

プレプリントとジャーナル論文の差異：bioRxiv を用いた試行

文部科学省 科学技術・学術政策研究所 データ解析政策研究室

要旨

エビデンスに基づく政策立案 (以下, EBPM: Evidence-based Policy Making) 機能の強化や、オープンサイエンスの潮流を踏まえ、本稿では、プレプリントを用いた分析を活用することで、原著論文を基とした研究力の分析に対して相補的に新しい知見を得ることができるのではないかと考えた。特に、プレプリントとそれが最終的に出版された OA ジャーナル論文とを比較することで知見を得られないか試行した。

まず、近年のオープンジャーナルなどの流れもあって、bioRxiv のプレプリントと OA ジャーナル論文の全文 XML を一定量確保することができ、プレプリントとジャーナル論文の比較に関する技術的な可能性は検証できた。他方、参考文献数や単語数などの外形的な基準や、簡単な文書類似度から両者の差分を明らかにしようとした今回の試行の範囲では、プレプリントとジャーナル論文、ジャーナル論文になったプレプリントとそうではないプレプリントの間で明確な違いを見いだすことはできなかった。機械学習手法を用いても分類精度は 47% 程度と高くなかった。

プレプリントとジャーナル論文の間に大きな差はないという結果は先行研究でも示されている知見と一致するものであるが、より大規模かつ比較的最近の状況でも再現性が確認できた。これらに加えて本稿のもたらした新たな知見としては、著者数などの多くの外形的基準においても差は小さい、ジャーナル論文になっていないプレプリントとの違いも大きくない、といった点が挙げられる。

Differences between preprints and journal articles: a trial using bioRxiv

Research-Unit for Data Application, National Institute of Science and Technology Policy (NISTEP), MEXT

ABSTRACT

In light of the strengthening of the evidence-based policy making (EBPM) function and the trend toward open science, this paper considers whether analysis using preprints can provide new knowledge in a complementary manner to the analysis of research power based on original papers. In particular, the preprint and the final version of the preprint are used as the basis for the analysis. In particular, we attempted to see if we could gain insights by comparing preprints with OA journal articles in which they were finally published.

Translated with www.DeepL.com/Translator (free version)

First, due to the recent trend of open journals, we were able to secure a certain amount of full-text XML of bioRxiv preprints and journal articles, and verified the technical feasibility of comparing preprints and journal articles. On the other hand, within the scope of this trial, which attempted to identify differences between preprints and journal articles, and between preprints that became journal articles and those that did not, we could not find any clear differences. Even using machine learning

methods, the classification accuracy was not high at about 47%.

The result that there is no significant difference between preprints and journal articles is consistent with the findings of previous studies, but it is also reproducible in a larger and relatively recent context. In addition to these findings, the new findings of this paper are that the differences are small in many external criteria such as the number of authors, and the differences between preprints and journal articles are not large.