

科学ニュース記事で取り上げられる学術雑誌に関する予備調査

吉田 光男[†] 菊地 真人[†] 梅村 恭司[†]

[†] 豊橋技術科学大学 情報・知能工学系 〒441—8580 愛知県豊橋市天伯町雲雀ヶ丘 1—1

E-mail: yoshida@cs.tut.ac.jp, m143313@edu.tut.ac.jp, umemura@tut.jp

あらまし 科学ニュース記事では、研究成果が分かりやすく記述される一方、その典拠となる学術文献が明記されない傾向がある。そのため、読者がより詳細を調べることが困難であるほか、第三者による記事の検証も難しくなっている。本研究では、どのような論文がニュース記事として報道されているのかという分析を可能にする、ニュースと論文とを対応付けたデータセットの構築を目指している。本稿では、そのデータセット構築の前段階として、科学ニュース記事から学術雑誌名を自動抽出した結果を用いた基礎調査について報告する。

キーワード Science of Science, 科学の科学, 学術情報流通

1 はじめに

科学コミュニティとマスメディアとの関係 [1] が強まった今日、多くの科学ニュース記事が掲載されている。しかし、科学報道においては、しばしば科学的に不正確な報道がなされたり、誤った印象を与える報道（特に少数派学説が多数派であるかのように報道される [2]）がなされたりしている。このような問題が、どのような学術研究によって引き起こされているのかという分析を行なおうとしても、一般的に、科学ニュース記事にはその典拠となる学術文献が明記されていない。

科学が担う重要性が認識されることに伴い、学術文献の読者が研究者に限られなくなってきた。一方、現在の学術文献の評価は、被引用数をもとにしており、研究者で閉じた評価となっている。研究者以外に与えたインパクトは不明瞭であることから、「オルトメトリクス」と呼ばれる補完指標が提案されている [3], [4], [5]。大学や研究機関での広報・研究支援においても、研究成果が新聞等に露出するように支援し、さらに、その露出を重要視するようになってきている。ここでもやはり、どのような学術研究が社会にインパクトを与えているのかを分析しようとしても、人手の介入（典拠文献の探索）が必要になる。

以上のように、科学ニュース記事の重要性が高まっているものの、典拠文献との関係が不明瞭であるが故に、報道と研究内容とを関連付けた分析が困難である。科学ニュース記事と学術文献とを関連付ける研究はこれまでも行われている [6], [7], [8]。しかし、これらの研究は、ニュース記事の品質や影響力の推定を目的としており、その目的を達成するために“関連する”学術文献の探索は行うものの、“典拠となる”学術文献の探索は行っていない。Science of Science と呼ばれる研究領域が立ち上がるなど [9]、研究そのものの成り立ちを研究する試みもなされており、本研究では、どのような論文がニュース記事として報道されているのかという分析を可能にするデータセットの構築を目指している。本稿では、そのデータセット構築の前段階として、科学ニュース記事から学術雑誌名を自動抽出した結果を用いた基礎調査について報告する。

2 データ

本稿では、雑誌名とニュース記事とを対応付けた公開データと、さらに、そのデータを名寄せ処理して分析に利用する。

2.1 公開データ

我々のこれまでの研究 [10], [11] で作成し、公開しているデータ¹を本稿の分析対象とする。このデータは、「学誌 OR 論文誌 OR 学術誌」という検索条件で絞り込んだ 30,076 件のニュース記事（2004 年 1 月～2017 年 5 月）²から抽出された、12,872 件の雑誌名とニュース記事との対からなる。ニュース記事の情報として、記事 URL と記事公開日の情報が記録されており、ニュースサイトごとの頻出雑誌名を調べたり、出現雑誌名の時系列変化を調べたりすることができる。

データの基礎統計として、表 1 にニュースサイト別の記事数を、表 2 に学術雑誌名別の記事数を示す。AFPBB News、マイナビニュース、時事通信は一般向けにはネット媒体を主とするニュースメディアであり、毎日新聞などの紙媒体を主とするニュースメディアよりも比較的、記事数が多いことが分かる。また、ニュース記事に出現する学術雑誌としては、ネイチャーとサイエンスという歴史ある著名な雑誌の記事数が多いことが分かる。ただし、歴史ある雑誌ばかりではなく、ネイチャー・コミュニケーションズやプロスワンのような、最近創刊されたオープンアクセスメガジャーナルも上位に連ねている。

