

Title	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
Author(s)	山口, 佳和
Citation	年次学術大会講演要旨集, 30: 894-897
Issue Date	2015-10-10
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/13418
Rights	本著作物は研究・技術計画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 Science Policy and Research Management.
Description	一般講演要旨

2H11

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

○山口佳和（千葉工業大学）

1. はじめに

科学技術に関連する政策や研究において、イノベーションが重要なキーワードになっている。かつては、昭和31年年次経済報告(1956年経済白書)[1]に「このような投資活動の原動力となる技術の進歩とは原子力の平和的利用とオートメーションによって代表される技術革新(イノベーション)である」との記述があったように、イノベーションと技術革新が区別されないまま用いられることがあった。現在では、政策だけでなく法令用語としても用いられ、また新聞などでもよく見かけるようになり、政策や社会に広く普及した用語となっている。

本研究では、このイノベーションという用語が、政策や社会にどのような経緯を経て普及してきたかに焦点を当てる。新聞記事を資料として用い、イノベーションという用語がどのように登場してきたかを分析することを試みる。それにより、イノベーションの普及の経緯を明らかにするとともに、イノベーションの利用や政策を検討する上で有用な示唆が得られることを期待する。

2. 研究目的

本研究の目的は、イノベーションを含む1990年以降の新聞記事を収集し、テキストマイニングなどを用いて分析して、イノベーションがどのようなトピックスの中で登場し普及してきたかという経緯を明らかにすることである。

3. 研究方法

ヨミダス歴史館[2]を検索して、イノベーションを含む記事を収集する。テキストマイニングツールであるKH coder[3]を用いて、収集した記事から頻出単語を抽出する。頻出単語を記事との関係から、いくつかのクラスターに分類する。クラスターと年との関係を分析する。以上の分析結果に基づいて、イノベーションが普及してきた経緯を明らかにする。

4. 分析結果

全文検索、全国版記事、1990年1月1日から2014年12月31日までを条件に検索したところ、1,051件を収集することができた。図1に記事数の推移を示した。記事数は2004年以降に増加し、2007年に



101 件に達した後、減少して 2009 年に 32 件まで落ち込んだ。2010 年以降は増加して 2013 年にピークの 160 件に達し、2014 年は 125 件とやや落ち着いた。落ち込んだ時期はあるものの、ここ 10 年間ほどは、新聞記事にイノベーションが登場することが多くなったと言える。

1,051 件の記事から、頻出単語を抽出した。最小出現回数と最大登場文書数による抽出単語数を表 1 に示した。多くの文書に登場するトピックスを表さない一般的な単語をなるべく抽出しないようにすること、少ない登場回数であってあまり重要でないトピックスを表す特殊な単語をなるべく抽出しないようにすること、さらに重要なトピックスを表す単語を漏れなく抽出することを考慮した結果、最小出現回数を 280、最大登場文書数を 385 とした。その結果、157 語を抽出することができた。

表1 最小出現回数と最大登場文書数による抽出単語数

		最大登場文書数									
		210	245	280	315	350	385	420	455	490	526
最小出現回数	105	499	525	534	548	552	555	559	561	563	565
	140	345	371	380	394	399	401	405	407	409	411
	175	237	263	272	286	291	293	297	299	301	303
	210	174	200	209	223	228	230	234	236	238	240
	245	128	154	163	177	182	184	188	190	192	194
	280	101	127	136	150	155	157	161	163	165	167
	315	76	102	111	125	130	132	136	138	140	142
	350	58	82	91	105	110	112	116	118	120	122
	385	39	62	71	85	90	92	96	98	100	102
	420	31	53	62	76	81	83	87	89	91	93

157 語にはまだ一般的な単語や特殊な単語が依然として含まれているため、記事の内容などから判断して 55 語を除いた。表 2 に 102 語の抽出単語一覧を示した。登場回数の最大は大学の 1,695、次いで

