1. Machine Learning, 5G and Beyond: Interplay and Synergies

Sergio Barbarossa[1], Andrea Zanella[2]

[1]Sapienza University of Rome - Italy

[2]University of Padua - Italy

Abstract: *Machine Learning (ML) is already playing an important role in wireless networks as an effective way to handle the complexity of the network, overcoming some of the limitations of traditional model-based approaches. The importance of ML tools in wireless network design and management is further increasing in 5G networks, where a single communication platform is designed to enable a plethora of new services, ranging from autonomous driving, virtual reality or Industry 4.0, with a wide variety of requirements and constraints. From the 5G perspective, besides looking at ML as a fundamental tool to improve the network design, the network itself becomes a fundamental driver to enable a truly pervasive deployment of intelligent mechanisms, especially useful in context-aware and delay-critical applications. The role of ML and, more generally, artificial intelligence (AI) is expected to further increase and become an essential part of next generation (6G) networks, which are supposed to be AI-native. The goal of this chapter is to highlight the synergies resulting from merging wireless network technologies with advanced ML tools. We start from the use of ML tools to enable proactive resource allocation in 5G networks, from physical up to network layers, and then we will provide a vision about 6G networks, where a native-AI network design, incorporating for example semantic and goal-oriented communication ideas, can represent a significant leap forward.*

1.1 Introduction

Artificial Intelligence (AI) is permeating our life as a powerful framework able to extract meaning from the deluge of data that humans and machines are continuously generating and to handle the complexity associated to new societal living models, including smart cities, intelligent transportation, smart grids, Industry 4.0, etc. AI is a very big framework encompassing a variety of fields, including knowledge representation, first-order logic, probabilistic reasoning, semantics and machine learning. Among these tools, Machine Learning (ML) is already having a big impact in enabling machines to learn from data, with or without the supervision of humans. An astonishing improvement of ML capabilities has come in the last decade from the inception of deep neural network (DNN) algorithms. A breakthrough has arrived when Deep Learning (DL) has shown to possess image recognition capabilities better than human capabilities [1]. This was a remarkable achievement, going beyond the expectations raised from the victory of IBM Deep Blue victory over the world chess champion Garry Kasparov in 1997. In fact, chess game is a

2. Model-aided Deep Learning for Radio Resource Allocation in Future Wireless Networks

Alessio Zappone[1], Marco Di Renzo[2]

[1]University of Cassino and Southern Lazio - Italy

[2]Université Paris-Saclay, CNRS, CentraleSupélec - France

Abstract: *This chapter reviews techniques for radio resource allocation in future wireless networks, based on an interplay of deep learning and traditional optimization-oriented techniques. In particular, it will be shown how to embed expert knowledge coming from mathematical models of wireless networks, into artificial neural networks. This will lead to two model-aided deep learning frameworks for radio resource allocation, namely deep transfer learning and deep unfolding. Specific case-studies will be analyzed to show the performance of the described methods and how they can significantly reduce the amount of live data that needs to be acquired/measured to train and test artificial neural networks.*

2.1 Introduction

With the roll-out of 5G wireless networks, the demand for wireless connectivity is expected to grow significantly in the timeframe between 2020 and 2030, reaching 5,016 exabytes and data-rate requirements up to 1 Tb/s [1]. At the same time, future wireless networks will have to provide new and heterogeneous services. As a result, future wireless networks will have to provide not only high data-rates, but also extreme reliability, very low latency, and very high energy efficiency. It is unlikely that this vision can be implemented by focusing only on the development of new transmission and reception technologies. Instead, a new approach to network management and operation is required, integrating the network infrastructure in the communication environment through the use of intelligent devices with data storage, transmission, and reception capabilities. This paradigm leads towards the concept of smart radio environments [2, 3], in which intelligence will be placed not only in the network core, but also distributed across all network segments.

