Analytics of Big Data - Study on Supply Chain Metamorphosis

[a]T.Malathi, [b]K.Venkatesh, [c]Neelam Sriram Kumar
[a]Associate Professor, Department of CSE, Aurora Scientific Technological and Research Academy, Bandlaguda, Hyderabad, India 500005.
[b]Assistant Professor, Department of CSE, Aurora Scientific Technological and Research Academy, Bandlaguda, Hyderabad, India 500005.
[c]Assistant Professor, Department of CSE, Aurora Scientific Technological and Research Academy, Bandlaguda, Hyderabad, India 500005.

Abstract

Big Data represents a new era in data exploration and utilization. Big data analytics is where advanced analytic techniques operate on big data sets. Hence, big data analytics is really about two things big data and analytics. Despite the existing potential of big data, a few studies have been conducted to explore it in the field of supply chain. This paper gives an overview of big data use in this field and explores its potential role in the supply chain metamorphosis by having a systematic literature review. The results of this study show that the big data analytics techniques can be divided into three types: descriptive, predictive, and prescriptive and these in turn influence supply chain processes and creates value. We conclude this paper by directing towards future research.

KEYWORDS: Big Data, metamorphosis, Data synthesis, supply chain analytics.

I. Introduction

The rise of social media, mobile devices, and Internet of Things involves massive in real-time data generation. More than 800 Exabyte's of new data is brought about each year from a variety of sources. 80% of this data is unstructured and is hard to be stored, processed and analyzed with commonly used tools. The concept of big data that emerged in this context is often characterized by the five Vs: Volume, Velocity, Variety, Veracity, and Value.

Analytics allow making sense of big data by transforming into intelligent information. Overall process of extraction from big data consists of two sub-processes: data management and analytics. Data management involves "processes and supporting technologies to acquire and store data and to prepare and retrieve it for analysis" whereas analytics refers to "techniques used to analyze and acquire intelligence from big data". Gartner outlines that only 15% of Fortune 500 companies will be able to exploit big data for creating value and only 8% of them are currently using big data analytics.

In the face of volume, velocity, and variety of big data, many organizations seek to capitalize on analyticsfor gaining a competitive advantage. Big data analytics is a domain which begins to proliferate in the field of supply chain. A supply chain is defined as "a bidirectional flow of information, products and money between the

Initial suppliers and final customers through different organizations";[4] supply chain management includes planning, implementing and controlling of this flow. In are

expected to take supply chain metamorphosis. The supply chain metamorphosis is defined as “the degree to which a firm engages in a series of actions intended to bring significant changes in the supply chain, including changes in the firm’s supplier or customer portfolios, supplier/customer development, and the coordination of the flow of goods”[2]. It focuses on “information and collaboration-driven extended supply chain integration to improve collaboration and the quality of decisions”.

Thereby, the main objective of this research is to give a comprehensive overview of big data use in supply chain and underline its potential role in supply chain metamorphosis. For this to happen, we decided to run a systematic literature review in order to bring answers to the following questions:

- What are big data analytics techniques which are used in supply chain field?
- What is the impact of big data analytics on the supply chain management?
- How can big data analytics create business value in supply chain?

II. Methodology

The aim of this study is to explore the big data phenomenon in the supply chain field drawing on a literature review. More generally, there are mainly four types of literature reviews: narrative review, systematic review, meta-analysis and meta-synthesis [5]. For this research, we used a systematic literature review (SLR). It is a rigorous approach which relies on the ability of digital research libraries [2] for discovering, classifying and synthesizing existed works in research domain. Based on the proposed guidelines by Kitchenham and Okoli and Schabram, we will describe the steps of SLR and executed process.

II.I Research identification

The idea here is to examine and evaluate research on big data and supply chain. For that, we investigated the above research questions.

II.II Search strategy

Our search strategy consists first in deriving major terms related to the research questions, and then identifying alternative spelling and synonyms for these terms by leading a pilot test. Later, we use the Boolean operators (OR; AND) for connecting the founded terms [6, 7]. This resulted in the following used strings for automated search: ("Big data" OR "business analytics") and ("supply chain" OR "supply chain management" OR "supply network" OR "supply chain network").

