JAIST Repository
https://dspace.jaist.ac.jp/

Title	科学生産性指標を導くプロセスベースの数理モデリング
Author(s)	別所, 和博
Citation	年次学術大会講演要旨集, 39: 937-938
Issue Date	2024-10-26
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/19474
Rights	本著作物は研究・イノベーション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 Research Policy and Innovation Management.
Description	一般講演要旨

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

２Ｅ２４

科学生産性指標を導くプロセスベースの数理モデリング

○別所和博（埼玉医科大学）
besshokazuhiro.research@gmail.com

緒言

科学の生産性(Science productivity)が示す性質についての研究は 1926 年の Lotka の報告に遡る[1]。そこで Lotka は科学者による論文出版数にある種の統計的性質、具体的には「z本の論文を発表する研究者の数が$1/z^2$に比例する」という傾向が成立することを報告した。その後、彼の見出した統計則自体の頑健性、彼や以降に提案された新たな統計則などが実現されるメカニズムについて様々な研究が行われてきている。本研究では、先行研究の多くでは考慮されていない論文発表のプロセスを明示的に考慮した数理モデルを新たに考え、シミュレーションから得られる生産性指標に関する統計量の性質を調べることで、この数理モデルの妥当性を検討し、さらに科学の生産性について理解を深めることを目指した。

モデルの概要

科学の生産が研究指導者（以下 PI, Principal Investigator）とそこに属する研究者（以下 Student）らによる共同作業で行われるフレームワークを想定した。そこでは PI の人数をn人、一人の PI の下に属する Student の人数をm人として、これらは常に一定だと仮定する。今、科学の生産性が個々人のもつ 1 次元のアビリティx, yと資源rで決まると仮定し、i番目の PI の下につくj番目の Student がある 1 単位時間あたりの生産性を$Z_j^{(i)}\left(x^{(i)}, y_j^{(i)}, r_j^{(i)}\right)$と表記する。ここで$x^{(i)}$は$i$番目の PI がもつアビリティ、$y_j^{(i)}$は$i$番目の PI の下につく$j$番目の Student がもつアビリティ、$r_j^{(i)}$は$i$番目の PI の下につく$j$番目の Student が研究に費やすことができる資源を意味する。例えば、具体的な関数形として

$$\Pr\left[Z_j^{(i)} = k\right] = \begin{cases} \left(1 - p_j^{(i)}\right)^{r_j^{(i)}} & k = 0 \text{ のとき} \\ 1 - \left(1 - p_j^{(i)}\right)^{r_j^{(i)}} & k = 1 \text{ のとき} \end{cases}$$

$$p_j^{(i)} = \alpha x^{(i)} + \beta y_j^{(i)}, \qquad r_j^{(i)} = r_{tot}^{(i)}/m$$

を考えることができるが、これは科学の生産性を論文出版の成功で評価し、アビリティをその成功確率と捉えた場合に相当する（つまり$x^{(i)}, y_j^{(i)}$は 0 以上 1 以下の値をとる）。即ち、i番目の PI の下につくj番目の Student が論文出版に成功する確率は PI と Student のアビリティの重み付き平均（つまり$\alpha + \beta = 1$）$p_j^{(i)}$で決まり、その Student が利用可能な資源$r_j^{(i)}$はi番目の PI がアクセス可能な資源$r_{tot}^{(i)}$をm人の Student に均等に配分することで決まる。さらに、このモデルで資源$r_j^{(i)}$は研究の試行回数という形で論文出版の成功に貢献しており、$r_j^{(i)}$回の独立な試行が一度でも成功した場合、論文が 1 報出版され、これは確率変数$Z_j^{(i)}$が確率$\left(1 - p_j^{(i)}\right)^{r_j^{(i)}}$の Bernoulli 分布に従うという形で表現される。

本研究ではこのような確率モデルにより 1 単位時間あたりの論文出版が表現され、このプロセスをt_{max}回繰り返すことで、アビリティが固定された PI と Student らがその期間内に出版する論文の総数が決まるとした。さらにシミュレーションではt_{max}回の試行を 1 世代と規定し、1 世代経過後に PI と Student の更新を行った。そこではこれまで Student であった合計nm人の研究者から一定の規則でn人を次世代の PI として選び、Student はそれら新 PI のもとに新たに配置されるとした。具体的な規則としては、t_{max}回の試行内に出版された論文本数が多い順に Student を選ぶことなどが考えられるであろう。なお、Student のアビリティは乱数で決めた。

以上の数理モデルを使えば与えられたパラメータ値のもとでシミュレーションを行うことで、人工的に科学生産性データを人工的に生成できる。本研究ではこの生成データをもとに、考案した数理モデルがどのような性質を持っているのかを様々な側面から検討した。なお、科学の生産性についての先行研究ではしばしば、科学者が論文を出版すればするほど、そのあとに論文を出版する確率が高くなる、といったフィードバックの重要性が指摘されている。上記のモデルにおいて、Student が利用可能な資源$r_j^{(i)}$がi番目の PI がアクセス可能な資源$r_{tot}^{(i)}$により決まるとしているが、$r_{tot}^{(i)}$が PI による論文出版数に依存して決まるという仮定をシミュレーションに組み込むことで、生産性が高い PI（とその下の Student）が以降に論文出版する確率が高くなる、というフィードバックを実装することが可能である。

結果と考察

まず、世代更新を行わずにプロセスを繰り返したときに論文出版数の時間変化を調べ、さらに PI がアクセスできる資源量が論文出版数に依存するときに資源量の時間変化も調べた。特に、資源配分が論文生産数に強く依存する仮定のもとでは、資源にアクセス可能な PI とアクセス不可能な PI への二極化が生じ、新たな論文出版が可能な PI とほぼ不可能である PI への分離が生じる結果が得られた。

次に、世代更新を伴うシミュレーションのもとで全 PI のアビリティ平均と総論文出版数の時間変化を調べたところ、資源配分が論文出版数依存であるか否かにより大きく結果が異なることがわかった。特に、時間とともに平均アビリティや総論文出版数が低下する非直感的な傾向が得られ、これは資源配分のバイアスが世代更新時の PI 更新により期待される平均アビリティ増加を阻害する効果を持つからであると予想された。

最後に、数理モデルから得られるデータから先行研究で報告されている論文出版数についての統計的パタンが再現できるか確認したところ、先行研究との接続性は悪かった。特に、PI と Student という立場の非対称性があるため、出版論文数は異なる分布を混ぜ合わせたものになってしまい、滑らかに減少するような分布の再現は難しい。また、論文出版数が非常に少ない著者が最も多く、そこから単調減少するパターンも現在のモデルではうまく再現できていない。数理モデルの妥当性については更なる検討が必要だと考えられる。

参考文献

[1] Lotka, A. J. 1926. The frequency distribution of scientific productivity. Journal of the Washington Academy of Sciences 16: 317-323.