Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 172 (2017) 544 – 549

Procedia Engineering

www.elsevier.com/locate/procedia

Modern Building Materials, Structures and Techniques, MBMST 2016

Big data in building energy efficiency: understanding of big data and main challenges

Natalija Koseleva*, Guoda Ropaite

Faculty of Civil Engineering, Department of Construction Technology and Management, Vilnius Gediminas Technical University, Sauletekio al. 11, 10223, Vilnius

Abstract

Data generation has increased drastically over the past few years. Data management has also grown in importance because extracting the significant value out of a huge pile of raw data is of prime important thing to make different decisions. One of the important sectors nowadays is construction sector, especially building energy efficiency field. Collecting big amount of data, using different kinds of big data analysis can help to improve construction process from the energy efficiency perspective. This article reviews the understanding of Big Data, methods used for Big Data analysis and the main problems with Big Data in the field of energy.

Peer-review under responsibility of the organizing committee of MBMST 2016

Keywords: Energy Efficiency; Big Data; Characteristics of Big Data; Big Data analysis; Construction; Web of Science.

1. Introduction

The building sector is evolving to be the greatest energy consumer around the world, accounting for 40% of the global energy use and one third of the global greenhouse gas emissions. As a result, building energy efficiency has become one of the top concerns of a sustainable society and attracted increasing research and development efforts in recent years [1-2].With the rapid development of sensor technology, wireless transmission technology, network communication technology, cloud computing, and smart mobile devices, large amounts of data has been accumulated

* Corresponding author.
E-mail address: natalija.koseleva@vgtu.lt

Peer-review under responsibility of the organizing committee of MBMST 2016
doi:10.1016/j.proeng.2017.02.064

in almost every aspects of our lives. Moreover, the volume of data is growing rapidly with increasingly complex structures and forms. A research report of International Data Corporation (IDC) pointed out that 1.8ZB data were created and replicated in 2011 worldwide, and it is estimated that this figure will increase by 50 times by the year 2020. The energy big data provides a new way to analyze and understand individuals' energy consumption behavior, and thus to improve energy efficiency and promote energy conservation [3-5].

2. The understanding and the use of Big Data

Big data is a popular phenomenon that aims to provide an alternative to traditional solutions based on databases and data analysis. Big data is not just about storage or access to data; its solutions aim to analyze data in order to make sense of them and exploit their value. Big data refers to datasets that are terabytes to petabytes (and even exabytes) in size, and the massive sizes of these datasets extend beyond the ability of average database software tools to capture, store, manage, and analyze them effectively.

The concept of big data has been defined through the 3V model, which was defined in 2001 by Laney as: "high-volume, high velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making". More recently, in 2012, Gartner updated the definition as follows: "Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization". Both definitions refer to the three basic features of big data: Volume, Variety, and Velocity. Other organizations and big data practitioners have extended this 3V model to a 4V model by including a new "V": Value. This model can be even extended to 5Vs if the concept of Veracity is incorporated into the big data definition [6-9]. The concepts in the energy sector can be briefly described as follows (Fig. 1):

