

## Editorial

# Big Data and Knowledge Extraction for Cyber-Physical Systems

**Xiuzhen Cheng,<sup>1</sup> Yunchuan Sun,<sup>2</sup> Antonio Jara,<sup>3</sup> Houbing Song,<sup>4</sup> and Yingjie Tian<sup>5</sup>**

<sup>1</sup>The George Washington University, Washington, DC 20052, USA

<sup>2</sup>Beijing Normal University, Beijing 100875, China

<sup>3</sup>The University of Applied Sciences Western Switzerland (HES-SO), 3960 Sierre, Switzerland

<sup>4</sup>West Virginia University, Montgomery, WV 25136, USA

<sup>5</sup>Chinese Academy of Sciences, Beijing 100190, China

Correspondence should be addressed to Xiuzhen Cheng; [cheng@gwu.edu](mailto:cheng@gwu.edu)

Received 2 July 2015; Accepted 5 July 2015

Copyright © 2015 Xiuzhen Cheng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cyber-Physical Systems (CPS) have come into our view and will be applied in our daily life and business process management. Emerging CPS are coordinated, distributed, and connected and must be robust and responsive. Tomorrow's CPS need to far exceed the systems of today in capability, adaptability, resiliency, safety, security, and usability. Big data are large, complex, or rapidly generated datasets that cannot be processed by traditional technologies. Decision makers, resource managers, engineers, first-responders, scientists, and citizens are faced with a multitude of constantly flowing data streams coming from many sources in many formats every day. Understanding the volumes of big data requires cutting-edge tools and techniques that can analyze and extract useful knowledge from vast and diverse streams of information. The wonderful living of human beings and the high efficiency of business rely mostly 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 and the physical world.

This special issue aims to explore the above challenges through papers that address (1) big data analytics and information processing in CPS, (2) knowledge extraction from the big sensing data in CPS, and (3) other issues (such as security and privacy) relevant to big data and knowledge processing in CPS. We received 60 submissions from both academia and industry. Each paper was peer-reviewed by at least two experts in the field. In the following, we provide a brief introduction to each accepted paper.

The paper “Data Mining for the Internet of Things: Literature Review and Challenges” by F. Chen et al. presents a systematic overview on data mining in view of knowledge, technique, and application, including classification, clustering, association analysis, time series analysis, and outlier analysis. Some new application cases and new algorithms are also reviewed. Challenges and open research issues in this field are discussed and a big data mining system is suggested.

The paper “A Phoenix++ Based New Genetic Algorithm Involving Mechanism of Simulated Annealing” by L. Hu et al. provides a novel MapReduce enabled simulated annealing genetic algorithm to increase the efficiency of the high-performance distributed computing platforms, which are characterized by the synthesis of the genetic algorithm and the simulated annealing algorithm and parallelism. The experiments on Phoenix++ indicate that the convergence speed of the proposed algorithm outperforms the traditional ones significantly.

The paper “Privacy Information Security Classification for Internet of Things Based on Internet Data” by X. Lu et al. focuses on studying the attributes of the privacy and proposes the Privacy Information Security Classification (PISC) model. The privacy issues are classified into four security categories by PISC, and each category has its own security goal.

The paper “A Novel Dynamic Weight Neural Network Ensemble Model” by K. Li et al. proposes a novel dynamic weight neural network ensemble model (DW-NNE). The

bagging algorithm generates certain neural network individuals that are selected by the  $k$ -means clustering algorithm.  $K$ -value optimization algorithm is put forward based on a distance cost function to find the optimal  $K$ -values. It also proposes a dynamic weight model based on fuzzy neural network in accordance with the ideas of dynamic weight.

The paper “A Reliable Broadcast Protocol in Vehicular Ad Hoc Networks” by C. Wang et al. proposes a MAC protocol based on the  $s$ -disjunct code to achieve reliable and real-time broadcasts via assigning a channel to each vehicle as the vehicle enters the network, which is adaptive as vehicles are required to adjust their communication ranges according to the network density in order to avert the interference among them. Even at a high vehicle density, the proposed scheme guarantees high success rates.

The paper titled “Online Optimization of Collaborative Web Service QoS Prediction Based on Approximate Dynamic Programming” by X. Luo et al. presents an algorithm that can incorporate approximate dynamic programming- (ADP-) based online parameter tuning strategy into the QoS prediction approach. The prior knowledge or identification of the prediction model is not required as the proposed approach can reach the QoS prediction with the automatic parameter tuning capability.

The paper “Device-to-Device Users Clustering Based on Physical and Social Characteristics” by L. Wang et al. proposes a novel method for device-to-device (D2D) user clustering that allows wireless users in proximity to share common resources to save both system bandwidth and energy resources, based on two proposed clustering approaches: the Chinese Restaurant Process (CRP) and the distance-dependent Chinese Restaurant Process (DCRP). Numerical simulation results show the superiority of the proposed clustering schemes in terms of energy consumption and energy efficiency.

