Volume 8, No.4, July – August 2019

International Journal of Advanced Trends in Computer Science and Engineering

Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse18842019.pdf

https://doi.org/10.30534/ijatcse/2019/18842019



Classification and Big Data Usages for Industrial Applications

Sang Boem Lim

Department of Smart ICT Convergence and Social Eco-Tech Institute, Konkuk University, Seoul, Korea, sblim@konkuk.ac.kr

ABSTRACT

Recently, Big Data has become a significant concept in various areas, especially with business sectors. Digital devices, social media, and simulation results are producing a vast amount of data which are used by various applications to find insights. However, conceptual work integrating the Big Data applications into categories has been limited. The goal of this paper is to classify Big Data use cases by application's characteristics and to provide a systematic review of Big Data usages in each application classification which is based on examination of the available use cases and survey of related works. Big Data applications are classified into three categories such as industry and business, scientific research, urban planning. The critical aspects of the classified areas are described to determine the importance of Big Data in these areas. Fog computing and cloudlet are presented as the core technology of collecting data. Twitter, LinkedIn, Facebook, and Netflix are discussed as industry and business cases of Big Data.

Key words: Big Data application, IoT, Smart city, Healthcare Big Data

1. INTRODUCTION

In recent years, importance of Big Data is increasing in almost every aspect of modern society. This applies not only in research areas such as engineering, natural science, and life science, but also in business and government sectors [1]. Previously, the importance and challenges of Big Data already recognized in Nature and Science [2] [3]. For the industrial and business sector, McKinsey presents how Big Data is transforming traditional business aspects to the new paradigm [4].

Unfortunately, common definition of Big Data there is not avaliable. The first appearance of the term "Big Data" was presented in a 1997 paper by scientists at NASA [5]. In this paper, authors defined Big Data as "provides an interesting challenge for computer systems: data sets are generally quite large, taxing the capacities of main memory, local disk, and even remote disk. We call this the problem of big data." Since then, various definitions are available. Currently, the most widely used definition described by McKinsey in 2011 [6] as "datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze."

The next generation of IT industry and research community strongly impacted by Big Data. According to International Data Corporation (IDC) [7], global revenues for Big Data and business analytics (BDA) solutions in 2019 will increase 12.0%, which reach \$189.1 billion, over 2018. This pace of growth will be maintained throughout the year 2022 and worldwide BDA revenue will reach \$274.3 billion. Throughout all the BDA revenue forecast, the fastest growth area is IT and business services with accomplishing more than half of revenues. The increasing importance of Big Data services motivates this literature survey.

This paper presents a literature review of Big Data applications. The purpose is to present a classification of research works, findings, and practical implications. This is necessary to provide an update about the state of use cases, identify active research areas in both academia and industry community, and understand existing challenges for the benefit of the future development of Big Data applications. For this research, the literature was selected from three major databases, Scopus, ScienceDirect, and EBSCO. Most of the literature focusing on Big Data application ranges from 2017-2019.

Section 2 explains significant of the Big Data in various research and industry communities. Section 3 discusses the IoT technology, which is the core digital devices of Big Data. Section 4 presents the business aspects of Big Data by evaluating the companies that are using Big Data heavily in their businesses. Finally, section 5 draws concluding remarks.

2. CATEGORIES OF BIG DATA APPLICATIONS

In this section, Big Data use cases are categorized in various perspectives, and we also describe the significance of Big Data in each category.

2.1 Industry and Business

Many companies are using Big Data as a common method. Collection of business performance information creates significant value to the business strategies such as price optimization, product design, equipment replaces, and labor inputs [6]. For example, manufacturers are able to predict the optimal time to maintain or replace their equipment to save expenses by collecting and analyzing vibration data. Financial service organizations can prevent financial fraud by Big Data analysis of customer behaviors and interactions.

Especially, the social network companies, such as Facebook, Instagram, Twitter, and LinkedIn, and video-streaming company like Netflix are generating and using a great deal of Big Data in their businesses. Usage of social big data is discussed in section IV in-depth.

2.2 Scientific Research

For the past decade, the importance of Big Data is recognized in various scientific research areas. The perspective of scientific research has changed because of Big Data. Scientists mainly focus on analyzing data to find knowledge and intelligence rather than scientific equipment operation [8].

