Invited Article
Challenges and Opportunities for Data Science and Machine Learning in IoT Systems – A Timely Debate: Part 2

Sumi Helal, Flavia C. Delicato, Cintia B. Margi, Satyajayant Misra, and Markus Endler

Abstract

In Part 1 of this two-part article, we presented the first installment of a panel that was held within the Very Large Internet of Things Workshop (VLIoT) held in conjunction with the Very Large Data Bases (VLDB) conference in Los Angeles, California, in August of 2019. The panel addressed the challenges and opportunities in the intersecting areas of IoT, Data Science and Machine Learning. We presented the opening statements of four panelists, and a discussion of numerous questions addressing infrastructure issues and how it can support the growing demands for integrated intelligence, including communication, coordination, and distribution. In this second installment, we cover the remainder of this important debate focusing on the critical issues of scalable information and event processing, embedded machine learning, security and privacy, and speculating about the new business models that could soon be emerging as IoT meets Data Science and Machine Learning.

Introduction

In this second part of this two-part article reporting on the panel on Internet of Things (IoT), Data Science and Machine Learning, we go past infrastructure issues to cover the heart of generating intelligence in these inter-related systems. The panel debate covered in this part focused on application-level issues, particularly information and event processing, embedded machine learning, and security and privacy. It also analyzed the challenges of embedding intelligence at all levels of an IoT ecosystem. Markus Endler moderated the panel by posing a number of questions surrounding these issues to the panelists: Flavia, Sumi, Cintia and Jay. At the close of the panel, the new economic models that can be envisaged when reaching the full potential of these emerging technologies have been briefly discussed. We hope the discussions presented in this and the first part of the article will shed light on practically important IoT-related issues in the context of Machine Learning and big data. Ultimately, we hope we contribute to setting a meaningful research agenda for the future of IoT.

Event Processing

Markus: I understand that one of the central issues in Data Science is data curation, i.e., the selection of useful data, its completion, transformation, and summarization. Since many IoT middleware systems have adopted (Complex) Event processing as a core functionality provided by cloud and edge nodes (e.g. [1, 2, 3, 4]), what do you think are the main benefits and challenges of doing such data curation while data is being collected.

Cintia: In many cases data captured at the IoT devices is transmitted without any associated information, although metadata is essential for the overall data analysis at the core of the network (i.e., at the cloud). So, at some place on the way from the Edge to cloud, the enrichment of the "raw" data of the IoT device must happen. But this enrichment with metadata should be done in a consistent way, since otherwise the data analysis may misunderstand the inherent spatial-temporal-contextual relation between data pieces from different IoT devices, that for example share a same place. Another issue is that data transmission is the most costly task in terms of energy consumption, therefore it is important to evaluate if a particular data should be locally processed or transmitted. Furthermore, pub/sub approaches (such as the Constrained Application Protocol (CoAP)) enable a particular piece of data to be transmitted more than once to different destinations. Therefore, a minimum amount of data curation is necessary to support data science and to reduce the energy consumption due to unnecessary data transmission.

Sumi: While I agree that event processing is popular and of high utility in distributed systems, I feel that as we ponder over the future of IoT-based systems, we should not assume that events will be the main or only "sentience abstraction" on the scene. I would like to respond to this question but allow me to alter it slightly, asking how should we prepare for data curation in the age of IoT? I believe we have a unique opportunity to advance the underpinning data science into new territories. Starting off with raw data that can be curated dynamically, it will help if we can think of the raw data as the most basic element in an emerging new calculus, data calculus. For instance, in the domain of daily living and lifestyle data, a first derivative, perhaps is events, which fold large episodic data streams into higher entropy elements much richer in their semantics, and also of higher utility, as mentioned before. But we have seen how events may be used to define and recognize activities (e.g., [5]), and hence perhaps activities could be a second derivative sentience abstraction. Perhaps behaviors are a third derivative. Domain-specific calculi of data may very well be one future direction to explore. If we continue with this view for a second, and back to Markus' question, how do we better prepare for IoT data curation, I would argue that the best preparation is to understand the possible query systems that will be needed for each derivative of a given calculus, and on this basis, find the magical data graph that can take raw data input while dynamically synthesizing the structures relevant to each and all derivative sentience abstractions, simultaneously. Doing so simultaneously is very crucial because we do not know, and cannot guess, how applications will be developed. An app may better utilize raw curated data, while another may utilize an activity, a behavior, or a phenomena cloud. In fact, several sentience abstractions may need to be used within the same single application. What exactly is this magical data graph? Can graph databases such as TigerGraph be used? What are the query languages of the future under this view? And how can we optimize such a magical data graph so its overhead and runtime resource needs are sustainable while catering to a full data calculus or even multiple data calculi?

