Machine Learning/AI as IoT Enablers

Yue Wang, Maziar Nekovee, Emil J. Khatib, and Raquel Barco

1 Introduction

Recent years have evidenced a rapid growth in the application of advanced Artificial Intelligence (AI) technologies in numerous fields, such as industry, healthcare, transportation, and domestic appliances. AI is a form of computing that allows a machine to perform cognitive functions, such as adapting their behaviour and modifying their decisions according to changing environment and conditions. Machine learning (ML) is an application of AI that provides a system the ability to automatically learn and adapt to the environment through experience. In particular, machines and tools that support AI are designed to react and learn from data collected from the environment, and the knowledge and insights created from them, through data analytics. Data analytics discovers new knowledge and creates new value through the exchange, selection, integration, and analysis of massive data. It provides a technology that reveals the knowledge and correlation in systems that may not be discovered or fully described with conventional mathematical models. The properties and problems of data analytics vary when the volume, generation velocity, and variability of the collected data grow above a certain threshold, entering into the Big Data analytics realm. To support these conditions, novel technologies have entered the market, such as cloud computing and NoSQL databases. Big Data analytics, combined with the underlying AI technologies, have

Y. Wang (✉)
Samsung Research UK, Staines, UK
e-mail: yue2.wang@samsung.com

M. Nekovee
University of Sussex and Quantrom Technologies Ltd, Brighton, UK

E. J. Khatib · R. Barco
University of Málaga, Málaga, Spain

N. H. Mahmood et al. (eds.), *Wireless Networks and Industrial IoT*,
https://doi.org/10.1007/978-3-030-51473-0_11

found their applications in all aspects of business, society, and life, which are reshaping our future technology landscape.

Industry 4.0 is one of the main consequences of this data-centric revolution. Information on the processes, consumer demands, supply chain, etc. become a necessity to achieve the flexibility and agility required in Industry 4.0. To obtain this information, data must be intensively collected by different kinds of IoT sensors (product tracking, environmental monitoring, etc.) and processes (online shopping trends, machine status information, network traffic information, etc.). The collected information is then stored and processed using the aforementioned Big Data storage and analytics technologies, etc. Wireless connectivity plays an even more important role in industrial environments, due to the ease of deployment, low maintenance costs, and the high flexibility they offer.

Future wireless networks are data-intensive and service-driven. The adoption of wireless technologies has enabled a new paradigm in connectivity and computation, where machines have access to the Internet to autonomously send data and receive instructions. These machine-to-machine (M2M) communications, which have varying characteristics and requirements, have enabled a rich set of novel applications, and, combined with mobile computing devices, shaped the Internet-of-Things (IoT). In IoT, novel applications have appeared, such as smart wearables, smart mobility, smart utility management, eHealth, virtual/augmented reality, ultra-high definition (UHD) video, driverless cars, etc. It has been predicted that around 25 billion IoT devices will be connected by 2025 [1]. Specifically, cellular technologies are seeing a great adoption by the IoT market, thanks to the ubiquitous connectivity they offer, plus their ease of use and maintenance from the point of view of the clients. 5G technologies, with their capability of providing high data rate, low latency, and guaranteed services through network slicing, are designed to cater for the needs of different IoT applications. The connection of the massive numbers of devices will generate a huge amount of data, gathered by individual devices, and shared over the IoT network in near real time.

An important point to take into account in industrial IoT networks are the particularities of the scenarios where connectivity occurs. Industrial environments such as factories or distribution centres are especially harsh for radio propagation, due to the presence of large metallic structures that cause shadowing and a large number of transmitters that produce interference. Therefore, a key point in deploying intelligent connectivity in industry, and a major differential factor with respect to the general use cases, is to use the appropriate Radio Access Network technologies and be especially careful with their dimensioning.

The accumulation and sharing of massive amounts of data and knowledge will in turn facilitate AI and ML, enabling the so-called intelligent connectivity – a vision of future network empowered by the combination of emerging technologies, including 5G, AI, ML, Big Data, and IoT [2]. Underpinned by ubiquitous hyperconnectivity, as well as real-time decision making with collective intelligence, intelligent connectivity is foreseen to transform industries such as energy, transportation, and manufacturing, as well as every aspect of our daily lives.