For example, the meteorologists are dealing with hundreds of terabyte data, which are generated by the earth observation satellites, radars, and sensor networks, to simulate climate changes [9]. The main focus of climate simulations is to deliver advanced warnings in term of severe weather changes by interpreting the huge amount of data.

Biomedical and healthcare area is one of the most active Big Data research areas. The main aim of these areas is to provide better services, medical therapies, personalized medicine, and financial advantages [10].

2.3 Urban Planning

Smart city is one of the characteristic examples of Big Data to the people. The goal of smart city is to provide better and safer services throughout integration of various urban subsystems such as business, transportation, and communication technology [11]. Big Data plays an important role in these services.

For example, urban planning for unused public properties can get benefit from Big Data. Big Data analysis based on demographics, facility usage, and social welfare indicators can fulfill the overall needs of society and urban development [12]. People also can reduce fuel consumptions and emissions by using dynamic Big Data traffic analysis [13].

3. THE INTERNET OF THINGS (IOT)

The Internet of Things (IoT) changes our daily life dramatically [14] [15]. Internet of Things Global Standards Initiative [16] defines the IoT as "a global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies." The physical environment provides information that can be used to improve the quality of life. The objective of this section is to deliver up to date research subjects of Big Data with the IoT. In the following subsections, IoT issues focusing on big data aspect are discussed. Categories of Big Data complexity handling are described in the first subsection. Fog computing and Cloudlets, that are new concepts in the computing landscape, are analyzed as IoT Big Data applications.

3.1 Complexity Handling

According to [17], the complexity of data that is produced by the IoT devices can be handled by local processing, cloud computing, and Edge computing. The local processing approach uses smart sensors to analysis the raw data that are collected. Smart sensors have the capabilities of computing and communication [18, 19]. In this approach, smart sensors play a significant role to reduce frequency of communication and workload of remote servers.

The cloud computing with IoT generates IoT-as-a-Service (IoTaaS) [20]. In this model, Big Data is used to analyze the data in the cloud computing and valuable knowledge is extracted using data mining methods to deliver rich services. Most suitable example of IoTaaS is the smart city. This application uses a centralized cloud facility to analyze data and provide services to citizens [21]-[23]. The cloud computing paradigm can easily expand services with newly developed technologies like 5G protocols [24].



Figure 1: Edge computing applications [26].

Edge computing (Figure 1) provides fast processing and quick response time to accomplish requirements of mobility-related issues [25] [26]. Many functionalities of the Cloud computing like data computation and storage services are relocated at the edge devices to reduce latency for delay-sensitive user services. Characteristics of Edge computing can be categorized as dense geographical distribution, mobility support, location awareness, proximity, low latency, and context-awareness [27]. There are several ways to implement Edge computing. This paper analyzes the literature available on Fog computing and Cloudlets.

3.2 Big Data on Fog computing

In this subsection, we provide a comparison of Fog computing solutions focused on Big Data aspect. The concept of the Fog computing, that has been introduced by Cisco, enables computational capability at the network edge [28] [29].

J. Santos *et al.* [30] proposed a Fog computing-based data analytics framework in 5G networks to overcome the challenges involved in fog node management for the exchange of provisioning information between fog nodes.

B. Qin *et al.* [31] introduced applications of the fog computing model based on big data. The industrial robot application is mainly discussed. This paper describes how to use the characteristics of big data and fog computing model in the robotics area. The 5G network application is also discussed as a future application area. A.

Yassine *et al.* [32] sought to apply big data analytics for smart homes with IoT and fog computing. The challenges of complexities and resource demands and the requirements and the design issues of big data analytics are discussed and proposed a relevant system architecture. In this study, the authors perform a case study with a dataset that acquired from real smart home.

3.3 Big Data on Cloudlet

Cloudlet was introduced by Satyanarayanan *et al.* [33] to overcome performance issues such as latency and packet losses of Cloud services by adding a small-scale data center and computation capabilities to the IoT devices.

U. Shaukat *et al.* [34] described the motivation, architectures, applications, and open challenges issues with cloudlet frameworks. This paper has presented a hierarchical taxonomy for classification of cloudlet solution, identification of cloudlet application, strengths and weaknesses analyzation, and requirements for deploying cloudlet.

