
Integration of Cloud and Big Data Analytics for Future Smart Cities

Jungho Kang* and Jong Hyuk Park**

Abstract

Nowadays, cloud computing and big data analytics are at the center of many industries' concerns to take advantage of the potential benefits of building future smart cities. The integration of cloud computing and big data analytics is the main reason for massive adoption in many organizations, avoiding the potential complexities of on-premise big data systems. With these two technologies, the manufacturing industry, healthcare system, education, academe, etc. are developing rapidly, and they will offer various benefits to expand their domains. In this issue, we present a summary of 18 high-quality accepted articles following a rigorous review process in the field of cloud computing and big data analytics.

Keywords

Big Data, Cloud Computing, Smart City

1. Introduction

According to a recent forecast report, 65% of the world's population is expected to live in cities by 2040, with 90% of this urban population growth to occur in Asia and Africa. The smart city creates new sources of data by deploying smart sensors in city infrastructures to monitor and anticipate urban phenomena. The ability to obtain business value from data consistently is now a characteristic of successful organizations in all sectors such as manufacturing industry, healthcare, financial services, and so on. Big data analysis has the potential to extend and use smart city services to generate value by discovering deep business information as we access more data.

Technology advancement has played a major role in reaping the benefits of streamlined processes and profitable operations. The availability of data from all data sources such as city infrastructure, social media, etc. has changed the game for businesses. These huge volumes of storage and data processing give rise to the need for technology integration such as big data analytics and cloud computing. Many organizations are trying to leverage the power of these technologies to advance their business, but only some of them have succeeded. To improve computing operations, many organizations have embraced cloud computing technology. The integration of cloud computing and big data analytics is delivering benefits for organizations. It provides cost-effective and scalable solutions for future smart cities.

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Manuscript received November 25, 2019; accepted November 27, 2019.

Corresponding Author: Jong Hyuk Park (jhpark1@seoultech.ac.kr)

* Dept. of Information Security, Baewha Women's University, Seoul, Korea (kangsamm@baewha.ac.kr)

**Dept. of Computer Science and Engineering, Seoul National University of Science & Technology (SeoulTech), Seoul, Korea (jhpark1@seoultech.ac.kr), <http://www.parkjonghyuk.net>

The *Journal of Information Processing Systems (JIPS)* is the official international journal with indices such as ESCI, SCOPUS, EiCompendex, DOI, DBLP, EBSCO, and Google Scholar and is published by the Korean Information Processing Society (KIPS). There are four divisions: Computer system and theory, Multimedia systems and graphics, Communication systems and security, and Information systems and application. This issue includes 18 peer-reviewed papers following a rigorous review process.

2. Integration of Cloud and Big Data Analytics for Future Smart Cities

Yang et al. [1] presented a secure deduplication scheme for cloud storage based on Bloom filter, thereby dynamically extending the standard Bloom filter. Data security deduplication in a cloud environment is a very active research direction. A public dynamic Bloom filter array (PDBFA) is utilized in this scheme, which improves efficiency of ownership proof, realizes the fast detection of duplicate data blocks, and reduces the false positive rate of the system. The experimental results show that the computational overhead of file upload is relatively small, and it can effectively control the false positive rate of the system. Moreover, the experimental results suggest that the PDBFA scheme has the characteristics of low computational overhead and low false-positive rate.

In the mass production of garment factories, digital printing is universally introduced with technological improvements. Yuan and Huh [2] presented a digital printing method in a clothing model to reflect a client's design in real time by matching the area of the diagram where a client designs on a given clothing model and the area where a model standard reflects the actual design information of the customer.

Feng and Hu [3] proposed an infrared and visible light image fusion method based on variational multiscale decomposition. Guided filtering was used to fuse the structural components. The proposed method was compared with the traditional techniques for infrared and visible image fusion. It not only effectively overcomes the noise interference in the fusion process but also obtains better texture details while maintaining the edge structure and displays certain subjective and objective quality. Note, however, that the proposed method has one demerit, i.e., if guided filtering algorithm is used in the method, then the algorithm will have high complexity. Thus, the next study needs to improve the computational efficiency of the algorithm further.

