Industrial Applications of Big Data: State of the Art Survey

Amir Mosavi[1(✉)], Alvaro Lopez[1], and Annamária R. Varkonyi-Koczy[1,2]

[1] Institute of Automation, Óbuda University, Bécsi Way 94-96, Budapest 1034, Hungary
{amir.mosavi, Alvaro.Lopez, koczy.annamaria}@kvk.uni-obuda.hu
[2] Department of Mathematics and Informatics, J. Selye University, Komarno, Slovakia

Abstract. Big data analytics has become an important tool for the progress and success of a wide range of businesses and industries. Its diversity and flexibility offer a steady increasing scope for the several applications to stay competitive in the market. For that, big data approach provides several advantages such as advanced analytics, intelligent optimization, informed decision making, large-scale modeling, and accurate predictions. Due to the numerous advantages, it has been particularly possible to find more accurate and feasible solutions for the current engineering problems. Hence, the impact of big-data analytics in the engineering realm and applications is increasing more than ever. This article presents a survey to investigate how engineering community has adopted big data technologies to stay competitive. To conduct the investigation a state of the art survey of the academic literature on the big data applications to engineering is presented.

Keywords: Big data · Data science · Big data analytics

1 Introduction

Over the last years, data has been growing in a large scale. Due to different activities, every day society is generating data in several fields. Nowadays there is a dramatic increase in data generation. It brings new challenges for data storing, data processing and the way to get useful information from big datasets. The challenge is to analyze big datasets and the aim is to find patterns within them. Hence, performing big data analysis is a suitable solution for current problems in industry [1].

Big data is mainly used to describe a huge dataset generated with a high rate. Compared to the traditional datasets, big data includes a significant amount of structured and unstructured data which demands more time and complex resources to analyze it. Big data analysis devolves upon the introduction of new process and new technologies to find the potential value in data. Furthermore, people are looking for trends in enormous quantities of data and they are performing analytics on it to discover what is happening. By applying big data, it is possible to simplify data analysis and understand some features from big datasets [2].

D. Luca et al. (eds.), *Recent Advances in Technology Research and Education*, Advances in Intelligent Systems and Computing 660, DOI 10.1007/978-3-319-67459-9_29

Social human activities can be analyzed by algorithms. After an algorithm is performed, a set of attributions will appear [3]. These attributions will help to create new models for services, launching a new product or even create a new technology. Previously, this procedure was based on guesswork or not appropriate procedures; but now it can be created based on the data itself. Furthermore, big data analysis will bring a new lifestyle for people, and companies can improve their processes [4]. Consequently, companies are able to better understand their business, customers, manufacturing products and market; which will enhance efficiency related to sales, production costs, customer service and other aspects.

Big Data is present everywhere in numerous applications. Several fields in industry are demanding to analyze their data in order to better understand their environment. Furthermore, big data is becoming a main part of the IT department in companies. Regarding decision making, big data analysis will be the next challenge for the innovation, competition, and productivity. In addition, many solutions will be generated to support all the issues involved in this context.

2 Big Data

In the present day, big data has become an important topic which is discussed by several researchers. Before defining what big data is, it is necessary to define information and the difference between dataset types. Data can be defined as structured and unstructured elements which in a suitable manner can be processed by intelligence systems and computers. Yet datasets may not necessarily have a particular meaning before processing. Data becomes information only when it communicates a meaning [5].

Big Data is a concept about digital information in a big scale. Hence, having a suitable definition for big data is necessary. According to [3, 6] big data is a larger collection of heterogeneous datasets which traditional databases or software are not able to process. In order to process big datasets, clusters sets of super computers are needed working toward the same aim.

Understanding the nature of big data and its features, open the possibility for new technology development, architecture model, and algorithms. Regarding dataset features and properties, there is other definition about big data. As described in [7], big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making". Hence, volume, velocity, variety, and complexity of big data continue to be an important challenge for computer systems.

According to [8], there are three important big data features: velocity, volume, and variety. Velocity refers to how fast data is generated or transmitted. Volume feature refers to the amount of data that has been generated in a period of time. Variety refers to the heterogeneity embracing different types of the data diversity, regarding data sources. Within big data framework, there are three important levels: Data Acquisition, Data Processing, and Data Services. These levels should be characterized by efficiency, processing time, flexibility and scalability [9].