Digital Object Identifier: 10.1109/IOTM.0022.2000002

Markus: Sumi, yes, these are all very relevant and unsolved questions. Data curation has to be, ultimately, specific to the application and will demand new forms of dynamically defining the data and event processing functions and distributing them correctly and synchronously over all the processing components of the IoT system. But let's hear Flavia's opinion on this issue.

Flavia: I fully agree with Sumi's opinion on the need to broaden our view of the relevant entities in a data curation process for IoT. In many traditional data analysis systems, events are the final entities, to be detected and acted upon, preferably on the fly. With the new perspectives brought by IoT, I really like the idea of seeing such events as a first derivative. Their enrichment with contextual data and metadata through powerful ML techniques will lead to much more complex and useful entities to actually support decision-making processes and the activities of human beings, in a personalized, application-tailored way. However, regardless of where we want to get to at the end of the process, there is clearly a continuous transformation that begins with the collection of raw data on instrumented things and ends with the delivery of information (high level knowledge, decisions, detected events, activities/behaviors) relevant to end users (or their applications). Such a transformation process can be viewed as a workflow, but one where the sequence of activities must be performed in a hierarchical manner and considering the different nature of the computational nodes involved, from IoT devices to the cloud across multiple tiers at the edge and through the core of the network. The data curation process, which involves the organization and processing of dynamic and distributed IoT data, needs to be done at all levels of this hierarchical workflow. The challenges involved in this process of knowledge production are related to the nature of the data, but also the nature of IoT applications. Regarding the data nature, IoT devices used in several application domains generate data in a continuous way (as a stream), the data vary as a function of time and space, and also in terms of their statistical distribution/properties, since they relate to physical phenomena with intrinsically variable and often unpredictable behavior. Moreover, IoT data are often highly volatile and their relevance for the application depends on the timely processing. In this regard, Forrester, Inc. [6] coined the term "perishable insights" to refer to information that must be used quickly at the expense of losing its value. Typically, the knowledge produced by IoT in several domains, such as all kinds of monitoring systems, is perishable, and events must be detected (and reacted upon) preferably over data in motion, as close to the moment of its occurrence as possible. On-line learning techniques [7], distributed complex event processing [8], and stream analytics [9] are promising approaches to deal with such a nature of IoT data. Regarding the data transformation process, it involves increasing the data abstraction level, including metadata and semantics, at each processing step, and the outcome needs to be accessible, understandable by humans, or interpretable by machines and decision-making systems [10]. However, it is not just the final outcome of the transformation process that will be effectively useful to the user. I fully agree with Sumi that it is necessary to make different sentience abstractions available as output entities of a knowledge generation process. For some applications, events or event streams may be the entity of interest. For others, intermediate results in the path from the start to the end of the data processing workflow will be the object of interest. Therefore, mechanisms and tools must be made available for end users or their applications to be able to represent their functional and Quality of Service (QoS) requirements in order to clearly express what they expect as a result of the data science process. These requirements may be better stated in a declarative way as queries, or perhaps better represented as services and service level agreements, but a meaningful and flexible mechanism is needed that provides abstractions related to the domain of the user/application, not to the computational infrastructure that will provide the service. Considering the need to combine data from multiple sources in a hierarchical manner, I believe that an interesting approach to explore in the context of the data transformation process is multilevel information fusion. A recent and interesting work is reported in [11], where the authors proposed a hierarchical automated data fusion architecture for Smart Healthcare ecosystems. The various elements in a new-generation healthcare system (incorporating body sensors, edge devices, and cloud platforms) contribute to processing the generated data according to their computing capabilities and for different purposes. Lower-level, more resourceful elements perform simple operations of data fusion and generate quick responses, while as we move up the hierarchy there is a transfer of aggregate knowledge produced at the lower levels and enrichment of the information. The proposed architecture is implemented using the Complex Event Processing (CEP) technology. In the context of representing the interests of applications, I would bet on building domain-specific languages (DSL). With such languages, it is possible to provide a representation in terms of the problem space (thus more familiar to the end user) rather than the in the solution space. In addition, existing mechanisms and tools from the model-driven development field can be inherited to do the translation and mapping into the problem space. Mechanisms for automatic translation and code generation help address the inherent heterogeneity of both IoT and ML platforms. Therefore, it would be possible on the one hand to expose to the end user a flexible language with elements representing the various application domains. On the other hand, expose to the data scientist a language with elements representing the various mechanisms and processing steps of a typical ML life. Automated mapping processes would reduce the gap between different types of stakeholders and accelerate both the commissioning of ML models and IoT systems as well as any system evolution resulting from changes in requirements, domain or IoT technology. There are some incipient attempts to propose DSLs for IoT, as reported in [12, 13, 14]. However, to the best of my knowledge, there is still no DSL initiatives focused on ML in general let alone ML for IoT. So, this seems like an interesting research avenue to explore.