表2 抽出単語一覧

単語	登場回数	登場文書数	語	登場回数	登場文書数	単語	登場回数	登場文書数	単語	登場回数	登場文書数
大学	1,695	306	安倍	698	189	行政	455	158	世紀	336	134
科学	1,513	333	教授	694	299	エネルギー	452	176	危機	333	144
医療	1,449	256	再生	685	239	細胞	441	87	景気	329	127
改革	1,381	303	連携	678	259	文化	438	140	外交	328	94
政策	1,324	359	社長	667	200	生活	435	182	ベンチャー	328	92
教育	1,086	274	米国	637	214	学生	424	113	女性	325	113
地域	1,074	284	市場	627	223	国家	423	166	土壌	322	223
会議	993	334	消費	620	201	人間	413	111	機器	320	223
制度	929	273	予算	617	181	視点	408	173	研究所	316	177
大阪	914	214	中国	616	185	外国	400	144	体制	315	177
環境	891	304	雇用	586	178	委員	395	193	ビジネス	313	178
支援	870	320	規制	572	221	サービス	381	176	オープン	310	66
事業	862	341	政権	569	195	国内	379	216	組織	308	171
情報	861	303	人材	567	224	年金	379	81	仕事	308	152
分野	843	357	金融	560	177	評価	377	210	復興	301	86
経営	843	274	地方	546	152	共同	373	192	製品	298	155
首相	842	244	会長	537	205	法人	370	184	大学院	296	127
財政	792	206	機関	531	268	育成	363	190	全国	291	164
国民	777	188	時代	518	212	国会	363	109	施設	291	137
投資	759	272	海外	516	223	労働	362	173	都市	285	122
競争	755	304	生産	502	240	特区	356	107	ファンド	285	30
アジア	755	196	センター	468	228	資金	350	159	仕組み	284	166
政治	752	178	内閣	465	180	高齢	350	170	価値	282	174
東京	743	200	システム	464	210	民間	340	178	中小	281	122
関西	733	151	計画	456	235	構造	348	165			
保障	720	149	会社	456	221	専門	343	181			

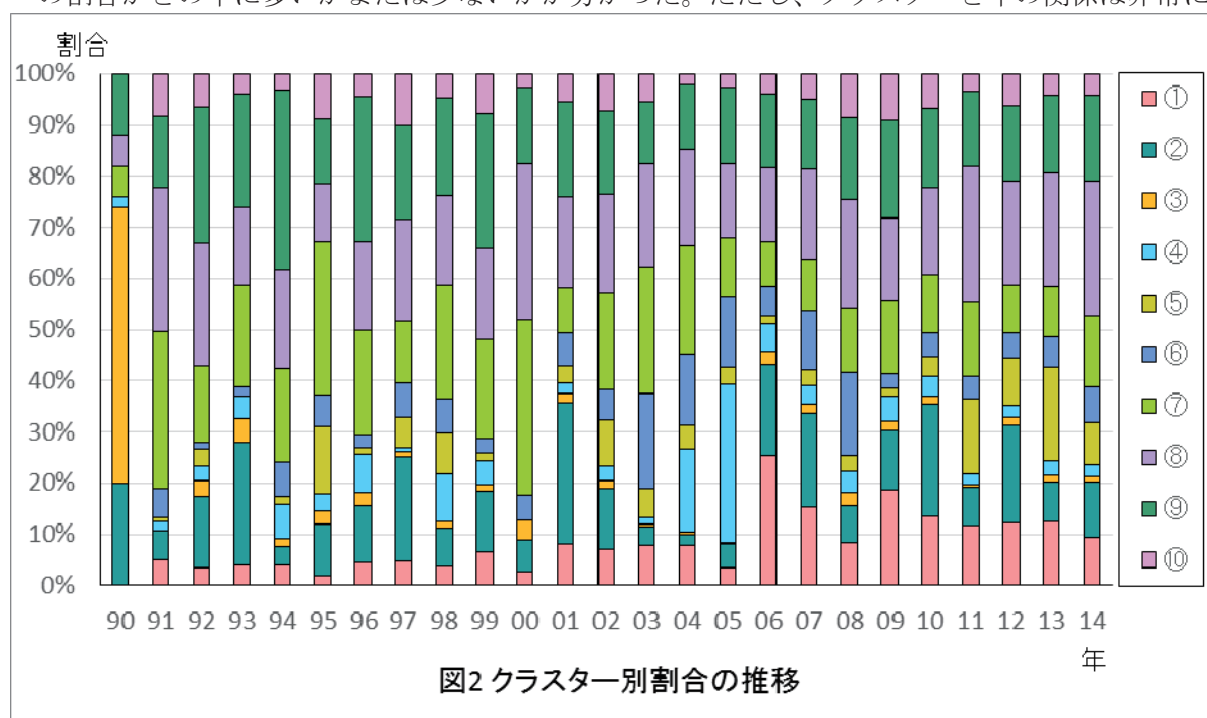
科学の 1,516、医療の 1,449 であった。登場文書数の最小はファンドの 30、次いでオープンの 66、年金の 81 であった。