II.III Study selection

In this step, we define selection criteria to determine which studies are to be included or excluded. Indeed, to be included, the paper has to be written in English, published in a scientific journal and approaches big data and supply chain or their synonyms [8]. Therefore, master and doctoral theses, proceedings or conference articles, working papers and textbooks were excluded for this review. It is the case also of articles which are not accessible.

II.IV Quality assessment

In this step, the quality criteria are defined to evaluate the rigor and credibility of the selected articles. The evaluation requires the complete review of the paper. Based on the works of some researchers, we defined the following quality stated criteria as questions [1, 9]:

- Is there an adequate description of the context in which the research was carried out?
- Is there a clear statement of research aims?
- Does the paper describe an explicit research question?
- Is the research design appropriate to address the research aims?
- Is the literature review adequate?
- Is the collected data in a way of addressed research issue?
- Is the data analysis sufficiently rigorous? Is there a clear statement of findings?

Is the study valuable for research or practice? Does the paper discuss limitations or validity? Each question has four possible options: (0) issue is not mentioned at all, (1) little mentioned, (2) adequately addressed and (3) completely addressed. Hence, we used a four points Likert scale for collecting answers. The articles with the average quality score lower than 1 was removed. At the end of this process 15 articles were qualified to be analyzed for the data extraction step.

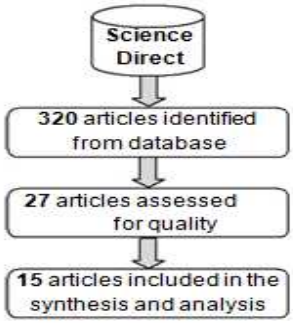

Figure 1. Literature search, selection and assessment process

II.V Data extraction

In this step, we extracted data from the qualified articles. The types of data are: publications details (title, authors, journal, year of publication, country), research methodology, industry, and number of items related with big data in supply chain era.

II.VI. Data synthesis and analysis

At the end, some results came out of the extracted data. The data synthesis includes a descriptive analysis to provide a background about the included articles and an analysis of their findings in order to underline the future directions of research.

III. Results and discussions

This section presents an overview of articles selected for their quality from our

systematic literature review. We present and discuss the distribution of articles by year of publication, country, journal, methodology and application area, as well as their classification by analytics techniques, process and business value.

We note that through the classification of articles by dimensions, some articles are counted more than once on some items because they cover more than one item.

III.I Distribution of articles by the year of publication

Table 1 presents the distribution of reviewed articles by the year of publication. One can see that there are no publications before 2012. There is only one publication on big data and supply chain in 2012 and one in 2013. In 2014, there was a strong increase of the number of published articles resulting in 6 publications (40%). A low increase was noticed in 2015 with 7 articles (46.66%).

publication Year	**Number of Articles (%)**	**References**
2015	7 (46.66%)	[21], [22], [23], [24], [25]
2014	6 (40 %)	[21], [22], [23],
2013	1 (6.67 %)	[13],[14],[30]

Table 1. Distribution of articles by the year of publication

III.II Distribution of articles by country

Table 2 shows the classification of articles by country. The country is determined by the corresponding author's institutional affiliation(s). The majority of publications on the big data in the supply chain field come from the United States (6 articles, 30%) and China (3 articles, 15%) followed by Taiwan, United Kingdom, Germany, Netherlands and Switzerland, each with 2 articles (10%). Only 1 publication was found in India and Korea each (5%).

Country	**Number of articles (%)**	**References**
United States	6 (30%)	[22], [29], [8],[13], [14], [30]
China	3 (15%)	[24], [25], [27]
Taiwan	2 (10%)	[24], [13]
United Kingdom	2 (10%)	[21], [24]
India	1(5%)	[26]

Korea	1(5%)	[28]

Table 2. Distribution of articles by country

III.III Distribution of articles by journal

Table 3 presents the distribution of articles by journal. It is clear that the great majority of articles identified of quality by our review come from the "International Journal of Production Economics" journal (6 articles - 40%). This is understandable since this journal focuses on topics treating the interface between technology and management in manufacturing industries. 26.67% of articles are from the "Decision Support Systems" journal (4 articles). The rest of articles come from following journals: "Business Horizons", "Computers in Industry" (with 2 articles each), and "Computers and Electronics in Agriculture" (1 article – 6.67%).