"Curation can summarize the data or quantize it, as possible, to enable reduction and compression of the data, thus significantly reducing the total data transfer load of the network. This strengthens the arguments of curation."

Markus: What about you, Jay, what is your view on the importance of the data curation activity and where it should be done in a multiple tiers IoT-edge system?

Jay: Data curation will be extremely important at the source or as the data migrates through the edge. The volume of the data flowing in the network is going to be very large and that curation can, in fact, reduce the amount of transmitted data. Curation can summarize the data or quantize it, as possible, to enable reduction and compression of the data, thus significantly reducing the total data transfer load of the network. This strengthens the arguments of curation. Another reason for curation would be the need for data privacy. The applications will need to transmit data to the cloud from the nodes or vice versa, but in each case privacy and security of the content will be essential, and curation approaches resulting in *k*-anonymity or some form of differential privacy (particularly dynamic differential privacy) can be possible with data curation. The challenge is to assess what data to curate and to what extent. This is primarily application driven, and without visibility of how the pre-processed data and post-processed information can be used to

affect user privacy, the edge computing nodes are uninformed to provision for privacy.

Built-in Machine Learning

Markus: AI and Machine Learning (ML) are certainly the natural next steps toward enabling IoT systems to better react to changes in the (monitored and controlled) physical environment as well as to the needs of the application's user. Moreover, it appears to be the cornerstone to support autonomy and long-term evolvability. However, one of the most remarkable characteristics of sensor data obtained through IoT systems is its crudeness. Because IoT devices collect data through various complex sensors, the data is typically raw. This means that major data processing is required before valuable information can be inferred for input to powerful AI applications. In fact, separating the meaningful signal from the noise and transforming the unstructured data flows into useful, structured data is the most paramount, yet complex, step when building a smart IoT application. So, to what extent and for which tasks do you think that AI/ML can, and should, be used? Would you rely on an IoT system that continuously learns how to control industrial, medical or unmanned aerial vehicle (UAV) equipment?

