The 6th International Conference on Construction Engineering and Project Management (ICCEPM 2015)
Oct. 11 (Sun) ~ 14 (Wed) 2015 • Paradise Hotel Busan • Busan, Korea
www.iccepm2015.org

The Preliminary Feasibility on Big Data Analytic Application in Construction

Yongho Ko[1] and Seungwoo Han[2]

Abstract: Along with the increase of the quantity of data in various industries, the construction industry has also developed various systems focusing on collecting data related to the construction performance such as productivity and costs achieved in construction job sites. Numerous researchers worldwide have been focusing on developing efficient methodologies to analyze such data. However, applications of such methodologies have shown serious limitations on practical applications due to lack of data and difficulty in finding appropriate analytic methodologies which were capable of implementing significant insights. With development of information technology, the new trend in analytic methodologies has been introduced and steeply developed with the new name of "big data analysis" in various fields in academia and industry. The new concept of big data can be applied for significant analysis on various formats of construction data such as structured, semi-structured, or non-structured formats. This study investigates preliminary application methods based on data collected from actual construction site. This preliminary investigation in this study expects to assess fundamental feasibility of big data analytic applications in construction.

Keywords: Big Data, Construction Data, Construction Management

I. Introduction

The quantity of data existing in the world has shown a dramatic increase during the past decade. It has been investigated that 90% of the data existing in the world has been produced during the last three years [1]. The construction industry has also shown a dramatic increase in in stored data. Construction projects have been increasing in size, difficulty and have become more complicated. Therefore such projects demand reliable process planning and monitoring based on actual performance that can be compared with the construction plan. Accordingly, analytic theories and applicable methods for processed data, which have not been considered systematically in the construction industry, must be investigated. This study investigates data existing in construction sites as a preliminary step to analyze the feasibility of the application of big data analytic methods. The purpose of this study is to develop a modified definition of big data that can be applied to the construction industry. Moreover, this study has conducted a case study on an actual construction site to apply various analysis methodologies on data existing in construction site. We also provide a feasible big data application concept that is expected to provide useful information based on data produced at a construction site.

II. Literature Review

According to McKinsey (2011), big data is defined as data that that are too big to manage using a traditional system. It has been found that most existing data is comprised of a huge quantity of unnecessary parts. Deducting the appropriate data and analyzing the results has been a top issue in the big data research field over the past decade [2].

Big data is not a new concept. However, the reason for the sudden dramatic increase in big data research is due to the appearance of big data analysis methodologies that did not exist in the past. Big data consists of three major parts: volume, velocity, and variety [3]. Volume is defined as the data size of location information, video, media, etc. The volume of the data is defined both by the actual size and by the possibility of analyzing the data. The data is defined as big data if it cannot be analyzed using traditional analysis systems. Velocity is defined as continuously produced data that deals with input and output. Various researchers have suggested methodologies of analyzing data. However, these methodologies have shown limitations in terms of real time analysis and streaming, which indicates that the problem can be defined as big data. Variety is defined as data that have not been systemized, and can be comprised of actual numerical data, text data, images, and videos. Big data requires the use of resources, technology, and specialists. Resources are needed for data collection and quality management, combined with a data collecting strategy. Technology is needed to analyze the processes, infrastructure, and platforms (e.g., technologies such as Apache Hadoop, NoSQL, data mining, etc.). Specialists in fields such as mathematics, engineering, statistics, and psychology are required to develop such technology.

The interest in big data has been changing the focus from analysis and supportive infra-facilities to value production. Countries and companies have been conducting numerous studies in attempts to develop efficient methodologies [4]. Studies have shown that big data has been existing among the world in the past but it has been issued in the recent years due to the appearance of new methodologies that overcome traditional system limitations. Business modeling and decision-making have achieved with the significant current trend that the Cloud and SNS(Social Network Service) are recognized as effective analysis tools. Therefore, developing more efficient and optimized decision-making tools has become a new priority [5]. Currently, various big data analysis programs are targeting

numerous organizations. According to J. M. Park (2012), the major trends of big data researchers are technologies such as advanced data visualization, in-memory databases, real-time reports and dashboards, text mining, data mining and predictive advanced analysis, visual discovery, data warehousing, protest mining and predictive analysis, statistical analysis, and complex SQL. Of all these technologies, the greatest growth potential is given to advanced data visualization (ADV) with emphasis on developing self-serviced user-optimized tools. Numerous organizations have chosen ADV as a visual analysis tool. Numerous big data streaming applications exist currently. However, only a few real-time analysis applications exist; therefore, real-time analysis tool development is the current research trend.

