
Machine Learning: The Panacea for 5G Complexities

N. Hari Kumar¹ and Sandhya Baskaran²

¹Senior Researcher, Ericsson Research, India

²Researcher, Ericsson Research, India

E-mail: n.hari.kumar@ericsson.com; sandhya.baskaran@ericsson.com

Received 01 March 2019;

Accepted 20 May 2019

Abstract

It's not a myth that transition in next generation technology brings with it a set of exciting applications as well as challenges to the telecom ecosystem and in-turn paves way for new revenue streams. 5G enables ultra-high data rates, exceptional low latencies which enables the telecom operator for the facilitation of interesting parallels like IoT and Next-Gen Industrial enhancements like autonomous vehicles, connected mines, connected agriculture and mission critical communications by enhancing infrastructure, software and hardware components of the 5G system. As imminent new features of 5G like Multiple Input Multiple Output (MIMO), network slices, virtual network functions, indoor localization, Machine to Machine (M2M) capabilities are highly appreciated, it also opens new set of challenges like real time dynamic configurations, low latency handovers. These challenges can be addressed with the application of AI technologies to components at the crux of 5G system. Here in this paper, we discuss some of the major challenges such as data burst, improving performance, fault tolerance and traffic management with new components appended to the 5G system, required upgrades to existing technology and how Machine Learning (ML), Artificial Intelligence (AI), becomes the self-evident answer to these stumbling blocks.

Journal of ICT, Vol. 7.2, 157–170. River Publishers

doi: 10.13052/jicts2245-800X.726

This is an Open Access publication. © 2019 the Author(s). All rights reserved.

Keywords: Artificial Intelligence, Machine Learning, ML, Deep Learning, Automation, 5G, Open Source, Virtualization, Management, Automation, 5G, 4G, multi-access, access independent, Datacentre, Information Communication Technologies (ICT).

List of Abbreviations

MIMO	Multiple Input Multiple Output
M2M	Machine to Machine
AI	Artificial Intelligence
OAM	Operations, administration and management
RAN	Radio Access Network
SLA	Service-Level Agreement
KPI	Key Performance Indicators
QoE	Quality of Experience
RTT	Round-trip time
NWDAF	Network Data Analytics Function
QoS	Quality of Service
AR	Augmented Reality
VR	Virtual Reality
AMF	Access and Mobility Management function
SMF	Session Management function
PCF	Policy Control Function
UE	User Equipment

1 Introduction

Data being the prime focus for the next gen-experiences and with 5G spawning colossal amounts of data we need a structure that is suited to derive meaningful insights while eliminating uneventful knowledge. Machine Learning is one such prime tool that aids in adaptive learning and intelligent decision making. At its core, Machine Learning integrates various models to progressively learn tasks by observing and training on the arrangement of data and its behaviour.

The employment of ML techniques in 5G is a compelling decision to make as the essential characteristics of the system requires inherent dynamism and self-governance in its mechanics. In this paper, we discuss about some of the elementary and crucial applications of Machine Learning into the nucleus of 5G and explore motivation and methodology of the implementations.

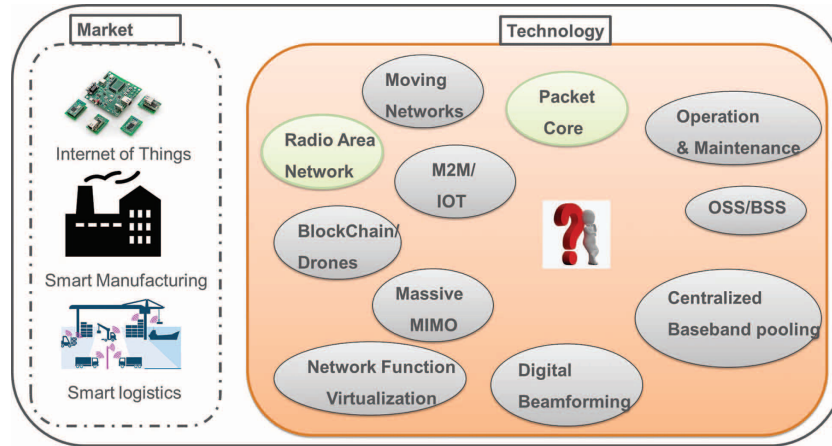


Figure 1 5G technologies and market opportunities.

Although Machine Learning as modules can be accommodated on top of the existing infrastructure and components, and can sometimes be deemed optional, the implications of such a system on a long haul can prove to be extremely inefficient [1]. Thus, a powerful system can be constructed with a ML framework at its core catering for all the 5G complexities by being a core entity of standardization.

