Guest Editorial: Advanced Complex Data Analytics for Smart City Industrial Environment

WITH the continuous increase in size of populations living in cities, those residents face increasing environmental pressures and infrastructure needs. To deliver a better quality of life for these residents to address these demands at a sustainable cost, smart technologies can help cities meet these challenges, and have been become the next wave of public investment. It all starts with data to be generated by the residents living in the cities. This special section will focus on the ever-increasing challenges of big complex data like social network data, traffic network data, and IoT network data altogether in the industrial environment of smart city. Finding the insights in all that data helps municipal governments respond to fluid situations, allocate resources wisely, and plan for the future. Furthermore, putting real-time information into the hands of individuals and companies empowers them to make better decisions and play a more active role in shaping the city's overall performance. As cities get smarter, they become more livable and responsive.

The smart city data processing technology is experiencing revolutionary changes in each stage, including data collecting, cleaning, organizing, interpreting, analytics, utilizing, and visualization. Those changes lead to a globally noticeable development trend of the convergence with big data frameworks, network analytical modeling, link or route prediction, recommendation systems, IoT, and edge computing. This special section aims at providing a forum to present recent advancements in the convergent research about big, complex, and domain-cross information network data within the smart city environment. Based on the reviewers' feedbacks as well as the evaluations of editors, 8 articles are selected in this special section from 53 submissions. The selected papers that cover broad topics are introduced briefly as follows.

The article "Deep Reinforcement Learning for Smart City Communication Networks" by Xia *et al.* proposed a new end-to-end congestion control protocol for smart city communication networks. The novel protocol supports high throughput, low queuing delays, and low packet loss for application services in a smart city network. When congestion is high, agents in a deep reinforcement learning network find the optimal policy for adaptively adjusting the congestion window according to the prevailing network conditions. Furthermore, a fast growth algorithm rapidly increases the congestion window at the network initialization stage to utilize free bandwidth, maximize throughput, and drastically speed up the learning process. By comparing with the other six protocols, the proposed protocol can make the best tradeoff between throughput, delay, and loss rate.

The article "A Trust-based Security System for Data Collecting in Smart City" by Fang *et al.* proposed a novel trust model by utilizing binomial distribution (BTM), which can better conform with the rule of trust, this is, "hard to get, easy to lose." Based on BTM, we design an efficient scheme to identify and resist a typical internal attack—ON–OFF attack, called O2TMS, which takes the time-varying characteristics of the wireless channel in the data collection stage into consideration. We devise a trust value-based secure routing protocol (TV-SRP), which combines other decision factors to solve a security-transmission–energy trade-off. These three components can construct a trust-based security system to guarantee data collecting security in the smart city industrial environment. The more objective trust value is generated in BTM to be used in O2TMS for preventing ON–OFF attacks, to be also introduced into TV-SRP for selecting the secure next-hop nodes, so that sensed data are collected and aggregated in security—Experiments demonstrate that our schemes have better performance than others.

The article "Recommendation by Users' Multi-modal Preferences for Smart City Applications" by Xu *et al.* proposed to explicitly mine users' view-level visual preference from their reviews and show that this idea can significantly improve the performance of recommendation. In addition, they developed a novel deep learning method, UMPR, together with proper training strategies to implement the proposed idea, which can well capture textual and visual matching features between users and items for recommendation. They applied the proposed novel model on two real applications related to smart city such as restaurant recommendation and product recommendation, and demonstrated the competitive performance of the model by comparing with the existing methods.

The article "Efficient and Exact Multi-graph Matching Search" by Pang *et al.* proposed a definition of exact multigraph matching by combining the definitions of the graph isomorphism and multiset equality. Graph models have been widely used to represent the city-related data, such as road network and social network. To solve the problem of exact multigraph matching search, they designed an efficient algorithm under the filtering-and-verification framework. To accelerate the filtering irrelevant results, they proposed multiple filtering conditions based on which an offline and multilayer inverted index is constructed to further improve the query speed. Thus, the basic

Digital Object Identifier 10.1109/TII.2020.3033855

detection approach can be optimized using the filtering results. The time complexity of all the above algorithms is analyzed and the extensive experiments have been conducted to verify the performance of the proposed techniques.

