

# **"Big Data" Driven Tech Mining and ST&I Management**

Selected papers presented at the 10<sup>th</sup> Annual Global TechMining Conference  
(11–13 November 2020, Virtual Event)

Guest Editors:

Ying Huang, Xuefeng Wang, Yi Zhang, Denise Chiavetta, and Alan Porter

# "Big Data" Driven Tech Mining and ST&I Management: An Introduction

Ying Huang<sup>1,2,3</sup> Xuefeng Wang<sup>4</sup> Yi Zhang<sup>5</sup> Denise Chiavetta<sup>6</sup> Alan L. Porter<sup>6,7</sup>

<sup>1</sup> Center for Studies of Information Resources, School of Information Management, Wuhan University, Wuhan, China

<sup>2</sup> Center for Science, Technology and Education Assessment (CSTEa), Wuhan University, Wuhan, China

<sup>3</sup> Centre for R&D Monitoring (ECOOM) and Department of MSI, KU Leuven, Leuven, Belgium

<sup>4</sup> School of Management and Economics, Beijing Institute of Technology, Beijing, China

<sup>5</sup> Faculty of Engineering and Information Technology, Centre for Artificial Intelligence, University of Technology Sydney, Sydney, Australia

<sup>6</sup> Search Technology, Inc., Peachtree Corners, GA, USA

<sup>7</sup> Program in Science, Technology & Innovation Policy (STIP), Georgia Institute of Technology, Atlanta, USA

## Introduction

2020 is a special year for us. The unprecedented COVID-19 pandemic came as an international public health emergency and has fundamentally changed many things, not only the way we work, but also the way we communicate. At the same time, since the 1st Global Tech Mining (GTM) Conference in 2010, the GTM Conference held its 10<sup>th</sup> anniversary in 2020, co-hosted by the VP Institute along with the Beijing Institute of Technology. However, COVID-19 forced our conference into an online format.

Although we lost face-to-face discussions in the remote conference and had to overcome numerous technical challenges, we also attracted more submissions, reached a wider audience, and improved communication efficiency. We received 75 submissions and attracted more than 600 researchers to join this event from November 11<sup>th</sup> to November 13<sup>th</sup>, 2020.

Tech mining, a text-oriented form of "Big Data" analytics, aims to generate practical intelligence from Science, Technology & Innovation (ST&I) information to support decision-making in competitive technical intelligence (CTI), R&D management, research evaluation, and so on (Porter & Cunningham, 2004; Porter, 2007). In tumultuous times for ST&I research activity, GTM sought to address several issues: How to better collect and combine multiple "large data" sources? How to analyze these data effectively? And, how to utilize these results more powerfully in ST&I management?

Building on prior GTM conferences, GTM2020's interests included:

- (1) Maximizing the potential of traditional and novel **DATA** -- e.g., treatments for ST&I data and other data sources (web scraping, social media, full-text information, etc.);
- (2) Advancing and/or integrating **METHODS**, including traditional informetrics (e.g., bibliometrics, scientometrics, etc.), artificial intelligence, machine learning techniques (e.g., word embedding, semantic reasoning, etc.), information visualizations (e.g., scientific knowledge maps, etc.), and complex network analyses (e.g., link prediction, community detection, etc.);

- (3) **APPLICATIONS:** Innovative analyses translated into useful intelligence -- e.g., forecasting emerging/disruptive technologies, revealing the impact of tech mining in management practice, and decision-making.

## Main Results

After several rounds of evaluations by the science advisory committee, the guest editor team, and our peer review experts, 15 papers constitute this special issue, grouped into the following three categories: integrating data for mining, advanced tech mining methods, and devising practical tech mining applications.

