Big Data Analytics in Civil Engineering: The Case of China

YouseokKang[1], JiayanYu[2], JiaruiChang[3]
[1]Hanyang University (Korea), [2]Johns Hopkins University (USA), [3]Dalian No. 24 (China)

Abstract

China has the world's largest construction market and exports a wide range of construction products to the world. With the increase of investment each year, the construction industry has played a crucial role in China's economic development, contributing 6.7% to China's GDP. In 2016, the construction industry achieved 17% growth in industry value, an increase of 5% over 2015. It is evident that China's construction industry has given rise to China's vast sea of high-rise structures, particularly with ICT advances. This paper presents how big data analytics can be applied to civil engineering and for various predictions in the construction sector for further improvement and for budget estimates and tender participation. Further, the study shows how improvements to the construction industry can be achieved through the estimation of the life and health of structures.

Keywords: *Civil Engineering, Construction Industry, Big Data, Data Mining, Construction Management, Civil Engineering Use Cases*

I. INTRODUCTION

China is the world's largest marketing country producing wide range of products especially electronic goods. But, in the recent years, with the increase in urbanization and national commitment to conserve the resources, the construction market in China is gaining investments and showing steady commitment for the expansion of construction market. (2016 ITA Building Products and Sustainable Construction Top Markets Report, 2016) Though, the economy growth of China is coming down gradually, its construction market maintains steady growth and it is expected to become world's largest construction market in the future. A recent survey on the Chinese construction industry has shown that, the construction industry stakeholders have increased from 5 percent to 28 percent. Also, the construction of huge skyscrapers and buildings were completed in the recent years at a record duration.

The Chinese construction industry is moving towards the green building market i.e., reducing the pollution and generation of wastes. With the rapid changes in technology and communication, the construction industry is also taking new shape in the design of new policies and focusing on the energy efficiency, building health monitoring, increase on building standards, utilization of by-products in the construction materials etc. The other most important cause for the growth in construction industry is the implementation of government decisions for the construction (conversion) of smart cities, construction of high quality roads, expansion of railway network etc.

In the recent years, China's construction sector continued to grow at a rapid pace and achieved 17% growth in 2016 which is 5% more than the growth in 2015. The primary causes for the growth of construction sector are the government considering increasing the infrastructure growth to sustain economic growth and the relaxation of government rules and regulations on the construction sector, purchase criteria, credit availability and encouragement for the boosting of real estate market.

This paper presents various developments in civil engineering constructions with the changes in Information technology, importance of storing data, mining for the prediction of various parameters in the future constructions, how the quality of civil engineering projects can be assessed, how others use cases where big data analytics may be applied to civil engineering.

The rest of the paper is organized as follows: section II Construction Sector in China, Section III discusses Literature Review. Section IV Big Data Analytics for Civil Engineering, Section V – Civil Engineering Use cases, Section VI concludes the work.

II. CONSTRUCTION SECTOR IN CHINA

As the investments in the construction sector are growing year by year, the construction industry has become the crucial player in China's economic development. This industry includes the construction of smart cities, lay of high ways, installation of pipe line system, building constructions, interior and exterior decorations of huge structures. (Zhang S.B, 2003)

Despite the increase in the real-estate business, construction sector contributes large percentage of China's Gross Domestic Product (GDP) (The Construction Sector in China, 2013). The construction sector consists of different categories of products and services. In general, this sector can be defined into three areas viz., Construction, Services

and Materials. Usually, the Construction activity includes the project management and building the structure by contractor companies. The major responsibility here lies in the completion of the project either by single contractor or through sub-contractors. The next activity services include the works related to the design of drawings, preparing calculations, management of resources, supervision task and advice. The last part of the construction is the Materials, the major task here is to provide the actual building materials, supply of machinery, man power and other equipment required for the construction activity.

In China, the construction industry includes Building engineering, Planning, Transportation, Production engineering, Civil engineering, Construction decoration and the people working in this sector focus on the performance parameters like increasing labor efficiency, energy efficiency, improvement of quality and strength of the construction. The construction enterprises handling construction tasks in China are classified in to small-size, middle-sized and large-sized.

III. LITERATURE REVIEW

Hore and Alan (Hore, 2006) in their article, Use of IT in construction projects, presented that, the use of Information Communication Technology (ICT) in the field of construction was not so popular like other industries due to the complex components of construction process, but now many construction sectors are increasingly recognizing the importance of ICT because ICT is improving the capability and efficiency of specific aspects within the construction process and it can also improve communication. Further, they discussed that, different kind of construction applications are available which help managing and integrating the data, visualizing the design, estimating the cost, and planning the maintenance.