Four major big data analysis technologies are currently being developed: 1) Text mining technology has been used in database structures for information searching. NEC Corporation and the Korean Institute of Science and Technology Information (KISTI) are the main patent-pending organizations. 2) Opinion mining has shown an increase in log analysis researches. Also, conversation analysis, brand monitoring, and market impact analysis-related patents have been filed related to unstructured data analysis methods. 3) The field of social network analysis-related patents is information searching database structure and application systems pended by Yahoo, SMART Link, OWAVE, HP, etc. 4) Cluster analysis has shown a significant increase compared to the other methods; however, is has shown a decrease since 2011. The main field of patent in this field is also information searching database [6].

III. Construction Data

One of the most noticeable characters of the construction industry is uniqueness which prevents collecting precise raw data by consistent collection systems since the environment of construction activities occur in difference conditions for each project. Numerous researches have been suggesting efficient data analysis methodologies, however have shown limitations of reliability due to the reasons explained above. [7].

This limitation in construction has brought other methodologies for data generations such as experimental designs based on statistic knowledge, and simulation techniques based on operational knowledge [8,9,10]. However, this temporary solution have not activated without collected raw data that reflects feasible cases under various conditions. We note that a new approach, based on new concepts such as big data, is required in order to use decision-making tools of proven applicability from fields other than the construction industry. We studied various forms of data produced at a construction site in order to analyze the possibility of applying big data concepts to the site. According to the research by Lee et al., an enterprise resource planning (ERP) system can integrate the major business management functions of the enterprise with a single common database to allow sharing of all information, and achieve efficient communication between management functions [11]. Based on the needs of running a construction enterprise, Lee et al. have shown that ERP has potential applicability to the construction industry. They focused on materials management and illustrated how ERP systems could be implemented, and demonstrated the efficiency of the construction materials management system quantitatively. According to other studies conducted in the construction data analysis field, the simulation technique and statistical methodologies have shown great efficiency in construction data analysis and predictions. Han et al. (2011) suggested a method for data generation based on sensitivity analysis that provides different results depending on the changed resource information. The sensitivity analysis used in their study was based on discrete event simulation, which is a processing simulation modeling technique. Generated datasets functioned as input data for multiple regression analysis (a statistical prediction method). Han et al. (2011) showed that the results based on generated data were more reliable and precise than those based on a small amount of raw data. However, the range of variance of the resources used in the sensitivity analysis is questionable [7]. Lee et al. (2011) collected data from a construction site and developed a simulation model based on the web-cyclone methodology. This model has shown efficiency in predicting productivity; however, it has data collection limitations with respect to time and size. The results of Lee et al. (2011) suggest that there is a need for developing a real time data collecting device or methodology that applies the changing conditions of the construction site. Melzner et al. (2011) presented various data sources functioning as input data in a construction process simulation. They suggested that data ranges were different depending on the planning phases. To achieve reliable planning results using the simulation technique, the establishment of information using building information modeling (BIM) and a vast of amount of data from dynamic data sources (such as construction materials, construction regulations, and construction practices) are required. However, their study provided no specific information about dynamic data sources, a crucial point of the data utilized. In addition, they did not provide a detailed method for integrating the data and simulation [12].

IV. Feasible Application of Big Data Analytics

The concepts of big data have strong possibilities for applications in the construction industry in terms of efficient decision-making tools and analysis methods. We developed a decision model based on the continuously produced job site data. We determined that the job site data consists of various components. Information on productivity and the unit cost of crews and equipment can be found in the literature, such as the series of R. S. Means, and performance books [13]. A large amount of data related to job site labor and equipment has been produced. Daily labor and equipment input and output data have been

produced for the job site using various methodologies. As shown in Figure 1, these data could be closely related to each other [14]. For example, the number of laborers working at the job site on a specific day could affect the productivity and unit cost of a specific process. Figure 1 shows an example of the potential data relationships of data produced in the job site.