Thus, smart radio environments will be composed of a much larger number of devices compared to present network infrastructures. While this allows a dramatic increase of the network capabilities, it also requires a large complexity to exploit the increased potentialities of the network. Increasing the number of devices provides a much larger quantity of resources to be allocated and optimized, but this also poses a complexity issue. In this context, a promising approach lies in the use of artificial intelligence (AI), and specifically deep learning by artificial neural networks (ANNs) [4], as a tool for wireless network design. ANNs are capable of learning from previous experience, and inferring in real-time the resource allocation policy to employ, with a much lower computational complexity than traditional model-based designs. On the other hand, in order to operate

3. How cooperation can boost learning: an overview of Federated Learning

Fabio Busacca[1,3], Laura Galluccio[2,3], Sergio Palazzo[2,3], Francesco Restuccia[4]

[1]University of Palermo - Italy

[2]University of Catania - Italy

[3]CNIT Research Unit Catania - Italy

[4]Northeastern University Boston - MA, USA

Abstract: *Recent emerging technologies such as the Internet of Things and 5G communications are changing the nature of mobile networks. The increased numbers of users, and the new variety of available communication technologies as well, bring a new level of complexity in mobile networks. Traditional optimization algorithms cannot deal with such complex environments, and need to be replaced by more flexible machine learning algorithms. Indeed, the integration of machine learning with the current communication systems is the key to the successful deployment of new generation mobile networks. In such a view, Federated Learning represents a promising approach, as it brings network intelligence from edge and cloud servers to the devices. This chapter introduces the basic concepts of Federated Learning, as well as its main advantages and drawbacks. The chapter also describes some possible use cases for Federated Learning in mobile networks. Finally, the chapter illustrates some useful techniques for the implementation of Federated Learning in such an extremely constrained scenario as the Internet of Things.*

3.1 Introduction

The last decades have witnessed an unprecedented evolution of wireless and mobile networks. The number of interconnected nodes has grown considerably, and their nature has changed as well. Indeed, more and more people can access mobile networks and terminals all over the world. As of 2018, 3.54 billions of individuals had used internet mobile services [1]. Moreover, in the past, mobile nodes were either cellular phones or computers. The recent growth of the Internet of Things (IoT) paradigm has completely changed this perspective. In such a vision, even the simplest object can be regarded as a network node connected to the Internet, able to transmit and receive data and even interact with the surrounding environment. A single IoT network can consist of hundreds or thousands of interconnected nodes, spanning from simple nodes barely capable of gathering and transmitting sensor data, to more powerful devices able to perform complex operations. The main consequence is the tremendous growth in the number of users in mobile networks, and, consequently, in the volume of generated data. In other words,

4. Who Can Learn Best: Distributed AI at the Network Edge

Carla Fabiana Chiasserini[1,2,3], Francesco Malandrino[2,3]

[1] Politecnico di Torino - Italy
[2] CNR-IEIIT - Italy
[2] CNIT - Italy

Abstract: *Artificial intelligence (AI) and machine learning (ML) will be an essential component of 6G, as they are becoming the cornerstone of both user applications and fully-automated management of next-generation networks. To cope with the ubiquitous demand for AI/ML, it is evident that machine learning tasks need to be distributed, so that the computing, memory, and energy resources of the plethora of devices in our daily life can be effectively exploited. At the same time, not all devices can contribute to the distributed learning effort equally; indeed, identifying the devices that can learn best is one of the main challenges of AI/ML in 6G. In this chapter, we focus on a popular approach to distributed learning, namely, Federated Learning (FL), showing how it can be conveniently implemented in an edge networking scenario, and discuss how identifying the best learners is key to ensure a swift, yet accurate, training of ML models.*

4.1 Introduction

Distributed machine learning is an emerging paradigm that aims at exploiting different *learning nodes*, each contributing to the training of machine learning (ML) models using its data, computing and storage resources. Federated learning (FL) stands as one the most popular approaches to distributed learning, where learning nodes cooperate in training the same ML model with the help of a centralized server, the so-called *coordinator*. FL has been proposed in [1], with the goal of involving mobile devices such as smartphones in the training process of ML models, without the need for such devices to share their own data; thus, beside allowing for the exploitation of distributed computing and storage resources, FL also provides data privacy.