2.2 名寄せ処理

表 2 のとおり、公開しているデータの雑誌名は記事中の表記に従っており、同じ雑誌であっても、和文表記と英文表記との統合は行われていない。例えば、「Nature Communications」「ネイチャー・コミュニケーションズ」「ネイチャーコミュニケーションズ」の 3 表記は同一の学術雑誌を指し示すにもかかわらず、表記の異なりによって、別の雑誌かのように集計されうる。

1 : <https://doi.org/10.5281/zenodo.2562355>

2 : Ceek.jp News (news.ceek.jp) で収集したニュース記事から抽出した。

表 1 ニュースサイト別の記事数

ニュースサイト	記事数
AFPBB News (www.afpbb.com)	2440
マイナビニュース (news.mynavi.jp)	1690
時事通信 (www.jiji.com)	932
毎日新聞 (mainichi.jp)	562
朝日新聞 (www.asahi.com)	534
読売新聞 (www.yomiuri.co.jp)	509
産経新聞 (sankei.jp.msn.com)	470
共同通信 (www.47news.jp)	327
日本経済新聞 (www.nikkei.com)	278
京都新聞 (www.kyoto-np.co.jp)	252

表 2 名寄せ前の学術雑誌名別の記事数

学術雑誌名	創刊年	記事数
ネイチャー	1869	1296
サイエンス	1880	915
Nature Communications (OA)	2010	280
PLoS ONE (OA)	2006	237
ランセット	1823	210
プロスワン (OA)	2006	209
ネイチャー・コミュニケーションズ (OA)	2010	206
ネイチャーコミュニケーションズ (OA)	2010	198
Nature	1869	176
ネイチャー・ジェネティクス	1992	163

(OA はオープンアクセスジャーナルであることを表す)

このままでは、ニュース記事においてどのような表記で学術雑誌が出現したかを調査できるものの、どの学術雑誌がどれだけ出現したかを調査することは困難である。

今回、Scopus³に収録されている雑誌のタイトルリスト (2019年10月版)⁴をマスターとし、主に2回以上出現した学術雑誌表記を Scopus 収録誌に対して人手によって名寄せした。その結果、12,872 件のデータの内、12,096 件 (94.0%) の雑誌名が Scopus 収録誌と関連付けられた。名寄せ後の出現集計結果を表 3 に示す。表 2 と比べると、出現順位が入れ替わっている雑誌もあり、学術雑誌の出現を調査する上での名寄せ処理が重要であると分かる。また、出版者を見ると、Springer Nature 社の学術雑誌が比較的良好にニュースで取り上げられることも分かる。

3 調査

本稿では、サイトごとに頻出雑誌名が異なるかどうか、出現雑誌名の時系列変化としてオープンアクセスメガジャーナルの盛衰、そして、ニュース掲載数と被引用数との関係について調べた。

3.1 サイト別の出現学術雑誌

全体としてネイチャーとサイエンスという歴史ある著名な雑誌がニュース記事に頻出していたが、この傾向が、どのサイトでも同じかどうかを、記事数の多い 5 サイトを対象として調べ

た。表 4 にサイトごとに集計した学術雑誌名別の記事数を示す。なお、表記方法の異なりも調べるために、名寄せ前のデータを用いた。

雑誌名の表記方法に着目すると、マイナビニュースとほかのサイトとは異なる傾向が見られる。マイナビニュースは英文 (学術雑誌の正式名称) で表記する一方、ほかのサイトでは和文で表記している。マイナビニュースはネット専門メディアであり、新聞等へニュース配信を行う通信社や紙媒体を主とするニュースメディアとは、異なる表記基準を採用していることがうかがえる。

