Big Data: A New Way for Today's Business Prospects and Challenges

Big Data in today's business

Ernest Johnson
Paul J. Hill School of Business
Faculty of Business Administration, University of Regina,
Regina, Saskatchewan, Canada
ernest.johnson@uregina.ca

S. Silas Sargunam
Department of Management Studies
Anna University
Tirunelveli, Tamil Nadu, India
silassargunam@gmail.com

Abstract— **The concept of big data is introduced along with its definitions and significance along with its dimensions namely volume, velocity, variety and veracity. Business applications of big data are explored in the various functional areas of management such as E-commerce, Inventory Management, Service, Customer Segmentation, Product Reputation Management, Recommender Systems, Pricing Strategy Formulation, Location-Based Advertising, and Machine Mediated Analysis. Subsequently, the big data challenges such as Data Complexity, Computational Complexity, System Complexity, Data Integration, Human Resources Requirement, and Privacy Issues, faced practically during the implementation have been elaborated. Also, user orientation should have priority over technical supremacy has been addressed. Thrust should be given to convenience and ease with which the business executives who have limited expertise in data management handle big data analysis for enhancing their functional domain decisions. However, it does not discuss the ethical aspect of collecting big data which is beyond the scope of this paper.**

Keywords- Big Data; e-Business; Business Prospects and Challenges;

I. INTRODUCTION

In recent years, the rapid development of Internet, Internet of Things (IoT), and iCloud Computing have led to the explosive growth of data in almost every industry and business area. According to the estimation by United Nations Initiative on Global Geospatial Information Management (UN-GGIM), 2.5 quintillion bytes of data are being generated every day. Also, in Google, about 25 PB of data is being generated per day. This trend will be even accelerated since the world becomes more and more mobile these days. "Big Data" is defined as "the amount of data just beyond technology's capability to store, manage and process efficiently" [1], [5]. Big data defined in Wikipedia as "an all-encompassing term for any collection of data sets so large and complex that it becomes difficult to process using traditional data processing applications". Big data can be regarded, from a macro perspective, as a bond that subtly connects and integrates the physical world, the human society, and cyberspace [5]. Big data can be characterized along four important dimensions, namely volume, velocity, variety and veracity [3], [12]. However, while accessibility to big data creates unprecedented opportunities for business, it also brings challenges to practitioners and researchers. Big data analytics is mainly concerned with three types of challenges namely data storage, management and processing.

II. BUSINESS PROSPECTS

There are many prospects while using big data. Some are caused by the characteristics of big data, some, by its current analysis models and statistical analytical methods. They are shown below

A. E-commerce

Leading e-commerce vendors such as Amazon and eBay have developed their innovative and highly scalable e-commerce platforms and product recommender systems. Major Internet firms such as Google, Amazon, and Facebook continue to lead the development of web analytics, iCloud Computing and social media platforms. The emergence of customer-generated Web 2.0 content on various forums, news groups, social media platforms, and crowd sourcing systems is helpful in collecting data from customers, employees and the media. The data that e-commerce systems collect from the web are less structured and often contain rich customer opinion and behavioural information [14], [18]. For social media, analytics of customer opinions, text analysis and sentiment analysis techniques are frequently adopted. Various analytical techniques such as association rule mining, database segmentation and clustering, anomaly detection, and graph mining have also developed for product recommender systems.

B. Inventory Management

The in-store movements of customers and their interactions with the products are recorded. Such data are combined with the transaction records and analysed to make decisions such as which products to carry, where to stock them

and how and when to offer a price discount. This has helped retailers to reduce the inventory and to increase the revenue. The benefit of the analysis of geospatial data was realised by Wal-Mart when Hurricane Frances was approaching Florida's Atlantic coast across the Caribbean island. By analyzing the transaction records stored in Wal-Mart's data warehouse, the company identified which items were bought when Hurricane Charley landed several weeks ago. This kind of predictive analytics can be used for reducing the cost of maintaining the inventory and shipping items between warehouses [3]. A retailer tracks the behaviour of individual customers from Internet click streams, updates their preferences and models their likely behaviour in real time. It is helpful in recognizing when customers are nearing a purchase decision and nudge the transaction to completion by bundling preferred products, offered with reward program savings. This real-time targeting increases the purchase of higher-margin products by its most valuable customers. The volume and quality of data available from internet purchases, social network communications and location-specific smart phone interactions have also led to data driven customization.