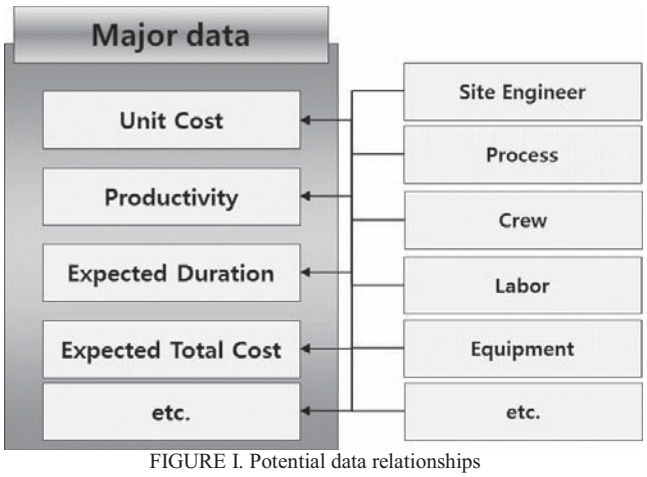

FIGURE I. Potential data relationships

There is a need to divide the data into smaller objects when possible. Based on the relationships shown in Figure 1, the main purpose of the proposed model is to analyze, predict, or monitor the unit cost, productivity, expected duration, and expected total cost. We assumed that each process has a site engineer responsible for it, and its own crews and equipment. Each crew has laborers and a unit cost. Operators usually work on specific equipment. Considering such data relationships at the job site, we note that unit cost and productivity can be affected by the number of laborers in the crew, labor experience, operator experience, and site engineer management skills. The expected duration of the target construction operation can be predicted approximately by analyzing the relationship between productivity and process quantity. The expected total cost can be predicted approximately by analyzing the relationship between unit cost, productivity, and the total quantity of work at the job site. The relationship shown in Figure 1 was assumed; however, it is a common structure for the job site construction process in most cases. Our study suggests concepts for applying big data analysis methods to the job site. In further studies, there is need to analyze relationships by using quantitative data from actual job sites, and add more objective data to the relationship diagram. As we have described, a huge amount of data is produced at the construction site. Our goal is to find an efficient methodology with which to solve numerous construction site problems by analyzing data mined from the site. Moreover, the data could be used to predict values such as productivity and unit cost. According to previous studies regarding construction data, data collection and analysis have shown limitations in terms of size and time. Therefore, the application of big data concepts in construction can be illustrated in Figure 3 [14]. We developed a decision-making and analysis tool that uses construction raw data. The construction raw data was considered to be big data, and was analyzed using methodologies applied in the information technology field for solving big data problems.

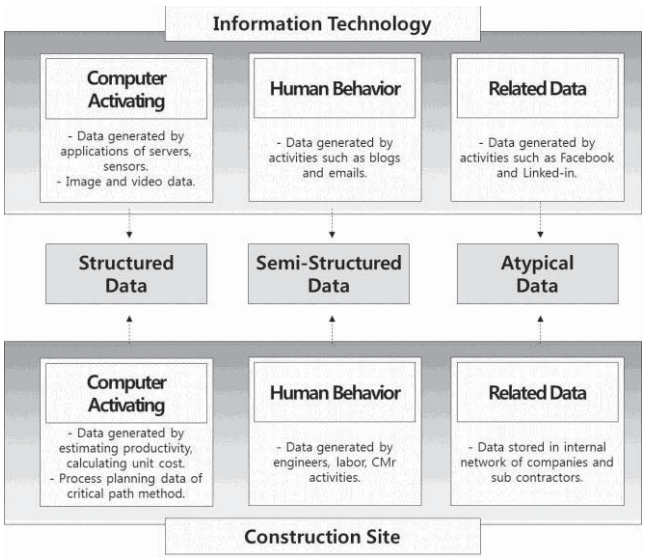

FIGURE II. Big data concept application

According to Bryant et al. (2008), big data technology can be utilized for sensing, collecting, storing, and analyzing. They suggest that it requires digital data by digital sources, a localized sensor computer network such as the Internet, data storage using magnetic disk technology, data-intensive computing systems, and data analysis algorithms that detect patterns and extract knowledge [15]. These big data technology concepts can be applied to the construction industry. The sensing of construction activities can be made available by data collection methods such as radio frequency identification (RFID). The sensor network, data storage, and data computing systems requirements can be accomplished with emerging technologies. However, analytic algorithms have been a significant challenge to the construction industry in that each construction project has a unique process with distinguishing features. Based on the statement above, data collecting and data analyzing can be expressed with data sensing and data analysis in the big data concept, prospectively.

A case study has been conducted based on data collected from an on-going construction project. The data used for the analysis is the daily labor, material and equipment report. The relationship between all labor, material and equipment has been investigated by correlation analysis. Accordingly, it has been found that numerous materials such as reinforced bars, concrete and form have shown statistical significance with various types of labors. The following figure shows an example of the research results. Regression analysis and neural network has been used to develop prediction models based on the results of correlation analysis. Distribution fitting and random

number generation have been applied also to predict total duration of rebar placement operation based on the quantity of rebars measured in tons.

