Preliminary Study on Literature-Based Classification of Space Missions

TSUNODA Takehiro[*1], DEMURA Hirohide[*1], HASHIMOTO Yasuhiro[*1], OKUDAIRA Kyoko[*1]

ABSTRACT

This research explores methods to enhance the long-term value of space missions and projects, particularly beyond their operational phases. Grounded in a review of existing literature, the study introduces a novel approach to the missions by analyzing their publication histories, with a specific emphasis on the temporal pattern of publications. This vectorization process is designed to quantify the influence and progression of space missions over time. While the current research focuses on this temporal aspect, future work aims to incorporate a graph-based citation network structure. This more complex model will examine the detailed interconnections among publications through their citation links, potentially revealing deeper insights into the development and impact of space missions. The present study, therefore, serves as a preliminary exploration, setting the stage for further advanced analyses that will investigate the complex network of citations to understand the broader impact and evolution of space missions.

Keywords: Space Missions, Data Science, Project Management, Digital Object Identifier

1. Introduction

The advancement of large-scale space development projects, such as NASA's Apollo Program [1], underscores the critical need for effective project management. Recognized frameworks like the Project Management Body of Knowledge (PMBOK) [2] have been developed to guide project initiation, management, and completion. This research extends these discussions by focusing on strategies for enhancing the long-term value of space projects, even after their operational phases have ended.

To measure a project's value, we propose a “ripple effect” metric, measured by both the number and interconnectedness of academic papers generated from the inception of a project to the present. This study represents a preliminary step for future work, which aims to identify and recommend [3,4,5] untapped areas in space missions by clustering missions using a graph-

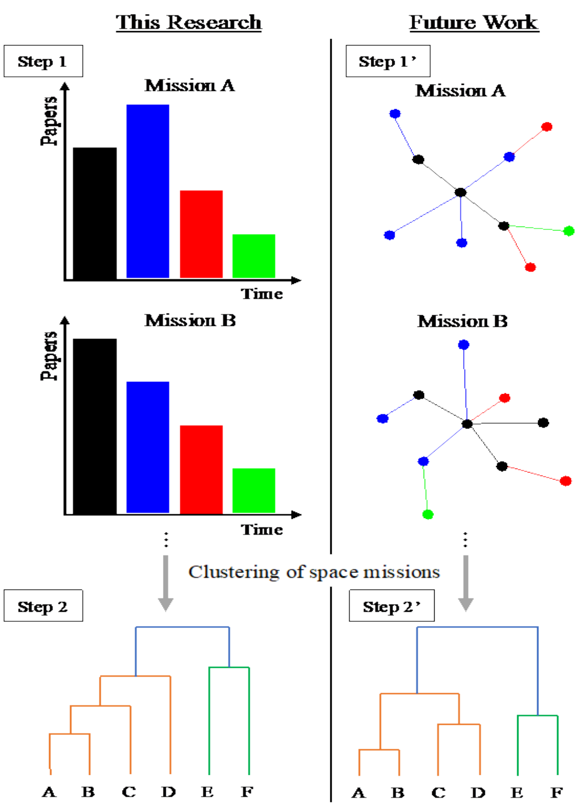

Fig. 1. Clustering analysis of space missions

* 2023 年 11 月 30 日受付 (Received November 30, 2023)
*1 会津大学 (The University of Aizu)

based citation network structure.

Figure 1 delineates the progression from the current study (on the left) to the anticipated future work (on the right), illustrating our methodological evolution in classifying space missions to pinpoint unexplored opportunities within the field. The present research lays the groundwork by clustering missions based on the temporal sequence of their scholarly output. The initial step, Step 1, presents a bar graph for each mission, where the bars' height represents the quantity of published papers over time, segmented into phases from the mission's start to the current period, each distinguished by a unique color—black, blue, red, and green—indicating distinct operational phases. The subsequent step, Step 2, involves hierarchical clustering to identify groups of similar missions based on their publication patterns. Future work intends to augment this approach by analyzing space missions through the lens of their citation networks. This expanded analysis will visualize the connections between publications, denoted by color-coded periods akin to the first graph, Step 1'. This addition will allow for the incorporation of citation counts and nuanced paper characteristics into the clustering process, thereby offering a more detailed comparison between different missions' influences.

