altice

labs

InnovAction

#4 | 2019

The value of an idea lies in the using of it.

Thomas Edison

InnovAction

#4 | 2019

Time for a new landscape

Digital service providers (DSP) are, once again, at the epicentre of a new relevant digital transformation of the industry! Boosted by the vast data heritage accumulated or accessed to over the years, this new wave of change is the perfect opportunity for the DSP to redefine itself in three key areas: the creation of new services, operational efficiency and user experience.

Willing to thrive and be the best, DSP are committed with the change, defining and designing new strategies and information architectures powered by a new silver bullet: Artificial Intelligence! New operational insights, produced by new decision tools, will anticipate and create new market scenarios and realities, leading to new products and services and, ultimately, to new business value.

And this is just the beginning because the possibilities associated with Artificial Intelligence and in particular Machine Learning are limitless: from process automation to recommendation systems supported on previously unknown patterns, solutions that support the DSP's activity should migrate to "cognitive-infused" models! These models will create cognitive organisations where intelligence is applied across multiple domains in an entirely new landscape, a cognitive landscape!

Year over year, aware of the most relevant industry trends and supported on their own R&D, Altice Labs publishes InnovAction, a technological magazine strategically designed to share the most relevant research and technical knowledge that will help DSP to reinforce their innovative capabilities.

I hope you enjoy reading it, as much as we enjoyed writing it.

Alcino Lavrador
General Manager of Altice Labs

Editorial note

In the profitable race for (best) digital transformation, a new additive has reborn to accelerate the speed of all players: artificial intelligence.

Responsible for enhanced decisions, powered by data, artificial intelligence is outlining a new landscape, a cognitive one where digital service providers define algorithm-driven strategies and action plans that level up their competitive advantages by creating new (digital) services and new business value.

Emerged in this cognitive landscape, Altice Labs publishes the fourth edition of InnovAction, whose articles focus on the following themes

- **Cognitive landscape: Artificial Intelligence redefining the Industry**: highlights the business impact and potential of artificial intelligence as a new toolset, especially for the digital services providers. It also presents Altice Labs' strategy to extract additional value from managed data and to create new use cases on top of its solutions by using artificial intelligence technology;

- **Cognitive ops services:** presents insights on how artificial intelligence and machine learning can help three main OSS domains (inventory, fulfilment and assurance) making the change to cognitive;

- **Predictive fault management:** describes the cognitive approach followed by Altice Labs to implement an alarm prediction use case using operational data, as well as it summarises the main operational benefits of the chosen approach;

- **Root Cause Analysis of Reduced Accessibility in Cellular Networks:** evaluates the KPI related with reduced accessibility in cellular networks to promote a more proactive network management and the prediction of an eventual future drop in network accessibility;

- **Access network failure prediction powered by cognitive techniques:** an overview: suggests some areas of research and development where artificial intelligence algorithms can correlate data generated by network elements to efficiently replace human analysis and prediction in the domain of passive optical networks.

- **5G Intelligent Communications for V2X ecosystems:** synthesises the emerging 5G

ecosystem characteristics, positioning it to be the service platform for advanced cellular vehicular use cases enhanced by artificial intelligence;

- **A Recommender System for Service Providers' Campaigns:** exposes a recommender system applied to a service provider's advertisement campaign. It also focuses on the extent to which it is possible to characterise the customers, using implicit feedback and state-of-the-art recommendation algorithms;
- **The smart home:** voice, machine learning and proactivity as innovation drivers: analyses the evolution that the smart home concept had until now, and the various efforts already put in place to fulfil this wish. It also presents what is perceived as the driving forces that can crack the persistent unfulfilled promise of the smart home and the way Altice Labs products and services can contribute to it;
- **HCI boosted by AI:** from smart interfaces to immersive cognitive environments: addresses the recent evolution of Human-Computer Interaction through the inclusion of technologies and features that envision the rise of immersive cognitive environments with seamless interaction, as well as highlights the exploratory research carried out by Altice Labs under this scope;
- **ENTERing the future through open innovation:** provides an overview of the ENTER programme's value proposition and the key challenges it faces when scouting for the best collaborations with startups and scaleups. It also shows the power that startups focused on artificial intelligence and machine learning solutions have to open new opportunities and revenue streams in the industry.

Thus, InnovAction 2019 not only highlights what Altice Labs is doing under the cognitive umbrella to optimise and leverage its customer experience but also to bring some light to its research work that will help Altice Group, its Customers and Partners to stay ahead of the competition.

Ana Patrícia Monteiro
ana-p-fonseca@alticelabs.com

Contents

8 Cognitive landscape: artificial intelligence redefining the industry

01 Cognitive landscape: artificial intelligence redefining the industry

Mário Rui Costa, Altice Labs
mcosta@alticelabs.com

Paulo Pereira, Altice Labs
paulo-s-pereira@alticelabs.com

The transition to data-driven digital businesses leveraged by a growing “data and algorithmic economy” is already a reality among many organisations across several industries, with businesses using algorithms daily to influence and support decisions and optimise operational efficiency. It is undeniable; we now live in an AI spring, a moment where the information, infrastructure and toolset virtuously combine to boost the AI infusion in digital processes across industries.

Keywords

Artificial intelligence; Cognitive; Digital transformation; Digital services provider; Data science

Overview

In past decades the industry has been evolving with the aid of computers and information systems to manage information and automate tasks, aiming to increase productivity and business value. Computers and information systems have been evolving exponentially in capacity and complexity to process ever-increasing amounts of data and to handle ever-increasing complexity in tech industries. However, business productivity is not by far increasing proportionally; in fact, when comparing with the evolution of computation power, productivity increase can be considered marginal.

The gap above mentioned can be explained by the inability of information systems to overcome two major constraints that drain productivity: a) they are unable or struggle to implement some tasks in critical business and operational processes, so they need to be delegated to humans; b) they are not able to address exceptions to what has been explicitly coded in the system. These constraints express the intrinsic limitation of the paradigm which information systems have been relying on: they are rule-based systems, limited in its capacity to handle some real-life complexity and mutability of industrial processes. In this sense, the general availability of artificial intelligent (AI) technology and high-performance computational resources bring the opportunity to create information systems able to overcome the aforementioned productivity constraints.

By learning from data and human experience, by providing fully autonomous processes (capable of continuously learn to adapt to changing environments), and by having the extraordinary ability to create insights unattainable by traditional rule-based systems, able to express new business value, AI brings the toolset to complement the existing rule-based operations with new pattern-based capabilities, creating a new operational landscape. A cognitive one.

Introducing artificial intelligence

AI as a new toolset

In past decades humans used their intelligence to develop the means required to create, at scale, intelligent artificial systems able to mimic human cognitive capabilities, i.e., able to perceive, understand and use acquired knowledge to act according to a specific scenario. Those systems are new cognitive tools that, in one hand, promise to boost the evolution of two strongly related business dimensions: performance and productivity, and, in the other hand, when coupled with process automation will allow making more, faster and efficiently.

In the current landscape, near to all industries still rely on humans to perform some industrial process activities, due to: a) functional gap in existing systems; b) the activities performed by humans are too complex to be coded into a system. While the first case can be handled through incremental evolution of systems and processes using traditional technology, the second cannot, and that's where AI systems will step in and reveal its potential in short to medium term. A paradigmatic example of such case are the customer care processes at call centres. The handling of customer complaints involves an enormous contextual diversity and human sentiments that make this activity near to impossible to be handled by a pre-coded rule-driven algorithm. AI systems using natural language processing (NLP) capabilities can mimic human interaction when relating to customers and promote the implementation of fully automated processes.

Beyond productivity, business performance is also closely related to the creation of something new and relevant, i.e., something that will allow to sell more or to decrease operational expenses (OPEX). In this dimension, AI systems can have a bigger impact in the medium to long term, since their cognitive capabilities will allow to extract new insights from the huge amounts of data produced by industrial processes and to act accordingly, in a more powerful way. A paradigmatic use case on the sell uplift side is the use of AI systems to create personalised one-2-one product recommendations considering the whole universe of characteristic and behavioural data associated with it. A paradigmatic use case on the OPEX reduction side is the use of AI systems to support predictive maintenance of industrial equipment, i.e., able to predict future anomalies considering continuous time series of operational parameters obtained from that equipment.

Technology behind AI

The creation of an AI system can be made using a very broad set of technologies and approaches. In the Industrial arena, machine learning (ML) has been the approach to create AI systems.

ML consists of using mathematical algorithms created to solve specific types of tasks (e.g., classification, clustering and regression) and to train those algorithms using datasets. Before the training stage, one must define the data that must be feed into the algorithm so it can learn (the model features) and the data the model should be able to produce (the model targets). The training phase roughly corresponds to perform N iterations to optimise the model parameters until reaching acceptable error values when generating its targets. Depending on the size of input data and the number of iterations required to converge, the time and computing power to train a model may increase considerably. And of course, the quality of the input data will, for sure, limit the quality of the ML model or even the ability to create it.

In fact, as illustrated in **Figure 1**, in most ML projects there is a substantial amount of hard work to ingest, clean, explore, transform and structure raw data to make it suitable as inputs for training and run the models. Effort aside, those activities are critical for the quality and time to deliver ML models, as they provide the means to create high-quality data sources and to extract knowledge from them, the true fuel for ML.

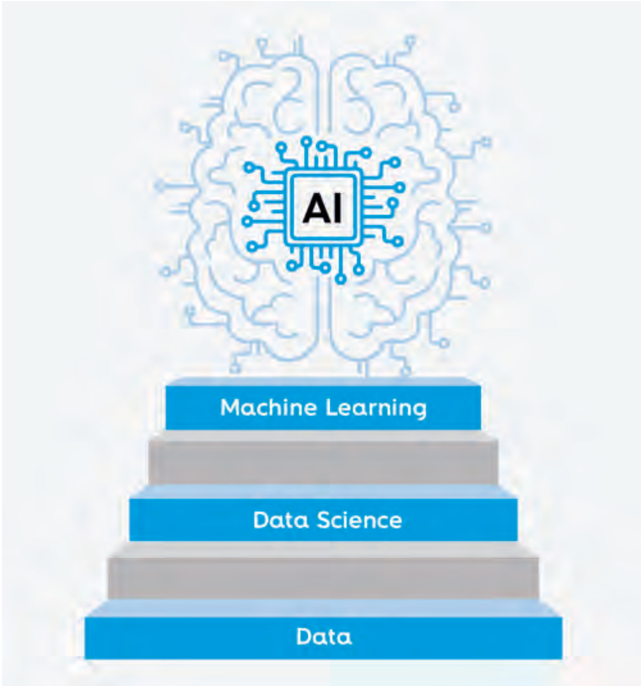


FIGURE 1 – The AI ladder

Considering the techniques behind ML domain, currently, the most popular one used to implement AI systems is deep learning (DL). This technique is a very young field of ML (when compared with other traditional algorithms, decades-old), based on artificial neural networks. DL performs better with large amounts of raw data (as illustrated in **Figure 2**), i.e., tends to increase its accuracy with the increasing amount of training data, where traditional ML models stop improving after a saturation point. This is probably the distinctive aspect of DL responsible for its hype and massive adoption.

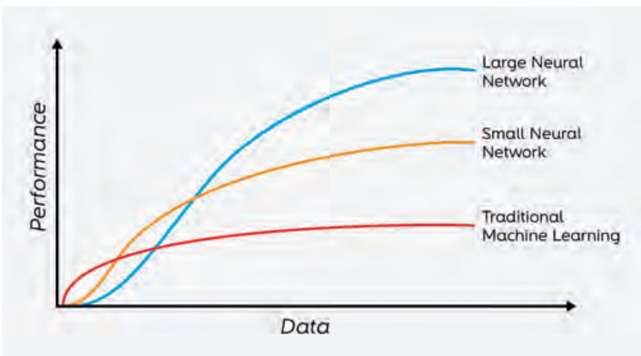


FIGURE 2 – Deep learning benefits

In the last years, thousands of AI and ML technical articles established a new slogan – "There is no AI without IA (information architecture) ", which states a very truthful dependency on data engineering and data science practices and technologies when implementing ML models. Considering this, if a company does not specify and operationalises an efficient data management infrastructure to support the required data engineering and data science activities, it will probably compromise the ability to deliver quality, in time and at scale.

AI spring - the age of implementation

Algorithms are not new. Traditional ML algorithms exist for decades, and even the theory behind artificial neural networks exists since the forties of the XX century. So what changed or is changing to boost its adoption?

FIGURE 3 – Factors allowing AI to accelerate digital business [1]

According to a report released by Gartner [1], as illustrated in **Figure 3**, four major factors stand out today, such that their confluence has enabled a significant tipping point in the potential for algorithms to be broadly deployed across industries and become instrumental in delivering business value and competitive differentiation:

- **Information explosion:** The number of sources of information to which AI technology has access is growing all the time. These sources include sensors, user equipment and other devices, which means that AI technology can now access the essential data to fuel its algorithms;
- **Increases in compute power:** Advanced system architectures, in-memory storage, and more powerful chipsets, combined with highly scalable cloud-based architectures are now widely available. This obviates several infrastructure constraints founded in an enterprise, making the required infrastructure more powerful and affordable, thus accessible for enterprises of all dimensions;
- **Infrastructure and toolset availability:** Cloud-based services combined with cloud infrastructure (eventually powered by GPU and TSU) make available the toolset required to process all types of data sources and to apply ML algorithms. This scenario, along with the fact that most of this toolset is made available as open-source technology, truly democratises the access to data science and ML;
- **Advanced algorithms:** The most relevant ML algorithms (if not all) are today implemented and made available as libraries for widespread programming languages, like Python, R and Java. The inner complexity of those algorithms is highly abstracted with higher-level primitives that make possible to work in data science and ML without the mandatory need of a PhD or decades of research.

In this context, the transition to data-driven digital businesses leveraged by a growing "data and algorithmic economy" is already well-established in many organisations across many industries, with businesses using algorithms daily to influence and support decisions (fact-based decision making) and optimise operational efficiency. It is undeniable; we now live in an AI spring, a moment where the information, infrastructure and toolset virtuously combine to boost the AI infusion in digital processes across industries. It's the age of AI implementation!

Industry impact of AI

Business value

When introducing new technologies or new paradigms in any industry, one of the most relevant aspects that will dictate adoption is the business value it can add to the industry since that introducing novelties on existing processes usually requires a significant level of investments. AI is no exception.

However, there are not so many technologies that can be as disruptive to a business as AI. By integrating human innovation capabilities with AI, hence introducing cognition into any processes, the span of possibilities where this technology can be applied is huge and so are the results that can be achieved. Ranging from operational breakthroughs on existing processes to enhancements on the way the business is promoted and passing through the experience a customer can be presented, AI can disrupt almost every domain where it is applied while creating new opportunities along the way - see **Figure 4** as an example [2]. Resources can be freed to innovative tasks, time can be better allocated, and innovation intersections may be created.

Moreover, having been named by major research companies as one of the top strategic technologies of the decade, it is with no surprise that one can see major industry players making large investments to secure their positioning as

FIGURE 4 – AI reach considering CSP business [2]

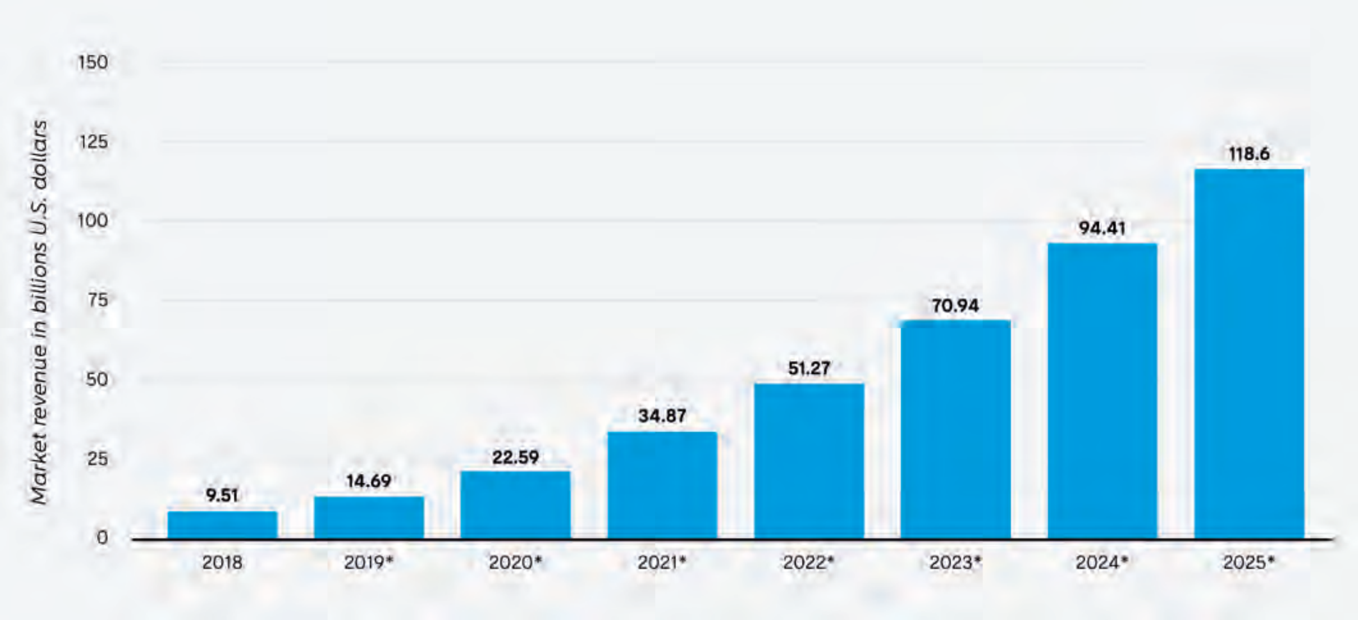


FIGURE 5 – Revenues from the AI software market worldwide from 2018 to 2025 (in billion U.S. dollars) [3]

technology enablers, driving others to apply AI on behalf of their businesses. A new industry is born (see **Figure 5** [3]), promising a strong revenue growth in the forthcoming years.

On the era of the digital transformation at every single industry (logistics, transportation, automotive, healthcare, social care, retail, education, city management) there are clear shreds of evidence of the improvements AI introduction can create as well as of its endless potential, either by massively processing huge amounts of data to infer a result or a trend or by leveraging natural language interaction to create proximity and interaction. Having the capability to leverage valuable information sets, by introducing AI combined with other disruptive technologies (augmented reality, automation, high capacity computing and high throughput networks) is expected to lead us to digital innovations capable of solving some of the major challenges of the society.

Organisation impacts

Digitalisation and AI by themselves are already quite demanding on any industry when looking from the inside perspective, and it only becomes more demanding with the breadth of new technologies sustaining their main fuel: data. This means that additional knowledge will have to be introduced into organisations, new programming languages will have to coexist with traditional ones, and the type of infrastructure being required is just changing. Furthermore, the human resources and skills required to embrace these technologies are also different, requiring a change on the way resources are on-boarded and the approach required to keep them engaged.

This change means that industries, digital service providers (DSP) included, need to evolve their organisations into a more fluid, dynamic and daily interactive working model, where cooperation between teams is absolutely mandatory, and a Medici-like approach is advisable when creating teams. In fact, multidisciplinary, flexible, quick reaction and customer-focused teams will be differentiating on any industry willing to thrive through this new era and failing to adapt to this specific approach will not allow to fully benefit from the assets and innovations driving this technological transformation.

The case for the digital services providers

DSP nowadays play a fundamental role in the overall digital transformation of society. Being at the interception of the communication path of all major industries, consumer profiles and the enormous data hurricane being generated every second makes DSP a critical enabler. Of course, being a central piece on such a disruptive environment and moment means that there is a need to look inside DSP internal frameworks and make them agile and resilient enough to maximise the benefits. As seen in **Figure 6**, having AI assisting network assets decisions or predicting artefacts behaviour are some of the possibilities that are becoming a reality.

FIGURE 6 – Key enablement dimensions of AI on DSP

The network evolutions will provide the DSP with the capability to address a much bigger area in terms of a business domain. This means the network complexity will increase due to the need to handle new elements, additional proximity, multiple combined access technologies, exponential growth of connected elements and, above all, multiple demand needs that metamorphose through time and space. Coping with this new reality means that networks will need to adapt, scale and adjust in real-time or, even better, in advance of the needs. AI thus become a powerful asset in assisting engineering such capabilities.

As a consequence, managing these AI-powered networks and operations is a challenge by itself. Multiply by several orders of magnitude the number of managed elements, the changes in self-configuration and the self-healing capabilities and one can understand the level of complexity involved! Doing it the traditional way won't work, and this is where AI will introduce new levels of productivity into the operations: creating algorithms that will assist humans by predicting when and where an issue will occur, by combining multiple sources from multiple domains, is now becoming at reach; working on huge data sets to identify the root cause of a problem can now be significantly reduced, and the optimisation opportunities keep showing up.

Moreover, when considering the business engineering and the users experience, introducing AI to improve it (not only when using DSP services, but also by applying cognition algorithms to the way services are used to create improved and personalised offer or campaigns) is considered the low-hanging fruit for DSP willing to quickly secure their positioning on the market. See, for example, the astonishing progress that has been achieved on AI-enabled natural language interactions, especially if one takes into account the cost-benefit relation. It is with no surprise that industry is seeing the spread of digital assistants for the most diverse aspects of customer interaction and with increasing levels of success and adoption. Also, profiting from the vast amount of information about service usage patterns existing within DSP to provide the customer with its dedicated experience is not a novelty. However, AI introduces a new stage on the agility that can be imposed on adapting to the needs of the customer, and on the flexibility that can be sought, since the depth of use of the existing information is much superior now.

Therefore, is covering experience or business opportunity depleting the opportunity space for AI? Absolutely not.

Altice Labs AI strategy

Altice Labs is a long term provider of solutions for communications and digital services industries, either for Altice Group operations and the external market. In this context, Altice Labs has a portfolio of solutions that addresses several activity domains, including network solutions, service platforms, operations support solutions and business support solutions, exactly the domains where service providers are investing more in data science and AI, going beyond experience and business.

More than an opportunity, this context drives a necessity. On the opportunity side, Altice Labs is in a privileged position to leverage its experience and expertise in the areas addressed by its solutions, thus to extract additional value from managed data and to create new use cases on top of its solutions using AI technology. On the necessity side, not making this investment will create a significant gap in existing solutions in the years to come.

As illustrated in **Figure 7**, betting in AI infusion into the existing portfolio is one of the more relevant dimensions of Altice Labs current strategy, which by itself drove Altice Labs to increasingly invest in data science and AI internal knowledge to define work methodologies and to establish the required infrastructure and toolset. This bet not only unleashes the infusion of AI into existing portfolio but also enables Altice Labs to address AI use cases in functional and data domains previously unexplored and not only on the ones being addressed by existing portfolio, empowering Altice Labs to start positioning itself as an AI competence centre for the Altice Group.

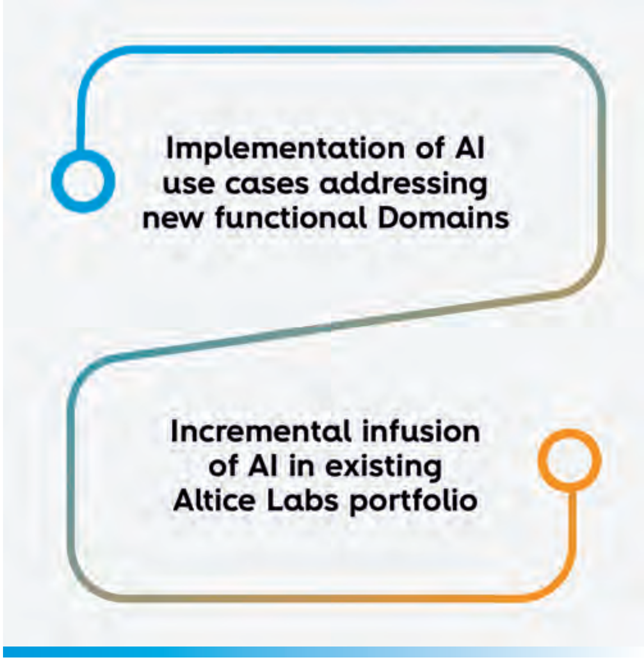


FIGURE 7 – Strategy for AI-infused portfolio

Altice Labs cognitive team

To support the operationalisation of its AI strategy, Altice Labs created the cognitive team (see **Figure 8**). This team is an instrument to catalyst the implementation of AI use cases in Altice Labs portfolio in cooperation with all Altice Labs business units, and also to address use cases that flow from within the operations and that can be addressed by the type of knowledge hereby created.

As illustrated in **Figure 8**, the cognitive team encapsulates key work dimensions:

- **Use cases:** prospect and implement, in strict cooperation with business units and customers, business-relevant AI use cases;
- **Methodology:** specification of a work methodology to implement AI use cases;
- **Human resources:** Define roles and competencies that a team should have to tackle previous defined AI use cases;
- **Infrastructure:** specify and manage the physical infrastructure and toolset required to support the implementation of those AI use cases.

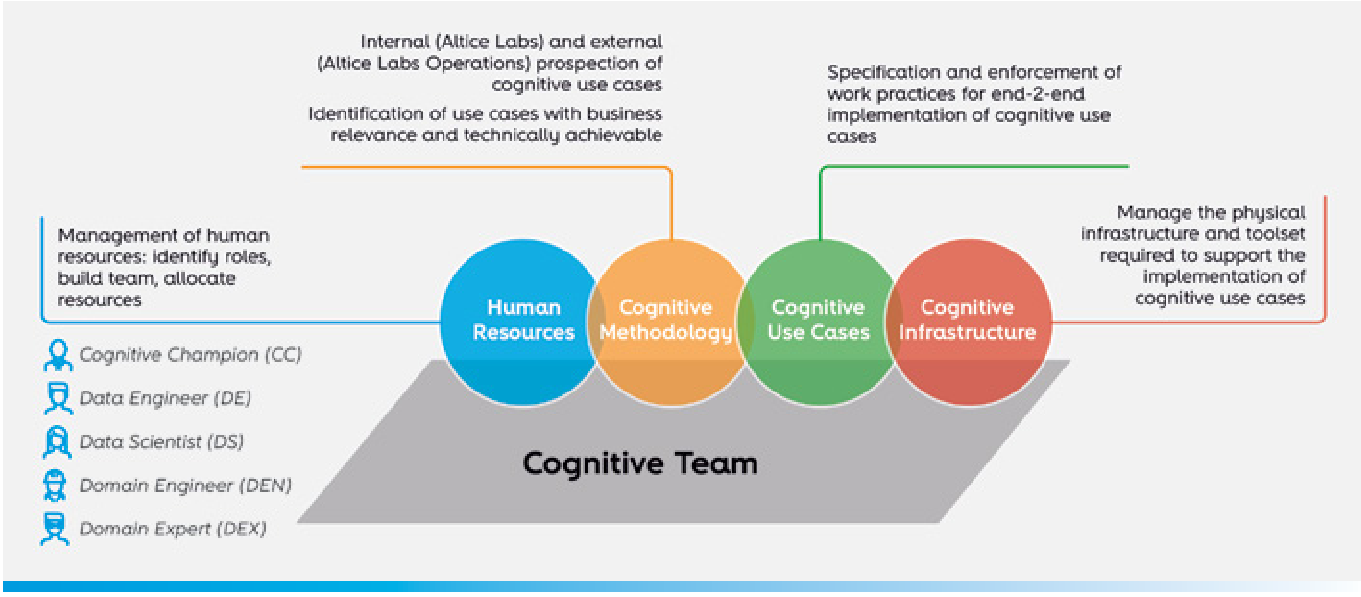


FIGURE 8 – Key work dimensions of Altice Labs cognitive team

For this team, the cognitive methodology is its cornerstone. As illustrated in **Figure 9**, the cognitive methodology specifies the workflow of stages and activities that must be conducted for a successful use case implementation, as well as the competences required to lead and execute each work stage.

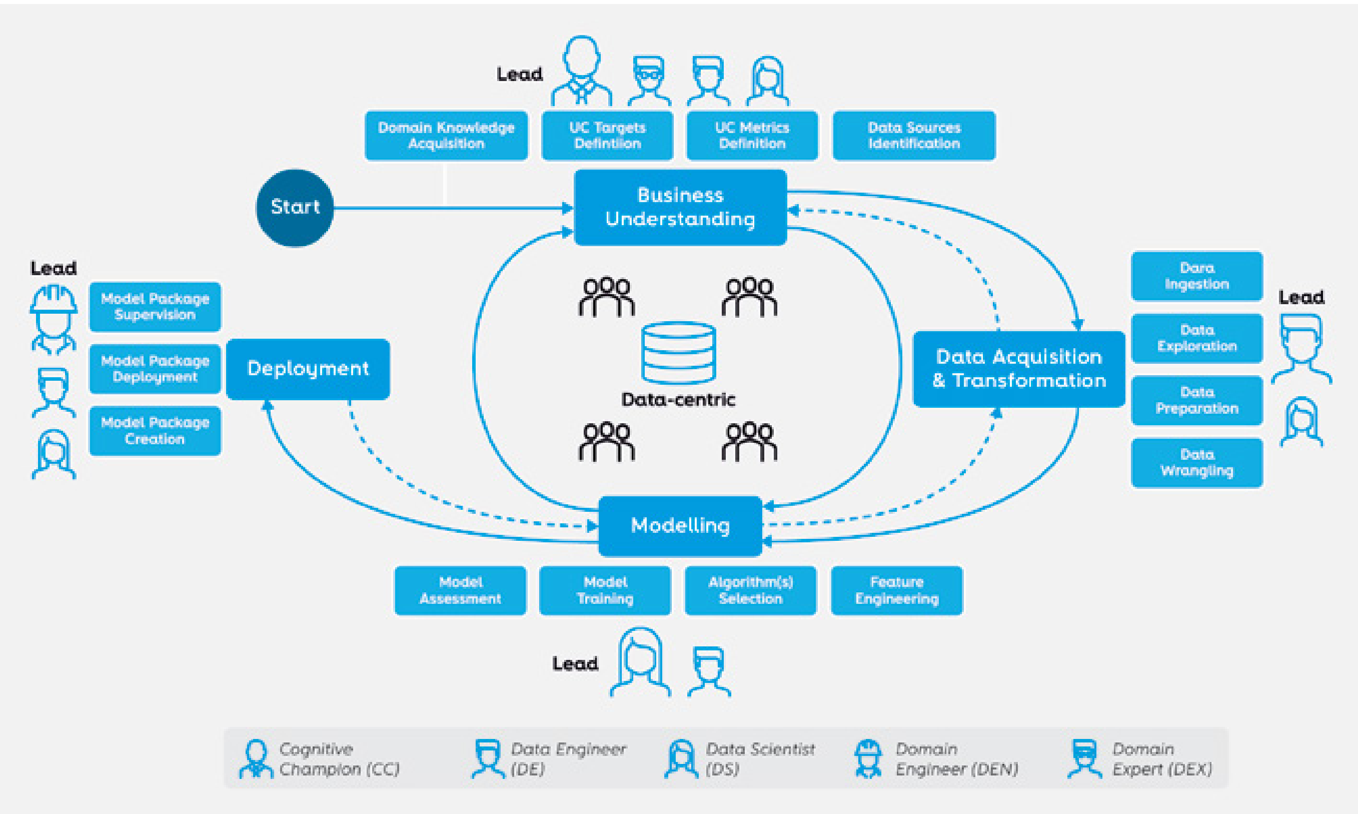


FIGURE 9 – Cognitive methodology of Altice Labs cognitive team

Using agile principles as its workstyle, closely interacting with domain experts and Customer teams allows for a quick-try-fail-fast mentality that keeps the team moving forward into consistent progress that can quickly move into production and produce fast results.

