International Journal of Information Management 42 (2018) 78–89

Contents lists available at ScienceDirect

International Journal of Information Management

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

Big data with cognitive computing: A review for the future

Shivam Gupta[a], Arpan Kumar Kar[b,*], Abdullah Baabdullah[c], Wassan A.A. Al-Khowaiter[d]

[a] Montpellier Research in Management, Montpellier Business School, 2300 Avenue des Moulins, 34185, Montpellier, France
[b] Department of Management Studies, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India
[c] Department of Management Information Systems, Faculty of Economics and Administration, King Abdulaziz University, Jeddah, Saudi Arabia
[d] Department of MIS and PM, Qassim University, Saudi Arabia

ARTICLE INFO

Keywords:
Big data
Cognitive computing
Literature review
Resource based View
Institutional theory

ABSTRACT

Analysis of data by humans can be a time-consuming activity and thus use of sophisticated cognitive systems can be utilized to crunch this enormous amount of data. Cognitive computing can be utilized to reduce the shortcomings of the concerns faced during big data analytics. The aim of the study is to provide readers a complete understanding of past, present and future directions in the domain big data and cognitive computing. A systematic literature review has been adopted for this study by using the Scopus, DBLP and Web of Science databases. The work done in the field of big data and cognitive computing is currently at the nascent stage and this is evident from the publication record. The characteristics of cognitive computing, namely observation, interpretation, evaluation and decision were mapped to the five V's of big data namely volume, variety, veracity, velocity and value. Perspectives which touch all these parameters are yet to be widely explored in existing literature.

1. Introduction

Advancement in technology has always been a natural phenomenon. Drastic technological advancements often tend to miss out to find itself a mass market where it can be sold. The population of the world stands at more than 7.4 billion people (Worldometers, 2016) and more than 3.1 billion of these people are connected to the internet (eMarketer Report, 2015). More than 4.4 billion people are using mobile phone, out of this 1.86 billion (42%) of people are using smart phone (eMarketer Report, 2015). The number of devices and connections will keep increasing year on year basis and thus this had led to an explosion of data (Yaqoob et al., 2016). This enormous amount of data that is produced on a continuous basis is termed as big data (Kreps & Kimppa, 2015). Fosso-Wamba et al. (2015) have discussed the concept of big data in detail and have brought out the various definitions by other researchers in this domain. Big data analytics has gained significant amount of importance as it enables organizations to be ahead of their competitors (Li, Tao, Cheng, & Zhao, 2015; Habib ur Rehman, Chang, Batool, & Wah, 2016; Bumblauskas, Nold, Bumblauskas, & Igou, 2017). Companies want to analyze this raw data as they want to spot the trends that they can use for further profit maximization (Frizzo-Barker, Chow-White, Mozafari, & Ha, 2016; Sun, Strang, & Firmin, 2016). The dire need to address the concerns of data deluge has led to the emergence of big data analytics (Dubey et al., 2015; Liberatore, Johnson, & Clain, 2016).

Analysis of data by humans can be a time-consuming activity and thus use of sophisticated cognitive systems can be utilized to crunch this enormous amount of data (Kim, Chan, & Gupta, 2016). Intel CEO Brian Krzanich explains in his editorial article on artificial intelligence (AI) in 2016 that *"AI is based on the ability of machines to sense, reason, act and adapt based on learned experience"* (Krzanich, 2016). In a system that is based on AI, it works on the rules and parameters that are fed inside it whereas a cognitive computing based system works by intercepting the command and then drawing inferences and suggesting possible solutions. Cognitive computing is an AI based system that enables it to interact with humans like a fellow human, interpret the contextual meaning, analyze the past record of the user and draw deductions based on that interactive session. Cognitive computing helps the humans in decision making whereas AI based systems works on the concept that machines are capable of making better decisions on the human's behalf. Cognitive computing is a sub-set of AI and anything that is cognitive is also AI.

The goal of cognitive computing is to build a rational, combined and collective mechanism motivated by the capability of human mind (Kwon, Lee, & Shin, 2014). Cognitive computing will develop unique learning systems, non-von Neumann computing structural design,

* Corresponding author.
E-mail addresses: connectwithshivamgupta@gmail.com (S. Gupta), arpan_kar@yahoo.co.in (A.K. Kar), baabdullah@kau.edu.sa (A. Baabdullah), w.alkhowaiter@gmail.com (W.A.A. Al-Khowaiter).

https://doi.org/10.1016/j.ijinfomgt.2018.06.005
Received 22 April 2018; Received in revised form 27 June 2018; Accepted 27 June 2018

programming archetype and functions that can combine, examine and work on large volumes of data from variety of sources at once (Modha et al., 2011).

Raghupathi and Raghupathi (2014) have showcased the various sources that lead to generation of big data and it encompasses nearly every spectrum of academic and scientific domain. Cognitive computing can be utilized to reduce the shortcomings of the concerns faced during big data analytics (Hurwitz, Kaufman, & Bowles, 2015). In cognitive computing, a computerized model captures the human thought process and improvises on the mistakes the system commits every time (Modha et al., 2011). This self-learning mechanism can greatly benefit the way large amount of data is analyzed for better decision making (Gudivada, Raghavan, Govindaraju, & Rao, 2016). As there is an emergence in the implementation of cognitive computing to provide insights by analyzing large data sets, it is imperative to understand this subject matter.

