ビブリオメトリックスによる研究評価

正橋直哉* 勝本加奈子**

1. はじめに

私たちが研究計画を考案したり成果を公表する際には、必ず過去の文献を調べるが、引用索引 SCI (Science Citation Index) はこの目的を満たす極めて有用なツールである。SCI は創始者である Eugene Garfield 博士が、「膨大な文献から研究者が興味を持った論文をいかに効率よく探し出すか」という目的と、「ある分野の論文がどの分野の研究に引用されるか」といった学際間の相関を調査するために1963年に創刊した。このツールを用いることで自分の公表した論文が、どの論文に引用されたか、またどれくらい引用されたか (Citation) を知ることができ、自分の研究に興味を持つ研究者仲間や、論文に対する反応を知ることができる。一方、学術誌においても、「特定の1年間に掲載された論文が、どの程度引用されているかを示す尺度」として IF (Impact factor) なる評価指標が作られた。IF も含め、発表論文数や被引用文献数といった定量可能な指標は、WoS (Web of Science) や JCR (Journal Citation Report) 等の Garfield が設立した Institute for Scientific Information (現: Thomson Scientific) 社の提供するサービスで入手することができる。これらのサービスは「文献」を統計対象として計測する計量書誌学 (Bibliometrics) や科学計量学 (Sciencemetrics) の発展に寄与し、この分野が学問として確立するにおよび、計量指標による様々な使用法が専門家の間で活発に取り上げられている⁽¹⁾。わが国ではこれらの計量指標を用いて、研究者や研究組織の評価に用いられ、研究機関によっては研究費の配分や研究者の処遇にまで利用され、研究の発展に影響を及ぼしかねない状況となっている。しかしこうした使用法の大部分は、サービスを提供する Thomson Scientific 社の意図するものではなく、Garfield 自らが度重なる注意を喚起している⁽²⁾。すなわち、計量指標の研究評価への安易な利用に対して、Bibliometrics への理解は著しく皮相的と言える。日本

金属学会会報では、すでに20年以上前に SCI の功罪を指摘し⁽³⁾、最近でも IF についての記事を掲載している⁽⁴⁾⁽⁵⁾。本稿では Bibliometrics による研究評価の現状について、書誌学・情報学の専門家の意見を紹介すると共に、本会会員に関連する分析結果を紹介することで、計量指標による研究評価への理解を深めることを試みる。

2. Impact factor と Citation

学術誌の IF は、学術誌に掲載された論文の1論文あたりの被引用回数の平均値で、直前の2年間の論文から算出される。つまり A 誌の2004年の IF は、2003年と2002年に A 誌に掲載された論文がその後の1年間(つまり2004年)に引用された回数 (A 誌も含む Thomson Scientific 社に登録されている全学術誌に引用された回数) を、2003年と2002年に A 誌に掲載された論文の総数で除した値である。一般的に、その算出が直前の2年間のデータに限定するため、研究の拡がりの速度により数値が変動し、直近の論文が多く引用されるような学術誌は IF は高く、比較的古い論文も含め発行年度に関わり無く万遍なく引用される学術誌は IF は低い。また応用研究に比べ基礎研究では研究の進展速度が遅く、そのような論文の掲載比率が高い学術誌の IF は必然的に低下する。さらに Nature 等の総合誌や Review 記事の掲載比率の高い学術誌ほど、IF は高い。図1は、材料系主要誌の中からそのうちの5誌について、2004年に引用された論文の被引用数の変化をプロットしたものである(5学術誌は傾向を顕に説明するために選択しており他意はない)。日本金属学会欧文誌(以下欧文誌と称する)は最近の論文が比較的多く引用されていることから、注目が集まっている傾向にあるが、2001年1月号からの誌名変更の影響も否定できない。Thomson Scientific 社によると、誌名変更に伴う名寄せ(論文執筆者が誌名変更に気づかず、誌名を間違えて記述した時の修正)は行っておらず、IF を吟味する上で注意すべき点で

* 東北大学教授; 金属材料研究所(〒980-8577 仙台市青葉区片平2-1-1)

** 東北大学附属図書館農学分館・図書係員

Research Evaluation by Bibliometrics; Naoya Masahashi*, Kanako Katsumoto** (*Institute for Materials Research, Tohoku University, Sendai. **Tohoku University Agricultural Library, Tohoku University, Sendai)

Keywords: research evaluation, bibliometrics, citation, Web of Science, impact factor

2006年2月9日受理

図1 2004年に引用された論文の被引用数の年度ごとの変化。

あるといえる。この点を踏まえて分析を進めるが、2004年のIF算出は2002年と2003年に出版された論文の被引用回数で算出され、この2年間の被引用数は全被引用数に対して欧文誌は25.4%で、Acta Mat.やAdv. Mat.と同様に高い傾向にあることから、現在のIF算出方法は欧文誌にとって有利に働いている。それに対し、J. Am. Ceram.やJ. Mat. Sci.は古い文献も長期にわたって引用される傾向にあり、上記の値は共に7.1%である。山崎によると、現在のIFの算出は、今日常話を提供している学術誌ほど有利で、最近の研究の「流行」をよく反映している指標と分析している⁽⁶⁾。