In a nutshell, FL enables the training of an ML model through subsequent training phases, called epochs. Each epoch includes the steps depicted in Fig. 4.1 and reported below:

1. each learning node participating in the process trains the model locally, using its own data;
2. the learning nodes send the parameters of the locally trained model to the coordinator;
3. the coordinator combines the parameters received from the learning nodes;

5. Decentralized Federated Learning for Extended Sensing in 6G Connected and Automated Vehicles

Luca Barbieri[1], Stefano Savazzi[2], Monica Nicoli[1]

[1]Politecnico di Milano - Italy

[2]Consiglio Nazionale delle Ricerche - Italy

Abstract: *Federated Learning (FL) techniques have been emerging in the last few years to provide enhanced learning functionalities and facilitate the decision-making process in connected automotive tasks. Yet, much of the research focuses on centralized FL architectures, which have been shown to be limited by latency and scalability. Decentralized FL tools, on the other hand, are based on a distributed architecture: rather than relying on a central orchestrator, vehicles are able to autonomously share the parameters of the Machine Leaning (ML) model via Vehicle-to-Everything (V2X) connections. In this chapter, we present an overview of FL potentials in 6G vehicular networks for automated driving and we propose a modular FL approach for road actor classification in a cooperative sensing use case. Lidar point clouds are used as input to a PointNet compliant architecture. At training time, a subset of the model parameters is mutually exchanged among interconnected vehicles, namely selected ML model layers, to optimize communication efficiency, convergence and accuracy. Real data extracted from a publicly available dataset are used to validate the proposed method. Data partitioning policies target practical scenarios with highly unbalanced local dataset across vehicles. Experimental results indicate the FL complies with the extended sensors use case for high SAE levels, and outperforms ego approaches with minimal bandwidth usage.*

5.1 Introduction

Distributed Machine Learning (DML) is a key enabling technology for Connected Automated Vehicles (CAV) where networked vehicles, with increased level of intelligence and autonomy, cooperate to improve safety, efficiency and driving comfort. CAD relies on big-data-driven training of large-size Machine Learning (ML) models for several automated functions [2], as well as high-rate ultra-reliable low-latency Vehicle to Everything (V2X) interactions with road infrastructure (V2I) and other vehicles (V2V) for cooperative sensing/maneuvering tasks. In such context, the integration of DML techniques [3] with vehicles acting as distributed learners is expected to enable faster, more accurate and flexible training as well as novel decision-making opportunities.

In conventional DML paradigms, the training procedure is supervised by a central orchestrator, e.g., a road side unit (RSU) or a mobile edge cloud (MEC). This is in charge

6. Enabling Mobile Edge Intelligence Through Deep Learning Techniques

Arcangela Rago[1,2], Giuseppe Piro[1,2], Gennaro Boggia[1,2], Paolo Dini[3]

[1]Department of Electrical and Information Engineering (DEI), Politecnico di Bari - Italy

[2]Consorzio Nazionale Interuniversitario per le Telecomunicazioni (CNIT) - Italy

[3]Centre Tecnologic de Telecomunicacions de Catalunya (CTTC/CERCA) - Spain

Abstract: *Deep learning techniques are emerging as powerful instruments for the design of advanced services at the edge of current and future generations of mobile networks. Thanks to their native ability to extract valuable information from data coming from heterogeneous sources, while automatically unveiling hidden correlations, they promise to really meet the mobile edge intelligence paradigm. This book chapter describes recent solutions conceived for classifying and predicting mobile radio patterns and to anticipatory allocate communication and computational resources at the network edge. In the first case, a Multi-Task Learning model, running directly at the edge of the network, is conceived to perform data mining from the control channel of an operative mobile network. Two configurations of neural networks, based on Undercomplete Autoencoder or Sequence to Sequence Autoencoder, are exploited to obtain common feature representations of traffic profiles. Then, softmax and fully-connected layers are used to anticipate information on the type of traffic to be served and the resource allocation pattern requested by each service during its execution, respectively. In the second case, a Convolutional Long Short-Term Memory architecture is exploited to predict both users' mobility and communication and computational resources they request over different look-ahead horizons. Obtained information is used for optimally allocating and fairly distributing Multi-access Edge Computing resources through Dynamic Programming.*

6.1 Introduction

ML is the branch of Artificial Intelligence (AI) that investigates algorithms able to learn and improve their experience and performance over time directly from data examples, without being explicitly programmed. With these algorithms, a system can scrutinize data and deduce knowledge: hidden patterns in the training data are identified and used to analyze unknown information and drive the execution of a given task (typically classification, prediction, or clustering). To improve these capabilities, deep learning further enables the mining of valuable information of data coming from heterogeneous sources and unveils hidden correlations automatically, which would have been too complex to extract by human experts. Therefore, this is a powerful instrument in the mobile networking domain, where the growing diversity and complexity of the mobile network

7. Machine-learning-aided resource allocation in 5G metro networks

Ligia Maria Moreira Zorello[1], Sebastian Troia[1], Guido Maier[1]

