Predicting Scientific Evolution by Understanding its Driving Factors

Jiangen He
Department of Information Science, Drexel University, United States
jiangen.he@drexel.edu

ABSTRACT

Background: The increasing availability of big scholarly data and computational methods offer unprecedented opportunities to understand the mechanism of how science advances and predict critical scientific dynamics. However, our understanding of driving factors behind scientific development is limited and predictions driven by big scholarly data are rarely made.
Goals: This proposed research aims to (1) provide computational approaches to measuring driving factors (novelty and uncertainty) in science from multiple perspectives, (2) build predictive models for future trends and dynamics in science based on representations of scientific knowledge, and (3) build a context-environment by visual analytics to support qualitative and in-depth analytics of the driving factors and predictions.
Methods: The research represents scientific knowledge as a system of scientific discoveries, assertions, and topics along with their status of novelty and uncertainty. Novelty and uncertainty are identified and measured from the full text of scientific publications and scholarly networks using representation learning and computational linguistics. Their roles and interactions in scientific development are examined over time. Based on the representations of scientific knowledge, I build predictive models for the trends of rising and falling and for the dynamics of convergence and divergence in scientific development. In addition, the measured factors and prediction results are integrated into temporal visualizations so as to effectively and intuitively validate, analyze, and understand the roles of the driving factors and prediction results in a concrete context.
Significance: The metrics and empirical studies of driving factors in science will offer a deeper understanding of how science advances at multiple levels of granularity. The data-driven evidence derived from the studies can complement related theories from philosophy and sociology studies of science. Predictions made based on the understanding will have practical implications for individual scientists to identify research opportunities and for research policymakers to optimize the allocation of research resources. The visualization designs and tools will serve as platforms for validating and analyzing not only metrics and predictive models developed in this research but also other existing and newly developed ones.

TOPICS

big data; machine learning; informetrics; data visualization; bibliometrics; natural language processing