We will discuss the existing challenges in all the sub systems and a rudimentary notion on Machine Learning as a viable solution in the rest of the paper. We will touch base on three main entities of telecom: Radio Access Network (RAN), Packet Core and Operations, administration and management (OAM) in the first half and the next half will contain the Machine Learning based solutions to the existing problems in those entities.

2 Telecom Entities

2.1 Communication Services

2.1.1 Network slice as a service

Network slice is a dynamic self-contained and independent entity that has evolved preliminarily out of network function virtualization with autonomic characteristics. Each slice is allocated a group of physical and virtual resources which is self-managed by the slice. Management and orchestration in a multi-domain system are a challenge as each domain needs different constraints including the technology to be utilized and levels of security to be established.

Decision rules could be negotiated among network elements. Slice-aware VM (Virtual Machine) Management and Slice-aware VM Migration based on traffic flow schedules and time-series Machine Learning based mechanisms [2] significantly reduce the number of hops between VMs across slices and balances the network load.

2.1.2 M2M services

With the evolution of 5G, it opened a new market with billions of devices and sensors unrolling to be prime players, each with different performance criteria – connectivity, low latency, bandwidth. M2M devices communicate in varying degrees of data transactions adapting to the dynamic requirements of each applications. M2M devices also need to share the spectrum with human owned user equipment's and thus must be highly spectrum efficient.

3 Packet Core

3.1 Radio Congestion Impact in Slice Key Performance Indicators (KPI)

With 5G, a significant increase of data rates to and from devices is expected; traffic that often will have tight requirements on KPI, associated to a Service-Level Agreement (SLA) between operator and application provider. The 4G packet core already today resorts to various techniques to manage traffic, to optimize network utilization. With the 5G system and the increase data rates and SLA related requirements on connectivity service KPI's, ongoing 3GPP standardization (eNA) implies 5G Core (5GC) to be part of a system wide automated framework to manage the traffic associated with a Slice. For instance, 5G Core may, on delegation by OAM, curb and isolate data hungry applications. All these techniques work based on a predefined threshold or parameters for modifying the flow of data packets within the core. Embedded intelligent applications with data congestion as a perspective, can help to intelligently sample traffic and make Quality of Experience (QoE) estimates and in a local closed loop, dynamically modify the applications' data pipeline and handling techniques to reach target KPI for the service (slice).

3.2 Packet Routing when Changing UPF

In 4G Evolved Packet Core (EPC), the Packet GW (PGw) is the gateway to external data networks and the anchor point for mobility. The PGw is typically not changed if the device is registered in the mobile network even if it moves.

In 5GC, the UPF (User Plane Function) – which is a part of what is PGw in 4G EPC – can be changed and there may even be several UPFs, for instance an UPF geographically close to the device for traffic to a specific customer applications server close to the device (part of Edge computing), and another UPF for traffic to the Internet. This increased flexibility is an opportunity for optimizing the traffic routing, by selecting optimal set of UPFs from a transport and Round-trip time (RTT) perspective for instance, a possibility that is naturally solved using automation and Machine Learning. Related to this, work is ongoing with the 5GC Network Functions, e.g. the Network Data Analytics Function (NWDAF), to learn to predict a device’s movement and use this to proactively change UPF including alerting the application about the change, thus enabling to make the mobility ‘seamless’, decreasing mobility related latency variations.

4 OAM

4.1 Network Operations Center

NOC centers operate across each geographic region for performing operational and maintenance tasks after a network rollout. With new alarm structures and notifications coming up with 5G and introduction of network slice management, we could expect drastic changes in the organization and structure of NOC centers.

4.2 Edge Computing

Quality of experience is a pivotal aspect when it comes to real time services. Edge computing becomes a potential dimension in 5G realms, harnessing the capabilities of smart devices while efficiently utilizing the aspects of distributed computing. Management of Network Slices becomes a vital part in E2E QoE for edge computing.

5 Machine Learning in 5G Telco Entities

5.1 Network Slice as a Service

5.1.1 Autonomic triggering of slice elasticity

Machine Learning capabilities can be used to predict the traffic and the bandwidth requirements of devices which makes network slice allocation autonomous and dynamic. The slice selection function within the Network

Slice Management Function allocates or deallocates resources to service request characteristics. The sharing of radio resources across slices is based on traffic load. Reactive evaluation of resources to be allocated can prove to be inefficient in terms of performance as the traffic patterns are growingly dynamic. An ideal solution here would be to utilize the Radio scheduling algorithms [3] which could incorporate a slice aware ML information model to allocate radio resources based on the bandwidth predictions.

