

Title	米国特許が引用する学術論文の計量書誌学的分析
Author(s)	調, 麻佐志
Citation	年次学術大会講演要旨集, 27: 746-749
Issue Date	2012-10-27
Type	Conference Paper
Text version	publisher
URL	<a href="http://hdl.handle.net/10119/11129">http://hdl.handle.net/10119/11129</a>
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

## 米国特許が引用する学術論文の計量書誌学的分析

○調 麻佐志（東京工業大学）

## はじめに

非特許引用 (Non-patent references; 以下、NPRs と略)、とりわけ学術文献に対する引用 (以下、science linkage と呼ぶ) については、技術と科学 (あるいは産業技術と学術研究) の連関を分析する有力な手がかりとして、1980 年台より米国 CHI 社 (当時) が調査分析を進めており、1990 年台以降急速に研究が進んできた (e. g., Narin & Olivastro (1992))。その中でも重要なのが Narin et al. (1997) であり、当該論文は science linkage のデータから、米国の特許が学術研究の成果である論文を引用する割合が高まり、かつはその学術研究の多くが公的資金の提供を受けていることなどを明らかにした。この結果から即座に linear model 的結論を導くことはすべきでないが、何らかの意味で産業技術と学術研究のリンクが強化されていることは否定しがたく、科学技術イノベーション政策にとって重要な示唆を与えた。さらに、この成果により science linkage データの分析の重要性が高らかに示されたともいえるかもしれない。

我が国でも、science linkage の分析は様々に行われてきた。たとえば、玉田 (2010) は日本および米国の特許データから無作為抽出したデータを使って、日米の science linkage の特徴を明らかにするとともに、産学連携によるイノベーションの分析を行なっている。本研究で分析を行う特許に引用された論文を文献データベースに紐付けした研究としては、(他の特許に引用される件数が多い) 有力特許 top500 に引用された学術論文の分析 (富澤・他 (2005)) もあり、他国と比較しても我が国では大学セクター所属の著者による論文の引用が多いことなどが示された。

一方、近年の関連する政策・研究動向として、SciSIP あるいは「科学技術イノベーション政策のための科学」などと呼ばれる政策および学術研究が有機的に結合したムーブメントが台頭したことは重要である。2005 年にワシントンで開催された AAAS のフォーラムの基調講演で、当時の科学技術政策局局長であり大統領科学顧問であった Marburger 氏は、「歴史的にみても巨額の研究開発資金が現状において投入されていることは適切であるか」あるいは「防衛や福祉、経済のために必要な研究開発領域へ投資がなされているか」といった疑問に答える信頼できる方法が必要なことを強調した。さらに、様々な指標を使ったベンチマーキングが政策形成には役に立たず、経済政策担当者が持つ計量ツールのようなものが必要であると指摘し、その方向に関連領域の研究者が集中することを求めた。これをきっかけに米国では SciSIP プログラムがスタートし、呼応するかの如く我が国を含め先進国で科学技術政策の効果を事前・事後に計量する手法の研究が官学一体となって進められるようになった。このような動向の中で、science linkage は重要なトピック一つとして改めて注目されている。なぜなら、それが学術研究への投資効果の一部 (≒ 産業技術への寄与) を評価する手がかりの一つとみなされるからである。

ただ、残念なことに science linkage の分析には弱点がある。理論上ないしデータそのものが示す連関の意味、内実において science linkage には様々な限界があることは調・他 (2007) で指摘されている。しかし、そこで指摘されるような限界は確かに弱点であるものの、畢竟許容される解釈のレベルの問題にしか過ぎない。より深刻なのは分析に必要なデータの入手可能性にある。

本研究で取り扱う米国で登録された実用特許 (utility patent) は USPTO のウェブページより無償でダウンロードできる。その中から NPR に相当する箇所を抽出することは容易であるが、そこからさらに science linkage を同定するには非常な手間となる作業が求められる。なぜなら、① NPRs として特許に記載されている文献には様々なものがあり、学術文献だけではなく一般文献、パンフレット、製品仕様書から海外に登録された特許の要旨や翻訳などが含まれている。しかも、② 各項は自由書式かつ index 化されていないため、その中から学術文献だけをすぐに取り出すことはできないからである。さらに、学術文献を同定するには、何が学術文献であるかを定義する必要があり、そのことを一定程度保証された精度で実現する唯一の方法は、③ 既存の学術文献データベース (以下、DB) に収録された論文と照合