[1]Politecnico di Milano - Italy

Abstract: *The 5G is emerging to address very stringent and heterogeneous requirements of numerous new services and applications. In particular, urban areas introduce challenging requirements in terms of throughput, latency, and reliability to enhance the performance of the new use cases related to enhanced Mobile Broadband (eMBB), ultra Reliable Low Latency Communication (uRLLC) and massive Machine Type Communication (mMTC). In addition, mobile data traffic exhibits repetitive patterns with spatiotemporal variations thanks to the highly predictable daily movements of large populations of citizens in urban areas. Hence, the upcoming applications require a dynamic and flexible optical metro network capable of integrating network and processing resources to carry the network demands ensuring service performance and network efficiency. Such flexibility can be achieved thanks to software-based technologies such as Network Function Virtualization (NFV) and Software-Defined Networks (SDN); however, the network reconfiguration is not immediate because of the time to compute and assign the required resources. Machine learning techniques can therefore be used to help the allocation of resources by predicting the traffic expected ahead of time.*

This chapter analyzes two use cases that exploit machine-learning-aided optimization to allocate resources. The first use case is an optical metro network used as the backbone for the mobile service, we present a dynamic resource allocation exploiting the programmability leveraged by SDN. We use the predicted traffic variation to solve offline mixed-integer linear programming instances of an optical routing and wavelength assignment optimization problem. The results demonstrate the effectiveness of the method, such that the prediction-based optical routing reconfiguration optimization matches almost perfectly the behavior with an oracle-like traffic prediction. The second use case is the allocation of Virtual Network Functions (VNFs) that implement the Radio Access Network (RAN) baseband functions over a metro network. We propose a mixed-integer linear programming optimization that uses the hourly predicted traffic to place the VNFs with the goal of minimizing the network operators' costs. The results show that the proposed machine-learning-based optimization is able to efficiently compute the resource assignment in advance without significant losses for the operators in terms of costs and performance degradation when compared to the optimal solution.

7.1 Introduction

Telecommunication networks are continuously evolving, pushed by an increasingly connected society. Millions of people rely every day on different services provided by such

8. Machine Learning Techniques for Context Extraction and User Profiling in 5G Mobile Systems

Francesca Meneghello, Giovanni Perin, Michele Rossi

Department of Information Engineering, University of Padova - Italy

Abstract: *In this chapter, we discuss the use of machine learning technology to provide context- and user-aware services in 5G, and beyond 5G, mobile networks. First, we overview the advantages of integrating side knowledge into the service management plane, from both a network and a user perspective. In fact, information about the network load, the type and distribution of users and their mobility can be utilized to control the quality of service/experience, and to intelligently allocate communication and computing resources. Such benefits are here discussed through three practical use cases, by detailing the scenarios, the data captured from the network entities, and the machine learning techniques that are exploited for context extraction. In the first use case, passive control traffic from the end users of a mobile network is analyzed to obtain information about the user's identity, and the type of applications running on their smartphone, without having to decrypt the packets exchanged with the serving base stations. In the second example, a passive monitoring device is exploited to infer the type of data, applications, user behavior and distribution in a city-wide base station deployment. Finally, mobility-aware control for edge computing resources is considered, tracking user mobility at the base stations through the analysis of beam steering patterns and the power received from the mobile users.*

8.1 Introduction

The *context* is of paramount importance in our lives. As humans, we adapt our behavior to the social context and approach problems in different ways depending on the surrounding environment and on the people around us. A similar concept is being intensively investigated as a means to manage communication networks. The increasing demand for new services characterized by stringent constraints in terms of quality of experience and delay, requires an intelligent allocation of network resources and a reasoned policy for the control of services. This is enabled by the integration and exploitation of context information in the related decision processes. In a mobile network scenario, *context* refers to the data collected from both the physical environment and the communication network, which provides information about the status of the network entities. For instance, context includes terminals' locations, transmission and mobility patterns, network energy consumption and neighboring cells' load, among others. These pieces of information enable the implementation of *context-aware* management strategies at different functional and control blocks. For example, based on the context, network resources

9. Smart Data Gathering for Network Optimization

Francesco Pase, Federico Mason, Paolo Testolina, Mattia Lecci, Andrea Zanella, Michele Zorzi

Department of Information Engineering (DEI), University of Padova, Padova, Italy

Abstract

This chapter highlights the central role that data will play in future cellular networks, considering practical issues regarding their collection and exploitation.