3 Big Data in Industry

Nowadays, several applications, services, and products are offered by different sectors and people are demanding access to these services or try to obtain new products according to their preference and tendency. Regarding people's culture, age, gender, education, and other aspects; their preferences for a certain service or product change. An industry responsibility is to find better solutions for a specific problem taking into account the aspects mentioned above, achieve a fault-free and a cost efficient process within the companies. Solving problems in engineering fields is a demanding decision making process and should be considered from different perspectives [10].

Although for enterprises finding a solution for a given problem is important, they are also identifying how to increase their revenue and processes optimization. Therefore, decreasing costs during the manufacturing process is imperative. Therefore, big data analysis plays an important role during the fourth industrial revolution as it helps to find a low cost strategy for companies to be more competitive [11].

4 Big Data in Mobile Networks

Applying big data in mobile networks is possible in two scenarios. One of them is for business support, business modeling and marketing [11]. The second scenario is to apply big data analysis for optimizing the mobile network design. Big data analysis provides three capabilities for fifth generation (5G) design [7]. The first is full intelligence of the current network status, the second is predicting user's behavior, and the third is related to the dynamic response association to the network parameters [12]. The 5G technology will allow users to have high speed Internet access consequently new applications and services will emerge.

For launching 5G, researchers have to investigate big data sets and find information to enhance technical parameters such as bandwidth, mobile network distribution, mobile network architecture, and others. With the advances of big data analysis in mobile networks, operators can have a deep insight when different events occur in the network. Thus, correlations can be determined to be within these events, and they will help in resources optimization and operational costs. Considering Quality of Experience (QoE) [7], there are important challenges regarding 5G standardization, the way to enhance the services and make the network more efficient than 4G networks. Mobile network operators are looking for the best way to adjust traffic requirements and optimize resources allocations. All these activities are performed by big data analysis intelligent usage. The data is collated from different layers of 4G networks. The data collated from these networks is very complex; it has features such as multisource data, high volume, unstructured, and real time data [13].

In a mobile network there are a variety of parameters which can be analyzed by a big data platform. These parameters are divided within four groups: Subscriber level, Cell level, Network core level, and others sources [14]. In the subscriber level, parameters such as call success rate, call drop ratio, speech quality, and hand over success are analyzed. In the same level, IP traffic flow, IP throughput and data streaming quality are also collated. Radio measurements are collected on cell level such

as: received interference power, thermal noise power, channel baseband power, etc. Additionally, the number of user per cell, received random access, preamble per cell are collected Regarding Network core level, historical alarms logs devices configuration, network performance, call data records are collected [15]. In other sources, customer retention manager, costumer complaint center and spectrum utility maps are collated as well. The analysis of all parameters enumerated above will help to determine a proper decision. Thus, big data visualization plays an important role after the big data analysis/ Good visualization techniques should be applied in order to avoid information lost. Finally, a suitable big data optimization helps managers to take the best decision for a certain problem [16].

Figure 1 illustrates the number of literature on big data analysis in mobile networks publications from different publishers from 2013-2016. It shows that in 2016, there has been a major increase in the number of publications in this field, this trend continues in 2017. Figure 1 well represents the evolution of big data applications throughout the past few years and, how this technology is used for optimizing processes in mobile networks.

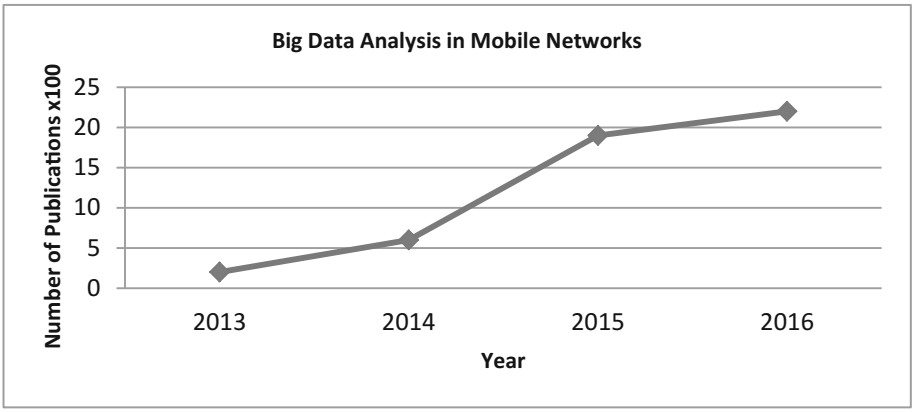

Fig. 1. Number of publications on big data analysis in mobile networks application. (source: village database www.engineeringvillage.com)

