

Title	特許における科学論文引用の機能について : 引用のカテゴリー化による分析
Author(s)	富澤, 宏之
Citation	年次学術大会講演要旨集, 29: 367-372
Issue Date	2014-10-18
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/12465
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

○富澤宏之（文科省・科学技術政策研）

1. はじめに

特許による科学論文の引用は、特許発明において科学知識が果たす役割に関する貴重な情報源と考えられる。そのため、科学研究がイノベーションに及ぼす影響の分析という科学技術政策研究だけでなく、今や経済学においても重要となった課題に取り組むための手がかりとして、特許による科学論文の引用について様々な研究が試みられてきた。

しかし、そもそも、特許が科学論文を引用することの意味は解明されておらず、したがって、そのデータが何を示すのか、必ずしも明確ではない。この問題の検討は、特許における科学論文引用のデータという豊富な情報源を活用するためには極めて重要である。なお、科学論文を対象とした計量書誌学研究においては、論文引用の動機についての類型化が以前から行われており、論文の引用データを解釈する際の基礎となっている。また、特許における引用の動機についても、多少の研究がなされている¹⁾。

本研究は、このような問題意識のもとで、特許において科学論文の引用が果たす機能についての検討を行うことを目的としている。特許による科学論文の引用は異なるカテゴリーのものが混在しており、それぞれの意味づけは異なると考えられる。本研究では、引用者の種類（発明者か審査官か、発明者と論文著者が関係者であるかないか、等）のカテゴリー別の統計データの分析を通じて、それぞれのカテゴリーごとの引用の機能の違いについて考察する。統計データとしては、特許がどのような論文を引用しているのかを示唆するデータである引用年齢（引用対象の古さ／新しさ）のデータを用いる。

2. 用いたデータセット

本研究では、2005～2012年の8年間に米国特許商標庁に登録された特許のデータセットを用いた。米国特許を用いた理由は、米国の特許法では、必要な文献等の引用を全ての特許に付す事（ただし必要

な文献に限る）が要求されており、引用のデータの質が高いためである。また、2005年以降のデータに限定した理由は、引用を付与したのが審査官であるか、出願者等であるかが区別されているためである。

このデータセットには、各特許のフロント・ページに記載された引用文献の情報が収録されている。これらの文献は、特許の内容の説明や、既存の科学技術との関係を示すために、特許の出願者や審査官が示した文献である。引用文献の多くは先行特許であるが、非特許文献も含まれている。非特許文献には、科学論文、単行本、会議資料、新聞記事、各種雑誌記事、広告やカタログ、あるいはwebサイトなど、様々なものが含まれている。

2005～2012年に米国特許商標庁が発行した特許（utility patent）1,499,205件のうち、非特許文献を引用している特許は774,896件（全体の51.7%）あり、そこで引用された非特許文献は10,569,754件である。そのうち、今回、Thomson Reuters社の科学論文データベースWoS（1981～2012年）に収録されている科学論文と同定できたものは3,548,183件である。これが次節以降で述べるデータ分析の対象である。

この同定には、筆者が開発してきた書誌同定アルゴリズムを使用した²⁾。一般に、このようなデータ抽出においては、適合率（Precision; 同定されたデータのうち正確であるものの割合）と再現率（Recall; 見落とされずに同定されたデータの割合）はトレードオフの関係にある。特許による科学論文引用という現象の規模を量的に把握したい場合は、見落としの少ない事が重要であり、再現率を十分に確保することが必要となるが、今回は、特許による科学論文引用という現象の性質に関する分析が目的であるため、適合率を重視したデータを作成した。今回の同定結果の精度の評価は行っていないが、これまでの研究によると、適合率は90%台後半であると考えられる。

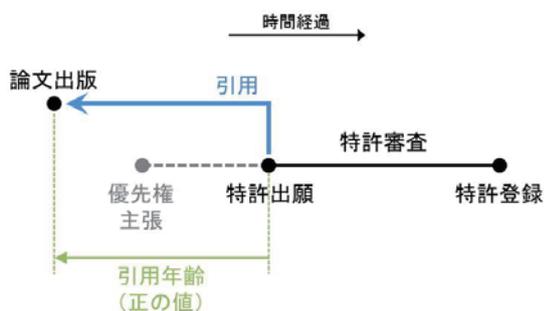
3. 引用年齢の考え方

本研究では、特許において科学論文の引用が果たす機能を理解するために、どのような論文を引用しているのかを知る手掛かりとなる引用年齢に注目する。引用年齢は、ある引用について、それが“どのくらい新しい／古い文献を引用しているか”を示す尺度として計量書誌学において用いられており、具体的には、引用元文献の出版時点とそれが引用している参考文献の出版時点のインターバルの長さとして定式化される。

