Current Status and Future Opportunities for Big Data Research in the Construction Industry

Donghoon Lee, Kyeong-Tae Jeong, Boong Yeol Ryoo

***Abstract*: Big data technology is expected that these technologies will bring new opportunities and values to various industries. However, the construction industry hasn't put in much effort in using big data.**

This paper investigates past and current studies on big data analytics to identify the future technology for the construction industry. The research trend within the industry is examined for prediction and proposal of future opportunities. The study is conducted in stages as described below. First, this study investigates the implementation cases. After that, we determine the past and current research trends. Finally, this study analysis current trends and future directions of big data implementation.

There is room for research in construction economics and market, business development, and portfolio management. Research activities in non-construction areas such as biology, management, and social science are the same level as the construction industry. As the proposed model for construction projects was generated based valid research activities, a big data analytics model can be used for implementing big data projects. This model is to divide construction data into project-specific (common) data and industry-specific (spatial) data. The model categorizes a variety of data contents based on characteristics of the data, either project specific data or industry-specific data. The data are then grouped by data analysis methods such as statistics, data mining, or machine learning methods. Finally, the data are grouped by the nature of data, either common information or regional information. The benefit of this approach is to assure the success of big data implementation projects by reconciling project information and industry information together since the success of a big data implementation relies on the integration of both data.

Using this study, realistic resource planning and allocation for future can be projected in advance. Moreover, dynamic property of this model can provide to construction companies with real-time industry and project data.

***Index Terms*: big data, implementation cases, construction industry, big data analytics model, big data trend**

I. INTRODUCTION

Recently, there is a great interest in big data due to the drastic data increase. When expressing the volume of data, it is no longer in kilobyte or megabyte. As data sizes increase rapidly, the unit of data size has been increased to gigabyte, terabyte, petabyte or zettabyte. According to a recent article, in the past two years 90 percent of data available have been produced [1] (Sintef 2013). It is possible because there are tools available to ordinary users to create and manage contents without technical training or experience. These tools allow them to personalize information through social media. In addition, the definition of data has been changed. Not only the traditional contents such as text, images, videos, or documents but also business process and rules, software components, data sources, legacy applications, visitor profiles [2] (Ryoo and Skibniewski, 2007). Inexpensive data storage devices and easy social media portal encourage people to participate data creation and share their data. Furthermore, the rapid development of smartphones and social networks led to convenient data sharing. This powerful integration tool is needed to efficiently extract meaningful information from data from different sources. The development of hardware and software brought about rapid data processing, making it easier to perform analyses which had not been possible in the past. This phenomenon provides an opportunity for companies or individuals to extract valuable information than others.

For the construction industry, as the technologies for storing and sharing data have developed, construction companies may store a wide range of data from the corporate level to the project level. In addition, a tremendous amount of data shared on websites can be used to obtain the regional information of construction projects, such as the market, regulation, climate, custom, traffic, technology and customer. Since construction projects adopt an in-situ production method, such regional information is extremely helpful in winning a contract and implementing the project. An increase of such data does not imply an increase of information or knowledge. When the data increase, there is an increase in opportunities to obtain better information. However, the quality of information differs by analysis ability.

Big data analytics is the advanced data analysis process to discover behaviors or patterns and extract valuable and unpredicted patterns, correlations, and information from extremely large data sets [3] (Adams 2009). Komatsu, a global construction machinery company, installed a remote monitoring system on all devices and equipment to identify any stolen construction machinery. The equipment and operation information sent from their system is used to not only prevent the machinery from being stolen but also to accurately learn the trend of world's construction

Revised Manuscript Received on May 22, 2019.

Donghoon Lee, Dept of Architectural Engineering, Hanbat National University, Daejeon, Korea.

Kyeong-Tae Jeong, Dept of Architectural Engineering, Hanbat National University, Daejeon, Korea.

Boong Yeol Ryoo, Dept of Consrtuction Science, *Tex as A&M University, Texas, U.S.A*

business by reading the rate of machinery operation [3] (Adams 2009). As decision-makers need such market and operation information, the system serves as not only a knowledge management system but also a data management system.

It is expected that these technologies will bring new opportunities and values to various industries. However, unlike other industries, the construction industry has not put in much effort in using big data. This paper investigates past and current studies on big data analytics to identify the future technology for the construction industry. The research trend within the industry is examined for prediction and proposal of future opportunities. The study is conducted in stages as described below.

- Investigation of implementation cases
- Determination of past and current research trends, and

Current trends and future directions of big data implementation

II. STATUS OF BIG DATA IMPLEMENTATION

When data are generated and need to be analyzed in order to extract meaningful information. Extracting meaningful information from loosely related data requires new data management analytics. With the increase in data storage technology and methods of data sharing, enormous data have become available easily. As it has become cheaper to store data, decision-makers need to obtain much value from the enormous amount of data extracted [4] (Elgendy and Ahmed, 2014). The size, variety, and rapid change of such data require a new type of big data analytics, as well as different storage and analysis methods. Such sheer amounts of big data need to be properly analysed, and pertaining information should be extracted [4] (Elgendy and Ahmed, 2014). Big data technique refers to a technology used to obtain insight from a vast amount of data that are uncontrollable with the existing data management and analytics systems [5][6] (Fan et al., 2014; LaValle et al., 2011). The technique allows data collection, storage/management and processing within the permissible time limit for users, which is beyond the limitation of general hardware environment and software tools [7] (Mathew et al., 2015). It is because the technique has to deal with the different types (Variety) of huge data (Volume) for a timely data collection/processing/analysis. Unlike the other data analyses, the big data technique includes the analysis on a vast amount of unstructured data that are not cleansed as well as the data cleansed according to a specific form. In addition, its meaning encompasses a series of processes as storing, collecting, analyzing and managing mass data [8][9][10][11][12][13] (Agrawal et al., 2011; Bughin et al., 2010; Buhl et al., 2013; Demchenko et al., 2013; Dumbill et al., 2013; Floridi, 2012).

