JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Advanced Technologies in Blockchain, Machine Learning, and Big Data

Ji Su Park* and Jong Hyuk Park**

Abstract
Blockchain, machine learning, and big data are among the key components of the future IT track. These technologies are used in various fields; hence their increasing application. This paper discusses the technologies developed in various research fields, such as data representation, Blockchain application, 3D shape recognition and classification, query method, classification method, and search algorithm, to provide insights into the future paradigm. In this paper, we present a summary of 18 high-quality accepted articles following a rigorous review process in the fields of Blockchain, machine learning, and big data.

Keywords
Big Data, Blockchain, Machine Learning

1. Introduction

Software development led to a period of significant social transformation. The current era is constantly changing technology as well as the environment. Therefore, research and implementation strategies should be adopted by the information technology (IT) environment to prepare for future changes regardless of the cause and speed of change.

The recent changes in IT can be summarized as big flows, such as Blockchain, machine learning, and big data. Blockchain is a data distribution processing technology that allocates and stores all data managed by users participating in the network. This technology suggests that all users manage the data controlled by the existing central administrator. Bitcoin is a typical Blockchain technology application. In this paper, Blockchain technology is applied to patch management and electronic voting systems.

Machine learning is a field of artificial intelligence (AI) that lets people acquire new knowledge by providing data to computers and learning in the same way as humans. This technology has started to attract attention due to its recent development into deep learning. Machine learning-based applied technologies are also emerging. Featured in this paper is a paper that focuses on the identification and analysis of the fuzzy probability of tea diseases and early Bayesian games.

Expressed differently from existing data in terms of size, speed, and diversity, big data has recently drawn attention in various fields. Moreover, various techniques for collecting, storing, and analyzing

structured and unstructured data are developed. This issue features papers that analyze some of the methods applied in different fields, such as subject extraction and classification, complex keyword extraction, offline-to-online (O2O) service analysis, and movie recommendation system.

The *Journal of Information Processing Systems* (JIPS) is an official international journal with indices such as ESCI, SCOPUS, Ei Compendex, 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 features 18 peer-reviewed papers following a rigorous review process.

2. Advanced Technologies in Blockchain, Machine Learning, and Big Data

Woo et al. [1] conducted a study to apply ITSM (IT Service Management), a framework for the integrated management of IT services, to the national defense acquisition system in Korea. Through the analysis of the existing obsolete national defense acquisition system and by applying ITSM to the existing obsolete national defense acquisition system for faster processing and response, the necessity of the ITSM system is explained, and the service satisfaction level for ITSM was investigated. In addition, the data processing procedure according to the application of ITSM to the defense acquisition system is expressed using the UML diagram, and the efficient ITSM application method is described.

Khamis et al. [2] proposed a new formal representation of Digital Contents domain that uses ontology-based model and semantic vector to redefine digital content data and combine with media segmentation methods. They also suggested an ontology-based digital contents query solution to provide a faster access mechanism of digital contents data stored under the persistent database. To do this, the authors classified the digital contents data into different types and introduced a formal representation method of digital contents data by redefining the digital contents data based on OWL/RDF and combining with media segmentation methods.

Knogwudhikunakorn and Waiyamai [3] proposed a clustering technique to group short-text documents such as news headlines, social media status, and instant messages into multiple related clusters. According to them, the combination of document representation, document distance, and document clustering needed to be identified in order to provide the best clustering quality. To this end, the k-mean partitioning-based clustering technology was applied to the proposed scheme. To verify the efficiency of the proposed method, they presented the experiment results for clustering quality in terms of accuracy, recall, F1-score, and adjusted Rand index.

Song et al. [4] proposed a Hyperledger Fabric Blockchain-based distributed patch management system to ensure the stability of enterprise systems through security, log management, and up-to-date status supervision and monitoring functions by improving the problems of the centralized structure of enterprise patch management systems. The authors designed the system using the Blockchain's distributed database storage method and PBFT (practical Byzantine fault tolerance) consensus algorithm technology and implemented a test environment for the patch management system. The technical validity of the proposed scheme was verified based on the test scenario so that the patch could be executed normally; through this, the integrity of the patch application status database was verified.