To reduce data dimensions for faster and better prediction by using sub-modular optimization for selecting the optimal number of features, a correlation-based method was proposed by Attigeri et al. [4]. The proposed model seeks to find the right subset of representative features from which the predictive model can be constructed for a particular task. With optimal subsets, the proposed model achieved considerable accuracy in a significantly shorter execution time.

Yang and Gong [5] focused on the knowledge of molecular biological information extracted from massive biomedical texts with the aid of machine learning methods whose use was widespread. Biological molecular information has six classes: protein, DNA, RNA, cell line, cell component, and cell. The probability statistics method, CRF (conditional random fields), was utilized to discover these pieces of knowledge. As stored information in any library, computer system, or human mind, knowledge base can help biologists in etiological analysis and pharmacists in drug development. In this work, first, protein, DNA, RNA, cell line, cell type, cell component, and feed were extracted into the CRF++ tool. The six classes of molecular information were then derived from the kinds of literature related to autism. Finally, the knowledge discovery method was used.

Kim and Yun [6] proposed predicting the total saturation method based on sensing data in the information processing module of crowdsensing for a smart parking system. It implemented a predictive model and performed prediction model learning. Crowdsensing technology is used for improving the efficiency of the smart parking system because it has a low sensor installation price and no restrictions. The comparison of mean squared error between the time-based prediction model and the combined prediction model showed that sensing the data-based prediction model has high prediction accuracy because the sensing data provides insufficient information in the time-based prediction model.

Gong et al. [7] studied the contact-impact algorithm of the drift ice crashing diversion tunnel, which was based on the symmetric penalty function in the finite element theory. ANSYS/LS-DYNA was adopted as the platform for establishing a tunnel model and a drift ice model. LS-DYNA SOLVER was used as the solver, and LS-PREPOST was used to do post-processing, analyzing the degree of damage of drift ice on the tunnel. The software simulation results and the experiment results show that the tunnel lining surface will have varying degrees of deformation and failure when drift ice crashes with the tunnel lining at different velocities, different plan sizes, and different thicknesses of drift ice. The impact between drift ice and tunnel lining is irregular.

Wu et al. [8] presented an approach to summarizing the differences between the Chinese and Vietnamese bilingual news using a graph model. They tackled cross-language issues and analyzed the differences between the bilingual Chinese and Vietnamese news. Based on Wikipedia's multilingual concept description page, they bridge different languages by extracting elements to represent sentences. The experimental result demonstrates promising improvement over the current baselines method by 3 percentage points.

Kim and Lee [9] presented a general context-based model in an RNN to generate a summary that reflects the overall context of a document. In the experimental results, the proposed model has been shown to outperform the state-of-the-art approaches.

Based on the time series segmentation linear representation method and the k-nearest neighbor local anomaly detection algorithm, an anomaly detection algorithm based on a sliding window was proposed to detect the possible anomalies in the electrolytic cell [10]. It is relatively stable under normal conditions but fluctuates considerably when it has an anomaly. The length, slope, and mean of each line segment pattern are calculated by segmenting the cathode voltage time series and mapping them into a set of spatial objects.

Gookyi and Ryoo [11] analyzed ten RISC-V processors and determined the ideal processor core for low cost and low power for IoT sensing and actuating devising. Various processors were used, such as Roa Logic RV12, ORCA, SiFive E31, SCR1, Rocket, BOOM, MRISCV, PICORV32, Shakti-E, and Hummingbird. Xilinx Vivado v2018.02 Integrated Software Environment was used to synthesize the processor cores using Zynq-7000 XC7Z020 FPGA devices. PICORV32 yielded better synthesis results compared to other processor cores, and it is implemented in different ISAs such as ARM, LatticeMicro, OpenRISC, and SPARC. It has a vital role because the first step in selecting a processor incurs a low cost.