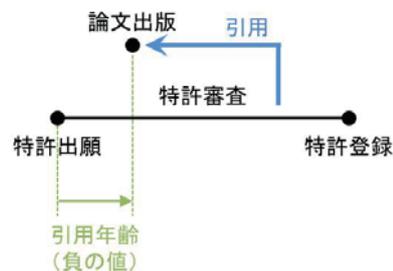
特許における引用の場合、一般的な科学論文の引用よりも様々な要素が絡んでおり、引用年齢の概念はより複雑である。図 1(a)は、特許によって論文が引用されるプロセスの典型を示している。特許における論文の引用は、多くの場合、特許出願者によって特許出願書に記載される形で行われる。この場合、引用が行われた時点は、特許出願時と見なすことができる。また、引用される論文は、多くの場合、出願時点よりも過去に出版された論文である。このような時間的関係がある場合、引用年齢は図 1(a)中に示されているように正の値となる。例えば、特許出願年より 5 年前に出版された論文を引用している場合、引用年齢は「5 年」となる。しかし、論文が出版予定であるなど、特許出願時点よりも後に出版される論文を引用する場合があります、引用年齢は負の値となることもある。

図 1. 特許における論文引用の時間的関係のモデル

(a) 典型的な例(過去の論文を引用)



(b) 審査期間中になされる引用の例



次に、論文引用が審査期間中に行われた場合のモデルを図 1(b)に示した。審査官による論文の引用は、原則的に全てここに該当する。審査官は、特許出願内容が特許の要件を満たすかどうか判定する際に、必要に応じて文献を引用する。この場合、実際に引用を行ったのは特許審査期間のいずれかの時点であるが、それがいつかを明確に示す情報が無いため、便宜的に特許出願時点を引用年齢の起点とする方法が考えられる。図 1(b)では、この引用対象論文は実際の引用時点から見ると過去に出版されたものであるが、特許出願よりは後であるため、この方法では引用年齢が負の値となっている。この点では必ずしも妥当な扱いとは言えないが、出願者も含めて総合的に見た場合、特許出願という重要な時点より前に出版された論文か、後に出版された論文かを区別できるという利点もあるため、本研究では、このような計算方法を採用した。

なお、引用年齢の起点として、「優先権主張」がなされた時点とする方法も考えられる。優先権主張とは、米国特許商標庁への出願に先立って他国の特許庁に出願しており、その出願時点を特許要件の判断基準時点と主張した場合を意味している¹。図 1(a)には、優先権主張があった場合の時間的経緯も示している。優先権主張があった場合は、その基準時点が実質的な出願時点であるので、そこを引用年齢の起点とすることが妥当と考えられる。しかし、本稿では基本的に優先権主張の有無を区別して分析したため、引用年齢の起点として優先権主張を用いていない。

¹ 特許要件の判断基準時点とは、特許が成立する要件である新規性・先願性を判断するための基準時点のことである。

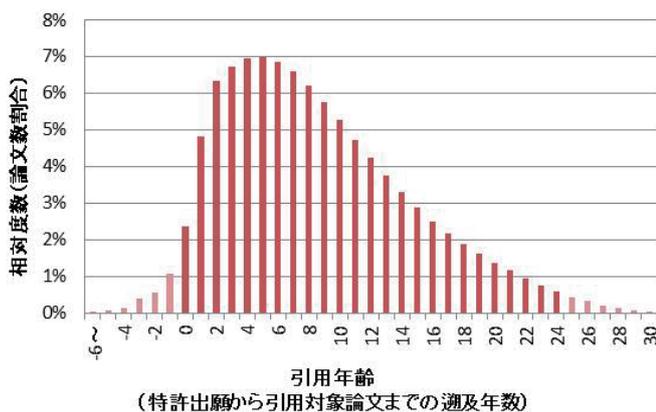
4. 引用年齢に関する集計結果

引用年齢に関する集計に際しては、引用元特許の出願年から引用された論文の出版年を引いた値を引用年齢とした²。

図2に、集計対象全体について、引用年齢別の論文数の分布を示した。引用年齢が5年の論文が最も多く、それよりも引用年齢が大きい場合、引用年齢が大きくなるにつれて、論文数が減っていくという法則が強く成り立っていることが分かる。言い換えれば、論文出願より5年以上経た科学論文は、引用される回数が年々、必ず減っている。このことは、出版から年数を経た論文は、少なくとも引用文献としての価値が減少していくことを意味していると解釈できる。ただし、この場合でも、必ずしも論文としての全体的な価値が減っているとは限らない。ある論文が次第に引用されなくなったとしても、その理由は、そもそも、なぜ引用されていたのか、すなわち引用の目的によって異なるであろう。

