

Title	テキストマイニングを用いた科学技術・イノベーション基本計画の分析
Author(s)	山口, 佳和
Citation	年次学術大会講演要旨集, 36: 800-803
Issue Date	2021-10-30
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/17946
Rights	本著作物は研究・イノベーション学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Research Policy and Innovation Management.
Description	一般講演要旨

2 G 1 5

テキストマイニングを用いた科学技術・イノベーション基本計画の分析

○山口佳和（千葉工業大学）
yamaguchi.yoshikazu@p.chibakoudai.jp

1. はじめに

科学技術・イノベーション基本計画[1]は科学技術政策を体系的に記述するとともに対象期間中の政府研究開発投資額を明記していて、日本の科学技術政策の推進の基盤となっている。2021年3月には第6期の基本計画が策定された。本研究では第1期から第6期までの基本計画の記述に着目する。第31回大会で山口[2]は第1期～第5期の基本計画の記述をテキストマイニング分析して、頻出単語のクラスターによるテーマと各基本計画との関わりの強さを分析して報告した。

本研究の目的は、第1期～第6期の基本計画の記述をテキストマイニング分析して、どのような頻出単語とそれらが組み合わさったテーマで構成されているか、テーマはどのように変遷してきているか、研究、イノベーションがどのように位置付けられるかを明らかにすることである。

2. 研究方法

表1に示す第1期～第6期の基本計画の記述をデータとして収集する。記述内容を把握するのに適していると考えられる単語の出現回数と登場文書数の制限を設定してデータ中の頻出単語を抽出する。テキストマイニングツールとしてはKH Coder[3]を用いる。基本計画の記述を内容のまとまりによるサブ文書に分割してどの単語がどの文書に何回登場するかを求める。このデータに基づくクラスター分析を行って単語をクラスターに分類する。クラスターが記述内容によるテーマを表していると考え、単語の構成や文書の記述に基づいてテーマ名を付ける。これらのテーマが第1期から第5期にかけてどのように変遷してきているかを延べ単語数の集計や双対尺度法により分析する。さらに、「イノベーション」がテーマと各期の関係においてどのように位置付けられるかを分析する。

表1 第1期～第6期科学技術基本計画(注)の概要とデータ量

期	対象期間	目次	分析対象		
			文字数	単語数	文書数
第1期	96～00年度	第1章 研究開発の推進に関する総合的方針 第2章 総合的かつ計画的な施策の展開	25,751	15,352	73
第2期	01～05年度	第1章 基本理念 第2章 重要政策 第3章 科学技術基本計画を実行するに当たっての総合科学技術会議の使命	46,511	28,057	89
第3期	06～10年度	第1章 基本理念 第2章 科学技術の戦略的重点化 第3章 科学技術システム改革 第4章 社会・国民に支持される科学技術 第5章 総合科学技術会議の役割	51,034	30,627	93
第4期	11～15年度	I. 基本認識 II. 将来にわたる持続的な成長と社会の発展の実現 III. 我が国が直面する重要課題への対応 IV. 基礎研究及び人材育成の強化 V. 社会とともに創り進める政策の展開	59,299	35,018	95
第5期	16～20年度	第1章 基本的考え方 第2章 未来の産業創造と社会変革に向けた新たな価値創出の取組 第3章 経済・社会的課題への対応 第4章 科学技術イノベーションの基盤的な力の強化 第5章 イノベーション創出に向けた人材、知、資金の好循環システムの構築 第6章 科学技術イノベーションと社会との関係深化 第7章 科学技術イノベーションの推進機能の強化	65,757	38,124	99
第6期	21～25年度	第1章 基本的な考え方 第2章 Society 5.0の実現に向けた科学技術・イノベーション政策 第3章 科学技術・イノベーション政策の推進体制の強化	105,471	68,184	117

(注)第1～5期は科学技術基本計画、第6期は科学技術・イノベーション基本計画。

3. 研究結果

頻出単語の最小出現回数が小さいと重要でないテーマまで拾ってしまい、大きいと重要なテーマが拾えない可能性がある。最大登場文書数が小さいと重要なテーマが拾うことができず、大きいと特定のテーマを表さない単語まで拾ってしまう可能性がある。表 2 に示すとおり最小出現回数を 50 語、最大登場文書数を 100 文書とした。