The article "Privacy aware Data Fusion and Prediction with Spatial Temporal Context for Smart City Industrial Environment" by Qi *et al.* proposed a novel privacy-aware data fusion and prediction approach for smart city industrial environment, which can balance user privacy, data availability, and time efficiency well. This novel solution is particularly suitable for the smart city service decision-makings (e.g., intelligent tourism planning and smart traffic scheduling) that are based on multisource industrial data with sensitive spatial–temporal context information and fragmented across different platforms. It makes the technical contribution from two sides. On the one hand, at the edge layer, they employ locality-sensitive hashing to help edge servers to transform the big volume of user-service1 quality data from user layer (containing sensitive spatial–temporal contexts) into a small set of item indices with little privacy. Only the item indices are transmitted to the cloud center for further processing; this way, the transmitted data amount is reduced and user privacy is secured accordingly. On the other hand, at the cloud layer, all the item indices produced by edge servers are gathered by a central cloud platform, according to which a recommender system hosted in the cloud platform is used to discover the similar items and use them for further data analysis, prediction, and recommendation. This way, all the less-sensitive item indices (from edge servers) take part in the centralized cloud decision-makings so as to ensure the accuracy of prediction and recommendation. A set of experiments are enacted based on a real user service rating dataset containing spatial–temporal context. Reported experimental results compared to other solutions indicate better performances of their solution, in the aspects of time cost and accuracy while securing user privacy.

The article "Modelling of Extreme Vulnerability Disclosure in Smart City Industrial Environments" by Tang *et al.* investigated the extreme vulnerability disclosure events based on monthly vulnerability disclosure time series data. Specifically, they proposed a rigorous statistical framework to model and predict extreme vulnerability disclosure events. The proposed model leverages the peak over threshold method in extreme value theory to model the distributions of extreme and rare vulnerability disclosure events. Different from existing works, they mainly addressed an important and practical cyber trust and risk problem in the domain of modeling the probability of extreme vulnerability disclosure events and its probable extreme values in the future. This is the pioneering work to build quantitative models fitting the extreme vulnerability disclosure events and answering some of the key cyber-risk management questions, such as what is the probability that the maximum vulnerability disclosures for the next year will exceed a level x? What is the expected length of time (e.g., in years) before the occurrence of a specific high quantity of vulnerability disclosures? What is the probability that the maximum vulnerability disclosure for the next year exceeds the maxima of all previous years? Therefore, this article focuses on modeling the extreme vulnerability disclosure incidents in Smart City industrial environment, from the cybersecurity perspective.

The article "Adaptive Non-Convex Sparsity based Background Subtraction for Intelligent Video Surveillance" by Li *et al.* proposed a generalized shrinkage thresholding operator (GSTO) by noticing the adaptive ability of p-shrinkage thresholding operator and the advantages of soft/hard thresholding operators. Thus, GSTO produces adaptive shrinkage performance that perfectly matches the compositions of singular values in surveillance system. Besides, GSTO makes the shrinkage start line and the shrinkage bias baseline that are fixed in p-shrinkage thresholding operator, controllable. In addition, they constructed a new nonconvex sparsity function as the objective function of GSTO. It can directly act on the singular values and no convex relaxation is required. Thus, more information is saved when building up an effective intelligent surveillance system. Experimental results show that the proposed model is more appropriate for the unchanged video background. Another main contribution of this article is to design a residual purifying operation for the self-repeating detail textures in the background, which is particularly essentially in the framework as the shrinkage operation will drop all the dynamical issues in the background. The employed dictionary learning based operation can effectively structure the principle atoms of frame patches. Experimental results have shown that, as to foreground detection task, when compared against current state-of-the-art techniques, the proposed model achieves comparable and often superior performance in terms of F-measure scores in most cases.

