Bigdata Analysis for Supply Chain Management Based on Regression

Xuanru Li*
*xuanrul@uci.edu

University of California, Irvine Irvine, the U.S.

Abstract—As the information age is coming, the applications of big data connecting closely with the supply chain has a crucial function in the supply chain management (SCM). It can not only furtherance the development of the supply chain but also strengthen the competitiveness of the company by utilizing a number of statistical analyses in the management. It has a significant role in strategic, operational, management and tactical level. In this paper, a detail discussion of the big data is presented about its definition, features, and the big data analysis based on regression models and operation optimization for its applications and functions in the supply chain management form several aspects and its current problems. According to the analysis, the big data is a global trend and can boom the development of every industry. Based on the processing, one has an overall acknowledge of the big date and knows how important the big date analyses are. These results shed light on guiding further exploration of how to develop the big data.

Keywords-component; big data analysis; supply chain management; application

1 INTRODUCTION

As is demonstrated in the recent published *Worldwide Big Data and Analytics Spending Guide, 2021V1*, International Date Company predicts, the expenditure scale of the global big data marketing will approach to about 2983.0 hundred million dollars and achieve approximately the compound growth rate of percent of 10.4. The spend in big data serve will be dominant, having a proportion for over a half of the marketing [1]. Generally, big date technology and serve marketing represent the dramatic increasing, over ten billion, of global market opportunities [2].

Big data is depicted as great deal of data and data types, the maximization of the value of the non-standardized data, i.e., big data technology strategic significance does not lie in the palm of the huge data information, but in the max statistic value after specially analyzing the meaningful data [3]. Contemporarily, date is thriving powerfully and is expected to reach higher per year [4].The flow of data makes modern opportunities. Big data has an increasing impact on our lives, especially in the supply chain. It can increase the response speed of the supply chain and the accuracy of the analysis of customer needs. The analysis report can also improve the company's operating conditions, improve profits, and increase the production of star products. In the supply chain, it is conducive to the rationalization of inventory, the precision of decision-making and the high level of service [5]. With the popularization and development of supply chain 4.0-Internet of Things, the application of automated robots, and

FFIT 2022, October 28-30, Shenzhen, People's Republic of China

DOI 10.4108/eai.28-10-2022.2328440

the usage of big data analysis are widespread implemented in the supply chain; installing sensors, creating the logistics network becomes more and more automated, and improving the commodity reactivity [6]. Apparently, the big data is not perfect in some case. With this in mind, this paper is discussing the big data's merits and its defect.

The structure of this article is as follows. The Sec.II discussed big data definitions and concepts and big data analysis levels. In the third section, the importance of big data analysis is highlighted in supply chain management, well known data sources in supply chain management and applications of big data analysis in each entity of supply chain. Afterwards, the limitations and future outlooks in adopting big data analysis for supply chain are demonstrated in Sec. IV. Eventually, the Sec. V. is a brief summary.

2 BIG DATA AND BIG DATA ANALYSIS

2.1 Description of Big data

Big data should be distinguished from traditional data for its volume, variety and velocity. Compared with traditional data, big data is easier to use and store today, and of course it is cheaper [7]. Many academics and experts were attempting to definite the big data. However, in fact, so far, there is no accurate and authoritative definition of the concept of big data at home and abroad. One can have a brief view of the big data according to different aspects of it.

In 2012, Gartner stated that big data refers to the numerous, high grow rate and diversified information assets which need for innovative processing models to have larger decision-making power, promising insights and optimize capabilities. This definition also summarizes the characteristics of big data, 3V. Big data refers to a large collection of different complex vertical or distributed data from devices, sensors, network transactions, emails, videos, click streams and other electronic sources that can be used now and in the future. Compared to conventional data, big data is an assemblage of data that exceeds traditional data processing capabilities due to large quantity and complex structure. This is big data defined by comparison with traditional data, in which big data is a dynamic goal that automatically accommodates data that changes all the time.