C. Services

Another sector that can benefit from analysis of human mobility is services. This field is supported by Tobler's first law of geography which states that everything is related to everything else, but near things are more related than distant things. This becomes relevant while suggesting the services or stores close to the current location of a user than suggesting those far-away. Location functionality has been combined with social networking, news, information, search and entertainment services. Location targeting is known to improve the performance of mobile advertising by over 200 percent. Many geo-social networking services such as Foursquare, when a user connects to the services, are suggesting recreational and other services close to his or her current location. Video rental service provider Netflix generated the data from the top rental units in each territory using the Zip code [5]. Popularity of movies rented in each region was analysed. Such patterns were used for suggesting movies for customers from a specific area. As the availability of pricing data proliferates on the Web price comparison services that automatically compile information across millions of products are provided. Such comparisons have created substantial value for consumers resulting in savings up to 10 percent [17]. Likewise, Insurance companies identify insurance policies for each customer, using fine-grained, constantly updated profiles of customer risk, changes in wealth, home asset value and other data inputs.

D. Customer Segmentation

For effective marketing, it is essential to identify a specific group of customers who share similar preferences and respond to a specific marketing signal. Customer related data is used to segment and target customers. Big data permits a major step beyond what until recently was considered state of the art, by making real-time personalization possible. For instance, to differentiate among customer groups for telecommunication applications, their call data is analyzed apart from their demographics [3]. In case of targeted marketing, profiling of each individual customer is performed so that the most suitable products or services are marketed to the most appropriate individual.

E. Product Reputation Management

Clustering and association mining techniques are being employed to monitor the reputation of a variety of products by data mining Web contents. An extraordinary size of images is being archived online, which is extremely challenging. Therefore, it is essential to develop a rich computer-based representation of product information for subsequent product reputation analysis [15]. The automatically constructed product ontologies can be used as the basis to support product reputation management applications and other marketing intelligence applications.

F. Recommender Systems

Effective promotional strategies are one of the key success factors for companies to increase their sales and revenue. The focus is on analyzing how different types of customers respond to different promotional strategies. Regression analysis methods are used to study promotions in different contexts. In the big data environment, more log data is obtained from Word of Mouth derived from both customer reviews and promotions [3]. Factors from other perspectives, such as price and place are also used in Promotional marketing analysis. Recommender systems have been widely used to improve product awareness and promote products to potential target customers in the e-commerce context. To develop such systems, User rating-based collaborative filtering methods or content-based association data mining methods are commonly used.

G. Pricing Strategy Formulation

Pricing research is an important element of marketing research. Normally, empirical research on pricing strategies employs survey data and regression methods. However, as price information is now available on websites, researchers have started using log data in e-commerce websites to study pricing strategy [11]. Based on the data derived from various log data sources, regression analysis methods are widely used for price prediction applications and association data mining methods are applied for competitor analysis.

H. Location-Based Advertising

Location is a valuable source for personalized marketing information. Location-based services (LBS) can provide personalized information about the users in a specific location at a specific time with the help of mobile technology. Location-based advertising ensures that the potential target customers get timely advertisements or product recommendations based on their current position.

I. Machine Mediated Analysis

Pricing research is an important element of marketing research. Normally, empirical research on pricing strategies

employs survey data and regression methods. However, as price information is now available on websites, researchers have started using log data in e-commerce websites to study pricing strategy [6], [13], [16]. Based on the data derived from various log data sources, regression methods are widely used for price prediction applications and association mining methods are applied for competitor analysis.