することと考えられる。それには、非常な手間となるかつ不定形の照合の技術と作業が要求される。

この種の問題を回避する方法が無いわけではない。すなわち、照合済みのデータを購入するという選択がある。実際、筆者の知る限りでも、学術文献 DB の提供元である Elsevier 社と Thomson Reuters 社、および初期から science linkage の調査を行なっている The Patent Board 社 (旧 CHI 社) からデータを入手できる。

その際に問題となるのが、価格と付随して発生する制約である (もう一つ重要な問題として照合の精度が明らかでないことも指摘できる)。相対的に安価な道は、「Web ベースのインターフェースを通じて DB にアクセスする権利」を購入することであるが、この場合、特定の特許が引用する論文、あるいはその逆を確認できても、計量書誌学的分析を実施するのに必要なデータセットを作るのには手間がかかる。しかも、仮に期間を限定しても、特許全体を対象とした分析を行うことはできない。せめていくつかの統計量を推計するために無作為抽出を行うにも、サンプル数に比例した手数がかかる。かといって、(極小さなサブセットでない限り) データを購入するのは、研究予算の観点から通常まず許されない。

いずれの経路をたどるにせよ、結果として通常、研究者が実施できる science linkage の分析は、特定範囲の特許とそれが引用する文献の枠内に留まらざるを得ない。もちろん、そこから得られる知見が重要なことは否定できないものの、大域的な統計情報との比較を伴わないことの弊害がある研究テーマも様々に考えられる。多様な技術に応用される汎用性の高い学術研究の同定などもその一例であろう。

このような状況を解決するのは、すでに述べた NPRs と学術文献 DB に収録された論文の照合を自動化することである。富澤 (2008) は、米国特許データと Scopus (および PubMed) 収録論文との照合を行い、およそ 60% の達成率を得たと判断しており、欧州委員会 (2002) が米国特許データで行った結果の 39% より高い結果を得ている。これらの達成率は必ずしも十分とはいえないものの、その値以上に照合精度が適切に計測されていないことに深刻な問題がある。しかし、欧州委員会が指摘するように精度を適切に計測しようとする行為自体が矛盾を抱えている。すなわち、その照合精度の測定を行うには照合結果についての正解が必要であるが、正解があればそもそも照合という作業自体が必要ないのである。この矛盾を直接解決することは困難であるが、それでも (限定的な意味ながら) 精度を推測することは可能である。すなわち、NPRs を無作為抽出し、極めて高い精度で照合を行う (すなわち、人手に頼って照合する) ことにより、「正解の部分集合」を作り、機械的な照合結果と比較することで、ある意味で精度を推測することができるのである。

無作為抽出によって得られたサブセットについて「正確」な照合を行うことには、精度の推計に必要なベースを提供するのに加えて、もう一つメリットがある。すなわち、サブセットの性質から母集団の性質が推測できる。本研究で分析するデータは、米国特許データ全体を対象とする照合アルゴリズム開発を目的として作成したものであるが、本研究報告では中間段階の成果報告として、サブセットの照合の方法および同定結果から得られた幾つかの性質を紹介する。

## データ

本研究で用いる特許データは、2000 年から 2009 年に登録された米国の実用特許で、そのフロントページから NPRs のみを抽出した。一方、照合を行う文献データベースとしては、Web 版の WoS (SCIEx の 1965~) を利用した。さらに、同定された論文はその ID 番号を活用して WoS のデータセット (1992-2011 年収録) とリンクし、書誌情報を付与した。なお、母集団となった実用特許の件数および NPR の件数は次表の通りである。ちなみに、Narin et al. (1997) によると、1993~94 年の登録特許では特許一本当たり 1.5 件の NPR があり、10 年で 2 倍強に増加していることがわかる。

