測データを用いた論文群であった。トピック 2 は、snr, plasma, Wind から、超新星爆発後の snr (超新星残骸) やプラズマ、エネルギー、恒星風の論文であった。トピック 3 は、halo, Chandra, temperature, shock からハロー銀河の研究でハロー銀河内に存在する物質の衝突、ガス温度について、Chandra (NASA の人工衛星) で得た情報と共にまとめられていた。

## 6 結論

本稿では、先行研究として中村ら[1]を挙げ、その分析手法の限界とその解決方法を示した。

適用例として、X 線探査衛星 SUZAKU に関する論文の要旨分析を行い、中村らが年代別の研究トレンドの提示にとどまっていたことに対して、その分析の詳細を示すことができた。

本研究の制約として、トピック数の手動判別

がある。本稿では、単語群の総合的意味を分析者が判断、自らトピック数の指定し分析を行っている。今後の研究の方向性として、トピック数の判断アルゴリズム HDP (Hierarchical Dirichlet Process) を導入することがある。また、本稿が提示した分析手法の適用範囲を広めることがある。

## 参考文献

- [1] 中村匡佑、水上祐治、大畠 昭子、本多啓介、中野純司 「論文要旨のテキストマイニング分析による探査衛星の測定データを基にした研究動向の把握」
- [2] 衛星利用推進サイト, JAXA
- [3] Mio Takei, "Clustering of research subject based on stochastic block model", ISI 2019, 6 pages, Malaysia
- [4] 小林雄一郎 「R によるやさしいテキストマイニング」
- [5] X 線天文衛星「すざく」の状況について, JAXA