The big data analytics software eco-system includes distributed file systems, data programming languages, and data analysis platforms [14] (Alexandrov et al., 2014). The big data processing and analysis technique are not heated competition of different techniques, but it is a platform technique focused on core techniques. Major information technology (IT) enterprises also strategically support open source software, and it is incorporated into a platform to promote securing of developers and expanding the base.

Big data analytics platform is based on the big data processing infrastructure, and it is composed of the techniques for data collection/integration, data preprocessing, data storage/management, data analysis and visualization of the data analysis. To take the initiative in the big data market, there is a fierce competition in ensuring a platform integrated with the big data analytics platform techniques, involving data collection/integration, preprocessing, storage/management, analysis/prediction and visualization, and the big data processing infrastructure technology [15-17] (Chen and Zhang, 2014; Habbal 2014; Chen et al., 2016).

As a big data processing framework, Hadoop MapReduce has rapidly become the de facto standard in both industry and academia. Its popularity largely results from the ease-of-use, scalability, and failover properties of Hadoop MapReduce [18] (; Dittrich & Quiané-Ruiz, 2012). This new platform surrounding the big data analytics platform has already started, and Hadoop has consolidated its position as the major platform in the big data platform market of open source software [17][19][20] (Chen et al., 2016; O'Driscoll et al., 2013; He et al., 2011).

Big data are remarkably diverse in terms of sources, data types and entities represented due to different intended purposes. Furthermore, various data that need to be processed continue to increase rapidly due to the emerging applications such as a social network analysis [21] (Abawajy 2015). Big data collection/integration techniques are to logically link data without sacrificing their intended purposes regardless of their data types and formats, contents. New data generation, a collection of external data scattered in network and integration of internal/external heterogeneous data can eventually be achieved. Data preprocessing techniques are to cleanse unstructured stream data are constantly generated. For example, dynamic contents from such as the sensing information and social networking service (SNS) can be structured for analyses. It enhances the accuracy of analyses and allows an in-depth analysis. Data storage/management techniques are distributed computing techniques that store and manage various types of data. Even though the production of data would be increased that have increased sharply such as web data, social media, business data and sensing information [17, 22-24] (Chen et al., 2016; He et al., 2013; Hashem et al., 2015; Ji et al., 2012).

Data analytics is an in-depth analysis technology that uses data analysis methods including machine learning and artificial intelligence of large-scale statistical processing, data mining and graph mining in real time so as to extract the underlying values of big data. Visualization of data analytics is composed of analysis tools that provide an environment for data analyses by non-experts and an

infographics technique that implicitly expresses the analysis result and provides intuitive information. Data processing infrastructure techniques are computing system and system software technologies that offer cloud computing and fabric computing for a real-time processing of mass data, securing high-performance, high-availability and high-scalability [17, 25-27] (Chen et al., 2016; Tirunillai, and Tellis, 2014; Wu et al., 2014; Ofli et al., 2016).

Hadoop started off with the distributed computing, and it has evolved from the techniques to build cloud computing infrastructures to the big data infrastructure technique. It has become an essential technique to constitute the big data analytics platform along with the analytics package tool, R. The Apache Hadoop provides the distributed and parallelised data analysis of petabyte (PB) scale data sets [19, 20, 28] (O'Driscoll et al., 2013; He et al., 2011; Simmhan et al., 2013;). Hadoop was developed to implement the algorithms of Google MapReduce. Hadoop is a large-scale distributed computing support framework for data processing analyses, and Hadoop Distributed File System (HDFS) and MapReduce, the framework for distributed processing are the core components. It is being improved into an integrated solution for big data, as it incorporates HBase, a distribution database, Nutch, a search engine and Hive that supports SQL [18, 29-31] (Dittrich and Quiané-Ruiz, 2012; Chen et al., 2012; Lee et al., 2011; Shim 2012).

III. RESEARCH TREND OF BIG DATA ANALYTICS

A. Research publication in big data

During the 1990s and till the early 2000s, "Information Technology (IT)" was a very important keyword. This is still true, yet the public's interest has changed. Three major information technology trends such as information technology, cloud computing, and big data are targeted to monitor interests of IT researchers. Figure 1 shows the number of search per year as of 2003 when the IT search count was the highest according to the Google data [32] (Google 2016). Information technology is a generic term used to categorize research trends and it has drastically reduced since 2003. On the contrary, research on "Cloud Computing" has been increased after since 2007. However, its search was cut by half after 2010. "Big data" was the last keyword to be searched among the others, and it has continuously increased since 2011, ranking the highest as of now. In this regard, it can be implied that there is an increased curiosity and interest on "Big data." As Figure 1 shows, big data-related research in construction suddenly increased from 2% in 2013 to 16% in 2015. IT related research has been shifted from general information technology to data-oriented research as researchers have understood general information technology for implementation. As new information technologies emerge, as in the forms of more solution, researchers broke into separate research groups to understand new technologies to seek advanced frameworks.