表1 米国の実用特許登録件数および NPR 件数の推移

登録年	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
登録件数	157496	166038	163518	169035	164291	143806	173770	157283	157772	167349
NPR 件数	466056	519743	549741	585150	557524	557780	851232	868929	936926	1139407
特許一本当たりの NPR 数	2.96	3.13	3.36	3.46	3.39	3.88	4.90	5.52	5.94	6.81

## 照合手順

① perl の乱数関数を用いて、10 年分の NPRs から各登録年につき 2,000 件、計 20,000 件の NPRs を「無

作為」抽出した。登録年による層化以外は行なっていないため、僅かであるが1つの特許から複数の NPRs が抽出されたケースがある。このことにより母集団の特性がサンプルに反映されたと考える。

- ② 抽出した NPRs には、明らかに学術文献ではない特許関係の引用（外国特許（抄録）、商用特許データベースなどの引用）や遺伝子 DB のレコードへの引用が含まれているので、これらについては正規表現によるパタンマッチを機械的に行い、人手による照合の対象から外した。
- ③ ②で残った NPRs について WoS の Web 版で検索を行い、合致したものの ID を記録するとともに、作業者の照合結果の確信度（有り／無し二段階）を併せて記録した。
- ④ ③の確信度を参照しながら、2万件の NPRs について確認を行った。なお、とくに確信度の低い結果については、再度の検索を必ず行った。

WoS との間でパタンマッチを行ったことから明らかなように、本研究では、実質上、学術文献を WoS 収録文献と定義している。

## マッチング結果および science linkage の計量文献書誌学的性質

以下、マッチングの結果および science 計量書誌学的性質を示す。現段階でダブルチェック（上記④に相当）が完了している件数が各年度 1,700 件（計 17,000 件）であるため、数値等もその範囲のものに限られていることには注意していただきたい。

表 2 は各年 1,700 件の NPRs をマッチングした結果を示す。当該期間中の NPRs に占める science linkage のパーセンテージは年々低下している。ただし、表 1 に示されるように同期間中にほぼ一貫して NPRs 件数が増加しているためであり、NPRs の母集団に含まれる science linkage 数および特許一本あたりの science linkage 数の推計値も増加している（図 1）。ただ、このことから、必ずしも産業技術と学術研究の関係強化が進展していると解釈できるわけではない。その結論を導くためには技術分野毎の変化を確認する必要があるとともに、（厳密には）文献データベース収録論文数の増大の影響を排除する必要がある。

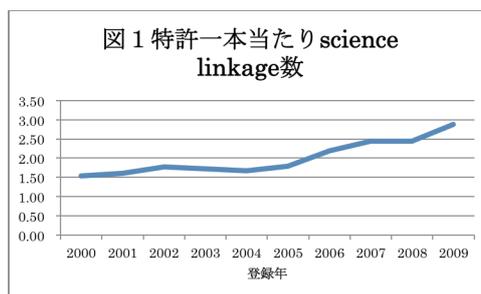
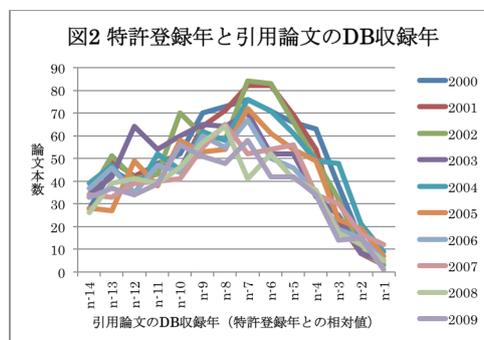


表 2 マッチング結果：米国実用特許の NPR 中の学術文献の推移

登録年	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
NPRs に含まれる学術文献 (Science Linkage 数)	887	878	898	848	841	785	765	752	701	721
パーセンテージ	52.2%	51.6%	52.8%	49.9%	49.5%	46.2%	45.0%	44.2%	41.2%	42.4%
母集団の SL 数 (推計値)	243172	268432	290393	291887	275810	257563	383054	384373	386344	483243
特許一本あたり SL 数 (推計値)	1.54	1.62	1.78	1.73	1.68	1.79	2.20	2.44	2.45	2.89