This research has two primary goals: First, to define a characterization vector for space missions based on existing literature; and second, to classify space missions for future identification.

Table 1. Mission names and associated research fields

Mission Name	Field in Space Science
Akari[9]	Infrared astronomy
Akatsuki[10]	Planetary atmosphere
HALCA[11]	Radio wave astronomy
Hayabusa[12]	Primitive asteroid sample return
Hiten[13]	Orbit engineering
Kaguya[14]	Lunar science, interplanetary science
Sakigake / Suisei[15]	Cometary science, interplanetary science
Suzaku[16]	X-ray astronomy
Tanpopo[17]	Astrobiology
Yohkoh[18]	Solar science

2. Method

2.1. Data Source

This study focuses on selected space missions listed in Table 1, chosen for their representative roles in Japanese space science. These missions span a range of research fields, including infrared astronomy, planetary atmosphere, radio wave astronomy, primitive asteroid sample return, orbit engineering, lunar science and interplanetary science, cometary science and interplanetary science, X-ray astronomy, astrobiology, and solar science.

We have compiled a literature dataset presented in Table 2, sourced from the Astrophysics Data System (ADS) [6], IEEE Xplore® [7], and ScienceDirect® [8]. Papers were extracted from these databases by including mission names in the title or the abstract. This selection method indirectly considers the citation counts of initial

Table 2. Space Mission Dataset

DOI	Title	Year	Journal Name	Field
10.1016/j.icarus.2022.115370	Sample studies and SELENE (Kaguya) …	2023	*Icarus*	S
10.1109/JDT.2010.2052452	Conversion Method From Moving Pictures …	2010	*Journal of Display Technology*	E
10.1038/nature08317	The global distribution of pure anorthosite …	2009	*Nature*	H

result papers, thus facilitating the identification of papers related to space missions. The dataset comprises five key attributes: DOI and Title are used for identifying individual papers; Publication Year measures the timing of each publication; and Journal Name and Journal Field are employed for classifying the papers; The Journal Field is categorized into two primary fields and one cross-field based on Journal Name. Two primary fields are 'Engineering,' which predominantly publishes papers related to engineering, and 'Scientific,' which focuses on papers from various scientific disciplines. However, this classification encounters difficulty when categorizing journals that publish papers in both engineering and scientific fields, such as 'Nature' and 'Science.' To address this issue, we introduce the 'High Impact Factor' class, as these journals typically exhibit higher impact factors compared to other journals. These categories are combined with Journal Names to address the issue of sparse data distribution.

2.2. Visualizing Space Mission Features

This research categorizes each mission into distinct phases. Phase 0 extends from the start of the mission to the satellite launch. Phase 1 spans from the satellite launch to the timing of initial reports plus an additional two years. Phase 2 starts from this point and lasts for the same duration as Phase 1. Subsequent phases are defined in the same manner.

These phases, along with three categories—*High Impact Factor*, *Engineering*, and *Scientific*—are visualized in bar plots, as exemplified by the HALCA mission shown in Fig. 2. The horizontal axis represents the summation of all phases as well as individual phases. The vertical axis displays the ratio of papers published in all phases and the percentage of the total number of papers in each phase. Dotted lines indicate the end of operations, while solid lines represent the current time. Bars are colored green, red, and blue to indicate *High Impact Factor*, *Engineering*, and *Scientific* categories, respectively.