To fully unleash the potential of intelligent connectivity, there are some challenging topics that must be addressed, for instance, data and intelligence sharing, scalability of the existing solutions, security, and the underlying transformation in infrastructure. In particular, data analytics and AI face some unique challenges when applied to IoT networks. Firstly, a great variety of device types exist, and data collected from different devices may follow different format with different data types. It is a huge task to harmonize the collected raw data into a universal language where insights and knowledge can be shared. Secondly, the constantly changing network conditions and surrounding environment needs to be detected and the connectivity methods adapted by the AI algorithms, ideally in real time. This implies a fast exchange of information and knowledge, as well as a need for selecting what data to pass on and what data to retain locally in the device. In addition, there is a problem of model applicability. The majority of the current AI models deployed in IoT networks are based on exhaustive experimenting over available data, so these models are highly adapted to the existing datasets. There is a significant problem of reusing and scaling the existing AI models extracted in one scenario to a different scenario, or a different part of the network for the same application. For example, an AI model extracted from a specific industrial process in a small factory cannot be easily scaled to larger factories.

This chapter provides an overview of the current and future applications enabled by the merging of AI, 5G, and IoT, and their future looking technologies.

2 The Role of AI and Big Data

AI and Big Data, as the key enablers of intelligent connectivity, have been evolving hand in hand with emerging IoT technologies, where the most significant sources of data are generated. Considerable interest from the industry and research efforts have been attracted to this field. In the developments both from academia and industry, data mining and ML are used to extract the insights and knowledge from the data collected by IoT networks.

AI is a set of techniques and algorithms that are meant to perform actions that usually would require human intervention. AI algorithms are ultimately functions that, given a certain input, return a corresponding output through a non-linear relation. The inputs are usually a set of complex observations, which may require a pre-processing with operations such as quantification or normalization. The outputs are dependent on the application where the algorithm is used and the nature of the algorithm. The type of output defines a taxonomy where AI algorithms can be grouped as classifiers, regressors, etc. The non-linear relation between the input and the output is shaped by a set of parameters that are commonly complex and hard to adjust. Some examples of AI algorithms are Recurrent Neural Networks, Fuzzy Logic Controllers, and Bayesian Networks. Figure 1 summarizes how AI methods are used in intelligent connectivity, and their relation with ML and Big Data Analytics. In intelligent connectivity, AI algorithms take as inputs the data

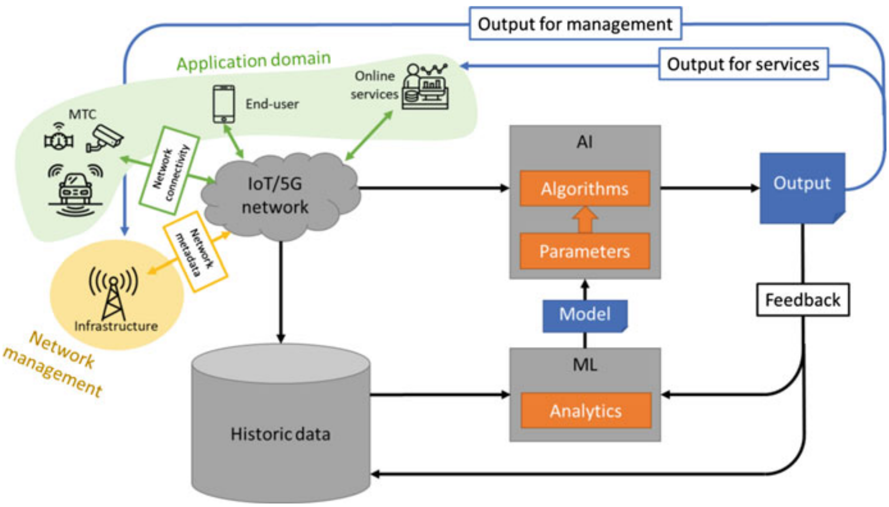

Fig. 1 General scheme for intelligent connectivity

collected by the IoT devices in the network, along with data from other sources such as the network infrastructure or external online services; and produce outputs that can be used to interact with the IoT applications and services, or to improve the connectivity by modifying network configuration.

AI algorithms have a large number of configuration parameters that must be fine-tuned to a specific scenario to work correctly. Although there are AI systems where these parameters are adjusted manually, ML is more often used to do this. In these kinds of setup, ML algorithms take as input large historic datasets similar to those that the AI algorithm will process once deployed and return as output optimal sets of configuration parameters.