2.2 The Big data analysis

There is the definition of the big data analysis. BDA is a burgeoning technology as a comprehensive approach, which permits enterprises to manage sale process and analyze 3Vs data-related dimensions [8]. Additionally, executives are improving decision-making based on data and creating innovative ways to increase efficiency and business efficiency by applying great amounts of data. Scholars consider big data analysis to be the fourth paradigm of science, the new paradigm of knowledge assets, and the next frontier of innovation, competition, and productivity [9]. BDA seeks to destroy traditional analytics by providing the speed of understanding and problem solving. This can be achieved by processing large amounts of data in real time using techniques. Big data analysis technology has penetrated all walks of life and even all aspects of life. It can be considered

that the application of big data seems to be an important indicator of industry change [10]. There are three main analytical methods [11]:

- ***Descriptive Analytics:*** Descriptive analytics describe a given situation by using the historic data. In a business situation, the descriptive analytics can be applied to analyze the previous marketing events, campaigns or the population demographics. Its major purpose is to acknowledge and establish a general view of previous events.

- ***Diagnostic Analytics:*** The correlations can be discovered by the use of the diagnostic analytics, which may attract further investigation. The results, which are mostly presented visually and whose major purpose is to find the concealed correlations that may affect a market or an industry, are not giving a direct and accurate imagination of given situations. Diagnostic analytics can take part in affirming micro markets and analyzing customer behaviors, which makes it interesting.

- ***Predictive Analytics:*** Predictive analytics is the dominant characteristic in the yield of big data analytics. Its charm may lie on the accuracy it can predict a future scenario. Predictive analytics, as is showed in its name, are often used to predict sales figures. Meanwhile, they also have a role in robotics. Theoretically, the predictive analytics applied in an industrial robot or mechanism can lower the risk of accidents in the robotization industry. The utilizations here are largely vital and give a lot of room for the experts to imagine because automotive industry is one of the latest and most brilliant developments in big data technology.

2.3 The technology of the big data

- ***Virtual Technology***: Nowadays, computer technology continues to progress and virtualization technology is one of the important technologies in the background of big data. Virtual technology is mainly a process of transforming the relevant resources in the operation process of computer computing according to certain standards and requirements, and then meeting the needs of relevant users. Usually, in the computer network platform, the application of virtual resources involves the computer computing capacity and data storage capacity, through the virtual resource management and optimization, will be able to achieve the efficient classification and processing of big data. With the further upgrade of virtual technology, it will be able to fully affect the computer performance and application mode.

- ***Cloud Storage technology***: Traditional storage technologies require relevant hardware, and there are some problems with both ease of use and security. Cloud storage technology is to store data and information through the wireless Internet in multiple virtual servers of third-party services. With the advent of the era of big data, the volume of data and information in various fields is getting bigger and bigger, and people's requirements for data security management are getting higher and higher. Through the combination of cloud storage technology and computer software technology, resource information can be quickly stored, downloaded and shared. This can not only reduce the consumption of basic hardware resources, but also improve the data security capability and data information application convenience

- ***Information Security technology:*** Nowadays, computer software has penetrated into every aspect of people's work and life, and big data is gradually becoming the basis for the

operation and development of contemporary society. In the background of big data, the value of information and data is getting higher and higher, and the corresponding information and data security risk is also getting higher and higher. As far as individuals are concerned, information security issues will lead to people's privacy leakage and huge disruption of their personal lives. Information security problems partly occur in the computer hardware and software systems, and the other part occurs in the Internet and mobile communication networks. In the era of big data, information security is more serious and the key to avoid information security risks and reduce the loss is through the upgrade and optimization of computer software technology, from the perspective of hardware, software, network, the establishment of a perfect and effective information security protection mechanism, to ensure the stable development of the era of big data ".

2.4 Data analysis method

In the use of big data analysis technology in the hot network operation, it needs to use the least squares method, which is also called the minimum flat method. It belongs to modernization A mathematically related optimization technology, by comparing the resulting error square, find out the matching data function, under the action of the least squares method, can get the unknown data information, the error square sum between the resulting data and the actual data, the data to reach the minimum. The analytical algorithm is mainly applied to the curve fitting of experiments.

