

論文データにもとづく研究活動の把握

分析結果を読む際の最低限の注意点

本講座では、科学技術イノベーション政策を考える基礎として、重要テーマごとに、どのようなデータがあるか、そのデータから何がいえるのか、どのような限界があるのかについて、専門家が解説する。第3回目は、「研究開発のアウトプット」としての論文データを取り上げる。

伊神正貴 いがみ まさつら

科学技術・学術政策研究所 科学技術・学術基盤調査研究室長

専門は科学計量学。科学技術システムの定点観測、科学における知識創出プロセスの分析、科学研究のマッピングなどに従事。SciREXのデータ・情報基盤構築にもかかわっている。博士(工学)。



論文データは、科学研究における我が国の位置づけの把握、説明責任にかかる公的投資の成果の把握、大学や公的研究機関のランキング等で、研究活動のアウトプットを計測する手段として、活用されている。SciREX事業においても、論文データを活用した調査研究が実施されている。このように色々な場面において活用されている論文データであるが、調査研究の実施やその結果の活用に際しては、最低限の背景を理解していないと、誤った活用につながる可能性がある。そこで、本講座では、論文データにもとづく分析結果を読む際の最低限の注意点を示したい。

メタデータの重要性

論文分析といつても、分析に用いたデータベースや手法によって、結果の見え方は異なる。既存の分析データを活用する際は、データの出典や分析がどのような条件で行われているかの確認は最低限必要である。出典や分析の条件が明確でないデータの活用は説明責任という観点からも避けるべきだろう。どの範囲まで確認するかは分析結果の利用シーンに依存するが、例えば、分析に用いたデータベース(Web of Science, Scopus, Google Scholarなど)、データ抽出を行った時点、論文のカウント方法(整数カウント/分数カウント)、分析対象の分野やドキュメントの種類(Article, Review, Letter, Proceedingsなど)については、日ごろから意識して確認することが必要である。

分数カウントと整数カウント

图表1は科学技術指標2016^[1]に掲載している論文数のデータである。图表1(A)が整数カウントによる順位、图表1(B)が分数カウントによる順位を示している。科学技術指標では、分析にWeb of ScienceのScience Citation Index Expanded(自然科学系)を用いているが、同じデータベースを用いてもカウントの仕方によって結果が異なることがわかる。

整数カウントとは、ある論文の著者の所属機関にA、B、Cという3カ国が含まれている場

合、それぞれの国について1件と数える方法である。一方、分数カウントとは、それぞれの国について1/3件と数える方法である^{*}。科学技術指標では、整数カウントによる結果を論文生産への関与、分数カウントによる結果を論文生産への貢献として解釈を行っている。どちらのカウント方法による測定が好ましいかは、分析の目的や分析結果を利用するコンテキストに依存する。

* 分数カウントの分母については、国単位、機関単位、著者単位など異なる単位で計算できる。ここでは、機関単位で計算した場合の例を示した。

图表1 国・地域別論文数: 上位15か国・地域

(A) 整数カウント

全分野	2012 – 2014年(PY)(平均)論文数		
	整数カウント		
国・地域名	論文数	シェア	順位
米国	338,635	25.8	1
中国	217,793	16.6	2
ドイツ	95,439	7.3	3
英国	92,347	7.0	4
日本	77,460	5.9	5
フランス	67,646	5.2	6
イタリア	59,136	4.5	7
カナダ	56,978	4.3	8
インド	53,238	4.1	9
スペイン	50,803	3.9	10
韓国	50,594	3.9	11
オーストラリア	46,606	3.5	12
ブラジル	37,987	2.9	13
オランダ	33,229	2.5	14
ロシア	28,941	2.2	15

(B) 分数カウント

全分野	2012 – 2014年(PY)(平均)論文数		
	分数カウント		
国・地域名	論文数	シェア	順位
米国	269,016	20.5	1
中国	191,043	14.5	2
日本	64,730	4.9	3
ドイツ	64,072	4.9	4
英国	58,208	4.4	5
インド	46,426	3.5	6
フランス	44,973	3.4	7
韓国	42,747	3.3	8
イタリア	42,513	3.2	9
カナダ	38,852	3.0	10
スペイン	35,705	2.7	11
オーストラリア	31,567	2.4	12
ブラジル	31,502	2.4	13
イラン	23,095	1.8	14
ロシア	22,898	1.7	15

