

A review of big data analytics and healthcare

Seok-Jae Moon · Namju Lee[†]

*Department of Information Security Engineering, Institute of Information Technology,
Kwangwoon University, Seoul, 01897, Republic of Korea*

[†]*Department of Physical Education, Institute of Information Technology,
Kwangwoon University, Seoul, 01897, Republic of Korea*

(Received January 29, 2019; Revised February 22, 2019; Accepted February 25, 2020)

빅데이터 분석과 헬스케어에 대한 동향

문석재^{1,*} · 이남주^{2,†}

¹광운대학교 정보과학교육원 정보보호학과, 교수

²광운대학교 정보과학교육원 체육학과, 교수

(2020년 1월 29일 접수: 2019년 2월 22일 수정: 2020년 2월 25일 채택)

Abstract : Big data analysis in healthcare research seems to be a necessary strategy for the convergence of sports science and technology in the era of the Fourth Industrial Revolution. The purpose of this study is to provide the basic review to secure the diversity of big data and healthcare convergence by discussing the concept, analysis method, and application examples of big data and by exploring the application. Text mining, data mining, opinion mining, process mining, cluster analysis, and social network analysis is currently used. Identifying high-risk factor for a certain condition, determining specific health determinants for diseases, monitoring bio signals, predicting diseases, providing training and treatments, and analyzing healthcare measurements would be possible via big data analysis. As a further work, the big data characteristics provide very appropriate basis to use promising software platforms for development of applications that can handle big data in healthcare and even more in sports science.

Keywords : *Big data, Healthcare, Convergence, Analysis, Application*

[†]Corresponding author
(E-mail: namju1210@gmail.com)

1. Introduction

The Fourth Industrial Revolution has begun with the revolutionary change of high-tech information and communication technology in Korea and all over the world, and the ways of accumulating, accepting and analyzing new knowledge have been diversified based on the Fourth Industrial Revolution. Big data is being studied in a variety of fields in terms of usability and value, and is considered to be an applicability to new academic convergence and a transition to a new paradigm using big data [1,2].

At the 2016 World Economic Forum, the Fourth Industrial Revolution was to bring about a major change in the society we live in [3]. Today's Fourth Industrial Revolution is a revolution in the convergence of technologies where the boundaries of physical, digital and biological space are diluted based on the digital revolution (the Third Industrial Revolution) such as internet Technologies (IT) [3]. The basic framework of the Fourth Industrial Revolution is the creation of big data based on the internet of things (IoT) for hyper-connected, super-intelligence, and convergence. The artificial intelligence interprets the big data created in this way, leading to the automation and intelligence of the process [4,5]. In response to these trends, Korea has also announced the "Medium and Long-term Comprehensive Measures for the Intelligent Information Society," which includes technology, industry, and social policy to prepare for the fourth industrial revolution based on intelligent information technology [6]. Therefore, research has been conducted in many academic fields, and the field of physical education and sports science also requires much research in preparation for the Fourth Industrial Revolution, but so far, there are not many studies related to big data in relation to the field of physical education and sports science in Korea [6].

Based upon previous studies, the Korean

government is actively expanding the policy support for expanding investment in sports research and development (R & D), but the national policy related to the convergence of sports science and technology has many problems in terms of industrialization and commercialization compared to other countries in the world [7,8]. In addition, convergence researches that have adopted IT technology in sports science and industry tend to focus on technological value development without going through necessary processes and procedures [8] and thus, it seems that big data analysis suitable for the field of physical education and sports science would be needed.

According to Kim's research [9], the objectives of applying the technologies of the Fourth Industrial Revolution to sports-related research and industry were to improve the performance of elite athletes, to improve the health of the people, to integrate sports spaces through the media, and to informatize sports. As such, big data analysis in sports and sports science research seems to be a necessary strategy for the convergence of sports science and technology in the era of the Fourth Industrial Revolution. In particular, big data may have a significant difference in the analysis results depending on the nature of the data. Therefore, in order to understand big data analysis technologies and integrate them into sports science, understanding of big data must be preceded [10].

At the present time, big data is regarded as a new paradigm of the era, not only research combining sports, sports science, society and culture, but also research areas related to literacy and sports industry for elite athletes' performance enhancement. Therefore, in-depth discussion of new research methods combined with big data and sports science would be needed. The purpose of this study was to provide the basic data to secure the diversity of big data and exercise and sports science convergence by discussing the concept, analysis method and application examples of big data

and exploring the application of exercise and sports science. This review focused on the theoretical concept of big data and convergence of big data analysis and exercise science.

2. Theoretical concept of big data

Big data is defined by many scholars. In general, big data refers to huge amounts of structured or unstructured compounded data sets that go beyond the ability to collect, store, manage, and analyze data with existing database management tools [11, 12].