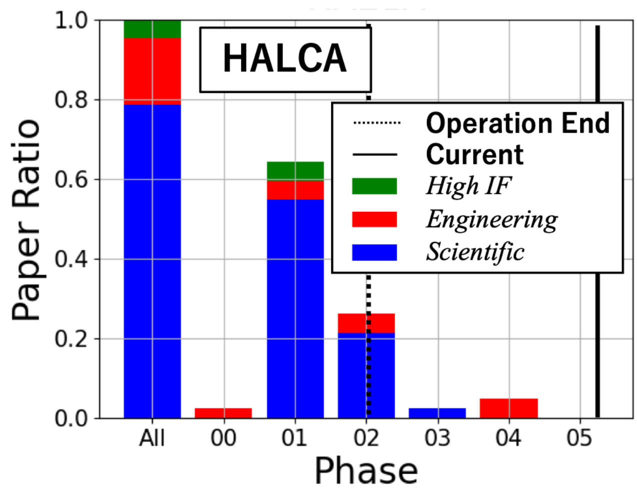

Fig. 2. Visualization of HALCA

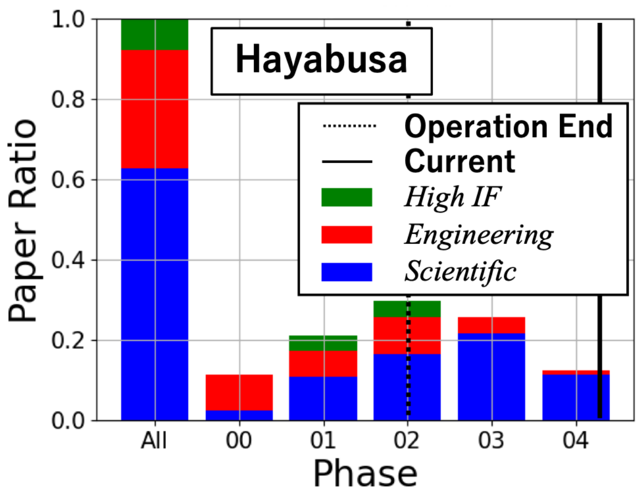

Fig. 3. Visualization of Hayabusa

2.3. Definition of Space Mission Vector

In this section, we introduce a methodology to represent space missions as vectors based on their scholarly output, which serves as an initial phase towards adopting a graph-based citation network for more complex clustering analyses in the future.

Our method categorizes space missions based on the timing of their research outputs. The first category includes missions that have the bulk of their publications between the launch (Phase 0) and up to Phase 2, indicative of early output. The second category consists of missions that continue to generate significant

Table 3. Space Mission Vector

Mission Name	$n \geq 2$	$\sum_{i=0}^{l-2} P(2+i) \geq 0.5$	$\sum_{i=0}^{1} P(l-i) \geq 0.1$	All_H	All_E	All_S	…	P03_H	P03_E	P03_S
Akari	1	1	1	0.0742	0.00606	0.92	…	0.0106	0.00152	0.185
Akatsuki	1	1	1	0.161	0.119	0.72	…	0.0424	0.0593	0.195
HALCA	0	0	0	0.167	0.0476	0.786	…	0	0	0.0238
Hayabusa	1	1	1	0.296	0.0788	0.626	…	0.0394	0	0.217
Hiten	1	1	1	0.625	0	0.375	…	0.0312	0	0.0625
Kaguya	0	1	1	0.248	0.0254	0.727	…	0.0212	0.00212	0.127
Sakigake/ Suisei	0	1	0	0.0769	0.135	0.788	…	0	0	0.0192
Suzaku	1	1	1	0.0336	0.00468	0.962	…	0.00781	0.000781	0.23
Tanpopo	0	0	1	0.217	0	0.783	…	0	0	0
Yohkoh	1	1	0	0.0249	0.00277	0.972	…	0.00461	0	0.291

research outputs well into Phase 3 and beyond, suggesting a late surge in scholarly interest. For our analysis, we consider the 'active lifetime' of a mission to typically conclude around Phase 2, the point by which primary objectives are often achieved, while also acknowledging that publications can continue to accrue long after.

The 'space mission vector' contains 18 elements that quantitatively capture various aspects of a mission's publication record. The initial three elements assess the temporal distribution of publications, represented as binary (0 or 1) values based on observed patterns:

- **Full bloom** ($n \geq 2$): This indicates whether the mission's peak publishing activity occurs no earlier than phase 2. It is set to 1 if the peak is in phase 2 or beyond.
- **Long bloom** ($\sum_{i=0}^{l-2} P(2+i) \geq 0.5$): This checks whether over half of the mission's publications are released from phase 2 onwards. A 1 is assigned if this criterion is met.
- **Late bloom** ($\sum_{i=0}^{1} P(l-i) \geq 0.1$): This determines if the final phases account for more than 10% of the total publications. A 1 is assigned for a significant late-stage output.