また、新しい論文の方が引用文献としての価値が高いということは、特許において、なぜ、論文が引用されるのか、という問題の手掛かりとなるかもしれない。この点については、第5節で考察する。

図2. 引用年齢別の論文数の分布



² 引用年齢を、より詳細に月数や日数で測る方法もあるが、WoSに収録された論文書誌データでは、論文の出版時点として年と月のみ、あるいは年のみが記載されているものも多く、また、そもそも科学論文（の掲載誌）の出版日は必ずしも正確に実態を示しているわけではないため、このように年を単位とした集計方法とした。

なお、今回分析対象としたデータでは、論文の出版年が1981年～2012年に限定されているため、図2における引用年齢は0～24年の部分のみが正確であり、その両側の部分は過小計上となっている。なぜなら、引用元の特許のうち最も古い2005年の特許では、1981年の論文を引用した場合の引用年齢24年が正確に計数できる限界となるためである。また、逆に、最も新しい2012年の特許では、2012年より後の論文のデータが集計対象でないため、引用年齢が負の値の部分は集計されていない。この部分については何年か後に新しいデータを用いて集計することによって正確な値を得る必要がある。ただし、他の年のデータによると引用年齢が負の値となる論文は少数であるので、今後、図2の引用年齢が負の部分が大きく変化することは無いであろう。

次に、引用の 카테고리によって引用年齢に違いがあるかを検討する。まず、審査官による引用と、それ以外の引用（出願者引用）を区別し、両者について、引用年齢別の論文数の分布を比較した（図3）。なお、第3節で述べた「優先権主張」のある場合の扱いを簡単にするために、ここでは「優先権主張」の無い場合の引用3,196,122件（全体の90%）に限定して集計した。このうち、出願者引用は3,101,969件（集計対象の97.1%）、審査官引用は94,151件（同2.9%）である。

図3. 審査官引用と出願者引用を区別した引用年齢別の論文数の分布

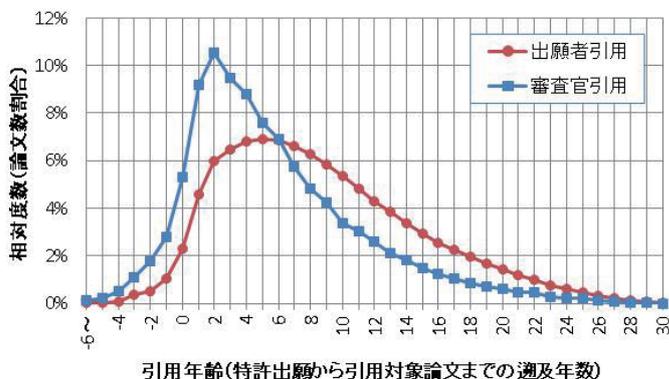


図3によると、審査官引用と出願者引用で、引用年齢の分布に明確な違いがある。出願者引用については、図2に示された分布と同様に引用年齢の最頻値は5年であるが、審査官引用では2年である。また、審査官引用では、最頻値の2年の部分に鋭いピークがあることも特徴である。これらの違いには、引用の目的が出願者と審査官で異なることが反映されていると考えられる。

次に、出願者引用に限定し、出願者が自分が関与した論文を引用した場合（自己引用と呼ぶ）と、それ以外の引用を区別し、それぞれについて、引用年齢別の論文数の分布を調べて図4に示した。

この集計においても、図3と同様に「優先権主張」の無い引用に限定して集計した。また、自己引用であるかどうかの判定は、いくつかの判定基準や方法が考えられるが、ここでは単純に、特許の発明者・出願者名と引用された論文の著者名を比較し、両方に同一の名前が1名以上含まれていた場合を自己引用とした。ただし、名前の一致は、last name 全体と first name の冒頭一文字が一致しているものとした。同性同名や類似の名前の他者の場合など、必ずしも正確に自己引用を同定していない場合もあり得るが、統計的な集計対象としての精度を大きく損ねるものではないと考えられる。この集計対象データのうち、自己引用は246,108件（7.9%）、それ以外の引用は2,855,861件（92.1%）である。

