Guest Editorial: Special Section on Emerging Trends Issues and Challenges in Edge Artificial Intelligence

EDGE computing has the advantages of real-time response and less network demand for computing closer to the edge of the network, while bridging the physical and digital worlds. The core of the edge computing is to provide the edge intelligent service. Therefore, edge artificial intelligence is becoming a popular trend for the future, such as intelligent sound box, and so on. Edge artificial intelligence combines edge computing with artificial intelligence, while taking both advantages.

However, there are some problems, which need to be solved for the edge artificial intelligence.

1) Artificial intelligence generally requires high-performance computing, especially for training, but, normally, the edge nodes have limited computing power. Therefore, it is important to manage resources in edge artificial intelligence for providing enough computing power.
2) Data placement, in the artificial intelligence of cloud computing, all data are placed in the cloud. But in the edge artificial intelligence, the data need to be migrated between cloud and edge for limited storage capacity in edge.
3) Edge computing is a typical distributed computing, how to solve the distributed models will be a difficult problem. Most of the models need to be distributed in the edge.
4) For providing the better edge intelligent service, more artificial intelligence applications need to be optimized in edge. Therefore, some new algorithms and methods for edge artificial intelligence needs to be proposed.
5) Privacy protection is more important in the edge circumstance for its easier to divulge personal information. Therefore, the edge artificial intelligence needs to pay more attention to privacy protection.

This special section aims to provide a platform for the communities to report recent findings and emerging research developments for addressing above-mentioned problems. Five papers have been selected for publication in this part of the section, following several rounds of reviews. They cover the range of subjects, from resources management for satisfying demands, new algorithms and methods for edge artificial intelligence applications, to new privacy protection policies for edge artificial intelligence.

Besides delay and energy, one critical performance indicator of an edge artificial intelligence service is the quality of result (QoR). The QoR can be represented by regression precision or classification accuracy. Moreover, the QoR is related to both algorithm complexity and input data. Zhang *et al.* propose a method to optimize energy and delay for computation offloading in edge artificial intelligence in this paper "MASM: A Multiple-Algorithm Service Model for Energy-Delay Optimization in Edge Artificial Intelligence." It proposes a multiple algorithm service model (MASM) specifically for edge artificial intelligence applications. Different VMs in the cloudlet are equipped with heterogeneous algorithms with different computation complexities and upload various data sizes in the MASM. The VM is chosen according to the demands of energy and time cost. Based on the MASM, an optimization model for minimizing the delay and the energy costs of the computation offloading process by jointly assigning proper workload assignment weights and the computing capacities of the VMs is proposed. Finally, a tide ebb algorithm (TEA) is developed for finding robust solutions to the proposed energy-delay optimization problem, while proving the Parato optimality of TEA.

Multiobjects tracking (MOT) have been proven usefully for its wide applications in visual surveillance, traffic safety, automatic video content analysis, and virtual reality. But pedestrians tracking is challenging for similar appearance with occlusions, intersected trajectories, missing data, and camera motion. Fortunately, tracking-by-detection (TBD) based MOT methods are the state-of-the-art method for developing pedestrian detectors. TBD tracking mainly has two kinds, the online tracking and the batch tracking. Both have one key issue, data associate, consisting of association affinity model, and association optimization model. To improve association optimization in MOT, Yang *et al.* propose an online MOT method, in which the association optimization problem is formulated into a two-stage strategy, i.e., local and global association stages, by relying on the rank-based association affinity model.

The automotive industry demands more personalized, integrated, and on-demand services including shared, connected, and autonomous environment, which is a typical edge application. However, one of the biggest issues in the automotive supply chain industry is counterfeiting products, which leads to dissatisfied customers and affects trust in the brand for poor quality. In order to effective plan of production capacity, track and trace of individual parts across the supply chain, and accurate and realtime information, Sharma *et al.* propose blockchain-based distributed framework model for the automotive industry

Digital Object Identifier 10.1109/TII.2019.2916875

in the smart city to build a sustainable platform for automotive ecosystem.

With the propagation of smart mobile devices, the position-based services (PBSs) have become progressively popular to deliver networked amenities based on roaming user's positions. For effective PBSs, extraction and recognition of meaningful positions and estimating the subsequent position are fundamental procedures. However, position-based real-time applications lack of concern. In the paper, "Enforcing Position-Based Confidentiality with Machine Learning Paradigm Through Mobile Edge Computing in Real-Time Industrial Informatics" proposes the method for conserving position confidentiality of roaming PBSs users using machine learning techniques. It makes use of prevalent machine learning techniques for implementation and execution of three-phase procedure (position extraction, position recognition, and position prediction) at the PBS intermediate server. Moreover, a mobile edge computing service policy is followed in the proposed paradigm, which ensures the timely delivery of PBSs, while offering position confidentiality and low latency by means of networking and computing services at the vicinity of roaming users.