S. Valluripally *et al.* [35] sought to introduce a community cloud architecture addressed implementation of health big data application using cloudlet. This architecture uses cloudlet to provide application performance in high-speed processing.

C. Shen *et al.* [36] proposed an energy-efficient cloudlet placement method to reduce the energy consumptions by rearranging the clustering of mobile devices. For a dynamic placement method of energy-efficient cloudlet, the authors have proposed three-step processes such as device center location recognition, cloudlet location determination, and dynamic placement of cloudlet.

4. SOCIAL BIG DATA

Social media companies play a main role as both major sources of contributing data and heavy user of the data (Figure 2). To understand various research topics and technical challenges in area of social big data is important.

There are no common definition and the related terms of social big data. The most relevant definition to this paper is presented by E. Olshannikova et al. [38] which is "any high-volume, high-velocity, high-variety and/or highly semantic data that is generated from technology-mediated social interactions and actions in digital realm, and which can be collected and analyzed to model social interactions and behavior." In the following subsections, current researches and use cases of social big data are presented.



Figure 2: Growth of Social Media Users [37].

4.1 Social Network Providers

Social network providers, such as Facebook, Twitter, and LinkedIn, are examples of Big Data use cases. Facebook perform batch-based analysis based on structured and stream-based user data to provide quality of service [39] [40]. Insight of the data is extracted by performing ad hoc analysis tasks in production or development environments.

Characteristics of Twitter has both a social network and an informational network and generates more than 143K tweets per second [41]. The nature of Twitter requires real-time data processing. Insight information is used to provide user-oriented services like Who to follow [39].

LinkedIn is a business networking service for industry professionals. LinkedIn provides an experience level option to select a category of professional level [42]. Structured and stream-based data are collected for analysis. Data analysis results are used to provide user-centric services like People you may know [39].

Netflix is a company to provide online or stream video service to home appliances and mobile devices. User event data are used to provide video recommendations for end-users, and network response time data of client applications are used to determine network condition [39]. Event data are analyzed in real-time using online environments. Netflix uses Amazon cloud to provide services.

4.2 Social Big Data Use Cases

Social network data can provide benefits to people in many ways. E. Olshannikova et al. [38] presented a definition of social big data and conceptualized it. The authors present four main related science fields of social big data such as social computing, Big Data science, data analytics, and computational social science.



J. C. Kurian et al. [43] studied themes of Facebook post on the federal emergency management agency Facebook page to identify the core principles of disaster recovery. By analyzing Facebook post, the authors claimed to maximize disaster recovery by achieving pre-disaster recovery planning, public information messaging, and psychological recovery.

Y. Liang et al. [44] presented digital phenotyping of mental health using social media data along with data from ubiquitous sensors and the healthcare system. The authors address the challenges, open issues, and solutions of mental health by Big Data analysis.

P. Del Vecchio et al. [45] sought to open up innovation processes to support tourism experiences. Innovation processes are evaluated based on a huge amount of travel experience social media data which are generated by tourists.

4.3 Reference Architecture of Social Big Data

P. Pääkkönen et al. [39] evaluated various social network

companies' service architectures, such as Facebook, Twitter, LinkedIn, Netflix, in order to provide a reference architecture (Figure 4) of social big data. The authors also provide a description of how to reference architecture is used in real use cases.

This reference architecture is divided into nine parts: data source, extraction, loading and pre-processing, processing, analysis, loading and transformation, data storage, job and model specification, and visualization. Some key features are described as follows. In the data sources part, the authors are defined a source of data as stored data (in situ), data flow (streaming), a strict data model (structured), and raw data (unstructured). The role of the extract is to collect data from the data sources and to store temporarily for pre-processing. Batch-processing is performed using pre-processed data in the deep analytics part. The data analysis results are sent to the visualization part for user presentation.