出典：科学技術指標2016、科学技術・学術政策研究所、調査資料-251、2016年8月

注：分析対象は、Article, Reviewである。年の集計は出版年(Publication year, PY)を用いた。

資料：トムソン・ロイター Web of Science XML (SCIE, 2015年末バージョン)を基に、科学技術・学術政策研究所が集計。

分析対象の分野やドキュメントの種類

研究活動のアウトプットの形態は分野に大きく依存する。したがって、分析対象となっている分野やドキュメントの種類についても確認が必要である。前出の科学技術指標の場合、分野としては自然科学系、ドキュメントの種類としてはArticle, Reviewを対象としている。

人文・社会科学系については、論文ではなく図書としても成果が公表される、日本語によって論文が記述される等の状況もあり、科学技術指標の論文分析では対象とはしていない。しかしながら、説明責任にかかる公的投資の成果の把握が求められることは自然科学系と同じであろうから、今後は人文・社会科学系の研究活動のアウトプットの把握も進めていく必要がある。

また、情報通信にかかる研究については、会議録(Proceedings)が重視される。人工知能等の動向把握については会議録の分析も有効と考えられる^[2]。

研究計量による評価の危うさ

論文の被引用数等の計量データは、適切に利用されれば専門家による評定をより妥当、公正にするための補完となり得る。しかし、データに主導された評価や、指標の意味・性質の不十分な理解による誤用がしばしば見られる。このような状況に対して、研究評価における計量データの利用についてのベストプラクティスを示した「研究計量に関するライデン声明」("The Leiden Manifesto for research metrics")が、2015年にNature誌上で公表された^[3]。このマニフェストは、研究者、管理者、評議者の全てにとって、計量データに立脚した研究評価のガイドラインとなるものと考えられるので、10の原則を以下に紹介する。ライデン声明自体のホームページ^[4]から、Nature記事、各國語への翻訳記事やビデオへのリンクが張られているので、それらの原則の詳細については、そちらを参照願いたい。

THE LEIDEN MANIFESTO FOR RESEARCH METRICS

- 1 定量的評価は、専門家による定性的評定の支援に用いるべきである。
- 2 機関、グループ又は研究者の研究目的に照らして業績を測定せよ。
- 3 優れた地域的研究を保護せよ。
- 4 データ収集と分析のプロセスをオープン、透明、かつ単純に保て。
- 5 被評議者がデータと分析過程を確認できるようにすべきである。
- 6 分野により発表と引用の慣行は異なることに留意せよ。
- 7 個々の研究者の評定は、そのポートフォリオの定性的判定に基づくべきである。
- 8 不適切な具体性や誤った精緻性を避けよ。
- 9 評定と指標のシステム全体への効果を認識せよ。
- 10 指標を定期的に吟味し、改善せよ。

参考文献

- [1] 科学技術・学術政策研究所 科学技術・学術基盤調査研究室、科学技術指標2016、科学技術・学術政策研究所 調査資料-251、2016年8月
- [2] 小柴 等、国際・国内会議録の簡易分析に基づく我が国の人工知能研究動向把握の試み、科学技術・学術政策研究所 調査資料-253、2016年8月
- [3] Hicks, D., Wouters, P., Waltman, L., de Rijcke, S. and Rafols, I. The Leiden Manifesto for research metrics. Nature, 2015, 520 (7548), 429–431. Leiden manifesto for research Metrics.
- [4] <http://www.leidenmanifesto.org/> (2016年12月閲覧)
- [5] OECD Patent Statistics Manual, OECD, Paris, 2009
- [6] 元橋一之、池内健太、党建偉、意匠権及び商標権に関するデータベースの構築、科学技術・学術政策研究所 調査資料-249、2016年4月