Wang and Guo [12] studied an evaluation model of rumor spreading based on social circle chat. The dynamic propagation equation helps depict the rumor spreading's evaluation, with the evaluation characteristics of the four nodes analyzed. The rumor communication study is based on various indicator terms such as degree, clustering coefficient, distance, and others of the network structure. It has a realistic basis, but platforms for spreading rumors are based on social media such as Facebook, Twitter, and so on. Thus, the rumor injection and the refuting-rumor-information injection evaluation process was simulated by the structure of virtual social network and social chats in this study. Finally, as a result of

this study, the spread and evolution of the rumors are related to the node degree on social web chat.

Shinde et al. [13] presented very effective debugging (VED), a new technique for the detection and correction of a division by zero error in all types of .NET application. It used C# applications because they are distributed through the Internet and used extensively in the form of executable format. The VED technique showed integer division by zero error as well as the location of the error-causing code in assembly language, including error recovery by the user's preference. It is useful to the developer by reducing the cost and effort of error recovery. The proposed technique is more suitable for small-scale software enterprises.

To solve the technological bottlenecks of the patch representation scheme in traditional learning-based methods, a method of synthesizing face sketches via the regularity terms of local similarity and non-local similarity was presented by Tang et al. [14]. They selected the most relevant patch samples to reconstruct the face sketch versions of the input photos by incorporating a local similarity regularization term into the selection of neighbors.

Choi et al. [15] proposed a software testing system for transforming a software to provide the runtime monitoring information from the software in real time. They implemented the testing system using a task/function-based lapse-time monitoring module to add modular monitoring targets. The software testing system generates monitorable binary code to reduce the load of runtime software testing.

Wen et al. [16] presented a case study related to the teleoperation system for interplanetary transportation. The interplanetary transportation mission of the teleoperation system has a vital role in the entire lunar exploration mission. Thus, in this paper, the researchers proposed a net assessment model for China's CE-3 "Jade Rabbit-1" and CE-4 "Jade Rabbit-2" mission verified for availability and practicability and which acquired the Weapons and Equipment Quality Management System Certification successfully. Finally, the net assessment model was found to be more comprehensive, forward-looking, and pure compared to the traditional evaluation methods for interplanetary transportation.

For business failure prediction, Xu and Yang [17] presented a new unweighted combination method using the integrated results of each basal classifier without weighting based on the soft set theory. The experimental results show that, regardless of sample size, prediction of business failure is done well compared to other selected benchmark methods.

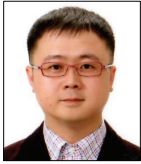
Yuan et al. [18] proposed a sentence similarity calculation method for improving the accuracy of long sentence similarity calculation. It was based on a system similarity function that used word2vector as the system element for calculating sentence similarity. The proposed methods have two characteristics: one is the negative effect of penalty item, and the other is that SSF (similar sentence function based on word2vector similar elements) does not satisfy the exchange rule. Finally, the experimental results show that the proposed method has higher accuracy than the WMD (word mover's distance) and has the shortest query time among the three calculation methods of SSF.

3. Conclusion

This issue features 18 high-quality articles following a rigorous review process. They include: contributions to theoretical research that present new techniques, concepts, or analyses; experience reports; experiments involving the implementation and application of new theories; and tutorials on state-of-the-art technologies related to integrating cloud computing and big data analytics.

