Editorial Preface

Big Data Analytics and Artificial Intelligence in Cyber-Physical Systems:
A Review of Challenges and Advances

Junsheng Zhang, Institute of Scientific and Technical Information of China, Beijing, China
Yunchuan Sun, Business School, Beijing Normal University, Beijing, China

INTRODUCTION

We have entered the era of big data. With the development of information and communication technologies in recent years, more and more data are sensed, collected and stored. Sensors are widely distributed to collect distributed objects, while wireless network, mobile network and Internet have made the data transfer quickly and easily. Data sensing, collecting and using have led to the occurrence of Cyber-Physical Systems (CPS) which are the infrastructure of big data. To manage and utilize the collected data, cloud computing, fog computing and edge computing enable the storage, processing and analytics for large-scaled data. After processing and analysis on data in CPS, information and knowledge are identified and discovered that are preliminary and basis for intelligent applications.

Big data is more than simply a matter of size; and it is an opportunity to find insights in new and emerging types of data and content, to acquire intelligence and to realize more intelligent applications. Intelligent applications need big data analytics and artificial intelligence studies. Achieving the benefits of CPS requires the integration and management of vast amounts of heterogeneous data for knowledge-based decision systems which need as much as data in limited time. With the occurrence and advances of CPS, big data has become an emerging paradigm applied to datasets whose size is beyond the ability of commonly used software tools to capture, manage, and process the data in a tolerable elapsed time. The datasets are often from various, heterogeneous and unstructured sources such as social media, sensors, scientific applications, surveillance, video and image archives, Internet texts and documents, medical records, business transactions and web logs. Furthermore, big data will be of high value and establish trust in it for artificial intelligence research and business decision-making.

Artificial intelligence (AI) has become the popular research area again. AI is a kind of advanced technology which has caused a lot of disruptive changes in many fields. During more than 60 years from the birth of AI, it has undergone several stages of success and failure. Recent years, AI technology with machine learning, especially in-depth learning as the core, has developed rapidly in the fields of vision, speech, natural language and other applications. It has begun to be endowed with various industries like hydropower and coal. This new hot research trend of AI with Google *AlphaGo* as the landmark achievement comes back with the benefits of big data in CPS and powerful computing in cloud computing. On one hand, without enough data for learning and training, machine is hard to become intelligent; on the other and, without high performance computing, it is hard to enable the big data-based machine learning in acceptable time. There artificial intelligence studies have been closely connected and combined big data.

In this work, we focus on big data analytics and artificial intelligence studies in CPS and give a review and advances in this research area. We introduce the major challenges for big data analytics and AI in CPS, and then we introduce the research works included in this special issue.

MAJOR CHALLENGES FOR BIG DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE IN CYBER-PHYSICAL SYSTEMS

Realizing the full potential of big data in CPS requires solving technical and business challenges including the identification of things, data organization and integration, big data processing and mining, security and privacy in big data utilization. Various technologies have been studied, developed and discussed to support the handling of big data such as massively parallel processing databases, scalable storage systems and cloud computing platforms. Various applications based on big data analytics have occurred in different domains which cover CPS and social society (Wang, Kung & Byrd, 2018; Zhang, Yao, Sun & Fang, 2016). However, many challenges still exist, including how to correctly identify things in CPS, how to integrate and transfer enriched data, how to collect and process information, how to retrieve the knowledge automatically from the huge volumes of data and information, how to share the knowledge among smart things and persons, and how to ensure the security of the information and protect the private information.

Identification of Things in Big Data

Sensors in CPS and users of social media software have collected massive data about various objects observed from all over the world. The collected data have different formats such as digital numbers, short texts, images, audios and videos. The massive data are collected and transferred in the Internet. However, there are large scale data have no definite metadata or even miss the description, which has hindered the utilization and analytics of data. So identifying things from big data has become an urgent problem.

Identification technologies at the hardware level include passive and active RFID tags, Near Field Communication systems, and geo-location beacons based on Bluetooth Low Energy communication, bar codes and QR codes. Unknown tag identification still remains open, and there are two major challenging issues: (1) How to avoid tag-tag collision efficiently. (2) How to collect unknown tags' IDs efficiently (Zhang, He, Qian, Li, & Zhao, 2016). A multi-dimensional approach from the content was proposed to identify places, objects, people, and animals based on multidimensional modeling of information contents (Rykowski, 2017).