Here, 'n' denotes the phase with the highest publication count (peak phase), 'l' the latest phase, and '$P(i)$' the number of publications in phase 'i'.

The remaining 15 elements of the vector provide a phase-by-phase ratio of publication types, capturing the impact and focus of the research output across High Impact Factor, Engineering, and Scientific categories:

- **All_H / All_E / All_S**: These elements reflect the overall ratio of High Impact Factor, Engineering, and Scientific publications across all phases.
- **P00_H / P00_E / P00_S to P03_H / P03_E / P03_S**: Each set of these elements corresponds to the ratios within specific phases, from the launch (phase 0) up to phase 3.

This detailed vectorization allows for the comparison and assessment of missions with varied durations and research outputs, such as Sakigake/Suisei's extensive 14-phase timeline versus Tanpopo's shorter 2-phase lifecycle.

3. Results

3.1. Identification of Space Mission Types

This study identifies two primary types of space missions based on the distribution of academic papers published at different phases of the mission lifecycle. These are:

Short-Run Type (SR): Missions like HALCA show a high paper output ratio during Phase 0 and Phase 1 (see Fig. 2).

Long-Run Type (LR): Missions like Hayabusa exhibit a high paper output ratio from Phase 2 onward (see Fig. 3).

3.2. Hierarchical Clustering Results

Figures 4 and 5 display the outcomes of a hierarchical clustering analysis. We applied the Ward method and Euclidean distance metrics to analyze space mission vectors, which encompassed data on the bar graph's shape and publication ratios for each phase. The clustering analysis successfully separates LR missions from SR missions at a distance of 1.7.

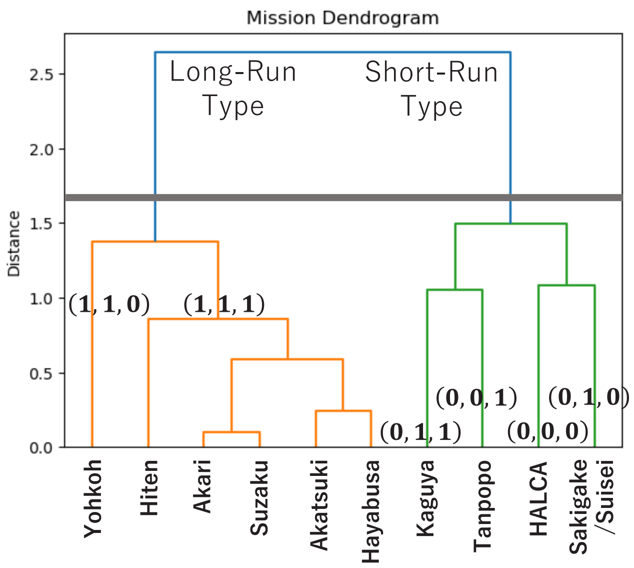

Fig. 4. Hierarchical Clustering of Space Mission Vector

4. Discussion

This study discovered two types of space missions: SR and LR types. These categories are based on the timing of academic publications related to the missions. SR missions are those that generate most of their papers before the initial report is published, while LR missions do so afterward.

We use these classifications to categorize missions into three distinct patterns: Full Bloom, Long Bloom, and Late Bloom, as shown in Table 3's initial three elements. A mission with more zeros in its vector indicates an SR tendency, while more ones suggest an LR tendency. The subsequent 15 elements depict the ratio of papers published during all phases, along with a percentage breakdown from Phase 0 to Phase 3.

Figure 4 and 5 indicate future work focused on identifying and recommending untapped areas within the realm of space missions. For example, in Figure 4, Akari, Suzaku, and Yohkoh are classified as LR type with a distance of 1.7. However, within each cluster, both Akari and Suzaku have a cluster distance of 0.2, indicating their higher similarity compared to Yohkoh within the same cluster. This research introduces a method of target comparison through hierarchical clustering. In our future work, we will define graph-based citation network structures and employ hierarchical clustering to achieve more detailed mission classification.