Figure 4: High-level design of the reference architecture of social Big Data [39]

5. CONCLUSION

Recently, Big Data became one of the most important technologies in industry and research area. This paper focused on Big Data applications by the literature reviews. First, a classification of Big Data is presented to identify relevant applications. IoT is described as a core technology since IoT sensors are one of the most data-producing sources. The other data source of Big Data is a social network. Social network companies use cases of social network data are presented. A social network reference architecture is also discussed. Sang Boem Lim, International Journal of Advanced Trends in Computer Science and Engineering, 8(4), July- August 2019, 1117 - 1122

REFERENCES

- C. Yang, Q. Huang, Z. Li, K. Liu, and F. Hu, Big Data and cloud computing: innovation opportunities and challenges, *International Journal of Digital Earth*, Vol. 10, no. 1, pp. 13-53, Jan. 2017. https://doi.org/10.1080/17538947.2016.1239771
- V. Marx, The big challenges of big data, *Nature*, Vol. 498, no. 7453, pp. 255-260, 2013. https://doi.org/10.1038/498255a
- Dealing with data, Science, Vol. 331, no. 6018, pp. 639-806, Jun. 2011. https://doi.org/10.1126/science.1197872
- 4. G. M. Asaftei, S. Doshi, J. Means, and A. Sanghvi, Getting ahead of the market: How big data is transforming real estate, *McKinsey Global Institute*, Oct. 2018.
- 5. M. Cox and D. Ellsworth, Application-controlled Demand Paging for Out-of-core Visualization, in Proceedings of the 8th Conference on Visualization '97, IEEE Computer Society Press, pp. 235-244, 1997.
- J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, Big data: The next frontier for innovation, competition, and productivity, *McKinsey Global Institute*, May 2011.
- 7. International Data Corporation (IDC), **IDC Forecasts Revenues for Big Data and Business Analytics Solutions Will Reach \$189.1 Billion This Year with Double-Digit Annual Growth Through 2022**, *IDC*, Apr. 2019.
- X. Jin, B. W. Wah, X. Cheng, and Y. Wang, Significance and Challenges of Big Data Research, *Big Data Research*, Vol. 2, no. 2, pp. 59-64, Jun. 2015. https://doi.org/10.1016/j.bdr.2015.01.006
- O. Y. Al-Jarrah, P. D. Yoo, S. Muhaidat, G. K. Karagiannidis, and K. Taha, Efficient Machine Learning for Big Data: A Review, *Big Data Research*, Vol. 2, no. 3, pp. 87-93, Sep. 2015. https://doi.org/10.1016/j.bdr.2015.04.001
- S. Dash, S. K. Shakyawar, M. Sharma, and S. Kaushik, Big data in healthcare: management, analysis and future prospects, *Journal of Big Data*, Vol. 6, no. 1, Dec. 2019.

https://doi.org/10.1186/s40537-019-0217-0

- 11. G. Fahmfam, and H. Hamidi, Factors affecting the development and management of smart city approach using a combination of big data and the internet of things and cloud computing technologies, *Iranian Journal of Information Processing Management*, Vol. 34, no. 2, pp. 557-584, Dec. 2019.
- J. W. Huang and W. M. Wey, Application of Big Data and Analytic Network Process for the Adaptive Reuse Strategies of School Land, Social Indicators Research, Vol. 142, no. 3, pp. 1075-1102, Apr. 2019. https://doi.org/10.1007/s11205-018-1951-y
- 13. D. Guo, J. Wang, J. B. Zhao, F. Sun, S. Gao, C. D. Li, M. H. Li, and C. C. Li, A vehicle path planning method

based on a dynamic traffic network that considers fuel consumption and emissions, *Science of the Total Environment*, Vol. 663, pp. 935-943, May 2019. https://doi.org/10.1016/j.scitotenv.2019.01.222

14. M. Pannu and I. Kay, Internet of things: Analyzing the impact on businesses and customers, Advances in Intelligent Systems and Computing, Vol. 953, pp. 322-327, 2020.

https://doi.org/10.1007/978-3-030-20473-0_31

- 15. J. Sasi Bhanu, J. K. R. Sastry, P. Venkata Sunil Kumar, B. Venkata Sai, and K. V. Sowmya, Enhancing performance of IoT networks through high performance computing, International Journal of Advanced Trends in Computer Science and Engineering, Vol. 8, no. 3, pp. 432-442, May-June 2019. https://doi.org/10.30534/ijatcse/2019/17832019
- 16. Internet of Things Global Standards Initiative, https://www.itu.int/en/ITU-T/gsi/iot/Pages/ default.aspx
- D. Gil, M. Johnsson, H. Mora, and J. SzymaNski, Review of the Complexity of Managing Big Data of the Internet of Things, *Complexity*, Vol. 2019, Article ID 4592902, 12 pages, 2019. https://doi.org/10.1155/2019/4592902
- 18. G. Meijer, *Smart sensor systems*, John Wiley & Sons, 2008.