Integration and Organization of Big Data

After things are identified, how to integrate and organize the heterogeneous and distributed data becomes a new research problem. The same thing may have different sensors or observers that collect data from different aspects. For example, some sensors collect the data of temperature, size and quantity, others are focusing on the data of color and position. Data with different content about the same object are collected from different sensors or observers, and together compose the complete information of the object. During the integration, data from various sources may have different schemas, and then it is necessary to build the mapping rules between attributes in different schemas for the unified schema which is necessary for the big data analytics in the global scope. Big data may be distributed in different databases and files, how to connect these data is the aim of the big data organization.

Organic Streams was proposed as a unified framework for personal big data integration and organization towards social sharing and individualized sustainable use (Jiang, Li, Yang, & Cuzzocrea, 2015). The role of conceptual modeling for big data is then analyzed and suggestions made for effective conceptual modeling efforts with respect to big data (Storey & Song, 2017).

On-Demand Big Data Storage and Processing

There are different technologies have between studied and developed to store and process big data such as cloud computing, fog computing, edge computing and mobile computing. However, there is no technology can be suitable for all kinds of scenes. Cloud computing could offer on-demand and scalable storage and processing services that can scale to IoT requirements. However, the delay caused by transferring data to the cloud and back to the application is unacceptable for time sensitive applications such as health-monitoring, emergency-response, and other latency-sensitive applications. To solve this problem, edge computing was proposed to use computing resources near IoT sensors for local storage and preliminary data processing. Edge computing could decrease network congestion. However, edge devices can't handle multiple IoT applications competing for the limited resources. Fog computing is proposed to overcome these limitations which avoids resource contention at the edge by leveraging cloud resources and coordinating the use of geographically distributed edge devices (Dastjerdi & Buyya, 2016). Therefore, it is necessary to study the on-demand big data storage and processing technologies for specific requirements of storage and computing. The computing resources and storage are assigned according to the specific application requirements.

Big Data Mining and Knowledge Discovery

Understanding the meaning implied big data requires cutting-edge tools and techniques that can analyze and extract useful knowledge from big data. Both the wonderful living of human beings and the high efficiency of business rely heavily on how to intelligently and correctly use the big data and how to retrieve useful knowledge from the massive data, such that it would be possible to seamlessly integrate the virtual world, the physical world and social society.

Wu, Zhu, Wu and Ding (2014) presented a HACE theorem that characterizes the features of the Big Data revolution, and proposes a Big Data processing model, from the data mining perspective. Chen et al. (2015) gave a systematic way to review data mining in knowledge view, technique view, and application view, including classification, clustering, association analysis, time series analysis and outlier analysis, and the latest application cases are also surveyed.

Security and Privacy During Big Data Utilization

Security and privacy are important issues in Big Data. The concern on security and privacy protection exist all the process from collection, transmission to analysis and applications (Sun, Zhang, Xiong & Zhu, 2014). The features of Big Data have brought unprecedented challenges to security and privacy protection together with the convenience and surprise they have provided to us. To protect the confidentiality, integrity, availability, traditional security measures such as cryptography, event analysis, intrusion detection, prevention and access control have taken a new dimension in the Big Data utilization. To protect the privacy, new pattern of measures such as privacy-preserved data analytics need to be explored.

Cuzzocrea (2014) provided an overview of research issues and achievements in the field of privacy and security of big data, by highlighting open problems and actual research trends, and drawing novel research directions in this field. Sadeghi, Wachsmann and Waidner (2015) introduced the security and privacy challenges in industrial Internet of Things.

A BRIEF REVIEW OF ACCEPTED ARTICLES OF THIS SPECIAL ISSUE

The purpose of this special issue is to collect analytical, simulation and empirical research papers that explore data, information and knowledge in CPS: data models required, data organization, information processing, knowledge representation, knowledge management, knowledge acquisition, and knowledge mining techniques to analyze interactions across spaces, and how collaboration and interaction in CPS can be facilitated leveraging the best practices developed in including artificial

intelligence, social computing, social and community intelligence, ubiquitous computing, wireless sensor networks and service oriented computing. We have received a large number of submissions from many countries. This issue will collaborate with 2017 International Conference on Identification, Information and Knowledge in the Internet of Things, and selected papers were recommended to this special issue for publication after extension. After a rigorous review with two rounds and revision process, we assembled five papers on the latest advances in state-of-the-art of big data analytics and artificial intelligence in CPS. This special issue contains research works from the data processing technologies such as storage and query of large-scale remote sensing images, sensing data classification to applications of big data analytics such as dietary preference analysis, stickiness of mobile payment and correlation analysis on massive scientific and technical literature.