- *Volume*: refers to large amounts of any kind of data from any different sources, including mobile digital data creation devices and digital devices. The benefit from gathering, processing, and analyzing these large amounts of data generates a number of challenges in obtaining valuable knowledge for people and companies [6]. In energy sector, the wide deployment of smart metering devices (e.g., smart meters) created massive amounts of data. For instance, for a distribution network with 1 million smart meters deployed, the amount of electricity consumption data collected in one year is very large. According to Zhou ant other scientists, the amount of electricity consumption data collected by 1 million smart meters in a distribution network within one year, the amount of electricity consumption data collected once every 15 mins by 1 million smart meters within one year will be up to 2920 TB. This presents not only a storage problem, but an analytic problem of making sense of all that data.
- *Velocity*: refers to the speed of data transfers. The data's contents are constantly changing through the absorption of complementary data collections, the introduction of previous data or legacy collections, and the different forms of streamed data from multiple sources. From this point of view, new algorithms and methods are needed to adequately process and analyze the online and streaming data [6]. Velocity mainly means the speed of energy big data collection, processing and analysis. Different from traditional post-processing type business intelligence and data mining, the collection and processing of energy big data need surprising speed. To support the near real-time decision-makings in energy system, the speed of data collection and processing ranges from sub-second to 5 or 15 mins intervals [3].
- *Variety*: refers to different types of data collected via sensors, smartphones or social networks, such as videos, images, text, audio, data logs, and so on. Moreover, these data can be structured (such as data from relational databases) or unstructured in format [6]. Energy big data has a high degree of variety. Generally, it is a mix of structured (e.g., the energy consumption data), semi structured (e.g., data exchanged between smart energy management platform and third-party data aggregators using XML, Web services), and unstructured data (e.g., email or SMS notification about energy use, interactions of consumers on social media about their energy use). In addition, there are also some inter-industry data (e.g., electric vehicle-related data) and outside-industry data (e.g., weather data) in the energy big data. These different types of data all combined will result in a significant increase in the complexity of energy big data applications [6].
- *Value*: refers to the process of extracting valuable information from large sets of social data and it is usually referred to as big data analytics. Value is the most important characteristic of any big-data-based application,

because it allows generating useful business information [6]. Energy big data is meaningless unless its value is explored and mined, to support either the business decisions or customer services. For energy products and service providers, it helps them to better understand the energy consumption patterns of different consumers and thus to develop more competitive marketing strategies. For customers, the value typically translates to energy savings, operational efficiency, and improved visibility into how they are using energy [3-6].

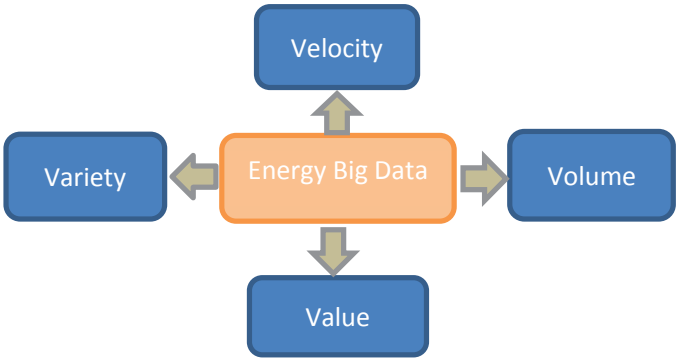

Fig. 1. 4 Characteristics of Big Data [3].

3. The main methods of Big Data analyses

Big Data is used efficiently in many fields of activity, such as: automotive industry; hi-tech; oil and gas industry; telecommunication sector, medicine and healthcare; media and show business; travel and transport sector; social media and online services; information and communication sector. One of the fields where Big Data can be gained and transformed into the useful information is Energy.

In the academic field Big Data can be accepted as a new concept. The analysis of the Big Data topic has been done online. It was found 5160 publications, including articles (1619) in different databases (Fig. 2).