Altice Labs AI initiatives

Altice Labs has been researching on AI arena for the past years, and beyond the creation of the cognitive team, several use cases were and still are subject of investigation in collaborative projects with universities and other research institutes. Nevertheless, since its inception in the first half of 2019, and together with Altice Labs business units, the Altice Labs cognitive team start to prospect candidate use cases having two main characteristics in mind:

- **Business value:** what is the expected business value of the use case to Altice Labs Customers? Will its outcome express it?
- **Feasibility:** Do we have the (best) data? Is it addressable using available technology? Do we have the required proficiency to go for it?

As a result, new projects were created to support the implementation of, among others, the following use cases:

- **Cognitive network operation centre (NOC):**
  - Prediction of mobile network critical faults;
  - Prediction and root cause analysis of low mobile networks accessibility.
- **Cognitive call centre:**
  - Pattern-based diagnostics of customer services issues and automated corrective actions recommendations.
- **Cognitive infrastructure maintenance:**
  - Prediction of customer premises equipment's anomalies (set-top-boxes & home gateways) and optimisation of repair processes.

Also, from the results of previous R&D initiatives, a new product line was created bringing digital assistants to Altice Labs portfolio. Its value proposition is entitling non-expert users in NLP to create by themselves digital assistants for their businesses.

Conclusion

AI is profoundly impacting every industry: automotive, media, finance, insurance, travel, healthcare, online gaming, communications, digital services and communication providers; you name it. We live in the age of implementation, where knowledge, tools and infrastructure combine virtuously to make possible the implementation of AI-enabled systems at scale, by large and small companies. This democratisation of AI is what is making possible the wave of transformation seen every day.

Industrial processes will improve significantly in future due to the adoption of data science and ML technologies, becoming fully automated and eventually autonomous. New processes for problems not attainable today will emerge. All these transformations will have business drivers. Enterprises ignoring this will stay aside from this industrial revolution and fall.

To take advantage of the wonders around data science and AI technologies a company must transform itself and evolve from an organisation to an ecosystem where human resources, infrastructure and work methodologies leverage the company growth by creating solid results at speed demanded by the market. Altice Labs defined and operationalised a strategy to make it happen and is now starting to deliver the results of it.

References

[1] Gartner, "Maximize Digital Business Value Creation Through AI Algorithms," 02, Stamford, 2018.

[2] Ovum, "The AI Opportunity for Communications Service Providers," Ovum, 2019.

[3] Statista, "Statista," Statista, 25 11 2019. [Online]. Available: https://www.statista.com/statistics/607716/worldwide-artificial-intelligence-market-revenues/.

02 Cognitive ops services

Carlos Guilherme Araújo, Altice Labs
carlos-guilherme-araujo@alticelabs.com

Gonçalo Nuno Gaspar, Altice Labs
goncalo-n-gaspar@alticelabs.com

João Tiago Guerrinha, Altice Labs
joao-t-guerrinha@alticelabs.com

The future of operations is data-driven, intelligent and autonomous. In a word, it is cognitive. Following this vision, Altice Labs OSS suite is becoming a cognitive-infused service platform, able to create new operational insights and create conditions for autonomous decision or advanced human decision support, a new dimension of operational efficiency.

Keywords

ASOP; Cognitive network planning & design; Provisioning process; Assurance; Cognitive operations

Introduction

The communication/digital service providers (CSP/DSP) are always on a challenge between giving the best service to the customer, reduce costs, evolve the technology and build their network faster than its competitors. As networks are becoming more and more complex, the human effort has proved to be insufficient to deal with this complexity increase. With the current tools and platforms, mostly human-driven, it is hard for operations to keep their goals at a high level of quality, speed/time-to-market and satisfaction while reducing the costs. Computational processing is cheaper, more powerful and incomparably faster when compared with human resources, and consequently, CSP/DSP are now looking towards automation to cope with it.

Network operations, such as planning and construction, are complex by nature, and network maintenance tasks are becoming more effort-consuming. With the help of artificial intelligence (AI), and particularly with machine learning (ML) techniques, it's possible to trust machines on planning and designing the network. This level is achievable by using multiple datasets or conditions from business/operations support systems (B/OSS) to build the network by itself and deliver it to construction on the field.

Designing provision or other operational processes on service providers' ecosystems is currently a manual process performed by engineers and analysts that tailor B/OSS to a particular operational and technological reality. Both continuous delivery (CD) and continuous integration (CI) provide the framework for the CSP/DSP to be fast and reliable on the delivery of process improvements. Nevertheless, those improvements are designed solely by humans that are looking at process analytics, trying to find patterns and refinement opportunities. Using the tools already available today (strong analytics, a centralized service and resource catalogue, the CD/CI framework, and a multipurpose and multi-technology workflow engine) AI/ML can be used to automate B/OSS process improvement. One significant advantage is being able to deliver an impressively efficient continuous improvement framework that can have a dramatic effect on operational costs, resource usage and service deployment time.

Today's network/service operation centres (N/SOC) are still very reactive, but the shift to a proactive mode has already begun. By using AI/ML, it's possible to gather and process network data in real-time and automate network functions, enabling faster decisions. By using these techniques to find patterns and anticipate network issues, a self-healing network can be set up for the customers' service to be fixed before the customer is even impacted [1]. These fixes can be applied automatically, or semi-automatically after validation from a network engineer. Another use-case of AI/ML is the ability to distinguish between real problems, that must be addressed, and noise, which can be ignored, helping human operators to focus their attention where needed.

The future of operations is intended to be data-driven, autonomous, and intelligent – in one word, cognitive. To cope with these requirements, inventory, fulfillment and assurance domains on OSS must evolve in that direction, creating an agile, flexible and cognitive-infused architecture. It is also essential to consider critical business use cases that can take advantage of AI and in particular, ML.

In the next sections, we will present a brief insight on how AI/ML can help OSS make the change to cognitive. We will also introduce the target architecture to achieve it and describe three use cases that apply to the OSS domains that unleash the potential of cognitive operational service.

Towards a cognitive network planning, fulfillment and assurance

Both AI/ML and cognitive computing rely on the use of a machine's capacity to learn from past experiences. However, the main difference is that AI/ML makes use of its detailed inspection to automate the decision, while cognitive computing provides insights to an operator so a better-informed decision can be made [2].

CSP/DSP must move from a reactive to a proactive, predictive, and cognitive operations mode through the adoption of AI/ML in network operations. Doing so would significantly lower operational expenditure, improve customer satisfaction and enhance resource utilization [3].

Figure 1 illustrates the ongoing evolution of how to perform actions, from a rule-based approach to an AI/ML scenario. Regarding the existing rule-based automation, operators handle the information provided by the intelligence information systems and perform actions based on it. That same data can be used to decide on how to automate actions using fixed rules. Using AI/ML automation allows for the use of self-improving models to enable machines to suggest actions to the operators or even to perform them autonomously.

On the inventory domain, the core of engineering in a communications network is the technology-independent question: how to plan, design and upgrade the network to its maximum capacity, to meet customer needs and reduce costs? In the network design problem, associated with this question is also the need to decide the link capacity to find a solution with minimum effort and cost.

Network planning tasks consider the type of technology, the number of customers/services to attend, and the experience of the designer to plan the best routes and reduce the bill of materials (BOM), while keeping in mind that more recent network designs are more easily updatable. From the network operator perspective, the main goal is, invariably, to maximize net revenue, i.e. the revenue generated by customers minus equipment, software and operational costs of the network. It's of paramount importance for the network operator to plan the detailed evolution of investments over time as accurate as possible,

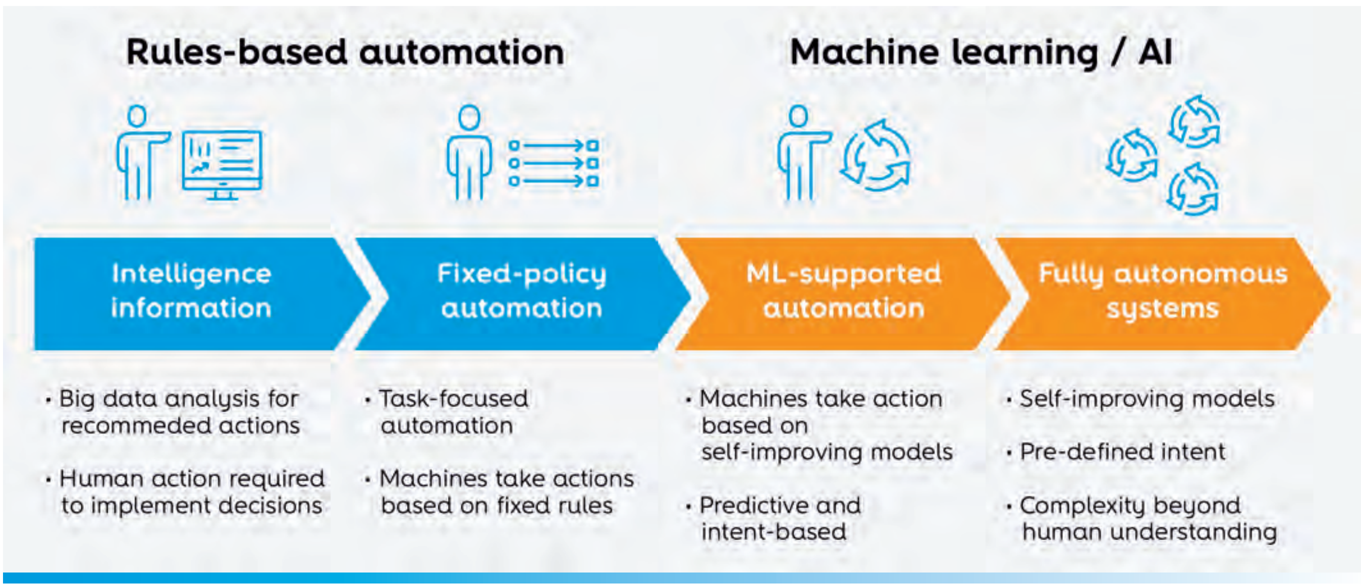


FIGURE 1 – Proactive to cognitive evolution [4]

which not only maximizes the net revenue but also creates a significant advantage over the competitors [5].

In the fulfillment domain, the best way to know if the organization has a well-implemented process is when no one notices its presence. Most of the time, provisioning and other related operational processes are invisible to customers, operators, technicians and other parties that make use of it. Typically, they are noticed only after a complaint from any of those parties, and some detailed analysis detects a problem. On the other way, automated and comprehensive processes are one of the main building blocks of a fully automated telco operation. They represent the means by which autonomous and closed-loop operations can effectively perform the changes that are going to improve the network and the customer's services.

With cognitive operations, assurance systems can continuously sensor the network and autonomously trigger diagnostics whenever a fault or degradation occurs. If the system is confident about the accuracy of the diagnostic, the automated resolution process can be triggered. The resolution must be confirmed once again by another automatic diagnostic process.

To achieve these goals, change the actual architectures to a new concept based on services is a mandatory task. The autonomous service operations platform (ASOP) architecture main objective is to serve that purpose.

The ASOP architecture

ASOP is Altice Labs reference OSS functional architecture for product and solution implementation, represented in the following **Figure 2**.

This architecture is microservices-based and cloud-ready, following an everything-as-a-service (XaaS) model that includes, among others, inventory-as-a-service, fulfillment-as-a-service

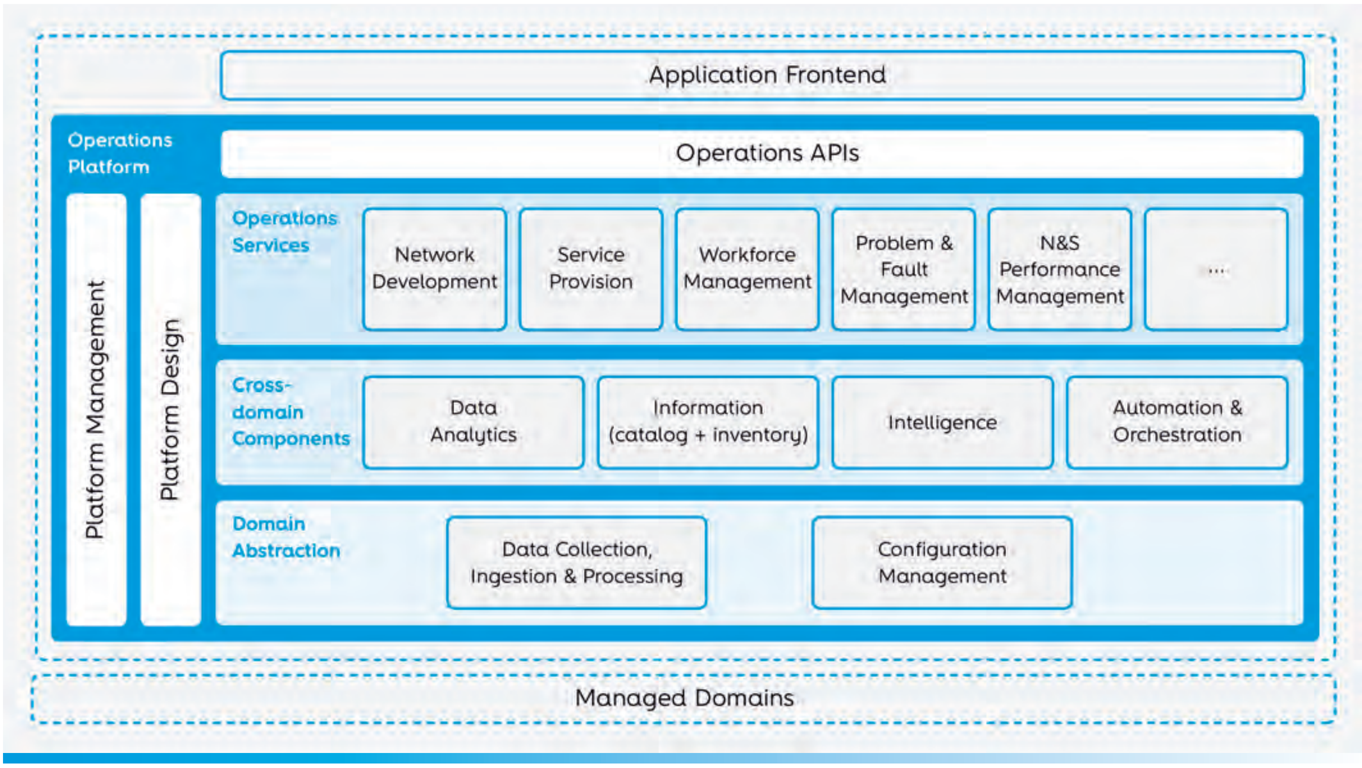


FIGURE 2 – ASOP functional architecture

and assurance-as-a-service, privileging the use of open source software and products.

ASOP serves both physical and virtualised managed domains, like network elements (NE), network management systems (NMS), virtualised infrastructure managers (VIM), software-defined networks (SDN), etc. The platform comprises the following main layers and components:

- **Operations APIs**: the API layer, based on TM Forum Open APIs to increase systems interoperability;
- **Operations Services**: this layer offers a set of the essential OSS services that deliver value to telcos and address their operational needs, including Network Development, Service Provision, Workforce Management, Problem & Fault Management, and Network and Services (N&S) Performance Management components, among others;
- **Cross-domain Components**: this layer comprises general-purpose building blocks to support OSS services, like Data Analytics, Information (catalog+inventory), Intelligence, and Automation & Orchestration components;
- **Domain Abstraction**: consists of a group of components that abstract the managed domain, including Data Collection, Ingestion & Processing, and Configuration Management components;
- **Platform Design**: this component provides support for onboarding new use cases;
- **Platform Management**: manages infrastructure and application lifecycle;
- **Application Frontend**: the presentation layer, including graphic users interfaces (GUI) and other user interfaces.

All ASOP components can be organized to respond to various functional use cases, and deliver the existing functions provided by the current OSS domains as well as new scenarios like cognitive operations. Some use cases that will be described more thoroughly in the next sections.

The cognitive network planning and design use case

As referred before, inventory may have many challenges that can be solved using a cognitive approach. In this section, we will focus on a particular use case of network planning and design. Many organisations consider fiber-based access networks as a major solution to make the most out of the higher-speed available for service usage. Fiber-based networks can be delivered to customer premises through point-to-point (P2P) and point-to-multipoint (P2MP) technologies, which increase the difficulty of planning and design the network, due to the large number of variables to consider. So, to create a cost-effective gigabyte passive optical network/fiber-to-the-home (GPON/FTTH) requires considering as many factors as:

- Headend position;
- Optical splitter position;
- Maximum splitter ratio;
- Optical distribution point position;
- Maximum distance;
- Routes;
- Number of surveys to attend;
- Accomplish the optical budget.

The two main advantages of automating/optimising network design are minimising the capex and reducing time-to-create from days to hours, and as so, using a cognitive AI-based approach allows the operator to automate the process of planning and design the network. In this use case, the input data can be the headend position, optical splitter position, etc. Business rules may be the optical budget, or cost of construction, among others. Then, by parsing the information

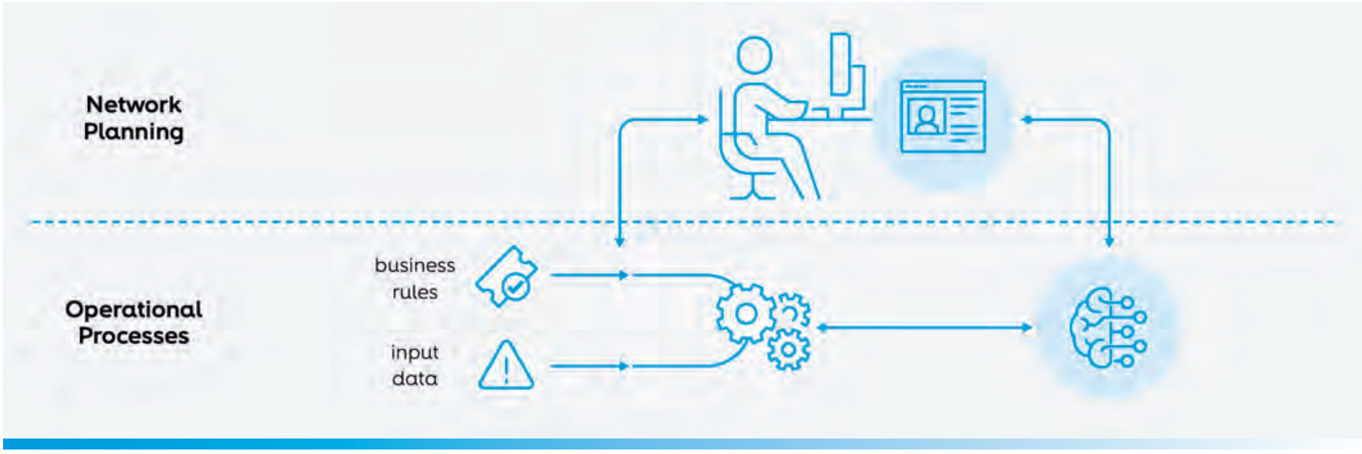


FIGURE 3 – Data flow

and compute it using an AI/ML approach, the operator will have the hardcopy outputs to analyse the results and change them if needed.

More generally, the system's input data can come from a geographic information system (GIS) database or other sources. This input data contains the infrastructure information: roads, installation points, routes, surveying, among others. Business rules define cost constraints (on placing cable, ducts, equipment, etc.), and the desired ratio of coverage customer.

Given the inputs and business rules, the information is parsed and prepared to be computed by AI/ML algorithms that will give the operator multiple possibilities of how to design the network. The outputs consist of a presentation of the georeferenced view, and the auto-generated results (like the surveying, routes, cable network,

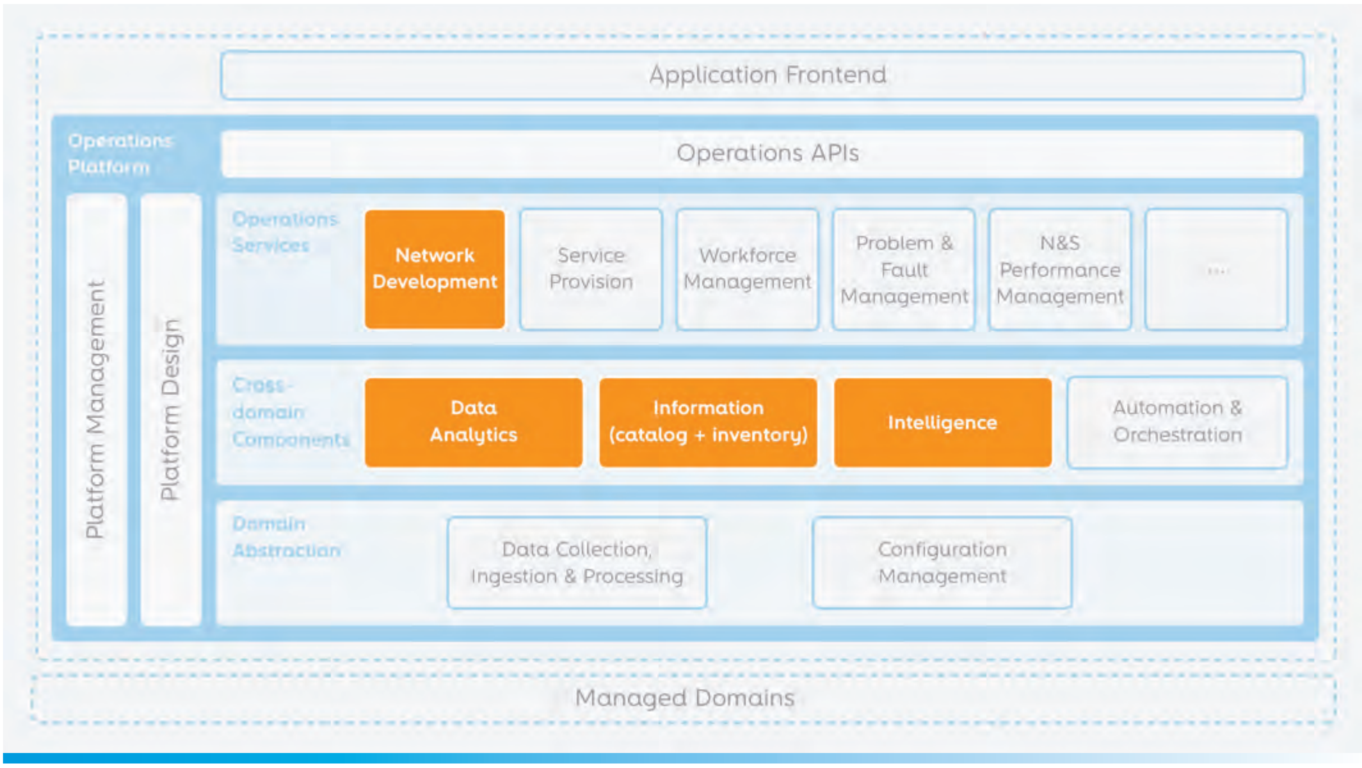


FIGURE 4 – ASOP modules used in the cognitive network planning and design use case

optimised BOM, etc.) of the desired network. The operator can change what is presented or even alter business rules and make other adjustments, to have different solutions. The acceptance and adaptations of each project will feed the AI/ML engine to improve the outcome of next projects, leading to a "smarter" process. **Figure 3** illustrates the process described before.

According to **Figure 4**, the ASOP components needed in this use case are the Network Development, as the main operation service component, and the cross-domain components Data Analytics and Information (catalog+inventory) for providing the necessary inputs and business rules. The Intelligence component is responsible for information parsing and computation to produce the desired outputs.

The provisioning process optimization use case

Provisioning processes are a critical component of the CSP/DSP ecosystem. They ensure that service instance creation, modification, and termination are successful and done in the most efficient way possible. These processes must be resilient enough to withstand any performance issue, failure or unforeseen behaviour of any of the systems with which they interact. Those systems include other CSP/DSP support systems, service platforms and the managed network that delivers the services to the customers. Additionally, all processes that are related to the workforce must be bulletproof to ensure they spend the least amount of time in the customer premises, to minimize the inherent costs, but also to increase customer satisfaction.

Currently, all these processes and their improvements are designed by analysts and engineers, experts in this field of work. But although having access to a rich set of tools that make their job easier, this is always done through manual analysis. Another aspect is that these processes, being so elaborated and mission-critical, are not modified unless it is absolutely necessary due to the associated risks of such changes. By introducing the concepts of CI/CD to the design, it is possible to automate the process development and release it faster. The process designers are responsible for the management of the CD pipeline, as depicted in **Figure 5**.

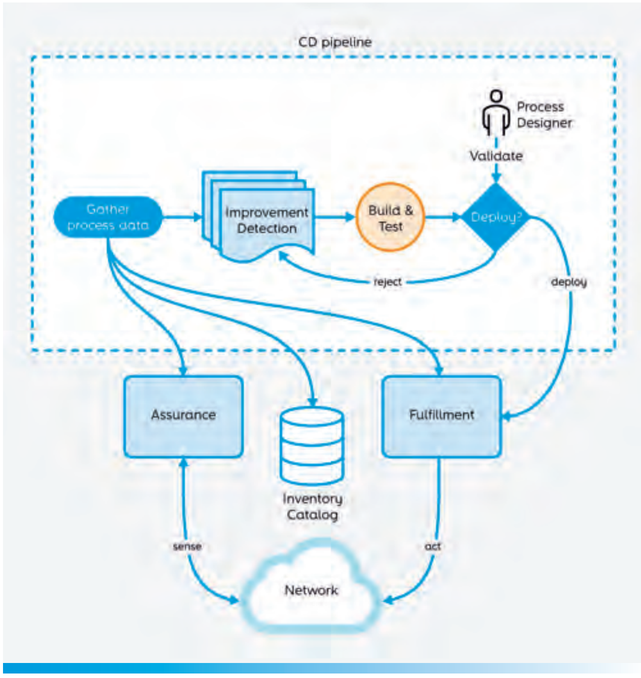


FIGURE 5 – CD pipeline

The objective of cognitive process optimization is to allow these improvements to be added using an automated approach, shifting from a reactive to a proactive perspective, as mentioned before. This shift is achieved via AI/ML, by continuously monitoring the overall performance of the processes associated with on-boarded services, and issuing improvement suggestions to the process designers along with the justification for these suggestions [6]. For example, an improvement can be changing the order in which specific network elements are configured, or adjust a simple communication timeout. If the designer approves the change, it will be propagated directly to the pipeline and deployed into the live environment.

Engineers and analysts responsible for those processes will have access to a potent tool to aid them in the difficult task of process optimization, being able to focus more on the validation and approval of the suggested improvements rather than discovering and implementing these improvements themselves.

Figure 6 highlights how this use case maps on the components provided by ASOP. Service Provision, from the operation services domain, holds the knowledge on how to manage a service for a specific client. Data Analytics, Information (catalog+inventory), and Automation & Orchestration components, on the cross-domain components layer, are used to contextualize the network data with the entities and the executed processes, along with their results. Finally, the Intelligence component identifies and detects the improvements that can be added to the Automation & Orchestration component. As for the abstraction domain, the Configuration Management component can be affected by any identified improvement in its own processes. Finally, the Platform Design component allows for the onboarding of any entity or workflow change resulting from any improvements, and aids in its deployment.

The closed-loop use case: from sensing to acting through 360-degree vision

It is undeniable that our world is continuously changing and the dynamics in which we spend our daily lives will take us into unimaginable paths. New technologies like 5G and mass internet of things (IoT) devices will introduce new paradigms in the assurance field, supported by increasingly virtualized and programmatic networks and a massive array of data analytics. The most advanced NOC have started, in recent years, the path of automation, but supported by basic rules and constraints. Nonetheless, this automation is

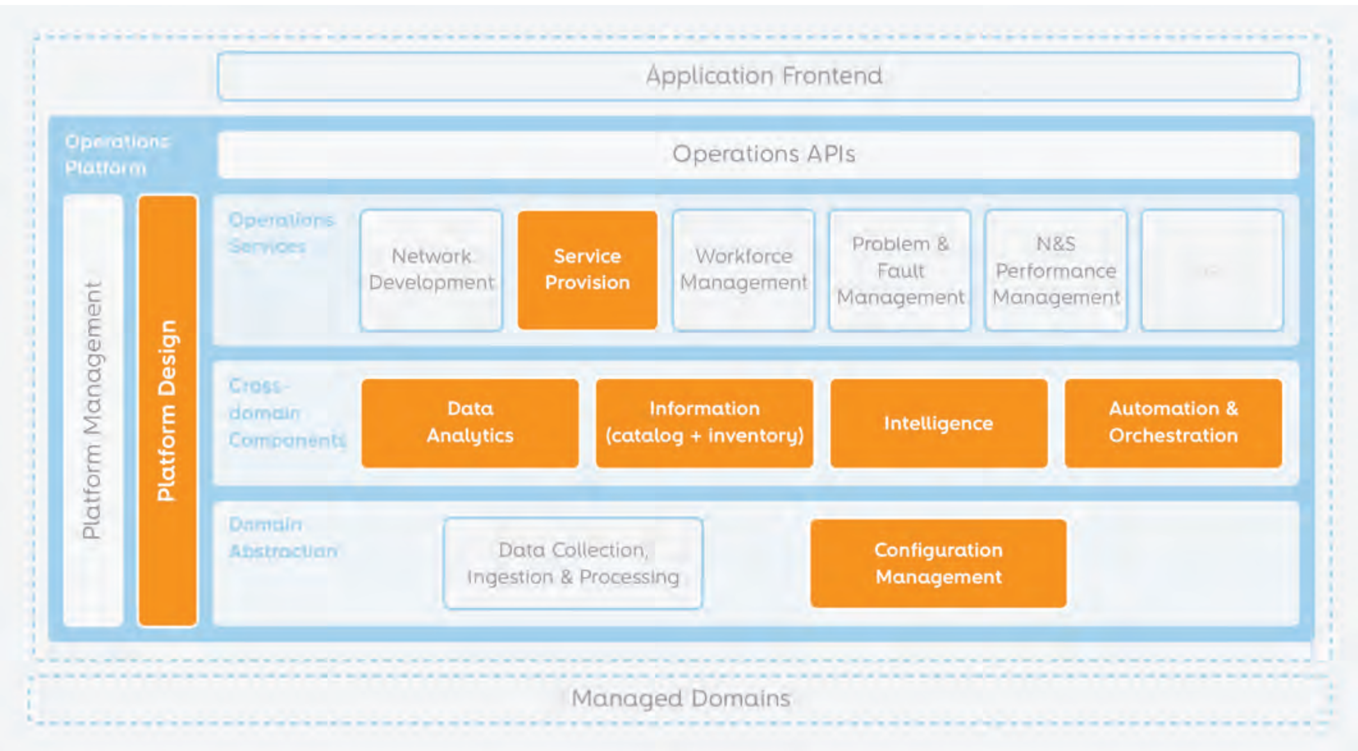


FIGURE 6 – ASOP modules used in the provisioning process optimization use case

not yet able to, autonomously, detect the cause of a failure, trigger more advanced troubleshooting processes or even act on the network to provide a better quality of experience.