The paper “An Approach for Prediction of Acute Hypotensive Episodes via the Hilbert-Huang Transform and Multiple Genetic Programming Classifier” by D. Jiang et al. presents a methodology to predict AHE for ICU patients based on big data time series. The experimental data is Mean Arterial Pressure (MAP) transformed from the Arterial Blood Pressure (ABP) data. The methodology is applied in the datasets of the 10th PhysioNet and Computers Cardiology Challenge in 2009 and Multiparameter Intelligent Monitoring for Intensive Care (MIMIC-II). The accuracy of 83.33% in the training set and 91.89% in the testing set of the 2009 Challenge’s dataset and 84.13% in the training set and 82.41% in the testing set of the MIMIC-II dataset are achieved.

A scheme using Extended Service Set (ESS) based architecture is presented to implement the proxy mobile IPv6 protocol, that is, PMIPv6, for IEEE 802.11 Infrastructure Wireless Local Area Networks (WLANs) in the paper titled “Implementing PMIPv6 Protocol Based on Extended Service Set for IEEE 802.11 Infrastructure WLAN” by D. Zhu et al. The key signaling packets together with their time sequences for the mobility management in the proposed scheme are proposed. Numerical analysis indicates the proposed scheme outperforms the existing one that uses the Basic Service Set (BSS) based architecture in terms of handoff delay in the case

when the delay between Mobile Access Gateway (MAG) and Local Mobile Anchor (LMA) is relatively large.

The paper “New Benchmarking Methodology and Programming Model for Big Data Processing” by A. Kos et al. presents a possible solution for the coming exascale big data processing: a data flow computing concept. A new benchmarking methodology is proposed, which integrates the performance issues of speed, area, and power needed to execute the task. According to the authors, the next step of data flow computing development should be a move from specialized to more general algorithms and applications.

The paper “Context-Aware Recommendation via Graph-Based Contextual Modeling and Postfiltering” by H. Wu et al. proposes a graph-based framework to model and incorporate contextual information into the recommendation process in an advantageous way to help the development and use of context-aware recommendation capabilities. A contextual graph-based relevance measure (CGR) is specifically designed to assess the potential relevance between the target user and the items further used to make an item recommendation. A probabilistic-based postfiltering strategy is proposed to refine the recommendation results. Experimental results show the superiority and the effectiveness in context-aware recommendation scenario.

The paper “Waypoint Graph Based Fast Pathfinding in Dynamic Environment” by W. Zhu et al. proposes a fast approach for waypoint graph-based pathfinding. Unnecessary waypoints and edges are eliminated to make the graph sparse. A prediction-based local method is also designed to handle the dynamic change in the environment. Extensive simulation results show the proposed approach outperforms existing ones.

The paper “Model-Based Sensitivity Analysis on Aerosol Optical Thickness Prediction,” by B. Han et al., proposes a support vector regression (SVR) model-based sensitivity analysis approach to order 35 MODIS input attributes according to their sensitivity to prediction outputs. The attribute sensitivity orders are used for feature selection in the context of regression by removing insensitive attributes one at a time or by removing attributes whose sensitive orders are larger than a number  $k$ . Experimental results show that the top 10 insensitive attributes can be screened to speed up prediction model computation with very little loss of accuracy. The proposed approach would be useful for remote sensing scientists or atmospheric scientists to optimize the design precision of top sensitive attributes in scanning equipment like MODIS to improve the AOT retrieval accuracy.

The paper “Enterprise-Oriented IoT Name Service for Agricultural Product Supply Chain Management” by Y. Liu et al. presents a simple and efficient enterprise-oriented name service, namely, *iotNS*, which is designed for agricultural products with faster response than the standard GSI ONS system. The *iotNS* has been deployed across five cities to enable the efficient storage and retrieval of information about agricultural products as they are tracked from the farm to the market. The deployment shows that *iotNS* is a mechanism that fosters cooperation and supply chain improvements among the enterprises using the *iotNS*.

The paper “A Dynamic Processing System for Sensor Data in IoT” by M. Li et al. proposes a dynamic sensor data processing (SDP) system to capture and process sensor data continuously on the basis of data streaming technology according the characteristics of the sensor data in IoT, which uses Particle Swarm Optimization (PSO) algorithm to train threshold dynamically for data compression avoiding redundancy. The experimental results show that SDP can compress sensor data through dynamically balancing the accuracy and compression rate.

The paper “Enhanced Asymmetric Bilinear Model for Face Recognition” by W. Gong et al. proposes an enhanced asymmetric model for illumination robust face recognition which initializes the factor probabilities with nearest neighbor method and optimizes it for the test data. The experiment results show that the enhanced asymmetric model can reach a high recognition accuracy.

Addressing MAC protocol for wireless network based on ultrawide band (UWB) to solve the problems of real-time services and reliability in wireless transmitting, the paper titled “The Study on Media Access Control Protocol for Wireless Network in Library” by W. Zeng et al. proposes a multichannel MAC protocol (UWBIMAC) based on status-adaptiveness to choose the channel mechanism. Simulation results show that the protocol has the ability of stronger anti-jamming, improves the network throughput, and provides better average delay performance than IEEE 802.15.4MAC, IEEE 802.11MAC, and other UWB MAC protocols.