- *Supervised learning:* The ML algorithm has access to sets of input variables of the AI and the expected output (labeled data). In this case, the ML needs to configure the AI so that it imitates the process that generated the training samples in the first place. Although supervised learning usually produces AI systems that need less post-processing and that have higher accuracy, one major issue is the availability of training data. Some common examples of supervised learning algorithms are Deep Learning and Support Vector Machines.
- *Unsupervised learning:* The ML algorithm only has access to sets of input data (unlabeled data). In this case, the ML will search for patterns and train the AI to find them. These systems will usually produce less accurate AIs, but accessibility of unlabeled data is much easier. Some examples of unsupervised learning are clustering and anomaly detection.
- *Reinforcement learning:* In the third kind, which is sometimes classified as a kind of supervised learning, the ML algorithm has access to the input data of the AI and, although it cannot access the expected output, there is a certain

feedback that indicates whether the output produced by a trained AI is correct or not. In this case, the ML will train the AI algorithm in a case-by-case fashion. Q-learning is an example of reinforcement learning.

AI and ML algorithms are based on the processing of large amounts of data. Both the storage and the processing of these data consumes a lot of resources. In fact, above a certain threshold, traditional computing techniques are insufficient for the successful execution of some AI/ML systems. This is where Big Data technologies come into play. Three features determine whether a problem can be considered part of this Big Data domain [3]:

- High volume: A very large amount of sources produce data. This is true for IoT networks, where a very high number of devices produces large amounts of data.
- High variability: The data from different data sources come in different formats that requires harmonization. In an IoT network, devices of different models, manufacturers, and purposes operate, producing data in many different formats (numerical records, audio/video files, etc.).
- High velocity: Data is generated quickly, that is, faster than it can be processed by traditional methods. In applications where speed is important (such as self-driving cars), processing the data fast is critical; and this is difficult when the maximum allowed delay is close to the minimal processing time.

In the case of intelligent connectivity, the collected and processed data has all the three features. Big Data techniques help to overcome these challenges by offering special storage and processing methods. NoSQL databases improve the storage and retrieval of data with high variability (i.e. data that may have different formats at different moments). Cloud computing is a set of Big Data technologies for improving the speed of processing. In cloud computing, tasks are divided into many parallel processes, reducing the overall computing time. Schemes such as Map-Reduce [4] and the Lambda architecture [5] are examples of Big Data processing techniques.

3 Use Cases of AI-Enabled Intelligent Connectivity

So far, the mainstream applications of AI in technologies include computer vision, natural language processing, voice recognition, and prediction. These technologies can be widely used by end consumers and businesses. Table 1 below gives an overview of the applications of AI in different technologies and their application scenarios for consumers and enterprise customers. In Fig. 2, we provide an overview of some of the use cases on AI-enabled intelligent connectivity. Next, we describe a few use cases of AI-enabled intelligent connectivity.

Table 1 AI algorithms and their applications

Technology	Overview	Application Scenarios
Computer vision	Computer replaces human vision to recognize, follow and measure the objects	Smart home AR, VR Shopping via image searching Intelligent home security 3D analytics
Natural language processing	Interpret meanings of texts and extract abstracts from articles	Search engine Recommendations and advertisement Machine translate
Voice recognition	Translate human instructions to texts and commands to machines	Smart TV Call centre Voice assistant Smart home assistant
Enterprise applications	AI applications for third-party, business customers	Network management Stock exchange Production planning

Fig. 2 Overview of intelligent connectivity use cases

3.1 Smart Manufacturing

In the last years, market trends have driven to a demand of highly customized manufacturing goods. To efficiently serve this new customized market, where production volumes of a single product are low, but total sales keep increasing, factories need to adopt agility as a basis for their operation. Agility is achieved with a vast set of novel technologies collected under the umbrella of Industry 4.0 [6]. Wireless connectivity, Big Data, robotics, and sensors are the four pillars of Industry 4.0. intelligent connectivity, as a combination of Wireless connectivity and Big Data, plays a major role in many Industry 4.0 applications. In this section, two of these applications will be described: predictive maintenance and hazard detection.

Predictive Maintenance As costs of production grow, the need for cutting expenses is an ever-increasing need in industry. In industrial machinery, there are two sources of expense: waste of unspoiled elements and machine breakdowns. To be more specific, some industrial machinery require a periodic maintenance, which can be done proactively or reactively. The first approach implies that some wear items (such as metallic pieces that are subject to stress, or parts that perform abrasive processes) may be changed before their lifespan is consumed, increasing the expense in replacements. On the other hand, the reactive approach consists of only replacing parts once they wear off and cause a malfunction. Although this means that the expendable elements are fully used, they may cause machine breakdowns that increase the cost with the need of repairs. Therefore, there is a need to optimize the scheduling of predictive maintenance so that the wear elements are fully used without causing breakdowns.