図4. 自己引用とその他引用を区別した引用年齢別の論文数の分布

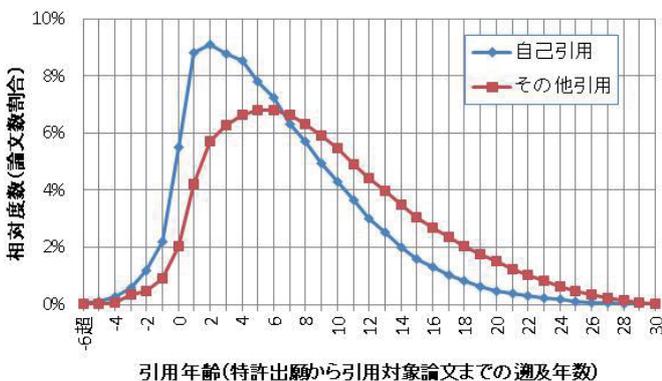
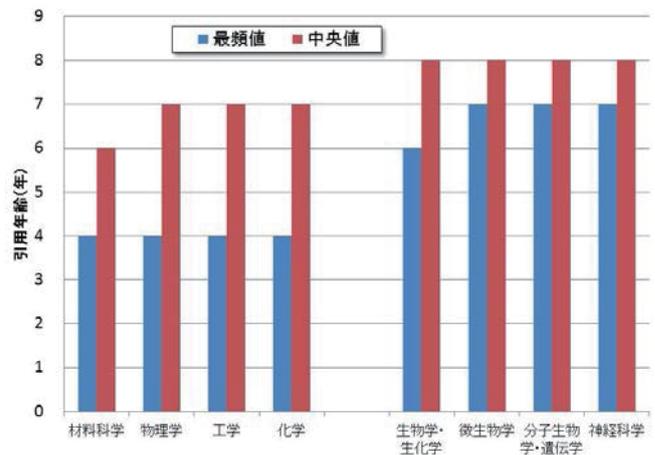


図4を見ると、自己引用とそれ以外の引用で、引用年齢の分布に明確な違いがある。自己引用については、引用年齢の最頻値は2年、それ以外の引用の引用年齢の最頻値は5年である。自己引用の場合、なぜ引用年齢が短い傾向があるのかは、第5節で検討する。

次に、科学分野の違いが引用年齢にどのように影響するかを調べた。そのため、引用年齢を22の分野別ごとに集計して、分野による引用年齢の違いを分析した。図5には、そのうち物質科学と生命科学のそれぞれから4分野選び、引用年齢の最頻値と中央値を示した。

これによると、物質科学の4分野では、引用年齢の最頻値が4年、中央値が6~7年であるのに対し、生命科学の4分野では、最頻値が6~7年、中央値が8年となっており、物質科学よりも生命科学の方が引用年齢が長いという傾向が表れている。

図5. 分野別の引用年齢の統計値についての物質科学と生命科学の比較



5. 特許における論文引用の意味に関する考察

前節に述べたデータ分析の結果を踏まえて、特許における論文引用の意味について考察する。

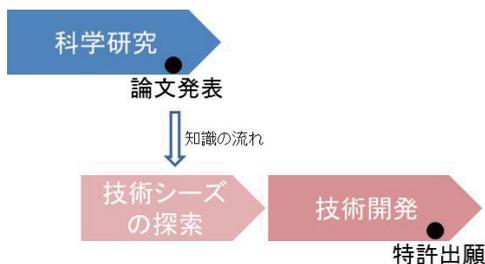
特許が科学論文を引用する理由として、従来、よく主張されていたことは、「引用された論文は、特許発明の源泉である」という解釈である。このような考え方を図6(a)に示した。まず、ある研究者・研究者グループが実施した科学研究の成果として論文が

発表される。次に、それとは異なる主体が、その論文に記述された新しい科学知識を基に技術開発を行い、その成果として特許発明が生まれる、というモデルである。このモデルは、イノベーションのプロセスに関するリニアモデルと整合的である。

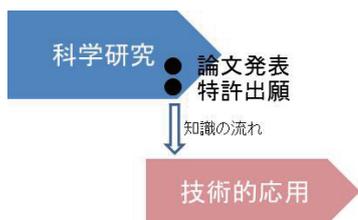
このモデルでは、図 6(a)にも描かれているように、論文発表から特許出願までの期間に、技術シーズの探索も含めた技術開発が行われる。この場合、引用年齢は、この技術開発の期間の長さの指標と見なすことができる。