頻出雑誌に着目しても、マイナビニュースとほかのサイトでは異なる傾向が見られる。上位 2 雑誌に着目すると、4 サイトはネイチャーとサイエンスであるものの、マイナビニュースは Nature Communications (ネイチャーコミュニケーションズ) と PLoS ONE (プロスワン) であった。これらの 2 誌はいずれもオープンアクセスジャーナルであり、誰もが無償で読むことができる。ここでも、編集方針や取材方法 (記事執筆方法) について、マイナビニュースはほかのメディアと異なる方針を採用していることがうかがえる。

表 1 と表 4 の記事数を比較すると、5 サイトにおける、上位 5 雑誌の寡占率を計算することができる。朝日新聞の寡占率はほかのサイトと比べて低く、出現する雑誌に多様性があることが分かる。

3.2 オープンアクセスメガジャーナルの盛衰

科学ニュース記事を閲覧した読者が、その記事で取り上げられた論文を読みたいと考えたとき、その障壁の一つは、論文へのアクセスが有償であることであり、逆に、無償でアクセスできれば障壁が低くなる。このような、誰もが無償でアクセスできる論文としてオープンアクセス (OA) 論文があり、掲載数が多い雑誌を特にオープンアクセスメガジャーナルと呼ぶ。オープンアクセスメガジャーナルの出版数に盛衰があることが知られており [12]、同様に、ニュース記事への掲載にも盛衰があるかどうかを調べた。なお、本節では、名寄せ後のデータを用いた。

図 1 にオープンアクセスメガジャーナル 3 誌の年別ニュース掲載数を示す。横軸は年、縦軸はニュース掲載数を表し、2017 年分については、5 月までの掲載数である。佐藤による調査 [12] は、2016 年に PLoS ONE と Scientific Reports との出版数が逆転したことが明らかにしており、ニュース記事への掲載も同様の傾向が確認できた。

本調査は、佐藤による調査 [12] を基礎として調べたものの、ニュース掲載数の観点からは Nature Communications は他の 2 誌よりも多い傾向にある。図 2 に示すように、掲載数が他誌を超えた 2013 年においては、Nature Communications, PLoS ONE, Scientific Reports の出版数はそれぞれ 1,661 件、32,055 件、2,567 件であり、2017 年においても、それぞれ 4,694 件、21,185 件、25,864 件と、Nature Communications の出版数は他誌に比べて少ない。このことから、ニュース記事における出現について、出版数は必ずしも連動しないことも分かる。

3 : <https://www.scopus.com/>

4 : <https://www.scopus.com/sources>

表 3 名寄せ後の学術雑誌名別の記事数

学術雑誌名 (Scopus 収録表記)	ISSN	出版者	記事数
Nature	0028-0836	Springer Nature	1480
Science	0036-8075	American Association for the Advancement of Science	1076
Nature Communications	2041-1723	Springer Nature	698
PLoS ONE	1932-6203	Public Library of Science	480
Nature Medicine	1078-8956	Springer Nature	296
Nature Genetics	1061-4036	Springer Nature	278
New England Journal of Medicine	0028-4793	Massachusetts Medical Society	261
Nature Geoscience	1752-0894	Springer Nature	258
The Lancet	0140-6736	Elsevier	252
The BMJ	0959-8146	BMJ Publishing Group	233

表 4 サイトごとの学術雑誌名別の記事数

ニュースサイト	学術雑誌名	記事数
AFPBB News	ネイチャー	183
	サイエンス	164
	ブリティッシュ・メディカル・ジャーナル	105
	ニューイングランド医学ジャーナル	85
	プロスワン	74
マイナビニュース	Nature Communications	217
	PLoS ONE	127
	Nature	101
	Science	98
	Physical Review Letters	44
時事通信	ネイチャー	163
	サイエンス	83
	ネイチャー・ジェネティクス	40
	カレント・バイオロジー	39
	ネイチャー・メディスン	36
毎日新聞	ネイチャー	66
	サイエンス	54
	プロスワン	23
	ネイチャーコミュニケーションズ	17
	セル	15
朝日新聞	ネイチャー	61
	サイエンス	57
	ネイチャーコミュニケーションズ	23
	ネイチャー・ジェネティクス	17
	ネイチャー・メディスン	16
	プロスワン	16
	ニューイングランド・ジャーナル・オブ・メディスン	16