Sensors are one of the key technologies in Industry 4.0. In recent years their cost has dropped, making vast deployments affordable from an economic standpoint. Novel IoT technologies, such as cloud-based platforms for collecting and processing sensor data, greatly simplify the process of sensorization at large scale. Information from many kinds of sensors can be collected, and models of the monitored processes can be extracted with ML processes. These models can then be used to perform predictions.

This scheme of heavy monitoring, modelling, and prediction can be used for predictive maintenance. Depending on the specifications of the machine, magnitudes such as vibrations, flow of fluids, electric current, conductivity, and thickness of certain pieces can be measured. Supervised ML using the collected data from reactive maintenance, can determine which variables contain information on an immediate breakdown, leading to an interruption in the operation of the machine next time maintenance is required. This achieves both objectives of fully utilizing wear elements and preventing breakdowns.

Hazard Detection Factories, having large and powerful machines, are dangerous places. Some dangers to personnel are fires, accidents with machinery, accidents with vehicles, toxic fumes, falling objects, and incorrectly isolated electric lines. According to Eurostat, over 3000 deaths per year are registered in the EU in the sector of agriculture, construction, and manufacturing. Strict regulations are in place to minimize these fatalities, determining how machines and buildings must be built in order to prevent hazards and facilitate danger mitigation measures. Active safety systems, such as automatic fire extinction systems, also help reduce the risk and incidence of accidents. A critical aspect of safety in factories is to detect hazards early, before they cause accidents or irreversible situations. These hazards may be varied in nature, determining the mechanisms that can be used for their detection. Smoke detectors, radiation detectors, or temperature sensors are some common examples.

Video analysis is a function that can be performed with intelligent connectivity. Processes such as object recognition or movement detection can be performed in the cloud with AI using video feeds from connected cameras. To train the analysis AI algorithms, videos of known activities can be used to feed supervised learning processes to train a model. Another option is to model normal behaviour and train the AI to recognize when abnormal activity occurs. The output of video analysis can then be fed to other systems to perform certain activities, such as raising alarms or activating actuators through integrated IoT platforms.

Video analysis can be used to detect hazards such as smoke, fire, or even sabotage in factories. Since surveillance cameras are usually deployed in factories, video analysis can be deployed over their feeds to increase the coverage of hazard detection, reducing the reaction time for active safety systems.

3.2 Connected Cars

As novel technologies arrive to the mobile communications market, an increasing demand to integrate them into vehicles is growing [8]. Services such as video streaming are starting to be part of on-board entertainment systems. But, beyond entertainment for passengers, mobile communications can offer some very interesting services to assist drivers: collision avoidance systems, navigation, predictive maintenance, etc. Mobile communications will also play a major role in the future of self-driving cars, where it is expected that autonomous cars will communicate among each other to create self-organizing traffic patterns. Two examples where intelligent connectivity has a central role are the transmission of traffic-related warnings to drivers and remote driving.

Connected cars also has the potential to transform logistics, which is a key aspect of supply chain management in industries.

Traffic-Related Warning Transmission Although fully autonomous self-driving cars are quickly becoming a realistic possibility, there is yet a long way to go in its social and legal aspects. In the first place, the psychological implications of not having any control over a self-driving vehicle cause a general rejection over the wider public. Also, issues like the coexistence between autonomous and non-autonomous traffic, the ethical dilemmas on what decisions should AI systems take in case of emergency, the liabilities in case of accident, etc. are debates that are still open to resolution. Therefore, the implantation of these technologies is taking place in a gradual manner; starting with technologies that assist drivers and provide information for better decision making. Information on traffic, road conditions, tolls, accidents, or weather all help drivers to plan their routes or be especially cautious at certain moments.