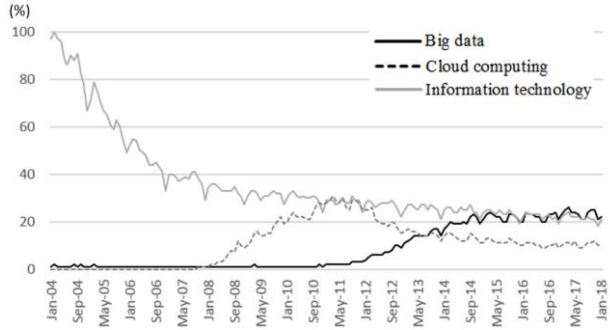

Figure 1. The number of research publications on three major information technologies (Data from www.google.com; 2018)

The studies related to big data analytics have dramatically increased since 2012 as in Figure 1. An investigation of around 200 publications with relatively more quotes shows the level of research activities in various fields between 2012 and 2015 as in Figure 2. The following can be inferred from the studies that lasted for 4 years. Firstly, the studies related to big data analytics have expanded into more areas and topics as big data research matures. Although the field of computer science was the most active, the methods of big data analytics were applied in a wide range of fields, like medicines, management, and economics. Secondly, the number of research publication in non-computer science fields has increased. As illustrated in Figure 2, the ratio of computer science that accounted for approximately 70% in 2012 declined to 44% in 2015, and that of other fields has been increasing. Interestingly big data research in construction has increased from zeros to 16% in 2015. This shows construction understood its potential.

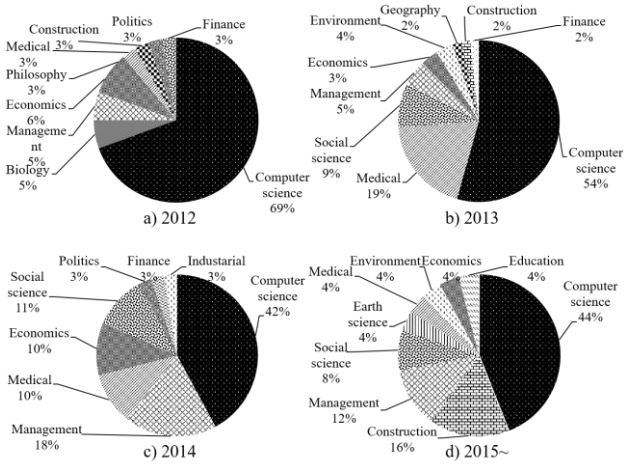

Figure 2. Yearly ratio of the field of bi data related publications

B. Frequency of words sued with big data

The study created Microsoft Excel macros and analyzed the words in sentences that include big data on web. These surveys identified the perception of big data. The search portals were Google scholar, Google news, Yahoo news and university libraries, and the data after

2011 were collected without repetitions. Searching for around 60 hours, 165,062 words were displayed as the search result and the number of characters was 1,017,631. Some supplemental words were used to understand the level of awareness of big data research and implementation among big data researchers. The frequency of words that supplemental words that are used with "big data" is as represented in Table 1. The most frequently used modifier was "new" and "hot", "complex" and "important" were also used to see the level of understanding. This combination showed how people accept big data as a new trend. As Table 1 shows, the big data researchers take big data as a newly introduced technology and it shows the research is in its early stages.

Table 1. The words for level of difficulty used with "big data"

Word	Frequency (ea)	Percent of Total (%)
new	498	48%
hot	151	15%
good	129	13%
complex	90	9%
important	60	6%
hard	37	4%
difficult	26	3%
easy	15	1%
simple	14	1%
bad	12	1%
Total	1,032	100%

Then the searches were also extended into two categories as positive words ("hard" or "difficult" or "complex") and negative words ("easy" or "simple"). As shown in Table 2 it was found that the ratio of negative words was higher than that of positive words. Due to a lack of understanding or implementation procedures, big data is still difficult and complicated for most researchers

Table 2. Positive and negative words used with "big data"

Category	Frequency (ea)	Ratio (%)
hard, difficult or complex	153	84%
easy or simple	29	16%

Table 3 and Figure 3 show the result of the frequency of additional words used in big data research to find application fields. The most frequently used words were "computer" and "science", which were referred to in 2,562 cases, 43 percent of all cases, and they are the words used in the field of big data analytics. Then, "economics" or "business" was used 755 times, and "customer" or "marketing" 579 times. Many firms are interested in consumer tendency, and since all purchase statements are quality data for corporate management, it is believed that such an interest is reasonable. Also, big data set may change the statistical methods used by economists [33-35] (Einav and Levin, 2014; Doornic and Hendry, 2015; Chow-White and Green, 2013). The field of "bio" and "medical" require a lot of data analyses [36] (Aronova et al., 2010), so there is a great interest on big data. In addition, many researchers examine cloud-based services in bioinformatics [37, 38] (Dai et al., 2012; Savage 2014). This is because they are greatly interested in methods to obtain useful information from the social networks and social media [39-42] (Asur and Huberman, 2010; Bello-Orgaz 2016; Manovich 2011; Tsou 2015). Also, the word "social" was frequently used. Even though numerous data are generated from a single construction project, there is very few interest or are not many researches on big data compare to other areas.