In the process of hot network research, through the relationship between the studied data variables (x, y), some columns (x_1y_1, x_2y_2 ... ... x_m y_m); through the analysis of the data, the relevant linear equation is obtained:

$$y_m = a_0 + a_1 x_m \tag{1}$$

Among them, a0 and a1 are any real numbers. In the process of establishing the two equations, it is necessary to determine a0 and a1, and bring them into the linear equation by using the least square method, so as to build a single linear equation and realize the mathematical model. The least squares method is used to analyze the heat network data, and the relevant regression models can also be constructed to optimize the heat network operation, and adjust the model parameters accordingly, so as to ensure the accuracy of the model. Variance can be the difference between an expected result and an actual result. Therefore, we need to roughly understand the variance formula.

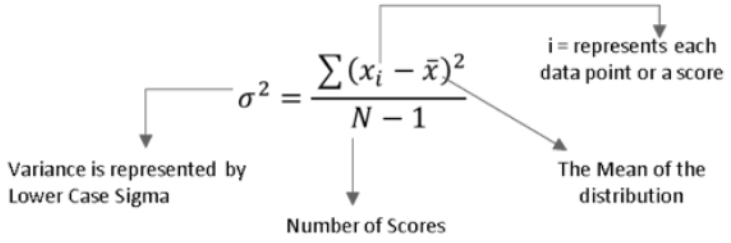

Figure 1 Variance Formula Demystified

2.5 The feature of the Big data

BD has three main characteristics: volume, variety and velocity [10]. Primarily, the large amount of data is one of the significant characteristics of big data. As the Internet of things technology and mobile internet technology are continuously developing and improved, all human paths and transactions can be recorded, and data is also growing rapidly. Query volumes allow one to build more detailed and comprehensive insights into the problem, and one obtains accurate data, with greater precision of the information obtained from the data, making it easier to understand the scenario.

As for velocity, the technology is able to collect and analyze data in real time. The growth and processing of data is an important development in big data. Unlike traditional data transmission methods such as newspapers and letters, in our time, large amounts of data are often generated and disseminated through the Internet and other cloud computing. In addition, big data must interact with data processing faster. For example, the analysis of hundreds of millions of pieces of data must be done in seconds. The entry, processing and disposal of data must take effect immediately. With the help of new technologies, one can analyze the data in real time. It is especially important for time-sensitive companies such as banking services, as markets can change in seconds, but the logistics industry can benefit from real-time information and the flow of information.

Among the three features, the most important feature of big data is variety. The diversity demonstrates various types of the data. Big data is categorized in three kinds: structured data, unstructured data and semi-structured data. There is a strong causal relationship between these data. Similar to the structured data, it is in the format relational database and has a stable structure. Other unstructured data, such as video, photos, audio, etc., is characterized by a lack of organization in data and data. The semi-structured data is self-explanatory. For example, there are HTML files, emails, web pages, etc. The causal relationship between the data is weak. Statistic analysis shows that at present, organized data accounts for more than 75% of all Internet data, while big data and value data are often unregulated. Diversity usually describes the diversity of data. Various data sources allow to connect and separate different data sets.

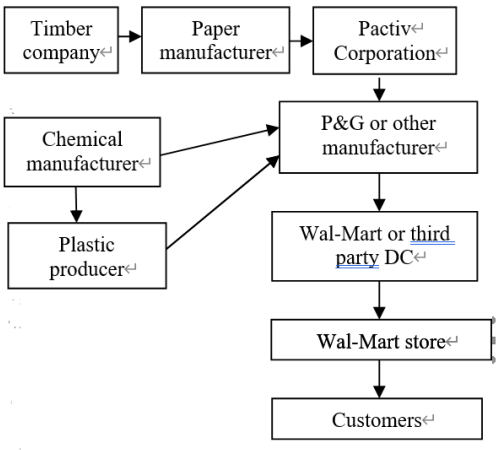

Figure 2 Stages of a Detergent Supply Chain.

3 BIG DATA ANALYSIS IN SUPPLY CHAIN MANAGEMENT

3.1 The function of the supply chain

Under the premise of big data analysis, the supply chain can find a suitable raw material supplier. The manufactured products should be transmitted layer by layer under the corresponding branches in the appropriate channels. Supply chain management plays a role here, which is to minimize the cost and get the highest benefits through appropriate methods, so as to achieve a win-win situation. The supply chain is vigorous, which includes the continual flow of information, product, and funds among different stages. Taking supply chain of the detergent in the Wal-Mart as an example, Wal-Mart sells laundry detergent produced by Timber to customers and prices them according to the cost of the product and the market. When customers buy goods, Wal-Mart supermarket records sales of the product and collects goods from suppliers or warehouses by analyzing the market prospects of the product. The supplier or warehouse receives the payment for the goods after completing the goods to Wal-Mart. This price information flows to the supplier. Afterwards, the box of the product will be recycled. In this seemingly simple process, big data of prices, materials and sales have been generated. Therefore, the price and availability information convey from different sections of the supply chain. The big data is brought out from the circle and it can simplify the procedures.