In order to address the problem of the privacy-preserving frameworks being computation intensive, Lyu *et al.* propose a distributed approach, called fog-embedded privacy-preserving deep learning (FPPDL), to optimize the service capacity of fog computing for distributed deep learning. In the FPPDL, end nodes take the role of data contributors, who never upload original training data to both the nearby fog node and the cloud. Instead, each end node only forwards the transformed data using random projection to a nearby fog node. The intermediate fog nodes compute fog-level model gradients based on all the received data, then a fraction of gradients are forwarded to the cloud, which maintains the cloud-level model by aggregating fog-level updates, and only this update is communicated. So, FPPDL leverages the intermediate fog nodes, which exchange less sensitive gradients in a differentially private manner, without revealing more sensitive observation-level data.

Cloud computing provides a more cost-effective way to deploy scientific workflows, but each task of a scientific workflow requires several large datasets that are located in different datacenters, resulting in serious data transmission delays. Edge computing reduces the data transmission delays and supports the fixed storing manner for scientific workflow private datasets, but there is a bottleneck in its storage capacity. Therefore, one challenge is emerged, which is to combine the advantages of both edge computing and cloud computing to rationalize the data placement of scientific workflow, and optimize the data transmission time across different datacenters. Lin *et al.* propose a self-adaptive discrete particle swarm optimization algorithm with genetic algorithm operators (GA-DPSO) to optimize the data transmission time when placing data for a scientific workflow. The proposed GA-DPSO considered the impact factors on the transmission delay, such as the bandwidth between datacenters, the number of edge datacenters, and the storage capacity of edge datacenters. The crossover and mutation operators of the genetic algorithm were adopted to avoid the premature convergence of traditional particle swarm optimization algorithm, which enhanced the diversity of population evolution and effectively reduced the data transmission time.

Real-time feature of data flow of the whole industrial system, directly affecting the production efficiency and normal operation of the system, plays a critical role in a smart factory and Industry 4.0. Therefore, reducing the overhead of the data processing and transmission in computing-extensive tasks [e.g., Artificial Intelligence (AI) and deep learning tasks] is another aspect to guarantee the real-time performance. The current latency-constraints methods concentrate on the network optimizations, data fusion, computing-task simplification, and computing resource decentralization. Edge/fog computing frameworks, close to producing equipment consequently leading to the decrement of the latency for data communications between servers and machines, are good candidate strategies for smart manufacturing to tackle the above-mentioned problems. However, some challenges we still need to face: 1) how to propose a computing system to handle and integrate the historical heritage of computing resources; and 2) how to construct some novel and efficient strategies and algorithms to ensure the real-time performance. Li *et al.* propose a hybrid computing framework and design an intelligent resource scheduling strategy to fulfill the real-time requirement in smart manufacturing with edge computing support. First, a four-layer computing system in a smart manufacturing environment is provided to support the artificial intelligence task operation with the network perspective. Then, a two-phase algorithm for scheduling the computing resources in the edge layer is designed based on greedy and threshold strategies with latency constraints. Finally, a prototype platform was developed.

The challenging issue in the power hungry, short battery lifetime, and delay-intolerant portable devices is inappropriate and inefficient classical trends of fair resource allotment. Moreover, it is interpreted through extensive industrial datasets that dynamic wireless channel could not be supported by the typical power saving and battery lifetime techniques, for example, predictive transmission power control (TPC) and baseline. Therefore, Sodhro *et al.* propose the following conditions.

1) A forward central dynamic and available approach (FCDAA) by adapting the running time of sensing and transmission processes in Internet of Things (IoT) based portable devices.
2) A system-level battery model by evaluating the energy dissipation in IoT devices.
3) A data reliability model for edge AI-based IoT devices over hybrid TPC and duty-cycle network.

The proposed FCDAA tunes transmission power level and duty cycle of IoT devices in industrial applications by adopting static (product processing) and dynamic (vibration and fault diagnosis) platform at acceptable reliability or packet loss ratio.