Table 3. The professional words used with "big data"

Category	Frequency (ea)	Percent of Total (%)
"computer" or "software"	2,562	43%
"economics" or "business"	755	13%
"customer" or "marketing"	579	10%
"engineering" or "industry"	541	9%
"bio" or "genome"	493	8%
"social" or "history"	391	6%
"government" or "politics"	271	5%
"medical" or "medicine"	200	3%
"statistics" or "mathematics"	160	3%
"construction" or "buildings"	65	1%
Total	6,017	100

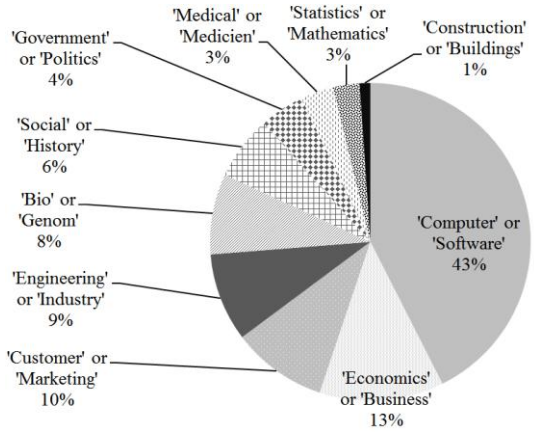

Figure 3. Percent ratio of the professional words which were used with "big data" (2012~2016)

As summarized in Table 4, the research interests related to big data analytics differ by field. When it comes to the field of economics that is relatively researched more, the studies on topics such as prediction of the stock market, business intelligence, bank implements and customer service and Implications for financial managers were conducted. In addition, the studies on disease prevention, personalized healthcare, global infectious disease surveillance, human well-being and public health in the medical field were conducted in connection with big data.

There are only a few studies in the field of construction. Considerable study areas are limited to Building Information Modeling (BIM), smart cities and city planning, and construction waste management performance. As shown in Table 4, there are only a few studies in the field of construction. Representative authors include Hedges [43] (2012), Batty [44] (2013) and Zhang et al. [45] (2015), and the studies on topics such as Building Information Modeling (BIM), Smart cities and city planning and Construction waste management performance were conducted. Other recent studies in biology, management, and social science are also in progress as shown in Table 4. When it comes to the field of economics that is relatively researched more, the studies on topics such as prediction of the stock market, business intelligence, bank implements and customer service and implications for financial managers have been conducted. The studies on disease prevention, personalized healthcare, global infectious disease surveillance, human well-being and public health in the medical field have been conducted in connection with big data.

Table 4. Summary of topics by field

Field	Authors	Focused Areas
Construction	Hedges, K. E.(2012); Batty, M.(2013); Correa, F. R.(2015); Lu, W.(2015); Mathew, P. A.(2015); Zhang, Y.(2015); Bilal, M.(2016); Lu, W.(2016)	Building Information Modeling (BIM), Smart cities and city planning, Construction waste management performance
Biology	Aronova, E.(2010); Schadt, E. E.(2011); Birney, E.(2012); Dai, L.(2012)	Big data problems in biology, Informatics clouds for big data manipulation
Economics	Diebold, F. X.(2000); Chen, H.(2012); Fanning, K.(2013); Mestyán, M.(2013); Einav, L.(2014); Habbal, M.(2014); Newman, N.(2014); Doornik, J. A.(2015); Johnson, J. E.(2012); Bennett, M.(2013); Bhimani, A.(2014)	Prediction the stock market, Business intelligence, Bank implements and customer service, Implications for financial managers
Environment	Hampton, S. E.(2013); Nativi, S.(2015)	Future of ecology, Urban air pollution Energy management
Management	LaValle, S.(2011); Manyika, J.(2011); McAfee, A.(2012); Rabl, T.(2012); Lee, J.(2013); Provost, F.(2013); Gabel, T. J.(2014); GalbRaith, J. R.(2014); George, G.(2014); Milliken, A. L.(2014); Power, D. J.(2014); Tirunillai, S.(2014); Russell, C.(2015)	Enterprise application performance management, Predictive manufacturing systems, Data-driven decision making, Supply chain design and management Marketing, Intelligent manufacturing

Medical	Kubick, W.R.(2012); Barrett, M. A.(2013); Chawla, N. V.(2013); Chute, C. G.(2013); Groves, P.(2013); Hay, S. I.(2013); Kayyali, B.(2013); Murdoch, T. B.(2013); Neff, G.(2013); O'Driscoll, A.(2013); Hoy, M. B.(2014); Khoury, M. J.(2014); Lichtman, J. W.(2014); Raghupathi, W.(2014)	Disease Prevention, Personalized healthcare, Global infectious disease surveillance Human well-being public health
Social Science	Bughin, J.(2010); Manovich, L.(2011); Buhl, H. U.(2013); Chow-White, P. A.(2013); Graham, M.(2013); Kitchin, R.(2013); Scott, S. L.(2013); Bettencourt, L. M.(2014); Kitchin, R. (2014); Kitchin, R.(2014); Lyon, D.(2014); Amoore, L.(2015); Lewis, S. C.(2015); Tsou, M. H.(2015); Ansolabehere, S.(2012); McNeely, C. L.(2014); Höchtl, J.(2015)	Social media, Social communication, Smart urbanism, Journalism, Epistemology, Expertise, Economics, and ethics, Social data, Policy decision making
Concept of big data analytics	Frankel, F.(2008); Brown, B.(2011); Russom, P.(2011); Borkar, V.(2012); Davenport, T. H.(2012); Durand, M.(2012); Lohr, S.(2012); MacLean, D.(2012); Dumbill, E.(2013); Harford, T.(2014); Chen, Y.(2012)	Meaning of big data, Big data management

The most studies have been conducted in the field of computer science, and the major study results and topics per field are as summarized in Table 5. Noticeable activities are: understanding the fundamentals and concepts of big data, trending, and forecasting of big data implementation, case studies, platforms and frameworks for big data processing, big data analytics and processing in the cloud computing environment. Table 5 shows that big data research activities have been driven by computer science and studied by those who need to work with the sizable amount of data. Tale 5 concludes that big data research is in its early stage.