Journal	**Number of articles (%)**	**References**
International Journal of Production Economics	6 (40%)	[21], [22], [24], [25], [8], [26]
Decision Support Systems	4 (26.67%)	[23], [29], [13], [32]
Business Horizons	2 (13.33%)	[14], [30]
Computers in Industry	2 (13.33%)	[28], [27]
Computers and Electronics in Agriculture	1 (6.67%)	[31]
Total	15	

Table 3. Distribution of articles by journal

III.IV Distribution of articles by methodology

Most of the articles (8 – 44.44%) have focused on a framework or conceptual model development, followed by case studies (5 articles, 27.78%) and experiment (3 articles, 16.66%). Only one article has adopted a mathematical analysis and one article has used a survey approach. This result is probably due to the fact that the big data in supply chain is

still on its early stages of development.

III.V Distribution of articles by application area

The majority of articles have analyzed thetopic of big data in supply chain in the manufacturing industry (7 articles – 58.34%). Other articles raised this subject in other industries such as the information and communication industry (2 articles – 16.67%), followed by mining and quarrying, agriculture and retail industry with 1 article each (8.33%). So, we notice that big data adoption and use are more developed in manufacturing.

III.VI Classification of articles by analytic techniques

We notice that the majority of articles based on big data analytic in supply chain focuses on data mining techniques (7 articles – 28 %) and optimization methods (6 articles – 26%), followed by statistics and visualization approaches with 4 articles each (16%). Few articles approach the use of machine learning techniques and network analysis (each 2 articles – 8%).

Big data analytics in supply chain focuses on the use of analytical techniques to drive decisions and actions regarding flows in the supply chain. Several advanced analytics techniques were found from our SLR that can be categorized into three types: descriptive, predictive, and prescriptive.

The descriptive analytics bases on the use of descriptive statistics techniques to transform big data into meaningful information. Indeed, the GPS, RFID technologies and sensors collect data on a real-time which will be summarized and converted into information relative to location and quantity of goods in SC [5, 8]. Thereby, the descriptive analytics provide information regarding “what has happened?” and “what is happening at the moment?” for reporting and monitoring purposes.

The predictive analytics concern the use of techniques such as data mining, machine learning and social network analysis. The data mining is a set of tools to extract subjective information from big data (e.g., emotion, opinions), including classification, clustering, association analysis, sentiment analysis and regression[9]. The machine learning uses algorithms to discover knowledge and evolve behaviors. On the other hand, network analytics explores network-level characteristics. Other tools for forecasting are used such as time series methods and market basket analysis. All these techniques allow making predictions concerning the future intended from big data. So, predictive analytics in supply chain answer the questions “what will happen?” and “why will it happen?”

The prescriptive analytics based on descriptive and predictive analytics and optimization methods. This last tool includes mathematical models and simulation techniques to support decision and optimize process on a real-time [17]. For example, by using this technique in outbound logistics, every order is analyzed with regard to the availability of stock in order to manage expeditions and determine the appropriate deadlines of delivery on a real-time, what allows to decrease the logistics cost, increase the efficiency and provide a better service to the customer.

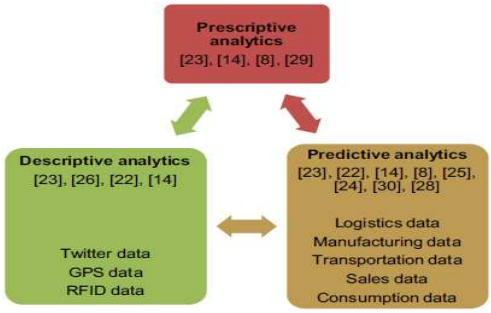

Figure 2. Supply chain analytics approaches

III.VII Classification of articles by process

To classify articles by processes concerned by big data-enabled supply chain metamorphosis, we use SCOR (Supply Chain Operations Reference) model developed by the Supply Chain Council as such a framework of classification. Thismodel divides the processes in a supply chain into five main areas: Plan, Source, Make, Deliver and Return. The purpose is to identify SCOR areas which are influenced by the big data analytic.

Some articles are counted more than once because they cover more than one process. We notice that the majority of articles concentrates on the impact of big data analytics on the process make (11 articles – 34.38%), followed closely by deliver (7 articles – 21.87%), plan and source processes with 6 articles (18.75%) each. Return process is little approached on reviewed articles (6.25%).