If, on the one hand, network operators are incessantly overflown with information such as alarms, performance measures, network topology, network and service transactional data, on the other, they suffer the absence of an effective way to diagnose and resolve issues rapidly. Without a quick and effective diagnose, mean time to repair (MTTR) is penalized and can cause not only network and service unavailability but also affect customer's satisfaction. For example, in a typical scenario, a network operations engineer must access multiple systems and knowledge bases to diagnose a problem and to identify the root cause. Simultaneously, network technicians must search if there is already a trouble ticket (TTK) opened for that problem and, if not, a TTK must be created. Additionally, the right resolution must be applied or, if not possible, the ticket must escalate to the next support tier. Not only is this process lengthy, inefficient and tedious, but it also presents a more dramatic issue – it's not scalable [7].

Faced with this dilemma, cognitive operations appear as a new era and the necessary path to take. In this new paradigm, the implementation of closed loops enables the full path automation, from sensing to confirming and then to acting.

Network technician's workforce is supported by a new set of exceptional tools that will [7]:

- Be able to predict problems before they impact customer service;
- Provide an integrated 360-degree view of the problem, with details of the alarms, performance measures, related TTKs, recent change requests on the component, weather information, social buzz, service impact analysis and more (as shown on **Figure 7**);
- Track actions that are taken autonomously on the network by reconfiguring or repairing in case of failure;

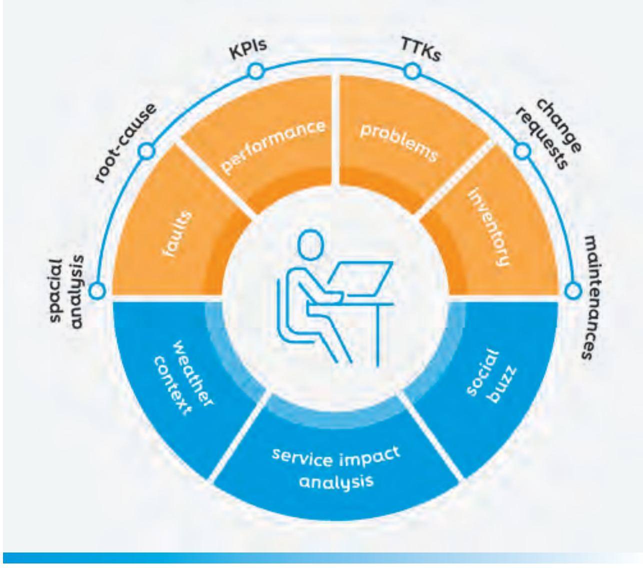


FIGURE 7 – 360-degree view

- Access advanced troubleshooting issues that autonomously diagnose problems by pointing to root causes and possible action for resolution;
- Make use of bots that guide the screening and resolution process.

From sensing to acting, the whole process can be completely autonomous or require human action in case of uncertainty or confirmation of the next step to take.

However, the decision to depend on whether or not a human intervention is necessary should be based on analytical criteria and the level of confidence, which can be obtained from the feedback of the sense-act-confirm cycle. If this value is above a predefined threshold, full automation is suggested. On the other hand, a degradation on this rate might indicate that the flow may require a human review.

This use case illustrates the potential of the new ASOP architecture by taking advantage of most of its modules.

Regarding the operations services level, the 360-degree view takes advantage of the Workforce management, Problem & Fault Management and N&S Performance Management components. In the

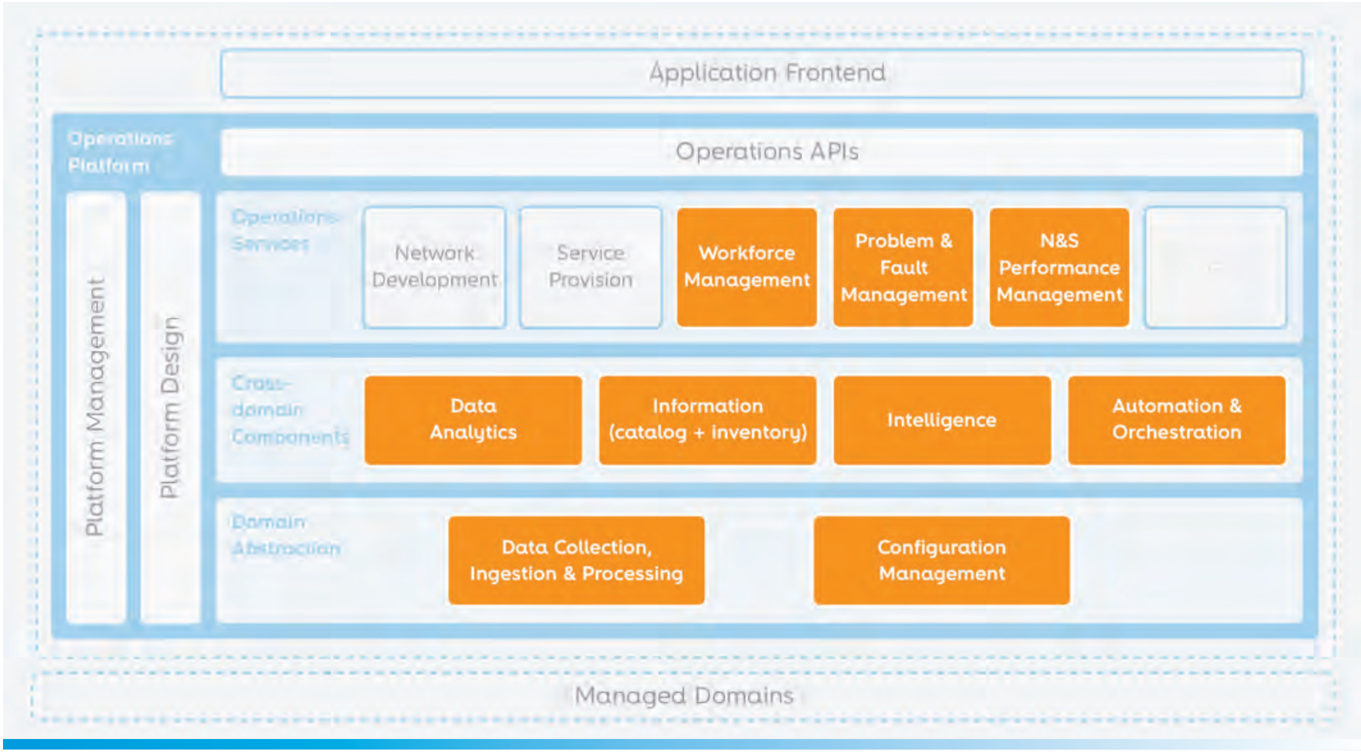


FIGURE 8 – ASOP modules used in the closed-loop use-case

cross-domain components level, all modules are employed in the sense-act-confirm cycle. As far as abstraction domain is concerned, Data Collection, Ingestion & Processing is involved in making data available to the operations services layer. The Configuration Management component is one of the inputs of the 360-degree view. **Figure 8** illustrates the use of the ASOP modules, as described above.

Conclusion and future work

Tomorrow's network management systems will undoubtedly need to be able to auto-provision, auto-scale and auto-heal for the telco industry to lower costs and improve performance. The changes will happen through a closed-loop process that collects data, identifies problems, recommends or makes decisions, and then takes action [8].

This evolution is vital, but it is not possible without a shift from reactive operations to proactive ones, from those to a predictive understanding, and subsequently to cognitive operations.

Nevertheless, cognitive operations are not the last stage. Instead, they are the intermediate step between predictive and prescriptive ones. The major difference between these two concepts is that while the previous forecasts the potential future outcomes, the latter helps you draw up specific recommendations. In fact, prescriptive operations use predictive ones to arrive at the different options available along with their anticipated impact on specific key performance indicators [9].

The adoption of ASOP's architecture is a crucial stage to guarantee that there is a functional alignment between all the components to attend any future technological challenge. This architecture will allow the creation of new tools for decision support while creating new value. Such is the ultimate paradigm of cognitive organizations: to be self-organized, data-driven, intelligent, adaptable to change and eventually autonomous. In this new generation, operators need not only to be intelligent and highly autonomous but also lean, agile, predictive and showing real-time awareness.

References

[1] A. Yazdan, "Taking Telecom to New Heights with Artificial Intelligence," *Intel Corporation*, 2018. [Online]. Available: https://www.intel.ai/taking-telecom-new-heights-artificial-intelligence/#gs.fsyz9r.

[2] M. Newman, "AI and Its Pivotal Role in Transforming Operations," *TM Forum*, 2018.

[3] R. Das, "Cognitive Operations: The Future of Telecom Networks is Here," Tata Consultancy Services Ltd, 2019.

[4] STL Partners, "Telco AI: How to Organise and Partner for Maximum Success," *STL Partners*, 2019. [Online]. Available: https://stlpartners.com/research/telco-ai-how-to-organise-and-partner-for-maximum-success/.

[5] M. Pickavet, C. Develder, E. Baert, and P. Demeester, "AI Techniques for Planning Telecommunication Networks," *Proc. Int. Conf. Artif. Intell.* (IC-AI 2002), 2002.

[6] T. McElligott, "Frameworks and Standards are Required for AI," *TM Forum*, 2019. [Online]. Available: https://inform.tmforum.org/insights/2019/09/frameworks-and-standards-are-required-for-ai/.

[7] U. Mangla, S. Sadagopan, and M. Thomas, "Cognitive Network and Service Operations," *IBM Telecommunications, Media and Entertainment*. 2017.

[8] Y. Stein, "5 Overlooked Principles in the Race for Autonomous Networks," *TM Forum*, 2019. [Online]. Available: https://inform.tmforum.org/insights/2019/10/5-overlooked-principles-in-the-race-for-autonomous-networks/.

[9] Cigniti Technologies, "Moving from Descriptive Metrics to Predictive & Prescriptive Metrics," *Cigniti Technologies*, 2019. [Online]. Available: https://www.cigniti.com/blog/moving-from-descriptive-metrics-to-predictive-prescriptive-metrics/.

03 Predictive fault management

Carlos Guilherme Araújo, Altice Labs
carlos-guilherme-araujo@alticelabs.com

Guilherme Cardoso, Altice Labs
guilherme-j-cardoso@alticelabs.com

José Sousa, Altice Labs
jose-nuno-sousa@alticelabs.com

Manuela Coutinho, Altice Portugal
manuela-f-coutinho@telecom.pt

Pedro Neves, Altice Labs
pedro-m-neves@alticelabs.com

Following the network and services evolution, operational management systems are also evolving from a reactive towards a proactive paradigm. Anticipating and mitigating potential issues in the network is crucial to maintain customers satisfaction. This article addresses one of the key areas in operations that are currently using AI/ML as a toolset to support their evolution - proactive alarms management. An overall perspective of the alarms evolution is provided, followed by the design and implementation of a specific alarm management use-case applied to the mobile domain.

Keywords

Alarms; Proactive; Neural networks; Alarm prediction; Cognitive

Introduction

Alarm management systems are critical in today's operations to enable the prompt resolution of network problems, therefore minimizing service interruption, impact in the quality of service (QoS) and impact in customer satisfaction (CeX).

From an operational perspective, existing alarm management solutions consume events from network elements (NE) – in the radio access network (RAN), passive optical network (PON) or other domains - and, based on a set of predefined human rules, convert those events into alarms. Typically NE are only able to report events when the problem is already affecting operations. This state of things defines the current fault management paradigm as a reactive one based on a pipeline of diagnosis followed by damage mitigation and resolution actions.

Evolving from a reactive towards a proactive approach is paramount for the operational management systems. Next-generation operations will be mostly machine-based and human-assisted, relying on advanced mathematical algorithms and high-capacity computational systems to enable early detection of network problems and allowing immediate mitigation actions. All this will, therefore, reduce the impact on CeX.

Machine learning (ML) technology provides the required toolset to evolve from a reactive paradigm to a proactive one. Applying ML techniques to available operational data makes it possible to predict future problems and allows the implementation of new processes to prevent degradations from occurring, creating a new pipeline of precocious diagnosis followed by preventive actions. This new paradigm will help to enable the nirvana of improved QoS and zero CeX impact due to operational issues.

In this article, we describe the cognitive challenge, and technical approach followed to implement an alarm prediction use case (UC) using real-life/production operational data, provide an in-depth analysis of the results obtained and elaborate on the operational benefits of such approach.

Alarm prediction use case

This section provides an overview of the alarms management procedures in operations, its existing limitations and introduces the cognition problem to be addressed by defining a representative UC. Additionally, a technical perspective of the cognition ecosystem used to address the UC is also presented.

Cognition-enabled alarm management

Alarm management is critical in today's operations to enable an adequate reaction and mitigation of potential network problems, therefore minimizing the service interruption and avoid impacting customer satisfaction. Currently, such issues are typically addressed by a set of rules that can be very simple/direct (e.g. one network event mapping directly to one alarm) or more complex, involving several correlation levels to produce the alarms. This last technique is known as root-cause analysis (RCA) and can use as its input several network events, consumed in different time instants, to produce a synthetic alarm - the root-cause. Additionally, the produced alarms can automatically open a new ticket on an external trouble ticket (TTK) management platform, by using predefined human rules to map the alarm into the TTK.

The existing alarm management solutions are stable and able to assist the operations in mitigating network issues (e.g. through remote actions or field force). Nevertheless, they are reactive - alarms are produced after the potentially faulty situation already occurred. A reactive alarmistic operation mode does not allow service providers to solve network problems in advance and minimize service interruption impact.

Moving towards a cognition-based alarm management paradigm is paramount [1]. It will enable evolving from a reactive to a proactive alarm management approach, in which artificial intelligence (AI)/ML technologies and algorithms are fundamental to anticipate problems and pave the road for optimizing service operations. Such is the cognition problem to be addressed in this particular UC – being able to predict a network alarm.

Figure 1 depicts the cognition problem to be addressed from a very-high and simplistic perspective, emphasizing the cognition added-value on the overall alarm management solution. As exposed, the cognition ecosystem:

- **a)** consumes all produced alarms (e.g. direct/simple alarms and/or synthetic/complex alarms);
- **b)** identifies alarms patterns through AI/ML;
- **c)** and outputs events predicting alarms.

From now on, for the sake of simplicity and easier document readability, we will refer to the predicted artefact as the predicted alarm.

The alarms produced by the Cognition Framework are consumed by the Alarm Management platform as a new predicted alarm, enhanced with the required additional information and added to the graphic user interface of the Alarm Management Platform.

Data source details

As described before, the identified cognition problem is being able to predict an alarm. To achieve this, at this stage, all the existing alarms are used as input data. Other data sources, such as key performance indicators (KPI), weather information, inventory and planned interventions, could as well be used to, potentially, improve the prediction results. The integration of these data sources in the alarm prediction UC is one of the next key steps.

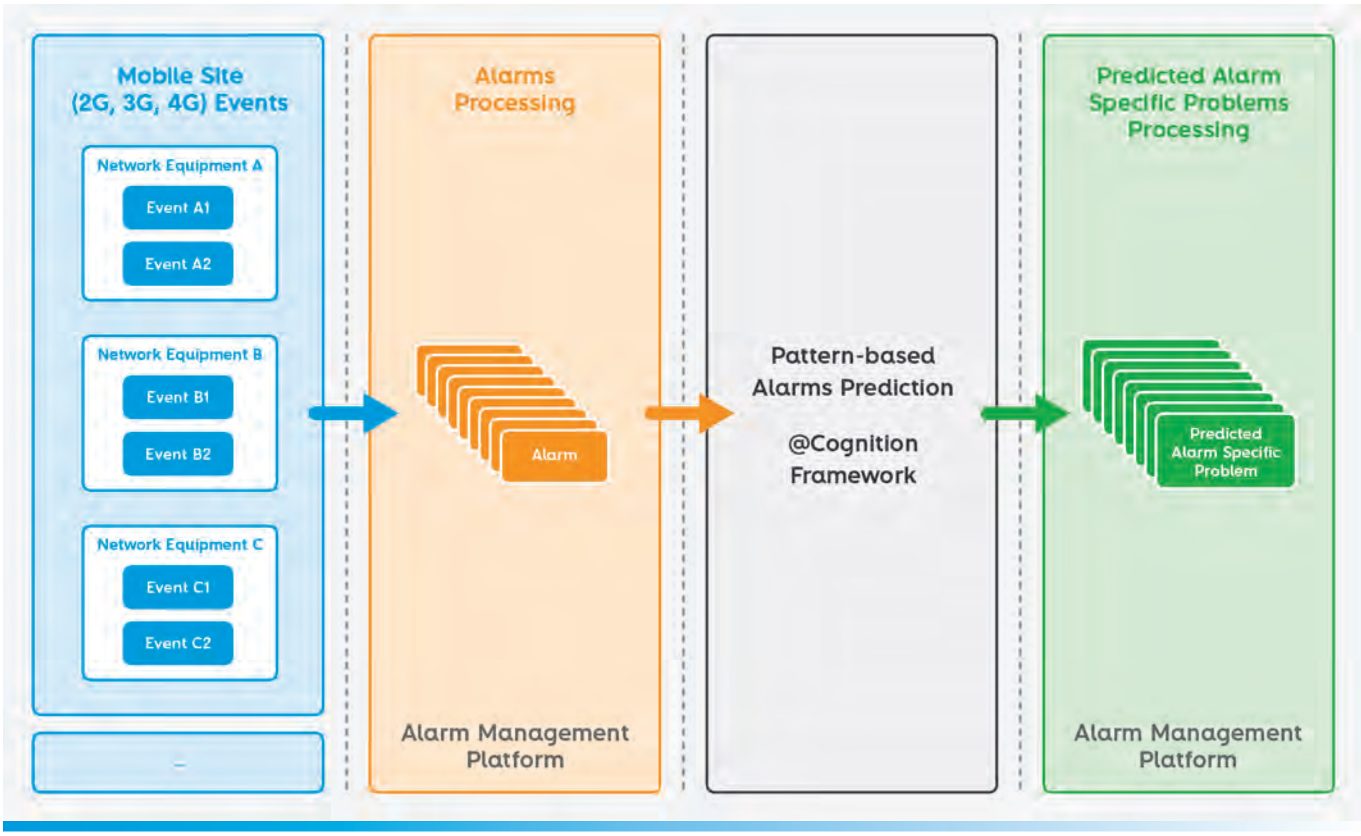


FIGURE 1 – Cognition-enabled alarm management

The provided alarms represent a very wide network footprint in terms of the supported technology domains. Specifically, used alarm data includes mobile and fixed access networks alarms, as well as backhaul and core alarmistic information. Besides the technology information, each alarm instance also includes the related inventory entry and the associated geographical area. The alarm specific problem, which details the detected anomaly, is added as well.

Figure 2 illustrates the above-described information. Besides spatial and domain information, alarms also include temporal information indicating all alarm instance state changes (e.g. start, update, end).

Cognition ecosystem

To address the needs raised by the predictive alarm management, a definition of an ecosystem of environments and processes was made (as illustrated in **Figure 3**). It is composed of the Alarm Management Platform's alarms, a Cognition Framework and an Execution Environment. The Cognition Framework includes the data management (the Data Acquisition & Transformation block) and models generation (in the Modeling block). The Execution Environment, on the other hand, hosts the deployment of the created AI/ML model for runtime/real-time alarms predictions (in the Cognition Model).

As shown in **Figure 3**, the cognition ecosystem must be able to accommodate the algorithms training and real-time predictions phases. Further details about the cognition ecosystem's logical architecture, as well as its operating mode during training and prediction, are covered in the next sections.

The results provided by the Execution Environment are wrapped in a message "Predicted Alarms", which contains the full details about each prediction. This metadata is used to add context for the systems consuming alarm predictions.

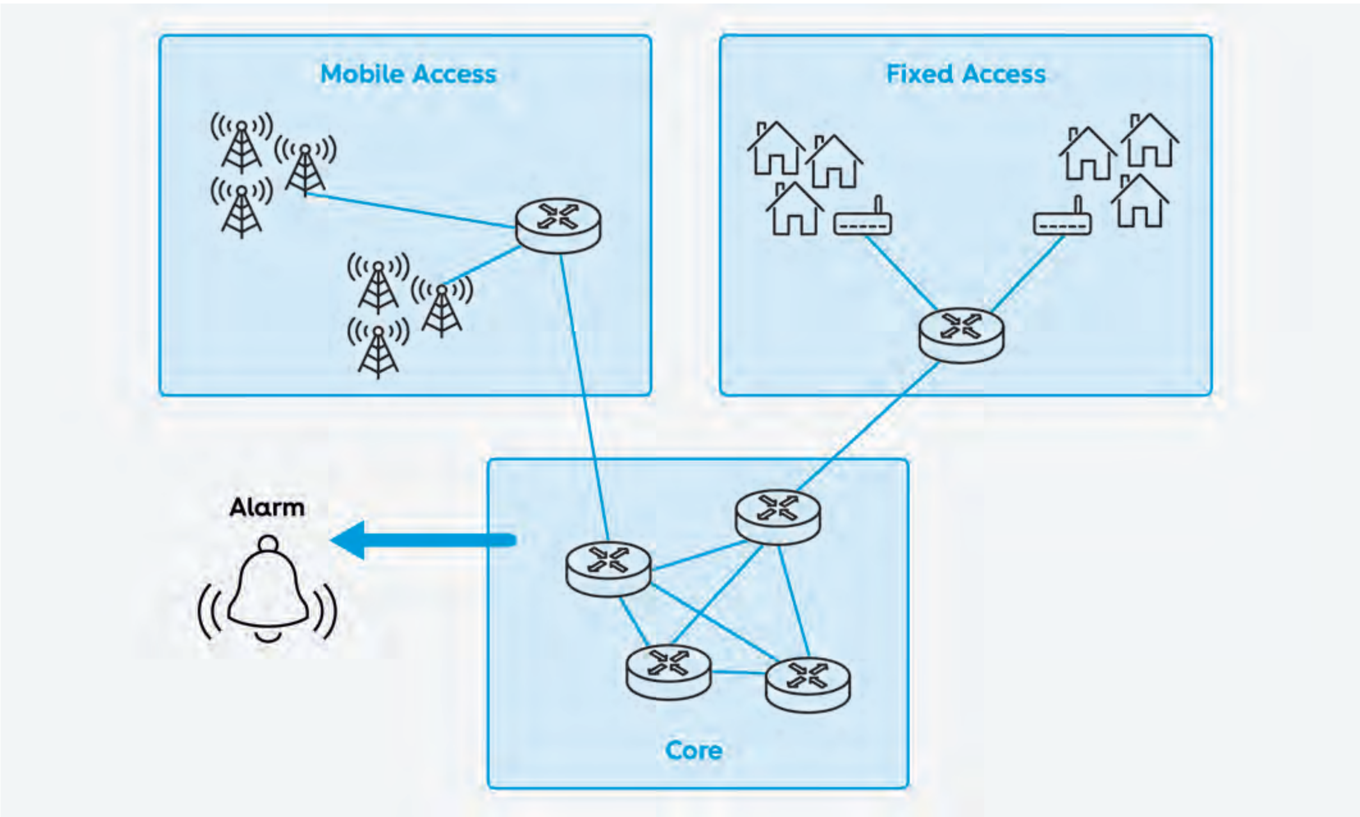


FIGURE 2 – Alarm characterization – spatial perspective

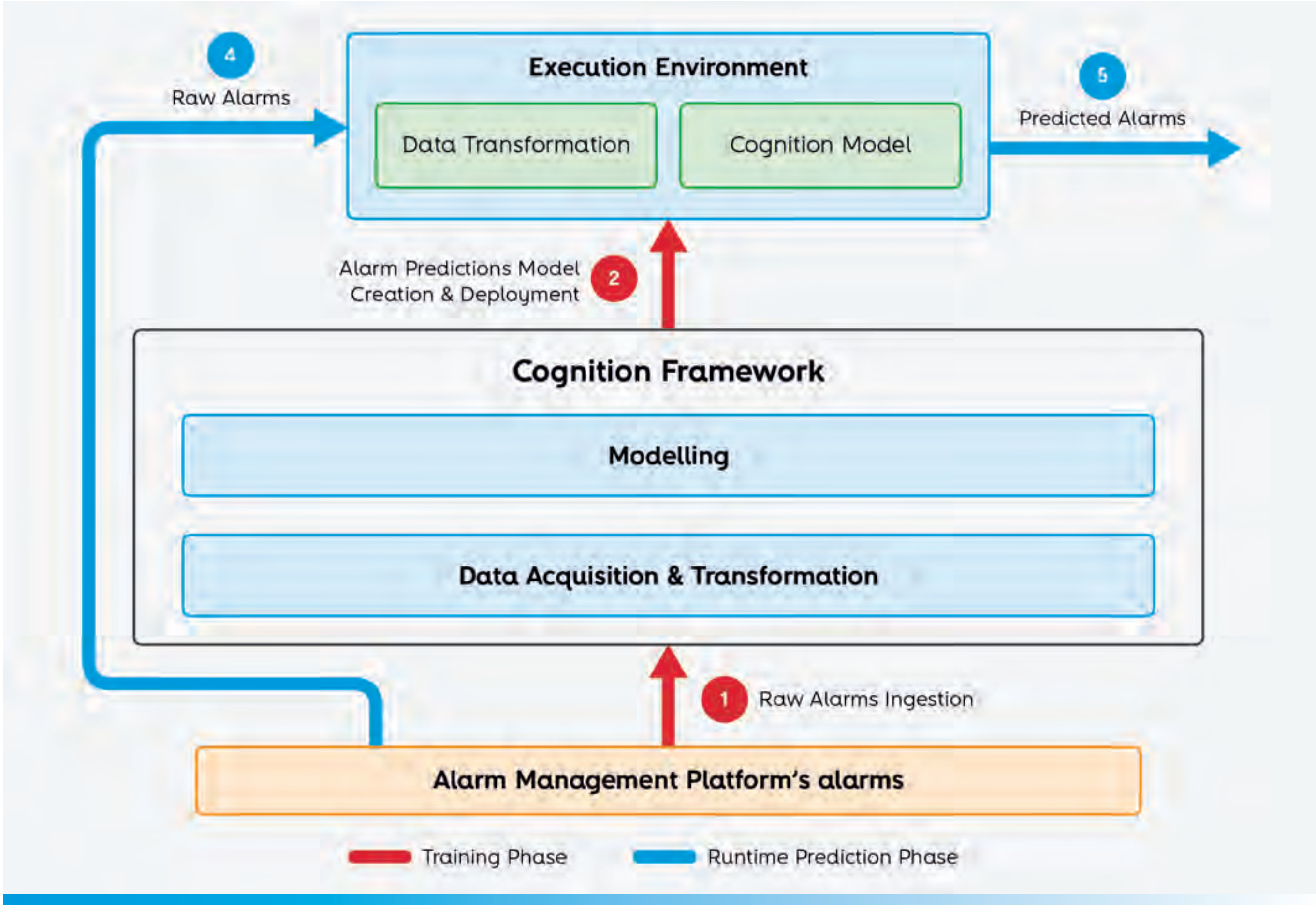


FIGURE 3 – Cognition ecosystem

The predicted alarms include the following two sections:

- **Prediction targets:**
  - Specific problems name list: a list of alarms being predicted for a particular context (defined by the prediction context, below).
- **Prediction context:**
  - **Prediction time**: the timestamp of the prediction made by the cognition model;
  - **Prediction location**: the code that identifies the geographical area for which the predictions are being made (i.e., where the reported prediction may occur);
  - **Active alarms list**: the list of the alarms that were active in the prediction location when the prediction was made;
  - **Occurrence probability list**: a list of the occurrence probability for each predicted alarm.

Data acquisition & transformation

The Data Acquisition and Transformation procedure is a stage within the Cognition Framework ecosystem, and consists of three tasks: ingestion, preparation, and transformation. This stage is responsible for obtaining, normalizing and consolidating the alarmistic data used in other stages.

Currently, the acquisition and transformation are divided into well-defined isolated steps, and it is intended to work in a fully automated pipeline after the iterations of data discovery and transformation are stable and defined.

Data ingestion, preparation & transformation

The data is ingested (via Apache Streamsets [2]) by consuming events published by the Alarm Management Platform in a Kafka bus [3]. After that, the retrieved alarms are stored in the data lake cluster, implemented through Hadoop [4].

Like many other ML projects, the data read from a source contains errors, redundant or useless data, and mislabeled fields. This stage processes the data according to a given configuration and prepares it for further processing. In our case, this includes, for example, discarding a set of event types used for alarm management control, and filtering fields needed for the data-wrangling stage.

In the alarm prediction UC, the data transformation phase converts alarms into a new representation. Data is modelled into network states, where it tracks the active alarm instances of a given set of problems occurring in a given location at a precise moment in time. Each network state results in a new contained and stateless snapshot capture on every change.