4.1 Big Data in Industry 4.0

According to [17], in 2050 about 70% of the world population will live in cities. Consequently, new challenges will appear for different sectors. Companies, who provide water supply, electricity, transportation, and health, should deal with the city overcrowding population. Furthermore, governments want to have a better control of natural resources and also they are looking for better life conditions for their citizens. All these areas can be addressed by an important technology which will help to optimize all the processes within a city; this technology is called the Internet of Things (IoT) [18]. It allows connecting multiple devices within a common network. Devices like sensors, actuators, and others, can be installed in different elements or other

electronics devices [19]. Consequently, it is possible to have an integral management of them, but the most fascinating advantage is that all this management can be possible through the Internet. Authors in [20] describes how big data analysis in IoT works beside an intelligent traffic controllers system and the big advantages this technology can bring into a smart city. People can decide the suitable route depending on the current traffic status. Likewise, traffic optimization network and traffic forecast can be accomplished and it will help in decision making. On the other hand, there are advantages related to air pollution. By big data analysis it is possible to create a suitable traffic model which helps to reduce the fuel consumption in vehicles.

The IoT is also related to Industry 4.0 [21], where it supports factories with a rapid product developing, flexible production and complex environment. The age of smart factories is coming, and intelligent and customized products can be manufactured in a shorter period of time with real-time considering the customer's preferences [21]. Such automation process is generating big datasets. It is worth to analyze these datasets in order to find new strategies for the supply chain processes to increase profitability. The big amount of data generated by all these devices connecting to the Internet will generate a big amount of structured and unstructured data. This sharp increase converges once again in a big data analysis [20]. Nevertheless, the data gathered by IoT have different features comparing to traditional big data because of data generation, data interoperability, and data quality [4]. Interpretation of big datasets from IoT is a challenge because the data sources are ubiquitous; the transmitted data is noisy, heterogeneous, and spatiotemporal dependent [18]. Figure 2 illustrates the progress in number of publications on big data in IoT over the past four years.

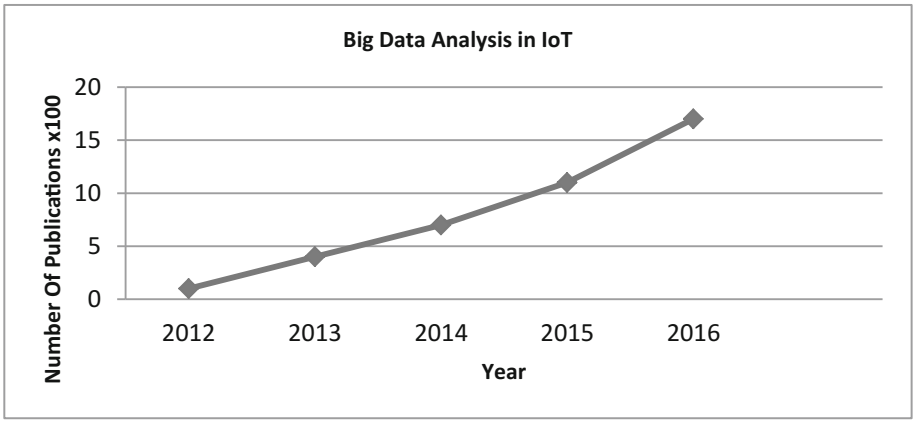

Fig. 2. Number of publications about big data analysis in IoT (source: village database www.engineeringvillage.com)

4.2 Big Data in Mechanical Engineering

Over the last decade, the automobile industry has developed important solutions for car driving, and all the mechanical and electronic systems within Electronic Vehicles (EV) [19, 22]. These developments allowed manufacturing autonomous vehicles,

which are equipped with advanced sensing, navigation devices, communication capabilities and computer vision. Thus, these entire new characteristic in vehicles are a potential support for all the user and transportation systems since it can avoid crashes, reduce the travel time, assisting traffic flows, among other benefits [17]. There are several sensors installed in vehicles, which help in different vehicle functions and, at the same time, provide large amount of data for research purposes. By the analysis of data collected from vehicles, it is possible to improve all the systems inside. Using Big Data techniques, all the function in the vehicle can be substantially improved [22, 23].

