Pin-Han Ho Shih Yu Chang James She Yuren You Min Chen János Tapolcai

Data Science Driven Intelligent IoT

The Internet of things (IoT) is an interconnected system of computing devices, machinery, and digital machines that digitize the real world. The IoT has already affected people's lives, including transportation, housing, food, clothing, health, and remote monitoring. Many home appliances can be controlled through mobile phones and voice. Many applications allow users to improve their quality of life, and even enable the elderly and the disabled to live more conveniently. Data science (DS) is a multidisciplinary approach to discovering, extracting, and presenting insights in data by focusing on data collection, data store and access, data analysis, and data communication techniques. Data science includes descriptive, diagnostic, predictive, and prescriptive capabilities. This means that through data science, administrators can use data to figure out what happened, why it happened, what happened, and what they should do with expected outcomes. Since the automation of an intelligent IoT system requires all tasks of DS, DS will be the most proper candidate technology ready to solve those issues faced by intelligent IoT systems. To collect data for IoT applications features, how to design sensor deployment and their connections via communication networks is the first main problem for intelligent IoT. The next step is how to apply machine learning (ML) and artificial intelligence (AI) algorithms to analyze and interpret insights concerning collected intelligent IoT data. Finally, it is also very crucial to communicate analysis results effectively to users of intelligent IoT devices.

From the viewpoint of DS, we believe that the following categories of problems should benefit from DS related technologies in developing future intelligent IoT systems. The first problem is how to deal with intelligent IoT Big Data. The amount of generated data in each application unit of an intelligent IoT system is at least the scale of Terabytes (TB). The collection, exchange, storage, and access of such a huge amount of data at an intelligent IoT device is an exceedingly challenging task, as the computational and communication resources in intelligent IoT devices are extremely limited. Deep learning is a breakthrough technology in ML/AI. Deep-learning (DL) applications on IoT devices often have an extremely strict real-time requirement. For example, security camera–based object-recognition tasks usually require a detection latency of less than 400 ms to capture and respond to target events—for example, abnormal targets (identified by DL technology) appearing inside a building—in a short response time. Current IoT devices often offload intelligence computation to the cloud. However, consistent and reliable wireless communication links, which are only available at limited locations with high cost, become one of the main difficulties for these intelligent IoT devices to fulfill real-time requirements. Hence, the second category of problems about intelligent IoT is to have advanced ML/AI algorithms which can perform data analysis with input data impacted by unreliable communication links. However, enabling ML/AI capabilities on the intelligent IoT device side is not an easy assignment. The main properties of intelligent IoT devices are small memory size, low power and distributed. The third category of problems is to design new ML/AI algorithms that can be implemented at IoT devices in a distributed manner under small memory size and low power constraints. Finally, security, trust and privacy of intelligent IoT users are always main considerations for any new technologies. With the huge number of intelligent IoT connected devices, how to apply DS to enhance access control systems, trust management, and secure data sharing with privacy considerations over intelligent IoT systems is a challenging problem.

This Special Issue (SI) called for research contributions on applying DS techniques to design intelligent IoT systems and their applications from many perspectives, including architecture design, algorithm invention and analysis, fundamental theories, and practical considerations. Thanks to the extensive efforts of the reviewers and the tremendous support from the Editor-in-Chief, Dr. Rath Vannithamby, we could accept eight contributed articles covering several aspects related to the topics about Data Science Driven Intelligent IoT. A brief review follows:

The first article "Impact and Challenges of Intelligent IoT to Meteorological Science" by Chen *et al.* aims to introduce the application of intelligent IoT technology in the field of meteorological science and the challenges it faces. Along with the introduction of intelligent IoT technologies, a comprehensive review of current studies related to meteorological observation, forecast, and services with intelligent IoT is provided. Correspondingly, the impact of intelligent IoT on meteorological businesses is analyzed. Following that, the authors highlight recent challenges along with a few directions for potential research that could fill gaps in the meteorological domains for researchers and developers.

The second article "Clustered and Multi-Tasked Federated Distillation for Heterogeneous and Resource Constrained Industrial IoT Applications" by Hamood *et al.* presents CMFD as a promising solution for service heterogeneity and enabling a collaborative and resource efficient AI between different IIoT environments. CMFD, which utilizes a combination of CFL and

KD, allows for efficient and cooperative training and more personalized models in data and resource-starved IIoT. The effectiveness of CMFD is demonstrated through the case study of heterogeneous IIoT environments using a realistic dataset. Our empirical results indicated that CMFD outperforms baseline techniques in terms of accuracy, communication overhead, and computation complexity. Further research is needed to test CMFD on a wider range of IIoT applications and investigate its potential for use in other IIoT environments.

