

クロスフィールドを考慮した スター・サイエンティストの特徴の分析

Analysis of Features of Star Scientists Considering Cross-Fields research

藤田 正典¹ 菅井 内音² 隅藏 康一¹ 牧 兼充³
Masanori Fujita¹ Naito Sugai² Kouichi Sumikura¹ Kanetaka Maki³

¹ 政策研究大学院大学 ² 東京工業大学, ³ 早稲田大学
¹National Graduate Institute ² Tokyo Institute of Technology ³ Waseda University
for Policy Studies

Abstract: For the development of science and the realization of innovation, the analysis of researchers who contribute them, especially “Star Scientists” who have made remarkable achievements in various aspects including academics, is important to make science and technology policies. In this paper, researchers with highly cited papers are extracted from Clarivate Analytics Web of Science, and their characteristics are analyzed from the viewpoint of interdisciplinary research in each research field. Furthermore, after clarifying the characteristics of researchers with highly cited papers and the characteristics of star scientists with outstandingly highly cited papers, we will try to obtain implications for science and technology innovation policies and corporate technology management strategies.

1. はじめに

イノベーションの源泉である科学を支えているのはサイエンティスト（科学者）であり、科学の発展とイノベーションの実現に向けて、これらを支える科学者、特に学術面をはじめとして様々な点で著しい成果を挙げた「スター・サイエンティスト」の分析は、科学技術政策の策定に重要である。スター・サイエンティストは、研究者として、多くの高被引用論文を発表するだけでなく、特許を多く出願し、更にはベンチャーを設立する等、積極的にビジネスに関与することもある。彼らによる科学的な知識が技術につながり、社会・経済にインパクトを与えるようなイノベティブな製品・サービスに繋がる。

イノベーションに関して、Schumpeter(1926)は、5つのタイプの「Neue Kombination (New Combination)」(=新結合)を定義し、経済発展にはこれらの新結合が必要であるとした。また、内閣府(2016)によれば、第5期科学技術基本計画において、科学技術イノベーションの基盤的な力の強化に向けて「学際的・分野融合的な研究」の推進が謳われている。また、文部科学省(2017)では、研究開発評価の4つの特筆課題の中の1つとして、学際・融合領域・領域間連携研究等の推進を挙げている。

このような背景のもと、本稿では、科学技術文献データベースから高被引用論文を持つ研究者(Main Cited Researcher : MCR)を抽出し、研究分野ごとの学際性の観点から、彼らの特徴を分析する。更に、高被引用論文が著しく多い研究者をスター・サイエンティスト(Star Scientist : SS)と定義し、彼らの特徴を、高被引用論文を持つ研究者の特徴と比較分析する。

以下本稿では、SS研究の課題について述べた後、本研究の分析手法について説明し、MCRやSSの特徴についての分析結果を示す。その上で、日本のMCRやSSについて考察し、科学技術イノベーション政策、及び企業の技術経営戦略に対する示唆を得ることを目指す。

2. 関連研究

SSに関する著名な研究であるZucker et al. (2002)によると、SSは、通常の研究者と比べ、ベンチャー企業を設立する傾向にあり、またその関わっているベンチャー企業は他のベンチャー企業に比較して、高い業績を生み出しているという。更にZucker and Darby (2007)は、SS自身も企業にコミットすることで、その業績が更に向上していることを指摘している。しかしこれらの研究は、生命科学分野等の特

定領域の分析結果をもとにしており、全ての分野を網羅的に分析したものではない。また、米国では SS に注目した実証研究が蓄積されているのに対して、日本の SS の研究はまだ十分とは言えない。

そこで、本稿では、学術文献データベースで高被引用論文 (Highly Cited Paper : HCP) を持つ MCR の特徴と、高被引用論文が著しく多い SS の特徴を、学際によるイノベーションの実現という観点から分野ごとに分析した上で、日本の MCR 及び日本の SS とそれ以外の MCR 及び SS との特徴を比較して分析する。

3. 対象データ

本稿では、Clarivate Analytics 社の学術文献データベースである Web of Science (WoS) を分析対象とする。WoS は、Biology, Computer 等の 21 分野と Multidisciplinary を含む 22 分野に分類されている。分析対象期間を 2008 年から 2016 年までの 9 年間とし、当該期間中の 19,536,453 件の文献を対象として、分野ごとに高被引用論文を持つ研究者を抽出した後、その特徴量を算出して分析する。