Thomson Scientific社はSCIにおける恣意的(artifact)なデータを排除するために、version-upごとに種々の機能を取り込んでいる。最新のWoSではAnalyze機能を使ってCiting Articlesを分析することにより、論文を引用している文献の中で、どの著者やどの研究機関がどれくらい引用しているかをrankingとして分布表記することが可能となり、自己引用の多い論文は統計的に顕わとなった。IFについても同様で、IFの算出根拠となるCited Journalの生データをJCRに収録し、IFを高めるために自誌引用の比率を高める行動を行えば、その行為の疑義が読みとれるようになった。Thomson Scientific社によると、適切なピアレビューを行う学術誌は過剰な自己引用を含む論文はとりあげないはずで、必然的にそのような論文は水準の低い学術誌に集まるとしているが⁽⁷⁾、こうした効果の具現化には時間を要する。Nature編集長のMaddoxは、“Restoring good manners in research”と題する記事にて、自己評価を高めるための恣意的な行為を強く批判したが⁽⁸⁾、そのNature誌でさえも2000年～2001年に複数のフィクション論文を掲載し、結果的にIFを高めるための所作との批判を浴びた。計量指標の誤った使用とそれを利用するユーザーのモラルが是正されない限り、本来の目的以外に計量指標を使用するのは危険と言える⁽⁹⁾⁻⁽¹²⁾。

一方、計量指標自体も進化している。IFについては、従来の2年間の論文を対象にした算出から、7年とか15年の長期間の論文を対象にした「累積型IF」や⁽¹³⁾、総合誌に対する専門誌のIFの健全性を目指した「SAIF(Scope Adjusted Impact Factors)」などの新しい指標が提唱されている⁽¹⁴⁾。

図2 材料系25誌のIFとSAIFの相関プロット。

特定の専門分野を中心に少数の学術誌から引用されている学術誌ほどSAIFは高く、学際的に広く引用されている学術誌ほど低くなる。図2は材料系の主要25誌のIFとSAIFの相関をプロットした図であるが、欧文誌は調査学術誌中でSAIFは第7位で、若干ではあるが特定の専門分野を指向する傾向にあると読み取ることができる。この図はそれぞれの学術誌が総合誌を目指すか、あるいは専門誌を目指すかの傾向もうかがい知ることができ、投稿先を決める際の一つの指標にもなりうる。また、Narinにより提唱された科学コミュニケーション構造を図示したステップマップは、学術誌間の引用・被引用の相互関係を知る上で有用な手法である⁽⁶⁾。詳細は省くが、材料科学・総合領域に属するIF上位50誌についてステップマップを描くと、物理系の学術誌がコア・ジャーナル群を形成し、欧文誌はそれらと独立した約10誌からなるサブグループを形成するという興味深い結果が得られた。Thomson Scientific社によると、サブグループ形成は歯科学分野でも見られ、同分野内の雑誌同士の引用が活発であるという。これらの新しい計量指標は、私たちが論文の投稿先や研究の拡がりを検討する上で、新しい切り口になりえると考える。

3. 計量指標による研究評価

2001年度ノーベル化学賞受賞の白川英樹博士の引用分析調査を例にした「論文評価・業績評価の難しさ」と題した小松による興味深い報告がある⁽¹⁵⁾。それによると2000年10月のISI Symposiumにおいて、Japan's Citation Laureatesの発表30名の中に白川博士は含まれておらず、データの信憑性に疑問が生じたという。通常の論文評価はCitation Impact(総被引用回数を論文数で除した値)を使用し、この30名はHigh-Impact Paperを少なくとも12論文発表した研究者として選出されたが、この評価法では白川博士は選ばれない。しかしSTN(The Scientific & Technical Information Network)のSciSearchによる全件引用分析の結果、論文数でも最高引用頻度でも白川博士はトップクラスにあることが判明した(表1参照)。この原因はISI社によるとJapan's Citation Laureatesは1981～1998年の文献に限定したためとしており、評価年代の限定方法によりノーベル賞の研究も選

研究者	論文数	総被引用数	平均引用数	第一著者名分	最高引用頻度数	第二世代ヒット
白川英樹	246	3052	12.4	988	709	546
福井謙一	199	3741	19.0	1400	419	317
中村修二	174	3578	20.6	3341	739	559
岸本忠三	266	11484	43.2	2732	1303	898
野依良治	328	8570	26.1	4327	575	493

専門家によると、ピアレビューは「質の評価」で、定量評価は「成果の有用性あるいは影響力に対する評価」という。前者が専門領域の主観に依存するのに対し、後者は客観的な指標に基づき、研究の「水準」「独創」「発展」は吟味できない。しかし定量評価が不可能なこれらの項目に対する精査なくして真の研究評価は可能であろうか？ 専門が多岐細分化し、成果の公表機会が飛躍的に増加した昨今において、今後とも研究分野に熟知した専門家が時間をかけて審査するという