

ほらいずん

## 科学技術基本計画のテキスト分析

第2研究グループ 研究員 川島 浩誉

科学技術・学術政策研究所 (NISTEP) では、2016年1月22日の第5期科学技術基本計画 閣議決定を受け、第1期から第5期までの基本計画の変遷の分析と可視化に取り組んでいる。本記事ではその試行的分析の一部を示す。

当研究所 (NISTEP) では、2016年1月22日の第5期科学技術基本計画 閣議決定を受け、第1期から第5期までの基本計画の変遷の分析と可視化に取り組んでいる。科学技術基本計画は、5年ごとに策定される科学技術の振興に関する政府の基本的な計画である。このような基本的な政策文書のテキスト分析について、NISTEP においてはこれまでに、藤垣・永田 (2000) が科学技術会議の1960年から1996年までの答申を分析し政策イシューの変化を明らかにすることを試み、近藤・山本 (2005) が第1期基本計画と第2期基本計画の出現頻度上位の単語を比較し政策文書が持つメッセージの顕在化を試みている。本研究は、それらの先行研究を踏まえ、第1期から第5期までの論点の変遷を捉え、第5期基本計画の外形的な特徴を計量分析 (文章内における特定の単語の出現頻度や、他の単語との同時出現 (共起) 頻度など) によって示すことを目的としている。以下に、現時点での試行的な分析の結果の一部を示す。

図表1は、第1期から第5期までの基本計画の本文から名詞 (キーワード) のみを取り出し、それらを共起関係でクラスタリングした後に、筆者が選択したキーワードについて「各基本計画の文の総数に対して、それらのキーワードを含む文がどの程度の割合で存在するか」を円の半径の大きさで表した図である。図表1を見ると、第1期・第2期には存在しなかったキーワード「イノベーション」が第3期における出現以降、急速にイシューとしての重要性を伸ばしていること、「経済」や「企業」等の経済活動に関するキーワードが (外形的には) 第5期の大きな関心対象の一つであるというメッセージを持っていることが読み取れる。

図表2は、各基本計画とそれらに含まれるキー

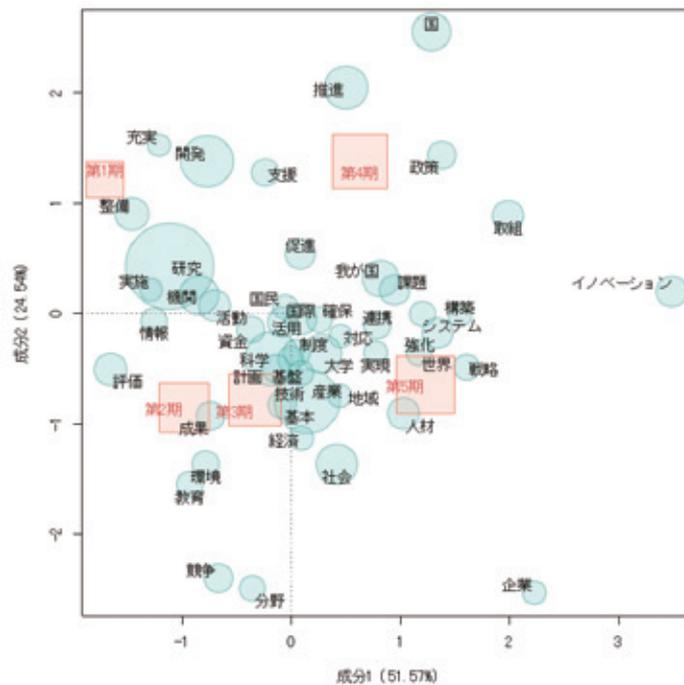
ワードの関係を、対応分析と呼ばれる手法で分析した結果の図である。円の大きさが第5期までの基本計画全体におけるキーワードの出現頻度を、円の位置関係はキーワードと各基本計画の関係及びキーワードの相互の関係 (共起頻度から算出したある種の“近さ”) を示している。図中、赤い四角によって示された各基本計画とキーワードを表す円が近いほど、その計画で相対的に多く用いられたことを意味している。これにより、基本計画間の相対的な位置付けや、各基本計画を特徴付けるキーワードを読み取ることができる。「イノベーション」は、原点から見て第4期と第5期の中間の方角にあり第1期・第2期・第3期から遠いため、第4期と第5期を特徴付けるキーワードであることを示している。また、「企業」は原点から見て第5期の方角にあり、他と離れているため、第5期を特徴付けるキーワードであることを示している。その他、第4期に着目すると、図中のキーワードは全期を通して出現頻度が上位のものであるが、それらのキーワードが密に配置されている原点付近から離れているため、第4期は他の基本計画とは異なる構造であることも見て取れる。

科学技術基本計画は最も分量の多い第5期においてもただか53ページ程度の分量であり、人間が精読することは困難ではない。また、策定の議論に関与した識者等は文言や行間の含意の詳細を熟知している。しかし、様々な施策の根拠となる政策文書は、その文書自体の外形が書き手の意図を離れ、政策イシュー間の重み付けや政策メッセージを持ち得る。本研究は、そのような重み付けやメッセージを明らかにすることによって、科学技術・学術政策の構造化に資するものである。

図表 1 主な単語の出現頻度



図表 2 対応分析による各期の関係と特徴の可視化



[カラー図] カラーの図表はウェブ版を御覧ください。

参考情報

- 1) 藤垣裕子・永田晃也 (2000). 科学技術政策コンセプトの進化プロセス - 科学計量学的アプローチによるダイナミクスの分析 -, 科学技術政策研究所 POLICY STUDY; 005, <http://hdl.handle.net/11035/726>
- 2) 近藤正幸・山本桂香 (2005). 科学技術政策文献の構造分析・内容分析 - 第1期科学技術基本計画及び第2期科学技術基本計画を対象として -, 科学技術政策研究所 調査資料; 115, <http://hdl.handle.net/11035/843>
- 3) 図表 1 の元になる数値の算出及び図表 2 の作成において、立命館大学 樋口耕一 准教授が開発したテキストデータの分析ツール KH Coder を使用している。