https://doi.org/10.1002/9780470866931

- P. Pratim Ray, D. Dash, and D. De, Edge computing for Internet of Things: A survey, e-healthcare case study and future direction, *Journal of Network and Computer Applications*, Vol. 140, pp. 1-22, Aug. 2019. https://doi.org/10.1016/j.jnca.2019.05.005
- 20. J. L. Perez, D. Carrera, **Performance characterization** of the servioticy API: An IoT-as-a-service data management platform, in *Proceedings of the 2015 IEEE 1st International Conference on Big Data Computing Service and Applications*, pp. 62-71, 2015. https://doi.org/10.1109/BigDataService.2015.58
- 21. R. Casadei, G. Fortino, D. Pianini, W. Russo, C. Savaglio, and M. Viroli, A development approach for collective opportunistic Edge-of-Things services, *Information Sciences*, Vol. 498, pp. 154-169, 2019. https://doi.org/10.1016/j.ins.2019.05.058
- A. M. Osman, A novel big data analytics framework for smart cities, *Future Generation Computer Systems*, Vol. 91, pp. 620-633, 2019. https://doi.org/10.1016/j.future.2018.06.046
- N. Shoaib and J. A. Shamsi, Understanding Network Requirements for Smart City Applications: Challenges and Solutions, *IT Professional*, Vol. 21, no. 3, pp. 33-40, 2019.

https://doi.org/10.1109/MITP.2018.2883047

24. M. Afrin, M. Razzaque, I. Anjum, M. Hassan, and A. Alamri, Trade of between User Quality-Of-Experience and Service Provider Profit in 5G Cloud Radio Access Network, *Sustainability*, Vol. 9, no. 11, p. 2127, 2017. https://doi.org/10.3390/su9112127

25. I. Sittón-Candanedo, R. S. Alonso, S. Rodríguez-González, J. A. García Coria, and F. De La Prieta, Edge Computing Architectures in Industry 4.0: A General Survey and Comparison, Advances in Intelligent Systems and Computing, Vol. 950, pp. 121-131, 2020.

https://doi.org/10.1007/978-3-030-20055-8_12

26. S. Sthapit, J. Thompson, N. M. Robertson, and J. R. Hopgood, Computational load balancing on the edge in absence of cloud and fog, *IEEE Transactions on Mobile Computing*, Vol. 18, no. 7, pp. 1499-1512, Jul. 2019.

https://doi.org/10.1109/TMC.2018.2863301

- 27. W. Z. Khan, E. Ahmed, S. Hakak, I. Yaqoob, and A. Ahmed, Edge computing: A survey, *Future Generation Computer Systems*, Vol. 97, pp. 219-235, Aug. 2019. https://doi.org/10.1016/j.future.2019.02.050
- 28. Cisco fog computing solutions: Unleash the power of the Internet of Things. Available at: https://www.cisco.com/c/dam/en_us/solutions/trends/iot /docs/computing-solutions.pdf.
- S. Karimunnisa and V. S. Kompalli, Cloud computing: Review on recent research progress and issues, International Journal of Advanced Trends in Computer Science and Engineering, Vol. 8, no. 2, pp. 216-223, Mar.-Apr. 2019.

https://doi.org/10.30534/ijatcse/2019/18822019

- 30. J. Santos, B. Volckaert, T. Wauters, and F. de Turck, Fog Computing: Enabling the Management and Orchestration of Smart City Applications in 5G Networks, *Entropy*, Vol. 20, no. 1, Dec. 2017 https://doi.org/10.3390/e20010004
- 31. B. Qin, X. Lin, S. Li, Q. Luo, F. Zheng, J. Cai, and Y. Luo, Design and application of fog computing model based on big data, in 2019 IEEE 2nd International Conference on Information and Computer Technologies, ICICT 2019, pp. 93-97, 2019. https://doi.org/10.1109/INFOCT.2019.8711436
- 32. A. Yassine, S. Singh, M. S. Hossain, and G. Muhammad, IoT big data analytics for smart homes with fog and cloud computing, *Future Generation Computer Systems*, Vol. 91, pp. 563-573, Feb. 2019. https://doi.org/10.1016/j.future.2018.08.040

 33. M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies, The case for VMbased cloudlets in mobile computing, *IEEE Pervasive Computing*, Vol. 8, no. 4, pp. 14-23, Oct.

