

# 衛星観測データのオープンデータ化とその意義に関する一考察

## — テキストマイニング分析による論文要旨分析 —

日大生産工 (学部) ○中村 匡佑 日大生産工 水上 祐治

### 1 はじめに

宇宙科学分野では、慣習として一定期間を経た情報は一般に公開される。技術革新や新規参入企業の増加等を背景に、宇宙由来の様々なデータの質の向上と量の増加が進んでおり、これらのデータはビッグデータの一部として様々な情報と組み合わせることで、幅広い分野で課題を解決することが期待されている。

本稿は、JAXA SUZAKU 衛星プロジェクトの観測データの公開化を題材として、研究の広がりや焦点を当て、オープンデータ化について考察を行うものである。本稿では、書籍データの論文要旨情報に対して分析を施し、オープンデータ化による研究動向の変化について考察を行う。

### 2 研究方法

論文情報は、Clarivate Analytics 社の論文データベース Web of Science を用いて収集した。

分析では、KH Coder を用いてテキストマイニング分析を行った。書籍データの論文要旨情報にテキストマイニングの対応分析を用いた。

対応分析は、一般的にコレスポンディング分析と言われ、クロス集計結果を2次元の散布図に展開する手法である。質的多変量分析の一種であり、数量化Ⅲ類と同等の分析が可能である。

分析には、名詞の抽出語と各論文の掲載年情報を外部の付加する情報 (外部情報) として用いる。

### 3 分析データ

SUZAKU に関する論文 2001 年～2018 年

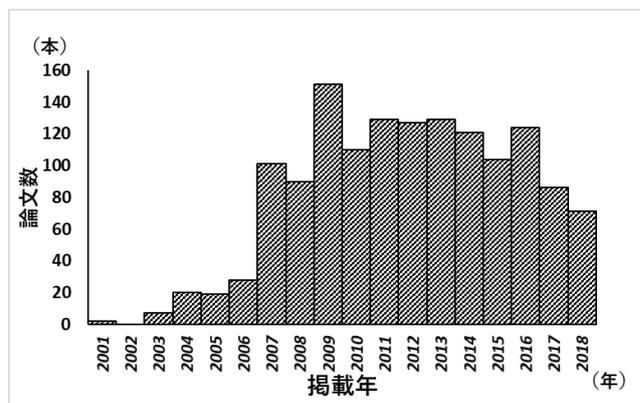


図1 論文数の年次変化

まで 1419 本を分析した。図 1 に論文数の年次変化の 3 つの特徴を示す。2007 年に急激に増えたこと、2009 年が最大であること、2011 年あたりからは減少傾向だったが、運用終了翌年の 2016 年に増えたことである。

### 4 分析結果

図 2 に論文要旨情報に対応分析を行った結果を示す。なお、2001 年や 2003 年など論文数が少なく結果が分かりにくいいため、2001 年から 2005 年を削除した対応分析を示す。散布図は、掲載年に着目した場合、原点 (0, 0) 付近の年は含まれる語の特徴がないこと、原点 (0, 0) から遠いほどその分布に特徴があることを示している。さらに、各年付近に位置する語は、それらの年に特徴的な語であることを示している。2006 年、2007 年といった年代は原点から離れており、2010 年あたりからは原点付近に位置していることがわかる。