The paper "Research on Improved Method of Storage and Query of Large Scale Remote Sensing Images" by Jing, Tian and Chen et al. introduce the storage and large scale remote sensing images. The traditional method is used to deal with massive remote sensing data stored with low efficiency and poor scalability. The authors present a parallel processing method based on MapReduce and HBase. The filling of remote sensing image by Hilbert curve makes the MapReduce method construct pyramids in parallel to reduce network communication between nodes. Then the authors design massive remote sensing data storage model composed of metadata storage model, index structure and filter column family. Finally, this paper uses MapReduce framework to realize pyramid construction, storage and query of remote sensing data. The experimental results show that the proposed method can effectively improve the speed of data writing and querying, and it has good scalability.

The paper "Towards Real-time Multi-sensor Golf Swing Classification Using Deep CNNs" by Jiao, Wu and Bie et al. use deep CNNs for realizing the real-time multi-sensor golf swing classification. In recent years, smart sports equipment and body sensory systems have become popular in professional and amateur sports. A natural way of their use is the biofeedback systems, especially the real-time variants. One of a few remaining problems in real-time applications is the discovery of knowledge from the embedded sensors data. In sports training, such knowledge helps at accelerated motor learning. The authors start by exploring the possibilities of classification of golf swing performance with the 1-D convolutional neural network (CNN) in real-time. They investigate multiple golf swing data classifiers based on CNNs fed with multi-sensor signals. They test the possibilities of real-time performance of CNN methods on the multi-length sequences. In addition, the authors evaluate the performance of their well-trained CNN-based classifier on the aforementioned test set in terms of common indicators. Experiments and corresponding results show that the proposed models can satisfy the real-time requirement of the accuracy of the classification and outperform support vector machine (SVM).

The paper "Detecting Users' Dietary Preferences and Their Evolutions via Chinese Social Media" by Zhou and Zhang studies the dietary preferences and evolution based on the data collected from Chinese social media. Dietary preferences are linked to human life and region culture. With the rapid development of Internet, people are becoming frequently interested in sharing their opinions about dietary in social media. This paper aims to mine social media users' dietary preferences and their evolutions with user generated content. The authors use microblogs from weibo.com to detect dietary preferences and their evolutions of social media users in China via sentiment analysis. Specifically, firstly, the authors compare four aspect extraction methods to obtain dietary aspects. Secondly, sentiment polarities of aspects and dishes are identified. Finally, dietary preference evolutions are analyzed. Empirical analysis presents that social media users in Weibo are not satisfied with status quo of Chinese dietary. Meanwhile, gender and region have significant effects on dietary preferences, and users' dietary preferences change over time. In addition, experimental results show that contextual information is useful in extracting dietary aspects.

The paper "Consumers' Stickiness to Mobile Payment Applications: An Empirical Study of WeChat Wallet" by Matemba, Li and Maiseli studies the stickiness to mobile payment with WeChat wallet as a case. WeChat wallet has gained wide acceptance because it incorporates compelling

features facilitating users to conduct financial transactions conveniently. Previous studies fail to explain why consumers remain sticky to this technology. Scholars have only attempted to explore factors for WeChat wallet acceptance—an unused contribution from developers intending to generate higher profits through consumers' sticky behaviors. This article establishes a theoretical model that integrates novel constructs predicting consumers' stickiness to WeChat wallet. The authors administered a questionnaire to 450 Chinese and foreigners in China. Internet infrastructure assisted to collect responses from users of WeChat wallet. Using structural equation modeling and SPSS for data analysis, the authors found that perceived availability of merchant support and red envelope preference promote consumers' stickiness. In addition, perceived security insignificantly moderates the relationships between convenience/social influence and consumers' stickiness. The results may help developers to design customer-centered products.