なお、科学技術政策研究所作成資料 (<http://goo.gl/yVrCc>) に示された science linkage 数の推移と比較すると、推計値はほぼ同じ傾向で推移しているものの、一貫して値が小さい。筆者の CHI 社データを利用した経験から判断すると、The Patent Board 社のサイエンスリンケージ計量において論文の引用とされるその論文の範囲が WoS 収録論文にとどまっていないためと推測される。

図 2 は、特許が登録された年とその特許が引用した論文が WoS に収録された年（多くの場合、掲載誌の発行年と一致するが例外はある）の関係を示す。サンプル数が必ずしも多くはないためそれぞれの登録年でグラフに波はあるものの、グラフ形状はどの登録年も同様の傾向を示しており、とくに登録の 6~7 年前に WoS に収録された論文（≒6~7 年前に刊行された論文）がよく引用されている。特許審査官による引用があるとはいえ、特許の多くが出願後 2~4 年の間に登録されることを考慮すると、必ずしも最新中の最新の科学的成果ではなく、世に出て 2~3 年経過した成果との関連が



強いことが示唆される。また、8年以上前の論文であっても、決して急激に引用が減少することはないが、この点に関連しては技術分野毎の違い等に着目した分析が今後必要と考えられる。

図3は米国特許に引用された論文の分野毎のシェアの特許登録年による推移を上位10分野に限って示したものである(縦軸は対数)。分野区分はThomson Reuters社による排他的分類であるESI区分にしたがっている。医療、ついで化学分野の論文の引用が安定して多いが、それ以外の分野については、変動が激しい。その変動が現実の変動を反映したものなのか、サンプルサイズに起因する確率的な変動なのかは定かではない。

図4は、2000-2009年に登録された米国特許に引用された論文の著者の所属機関がある国上位6カ国のシェアの変動を示す(縦軸は対数)。日米のシェアが多少なりとも低下傾向にあるように見えるが、学术论文全体に占める当該国シェアの低下を反映した可能性もあり、それぞれの学术研究の産業リンクが弱化したと結論することはできない。

図5は、図3と同内容のものを期間全体で我が国に関してのみ集計した。各分野の引用のされやすさは図3の傾向と大きくは異なる。このことは、学术研究の全世界への貢献という観点からは望ましい状況ながら、自国産業への貢献という視点では手放しに望ましいとはいえない。

## まとめ

本報告では簡単な分析結果しか紹介できなかったものの、その結果からも、science linkageは科学技術イノベーション政策にとって貴重な情報を提供することが示されたと考えられる。また、無作為抽出したサンプルデータにより示された結果のいくつかは先行する調査、研究の成果とも整合的であり、(当然ながら)人手による照合の正確性が確認されたと考えられる。

## 謝辞

本研究の一部は科学研究費補助金(基盤C:22500848)およびRISTEX「科学技術イノベーション政策のための科学 研究開発プログラム」からの研究資金により実施された。

## 参考文献

- European Commission, Directorate-General for Research (2002): *Linking Science to Technology, Bibliographic References in Patents Vol.1-9*, Project Report EUR20492/1-20492/9.
- Narin, F. and Olibastro, D. (1992): "Status Report: Linkage between technology and science" *Research Policy* 21 (3), 237-249.
- Narin, F. et al. (1997): "The increasing linkage between US technology and public science" *Research Policy* 26 (3), 317-330.
- 調麻佐志・他(2007):「科学研究と技術の連関」『NISTEP REPORT No. 103 イノベーションの測定に向けた基礎的調査報告書』, 110-148.
- 玉田俊平太(2010):『産学連携イノベーション』, 関西学院大学出版会.
- 富澤宏之(2008):「特許における論文引用データと論文書誌データのマッチング」『NISTEP REPORT No. 111 イノベーション測定手法の開発に向けた調査研究報告書 第二部』, 1-7.
- 富澤宏之・他(2005):「有力特許に引用された科学論文の定量分析」、研究・技術計画学会第20回年次学術大会・講演要旨集, 228-231.