Nowadays, smart sensors equipped with face, gait, or voice recognition modules and machine learning algorithms are prevalent. However, confidentiality of biometric traits in these systems cannot be met because the biometric traits used in these systems are unprotected and usually exposed outside human

bodies; for example, our faces can be captured easily by someone malicious. In contrast to some commonly used biometric identifiers, e.g., fingerprint and face, finger vein is relatively robust in the sense that it is located inside a finger and immune to forgery. The finger-vein pattern can only be acquired by using an infrared camera in which the infrared light is absorbed by the hemoglobin in the blood inside the finger vein. As a result, finger veins are shown as dark lines in the image. Because of finger vein's high uniqueness, stability, and difficulty of being forged, finger vein based biometric recognition systems have drawn an increasing attention in recent years. However, the existing machine learning based methods do not consider the security aspect of the biometric traits used in the authentication system. To address the issue, Yang *et al.* develop a novel biometric template protection algorithm using the binary decision diagram (BDD) for deep learning based finger-vein biometric systems. The proposed algorithm is capable of creating a new noninvertible version of the original finger-vein template, which is stacked with an artificial neural network. The multilayer extreme learning machine (ML-ELM) to generate a privacy-preserving finger-vein recognition system, named BDD-ML-ELM. The proposed BDD-ML-ELM ensures the safety of the original finger-vein template even if its transformed version is compromised.

Facing the huge amount of data, recommender systems are increasingly playing an important role in our daily life. Su *et al.* propose a novel user-centered recommendation strategy for cultural items suggestion with the help of edge intelligence in the paper "An Edge Intelligence Empowered Recommender System Enabling Cultural Heritage Applications." After analyzing the challenges and drawbacks of previous recommender systems, Su *et al.* design a new big data infrastructure for the management of cultural items, which offers APIs to easily create new applications in a multilayer architecture way including query engine, semantic search, and machine learning module. Besides, to improve the recommendation of cultural contents, the paper proposes to better understand user using machine learning algorithms before prefiltering the potential candidates. In order to comprehensively evaluate the proposed idea, the authors build an edge intelligent Smart Search Museum application to perform semantic searches and ML-based inference.

As we enter the era of Industry 4.0, the smart factory can autonomously make decisions on complex and real-time manufacturing, maintenance, and logistics decisions with the help of advanced ICT technologies. Semiconductor manufacturing, due to its high investment and high payoff, has been one of the most competitive industries in the world. Semiconductor manufacturing can be referred to as a job shop (JSP) industry with the long manufacturing process, which is a classical NP-hard problem with numerous applications. Lin *et al.* propose a smart semiconductor manufacturing factory based on an edge computing framework and solve the JSP problem using the modified deep Q network algorithm, which combines deep learning and reinforcement learning. The experimental results show that the novel approach can address the decisions of multiple edge devices in an efficient way.

Edge computing provides high-class intelligent services and computing capabilities at the edge of the networks to ease the backhaul impacts and offer an improved user experience. However, the edge artificial intelligence exacerbates the security of the cloud computing environment due to the dissociation of data, access control, and service stages. Lateral movement techniques are frequently used to launch cyber attack, especially in the hierarchical architecture system. But the following two challenges make the existing approaches hard to apply in edge environment: Since the edge nodes are often the memory limited and computationally constrained devices, the traditional lateral movement detection methods that need a significant manual effort and business correlated knowledge are not suitable in edge-cloud environment. Besides, the lateral movement detection methods that rely on detecting changes in node behaviors may not be applied successfully on the edge environment due to its dynamic nature. Hence, Tian *et al.* propose CloudSEC, a lateral movement detection method for the edge-cloud computing environment, to supply a complete and credible evidence chain as well as supports the capacity of real-time lateral movement reasoning to enhance the system security.

Real-time data processing applications demand dynamic resource provisioning and efficient service discovery, which is particularly challenging in resource constraint edge computing environments. Network embedding techniques can potentially support effective resource discovery services by achieving a proximity-preserving representation of the network resources. However, most of the existing techniques fail to capture accurate proximity information among the network nodes and further lack exploiting information beyond the second-order neighborhood. The paper "An Inductive Content-Augmented Network Embedding Model for Edge Artificial Intelligence" proposes inductive content augmented network embedding (ICANE) model, to preserve higher order structural and semantic content proximity in large-scale decentralized networks leveraged by artificial intelligence techniques. ICANE can effectively learn the embedding function to generate a low-dimensional vector representation of complex networks in an inductive way. Moreover, a semantic proximity search method has been proposed to locate the top-k relevant nodes using the learned network representation.