As shown in Table 1, data attributes are characterized by 3 things: Structured data stored in fixed fields, semi-structured data not stored in fixed fields but containing metadata or schemas, unstructured data such as images, videos, and voices, not stored in fixed fields [2,10]. In particular, as the unstructured data has increased due to the recent spread of digital information, the analysis method has become a major concern. Therefore, there is a need to continuously discuss the effective management and analysis of big data.

Big data can be described in terms of 6 V's as follows: The first V is volume in reference to the size of the data and the data is expected to raise intensely in the years ahead, usually measured in terabytes, petabytes even yottabytes. The second V is variety in reference to the different types and/or formats of the data, the third V is velocity in reference to the speed of data generation, and the fourth V is veracity in reference to the

reliability of the data. While variability regards about consistency of the data over time. The last V, value of the big data, refers to their coherent analysis [13,14]. The data generated during manufacturing is increasing daily with different types and formats such as process logs, events, images, and sensor data. Therefore, the processing and storage of these data is becoming a challenging issue in the big data processing.

Critical issues in big data processing is shown in figure 1 [15,16]. Data growth, data infrastructure, data governance/policy, data integration, data velocity, data variety, data compliance/regulation, and data visualization would be carefully considered in handling big data. Firstly, big data is based on analysis based on large amounts of data, unlike conventional data processing, which enables instant processing speed. Therefore, fast decision making requires relatively less time because it requires a long-term and strategic approach. Second, since big data processes various data sources, complex logic processing, and large data processing, processing complexity is high, and for this reason, distributed processing technology is generally required. Third, big data has a high proportion of unstructured data, which increases processing complexity, because most of social media data, log files, clickstream data, call center logs, and communication logs are unstructured data. Fourth, big data is highly flexible in processing and analysis because it is essential for accommodating new and diverse processing methods. Lastly, because of the large capacity and complex processing,

Table 1. Properties of data

Attributes	Definition and Contents
Structured data	Data stored in fixed fields, such as relational databases and spreadsheets
Semi-structured data	Although it is not stored in a fixed field, it refers to data including metadata or schemas, such as XML and HTML text
Unstructured data	Data not stored in fixed fields and text analysis is possible, such as text documents, images, videos, audio data, etc.

the amount of data that needs to be processed at the same time is lower than that of conventional data processing [10,14,16].

In addition, the big data effect is that the larger the scale, the more likely the data is to be useful, and the more likely it is to find a whole new pattern of information. The big data is that as the cognitive information distribution such as personal experiences, perceptions, and preferences increases, the reality information such as daily data records increases, and the prospect of real-time information accumulating rapidly. The trend effect of big data is that the historical data is maintained and the data with time series continuity is composed so that the analysis of time flows such as the past, present, and the future is possible, and the combination effect is heterogeneous data. The new meaning of information is discovered through the combination of the two data, and further advances in the field of safety verification are possible through pre-simulation through data without the actual physical coupling [10,14,17].

3. Big data handling mechanisms in the healthcare applications

Previous exercise and sports science related to big data studies used several types of big data analysis based upon their data attributes. Based upon Pashazadeh and Navimipour's study [18], reviewed mechanisms in the big data related to healthcare applications in five categories, including machine learning, cloud-based, heuristic-based, agent-based, and hybrid mechanisms. The machine learning, which has been used recently, is a practical and effective approach for big data handling based on both artificial intelligence techniques and historical databases. An agent-based method is to provide the suitable software support for coordinating healthcare management and is based on the medical care management domain, associated with intelligent agents. The efficiency of healthcare systems is improved through the integration of distributed monitoring with centralized analytics and the long-term care of the population can be enhanced by an agent-based method in

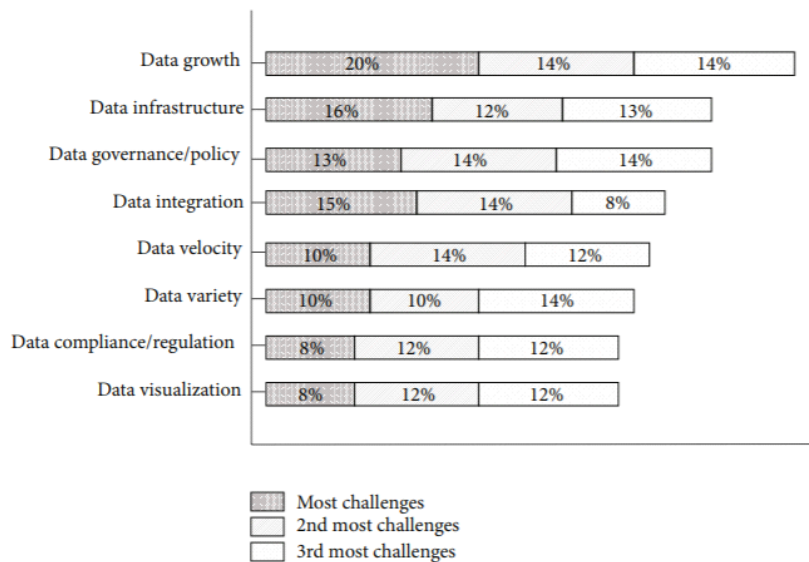


Fig. 1. Challenges in big data [15]