References

- [1] X. Yan, W. Shi, and H. Tian, "Cloud storage security deduplication scheme based on dynamic bloom filter," *Journal of Information Processing Systems*, vol. 15, no. 6, pp. 1265-1276, 2019. <https://doi.org/10.3745/JIPS.04.0143>
- [2] Y. Yuan and J. H. Huh, "Automatic pattern setting system reacting to customer design," *Journal of Information Processing Systems*, vol. 15, no. 6, pp. 1277-1295, 2019. <https://doi.org/10.3745/JIPS.04.0148>
- [3] X. Feng and K. Hu, "Perceptual fusion of infrared and visible image through variational multiscale with guide filtering," *Journal of Information Processing Systems*, vol. 15, no. 6, pp. 1296-1305, 2019. <https://doi.org/10.3745/JIPS.04.0144>
- [4] G. Attigeri, M. Pai MM, and R. M. Pai. "Feature selection using submodular approach for financial big data," *Journal of Information Processing Systems*, vol. 15, no. 6, pp. 1306-1325, 2019. <https://doi.org/10.3745/JIPS.04.0149>
- [5] R. Yang and L. Gong, "Knowledge base associated with autism construction using CRFs learning," *Journal of Information Processing Systems*, vol. 15, no. 6, pp. 1326-1334, 2019. <https://doi.org/10.3745/JIPS.04.0146>
- [6] M. Kim and J. Yun, "Saturation prediction for crowdsensing based smart parking system," *Journal of Information Processing Systems*, vol. 15, no. 6, pp. 1335-1349, 2019. <https://doi.org/10.3745/JIPS.03.0123>
- [7] L. Gong, Z. Wang, Y. Li, C. Jin, and J. Wang, "Damage mechanism of drift ice impact," *Journal of Information Processing Systems*, vol. 15, no. 6, pp. 1350-1364, 2019. <https://doi.org/10.3745/JIPS.04.0147>
- [8] J. Wu, Z. Yu, S. Liu, Y. Zhang, and S. Gao, "Summarizing the differences in Chinese-Vietnamese Bilingual news," *Journal of Information Processing Systems*, vol. 15, no. 6, pp. 1365-1377, 2019. <https://doi.org/10.3745/JIPS.02.0126>
- [9] H. Kim and S. Lee, "Document summarization model based on general context in RNN," *Journal of Information Processing Systems*, vol. 15, no. 6, pp. 1378-1391, 2019. <https://doi.org/10.3745/JIPS.02.0123>
- [10] D. Cao, Y. Ma, and L. Duan, "An anomaly detection algorithm for cathode voltage of aluminum electrolytic cell," *Journal of Information Processing Systems*, vol. 15, no. 6, pp. 1392-1405, 2019. <https://doi.org/10.3745/JIPS.04.0150>
- [11] D. A. N. Gookyi and K. Ryoo, "Selecting a synthesizable RISC-V processor core for low-cost hardware devices," *Journal of Information Processing Systems*, vol. 15, no. 6, pp. 1406-1421, 2019. <https://doi.org/10.3745/JIPS.03.0129>
- [12] L. Wang and Y. Guo, "An evolution model of rumor spreading based on WeChat social circle," *Journal of Information Processing Systems*, vol. 15, no. 6, pp. 1422-1437, 2019. <https://doi.org/10.3745/JIPS.04.0152>
- [13] R. Shinde, M. Choi, and S. H. Lee, "Error recovery script of immunity debugger for C# .NET applications," *Journal of Information Processing Systems*, vol. 15, no. 6, pp. 1438-1448, 2019. <https://doi.org/10.3745/JIPS.04.0151>
- [14] S. Tang, X. Zhou, N. Zhou, L. Sun, and J. Wang, "Face sketch synthesis based on local and nonlocal similarity regularization," *Journal of Information Processing Systems*, vol. 15, no. 6, pp. 1449-1461, 2019. <https://doi.org/10.3745/JIPS.02.0125>
- [15] K. Choi, S. Kim, D. Park, and J. Cho, "Runtime software monitoring based on binary code translation for real-time software," *Journal of Information Processing Systems*, vol. 15, no. 6, pp. 1462-1471, 2019. <https://doi.org/10.3745/JIPS.01.0047>
- [16] J. Wen, Z. Zhao, and Q. Zhong, "Study on net assessment of trustworthy evidence in teleoperation system for interplanetary transportation," *Journal of Information Processing Systems*, vol. 15, no. 6, pp. 1472-1488, 2019. <https://doi.org/10.3745/JIPS.04.0153>
- [17] W. Xu and D. Yang, "A novel unweighted combination method for business failure prediction using soft set," *Journal of Information Processing Systems*, vol. 15, no. 6, pp. 1489-1502, 2019. <https://doi.org/10.3745/JIPS.04.0154>
- [18] X. Yuan, S. Wang, L. Wan, and C. Zhang, "SSF: sentence similar function based on word2vector similar elements," *Journal of Information Processing Systems*, vol. 15, no. 6, pp. 1503-1516, 2019. <https://doi.org/10.3745/JIPS.02.0124>



