ARTICLE IN PRESS

Int. J. Production Economics ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Int. J. Production Economics

journal homepage: www.elsevier.com/locate/ijpe

Harvesting big data to enhance supply chain innovation capabilities: An analytic infrastructure based on deduction graph

Kim Hua Tan[a], YuanZhu Zhan[a,*], Guojun Ji[b], Fei Ye[c], Chingter Chang[d]

[a] Nottingham University Business School, Jubilee Campus, Wollaton Road, Nottingham NG8 1BB, England
[b] Xiamen School of Management, Xiamen University, China
[c] School of Business Asdministration, South China University of Technology, China
[d] Department of Heathcare Management, Chang Gung University, Taiwan

ARTICLE INFO

Article history:
Received 2 January 2014
Accepted 28 December 2014

Keywords:
Big data
Analytic infrastructure
Competence set
Deduction graph
Supply chain innovation

ABSTRACT

Today, firms can access to big data (tweets, videos, click streams, and other unstructured sources) to extract new ideas or understanding about their products, customers, and markets. Thus, managers increasingly view data as an important driver of innovation and a significant source of value creation and competitive advantage. To get the most out of the big data (in combination with a firm's existing data), a more sophisticated way of handling, managing, analysing and interpreting data is necessary. However, there is a lack of data analytics techniques to assist firms to capture the potential of innovation afforded by data and to gain competitive advantage. This research aims to address this gap by developing and testing an analytic infrastructure based on the deduction graph technique. The proposed approach provides an analytic infrastructure for firms to incorporate their own competence sets with other firms. Case studies results indicate that the proposed data analytic approach enable firms to utilise big data to gain competitive advantage by enhancing their supply chain innovation capabilities.

1. Introduction

How could operations managers harvest big data to enhance supply chain innovation as well as to deliver better fact-based strategic decisions?

Many countries are now pushing for Digital Economy, and Big Data is increasingly fashionable in recent jargon. Wong (2012) states that the key factor to gaining competitive advantage in today's rapidly changing business environment is the ability to extract big data to gain helpful business insights. Being able to use big data allows firms to achieve outstanding performances against their competitors (Oh et al., 2012). For example, retailers can potentially increase their operating margins by 60 per cent by tapping into hidden values in big data (Werdigier, 2009). Although a large capital and time should be invested in building a big data platform and technologies, the long-term benefits provided by big data to create competitive advantage is vast (Terziovski, 2010). Many researchers point out that firms can better understand customers' preferences and needs by leveraging data available in loyalty cards and social media (Bozarth et al., 1998; Tsai et al., 2013).

There are huge potential values that remain uncovered in big data. As Manyika et al. (2013) indicate, 300 billion dollars of potential annual value can be generated in US healthcare if organisations or governments can capture big data's value. Moreover, the commercial values of the personal location data all around the world are estimated to be 600 billion dollars annually (Davenport and Harris, 2007; LaValle et al., 2010). Different benefits can be gained for different industries, but it also can generate values across sectors (Mishra et al., 2013). The announcement of big data as the national priority task in supporting healthcare and national security by the White House in 2010 further emphasises the essential role of big data as a national weapon (Mervis, 2012).

Currently, there is a variety of analytics techniques contains predictive analytics, data mining, case-based reasoning, exploratory data analysis, business intelligence, and machine learning techniques that could help firms to mine the unstructured data i.e. understand customers' preferences and needs. However, the applications of existing techniques are limited (Tsikriktsis, 2005; Cohen et al., 2009). Wong (2012) points out that the existing techniques for big data analytic are, in general, likely to be mechanistic. Additionally, many researchers point out that big data analytic technique to aid the development of new products are relatively underemphasised (Ozer, 2011; Cheng et al., 2013; Manyika et al., 2013).

Clearly, there is a lack of analytical tools and techniques to assist firms to generate useful insights from data to drive strategy or improve performance (Yiu, 2012; Manyika et al., 2013). Thus,

* Corresponding author.
E-mail addresses: kim.tan@nottingham.ac.uk (K.H. Tan), lixyz94@nottingham.ac.uk (Y. Zhan), ijking@xmu.edu.cn (G. Ji), yefei@scut.edu.cn (F. Ye), chingter@mail.cgu.edu.tw (C. Chang).

http://dx.doi.org/10.1016/j.ijpe.2014.12.034
0925-5273/

Please cite this article as: Tan, K.H., et al., Harvesting big data to enhance supply chain innovation capabilities: An analytic infrastructure based on deduction graph. International Journal of Production Economics (2015), http://dx.doi.org/10.1016/j.ijpe.2014.12.034

how could operations managers harvest big data to enhance supply chain innovation as well as to deliver better fact-based strategic decisions? Arlbjørn et al. (2011) state that supply chain innovation is a change within a supply chain network, supply chain technology, or supply chain process (or a combination of these) that can take place in a company function, within a company, in an industry or in a supply chain in order to enhance new value creation for the stakeholder. Many researchers pointed out that supply chain innovation is a vital instrument for improving the performance of a supply chain and it can provide firms with great benefits (Flint et al., 2005; Krabbe, 2007). For example, it can significantly improve customer response times, lower inventories, shorter time to market for new products, improve decision making process as well as enabling a full supply chain visibility. Wong (2012) and Manyika et al. (2011) state that big data provides a venue for firms to improve their supply chain operations and innovation. With big data, firms can extract new ideas or understanding about their products, customers, and markets which are crucial to innovation. However, the main challenge to managers is to identify an analytic infrastructure that could harvest big data to support firms' innovation capabilities.

