

重回帰分析を用いた論文の質と影響度に関する考察

－ 社会科学分野の論文を題材として －

日大生産工（学部） ○林 直弘 日大生産工 水上 祐治

1 要旨

本稿は論文の被引用数に着目し、研究活動の見える化を図る。具体的には、論文の被引用数を従属変数とし、抄録の情報量等の各要素と研究分野を説明変数とすることで、論文の各要素がどの程度、論文の質へ影響を与えているのかを定量的に示す。Clarivate Analytics社のWeb of Scienceから82980編の論文抽出を行い、3つの分析モデルを作成した。このデータから論文のドキュメントタイプにおける影響度、そして社会科学分野の論文の質の低下に着目した考察を行った。

2 先行研究と仮説

日本の研究機関では競争力向上のため、研究の成果の共有や研究手法の標準化を目指している。研究機関及び国の学術的研究力を評価する場合、論文の量と質で評価する手法がある¹⁾。量とは論文数、質とは論文の出来栄であり多くの評価軸がある²⁾。また、インパクトファクターやh指数等の定量的な評価では、被引用数が重要な要素とされており、被引用数を論文の質とみなす見方も存在する³⁾。そこで本稿では、被引用数の多い論文を質の高い論文として定義する。

また本稿では、「社会学の現状が、社会科学の論文の被引用数（論文の質）へ影響を及ぼしているのではないか」という仮説を基に議論を進める。社会科学の下位分野に位置する社会学が、どのような学問であるかという問題は、つねに社会学者を悩ましてきた問題であり、その背景には断片化や細分化といった現象が考えられる⁴⁾。断片化とは、社会学の研究における傾向を示しており、論文において単一著者の質的なものが多く、比較の対象や基準が明瞭でないことがわかっていいる。他にも社会学の下位分野において、「ドイツ社会学」や「アメリカ社会学」と呼ばれるように、国や言語における学問の細分化が起きており、このような現象から社会学の論文は他の研究分野と比べて、引用文献が少なく、新しい文献が引用されにくいことが指摘されている。

3 重回帰分析

3. 1 分析データの抽出

論文の抽出は書誌データベースのWeb of Scienceから行った。言語タイプは英語、ドキュメントタイプはArticle もしくは Review、そしてトピックに「social

issue」が含まれる 82980 件（2018 年 8 月 26 日時点）を対象としている。

3. 2 分析モデルと変数

(1)分析モデル

従属変数を合計被引用数（Z9）とし、その他の項目ドキュメントタイプ（DT）、抄録（AB）、引用文献数（NB）、直近180日の利用回数（U1）、出版年（PY）、ページ数（PG）、研究分野（WC）を説明変数とし、論文の質に対する各々の説明変数の影響度を示した。分析は3種類のモデルに対して行った。まず直近180日の利用回数と研究分野ダミー変数以外を適用した最小のモデル（モデル①）、次にモデル①に研究分野を加えたモデル（モデル②）、最後にモデル②に直近180日の利用回数を加えたモデル（モデル③）である。データ数はモデル①、②は82980件、モデル③は直近180日間に利用実績がない論文を削除した32015件である。

(2)抄録の数値化

抄録に含まれる名詞をカウントして「情報量」と定義した。分析モデル別に上位101種類の名詞を抽出後、出現回数をカウントする。そして論文ごとに抄録に含まれる名詞の合計値を量的データとして扱った。抄録の名詞をカウントして「情報量」と定義する手法は、本稿の新規性の一つである。

(3)ダミー変数の作成

ドキュメントタイプはArticleを0、Reviewを1とするダミー変数を作成した。また研究分野はMizukami et. al. (2017) ⁵⁾を基に、小分類264種を大分類23種に変換した。その後、それら23種のダミー変数を作成し、該当しない分野を0、該当する分野を1とした。なお、一つの論文に対して、複数の分野が該当する場合もある。

4 分析の手順

重回帰分析はそれぞれのモデルに対して施す。まず、既定の説明変数をすべて投入した分析を行った。そして、有意水準10%において有意でない変数を削除した分析を行った。表1に重回帰分析の結果を示す。ここでは、有意水準10%において有意でない変数を削除した分析の結果を掲載している。

Consideration on the quality and influence of articles using multiple regression analysis
－ Subjects of social science papers －

