Guest Editorial
Data Science for the Internet of Things

THE INFLUENCE of the Internet of Things (IoT) and related produced data is destined to revolutionize economic and social society, even more incisively than the advent of digital. The creation of the IoT world would have been much more complicated if a big data structure had not been followed since the latter allows to analyze vast amounts of data. The IoT is the most significant flow of information collected on the Internet and, therefore, the largest supplier for big data systems, artificial intelligence (AI), and data science. Advanced statistical analysis techniques, neural networks, and AI algorithms, but also the ability to create mathematical models that best represent a physical phenomenon or social behavior, these are the new strategic assets in the digital transformation process that is facing the world of industry and services. In fact, collecting data is not enough: they must be managed, integrated, and compared with a mathematical model that formalizes the intrinsic knowledge in the experience and competence of people.

The IoT is one of the sectors of the data economy with the highest potential. Over the next few years, the boom in connected objects and the applications they enable will change people's lives and give an extra gear to both companies who want to increase competitiveness on the market and to public administrations who want to reduce costs and better serve the citizen. The so-called "integrity" of IoT data is the real added value. The data communicated by the sensors do not represent opinions, but facts; for example, the object X is on the shelf Y, or the user X has purchased the item Y. The impact on the business world is evident, as long as you know how to extract knowledge from the information. In essence, from the collection and insertion phases of the data in the so-called "data lake," it is necessary to move on to discovery and cleansing (or cleaning), which allows you to filter the data by isolating the useful ones, to arrive at the actual analysis and production of knowledge.

The fields already considered and those that will be affected soon by the IoT are numerous: smart cities, building and smart home management, manufacturing and logistics, agriculture, connected vehicles and public transport, healthcare, insurance, retail, and consumer electronics. People themselves have transformed into "data points": thanks to our smartphone, we all send data on location, behavior, and consumption practices.

The applications capable of integrating IoT and data science are innumerable and will have a radical impact on companies, public administrations, and consumers. The use of machine learning and other learning techniques at the basis of data science will acquire an ever-increasing role to complement the more traditional services with new logics capable of satisfying—and in many cases anticipating—the needs of companies and consumers. Hence, the application of machine learning is necessary and indeed fundamental, only in the case of large quantities of heterogeneous data that change rapidly: and this is one of the essential characteristics of the IoT.

Data science, i.e., the set of techniques for collecting, managing, and analyzing information to extract valuable insights, is the driving strength of any business organization and a recurring element of personal experience. Organizations are preparing to become data driven, aiming to manage any operational and decision-making process through the analysis of information.

The future of data science will see an increasingly pervasive use of analytical tools, an increase in data monetization initiatives and business transformation, and extensive use of AI.

In particular, machine learning techniques will be refined to automate algorithm correction processes and obtain an ever-higher level of accuracy. Further progress will concern the technologies for model management to ensure the rapid deployment of projects from the laboratory to the production environment. Data visualization will be a cornerstone of any data science initiative to extend the use of analytics to users without specific backgrounds. Data science projects are gradually shifting the focus. Initially, the analyses were performed with a purely descriptive function (monitoring of results) and diagnostics (to identify and correct contingent problems). Today, the purposes are becoming increasingly predictive (data and algorithms allow to anticipate future situations) and prescriptive (AI suggests and even activates response mechanisms to any critical issues).

The cities become smarter and smarter to deal with the incredible pressure they are facing. The integrated data of personal devices and IoT sensors will improve the management of urban spaces to ensure more efficient services and more population wealth. A smart city must, therefore, be instrumented, interconnected, and intelligent, that is, capable of collecting data in real time from sensors and personal devices, integrating the data collected in a single platform accessible to the various urban service providers, and capable of analyzing and visualizing the data to optimize the decision-making process. What does "smart city" mean? There are many definitions of smart city, often referring to the number of connected objects. Our idea of a smart city is a dynamic structure that improves the quality of life in a high-density urban environment, convenient for the administrations in charge of managing it and for users who