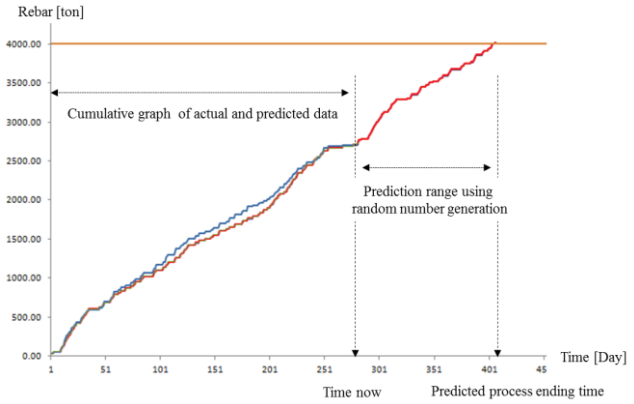

FIGURE III. Case study results on reinforced bar placement operation

V. CONCLUSION

Concepts and analysis methods studied in the information technology industry have been investigated. It has been shown that the concept of big data has brought a large impact not only in the information technology industry but also in various other industries such as the construction industry. The construction industry has also shown a dramatic increase in size and complexity. More efficient and reliable management and predictions are demanded for complex construction projects to prevent cost overruns and construction delays. However, the decisions made at the job site have typically been made based on the site engineers' experience. Therefore, this study investigates the analysis methods and concepts of big data applied in the information technology field to suggest a reliable decision making tool in the construction site. It has been investigated that massive amount of data is produced in the construction site; however, it has been analyzed that engineers in construction sites have been neglecting the use of the data and focused on decision making based on experience only. Numerous studies in the construction field have been made in order to develop methodologies dealing with construction data, and to develop decision-making and prediction models. However, these studies have revealed limitations in data collection with respect to size and time. Thus, there is need for more efficient and automated approaches. Continued research on the application of big data concepts for the construction site is expected to contribute to more objective and accurate decision-making.

ACKNOWLEDGEMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education, Science and Technology (MEST) of the Korean government (NRF-2012R1A2A2A01046193), and Inha University research grant. The authors gratefully acknowledge this support.

REFERENCES

[1] M.M. Kang, S.L. Kim and S.M. Park, "The use and analysis of big data", Special Issue (2012), The Korean Institute of Information Scientists and Engineers, South Korea, 2012.
[2] McKinsey, "Big Data, The Next Frontier for Innovation, Competition, and Productivity", McKinsey & Company, U.S, 2011.
[3] B. Choi, H. Kim, J. Kim and S. Jin, "Data Analytics for CRM in the Age of Big Data", *Entrue Journal of Information Technology*, vol. 11, no. 1, pp.19-27, 2012.
[4] S. Kim, "Recent Global Trends and Issues of Big Data", ISSUE&TREND (2012), KKT Economics and Management Laboratory, South Korea, 2012.
[5] K. Lee, "Utilizing Big Data implementation for smart government", President's Council on National ICT Strategies, vol.146, no. 3, pp.3-4, 2011.
[6] J. Park, T. Um and H. Kim, "Big data analysis technology trends and possible challenges", Korea Institute of Science and Technology, South Korea, Yoohan University, pp. 55-65, 2012.
[7] S. Han, T. Hong, G. Kim and S. Lee, "Technical comparisons of simulation-based productivity prediction methodologies by means of estimation tools focusing on conventional earthmovings", *Journal of civil engineering and management*, vol.17, no. 2, pp.265-277, 2011.
[8] S. Han, T. Hong and S. Lee, "Production prediction of conventional and global positioning system-based earthmoving systems using simulation and multiple regression analysis" *Canadian journal of civil engineering*, vol. 35, no. 6, pp,574-587, 2008.
[9] S. Wang and D. Halpin, "Experiment for Improving Construction Processes", Proceedings of the 2004 winder simulation conference, pp.1252-1259, 2004.
[10] W. Huhnt, S. Richter, S. Wallner, T. Habashi and T. Kramer, "Data management for animation of construction processes", *Advanced Engineering Informatics*, vol. 24, no. 4, pp.404-416, 2010.
[11] S. Lee, A. Arif and H. Jang, "Quantified benefit of implementing enterprise resource planning through process simulation", *Canadian Journal of Civil Engineering*, vol. 31, no. 2, pp.263-271, 2004.
[12] J, Melzner, S. Hollermann and H. Bargstadt, "Detailed Input data source for construction process simulation", SIMUL 2011, The Third International Conference on Advances in System Simulation, pp.132-135, 2011.
[13] S. Han, S. Lee and D. Halpin, "Productivity Evaluation of the Conventional and GPS-Based Earthmoving Systems using Construction Simulation", Construction Research Congress 2005, American Society of Civil Engineers, 10. 1061/40754(183)35, 2005.
[14] Y. Ko, K. Kim, M. Cho and S. Han, "Assessment of Preliminary Concept for Big Data in Construction Management", International Research Symposium on Engineering and Technology, Kuala Lumpur, Malaysia, 2014.
[15] R. Bryant, R. Katz and E. Lazowska, "Big-Data Computing: Creating revolutionary breakthroughs in commerce, science, and society", http://www.cra.org/ccc/docs/init/Big_Data.pdf, 2008, 1-15.