Advanced Engineering Informatics 30 (2016) 500–521

Contents lists available at ScienceDirect

Advanced Engineering Informatics

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

Review article

Big Data in the construction industry: A review of present status, opportunities, and future trends

Muhammad Bilal [a], Lukumon O. Oyedele [a,*], Junaid Qadir [b], Kamran Munir [a], Saheed O. Ajayi [a], Olugbenga O. Akinade [a], Hakeem A. Owolabi [a], Hafiz A. Alaka [a], Maruf Pasha [c]

[a] Bristol Enterprise, Research and Innovation Centre (BERIC), Bristol Business School, University of West of the England, Bristol, United Kingdom
[b] School of Electrical Engineering & Computer Science (SEECS), National University of Sciences & Technology (NUST), Islamabad, Pakistan
[c] Department of Information Technology, Bahauddin Zakariya University, Multan, Pakistan

ARTICLE INFO

Article history:
Received 2 February 2016
Received in revised form 3 July 2016
Accepted 6 July 2016

Keywords:
Big Data Engineering
Big Data Analytics
Construction industry
Machine learning

ABSTRACT

The ability to process large amounts of data and to extract useful insights from data has revolutionised society. This phenomenon—dubbed as Big Data—has applications for a wide assortment of industries, including the construction industry. The construction industry already deals with large volumes of heterogeneous data; which is expected to increase exponentially as technologies such as sensor networks and the Internet of Things are commoditised. In this paper, we present a detailed survey of the literature, investigating the application of Big Data techniques in the construction industry. We reviewed related works published in the databases of American Association of Civil Engineers (ASCE), Institute of Electrical and Electronics Engineers (IEEE), Association of Computing Machinery (ACM), and Elsevier Science Direct Digital Library. While the application of data analytics in the construction industry is not new, the adoption of Big Data technologies in this industry remains at a nascent stage and lags the broad uptake of these technologies in other fields. To the best of our knowledge, there is currently no comprehensive survey of Big Data techniques in the context of the construction industry. This paper fills the void and presents a wide-ranging interdisciplinary review of literature of fields such as statistics, data mining and warehousing, machine learning, and Big Data Analytics in the context of the construction industry. We discuss the current state of adoption of Big Data in the construction industry and discuss the future potential of such technologies across the multiple domain-specific sub-areas of the construction industry. We also propose open issues and directions for future work along with potential pitfalls associated with Big Data adoption in the industry.

Contents

* Corresponding author at: Bristol Enterprise, Research and Innovation Centre (BERIC), Bristol Business School, University of West of England, Bristol, Frenchay Campus, Coldharbour Lane, Bristol BS16 1QY, United Kingdom.
E-mail addresses: Ayolook2001@yahoo.co.uk, L.Oyedele@uwe.ac.uk (L.O. Oyedele).

http://dx.doi.org/10.1016/j.aei.2016.07.001
1474-0346/

1. Introduction

The world is currently inundated with data, with fast advancing technology leading to its steady increase. Today, companies deal with petabytes (10^{15} bytes) of data. Google processes above 24 petabytes of data per day [1], while Facebook gets more than 10 million photos *per hour* [1]. The glut of data increased in 2012 is approximately 2.5 quintillion (10^{18}) bytes per day [2]. This data growth brings significant opportunities to scientists for identifying useful insights and knowledge. Arguably, the accessibility of data can improve the status quo in various fields by strengthening existing statistical and algorithmic methods [3], or by even making them redundant [4].

The construction industry is not an exception to the pervasive digital revolution. The industry is dealing with significant data arising from diverse disciplines throughout the life cycle of a facility. Building Information Modelling (BIM) is envisioned to capture multi-dimensional CAD information systematically for supporting multidisciplinary collaboration among the stakeholders [5]. BIM data is typically 3D geometric encoded, compute intensive (graphics and Boolean computing), compressed, in diverse proprietary formats, and intertwined [6]. Accordingly, this diverse data is collated in federated BIM models, which are enriched gradually and persisted beyond the *end-of-life* of facilities. BIM files can quickly get voluminous, with the design data of a 3-storey building model easily reaching 50 GB in size [7]. Noticeably, this data in any form and shape has intrinsic value to the performance of the industry. With the advent of embedded devices and sensors, facilities have even started to generate massive data during the operations and maintenance stage, eventually leading to more rich sources of Big BIM Data. This vast accumulation of BIM data has pushed the construction industry to enter the Big Data era.

Big Data has three defining attributes (a.k.a. 3V's), namely (i) volume (terabytes, petabytes of data and beyond); (ii) variety (heterogeneous formats like text, sensors, audio, video, graphs and more); and (iii) velocity (continuous streams of the data). The 3V's of Big Data are clearly evident in construction data. Construction data is typically large, heterogeneous, and dynamic [8]. Construction data is voluminous due to large volumes of design data, schedules, Enterprise Resource Planning (ERP) systems, financial data, etc. The diversity of construction data can be observed by noting the various formats supported in construction applications including DWG (short for drawing), DXF (drawing exchange format), DGN (short for design), RVT (short for Revit), ifcXML (Industry Foundation Classes XML), ifcOWL (Industry Foundation Classes OWL), DOC/XLS/PPT (Microsoft format), RM/MPG (video format), and JPEG (image format). The dynamic nature of construction data follows from the streaming nature of data sources such as Sensors, RFIDs, and BMS (Building Management System). Utilising this data to optimise construction operations is the next frontier of innovation in the industry.

To understand the subtleties of Big Data, we need to disambiguate between two of its complementary aspects: Big Data Engineering (BDE) and Big Data Analytics (BDA). The domain of BDE is primarily concerned with supporting the relevant data storage and processing activities, needed for analytics [9]. BDE encompasses technology stacks such as Hadoop and Berkeley Data Analytics Stack (BDAS). Big Data Analytics (BDA), the second integral aspect, relates to the tasks responsible for extracting the knowledge to drive decision-making [9]. BDA is mostly concerned with the