The objectives of this literature review are:

- To propose a link between Cognitive Computing and Big Data.
- To perform a literature review and showcase the academic research work that has been published in the domain of Big Data and Cognitive Computing.
- To propose a conceptual model based on theoretical understanding to link the characteristics of Cognitive Computing by using the benefits offered by Big Data.

Section 2 showcases the background and motivation for this study. Section 3 brings out the methodology employed to undertake the literature review for big data and cognitive computing. Section 4 focusses on the analysis and discussion and Section 5 concludes this study with limitations and scope for future research. Appendix A in Table A1 shows the tabular representation of the journals and the years in which papers have been published. Appendix B presents a list of the selected papers that have been considered for the review.

2. Background and motivation

The history of computing is marked with events that very few had envisaged. The era of computing can be broadly classified into three categories:

2.1. The Tabulating Era (1900 s - 1940 s)

The computing devices consisted of calculators which were in the form of single-purpose mechanical systems where counting used to take place by using punched cards. The use of such systems was limited to large scale companies and academic and scientific institutions (IBM White Paper, 2015).

2.2. The Programming Era (1950 s - present)

During the World War II, the need for computing was driven by military and scientific requirements. Invention of transistor and the microprocessor lead to the creation of computers which were bulky machines compared to the modern day computer. The capacity and speed of the computer kept increasing at a drastic pace for the six decades and it has led to the creation of modern day compact computers to smart phones (IBM White Paper, 2015).

2.3. The Cognitive Era (2011 -)

The amount of data will become unsustainable for humans to process and thus there is a need for the computer systems to assist the humans in decision making (Changchit & Chuchuen, 2018). Having access to the vast amount of data will be a daunting task for any individual or an organization but a cognitive system can make sense of the raw data and turn into actionable information and thereafter valuable knowledge (IBM White Paper, 2015).

The origins of big data can be trailed back to the mid 1990′s when John Mashey, working at Silicon Graphics, USA was involved in processing and analysis of large datasets (Diebold, 2012). During a conference in 2010, Eric Schmidt, then Google CEO, had commented that the amount of data created in two days of time is equal to the amount of data created from the dawn of civilization till 2003 (Schmidt, 2010). This amount of data that he was referring was five Exabyte of data in 2010. Since then the amount of data created has only gone up exponentially. The ever-increasing challenges of data deluge have led to the rise of big data analytics (Raguseo, 2018). With more and more IoT (Internet of Things) connected devices, the problem of data deluge will further aggravate (Cintra White Paper, 2016; Hashem et al., 2016). Cognitive computing is considered to be a convergence of cognitive science, neuroscience, data science, and cloud computing (Gudivada et al., 2016). IBM has been pioneering in making commercial use of cognitive computing by launching Watson in 2011. There has been widespread of adoption of cognitive systems and the worldwide revenue for 2016 is touted at USD 8 billion and furthermore it is expected to rise to USD 47 billion by 2020 (IDC, 2016). The compound annual growth rate (CAGR) in this domain will be more than 55% over the five-year period from 2016 to 2020 (IDC, 2016). North America (the United States and Canada) constitute the biggest share (78%) in this USD 8 billion expected revenue and Europe, the Middle East and Africa (EMEA) comprise of the second largest region in the 2016 forecast of IDC. By 2020, the demand for cognitive systems from Asia-Pacific region including Japan will be similar to the EMEA region (IDC, 2016).

The intersection of mixed research studies using cognitive computing and big data analytics suffers from multiple challenges (Aswani, Kar, & Ilavarasan, 2017; Aswani, Ghrera, Kar, & Chandra, 2017). Cognitive computing approaches typically work on vectors having objective data where machine learning approaches learn through mathematical operations on these data. However, big data analytics can analyze and provide visualization of information in unstructured data (Ragini, Anand, & Bhaskar, 2018). The integration of these approaches first requires the conversion of unstructured data into some objective or quantitative parameters which are derived as vectors. Then these vectors over numerous instances may be used to train an algorithm in cognitive computing, so that predictions would be possible on the basis of the intent hidden in the content on test cases.

In order to extract the benefits of cognitive systems, it has to be fed with massive amount of data so that the systems are capable to find hidden patterns and relationships between the various variables which may be present in structured, semi-structured and unstructured data of high variety and veracity arriving in high volume and velocity. Thus, the field of big data analytics and cognitive computing will go hand-in-hand. The next section will highlight the methodology employed to undertake a literature review study for big data and cognitive computing as a combined subject matter.

3. Literature Review: Big Data and Cognitive Computing

The literature review for this study has been based on the guidelines that have been outlined by Tranfield, Denyer, and Smart, (2003) and Kitchenham and Charters (2007). There are three broad steps and they are namely: (i.) Planning the review, (ii.) Conducting the review and (iii.) Reporting the review. They are explained below:

3.1. Planning the Review

A review protocol has been developed that brings out the criterion for the selection of studies. The keywords that were used in this study were: **“big data”** and **“cognitive computing”**. An online digital database has been used to gather the information for the papers that are required for the review process. The digital databases that were