5. Conclusion

In conclusion, this study has successfully met its two primary objectives. First, based on existing literature, we have defined a characterization vector for space missions. This vector serves as a multidimensional representation that encapsulates the diverse impacts of various missions, offering valuable insights into their strengths and weaknesses.

Second, we have applied this characterization vector to classify space missions. Our classification framework employs Short and Long-Run categories to account for the timing and volume of academic publications generated by these missions. This has revealed the process of evaluating space missions through vectorization. It has also enabled us to lay the

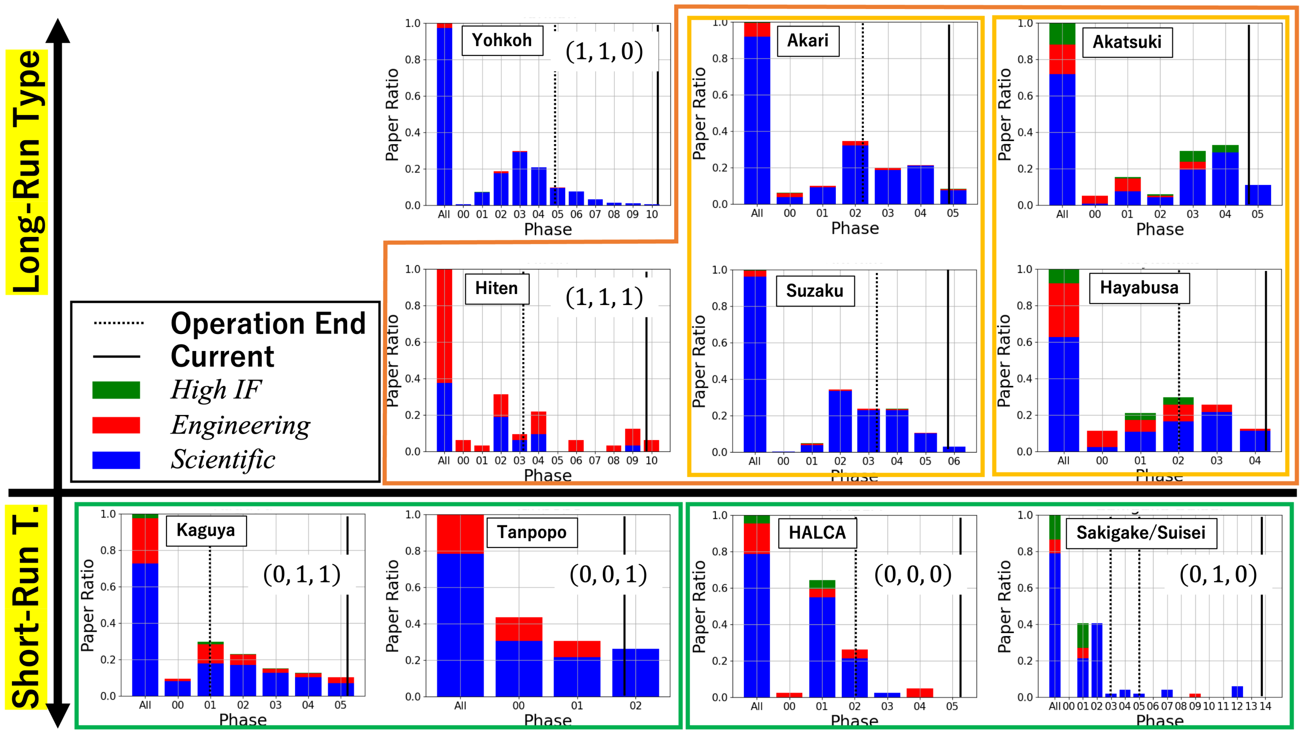

Fig. 5. Result of hierarchical clustering

foundation for evaluating space missions using citation network structures in future research.

This study represents a step toward understanding how to measure and enhance the long-term value of space missions. It also provides a foundation for developing recommendation systems that have the potential to inform decision-making in space research and development.

However, it's worth noting that this research has some limitations, including the need for a more extensive dataset for validation and the challenges associated with categorizing highly interdisciplinary or evolving missions.