III. Business Challenges

There are many challenges faced while using big data. Some challenges are caused by the characteristics of big data, some, by its current analysis models and methods, and some, by the limitations of current data processing systems. The main difficulty in coping with big data lies in its huge volume. However tougher challenges are caused by the diversified data types (Variety), timely response requirements (Velocity) and uncertainties in the data (Veracity) [7], [8], [9]. [12]. The challenges are categorized as shown below

A. Data Complexity

The typical characteristics of big data are diversified types and patterns, complicated inter-relationships, and greatly varied data quality. The inherent complexity of big data makes its perception, representation, understanding and computation far more challenging [10], [17]. Traditional data analysis and data mining tasks, such as retrieval, topic discovery, semantic analysis, and sentiment analysis, become extremely difficult when using big data. Moreover, by modelling and analyzing the intrinsic mechanisms of data complexity, we will be able to develop the principles and mechanisms for processing big data.

B. Computational Complexity

The key features of big data, namely, multiple sources, huge volume, and fast-change, make it difficult for traditional computing methods to effectively support the processing, analysis and computation of big data. Such computations cannot be performed by the analytical tools and iterative algorithms used for handling small amounts of data. Their computability and computational complexity have to be assessed in the context of big data. Insufficient samples, open and uncertain data relationships and unbalanced distribution of value density in big data should be taken into account while developing new computing paradigms [17]. Appropriate computing framework needs to be developed where communication, storage, and computing are well integrated and optimized. Non-deterministic algorithmic theory suitable for big data should be developed. Reduction-based computing methods where big data is reduced on demand should be explored. Bootstrapping and sampling based local computation and approximation methods are to be developed.

C. Systems Complexity

Big data processing systems suitable for handling a diversity of data types and applications are necessary for supporting scientific research of big data. For data of huge volume, complex structure, and sparse value, its processing is challenged by high computational complexity, long duty cycle and real-time requirements [10]. These requirements not only pose new challenges to the design of system architectures, computing frameworks and processing systems, but also impose stringent constraints on their operational efficiency and energy consumption. The issues to be addressed in system complexity include design of system architectures, computing frameworks, processing modes and benchmarks for highly energy-efficient big data processing platforms. Solving these problems will lead to the development of hardware and software system architectures with efficient distributed storage and processing. Factors such as system throughput, parallel processing capabilities, computational accuracy, and energy consumption per unit should be accurately measured and monitored. Through an iterative process of design, implementation and validation, it is possible to develop big data processing systems with a high data acquisition throughput, low energy consumption and high efficiency.

D. Data Intergration

Another challenge facing the big data is the accumulation of data in departmental "silos," such as research and development (R&D), engineering, manufacturing, or service operations. This prevents timely sharing of information leading to information hoarding within business units. Many organisations suffer from their own failure to share data among the diverse lines of their own business [17]. This prevents these companies from forming a coherent view of individual customers or establishing understanding links among their departments.

E. Human Resource Requirement

The demand for people with the deep analytical skills in big data outstrips current supply by 50 to 60 percent. In 2018, as many as 140,000 to 190,000 additional specialists may be required. An additional 1.5 million managers and analysts with a sharp understanding big data applications are needed. The challenge of recruitment and retention of key data personnel loom large in the face of the companies [10]. There is also a steady demand for data aggregators, who combine and analyze information from multiple sources to generate insights for clients.

F. Privacy Issues

The greater access to personal information that big data often demands will lead to issues of privacy. As a larger amount of data on the buying preferences, health, and finances of individuals is collected, privacy concerns will grow. More open access to information, development of new devices for data gathering and cloud computing to support big data's weighty storage and analytical needs will raise data security issues [8]. The implication is that Information technology (IT) architectures will become more integrated and will pose greater risks to data security and intellectual property (IP).

IV. Suggestions

In organizations, an executive is responsible for making

decisions based on the result of the data analysis. It is estimated that an additional 1.5 million managers and data analysts with a sharp understanding of big data analysis and applications are required. The business executive is a domain expert but has limited database expertise as such efforts should be taken to make him or her understand and appreciate the data and the big data analysis. User oriented approach has to be followed in data analysis and interpretation of result. As the demand for people with the deep analytical skills in big data is huge, steps must be taken to enhance the effectiveness of recruitment and retention programs. Substantial investments must be made for the education and training of big data related skills. Therefore all companies must invest in big data infrastructure including data scientists and big data platforms. The success of business intelligence (BI) is critically dependent on data selection. Therefore, there must be perfect alignment between data and business intelligence (BI) goals.