図 6. 科学論文と特許の関係のモデル

(a) “特許発明の源泉”型



(b) コンカレント型



この考え方のもとで、科学分野によって引用年齢が異なる傾向を示すという図 5 の結果を改めて考察してみる。図 5 には分野による技術開発の長さの違いが表れていると単純に解釈することは不適切である可能性がある。なぜなら、技術開発の期間は、生命科学分野の方が物質科学分野よりも短い場合も多いと考えられるためである。

図 5 に示された傾向を説明する一つの仮説として

は、次のようなものが考えられる。生命科学分野では、引用された論文に書かれた科学的知識が実際に発明特許の源泉となっていることが多いと考えられる。一方、物質科学分野では、企業等によって様々な技術開発が行われているところに、部分的に参考となる科学論文が発表された場合には、比較的、短い期間のうちに引用されると考えられる。この場合、引用される論文は、発明特許の源泉といっても、技術の核心ではなく要素技術の源泉ということになる。

この仮説の場合、物質科学については、図 6(a)のモデルは根本的には変わらないものの、図中の「科学研究」と「技術開発」を、時間的により近くに配置し、また、「知識の流れ」と記した下向きの矢印は、「技術シーズの探索」に向かうのではなく、直接、「技術開発」に向かうように修正する事が適切であろう。

特許が科学論文を引用する理由を説明するモデルとして、図 6(a)とは異なる考え方のものを図 6(b)に示した。これは、科学研究の成果が論文として発表されるだけでなく、同時に特許としても出願されるような場合である。第 4 節の分析で、引用年齢が 0 年であるような場合は、このモデルで説明できるかもしれない。ただし、このモデルにおいても、論文と同時に特許が出願されるのは最初の特許の場合であり、技術的応用が進むにつれて、第 2、第 3 の特許が出願される可能性もある。

最後に、特許が科学論文を引用する理由を説明するモデルとして、背後にある科学研究や技術開発とは全く異なる要素に関するモデルも必要であることを指摘したい。

その一つは、特許の審査官による論文の引用の場合である。審査官は、特許の内容の明確化や、既存の科学技術との関係を示すために、論文を引用する場合があると考えられる。この場合、引用対象となる論文は、審査対象の発明の源泉知識ではなく、当該発明のある側面を説明し得る科学知識の情報源ということになる。この解釈の難点は、引用する科学論文は必ずしも新しい論文である必要性は高い点にある。

そこで、別のモデルを考えてみると、審査官は、当該発明に本当に新規性があるかどうかを判定する

ための根拠として論文を引用する、というモデルが考えられる。この場合、特許とそれが引用する論文の関係は、出願者による自己引用の場合のように、そもそも同一の主体によって生み出されたものであるか、そうでなくとも、特許と論文が内容的に極めて近い関係にあると考えられる。このモデルは、審査官の行動を説明するだけでなく、背後にある研究開発についても、図 6(b)に示した「コンカレント型」モデルが成り立っていることを示唆していると解釈できそうである。

6. 今後の展望

本稿では、いくつかのモデルを用いて、引用年齢に関するデータ分析結果の解釈を試みたが、実際の引用は、様々なタイプのもが多様にあり、また、引用の諸属性は、散らばりの大きい統計的分布に従っていると考えられる。従って、モデルとデータ分析結果の関係は、それほど明確ではなく、ある程度の傾向を示すものに過ぎない。そのため、今後、当面は様々なモデルの精緻化を進める必要はあるものの、それぞれのモデルについてケーススタディ等の詳しい研究を行うことが必要であると考えられる。

参考文献

- [1] 鈴木 裕, 芳鐘 冬樹, 「特許間引用のカテゴリ化：被引用数計測の精緻化に向けて」, 研究・技術計画学会第 27 回年次学術大会・講演要旨集, pp. 760-764, 2012 年 10 月.
- [2] 富澤 宏之, 「科学論文を引用することは特許の影響力を増大させるか」, 研究・技術計画学会第 25 回年次学術大会・講演要旨集, pp. 499-501, 2010 年 10 月.
- [3] 富澤 宏之, 「引用データによる科学技術知識フローの測定：科学技術知識の国際的流通とスピルオーバー」, 研究・技術計画学会第 27 回年次学術大会・講演要旨集, pp. 739-742, 2012 年 10 月.