
Rita Wouhaybi


Ravikumar Balakrishnan

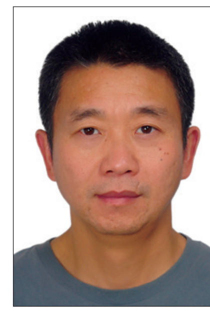

Navid NaderiAlizadeh


Ahmad Beirami


Marco Di Felice


Shadi Noghabi


Anirudh Badam


Shugong Xu

An End-to-End Machine Learning Perspective on Industrial IoT

The world is witnessing the emergence of billions of IoT devices fueled by tremendous leaps in computing paradigms spanning clients, the edge, and the cloud, as well as the ultra-low-latency and gigabit-per-second connectivity offered by 5G networks. It is estimated that IoT devices will generate 90 Zettabytes of data at the network edge by 2025 (IDC Data Age 2025 Whitepaper). The Industry 4.0 revolution is one of the major drivers for this large-scale adoption of IoT and is taking place rapidly, transforming a wide range of markets including manufacturing, energy, agriculture, transportation, and logistics. For example, in smart manufacturing, predictive maintenance for minimizing downtime, location tracking for inventory, and automation are some key areas that can boost yield and decrease time to market. Similarly, in energy, smart metering and detecting defects in oil/gas extraction stations as well as in the extracted matter are leading to significant improvements in operational and capital expenditures.

Access to abundant multi-modal data in such industrial use cases means that machine learning (ML) and artificial intelligence (AI) will play an integral role in analysis and understanding of data, improvement of operational efficiency and product quality, as well as reduction of down time by several orders of magnitude. Successfully implementing an ML-based solution for Industrial IoT (IIoT) involves many aspects including but not limited to (i) adequate infrastructure placement for timely and accurate collection of data from multiple sensors, (ii) aggregation of data at gateways/cloud, (iii) development and adoption of advanced AI algorithms, and (iv) hosting ML-related tasks at the edge for the purpose of reducing the network load. Industry 4.0 will change forever factories as we know them, with distributed manufacturing and on-demand personalized products reaching a wide scale.

This Special Issue (SI) focuses on recent advances, challenges, and trends of applying ML in IIoT settings. The articles went through careful curation and span two broad thrusts, namely, scalable and efficient ML technologies to enable IoT as well as applications, system design, and/or software optimization of ML applied to three different IIoT scenarios.

In "Federated Learning for the Internet of Things: Applications, Challenges, and Opportunities," the application of federated learning (FL) to IoT where ML models are collaboratively learned from devices without the need to share their data is discussed. Particular emphasis is placed on the challenges and opportunities of adopting FL for IoT platforms including device and network resource constraints, intermittent connectivity, temporal dynamics of data, trustworthiness, as well as the difficulty and opportunities for standardization and building scalable systems to realize such a distributed learning paradigm.

In "Toward Scalable and Robust AIoT via Decentralized Federated Learning," the authors introduce a generic decentralized federated learning (DFL) architecture that can transition between synchronous and asynchronous modes. Such flexible architecture is highly suited for IoT environments where over-reliance on a central server is a burden due to stragglers (slow devices) and device availability/connectivity potentially being intermittent. The authors implement, deploy and evaluate the DFL architecture in both simulations and testbed environments with results indicating improved model convergence by effectively combating the straggler issue without affecting model performance.

Ultra-reliable and low-latency communications are critical for enabling IIoT, and in "ML-Assisted Beam Selection via Digital Twins for Time-Sensitive Industrial IoT," ML technologies applied to communication between IoT devices with particular focus on reducing the complexity and overhead for 5G millimeter-wave (mmWave) communications are proposed. The authors rely on the digital twin paradigm and demonstrate the capability of lightweight ML models in predicting optimal mmWave beam configurations without compromising the communication performance.

In "Re-Inventing the Food Supply Chain with IoT: A Data-Driven Solution to Reduce Food Loss," an end-to-end solution of ML for the food supply chain is addressed. The authors envision using sensors in different stages of the food supply chain, namely farming, warehouse, transport, and retail, and discuss how ML solutions can be applied to reduce global food loss. The authors cast the food supply chain management as a multi-agent optimization framework and discuss some potential ways to solve the problem to incentivize practices that can help sustainably meet future food goals.