Digital Object Identifier 10.1109/JIOT.2020.2985598

dwell there to live or do business. Different buildings and public spaces should be a part of a highly responsive network. Inside, a series of sensors collect data, which are analyzed and deployed to meet the needs of the inhabitants. Food, energy, and water resources are produced to meet demand and used more efficiently. In smart cities, communication channels through social networks become increasingly open, which allows administrations to share policy changes and success stories, thus encouraging feedback. In this way, they can involve their stakeholders to make them feel part of the decision-making process. The article titled "Dominant data set selection algorithms for electricity consumption time-series data analysis based on affine transformation" presents a cohort of dominant data set selection algorithms for electricity consumption time-series data with the focus on discriminating the dominant data set. From a computational point of view, the authors prove that the selection problem of the minimum dominant data set is an NP-complete problem. The proposed analysis and experimental results assess that the algorithms have high performance in terms of effectiveness and efficiency. The article titled "Geography-aware inductive matrix completion for personalized point-of-interest recommendation in smart cities" proposes a geographical-aware inductive matrix completion approach for POI recommendations. The experimental results on two data sets show that the proposed approach can achieve a significant improvement compared to state-of-the-art POI recommendation methods. The article titled "A survey of Internet of Things (IoT) for geohazard prevention: Applications, technologies, and challenges" discusses a comprehensive survey of relevant research and technological developments of the IoT applied in geohazards prevention. The authors assess that compared with the conventional human-based monitoring and early warning systems for geohazards prevention, the IoT-based technology is more accurate, faster, safer, timelier, and smarter. To deal with the "big data" issues in a smart city scenario, the article titled "Online distributed IoT security monitoring with multidimensional streaming big data" proposes an online distributed IoT security monitoring algorithm (ODIS) for multidimensional streaming big time-series data. The proposed approach is a general IoT cybersecurity solution and shows promising performances in terms of cyberattack detection and monitoring. The article titled "Generating situational awareness of pedestrian and vehicular movement in urban areas using IoT data streams" presents a novel algorithm, the deep growing self-organizing map (Deep GSOM) algorithm that addresses the challenge of managing the scale, velocity, and magnitude of IoT data streams. The authors applied the Deep GSOM and congestion metric to profile vehicular and pedestrian movement in highly urbanized environments, using real-life data from the city of Melbourne, Australia and the city of Aarhus, Denmark. The article titled "FANN-on-MCU: An open-source toolkit for energy-efficient neural network inference at the edge of the Internet of Things" presents FANN-on-MCU, an open-source framework for easy deployment of NNs trained with the FANN library on both ARM Cortex-M cores and new parallel ultralow-power (PULP) RISC-V-based processors. Since the huge amount of IoT sensors in the smart city framework, they have shown performance comparisons of neural network inference between the two classes of processors and analyzed the parallel speedups and degradations due to parallelization overhead and memory transfers. The article titled "Big data analytics for event detection in the IoT-multicriteria approach" discusses advanced data science analysis methods used for ensuring the security of IoT in a smart city scenario. IoT security is one of the main challenges that cybersecurity investigations have been facing over the past few years; the three novel processes shed some light on the complex problem of anomaly detection, relevant for IoT security. The article titled "Deep-learning-based pedestrian inertial navigation: Methods, data set, and on-device inference" proposes L-IONet, a lightweight deep neural networks framework to learn inertial tracking from raw IMU data. The authors also conduct systematic research into the performance of deep-learning-based inertial odometry models on low-end devices. Finally, the article titled "Understanding crowd behaviors in a social event by passive WiFi sensing and data mining" presents a comprehensive data analysis framework that is proposed to extract multiaspect patterns regarding crowd behaviors from WiFi probe request records collected by passive WiFi sensor networks in a significant social event. The proposed framework is assessed by using real-world data collected in a significant social event held in a popular tourist area.

How does this virtuous integration work? We can attempt an automotive comparison: the IoT is the fuel and machine learning is the engine. The IoT collects the data that feed the ML, then allowing the transformation of the power supply into movement. ML benefits from the IoT, but it is a two-way relationship: the volumes of data generated by IoT devices have limited value without ML technologies capable of finding valuable insights and predictive models in the information. Their combination is increasingly dominant, with the new IoT solutions that integrate advanced data analysis platforms, and ML algorithms can simplify the management of connected devices, despite the continuous increase in smart objects.