Flavia: In my point of view, the tasks for which AI/ML can be used in IoT are limited only by human imagination. The potential of IoT systems will only be realized when interconnected objects are truly intelligent, and collaborate with each other in an autonomous, organic way to effectively contribute to producing useful knowledge and providing services that enhance the lives of human beings. ML techniques are the key enablers of the vision of intelligent environments, where every interaction of humans with their surroundings, from inanimate objects to other human fellows, will be personalized and optimized. Processes will not only be automated, but tailored to users' personal tastes, schedules, habits, health conditions, preferences, and even moods, to make the human experience itself richer. However, this implies delegating day-to-day tasks to machines, which will need to have access to personal data, and therefore almost blindly trust that these machines will use that data and perform such tasks always and only for their human-owners' benefit. This requires absolute trust and delegation of control over part of our lives; obviously, the challenges are numerous on the way to create such intelligent ecosystems of sentient objects and benefactor applications. One of the first challenges lies in circumventing the resource limitations of IoT devices to have built-in intelligence. In this sense, on the one hand, there has been a great advance in the processing boards, while initiatives are emerging for optimizing ML environments and algorithms. Thinking about the potential that ML can bring to IoT applications, the industry has recently created new hardware specifically tailored to develop ML-based solutions. For example, OpenMV (https://openmv.io) is a tiny open hardware kit for IoT developers with an embedded camera that can detect faces and find eyes using built-in feature detection algorithms and consuming very little energy. On the other hand, the resource constraints of some devices can be offset by collaboration between them, provided that interoperability and efficient coordination issues are resolved. Either way, the efficient use of both device and network bandwidth resources will always have to be considered. Therefore, the data preparation and preprocessing step, typical of an ML life cycle, has a second goal, which is not only to improve data quality but also to filter out unnecessary data, reducing the amount of traffic and the processing in devices upstream. There is a need to intelligently reduce data on the way from its source, with a balance between quality and quantity. While redundant and spurious data is eliminated, raw data must gradually be augmented with semantics as it moves toward the applications that will consume it, so as to generate useful and actionable knowledge as the process outcome. Another challenge is related to the nature of data in IoT, in particular, that it is highly perishable. To perform day-to-day tasks, IoT applications need to make real-time decisions backed by data that quickly becomes obsolete and loses its value if not immediately used in processes. Therefore, intelligence must not only be embedded in objects but generated over moving data. In this context, a promising approach is adopting incremental and on-line learning algorithms. Losing *et al*. [15] define on-line learning algorithms as incremental learning algorithms which are additionally bounded in model complexity and run-time, capable of lifelong learning on a constrained device. Incremental and online learning algorithms aim for minimal processing time and space, and thus fit in IoT and CPS data processing environments [7]. In addition to built-in and online intelligence, a high degree of context-awareness and adaptability is required for humans to be able to rely on the support of IoT objects and applications, and trust that their needs will be met. The complexity of intelligent environments and the number of parameters that need to be taken into account in decision making and inference processes make constant context monitoring necessary. By context, we understand everything that affects the extraction of useful and actionable knowledge and the behavior of smart applications, from the user's agenda to the resources available on computational nodes. Therefore, software agents should monitor all relevant contextual aspects, and provide them as inputs to other agents who continually adapt system behavior in feedback loops. At the top of decision-making processes are a set of adaptation policies. To achieve the desired customization and allow sophisticated (re)configuration of parameters by end users, not necessarily programmers, requires tools to express needs and policies, and then translate to system commands. Last but not least, we need to ensure the robustness and reliability of intelligent IoT systems and the privacy of data. These are critical requirements that must be considered from the design phase of an IoT system, and dealt with not in isolation, but holistically in order to build not only an intelligent but a secure and resilient infrastructure. This is certainly an open issue that will challenge the community in the medium term.

> "Intelligence must not only be embedded in objects but generated over moving data. In this context, a promising approach is adopting incremental and on-line learning algorithms."

Cintia: IoT devices are usually very constrained, and it is not likely that complex AI/ML algorithms will be executed on these types of nodes. On the other hand, the edge of the IoT could gather these data and apply such algorithms. AI/ML algorithms could support decisions concerning network management.

Markus: So, Sumi, do you agree with Flavia's and Cintia's points of views?