表2 最小出現回数と最大登場文書数による抽出語数

		最大登場文書数						
		40	60	80	100	120	140	160
最小出現回数	20	350 ○ ×	420 ○ ×	461 ○ ×	483 ○ ○	500 ○ ○	510 ○ ○	518 ○ ○
	30	199 ○ ×	269 ○ ×	310 ○ ×	332 ○ ○	349 ○ ○	359 ○ ○	367 ○ ○
	40	102 ○ ×	172 ○ ×	213 ○ ×	235 ○ ○	252 ○ ○	262 ○ ○	270 ○ ○
	50	49 ○ ×	117 ○ ×	158 ○ ×	180 ○ ○	197 ○ ○	207 ○ ○	215 ○ ○
	60	20 × ×	73 × ×	114 × ×	136 × ○	153 × ○	163 × ○	171 × ○
	70	12 × ×	49 × ×	90 × ×	112 × ○	129 × ○	139 × ○	147 × ○
	80	4 × ×	31 × ×	67 × ×	89 × ○	106 × ○	116 × ○	124 × ○

(注)最小出現回数では「プログラム」、「ポスト」、最大文書数では「ネットワーク」、「仕組み」が含まれると○、含まれないと×。

表 3 に示すとおり抽出した 180 語から特定のテーマを表さないと考えられる 74 語を除いて、106 語を分析対象とした。イノベーションは登場文書数が多いため抽出した 106 語には含まれていなかった。

表3 抽出語の一覧

抽出語	出現回数	登場文書数	抽出語	出現回数	登場文書数	抽出語	出現回数	登場文書数	抽出語	出現回数	登場文書数
データ	322	80	拠点	108	60	公共	80	64	挑戦	61	35
投資	223	76	改善	107	79	先端	80	55	ポスト	59	34
施設	203	79	サービス	107	50	循環	80	43	AI	59	31
知的	200	89	提供	106	81	空間	79	28	海洋	59	22
改革	189	89	民間	106	65	革新	78	52	保障	58	41
国立	170	67	海外	106	58	リスク	76	53	学術	58	27
基礎	159	91	協力	104	67	学校	76	31	還元	57	46
配分	159	83	養成	101	56	発信	75	54	感染	56	39
共同	156	87	知識	100	61	Society	75	42	審査	56	24
設備	151	65	融合	98	70	普及	74	61	国内	55	44
創造	140	93	ベンチャー	98	28	共有	70	53	長期	55	41
エネルギー	137	61	水準	96	68	成長	70	42	流動	55	39
標準	137	48	試験	96	54	インフラ	70	36	高齢	54	35
能力	136	83	外国	96	50	都市	70	20	財政	54	33
持続	131	89	若手	96	41	再生	69	35	材料	54	25
拡充	123	72	機会	94	65	教員	69	26	特性	53	44
重点	121	61	通信	94	56	活性	68	53	統合	53	40
ネットワーク	120	82	実装	93	54	経営	67	36	人類	53	29
生活	120	71	スマート	91	32	オープン	67	32	構造	52	41
領域	117	67	災害	88	50	サイバー	67	29	プログラム	52	37
活躍	117	66	交流	85	55	経験	66	49	先進	51	39
仕組み	116	90	プロジェクト	85	54	市場	66	45	見直し	50	36
法人	116	48	大学院	85	30	未来	66	41	早期	50	36
地球	115	64	専門	84	53	キャリア	64	39	倫理	49	19
ニーズ	110	80	マネジメント	83	59	分析	63	49	方針	45	28
財産	110	41	課程	82	30	導入	62	47			
医療	109	50	デジタル	81	41	加速	61	48			

表 4 に示すとおり 106 語は 20 クラスターに分類することができた。各クラスターのテーマ名としてデータ、知的基盤、研究環境、研究交流、エネルギー、マネジメント、政策方針、市場ニーズ、ライフサイエンス、社会貢献、オープン化、研究プロジェクト、環境問題、Society 5.0、海洋開発、ベンチャー、スマートシティ、人材育成、研究費配分、研究投資を付けた。