4. 手法

分析にあたっては、以下のように、HCP, MCR, SS, 及び研究者の主研究分野 (Main Field : MF) を定める。その上で、研究者の学際性特徴量として、研究分野の多様性 (Field Variety : FV), 研究分野の均等性 (Field Balance : FB) 所属機関の移動頻度 (Institution Change : IC), 単一研究分野比率 (Single Field Ratio), 及び単一所属機関比率 (Single Institution Ratio) を算出し分析する。

4.1. MCR 及び SS の抽出

- (1) 「研究者」: 上述の WoS の 2008 年から 2016 年まで 9 年間 (全期間) の文献から「氏名」より同定された研究者
- (2) 「論文被引用数」: 当該論文発表以降 2016 年までの他論文から当該論文への引用数の合計
- (3) 「HCP」: 発表年ごと、分野ごとに、(2) の論文の中で論文被引用数が上位 1% の論文
- (4) 「MCR」: 全期間、全ての分野で、(1) の研究者の中で(3) の HCP を 1 本以上発表した研究者
- (5) 「SS」: 全期間で、分野ごとに、HCP の本数に基づき(4) の MCR 数の平方根の順位より上位の研究者、及びクロスフィールド (複数分野) での HCP の本数に基づく順位が同等に上位の研究者
- (6) 「SS/JP」: (5) の SS うち、所属機関の所在地が日本である研究者

「SS/exJP」: (5) の SS うち、「SS/JP」を除いた研究者

- (7) 「MCR/JP」: (4) の MCR から(5) の SS を除いた研究者うち、所属機関の所在地が日本である研究者 (「MCR/exSS/JP」と表記してもよいが、本稿では「MCR/JP」と表記する)

「MCR/exJP」: (4) の MCR から(5) の SS を除いた研究者うち、「MCR/JP」を除いた研究者 (「MCR/exSS/exJP」と表記してもよいが、本稿では「MCR/exJP」と表記する)

4.2. 研究者の主研究分野の決定

- (8) 「MF (年)」: 発表年ごとに、分野ごとの HCP 数が最大である分野 (もし最大分野が複数ある場合、発表年ごと、分野ごとの HCP の論文被引用数合計を分野ごとの標準化した順位が最高である分野)
- (9) 「MF (全期間)」: 全期間での HCP 数が最大である分野 (もし最大分野が複数ある場合、全期間で、分野ごとの HCP の論文被引用数合計を分野ごとの標準化した順位が最高である分野)

4.3. 研究者の特徴量の算出

- (10) 「FV (研究者, 年)」: 研究者ごと、発行年ごとに、HCP が 1 件以上存在する分野の数の合計
- (11) 「FB (研究者, 年)」: 研究者ごと、発行年ごとに、HCP が 1 件以上存在する分野ごとの HCP の数の標準偏差
- (12) 「FV (研究者, 全期間)」: 研究者ごとに、全期間に HCP が 1 件以上存在する分野の数の合計
- (13) 「FB (研究者, 全期間)」: 研究者ごとに、全期間に HCP が 1 件以上存在する分野ごとの HCP の数の標準偏差
- (14) 「IC (研究者, 全期間)」: 研究者ごとに、全期間に所属した研究機関の合計

4.4. 主分野ごとの研究者の特徴量の算出

- (15) 「FV (MF, 年)」: MF (年) ごと、発行年ごとに、FV (研究者, 年) の平均
- (16) 「FB (MF, 年)」: MF (年) ごと、発行年ごとに、FB (研究者, 年) の平均
- (17) 「SF (MF, 全期間)」: MF (全期間) ごとに、研究分野が一つである研究者の比率
- (18) 「SI (MF, 全期間)」: MF (全期間) ごとに、所属機関が一つである研究者の比率
- (19) 「FV (MF, 全期間)」: MF (全期間) ごとに、FV (研究者, 全期間) の平均
- (20) 「FB (MF, 全期間)」: MF (全期間) ごとに、FB (研究者, 全期間) の平均

表1 MCR/exJP, SS/exJP, MCR/JP, SS/JP の学際度

区分	研究者数	研究者比率	MCR数	MCR比率	SS数	SS比率	FV平均	FB平均	IC平均	単一分野比率	単一機関比率
MCR/exJP	624,003	96.51%	624,003	97.03%			1.21	1.76	1.40	84%	71%
SS/exJP	9,557	2.94%			9,557	98.15%	3.84	10.16	5.26	11%	15%
MCR/JP	19,107	0.54%	19,107	2.97%			1.21	1.71	1.38	84%	75%
SS/JP	180	0.01%			180	1.85%	3.69	9.57	5.02	15%	10%
合計	652,847	100.00%	643,110	100.00%	9,737	100.00%	1.25	2.41	1.46	83%	70%

