

# テキストマイニング分析による 研究動向把握の簡易的手法

## -衛星プロジェクトの関連論文の要旨分析-

水上 祐治\*（日本大学）・ 中村 匡佑（日本大学）

家富 紗妃（日本大学）・ 大畠 昭子（JAXA）・ 中野 純司（統計数理研究所）



大学共同利用機関法人 情報・システム研究機構  
統計数理研究所  
The Institute of Statistical Mathematics

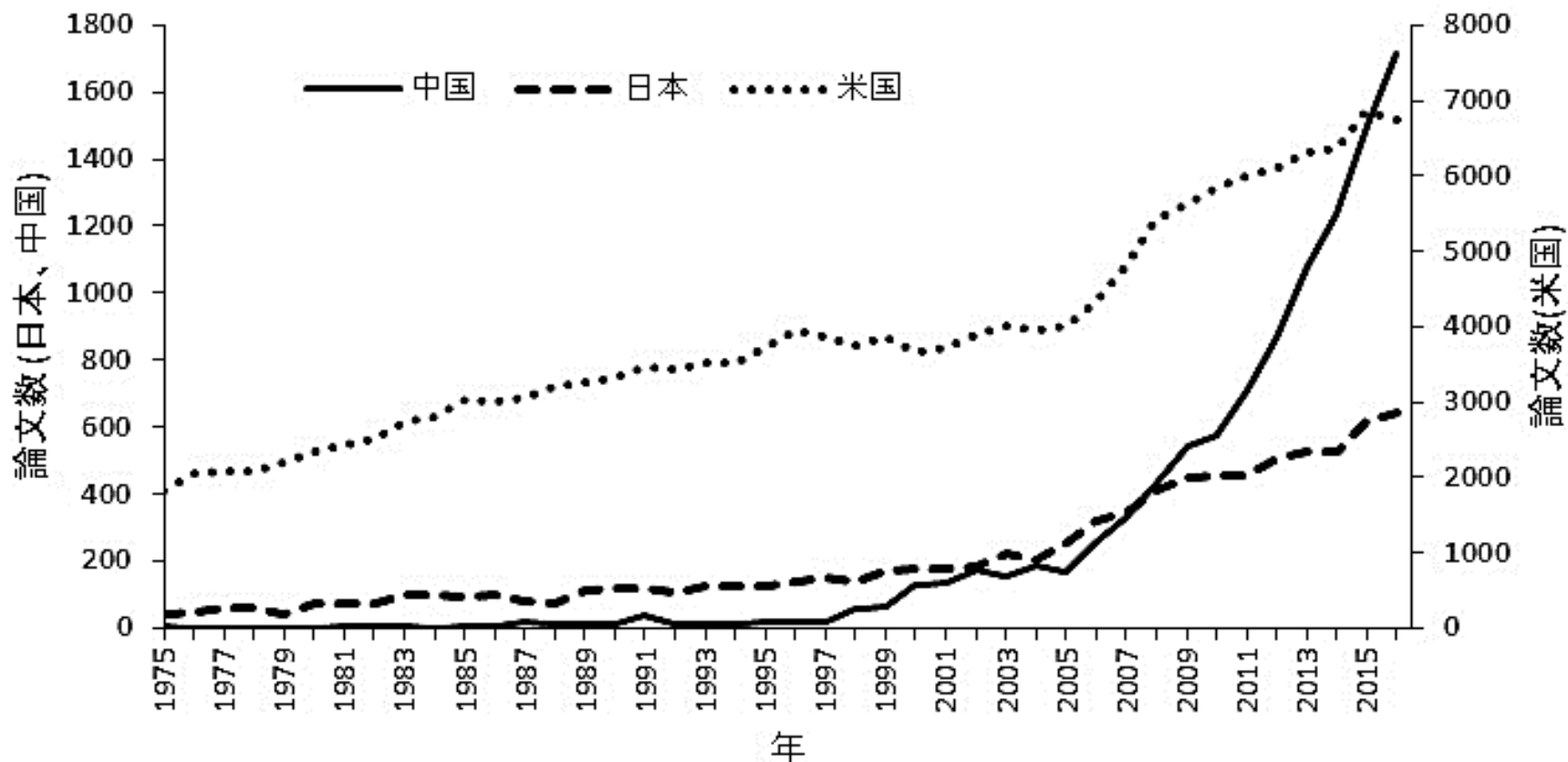


# まえおき

本稿は、研究IR（Institutional Research）に関する研究です。

研究者の**実験や研究の成果を「見える化」**したり、さらに活用するための戦略を立案するための情報を提供する活動です。

# まえおき： 研究IR活動の例



経済学分野の論文数の国際比較 (WoSを使用)