The article titled "Cost-effective video summarization using deep CNN with hierarchical weighted fusion for IoT surveillance networks" presents a new solution of designing an effective deep CNN framework with a hierarchical, weighted fusion operator for surveillance video summarization in an IoT framework. The authors validate the performance of the framework by using two benchmark data sets, verifying its importance and superiority for surveillance versus comparison with other state-of-the-art methods. The article titled "Internet of Things (IoT) and machine-learning-based leaching requirements estimation for saline soils" discusses the role of the IoT in the efficient use of irrigation water for leaching purposes by accurate estimation according to the microenvironment conditions of the crop field. The authors assess that IoT-based data provide high accuracy with real-time field microenvironment conditions. The article titled "A semantic collaboration method based on uniform knowledge graph" first analyzes the data characteristics of the semantic IoT (SIoT). Then, the authors consider the uniform knowledge graph for the organization form of semantic data and combine with the advantages of the knowledge graph in knowledge representation and processing. Besides, the basic function of semantic collaboration based on the uniform knowledge graph in the SIoT is achieved. The article titled "Clustering analysis for Internet of Spectrum Devices: Real-world data analytics and applications" focuses on clustering analysis for the Internet of Spectrum Devices, which finds applications in the joint time–frequency spectrum inference. Real-world inference experiments demonstrate that the proposed approach can improve the inference performance on both the inference accuracy and the runtime overhead than the previous one. The article titled "Countering malicious URLs in Internet of Things using a knowledge-based approach and a simulated expert" proposes a novel methodology to detect malicious URLs using a simulated expert (SE) and a knowledge-based system (KBS). The proposed approach is capable of detecting and classifying the URLs from handheld devices if it can support the Web browser. Finally, the article titled "A hybrid deep learning architecture for privacy-preserving mobile analytics" presents and evaluates a hybrid architecture where the local device and the cloud system collaborate on running a complex neural network that has previously been trained and fine-tuned on the cloud. The authors show that we can highly decrease the computational complexity on the user side, as well as the communication cost between the user's device and the cloud.

In the last decade, IoT technologies have multiplied and developed, as well as the application areas have evolved. Just think of the smart home solutions, which today allow you to turn on the house lights with a simple voice command, or even the smart car sector, which equips cars with onboard connectivity during their production (while initially they were only connected via box GPS-GPRS). But the evolutionary trends are manifold and concern various aspects of this area. We are not only talking about new standards or new products but how to implement its features and how to improve services for users. The Internet of Services arises from the integration of IoT devices, to increase the added value of the latter for the benefit of the various players in the supply chain. In particular, the Internet of Services offers a complete framework for managing the functionality of connected objects simultaneously. The article titled "Multiuser multivariate multiorder Markov-based multimodal user mobility pattern prediction" presents a multiuser multivariate multiorder Markov model including the influence model of multiple users and the multiuser Markov trajectory transition model. The authors show a series of experiments based on real-world GPS trajectory data set to assess the performance of the proposed approaches.