The article "Data Augmentation based Cellular Traffic Prediction in Edge Computing Enabled Smart City" by Wang *et al.* proposed the ctGAN-S2S model to effectively achieve data augmentation and improve the cellular traffic prediction performance. This new model can protect privacy by using the data augmentation model to generate close-to-real cellular traffic data, which is used to reduce direct data sharing between base stations for training the prediction model. Moreover, they also developed a cellular traffic data augmentation submodel ctGAN, where the potential time-series samples of target cellular traffic can be generated for data augmentation through the generative adversarial process. After that, they designed a long short-term memory (LSTM) based S2S prediction submodel, which can achieve flexible multistep prediction on time-series cellular traffic data. Finally, they evaluated the proposed ctGAN-S2S model on real-world cellular traffic dataset, and the results show that this new model achieves up to 48.49% improvement of prediction accuracy compared with four reference models.

The Guest Editors would like to thank all the authors who submitted their articles and anonymous reviewers who carefully reviewed and helped evaluate these articles. We would also like to extend our sincere thanks to the Editor-in-Chief of IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, Prof. Ren Luo, for providing this opportunity and for his guidance throughout the process, and the editorial staff for their continuous support on the special section.

JIANXIN LI, *Guest Editor*
School of Information Technology
Deakin University
Geelong, VIC 3220, Australia

YONG XIANG, *Guest Editor*
School of Information Technology
Deakin University
Geelong, VIC 3220, Australia

TIMOS SELLIS, *Guest Editor*
The Data Science Research Institute
Swinburne University of Technology
Hawthorn, VIC 3122, Australia

JIE XIONG, *Guest Editor*
College of Information & Computer Sciences
University of Massachusetts at Amherst
Amherst, MA 01003 USA

Jianxin Li received the Ph.D. degree in computer science from Swinburne University of Technology, Melbourne, VIC, Australia, in 2009.

He is currently an A/Professor of Data Science, and Director with the Smart Networks Lab, School of IT, Deakin University, Melbourne, VIC, Australia. He has authored/coauthored 90 high quality research papers in top international conferences and journals. His research interests include social computing, query processing, and optimization, and big data analytics.

Dr. Li is the Associate Editor for *Information Systems*, IEEE SIGNAL PROCESSING LETTERS, and the Editor-in-Chief of the *International Journal of Knowledge Science & Engineering*. His research has been funded by Australia Research Council Discovery Project, Linkage Project, and several industry grants.

Yong Xiang (Senior Member, IEEE) received the Ph.D. degree in electrical and electronic engineering from the University of Melbourne, Melbourne, VIC, Australia, in 2003.

He is a Professor with the School of Information Technology, Deakin University, Melbourne, VIC, Australia. He has authored/coauthored five monographs, more than 160 refereed journal articles, and numerous conference papers in these areas. His research interests include information security and privacy, signal and image processing, data analytics and machine intelligence, Internet of Things, and blockchain.

Prof. Xiang was invited to give keynote speeches at a number of international conferences. He is the Senior Area Editor of the IEEE SIGNAL PROCESSING LETTERS and the Associate Editor of the IEEE Communications Surveys and Tutorials. He was the Associate Editor and Guest Editor of several reputed international journals. He served as Honorary Chair, General Chair, Program Chair, TPC Chair, Symposium Chair, and Track Chair for many international conferences.

Timos Sellis (Fellow, IEEE) received the Ph.D. degree in computer science from the University of California, Berkeley, CA, USA, in 1986.

He is currently a Research Scientist with Facebook (USA) and an Adjunct Professor with Swinburne University of Technology, Melbourne, VIC, Australia. Till the end of 2012, he was the Director with the Institute for Management of Information Systems and a Professor with the National Technical University of Athens, Athens, Greece. Between 2013 and 2015, he was a Professor with RMIT University, Melbourne, VIC, Australia. His research interests include big data, data streams, personalization, data integration, and spatiotemporal database systems.

Prof. Sellis is a Fellow of the ACM.

Jie Xiong received the Ph.D. degree in computer science from the University College London, London, U.K., in 2015.

He is currently an Assistant Professor with the College of Information and Computer Sciences, University of Massachusetts Amherst, Amherst, MA, USA. His research interests include wireless sensing, mobile health, and wireless networking.

Dr. Xiong was the recipient of a Google European Doctoral Fellowship and a British Computer Society Distinguished Dissertation Award runner-up. His recent work received several awards including the MobiCom 2020 Best Paper Honourable Mention, SenSys 2019 Best Paper candidate, and CoNEXT 2014 Best Paper.