Snapshots characterization

Snapshots describe the reality within a given location. We use this data representation to try to assess if the current status is known to lead to a future alarm that should be predicted. The snapshot creation process is the following:

1. Track all changes by location: every alarm state change leads to a new capture;
2. A buffer mechanism keeps a timed window of the states so that future occurrences (labels/targets) can be matched;
3. The state is translated into ML features (e.g., encode categorical fields) where relevant data is created from statistical analysis of the data.

Figure 4 illustrates the snapshot generation process. In this representation, four alarms' state variations happen along the time: alarms 19, 7, 3, and the alarm targeted for prediction ULTRAN Cell Unavailable. Each one has different instance durations that overlap, and depending on the observation time, there are different active alarms.

Each snapshot (depicted by a camera icon) captures each change:

1. Alarm **19** opens, so the snapshot captures it as active;
2. A few moments later, alarm **7** also opens, meaning that the snapshot now has the information that two alarms are open;
3. In a third moment, alarm **3** also opens meaning that the snapshot now tracks three open alarms;
4. Finally, when alarm **19** closes it is consequently removed from the snapshot.

For each one of these moments, a label is also generated: how long until a future ULTRAN Cell Unavailable. This process builds a representation of the relationship between the alarm opening and closing events and a possible future targeted alarm.

Hence, each snapshot illustrates the relationship between each network state in a precise moment in time, and its relationship to future alarms. Snapshots are engineered into a set of ML features and labels, batched together and handed to the model for training. The labels contain information about alarms that occur after the provided features, thus requiring the use of historical data. The model then tries to generalize known observations as precisely as possible when confronted with new never-seen ones.

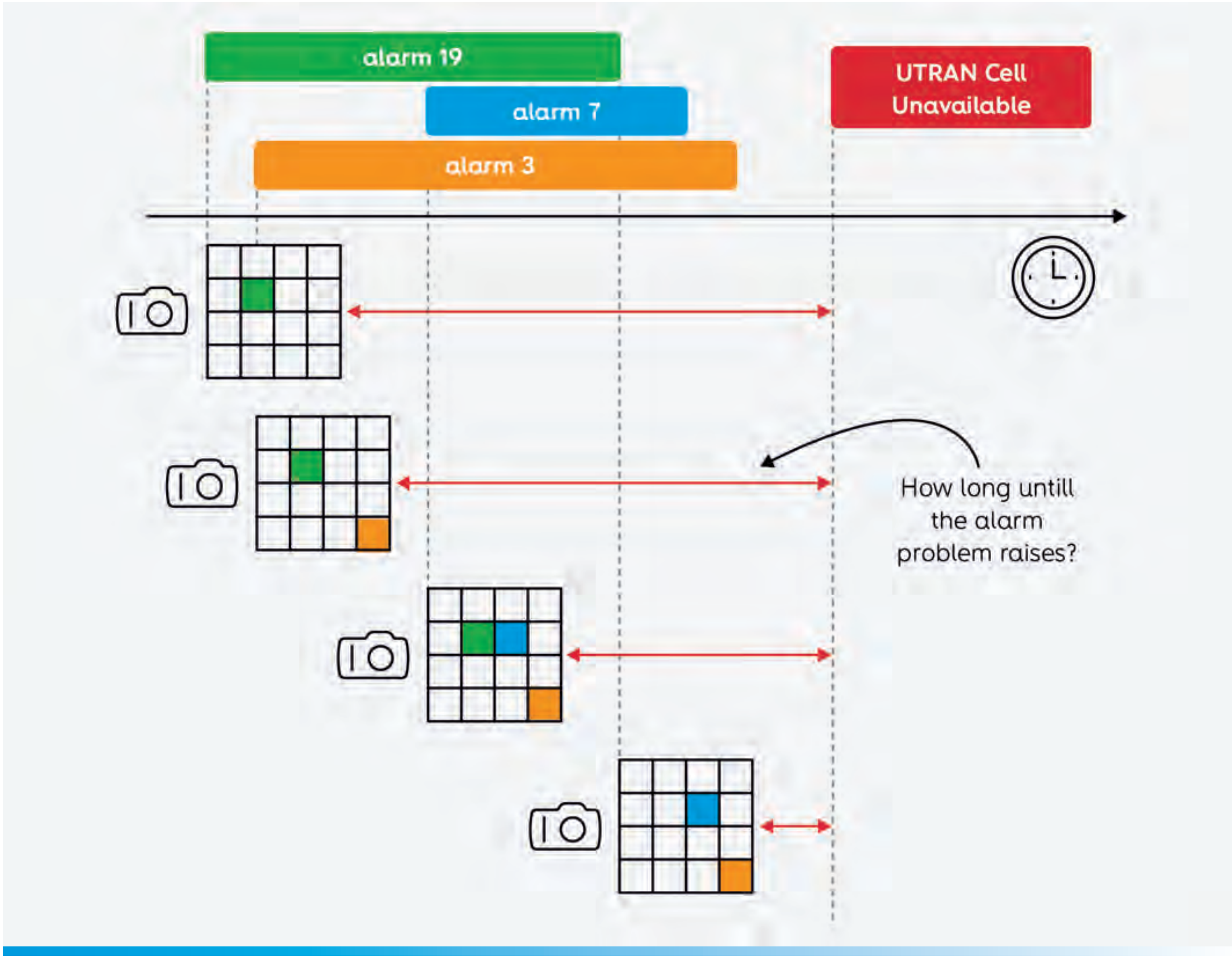


FIGURE 4 – Snapshots creation process

Modelling

Modelling is the second subtask in the Cognition Framework. This stage assumes the existence of a dataset that represents the knowledge to be learned, with as many examples as possible gathered from past observations so that it can be generalized to future alarms. The Modelling phase expects processed and clean data as a result of the Data Acquisition & Transformation step. The next step is translating snapshots into their numeric representation so that models can learn data patterns.

Feature engineering

We selected and identified a set of features that transforms the snapshot and its labels into a numeric matrix compatible with our selected ML framework: the open-source ML platform Tensorflow [5]. The feature list includes information assembled from the state:

1. If the alarm is statistically representative, then if it is active or not;
2. If active, for how long;
3. What time of the day was the snapshot captured;
4. What region is the snapshot from.

FIGURE 5 – Feature vector

The categorical fields are encoded according to the lookup tables obtained from statistical analysis of the data. **Figure 5** shows a final feature vector with one snapshot encoded in its numeric representation.

When all these snapshots are processed, then we have the dataset ready for digestion by the ML algorithms.

ML algorithm selection & configuration

This UC can be modelled in different types of ML architectures: binary, binary multi-label and multi-label multi-class. Each one has different ML restrictions, requirements and challenges. In the predictive fault management UC, it is being addressed in three incremental work stages:

1. **Binary decision**: we predict if one single alarm is going to happen within a timeframe, without any prediction of when it is going to occur;
2. **Binary multi-label problem**: we predict if a series of different alarms will happen within a timeframe, with the possibility of predicting more than one alarm for the same event;
3. **Multi-label regression problem**: Each prediction alarm also identifies, for each prediction, the estimated time to the occurrence.

As for ML models and frameworks, we started by following the most well-known and straightforward solutions with the advantage of faster development. However, the alarm data subtleness requires a different type of approach, and deep neural networks (DNN) [6] are, in our opinion, a suitable solution for this use-case for the following reasons:

1. **Data volume**: there are 2 million newly generated alarms per day coming from the network elements. DNN provide the required scalability to handle this volume of data both in performance and accuracy. There's an increased difficulty for this UC as we are dealing with an unbalanced dataset. This means that most of the data ingested by the algorithm lack the target alarms, as well as the fact of those occurring at different intervals from each other.
2. **Opacity is acceptable**: DNN are hard to explicit the reasoning behind each decision, but in the current UC objectives, there is no need to decipher what factors are contributing more for the prediction of a failure

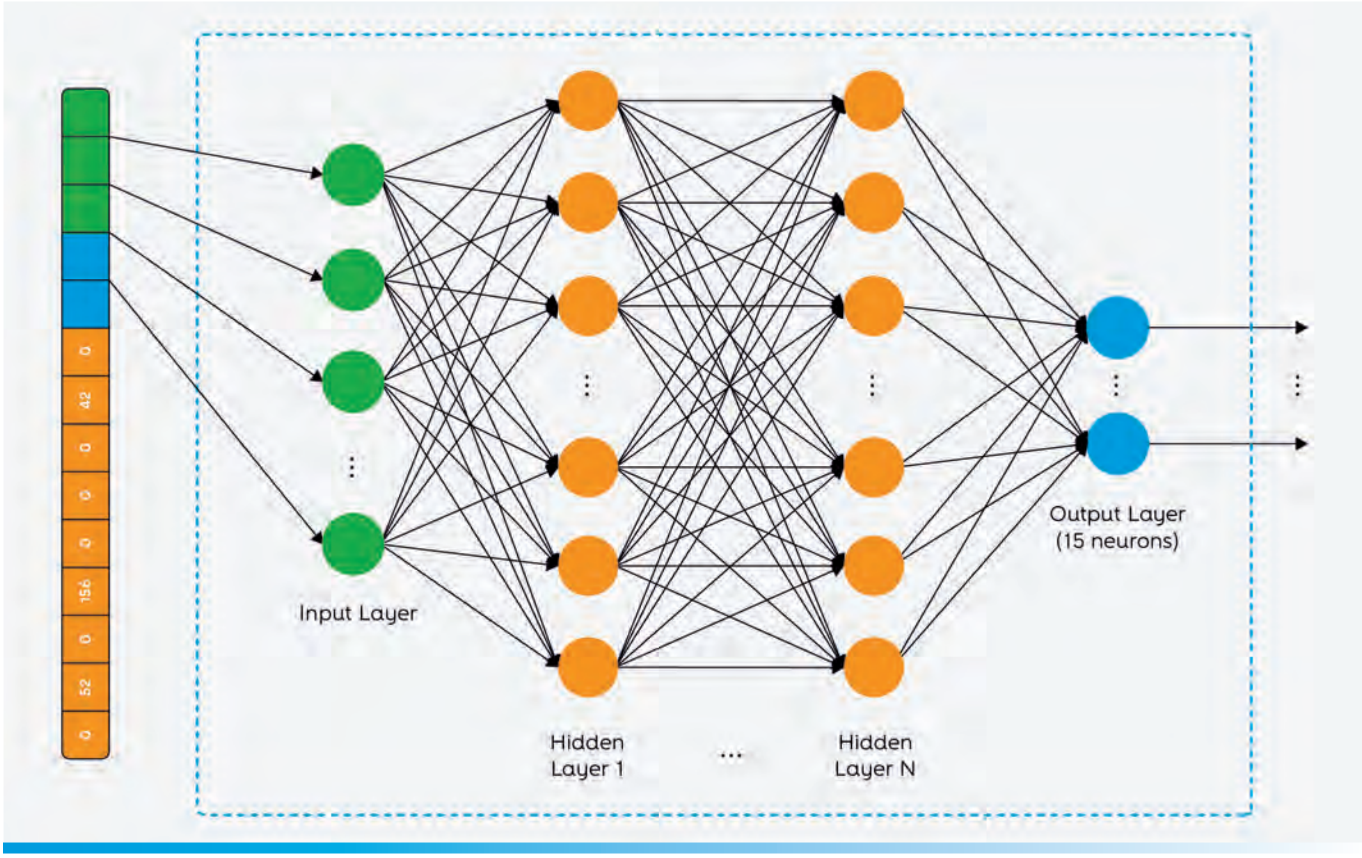


FIGURE 6 – Neural network layers and characterization

The UC is currently implemented in pure TensorFlow [7], with five fully connected hidden layers. **Figure 6** maps the representation between the work done in the feature engineering stage and the model architecture.

Deployment

This section highlights the AI/ML model Deployment related aspects.

(Real-time) data transformation

To allow the data to be consumed in such way to obtain predictions through a trained model, it is necessary to consume the raw alarms provided by the alarm management system and transform them in a similar way to that of the training data. It differs from the training data transformation as it does not require the labelling applied for the training. As such, the transformation process maps the alarms into snapshots (in real-time) by maintaining the status of the active alarms and providing snapshots according to the changes given by the alarms.

Cognition model artefacts

Regarding the cognition model, it is instantiated using a set of defined artefacts:

1. **Feature set transformation**: mappings from snapshot attributes to the feature vector ingested;
2. **Computation graph**: the actual neural network;
3. **Train variables**: the DNN weights optimized through training;
4. **Configuration parameters**: parameters used in the DNN.

Real-time prediction process overview

Herein are briefly explained the major steps executed during the real-time prediction process:

1. Alarms are published by the alarm management data source in the Kafka bus;
2. The published alarms are consumed by the Data Transformation artefact deployed in the Execution Environment, which creates the data snapshots and delivers them to the Cognition Model;
3. The Cognition Model, in which the generated prediction model is running, receives the snapshots and outputs the predictions to a Kafka bus. The alarm management system consumes the alarm predictions, enriches them, and, as a result, opens a new alarm instance: the predicted alarm.

Real-time cognition model supervision

Upon deployment of the model in the run-time environment, and as its results – the predicted alarms – become a part of the daily business, it is necessary to monitor and keep track of the quality of those predictions. A carefully designed maintenance and monitoring strategy will help avoid keeping a model with results that drifted from the baseline metrics (for instance, due to changes in the underlying network infrastructure or some model overfitting) and producing sub-quality predictions over long periods.

The supervision process will consider the previously defined model validation metrics and, upon identification of drifts above some defined acceptable thresholds, will deploy a newly trained model. This new model will be more adjusted to identified problems - for instance, more adapted to network infrastructure evolving nature or reducing model overfitting. In case the newly trained models are unable to cope with the changes, the process will notify the system that human intervention is required to maintain the execution environment expected behaviour.

Future work

Altice Labs ran the models in a real-world experiment for one month with real-time alarm data coming from the Altice Portugal network. Even though Keras [8] and Tensorflow do not transparently support model pruning or optimizations for deployment, the ongoing experiments run with a relatively low footprint, making the model response fast enough for the volume observed in the production servers. We have chosen to deploy models tuned to be very silent by subsampling the predictions via a strict threshold in the output, which impacted the number of observable predictions.

While not unexpected, one of the most striking conclusions for obtaining good predictions is the need to add more data sources for a complete world representation, such as weather information (current and forecast), relevant spatial relationship of entities or/and better segmentation of related alarms.

Since the start of these experiments, novel and interesting methods and algorithms have been proven quite successful in tackling classification and regression problems [9]. These are notably different from the DNN used here and would require some research in future works to validate their integration in this or a similar UC.

As a final note, the training dataset is always growing from the previous three months of historical data used in this experiment. As more data is available, we believe it will help in the detection of alarm recurrences that happen less frequently as well as establish stronger confidence in the results obtained.

References

[1] J. Zhong, W. Guo, and Z. Wang, "Study on network failure prediction based on alarm logs," in *2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC)*, 2016, pp. 1–7.

[2] StreamSets, "StreamSets DataOps Platform," 2019. [Online]. Available: https://streamsets.com/.

[3] Apache, "Apache Kafka," 2019. [Online]. Available: https://kafka.apache.org/.

[4] Cloudera, "Apache Hadoop Ecosystem," 2019. [Online]. Available: https://www.cloudera.com/products/open-source/apache-hadoop.html.

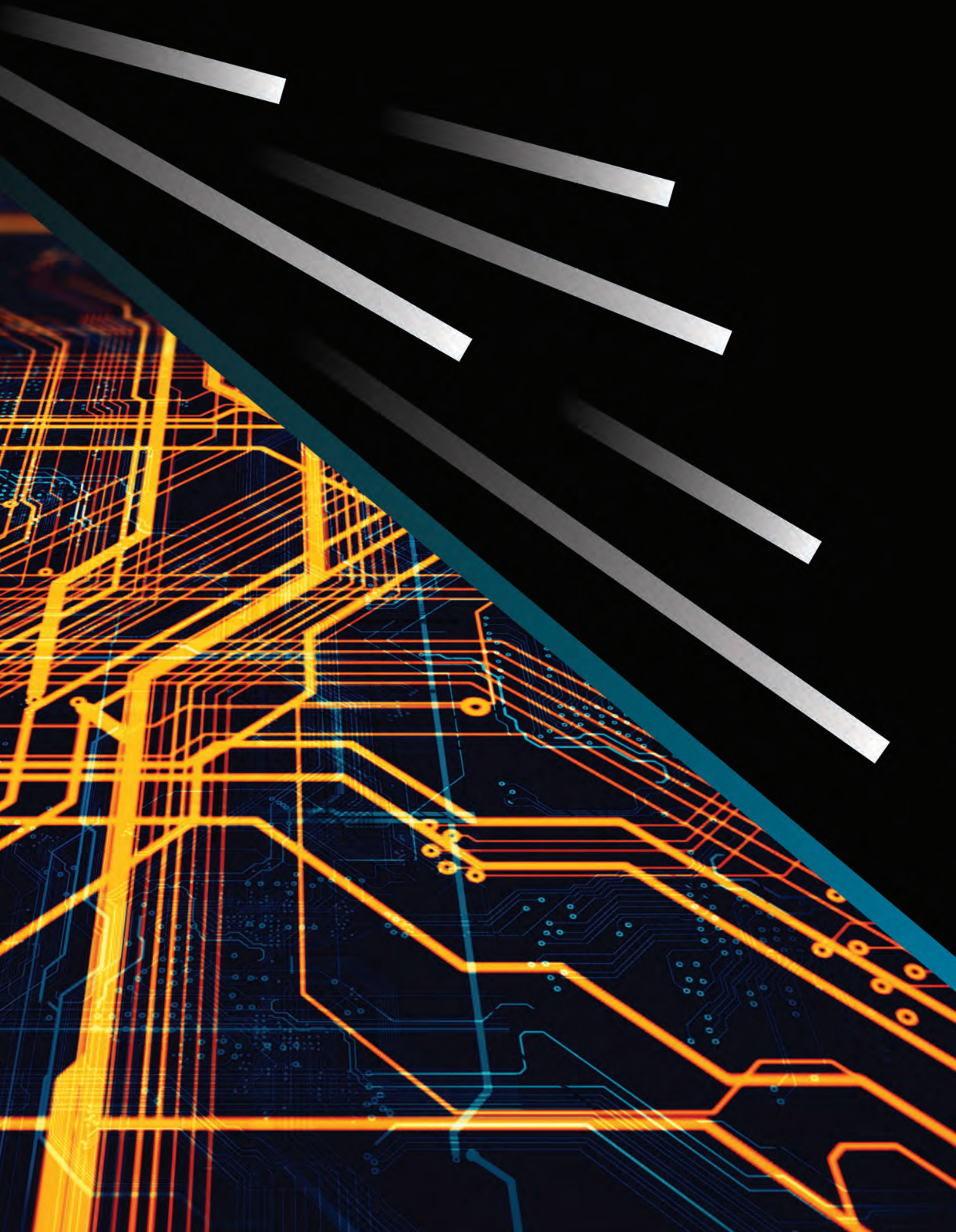
[5] F. Ertam and G. Aydın, "Data classification with deep learning using Tensorflow," in *2017 International Conference on Computer Science and Engineering (UBMK)*, 2017, pp. 755–758.

[6] H. Yi, S. Shiyu, D. Xiu-sheng, and C. Zhi-gang, "A study on Deep Neural Networks framework," 2016 IEEE Adv. Inf. Manag. Commun. Electron. Autom. Control Conf., pp. 1519–1522, 2016.

[7] "TensorFlow," 2019. [Online]. Available: https://www.tensorflow.org/.

[8] "Keras," 2019. [Online]. Available: https://keras.io/.

[9] F. Pereira et al., "Feature-Based Time Series Classification for Service Request Opening Prediction in the Telecom Industry," in Progress in Artificial Intelligence, 2019, pp. 120–132.



04 Root cause analysis of reduced accessibility in cellular networks

Cristina João Pires, Altice Labs
cristina-j-pires@alticelabs.com

Diogo Ferreira, Universidade de Aveiro
diogodanielsoaresferreira@ua.pt

Luís Cortesão, Altice Labs
luis-m-cortesao@alticelabs.com

Rui Filipe Pedro, Altice Labs
rui-d-pedro@alticelabs.com

Susana Sargento, Universidade de Aveiro
susana@ua.pt

Nowadays, the increased programmability of communication networks is driving such networks to become autonomous, able to provide fast actuation in response to users and networks' events. However, to achieve autonomous operations, the difficult task of understanding the root cause of network problems must be tackled. Establishing the cause-effect relation between network events and their origins provides network operators with the capability to identify, predict and mitigate these situations, by proactively manage network resources.

Keywords

Cellular networks; Root-cause analysis; Feature importance; Network correlated KPI

Introduction

One of the most important and critical operational processes of service providers is related to compliance with the service level agreements (SLA) established with their customers. In the ongoing digital transformation process, service provider assets are increasingly exposed to their customers, making SLA more relevant instruments whose non-compliance may seriously compromise the business relationship.

In this business relationship context, to ensure that SLA are not breached, service providers need to continuously monitor and evaluate the performance of all their assets (network infrastructure, services, etc.). This monitoring and evaluation can be done by using, among other tools, key performance indicators (KPI) built on information obtained from the network, allowing for the continuous evaluation of network and service behaviour.

Currently, rather than actively monitoring network/service performance through KPI, the aim is to anticipate, as far as possible, performance degradation and, consequently, failures that may result. Through predictive performance degradation, it is possible to anticipate and mitigate the underlying impact on customer services by applying actuation strategies (also known as tactics or policies) over the network.

In network management, root cause analysis (RCA) is an essential part of quickly solving network failures. Despite this, for more abstract metrics such as reduced accessibility, it is hard to understand their causes, preventing its premature detection. However, RCA processes enable the network operator to understand what causes originated a particular problem, allowing it to find where to actuate in the network to mitigate that specific problem.

With the increased requirements proposed for the 5G networks (e.g., 1 000 000 devices per km^2, 20Gbit/s of download peak data rate), the new generation of cellular networks promises to handle more traffic than ever before. The incorporation of network slicing, as well as software-defined networking (SDN) and network function virtualization (NFV) in the 5G architecture, overly increases the management complexity of those networks. With so many metrics to monitor, it is becoming harder to detect the cause of an event due to the complex combinations of various KPI.

Traditional software for detecting root cause analysis, with a knowledge base and a set of rules, is becoming obsolete due to the increased flexibility of the network. With the advances in machine learning (ML), it is easier to indirectly analyze dependent variables with reduced complexity, but with increased uncertainty.

In this work, we will evaluate the KPI related to reduced accessibility in cellular networks, using ML techniques. Identifying the most correlated KPI will allow more proactive management of the network, enabling prediction of an eventual future drop in network accessibility and identifying the more adequate actions to prevent it (e.g., adjusting the resources that have the most impact on those KPI). We will present and discuss two different approaches for root cause analysis using ML techniques. The first one will measure feature importance using model weights to determine the importance of each KPI in a reduced accessibility event. In a second approach, we will also analyze the KPI originating reduced accessibility per cell, enabling a more fine-grained network monitoring.

Problem statement

The objective of this work is to understand what are the most critical KPI to anticipate reduced accessibility in a cellular network. A dataset composed of KPI from a 4G network will be used, but the problem addressed here extents itself to future 5G networks with higher complexity.

The evolved universal mobile telecommunications system terrestrial radio access network (E-UTRAN) is the air interface in an LTE cellular network, and the most common metric used to indicate low accessibility in the network is the number of E-UTRAN radio access bearer (E-RAB) network setup failures per hour. In a 4G network, E-RAB setup is a major KPI for accessibility. The E-RAB is a bearer that the user equipment (UE) needs to establish communications in the network. **Figure 1** shows the E-RAB setup phase - after the UE has established a connection with the E-UTRAN node B (eNB), it sets up a context with the mobility management entity (MME), to enable the UE to communicate and send data to the network.

Usually, there are more E-RAB setup attempts than successes, particularly when the network is congested since messages after the E-RAB setup attempt are generally lost due to network problems.

To provide a better measure of low accessibility in the network, a new accessibility metric is used: the number of E-RAB setup failures. The network has congestion if the number of E-RAB setup failures is high.

The E-RAB setup failure daily distribution is presented in **Figure 2**. The most significant congestion occurs on day 17. The objective is to understand which KPI most contributed to this specific metric.

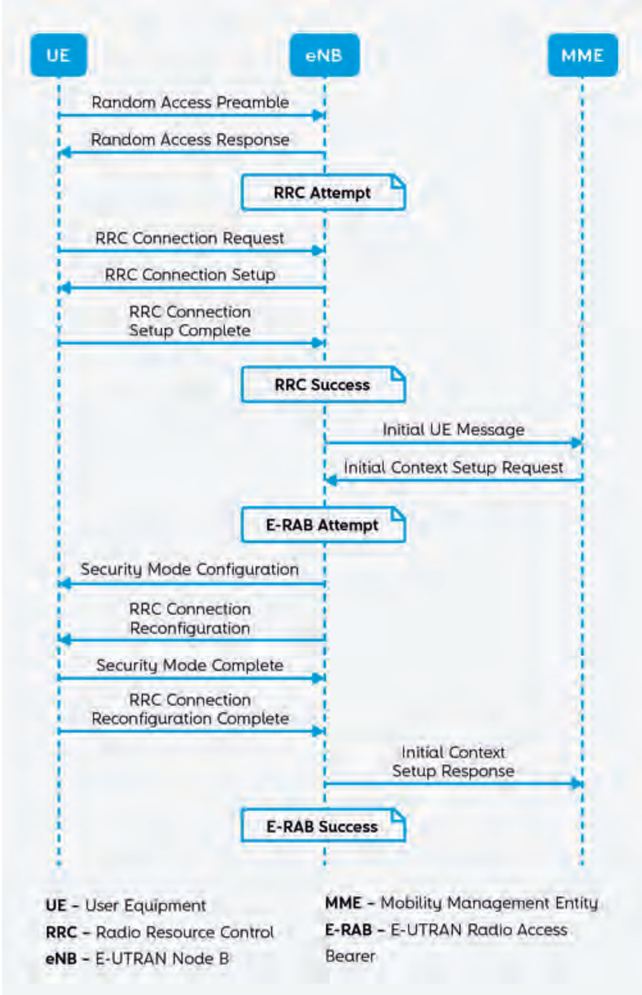


FIGURE 1 – Sequence diagram: E-RAB setup phase

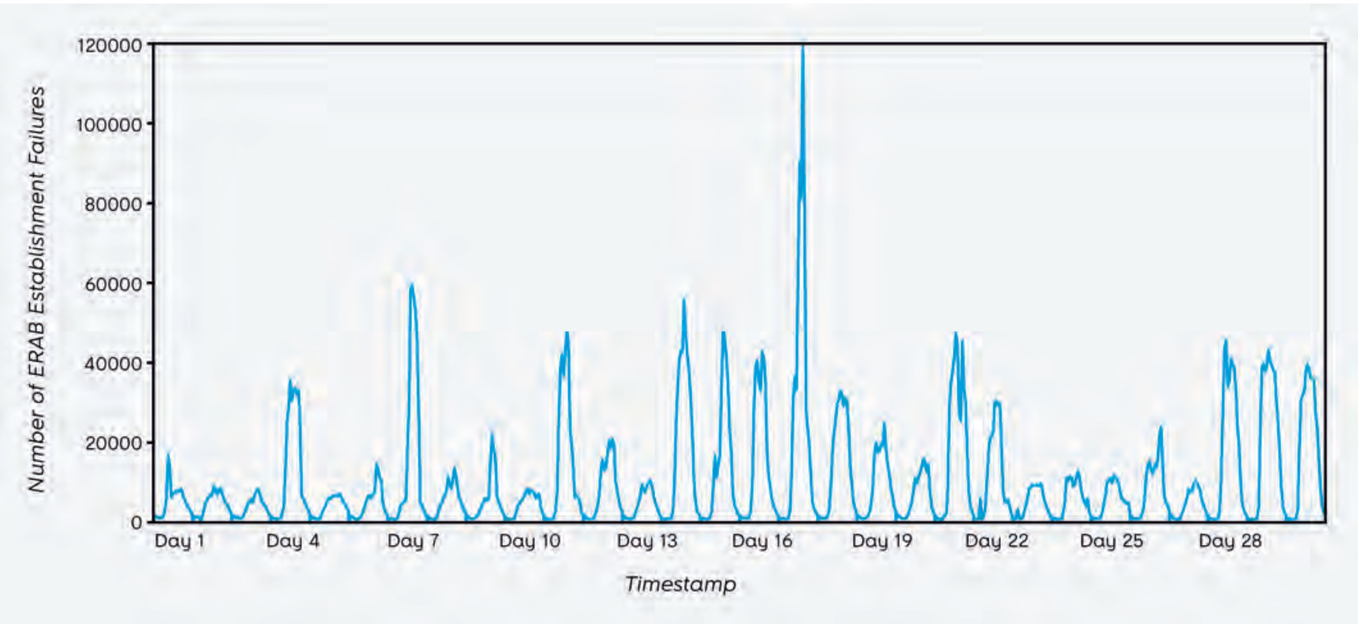


FIGURE 2 – Number of E-RAB setup failures

Methodology

This section covers the practical details regarding this work's methodology: how can reduced accessibility root cause be traced to, ultimately, be able to understand which cellular network KPI affect accessibility the most.

Taking a look into the RCA topic, it is relevant to comprehend that there are several techniques which address this analysis process in different ways, from more fundamentally mathematical solutions, like Bayesian Networks, to more recent methods that rely on ML.

Bayesian Networks have the capability to depict network metrics and events in nodes [1], with their relations represented as dependencies, alongside with their respective conditional probability, enabling the network operator to perform probabilistic inference methods to obtain the most probable cause of a network fault. Moreover, this method can be integrated with other techniques, as seen in [2], where a generic framework for large IP networks RCA was proposed. To determine the root cause of events, two reasoning engines were included: Bayesian inference and rule-based reasoning. The authors discuss that rule-based logic is often preferred over Bayesian inference because it is easier to configure, it provides an easier interpretation of results, and it is effective in most applications. However, Bayesian networks are preferred when the root cause is unobservable (no direct evidence can be collected), a characteristic present in our work.

In [3], the authors argue that a Bayesian network is not suitable for large-scale systems with a large number of components, because the complexity of inference exponentially increases with the number of nodes and dependencies between them. To solve that, they combine the Bayesian network with case-based reasoning techniques to prune the nodes needed to analyze in the network. The results show that the technique used drastically reduces the inference time, as well as the need for human intervention.

The work presented in [4] uses the concept of "variable importance" ("feature importance" in our work) to measure how much a feature contributes to predicting an objective variable on a ML model. This technique was adopted in our work and is described later, in section Defining the input and the output of the algorithms.

The remainder of this section will cover important aspects related to the usage of a ML-based root cause analysis for our particular case.