2009. https://doi.org/10.1109/MPRV.2009.82

34. U. Shaukat, E. Ahmed, Z. Anwar, and F. Xia, Cloudlet deployment in local wireless networks: Motivation, architectures, applications, and open challenges, *Journal of Network and Computer Applications*, Vol. 62, pp. 18-40, Feb. 2016.

https://doi.org/10.1016/j.jnca.2015.11.009

35. S. Valluripally, M. Raju, P. Calyam, M. Chisholm, S. S. Sivarathri, A. Mosa, and T. Joshi, **Community cloud** architecture to improve use accessibility with security

compliance in health big data applications, in *20th International Conference on Distributed Computing and Networking, ICDCN 2019*, pp. 377-380, 2019. https://doi.org/10.1145/3288599.3295594

- 36. C. Shen, S. Xue, and S. Fu. ECPM: An Energy-Efficient Cloudlet Placement Method in Mobile Cloud Environment, Eurasip Journal on Wireless Communications and Networking, Vol. 2019, no. 1, Dec. 2019. https://doi.org/10.1186/s13638-019-1455-8
- 37. T. Veeramani, P. Srinuvasarao, B. Rama Krishna, and R. Thilagavathy, Impact of social media networks big data analysis for high-level business, International Journal of Recent Technology and Engineering, vol. 7, no. 5, pp. 87-92.
- 38. E. Olshannikova, T. Olsson, J. Huhtamäki, and H. Kärkkäinen, Conceptualizing Big Social Data, *Journal of Big Data*, Vol. 4, no 3, Jan. 2017. https://doi.org/10.1186/s40537-017-0063-x
- 39. P. Pääkkönen, and D. Pakkala, Reference Architecture and Classification of Technologies, Products and Services for Big Data Systems, *Big Data Research*, Vol. 2, no. 4, pp. 166-186. Dec. 2015. https://doi.org/10.1016/j.bdr.2015.01.001
- 40. S. Ravi Kumar, V. K. and Anurag, A survey on prediction approaches for epidemic disease outbreaks based on social media data, *International Journal of Advanced Trends in Computer Science and Engineering*, Vol. 8, no. 3, pp. 897-908, May-June 2019. https://doi.org/10.30534/ijatcse/2019/86832019
- 41. A. M. Sadri, S. Hasan, and S. V. Ukkusuri, Joint inference of user community and interest patterns in social interaction networks, *Social Network Analysis* and Mining, Vol. 9, no. 1, Dec. 2019. https://doi.org/10.1007/s13278-019-0551-4
- 42. H. Chang, C. Wang, and S. Hawamdeh, Emerging trends in data analytics and knowledge management job market: extending KSA framework, *Journal of Knowledge Management*, Vol. 23, no. 4, pp. 664-686, May 2019.

https://doi.org/10.1108/JKM-02-2018-0088

- 43. J. C. Kurian, B. M. John, User-generated content on the Facebook page of an emergency management agency: A thematic analysis. *Online Info Rev* Vol. 41, no. 4, pp. 558-579, 2017. https://doi.org/10.1108/OIR-09-2015-0295
- 44. Y. Liang, X. Zheng, and D. D. Zeng, A survey on big data-driven digital phenotyping of mental health, Information Fusion, vol. 52, pp. 290-307, Dec. 2019. https://doi.org/10.1016/j.inffus.2019.04.001
- 45. P. Del Vecchio, G. Mele, V. Ndou, and G. Secundo, Open innovation and social big data for sustainability: Evidence from the tourism industry, *Sustainability (Switzerland)*, Vol. 10, no. 9, Sep. 2018. https://doi.org/10.3390/su10093215