Further, complicated business problems require combining data from different sources such as social media and transaction records. This, coupled with the computational complexities, requires the development of more efficient computational methods. In order to meet the challenges for more advanced business intelligence in the next generation of big data management, new computational methods must be developed to cope with the volume, velocity, variety and veracity issues of big data.

V. CONCLUSION

Big data has made a strong impact in the way business is carried out. Almost every functional area has immensely benefited from the use of big data. This paper has explored the benefits of applying big data analysis in the context of business. The major challenges faced while employing big data analysis has also been probed. Suggestions are also given to improve the effectiveness of applying big data analysis to business decision making. Big data goals should be aligned with the domain goals of business. Emphasis should be on user orientation rather than on technical supremacy. Also as you might have noticed, the ethical aspect of collecting big data is not discussed and it is beyond scope of this paper.

REFERENCES

[1] J. Manyika, M. Chui, B.Brown, J.Bughin, R.Dobbs, C. Roxburgh., A. Hung, Big Data: The next frontier for innovation, competition, and productivity, Tech. rep., McKinsey Global Institute, available at http://ww.mckinsey.com/insights/business technology/big data the next forntier for innovations (2011).

[2] JC/O'Neil, R. Schutt, Doing Data Science: Straight Talk from the Front-line. O'Reilly Medica, Inc., 2013

[3] Big Data, http://en.wikipedia.org/wiki/Bigdata (2014).

[4] G.Li, X. Cheng, Research status and scientific thinking of big data, Bulletin of the Chinese Academy of Sciences 27 (6) 647-657, (2012).

[5] Y.Wang, X.Jin, Xueqi, Network Big Data: Present and future, Chinese Journal of Computers 36 (6), 1125 -1138, (2013).

[6] X/Q.Cheng, X.Jin, Y.Wang, J.Guo, T.Zhang, G.Li, Survey on big data system and analytic technology, Journal of Software 25 (9), 1889-1908, (2014).

[7] Challenges and opportunities with Big Data, a community white paper available at http://cra,org/ccc.does.init/bigdatawhitepaper.pdf

[8] H.V.Jagadish, Johannes Gehrke, Alexandros Labrinidis, Yannis Papakonstantinou, Jignesh M. Patel, Raghu Rama krishnan, Cyrus Shahabi, Big data and its technical challenges, Commun, ACM 57 (7), 86-94, (July, 2014). http://dx.doi.org/10.1145/2611567.

[9] S.Kaisler, F.Armour, J.A.Espinosa, W.Money, Big data:Issues and Challenges moving forward, in 6th Hawaii International confrence on System Sciences, (HICSS), IEEE, 2013, 995-1004, 2013.

[10] P. Zikopoulos, C.Eaton, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, McGraw Hill/Osborne Media, 2011.

[11] Pattern-Based Strategy:getting value from Big Data, Gartner Group press release, available at http://www.gartner.com/it/page.jsp?id+1731916, July, 2011.

[12] The 4 V's of Big Data, http://www.ibmbigdatahub.com/tag/587.

[13] A.Dasgupta, Big data: The future is in analytics, http://www.geospatialworld.net/Magazine/MArticleView.aspx?aid=30512, geospatial world April, 2013.

[14] M. Meeker, 2012 KPCB internet trends year-end update, http://www.slideshare.net/kleinerperkins/2012-kpcb-internet - trends-yearend-update, Dec. 2012.

[15] H.Hedin, I.Hirvensalo, M.Vaarnas, The Handbook of Market Intelligence: Understand, Compete and Grow in Global Markets, John Wiley and Sons, 2014.

[16] W.Tan, M.B.Blake, I.Saleh, S.Dustdar, Social Network Sourced Big Data Analytics, IEEE Internet Computer 17(5) (2013) 62-69.

[17] S.Kaisler, F.Armour, J.A.Espinosa, W.Money, Big Data: Issues And Challenges Moving Forward, 6th Hawai International Conference on System Sciences, IEEE, 2013, 995- 1004.

[18] Hsinchun Chen, Business Intelligence and Analytic: From Big Data to Big Impact, MIS Quarterly, Vol.36, No 4, 1165-1188, 2012.