### 5 考察

オープンデータ化の境界を特定するための 3 つの要素「掲載年別の傾向の変化」、「掲載

A Study on Open Data of Satellite Observation Data and Its Significance

— Analysis of abstracts by text mining analysis —

Kyosuke NAKAMURA and Yuji MIZUKAMI

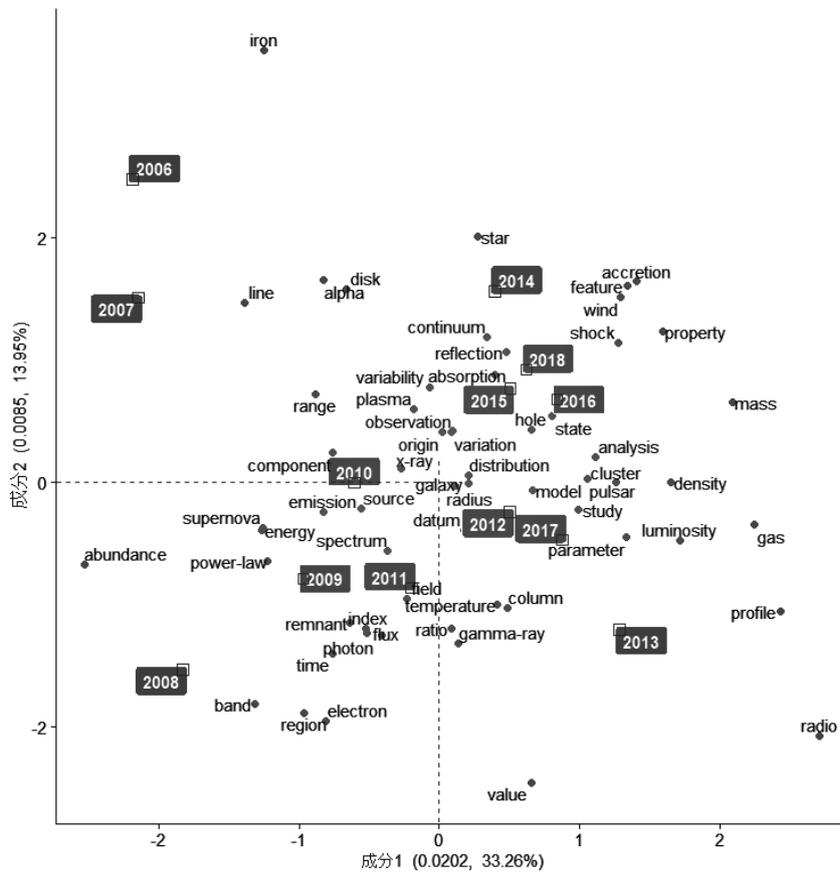


図2 論文要旨情報の対応分析

年と語の位置関係の変化」、「分析用データ入手から査読論文掲載までの期間」について述べる。

まず、「掲載年別の傾向の変化」について述べる。対応分析の結果は、原点に近いものはその分布に偏りがなく、原点から遠いものはその分布に偏りがあり特定分野に特化していることを示している。つまり、本稿の分析では、掲載年の位置に着目すると、その位置が原点に近いほど幅広い分野の研究で行われており、オープンデータ化後に情報が広く活用されたことが示されていると考えられる。本稿の分析結果では、2009年を境にして掲載年の位置が変化して原点に近い位置に集まる傾向が示されている。

次に「掲載年と語の位置関係の変化」について述べる。2009年以前をオープンデータ化前、2009年以降をオープンデータ化後のデータを使用した論文として、掲載年と各語の位置関係に着目する。オープンデータ化前は、掲載年の付近に位置する語が少ない。また、「形状」に関することなど衛星自体に関する

語が中心であると考えられる。そして、オープンデータ化後では、x-ray, galaxy を中心とした宇宙分野の X 線観測に関する論文と考えられる。

そして、「分析用データ入手から査読論文掲載までの期間」について述べる。分析用データを手入してから査読論文が掲載までの期間にはタイムラグが存在することを考慮する必要がある。つまり、分析用データを手入後、分析等の研究、執筆期間、投稿後の査読期間、掲載までの編集機関等があり、その間に数か月から数年を要することが

あり、また、その期間は論文ごとに異なる。

今後は分析する論文数を増やし、テキストマイニング分析の精度を高める、オープンデータ化についてよりの確に表すことが課題だと考える。

### 「参考文献」

- [1] 政府衛星データのオープン&フリー化 及び利用環境整備に関する検討会 報告書  
<http://www.meti.go.jp/press/2017/10/20171027001/20171027001-1.pdf>
- [2] 樋口耕一, 「社会調査のための計量テキスト分析—内容分析の継承と発展を目指して—」 ナカニシヤ出版, 2014
- [3] 水上 祐治, 平野未来, 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ホスピタリティ・マネジメント分野の研究傾向に関する一考察, 日本ホスピタリティ・マネジメント学会 第27回 全国大会予稿集, pp. 8-11, 2018