

## COVID-19 / SARS-CoV-2 関連のプレプリントを用いた研究動向の試行的分析

文部科学省 科学技術・学術政策研究所

小柴 等, 林 和弘, 伊藤 裕子

### 要旨

近年, 査読前の論文草稿であるプレプリントをとりまとめて公開するプレプリントサーバの活用が進んでいる。ジャーナル論文の投稿に先立つというプレプリントの性質上, ジャーナル論文を対象とした研究動向分析と比較して早期に研究動向を把握できる可能性があり, そのため新興の(エマージングな)研究領域の補足に有用と考えられる。

こうした背景のもと, 本報ではCOVID-19関連のプレプリントを対象に, 自然言語処理を用いたエマージング領域の把握を試行した。

その結果, 既存の分析と同様に, 疫学調査のステップに合致する動向を得ることができた。それに加えて, 通常の査読論文も含めて分析した先行研究では明確には検出ができていなかった, 医薬・ワクチン開発に関するトピックを抽出することができた。

今回の試行により, プレプリントを利用したエマージング研究内容のメタ把握が実現できる可能性が示唆された。

## A Trial of early detection system for research trends through the preprints data — Research status around COVID-19 / SARS-CoV-2

KOSHIBA Hitoshi, HAYASHI Kazuhiro, ITO Yuko

National Institute of Science and Technology Policy (NISTEP), MEXT

### ABSTRACT

In recent years, the use of preprint servers, which compile and publish preprints of pre-reviewed drafts of articles before peer review is in progress. Due to the nature of the preprint, which precedes the submission of a journal article, the research on journal articles in comparison with trend analysis, it is possible to understand research trends at an earlier stage, and therefore it can be used to supplement emerging research areas. It is thought to be useful.

In this report, we attempted to understand the emerging regions of COVID-19 related preprints by using natural language processing.

As a result, we were able to obtain the trend consistent with the epidemiological survey step as well as the existing analysis. In addition, we were able to detect the emergent regions, which had not been clearly detected in previous studies that also included ordinary peer-reviewed papers. In addition, we were able to extract topics related to drug and vaccine development.

This trial demonstrated the possibility of realizing a meta-understanding of emerging research content using preprints.