

Title	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
Author(s)	山口, 佳和
Citation	年次学術大会講演要旨集, 31: 652-655
Issue Date	2016-11-05
Type	Conference Paper
Text version	publisher
URL	<a href="http://hdl.handle.net/10119/13994">http://hdl.handle.net/10119/13994</a>
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

## 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

○山口佳和（千葉工業大学）

### 1. はじめに

科学技術基本計画[1]は、日本の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、科学技術政策の推進の基盤となっている。2016年1月には、第5期の計画が策定された。本研究では、第1期から第5期までの計画の記述に着目する。

前回大会で山口[2]は、イノベーションを含む新聞記事のテキストマイニング分析を行い、2000年までは民間との関わりが深い記事のテーマが多く、2001年以降は政府との関わりが深いテーマが多いことなどを明らかにした。また、山口[3、4]は、科学技術基本計画の記述における産学連携の位置付けをテキストマイニングを用いて分析し、第1期～第4期では産学連携はイノベーションや経済再生と結び付きが強いこと、第1期～第5期では産学連携は資源配分と結び付きが強いことなどを明らかにした。

本研究の目的は、第1期～第5期の科学技術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

第1期～第5期の科学技術基本計画の記述(表1)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、KHcoder[5]を用いる。

計画の記述を内容のまとまりによる文書に分割して、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

表1 第1期～第5期科学技術基本計画の概要とデータ量

期	対象期間	目次	分析対象	
			文字数	単語数
第1期	99年度～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	24,394	14,955
第2期	01年度～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	45,258	27,326
第3期	08年度～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	49,537	29,798
第4期	11年度～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	57,582	34,117
第5期	16年度～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造の取組 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	84,086	37,204

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、「研究」と「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。

### 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。ここでは、最小出現回数を40語、最大登場文書数を80文書とした(表2)。

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
最小 出現 回数	20	269	332	363	378	392
	30	151	214	245	260	274
	40	58	121	152	167	181
	50	28	91	122	137	151
	80	14	62	93	108	122

抽出した152語から特定のテーマを表さないと考えられる49語を除いて、103語を分析対象とした(表3)。なお、研究、イノベーションは登場文書数が多いため、抽出した152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
地域	283	72	水準	96	68	外国	78	34	国内外	51	45
企業	218	80	若手	96	41	プロジェクト	71	44	見直し	50	36
知的	186	78	領域	95	54	交流	70	43	学校	50	20
施設	180	66	試験	93	51	成長	70	42	公共	50	46
国立	170	67	改善	91	64	再生	69	35	審査	50	19
産学	165	76	ベンチャー	90	24	運用	68	48	特性	49	40
投資	141	52	活躍	90	49	提供	68	54	博士	47	22
配分	136	71	ネットワーク	89	59	先端	64	41	任期	46	13
基礎	133	73	ニーズ	87	62	融合	64	46	参画	46	33
設備	130	58	持続	87	61	データ	63	23	市場	45	31
生活	120	71	事業	86	51	導入	62	47	女性	45	11
組織	119	63	海外	85	46	サ・ビス	61	26	倫理	44	15
法人	116	48	大学院	85	30	災害	60	34	ポスト	43	23
形成	115	79	知識	85	49	国家	59	38	論文	43	20
貢献	113	71	専門	84	53	リスク	58	40	流動	43	29
重点	111	54	振興	84	60	活性	57	44	財政	42	24
共同	111	61	解決	83	55	共通	56	41	人々	42	29
エネルギー	109	44	通信	81	48	長期	55	41	学生	41	27
民間	106	65	仕組み	80	62	発信	55	41	機器	41	20
財産	105	36	地球	80	44	特許	55	20	雇用	41	34
創造	104	68	機会	79	53	高齢	54	35	課程	40	22
協力	104	67	経費	78	35	人類	53	29	アジア	40	22
標準	103	35	革新	78	52	還元	53	42	オ・ブ	39	16
価値	101	59	拠点	77	41	マネジメント	52	36	変革	39	26
養成	100	55	医療	75	32	材料	52	23	挑戦	38	21
政府	96	45	運営	74	46	普及	52	43			