Table 5. Research activities in computer science grouped by topic

Field	Authors	Topical Areas

Trend and forecast to the future of big data	Bryant, R.(2008); Howe, D.(2008); Lynch, C.(2008); Bell, G.(2009); Cohen, J.(2009); Jacobs, A.(2009); Nielsen, M.(2009); Asur, S.(2010); Crawford, K. (2011); Cuzzocrea, A.(2011); He, Y.(2011); Lee, R.(2011); Bakshi, K.(2012); Bizer, C.(2012); Chang, J.(2012); Chen, Y.(2012); Dittrich, J.(2012); Fisher, D.(2012); Kleiner, A.(2012); Labrinidis, A.(2012); Laurila, J. K.(2012); Madden, S.(2012); Mervis, J.(2012); Snijders, C.(2012); Asay, M.(2013); Baru C.(2013); Cukier, K.(2013); Demchenko, Y.(2013); Demirkan, H.(2013); Dong, X. L.(2013); Fan, W.(2013); Ghazal, A.(2013); Junqué de Fortuny, E.(2013); Kaisler, S.(2013); Katal, A.(2013); Mahrt, M.(2013); Mattmann, C. A.(2013); Mayer-Schönberger, V.(2013); Michael, K.(2013); Mouthami, K.(2013); Sagiroglu, S.(2013); Sathiamoorthy, M.(2013); Alexandrov, A.(2014); Fan, J.(2014); Hu, H.(2014); Jagadish, H. V.(2014); Kambatla, K.(2014); Poldrack, R. A.(2014); Wu, X.(2014); Abawajy, J.(2015); Luo, J.(2015); Ofli, F.(2015); Pääkkönen, P.(2015); Bello-Orgaz, G.(2016); Frizzo-Barker, J.(2016)	Challenges and opportunities with big data, Current status, and forecast to the future of big data, Software, Platform for big data analytics
Practical case study	Herodotou, H.(2011); Shim, K.(2012); Elgendy, N.(2013); Hensman, J.(2013); Liang, F.(2013); Liu, Z.(2013); Moniruzzaman, A. B. M.(2013); Dobre, C.(2014); Kim, G. H.(2014); Lazer, D.(2014); Savage, N.(2014); Sejnowski, T. J.(2014); Singpurwalla, N.(2014); Luo, T.(2015); Romanillos, G.(2015)	MapReduce algorithms, Hadoop system, Text mining, Gaussian processes, Decision making process
Cloud computing	Ananthanarayanan, R.(2009); Agrawal, D.(2011); Ji, C.(2012); Simmhan, Y.(2013); Assunção, M. D.(2015); Hashem, I. A. T.(2015); Sharma, S.(2016);	Big data processing in cloud computing environments, Cloud-based software platform for big data analytics, Cloud-based big-data analytics

IV. CURRENT STATE AND FUTURE NEEDS OF BIG DATA ANALYSIS TECHNOLOGY IN THE CONSTRUCTION INDUSTRY

Examining the research activities trend up to now, there are few studies on big data analytics techniques in construction. However, construction projects need new information that is suitable for the projects, and their risk is determined by the information quantity and quality. As the project progresses, there is a tremendous increase in information, including project, local economy and site condition data. Generating and sharing digital data become faster and cheaper. In general, the big data studies on construction industry have been conducted within the limited areas as shown in Table 4. The major topics of the research are related to BIM, project cost, or environment. Table 6 shows topical research areas in construction published after 2012.

Table 6. Current use of big data applications cases

Author (Year)	Title/Description
Bilal et al. (2016)	Construction waste minimization
Guo et al. (2015)	Workers behavior observation
Hedges (2012)	Closing the gap between big data in design and construction
Lu et al. (2016)	Analysis of the construction waste management performance in Hong Kong
Lu et al. (2015)	Construction waste management
Zhang et al. (2015)	Tender price evaluation of construction project

The study proposes a big data analytics model that is suitable for the construction field as shown in Figure 4. The big data in the industry are largely classified into 2 categories. Firstly, they are the data generated from the previous projects. These data include experiences, such the technologies and

cost related to construction. Due to the development of IT technologies, most data in industry are even digitalized and they are stored and archived. Thus, project data have drastically increased, and information that wasn't available is now obtained because of big data analytics.

Project data are consist of common data that are applicable to all projects and regional data that are only applicable to the project. The common data refers to the data such as technique and productivity, and the regional data (spatial data) refer to unit price, climate, transport, and market. Construction projects are greatly influenced by local environments where the projects are being executed, so such a classification is necessary. Since the internet is a primary tool for collecting and sharing project data, as searching the internet, the amount of construction data was increased rapidly. The collected data were divided into regional data (spatial data) and common data.

For construction economics, construction-related data can be collected through the internet and analyzed by data mining techniques or machine learning to extract meaning information. For business development, it can then be organized by specific keywords that are applicable to all projects and regional information that is the project specific. Such information was then categorized by project phases such as project planning, delivery, contract, and project management.