Leu et.al, (Leu, 2011) have proposed how to improve public construction (bridge) project quality and maintain a quality system. The proposed method deals with the construction quality management system i.e., Three-level Quality Management System (TQMS) whose final and critical process would be to audit all construction processes and results. Here, one of the clustering algorithms called KJ method is implemented to classify the defects. This method adopts the bottom-up sorting process and is very useful for classifying data. CKP sorting method is better because it can simultaneously handle user constraints and mixed data types.

Cheung et.al, (Sai On Cheung, 2004) proposed a Web-based construction Project Performance Monitoring System (PPMS) to help construction project managers for the project management. PPMS was developed with four executive sections: Data Entry, Database, Reporting, and Action. PPE and KPIs are valuable tools for measuring project performance over a period of time. One important function PPMS is that it can present graphical data in both Overall Quality Measures and Contracts Comparison. Based on the case study, it was concluded that PPMS is designed purely for the monitoring of several key aspects of project performance.

Bilal et.al. (Bilal M., 2016) discussed about the construction waste analytics. Construction waste have always been a problem and the usage of big data can be a great solution. Since the ultimate goal is to minimize the waste and design the construction from the beginning, therefore waste analytics is more suitable than waste intelligence. They proposed a big data platform for waste analytics architecture based on the Hadoop framework. After careful analytics and potential applications, they presented results and discussion about waste distribution by different categories. Also the paper is a part of an on-going research and development effort on developing a robust BIM-based construction waste simulation tool.

Jae-Gil Lee et.al (Jae-Gil Lee, 2015) presented that, Geospatial big data is increasing exponentially and receiving more and more attention now. Also the data benefits humans in many aspects, such as the eco-route proposed by Ford which applies geospatial big data. First, the power of location suggests that 'Everything is related to everything else, but near things are more related than distant things', which is validated by geospatial big data; therefore, business can make sales policies based on this rule. Second there are several forms of geospatial big data. Traditionally, geospatial data can be categorized into three forms: raster data, vector data, and graph data but with the advancements of sensor and communication technologies, new sources of geospatial big data are emerging. Processing these data is the most important issue right now.

Kunz et. al. (Kunz, 2014) presented that, the Information technology is now commonly used in construction projects and it helps coordinating and integrating information across disciplines and throughout several project phases. The paper focuses on the role and scope of IT in support of multidisciplinary planning and coordination of construction projects. From the renovation example, it is concluded that IT should help to make predictions of the anticipated performance of the design of a project's scope, schedule, and organization with respect to the business objectives. And most projects involve POP project models, which can be greatly supported by VDC (Virtual Design and Construction).

IV. BIG DATA FOR CIVIL ENGINEERING CONSTRUCTIONS

Big Data is a collection of large and complex data sets which are difficult to process using common database management tools or traditional data processing applications. According to zdnet.com, "Big data refers to the tools, processes and procedures which allow an organization to create and manage very large data sets and storage facilities".

Big data (V.Zicari) is getting more and more accessible and important in modern information analyzing and the paper looks it from three different perspectives: the business perspective, the technological perspective, and the social good perspective. “Big data” refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze. And there are several characteristics of big data: velocity, volume, variety, and so on, from which we can conclude that big data is a kind of unprecedented information which needs a brand new analytic method, which leads to the invention of big data platform technology, mainly known as Hadoop. Hadoop is widely used in business and large Internet companies like Amazon are advanced users of Hadoop. What’s more big data can also contribute to common good. Big data analytics can process economic or environmental data and therefore experts can make expectations and precautions.

A. *Characteristics of Big Data*

Big data generally refers to the social network data from the micro-blogging sites like Twitter, LinkedIn and social media platforms like Facebook, Traditional enterprise including transactional data, web store transactions etc. and machine generated / sensor data like call data records, smart meters, manufacturing sensors, trading systems, traffic data, air data etc. which keeps on increasing without the human intervention. Big data is not only driven by the exponential growth of data but also by changing user behaviour and globalization. Globalization provides competition among the participants in the market. As a result, organizations are constantly looking for opportunities to increase their competitive advantage by using better analytical models.