Measuring feature importance

As mentioned in the Problem statement section, the use case presented in this document focuses on RCA regarding reduced accessibility on cellular networks, particularly on 4G networks, which is measured by taking into account the number of E-RAB setup failures. This KPI plays a major role in detecting network access downtime; however, it may not be the only relevant KPI in the equation. Cellular networks display enough complexity, due to the high number of heterogeneous network elements: when we consider the vast amount of data they produce, it becomes imperative to consider the impact of other KPI when addressing this type of analysis. Having this in mind, the work presented here showcases the process in which we measure the KPI importance relative to the problem to solve. We call this exercise, measuring feature importance.

Since the goal is to understand which KPI can forecast low accessibility before there is an increase of failed network connections, the KPI values will be shifted (lagged) by one hour into the past (values are sampled hourly and one hour is the minimum time interval). Moreover, combinations of KPI that can be important must be taken into account, since we cannot assume that the KPI are independent of each other, especially when there are non-linear correlations between KPI.

Such considerations can be addressed by using ML techniques, which can be divided into two main categories: the ones that take into account the error of the model in a test set (to calculate the importance of the input features), or the ones that consider the algorithms internal weights associated to each input feature.

Take into account the model error

One of the approaches that fall into the first category is the "leave one out" method. In this approach, the importance of a feature is measured by comparing the test error of a model when all features are available as input, with the test error of a model when one feature is dropped for training. The higher the error for the model with one feature dropped, the more importance is given to that feature.

A significant disadvantage of this approach is that a separate training session is needed for every available feature, which causes the approach to be inefficient for datasets with a high number of features, or for models that take considerable time training.

An advantage of this method is that it is possible to apply it to black-box models [5], given that the feature importance is measured by the model error. It also takes into account all interactions between features, an advantage when compared with correlation tests.

Measure feature importance using model internal weights

Techniques that measure the feature importance by inspecting the internals of the models are algorithm-dependent. For some algorithms, like neural networks or Support Vector Machines (SVM), it is impossible to calculate the importance of each feature, due to the non-linear transformations applied. However, for other algorithms such as Logistic Regression, Extra Trees, Random Forest, Gradient Boosting or AdaBoost, it is possible to estimate the importance of each feature.

With this approach, it is essential that the features are all normalized within the same scale, and it is recommended that they are from the same type (continuous/categorical) for better importance estimation.

The advantage of these approaches is that they depend on the model, rather than on the test set. If the model is accurate and it is not showing signs of underfitting or overfitting, the feature importance can be calculated in a more reliable way, when compared to previous methods. Otherwise, the feature importance will be highly biased and will not be able to represent reality accurately. The biggest challenge in this approach type is to create accurate models that do not overfit the train set.

Despite this, since the dataset is complex enough to use approaches like Bayesian networks, the strategy to measure the importance of KPI in this work will be based on ML techniques.

Feature selection

An important step when building a model for training is selecting the relevant features from the dataset, which in this case are comprised of several distinct KPI. This selection process is necessary since some features are uncorrelated and do not contribute to the output classification, thus degrading the model performance. The ideal scenario is to test all feature combinations and determine the best ones. However, it is not feasible to perform this task due to the high number of combinations. Instead, the approach chosen was to use a dimensionality reduction algorithm to shorten the features number, which in this case was the Principal Component Analysis (PCA) [6] algorithm. With the chosen ML approach for the feature importance measurement and this dimensionality reduction algorithm, we are now able to train ML models and determine the relevant KPI for reduced accessibility forecasting.

Regarding the ML algorithms, we choose five to train the model and calculate the feature importance: Logistic Regression, Extra Trees, Random Forest, Gradient Boosting and AdaBoost. Each model provides a different set of results about the feature importance. For each KPI, the feature importance is calculated by taking into account the PCA results, thus achieving the KPI importance for each component.

To achieve the best model, the number of PCA components cannot be too small, otherwise relevant features can be lost; on the other hand, it cannot be too big, since it increases the risk of creating overfitted models. The number of components tested varies from one to the number of KPI, for the five algorithms. The model with a lower test error will be used to calculate the KPI importance.

Defining the input and the output of the algorithms

Naturally, the input values for model training are based on KPI, which are divided into two value types: normalized values and normalized variations. For each KPI, the normalized value of the previous hour will be used as an input, known as "lag". The number of lags is set to one and could be increased besides one hour; however, in this test, we considered that the low accessibility indicators appear at most one hour before network congestion. Similarly, for each KPI, a normalized variation is calculated, taking into account the ratio of variation between the normalized values of the current hour and correspondent lag value, to measure sudden variations on KPI.

Calculating the KPI importance reveals two classification problems that, from a business point of view, are different questions to what regards the accessibility problem: which KPI are most important when forecasting the possibility of a low accessibility event in the network, and which KPI are most important when forecasting sudden increases and decreases in network accessibility. For both cases, the data was transformed by using the 90th percentile has a threshold, values above this percentile become one, and zero otherwise, turning these cases into binary classification problems. By following this strategy, it is possible to analyze which KPI are most important for classifying low accessibility events, and also for forecasting bigger increases and decreases in data, which are important for resource allocation.

The two classification problems are tackled in different ways, according to the data split strategy. The tests are done with aggregated data (cells aggregated by region), network KPI are considered, and the notion of a complete network is taken into account. This enables us to forecast low accessibility of the network as a whole. Another way of performing these tests is to split the data per network cell. With this approach, the data is not aggregated, and 75% of it is considered for training and 25% for testing. With this in mind, the model built is capable of detecting low accessibility for each individual cell. Being able to forecast low accessibility per cell has a direct impact on how an operator manages its network. As an example, it allows the operator to issue the installation of temporary cells proactively, or change the current cells resource allocation strategy, to account for future low accessibility in certain geographical regions. Besides mitigating future lack of accessibility, this forecasting capability can be used to analyze broader regions and possibly conclude that certain geographical zones might need to see their cellular networks expanded.

Performance metric

For both classification problems, a metric must be used to measure the models' performance. Since both problems have a dataset where the number of positive samples is lower than the negative ones, it is necessary to choose a metric that considers false negative and false positive errors, while also considering the fact that the dataset is imbalanced. To address this issue, we have chosen the F1-score, which is the harmonic mean of precision and recall metrics [7].

Root cause detection results

Aggregated network tests

As explained in the Methodology section, two scenarios were explored:

a) the most correlated KPI for predicting if the number of E-RAB setup failures is above a threshold;

b) the most correlated KPI for predicting if the number of E-RAB establishment failures has high variations.

The best model for scenario a) was with the Extra Trees algorithm, achieving an F1-score of 86.6%, with 23 PCA components. With a higher number of PCA components, the performance of the algorithms starts to deteriorate. **Table 1** shows the ten KPI that were considered more correlated in scenario a). The handover failures and Circuit Switched FallBack (CSFB) was considered the most correlated KPI with accessibility problems.

The best model for the scenario b) was achieved through the AdaBoost algorithm. With 24 PCA components, the model achieved an F1-score of 40.0%. The F1-score is lower than in the scenario a) because the task of predicting variations is harder than predicting if the value is above a threshold. The number of PCA components is almost the same as in the scenario a). However, in **Figure 3** it can be seen that the algorithms' F1-score is very unstable and that it is hard to build a model to, accurately, predict the variations of the number of E-RAB establishment failures for the aggregated network.

KPI	Correlation
Handover Interfrequency Failure	0.194
Handover Intrafrequency Failure	0.188
Circuit Switched FallBack (CSFB) Prep Success	0.181
Packet Data Convergence Protocol (PDCP) Download TX Time	0.169
Variation Handover Interfrequency Failure	0.150
Cell Availability	0.131
E-RAB Normal Release	0.124
Handover Intrafrequency Success	0.121
Handover intrafrequency Attempts	0.121
PDCP Upload Volume (Mb)	0.115

TABLE 1 – KPI with the highest correlation factor for scenario a) for the aggregated network

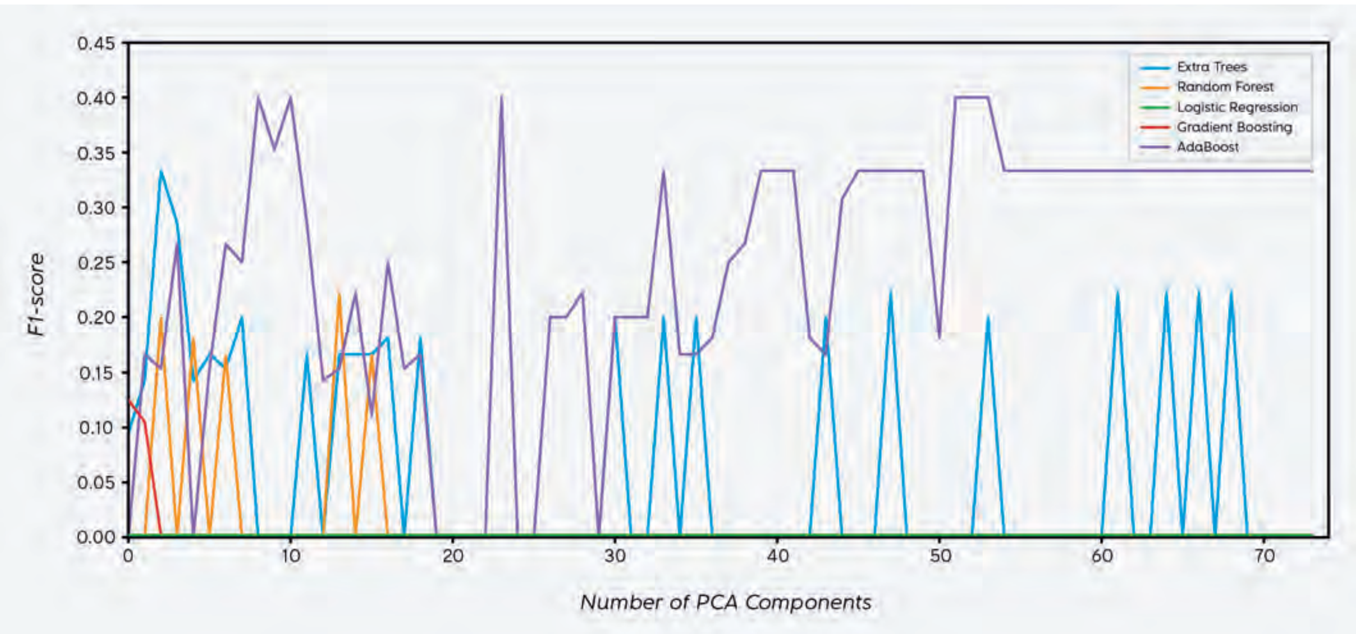


FIGURE 3 – F1-score with different algorithms varying the number of PCA components for the task b) for the aggregated network

Table 2 shows the ten KPI that were considered more correlated for the task b). The handover failures are still important, but they are not the most correlated KPI. The CSFB preparation success is the most correlated KPI to predict the variations in the number of E-RAB setup failures.

KPI	Correlation
Circuit Switched FallBack (CSFB) Prep Success	0.188
Handover Intrafrequency Failure	0.170
Packet Data Convergence Protocol (PDCP) Download TX Time	0.157
Handover Interfrequency Failure	0.148
Variation Handover Interfrequency Failure	0.141
Download Active Subscribers (Max)	0.138
Active Subscribers (Max)	0.125
RRC Setup Failure	0.121
Cell Availability	0.120
Upload Active Subscribers (Max)	0.119

TABLE 2 – KPI with the highest correlation factor for scenario b) for the aggregated network

The CSFB was the KPI that has more influence in availability, but the handover intra-frequencies and inter-frequencies failures also impact the availability of the aggregated network.

Individual cells tests

For individual cells test, two scenarios were analyzed:

a) the most correlated KPI for predicting if the number of E-RAB establishment failures is above a threshold;

b) the most correlated KPI for predicting if the number of E-RAB setup failures has high variations.

Using the Gradient Boosting algorithm, with 66 PCA components, the model for scenario a) achieved an F1-score of 30.79%. For cell prediction, more information is needed to obtain the best model when compared with the aggregated network. **Figure 4** depicts the F1-score varying with the number of components and the different algorithms, indicating that the F1-score increases when we consider more than 61 PCA components for all algorithms.

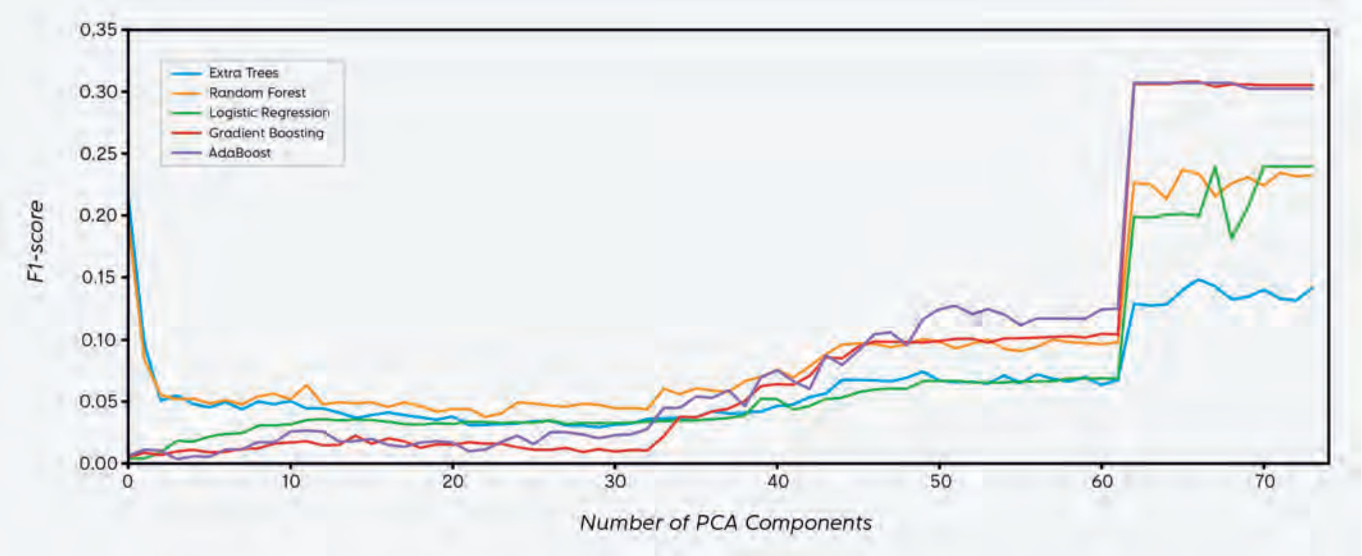


FIGURE 4 – F1-score with different algorithms varying the number of PCA components for the task a) for the individual cells tests

Table 3 shows the ten KPI that were considered more correlated for the task a). Two KPI have much more influence than all others: RRC Setup Success Rate and E-RAB Setup Success Rate.

KPI	Correlation
RRC Setup Success Rate	0.560
E-RAB Setup Success Rate	0.547
Connected Subscribers (Max)	0.069
Connected Subscribers (Avg)	0.066
Connected Active Subscribers (Avg)	0.043
Variation Connected Subscribers (Max)	0.034
Variation Connected Active Subscribers (Max)	0.032
E-RAB Normal Release	0.031
Variation Radio Bearers (Avg)	0.031
Variation Connected Subscribers (Avg)	0.031

TABLE 3 – KPI with the highest correlation factor for scenario a) for the individual cells tests

The best model for scenario b) has an F1-score of 20.39%, using the AdaBoost algorithm with 68 PCA components. The F1-score for different PCA components is similar to the scenario a), where the F1-score improved its performance significantly after 60 PCA components.

Table 4 shows the ten KPI that were considered more correlated for the scenario b). Like in the previous scenario, RRC Setup Success Rate and E-RAB Setup Success Rate are the most influence KPI for the output. In this task, other KPI have similar influence but with residual weight.

KPI	Correlation
E-RAB Setup Success Rate	0.147
RRC Setup Success Rate	0.174
Variation Cell Availability	0.138
Connected Subscribers (Avg)	0.102
Radio Bearers (Avg)	0.095
Variation RRC Setup Success Rate	0.093
Variation E-RAB Setup Success Rate	0.092
E-RAB Normal Release	0.088
Connected subscribers (Max)	0.087
Variation Radio Bearers (Avg)	0.081

TABLE 4 – KPI with the highest correlation factor for scenario b) for the individual cells tests

Discussion

Aggregated network tests

From the results in both tasks for the aggregated network tests, it can be concluded that most correlated KPI with E-RAB setup failures in a network to be above a threshold are the same that cause it to have high variations. Those KPI are the number of failure handovers (intra and inter-frequency), the CSFB preparation success (number of phone calls and SMS in the network), the PDCP download volume and the cell availability.

Interpreting the KPI, the results achieved are according to the intuition about lower network accessibility. When the number of failure handovers is high, the cells are crowded with user sessions and cannot accept any more sessions, which leads to lower network accessibility in the next hour. The high number of CSFB preparation success shows that there is a clear relationship between the high number of phone calls in the network with its lower accessibility. The KPI of PDCP download volume and the maximum number of active subscribers also show that the number of active subscribers and their download volume influence the network accessibility, more than the number of connected subscribers.

Finally, the cell availability indicates that, if many cells are unavailable in the current hour, it is likely that the network accessibility will be lower in the next hour.

Individual cells tests

The results of low network accessibility for individual cells had higher error than the results with the aggregated network. In the performed tests, the best models needed more data than the aggregated network models to achieve the best result, since more features were necessary to be able to generalize the predictions for different cells.

The results from both tasks show that, just like in the aggregated network tests, the most important KPI to predict the number of E-RAB setup failures in a cell to be above a threshold are the same that cause it to have high variations. However, the most correlated KPI that cause lower accessibility in a network are different from the KPI that cause lower accessibility in a cell. It is essential for a network operator to monitor the right KPI for the different tasks: prediction lower accessibility in a network or prediction accessibility per cell.

The two most important KPI for both tasks are the RRC and the E-RAB setup success rate. These KPI cannot be understood as the cause for lower network accessibility but as a consequence. Since they are intrinsically related to the E-RAB setup failure, being themselves accessibility metrics, they can be understood as an indicator of the high autocorrelation between consecutive hours. These results show that the network accessibility per cell is highly dependent on the network accessibility of that cell in the previous hour. If a cell has lower network accessibility in an hour, the network accessibility in the next hour for that cell will likely remain low, and vice-versa.

As opposed to the results for the aggregated network, the important KPI for low network accessibility are not related to the CSFB preparation success, or with any of the handover metrics. The results by cell show that, besides the RRC and the E-RAB setup success rate, the most correlated KPI have counters related to the number of users in a cell and its utilization. It is expected that, as these KPI have higher values, the accessibility of a cell decreases.

Conclusions

Understanding the causes of events in a network, such as low accessibility, helps the network operators to forecast and avoid them, by adjusting network resources that influence their causes. In this work, the goal was to determine the causes of reduced network accessibility in 4G networks, using only historical data.

Two different analysis were made. Besides analyzing the causes of reduced accessibility in the whole network, the causes of reduced accessibility for each cell were also analyzed. The results showed that the causes of reduced accessibility in each analysis are very different. For the overall network, the KPI that most influence the accessibility are:

- the number of failure handovers;
- the number of phone calls and SMS in the network;
- the overall download volume;
- the availability of the cells.

For each specific cell, the KPI that most influence the accessibility are related to the number of users in a cell and its download volume. A network operator needs to know if it is important to monitor low accessibility in a cell, in a network, or both, to make the right measurements and adjustments in the network.

As future work, the next step will be to detect the patterns of those KPI that indicate future low accessibility, to be able to predict it and adapt the network to prevent it from occurring. For example, if the network operator knows that the network accessibility will be lower in the next hour when the number of handover failures intra-frequency and the maximum connected users both exceed a threshold, he can take proactive measures to adapt the network and avoid the low accessibility.

The inclusion of these capabilities in Altice Labs' network performance analysis line of products is proving critical as, more than actively monitoring network/service performance through KPI, the aim is to anticipate, as far as possible, performance degradation and, consequently, failures that may result. This work will allow us to build the models to support this desideratum, as well as to develop the set of competencies need to apply them generically, to the multitude of existing network technologies.

References

[1] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1988.

[2] H. Yan, L. Breslau, Z. Ge, D. Massey, D. Pei, and J. Yates, "G-RCA: A Generic Root Cause Analysis Platform for Service Quality Management in Large IP Networks," *IEEE/ACM Trans. Netw.*, vol. 20, no. 6, pp. 1734–1747, 2012.

[3] L. Bennacer, Y. Amirat, A. Chibani, A. Mellouk, and L. Ciavaglia, "Self-Diagnosis Technique for Virtual Private Networks Combining Bayesian Networks and Case-Based Reasoning," *IEEE Trans. Autom. Sci. Eng.*, vol. 12, no. 1, pp. 354–366, 2015.

[4] J. M. N. Gonzalez, J. A. Jimenez, J. C. D. Lopez, and H. A. Parada, "Root Cause Analysis of Network Failures Using Machine Learning and Summarization Techniques," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 126–131, 2017.

[5] A. Fisher, C. Rudin, and F. Dominici, "All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously," no. Vi, 2018.

[6] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *London, Edinburgh, Dublin Philos. Mag. J. Sci.*, vol. 2, no. 11, pp. 559–572, 1901.

[7] K. P. Shung, "Accuracy, Precision, Recall or F1?," *Towards Data Science*, 2018. [Online]. Available: https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9.

05 Access network failure prediction powered by cognitive techniques: an overview

Gil Alves Serafim, Altice Labs
gil-a-serafim@alticelabs.com

Nuno Duarte, Altice Labs
nuno-p-duarte@alticelabs.com

Nuno Ricardo Farinha, Altice Labs
nuno-r-farinha@alticelabs.com

Pedro Luís Ferreira, Altice Labs
pedro-l-ferreira@alticelabs.com

Simão Oliveira Cardeal, Altice Labs
simao-o-cardeal@alticelabs.com

Nowadays, network devices provide additional monitoring data useful for operational teams' insights into network health. But humans are often overwhelmed by the amounts and complexity of data available! Access network, and particularly, passive optical networks and its network indicators and elements, are some of the areas where AI might be used to replace human analysis and prediction efficiently, aiming to anticipate failure, optimise and automate network operation, and overall improve network quality.

Keywords

xPON; Equipment optimisation; Artificial intelligence; Predictive operations

Introduction

Communications networks experience such a tremendous evolution in complexity that it is now virtually impossible to proactively keep them in optimal running conditions. Engineers in network operations centres always have reactive responses: networks are monitored and, once a failure is detected, actions are performed to eliminate the problem. This approach relies heavily on failure-prone human analysis, often resulting in a negative impact on customer experience and satisfaction.

Nowadays, network devices provide additional monitoring data. This data availability would allow those teams to perceive failure patterns and to some degree predict, based on their acquired experience, future failures. But such human insights are often overwhelmed by the amounts and complexity of data available. Therefore, the natural logic step is to lend some help from computing processing.

Cognitive computing allows us to approach the challenges mentioned above with new tools such as artificial intelligence (AI). The last decades have shown the evolution of both CPU and GPU processing power to a level where cognitive computing is now a feasible process to be executed on common computing frameworks, and no longer an exclusivity of supercomputers.

This article aims to present an area where AI might be used to efficiently replace human analysis and prediction in the domain of passive optical networks (PON).

We will identify, describe and analyse some of the network elements and variables available available to better understand how to use them. Indicators such as temperature, number of power cycles, received or transmitted optical power, among others, are essential to predict behaviours by comparing existing curves. These indicators are to be correlated with each other to produce algorithms capable of generating warning or alarm events or even to trigger autonomous network healing mechanisms.

We start by introducing GPON, followed by a description of the critical issues in this kind of networks. We will then describe each identified key indicator and point new areas of research and development.

xPON networks

Technology evolution

In the last decades, customers have seen a substantial increase in network bandwidth availability. The main reason for this achievement is the implementation of passive optical networks in the access network.

In 1995, a group of the world's leading telecommunications services providers, independent test labs, and equipment suppliers created the Full Service Access Network (FSAN) [1] to push the industry to produce a set of specification documents, particularly relating to the fiber-to-the-home architecture. In 1998 the FSAN specification was standardised by the International Telecommunications Union (ITU) that delivered the Recommendation for BPON optical interface specification as ITU-T G.983 [2]. Some years later the ITU-T G.984 gigabit-capable passive optical networks (GPON) [3] was delivered to improve the older ITU-T G.983, defining improvements on the total bandwidth, both downstream and upstream, and allowing variable packet length. The ITU has continued to improve his recommendation standards by delivering new documents. The ITU-T G.987 [4] that defines a 10-gigabit-capable asymmetric passive optical network (XG-PON), the ITU-T G.9807 [5] that describes a 10-gigabit-capable symmetric passive optical network (XGS-PON) and the ITU-T G.989 [6] that defines time and wave division modulation (TWDM-PON) also called NG-PON2. These standards are examples of the work made to improve the PON networks mainly by increasing the bandwidth availability. This evolution is represented in **Figure 1**.

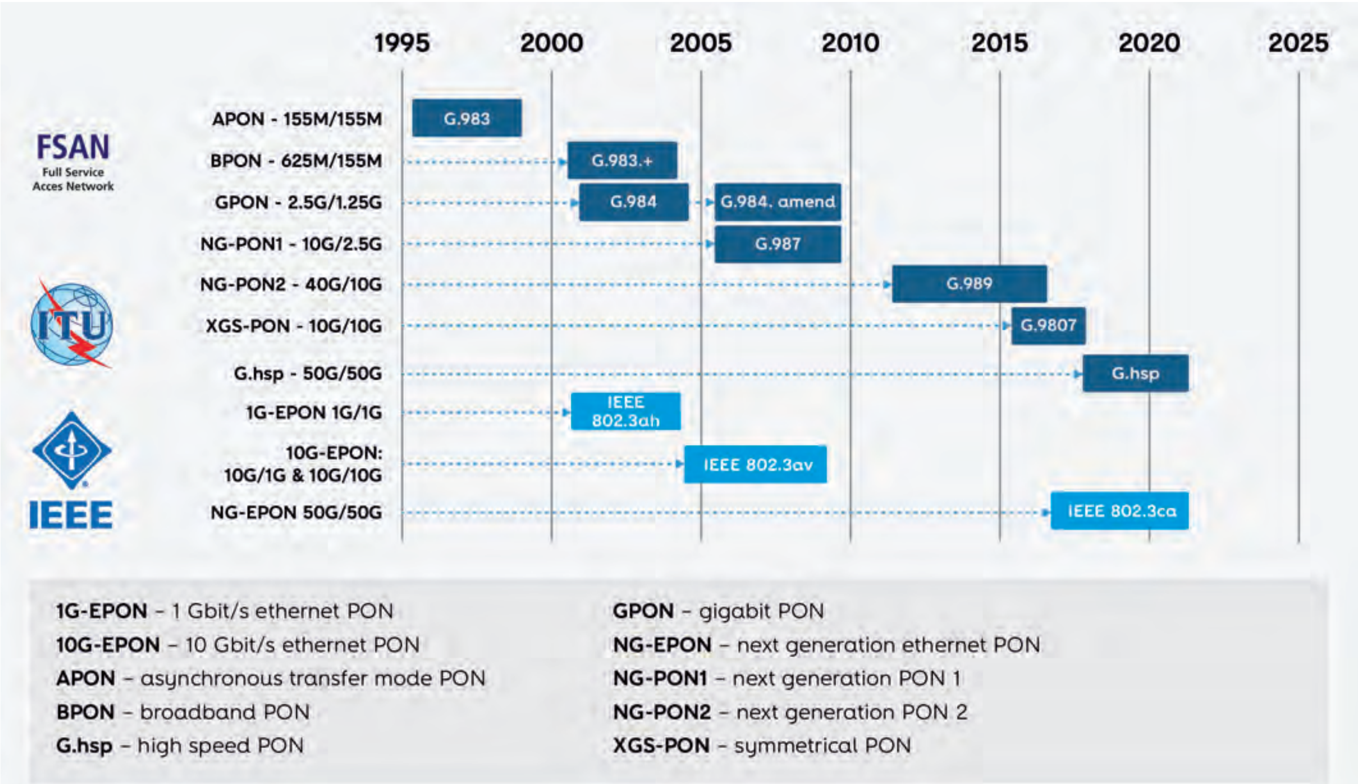


FIGURE 1 – PON technology evolution

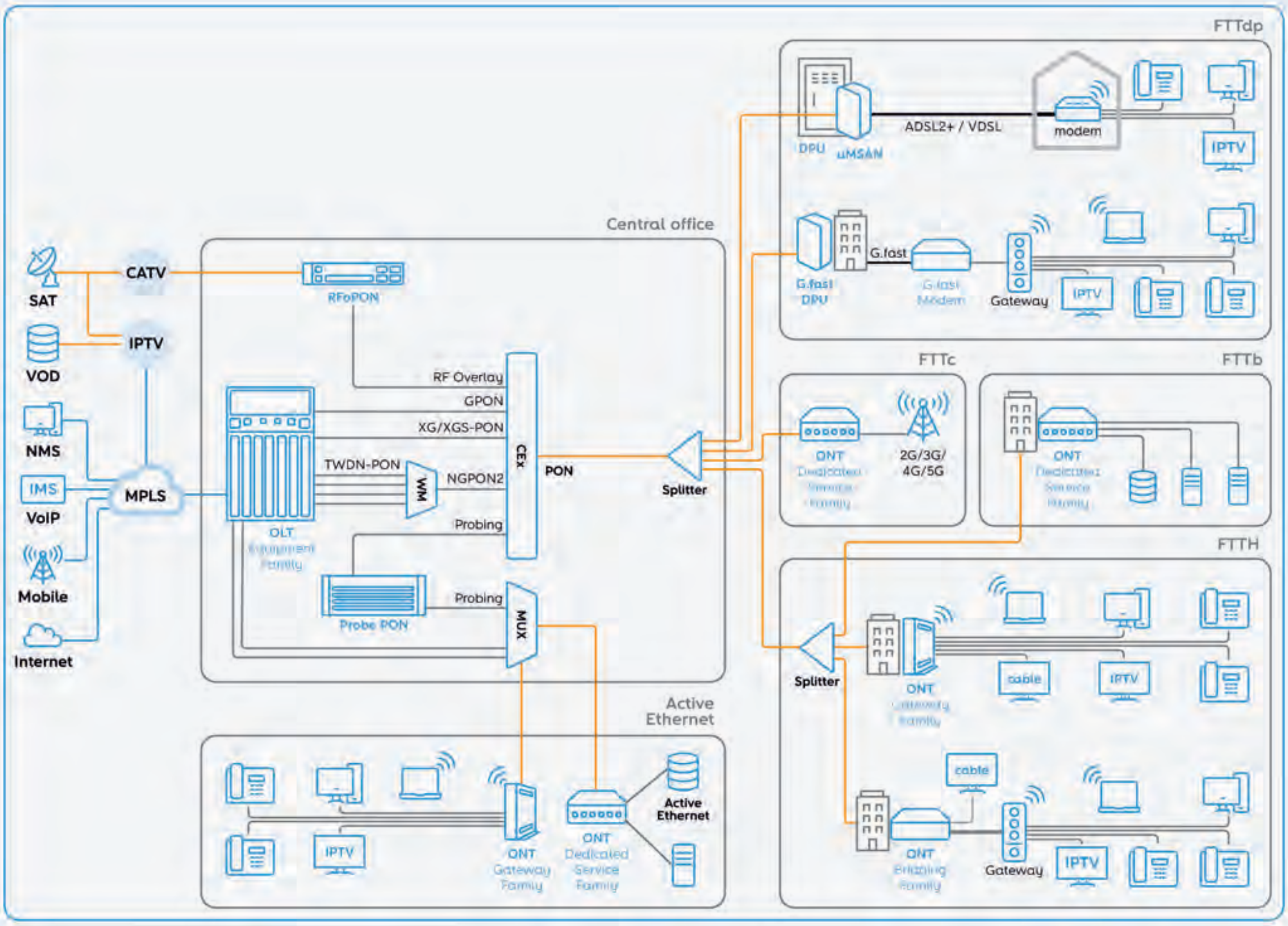
The architecture

xPON fiber-to-the-x (FTTx) network topologies, as represented in **Figure 2** (on next page), are based in optical power division along with a passive optical network (PON). Each one of the PON ports of the optical line termination (OLT) may be divided into a maximum of 128 optical network terminals (ONT), for a maximum length defined of 60Km, with a differential fiber distance between the farthest and the nearest optical network unit (ONU) of 20Km.