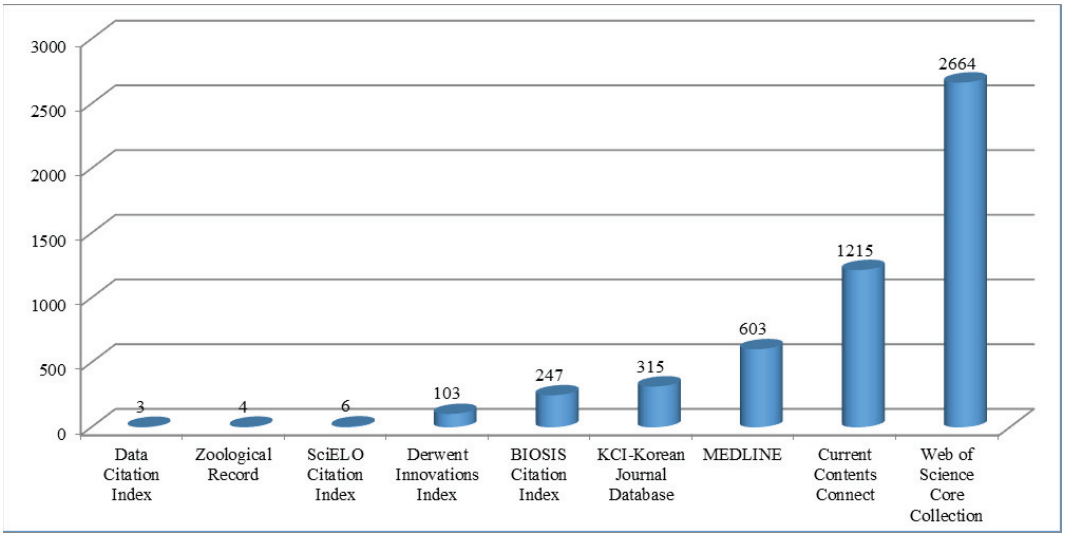

Fig. 2. Number of publications on the “Big Data” topic in different databases (total: 5160).

Mainly publication was found in Web of Science database. It contains 2664 referred publications on the topic of Big Data (15 January 2016), covering all document types, including articles (1060) (Table 1).

Table 1. Publications on the topic of Big Data in Web of Science database.

Publications on Big Data	Number of Publications
All	2664
Articles	1060

As depicted in Fig. 3, the first science research on the topic of Big Data was done in 1974. The extent of research in the area has been rapidly increasing during the last ten years. Numbers of publications on Big Data increased from one-to-two papers per year up to 110 in 2012. More than 90 percent publications were published in the last three years (2013-2015).

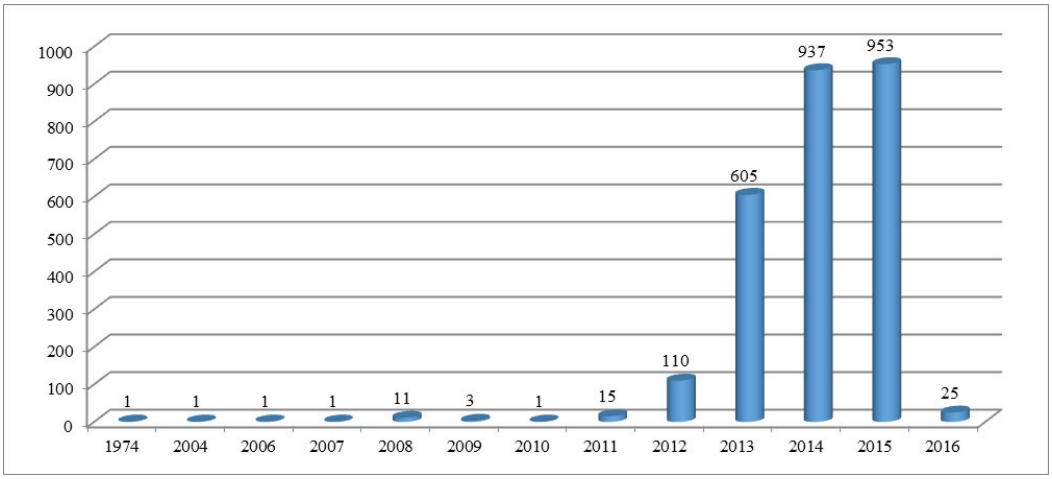

Fig. 3. Number of publications on the "Big Data" topic in Web of Science Core Collection database by year (total: 2664).

The several methods of the Big Data analysis in energy sector, based on data acquisition could be used: text analytics, audio analytics, video analytics, social media analytics, and predictive analytics. Ozkose and other scientists define those methods as follows:

- Text analytics is used for information retrieval from data. E-mails, blogs, online forums, news and call center records are all examples of text data. Text analytics involve machine learning, statistical analysis and computational linguistics. Text analytics enable to extract meaningful summaries from large scale data. Information Extraction, Text Summarization, Question Answering and Sentiment Analysis are some of the techniques used in text analytics.
- Audio Analytics is used to extract information from unstructured audio data. Call centers is commonly used utilization areas of audio analytics. Audio analytics can be used in numerous fields such as increasing the customer experience, the performance of customer representative and the sales rate; comprehending several tasks such as customer behaviors and the troubles of products.
- Video analytics is the usage of various techniques to extract meaningful information, track and analyze video streams. Marketing and operations management is the main application area of video analytics.
- Social media analytics is the analysis of the structured and unstructured data on the social media channels. Social media can be categorized as follows Social networks (Facebook, LinkedIn), Blogs (BlogSpot, WordPress), Microblogs (Twitter, Tumblr), Social news (Digg, Reddit), Social bookmarks (Delicious, StumbleUpon), Media

sharing (Instagram, YouTube), Wiki (Wikipedia, Wikihow), Question-and-answer sites (Yahoo! Answers, Ask.com), Review sites (Yelp, TripAdvisor) [10-12].

