

NSF データを用いた研究課題動向分析の試み

文部科学省 科学技術・学術政策研究所 データ解析政策研究室

要旨

本稿では、我が国における今後の研究開発戦略の在り方についての検討に資することを念頭に、米国の公的資金助成による研究プロジェクトレベルの課題情報、具体的には NSF (National Science Foundation) の課題データを分析した結果と、我が国における代表的な研究に関する公的資金助成である科学研究費助成事業（学術研究助成基金助成金／科学研究費補助金）の課題データを併せて分析した結果を紹介する。

従来、著者らは米国の研究課題動向を分析する上で、米国政府が提供する Federal RePORTER と呼ばれるサービスを利用してきた。Federal RePORTER は NIH (National Institute of Health) や NASA (National Aeronautics and Space Administration) など、複数機関の助成研究課題を取りまとめ、年ごとや機関ごとといった単位で件数や金額の推移を可視化する機能や、生データそのものを提供・公開するシステムで、例えば深層学習 (Deep Learning) の語を含む研究課題が、どの機関からどの程度出ているか、それらの年ごとの推移はどのようなものか、といったことを容易に調査することができた。

しかしながら、Federal RePORTER は 2022 年 3 月に運用を停止し、アーカイブサイトを用いて過去公開データを提供するのみの状態に移行した。これにより、米国の研究課題動向を分析するためのデータソースなどを検討する必要が生じた。そこで grants.gov および、NIH、NSF を候補に選定し、grants.gov と NSF については具体のデータ分析まで行った。

結果、今回の候補の中では NSF のデータを分析することが目的とコストに照らして適当であることがわかり、バイオ、量子科学、人工知能などの分野については、政策的な動向および、それらに現れない動向を観察することができた。

さらに、KAKEN のデータ（一部）を機械翻訳して比較し、NSF の研究課題と合致・相違する点も調査し、材料系などは NSF よりも件数が多い一方、情報系は少ないなどの傾向を観察できた。

Attempts to analyse research issue trends using NSF data

Research-Unit for Data Application, National Institute of Science and Technology Policy (NISTEP), MEXT

ABSTRACT

This paper describes the results of an analysis of project-level issue information on publicly funded research projects in the US, with a view to contributing to the consideration of future R&D strategies in Japan. Specifically, the results of the analysis of the National Science Foundation (NSF) proposal data and the results of the analysis of the “KAKEN” proposal data, which is a representative publicly funded research grant in Japan, are presented.

In the past, the authors have used a service called “Federal RePORTER” provided by the US government to analyse research proposal trends in the US. “Federal RePORTER” compiles funded research proposals from multiple organisations, such as the National Institute of Health (NIH) and the National Aeronautics and Space Administration (NASA), and visualises the number of proposals and the amount of money spent on them on a yearly or agency-by-agency basis. It is a system that provides and discloses the raw data itself, as well as functions for visualising changes in the number of cases and amounts of funding by year and by organisation. Using “Federal RePORTER”, it was easy to investigate, for example, how many research proposals containing the term “Deep Learning” were issued by which institutions and what their annual trends were like. However, the “Federal RePORTER” ceased operation in March 2022 and moved to a state where it only provides historical public data using an archive site. This made it necessary to consider data sources for analysing research agenda trends in the US.

Therefore, grants.gov, NIH and NSF were selected as candidates, and for grants.gov and NSF, specific data analysis was conducted.

As a result, it was found that analysing the NSF data was appropriate for the purpose and cost among the candidates, and policy trends and trends that did not appear in the fields of biotechnology, quantum science, artificial intelligence, etc. could be observed.

Furthermore, we machine-translated some of the data from “KAKEN” and investigated the differences and matches with the NSF research agenda, and were able to observe trends such as a higher number of cases for materials than the NSF, but fewer cases for information science.