

ARBITER: Toward the Automated Detection of Technoscientific Emergence from Full-Text Publications and Patents

Olga Babko-Malaya¹, David C. Brock², James Pustejovsky³, Patrick Thomas⁴, Sean Stromsten¹ and Fotis Barlos¹

¹BAE Systems, ² David C. Brock Consulting, ³Brandeis University, ⁴1790 Analytics
Email: olga.babko-malaya@baesystems.com

Abstract

This paper describes ARBITER (Abductive Reasoning Based on Indicators and Topics of EmERgence), a system for characterizing scientific and technological fields and detecting emergent fields. Guided by actant network theory, we define emergence as growth in the robustness of an actant network. To measure emergence, we define a set of characteristics of actant networks that measure their extent, resilience, growth, novelty, maturity, etc. We further separate between the scientific and technological fields by analyzing the presence and impact of the marketplace actant. Finally, we evaluate the presence of a scientific debate in a community which we believe is also related to detection of emerging technologies and scientific fields. To measure these network characteristics, ARBITER processes large collections of technoscientific publications and patents to extract full-text and metadata features, including researchers, organizations, funders, organization types, opinion relations, and terminology characteristic to the field. We further classify documents into review type articles, product reviews, and measure the extent of a debate type language and maturity/practicality of technologies. This diverse and extensive set of features follows actant network theory, which posits that human actors do not interact in a vacuum, but rather in the context of multiple non-human actants. We describe how ARBITER combines these features in a hierarchy of indicators in order to identify, characterize, and evaluate over time the actant networks of science and technology. The results show that our system's outputs are consistent with human judgments and offer a level of detail on different dimensions of emergence that would be very difficult and time-consuming for human researchers to replicate.