表2 MCRの分野ごとの学際度

分野	研究者数	研究者比率	FV平均	FB平均	IC平均	単一分野比率	単一機関比率
agricultural	13,885	2.1%	1.16	0.44	1.23	88%	81%
biology	29,936	4.6%	1.32	0.50	1.37	78%	74%
chemistry	36,732	5.6%	1.38	1.17	1.46	79%	75%
clinical	222,032	34.0%	1.16	1.58	1.49	89%	67%
computer	8,584	1.3%	1.22	0.74	1.30	85%	76%
ecology	13,443	2.1%	1.32	0.38	1.29	77%	79%
econbusiness	4,809	0.7%	1.11	0.54	1.30	91%	73%
engineering	26,581	4.1%	1.23	0.72	1.28	85%	81%
geosciences	19,651	3.0%	1.15	0.82	1.31	89%	78%
immunology	15,802	2.4%	1.32	0.64	1.42	77%	71%
materials	17,600	2.7%	1.54	0.56	1.42	66%	75%
mathematics	5,067	0.8%	1.17	0.93	1.29	88%	76%
microbiology	6,290	1.0%	1.35	0.41	1.33	77%	76%
molecularbiology	27,877	4.3%	1.45	0.79	1.54	71%	69%
multidisciplinary	17,862	2.7%	1.47	0.55	1.45	68%	71%
neuroscience	27,636	4.2%	1.25	0.98	1.46	81%	69%
pharmacology	14,718	2.3%	1.23	0.45	1.26	83%	78%
physics	54,699	8.4%	1.37	1.19	1.90	74%	60%
plantanimals	27,903	4.3%	1.18	0.75	1.31	87%	77%
psychiatry	14,690	2.3%	1.21	0.85	1.35	84%	68%
socialsciences	34,578	5.3%	1.19	0.56	1.30	85%	76%
space	12,472	1.9%	1.16	4.48	1.58	86%	63%
合計	652,847	100.0%	1.25	2.41	1.46	83%	70%