5.1.2 Mobility management

Predicting the next possible Network slice that a UE (User Equipment) would hand-over to using Machine Learning techniques like time-series correlation, enables the network to prepare the network slice prior to the occurrence of demand with the needed capabilities to perform a smooth transition. The tenant needs to be provided with capabilities to support mobility of the UE across slices. The slice selection information is mapped from one operator to the other to support roaming users. Rapid mobility hand over is required for crucial real time services. The Network Slice Management Function must have extreme high-speed capabilities for switch overs and information mapping. Prediction of next Network Slice would ensure that the time taken for mapping the slice information is reduced, thus providing better quality of service.

5.1.3 Network slice discovery

A 5G device may be enabled with discovery mechanisms to identify the appropriate end-to-end network slice in the future. Such a solution has been termed as Device Triggered Network Control mechanism. Management and orchestration unit must be embedded with capabilities for recommending an appropriate network slice for every user and a Machine Learning based recommender system that discovers user's requirements based on user behaviour, mobility, utilizations in the history [4]. A Machine Learning standard with a user level model is required to ensure a Quality of Service (QoS) or a fair usage policy.

5.1.4 QoS capabilities

Each industrial application will require a different configuration for the network slices, and these slices are often preprogrammed for allocation of resources. For example, a massive IoT static industrial application might require 5G core with limited handover and large number of connections. On the other hand, a broadband mobile Augmented Reality/Virtual Reality (AR/VR) application might require high capacity, complete mobility and low latency

capabilities all in one. And when there is bombardment of autonomous cars in a specified space, then we need to allocate the network resources dynamically. This dynamic provisioning of the resources while incorporating application requirements needs to be done in real time along with coordination of resources across slices. Reinforcement Learning kind of techniques [5] can be applied to measure the performance of each slice and learn the necessary configuration as the deployment progresses.

6 M2M Services

6.1 M2M Charging and Billing Specifications

An operator should be enabled to configure and manage billing services pertaining to different M2M services. The complexity encompassing the IoT systems with Adhoc device spawns coming into picture, which leads to more complex revenue chain and fund clearance requirements. Traditionally, only human users needed to be charged depending on a fixed set of services offered by the operators. With the M2M systems in the scenario, there must be a support for very high number of devices and the charging model could be designed to be based on per service consumption or bandwidth consumptions. There is also the scenario of human to Machine or vice versa interactions which would require complex scenarios for billing. Right partnering and settlements between different operators with a standard contract would be difficult to establish with the growing variety of services and the prospects each operator has to offer. Pay-per-use (for devices that intermittently utilize the bandwidth), flat rate based on criteria (for devices that consistently use minimal bandwidth), consumer specific models (where the pricing is based on the consumer and not the device), and enterprise specific models are some of the proposed models for billing. In each of these plans, the operators need to ensure customer satisfaction as well as profitability. For example, in the consumer specific models, the data is the prime commodity and the profitability is calculated based on the predicted amount of usage. Machine Learning algorithms can help model and identify local user behaviour through which an operator would recommend the right plan. Traditional static rules engines would fail to serve this purpose as there would be a wide range of discounting rules at various levels of abstraction. The pricing model generated using Machine Learning algorithms that can consider consumer satisfaction and profitability both at a consumer and global level as primary features is required to ensure a balanced revenue chain.

6.2 Bandwidth Adaptability

With an exponential increase in number of monitoring systems across the landscape, it is expected to accelerate the bandwidth utilization by IoT devices. While sensor values are only small messages, there is a huge amount of data being sent across in short intervals leading to congestion. Prioritization of data cannot be performed in a general abstraction as the requirements for each business application varies, e.g. with critical applications all the data needs to be dispatched in real time to the network as it is time crucial, while some recurrent duplicate information can be concealed from being dispatched. Device communication mechanisms should possess this intelligence at the edge or in the network. Machine Learning models assist in recognizing the demand from the consumer without revealing explicit details owing to growing privacy concerns [6]. Such optimization becomes extremely mandatory with magnifying usage of devices.

6.3 Autonomous Device Discovery

Device discovery mechanisms available today perform paging requests to initiate services for an idle user equipment. Paging request must be broadcast to all the base stations in the location area and bursting nature of packet traffic means more paging requests per mobile. Such device discovery mechanisms become inefficient with introduction of billions of devices into the system and an overcrowded cell system. Knowing the UE location with certain degree of accuracy is preferable over traditional circuit switched traffic. Novel discovery mechanisms such as device to device discovery have been proposed to handle IoT devices, especially with the introduction of capillary networks. Owing to the sprawling adoption of IoT devices and Adhoc device spawns it becomes tedious to identify suspicious unauthorized devices that could access and congest the network. Device behaviour modelling and anomaly detection Machine Learning algorithms with features extracted from previous incidents could help predict if a new device raises a red flag for spurious behaviour.