Sumi: In general, I believe this will greatly depend on the application. Once data hit the edge or the cloud, AI and ML can be applied as required by the specific applications as required. Having said that, there are a few opportunities where AI can tremendously help IoT architecturally, regardless of the applications. The first is understanding the application domain, whatever it is, and using Representation Learning techniques, for instance, to automate gleaning the salient features of the applications and its data queries. The second is understanding, modeling and predicting the data domains of the IoT, a process that traditionally relied on statistical approaches such as the Auto-Regressive Moving Average model (ARMA) [16] and the application-aware ARMA [17]. Applying AI techniques to better understand a data domain seems very promising, especially if such use of AI is possible at the thing level, which requires "embedded intelligence", a code name for a lighter-weight AI capable of running on a thing directly, trading little accuracy for massive savings in processing and power needs. The third opportunity is simply linking up the first two opportunities together and understanding the interplay

between the applications and the data. Using combined insights of both applications and data will potentially be a game changer in many ways, including enabling sentience-efficiency, which I mentioned earlier, which is a big deal. In which other ways could this be a game changer? That could be a panel by itself.

Jay: I have a different take on this. The idea that the data is noisy and hence will not be that useful with machine learning may not be true with the use of deep learning approaches. The main reason behind the popularity of deep learning is that given a labeled (also unlabeled) dataset, it can identify the patterns in the dataset to help with classification of newly generated data. This procedure requires a user to tweak the number of hidden layers and the other hyper-parameters in most of the cases to result in good outcomes. Deep learning seems to do pretty well in the presence of noise compared to the state-of-the-art, which makes it popular. In addition, the ideas of federated and transfer learning implies that ML models can be deployed at the edge of the network. These models can learn their enhancements (transfer learning) in the context of the environment. This data, from many edge nodes, will be sent to the central ML brain to update the model and deploy it as an extension back at the different edge ML locations. This iterative mechanism, over several iterations, will weed out the impact of the noise in the observation and test datasets.

"Edge computing presents interesting economic model design challenges for both deploying the infrastructure as well as managing the pre-processed and post-processed data."

Security and Privacy

Markus: Security and Privacy are fundamental overall concerns in networked systems and systems of systems, such as with IoT [18]. But when we regard the growing entanglement of IoT with data analytics and Machine Learning, which new impacts do you think this lack of security and privacy may have on the quality, the reliability, availability and the timeliness of smart learning systems?

Cintia: Security and privacy are important concerns for IoT. There are studies showing which algorithms and mechanisms could be applied (such as [19]), there are communication standards available (such as DTLS), and also several studies showing the DoS attacks that could be launched from IoT devices. On the other hand, security is often neglected when IoT devices are deployed. Thus, the main concern I see is how such algorithms and mechanisms will be employed when deploying IoT devices and applications.

Jay: In fact, the security of the users, their data, and also the data from the devices is of paramount importance in VLIoT. This is especially true because in VLIoT there will exist many different IoT devices in the vicinity of each other and sensing the environment in different ways (video, sound, heart beat, etc.). These IoT will result in the creation of multimodal data, which has been known to enable inferencing with more ease. But without being able to trust the data or being able to verify their provenance, such fusion of data will lead to low confidence in the data. With the diversity in the VLIoT universe, this will become a major security obstacle. The other aspect is the demonstrated techniques that show that an ML algorithm can be deceived into misclassifying an event. This is commonly done using an adversary crafting attack where an adversary trains their neural network (a surrogate model) to generate an input that when given to the base ML model will result in misclassification. This field of research is new and needs significant exploration to ensure we do not adopt and deploy ML without understanding the potential attack vectors. I believe that this will be a major challenge to the adoption of ML in a majority of the applications that use ML [20].

Markus: Yes, and I would add the additional peril that viruses or network infections may silently introduce data transformations and aggregations that "slightly manipulate" or twist the data so that data analytics will derive incorrect information from it and lead to incorrect behavior of the Cyber-Physical System. A similar threat is in place when this twisted data used to learn the ML algorithm is manipulated.