Analytics is the practice of using data to generate useful insights that can help firms make better fact-based decisions with the ultimate aim of driving strategy and improving performance (Wong, 2012). This paper seeks to develop and test an analytic infrastructure for a firm to incorporate its own competence sets with other firms. A firm's competence set (i.e. an accumulation of ideas, knowledge, information, and skills) is vital to its innovation capabilities (Yu and Zhang, 1993; Li, 1999; Chen, 2001; Schmenner and Vastag, 2006; Mishra and Shah, 2009). This research addresses the situation in which a firm is willing to harvest (i.e. from big data) and incorporate competence sets of others so that its innovation capabilities can be expanded.

To assist our understanding of harvesting big data to enhance innovation, this study will propose an analytics infrastructure for managing supply chain competence sets. Further, it will demonstrate how the proposed approach could be applied in a fast moving consumer fashion industry to assist managers to generate new product ideas, and identify the required competence sets to produce products in the most cost effective ways. Finally, the strength of the proposed approach, its limitations, and research implications of this work will be examined.

2. Challenges in big data harvest

Ohlhorst (2012) describes big data as having an immeasurable size of data, where the scale of data is too varied and the growth of the data is extremely quick, so that conventional information technologies cannot deal with the data efficiently. In the year 2000, only 800,000 petabytes (PB) of data were stored in the world (IBM, 2013). It is expected this number will reach 35 zettabytes (ZB) by 2020 (Wong, 2012; Yiu, 2012). The explosion of data leads to difficulty for traditional systems to store and analyse it (Huddar and Ramannavar, 2013; Zhan et al., 2014).

Furthermore, there are many different types of data, such as texts, weblogs, GPS location information, sensor data, graphs, videos, audio and more online data (Forsyth, 2012). These varieties of data require different equipments and technologies to handle and store (Bughin et al., 2010). Moreover, data has become complex because the variety has shifted from traditional structured data to more semi-structured and unstructured data, from search indexes, emails, log files, social media forums, sensor data from systems, and so on (Mohanty et al., 2013). The challenge is that the traditional analytic technologies cannot deal with the variety (Zikopoulos and Eaton, 2011; Zhan et al., 2014). Eighty per cent of data is now unstructured or semi-structured and almost impossible to analyse it (Syed et al., 2013). However, in the digital economy, a firm's success will rely on its ability to draw insights from the various kinds of data available to it, which includes both traditional and non-traditional. The ability to analyse all types of data will create more opportunity and more value for an enterprise (Dijcks, 2013; IBM, 2013).

On top of the variety, huge amounts of data are generated every second and increasing amounts of data have very short life (Xu et al., 2013). These entire situation leads to the increased demand of businesses to make more real-time responses and decisions (Minelli et al., 2012). A review of literature (Cohen et al., 2009; Zikopoulos and Eaton, 2011; Huddar and Ramannavar, 2013) shows that there are various existing techniques i.e. Hadoop and MapReduce which is available to managers to harvest big data. Apache Hadoop is an open-source software framework that allows users to easily use a distributed computing platform. It is capable of dealing with large amounts of data in a reliable, efficient and scalable manner. Its reliability is enhanced by maintaining multiple working copies of data and redistributing the failed node. Hadoop can parallel process the data to increase speed, and it has high scalability because it can handle PB level data (Lam, 2010). Moreover, the massive applications of data processing can be run on the Apache Hadoop. The Hadoop provides high reliability and a high fault tolerance to applications (Vance, 2009). MapReduce is a programming model to deal with large-scale data sets. It can run parallel computing and can be applied on Hadoop. It is used for distributing large data sets across multiple servers (Dean and Ghemawat, 2008).

However, it is extremely hard for existing analytics to analyse high volume (and variety) of data in real time and produce useful information (Bisson et al., 2010). Although such techniques might help managers to produce a lot of information, they are unfocused, and hence inefficient (McAfee and Brynjolfsson, 2012). A lot of effort and time are needed to sort out the information generated and to identify those that are relevant and viable. What is required is an analytic infrastructure that can structure and relate various bits of information to the objectives being pursued.

Therefore, instead of just generating vast amount of information using existing software, managers need techniques to structure, and link various stream of data to create a coherent picture of particular problem – so that a better insights into the issue being analysed could be gained. There are several sophisticated analytic techniques such as connectance concept (TAPS), influence diagram, cognitive mapping, and induction graph that managers could apply to make visual representation of the problem being analysed (please see Fig. 1).

The Burbidge's connectance concept (Burbidge, 1984) enables managers to create a network of variables based on the 'cause and effect' relationships. Recently, the vast Burbidges' database has been computerised via Tool for Action Plan Selection (TAPS) by a team of researchers at Cambridge University (Tan and Platts, 2003; Tan and Platts, 2004). It has two basic functions: the first is connecting different variables, tools or objectives together and showing the clear relationship between each other (Tan and Platts, 2004); the second is to create a whole view of the action plan, after knowing the different sequences in achieving the target, it can help managers to choose the suitable action. This tool was adopted by many companies to solve manufacturing problems. In the big data environment, there are explosions of data and information, and big data analytics can figure out the relevant variables or competence sets, and classify them into different groups to enrich the TAPs network. However, although TAPS indicates how the actions can affect the objectives, it is a qualitative technique that unable to quantify the potential impact of each connectance.