3.3 ニュース掲載数と被引用数との関係

前節で述べたように、ニュースの掲載数と出版数とは必ずしも連動していない。様々な批判はあるものの、一般的に、学術雑誌の評価には被引用数をもとにしたクラリベイト・アナリティクス社のジャーナルインパクトファクター (JIF) が用いられる [13]。ニュースへの掲載に際して、学術雑誌の重要度を考慮しているのであれば、掲載数は JIF のような被引用数をもとにした重要度と連動する可能性が高い。本節では、名寄せ後のデータをもとに、JIF 同様の指標であるエルゼビア社の CiteScore を用いて、ニュース掲載数と被引用数との関係を調べた。JIF

はクラリベイト・アナリティクス社の Web of Science⁵ のデータをもとに算出するのに対して、CiteScore はエルゼビア社の Scopus のデータをもとに算出する。対象とするデータが異なるものの、算出式は同様である。

図 3 にニュース掲載数と CiteScore との関係を散布図で示す。横軸は 2017 年版の CiteScore、縦軸は 2014 年から 2016 年までのニュース掲載数の合計を表し、CiteScore の算出期間とニュースの集計期間とが一致する。なお、2 回以上出現した雑誌名を名寄せした都合により、掲載数が 2 件以下のものは除外した。散布図の対象となった学術雑誌は 175 件である。

5: <http://webofknowledge.com/WOS>

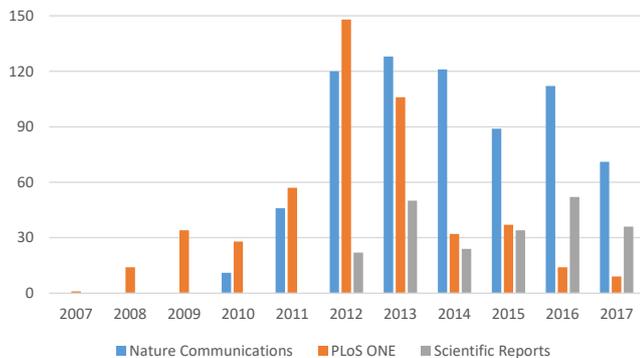


図1 オープンアクセスメガジャーナル3誌の年別ニュース掲載数

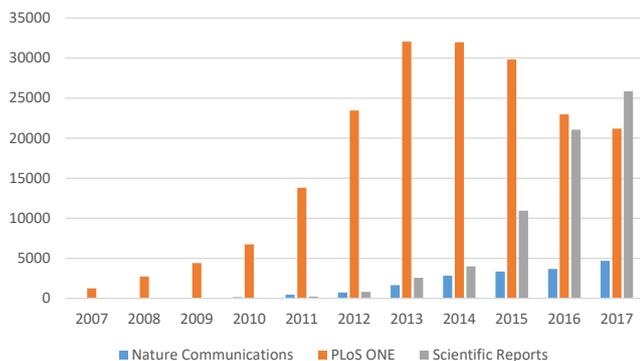


図2 オープンアクセスメガジャーナル3誌の年別出版数

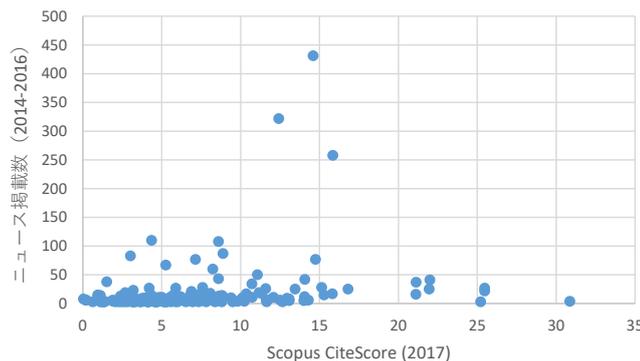


図3 ニュース掲載数と Scopus CiteScore との関係

図3より、ニュース掲載数と CiteScore との間にほぼ相関がないことが分かる。つまり、ニュースへの掲載は被引用数をもとにした重要度とは別に行われる傾向があることを示唆している。ただし、JIF や CiteScore を用いて分野をまたいだ重要度の比較はできないことから、この調査にあたっては、分野別に集計する方が望ましい。分野ごとの調査は今後の課題である。