表3 MCR/exJP, SS/exJP, MCR/JP, SS/JP の分野ごとの学際度

区分	MCR/exJP						SS/exJP						MCR/JP						SS/JP					
	研究者数	FV平均	FB平均	IC平均	単一分野比率	単一機関比率	研究者数	FV平均	FB平均	IC平均	単一分野比率	単一機関比率	研究者数	FV平均	FB平均	IC平均	単一分野比率	単一機関比率	研究者数	FV平均	FB平均	IC平均	単一分野比率	単一機関比率
agricultural	13,496	1.14	0.26	1.20	89%	82%	211	2.50	2.44	2.79	36%	35%	178	1.16	0.19	1.19	88%	85%	0					
biology	28,699	1.28	0.35	1.33	79%	75%	409	3.99	3.01	4.15	11%	11%	824	1.27	0.31	1.30	78%	77%	4	3.75	3.49	2.25	0%	0%
chemistry	34,457	1.29	0.80	1.31	80%	77%	710	6.09	4.74	8.26	3%	17%	1,532	1.26	0.98	1.42	81%	72%	33	4.82	5.70	7.21	12%	6%
clinical	213,694	1.14	1.22	1.45	89%	67%	1,686	4.17	7.37	7.03	14%	1%	6,637	1.13	1.11	1.27	89%	81%	15	4.00	4.49	4.07	0%	20%
computer	8,272	1.18	0.39	1.23	86%	78%	179	3.36	3.27	4.15	13%	2%	132	1.15	0.67	1.33	85%	74%	1	1.00		2.00	100%	0%
ecology	13,026	1.28	0.25	1.26	79%	80%	253	3.46	1.92	2.77	6%	33%	159	1.31	0.14	1.27	73%	79%	5	4.40	1.74	3.60	0%	0%
econbusiness	4,690	1.19	0.34	1.27	92%	74%	95	1.99	2.20	2.78	55%	23%	24	1.33	0.80	1.71	79%	63%	0					
engineering	25,517	1.19	0.43	1.22	86%	82%	355	4.31	4.12	5.57	11%	71%	702	1.19	0.33	1.23	86%	84%	7	1.86	4.60	2.86	57%	14%
geosciences	18,825	1.12	0.56	1.28	90%	78%	267	2.77	3.53	3.27	32%	27%	554	1.14	0.54	1.33	89%	75%	5	2.20	1.66	3.60	60%	20%
immunology	15,049	1.28	0.43	1.38	78%	72%	275	3.59	3.18	3.57	4%	17%	466	1.26	0.44	1.35	79%	75%	12	4.33	2.78	4.25	0%	8%
materials	16,476	1.47	0.39	1.33	67%	76%	325	5.43	3.34	5.73	0%	16%	783	1.36	0.47	1.47	74%	70%	16	4.31	3.79	5.88	6%	0%
mathematics	4,888	1.14	0.46	1.26	89%	76%	100	2.57	4.66	2.87	34%	32%	76	1.17	0.61	1.25	91%	83%	3	1.33	1.12	2.67	67%	33%
microbiology	5,993	1.27	0.22	1.29	79%	77%	202	3.79	4.64	2.71	6%	34%	94	1.26	0.17	1.21	80%	85%	1	1.00	3.79	7.00	0%	0%
molecularbiology	26,195	1.38	0.52	1.46	73%	70%	607	4.48	3.97	5.45	0%	12%	1,066	1.34	0.47	1.39	76%	73%	9	5.67	3.20	5.00	0%	0%
multidisciplinary	16,947	1.40	0.34	1.38	71%	72%	435	4.05	3.16	4.39	2%	15%	476	1.44	0.25	1.41	67%	70%	4	3.75	1.51	6.00	0%	0%
neuroscience	26,666	1.22	0.70	1.42	82%	69%	369	3.54	4.84	3.91	4%	19%	601	1.22	0.41	1.41	83%	73%	0					
pharmacology	14,023	1.22	0.31	1.25	83%	78%	239	2.10	3.60	2.73	58%	21%	454	1.18	0.51	1.20	86%	83%	2	3.00	3.54	1.50	0%	50%
physics	50,458	1.31	0.47	1.76	76%	64%	1,697	3.52	32.31	6.18	1%	2%	2,515	1.24	2.63	1.75	82%	60%	29	3.52	30.77	6.48	0%	10%
plantanimals	26,345	1.15	0.50	1.28	88%	78%	407	2.86	3.13	3.01	26%	29%	1,122	1.15	0.77	1.33	88%	76%	29	2.17	6.41	3.28	41%	17%
psychiatry	14,401	1.19	0.57	1.32	85%	69%	210	2.94	4.18	3.47	12%	21%	78	1.22	0.51	1.51	85%	65%	1	4.00	3.37	2.00	0%	0%
socialsciences	33,939	1.17	0.39	1.27	86%	76%	381	2.98	3.01	3.35	12%	23%	258	1.28	0.31	1.24	78%	79%	0					
space	11,947	1.15	3.14	1.54	87%	66%	145	2.15	20.55	6.56	6%	5%	376	1.24	1.81	1.63	82%	61%	4	3.50	12.26	5.00	0%	0%
合計	624,003	1.21	1.76	1.40	84%	71%	9,557	3.84	10.16	5.26	11%	15%	19,107	1.21	1.71	1.38	84%	75%	180	3.69	9.57	5.02	15%	10%

(21) 「IC (MF, 全期間)」: MF (全期間) ごとに, IC (研究者, 全期間) の平均

5. 結果

前節の手法で分析した結果, 2008年から2016年までの分析対象期間の論文数 19,536,453 件において, MCR/exJP が 624,003 名, SS/exJP が 9,557 名, MCR/JP

が 19,107 名, 及び SS/JP が 180 名, 合計 652,847 名であった. また, MCR/exJP と MCR/JP では clinical が全体の約 1/3 以上を占めている. これらの結果を表 1, 表 2, 表 3, 図 1 に示し, MCR/exJP, SS/exJP, MCR/JP, SS/JP の特徴量を比較分析した結果について以下に述べる.