# はじめに (1)

人工衛星・探査機から得られる様々な観測データは**広く活用が期待されている**。

一方、宇宙開発プロジェクトは大規模な事業であるため、その**学術的貢献**については、国会をはじめあらゆる場面で**説明が求められている**。

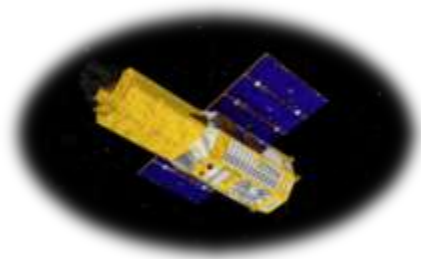
## はじめに (2)

本稿では、観測データを用いた論文に着目、その量や質をもとにして宇宙開発プロジェクトの**成果の見える化**を促進、学術的貢献の補足的な**説明資料となる客観的情報を用意**することを目標としている。

## はじめに (3)

本稿の分析では、宇宙開発プロジェクトの論文を収集、それら論文の要旨にテキストマイニング分析を施して特徴を抽出する。次に各論文の特徴を集め、宇宙開発プロジェクトの研究の広がりと方向性を示すことを目的としている。

# 分析対象



天文衛星  
X線天文衛星「SUZAKU」

打上げロケットの不具合によっ  
て軌道投入できなかった  
ASTRO-Eの再挑戦をかけたX線

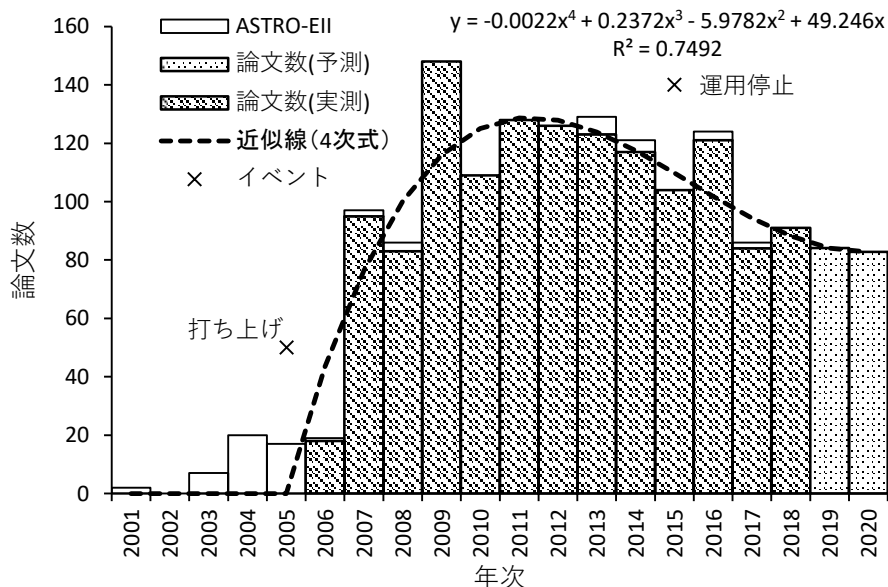
X線天文衛星SUZAKUプロジェクトの探査  
データを用いた論文を分析対象とした。

- 宇宙論的な遠距離にある天体のX線観測、  
宇宙の高温プラズマのX線分光観測
- 2005年に打ち上げ、2015年に観測終了

# 分析データ

WoS 検索条件 : (TS=(SUZAKU) OR TS=("ASTRO-E II") OR TS=("ASTRO-E 2") OR TS=("ASTROE II") OR TS=("ASTROE 2") OR TS=("ASTROEII") OR TS=("ASTROE2") OR TS=("ASTRO-EII") OR TS=("ASTRO-E2")) AND 言語: (English) AND ドキュメントタイプ: (Article OR Review)  
索引=SCI-EXPANDED, SSCI, A&HCI, ESCI タイムスパン=全範囲

