

A Quantitative Examination of Software-Method Packages in Psychological Research Articles

Kai Li

Drexel University, Philadelphia, USA

kl696@drexel.edu

ABSTRACT

Quantitative science studies have developed unbalanced interests in scientific knowledge, focusing on the impact of scientific outputs and distribution of knowledge products, rather than how such knowledge is produced. The lack of research in the latter aspect is an important reason for the growing gaps between the quantitative and qualitative communities of science studies, which have prevented quantitative researchers from developing deeper insights into the nature of scientific knowledge and communication. I propose to examine how research objects related to scientific knowledge production, including research method and software entities, are represented and distributed in full-text scientific publications.

This research will develop a machine learning pipeline to extract method and software entities from a large-scale full-text scientific publication corpus, as the foundation for the rest of this study. A semi-supervised learning approach will be adopted to learn these entities using an iterative manner. Based on the extracted entities, quantitative analyses will be conducted to examine the relationships among the presence and form of these research objects and the conditions in which scientific knowledge is produced, as informed by theories developed from both communities of qualitative and quantitative science studies. I am specifically interested in how the lifecycle of these scientific objects could influence the ways in which these objects are represented (as citation or mention) in scientific publications.

This research will broaden the scope of quantitative science studies by introducing new objects into the existing research frameworks and provide a new, science as practice perspective on the quantitative understanding of scientific knowledge. Both benefits will contribute to a more comprehensive view of science that are derived from an integration of different theoretical and methodological traditions of science studies.

TOPICS

bibliometrics; machine learning; natural language processing