Predictive analytics is based upon estimating future considering current or stale data. Predictive analysis is used to capture the relationships of data and discover the patterns. Predictive analytics which is primarily based on statistical methods is highly applicable on many disciplines [10-12].

4. The main problems with Big Data

The energy consumption behavior of households can be described in three dimensions, namely time dimension, user dimension and spatial dimension (Fig. 4).

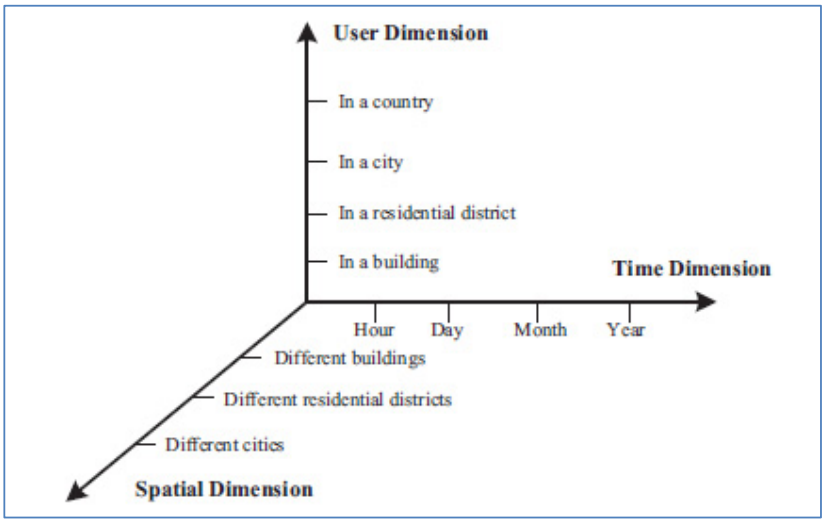

Fig. 4. Different dimensions of energy consumption.

According to Adam Jacobs, it's the big truth about big data in traditional databases: it's easier to get the data in than out. Most database management systems are designed for efficient transaction processing: adding, updating, searching for, and retrieving small amounts of information in a large database. The trouble comes when we want to take that accumulated data, collected over months or years, and learn something from it—and naturally we want the answer in seconds or minutes.

In the time dimension (Fig. 4), the energy consumption behavior of households can be in an hour, a day, a month or even a year. The energy consumption behaviors of households in a day often show some differences in different times of the day. In contrast, the monthly and annual energy consumption behaviors are usually affected by many external factors [3-5]. In the user dimension (Fig. 4), the energy consumption behaviors of different households also vary greatly. Individual's energy use behavior is generally influenced by various factors, including both internal and external ones. Internal factors are the subjective intentions, such as habit and environmental awareness. External factors mainly include housing characteristics, demographic characteristics, way of working, and other factors [3-5].

Other problem is that big data become bigger and bigger. The Web log records millions of visits a day to a handful of pages; the cellphone database stores time and location every 15 seconds for each of a few million phones; the retailer has thousands of stores, tens of thousands of products, and millions of customers but logs billions and billions of individual transactions in a year. Scientific measurements are often made at a high time resolution (thousands of samples a second in neurophysiology, far more in particle physics) and really start to get huge when they involve two or three dimensions of space as well. In addition, the energy consumption (Fig. 4) behavior of households in spatial dimension also shows certain differences. In different areas, householders' energy uses are

often affected by the geographical environment, level of economic development, climate characteristics and other factors. In a smaller spatial range, household energy use behaviors in different residential districts, or even different buildings, also show some differences, due to the impact of regional location, building structure and other spatial differences [3-5]. So the amount of data in energy sector is growing up every time.