In order to control the risk and uncertainty of the Letter of Credit (L/C), a payment method mainly used in international trade processes, Cheng and Huang [5] presented a risk assessment and decision-making method for the L/C settlement of listed companies based on fuzzy probability and Bayesian game theory. To solve the problem of incomplete information related to L/C, the FAHP and KMV methods were used, and an analytical model was designed for import and export companies based on fuzzy probability and Bayesian game theory. The authors presented reasonable measures to aid in L/C risk assessment and decision making through their own simulation and case study.

Tan [6] presented a new improved emotional text classification, which is one of the essential research topics in the area of natural language processing. In order to improve the weakness of the existing LDA topic model, the author proposed an improved weighted-LDA topic model that assigns weights so that words related to the subject are not pushed out to high-frequency words in the process of calculating the sample and distribution of words. Based on the experiment results, improved results were shown compared to the existing algorithms in terms of subject classification, precision, and F1-measurement.

Kanaan and Behrad [7] presented a new algorithm for 3D shape recognition using the local features of model views and its sparse representation. The algorithm processes include the normalization of 3D models and extraction of 2D views from uniformly distributed viewpoints. Support vector machine classifiers are also used to recognize the 3D models by applying Gabot filters as initial recognition, measuring the similarity, and representing the intermediate feature vectors of 3D models. An experiment using the Princeton shape benchmark databases yielded effective results with average recognition rate of 89.7% compared to other known algorithms.

Li et al. [8] presented a class-oriented attribute reduction (COAR) algorithm, an enhanced heuristic attribution reduction algorithm for providing the better match for multiclass datasets since the existing heuristic algorithms are not perfect for multiscale datasets. The authors proposed the new ensemble constructed algorithm based on class-oriented reducts with a customized weighted majority voting strategy considering the strong dependence in a reduct and target class. The experiment showed that the proposed algorithm was better in terms of four general evaluation metrics by using five actual multiclass datasets.

Selvaraj et al. [9] presented a new system architecture for O2O service and showed useful knowledge from previous real-time freight data for further business development in the Freight management vertical. A new business module for rapid decision making is needed to improve the business module, and the data analysis process is performed offline. According to the authors, the proposed system architecture is useful for the transport management companies in dynamically requesting the big data analysis results using O2O services for such kinds of predicted customer expectation, price, and overhead reduction by growing profit margins and load balancing.

Sun et al. [10] presented the mechanism for archiving an online estimation of transmission line full parameters to predict the influence on the electromagnetic environment. The authors propose a method that uses Phasor Measurement Unit and Supervisory Control and Data Acquisition and differs from the existing method, which is based on independent resistance estimation. The experiment result showed that the online estimation of transmission line full parameters was much more accurate.

Zhang et al. [11] presented the k nearest neighbor query method of a line segment in obstacle space to make up an existing method that cannot handle the nearest neighbor query effectively. The query process has two steps: (1) the filtering process uses the proposed corresponding pruning rules, and (2) the refining

process gets the final result by comparing the distance with the proposed corresponding distance expression method. The experiment result showed that the proposed algorithm could solve the problem of k nearest neighbor query of the line segment in the obstacle environment.

Roh and Lee [12] used Blockchain technology to ensure transparency among the participants. The existing electronic voting system works by applying various algorithms. As one disadvantage of this system, however, the content of the vote can be forged or changed by the administrator as all rights are granted to the administrator. Therefore, the electronic voting system uses a Blockchain technology that provides stability and data integrity. This technology satisfies the security requirements of the system and uses a private Blockchain algorithm that is 50 times better than the existing public one.

Zou et al. [13] identified tea diseases based on spectral reflectance and machine learning. Machine learning models can classify unknown objects, but using this technique to classify dimensions of hyperspectral data results in overfitting. Therefore, the authors improved the identification method of tea diseases and random forests based on the function selector and spectral reflection and decision tree. The experiment results showed that the recall rate and F1 score improved, with the accuracy of tea diseases recording average values of 15%, 7%, and 11%.