These types of equipment (OLT and ONT) are specially designed for fiber network infrastructures either point-to-multipoint (P2MP) or point-to-point (P2P) FTTx gigabit (or Ten Gigabit) xPON architectures. OLT reference P2MP FTTx network topology scenarios are fiber-to-the-home (FTTH), fiber-to-the-building (FTTB), fiber-to-the-cabinet (FTTC), fiber-to-the-cell (FTTc) and fiber-to-the-business (FTTb). Those are intended to solve most fiber access scenarios in terms of retail as well as wholesale client needs and are able to serve multiplay services, namely voice (VoIP), data (HSI), TV (IPTV and RF Overlay) of up to 128 clients per single PON port.

Data exchange

In PON networks all downstream traffic is broadcasted. However, in the upstream direction, each ONU is only allowed to transmit in a specific time managed by the OLT. For this reason, there are some specific mechanisms defined in these recommendations to respond to these requirements. The ONU discovery and ONU ranging mechanisms are two processes used by OLT to perform ONU registration. In the first, the OLT generates a quiet window to allow all not registered ONU in the PON to make his announce and let the OLT know about his existence in the PON. In the second, the OLT previously knows about ONU existence and gives a communication chance to a specific ONU to start his registration



ADSL2+ – asymmetric digital subscriber line for 2MHz
CATV – cable TV
CEx – coexistence element
DPU – distribution point unit
FTTb – fiber-to-the-business
FTTc – fiber-to-the-cell
FTTdp – fiber-to-the-distribution-point
FTTH – fiber-to-the-home
G.fast – ITU-T digital subscriber line protocol standard
GPON – gigabit passive optical network
IMS – IP multimedia subsystem
IPTV – internet protocol TV
MPLS – multiprotocol label switching
MUX – multiplexer
NGPON2 – next generation PON 2
NMS – network management system
OLT – optical line termination
ONT – optical network terminal
PON – passive optical network
RF Overlay – radio frequency overlay
RFoPON – radio frequency over PON
SAT - satellite
TWDM-PON - time and wavelength division multiplexed PON
uMSAN – micro multi-service access node
VDSL – very high speed digital subscriber line
VOD – video on demand
VoIP – voice over IP
WM – wavelength modulator
XG/XGS-PON – 10-gigabit-capable asymmetrical/ symmetrical PON

FIGURE 2 – PON network topology

process. Only after the end of the registration process, the OLT can configure the specific services on the ONU, accordingly to the service hired by the customer. This configuration is made by using the ONU management and control interface (OMCI) recommendation ITU-T G.988 [7]. Through the OMCI specification, the OLT can configure, manage and monitor performance indicators for optical access system operation.

At a physical level, the signals carried by the fiber between OLT and ONU are converted using optical transceivers. These transceivers have optical transmitters and receivers that operate at specific wavelengths. For example, a transceiver accordingly to the ITU-T G.984.2 [8] use a 1490nm continuous-mode transmitter and a 1310nm burst-mode receiver. Each transceiver has its own optical characteristics that the OLT should previously know. These characteristics are used by the OLT and by the ONU. This way, they know how the ONU should form the upstream bursts to meet the specific requirements of the transceiver used by the OLT.

Identified issues in xPON

Today network operators have ways to debug, identify and fix issues that are mostly triggered by incidents raised by clients or based on alarms. The equipment deployed is able to provide several key performance indicators that allow those operators to identify anomalies or failures.

Network issues

Rogue ONU

From what we mentioned before we can note that there are a large number of potential problems that can occur in GPON networks, mainly by being based in time-division multiplexing (TDM). For example, if the ONU is not transmitting in its time window, this could cause a set of alarms like a drift of window or transmission interference. On the worst-case scenario, we can have an ONU that transmits a burst out of its allocation opportunity. This type of behaviour is called a rogue ONU and can be of two types: one is when the ONU is transmitting over other ONU allocation; the other is when the out of time allocation is not overlapping any other allocation.

The rogue ONU issue is very hard to solve and is mainly caused by misbehaviour on the ONU transmission laser. For this reason, it is crucial to monitor laser performance metrics to control its correct operation function and to check if its lifetime is near the ageing point. Laser lifetime is affected by operating conditions, including injection current, optical output power, and temperature [9]. So it is theoretically possible to predict, or at least estimate, the nearing of its end of life by correlating ONU dying-gasp alarms, transmit power, laser current and operating temperature.

Through monitoring performance metrics from OLT and ONU, it can be possible to predict and mitigate potential transceiver failures.

By correlating the drift of window and transmission interference alarms, it is possible to predict if an ONU can potentially become rogue. If the drift increases over time, or if the compensation made by the OLT is not enough to set the ONU laser in the correct allocation window, it can be possible to expect that leaser may be in rogue behaviour.

Elements and variables to retain:
temperature; transmitted optical power; laser bias current; online time (transmitting)

Optical distribution network problems

The optical distribution network (ODN) is based on the physical elements that are needed to allow communication between OLT and ONU. It includes the optical fiber, splitters, fiber enclosures, ducts, etc. If some issue occurs with any of the ODN elements, it could cause many problems to the service providers. From wrong connections made by technicians, malfunction on passive components, and fiber cuts, there are a lot of places to look. The industry has developed systems potentially able to detect issues in ODN. This kind of probes can detect fiber cuts with high accuracy and allow the service providers to act faster if some of these issues happen.

Detect fiber cuts in P2MP is not an easy task because of its intrinsic architecture. For example, consider a basic ODN composed by an OLT PON port connected to a 1:8 splitter and 8 ONU. If a cut is made on one of the fibers connected to the ONU the probe will detect the cut after the splitter but cannot indicate which fiber is cut. With this example, we note that the probe, by itself, does not have full knowledge of the ODN. By correlating the fiber cut information from the probe with loss of signal alarms from the ONU, the system can detect where the cut occurred. The system can infer the severity of the problem by considering the number of clients affected by the cut and acting accordingly.

A cognitive network can predict the probability of a specific fiber to suffer a cut, through the interaction with other systems. By using data from both an inventory management system and a field intervention system, the network can calculate the probability of a particular optical fiber being cut. For example, we can postulate that it is less probable to cut a specific fiber in a duct than in aerial installation. However, if some intervention on the duct area is going to happen, that probability will increase.

Weather is another factor that can change the probability of a fiber being affected. Floods, strong winds and wildfires can create a lot of problems to ODN stability. For this reason, a cognitive network needs to be aware of these factors.

> Elements and variables to retain:
> *fiber cuts; power budget; infrastructure interventions*

Equipment issues

Since the beginning of the FTTx deployment projects, the ONU equipment has had several improvements which increased equipment complexity. It started with an ONU box converting the signal from optical to electrical (and vice-versa) and bridging the traffic to a local router. Later, to reduce the number of equipment on the client premises – thus reducing costs, the ONU become part of what is known today as fiber gateways. These are devices with the capability of acting as routers as well as integrating an ONU module. In some cases, it includes an android module for IPTV services. Unboxing a fiber gateway implies the analysis of several and different modules, such as optical transceivers, Wi-Fi chips, layer 2 logical modules, RF transceivers, electrical and physical components, among others.

Equipment installation and location

Installation conditions have a huge impact in equipment performance. The building location, the deployment quality or the environmental conditions are major factors. The installation process is usually performed by technical staff, but in some cases equipment is swapped or somehow tampered with by the clients.

A critical environmental factor is, once again, temperature, which influences the equipment longevity, particularly the lifetime of some ONU components.

As described above, and due to the components and modules diversity aggregated in a single box, it is highly recommendable to assess the critical components and identify those indicators that should be considered, to provide useful information for operators as well as vendors.

The installation location can also derive into poor equipment performance, causing unnecessary complaints early in the contract. Service interruption and Wi-Fi issues are so far the main reason for customer dissatisfaction. To avoid such negative feedback, installation details, such as room position, home areas, wall types and information about the usage (Wi-Fi or Ethernet), must be a part of customer available information. This will help data scientists (or AI processes) to find patterns, by correlating it with other indicators, such as the number of connected devices, type of connections, received Wi-Fi power by device, etc.

Elements and variables to retain:
environment classification (temperature, humidity); environment details (room type, wall type, home areas, number of floors, preferable connection types, potential number of users); Wi-Fi details (number of connected devices, received power)

Degradation of performance

Electrical components have a limited lifetime, depending on the conditions and usage. In this field, operators are once again dependent on the client's feedback to act, analyse and resolve the issue. Today's equipments have the possibility to be self-managed and self-diagnosed by storing and comparing indicators that degrade over time, such as input voltage, flash blocks, output optical power, or transmitted Wi-Fi power. If these indicators were available since the beginning of production, it would be possible to have an accurate idea of how components perform in specific environmental conditions.

Other indicators or information can also be considered, such as the production date, production lot, and key component models, like the bi-directional optical sub-assembly (BOSA), Wi-Fi, flash, etc.

By analysing and correlating all this information, it is possible to predict potential faults and replace equipment before costumers complain.

Elements and variables to retain:
input voltage; physical memory; Wi-Fi model; BOSA model; manufacturing details (date, production lot and initial performance)

Elements and variables analysis

This section briefly describes some elements and variables used in network operations. We believe that these listed here, in particular, show potential to be used in a cognitive process meant to anticipate failure, optimise and automate network operation, and overall improve network quality. Some indicators are already available on current operations, while others will require further data collection during production and installation. The following list presents some relevant examples of these indicators:

- **Dying-gasp event:** this event is sent by the equipment to indicate a loss of power. The number of events reported within a short time suggests an instability issue which might be caused by faulty power supply units (PSU), defective devices, local power supply issues, etc. The number of power cycles is also a value to take into account to estimate the end-of-life of PSU or power modules;

- **Optical transceiver** (laser): this element provides optical transceiver details, such as bias current, temperature and optical transmitted power. Laser ageing causes degradation of optical power, which can be worse in the wear of the photodiodes due to the increase of the temperature. In [10], there is a description of a process to estimate laser diode lifetime by a function of operation time and operational temperature. The operational temperature is the main cause of the laser lifetime decrease. It is stated in [10] that the lifetime of a laser diode

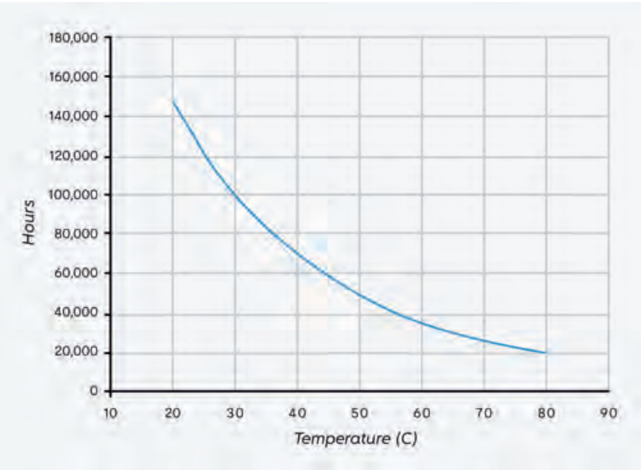


FIGURE 3 – Laser diode lifetime versus temperature [10]

decreases by approximately a factor of two for every 20 degrees Celsius rise in operational temperature. **Figure 3** illustrates this effect;

- **Physical memory** (flash): this component had a notable evolution over time, and today we can find various types of technology (NOR, NAND, SSD, eMMC ...). However, the limited number of write-cycles remains one of its main limitations. Problems such as bad blocks are also factors that may affect software performance and, consequently, equipment performance. Flash-writes count the number of operations performed in flash components and can help us understand how far the device is from the end of its lifecycle. An exponential increase of such counter might point an issue that correlated with other indicators might show up erroneous operations from local or remote systems;

- **Power amplifier:** used, for example, in RF modules, this is a key component that is also affected by temperature fluctuations;

- **Power supply unit** (PSU): there are frequent disturbances in the input voltage, which might influence the operating conditions (thus affecting some modules). Some equipment is already able to measure input voltage and internal voltage supplies. These values are likely to identify an imminent failure or faster hardware degradation, so it is essential to analyse them whenever available;

- **BOSA model:** well-identified information for each device, and it is already used in some operators' networks;

- **Manufacturing date:** this indicator may be used to calculate unit ageing;

- **Production lot:** this information can be used to group devices;

- **Transmitted power:** this value is measured in manufacturing tests for each interface (optic, Wi-Fi, CATV ...). It is stored in a local database to be used only for individual root cause analysis (RCA), but, if available, this information can be useful also if it is monitored for changes over time;

- **Optical received signal strength indicator** (RSSI): can also be used to detect or predict anomalies in the network or customer premises;

- **Environmental conditions:** provided by the installation report, may contain useful information like room type, wall type, home areas, number of floors, preferable connection types, potential number of users, temperature, humidity, etc.

Conclusion

AI provides the capability to go far beyond the human capacity to deal with the volume and complexity of data that networks generate today. In particular, for passive optical networks, there are numerous benefits to be collected from the additional insights that can be produced. More specifically, we have identified the following areas:

- **Predictive operation:** recognition of failure and pre-failure patterns, enabling proactive assurance and promoting the autonomous behaviour of the network;

- **Customer network awareness:** providing insights to the customer premises equipment and network, allowing for informed customer recommendations and local optimisation;
- **Network equipment optimisation:** equipment manufacturers can also benefit from intelligent operation data generated to improve both hardware and software quality.

This article results from an approach to some problems of PON in the light of AI, intending to identify opportunities that these new processes may bring to the area. Some of the use cases identified here may be addressed by Altice Labs in proof of concept initiatives, both in the laboratory or on the field.

References

[1] "FSAN - Full Service Access Network." [Online]. Available: https://web.archive.org/web/20091010143600/http://www.fsanweb.org/history.asp.

[2] ITU-T, "ITU-T G.983.1 (01/2005) - Broadband optical access systems based on Passive Optical Networks (PON)," 2005.

[3] ITU-T, "ITU-T G.984.3 (01/2014) - Gigabit-capable Passive Optical Networks (G-PON): Transmission convergence layer specification," 2014.

[4] ITU-T, "ITU-T G.987.3 (01/2014) - 10-Gigabit-capable passive optical networks (XG-PON): Transmission convergence (TC) layer specification," 2014.

[5] ITU-T, "ITU-T G.9807.1 (10/2017) - 10-Gigabit-capable symmetric passive optical network (XGS-PON)," 2017.

[6] ITU-T, "ITU-T G.989.3 (11/2018) - 40-Gigabit-capable passive optical networks (NG-PON2): Transmission convergence layer specification," vol. 3, no. 2015, 2018.

[7] ITU-T, "ITU-T G.988 (11/2017) - ONU management and control interface (OMCI) specification," 2017.

[8] ITU-T, "G.984.2 (03/2003) - GPON: Physical Media Dependent (PMD) Layer Specification," 2003.

[9] L. A. Johnson, "Laser diode burn-in and reliability testing," *IEEE Commun. Mag.*, vol. 44, no. 2, 2006.

[10] P. Gale, "Estimating Laser Diode Lifetimes and Activation Energy," *ILX Light. Appl. Noter#33*, 2008.


5G

06 5G intelligent communications for V2X ecosystems

Alberto Rocha, Altice Labs
alberto-m-rocha@alticelabs.com

Francisco Fontes, Altice Labs
fontes@alticelabs.com

Luís Cortesão, Altice Labs
luis-m-cortesao@alticelabs.com

Pedro Miguel Salgueiro Santos, CISTER
pss@isep.ipp.pt

Ramiro Sámano Robles, CISTER
rasro@isep.ipp.pt

Ricardo Severino, CISTER
rarss@isep.ipp.pt

5G associated to edge computing and the latest implementation technologies, is deemed to provide the required performance to address the most demanding scenarios like V2X use cases. However, if in ideal conditions, 5G has all the ingredients to fulfil those requirements, the real world is more challenging, with internal and external factors affecting network behaviour. To overcome that, AI and ML components play an essential role, sensing, mining, predicting and reasoning situations, being relevant building blocks of future autonomous 5G networks.

Keywords

5G; Intelligent communications; V2X; Platooning; Network slicing

Introduction

We are experiencing an increasingly connected and digital world. Industries are no exception, undertaking a digital transformation where external and internal communications are essential.

5G differentiates from previous cellular generations and other wireless technologies by providing business-to-business (B2B) vertical sectors (called "verticals") with enhanced private and public communications, over a common infrastructure. 5G high bandwidth, low latency and high reliability, are the enablers for a new generation of services, to be anchored in natively supported edge computing. Embedded slicing mechanisms provide tailored connectivity to be used, for instance, in e-Health, energy, education, agriculture or transportation use cases.

The emerging 5G ecosystem will be distributed and heavily based in virtualised environments. A complete vision of the entire system and a holistic orchestration will be required, eventually across different administrative domains. The adoption of artificial intelligence (AI) and machine learning (ML) mechanisms, at network and services levels, will be of paramount importance to fine-tune the system, forecast events and anticipate actions.

Vehicle-to-everything (V2X) is one of the most relevant and promising verticals to exploit 5G. The creation of the 5G Automotive Association (5GAA) [1] reflects this. Besides interconnected autonomous vehicles, many other use cases are possible by ensuring vehicles have the right connectivity to (1) other vehicles, (2) with the surrounding environment and (3) with remote service platforms.

This document starts presenting the emerging 5G ecosystem characteristics, positioning it to be the service platform for cellular vehicular (C-V2X) use cases. The following section describes the role of AI/ML at 5G network level and as an enabler of advanced vehicular use cases. The final section presents vehicular platoons as an example of the need and the advantages of having a close interaction of AI/ML mechanisms operating at those two layers.

5G: the next generation services platform

5G represents a new era in cellular communications, presenting the functionality and the performance improvements required to place it as a keystone in the digital transformation process. But exploiting all the 5G technology potential, impacts other components, like data centres, and requires the common adoption of AI and ML.

5G performance and functionality

Current 5G commercial deployments, based on 3GPP Release 15 specifications [2], add a 5G radio data plane to existing 4G networks, classed as a non-standalone (NSA) architecture, mainly providing more bandwidth to be used by the business-to-consumer (B2C) market. These deployments fall into the enhanced mobile broadband (eMBB) use cases. The 5G-PPP Vertical Engagement Task Force (VTF) identified the following industry sectors, or "verticals":

1. Automotive;
2. Manufacturing;
3. Media;
4. Energy;
5. e-Health;
6. Public safety;
7. Smart cities.

Some of these verticals are already organised around the 5G topic, such as automotive, with the 5GAA, and industry, with the 5G Alliance for Connected Industries and Automation (5G-ACIA) [3], reflecting the importance to influence 5G definition and regulations, and consistently take the maximum benefit of it.

Aiming at serving all those, requires 5G to improve performance over 4G significantly, beyond mobile Internet requirements. Besides higher bit rates (up to 20Gbps), a 5G standalone (SA) system, based on 3GPP Release 16 specifications [4], will support a larger number of simultaneous connections and devices (up to 1 million per square kilometre), provide ultra-low latency at radio level (down to 1ms) and high reliability (up to 99.9999%), opening the floor to massive machine type (mMTC) and ultra-reliable and low latency (URLLC) communications, addressing the B2B market.

Higher frequencies (above 24GHz) operation, a denser set of antennas and integrating mechanisms such as multi-user multiple-input and multiple-output (MU-MIMO), guarantees high bandwidth. However, 5G is also able to operate at lower frequencies (below 1GHz), with lower bandwidth but broader and deeper coverage. A new radio interface, e.g. via a new frame structure, and system architecture, clearly separating the control and user planes (CUPS), guarantee low latency. High reliability is achieved by incorporating mechanisms in the architecture and in the radio interface to provide spatial, frequency and time diversity.

Functionally, 5G emerges with native, built-in, support of slicing and edge computing. Slicing allows the creation of complete virtual end-to-end networks, from the radio interface to the service platforms, which can be managed individually (see **Figure 1**). These slices have the appropriate characteristics and are assigned with the right resources, tailored to the type and level of connectivity services they need to provide. Edge computing allows the placement of computing resources close to the network edge, as much as the 5G radio units, allowing services to be provided with minimal latency.

3GPP specified three main types of standard slices, later adding a V2X one, which is intended to address the specific requirements of the automotive vertical, resulting in the following [5]:

1. **eMBB**: providing high bandwidth, with faster mobility (e.g. fast trains or drones);
2. **mMTC**: for more connections, widespread and deep coverage, even if at lower bitrates;
3. **URLLC**: for extremely low latency communications and guaranteeing strong reliability;
4. **V2X**: for the specific support of cellular vehicular communications use cases.

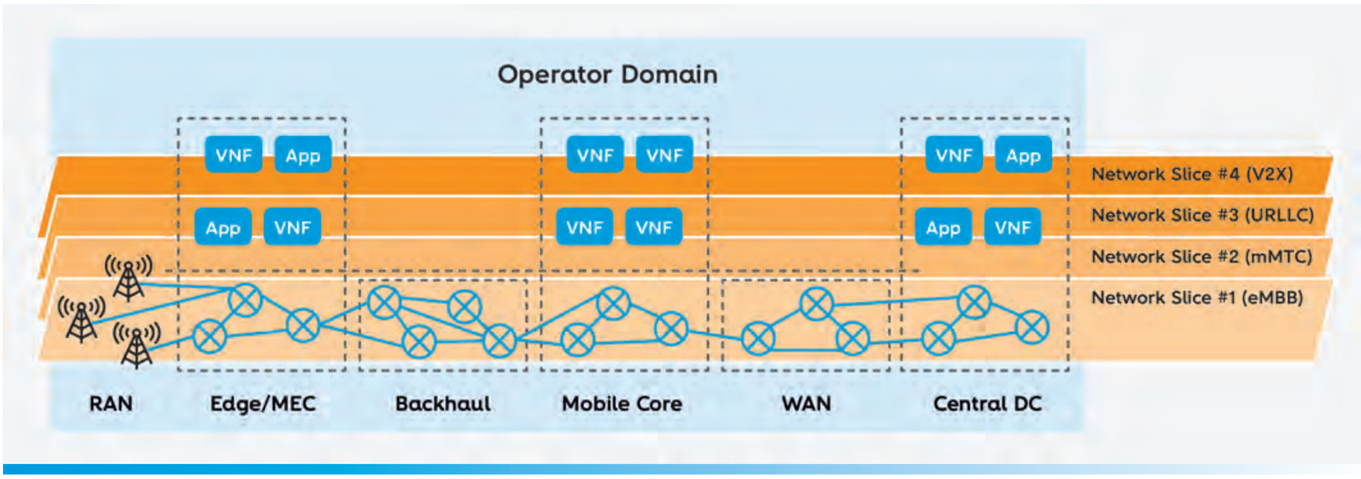


FIGURE 1 – 5G slices

Operators will instantiate multiple slices of these standardised types and other tailored ones, answering the needs of specific use cases and verticals. Those needs may span from performance (e.g. latency, throughput, availability), to functional (e.g. security and identity management) and operational (e.g. self-management) requirements.

The new 5G data centre paradigm

Aligned with current IT trends, 5G brings new technologies, new architectures and new topologies, associated with a new development culture, new deployment methods and new delivery challenges. The virtualisation of compute, network and storage resources has leveraged the success of cloud computing services, decoupling software functions from the physical bearer.

The degree of automation and orchestration (enabled by a programmable infrastructure) brings higher control and flexibility to the deployment and management of the infrastructure resources. In particular, network function virtualisation (NFV) service platforms, jointly with the software-defined networks (SDN) end-to-end network control capabilities, provide the tools for the creation and deployment of innovative network services and application in a short time, most likely organised as slices (as referred in the previous section), and addressing the time-to-market business demands.

Cloud computing capabilities (e.g. application and service orchestration, and device management) are gradually moved to smaller, distributed data centres with identical characteristics but placed near the end-user to fulfil the latency requirements for specific services. By moving the intelligence closer to the edge, enriching received raw data, creating metadata and sending back the information relating to the vertical or the actions/decisions to be taken, can be done in a much faster way. Operators also take own benefit of this infrastructure, running virtualised components of disaggregated wireless and wired access components for mobile, fiber or cable networks, on this infrastructure. Independently from 3GPP work on 5G, the Broadband Forum (BBF) is proposing a Cloud Central Office (CCO) architecture to address this [6].

Edge computing will benefit from solutions which bring efficient infrastructures, running on open commoditised platforms, enabling solutions that run on standard processors rather than hardware silos. The degree of openness of the platform and its provisioning, relying on open-source software and commodity hardware, are the drivers for the solutions being designed and are critical to success. The approach being adopted by the majority of manufacturers is to commoditise the hardware aspect and, at the same time, increasingly use software to differentiate and drive competition.

As it turned out, edge computing is a highly distributed network of infrastructure resources, possibly organised in network slices. NFV and SDN solutions can leverage its deployment, placement, rule-based autoscaling, self-healing, rollup updates, onboarding and the full life cycle management.

Several standard definition organisations, open-source communities and industry fora, like the OpenStack Foundation (OSF) [7], Cloud Native Computing Foundation (CNCF) [8], Telecom Infra Project (TIP) [9], Open Compute Project (OCP) [10] and the European Telecommunications Standard Institute (ETSI) multi-access edge computing (MEC) [11] and NFV [12], are actively working on the edge computing definition, architectures, APIs, guidance, best practices, management, interworking, security and business models.

ETSI NFV Release 3 [12] features related to 5G include support for network slicing in NFV, management over multi-administrative domains, and multi-site network connectivity. These features are essential to address the variety of applications expected to run on top of a 5G system, whether using distributed resources over multiple sites, centralised or a combination of both.

Actually, commercial solutions like Vapor IO [13], EdgeMicro [14] and Dell/EMC Micro Modular Data Center [15] were designed to deploy a nationwide network of small modular edge data centres, either on a fixed location or as nomadic units, to address unexpected calamity, traffic rush or sports event with thousands of consumers sited on a large stadium.

TIP is an engineering-focused initiative driven by operators, suppliers, developers, integrators, and start-ups to disaggregate the traditional network deployment approach. Although the various TIP Project Groups are independent, their work is complementary. Among others access network initiatives, the Edge Computing [16] is one of the working areas: focus on lab and field implementations for services/applications at the network edge, leveraging open architecture, libraries, software stacks and MEC.

5G-based cellular-V2X

In the vehicular communication ecosystem, there is a variety of required communications. The first ones are vehicle “direct links”, including vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I) and vehicle-to-pedestrian (V2P). The second ones are “network links”, for vehicle-to-network (V2N) communications, supported by LTE/5G technology, suitable for both local and long-distance applications. A single or a combination of such links need to be considered when designing a cooperative vehicular application, depending on factors like the application’s architecture (local or cloud-based), requirements (small latency, long-distance communication) and availability of the technologies supporting the various types of links.

Wireless local area network (WLAN) [17] based solutions have been proposed for a long time to address V2V and V2I needs. Specifically, IEEE 802.11p [18], often referred to as dedicated short-range communication (DSRC), provides ad hoc device-to-device communications. It is the technical foundation of two regional protocol stacks, the Wireless Access in Vehicular Environment (WAVE) stack in the USA [19] and the ETSI Intelligent Transportation System (ITS-G5) in Europe [20]. It leverages the physical (PHY) and medium access control (MAC) layers of IEEE 802.11 and operates in the 5.9GHz Industrial, Scientific and Medical (ISM) band, offering a maximum bitrate of 27Mbit/s, and typical line-of-sight ranges are up to 1km [21]. The current lack of dedicated 802.11p-enabled roadside infrastructure means that connections from the vehicle to the network need to be mediated by an onboard cellular interface. In 2018 IEEE started the Next Generation V2X study group [22] to address identified limitations.

C-V2X is a 3GPP reference for vehicular communications. It answers both performance and functional requirements, leveraging the defined cellular radio technology for the direct communications between vehicles (device to device or D2D). C-V2X guarantees communications to the network, providing ubiquitous connectivity via the built-in handover and roaming mechanisms.

While 3GPP Release 14 specified LTE-based C-V2X, Release 15 introduced 5G New Radio (NR) to C-V2X use-cases and Release 16 extended it. The most demanding requirements for autonomous driving [23], to be answered by NR-V2X, are presented in **Table 1**.

Use cases	E2E latency (ms)	Realiability (%)	Data rate (Mbps)
Vehicle platooning	10	99.99	65
Advanced driving	3	99.999	53
Extended sensors	3	99.999	1000
Remote driving	5	99.999	upload: 25 download: 1

TABLE 1 – NR-V2X requirements for autonomous driving [23]

These are organised in four areas [24]:

1. Vehicles platooning;
2. Advanced driving;
3. Extended sensors;
4. Remote driving.

An analysis produced by Qualcomm [25] informs about how LTE/C-V2X and IEEE 802.11p compare in a highway emergency breaking scenario, with C-V2X outperforming the IEEE standard.