Junggho Kang <https://orcid.org/0000-0002-5038-698X>

He received his B.S. degree in Mathematics and his M.S. and Ph.D. degrees in Computer Science and Engineering from Korea University in Seoul, Korea in 1987, 1989, and 1993, respectively. Professor Kang works in the Department of Multimedia Engineering of Dongguk University, Korea. His research interests include multimedia cloud computing, information security for cloud computing, mobile computing, the Internet of Things (IoT), and wireless sensor network applications. He was a professor in the Department of Computer Engineering of Wonkwang University, Korea, from 1993 to 2012. He worked and conducted research at Michigan State University and Wayne State University as his capacity as a visiting professor in 1997 and 2004, respectively. He is also an executive editor of the *Journal of Information Processing Systems*, an associate editor of the *Journal of Supercomputing* (JoS) and the *International Journal of Communication Systems* (IJCS), an editor of the *Journal of Internet Technology* (JIT), and, finally an associate editor of the *Journal of Human-centric Computing* (HCIS). In addition, he has been employed as a guest editor for various international journals by a number of publishers including Springer, Elsevier, John Wiley, Oxford University Press, Hindawi, Emerald, and Inderscience. He is also a member of the IEEE.



James J. (Jong Hyuk) Park <https://orcid.org/0000-0003-1831-0309>

He received Ph.D. degrees from the Graduate School of Information Security, Korea University, Korea and the Graduate School of Human Sciences of Waseda University, Japan. Dr. Park served as a research scientist at the R&D Institute, Hanwha S&C Co. Ltd., Korea from December 2002 to July 2007, and as a professor at the Department of Computer Science and Engineering, Kyungnam University, Korea from September 2007 to August 2009. He is currently employed as a professor at the Department of Computer Science and Engineering and the Department of Interdisciplinary Bio IT Materials, Seoul National University of Science and Technology (SeoulTech), Korea. Dr. Park has published about 200 research papers in international journals and conferences. He has also served as the chair, program committee chair or organizing committee chair at many international conferences and workshops. He is a founding steering chair of various international conferences including MUE, FutureTech, CSA, UCAWSN, etc. He is employed as editor-in-chief of *Human-centric Computing and Information Sciences* (HCIS) by Springer, *The Journal of Information Processing Systems* (JIPS) by KIPS, and the *Journal of Convergence* (JoC) by KIPS CSWRG. He is also the associate editor or editor of fourteen international journals, including eight journals indexed by SCI(E). In addition, he has been employed as a guest editor for various international journals by such publishers as Springer, Elsevier, Wiley, Oxford University Press, Hindawi, Emerald, and Inderscience. Dr. Park's research interests include security and digital forensics, human-centric ubiquitous computing, context awareness, and multimedia services. He has received "best paper" awards from the ISA-08 and ITCS-11 conferences and "outstanding leadership" awards from IEEE HPCC-09, ICA3PP-10, IEE ISPA-11, and PDCAT-11. Furthermore, he received an "outstanding research" award from SeoulTech in 2014. Also, Dr. Park's research interests include human-centric ubiquitous computing, vehicular cloud computing, information security, digital forensics, secure communications, multimedia computing, etc. He is a member of the IEEE, IEEE Computer Society, KIPS, and KMMS.