Another major challenge for data analysis is exemplified by applications with hard limits on the size of data they can handle. Occasionally the limits are relatively arbitrary; consider the 256-column, 65,536-row bound on worksheet size in all versions of Microsoft Excel prior to the most recent one. Such a limit might have seemed reasonable in the days when main RAM was measured in megabytes, but it was clearly obsolete by 2007 when Microsoft updated Excel to accommodate up to 16,384 columns and 1 million rows. According to Adam Jacobs, Excel is not targeted at users crunching truly huge datasets, but the fact remains that anyone working with a 1-million-row dataset is likely to face a 2-million-row dataset sooner or later, and Excel has placed itself out of the running for the job.

5. Conclusions

The building sector is a huge energy consumer and it is very important to construct and manage the buildings in a sustainable way. The building energy efficiency has become one of the top concerns of a sustainable society and attracted increasing research and development efforts in recent years. The Big Data analysis could be the one of the methods used to analyze and understand individuals' energy consumption behavior, help to improve energy efficiency in building sector and promote energy conservation.

First science research on the topic of Big Data was done in 1974, but the extent of research in this area has been rapidly increasing only during the last ten years. More than 90 percent publications were published in the last three years (2013-2015). It is a result of fast growth of new technologies, new methods of collection and storage of big amount of different kinds of data. Furthermore, it means that it is becoming more and more important not only collect the data, but also extract as much useful information from it as possible.

Three main problems with Big Data in energy field marked in this paper: taking out the accumulated data in a short time, very big amount of information when involving several dimensions, limited possibilities of existing applications to process big amount of data. In order to analyze and understand individuals' energy consumption behavior, to improve energy efficiency and promote energy conservation, it is necessary to solve challenges rising working with Big Data.

References

[1] C. Fan, F. Xiao, H. Madsen, D. Wang, Temporal Knowledge Discovery in Big BAS Data for Building Energy Management, Energy and Buildings 190 (2015) 75–89.
[2] P.A. Mathew, L.N. Dunn, M.D. Sohn, Big-data for building energy performance: Lessons from assembling a very large national database of building energy use, Applied Energy 140 (2015) 85–93.
[3] K. Zhou, C. Fu, S. Yang, Big Data Driven Smart Energy Management: From Big Data to Big Insights, Renewable and Sustainable Energy Reviews 56 (2015) 215–225.
[4] L. Mashayekhy, M.M. Nejad, D. Grosu, Energy-Aware Scheduling of MapReduce Jobs for Big Data Applications, Ieee Transactions on Parallel and Distributed Systems 26(10) (2015) 2720–2733.
[5] J. Cooper, M. Noon, C. Jones, Big Data in Life Cycle Assessment, Journal of Industrial Ecology 17(6) (2013) 796–799.
[6] G. Bello-Orgaz, J. Jung, D. Camacho, Social Big Data: Recent Achievements and New Challenges, Information Fusion 28 (2016) 45–59.
[7] H. Fang, Z. Zhang, C.J. Wang, A Survey of Big Data Research, Ieee Network 29(5) (2015) 6–9.
[8] B.T. Hazen, C.A. Boone, J.D. Ezell, Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications, International Journal of Production Economics 154 (2014) 72–80.
[9] W.Q. Meeker, Y. Hong, Reliability Meets Big Data: Opportunities and Challenges, Quality Engineering 26(1) (2014) 102–116.
[10] H. Ozkose, E. Ari, C. Gencerb, Yesterday, Today and Tomorrow of Big Data, Procedia - Social and Behavioral Sciences 195 (2015) 1042 – 1050.
[11] J. Jose Camargo-Vega, J. Felipe Camargo-Ortega, L. Joyanes-Aguilar, Knowing the Big Data, Revista Facultad de Ingenieria 24(38) (2015) 63–77.
[12] H. Li, K. Lu, S. Meng, Big Provision: A Provisioning Framework for Big Data Analytics, Ieee Network 29(5) (2015) 50–56.