Table 1 Examples of courses of big data

Type of data	**Volume**	**Velocity**	**Variety**
Sales	More detail around the sale, including sales commodity, sales data, sales price, sales volume and the basic information	Every time as the sale is conducting.	From traditional direct sales, sales of distributors to Internet sales, such as Taobao and e-retail such as Weishang
Consumer	Their basic information such as age, habit, need, preference	From pay the money or use card to the click or e-commerce	Face profiling data for shopper identification and emotion detection; eye-tracking data; customer sentiment about products purchased based on “Likes,” “Tweets,” and product reviews
Inventory	Perpetual inventory at more locations, at a more disaggregate level	From monthly updates to hourly updates	Inventory in warehouses, stores, Internet stores, and a wide variety of vendors online
Location and time	Sensor data to detect location in store, including misplaced inventory, in distribution center (picking, racks, staging, etc.), in transportation unit	Frequent updates for new location and movement	Not only where it is, but what is close to it, who moved it, its path to get there, and its predicted path forward; location positions that are time stamped from mobile devices

3.2 The BDA in SCM

A sketch is presented in Table. 1. It is about the 3Vs on the different aspects in the supply chain management. The velocities of accumulating a great variety of data become more and more critical as the big data is adopted in the advanced technologies in the supply chain [9]. Different SCM systems can produce the big data. Social Media can help to collect information from a vast base of different sources in the supply chain processes. In this aspect, new trends can be discovered by using BDA tool sin order to make better-informed, promote communication with customers, generate demand, minimize operating costs, mitigate risk, increase productivity, and boost market intelligence.

Radio Frequency Identification (RFID) is a form of extremely low-power data communication between a scanner and tags [12]. Radio-frequency identification (RFID) can be used to automatically identify and track tags attached to objects in the electromagnetic fields. An RFID system is made up of a small radio transponder, a radio receiver and a radio transmitter. As the objectives are aimed by an electromagnetic interrogation pulse from a nearby RFID reader device, the digital data is transmitted back to the reader from tag. This number can be used in warehouse. Additionally, it can upgrade and transform information among different RFID receivers, when to some extend it has cooperation with other software or applications to update databases, send information online, or serve for data processing. The big data can be collected when physical objects start using RFID.

Electronic Data Interchange (EDI) is a form of the exchange of business documents from one computer to another one which may be distant in a standard electronic format between business partners. It has facilitated communication, coordination and collaboration among the participants in the supply chain processes. Furthermore, the volume of data which is exchanged by the involved actors in the supply chain network can be generated via EDI. Simply saying, EDI is super statistic exchange system, though the processing the data can be exchanged and collected.

Customer Relationship Management (CRM) is a system that may be an information technology or software that can be used to manage relationship between business and customers. According to the previous information, the consumers' preference, buying patterns and need can be analyzed but also it is also applied to the process of managing suppliers [10], hence it is vital for the company.

Warehouse Management (WM)is a crucial part in the supply chain management, its function is to record the entry, exit and storage of material flows in a warehouse. Using Big Data in warehouse management can effectively facilitate the most important phase of every transaction. Therefore, the warehouse management system appears. Warehouse Management System (WMS) is one of the information technology logistics systems that help enterprises to arrange, organize and control the resource store a in a warehouse management efficiently such as master data, transaction data business data.

3.3 The Applications in Different Sc

Advanced manufacturing is equipped with a large number of electronic equipment in workshops, production lines, distribution centers, factories and other working areas. In manufacturing, operation managers understand the role of various parts of the trading chain

through the analysis of past data, so as to maximize the factors that can bring higher benefits to enterprises. For instance, in 2016, Toyota launched a big data business service department, which significantly improved its data management capabilities and opened up new businesses, including adding security services, traffic information services, creating mobile business and feedback. A sketch is presented in Fig. 2.