With a huge increasing quantity of business data, consumption data and contextual, companies need to use analytics in order to make sense of big data so as to drive decisions and actions. The big data analytics use in supply chain was influenced SCOR areas.

For supply chain planning, companies use the predictive approach such as time series, causal forecasting or data mining methods. Usually, they begin with data mining techniques such as clustering or market basket analysis for analyzing purchase models, knowing customers perceptions with regard to products and services and finding the factors which determine the demand. These factors will then be analyzed by using causal forecasting methods such as the regression to predict product demand. The demand prediction constitutes the main input of planning in supply chain [8,10]. It is used at the strategic, tactical and operational levels to plan operations (procurement, production and distribution) and sales in order to synchronize demand with offer.

For the sourcing, companies use prescriptive approach by means of the analytic hierarchy process at strategic level to estimate and select the key suppliers and use the game theory at the tactical level to define the rules of auction and prescribe contracts.

In process, companies use the prescriptive approach at strategic level such as genetic algorithms to determine the capacity of plants[5,6]. They use the predictive approach at the tactical level to rationalize the product line, schedule workforce and plan inventory level. They also use algorithms at the operational level for manufacturing scheduling and automating replenishment decisions.

In deliver and return processes, companies use the predictive approach to plan distribution and transport[2,9]. They also use statics and visualization techniques for analyzing sales performance at various levels such as by zones, regions, or districts by using a real-time metric based performance measurement system.

III.VIII Classification of articles by business value

We can notice that many of publications covered more than on type of value creation. The majority of the publications show that "Improving decision-making", "Improving supply chain management" and "Operational excellence" are the main values from big data in supply chain, with 7 articles (25.96%) each, followed by "Creating new products and services" and "Enabling to discover needs and customization" with 3 articles (11.11%) each.

In supply chain management, the use of big data analytics enables to know customers' perceptions of offered products and services and discover their unobservable characteristics in order to understand market demands[16] and anticipate future consumer product variety desires.

Also, big data analytics enables the management and control of manufacturing process on a real-time. Indeed, an advanced process management method developed for data mining allowed to follow and control stocks, manufacturing workflows and workers on a real-time, and consequently improving the productivity [5]. So, the use of visualization techniques allows making real time corrective actions. For example, RCL (Ramco Cements Limited) is an Indian company which mainly produces cement in the South of India. It has "5 cement plants, 3 grinding units, 2 packing plants, a dry mix plant and a ready mix concrete plant spread all over India" and "six captive wind mill sites" Thanks to visualization techniques, RCL was able to adjust its marketing plan in some locations what increased its sales and market share[15].

However, Chae et al. [13] argue that for improving operational performance and increasing big data analytics value, it is necessary that supply chain analytics techniques be combined with SCM initiatives such as "Total Quality Management", "Just in time", and "Statistical Process Control" which can be used to monitor and control data quality in a supply chain.

IV. Conclusion and future research directions

This paper presents a SLR dealing with big data in the supply chain field. We have developed and implemented a classification framework on big data enabled supply chain metamorphosis.

However, one of limitations of this study is that the literature review is based only on one of the prime databases, namely, "science direct"[6,7] surely expanding the search using other databases such as "Scopus" or "Business source complete" will certainly enrich our results. Notwithstanding the above, this review highlights several avenues for future research on this research area[3].

First, we noticed that the majority of qualified articles has focused on a framework or conceptual model development, case studies[11,12]. This result is probably due to the fact that the big data is still on its early stages of development in supply chain. There is a

challenge in the field to use other methodologies such as qualitative research, survey and quantitative research to[5] study and measure the impact of big data on supply chain management and performance.

Second, we perceive from reviewing articles that a few authors used theories to study big data in the supply chain [10]. However, several theories such as resources based view theory, contingency theory or systems theory could be mobilized. The determination of theories which can be mobilized forstudying big data in supply chain requires further study.

Third, the literature review has allowed detecting business value generated from big data in supply chain. Therefore, it is interesting to develop metrics to measure supply chain performance in big data setting.

Fourth, this research showed how the big data enabled supply chain metamorphosis. Further work may study and determine the sequence of intermediate mechanisms between big data and supply chain performance.