Regarding the range estimation for EV in [23], there are several parameters to analyze and this big amount of data has different levels of accuracy, relevance, and unstructured ways. Big Data analysis provides much better estimation for vehicle driving range [24]. In order to give an optimum solution for this particular problem, the data acquisition, from e.g., state of charge of the battery, battery manufactures model, driving history, the model of the vehicle, GPS location, weather conditions, traffic report etc. are collected and categorized according to their properties within three groups. The first is standard data, this category includes data such as GPS position, weather conditions, and estimation of the driving time to the destination. The second is historical data, within this group there are parameters like mile per gallon, and the data from other people who did the same trip are collected. The third is the real time data, which includes data collected when an event occurs unexpectedly, for instance a traffic jam due to an accident [24]. Figure 3 illustrates the progress in number of publications from different publishers for big data analysis over the past 8 years in mechanical engineering.

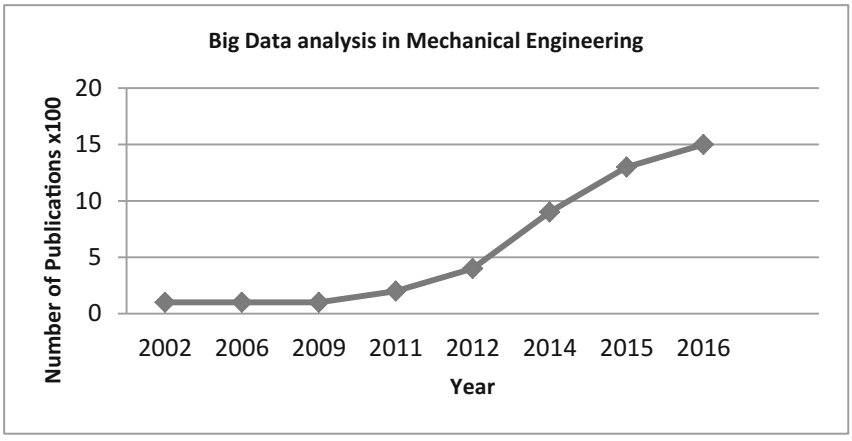

Fig. 3. Number of publications about big data analysis in mechanical engineering. (Source: village database www.engineeringvillage.com)

4.3 Big Data Future Challenges

Big Data analysis involves several challenges and nowadays the research at the industry and laboratories are in the beginning phase. There is much work to do and big efforts are needed to improve different features in big data analysis. Moreover, there

will be new fields in which a big data analysis will be the solution to tackle the problem. According to [20, 21] there are several problems to be solved as it is addressed below.

4.4 Standardization

The need of an evaluation standard for data computing efficiency and a system which can evaluate the quality of data to enhance big data features is an important issue now and in the future [25]. Although there are several big data solutions, there is no capacity to measure the big data performance by mathematical algorithms. This performance is evaluated by the implemented system and it shows the results, however it is not possible to evaluate and compare them before and after the big data analysis.

4.5 Big Data Computing Modes

Regarding computing modes, transferring data is a challenging aspect as there are several features that a network should offer for this type of applications. There are many features to analyze and transfer data within a network such as channels features, security, reliability, high network availability, and others. All these features should be ensured in order to avoid a bottleneck in this process. Nasser et al. [15] states that, the big data processes involve multiple phases of data acquisition, extraction of information and filtering, data integration, aggregation and visualization, query processing, data modelling, data analysis, and data interpretation and presentation. The disadvantage is that every phase has its own challenges and many difficulties to face [26].

5 Conclusions

In this article it is discussed that, how engineering community is adopting the big data technologies within their processes to stay competitive. In particular this article presents a state of the art survey of the academic literature on the big data applications in the engineering realm. It was well illustrated that, the big data technologies has become an important tool to manage big datasets generated in engineering applications. As it is addressed, the availability of a large number of scientific papers, reports and proceedings confirms the progress of big data analysis in a wide range of industrial applications and a diverse number engineering fields. It is concluded that, the world is adopting big data tendency because it is an efficient option to perform analysis of large the databases and conduct decision support analysis in an effective manner.