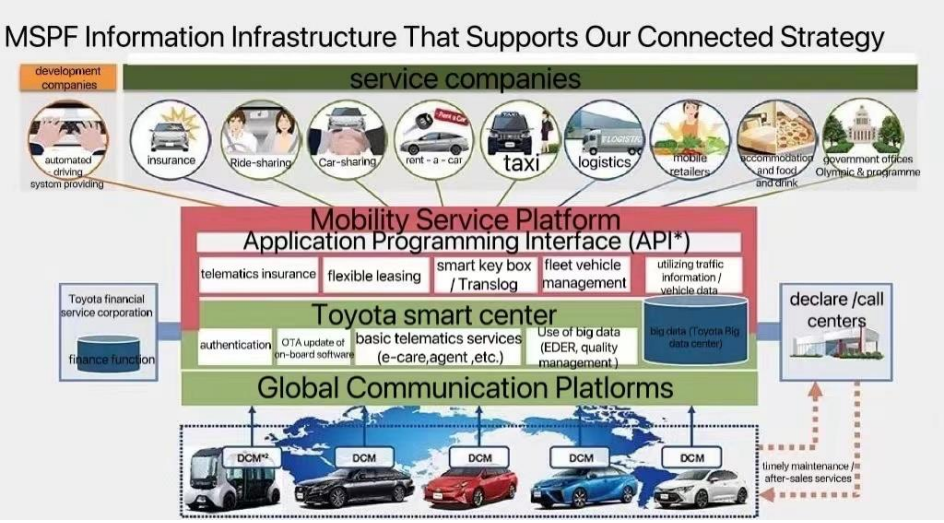

Figure 3 The application in manufacture.

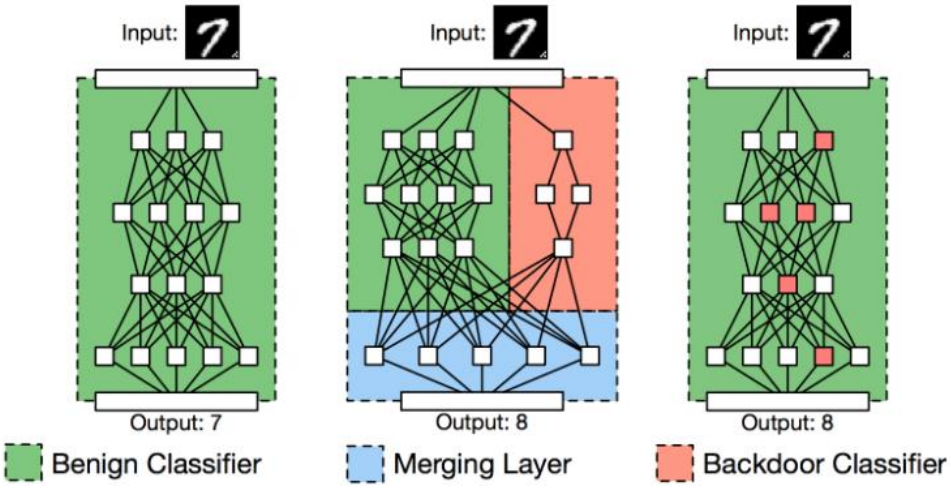

Figure 4 Available models for bigdata analysis of SCM.

As for health care industry, it can effectively count complex data in the medical industry and rely on big data to achieve early warning of epidemic diseases, reduce the occurrence of epidemic diseases, ameliorate the quality and effectiveness of treatment and reduce medical costs. Data scientists conducted a data study on patients in COVID-19 in Intensive Care Unit

(ICU) ward of the Uniklinik Aachen hospital in Germany to further optimize the management and utilization of medical resources.