In "End-to-End Industrial IoT: Software Optimization and Acceleration," the significant speedup benefits of software optimization and acceleration in real-world IIoT ML use cases are revealed and extensively discussed. It is shown how the end-to-end inference runtimes of different ML tasks, such as predictive analytics and anomaly detection, can be accelerated by several orders of magnitude through the use of optimized versions of popular ML frameworks and libraries, including scikit-learn, TensorFlow, and PyTorch.

In "Touchless Control of Heavy Equipment Using Low-Cost Hand Gesture Recognition," end-to-end touchless control of heavy equipment using an ML pipeline is considered. In particular, a hand landmark detection model is used to map the input RGB image into a set of 3D landmarks. The estimated landmarks are then fed into a multi-layer perceptron (MLP), which at its output classifies the hand gesture given a history of prior input images. The proposed method is implemented and demonstrated to work notably well in a real-world use case of excavating in a rock pile by remotely controlling a robotic 1-tonne wheel loader.

In summary, this Special Issue contains six peer-reviewed articles touching on different aspects of machine learning technologies that can transform the Industrial IoT landscape, and some of the articles can serve as examples of how to successfully apply machine-learning-based solutions in industrial environments. Our overarching goal is to also reach a larger body of academic researchers, engineers/practitioners, and businesses that can benefit from understanding this emerging area, which is made possible through this Special Issue.

We would like to express our sincere thanks to all the authors and reviewers as well as the editorial team for providing us with the opportunity to put together this Special Issue on this exciting topic.

Biographies

Rita Wouhaybi (rita.h.wouhaybi@intel.com) is a senior principal engineer with the Industrial & Energy Solutions Division in the Internet of Things Group at Intel Corporation. She received her Ph.D. in electrical engineering from Columbia University in 2006. She also holds B.E. and M.E. degrees in computer and communications engineering from the American University of Beirut. Her career includes more than 15 years of industry experience, including engineering, management, and research positions. Her interests include peer-to-peer and distributed networks, game theory, and the use of machine learning in networking and social networks. She has filed over 150 patents and published over 20 papers in acclaimed IEEE and ACM conferences and journals. She was also the recipient of several awards for both academic and industry achievements.

Ravikumar Balakrishnan [M] (ravikumar.balakrishnan@intel.com) received his B.E. degree in ECE from SSN College of Engineering, India, in 2009, and his M.S. and Ph.D. degrees in ECE from Georgia Tech in 2011 and 2015, respectively. Since 2015, he has been a staff research scientist conducting research in the area of distributed machine learning and analyzing its behavior over wireless networks. Prior to this, he conducted research and development of model- and learning-based resource management strategies for wireless networks. He has over 15 peer-reviewed papers and over 30 pending U.S. patents. He has served as a Guest Editor for *IET Communications*, organized a special session at IEEE SPAWC 2020, and has served as a TPC member of several IEEE conferences.

Navid NaderiAlizadeh (nnaderi@seas.upenn.edu) is a postdoctoral scholar in the Department of Electrical and Systems Engineering at the University of Pennsylvania (UPenn). He received his B.S. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 2011, his M.S. degree in electrical and computer engineering from Cornell University, Ithaca, New York, in 2014, and his Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, in 2016. Prior to UPenn, he spent more than four years as a research scientist at Intel Labs and HRL Laboratories. His research interests include various areas of machine learning, including multi-agent reinforcement learning, graph representation learning, and self-supervised learning, and the applications of information theory and machine learning for resource allocation in wireless networks. He co-organized a special session at IEEE SPAWC 2021 on "The Interplay between Machine Learning and Resource Management in Wireless Networks." He currently serves as an Associate Editor for the *IEEE JSAC* Series on Machine Learning in Communications and Networks.

Ahmad Beirami (ahmad.beirami@gmail.com) received his B.S. degree in EE from Sharif University of Technology in 2007 and his Ph.D. degree in ECE from Georgia Institute of Technology in 2014. In 2018–2022, he was a research scientist at Facebook AI, where he led research on powering the next generation of virtual digital assistants with AR/VR capabilities. His research involves optimizing performance with robustness and fairness constraints in large-scale multimodal learning systems. Prior to that, he led the AI agent research program for automated playtesting of video games at Electronic Arts. Before moving to industry in 2018, he held a joint postdoctoral fellow position at Harvard and MIT, focused on problems at the intersection of core machine learning and information theory. He was the recipient of the 2015 Sigma Xi Best Ph.D. Thesis Award from Georgia Tech for his work on the fundamental limits of efficient communication over IoT networks.