The typical characteristics of big data are:

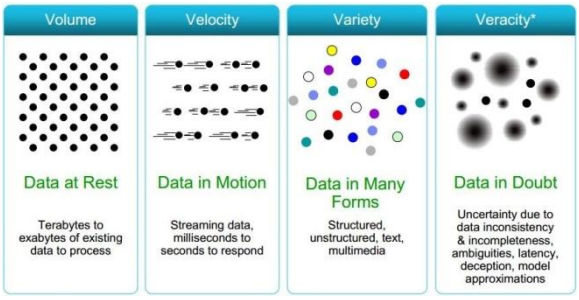

Fig. Characteristics of Big Data

1) *Volume:*

Data volume has been increasing exponentially: up to 2.5 Exabytes of data is already generated and stored every day. This is expected to double by 2015. Big data generated vast amounts of data being by organizations or individuals from Terabytes to Exabytes and Zettabytes of data.

2) *Velocity:*

Big data grows rapidly, which generated unprecedented quantities need to be stored, transmitted, and processed quickly. Velocity is the speed at which the data is being generated like streamed data from various smart devices into social media and also camera streamed data which stores the data in motion from huge number of closed circuit cameras.

3) *Variety*

In Big data, the variety and heterogeneity of data sources and storage has increased, fuelledby the use of cloud, web and online computing. Variety makes big data really big. Big data comes from a great variety of sources and generally has three types: structured, semi-structured and unstructured. Structured data inserts a data warehouse already tagged and easily sorted but unstructured data is random and difficult to analyse. Semi structured data does not conform to fixed fields but contains tags to separate data elements.

4) *Veracity*

Big data Veracity refers to the biases, noise and abnormality in data. And the problem also can be if the data that is being stored, and mined is meaningful to the problem being analyzed. Veracity in data analysis is the biggest challenge when compared to other characteristics like volume and velocity.

5) *HADOOP*

Hadoop is an open source framework which employs a simple programming standard that allows distributed processing of massive data sets on clusters of computers. The entire technology incorporates shared utilities, a distributed file system (DFS), analytics and information storage platforms, plus an application layer which manages the activities like

workflow, distributed processing, parallel computation and configuration management.

6) ***HDFS:***

The basic idea of Hadoop is to make use of the Distributed file system for storing and processing the data. This HDFS splits the file into blocks and these blocks are allocated in the Hadoop cluster nodes. The input data in HDFS is given once and it is processed by MapReduce and the outcomes are sent to HDFS. The HDFS data is safeguarded by duplication mechanism among the nodes which gives reliability and availability regardless of node failures.

In Hadoop, there are two types of HDFS nodes:
(1) Data Node (2) Name Node
Data Node stores the data blocks of the files
Name Node contains the metadata, with record blocks and a list of DataNodes in the cluster.

a) How Does HDFS Work?

An HDFS cluster is comprised of a NameNode which manages the cluster metadata and DataNodes that store the data. Files and directories are represented on the NameNode by inodes. Inodes record attributes like permissions, modification and access times, or namespace and disk space quotas.

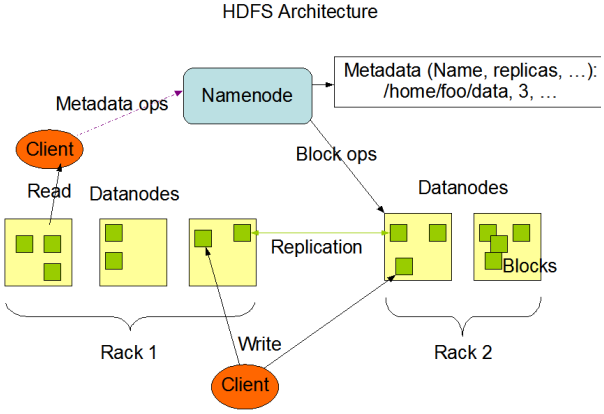

Fig. HDFS Architecture

The file content is split into large blocks (typically 128 megabytes), and each block of the file is independently replicated at multiple DataNodes. The blocks are stored on the local file system on the datanodes. The NameNode actively monitors the number of replicas of a block. When a replica of a block is lost due to a DataNode failure or disk failure, the NameNode creates another replica of the block. The NameNode maintains the namespace tree and the mapping of blocks to DataNodes, holding the entire namespace image in RAM.

The NameNode does not directly send requests to DataNodes. It sends instructions to the DataNodes by replying to heartbeats sent by those DataNodes. The instructions include commands to: replicate blocks to other nodes, remove local block replicas, re-register and send an immediate block report, or shut down the node.

b) MapReduce:

MapReduce is the programming paradigm that allows for massive scalability across hundreds or thousands of servers in the Hadoop cluster. MapReduce is the heart of Hadoop where the processing is carried out by assigning the tasks to various clusters.