One of the main use cases of intelligent connectivity is the obtention of information from sensors deployed over large geographic areas. The two basic building blocks of this use case are the sensors themselves, and the wireless technologies that provide connectivity. The reduction of the price in sensors and the availability of low power systems in the last decades has made the deployment of massive amounts of sensors economically feasible. But thanks to AI and ML, other devices, such as cameras, smartphones or network access points, can be used to extract additional information after some processing. Image recognition, location analysis or network traffic modelling are just some of the processes that can be used to obtain rich information from these devices. Regarding wireless networks, wireless access network (WAN) technologies, such as 5G or 6LoWPAN, enable low cost and low power connectivity both for sensors and for users of the information.

For traffic information, intelligent connectivity can be used to gather and centralize all the useful information. This information may come from sources such as road cameras, which can be used to measure the traffic and detect jams through image recognition. Other incidents, such as oil stains over the asphalt, can be recognized either by cameras or by sensors in the cars that upload this data to the cloud. Collision detection systems, which are currently being installed in modern vehicles and are being enforced by legislation, can report accidents. All this information can then be curated and customized for each driver based on the analysis of their trajectory, which can be obtained inspecting the geolocation information of smartphones. As a result, drivers will have an updated and simple newsfeed on their dashboards, that warns them of any important event they might encounter in the near future.

Remote Driving Another intermediate steps towards full automation is remote driving. In this stage, although there is still a human making the decisions, the driver is in a remote location, so there is a delay in the feedback that, if not appropriately dealt with, may cause accidents due to late decisions.

Mixed-Criticality Systems model devices where several different information paths coexist. Some are more proprietary, therefore they are processed earlier, having to wait less in queues, and receiving more resources (such as increased CPU frequency) when needed. The wireless network can also establish different policies for different kinds of traffic. Currently, 5G networks consider three main traffic types: Enhanced Mobile Broadband (eMBB), Massive Machine Type Communications (mMTC), and Ultra-Reliable Low Latency Communications (URLLC). This differentiation allows to adapt the resources available in the Radio Access Network carrier and the Core Network connections to better serve the needs of each type of message.

In remote driving, video feeds are a very important data source, allowing the driver to visualize the environment. To transmit a high-resolution view, eMBB connections are required in order to provide the required bandwidth. On the other hand, when a sudden obstacle, such as an animal, appears in view, eMBB may introduce a high latency, so a URLLC message showing the danger would be required. To differentiate when this warning must be sent, image recognition must be running at all times in the car's CPU using a high-priority process. Collision avoidance between cars must also be dealt with high priority. In this case, geolocation information must be sent regularly by vehicles to the network, and a collision prediction must be run in the network edge, where the information of neighbouring vehicles can be aggregated. Once a potential collision is detected, a warning can be sent to both drivers using URLLC.

3.3 Next-Generation Healthcare

Applications of IoT in healthcare seem to be endless: from remote monitoring and personal healthcare to smart sensors and medical device integration, as well as the pharmaceutical industry, healthcare insurance, healthcare building facilities, robotics, smart pills, and even treatments of diseases [7]. It has the potential to not only keep patients safe and healthy, but also to improve how physicians deliver care. In the following we will focus on a few prominent IoT use cases in health with the greatest potential from AI.

Remote Patient Monitoring Personal health and medical data are collected from an individual and transmitted to a provider for use in care and related support. In this way the provider can track healthcare data for a patient once released to home or a care facility, reducing readmission rates. Healthcare devices as insulin pumps, defibrillators, scales, continuous positive airway pressure machines, cardiac monitoring devices, and oxygen tanks are now connected in the IoT to ensure remote monitoring, providing patients and their caregivers valuable real-time information.

IoT-supported healthcare services can provide better and more efficient treatment to patients while also inducing cost saving for the providers. On the other hand, interconnectivity can provide for easy data collection, asset management, Over-the-Air updates, and device remote control and monitoring.

Assisted Living Demographics, public policy, and the labour market are driving an emerging market for IoT to deliver elder care services. By 2029, 20 percent of the U.S. population will be over the age of 65 and 70 percent of those individuals will need some form of assisted care, according to recent research [16].

AI and the IoT have the potential to shape a new collection of technologies to improve the quality and availability of elder care while helping to control its costs. Ambient intelligence, which combines AI and IoT, will provide real-time monitoring of an environment and event-driven response to changes in that environment. Sensors designed to detect changes in sound, motion, physiological signals, as well as more generalized image processing are core components of an ambient intelligent environment.