Shadi Noghabi (shadi.noghabi@microsoft.com) is a senior researcher in the Networking Research Group at Microsoft Research. She received her Ph.D. Degree in CS from the University of Illinois at Urbana-Champaign in 2018 where she worked on the design and development of several large-scale production systems. Specifically, in collaboration with LinkedIn, she developed Ambry, a geo-distributed storage system used as the main backed for media objects at LinkedIn, and Samza, an Apache open-source stream processing framework used by LinkedIn and 10s of other companies. She has received the David Kuck Outstanding Thesis Award and the Microsoft Research Dissertation Award for her work during her Ph.D. Her current focus at Microsoft is on edge computing and Industrial IoT. She is currently working on developing highly available and scalable management platforms for the edge, ranging from small edges in remote areas to large renewable-based edge data centers and edge computing in space.

Marco Di Felice (marco.difelice3@unibo.it) received his Laurea (summa cum laude) and Ph.D. degrees in computer science from the University of Bologna, Italy, in 2004 and 2008, respectively. He was a visiting researcher at Georgia Institute of Technology, Atlanta, and Northeastern University, Boston, Massachusetts. He has authored more than 120 papers on wireless and mobile systems, achieving three best paper awards for his scientific production. He is currently an associate professor of computer science at the University of Bologna, where he is the director of the PRISM IoT laboratory, which focuses its research on pervasive systems. From 2014 to 2020, he was an Associate Editor of *Ad Hoc Networks*. He is currently an Associate Editor of the *IEEE Internet of Things Journal*. He was General Chair of IFIP WWIC 2019 and Program Chair of several ACM/IEEE conferences related to wireless and mobile systems. His research interests include: the Internet of Things and the Web of Things, robotic wireless networks and unmanned aerial systems (UAS), self-organizing and learning-based wireless systems, and context-aware, mobile, and edge computing.

Anirudh Badam (anirudh.badam@microsoft.com) is a principal research manager at Microsoft Research Redmond where he leads the edge-cloud connectivity/storage and analytics projects. His Ph.D. research at Princeton University on memory technologies called HashCache was recognized as MIT TR-10, and the IP eventually became a part of Akamai. His Ph.D. work on SSD technologies called SSDAlloc was licensed exclusively by Fusion-io, which was later acquired by SanDisk. His work on energy-efficient mobile-storage work received the best paper award at USENIX ATC 2015. His work on designing new SSDs optimized for memory-mapping received an honorable mention at the IEEE Micro Top Picks 2016. His other recent work on energy-efficient mobile computing was featured in the research spotlightand research highlight articles of *IEEE Pervasive Computing* and *CACM*, respectively, in 2016, and the solution today runs on every Windows 10 laptop. His work on battery-backed DRAM at Microsoft Research formed the basis for commodity non-volatile memory that is implemented on every server in Azure. He received the MSR Redmond Rockstar Researcher (2017) award, and the project on storage systems that he led received the MSR Redmond Outstanding Research Project (2017) award. More recently, he received a best paper award at ACM COMPASS 2019 for his work on IoT-based digital agriculture.

Shugong Xu [F'15] (shugong@shu.edu.cn) is a professor at Shanghai University and head of the Shanghai Institute for Advanced Communication and Data Science (SICS). In his 20+-year career in research (over 15 years in industrial research labs, including Sharp Labs of America, Huawei 2012 Labs, and Intel Labs), he has had over 40 issued U.S./WO/CN patents and published more than 100 peer-reviewed research papers. His work was among the major triggers of the research and standardization of IEEE 802.11s. He founded and directed the green radio project GREAT at Huawei, led the China national 863 theme project on green networks as Principal Investigator, and served as Co-Chair of the Technical Committee of the Green Touch Consortium. He was awarded the National Innovation Leadership Talent award by the Chinese government in 2013. He was also the winner of the 2017 Award for Advances in Communication from IEEE Communications Society. He received his B.S. degree from Wuhan University, and his M.E. and Ph.D. degrees from Huazhong University of Science and Technology. His current research interests include wireless communication systems, computer vision, and machine learning.