With the rapid development of mobile phones, IoT, and mobile terminal devices, network resources are gradually insufficient and the burden on existing cellular networks is increasing. To improve the efficiency of network resources, the cloud of radio access network (C-RAN) architecture is considered in the fifth generation mobile networks standard. It splits the traditional base station into two parts, including the remote radio head and the baseband unit. For resource allocation in C-RAN, accurately predicting traffic is very important, which can reduce the number of migrations of C-RAN resources as well as transmission costs and energy costs. However, most of the existing methods for predicting traffic use statistical analysis, which cannot comprehensively consider global factors. The paper "Dynamic Resource Prediction and Allocation in C-RAN with Edge Artificial Intelligence" formally defines the resource

allocation problem in the C-RAN. Besides, to further improve the performance, the long short-term memory is used to predict dynamic throughput and GA-based resource allocation algorithm is used to optimize resource allocation.

In this part of the special section, the selected articles provide interesting approaches and promising views regarding the frameworks and modeling. We hope the readers benefit from the different perspectives in this part of special section and it will contribute to this frontier but fundamental research area.

The Guest Editors would like to thank the Authors for their contributions and the Reviewers for their valuable time. We are grateful to the Editor-in-Chief, Prof. R. C. Luo for the acceptance of the special section proposal. Warm thanks are extended to the Administrators of the journal, for their professional support of the full stages leading to publication.

G. HAN, *Guest Editor*
Department of Information and Communication Systems
Hohai University
Changzhou 213022, China

M. GUIZANI, *Guest Editor*
Department of Electrical and Computer Engineering
University of Idaho
Moscow, ID 83844-4332 USA

G. JIA, *Guest Editor*
Department of Computer Science
Hangzhou Dianzi University
Hangzhou 310018, China

J. LLORET, *Guest Editor*
Department of Communications
Universitat Politècnica de València
46022 València, Spain

Guangjie Han (S'03–M'05–SM'18) received the Ph.D. degree in computer application from Northeastern University, Shenyang, China, in 2004.

He is currently a Professor with the Department of Information and Communication System, Hohai University, Changzhou, China, and a Distinguished Professor with the Dalian University of Technology, Dalian, China. From 2004 to 2005, he was a Product Manager for the ZTE Company. From 2005 to 2006, he was a Key Account Manager for the Huawei Company. In 2008, he finished his work as a Postdoctoral Researcher with the Department of Computer Science, Chonnam National University, Gwangju, South Korea. From 2010 to 2011, he was a Visiting Research Scholar with Osaka University, Suita, Japan. From 2017, he was a Visiting Professor with the City University of Hong Kong, China. He has authored more than 330 papers published in related international conference proceedings and journals, including the IEEE COMST, IEEE TII, IEEE TMC, IEEE TVT, IEEE TIE, IEEE TPDS, IEEE TETC, IEEE IoT Journal, IEEE TETCI, IEEE TCC, IEEE Systems, IEEE Sensors, IEEE Wireless Communications, IEEE Communications, IEEE Network, etc, and is the holder of 125 patents. Currently, his H-index is 32 and i10-index is 91 in Google Citation (Google Scholar). His current research interests include Internet of Things, industrial Internet, machine learning and artificial intelligence, mobile computing, security, and privacy.

Dr. Han was a Co-Chair for more than 50 international conferences/workshops and as a Technical Program Committee Member of more than 150 conferences. He has served on the Editorial Boards of up to 16 international journals, including the IEEE JSAC, IEEE Network, IEEE Systems, IEEE ACCESS, IEEE/CCA JAS, *Telecommunication Systems*, etc. He has guest edited a number of special issues in IEEE journals and magazines, including the IEEE Communications, IEEE Wireless Communications, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, COMPUTER NETWORKS, etc. He has served as a Reviewer of more than 60 journals. He was awarded the IEEE International Conference on Computing, Management and Telecommunications (ComManTel), IEEE Conference on Computing, Communications and IT Applications (ComComAP), and EAI International Conference on Communications and Networking in China (Chinacom) in 2014, and International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (Qshine) Best Paper Awards in 2016.

Mohsen Guizani (S'85–M'89–SM'99–F'09) received the bachelor's (with distinction) and master's degrees in electrical engineering, and the master's and doctorate degrees in computer engineering from Syracuse University, Syracuse, NY, USA, in 1984, 1986, 1987, and 1990, respectively.

He is currently a Professor of Electrical and Computer Engineering and the Electrical and Computer Engineering (ECE) Department Chair with the University of Idaho. Previously, he was the Associate Vice President of Graduate Studies, Qatar University, Chair of the Computer Science Department, Western Michigan University, Chair of the Computer Science Department, University of West Florida. He was also in academic positions with the University of Missouri-Kansas City, University of Colorado-Boulder, Syracuse University, and Kuwait University. He has authored nine books and more than 400 publications in refereed journals and conferences. His research interests include wireless communications and mobile computing, computer networks, mobile cloud computing, security, and smart grid.