102 語をクラスター分析した。データはどの記事にどの単語が何回登場するかを用い、分析条件はユークリッド距離、ワード法とした。その結果、10 クラスターに分類することができた。表 3 にクラスターと含まれる単語を示した。クラスターに含まれる単語とそれらの単語が多く含まれる記事の内容から判断して、クラスター名を付けた。10 クラスターは、政策、制度改革、国際関係、海外投資、医療特区、人材教育、ビジネス、プロジェクト、市場、制度となった。最大のクラスターはプロジェクトで、単語数が 23、単語出現回数が 11,360 であった。最小のクラスターは国際関係で、単語数が 3、単語出現回数が 932 であった。

表3 クラスターと含まれる単語

クラスター番号	クラスター名	単語(単語数)	単語出現回数
①	政策	内閣、科学、政策、予算、政権、会議、首相、安倍 (8)	7,024
②	制度改革	国会、雇用、労働、景気、消費、国民、政治、年金、構造、地方、保障、財政、改革 (13)	7,955
③	国際関係	主義、外交、価値 (3)	932
④	海外投資	ファンド、オープン、アジア、米国、中国 (5)	2,603
⑤	医療特区	大阪、特区、関西、体制、医療、機器 (6)	4,090
⑥	人材教育	人材、育成、学生、大学、教育 (5)	4,135
⑦	ビジネス	生産、サービス、会社、社長、経営、東京、センター、世紀、専門、大学院、文化、人間、情報、ビジネス (14)	7,063
⑧	プロジェクト	事業、支援、計画、復興、細胞、再生、製品、規制、教授、研究所、機関、組織、会長、委員、システム、仕組み、拠点、施設、共同、地域、連携、全国、中小 (23)	11,360
⑨	市場	ベンチャー、民間、資金、投資、高齢、生活、女性、仕事、都市、海外、外国、市場、国内、競争、環境、時代、国家、分野、エネルギー (19)	9,293
⑩	制度	制度、行政、法人、評価、金融、危機 (6)	3,024

クラスターに含まれる単語の登場回数を分析した。図 2 にクラスター別割合の推移を示した。どのスターの割合がどの年に多いかまたは少ないかが分かった。ただし、クラスターと年の関係は非常に複雑な



ものになったため、さらなる明確化が必要であると考えた。

クラスターと年の関係をより明確にし可視化するため、双対尺度法を用いた分析を行った。図3に分析結果を示した。近い関係にある年またはクラスターが、近くにプロットされている。1991年とビジネス、1992年と市場、制度、1994年と人材教育、1997年と制度、2000年とビジネス、2001年と政策、制度改革、国際関係、2002年とプロジェクト、2005年と海外投資、2011年と医療特区、2012年と制度、2013年と医療特区、2014年とプロジェクトが近い関係にあることが分かった。すなわち、イノベーションを含む新聞記事の主要なトピックスが、以上のように変遷してきており、イノベーションという用語がそうした経緯を経て普及してきたと言える。