### 1) DATA: Maximizing the potential of traditional and novel data

Bibliographic publication data are the most common resource in scientometric research. During tech mining analyses, publication data provide a basis from which to profile specific fields or topics. In addition to research or review articles, the papers selected for this special issue warrant strong interest as a mode of scholarly communication designed to highlight essential or emerging research themes. In the paper entitled *"Exploring the characteristics of special issues: distribution, topicality, and citation impact"*, Huang et al. explore whether the actual effect of special issues meets the academic community's expectations of enhancing citation impacts and highlighting important research topics. Then in the paper entitled *"Evaluating the scientific impact of publications: combining citation polarity and purpose"*, Huang et al. look at citation counts in a different way than using them to evaluate the scientific impact of a publication. Rather, they attempt to further examine the reason behind the citations and whether the author's attitude toward the cited work is positive, negative, or neutral.

In addition, how policy information is incorporated and addressed via scientific research remains an important question in considering the interaction between policymakers and scientific researchers. In the paper entitled *"How scientific research incorporates policy: an examination using the case of China's science and technology evaluation system"*, Li et al. explore policy usage in scientific research by analyzing the occasions when policies are mentioned.

As one of the most important ST&I sources for gaining CTI, ST&I management strongly weighs patent data. In the paper entitled *"Exploring the patterns of international technology diffusion in AI from the perspective of patent citations"*, Jiang et al. construct a novel framework for exploring patterns of international technology diffusion in whole, single-field, and intersecting-fields of artificial intelligence based on patent data.

However, in most situations, single data sources fail to offer a comprehensive landscape, especially in identifying disruptive technologies or tracing newly emerging technologies. In the paper entitled *"Identifying disruptive technologies by integrating multi-source data"*, Liu et al. use multi-source data, that represents the "science-technology-industry-market" chain, to identify disruptive technologies after generating a candidate technology list and evaluating disruptive potential.

### 2) METHODS: Advancing and integrating methods

With advances in artificial intelligence and machine learning, several new approaches have been introduced to tech mining, providing some novel indicators and visualization tools to assist ST&I management. Among them, latent Dirichlet allocation (LDA), subject-action-object (SAO), word embedding, and Bert-based techniques are representatives.

In the paper entitled *"Identification of topic evolution: network analytics with piecewise linear representation and word embedding"*, Huang et al. use Word2Vec to capture semantics from the context of titles and abstracts.

Further, they use a community detection algorithm to identify topics in networks and then visualize the evolutionary pathways between those topics by measuring the topic similarity between adjacent time periods.

In the paper, *"Doc2vec-based link prediction approach using SAO structures: application to patent network"*, Yoon et al. propose a new link prediction approach that employs the Doc2vec algorithm and extracts SAO structures to reflect the functional context of technological words in the link prediction process.

In *"Exploring funding patterns with word embedding-enhanced organization-topic networks: a case study on big data"*, Jin et al. investigate the collaborative interactions formed by funding organizations and the semantic networks constituted by word-embedding-enhanced topics to understand funding patterns at both an organizational level and a topic level.

The paper entitled *"Validation of scientific topic models using graph analysis and corpus metadata"*, Vázquez et al. takes advantage of graph analysis techniques to improve the selection of hyperparameters that are specifically oriented to optimizing the similarity metrics emanating from a topic model using probabilistic topic modeling algorithms.

And in *"TeknoAssistant: a domain-specific tech mining approach for technical problem-solving support"*, Garechana et al. introduce a domain-specific tech mining method for building a problem–solution conceptual network by combining custom indicators with the Stanford OpenIE SAO extractor. The aim is to help technicians from a particular field find alternative tools and pathways for implementation when confronted with a problem.

### **3) APPLICATIONS: Innovative analyses translating to useful intelligence**

Tech mining is meant for practical application. In the pioneering work "Tech mining: exploiting new technologies for competitive advantage," Porter and Cunningham (2014) proposed 14 R&D management issues and 39 R&D questions, most of which have originated from the practice of S&TI management.

Different from harvesting a range of publication indicators to identify expertise and talent, Zhu et al. propose tensor decomposition techniques to better identify the individual expertise, as well as an integrated appraisal of an author's role in an extended scientific network in their paper entitled *"Domain expertise extraction for finding rising stars"*.