4 おわりに

本研究では、どのような論文がニュース記事として報道されているのかという分析を可能にするデータセットの構築を目指している。本稿では、そのデータセット構築の前段階として、科学ニュース記事から学術雑誌名を自動抽出したデータを用い、頻出雑誌名の傾向について調査した。その結果、ネット専門メディアはほかのメディアと比べて、雑誌名の記載方法や頻出雑誌の傾向が異なることが分かった。また、オープンアクセ

スメガジャーナルを対象とした年別記事数の調査では出版数と同様の盛衰が確認できたものの、必ずしもニュース掲載数が出版数と連動しないことも分かった。さらに、エルゼビア社の CiteScore を用いて、ニュース掲載数と被引用数との関係を調査したところ、その数値間に相関はなく、ニュースへの掲載は被引用数をもとにした重要度とは別に行われる傾向があることが示唆された。

謝 辞

この研究は 2019 年度国立情報学研究所公募型共同研究 (19FC06) と JSPS 科研費 JP19H04421 の助成を受けています。

文 献

- [1] National Academies of Sciences, Engineering, and Medicine: *Communicating Science Effectively: A Research Agenda*, National Academies Press, Washington, D.C. (2017).
- [2] Petersen, A. M., Vincent, E. M. and Westerling, A. L.: Discrepancy in scientific authority and media visibility of climate change scientists and contrarians, *Nature Communications*, Vol. 10, No. 3502 (2019).
- [3] Priem, J., Taraborelli, D., Groth, P. and Neylon, C.: *altmetrics: a manifesto* (2010).
- [4] 吉田光男: 計量書誌学の新たな挑戦—国産オルトメトリクス計測サービスの開発—, *情報の科学と技術*, Vol. 64, No. 12, pp. 501–507 (2014).
- [5] 佐藤 翔, 吉田光男: 日本の学協会誌掲載論文のオルトメトリクス付与状況, *情報知識学会誌*, Vol. 27, No. 1, pp. 23–42 (2017).
- [6] MacLaughlin, A., Wihbey, J. and Smith, D. A.: Predicting News Coverage of Scientific Articles, *Proceedings of the Twelfth International AAAI Conference on Web and Social Media*, pp. 191–200 (2018).
- [7] Ravenscroft, J., Clare, A. and Liakata, M.: HarriGT: Linking news articles to scientific literature, *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics-System Demonstrations*, pp. 19–24 (2018).
- [8] Smeros, P., Castillo, C. and Aberer, K.: SciLens: Evaluating the Quality of Scientific News Articles Using Social Media and Scientific Literature Indicators, *Proceedings of the Web Conference 2019*, pp. 1747–1758 (2019).
- [9] Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D. and Barabási, A.-L.: Science of science, *Science*, Vol. 359, No. 6379, eaao0185 (2018).
- [10] Kikuchi, M., Yoshida, M. and Umemura, K.: Journal Name Extraction from Japanese Scientific News Articles, *Proceedings of the 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, pp. 143–148 (2018).
- [11] 菊地真人, 川上賢十, 吉田光男, 梅村恭司: 観測頻度に基づくゆう度比の保守的な直接推定, *電子情報通信学会論文誌 D*, Vol. J102-D, No. 4, pp. 289–301 (2019).
- [12] 佐藤 翔: オープンアクセスメガジャーナルの興隆, と, 停滞, *情報の科学と技術*, Vol. 68, No. 4, pp. 187–188 (2018).
- [13] 逸村 裕, 池内有為: インパクトファクターの功罪: 科学者社会に与えた影響とそこから生まれた歪み, *月刊化学*, Vol. 68, No. 12, pp. 32–36 (2013).