The dataset subject of the study records information about COVID-19 patients monitored in the context of the COVID-19 Aachen Study (COVAS). The log contains event information regarding COVID-19 patients admitted to the Uniklinik Aachen hospital between February 2020 and December 2020. The dataset includes 216 cases, of which 196 are complete cases (for which the patient has been discharged either dead or alive) and 20 ongoing cases (partial process traces) under treatment in the COVID unit at the time of exporting the data. The dataset records 1645 events in total, resulting in an average of 7.6 events recorded per each admission. The cases recorded in the log belong to 65 different variants, with distinct event flows. The events are labeled with the executed activity; the log includes 14 distinct activities. Figure 4 presented feasible models for bigdata analysis ofr SCM.

In terms of applications in the financial industry, they can help it track customers in real time and get the feedbacks in time, improve the services and resources they need, and help the overall profitability of the finance industrial chain. With regard to applications, agriculture has subverted traditional perceptions, and many high technologies have been applied to agricultural production. It has utilized data in agriculture operations for numbers of years already. At present, agricultural big data has involved arable land, breeding, sowing, fertilization, plant protection, harvesting, storage and transportation, agricultural product processing, sales, animal husbandry production and other links, which can realize the management of crop cultivation, cultivation, maturity and sales. Under the circumstances of a decline in the market economy, fierce competition and difficult sales of agricultural products, agricultural big data is becoming more and more important. In the overall solution, satellites, drones and other means are used to collect crop data and information, and upload the data to the agricultural big data platform to penetrate all links of agricultural production and operation, so as to provide a basis for operators' management decisions.

As for Applications in e-retail, the e-retail industry is an industry that is developing as the big data is booming, very prone to take full great advantages of big data. Based on previous transaction data, e-retailers can easily explore new sales models and minimize operating costs. Though the Internet, the e-retail industry is thriving, and the CRM is easy to establish from the after-sale feedback. Moreover, one can find that the shopping application can display what one wants to buy by analyzing your skimming record and the retailers can adjust their commodity storage, which is related to the WM.

3.4 Descriptive Statistics

Descriptive statistics are summary statistics that quantitatively describe data using meas ures of central tendency and measures of variability or dispersion. The table below de picts the most commonly used descriptive statistics and visualization methods:

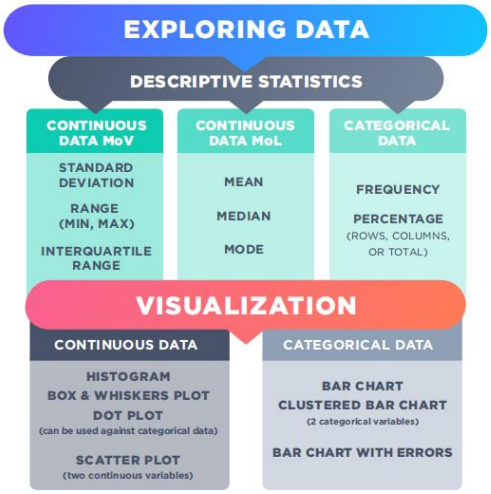

Figure 5 Descriptive Statistics

4 LIMITATION & FUTURE OUTLOOKS

4.1 BD complexity

In the SC, the generation and complex data collection and storage should be simplified by establishing multifunctional entity associations to form a larger database. On the one hand, these data are incomplete, inconsistent and inconsistent, which makes it difficult to obtain real-time data due to the lack of the necessary infrastructure. On the other hand, due to the large number of BDs established, this also raises another problem related to the time-consuming analytical process. As it known to all, data traffic is becoming more and more complex. In other words, the system must consider more factors and parameters than ever before. Since the raw data comes from multiple different sources, like a company's suppliers, processes, consumers, and other users, data structures are clearly more complex than ever. In addition, there are many technologies that can operate on multiple channels, such as not connecting to networks and mobile phones, thus taking complexity to a whole new level and adding to the challenges of big data.

4.2 Inexperienced professionals.

To run these modern technologies and big data tools, the company needs data professionals. As companies begin to shift to digital products that use big data, their employees may not be ready to use this advanced solution. As a result, implementing uninsured employees can result in a significant slowdown in work operations, often with workflow interruptions and a large

number of errors. Productivity can decline until employees understand all the benefits of innovation and learn to take advantage of it.

The difficulty in integrating data comes from a spread of source. As is demonstrated above, there are many sources, social media, Radio Frequency Identification, Electronic Data Interchange, Customer Relationship Management, and Warehouse Management etc. The links between the different sources is crossing. Integrating all the data can accelerate the development of Internet of things. However, combing all the date and finding the links are a hard job.