Ambient intelligence thus is poised to serve a range of functions with regards to elder care, but most applications will address three broad functions: maintaining routine activities and social connectedness, enhancing safety, and monitoring health. Routine activities and social support are especially suited for elders suffering cognitive decline. These systems detect changes in patients' location or environment and provide verbal assistance as needed, or if needed, notify caregivers. Safety-enhancing sensors are often wearable and provide early warnings of potentially threatening situations, such as falls. Health monitoring systems may combine wearable and stationary sensors to monitor blood pressure, pulse, and movement of the patient as well as environmental data, such as ambient temperature.

Unlike IoT applications that function primarily to monitor and control devices or environmental conditions, ambient intelligence systems are designed to monitor and support humans, creating an additional dimension of complexity. Developers of ambient intelligence systems face challenges common to IoT as well as some specific to this domain. Real-time processing, quality control, and data integration are especially important when making decisions about the physical well-being of a patient.

4 Architecture for AI-Enabled IoT

Intelligent connectivity encompasses a wide set of ML and AI algorithms for a very wide array of solutions applied over a great variety of use cases. To implement these solutions in practice, the first question to resolve is its architecture, that is, what elements will be used, and at which location in the IoT system they will be set up. Figure 1 provides an overview of such an architecture. In this section, we will delve into the details of the elements of this architecture.

4.1 IoT Network

The IoT network has a main role as a gatherer of information. By providing connectivity to IoT devices, it collects all the data and redirects them to the services they are connected to. It also plays the reverse role, that is, to send commands and responses from the services to the devices.

In intelligent connectivity, the IoT network adapts itself to the connectivity needs of the devices, ensuring that they have the resources they need for their operation. This means that the network is reactive to external changes, and in some cases even proactive, in the sense that it uses predictions to adapt to these changes beforehand. Therefore, a secondary role of the network is as a client of the AI services. For this, the IoT network must share its configuration parameters and performance indicators with the AI/ML blocks, and use their outputs for self-configuration.

This adaptability functionality is especially important in harsh environments such as those found in Industry 4.0. Shadowing and interference are major problems, as earlier stated. Intelligent connectivity solutions can help in tasks such as the detection of coverage holes (i.e. zones in an industrial premise where no wireless connectivity is present), interference mitigation and load balancing. AI techniques can perform these functions, and even do it in a predictive manner.

Although the network is a central component to intelligent connectivity, it is also the component over which the least control is usually feasible by the industrial installation owners. Large deployments are usually undertaken and operated by network operators, while often the applications are demanded and developed by external entities with very specific needs. There is a need, therefore, for coordination among the different entities.

4.2 Databases

In intelligent connectivity, the IoT network "knows" where all the information from all the devices is located. This knowledge can be modeled as a single, huge database, where the AI/ML blocks can query specific data. Since the central database is actually a set of disperse network services, common formats [9] such as XML or JSON, and normalized interfaces such as REST [10] or Graphql [11], are key technologies to retrieve the information when required. Technologies such as NoSQL databases are used in online services to store very large amounts of schema-less data, which are common when the data sources (IoT devices, smartphones, etc.) are from different vendors. This technology can also be used by the IoT network to centralize the data from different services once they are queried by the AI/ML block.

Some important features of databases in industrial applications are their reliability, their performance and their security. Databases play a central role in data-centric applications, therefore, it is important that they are accessible at all times, information is not corrupted easily, since otherwise the outcomes of ML/AI processes would be affected and the cost due to errors would escalate. Performant databases are the basis of performant ML/AI algorithms that can cater for large data-centric applications and processes with very high throughput. Security is key to avoid industrial information theft and sabotage; and it is also a major selling point for owners, which ultimately helps in the expansion of the intelligent connectivity market.

4.3 AI and ML Components

The elements that will be used for the AI block depend highly on the application, which imposes a certain output, for example, a classification label, or the prediction of a time series. Two boundary conditions must be set based on the requirements of the application:

- Selection of input data: to decide the datasets that will first train the AI algorithm with ML and afterwards be used as input of the trained AI algorithms, the main criteria is the availability of the information within the data. In other words, the first step to take is to assess what is the base dataset that contains the target information. The base dataset determines also the physical data sources that must be used (e.g. databases, file systems, devices, etc.), and the flow of data throughout the network. These are aspects that must be taken into account to ensure that requirements such as latency and reliability are fulfilled.
- Selection of the type of output: this decision depends on what the objective of the intelligent connectivity solution is. It defines what information will be extracted from the input data. There are many kinds of output; for instance, class labels, that classify a certain input dataset into one of a finite number of classes; or predicted values, that provide a value for a variable in the near future based on past values of that same variable or others. The output can also consist in model parameters, such as statistic indicators (averages, quantiles, etc.).