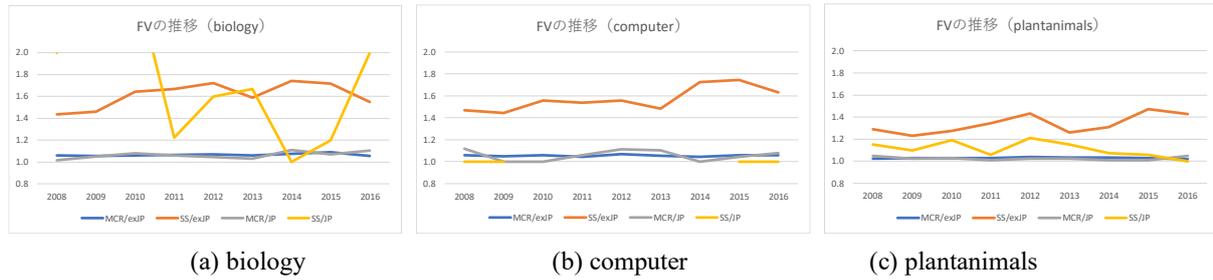


図 1 MCR/exJP, SS/exJP, MCR/JP, SS/JP の学際度の時間推移

5.1. MCR/exJP の特徴

FV (MF, 全期間) は, materials が 1.47 で他の分野に比較して高い。他に FV (MF, 全期間) が比較的高い分野として, multidisciplinary : 1.40, molecularbiology : 1.38, physics : 1.31, chemistry : 1.29, biology : 1.28, ecology : 1.28, immunology : 1.28, microbiology : 1.27 等があり, 生命科学系分野の FV が高い。一方, 事前に応用範囲が広いと考えた computer は 1.18 と比較的低かった。

FB (MF, 全期間) は, physics : 9.47, space : 3.14, clinical : 1.22 を除き, 全分野で 1.0 未満であった。

また, IC (MF, 全期間) は, physics : 1.76, space : 1.54 が比較的高かった。

5.2. SS/exJP の MCR/exJP に対する特徴

SS/exJP の FV (MF, 全期間), FB (MF, 全期間), IC (MF, 全期間) は, 全ての分野において, MCR/exJP のそれらより高い。

また, MCR/exJP の SF (MF, 全期間) 及び SI (MF, 全期間) はそれぞれ 84%, 71%であり, 多くの研究者が単一機関で単一分野の研究をしているのに対して, SS/exJP のそれらは 11%, 15%であり, 逆に多くの研究者が複数機関や複数分野での研究経験を持っている。

5.3. MCR/JP の MCR/exJP に対する特徴

MCR/JP の数は MCR 全体の 2.97%にしか過ぎないが, physics, materials, chemistry, molecularbiology, plantanimals 等は, 他の分野に比較して高い比率を示している。

一方, MCR/JP の FV (MF, 全期間), FB (MF, 全期間), IC (MF, 全期間) や, SF (MF, 全期間) 及び SI (MF, 全期間) は, MCR/exJP のそれらとで大きな差異は見られない。

5.4. SS/JP の SS/exJP に対する特徴

SS/JP は全 22 分野中で 18 分野でしか抽出できず, SS/JP の数も SS 全体の 1.85%にしか過ぎないが, plantanimals, materials, chemistry, immunology, 等は, 他の分野より高い比率を示している。SS/JP の数も SS 全体に対する比率 (1.85%) は MCR/JP の MCR 全体に対する比率 (2.97%) より低くなっており, SS/JP はより希少な存在となっている。

5.5. 学際性特徴量の時間推移

図 1 に, 分野の事例として, biology と computer, 及び SS/JP の数が多い分野の一つである plantanimals の MCR/exJP, SS/exJP, MCR/JP 及び SS/JP の FV (MF, 年) の時間推移を示す。

FV (MF, 年) は, biology, computer よりも, plantanimals が低く, 分野によって異なるが, いずれの分野の場合も SS/exJP の FV (MF, 年) が最も高くなっている。

6. 考察

本節では, MCR や SS, 取り分け日本の MCR や SS の特徴について考察するとともに, 科学技術政策及び技術経営戦略に対する含意について考察する。

6.1. MCR と SS の特徴

本稿の分析結果, SS の FV や IC は MCR のそれらよりも高いことが分かった。SS の数は MCR の数の 1.5%程度に過ぎず, MCR の多くも十分な研究経験を持っているが, 研究成果が高くインパクトを与えている MCR の中でも, 取り分け著しくインパクトを与えている SS は, 多様な分野の研究経験と複数の機関への所属経験を持っていることが分かった。

前述の通り, イノベーションの実現に向け「学際的・分野融合的な研究充実」が謳われているが, これらの分野融合的な取り組みを積極的に推進することは有効であるといえよう。

6.2. 日本の MCR や SS の特徴

本稿で分析対象とした 22 分野の中で、SS/JP が存在する分野は、生物学系分野を中心とした 18 分野のみである。しかも、例えば、日本政府が推進している AI 等の分野は僅かしか含まれていない。また、MCR/JP の数は MCR の 2.97% と少ないが、SS/JP の数は SS の 1.85% にしか過ぎず、その比率は更に低く、SS/JP はより希少な存在となっている。