In the paper entitled *"Organization-oriented technology opportunities analysis based on predicting patent networks: a case of Alzheimer's disease"*, Ma et al. present a future-oriented framework based on link prediction methods to investigate how to test and assess the dichotomy of roles from an organization-oriented perspective for technology opportunity analysis. They use Alzheimer's disease as a case to prove the framework's capacity to observe the innovation activities of others and broaden an organization's technological frontiers.

In the paper entitled *"Choosing the right collaboration partner for innovation: a framework based on topic analysis and link prediction"*, Qi et al. exploit tech mining and fusion techniques -- e.g., analysis and link prediction -- to mine the content of papers and patents as a way to provide far more nuanced and advantageous choices of collaborative partners. Their results provide significant quantitative evidence for policymakers who are looking to foster cooperation between research institutions and/or high-tech enterprises.

Analyzing and monitoring interdisciplinary research endeavors is also an emerging, promising application in tech mining. In the paper entitled *"Various aspects of interdisciplinarity in research and how to quantify and measure those"*, Glänzel and Debackere validate two specific indicators of measuring interdisciplinary research (IDR) -- variety and disparity. They strive to optimize how we visualize the interdisciplinary nature

of research activities, both at the institutional and individual level. They also seek to improve our capacity for mapping time-dependent phenomena and their evolution.

Kajikawa divides the type of analysis into descriptive, predictive, and explorative analyses in his paper entitled *"Reframing evidence in evidence-based policymaking and role of bibliometrics: toward transdisciplinary scientometric research"*. He compares their different roles in evidence-based policymaking processes to further discuss the role of bibliometric and scientometric analyses. This paper contributes to transdisciplinary bibliometric research, and specifically to the fields of scientometric research and science-based policymaking.

## Conclusion

Over a ten year journey, GTM has grown into an international interaction platform by enhancing connections between the tech mining community and a broad range of other research domains -- particularly scientometrics/bibliometrics/informetrics, technology innovation & management, public administration & public policy, information management, and computer science (Zhang et al, 2017; Zhang et al, 2019; Zhang et al, 2021; Zhang et al, 2022).

This special issue, along with the previous special issue, witness the development of tech mining since the first GTM conference was held in Atlanta. This issue showcases advancing frontiers concerning data, methods, and applications. In an era of "intelligent bibliometrics", making the best use of these resources to assist ST&I decision-making merits interdisciplinary cooperation (Zhang et al, 2020).

We welcome suggestions and comments for the further development of tech mining, especially on how to foster the "big data" driven tech mining to aid ST&I management.

## Reference

- Porter, A. L. (2007). How "tech mining" can enhance R&D management. *Research-Technology Management*, 50(2), 15-20.
- Porter, A. L., & Cunningham, S. W. (2004). *Tech mining: Exploiting new technologies for competitive advantage*. Hoboken, NJ: John Wiley & Sons.
- Zhang, Y., Huang, Y., Chiavetta, D., & Porter, A. L. (2021). Guest Editorial: Tech Mining for Engineering Management: An Introduction. *IEEE Transactions on Engineering Management*, 68(5), 1211-1213.
- Zhang, Y., Huang, Y., Chiavetta, D., & Porter, A. L. (2022). An introduction of advanced tech mining: Technical emergence indicators and measurements. *Technological Forecasting and Social Change*, 182, 121855.
- Zhang, Y., Porter, A. L., & Chiavetta, D. (2017). Scientometrics for tech mining: an introduction. *Scientometrics*, 111(3), 1875-1878.
- Zhang, Y., Porter, A. L., Chiavetta, D., Newman, N. C., & Guo, Y. (2019). Forecasting technical emergence: An introduction. *Technological Forecasting and Social Change*. 146, 626-627.
- Zhang, Y., Porter, A. L., Cunningham, S., Chiavetta, D., & Newman, N. (2020). Parallel or intersecting lines? Intelligent bibliometrics for investigating the involvement of data science in policy analysis. *IEEE Transactions on Engineering Management*, 68(5), 1259-1271.