The paper “Minimized Cost Gateway Deployment in Cyber-Physical Systems” by H. Wang et al. addresses the problem of gateway placement for satisfying the bandwidth-requirement of each node by using minimum gateways which can be formulated as a variant of Minimum Geometric Disk Cover problem. A heuristic gateway placement algorithm and one grid-based heuristic algorithm are proposed for this NP-complete problem. The result of simulation demonstrates that the heuristic algorithm can offer a good solution with big probability.

The paper “A Genetic-Algorithm-Based Approach for Task Migration in Pervasive Clouds” by W. Zhang et al. proposes a genetic-algorithm- (GA-) based approach that is effective in addressing multiobjective optimization problems. The preliminary evaluations show quite promising results by using one of the classical genetic algorithms and the GAs can be used for decision making in task migrations in pervasive clouds.

The paper “Development and Validation of a Portable Human Body Joint Power Test System” by L. Li et al. presents a portable human body joint power test system using inertial sensor technology and wireless Bluetooth acquisition technology. The test system can be applied into many areas such as athlete selection and daily strength training.

A two-dimension code spatiotemporal modeling method is proposed in the paper “Two-Dimensional Code Based Spatiotemporal Modeling and its Application in Object Tracing” by Y. Gao et al. Two-dimension code is used to store object’s characteristics in the smartphone to obtain the object’s real-time spatiotemporal data. A use case to

trace object’s spatiotemporal information is also developed to validate the proposed modeling method.

Due to the imbalanced energy consumption among nodes in wireless sensor networks, some nodes die prematurely, which decreases the network lifetime. Addressing this problem, the paper “IDUC: An Improved Distributed Unequal Clustering Protocol for Wireless Sensor Networks” by C. Chen et al. proposes an improved distributed unequal clustering protocol (IDUC) for wireless sensor networks, where nodes are energy heterogeneous and scattered unevenly. The cores of IDUC are the formation of unequal cluster topology and the construction of intercluster communication routing tree. Compared with previous protocols, IDUC is suitable for various network scenarios, and it can balance the energy consumption more efficiently and extend the lifetime of networks significantly.

The paper “Vehicular Ad Hoc Networks: Architectures, Research Issues, Methodologies, Challenges, and Trends” by W. Liang et al. provides an overview of the main aspects of VANETs from a research perspective including the architecture, research issues and general research methods, and challenges and future trends of VANETs.

In the paper “Identifying Missing and Spurious Interactions in Directed Networks” by X. Zhang et al., the authors make use of classical link prediction indices for undirected networks, adapt them to directed version which could predict both the existence and direction of an arc between two nodes, and investigate their prediction ability on six real-world directed networks. Experimental results demonstrate that those modified indices perform quite well in directed networks. Compared with bifan predictor, some of them can provide more accurate predictions.

The paper “A Multimodel Based Range Query Processing Algorithm for Information Collection in CPS” by G. Li et al. proposes a multimodel based range query processing algorithm, which utilizes multiple probability models to depict the data distribution of a sensor node. Compared with the single model-based algorithm, the data distribution depicted by the multimodel based algorithm is much more accurate than that of the single one. Experiments on real dataset show the efficiency of the proposed model.

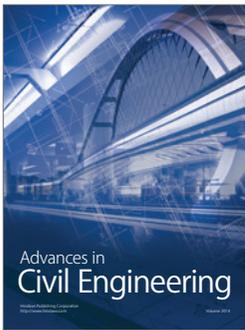
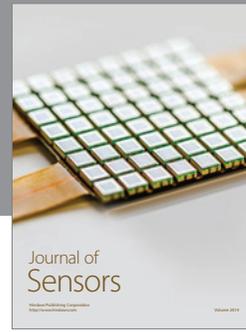
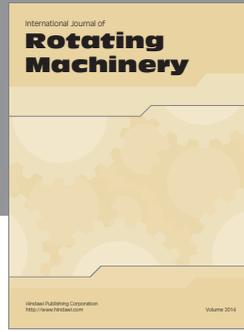
The paper “Deadline Aware Retransmission Threshold Setting Protocol In Cyber-Physical Systems” by R. Bi et al. investigates the problem in finding the optimal retransmission thresholds for the relay nodes along given delivery path, such that the summation of the probability of the packet being transmitted to the next relay node or destination node within the specified deadline is maximized. A distributed greedy algorithm is proposed for computing the optimal retransmission threshold. Experimental results show that the proposed protocols have better performance in terms of deadline success ratio and real-time ratio.

Addressing real-time transmitting in CPS, the paper “Handling Interservice Time Constraints in Wireless Networks” by X. Zheng et al. designs an interservice time guaranteed scheduling which defines a new capacity region of networks with a strict interservice time and a novel scheduling policy. Simulation results show the framework performs well in interservice time and throughput.

## **Acknowledgments**

We would like to thank all authors for their valuable contributions to this special issue. We are indebted to the journal editors and all anonymous reviewers for their hard work that helps the authors further enhance the quality of the manuscripts. It is also an honor for us to serve as the Guest Editors of this issue.

*Xiuzhen Cheng  
Yunchuan Sun  
Antonio Jara  
Houbing Song  
Yingjie Tian*



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