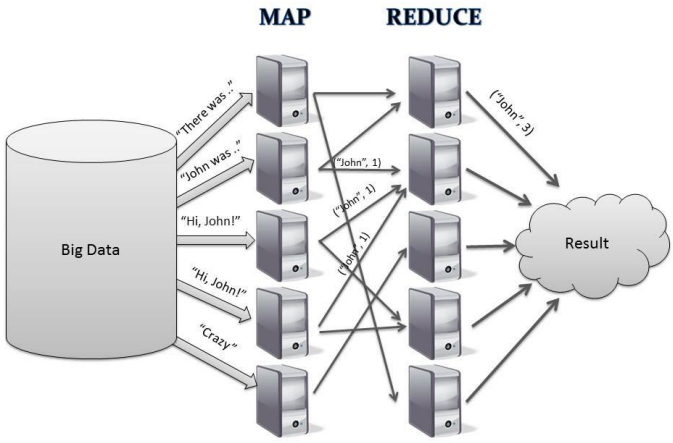

Fig. MapReduce

V. CIVIL ENGINEERING USE CASES

The word Big data has become the most common buzzword in every business sector, especially in the diversified business world, right from healthcare, insurance, finance, marketing and civil engineering. Among the engineering fields, Big data has found several applications in the civil engineering constructions. As presented in the literature review, the big data is being used in various operations of the constructions like transport management, inventory management, contract management, crew management etc.

The civil engineering operations and construction project management eco-systems are undergoing noticeable changes as a result of the huge amounts of data processed by big data tools and other platforms for the storage, testing and monitoring of the work flow systems.

Big data can be applied to any field of engineering where ever the data storage and the processing of data are involved. Big data will use the Hadoop Framework for both the storage and processing of data. Hadoop eco-system includes various tools for the storage, analytics, visualize, and execution, among others. The major components of the Hadoop framework are HDFS (Hadoop Distributed File System) and Map Reduce (MR). HDFS is the file system where the data will be stored and retrieved when ever required. Map Reduce is the processing engine where it will use the Map and

Reduce operations for the quick processing of big data.

The following are some of the use cases in the construction sector where big data can be applied for the better analysis of data and decision making.

1. Design of Construction Management System
2. Estimation and Management of Crew for the proposed project
3. Better analysis in monitoring the construction health
4. Big Data Analytics in Environmental Engineering
5. Design of Highway and Transportation Engineering for the prediction of Transport arrivals.
6. Application of Big Data on Geo Technical Engineering
7. Prediction of the Severity of Earthquakes
8. Big Data for the Coastal and Harbour Engineering
9. Use of Big Data analytics in Surveying and Geo-spatial engineering.
10. Application of Big Data analytics for the Geographical Information System (GIS)

A. Use Case 1:
Title:Design of Construction Management System

The Construction Industry generates huge amounts of information. But, due to lack of proper storage and analytics in the construction information, the Civil engineersering people are not using the past data. Still, most of the civil engineering peopleare using computers only designing the structures and for graphical illustrations of the designs.

Civil Engineers are not aware of the trends in the storage and processing of data. Here, we addressed how do we store the huge amounts of data and how those data can be processed with the commodity hardware. Design of Civil Construction is not different from the construction of any software. As the software has different phases / levels, the construction industry also requires initial information gathering, followed by planning and then designing before actual implementation and then goes to the next stage of implementation.

In the design of Construction Management System, the various people involved in the process were given access as per their access levels. The Management here involves managing the resources like raw-materials, inventory, crew (staff), budgeting, task management, scheduling etc. In big data analytics, the large amounts of data gathered from various resources will be stored in the HDFS (Hadoop Distributed File System) and then the data will be processed through Map Reduce to obtain the better results.

The construction management will store the data produced from the project construction and will be stored for the future projects. Without the use of computers, the data may be lost or the person handing the management tasks may not carry the past data to the present. But, with the use of computers in the construction management, every piece of information generated will be pushed on to the database and will be accessed in future for further processing to yield better results.

B. Use case 2:
Title:Estimation and Management of Crew for the Proposed project.

Estimation of the requirement resources is a major task for any construction company. The success of the project lies in the correct / approximate estimate of the required resources. Any civil engineering company will go for tenders to get the project. In the tender process, the main job for the key persons is to estimate the resources required like raw-materials, staff, workers, vehicles and the duration of the project. The success in the tender lies purely in the proper estimation. Without estimation, a vague or rough estimate may result in either loss in tenders or generate loss in the execution of the project. Therefore, the estimation is very crucial for the construction company.