Naohiro HAYASHI, Yuji MIZUKAMI

表 1. 各モデルの重回帰分析の結果

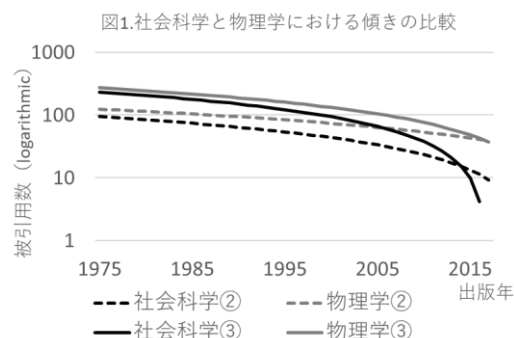
変数		意味	数値の取り扱い方	傾き		
タイプ	記号			モデル①	モデル②	モデル③
従属変数	Z9	合計被引用数	合計被引用数	-	-	-
説明変数	DT	ドキュメントタイプ	0.Article 1.Review	13.436 ***	12.816 ***	10.272 ***
	AB	抄録の情報量	名詞の数	0.162 ***	0.129 ***	-0.090 ***
	NR	引用文献数	合計文献数	0.306 ***	0.291 ***	0.165 ***
	U1	直近 180 日の利用回数	利用回数	-	-	5.923 ***
	PY	出版年	出版年	-2.025 ***	-2.025 ***	-5.616 ***
	PG	ページ数	ページ数	-0.153 ***	-0.085 ***	-
	WC	- 研究分野(大分類)*1	1-23	-	-	-
		1 農学	0.該当なし 1.該当	-	-2.802 *	-8.105 ***
		4 臨床医学	0.該当なし 1.該当	-	2.561 ***	4.984 ***
		5 コンピューターサイエンス	0.該当なし 1.該当	-	5.025 ***	-
		6 経済学 & ビジネス	0.該当なし 1.該当	-	7.332 ***	-
		7 工学	0.該当なし 1.該当	-	-	-5.177 ***
		8 環境/生態学	0.該当なし 1.該当	-	4.538 ***	-4.147 ***
		10 免疫学	0.該当なし 1.該当	-	-	-
		12 数学	0.該当なし 1.該当	-	-5.196 **	-
		15 総合	0.該当なし 1.該当	-	10.528 ***	11.944 ***
		16 神経科学 & 行動	0.該当なし 1.該当	-	- ***	5.001 ***
		17 薬理学 & 毒物学	0.該当なし 1.該当	-	-6.578 ***	-
		18 物理学	0.該当なし 1.該当	-	25.274 ***	32.998 ***
		19 植物 & 畜産学	0.該当なし 1.該当	-	-4.885 ***	-12.874 ***
		20 心理学/精神医学	0.該当なし 1.該当	-	5.267 ***	3.564 ***
		21 社会科学、一般	0.該当なし 1.該当	-	-4.795 ***	-5.390 ***
		23 芸術学と人文 切片	0.該当なし 1.該当 -	- 4071.390 ***	- 4071.625 ***	- 11300.000 ***
対象データ数				82980	82980	32015

*1: WoS Essential Science Indicators の分類を使用

***p<0.01, **p<0.05, *p<0.10

5 結果・考察

全てのモデルにおいて、「ドキュメントタイプ」が大きく正の相関を示しており、Review 論文が支配的に出ている。これは、Review 論文が各分野の論文の位置づけ等で参照されているため、論文の質へ大きく影響を与えていることが考えられる。「出版年」は、古い論文であるほど被引用数が多くなり、「直近 180 日の利用回数」は利用頻度の高い論文の被引用数が多くなる傾向が見られた。これらは研究者の目に触れることが多ためであると考えられる。



次に「研究分野」の「社会科学」に着目する。図 1 では表 1 で負の相関を示した社会科学と、大きく正の相関を示した物理学の傾きを比較したグラフとなっている。波線をモデル②、実線をモデル③としており、ドキュメントタイプは Review で統一した。そ

他の研究分野と出版年を除いた説明変数には平均値を適応している。社会科学の傾きは物理学を下回っており、その差はモデル③でより顕著に表れている。すなわち、最近注目されている論文において、社会科学の分野は更なる質の低下が懸念される。新しい研究結果で、新しい文献について物理学は引用されやすく、社会科学は引用されにくいことが明らかになっている 4)。

こうした社会科学の論文の質の低下には、下位分野である社会学の細分化や断片化が背景にあると考えることができるだろう。また今後は地域別で社会学の現状を見ていき、詳細な原因を究明するなど、ミクロな視点での分析をしていく方針である。

「参考文献」

- 1) Yuji Mizukami, Yosuke Mizutani, Keisuke Honda, Shigenori Suzuki, Junji Nakano, An International Research Comparative Study of the Degree of Cooperation between disciplines within mathematics and mathematical sciences, Springer, Behaviormetrika, Vol.44, issue 2, pp.385-403, 2017
- 2) 根岸正光, 山崎茂明, 「研究評価—研究者・研究機関・大学におけるガイドライン」, 丸善, 2001
- 3) 山本耕平, 太郎丸博, 「社会学の方法と引用文化の日英米比較」, 理論と方法, Vol.30, No.2, pp.165-180, 2015
- 4) Hargens, L. L. 2000. "Using the Literature: Reference Networks, Reference Contexts, and the Social Structure of Scholarship." American Sociological Review 65(6): 846-65.