論文情報は、書誌データベースの Web of Science Core Collection を使用する。



SUZAKU関連論文の年別変化

SUZAKU 関連論文  
は、2001年-2019  
年の18年間で**1419**  
本であった。



# 分析手法（データ抽出）

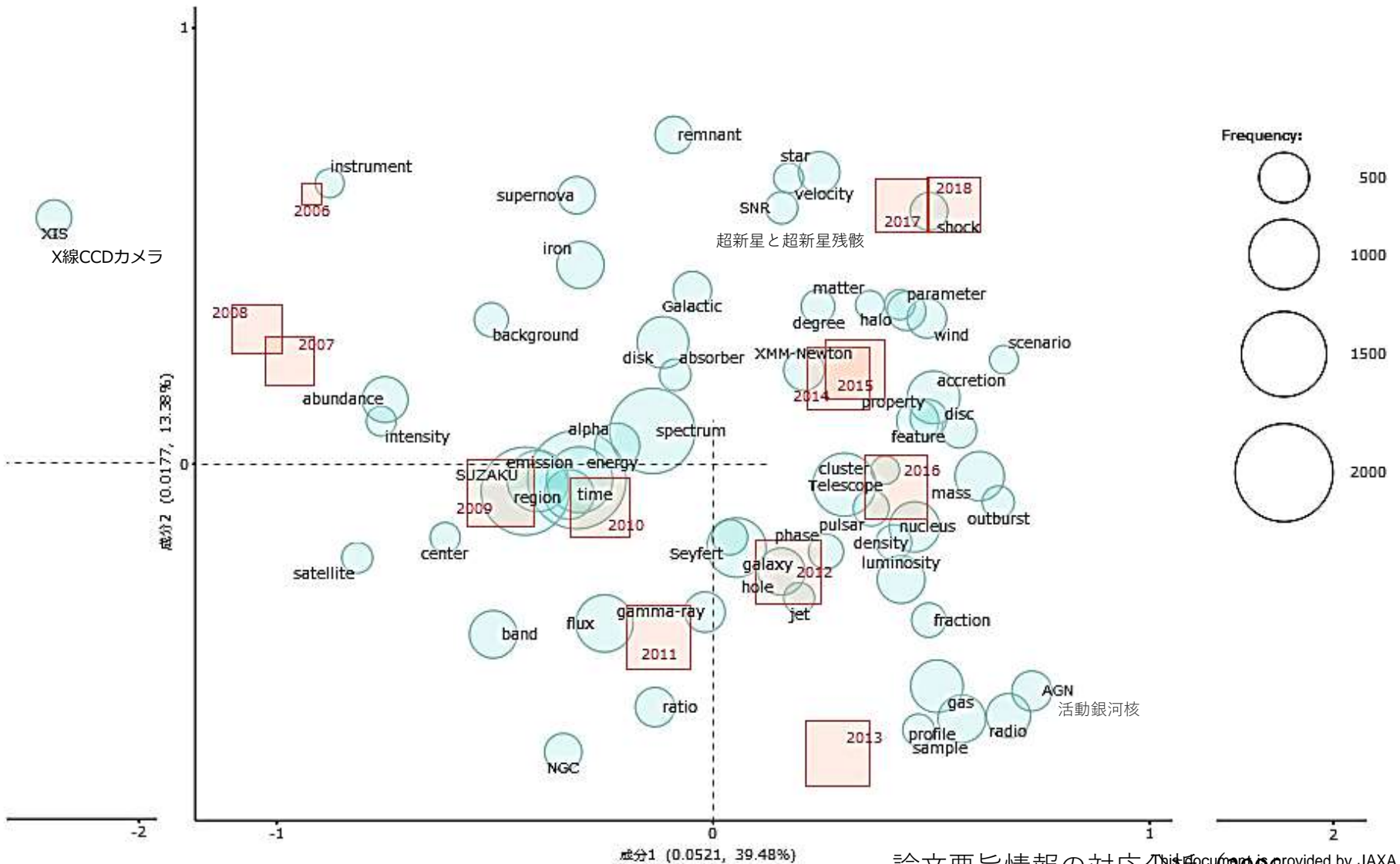
テキストマイニングを用いる。テキストマイニングでは、テキストデータに対して、自然言語解析の手法を使って、文章を単語（名詞、動詞、形容詞等）に分割し、それらの出現頻度や相関関係进行分析することで有益な情報を抽出する手法である。