Sumi: My fear is slowness in adoption and market hesitation. The sooner we are able to find out how to make the IoT more secure, privacy-preserving and identity-theft free, the sooner IoT will take off. In fact, Gartner's estimates of how many IoT devices we may have by 2022 may not at all be accurate as we have to first establish trust, and then estimate more accurately within a trust context conductive of adoption. Perhaps this explains Gartner's focus in 2019 on only Industrial IoT estimates rather than the prior years' generalized reports.

IoT Economic Models

Markus: As more and more applications become Cyber-Physical, requiring continuous monitoring of automated reaction to events of the physical world and the system resources itself, to me it seems clear that data collected by, and control issued by, the system will gain in social-economic value. And this will lead to an IoT or Cyber-Physical economy [21], where data-and-control might be sold, refined, curated, purchased and exchanged in a similar way as nowadays we do this with physical goods, services and commodities. So I would like to hear your ideas on such a forthcoming new digital monitoring and control economy.

Jay: Edge computing presents interesting economic model design challenges for both deploying the infrastructure as well as managing the pre-processed and post-processed data. In particular, at the edge the infrastructure may be deployed by the cities, the Internet service providers (ISPs), or the cellular network providers (AT&T, T-Mobile, etc.). The edge computing software (software-as-a-service) is likely to be deployed by software providers, such as Amazon or Microsoft. There is a need for a cost-sharing model between these entities. Further, an end user who is a subscriber of one particular company (e.g., T-Mobile) can be served by an edge server running on the AT&T network. In that case, a mechanism needs to be in place for AT&T to provide the service while getting paid by T-Mobile later after demonstrating that the service was provided. For these monetary exchanges, distributed ledger based technologies look promising. The end user's data is received and processed at the edge servers, who perform operations resulting in processed information (e.g., annotation of videos or pictures, overlaying of images, processing delays for travels, etc.) relayed to the end users. However, the data stays with the edge servers and can be used by the edge servers and other entities as needed. These entities should remunerate the end user for using their data. Mechanisms could be designed to monetize the use of the user data and also to seek user permission before the data is released to a new entity or purposed for new use.

There is potential for an independent entity serving as a data/financial clearing house at the edge. The clearing house entity can serve as the conduit for all the communications and ensure traceability of where the user data went and create a mechanism for user payments. The payments can happen on a cryptocurrency network or through some common financial clearing house. However, such a setup will only work if all entities are trustworthy and follow the protocols faithfully. If they do not, a technology based check and balance mechanism to enforce compliance of, say, the General Data Protection Regulation (GDPR) stipulations or to prevent unauthorized sharing, is difficult to design. Then the problem can potentially be addressed by the legal/judicial and regulatory systems.

Markus: I agree that general-purpose Edge computing is likely to give birth to new and interesting economic market models for IoT, where services such as traceability, authentication, access control, data curation, aggregation and context enrichment, etc.

will have a price and can be remunerated to companies that operate edge devices scattered in all corners of smart buildings, highways, and smart cities. But since data (and control) are the most important commodities in IoT applications, it is natural that besides basic resources (cloud, network and IoT devices) also big data services, ML models and verified, cleaned and enriched training data for Machine Learning will also become a valuable product in marketplaces for Cyber-Physical Systems (CPS). For example, the authors in [22] describe a platform where consumers and data providers can transact verified training data or learning settings, in order to enable fast learning of some behavior (a detection of reaction on an environmental event) from third party data sets. It thus facilitates transactions of data from several providers toward the IoT consumer's learning system components, by forming a dataset that is close/similar to the validation dataset, so that supports the quick operation of an IoT system with "fast"-learned behavior. Of course, all this transfer and monetization of the data has to be done in a trusted, fair manner while preserving data ownership and the consumer's privacy. Although not directly tailored to IoT, there is Agora [23], a scalable infrastructure (data ecosystem) for fine-grained AI, data science and high-quality data asset exchange, and a platform for access to distributed computing and storage resources.