The paper "Research on Methodology of Correlation Analysis of Sci-tech Literature Based on Deep Learning Technology in the Big Data" by Zeng, Xu and Li et al. studies the correlations between massive scientific and technical literature. In the big data era, it is a great challenge to identify high-level abstract features out of a flood of sci-tech literature to achieve in-depth analysis of data. The deep learning technology has developed rapidly and achieved good applications in many fields, but has rarely been utilized in the research of sci-tech literature data. This paper introduced the presentation method of vector space of terminologies in sci-tech literature based on the deep learning model. It explored and adopted a deep AE model to reduce the dimensionality of input word vector feature. And it put forward the methodology of correlation analysis of sci-tech literature based on the deep learning technology. The experimental results showed that the processing of sci-tech literature data could be simplified into computation of vectors in the multi-dimensional vector space, and the similarity in vector space could be used to represent similarity in text semantics. The correlation analysis of subject contents between sci-tech literatures of the same or different types can be made using this method.

CONCLUSION

Cyber-Physical Systems have been widely used in the daily life and business, which have led to the coming of big data era. Big data analytics and artificial intelligence applications have great influence on our life and work. In this paper, we give a review of challenges and advances in the areas including big data analytics and artificial intelligence in Cyber-Physical Systems. Data collection, transferring, storage and processing will still be hot research topics for CPS and Big Data, while how to analyze and discover the values hidden in the big data will be more exciting and attract more interests from both academic and industry. Based on the development of big data analytics and artificial intelligence research and development, we believe more intelligent applications will occur which will make the life better in the near future.

ACKNOWLEDGMENT

We would like to thank all authors for their valuable contributions to this special issue and all reviewers for their hard work which help us further enhance the quality of the issue. Also, we would also like to express our honor to serve as the guest editors of this issue.

Junsheng Zhang
Yunchuan Sun
Guest Editors
IJGCMS

REFERENCES

Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A. V., & Rong, X. (2015). Data mining for the internet of things: Literature review and challenges. *International Journal of Distributed Sensor Networks*, *11*(8), 431047. doi:10.1155/2015/431047

Cuzzocrea, A. (2014, November). Privacy and security of big data: current challenges and future research perspectives. In *Proceedings of the First International Workshop on Privacy and Secuirty of Big Data* (pp. 45-47). ACM. doi:10.1145/2663715.2669614

Dastjerdi, A. V., & Buyya, R. (2016). Fog computing: Helping the Internet of Things realize its potential. *Computer*, *49*(8), 112–116. doi:10.1109/MC.2016.245

Jiang, H., Li, K. C., Yang, L. T., & Cuzzocrea, A. (2015). Organic Streams: A Unified Framework for Personal Big Data Integration and Organization Towards Social Sharing and Individualized Sustainable Use. In Big Data: Algorithms, Analytics, and Applications (pp. 278-293). Chapman and Hall/CRC.

Rykowski, J. (2017, September). Multi-dimensional identification of things, places and humans. In *2017 IEEE International Conference on RFID Technology & Application (RFID-TA*) (pp. 152-157). IEEE. doi:10.1109/RFID-TA.2017.8098910

Sadeghi, A. R., Wachsmann, C., & Waidner, M. (2015, June). Security and privacy challenges in industrial internet of things. In *52nd ACM/EDAC/IEEE Design Automation Conference (DAC)* (pp. 1-6). IEEE.

Storey, V. C., & Song, I. Y. (2017). Big data technologies and management: What conceptual modeling can do. *Data & Knowledge Engineering*, *108*, 50–67. doi:10.1016/j.datak.2017.01.001

Sun, Y., Zhang, J., Xiong, Y., & Zhu, G. (2014). Data security and privacy in cloud computing. *International Journal of Distributed Sensor Networks*, *10*(7), 190903. doi:10.1155/2014/190903

Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, *126*, 3–13. doi:10.1016/j.techfore.2015.12.019

Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, *26*(1), 97–107. doi:10.1109/TKDE.2013.109

Zhang, D., He, Z., Qian, Y., Wan, J., Li, D., & Zhao, S. (2016). Revisiting unknown RFID tag identification in large-scale internet of things. *IEEE Wireless Communications*, *23*(5), 24–29. doi:10.1109/MWC.2016.7721738

Zhang, J., Yao, C., Sun, Y., & Fang, Z. (2016). Building text-based temporally linked event network for scientific big data analytics. *Personal and Ubiquitous Computing*, *20*(5), 1–13. doi:10.1007/s00779-016-0940-x