7 Data Congestion in Packet Core

7.1 Data Flow Pipeline

The data flow pipelines for read and write of data within the core network goes through the same channel for various application and various users. When the channel becomes unavailable for business priority users, regular

users are stacked from getting data during a brief congestion period. At present, congestion levels are determined based on a static threshold limit set by the operator or the business layer. Although this is effective in handling congested networks for business priority users, the isolation of regular users is applied even if the network can accommodate more traffic. Such a situation could lead to serious under-utilization of network resources and result in QOS lapses. Regression based functions embedded into the data flow pipeline could perform switch overs between stacking and releasing data flow based on the current congestion rate and based on predicted time interval.

7.2 Congestion Based Device Caching

Currently with a data request to the network, the network reacts to the request based on the available bandwidth to download the data. A user could be stacked if the network congestion happens in the mid-way. With an ongoing request, the end user device could be intimated on a possible predicted congestion which could let the user switch to a higher priority real time application for utilizing the current bandwidth.

8 Network Operations Center

8.1 Fault and Performance Management Rules

Rules engine and management is a key aspect of NOC centers as their rules decipher the next course of action given an incident. Any performance degradation or fault is met across with statically defined specific rules pre-programmed by domain experts. When a performance metric crosses a predefined threshold an alarm or a notification is raised and after expert supervision is converted to a ticket. Currently, a typical NOC is handling over 20 million notifications and millions of service desk emails every year. With billions of devices and cross hybrid structures invading the network, it becomes practically impossible to perform a manual monitoring and recording observations of the entire network. Incident Management with hand-written rules is extremely challenging both from technology and deployment point of view. Fault Management systems need to accommodate for dynamic network changes and unprecedented number of faults with the 5G systems. Automated processing of incidents in a data driven domain agnostic manner without the need for expert rules would help significantly enhance automation in NOCs. Machine Learning based components enable us to automate and accommodate the varying degree of changes within a network in which the model would

be capable of realizing a fault and converting into a trouble ticket (action on alarm) based on human reactions to previous such faults. A simple fault might lead to a major cascading effect and impact the entire network from which determining the root cause becomes a tedious process. Capabilities for discovering co-occurring patterns using pattern mining techniques enable us to identify the root cause of such cascading alarms. An autonomous end-end loop of problem resolution can be arrived at only by inclusivity of Machine Learning based intelligent components such as a pattern recognition ticketing system or a self-resolving remote operation.

8.2 Field Service Operations

Field service operations are bound to take up a new turn with the 5G Systems in place requiring major optimization changes. With few numbers of resources and exponentially increasing number of devices with new technology, the field services require additional support in terms of enhanced equipment and tool capabilities to identify and resolve the problems ever so quickly. Image recognition algorithms embedded within an AR/VR equipment can help speed up operational maintenance by multiple folds. A voice aided algorithm or a chat bot that could aid spot the exact area of fault could increase the sustainability and reliability of the system. Additionally, an ML based optimal work flow planner can help accommodate and engage the right skillset at the right place keeping in mind the priority of the problems at hand [7].

9 Illustrations on ML Based Optimized Network Coverage in Radio Self Organizing Networks

Ref. [1] talks about optimizing the network to increase the Signal Strength regions/minimizing the dead zones in the network area. ML based beam tilt algorithm identifies the right azimuth angle to adapt to varying channel conditions. The technique employs regression analysis and stochastic gradient descent to study the relationship between azimuth angle combinations and signal strength. This is one such example of application of Machine Learning based techniques in a basic configuration setup of antenna parameters.

10 Management Functions

A Management Data Analytics Service provider for the 5G systems needs to be equipped with capabilities for providing communication service requirements

related to the network including SLA. With these capabilities, new market verticals have opened that fall into categories such as:

- **Business to Consumer (B2C) services**
Business to consumer services often require nominal data rates and high availability.
- **Business to Business (B2B) services**
Business to business services requirements vary with the service being requested. For example, a logistics service might require low data rates and low availability while a hospital service might require high data rates and high availability.
- **Business to Household (B2H) services**
Household services majorly constitute of connected IoT devices or a capillary network which would require anomaly detection such as surveillance systems. In this case, very low data rates and high availability is required to intimate disaster with low latency.
- **Business to Business to Everything (B2B2X) service**
This is envisioned as a hybrid system with extremely fluctuating service requirements. For example, a multipurpose factory manufacturing requiring both real time and non-real time information with the same network slice.