The third article "Toward a Reliable Evaluation of Machine Learning Schemes for Network-Based Intrusion Detection" by Eduardo *et al.* demonstrate those existing machine learning techniques applied for network traffic classification fail when facing the characteristics of real-world environments. The experiments analyzed more than 30 TB of data spanning 10 years of real network traffic and 9 intrusion detection datasets. Besides the analysis, the authors also define a set of guidelines to build reliable application of machine learning for network traffic classification, which may guide future research and ensure the reliability of machine learning model deployment in production environments.

The fourth article "Intelligent Space-Air-Ground Collaborative Computing Networks" by Shahnila *et al.* discuss the space-air-ground collaborative computing networks (SAGCCN) and the intelligent technologies required in the data collection and offloading process to enable a smart and integrated system. The architecture of SAGCCN has been discussed in detail, along with the AI-based techniques for data collection and offloading between space, air, and ground. Additionally, this article has explored the benefits of using AI techniques for trajectory design in low-altitude UAVs and provided scenarios demonstrating the importance of a hierarchical architecture that can provide centralized training and computing on demand. Finally, several areas for future research have been discussed, highlighting the challenges and possible solutions associated with implementing SAGCCN in aerial computing.

The fifth article "Physical-Layer Counterattack Strategies for the Internet of Bio-Nano things with Molecular Communication" by Huang *et al.* presents the looming threats underlying the MC networks. Then, the biochemical counterattack strategies inspired by nature in multi-scale scenarios are exemplified. Against this background, the security approaches can be easily achieved via the PLS manner without the involvement of high layer encryption, which is reasonable for MC networks due to the lightweight implementations. Then, both the keyless and key-based PLS schemes for MC networks were discussed from the perspective of data science and conventional model-based techniques. Finally, some open problems and future directions for secrecy enhancement in MC networks, especially for IoBNT, are envisaged, which suggests that interdisciplinary efforts are required to reach this goal.

The sixth article "Data Liquidity Optimization for Digital Twin Empowered Internet of Things: A Federated Learning Approach" by Tsai *et al.* investigates the requirements for designing DT empowered IoT systems from the perspective of DT data liquidity, and discusses its optimization strategy, aiming to make multiple DTs complement each other for model training and inference through various DT communication schemes. In addition, the LiqDT architecture and FedDT method are proposed to improve DT data liquidity through cloud-edge device DT collaboration, which can realize task migration among multiple DTs for highly fault tolerant and highly reliable IoT applications. According to the experimental results, FedDT achieves faster convergence speed, higher accuracy, and better generalization ability than FedAvg and FedProx. Future works about applying DT concept are also discussed in this article.

The seventh article "A DT Machine Learning-Based Satellite Orbit Prediction for IoT Applications" by Xu *et al.* discusses a DT framework for satellite orbit prediction to improve the accuracy of traditional dynamic orbit prediction model. The application of container technology in the novel model provides a solution for frequently updating the orbit prediction model. Then, in the proposed DT model, the authors built an orbit error prediction model based on TCN to predict the error of traditional dynamic prediction model. Finally, the performance of the proposed orbit error prediction model is evaluated by simulation results to show its effectiveness and feasibility in improving the accuracy of satellite orbit prediction. This makes the novel DT model a promising fast satellite orbit prediction under IoT supporting with low consumption.

The eighth article "RIS-IoE for Data-Driven Networks: New Mentalities, Trends and Preliminary Solutions" by Zhuo *et al.* provides a comprehensive study for future 6G RIS-IoE networks with novel mentalities of AI, their corresponding design, deployments and optimizations, which possess an intelligent and controllable communication environment with lower costs, power consumptions and computation algorithms. Practically, RIS-assisted emerging techniques are also introduced to improve the security, spectrum efficiency, transmitting coverage and sustainable communications, consisting of PLS, SWIPT, NOMA and UAV. In addition, the authors consider a case study to address the problem of channel estimations, where such issue can be efficiently solved by our proposed decoding principles in the RIS-IoE networks with NOMA. Finally, the future research trends and open issues for proposed systems are summarized, aiming to satisfy the incoming requirements.

In summary, the collected articles offer innovative application scenarios related to topics about Data Science Driven Intelligent IoT and shed light on the underlying principles, ML/AI algorithms, architecture and applications of the IoT systems. We hope that this timely special issue will trigger more future work in the emerging area.