Once these boundary conditions are adjusted, the set of AI algorithms that can be used is narrowed down to those that can provide the expected output with the selected inputs. In some cases, some algorithms (such as Artificial Neural Networks) can be used for different kinds of applications (prediction and classification), but the mode of working with them, the set of selected inputs and their roles varies widely for each case.

Once the AI algorithm is fixed, the ML method that will train it must be selected. If the data available for training includes examples of the output, supervised learning can be used; otherwise, unsupervised learning must be chosen. In a system where ML is done online, that is, when the output of the AI is validated by an external factor and fed back to the ML algorithm, reinforcement learning can also be used.

Selecting the datasets and the ML/AI algorithm are the base of the intelligent connectivity solution design; but to actually implement the system other decisions must also be taken. Specifically these decisions affect where each of the functions composing the solution are implemented:

- Physical computing element: aspects such as the dimension of data or the complexity of the operations determine the required computing power. Also, the cost of such computing power must be taken into account.
- Centralized/decentralized architecture: this aspect of the architecture determines whether the algorithm is implemented in a single server (centralized) or aggregating the results of instances running in a distributed set of devices (decentralized).

These two architectural decisions comprehend a very large set of technologies that exist in the market nowadays. Also, the very broad range of ML/AI algorithms that exists in the market gives place to a very heterogeneous set of requirements. There are some algorithms that are very lightweight while others are demanding; some that are parallelizable while others are not; some that need a global view of the use case (for instance, they need to access data coming from many sensors as well as data saved in the cloud) while others only need a local view (for instance, only on the sensing device). All these considerations define the boundary conditions of the selected location for the implementation of the algorithm.

Considering the physical computing element and the centralized/decentralized decisions as a single issue, there are three main options for the implementation location:

- Local device: For algorithms that only need local visibility and are simple enough so that they can run within the computational resources offered by the device, Implementation in the same device is a possibility. The main advantage of this location is that the latency is very low. On the other hand for energy constrained devices this implies a higher consumption. This is an example of a decentralized implementation option. Different instances running in different devices can communicate among themselves using peer to peer communications.
- Remote device: The traditional client server scheme also has its place in intelligent connectivity. In this case the algorithms will not be running the devices but in a remote location. The devices only act as information collectors and actuators. Also data from other data sources (such as databases or the Internet) can be used in this scheme. Applications that require a global view must be implemented using this scheme. This is a traditionally centralized architecture, where a network connection is established between the devices (clients) and one remote computer (server). Nowadays this scheme is a little bit more complex but also more flexible, thanks to technologies such as virtualization (that allows to run virtualized servers and share physical resources between them) and cloud computing (that allows running a flexible number of parallel instances of a specific algorithm). These technologies combined allow the access to a high computing power with a low cost.
- Edge computing: In the last years, the separation between the communications infrastructure and the computing platform has started to vanish. In edge computing the implementation of the algorithms is done over computing elements located in the access network nodes (e.g. gNBs in 5G). This combines the advantages of a local implementation (low latency) with the advantages of a centralized implementation (global visibility, energy saving, computing power, etc.)

In IoT, one important development of the last years is the emergence of integral platforms, such as OneM2M [12], Fiware [13] or OpenStack [14] that offer premade solutions including data storage, processing, computation resource management, edge computing, security, etc. These platforms offer a scalable starting point for any new intelligent connectivity solution.

5 Future Outlook

The next few years are going to see the merging of the emerging technologies, including the convergence of big data, AI, and IoT. In particular, industrial IoT will harness the power of AI for optimized manufacturing process, including predictive maintenance and root cause analysis.

5.1 Digital Twins

As "things" becoming connected and with increased capability of producing data through sensing, virtual replicas of physical entities and processes can be produced to run simulation, before actual entities are built and deployed. Such virtual replicas are referred to as 'digital twins'. In essence, a digital twin is a computer program that takes real-world data and contexts about a physical system and process and reproduces how the system or the process will react to these inputs. Digital twins have been applied to manufacturing industry to facilitate production and proactive maintenance, and can include large items such as buildings, factories, and even cities.