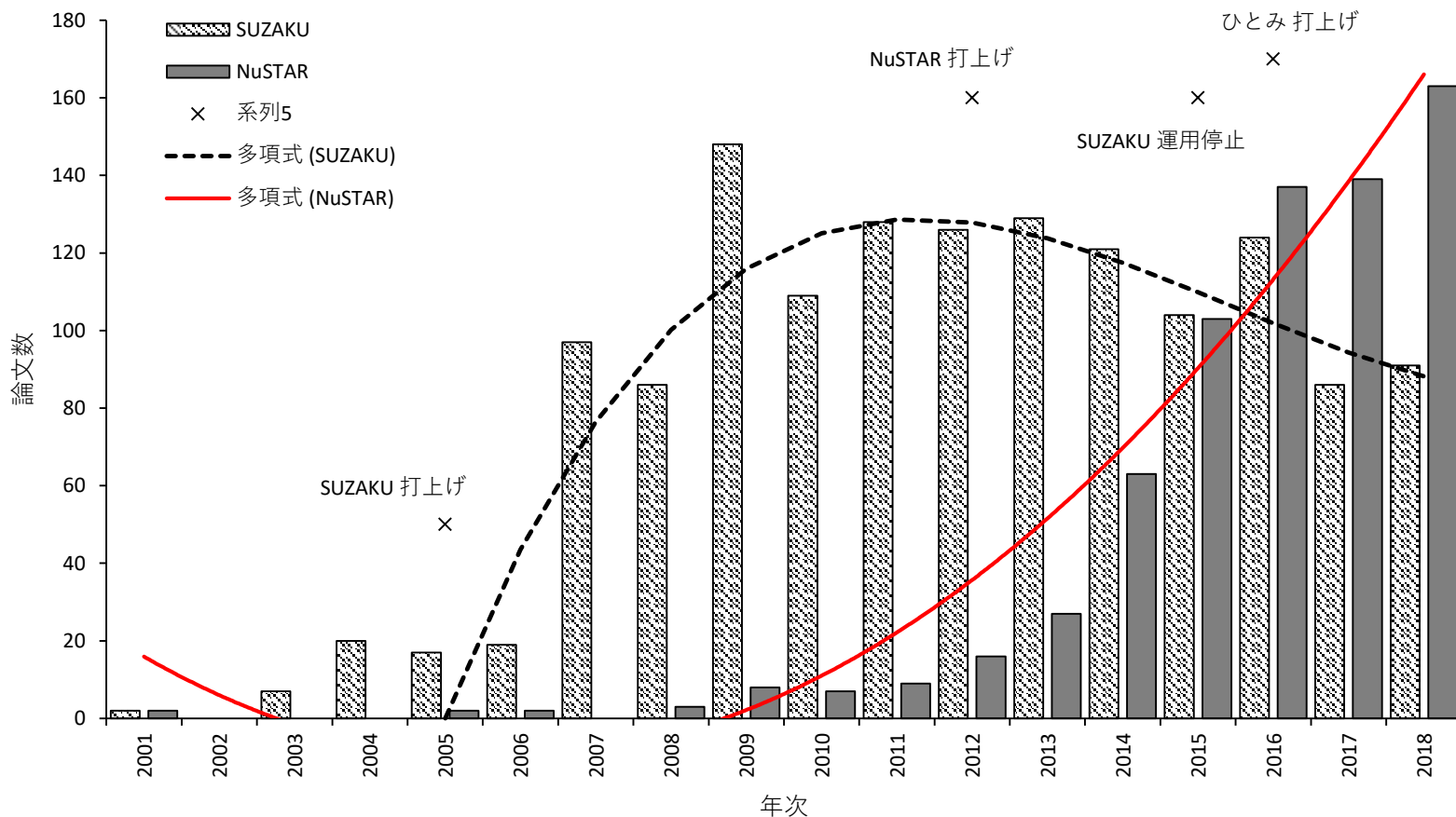
# 分析手法（統計分析）

テキストマイニングで得られた単語情報に対して**対応分析**を施す。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。質的多変量解析の一種であり、数量化Ⅲ類と同等の分析が可能である。

