Textual Content and Academic Journals Selectivity:
A Case of Economic Journals

Paweł Baranowski

Faculty of Economics and Sociology, University of Lodz, Poland
Institute of Economic and Financial Research, Lodz, Poland
pawel.baranowski@eksoc.uni.lodz.pl

Szymon Wójcika

Faculty of Economics and Sociology, University of Lodz, Poland
szymon.wojcik@eksoc.uni.lodz.pl

Abstract

Currently observed vast influx of papers obstructs the editorial procedures in scientific journals, especially in top-quality academic periodicals. Moreover, it stimulates the emergence of non-selective journals, attracting authors with short editorial procedures in exchange for high fees. We argue that introducing natural language processing (NLP) can help distinguish the papers worth reading by the editor from those whose scientific quality does not meet the standards. To test this hypothesis, we apply state-of-art large language models, i.e. bidirectional encoder representations from transformers (BERT). Our sample consists of 510 academic papers representing economics, finance or business. The papers were collected from journals of three levels of selectivity, namely: highly selective (top-tier journals), moderately selective (journals listed on DOAJ list), nonselective ("predatory" journals). More specifically, we applied a pre-trained Sci-BERT model on anonymised texts of academic papers. The results show that the pure textual content may give 82% out-of-sample accuracy in classifying texts into the three levels of selectivity. The outcomes of the study prove the usefulness of NLP in distinguishing the quality of the paper and supports Beall's classification of "predatory" journals. Even if our model is based only on the linguistic features of the academic paper, we believe it might be applied to assist journals' editors in the initial screening of submitted papers.

Keywords: Text Classification, Natural Language Processing, open access journals, predatory journals, BERT