Thus, 5G has the potential to become a game-changer in C-V2X rollout, benefiting from the ubiquity of the cellular infrastructure and mature industrial foundations, inherited from decades of standardisation work in 3GPP. It is able to support all V2X communication "flavours", either between vehicles, vehicles to pedestrians, vehicles to infrastructure or network, with services running at the edge or at centralised platforms.

Besides technology, policymaking organisations play an important role regarding C-V2X support by the networks and by vehicles, with the European Commission (EU) setting up the target for the main European transport paths to have 5G coverage by 2025 [26].

AI and ML for 5G V2X scenarios

AI and ML are nowadays present in many applications and services. Operators are adopting them to have better-performing networks and the vehicular industry to provide more secure and efficient transportation systems. Vehicular use cases, requiring communications with the surrounding environment, need AI/ML processes running at both levels to coordinate for optimal performance.

AI/ML in 5G networks

"Many 5G aspects require analytical capabilities and responsiveness beyond human capacity. For instance, if operators want to optimise the connections for every device and application, they need to shift their network planning and monitoring tools from a network-centric to a user-centric view. That requires a continuous, end-to-end perspective of real-time network behaviour, which, in turn, demands the ability to correlate vast amounts of network testing and statistical data towards an accurate picture of the quality of experience" [27]. Furthermore, the network topology, design and propagation models, along with user's mobility and usage patterns in 5G, will be complex.

To assure the performance of the diversity of the supported applications (e.g. connected cars and industrial Internet of things), it will require continuous network monitoring, troubleshooting, and optimisation based on an accurate end-to-end view of network behaviour. It will also require collecting and processing data from a multiplicity of sources, simultaneously and in real-time. Such a vast amount of information will need complex analytical capabilities, supported on dedicated tools and practices, including AI and ML processes that will ensure efficient management of network resources and flexibility to meet user demands.

AI and ML will prove to be crucial in supporting the desired evolution:

- From **adaptive networks** – able to perform intelligent analysis, real-time acquisition of network data and perception of network status, to generate optimisation strategies to enable closed-loop operations;
- To **autonomous networks** – able to realise self-* (optimisation, healing, organising, etc.) features, based on a robust knowledge representation including context-aware situation awareness as part of a comprehensive cognition framework, and use policy-based management to enable adaptive

and extensible service offerings that respond to changing business goals, user needs, and environmental constraints [28].

ML techniques are enablers for the automation of network functions, through its capability of sensing (e.g., anomaly detection), mining (e.g., service classification), predicting (e.g., forecasting traffic trend or anomalies), and reasoning (e.g., the configuration of system parameters for adaptation). In an end-to-end perspective, it will provide capabilities to analyse a huge volume of data in a very short time, learn to adjust the system to time-varying environments, make predictions of future events with reasonable accuracy and prescribe proactive solutions [29].

For instance, embedding ML algorithms and AI into 5G networks can enhance automation and adaptability, enabling efficient orchestration and dynamic provisioning of network slices [30]. ML and AI can collect real-time information for multidimensional analysis and create a panoramic data map for each network slice, based on user subscription, quality of service, network performance, events and logs.

Different aspects where ML and AI can be leveraged, include [30]:

- **Predicting and forecasting the network resources health** can enable wireless operators to anticipate network outages, equipment failures and performance degradation;
- **Cognitive scaling**, to assist wireless operators to dynamically modify network resources for capacity requirements, based on the predictive analysis and forecasted results;
- **Predicting user equipment mobility in 5G networks**, allowing for the update of mobility patterns data based on user subscription, historical statistics and instantaneous radio conditions for optimisation and seamless transition to ensure a better quality of service;
- **Enhancing the security in 5G networks**, preventing attacks and frauds by recognising user patterns and tagging certain events to prevent similar attacks in future.

In the scope of verticals with real-time requirements (e.g. ultra-low-latency sensitive V2X applications), 5G network operators need to provide efficient and precise tools, many running at the network edge, to cope with the requirements mentioned above. Operators will rely on AI/ML to process its massive and diverse historical data to build the models supporting this multiplicity of features. It ranges from pre-emptively informing the applications about issues with the predicted network service quality so that it can take appropriate measures (i.e., adjust its internal processes to rely less on network service), to plan and deploy network adjustments (e.g., service migration from the initial edge data centre) based on the information provided by the V2X application, like planned itinerary, traffic constraints, unplanned impacting events, etc.

V2X in intelligent 5G environments

V2X communications will be the technological basis to support an entirely new set of vehicular applications, operating as a service overlay that will extend the functionalities of existing or soon-to-be sensing and autonomous driving technology. In turn, AI is one of the quintessential elements of autonomous driving. V2X and AI will complement each other to support much more complex and mobility-disruptive interactions between vehicles, pedestrian users, and physical infrastructure.

The keyword is cooperative: V2X (e.g. through 5G) allows information sharing between vehicles, infrastructure and other road-users that is outside the reach of the sensors of individual road-users. The AI algorithms will take the wide variety of sensors available in the vehicles, processing the generated data to identify where other road users are, then sharing that information through a V2X channel; or receive information through a V2X link about the whereabouts of other road users and react accordingly.

In safety-critical applications, for instance, alerting for immediate collision danger, URLLC communication between vehicles will be required. Consider the following V2X use-cases (inspired by [25]):

1. **Forward collision warning** (FCW): a vehicle moving ahead informs a follower vehicle that a collision occurred and an emergency braking procedure should be initiated; for this, an AI algorithm needs to receive the video feed from the front-facing cameras of the vehicle, identify the event and decide to relay that information over V2X links;
2. **Do-not-pass warning** (DNPW): the vehicle warns the follower not to attempt an overtake as an oncoming car in the opposite lane approaches, after identification by the AI algorithm;
3. **Vulnerable road user** (VRU) indication: an AI algorithm receives the vehicle's front-facing cameras video feed and on-going wireless communications. The algorithm understands the vehicle is approaching a blind spot and relies less on the camera, assigning more weight to other sensors. The algorithm identifies the presence of a pedestrian behind the blind spot by detecting the pedestrian's smartphone beacons.

The correct, safe, stable and coordinated control of vehicles is another vehicular application in which AI and V2X communications will have to operate hand in hand. Consider the following examples for platooning:

- **Kinematic models**: kinematic specificities of each member (e.g. inferior braking power and leaning tendencies) can be sent to the leader through V2X links, and AI algorithms can incorporate the individual behaviours in the platoon's overall kinematic model;
- **Efficient driving techniques**: a platoon in a highway is preparing to engage in higher speed and smaller inter-vehicle distance. V2X links are used to coordinate the action, while AI algorithms must build situational awareness to greenlight the manoeuvre.

At a macroscopic scale, AI will also enable intelligent traffic management systems. By sharing information with other road-users and a centralised traffic management authority, goods freight will become more efficient, less risky and with reduced delays. Consider:

- **Kinematic-aware route selection**: AI mechanisms can identify the best routes, taking into consideration traffic density, received from a central system via V2X, and the truckloads (e.g. if trucks are loaded or not). This condition affects the kinematic properties of the platoon (e.g. if loaded, inferior speeds and less agility shall be expected);
- **Last-mile logistics with drones**: platoons equipped with drones, known as unmanned aerial vehicle (UAV), can optimise their delivery routes by dispatching drones to deliver goods to nearby delivery points. Trucks and the drones must keep real-time communication so that the drone can be aware of the truck's position;
- **Traffic conditions and services**: warnings about traffic jams, alerts about emergency vehicles and available parking spots and tariffs, can be shared.

Platooning use case

Platoons will be the basis of future automated transportation systems, including passenger buses, fleets, freight, and trucks. Transportation of goods (supply chain management) will become more efficient, less risky and faster [31]. With platooning solutions, urban traffic flows will be more efficiently managed by relying upon regular structures of vehicles with similar routes, exchanging information between them and with the roadside infrastructure to safely coordinate their actions.

To support the use cases mentioned in section "V2X in intelligent 5G environments", in particular platooning, the requirements to be fulfilled by network operators regarding 5G and edge computing technologies, are significant and challenging. The safety- and time-critical nature of many of those use cases, require ultra-reliable and low latency wireless connectivity with other vehicles, as well as with the infrastructure and with the edge cloud nodes. In turn, applications such as control (platoon coordination, efficient driving manoeuvres) and ITS (traffic coordination, route planning) may have a different set of requirements, e.g., demand continuous connectivity to edge data centres that provide updated information.

Let us contextualise the wide range of network requirements in a broader real-world application. Consider that a logistics company wishes to freight goods, leveraging on automated platoons for regular delivery missions to industrial or logistic locations for improved efficiency and savings. The platoon is supported throughout its itinerary by infrastructural the most technologies, such as 5G-enabled cell-towers, roadside units equipped with IEEE 802.11ac/ax/ad/p technologies, and edge data centres for low latency processing of offloaded applications. Under the framework of a VNF/slicing-enabled network, and tightly linked to edge-provided services, the network operator starts by deploying (one or more) network slices. These slices can be of different types, according to the operator strategy (e.g. URLLC, eMBB and V2X), devoted to providing tailored communication services to the V2X vertical. Applications requiring network-side processing, such as autonomous platoon driving support, and included in the network operator catalogue, are onboarded and instantiated at edge data centres, and the required network slices are activated. The selection of the target edge DC aims at simultaneously guaranteeing the V2X application performance requirements and efficient network usage. Finally, the vehicles connect to the network and the V2X network slices and initiate their mission.

To get a rough idea of the scale of delays at play, consider the example of a highway occurrence that requires emergency braking by the platoon, in the following conditions:

1. Platoon cruising at a reduced speed of 50km/h;
2. Distance between vehicles of 2 meters;
3. Maximum deceleration (breaking) capability of $6m/s^2$.

Upon identification of, for instance, a stopped vehicle involved in a crash, the leading vehicle triggers and emergency brake action and communicates this action to the followers. If relying on a typical 4G network, a 100 bytes message and often assuming peak situations, a successful transmission could take around 100ms. Within such timeframe, the follower would have cruised 1.4 meters and the leader 1.1m (due to breaking). If the follower does not receive a transmission up to approximately 817ms, the follower will crash into the front vehicle (see **Figure 2** on next page). 5G provides some milliseconds transmission delays, supporting even smaller inter-vehicle distances is possible, resulting in increased platooning efficiency. From the perspective of a 5G network, a lot is happening in the background so that this level of service can be provided (see section "AI/ML in 5G Networks").

For most of its uneventful itinerary, the platoon will perform handovers between the different mobile 5G cells and/or open networks (e.g. according to signal strength, exact location, etc.). Nevertheless, the network must provide continuous connectivity to the edge data centres supporting the platoons' internal control operations and coordination with the remaining traffic. Platoon operations include complex kinematics and trajectory prediction calculations that may be carried out at the edge nodes, and coordination with other traffic may consist of intersection coordination and route planning. On the one hand, by leveraging AI/ML technologies the operator will be able to correlate the planned itinerary of the platoon with the infrastructure

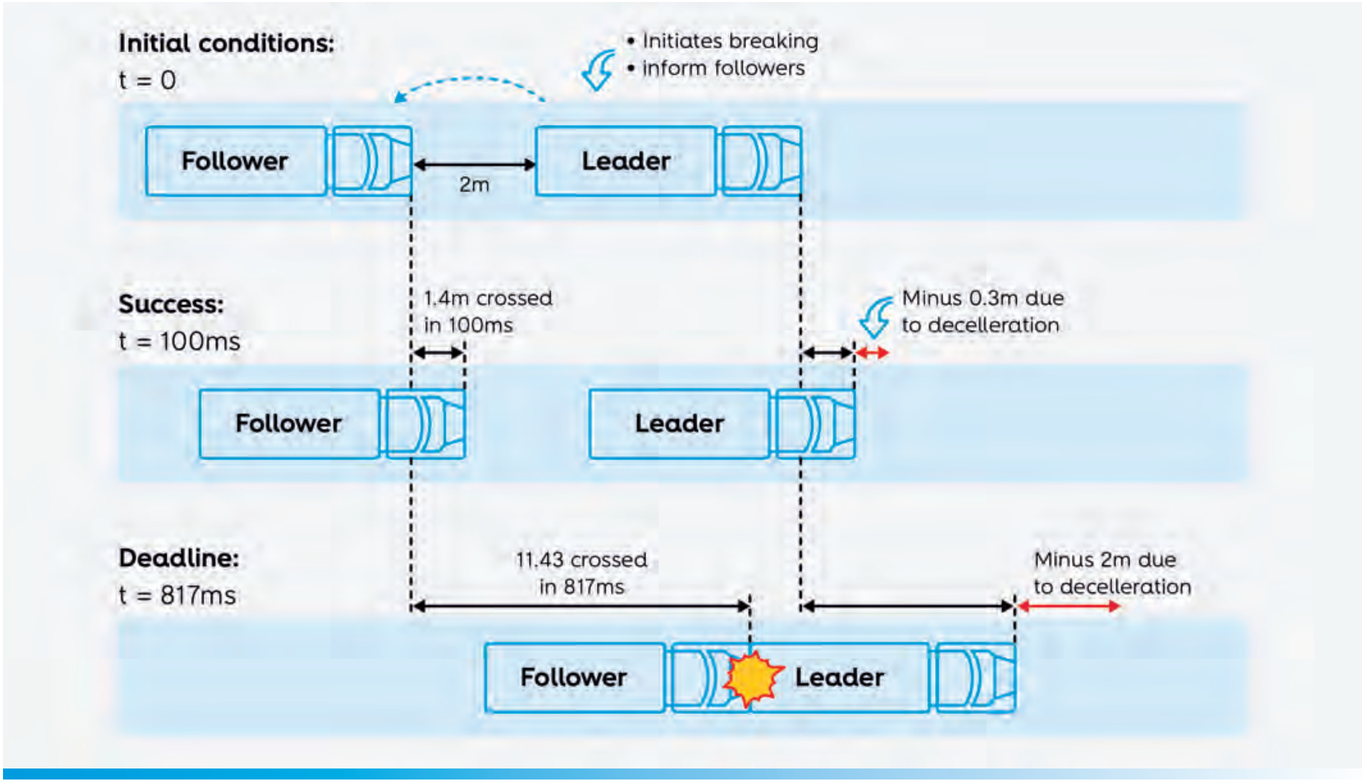


FIGURE 2 – Platooning emergency breaking

topology, time of day and expected network traffic, and also making the network to predict when to perform service migration from the initial edge data centre proactively. On the other hand, the network may continuously monitor and predict the connection quality to the platoon and its vehicles (e.g. if the platoon will enter an area of subpar connectivity). AI/ML will support the run-time operation, through the prediction of wireless channel variations and physical obstacles, the identification and mitigation of potential sources of interference or jamming attacks.

If the network's AI algorithms detects a potential loss of connectivity, it pre-emptively informs the platoon (the driver, if present and in control, and the remote human operator responsible for monitoring the status of the platoons) about the predicted network service quality. In turn, the platoon can adjust its internal processes to the network service level, reducing the connectivity performance requirements (latency and throughput). A safeguard may be in place, to safely immobilise the platoon or a subset of the platoon vehicles, in case this communication does not occur correctly.

In conclusion, this support infrastructure, enabled by 5G and edge computing, will be critical for a variety of applications where the platoon is required to engage in communication with the network and other vehicles, such as:

- **Transmission of tracking data to the logistics management site**, for safety and efficiency reasons, with rich multimedia content (e.g., live video feeds);
- **Engage with traffic management systems** (or ITS), for services such as lane merging, reversible lane or safe intersection management, and in coordination with other road-users;
- **Be alerted to unforeseen or dangerous events during these ITS processes**, such as unexpected obstacles, temporary roadblocks or accidents in a specific section of the itinerary.

Conclusions

As presented in the section "5G: the next generation services platform", 5G is expected to support use cases that need cellular connectivity. But it goes even further, defining a 5G Core able to connect other wireless (including non-3GPP, e.g. WLAN) and wired access technologies. Thus, 5G targets a wide range of use cases, even those currently suited to be served by fixed technologies. 5G, based in 3GPP Release 16 specifications, associated to edge computing and the latest implementation technologies, is deemed to provide the required performance to address the most demanding scenarios like the identified vehicular use cases (V2X). However, if in ideal conditions, 5G has all the ingredients to fulfil those requirements, the real world is more challenging, with internal and external factors affecting network behaviour. To overcome that, AI and ML components play an essential role, sensing, mining, predicting and reasoning situations, being relevant building blocks of future autonomous 5G networks.

Unexceptionally, many V2X applications will exploit AI and ML, complemented by advanced C-V2X communications. That will allow vehicles to sense and connect to the surrounding environment and remote service platforms, creating use cases that go well beyond the simple individual, isolated and autonomous vehicle, as presented in section "AI and ML for 5G V2X scenarios".

The marriage of a universal communications platform, enabled by 5G based C-V2X communications, with correlated AI/ML algorithms, running at the network and services levels, is the only possible way for the consistent exploitation and smooth execution of advanced use cases, like vehicles platooning, described in the "Platooning use case" section.

As so, presented C-V2X based platooning use case, is one of the most demanding applications of 5G, edge computing and AI/ML-based management and orchestration mechanisms.

References

[1] "5G Automotive Association (5GAA)," 2018. [Online]. Available: http://5gaa.org/.

[2] 3GPP, "TR 29.891 - 5G System - Phase 1; CT WG4 Aspects (Release 15)," 2019.

[3] 5G-ACIA, "5G Alliance for Connected Industries and Automation (5G-ACIA)," 2018. [Online]. Available: https://www.5g-acia.org/.

[4] 3GPP, "TS 22.186 - Enhancement of 3GPP support for V2X scenarios; Stage 1 (Release 16)," 2019.

[5] 3GPP, "TS 23.501 - System Architecture for the 5G System; Stage 2 (Release 15)," 2018.

[6] Broadband Forum, "TR-384 - Cloud Central Office (CloudCO) Reference Architectural Framework," 2018. [Online]. Available: https://www.broadband-forum.org/technical/download/TR-384.pdf.

[7] OpenStack Foundation, "Edge Computing Group/Edge Reference Architectures." [Online]. Available: https://wiki.openstack.org/wiki/Edge_Computing_Group/Edge_Reference_Architectures.

[8] CNCF, "Cloud Native Computing Foundation," 2019. [Online]. Available: https://www.cncf.io/.

[9] Telecom Infra Project, "Telecom Infra Project." [Online]. Available: https://telecominfraproject.com/.

[10] [OCP, "Open Compute Project." [Online]. Available: https://www.opencompute.org/projects.

[11] ETSI, "Multi-access Edge Computing (MEC)," 2019. [Online]. Available: https://www.etsi.org/technologies/multi-access-edge-computing.

[12] ISG NFV, "Network Functions Virtualisation (NFV)," ETSI, 2019. [Online]. Available: https://www.etsi.org/technologies/nfv.

[13] Vapor IO, "Vapor IO." [Online]. Available: http://vapor.io.

[14] EdgeMicro, "EdgeMicro." [Online]. Available: https://www.edgemicro.com/.

[15] Dell EMC, "Dell EMC Micro Modular Data Centers," 2019. [Online]. Available: https://www.sdxcentral.com/products/micro-modular-data-center/.

[16] Telecom Infra Project, "Edge Computing." [Online]. Available: https://telecominfraproject.com/edge-computing/.

[17] IEEE, "IEEE 802.11 Wireless Local Area Networks," 2019. [Online]. Available: http://www.ieee802.org/11/.

[18] IEEE, "IEEE 802.11p-2010-IEEE Standard for Information technology-Local and metropolitan area networks-Specific requirements-Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 6: Wireless Access in Vehicular Envir." 2010.

[19] IEEE, "IEEE P802.11p/D3.0, Draft Amendment for Wireless Access in Vehicular Environments (WAVE)." 2007.

[20] ETSI EN 302 663 V1.2.1, "Intelligent Transport Systems (ITS); Access layer specification for in the 5 GHz frequency band," vol. 1, pp. 1–24, 2013.

[21] A. Paier, D. Faetani, and C. F. Mecklenbräuker, "Performance evaluation of IEEE 802.11p physical layer infrastructure-to-vehicle real-world measurements," in *2010 3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL 2010)*, 2010, pp. 1–5.

[22] IEEE, "IEEE Announces Formation of Two New IEEE 802.11TM Study Groups," 2018. [Online]. Available: https://standards.ieee.org/news/2018/ieee_802-11_study_groups.html.

[23] 3GPP, "TR 22.886 - Study on enhancement of 3GPP Support for 5G V2X Services." 2018.

[24] 3GPP, "TR 38.885 - Study on NR Vehicle-to-Everything (V2X)." 2019.

[25] Qualcomm Technologies, "Leading the World to 5G: Cellular Vehicle-to-Everything (C-V2X) Technologies," no. June, pp. 1–39, 2016.

[26] European Commission, "Communication – 5G for Europe: An Action Plan and accompanying Staff Working Document," 2016. [Online]. Available: https://ec.europa.eu/digital-single-market/en/news/communication-5g-europe-action-plan-and-accompanying-staff-working-document.

[27] A. Jakobsson, "The 5G Future Will Be Powered By AI," Network Computing, 2019. [Online]. Available: https://www.networkcomputing.com/wireless-infrastructure/5g-future-will-be-powered-ai.

[28] R. Forbes, H. Wang, F. Feisullin, Y. Wang, K. Sylwia, and S. Liu, "ETSI ISG ENI - Defining closed-loop AI mechanisms for network management," *ITU-T FG-ML5G Work.*, 2018.

[29] V. P. Kafle, Y. Fukushima, P. Martinez-Julia, and T. Miyazawa, "Consideration On Automation of 5G Network Slicing with Machine Learning," *2018 ITU Kaleidosc. Mach. Learn. a 5G Futur.* (ITU K), pp. 1–8, 2018.

[30] O. Dharmadhikari, "Leveraging Machine Learning and Artificial Intelligence for 5G," CableLabs, 2019. [Online]. Available: https://www.cablelabs.com/leveraging-machine-learning-and-artificial-intelligence-for-5g.

[31] G. Domingues, J. Cabral, J. Mota, P. Pontes, Z. Kokkinogenis, and R. J. F. Rossetti, "Traffic Simulation of Lane-Merging of Autonomous Vehicles in the Context of Platooning," 2018 *IEEE Int. Smart Cities Conf. ISC2 2018*, no. March 2019, 2019.

07 A recommender system for service providers' campaigns

Diogo André Cadima Alves, Altice Labs
diogo-c-alves@alticelabs.com

Maria Manuel Castro, Altice Labs
maria-m-castro@alticelabs.com

Ricardo Ângelo Filipe, Altice Labs
ricardo-a-filipe@alticelabs.com

Achieving a complete customer characterisation to attribute better and customised products is a trend on the rise. This information filtering process, known as "recommendation system", increases companies revenues, and improves customer quality-of-experience. Our main goal is to understand the feasibility of a recommendation approach in one particular scenario: recommendation of campaigns. We aim to determine the extent to which it is possible to characterise the customers, using implicit feedback and state-of-the-art recommendation algorithms.

Keywords

Recommendation Systems; Machine Learning; Service Provider; Analytics; Advertising Campaigns

Introduction

Recommendation systems are information tools tailored to deal with information overload by suggesting items that are likely to match customers' needs and preferences. They have become a trend in the analysis and creation of customised profiles that are customer-oriented. Finding viable products and information and understand what is useful and relevant to a specific customer are key aspects to increase companies trustworthiness and revenues, decrease churn rate and ultimately increase the customer quality-of-experience.

Hence, recommendation systems take an essential role in nowadays companies and are vital to improve customer loyalty [1]. Furthermore, they help to increase the companies' product sales and create personalised advertising, inferring what product to advertise to the customer, according to their preferences and needs.

In this article, whose work is part of the scientific research article presented in the 19th EPIA Conference on Artificial Intelligence [2], we describe a recommender system applied to service provider's advertisement campaigns. This system uses historical data of customers reflecting their previous subscriptions and also data concerning customers' characteristics in terms of their behaviour and personal information. For customers that already had past subscriptions to campaigns, we applied collaborative filtering algorithms to determine what are the most suitable campaigns for them. To the customers that had not joined any campaign yet, we cross the customers' characteristics data with historical data of other customers to obtain the recommendations.

Our experiments show that the system can accurately infer recommendations for customers that purchased products or services in the past and also for the customers that don't. The results obtained show the feasibility of using recommendation algorithms to do personalised advertising. Furthermore, this study reinforces the possibility of having customer characterisation even without explicit feedback concerning the products proposed to them.

The following sections describe the methods we applied to the problem we tackle in this article and present and evaluate the meaning of the results, the strengths of this approach and its limitations. In the final section, we also describe future directions.

Proposed methodology

In this section, we explain how we built the data sets and the method used to obtain the recommendations.

Service providers' advertising campaigns

A service provider relies on advertising campaigns to increase its revenues and loyalty of its customers. The company sends notifications via SMS, interactive voice response (IVR) or e-mail to the customers advertising its products and services. Having received these notifications, the customers choose to join the campaign or not. Depending on the nature of the campaign, a customer can apply to the same campaign several times within a time frame, if the events that trigger the notification of the campaign occur. This process consists of the information we use to build the recommendations for our system.

Data used for recommendations

In this study, we used two types of customer data, their history of subscriptions to campaigns and their characteristics as customers of the company.

In the context of a service provider's campaigns, customers do not express their preferences in the form of ratings or likes, such as in Amazon [3]. Thus, providing campaigns' recommendations without this type of explicit feedback can be a difficult task. Regarding the historical data of customers, we transform the implicit feedback, specifically, how many times a customer joined a campaign, into explicit feedback. For this, we computed the ratio of the number of subscriptions to the number of notifications received by each customer. This operation allows us to obtain a numerical value that expresses the customer's interest in the service, i.e., a rating value for every campaign the customer received. We do not consider a binary value reflecting that the customer had joined the campaign or not because we want to distinguish the overall acceptance of each customer to a specific campaign. For instance, a customer that was notified three times and joined once is different from a customer who was notified fifteen times and also joined only once. With this ratio value, we have a more fine-grained idea of how interested a customer might be in a campaign.

This data set is very sparse since the majority of customers join very few campaigns, which is also a reason why a recommender system, able to provide more adequate campaigns for each customer, can be very advantageous for a service provider.

Recommendations

Here, we describe the process of obtaining the campaign recommendations for the two types of customers the system deals with: customers that joined campaigns in the past and customers that didn't. The collaborative filtering approach applies to the first type, while for the second type, we use customers characteristics and historical data.

Collaborative filtering

Recommendation systems use collaborative filtering algorithms for providing product recommendations based, solely, on the customers' history of purchases, searches or subscriptions in this case. Typically, these algorithms recommend to the customer products that similar customers had shown interest. The similarity is given by the common products that customers liked or bought. That is shown in **Figure 1**.

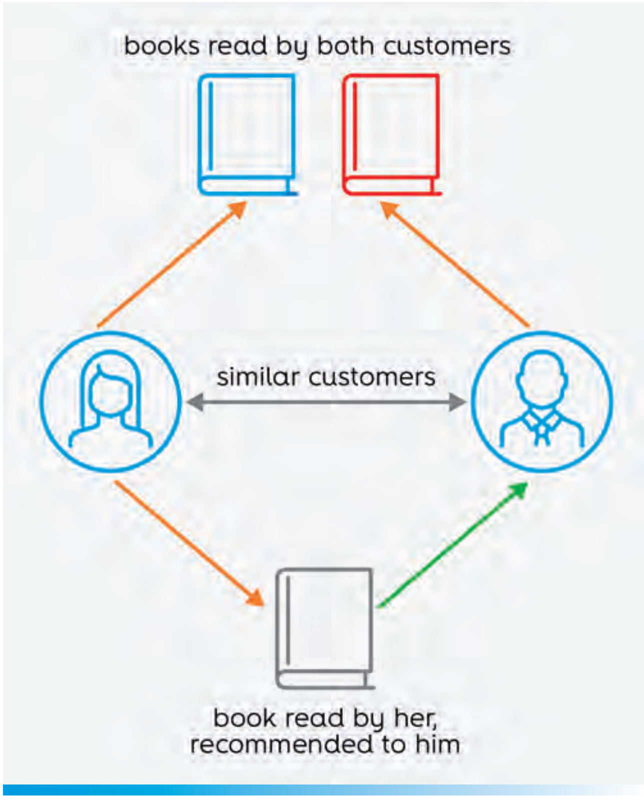


FIGURE 1 – Collaborative filtering

We use the data set of ratings mentioned before in this approach, dividing it into a training set, for training the collaborative filtering algorithms, and a test set, to evaluate the output obtained. This output consists of a list of campaigns, ordered by the predicted rating the algorithms calculated for the customer.

To evaluate this approach, and thus to understand the trustworthiness of the recommendations algorithms on predicting a rating value that a customer would have given to a particular campaign, we resort to several metrics. These are the mean absolute error (MAE), and the root mean squared error (RMSE), that measure how close the system predicted ratings

	predicted rating >= threshold	predicted rating < threshold
true rating >= threshold	true positive	false negative
true rating < threshold	false positive	true negative

TABLE 1 – Classification of predicted ratings and true ratings for a specific threshold

were from the true ratings given by the customers. We also use classification accuracy metrics like precision, recall, F1-Score [4] and specificity to evaluate the correctness of the predictions. These metrics have in consideration the number of true positives, false positives and false negatives present in the list of recommendations. For this kind of denomination, we defined a threshold corresponding to the value above which a rating is considered to be positive or negative, i.e., if the rating reflects the subscriptions of the campaign. The classification in true positives, false positives, false negatives and true negatives was done according to the table and respective threshold present in **Table 1**.

We also used the mean reciprocal rank (MRR) because we want to analyse the rank of the highest-rated campaign in each users' list of recommendations. The average ranking position in the recommendation lists of every campaign is also calculated to see in which place a campaign is ranked on average, considering several customers recommendation lists.

The algorithms used to test our data set are algorithms of collaborative filtering, namely matrix factorisation and baseline estimators, among others [5].

Customers characteristics

For customers that do not have expressed their preferences on any product or service, collaborative filtering recommendation systems cannot provide recommendations. The reason is that the system does not have enough information on that customer's past to determine what products he is interested in and what are his most similar customers. In the context of collaborative filtering algorithms, this is called the cold-start problem.