People read a document for the conceptual extraction of keywords from the document, and then construct a concept for information and set keywords to represent the material by merging several words. Lee [14] collected the titles and abstracts of journals about natural and auditory languages to verify and analyze the validity of the extracted keywords. The author proposed a new method to determine the importance of each keyword, excluding unrelated keywords. As a result of the experiment, the proposed system showed up to 96% accuracy.

Ma et al. [15] proposed an energy-aware virtual data center embedding using an energy consumption model to solve the energy consumption problem of embedding a virtual data center. The model quantitatively measures the energy consumption of virtual machine and switch nodes and utilizes heuristic and particle cluster optimization techniques. Their results suggest that energy is effectively conserved, and that the embedding success rate is guaranteed.

Liu and Li [16] developed a meta-heuristic algorithm called photon search algorithm (PSA) through mathematical formulas and models of the proposed algorithm based on physical knowledge, including light speed consistency, uncertainty, and Pauli exclusion principles. The evaluation of the algorithm is compared with 7 single and 23 multi-modal benchmark functions. Their results suggest that PSA obtained high efficiency with excellent convergence and robust global search function. In solving the optimal solution of a specific function, however, a part that is slightly inferior to the existing heuristic algorithm was observed.

A recommendation technology based on the personal information of a user or a best-selling product is generally used in a movie recommendation system. Vilakone et al. [17] proposed a movie recommen-dation system using k-click and normalized discounted cumulative gain methods to improve accuracy. They found that the most acceptable MAPE value was obtained at $k = 11$, which increased the accuracy to 87.28% and solved the cold-start problem.

In the paper by Zhang et al. [18], they proposed a quality index based on the new Lempel-Ziv complexity (ELZC) to evaluate the quality of multi-lead electrocardiogram (ECG) collected in real time. For the multi-lead ECG quality evaluation of the proposed technique, three algorithms with the same complexity were compared, and six artificial time series were calculated according to each algorithm to

compare performance in terms of randomness and irregularity within the time series. To do this, the authors analyzed the sensitivity of the algorithm according to the noise content within the ECG and performed an evaluation by reflecting the trend of changes to artificial synthetic noise containing different kinds of noise. The graph shows that it is more suitable for quality evaluation.

3. Conclusion

This issue features 18 high-quality articles following a rigorous review process. This paper has reviewed the technologies developed in various research fields, such as data representation, Blockchain application, 3D shape recognition and classification, query method, classification method, and search algorithm, to provide insights into the future paradigm. Published articles on the following topics are also featured in this issue: contributions to theoretical research, including 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 Blockchain, machine learning, and big data.

References

[1] H. Woo, S. J. Jeong, and J. H. Huh, "Improvement of ITSM it service efficiency in military electronic service," *Journal of Information Processing Systems*, vol. 16, no. 2, pp. 246-260, 2020. https://doi.org/10.3745/JIPS.03.0134

[2] K. A. L. Khamis, H. Song, and X. Zhong, "Formal representation and query for digital contents data," *Journal of Information Processing Systems*, vol. 16, no. 2, pp. 261-276, 2020. https://doi.org/10.3745/JIPS.02.0130

[3] S. Kongwudhikunakorn and K. Waiyamai, "Combining distributed word representation and document distance for short text document clustering," *Journal of Information Processing Systems*, vol. 16, no. 2, pp. 277-300, 2020. https://doi.org/10.3745/JIPS.04.0164

[4] K. T. Song, S. I. Kim, and S. H. Kim, "A design for a Hyperledger Fabric Blockchain-based patch-management system," *Journal of Information Processing Systems*, vol. 16, no. 2, pp. 301-317, 2020. https://doi.org/10.3745/JIPS.03.0136

[5] Z. Cheng and N. Huang, "Risk assessment and decision-making of a listed enterprise's L/C settlement based on fuzzy probability and Bayesian game theory," *Journal of Information Processing Systems*, vol. 16, no. 2, pp. 318-328, 2020. https://doi.org/10.3745/JIPS.04.0156

[6] X. Tan, "Topic extraction and classification method based on comment sets," *Journal of Information Processing Systems*, vol. 16, no. 2, pp. 329-342, 2020. https://doi.org/10.3745/JIPS.04.0165