今後高齢化が進み財政的制約が拡大する中でムーンショット型研究開発等、特定分野への集中化する政策が打ち出されているが、イノベーションの実現に向けては多様な分野の学際的・分野融融合的研究も有効であり、研究分野の裾野を拡大するとともに、研究者間の連携を推進することも必要と考えられる。更に、非常に少ないスター研究者の育成も課題であるが、所属機関の移動や学際的研究の推進は、この課題解決の手段の一つになるかもしれない。

6.3. 研究分野ごとの学際性の特徴

分野ごとの学際性について、本稿の分析結果では、molecularbiology, microbiology, biology, immunology 等の生物学系分野の学際性が他の分野に比較して高かった。また、応用範囲の広い computer の学際性が比較的高いと事前に想定したが、分析結果はむしろ逆であった。本稿では、分析対象とする学術文献データベースとして WoS を用いたが、WoS では一つの学術文献が複数の分野に振分されることがあり、上記の生物学系分野の学際性の高さは、この分野振分方法に起因する可能性もある。

6.4. 学際性分析の課題

本稿の分析を通じて、学際性の特徴として、以下のような事項も分かった。その要因等については明らかにできていない事項もあり、今後更に検討・分析を行う予定である。

6.4.1. 学際特徴量の時間推移

5.5 項において、biology, computer, plantanimals の FV (MF, 年) の時間推移を示したが、分野によって FV の大きさやその時間推移は異なっていた。学際性は時間とともに上昇していると事前に想定したが、これら 3 つの分野では時間推移に伴う FV の上昇を確認することができなかった、これらの時間推移に伴う特徴量の変化については今後更に分析を進める予定である。

6.4.2. 研究分野の均等性

5.1 項において、MCR/exJP の physics と space は、FV (MF, 全期間) が他分野に比較して高くないにもかかわらず、FB (MF, 全期間) が他の分野に比較し

て突出して高い。physics や space については大規模な設備を使い、学術文献の共著者数も多数になると考えられるが、その因果関係等について更に分析を進める予定である。

6.4.3. 研究分野の差異性

Rafols, et al. (2010)は、学際性を表す特徴量として、分野の多様性 (Variety)、均等性 (Balance)、差異性 (Disparity) を示している。本稿では WoS の分野区分をそのまま活用し、多様性 (FV) と均等性 (FB) を分析した。分野間の差異性については考慮しておらず、分野の区分とその差異性について今後検討する予定である。

7. おわりに

本稿では、WoS から抽出された MCR 及び SS について、研究分野の多様性や研究機関の移動頻度などの特徴量を定義・活用して分析した。その結果、① SS は MCR に比較して多様な分野や複数の機関での研究経験を持っていること、② SS/JP は SS/exJP に比較して研究分野が偏っており MCR に対する SS 比率も低いこと、等が分かった。したがって、日本におけるイノベーションの実現に向けて、研究分野の裾野を拡大するとともに、数少ない日本スター研究者の育成が望まれる。そのためには、所属機関の移動や学際的研究の推進が有効な手段の一つになるかもしれない。

今後、学際研究における研究分野の均質性や差異性等の学際性評価指標による分析、などを行いつつ、学際性のイノベーションに与える影響について分析を進める予定である。

謝辞

本研究は JST-RISTEX 政策のための科学「スター・サイエンティストと日本のイノベーション」の助成を受けたものである。

参考文献

- [1] Schumpeter, J. A. (1926) Theorie der Wirtschaftlichen Entwicklung
- [2] 内閣府. (2016). 第 5 期科学技術基本計画
- [3] 文部科学省 (2017). 文部科学省における研究及び開発に関する評価指針.
- [4] Zucker, L. G., Darby, M. R., and Armstrong, J. S. (2002). Commercializing knowledge: University science, knowledge capture, and firm performance in biotechnology. *Management science*, 48(1):138-153.
- [5] Zucker, L. G. and Darby, M. R. (2007). Virtuous circles in

science and commerce. *Papers in Regional Science*, 86(3):445–470.

- [6] Rafols, I., Meyer, A. M. (2010). Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience *Scientometrics* 82:263–287
- [7] 牧兼充, 菅井内音, 隅蔵康一, 原泰史, 長根裕美 (近刊) . スター・サイエンティストの検出とコホート・データセットの構築