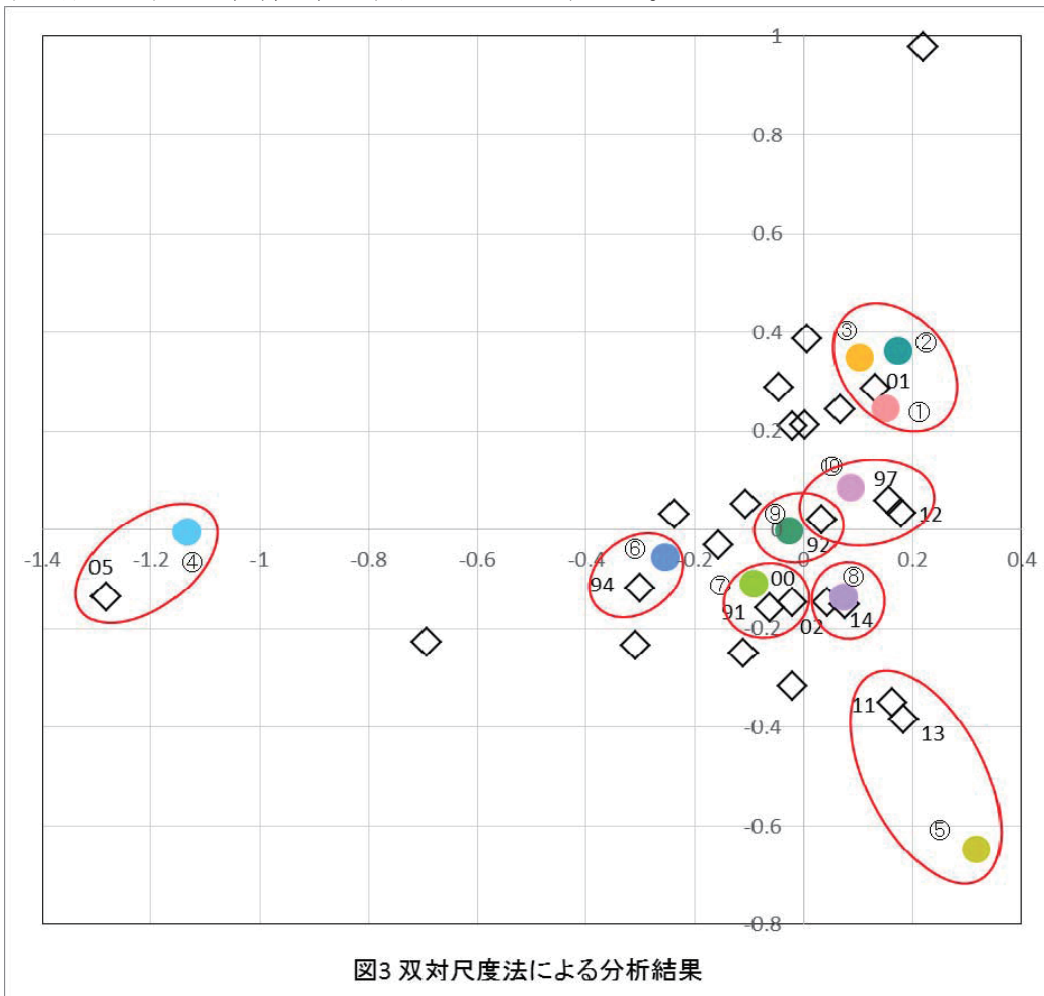


図3 双対尺度法による分析結果

5. まとめ

イノベーションという用語が、どのような新聞記事のトピックスの中で取り上げられてきたかを分析することにより、現在のように普及するに至った経緯を明らかにした。

今後は、予算書、計画書、報告書、プロジェクト評価書、白書などの政策に関する様々な文書について、イノベーションがどのようなトピックスの中で取り上げられているかを、テキストマイニングなどを用いて分析することが課題である。

(引用文献)

[1]経済企画庁、昭和31年年次経済報告(1956年経済白書)、<http://www5.cao.go.jp/keizai3/keizaiwp/wp-je56/wp-je56-0000i1.html>、1956年。

[2]読売新聞、ヨミダス歴史館、<https://database.yomiuri.co.jp/rekishikan/>、2015年。

[3]樋口耕一、KHcoder、<http://khc.sourceforge.net/>、2015年。