Future research will build on the insights gained in this study by integrating a graph-based clustering approach to analyze citation networks, aiming to uncover patterns and insights within space mission literature. Advanced machine learning techniques will play a pivotal role in the classification and analysis of missions, taking into account a multitude of variables such as budgetary allocations, technical complexities, and broader societal impacts. This multi-faceted approach is expected to yield a robust framework for assessing and understanding the full spectrum of influences on development of space missions.

ACKNOWLEDGEMNETS

We would like to thank Prof. Hirata, N., Prof. Honda, C., and Prof. Ohtake, M. for their fruitful discussions. All contributions were made in affiliation with The University of Aizu.

REFERENCES

[1] “Apollo”, https://www.nasa.gov/mission_pages/apollo/index.html (Feb. 17th, 2023 retrieved)

[2] PMBOK 7th edition (2021)

[3] D. Jannach, *et al.*, “Recommender Systems: An Introduction” Cambridge University Press (2010)

[4] Deepak K. Agarwal and Bee-Chung Chen, “Statistical Methods for Recommender Systems” Cambridge University Press (2016)

[5] M. Schrage, “RECOMMENDATION ENGINES” The MIT Press, (2020)

[6] "astrophysics data system", https://ui.adsabs.harvard.edu/ (Feb. 17th, 2023 retrieved)

[7] "IEEE Xplore®", https://ieeexplore.ieee.org/Xplore/home.jsp (Feb. 17th, 2023 retrieved)

[8] "ScienceDirect®", http://www.sciencedirect.com/ (Feb. 17th, 2023 retrieved)

[9] D. Ishihara, *et al.*, "AKARI Infrared Imaging of Reflection Nebulae IC4954 and IC4955" Publications of the Astronomical Society of Japan, vol. 59, issue sp2 (2007), pp. S443-S454, doi: 10.1093/pasj/59.sp2.S443

[10] M. Nakamura, *et al*., "Overview of Venus orbiter, Akatsuki" Earth, Planets and Space, vol. 63 (2011), pp. 443-457, doi: 10.5047/eps.2011.02.009

[11] H. Hirabayashi, *et al.*, "Overview and Initial Results of the Very Long Baseline Interferometry Space Observatory Programme" Science, vol. 281, issue 5384 (1998), pp. 1825-1829, doi: 10.1126/science.281.5384.1825

[12] A. Fujiwara, *et al.*, "The Rubble-Pile Asteroid Itokawa as Observed by Hayabusa" Science, vol. 312, issue 5778 (2006), pp. 1330-1334, doi: 10.1126/science.1125841

[13] K. Uesugi, *et al.*, "Japanese first double lunar swingby mission "Hiten"" Institute of Space and Astronautical Science, vol. 25, issue 7 (1991), pp. 347-355, doi: 10.1016/0094-5765(91)90014-V

[14] J. Haruyama, *et al.*, "Lack of Exposed Ice Inside Lunar South Pole Shackleton Crater" Science, vol. 322, issue 5903 (2008), pp. 938-939, doi: 10.1126/science.1164020

[15] K. Hirao and T. Itoh, "Project overview and highlights of Suisei and Sakigake" Advances in Space Research, vol. 5, issue 12 (1985), pp. 55-64, doi: 10.1016/0273-1177(85)90067-5

[16] K. Mitsuda, *et al.*, "The X-Ray Observatory Suzaku" Publications of the Astronomical Society of Japan, vol. 59, issue sp1 (2007), pp. S1-S7, doi: 10.1093/pasj/59.sp1.S1

[17] A. Yamagishi, *et al.*, "Environmental Data and Survival Data of Deinococcus aetherius from the Exposure Facility of the Japan Experimental Module of the International Space Station Obtained by the Tanpopo Mission" Astrobiology, vol. 18, issue 11 (2018), pp. 1369-1374, doi: 10.1089/ast.2017.1751

[18] Y. Ogawara *et al.*, "The Status of YOHKOH in Orbit: an Introduction to the Initial Scientific Results" Publications of the Astronomical Society of Japan, vol. 44, no. 5 (1992), pp. L41-L44