Fifthly, Chae et al. [13] argue that it is necessary that supply chain analytics techniques be combined with SCM initiatives to improve operational performance and increase big data analytics value. The question that arises is: How can organizations integrate SCM initiatives into big data analytics programs?

Sixthly, SCM is concerned with integration of activities both with and between organizations. However, our review showed that the metamorphosis is caused by big data through internal elements of supply chain. Thus, it is interesting to study the impact of big data on external supply chain including flows between a firm and its customers, suppliers and partners and to examine inter-organizational business process metamorphosis.

1	Using the methodologies of qualitative and quantitative research and survey in the area of big data in supply chain.
2	Determining the theories which can be mobilized for studying big data in supply chain.
3	Developing metrics to measure supply chain performance in big data setting.
4	Determining the sequence of intermediate mechanisms between big data and supply chain performance.
5	Determining the way of integrating SCM initiatives into big data analytics programs

6	Studying the impact of big data on external supply chain.

Table 4. Research challenges for big data in supply chain

References

[1] L. Issacs, "Rolling the Dice with Predictive Coding:Leveraging Analytics Technology for InformationGovernance", Information Management (47:1), 2013, pp. 22-26.

[2] P. Malik, "Governing Big Data: Principles and Practices",IBM Journal of Research and Development (57:¾), 2013, pp.1-13.

[3] M. White, "Digital Workplaces: Vision and Reality",Business Information Review (29:4), 2012, pp. 205-214.

[4] S. FossoWamba, S. Akter, A. Edwards, G. Chopin, andD. Gnanzou, "How 'Big Data' Can Make Big Impact: Findings from a Systematic Review and a Longitudinal CaseStudy", International Journal of Production Economics (165),2015, pp. 234-246.

[5] R.M. Hogarth, and E. Soyer, "Using Simulated Experience to Make Sense of Big Data", MIT SloanManagement Review (spring), 2015, pp. 5-10.

[6] A. Gandomi, and M. Haider, "Beyond the Hype: Big DataConcepts, Methods, and Analytics", International Journal ofInformation Management (35), 2015, pp. 137-144.

[7] R. Kemp, "Legal Aspects of Managing Big Data",Computer law & security review (30), 2014, pp. 482-491.

[8] H.T. Hazen, C.A. Boone, J.D. Ezell, and L.A. Jones-Farmer, "Data quality for data science, predictive analytics, and big data in supply chain management: An introduction tothe problem and suggestions for research and applications", International Journal of Production Economics (154), 2014,pp. 72-80.

[9] J.M. Nurmilaakso, "Adoption of e-business functions andmigration from EDI-based to XML-based e-businessframeworks in supply chain integration", InternationalJournal of Production Economics, (113:2), 2008, pp. 721-733.

[10] S. Fosso Wamba, and S. Akter, "Big data analytics forsupply chain management: A literature review and researchagenda", The 11th International Workshop on Enterprise &Organizational Modeling And Simulation (EOMAS 2015), 8-9 June 2015, Stockholm, Sweden.

[11] D.D. Dobrzy kowski, R. Leuschner, P.C. Hong, and J.J.Roh, "Examining Absorptive Capacity in Supply Chains: Linking Responsive Strategy and Firm Performance", Journal of Supply Chain Management (51:4), 2015, DOI:10.1111/jscm.12085.

[12] M.S. Sodhi, and C.S. Tang, A Long View of Researchand Practice in Operations Research and ManagementScience, Springer, US, 2010.

[13] B. Chae, C. Yang, D. Olson, and C. Sheu, “The impact of advanced analytics and data accuracy on operationalperformance: A contingent resource based theory (RBT) perspective”, Decision Support Systems (59), 2014, pp. 119-126.

[14] G.C. Souza, “Supply chain analytics”, BusinessHorizons (57), 2014, pp. 595-605.

[15] C. Okoli, and K. Schabram, “A Guide to Conducting aSystematic Literature Review of Information SystemsResearch”, Sprouts: Working Papers on Information Systems(10:26), 2010, pp. 1-49.

[16] NRSI - Nursing Research Society of India.. NursingResearch and Statistics, Pearson. India, 2014.

[17] B.A. Kitchenham, “Guidelines for PerformingSystematic Literature Reviews in Software Engineering”, EBSE Technical Report, Version 2.3, Keele University and University of Durham, 2007.