The article titled "Context-aware QoS prediction with neural collaborative filtering for Internet-of-Things services" discusses a holistic approach based on fuzzy c-means (FCM) and neural collaborative filtering (NCF) with two advantages: 1) contextual information clustering and 2) deep latent feature learning. The proposed method can simultaneously discover latent features in historical QoS data and latent cluster features. The article titled "An application development framework for Internet-of-Things service orchestration" proposes a framework supported by attribute-driven design and MDD to solve the challenges of IoT application development and service orchestration by providing layered abstractions and granularity. The authors demonstrate a transformation model in a smart vehicle system using the IADev framework, which enables the development of IoT applications by abstracting the complexities of the platform. The article titled "A group signature and authentication scheme for blockchain-based mobile-edge computing" proposes a group signature scheme designed for validating blocks of blockchain to address such issues. The authors demonstrate that their solution makes the BMEC architecture resistant to some blockchain consensus-level attacks compared to the architecture in the absence of such a solution. The article titled "A novel approach for efficient management of data lifespan of IoT devices" focuses on data storage at the things layer (wireless sensors). The authors evaluate the performance of several distributed data storage systems over WSN running over the ZigBee MAC protocol. The proposed scheme is evaluated against two prominent schemes, DEC-DS and DEC-EaD, using the IEEE 802.15.4 ZigBee MAC protocol. The article titled "Data integrity monitoring method of digital sensors for Internet-of-Things applications" models the data failure of digital sensors by dividing it into format failure, timing failure, and value failure based on the concept of data quality in IoT. The authors present the control mechanism named dual variable-length data monitoring window (VLDMW) to improve the credibility of data recovery. The article titled "Challenges to IoT-enabled predictive maintenance for Industry 4.0" discusses a comprehensive outlook of the current PdM issues, with the final aim of providing a deeper understanding of the limitations and strengths, and challenges and opportunities of this dynamic maintenance paradigm. The authors outline some main research issues to be addressed for the successful development and deployment of IoT-enabled PdM in industry. The article titled "Toward secure and provable authentication for Internet of Things: Realizing Industry 4.0" presents a robust, lightweight, and provably secure authentication and key agreement protocol specifically for the IoT environment based on a hierarchical approach. The computational and communication overhead analysis indicates the fact that the proposed mechanism is comparatively less expensive than the existing state of the art. The article titled "DLCD-CCE: A local community detection algorithm for complex IoT networks" proposes DLCD-CCE, a novel distributed local community detection algorithm based on community center expansion. The experimental results clearly demonstrate that for different scales of the network, compared with the existing local community detection algorithm, DLCD-CCE has higher accuracy and good scalability, as well as provides a more efficient method for measuring node centricity. The article titled "Mobile Internet of Things under data physical fusion technology" discusses AC4E, a model that is built around the information physical fusion technology and verified by simulation experiments. The performance of the model, such as the best reaction path, trust mechanism, and secret rate variation, is analyzed, and the model is proved to have good reactivity, correctness, security, and parallel processing. The article titled "A distributed CoRE-based resource synchronization mechanism" discusses how CoRE-related standards and ongoing research work documented in Internet-Draft proposals can be used to build a robust and reliable IoT synchronization mechanism. Here, a decomposition mechanism for the rule is proposed allowing it to be split into a set of small, chainable, and reusable components. The article titled "Federated learning with cooperating devices: A consensus approach for massive IoT networks" proposes a fully distributed (or serverless) learning approach: the proposed FL algorithms leverage the cooperation of devices that perform data operations inside the network by iterating local computations and mutual interactions via consensus-based methods. The proposed distributed learning approach was validated on an IIoT scenario where an NN model was distributedly trained to solve the problem of passive body detection inside a human–robot collaborative workspace. Finally, the article titled "Exploiting workflow languages and semantics for validation of security policies in IoT composite services" proposes a methodology that enables the analysis and validation of security properties of composite IoT systems in the presence of complex, dynamic data workflows. A case study from the e-health domain is presented to validate the approach.

One of the most salient features of the IoT is undoubtedly represented by the number of data that billions of sensors transmit on the network. A large amount of information (not surprisingly called big data) is the new added value that industries are preparing to manage in the best way to transform it into the business value. From big data to data science, the transition is the most natural. If it is true that the theme of enhancing data collected from connected devices goes hand in hand with the ability to extract useful information from them, data science can play, in this regard, an extremely important role.

ACKNOWLEDGMENT

The guest editorial team would like to convey their heartiest gratitude to all the authors who have submitted their knowledgeable contributions and to the highly qualified anonymous reviewers. They believe that the selected contributions, which represent the state of the art of data science for the IoT, will be of great benefit to the IoT and computer science community. The guest editors would also like to thank Prof. Xuemin (Sherman) Shen, Past-Editor-in-Chief of the IEEE INTERNET OF THINGS JOURNAL and Prof. Honggang Wang, Editor-in-Chief of the IEEE INTERNET OF THINGS JOURNAL, for allowing them to organize this special issue and for all the encouragement, help, and support given throughout the process, and Mariola Piatkiewicz, IEEE INTERNET OF THINGS JOURNAL peer-review support specialist, for her professional support and assistance during the preparation of this special issue.

FRANCESCO PICCIALLI, *Guest Editor*
Department of Mathematics and Applications "R. Caccioppoli"
University of Naples Federico II
80138 Napoli, Italy

SALVATORE CUOMO, *Guest Editor*
Department of Mathematics and Applications "R. Caccioppoli"
University of Naples Federico II
80138 Napoli, Italy

NIK BESSIS, *Senior Member, IEEE*
Department of Computer Science
Edge Hill University
Ormskirk L39 4QP, U.K.