Prof. Guizani is currently with the Editorial Boards of several international technical journals and the Founder and the Editor-in-Chief of the *Wireless Communications and Mobile Computing* (Wiley). He Guest Edited a number of special issues in IEEE journals and magazines. He was a member, chair, and the general chair of a number of international conferences. He was selected as the Best Teaching Assistant for two consecutive years at Syracuse University. He received the Best Research Award from three institutions. He was the Chair of the IEEE Communications Society Wireless Technical Committee and the Chair of the Transmission, Access, and Optical Systems (TAOS) Technical Committee. He served as the IEEE Computer Society Distinguished Speaker from 2003 to 2005.

Gangyong Jia (M'13) received the Ph.D. degree in computer architecture from the Department of Computer Science, University of Science and Technology of China, Hefei, China, in 2013.

He is currently an Associate Professor with the Department of Computer Science, Hangzhou Dianzi University, Hangzhou, China. He has authored and coauthored more than 50 papers in related international conferences and journals, such as IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE TRANSACTIONS ON CLOUD COMPUTING, IEEE Internet of Things (IoT), *Journal of Network and Computer Application*, IEEE System Journal, *Enterprise Information Systems, Sensors, Personal and Ubiquitous Computing*, IEEE ACCESS, and so on. He was a Reviewer for the IEEE TRANSACTIONS ON COMPUTER, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE Communications Magazine, IEEE TRANSACTIONS ON MODELING AND PERFORMANCE EVALUATION OF COMPUTING SYSTEMS, IEEE TRANSACTIONS ON SUSTAINABLE COMPUTING, *Applied Soft Computing, International Journal of Parallel Programming, Microprocessors and Microsystems*. His current research interests include IoT, cloud computing, edge computing, power management, and operating system.

Dr. Jia is a member of China Computer Federation (CCF).

Jaime Lloret (M'07–SM'10) received the M.Sc. degree in physics, the master's degree in corporative networks and systems integration, and the M.Sc. degree in electronic engineering from the University of Valencia, Valencia, Spain, in 1997, 1999, 2003, respectively, and the Ph.D. degree in telecommunication engineering (Dr.Ing.) from the Polytechnic University of Valencia, Valencia, in 2006.

He is a Cisco Certified Network Professional Instructor. He worked as a Network Designer and Administrator in several enterprises. He is currently an Associate Professor of Computer Communications with the Polytechnic University of València, València, Spain. He is the Chair of the Integrated Management Coastal Research Institute and he is the Head of the "Active and collaborative techniques and use of technologic resources in the education" Innovation Group. He is the Director of the University Diploma "Redes y Comunicaciones de Ordenadores" and he was the Director of the University Master "Digital Post Production" for the term 2012–2016. He was the Vice-Chair for the Europe/Africa Region of Cognitive Networks Technical Committee (IEEE Communications Society) for the term 2010–2012 and the Vice-Chair of the Internet Technical Committee (IEEE Communications Society and Internet Society) for the term 2011–2013. He has been the Internet Technical Committee Chair (IEEE Communications Society and Internet Society) for the term 2013–2015. He has authored 22 book chapters and has more than 450 research papers published in national and international conferences, international journals (more than 200 with ISI Thomson JCR). He has been the Co-Editor of 40 conference proceedings and a Guest Editor of several international books and journals.

Dr. Lloret is the Editor-in-Chief of the *Ad Hoc and Sensor Wireless Networks* (with ISI Thomson Impact Factor), the international journal *Networks Protocols and Algorithms*, and the *International Journal of Multimedia Communications*. Moreover, he is an Associate Editor-in-Chief of "Sensors" in the Section Sensor Networks, an advisory board member of the *International Journal of Distributed Sensor Networks* (both with ISI Thomson Impact factor), and a Board Chair for the International Academy, Research, and Industry Association (IARIA) Journals (8 Journals). Furthermore, he is (or has been) an Associate Editor of 46 international journals (16 of them with ISI Thomson Impact Factor). He has been involved in more than 450 Program Committees of international conferences, and more than 150 organization and steering committees. He has led many local, regional, national, and European projects. He is currently the Chair of the Working Group of the Standard IEEE 1907.1. Since 2016, he has been the Spanish Researcher with highest h-index in the *Telecommunications* journal list according to Clarivate Analytics Ranking. He has been the General Chair or Co-Chair of 45 International workshops and conferences. He is Association of Computing Machinery (ACM) Senior and IARIA Fellow.