4.3 Future outlooks

Maximizing the use of massive amounts of data is a big challenge, and extending the application of people is only a small part of the data execution. One of the biggest challenges with big data projects is the successful implementation of the collected vision. Many applications and systems collect data, but organizations often find it difficult to understand and effectively apply the value of this knowledge. In general, the big data development can be better. Besides, the solution of the challenge will come out. The big data gives us the information about what will happen and why. More and more industries rely on the data. Then, one will obtain the links among all aspects and integrate them to achieve the whole Thing Internet and the big data is the global trend. The data volumes will continue to increase and migrate to the cloud. It is still growing dramatically. Apparently, the problems come out as the big data is thriving. However, it is essential for the development as the first Industrial Revolution started.

5 CONCLUSION

In summary, the state-of-art applications of big data are demonstrated and discussed in the field of supply chain management. Based on the analysis, three major parts were announced in a targeted manner: big data and big data analysis; the applications in the SCM; the limitations and future outlooks of the big data. With the boom of big data, many companies have a wisdom decision to collect the big data and use it to analyze the target market. Additionally, more and more companies begin to use a variety of information and communication technologies for supply chain management such as RFID, Enterprise Resource Planning (ERP), CMS to collect, manage and store their data for the further development. Big data analysis capabilities are efficient tools. By using descriptive, predictive, and prescriptive analytics, companies can have invaluable insights from these data among SCM to improve data-driven decision-making and develop creative ways that enable to increase business efficiency and captivity. However, there are several limitations about BD that we pay attention to BD complexity, privacy and security problems, the lack of skills and expertise. Although the big data is facing a lot of challenges, its future must be brilliant. These results offer a guideline for further implementation of big data technology into SCM.

REFERENCES

[1] Worldwide Big Data and Analytics Spending Guide, 2021V1, online information, avaliable at: https://cdn.idc.com/getdoc.jsp?containerId=IDC_P33195.

[2] M. Framinghan, “New IDC Worldwide Big Data Technology and Services Forecast Shows Market Expected 80 to Grow to $32. 4 Billion in 2017”. IDC Press Release, 18 December, 2013.

[3] C. Yuan, L. Hui, and T. Yuan. "Study on Integration & Network Sharing of Big Data in Supply Chain Management Field." Hans Journal of Data Mining vol. 5(4), 2015, pp. 75-80.

[4] S. Tiwari, H. W. Wee, Y. Daryanto, “Big data analytics in supply chain management between 2010 and 2016: Insights to industries.” Computers and Industrial Engineering, vol. 115, 2018, pp. 319-330

[5] M. Indralingam, M. Wahab, and L. Fang. “Benefits of Big Data in Supply Chain Management.”Proceedings of the Seventh International Forum on Decision Sciences. Springer, Singapore, 2020, pp. 151-156.

[6] K. Alicke, J. Rachor, and A. Seyfert. "Supply Chain 4.0–the next-generation digital supply chain." McKinsey & Company, 2016.

[7] A. McAfee, and E. Brynjolfsson, “Big data: The Management Revolution”. Harvard Business Review vol. 90, 2012, pp. 60-68.

[8] 13ème CONFERENCE INTERNATIONALE DE MODELISATION, OPTIMISATION ET SIMULATION (MOSIM2020), 12-14 Nov 2020, AGADIR, Maroc. 2020.

[9] L. Tamym, et al. "Big Data for Supply Chain Management in Industry 4.0 Context: A Comprehensive Survey." 13ème CONFERENCE INTERNATIONALE DE MODELISATION, OPTIMISATION ET SIMULATION (MOSIM2020), 12-14 Nov 2020, AGADIR, Maroc. 2020.

[10] K. Blasiak, “Big Data; A Management Revolution The emerging role of big data in businesses” Metropolia Ammattikorkeakoulu,2014.

[11] Q. Li, amd A. Liu, “Big data driven supply chain management.” Procedia CIRP vol. 81, 2019, pp. 1089-1094.

[12] M. K. Özlen, and N. Hadžiahmetović. "Customer relationship management and supply chain management." World Applied Programming vol. 3.3, 2013, pp. 126-132.