handling data. The heuristic algorithm is commonly used to solve the optimization problems by choosing one objective and is cost-effective, easy to scale, and pluggable. In cloud computing, there are many information and computing resources such as operating systems, storage, networks, hardware, data-bases, and software applications and is appropriate health monitoring system. A hybrid mechanism is that two or more distinct types of methods are used by combining them to solve the big data handling problems [18]. Healthcare data including exercise and sports science data can be applied above 5 types big data handling methods based upon raw data attributes.

Based on the previous studies [2,18–22], text mining, data mining, opinion mining, process mining, cluster analysis, and social network analysis is currently used for analyzing exercise and sports science big data. Text mining is to use natural language processing techniques to extract useful information from text, or find information in the meaning hidden in big data such as linkage, classification, clustering, or summarization [2,19,21]. This is an analysis method necessary to grasp research trends in exercise and sports science by analyzing words related to a specific subject and this method can be used for analyzing awareness of sports injuries. Data mining helps to make decisions by discovering useful correlations hidden among many data and extracting actionable information in the future and this method can be used for analyzing motion measurement data to produce results that help prevent injuries [2,18]. Opinion mining means analyzing people's opinions, evaluations, attitudes, and feelings about an issue, person, and/or event [2,18]. This can be used for analyzing the opinions of athletes to help develop sports goods to enhance performance and prevent injuries. Process mining is a method of extracting and analyzing event log data collected from various information systems to provide useful information [2,18],

which can be used for simulation analysis to derive ways to improve the training process of athletes. Cluster analysis is used to analyze the differences between objects of similar groups and objects in other cluster groups before analyzing the structure, connection strength, and distance of social network analysis based on structured text through text mining [2,18,22]. This can be used for analyzing trends of sports medicine researches and interests of sports science. Social network analysis analyzes network connection structure and strength to identify the type of message, its propagation path, and quantify who will be able to get affected [2,20], which can be used for identifying research trends and their relationship to exercise and sports science field.

Big data analytics in exercise and sport science is very promising process of integrating, exploring and analysing of large amount complicated data with different nature including biomedical data, experimental data, electronic health records data, social media data, and so on [22]. Identifying high-risk factor groups, determining specific health determinants for diseases, monitoring bio signals, predicting diseases, providing training and treatments, and analyzing healthcare measurements would be possible via big data analysis [22,24]. As a further work, the big data characteristics provide very appropriate basis to use promising software platforms for development of applications that can handle big data in exercise and sport science and even more healthcare. In addition, personal data security is critical in all data administration process and thus, new strategic method would be needed to guarantee and protect the big data privacy and security.

4. Conclusion

Big data analysis in sports and sports science research seems to be a necessary strategy for the convergence of sports science and

technology in the era of the Fourth Industrial Revolution. At the present time, big data is regarded as a new paradigm of the era, not only research combining sports, sports science, society and culture, but also research areas related to literacy and sports industry for performance enhancement. Text mining, data mining, opinion mining, process mining, cluster analysis, and social network analysis is currently used for analyzing exercise and sports science big data. Identifying high-risk factor for a certain condition, determining specific health determinants for diseases, monitoring bio signals, predicting diseases, providing training and treatments, and analyzing healthcare measurements would be possible via big data analysis. As a further work, the big data characteristics provide very appropriate basis to use promising software platforms for development of applications that can handle big data in healthcare and even more exercise and sports science.

Acknowledgment

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit-sectors and authors confirm that no conflict of interest exists.

