

Title	日本の生命科学・医学分野のスター研究者におけるサイエンス・リンケージ分析
Author(s)	福澤, 尚美; 依田, 高典
Citation	年次学術大会講演要旨集, 29: 252-255
Issue Date	2014-10-18
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/12439
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

日本の生命科学・医学分野のスター研究者におけるサイエンス・リンケージ分析

○福澤尚美 (科学技術・学術政策研究所)、依田高典 (京都大学)

日本の生命科学・医学系分野のトップ研究者 100 名が、1996 年から 2009 年までに **Corresponding author** として発表した 4,763 本の学術論文を対象として、論文の論文、特許への被引用数についての分布形状や、論文-論文引用数が論文-特許引用数とどのような関係があるのかについて分析した結果を報告する。分析の結果、論文-論文引用のピークの期待値は 4 年後、特許へは 6 年後であり、特許に実用化されるまでにはある程度時間を要するが、その年数は近年の論文になるに従い短くなることが分かった。また、論文の質は特許に引用される際に重要であり、論文の質が高くなるほど特許に引用されることが分かった。さらに、論文の質と研究費には逆 U 字の関係、論文の質と量には U 字の関係があることが明らかになった。

1. はじめに

産業において研究開発や技術発展を促進するうえで学術界の知識を如何に取り入れていくかは重要である。Mansfield (1991)は、「企業の製品やプロセス(1975-85 年)で、近年の学術研究なしでは生み出されなかったのは何%か」という質問を実施し、アメリカの主要な企業の新製品の平均で 11%、新プロセスの 9%は開発されなかったとした。また、Zucker and Durby (2001)は日本における大学のスター研究者と企業のコラボレーションは、企業の研究生産性に大きな正のインパクトを与えることを示した。

また、Sanberg, et al., (2014)は、研究大学は技術移転において中心的な役割を求められており、大学でのテニユアや昇進においても特許化やライセンス等の活動についても評価すべきであると指摘している。しかしながら、基礎的な研究から産業への知識・技術移転を測ることは最もチャレンジングな研究の 1 つであり (Tijssen et al. 2000)、学術的研究と産業的技術にはどのような相互関係があるのかについては、多くの研究があるものの、その実態は未だ十分に明らかにされているとは言えない。

学術研究と産業技術の関係について分析する指標として、特許が申請書内で引用している特許以外の文献である **Non-patent references (NPRs)** が使用されてきており、“サイエンス・リンケージ”とも呼ばれる。NPRs は科学と技術間の関係の強さ“science-intensity”を測る指標である (Meyer, 2000; Tijssen et al. 2000)。先行研究では、特にバイオテクノロジー分野において科学と産

業との関係が強いことが示されてきた (Anderson, et al. 1996; Narin, et al. 1997; MacMillian, et al. 2000 等)。

これらの先行研究で、学術界での知識が産業技術の代理変数としての特許にとって重要な役割を果たしていることが明らかになったが、産業技術により多く引用される論文は、果たして学術界でもインパクトの強い、質の高い論文なのだろうか。先駆的な先行研究として、Tijssen et al. (2000)や Hicks et al. (2000)は論文への引用数が多い論文は特許にも引用される傾向があることを示したが、対象とする期間が短いことや相関をみる段階にとどまっている。さらに、これらの先行研究は特許を元にした後方引用のアプローチを使用しており、セレクションバイアスの問題が生じる可能性がある。

本研究では、論文への引用数が特許への引用数とどのような関係を持つのかを、各研究者の論文数や研究費をモデルに組み込んで分析した。さらに、学術界、産業界への引用のタイムラグを同時に分析した。その際には、学術論文を元にした前方引用のアプローチを使用し、日本の生命科学、医学系分野のトップ研究者 100 名を対象として、学術論文から論文への引用数、特許への引用数を抽出した。

2. 仮説

本研究で明らかにしたい仮説は 3 つある。

H1: 論文の論文への引用ラグは特許への引用ラグより短い。

論文が論文や特許に最も多く引用されるピー

クが訪れるのは何年であるのかを明らかにし、論文が発表されてからインパクトを与えるまでの分布形状を明らかにし、その形状には違いがあるのかについて分析する。Bouabid and Larivière (2013)、Finardi (2014)によると論文-論文間のタイムラグは2-3年後であり、Mansfield (1991)、Van Vianen et al. (1990)、Verbeek, et al. (2002)、Lo (2010)、Finardi (2011)によると論文-特許間のタイムラグは3年から10年後という傾向が示されているため、論文-論文間のラグの方が短いことが予想される。