> "There is potential for an independent entity serving as a data/financial clearing house at the edge. The clearing house entity can serve as the conduit for all the communications and ensure traceability of where the user data went and create a mechanism for user payments."

Conclusion

This article is the second part of a two-part article discussing the challenges and opportunities in the intersecting areas of IoT, Data Science and Machine Learning. The first part debated and reported on challenges and opportunities in the infrastructural aspects of such systems. This part debated and reported on issues related to event processing, embedded machine learning, security and the potential business models that could drive such intelligent and data-driven IoT systems in the future. We hope that you find the panel coverage useful and helpful in charting out future research directions.

Acknowledgments

Flavia Delicato is supported by CNPq grant number 306747/2018.9 and by FAPESP (São Paulo Research Foundation) grant number 2015/24144-7; Cintia Margi is supported by the ELIOT project, FAPESP (São Paulo Research Foundation) grant #2018/12579-7; Markus Endler is supported by CNPq proc. 433183/2018-7 and INCT of the Future Internet for Smart Cities funded by CNPq, proc. 465446/2014-0, CAPES – Finance Code 001, and FAPESP, proc. 2014/50937-1 and 2015/24485-9; Satyajayant "Jay" Misra is supported by U.S. NSF grant awards #1800088, #1719342, #1345232, EPSCoR Cooperative agreement OIA-1757207, and Intel grant #34627535. The opinions presented here are those of the author Misra and not the official position of the federal government and Intel Corporation.