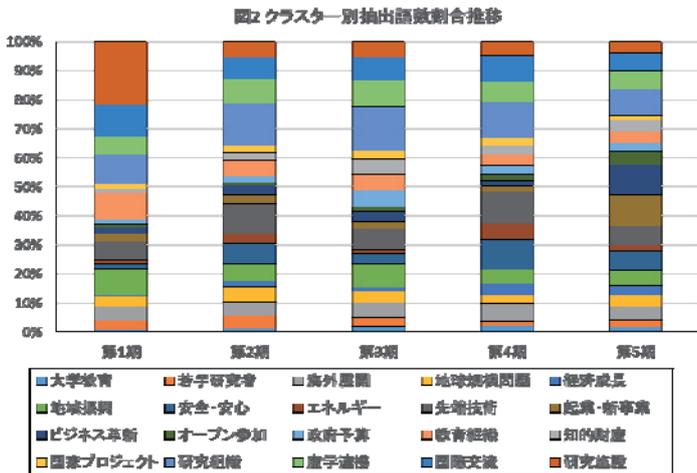
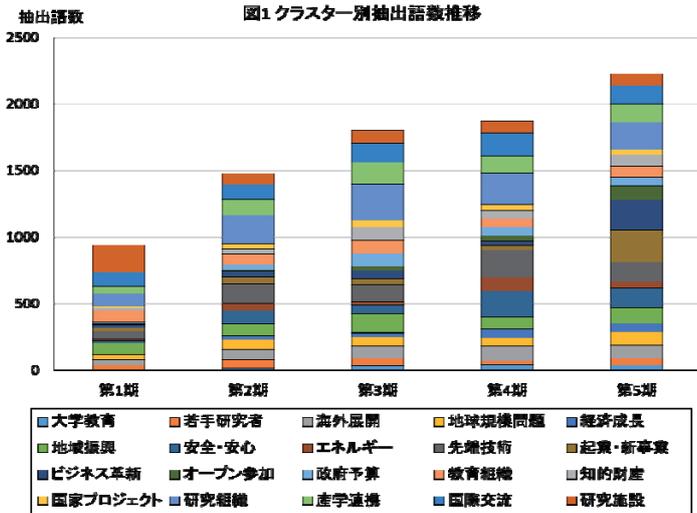
103語は20クラスターに分類することができ、各クラスターのテーマ名として、大学教育、若手研究者、海外展開、地球規模問題、経済成長、地域振興、安全・安心、エネルギー、先端技術、起業・新事業、ビジネス革新、オープン参加、政府予算、教育組織、知的財産、国家プロジェクト、研究組織、産学連携、国際交流、研究施設を付けた(表4)。

クラスター別の抽出語数の推移を見ると、第1期で研究施設、第5期で起業・新事業、ビジネス革新が大きいことなどが分かった(図1)。

クラスター別の抽出語数割合の推移を見ると、第1期で研究施設、教育組織、第2期と第4期で先端技術、第3期で産学連携、第4期で起業・新事業、ビジネス革新、オープン参加の割合が大きいことなどが分かった(図2)。

表4 抽出語クラスターの一覧

クラスター名	抽出語	抽出語数	出現回数
大学教育	学生、博士、課程	3	126
若手研究者	若手、ポスト、任期、進路	4	226
海外展開	機会、機会、海外、外国、国際	5	427
地球規模問題	人類、国際、地球、解決	4	329
経済成長	雇用、技術、成長	3	195
地域振興	活性、地域、発兵、製造、復興	3	166
安全・安心	課題、高齢、人々、生活、消費、リスク、管理、再生	5	522
エネルギー	普及、エネルギー、革新	3	239
先端技術	戦略、融合、ネットワーク、連携、共通、先端、情報、標準、材料、市場	10	639
起業・新事業	事業、企業、ベンチャー	3	394
ビジネス革新	関係、サービス、価値、変革、組織、変革	4	302
オープン参加	テーマ、オープン、女性、参加	4	185
政府予算	削減、政府、投資	3	279
教育組織	専門、学校、民間、法人、導入	3	418
知的財産	知的、財産	2	291
国家プロジェクト	プロジェクト、国際、基盤	3	183
研究組織	研究、組織、運営、基盤、重点、大学別、改善、配分、運営、特性、マネジメント、見直し	12	1088
産学連携	産学、コース、仕組み、特許、運用、運用、拡大、発展	3	600
国際交流	交流、能力、アジア、水準、国内外、環境、拠点、共同	3	664
研究施設	国立、試験、施設、設備	4	373



第1期～第5期の計画と20クラスターの双対尺度分析の結果から、第1期は研究施設、第2期は研究組織、国家プロジェクト、海外展開、政府予算、先端技術、第3期は産学連携、国際交流、第4期はエネルギー、第5期はオープン参加と関わりが強いことが分かった。さらに、第5期は第1期～第4期よりも、起業・新事業、ビジネス革新に近いところに位置付けられることが分かった(図3)。