H2: 学術的に質が高い論文ほど、特許に引用される。

Tijssen et al. (2000)、Hicks et al. (2000)が論文への被引用数と特許への被引用数には相関があるとしたことから、学术界で質が高い、インパクトがある論文ほど、特許にも引用されることを検証する。

H3: 論文の質と量にはトレードオフがあり、論文の質を最大にする研究費額がある。

論文数を多く出すこととその質の高さにはトレードオフがあるのかを検証する。また、論文の質を最も高くできる研究費額のレベルがあるのかについて検証する。

3. データ

本研究では、2002年度から文部科学省の研究拠点形成費等補助金として措置された21世紀COEプログラムの生命科学と医学系分野に採択された研究者から、Scopusを使用してデータを作成した。対象分野の研究者のうち論文データが存在した研究者は1,232名であり、これら研究者が1996年から2009年までに発表した論文の総被引用数合計が多い順から100名の研究者を抽出した。この100名の研究者が13年間に発表した全論文の中で、Journal articleとして発表した20,661本の論文から、各研究者がCorresponding authorとなっている論文4,763本を抽出し、これらの論文を分析の対象としている。なお、Corresponding authorはSupervision、Principal Investigatorsとしての位置づけである(Wren et al. 2007)。

これら4,763本の論文が、1996年から2012年までの16年間に、何年に発表された論文、特許に引用されているのかというデータを、論文1本ずつに対して作成した。特許情報については、European Patent Officeが作成しているEspacenet Patent searchを使用しており、特許への引用は特許文書内にその論文が引用されて

いる場合にカウントされる。年次は出願公開された年次(publication date)を使用した。

さらに、研究者の研究費情報として、科学研究費助成事業データベース(KAKEN)を使用し、各研究者が代表者となっている研究課題のみを対象とし、1996年から2009年の各年度の直接経費を使用した。

4. 分析結果

引用の分布形状と特許への引用数との関係についての分析手法と結果を以下に示す。

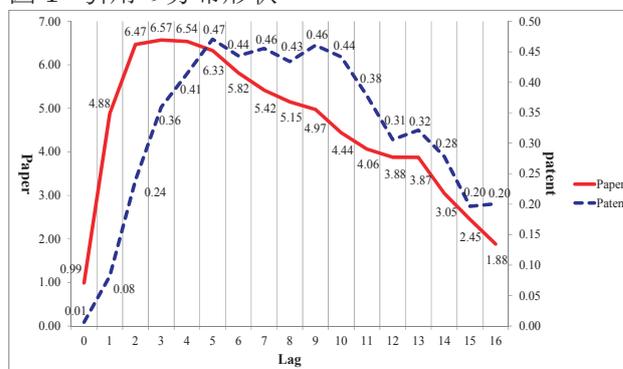
4.1. 引用分布形状

論文を発表した年を a 、引用された年を b し、 $a \leq b$ である。引用されるまでのラグは $l = b - a$ であり、 l は0から16までの値をとる。ここで、 $CITATION_{a,b}$ を a 年の論文が b 年に引用された回数、 $ARTICLES_a$ を a 年の論文数とする。この時、 $MEAN_{b-a=l}$ を a 年の論文が l 年のラグで b 年に引用される論文1本あたりの平均引用とすると、(1)式のように与えられる

$$MEAN_{b-a=l} = \frac{\sum_{b-a=l} CITATION_{a,b}}{\sum_{b-a=l} ARTICLES_a} \quad (1)$$

図1にラグ l に対する $MEAN_{b-a=l}$ をプロットした。

図1: 引用の分布形状



論文への引用については発表されてから2年から5年が最も多く、3年目に1本あたり最大6.57回引用され、その後緩やかに減少していく。特に発表してからの2年間で急激に引用が増加することから、学术界へはすぐにインパクトを与えることが分かる。一方、特許への引用については論文が発表されてから5年から10年で最も多くの引用を得ており、5年目に1本あたり最大0.47回引用され、論文への引用と比較すると、引用され続ける期間が長い特徴がみられる。特許に実用化されるまでには論文へと比べるとある程度の年数