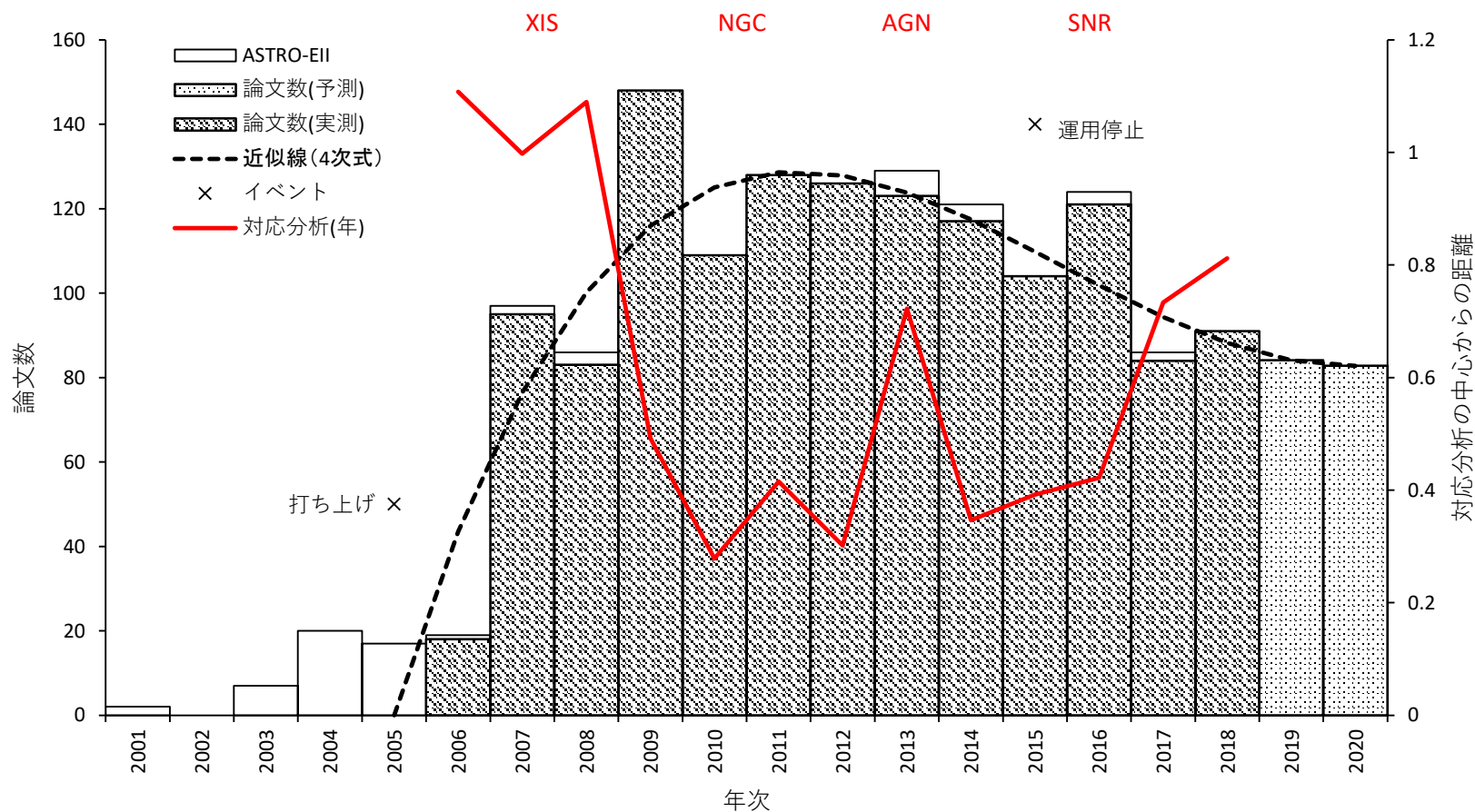
# 分析結果 (対応分析)



# 考察 (1)



# 考察 (2)



掲載年における論文数と対応分析の原点からの距離

# まとめ（1）

- 人工衛星・探査機の測定データの活用度を論文数とその特徴で表すことを目指しテキストマイニング分析の適用を試行した。
- X線天文衛星SUZAKUプロジェクトの情報を用いた。
- SUZAKU論文は、年代別にXIS,NGC,AGN,SNRにその特徴が移行している。

## まとめ（２）

- オープンデータの活用度が低いことがある。
- 2013年にAGN関連論文が特徴的に多く発表されている。
- SUZAKU論文の減少と共にNuSTAR論文が増加している。



## まとめ（3） 今後の方向性

- オープンデータの活用度を高めるにはどうすればよいのか検討する。
- 2013年にAGN関連論文が特徴的に多く発表されているが、その傾向が一過性であった理由を調査する。

## まとめ（４） 今後の方向性

- 本稿が適用したテキストマイニング分析は、分析が容易であり、**分析時間が短い特徴**がある。手法の精度向上などの深化が必要と考えるが、分析時間が短い特徴を活かして、月報用簡易分析など、その活用方法を模索することがある。

# 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ー内容分析の継承と発展を目指してー」 ナカニシヤ出版、2014

[3] 中村匡佑、水上祐治、大畠昭子、「JAXA SUZAKU衛星プロジェクトのオープンデータ化に伴う研究動向の変化ーテキストマイニング分析による論文要旨分析ー」、日本経営システム学会第61回全国大会、2018

[4] 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

[5] 平野未来、水上祐治、永井博昭、「論文要旨のテキストマイニング分析による研究トレンドの年次変化に関する一考察 - 大学等研究組織の研究力に関する組織比較研究 -」、第61回 日本経営システム学会 全国研究発表大会講演論文集、pp.256-257