As in Figure 4, a diagram for big data analytics model for implementation of construction projects has been created based on big data research patterns. Through the proposed model, project specific (common) data and industry-specific (spatial) data are integrated for project managers to make reasonable managerial decisions at various management levels. In the end, construction companies can reconcile their business plans and marketing plans with the construction market at the corporate level. Using the market information, realistic resource planning and allocation for future can be projected in advance at the corporate and project levels. Moreover, dynamic nature of this model can provide construction companies with real-time industry and project data. Thus an adjustment of the plans can be made in real time.

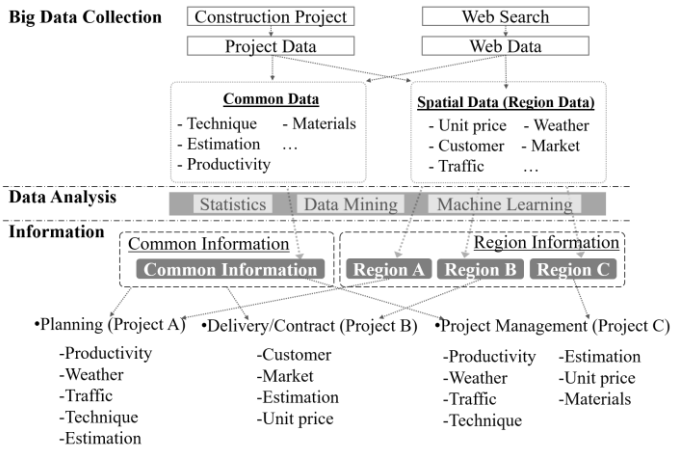

Figure 4. Proposed big data analytics model for implementation of construction projects

V. CONCLUSION

Unlike other areas, there have been relatively few studies of big data analytics in the construction industry. The study examined the research activities and publications since 2012 to find the research trends using a data mining method. The words used with "big data" from the text on web were analyzed to draw the words that are in connection with big data analytics to determine the areas related. As a result, the most frequent words used in the publications were related to "computer" "software", "economics" and "business." Less than 100 cases of "construction" and "building" were found in the publications which imply that there has been a small number of studies or is not much interest on big data related studies in the construction industry.

As presented earlier, the literature review also showed the same result, demonstrating that there have been only a few studies in relation to big data analytics in the industry except popular research areas such as Building Information Modeling (BIM), smart cities and city planning and construction waste management. There is room for research in construction economics and market, business development, and portfolio management. Research activities in non-construction areas such as biology, management, and social science are the same level as the construction industry.

As the proposed model for construction projects was generated based valid research activities, a big data analytics model can be used for implementing big data projects. This model is to divide construction data into project-specific (common) data and industry-specific (spatial) data. The model categorizes a variety of data contents based on characteristics of the data, either project specific data or industry-specific data. The data are then grouped by data analysis methods such as statistics, data mining, or machine learning methods. Finally, the data are grouped by the nature of data, either common information or regional information. The benefit of this approach is to assure the success of big data implementation projects by reconciling project information and industry information together since the success of a big data implementation relies on the integration of both data. In addition it defines the relationships among available data sources if this model is detailed further for the industry.

ACKNOWLEDGMENT

This research was supported by a grant (NRF-2017R1C1B5076057) from the National Research Foundation of Korea by Ministry of Science, ICT and Future Planning.

REFERENCES

1. Sintef, "Big Data, for better or worse:90% of world's data generated over last two years," *ScienceDaily*. May 2013 Available:www.sciencedaily.com/releases/2013/05/130522085217.htm
2. B. Y. Ryoo, M. J. Skibniewski, "e-Portal for Construction Engineers: Integrated Construction Information Portal using 4-Tier