Management Analytics function help the other functions in deriving network topology VNF configurations to provide the required service assurance. Continuous improvements and adjustments are required from both the user plane and control plane perspectives. Reinforcement Learning techniques incorporating customer feedback, KPI information, SLA information, resource requirements as reward functions serve the purpose of identifying the areas of improvements.

Additionally, each of the market verticals mentioned above require different configurations to cater for different performance requirements. From performance requirements perspective communication services are classified as eMBB (high data rates and high traffic densities), URLLC (low-latency and high-reliability services) and MIoT (large number and high density) and each category requires deriving suitable network topology (e.g. VNF chains, network configurations etc.). Management functions must have dynamic potential to balance and fine tune the system according to the various service requirements such as slice deployment or maintenance or changes in business requirements. For example, the QoS of user traffic and cell KPIs may be degraded and hence the operator shall have the capability configuring

unreserved resources in RAN, potentially to be used for a new network slice instance. Such Machine decisions are enabled using Machine Learning techniques based on data and many factors such as congestion, coverage issues, interference, shortage of radio resources. Recommendation systems with collaborative filtering could be considered as next steps in dynamically classifying a network slice for provisioning according to the various usage levels.

11 Data Collection Services

Data collection from various sources such as Access and Mobility Management function (AMF), Session Management function (SMF), Policy Control Function (PCF) and OAM global data forms the basis for a network analytics function computation. Behaviour data related to UE groups, with spatial and temporal dimensions, must be interfaced with the analytics function. Such data would have varying granularity and can have huge memory requirements for archival. Optimization of the data storage can be performed using Machine Learning techniques of pattern recognition. Pattern recognition can help identify redundancy in information and thus prioritize storage information.

Additionally, data collection systems include event subscribe/unsubscribe for utilizing services provided by the Network Functions. Dynamically discovering NF requires continuous polling to the systems, which results in overutilization of resources. Periodic discovery is highly efficient provided the right time window of polling is identified.

Identifying the right window becomes complicated with each Network Slice catering to different business structures and hence a behaviour-based polling window determination could have a better performance over the static window size. A Machine Learning based activity and intent recognition framework would help identify the customer service requirements and map them to the appropriate subscription window size.

12 Conclusion

We consider an intelligent telecom network, where various Machine Learning techniques and advanced data analytics can be integrated with telecom components to make them automated and perform efficient control and optimization. We present the various entities of the telecom domain and discuss how ML plays important roles in each of these entities, as well as various challenges and solutions for each entity like improving performance, fault tolerance and

traffic management. We present a set of optimization techniques with respect to Machine Intelligence. Finally, we discuss the benefits and challenges that the management of network functions encounter in adopting data analytics and ML in next-generation wireless networks.

Acknowledgments

We are thankful to our colleagues from Ericsson Research and Packet Core PDU whose research work provided additional value and increased understanding of ongoing trends and thus greatly influenced our paper, although they may not agree with all the interpretations/proposals provided in this paper. We are also grateful to Göran Eriksson AP for assistance with packet core who moderated this paper and, in that line, improved the manuscript significantly.

References

- [1] <https://medium.com/@daveevansap/so-whats-the-real-difference-between-ai-and-automation-3c8bbf6b8f4b>.
- [2] <https://ieeexplore.ieee.org/document/8321782>
- [3] <https://ieeexplore.ieee.org/document/483546>
- [4] <https://ieeexplore.ieee.org/abstract/document/6691621>
- [5] <https://ieeexplore.ieee.org/document/5158403>
- [6] <https://ieeexplore.ieee.org/document/7906930>
- [7] <https://www.comarch.com/telecommunications/blog/transform-field-service-management-with-machine-learning-and-artificial-intelligence/>

Biographies



N. Hari Kumar is a Senior Researcher at Ericsson Research, India, currently focusing on the areas of Big Data, IoT and Machine Intelligence. He joined Ericsson in 2008 as a Software Engineer. Hari holds 9 granted patents

and 7 more patent filings on his behalf. He holds his engineering degree in Information and Technology from Anna University, India.



Sandhya Baskaran is a Researcher at Ericsson Research, India, currently working on Machine Intelligence technologies and IoT. She joined Ericsson in 2013 as a Software Engineer and specializes in the areas of Operating systems and Virtualization. Sandhya holds a bachelor's degree in Information and Technology from Anna University, Chennai, India.