References

[1] M. Endler and F. S. Silva, "Past, Present and Future of the ContextNet IoMT Middleware," *OJIOT*, vol. 4, no. 1, 2018, pp. 7–23.
[2] L. E. Talavera *et al.*, "The Mobile Hub Concept: Enabling Applications for the Internet of Mobile Things," in *Pervasive Computing and Communication Workshops (PerCom Workshops)*, March 2015, pp. 123–8.
[3] E. G. Renart, J. Diaz-Montes, and M. Parashar, "Data-driven Stream Processing at the Edge," in *2017 IEEE 1st International Conference on Fog and Edge Computing (ICFEC)*, 2017, pp. 31–40.
[4] S. Choochotkaew *et al.*, "Edgecep: Fully-distributed Complex Event Processing on IoT Edges," in *2017 13th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, 2017, pp. 121–9.
[5] H. Storf *et al.*, "An Event-Driven Approach to Activity Recognition in Ambient Assisted Living," in M. Tscheligi *et al.*, *Ambient Intelligence*, AmI 2009, *Lecture Notes in Computer Science*, vol 5859. Springer Verlag, 2009.
[6] M. Gualtieri and R. Curran, "The Forrester Wave: Big Data Streaming Analytics, Q1 2016 Streaming Analytics Are Critical to Building Contextual Insights for Internet-of-Things, Mobile, Web, and Enterprise Applications, 2016.
[7] X. Fei *et al.*, "CPS Data Streams Analytics Based on Machine Learning for Cloud and Fog Computing: A Survey," *Future Generation Computer Systems*, 90, 07 2018.
[8] S. Choochotkaew *et al.*, "Edgecep: Fully-distributed Complex Event Processing on IoT Edges," in *2017 13th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, June 2017, pp. 121–9.
[9] I. Akkaya, "Data-Driven Cyber-Physical Systems via Real-Time Stream Analytics and Machine Learning," Ph.D. thesis, EECS Department, University of California, Berkeley, Oct. 2016.
[10] F. Ganz *et al.*, "A Practical Evaluation of Information Processing and Abstraction Techniques for the Internet of Things," *IEEE Internet of Things J.*, vol. 2, no. 4, Aug. 2015, pp. 340–354.
[11] R. Dautov, S. Distefano, and R. Buyya, "Hierarchical Data Fusion for Smart Healthcare," *J Big Data*, 2019.
[12] B. Costa, P. F. Pires, and F. C. Delicato, "Modeling SOA-based IoT Applications with soaml4IoT," in *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*, April 2019, pp. 496–501.
[13] M. Pitanga Alves, F. C. Delicato, and P. F. Pires, "IoTA-MD: A Model-driven Approach for Applying QoS Attributes in the Development of the IoT Systems," in *Proceedings of the Symposium on Applied Computing, SAC '17*, New York, NY, USA, 2017, ACM, pp. 1773–80.
[14] B. Costa, P. F. Pires, and F. C. Delicato, "Modeling IoT Applications with sysml4IoT," in *2016 42th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, Aug. 2016, pp. 157–164.
[15] V. Losing, B. Hammer, and H. Wersing, "Incremental On-line Learning: A Review and Comparison of State of the Art Algorithms," *Neurocomputing*, vol. 275, pp. 1261–12, 2018.
[16] ARM Mbed Operating System Architecture, https://www.arm.com/products/iot/mbed-os, 2018, accessed Nov. 19, 2019.
[17] Yi Xu *et al.*, "Energy Savings in Very Large Cloud-IoT Systems," *Open Journal of Internet of Things (OJIOT)*, vol. 5, no. 1, 2019, pp. 6–28, presented as the Keynote of the *Very Large Internet of Things (VL-IoT) Workshop*, in conjunction with the VLDB, Los Angeles.
[18] Z. Zhang *et al.*, "IoT Security: Ongoing Challenges and Research Opportunities," in *2014 IEEE 7th International Conference on Service-Oriented Computing and Applications*, Nov. 2014, pp. 230–4.
[19] G. C. C. F. Pereira *et al.*, "Performance Evaluation of Cryptographic Algorithms over IoT Platforms and Operating Systems," *Security and Communication Networks*, 2017:2046735:1–2046735:16, 2017.
[20] M. Biswal, S. Misra, and A. Tayeen, "Black Box Attack on Machine Learning Assisted Wide Area Monitoring and Protection Systems," in *IEEE North America Conference on Intelligent Smart Grid Technologies (ISGT)*, 2020, pp. 1–4.
[21] L. G. Pitta and M. Endler, "Market Design for IoT Data and Services, the Emergent 21st Century Commodities," in *IEEE Symposium on Computers and Communications*, June 2018.
[22] K. K. Sarpatwar, "Blockchain Enabled AI Marketplace: The Price You Pay for Trust," in *CVPR Workshops*, 2019.
[23] J. Traub *et al.*, "Agora: Towards an Open Ecosystem for Democratizing Data Science Artificial Intelligence," ArXiv, abs/1909.03026, 2019.

Biographies

Abdelsalam (Sumi) Helal is a professor and Chair of Digital Health in the School of Computing and Communication at Lancaster University. He is also a professor at the University of Florida. He has made significant contributions in the areas of digital health, pervasive and mobile computing, distributed databases and the Internet of Things.

Flavia Delicato is an associate professor at Universidade Federal Fluminense and also a collaborator researcher at the Centre for Distributed and High-Performance Computing (University of Sydney, Australia). Her primary research interests are IoT, WSN, middleware and Edge computing.

Cintia Borges Margi is an associate professor in the Computer and Digital Systems Engineering Department at Escola Politécnica, Universidade de São Paulo. She does research in Wireless Sensor Networks and Software Defined Networking.

Satyajayant "Jay" Misra is a professor in computer science at New Mexico State University. He has contributed to protocol design for anonymity, security, and survivable systems of the Future Internet, super-computing, and IoT/Cyber-Physical Systems architectures.

Markus Endler obtained his Dr. rer. nat. degree (Technical University of Berlin) in 1992, and has been a professor livre-docente (University of So Paulo) since 2001, and a CNPq Researcher. In 2001 he joined the Department of Informatics at Pontifícia Universidade Catlica in Rio de Janeiro (PUC-Rio), where he is currently an associate professor. His main research interests include mobile computing, distributed pervasive systems, context awareness and Internet of Mobile Things.