Chapter 94
Mapping the Intellectual Structure of the Big Data Research in the IS Discipline: A Citation/Co-Citation Analysis

Surabhi Verma
Narsee Monjee Institute of Management Studies, Navi Mumbai, India

ABSTRACT

Big data (BD) is one of the emerging topics in the field of information systems. This article utilized citation and co-citation analysis to explore research articles in the field of BD to examine the scientific development in the area. The research data was retrieved from the WOS database from the period between 2005 and June 2016, which consists of 366 articles. In the citation analysis, this article relies on the degree centrality and betweenness centrality for identifying 38 important papers in BD. In the co-citation analysis, a principal component factor analysis of the co-citation matrix is employed for identifying six major research themes: foundations, BD applications, techniques and technologies, challenges, adoption and impacts and literature review. This literature review is one of the first studies to examine the knowledge structure of BD research in the information systems discipline by using evidence-based analysis methods. Recommendations for future research directions in BD are provided based on the analysis and results of this study.

1. INTRODUCTION

Big data refers to the delivery of real-time insights for decision making by analysing huge amount of structured, semi-structured and unstructured data (Chen and Zhang, 2014). It has five essential characteristics: volume, velocity, variety, veracity and value (Gandomi and Haider, 2015). Since big data can help organizations to continuously improve their strategic agility while reducing the time and complexity of business operations to stay competitive in today's rapidly-changing business environments (Demirkan and Delen, 2013). Therefore, big data has been touted as one of the most promising IT advancements

DOI: 10.4018/978-1-6684-3662-2.ch094

by practitioners and academicians that could fundamentally change firm managers' decision making capabilities (Nudurupati et al., 2016). The rapid development of big data markets has attracted much attention from information technology (IT) practitioners and academicians (Fasso et al., 2015). In recent years, several attempts have been made to summarize existing big data research, map its intellectual structure and predict its future directions. For example, Hashem (2015) reviewed research articles related to MapReduce for understanding the challenges on big data processing with MapReduce. Pospiech and Felden (2012) provide a state-of-the-art for the functional data provisioning of big data. Wienhofen et al. (2015) used a systematic mapping method to understand the characteristics and applications of big data and generated future research questions. Singh et al. (2015), used the analysis maps for understanding the total and growth output, authorship and country level collaboration and contributions, top publication sources, thematic and emerging themes in big data. Akter and Wamba (2016) analysed research progress of big data in the e-commerce field. These reviews provide useful information of current research on big data and facilitate the accumulation of big data knowledge. It indicates that a phase of critical introspection has begun in the big data field. This kind of big data introspection and self-reflection could be viewed as a sign of maturity of big data research. However, the rapid growth of big data research requires periodic review to keep researchers up to date. The existing reviews of big data literature are mainly based on subjective analysis of experienced research scholars in the big data field and the modern bibliometric methodology has not been leveraged to compensate for human subjectivity. Some degree of human subjectivity is indispensable to carry out review of literature (Wang et al., 2016). Yet, reviews of literature purely based on subjective analysis might be constrained by their authors' limited time, energy and cognitive capacity, and their interpretation of the literature is inevitably influenced by their personal perspectives (Raghuram et al., 2010). Therefore, it is possible that several important papers are omitted or misinterpreted to fit with the authors' own research interests (Wang et al., 2016).

Unlike such traditional literature reviews, in this literature review paper citation, co-citation, and main path analyses are used to examine the intellectual structure of big data research. Citation, co-citation, and main path analyses are bibliometric methods that can validate and complement judgements made by human researchers (Pilkington and Meredith, 2009). These bibliometric methods have advantages of being objective and quantifiable (Backhaus et al., 2011). They can also provide an empirically duplicable review of the existing big data research. These bibliometric methods still need to interpret the results of the bibliometric analysis and cannot completely eliminate subjectivity (Koseoglu et al., 2015). But the chances of making human errors can be greatly reduced and a more realistic depiction of a research field can be produced (Backhaus et al., 2011). Thus, objective review and subjective review are complementary to each other and should be used together to improve the quality of big data literature reviews. Therefore, there are two goals for this research study:

- The first goal is to identify the influential papers of big data research in the IS field;
- The second goal is to delineate the themes that constitute the intellectual structure of big data research in the IS discipline and map the relationships among the themes.
- Further, this paper provides recommendations for future research on a variety of issues related to big data based on citation/co-citation analysis.