As we discussed in Usecase 1, if we store the past historical information about the projects with every piece of information, the big data analytics tools can be used for the prediction of the estimates and the duration of the project. By knowing the exact estimates, we can get the closer budget required for the completion of the project as well. The company which quotes with the predicted estimates from the past historical data has high chances of winning the bid and also successfully execute the project without any hurdles in the later stages.

C. Use case 3:
Title:Prediction of the Severity of the Earth quakes

Earthquake is a natural occurring phenomenon with uncertainty. Earthquakes are one of the natural disasters caused because of the disturbances in the geographical area and also because of civil issues. The occurrence of earthquakes influences the life and property of all human beings. The earthquake engineering provides an in-depth knowledge of earthquakes and also provides guidelines for the minimization of damage due to earth quakes.

In Earthquake engineering, the analysis of civil constructions and the consequences of earthquake are studied. Its like analyzing the past history of the earth quakes and the damages caused due to earthquakes. In this study, the main aim to

design the construction of building in accordance with the standard codes for the buildings so as to minimize the damages.

The proper analysis with the building data and the study of the building and geographical area can help us in the prediction of earth quakes before its occurrence. This prediction is possible with the big data analytics by storing the related variety of information in HDFS and thereby processing that huge amounts of information in MapReduce. Further, other techniques like Neural Networks, Data Mining will also help in the proper analysis and accurate prediction of Earthquakes.

In this paper, we have discussed about the 3 important usecases in civil engineering applied to Big Data Analytics, likewise other use cases also have lot of importance and those can be better analysed and proper decisions can be taken with the application of Bigdata analytics.

VI. CONCLUSIONS

This paper shows the various works related to the constructional engineering where the big data plays a very big role in the management of project, maintaining of resources, scheduling of jobs, estimation of time and cost involved for the civil construction works. Further, this paper discussed about what is big data and big data characteristics along with the big data use cases in civil engineering sector. Big data analytics along with data mining can make the construction industry fruitful and suggests the methods for improving the health of the construction.

VII.FUTURE WORK

The papers can be extended by implementing big data analytics and processing to the use cases discussed. Data available at various construction people/sites may be gathered to form big data and the data mining methods like clustering and classification algorithms can be implemented for the better analysis, decision making and prediction of the new estimates using the past data.

BIBLIOGRAPHY

1) (2016). *2016 ITA Building Products and Sustainable Construction Top Markets Report.* Department of Commerce, USA.
2) Aljumaily, H., Laefer Debra F, &Cuadra, D. (2015). A Big Data Approach for 3D Building Extraction from Aerial Laser Scanning. *Journal of Computing in Civil Engineering* .
3) Bilal M., O. L. (2016). Big Data Architecture for Construction Waste Analytics(CWA): A Conceptual Framework. *Journal of Building Engineering* , 144-156.
4) Hore, A. (2006, March 21). Use of IT in Managing Information and Data on Construction Projects - A Perspective for the IRISH Construction Industry. *Information Technology in Construction Project Management*.
5) Jae-Gil Lee, M. K. (2015). Geospatial Big Data: Challenges and Opportunities. *Big Data Research* , 74-81.
6) Jin, C. (2017). *Real-time Damage Detection for Civil Structures using Big Data*. Doctoral Dissertations.
7) Kunz, M. F. (2014). *The Scope an Role of Information Technology in Construction.* Center for Integrated Facility Engineering.
8) Leu, Y.-M. C.-S. (2011). Integrating data mining with KJ method to classify bridge construction defects. *Expert Systems with Applications* , 7143-7150.
9) Sai On Cheung, H. C. (2004). PPMS: A Web-based Construction Project Performance Monitoring System. *Automation in Construction* , 361-376.
10) Sung, H. K. (2017). Develoment of Road Traffic Analysis Platform Using Big Data. *International Conference on Advances in Big Data Analytics*, (pp. 65-66).
11) (2013). *The Construction Sector in China.* Beijing: EU SME Centre.
12) V.Zicari, R. Big Data: Challenges and Opportunities. In *Big Data Computing* (pp. 103-128).
13) Weisheng Lu, X. C. (2015). Benchmarking Construction waste management performance using big data. *Resources Conservation and Recycling* , 1-23.
14) Zhang S.B, Z. L. (2003). Construction Industry and Enterprises in China: A Comprehensive Review. *CIB TG 23 International Conference*, (pp. 1-12). Hong Kong.