Digital twin is a perfect example as the merging of emerging technologies including big data, AI, and IoT. The technology has been made possible due to the massive number of IoT sensors. In particular, construction of digital twins requires inputs from massive sensors gathering all relevant features—in the form of big data—of its physical counterpart, such that its digital twin can represent the physical entity, and reactions of these data can be simulated in real time. Representing a complicated physical entity (e.g., a factory, a bridge) may rely on the underlying features of the material and structure of the physical entity, and conventional method of modelling such an entity may not be sufficient. AI can serve as an effective tool in this case to reflect the underlying features of the physical entity, offering recommendations and insights to performance validation, with or without a specific modelling. It can also effectively react to the dynamic contexts of the twin, and provide enhancement in real time, according to the contexts. In many cases, a digital twin could serve as a prototype of the physical entity, before it is physically deployed in practice.

5.2 Next-Generation IoT

The next-generation IoT (NG-IoT) technologies and applications [8] will be human-centric. A human-centric IoT environment requires tackling new technological trends and challenges. The next-generation IoT development, including human-centred approaches, is interlinked with the evolution of enabling technologies (AI, connectivity, security, virtualization) that require strengthening trustworthiness with electronic identity services, services and data portability across applications and IoT platforms. This ensures evolution to platforms with better efficiency, scalability, end-to-end security, privacy, and resilience. The virtualization of functions and rule-base policies will allow for free, fair flow and sharing of data and knowledge, while protecting the integrity and privacy of data.

Intelligent/cognitive IoT networks provide multiple functionalities, including physical connectivity that supports transfer of information and adaptive features that adapt to user needs. These networks can efficiently exploit network-generated data and functionality in real time and can be dynamically instantiated close to where data are generated and needed. The dynamically instantiated functions are based on (artificial) intelligent algorithms that enable the network to adapt and evolve to meet changing requirements and scenarios and to provide context and content suitable services to users. The AI embedded in the network allows the functions of IoT platforms to be embedded within the network infrastructure.

Advanced technologies are required for the NG-IoT to provide energy-efficient, intelligent, scalable, and high-connectivity performance, with intelligent and dynamically adaptive infrastructure to provide high quality experience that can be developed by humans and things. In this context, the connectivity networks provide energy efficiency and high performance as well as the edge-network intelligence infrastructure using AI, ML, Deep Learning, Neural Networks, and other techniques of decentralized and automated network management, adaptive analytics, and shared context and knowledge.

The development of AI and IoT combined in NG-IoT enables new ways of interacting with connected objects through voice or gesture, while augmented reality (AR) and virtual reality (VR) are powered by the data generated by IoT. Furthermore, sensors and actuator technologies together with AI and connectivity will push the development of tactile IoT based on convergence of these technologies, where the boundaries between virtual and physical worlds blur.

Bibliography

1. Ericsson (2019) Ericsson mobility report
2. GSMA (2019) Intelligent Connectivity, how the combination of 5G, AI, Big Data and IoT is set to change everything
3. Russom P (2011) TDWI best practices report, fourth quarter: big data analytics. TDWI

4. Dean J, Ghemawat S (2004) MapReduce: simplified data processing on large clusters. Google Inc
5. Marz N, Warren J (2015) Big Data: principles and best practices of scalable realtime data systems. Manning Publications Co
6. Bundesministerium für Bildung und Forschung, «Industrie 4.0. Innovationen für die Produktion von morgen,» 2017
7. Zhu H et al (2019) Smart healthcare in the era of Internet-of-Things. IEEE Consumer Electronics Magazine 8(5):26
8. Everis (2014) everis Connected Car Report
9. Di Martino B, Esposito A, Nacchia S, Maisto SA (2018) Towards an integrated internet of things: current approaches and challenges. In: Internet of everything. Springer, Singapore, pp. 13–33
10. Fielding RT (2000) Architectural styles and the design of network-based software architectures. University of California Irvine, Irvine
11. GraphQL, «GraphQL specification,» June 2018. [En línea]. Available: https://spec.graphql.org/June2018/
12. «OneM2M,» [En línea]. Available: www.onem2m.org
13. «Fiware,» [En línea]. Available: www.fiware.org
14. «OpenStack,» [En línea]. Available: www.openstack.org
15. Vermesan O, Bacquet J (2018) The next generation Internet of Things – distributed intelligence at the edge and human machine-to-machine cooperation. River Publishing
16. Population Reference Bureu of the United States, fact sheet: aging in the United States, July 15 2019, https://www.prb.org/aging-unitedstates-fact-sheet/