References

1. D. K. Kim, H. S. Lee. "Study on the analysis of national paralympics by utilizing social big data text mining", *Korean J Phys Edu*, Vol.55, No.6 pp. 801-810, (2016).
2. H. K. Roh, D. C. Lee. "Exploring the applicability of big data analysis in adapted physical activity research", *Korean J Adap Phys Act*, Vol. 25, No. 4 pp. 1-15, (2017).
3. World Economic Forum, "New vision for education: Fostering social and emotional learning through technology", Colony/ Geneva: World Economic Forum (2016).
4. H. Chung, Y. Choi, S. Lee. "The fourth industrial revolution and change of health industry paradigm", *Korea Heal Indus Prom Inst*, 215, pp. 1-28, (2016).
5. S. G. Park, Y. C. Hwang. "The exploratory study for the application of the sports field in the fourth industrial revolution: Focus on the social big data", *Korean J Phys Edu*, Vol. 56, No. 4 pp. 397-413, (2017).
6. C. Kim. "The change and role of sport in the age of the fourth industrial revolution", *Korean J Phys Edu*, Vol. 57, No. 3 pp. 17-27, (2018).
7. J. D. Oh, J. D. Park, C. H. Park, B. H. Park, Y. S. Choi, H. K. Seong. "A study on the strategic for sports industry technology and IT technology convergence", *2015 Proceedings of Symposium of the Korean Institute of communications and Information Sciences*, pp. 293-294, (2015).
8. M. K. Kim. "Issues and tasks of sports science technology convergence policy on the fourth industrial revolution: Focusing on future tasks and response strategies", *Korean J Phys Edu*, Vol. 58, No. 4 pp. 349-362, (2019).
9. T. D. Kim. "Sports ICT industry development plan", *Korea Enter Indust Assoc*, Vol. 7, No. 1 pp. 21-34, (2015).
10. S. Y. Lee, H. J. Yoon. "The study on strategy of national information for electronic government of S. Korea with public data analysed by the application of scenario planning", *J Inst Elec Info Sci*, Vol. 7, No. 6 pp. 1259-1273, (2012).
11. S. Lavalle, E. Lesser, R. Shockley, M. Hopkins, N. Kruschwitz. "Big data, analytics and path form insights to value", *MITSI Manage Rev*, Vol. 52, No. 2 pp. 21-31, (2011).

12. R. Nambiar, A. Sethi, R. Bhardwaj, R. Vargheese. "A look at challenges and opportunities of big data analytics in healthcare", *2013 IEEE International Conference on Big Data*, pp. 17–22, (2013).
13. P. Zikopoulos, C. Eaton. "Understanding big data: Analytics for enterprise class Hadoop and streaming data", *McGraw-Hill Osborne Media*: Berkeley, CA, USA, 2011.
14. B. Ristevski, M. Chen. "Big data analytics in medicine and healthcare", *J Integ Bioinfo*, Vol.15, No.3, 20170030, (2018).
15. Intel, "Big Data Analytics," 2012, <http://www.intel.com/content/dam/www/public/us/en/documents/reports/data-insights-peer-research-report.pdf>.
16. S. Y. Lee, H. J. Yoon. "The study on strategy of national information for electronic government of S. Korea with public data analysed by the application of scenario planning", *J Korea Inst Elec Comm Sci*, Vol.7, No.6, pp. 1259–1273, (2012).
17. N. Khan, I. Yaqoob, I. A. T. Hashem, Z. Inayat, W. K. M. Ali, M. Alam, M. Shiraz, A. Gani. "Big data: survey, technologies, opportunities, and challenges", *Sci Wor J*, Vol. 2014, <http://dx.doi.org/10.1155/2014/712826>.
18. A. Pashazadeh, N. J. Navimipour. " Big data handling mechanism in the healthcare applications: A comprehensive and systematic literature review", *J Biomed Info*, Vol.82, pp. 47–62, (2018).
19. J. Lee, J. Lee, M. J. Kim. "A study on perception of ski resort using big data analysis", *Korean J Phys Edu*, Vol. 56, No. 4 pp. 415–430, (2017).
20. M. S. Kim, S. W. Oh, J. Han. "A trend analysis on leisure sports tourism using social big data" *Korean J Phys Edu*, Vol. 58, No. 4 pp. 259–270, (2019).
21. E. J. Park, S. H. Park. "A study on occupational therapy research trends by dementia using big data analysis: Text-mining technology", *J Korea Aging Fri Indus Assoc*, Vol. 9, No. 2 pp. 15–24, (2017).
22. N. Mehta, A. Pandit. "Concurrence of big data analytics and healthcare: A systematic review", *Int J Med Info*, Vol.114, pp. 57–65, (2018).
23. F. Firouzi, A. Rahmani, K. Mankodiya, M. Badaroglu, G. Merrett, P. Wong, B. Farahani. "Internet-of-Things and big data for smarter healthcare: from device to architecture, applications and analytics", *Fut Gene Comp Sys*, Vol.78 (part 2), pp. 583–586, (2017).
24. L. Wang, C. A. Alexander. " Big data analytics in healthcare systems", *Int J Math Eng Manag Scis*, Vol.4, No.1, pp. 17–26, (2019).