表4 テーマと抽出語の一覧

テーマ名	抽出語	抽出語数	出現回数
データ	データ	1	322
知的基盤	知的、財産、標準	3	447
研究環境	国立、法人、施設、設備、基礎、重点、領域、拠点	8	1145
研究交流	ネットワーク、海外、外国、拡充、民間、試験、交流、共同	8	888
エネルギー	エネルギー	1	137
マネジメント	仕組み、マネジメント、リスク、提供、先端、経営、ニーズ	7	638
政策方針	公共、活性、長期、見直し、方針、還元、早期、導入	8	467
市場ニーズ	サイバー、特許、構造、先進、国内、加速、普及、共有、分析、統合、プログラム、市場、発信、高齢、保障	15	904
ライフサイエンス	倫理、インフラ、感染	3	175
社会貢献	水準、人類、財政、改善、生活	5	430
オープン化	オープン	1	67
研究プロジェクト	知識、プロジェクト、学術、挑戦、融合、創造、材料、通信、革新	9	768
環境問題	地球、協力、実装、循環、持続、成長、再生、災害	8	750
Society 5.0	デジタル、AI、空間、未来、Society	5	360
海洋開発	海洋	1	59
ベンチャー	ベンチャー	1	98
スマートシティ	都市、スマート、サービス、医療	4	377
人材育成	能力、専門、養成、機会、学校、キャリア、経験、流動、ポスト、活躍、若手、課程、教員	13	1099
研究費配分	大学院、改革、配分、審査	4	489
研究投資	投資	1	223

図1から第1期で研究交流、第6期でデータ、Society 5.0が大きいことなどが分かった。図2から第1期で研究交流、研究環境、人材育成、第2期で人材育成、研究環境、研究プロジェクト、第3期で研究環境、人材育成、第4期で研究環境、環境問題、人材育成、第5期で市場ニーズ、研究環境、第6期で市場ニーズ、Society 5.0、データの割合が大きいことなどが分かった。

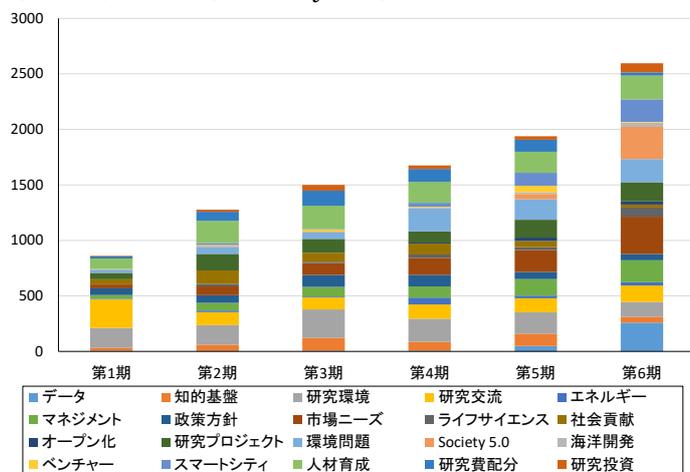


図1 テーマ別抽出語出現回数推移

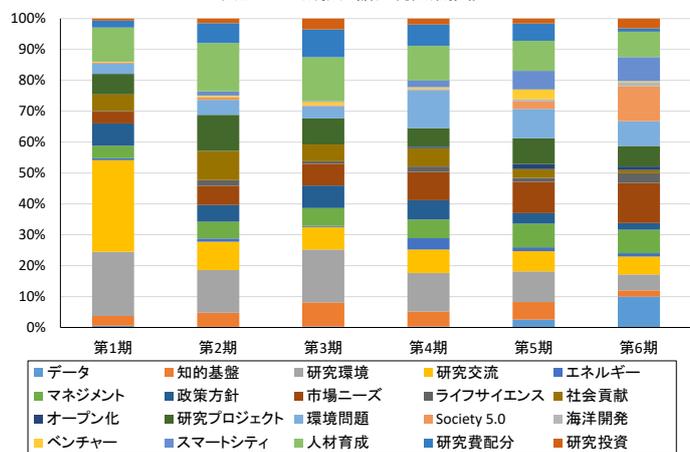


図2 テーマ別抽出語出現回数割合推移

図3に示す第1期～第6期の基本計画と20クラスターの双対尺度分析の結果から、第1期は研究交流、第2期は研究環境、政策方針、第3期は社会貢献、第4期は研究費配分、第5期は環境問題、第6期はSociety 5.0、データ、海洋開発、スマートシティと関わりが強い。第1期、第2～5期、第6期がそれぞれ離れたところに分布している。図4からイノベーションは第5期と関わりが強い。