On the one hand, gathering and learning from network data before deployment opens to innovative optimization strategies. As an example, we report the complex problem of designing antenna arrays for Millimeter Wave (mmWave) networks, which can be greatly simplified using Machine Learning (ML) algorithms, thus identifying antenna configurations that are optimal at the network level.

On the other hand, novel paradigms such as Network Slicing (NS) enable the network to offer a number of services to the users in real time. Within this context, we provide a viewpoint on how Deep Reinforcement Learning (DRL) can be leveraged to orchestrate NS services in a timely manner. Moreover, the concept of Machine Learning (ML) as a service, i.e., running ML algorithms in a distributed manner over the network, is gaining a lot of attention, thanks to the performance that data-driven models can obtain. Federated Learning (FL) has emerged as a promising solution to train centralized models with distributed data guaranteeing the secure transmission of sensitive information and reducing the communication burden. Consequently, the interplay between FL and wireless networks will be discussed at the end of the chapter.

9.1 Introduction

Next generation cellular networks are expected to support novel services, with higher bitrates, lower delays, higher reliability, and guaranteed Quality of Service (QoS).

Higher bitrates can be achieved with Millimeter Wave (mmWave) communications that exploit wide-bandwidth wireless channels at high frequencies, typically between 10 and 100 GHz, which allow for exceptionally high communication rates and efficient spatial multiplexing. Furthermore, the short wavelength enables the implementation of extremely small antennas and, thus, to pack large, high-gain antenna arrays in a relatively small space, boosting the system performance. However, the design of such antenna arrays is very complex. While general rules of thumb derived from the principles of antenna theory exist, their optimization has to take into account many correlated parameters, often making the problem non-trivial and possibly even non-convex.

While higher transmission rates are surely needed to enable future services, they may not be sufficient if not complemented by suitable resource management algorithms, able to properly differentiate the service based on the applications' needs. *Service differentiation*

10. Wireless Edge Machine Learning in 5G/6G Networks

Paolo Di Lorenzo[1], Mattia Merluzzi[2], and Sergio Barbarossa[1]

[1]DIET department, Sapienza University of Rome, via Eudossiana 18, 00184, Rome, Italy

[2]CEA-Leti, Université Grenoble Alpes, F-38000 Grenoble, France

Abstract: *With the advent of beyond 5G and 6G systems, wireless communication networks will evolve from a pure communication perspective to service enablers in different vertical sectors, incorporating machine learning (ML) mechanisms to build an effective complex system able to learn and dynamically adapt to the evolving network landscape. This book chapter describes some recent solutions for wireless edge machine learning, i.e., a novel class of distributed and reliable ML services that can be accessed by end-users via wireless communications. Differently from cloud-based ML, the edge machine learning process requires not only high learning accuracy and reliability, but also very low latency for autonomous decision making, while coping with communication and on-device resource constraints. In this context, considering both stand-alone and cooperative/federated learning scenarios, the goal of this work is to introduce a dynamic resource allocation framework to enable wireless edge machine learning, jointly optimizing radio parameters (e.g., set of transmitting devices, transmit powers, rates, etc.) and computation resources (e.g., CPU cycles at devices and at edge server) in order to strike the best trade-off between energy, latency, and performance of the inference and/or training task. The framework hinges on stochastic Lyapunov optimization, which enables adaptation to time-varying network conditions, without the need for a priori knowledge of the statistics of random context parameters (e.g., radio channels, data arrivals, etc.). Numerical results on both synthetic and real data assess the performance of the proposed resource allocation framework enabling ML at the wireless network edge.*

10.1 Introduction

The future of wireless networks is not only to connect humans, but to enable a new class of services that embed humans and machines into the same ecosystem, with a holistic view of communication, computation, caching and control [1]. This integration will be driven by two main pillars, namely, a performance boost over the radio interface, and a pervasive deployment of cloud resources at the edge of the network. The aim is to enable Machine Learning and Artificial Intelligence (AI) close to the end users, in order to experience low latency, high energy efficiency and, in general, sustainable deployment and operations of B5G/6G networks and services. In this context, ML/AI will work as enablers of flexible network optimization and control (AI/ML for networks), but will also benefit from the availability of distributed computing and storage resources, to be enabled more efficiently