Architecture," *The International Journal of Construction Management*. Jan 2007. 7(1). pp. 83-99. Available:https://www.tandfonline.com/doi/abs/10.1080/15623599.2007.10773097
3. N. M. Adams, "Perspectives on data mining," *International Journal of Market Research*. Apr 2009. 52(1). pp. 11-19. Available:http://wwwf.imperial.ac.uk/~nadams/Adams-talk.pdf
4. N. Elgendy, A. Elragal, "Big Data Analytics:A Literature Review Paper," *Advances in Data Mining. Applications and Theoretical Aspects*. July 2014. 8557. pp. 214-227. Available:https://link.springer.com/chapter/10.1007/978-3-319-08976-8_16
5. W. Fan & A. Bifet, "Mining big data: current status, and forecast to the future," *ACM SIGKDD Explorations Newsletter*. Dec 2013. 14(2). pp. 1-5. Available: https://dl.acm.org/citation.cfm?id=2481246
6. S. LaValle, E. Lesser, R. Shockley, M. S. Hopkins, N. Kruschwitz, "Big data, analytics and the path from insights to value," *MIT Sloan Management Review*. Nov 2011. 52(2). pp. 21-32. Available:http://tarjomefa.com/wp-content/uploads/2017/08/7446-English-TarjomeFa.pdf
7. P. A. Mathew, L. N. Dunn, M. D. Sohn, A. Mercado, C. Custudio, T. Walter, "Big-data for building energy performance: Lessons from assembling a very large national database of building energy use," *Applied Energy*. Feb 2015. 140. pp. 85-93. Available:https://www.sciencedirect.com/science/article/pii/S0306261914012112
8. D. Agrawal, S. Das, A. El. Abbadi, "Big data and cloud computing:current state and future opportunities," *In Proceedings of the 14th International Conference on Extending Database Technology*. Mar 2011. pp. 530-533. Available:https://dl.acm.org/citation.cfm?id=1951432
9. J. Bughin, M. Chui, J. Manyika, "Clouds, big data, and smart assets: Ten tech-enabled business trends to watch," *McKinsey Quarterly*. 2010. 56(1). pp. 75-86. Available:http://www.itglobal-services.de/files/100810_McK_Cloudsbig_data_and%20smart%20assets.pdf
10. H. U Buhl, J. Heidemann, "Big Data: A Fashionable Topic with (out) Sustainable Relevance for Research and Practice?. *Editorial, Business & Information Systems Engineering 2*. Apr 2013. 5(2). pp. 65-69. Available:https://link.springer.com/article/10.1007/s12599-013-0249-5
11. Y. Demchenko, P. Grosso, C. De. Laat, P. Membrey, "Addressing big data issues in scientific data infrastructure," *International Conference on IEEE*. 2013. Available:https://ieeexplore.ieee.org/abstract/document/6567203
12. E. Dumbill, "Making sense of big data," *Editorial, Mary Ann Liebert*. Feb 2013. 1(1). pp. 1-1. Availble:https://www.liebertpub.com/doi/abs/10.1089/big.2012.1503?journalCode=big
13. L. Floridi, "Big data and their epistemological challenge," *Philosophy & Technology*. Dec 2012. 25(4). pp. 435-437. Available:https://link.springer.com/article/10.1007/s13347-012-0093-4
14. A. Alexandrov, R. Bergmann, S. Ewen, J. C. Freytag, F. Hueske, A. Heise, et al, "The stratosphere platform for big data analytics," *The VLDB Journal*. Dec 2014. 23(6). pp.939-964. Available: https://dl.acm.org/citation.cfm?id=2691544
15. C. P. Chen, C. Y. Zhang, "Data-intensive applications, challenges, techniques and technologies A survey on Big Data," *Information Sciences*. Aug 2014. 275(10). pp. 314-347. Available:https://www.sciencedirect.com/science/article/pii/S0020025514000346
16. M. Habbal, "Big data/big compute technologies and the realm of possibilities: The emergence of real-time risk enterprises in the capital markets industry," *The Journal of Trading*. Dec 2013. 9(1). pp. 67-70. Available:https://www.sciencedirect.com/science/article/pii/S0020025514000346
17. Y. Chen, H. Chen, A. Gorkhali, Y. Lu, Y. Ma, L. Li, "Big data analytics and big data science: a survey," *Journal of Management Analytics*. Feb 2016. 3(1). pp. 1-42. Available:https://www.tandfonline.com/doi/abs/10.1080/23270012.2016.1141332
18. J. Dittrich, J. A. Quiané-Ruiz, "Efficient big data processing in Hadoop MapReduce," *Proceedings of the VLDB Endowment*. Aug 2012. 5(12). pp. 2014-2015. Available:https://dl.acm.org/citation.cfm?id=2367562
19. A. O'Driscoll, J. Daugelaite, R. D. Sleator, "'Big data', Hadoop and cloud computing in genomics," *Journal of biomedical informatics*. Oct 2013. 46(5). pp. 774-781. Available:https://www.sciencedirect.com/science/article/pii/S1532046413001007
20. Y. He, R. Lee, Y. Huai, Z. Shao, N. Jain, X. Zhang et al, "RCFile: A fast and space-efficient data placement structure in MapReduce-based warehouse systems," *In Data Engineering (ICDE), 2011 IEEE 27th International Conference on IEEE.* Aprill 2011. pp. 1199-1208. Available:https://ieeexplore.ieee.org/abstract/document/5767933
21. J. Abawajy, "Comprehensive analysis of big data variety landscape," *International journal of parallel, emergent and distributed systems*. Jan 2015. 30(1). pp. 5-14. Available:https://www.tandfonline.com/doi/abs/10.1080/17445760.2014.925548
22. W. He, S. Zha, L. Li, "Social media competitive analysis and text mining: A case study in the pizza industry," *International Journal of Information Management*. Jun 2013. 33(3). pp. 464-472. Available:https://www.tandfonline.com/doi/abs/10.1080/17445760.2014.925548
23. I. A. T Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, S. U. Khan, "The rise of "big data" on cloud computing: Review and open research issues," *Information Systems*. Jan 2015. pp. 98-115. Available:https://www.sciencedirect.com/science/article/pii/S0306437914001288
24. Ji. C, Li. Y, Qiu. W, Awada. U, & Li, K, "Big data processing in cloud computing environments," *2012 12th International Symposium on IEEE*. Feb 2012. pp. 17-23. Available:https://ieeexplore.ieee.org/abstract/document/6428800
25. Tirunillai. S, & Tellis. G. J, "Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation," *Journal of Marketing Research*. Aug 2014. 51(4). pp. 463-479. Available:https://journals.sagepub.com/doi/abs/10.1509/jmr.12.0106
26. Wu. X, Zhu. X, Wu. G. Q, & Ding. W, "Data mining with big data," *IEEE*. Jan 2014. 26(1). pp. 97-107. Available:https://ieeexplore.ieee.org/abstract/document/6547630
27. Ofli. F, Meier. P, Imran. M, Castillo. C, Tuia. D, Rey. N, et al, "Combining human computing and machine learning to make sense of big (aerial) data for disaster response," *Big data*. Mar 2016. 4(1). pp 47-59. Available:https://www.liebertpub.com/doi/abs/10.1089/big.2014.0064
28. Simmhan. Y, Aman. S, Kumbhare. A, Liu. R, Stevens. S, Zhou. Q, et al, "Cloud-based software platform for big data analytics in smart grids," *Computing in Science & Engineering*. Dec 2013. 15(4). pp. 38-47. Available:https://aip.scitation.org/doi/abs/10.1109/MCSE.2013.39
29. Chen. H, Chiang. R. H, & Storey. V. C, "Business intelligence and analytics: From big data to big impact," *MIS Quarterly*. Dec 2012. 36(4). pp. 1165-1188. Available:https://pdfs.semanticscholar.org/f5fe/b79e04b2e7b61d17a6df79a44faf358e60cd.pdf%3E.
30. Lee. R, Luo. T, Huai. Y, Wang. F, He. Y, & Zhang. X, "Ysmart: Yet another SQL-to-MapReduce translator," *2011 31st International Conference on IEEE*. June 2011. pp. 25-36. Available:https://ieeexplore.ieee.org/abstract/document/5961685
31. Shim. K, "MapReduce algorithms for big data analysis," *Proceedings of the VLDB Endowment*. Aug 2012. 5(12). pp. 2016-2017. Available:https://dl.acm.org/citation.cfm?id=2367563
32. Google, Trends. 2016. Available:https://trends.google.com/trends
33. Einav. L, & Levin. J, "Economics in the age of big data," *Science*. Nov 2014. 346(6210). pp. (6210, 1243089). Available:https://science.sciencemag.org/content/346/6210/1243089
34. Doornik. J. A, & Hendry. D. F, "Statistical model selection with "Big Data"," *Cogent Economics & Finance*. Apr 2015. 3(1). Available:https://www.tandfonline.com/doi/abs/10.1080/23322039.2015.1045216
35. Chow-White. P. A, & Green. Jr. S, "Data Mining Difference in the Age of Big Data: Communication and the social shaping of genome technologies from 1998 to 2007," *International Journal of Communication*. 2013. Available:https://ijoc.org/index.php/ijoc/article/view/1459
36. Aronova. E, Baker. K. S, & Oreskes. N, "Big Science and Big Data in Biology: From the International Geophysical Year through the International Biological Program to the Long Term Ecological Research (LTER) Network, 1957—Present," *Hist Stud Nat Sci*. 2010. 40(2). pp. 183-224. Available:http://hsns.ucpress.edu/content/40/2/183.abstract
37. Dai. L, Gao. X, Guo. Y, Xiao. J, & Zhang. Z, "Bioinformatics clouds for big data manipulation," *Biology direct.* Nov 2012. 7(1). pp. 43. Available:https://biologydirect.biomedcentral.com/articles/10.1186/1745-6150-7-43
38. Savage. N, "Bioinformatics: big data versus the big C," *Nature*. May 2014. 509(7502). pp. S66-S67. Available:https://www.nature.com/articles/509S66a
39. Asur. S, & Huberman. B. A, "Predicting the future with social media," *IEEE*. Sep 2010. pp. 492-499.