[7] H. Kanaan and A. Behrad, "Three-dimensional shape recognition and classification using local features of model views and sparse representation of shape descriptors," *Journal of Information Processing Systems*, vol. 16, no. 2, pp. 343-359, 2020. https://doi.org/10.3745/JIPS.02.0132

[8] M. Li, S. Deng, and L. Wang, "Ensemble of classifiers constructed on class-oriented attribute reduction," *Journal of Information Processing Systems*, vol. 16, no. 2, pp. 360-376, 2020. https://doi.org/10.3745/JIPS.04.0166

[9] S. Selvaraj, H. Kim, and E. Choi, "Offline-to-online service and big data analysis for end-to-end freight management system," *Journal of Information Processing Systems*, vol. 16, no. 2, pp. 377-393, 2020. https://doi.org/10.3745/JIPS.01.0051

[10] Z. Sun, X. Zhou, L. Liang, and Y. Mo, "Electromagnetic environment of transmission line based on full parameter online estimation," *Journal of Information Processing Systems*, vol. 16, no. 2, pp. 394-405, 2020. https://doi.org/10.3745/JIPS.04.0171

[11] L. Zhang, S. Li, Y. Guo, and X. Hao, "A method for k nearest neighbor query of line segment in obstructed spaces," *Journal of Information Processing Systems*, vol. 16, no. 2, pp. 406-420, 2020. https://doi.org/10.3745/JIPS.04.0167

[12] C. H. Roh and I. Y. Lee, "A study on electronic voting system using private Blockchain," *Journal of Information Processing Systems*, vol. 16, no. 2, pp. 421-434, 2020. https://doi.org/10.3745/JIPS.03.0135

[13] X. Zou, Q. Ren, H. Cao, Y. Qian, and S. Zhang, "Identification of tea diseases based on spectral reflectance and machine learning," *Journal of Information Processing Systems*, vol. 16, no. 2, pp. 435-446, 2020. https://doi.org/10.3745/JIPS.02.0133

[14] S. S. Lee, "Conceptual extraction of compound Korean keywords," *Journal of Information Processing Systems*, vol. 16, no. 2, pp. 447-459, 2020. https://doi.org/10.3745/JIPS.02.0131

[15] X. Ma, Z. Zhang, and S. Su, "Energy-aware virtual data center embedding," *Journal of Information Processing Systems*, vol. 16, no. 2, pp. 460-477, 2020. https://doi.org/10.3745/JIPS.02.0112

[16] Y. Liu and R. Li, "PSA: a photon search algorithm," *Journal of Information Processing Systems*, vol. 16, no. 2, pp. 478-493, 2020. https://doi.org/10.3745/JIPS.04.0168

[17] P. Vilakone, K. Xinchang, and D. S. Park, "Movie recommendation system based on users' personal information and movies rated using the method of k-clique and normalized discounted cumulative gain," *Journal of Information Processing Systems*, vol. 16, no. 2, pp. 494-507, 2020. https://doi.org/10.3745/JIPS.04.0169

[18] Y. Zhang, Z. Ma, and W. Dong, "Nonlinear quality indices based on a novel Lempel-Ziv complexity for assessing quality of multi-lead ECGs collected in real time," *Journal of Information Processing Systems*, vol. 16, no. 2, pp. 508-521, 2020. https://doi.org/10.3745/JIPS.04.0170

Ji Su Park https://orcid.org/0000-0001-9003-1131

He received his B.S. and M.S. degrees in Computer Science from Korea National Open University, Korea, in 2003 and 2005, respectively and Ph.D. degree in Computer Science Education from Korea University, 2013. He is currently a Professor in Department of Computer Science and Engineering from Jeonju University in Korea. His research interests are in mobile grid computing, mobile cloud computing, cloud computing, distributed system, domputer education, and IoT. He is employed as associate editor of *Human-centric Computing and Information Sciences* by Springer and *The Journal of Information Processing Systems* by KIPS. He has received "best paper" awards from the CSA2018 conferences and "outstanding service" awards from CUTE2019.

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* by Springer, *The Journal of Information Processing Systems* by KIPS, and the *Journal of Convergence* 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.