Acknowledgement. This work has been sponsored J. Selye University, ITMS 26210120042, and co-funded by the European Regional Development Fund. Authors would like to thank PhD candidate Lourdes Ruiz for her supports during the preparation of the manuscripts.

References

1. Chen, C.: Data-intensive applications, challenges, techniques and technologies: a survey on big data. Inf. Sci. **275**, 314–347 (2014)
2. Mosavi, A., Varkonyi, A.: Learning in Robotics. Learning **157**, 8–11 (2017)
3. Elarabi, T.: Big data analytics concepts and management techniques. IEEE Inventive Comput. Technol. **2**, 183–202 (2016)
4. Gubbi, J.: Internet of Things (IoT): A vision, architectural elements, and future directions. Future Gener. Comput. Syst. **29**, 1645–1660 (2013)
5. Demchenko, Y.: Defining architecture components of the big data ecosystem. Collab. Technol. Syst. IEEE **15**, 345–378 (2014)
6. He, T.: Big data analytics in mobile cellular networks. IEEE Access **4**, 1985–1996 (2016)
7. Imran, A.: Challenges in 5G: how to empower SON with big data for enabling 5G. IEEE Netw. **28**, 27–33 (2014)
8. Jati, G.: Big data compression using spiht in Hadoop. Big Data Inf. **34**, 64–73 (2016)
9. Rahimi-Eichi, H.: Big-data framework for electric vehicle range estimation. Ind. Electron. Soc., IEEE (2014)
10. Mosavi, A.: Optimal Engineering Design. Technical report University of Debrecen (2013)
11. Mosavi, A., Vaezipour, A.: Developing Effective Tools for Predictive Analytics and Informed Decisions. Technical report (2013). https://doi.org/10.13140/RG.2.2.23902.84800
12. Latinovic, T.: Big Data in industry. Mater. Sci. Eng. **144**, 36–59 (2016)
13. Mosavi, A: Predictive decision model (2015). https://doi.org/10.13140/RG.2.2.21094.63047
14. Mosavi. A.: Predictive Decision Making, Predictive Decision Model, Tech. Report. (2015). https://doi.org/10.13140/RG.2.2.21094.63047
15. Nasser, T., Tariq, R.: Big data challenges. J. Comput. Eng. Inf. Technol. **4**, 122–143 (2015)
16. Prasath, R.R., Ozturk, P.: An Approach to Content Extraction from Scientific Articles using Case-Based Reasoning. Res. Comput. Sci. **117**, 85–96 (2016)
17. Jin, J.: An information framework for creating a smart city through internet of things. IEEE Internet Things J. **2**, 112–121 (2014)
18. Lelwala, N.: Ensemble inference based framework for creating knowledge from big data in IoT. Adv. ICT Emerg. Reg. **4**, 84–94 (2016)
19. Mosavi, A., Varkonyi-Koczy, A.R.: Integration of machine learning and optimization for robot learning. Adv. Intell. Syst. Comput. **519**, 349–355 (2017)
20. Mosavi, A., Rabczuk, T.: Learning and intelligent optimization for material design innovation, Theoretical Computer Science and General Issues, LION11 (2017)
21. Kagermann, H.: Recommendations for Implementing the strategic initiative INDUSTRIE 4.0: securing the future of German manufacturing industry; Technical report of the Industrie 4.0 working group. Forschungsunion (2013)
22. Mosavi, A.: Decision-Making Models for Optimal Engineering Design and their Applications. Doctoral Dissertation, University of Debrecen, Hungary (2013)
23. Petit, J., Shladover, E.: Potential cyberattacks on automated vehicles. IEEE Trans. Intell. Transp. Syst. **16**, 546–556 (2015)
24. Mosavi, A.: Multiple criteria decision making integrated with mechanical modeling of draping for material selection of textile composites. Compos. Mater. **12**, 73–81 (2012)
25. Sowmya, R., Suneetha, K.: Data mining with big data. IEEE Intell, Syst. Control **11**, 342–360 (2017)
26. Yin, S.: Big data for modern industry. Challenges Trends IEEE **103**, 143–146 (2015)