Available:https://dl.acm.org/citation.cfm?id=1914092
40. Bello-Orgaz. G, Jung. J. J, & Camacho. D, "Social big data: Recent achievements and new challenges," *Information Fusion*. Mar 2016. pp. 45-59.Available:https://www.sciencedirect.com/science/article/pii/S1566253515000780
41. Manovich. Lm, "Trending: The promises and the challenges of big social data," *Debates in the digital humanities*. 2012 Jan 2012. pp. 460-475.Available:https://books.google.es/books?hl=ko&lr=&id=_6mo2tApzQQC&oi=fnd&pg=PA460&dq=Trending:+The+promises+and+the+challenges+of+big+social+data.+Debates+in+the+digital+humanities&ots=9YyinxG8Rz&sig=WJyfgjWlz3toRZtO-iGytR-R93Y#v=onepage&q=Trending%3A%20The%20promises%20and%20the%20challenges%20of%20big%20social%20data.%20Debates%20in%20the%20digital%20humanities&f=false
42. Tsou. M. H, "Research challenges and opportunities in mapping social media and Big Data," *Cartography and Geographic Information Science*. May 2015. 42(sup1). pp. 70-74. Available:https://www.tandfonline.com/doi/abs/10.1080/15230406.2015.1059251
43. Hedges. K. E, "Closing the gap between big data and the design and construction industries with Building Information Modeling (BIM) Schema," *J Archit Eng Tech*. Apr 2012. 1(1). Available:https://www.researchgate.net/profile/Keith_Hedges/publication/271098740_Closing_the_Gap_between_Big_Data_and_the_Design_and_Construction_Industries_with_Building_Information_Modeling_BIM_Schema/links/589ddf00aca272046aa930cf/Closing-the-Gap-between-Big-Data-and-the-Design-and-Construction-Industries-with-Building-Information-Modeling-BIM-Schema.pdf
44. Batty. M, "Big data, smart cities and city planning. *Dialogues in Human Geography*. Dec 2013. 3(3). pp. 274-279. Available:https://journals.sagepub.com/doi/full/10.1177/2043820613513390
45. Zhang. Y, Luo. H, & He. Y, "A system for tender price evaluation of construction project based on big data," *Procedia Engineering*. Oct 2015. 123. pp. 606-614. Available:https://www.sciencedirect.com/science/article/pii/S1877705815032154

AUTHORS PROFILE

First Author Donghoon-Lee, Professor, Department of Architectural Engineering, Hanbat National University, Daejeon, Korea.

Second Author Kyeong-Tae Jeong, The master's course, Department of Architectural Engineering, Hanbat National University, Daejeon, Korea.

Third Author Boong Yeol Ryoo, Professor, Department of Construction Science, Tex as A&M University, Texas, USA.