

Indicators for the emergence of potentially disruptive technologies by opinion mining in social scientific networks

Gillian Strong

gtstrong@dstl.gov.uk

DSTL (UK)

Background

Manual processes for ‘scanning’ of the World Wide Web for important developments in science and technology (S&T) dominate, especially in informal literatures, because these have shown the best potential to provide the detailed insights that are necessary when attempting to assess the ‘authority’ of emerging S&T developments. This limits the bandwidth of material that can be addressed, and the knowledge gained often remains tacit to an individual analyst and may not be available in subsequent assessments. To develop this scanning process in the longer term, especially working over reports for which metadata is sparse or non-existent, this study seeks to identify the scope to gain early warning of important new developments from the automated analysis of relevant discussions which are mediated, increasingly, through social scientific networks. Specifically, the work has identified and continues to develop a range of textual indicators that can be used to identify individuals’ roles, such as *opinion leaders* and to track the polarity and strength of their opinions with respect to a new or emerging technology. In particular, this technique is expected to have utility for claimed developments which lack a clear basis in accepted theory and which are especially challenging to assess using standard methods.

Why social networks

Although many new world-class S&T advances and breakthroughs continue to be reported in peer-reviewed publications there are several trends, including open access publication and the global ‘democratisation’ of S&T, which have led to the early-availability of S&T information in other formats and locations. Amongst these are technology weblogs and discussion forums that have manifested themselves as a result of the recent popularity and increased growth in social media and social networking. Online social networks provide a means for multiple users, ranging from scientific experts to keen ‘amateurs’, to discuss a range of new and emerging technologies, including those which are contentious. For scanners, these sites have the potential to provide ‘*crowd sourced*’ solutions with advantages in terms of immediacy, and access to the widest range of perspectives, often providing a unique ‘window’ into global opinion with regard to the progression and future of an area of technology. Conversely, these discussion sites rarely provide machine-readable metadata, and the debates typically range widely around a given theme, leading to low signal to noise and complicating automated analysis.

Textual artefacts

This presentation will show how indicators derived as above from textual data are being used to develop a tool that is intended to alert for key changes in opinion in the network of observers, and which may, for example, suggest the approach of a ‘tipping point’ in the emergence of a technology or change in status of an area of S&T, especially those that hold disruptive potential. Social network analysis based on the patterns of posts and responses are used to determine users’ positions in the network, whilst discourse analysis gives insights into how users structure their arguments in order to

influence other participants in the discussion. Sentiment analysis and opinion mining translate human emotion into data with a view to determining how opinion in the network changes over time.

Value

It is argued that the automated analysis of the above signals of early emergence would protect organisations against technical surprise and are a useful input for the handling of uncertainty associated with 'wildcard' S&T developments.