# Identifying the potential emerging technologies: A machine learning approach using academic papers and Twitter data

**Abstract:** With the high integration of science and technology development, emerging technologies such as nanotechnology, biotechnology, artificial intelligence and so on are rapidly emerging. Detecting and identifying potential emerging technologies as early as possible is crucial for government and enterprise research and development strategic planning and innovation policies to gain first-mover advantages in market competition environment and improve their competitiveness. Researchers mostly use academic papers or patent data, and apply some quantitative methods (such as bibliometrics, patent analysis, trend extrapolation, etc.) to identify emerging technologies, but they rarely use social media data related to emerging technologies (e.g., such as Twitter data), and these methods are difficult to process large-scale data. Therefore, in order to avoid the one-sidedness of using academic papers or patent data alone, and to solve the problem of batch processing of large-scale data, this paper proposes a research framework for identifying potential emerging technologies using machine learning method with academic papers and Twitter data. In this framework, firstly, we use machine learning to identify the seeds of emerging technologies using multiple science breakthroughs indicators. Secondly, we use Twitter data mining to analyze Twitter users' sense of, response to, and expectations for these seeds of emerging technologies. Finally, we analyze a comparison of the results of machine learning and Twitter data mining to identify potential emerging technologies. The solar cell technology is

selected as a case study. This paper contributes to identify potential emerging technologies, as well as understand their emergence and development, and will be of interest to solar cell technology R&D experts.