を要することが見てとれる。

さらに、各論文の引用のピークが得られるラグの期待値が論文と特許で異なるのかを検証した。まず、論文や特許に1回以上引用された論文のみを対象とし、引用数がピークとなるラグを各論文について求める。その際、各論文の最大引用数のピークが2回以上生じる場合には、それぞれを別々にカウントする。ラグ l を離散確率変数とし、確率分布を $P(X=l)$ とすると、引用ピーク時のラグの期待値は以下で計算される。

$$E(X) = \sum_l \{l \times P(X=l)\} \quad (2)$$

また、各論文の引用数のピークとなるラグの分布が正規分布に従っているかをKolmogorov-Smirnov検定で検定した結果、正規分布に従っているという帰無仮説が1%有意水準で棄却されたため、検定はt検定ではなく、ノンパラメトリックなWilcoxon rank-sum testを使用して中央値に差があるかを検定し、特許へのラグが論文へのラグよりも大きくなる確率を求めた。

その結果、論文、特許への引用がピークになるラグの期待値はそれぞれ、4.31、6である。中央値は論文へは4年後、特許へは5年後である。Wilcoxon rank-sum testの結果をみると、中央値に差が無いという帰無仮説が有意水準1%で棄却されるため、特許へ引用されるラグのピークの方が遅いことが分かった。さらに、特許への引用ラグの方が論文への引用ラグより長い確率は0.664であることから、特許に引用されるまでの方が論文に引用されるまでよりも時間を要することが分かる。

さらに、論文発表年を3年ごと(96-98年、99-01年、02-04年、05-07年)に区切って分析した。その結果、05-07年を除いて、論文への引用ラグの中央値は一貫して4年であるものの、特許への引用ラグの中央値は7年、6年、5年、3年と短くなる傾向がみられた。これらのことから、学术界へ与えるインパクトの速さは論文発表年次で大きな違いがないものの、学術論文が技術に活用されるまでの期間は、年次の新しい論文の方が短い傾向がみられ、近年のオープンイノベーションや産学連携への取組が、技術移転に要する期間が短くなることの背景の1つとして考えられる。

4.2. 特許-特許間引用数との関係

論文-論文間引用数と論文-特許間引用数の関係を分析するため、被説明変数には特許への被引用数、説明変数には論文への被引用数を使用して推定を行った。特許への被引用数は非負の値を取るため分析にはトービットモデルを使用した。説明

変数は内生変数である可能性が高いため、操作変数として各研究者の総研究費額と総論文数を使用してNewey(1987)の二段階推定で行った。

以下のモデルを推定する。 i は各研究者、 p は各論文とする。

$$\begin{aligned} y_{ip} &= \beta x_{ip} + u_{ip} \\ x_{ip} &= \pi z_i + v_{ip} \\ y_{ip} &= \begin{cases} y_{ip}^* & \text{if } y_{ip}^* > 0 \\ 0 & \text{if } y_{ip}^* \leq 0 \end{cases} \end{aligned} \quad (3)$$

y_{ip} は各研究者の各論文の特許への被引用数(y_{ip}^* は潜在変数である)、 x_{ip} は各研究者の各論文の論文への被引用数である。 z_i は操作変数である。研究費や論文数は論文被引用数と非線形の関係性があることを考慮し、それぞれの二次項も使用した(Model 1)。また同様に、論文への被引用数と特許への被引用数にも非線形の関係がみられることを考慮し、論文への被引用数の二次項を内生変数として入れたモデルも推定した(Model 2)。

第2段階推定の結果から、Model 1では係数は0.154で1%水準で統計的に有意であった。Model 2では1次項の係数は0.168で1%水準で統計的に有意だったが、2次項の係数は-0.0000162で非有意であった。よって、論文への被引用数は特許への被引用数と正に有意な関係があり、学术界で質の高い論文は特許にもより多く引用されることが分かった。中央値で評価した限界効果は0.036であり、論文への引用数が1回増加することにより特許への引用数が0.036回増加することが分かった。

次に、操作変数と論文被引用数との関係を見ると、各研究者の総額研究費は一次項では有意に正、二次項では有意に負に関係しており逆U字の関係がみられた。一方、各研究者の総論文数は一次項では有意に負、二次項では有意に正であり、U字の関係がみられた。

論文への被引用数に対して、研究費は年間1億8988.6万円(13年間で24億6851.4万円)の場合に最も論文の質が高くなる。それを超えると徐々に下降していく。一方、論文数は年間25.2本(13年間で327.0本)までは、量産するほど論文の質が落ちていき、それを超えると質が高くなることが分かった。なお、鞍点(z_1, z_2)=(246851.4, 327.0)での引用数は68.0である。