研究、イノベーションを含めた双対尺度分析の結果から、研究は第2期、第3期と関わりが強いが、第1期、第4期、第5期に対してはやや離れたほぼ等距離に位置付けられること、イノベーションは第5期と関わりが深く、第1～第4期からは遠距離に位置付けられることが分かった(図4)。研究は第1期～第5期のいずれにおいても、500回～600回強と安定して登場していること、イノベーションは第3期から登場し第5期に223回と特に多く登場していることが、これらの背景にある。

図3 第1～5期基本計画と20クラスターの双対尺度分析

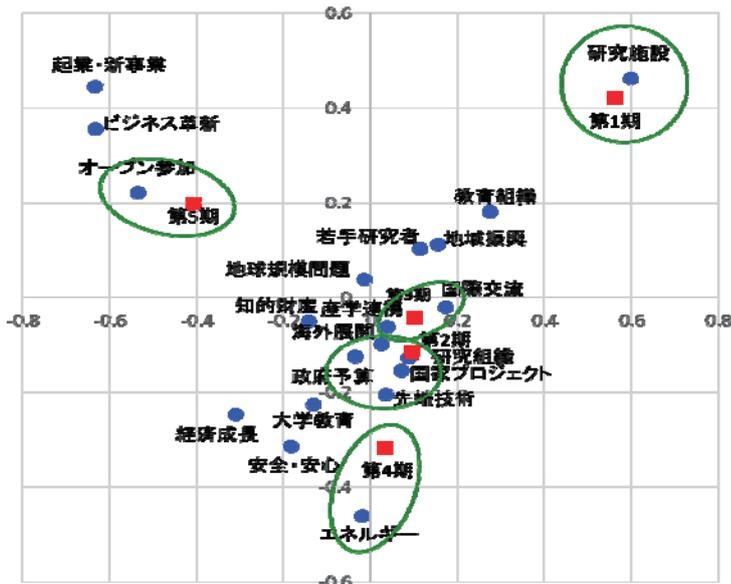
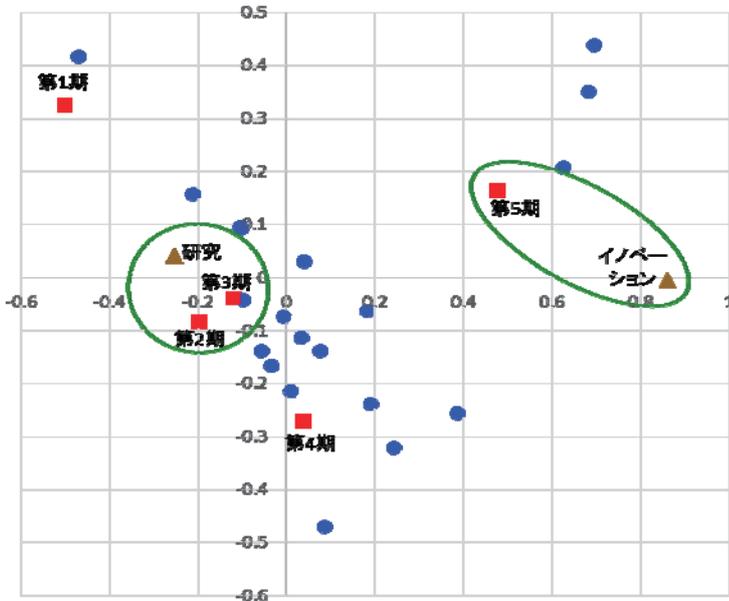


図4 研究、イノベーションを含めた双対尺度分析



#### 4. まとめ

科学技術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、研究と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5期科学技術基本計画、1996、2001、2006、2011、2016。  
<http://www8.cao.go.jp/cstp/kihonkeikaku/index5.html>
- [2]山口佳和、イノベーションを含む新聞記事のテキストマイニング分析、研究・技術計画学会第30回年次大会予稿集、pp. 894-897、2015。
- [3]山口佳和、科学技術基本計画の記述における産学連携の位置付けの定量分析の試み、産学連携学会第13回大会予稿集、pp. 193-194、2015。
- [4]山口佳和、科学技術基本計画の記述における産学連携の位置付けの定量分析の試み(その2)、産学連携学会第14回大会予稿集、pp. 226-227、2016。
- [5]樋口耕一、KHcoder、2016。 <http://khc.sourceforge.net/>