YUJI YOSHIMURA
Research Center for Advanced Science and Technology
University of Tokyo
Tokyo 113-8654, Japan

Francesco Piccialli received the Laurea (B.Sc. and M.Sc.) degrees in computer science and the Ph.D. degree in computational and computer science from the University of Naples Federico II, Naples, Italy, in 2012 and 2016, respectively.

He is currently an Assistant Professor (tenure track) of computer science with the Department of Mathematics and Applications "R. Caccioppoli," University of Naples Federico II. In 2018, he took the Italian Scientific Habilitation for Associate Professorship. He has been a Visiting Research Scholar with Chun-Ang University, Seoul, South Korea, in 2017, and a Visiting Professor with the University of Geosciences, Beijing, China, in 2019. He is also a Research Fellow with CINI (National Interuniversity Consortium for Informatics), Rome, Italy, in 2013. He is the Co-Founder and the Scientific Director with T.P.S. S.r.l., Treviso, Italy, an innovative startup company whose mission is the development of innovative systems and services based on IoT technology for cultural heritage. He is the Founder and the Scientific Director of the M.O.D.A.L. Research Group that is engaged in cutting-edge novel methodologies, applications and services in data science and machine learning fields and their emerging application domains. He gained funding for about €3.5 million. He has authored many papers (over 80) in international conferences and leading journals (IEEE, ACM, Elsevier, and Springer). He has been involved in research and development projects in the research areas of the Internet of Things, smart environments, data science, and mobile applications.

Dr. Piccialli currently serves as an Associate Editor and the Editorial Board Member of many international and top-ranking journals.

Salvatore Cuomo received the Ph.D. degree in applied mathematics and computer sciences from the University of Naples Federico II, Naples, Italy, in 2004.

He is an Associate Professor of numerical analysis with the Department of Mathematics and Applications "R. Caccioppoli," University of Naples Federico II. He has been a Visiting Researcher with Case Western Reserve University, Cleveland, OH, USA, in 2016, and a Visiting Professor with the University of Geosciences, Beijing, China, in 2019. He has been a Founder of Educabile S.r.l., Naples, an innovative startup company, dealing with innovative educational solutions and data analysis since May 2019. One of his passions is technology transfer, particularly in data analytics contexts and smart educational environments. He has authored numerous research papers (over 90) in international conferences and international journals indexed in Scopus and WOS databases. He has been involved in several research and development projects in the research areas of scientific computing, data science, and Internet of Things. His research interests are in applied mathematics topics, more in detail in numerical approximation problems (theory, practice, and applications), inverse problems, and machine learning and data science.

Nik Bessis (Senior Member, IEEE) received the B.A. degree from T.E.I., Athens, Greece, in 1991, and the M.A. and Ph.D. degrees from De Montfort University, Leicester, U.K., in 1995 and 2002, respectively.

He is a Full Professor with Edge Hill University, Ormskirk, U.K., from 2010, where he has been the Head (Chair) of the Department of Computer Science since 2015. He is involved in a number of funded research and commercial projects in these areas. He has published over 300 works. His research is on social graphs for network and big data analytics as well as developing data push and resource provisioning services in IoT, FI and inter-clouds for a number of settings, including disaster management.

Dr. Bessis won four best papers awards. He is an editor of several books and an editor-in-chief of a refereed international journal. He is a Fellow of HEA and BCS.

Yuji Yoshimura received the Ph.D. degree in computer science from Universitat Pompeu Fabra, Barcelona, Spain, in 2016.

He worked as the Postdoctoral Associate with the Massachusetts Institute of Technology, Cambridge, MA, USA. He is a Project Associate Professor with the Research Center for Advanced Science and Technology, University of Tokyo, Tokyo, Japan. He is also the Scientific Advisor with Louvre Museum, Paris, France. His speciality is the AI and bigdata analysis for architecture and urban planning. Supported by the combination of his expertise on architecture, mobile technologies and direct experience in mobility and environmental planning, he largely works on vehicle and pedestrian flow analysis and simulation, environmental analysis through wireless sensor technologies and mobile technology applied to urban planning.