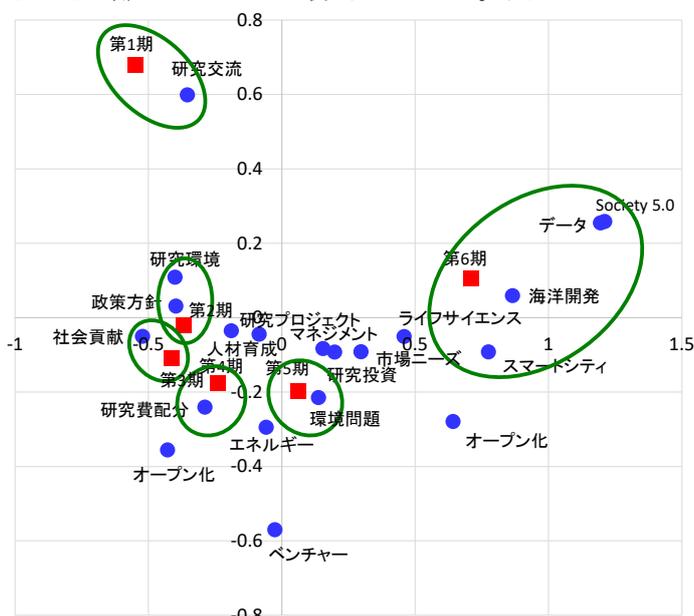


図3 テーマと基本計画の双対尺度分析

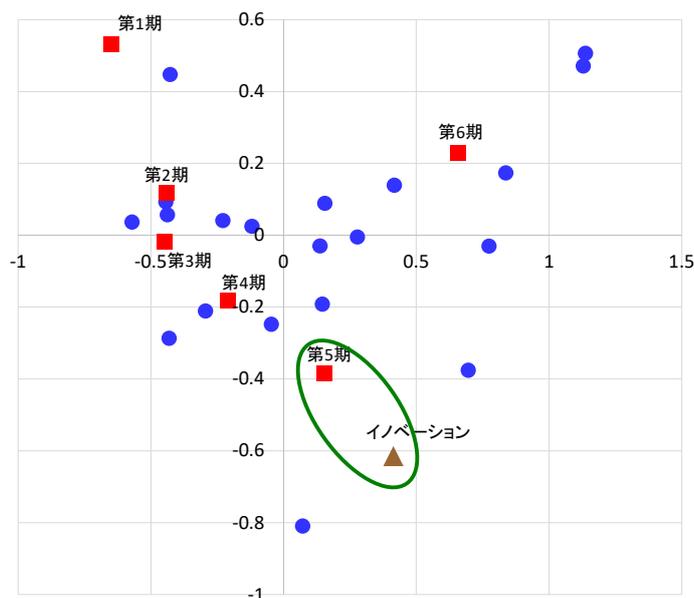


図4 イノベーションを含むテーマと基本計画の双対尺度分析

4. まとめ

科学技術・イノベーション基本計画の記述のテキストマイニング分析により、頻出単語によるテーマがどのように変遷してきているか、イノベーションがどのように位置付けられるかを明らかにした。

(引用文献)

- [1]閣議決定、第1期～第6期科学技術基本計画(科学技術・イノベーション基本計画)、<https://www8.cao.go.jp/cstp/kihonkeikaku/index.html>、1996、2001、2006、2011、2016、2021。
- [2]山口佳和、科学技術基本計画のテキストマイニング分析と研究、イノベーションの位置付け、研究・イノベーション学会第31回年次学術大会予稿集、pp.652-655、2016。
- [3]樋口耕一、KH Coder、<http://khc.sourceforge.net>、2021。