分析の結果を要約すると、質の高い論文を発表するには効率がよい研究費額や論文数があり、論文の質と量には年間約25本まではトレードオフ

があることが分かった。研究費については年間約2億円を超えるとマネジメントが上手くいかず効率が低下することが考えられる。

5. 結論

検証した仮説に対して以下の結果が得られた。

H1: 論文の論文への引用ラグの期待値4年(中央値4年)は特許への引用ラグの期待値6年(中央値5年)より有意に短く、特許への引用ラグは年次が新しい論文の方が短い傾向がみられる。

H2: 学術的に質が高い論文ほど、特許に引用され、論文への引用数が1回増加すると特許への引用数が0.036回増加する。

H3: 論文の質と量はU字の関係がみられ、ある一定の論文数まではトレードオフがある。論文の質と研究費には逆U字の関係がみられ、論文の質を最大にする研究費額がある。

ただし、本分析で得られた論文の質を最大にする研究費や論文数の水準は、あくまで日本の生命科学・医学系分野におけるトップ研究者を対象にした分析であり、各研究者がPIとして関わった論文を対象とした結果である。よって、この結果が他の分野や日本の研究者全体に必ずしも適用される訳ではない点に留意は必要である。研究分野やラボの環境によってはより多くの研究費が必要な場合や、より少額でよい場合が十分考えられるため、全ての研究者に最適額を配分すれば生命・医学分野のあらゆる分野において質が高い研究が得られるというわけではない。

参考文献

1. Anderson J, Williams N, Seemungal D, Narin F, Olivastro D (1996) Human genetic technology: Exploring the links between science and innovation. *Technology Analysis and Strategic Management* 8(2): 135-156.
2. Bouabid H. and Larivière V. (2013) The lengthening of paper's life expectancy: A diachronous analysis. *Scientometrics* 97(3): 695-717.
3. Finardi U (2011) Time relations between scientific production and patenting knowledge: The case of nanotechnologies. *Scientometrics* 89(1): 37-50.
4. Finardi U (2014) On the Time evolution of received citations, in different scientific fields: An empirical study. *Journal of Informetrics* 8(1): 13-24.
5. Hicks D, Breitzman A, Kimberly S, Narin F (2000) Research excellence and patented innovation. *Science and Public Policy* 27(5): 310-320.
6. Mansfiels E (1991) Academic research and industrial innovation. *Research Policy* 20(1): 1-12.
7. McMillian G, Narin F, Deeds D (2000) An analysis of the critical role of public science in innovation: the case of biotechnology. *Research Policy* 29(1): 1-8.
8. Meyer M (2000) Does science push technology? Patents citing scientific literature. *Research Policy* 29(3): 409-434.
9. Narin F, Hamilton K, Olivastro D (1997) The increasing linkage between U.S. technology and public science. *Research Policy* 26(3): 317-330.
10. Newey W (1987) Efficient estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics* 36(3): 231-250.
11. Sanberg P, Gharib M, Harker P, Kaler E, Marchase R, Sands T, Arshadi N, Sarkar S (2014) Changing the academic culture: Valuing patents and commercialization toward tenure and career advancement. *Proceedings of the National Academy of Sciences of the United States of America* 111(18): 6542-6547.
12. Tijssen R, Buter R, Van Leeuwen Th (2000) Technological relevance of science: An assessment of citation linkages between patents and research papers. *Scientometrics* 47(2): 389-412.
13. Van Vianen B, Moed H Van Raan A (1990) An exploration of the science base of recent technology. *Research Policy* 19(1): 61-81.
14. Verbeek A, Debackere K, Marc L, Andries P, Zimmermann E, Deleus F (2002) Linking science to technology: Using bibliographic references in patents to build linkage schemes. *Scientometrics* 54(2): 399-420.
15. Wren J, Kozak K, Johnson K, Deakynne S, Schilling L, Dellavalle R (2007) A survey of perceived contributions to papers based on byline position and number of authors. *EMBO reports* 8(11): 988-991.
16. Zucker L, Durby M (2001) Capturing technological opportunity via Japan's star scientists: Evidence from Japanese firm's biotech patents and products. *Journal of Technology Transfer* 26(1-2): 37-58.