To solve this problem, we use customers characteristics to get recommendations for customers that do not have historical data. We do this by applying clustering algorithms to the data set of customers characteristics to obtain groups of customers. The clustering algorithm assigns customers with similar characteristics to the same cluster. The rationale behind this approach is that similar customers will have similar preferences in terms of the products and services they like. Thus, we can recommend them the same campaigns that similar customers have joined.

A customer receives a list of recommended campaigns, depending on the cluster assigned to him, and the campaigns he was notified. The list of campaigns is sorted by a numerical score which reflects the likeability of the customer to join that campaign. This score has in account the campaign's popularity in the cluster, as well as it's representativeness. The campaign's popularity in a cluster is calculated by the average rating given by the customers in that cluster. The campaign's representativeness, which must reflect how well each campaign is represented in each cluster, is obtained by the subscriptions' ratio times the notifications' ratio. The final score of a campaign, which decides the position that they are going to be recommended, is obtained by multiplying the popularity and the representativeness values.

The goal of this second approach is to be able to generate recommendations for the customers that don't have historical data. However, this approach can also give recommendations to customers that already joined campaigns in the past. For these customers, the list of recommended campaigns they receive consists of the most popular and representative campaigns of their cluster, except those that they were notified of but did not join. The reason is that we do not want to bother customers with campaigns that they already know and chose not to join. Therefore, this second approach can be complementary to the first one, or it can be used as the only way to generate recommendations to the customers.

Preliminary results

In this section, we present the results obtained for the two approaches. The results are preliminary because the second approach is yet to be evaluated by a proof-of-concept test with a service provider.

Approach with collaborative filtering algorithms

The analysed collaborative filtering algorithms were singular value decomposition (SVD), SVD++, NormalPredictor, BaselineOnly, SlopeOne and CoClustering, which are implemented in Surprise recommender library, available in the prediction_algorithms package [6].

The algorithms were evaluated with the metrics indicated previously. The threshold mentioned in the Collaborative filtering section was tested for several values since this threshold can vary according to what rating value the service provider considers to reflect the interest in a campaign.

The values for some of the mentioned metrics, computed with a threshold value of 7.5, are shown in **Table 2**. We can observe that in terms of the error metrics RMSE and MAE, the BaselineOnly and SVD algorithms show slightly better results than the other algorithms, with predicted ratings deviating on average from true ratings approx. 2.5 and approx. 1.85, for RMSE and MAE, respectively, on a scale of 0 to 10.

Regarding the F1-Score metric, which combines precision and recall, it also shows good results, meaning that the predicted ratings correctly reflect the behaviour expressed in the corresponding true ratings, whether that behaviour means the customer is interested in the campaign or not. The specificity metric of the BaselineOnly and SVD algorithms show the best results, which tells us that they are more capable of identifying customers that should not get notifications about specific campaigns. Values for the MRR metric are all very similar, which implies that the customers' top-rated campaign is occupying roughly the same position in the recommendation lists of each customer. The exception is the SVD++ algorithm that places this best-rated campaign in a much higher position on the list.

algorithm	RMSE	MAE	Precision	Recall	F1-Score	Specificity	MRR
BaselineOnly	2.518	1.888	83.9%	81.7%	82.7%	63.2%	0.0286
SVD	2.558	1.908	82.8%	82.5%	82.7%	59.9%	0.0287
SVD++	2.762	2.312	73.4%	94.1%	82.5%	20.6%	0.2176
CoClustering	3.143	2.537	72.6%	89.9%	80.3%	20.7%	0.0132
SlopeOne	2.959	2.433	73.6%	91.1%	81.4%	23.4%	0.0261
NormalPredictor	3.769	2.864	70.1%	58.8%	63.9%	41.3%	0.0299

TABLE 2 – Metric values of the different algorithms

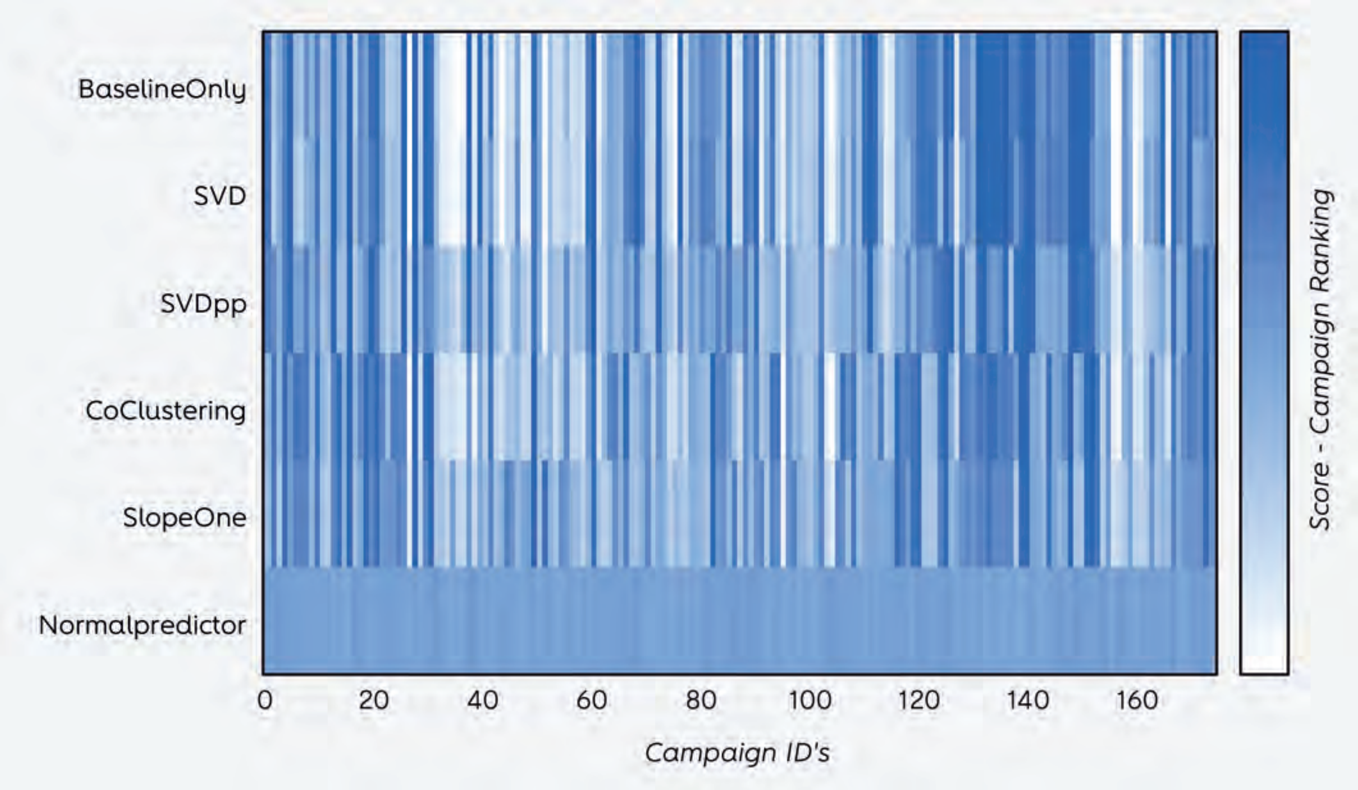


FIGURE 2 – Heatmap for all the campaign rankings and different algorithms

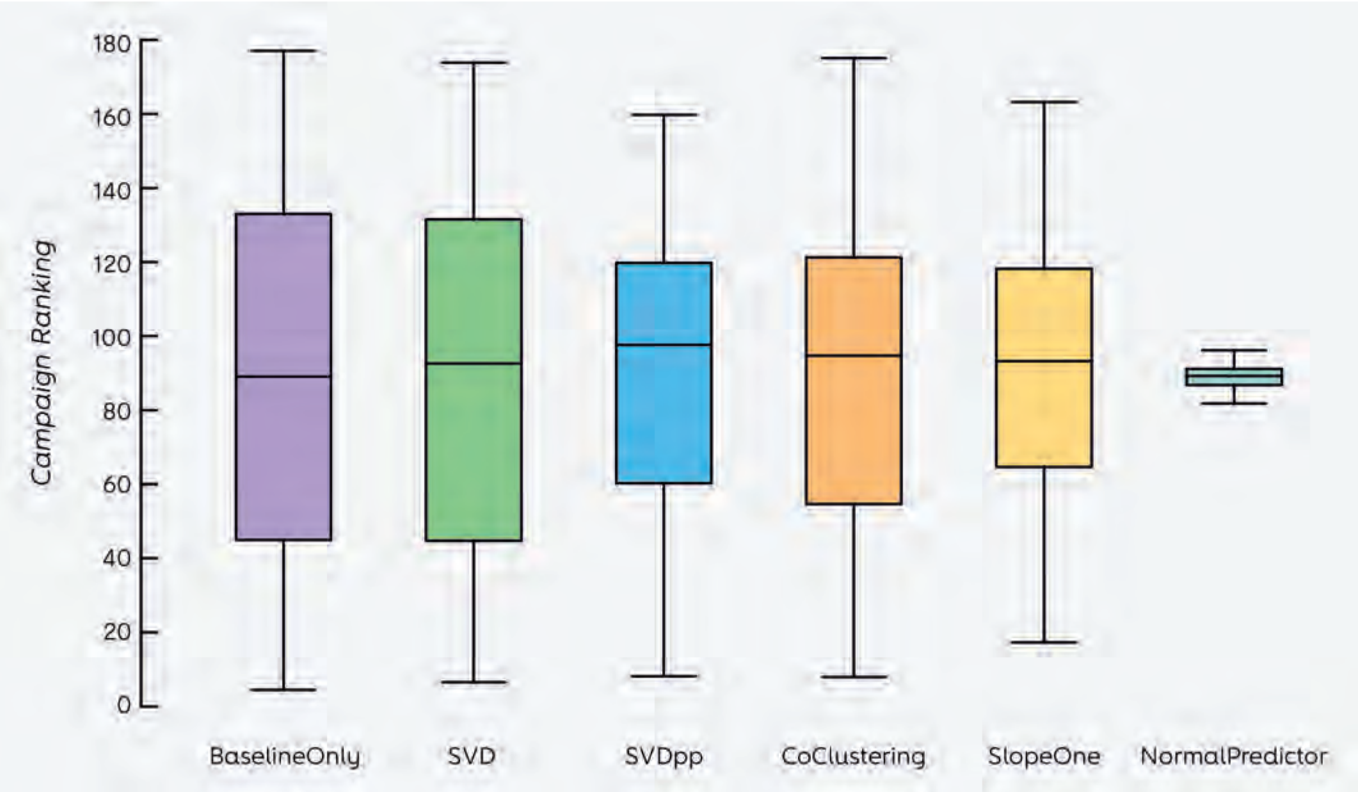


FIGURE 3 – Box plot for campaign ranking of different algorithms

As the customers receive a ranked list of campaigns, we analysed the dispersion of that campaign ranking for all the algorithms and presented the results in the heatmap graphic shown in **Figure 2**. The graphic's X-axis represents all the 177 campaigns this system deals with. The score bar represents the campaign ranking, with more saturated colours corresponding to the lower values for the ranking, i.e., the best-positioned campaigns. The graphic's Y-axis contains the six algorithms mentioned before. There are very saturated vertical lines in the graphic, which means that, for different algorithms, the same campaigns are recommended in top customers' recommendations lists. For example, around campaign 150, there are several vertical lines, meaning that some campaigns are highly recommended.

The box plot of **Figure 3** complements the analysis of the heatmap of **Figure 2**. These plots give us information about the range and distribution of the score values of the adapted ranking metric. We can see a similar behaviour between the algorithms, with the exception of the NormalPredictor algorithm, which is in compliance with the heatmap. The minimum value for the boxplot is always very low, regardless of the algorithm (exception made to the NormalPredictor) demonstrating that some campaigns are always appearing in the top recommendations.

We also measured the training and testing performance of the system. The NormalPredictor algorithm has the fastest training phase, but, as shown by the heatmap in **Figure 2**, this algorithm does not generate good recommendations. As shown in **Figure 2** and **Table 2**, the best algorithms are BaselineOnly and SVD. They present a good model training time performance and also have good results concerning our evaluation metrics.

Approach with customer characteristics

Before the recommendation phase, we evaluated the clustering model using some popular methods to execute this task, such as the elbow method [7] and the Davies-Bouldin metric [8]. The elbow method measures the intra-cluster variation, i.e., the distance of every sample in the data to their corresponding cluster centroid. The Davies-Bouldin metric measures the similarity between clusters, considering the intra-cluster and inter-cluster distances. With these two methods, we can infer the optimal number of clusters to choose for our clustering model, specifically for the data we used.

For evaluating the recommendations obtained with this approach, we intend to conduct a proof-of-concept with the service provider. This consists on notifying the company's customers with the recommended campaigns, and measure its performance, i.e., if the customers are joining the campaigns they are being notified of, and therefore, generating profit to the company.

As an example, **Table 3**, on next page, shows the campaigns considered for the recommendations of customers in one of the clusters. In this example, we only used a subset of six campaigns.

The columns represent. - notifications and represent. - subscriptions indicate the representativeness of the campaigns, in relation to notifications and subscriptions, respectively. These columns derive directly from the number of notifications and number of subscriptions columns, consisting of a ratio of the respective class (notifications or subscriptions). The mean rating column indicates the mean rating values given by the customers in that cluster. This value represents the popularity of the campaign.

As mentioned before, we obtained the score by considering the representativeness of the campaign and its popularity. The score indicated in the table uses a scale from 0 to 1. So, the top one campaign, the one with the highest score, represents the best recommendation for the customer in that cluster. If that customer already received notifications for that campaign and did not subscribe it, it is recommended with the next one on the list, and so on.

campaign	number of notifications	number of subscriptions	represent. - notifications	represent. - subscriptions	mean rating	score
campaign 1	412	280	0.12022	0.84337	7.28870	0.99129
campaign 2	1658	14	0.48380	0.04216	0.10035	0.00274
campaign 3	988	15	0.28829	0.04518	0.1567	0.00272
campaign 4	150	13	0.04377	0.03915	0.89655	0.00206
campaign 5	35	10	0.01021	0.03012	2.85714	0.00117
campaign 6	184	0	0.05369	0	0	0

TABLE 3 – Example of popular and representative campaigns in one of the clusters

Conclusion and future work

In a world with a plethora of products and services, it is a herculean task for a customer to identify good offers for his needs. Hence, recommendation systems have an essential role in ensuring the best experience and quality-of-service.

With this study, we explore the possibility of recommending advertising campaigns to the customers of a service provider. These campaigns will advertise products or services more suited to the customer's needs and interests.

The system bases its recommendations on collaborative filtering and clustering algorithms, using customer-related data, such as subscriptions history to campaigns and personal characteristics. We considered these two approaches because collaborative filtering may suffer from the cold-start problem, and the customer characteristics approach can address that.

We analysed and evaluated distinct state-of-the-art algorithms. Our results allow us to infer the best one to use for recommendations for customers that have historical data. We based this decision in metrics evaluating the quality of the recommendations and the performance of the algorithm. The recommendations obtained from the clustering approach, aimed at the customers that did not join any campaign in the past, will be evaluated in a real-world scenario, with real customers being notified with their recommended campaigns. This evaluation step consists of a proof-of-concept being executed alongside with the service provider that provided the data.

Although the second approach of this study is yet to be completed, we believe in the methodology's feasibility for our recommender system. Overall, we acknowledge the advantages of recommender algorithms applied to a service provider's advertising campaigns but also recognise the challenges required to obtain good recommendations with the data we have available.

References

[1] F. Ricci, L. Rokach, and B. Shapira, "Introduction to Recommender Systems Handbook," in *Recommender Systems Handbook*, Springer, 2010, pp. 1–35.

[2] D. Alves, B. Valente, R. Filipe, M. Castro, and L. Macedo, "A Recommender System for Telecommunication Operators' Campaigns," in *Progress in Artificial Intelligence*, 2019, pp. 83–95.

[3] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: item-to-item collaborative filtering," *IEEE Internet Comput.*, vol. 7, no. 1, pp. 76–80, 2003.

[4] K. P. Shung, "Accuracy, Precision, Recall or F1?," *Towards Data Science*, 2018. [Online]. Available: https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9.

[5] "Surprise prediction_algorithms package," 2019. [Online]. Available: https://surprise.readthedocs.io/en/stable/prediction_algorithms_package.html.

[6] "Surprise, a Python scikit for recommender systems," 2019. [Online]. Available: http://surpriselib.com/.

[7] R. Gove, "Using the elbow method to determine the optimal number of clusters for k-means clustering," 2017. [Online]. Available: https://bl.ocks.org/rpgove/0060ff3b656618e9136b.

[8] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224–227, 1979.


Smart Home
Home
22.6°
17
CLOUDY
45%
Monday
Tuesday
Tednesday
Music
Network
Security
Camera

08 The smart home: voice, machine learning and proactivity as innovation drivers

Bernardo Cardoso, Altice Labs

One of the most significant advantages of the current smart devices is that they are laser-focused on solving one problem. This characteristic allows the design and the UX to be really driven to provide a solution. That's fine if we only have one of these devices, but if we have more than a couple of those systems we need to deal with several different apps in the smartphone. This is where the cognitive layer could be much extended by collecting data from multiple sensors and systems and correlating them to create integrated models.

Keywords

Smart home; Smart devices; Wireless connectivity; Voice interface; Cognitive solutions

Introduction

The digitalization process and the resulting automation are pervasive forces in various dimensions of our lives; however, the impact hasn't been uniformly felt on all of them. In our jobs and in most industries, it is impossible not to notice a substantial evolution towards increased automation. Yet, in the residential domain, the dream of a fully automated and intelligent home usually pushed forward by technology visionaries, has resisted several attacks from different quarters, at least up to now.

According to the research report "Smart Home Market - Global Forecast to 2024" [1], the overall smart home market is expected to grow from USD 76.6 billion in 2018 to USD 151.4 billion by 2024. This research indicates that this market is expected to double in the next few years. The reasons and technologies responsible for that envisioned growth are significantly different from the ones proposed in science fiction. They have, however, been lingering in the market for a couple of years and are a lot more believable now.

In this paper, we will present some of the evolution that the smart home concept had until now, and the various efforts already put in place to fulfil this dream, most of them quite unfruitful. We will also present what we perceive as the driving forces that can crack the persistent unfulfilled promise of the smart home and the way Altice Labs products and services can contribute to the realization of this vision.

Initial efforts

About ten years ago, Altice Labs (at that time as PT Inovação) launched the Smart@Home initiative, whose ambition was to develop, create and certify home devices that could be used to build a credible smart digital home. We were not the first (nor of course the last) with this ambition. Recent history is full of forecasting attempts on the future of the smart house, particularly regarding what will be the equipment and services that will make homes even more intelligent. Genuine quests for automation at home started in the 3rd century BC when Philon of Byzantium created an automatic servant (robot) to mix and serve water and wine [2]. Since those times, most of the predictions for the home of the future have some kind of humanoid robot in mind, and the idea that the house will do more or less everything by itself [3], with these robots working as kind of slaves, unlocking huge blocks of free time and leisure to their owners.

Not following in the robot direction, our vision at that time was very focused on two elements considered central to a 21st century smart home: the Home Gateway, now the Fiber Gateway, which provided home-wide connectivity, and the set-top box, that together with a TV served as a home command and control centre. These devices were connected to an incipient set of other "home automation" equipment, which included connected plugs, that could remotely and selectively measure electricity consumption and turn devices on and off. We also supported video intercoms with TV integration, remote-controlled lamp switches, electric blinds, etc. Additionally, and with a more assisted living perspective, we integrated devices for measuring vital signs, such as weight scales and blood pressure meters.

But these connected devices didn't make a smart home. They just created a cluster of undeveloped technology with a tendency to give problems that, at times, instead of helping just bore. Even vital signs measurements are of little use if they are merely collected, without doing anything useful with them.

This vision for the digital home ended up having several problems. One of the most important issues was its ultra-centralist perspective, in which all equipment had to be connected to a single central system that served to control everything. More than that, it also anticipated that all those devices would have a standard way to communicate and to be accessed and remotely controlled. This vision never truly materialized. Besides, the implemented user experience (UX), although not intentionally, almost required the user to have super capacities, like being able to understand a lot of different parameters that must be configured. The user should also have the patience and ability to do these configurations despite intricate user interfaces.

This kind of scenarios, built like Legos and coming from the R&D of that time, were not commercially successful. Not even the large, professional, and fully integrated home automation systems, costing thousands of euros and installed from scratch on new buildings, were able to provide a simplified day-to-day life. It is very common to find remnants of these attempts in several residences, somewhat like industrial archaeology.

Inflexion point

In one aspect, however, we were right, Internet connectivity and its distribution inside the home, mostly thru Wi-Fi, was one of the keys to the resurgence of the smart home concept. In that sense, a Fiber Gateway having Wi-Fi 6 mesh capabilities, and a couple of extenders, like the ones developed in Altice Labs are, and will be, a key ingredient in the new smart home ecosystem.

Effective coverage of wireless Internet within the homes was just the start. Coupled with the onset of the smartphone massification era, it gave rise to an influx of smart home solutions focused on solving very specific problems. The first one, gathering a sufficiently high level of success in the United States, was the Wi-Fi-connected and smartphone-controlled Nest thermostat. Instead of trying to control and automate the whole house, Nest focused only on one thing, a smart thermostat for heating, ventilation and air conditioning (HVAC) systems. They claimed, not only an improvement on convenience, since it could be controlled from everywhere with an app, but also that the thermostat could pay for itself in less than two years due to energy savings.

With some of the same ingredients, Philips Hue is a line of colour changing LED lamps and white bulbs which can be controlled wirelessly. Again, in this case, we can see the increased convenience provided by the connectivity and a smartphone app, coupled with a promise of cost savings due to the energy consumption reduction. This pattern has been repeated a couple of times, for example, on the security area, with smart surveillance cameras and smart doorbells and locks. Likewise, in the health, wellbeing and fitness field there were many different offerings, both for the home and on the move, but following the same connectedness pattern.

Behind this increased adoption of smart devices are some driving forces that, we believe, will continue to lead this market in the upcoming years. The first one is strongly connected with the rising need for energy-saving and low carbon emission-oriented solutions. This justifies that the first successful products were related to energy savings like Nest and Philips Hue. The second force is a growing concern about safety, security, and convenience among young people, which increases the importance of home monitoring and justifies the rising success of surveillance and security devices.

The potential for energy savings and increased security is closely related to the smart part of these devices, that is, the smartness relies upon systematic data measurement and the capability to transform this data in smart actuation, mostly thru the usage of machine learning (ML) techniques. For instance, the Nest thermostat is a learning thermostat, that uses ML to model the personal usage pattern in a house and then, with that model and systematic temperature sensing, optimize the HAVC usage to maintain a comfortable ambient, while trying to reduce the energy consumption. It can also do the kind of things we keep forgetting about, like turning off the air conditioning (AC) before leaving the house. But, since it knows our routine, it can also turn the AC on automatically at the time we usually return home.

The same concept is used by the smart surveillance cameras, for instance, using ML to identify potential threats and signal or alarm the user, while also recording everything. The user can check in realtime if there is a real danger or if it's just the cat wandering around the house. Again, since this is a very competitive arena, the most advanced of these systems can already detect that it is indeed the cat, and simply record the occurrence, without generating any alarm.

This cognitive layer, on top of home devices, is a key ingredient not only in making them "smart" but also getting the devices' configuration and parametrization to a level that allows the design of the interface to be really simplified. This simplification allows the creation of minimalist applications on a smartphone with the goal of hiding all the complexity that lies behind the intricate set of algorithms that they need to work and only surface the bare minimum to the user interface.

Improvement opportunities

One of the most significant advantages of the current smart devices is that they are laser-focused on solving one problem and just one problem. This characteristic allows the design and the UX to be really driven to provide a solution. That's fine if we only have one of these devices, but if we have more than a couple of those systems, let us think of a Nest thermostat, a set of Hue bulbs and a smart doorbell, we need to deal with at least three different apps in the smartphone. Even with very well-designed apps, this will quickly get boresome.

So, one of the advantages quickly turns into a big disadvantage. However, since these are connected devices with a high reliance of cloud features, it also means that most of the devices are controlled and managed thru a set of APIs that, in some cases, are publicly available and very well documented. That fact allows the creation of systems like "if this then that" (IFTTT) [4], that enables the interconnection of disparate systems thru a set of simple rules. The problem here is that it represents the addition of one more item to the cogwheel, complicating the solution for the technologically challenged user.

This is where the cognitive layer could be much extended by collecting data from multiple sensors and systems and correlating them to create integrated models. For example, the Nest thermostat, instead of turning the AC on, based only on its internal model, could also have updated travel information for the house inhabitants, and just turn the AC on when one of these persons is approaching the house. In the same way, the Fiber Gateway could turn down the Wi-Fi when there is no one at home, and no devices using the network, and automatically turn the Wi-Fi back on when one of the households gets home. When applied systematically, this kind of algorithms can simultaneously improve security, besides reducing power consumption, by hiding the network when not needed. This is one opportunity still open in the smart home devices domain, and one that a platform like Altice Labs DataPlaxe could reap.

Introducing voice

As discussed, the current success of smart home devices has relied on a few ingredients: wireless connectivity, smartphones, simple mono function devices, cloud computing and machine learning. However, some of these ingredients are a double-edged sword. The biggest issue is the proliferation of different apps a user needs to navigate to control all of his devices. In the past, different attempts to create unified interfaces mostly aimed for the lowest common denominator UX, with low success rates.

A set of recently introduced devices, called smart speakers, could represent one way that might, at least partially, solve this unified interface issue. Smart speakers are wired or wireless connected speakers powered by a virtual assistant, which is driven by artificial intelligence. Initially, these devices were somewhat focused on playing music and answering simple questions, like “what will be the weather?”, or doing simple activities thru utterances like “set the alarm to go off in 10 minutes”.

The first smart speaker to make a splash in the market was Amazon Alexa. The strategy Amazon took with Alexa actively promoted the creation of “Alexa Skills” [5] by 3rd parties. A substantial slice of these skills was intended to manage smart home devices. In a short time, American users started to control their Nest thermostats and their Philips Hue LEDs by just speaking to them. The utterances where not very complicated, usually in the format “Alexa, turn on living room lights” or “Alexa, set the temperature to 70 degrees”, like in **Figure 1**.

The smart speaker addition to the home ecosystem represented, perhaps, the first-time users could use a simple interface to control disparate devices, each one provided by different suppliers. The user needed to learn the command structure to communicate with Alexa, but besides that, he was just required to use his voice. And voice is just voice, there was no need for an additional app, no need to learn how to use the app; the only thing needed was to talk to the smart device.

Amazon is not the only actor on this market, Google, with its Google Assistant, is pushing very hard, and the race is on. Unfortunately, both companies are competing in a winner-take-all market approach, trying to dominate the market, and Apple or Microsoft are lagging far behind. However, on top of the smart speaker resides the virtual assistant, and we believe that there will be space for specialized voice-enabled bots. These bots could piggyback on existent voice hardware, like Alexa Echo, Google Home or other smart devices. They will not focus on the voice recognition part but will add value and functionality in a process that needs more than a simple utterance, that is, they can specialize in knitting more complex dialogues.

It’s on this niche that Altice Labs’ BotSchool could help develop solutions like smart helpdesk for almost any kind of enterprise. The same platform also allows users to fully control a complex system like a Pay TV solution just by voice. That’s what Altice Labs and the University of Aveiro are doing in the CHIC project [6]. The voice capability coupled with additional machine learning frameworks is another key element on this new smart home ecosystem.

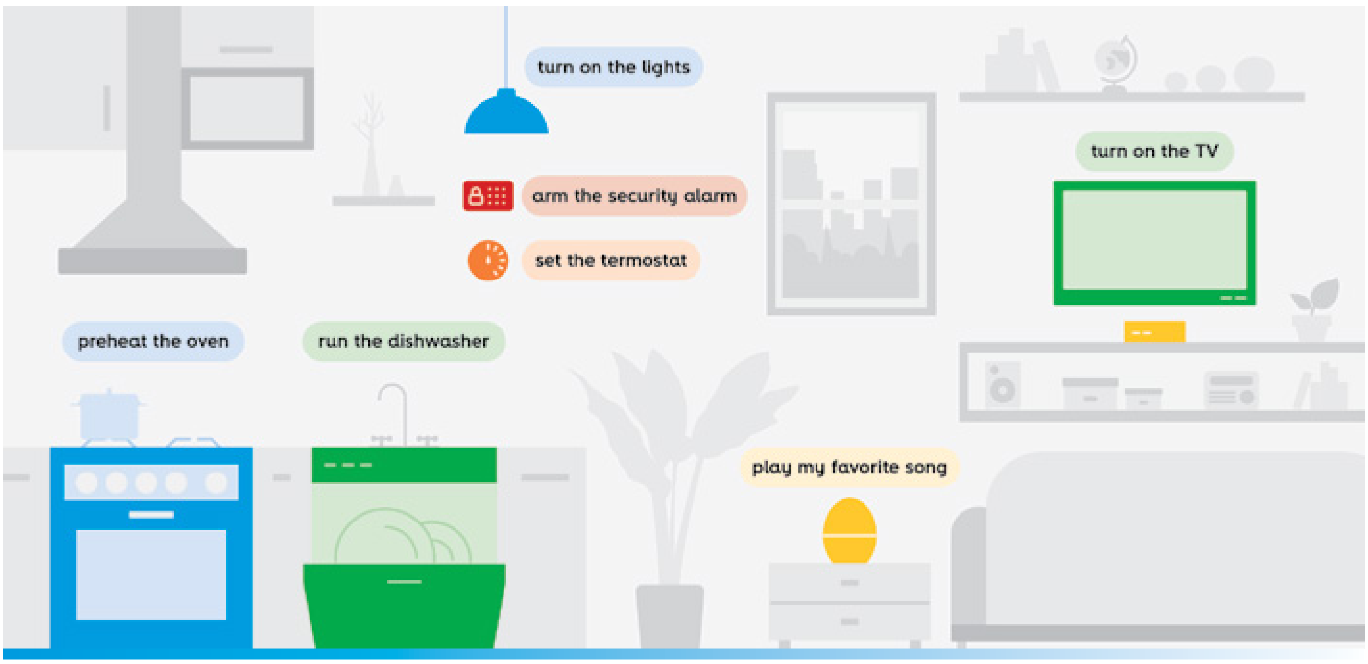


FIGURE 1 – Controlling the smart home with voice [13]

Potential for proactivity

However, using voice just for the sake of it only creates a new interface. Speaking and commanding a device through voice appears to be a leap in smartness, but there is no substantial breakthrough here. It is not smarter to say "turn on the TV" than to press the "on" button on the TV remote. It may be more accessible, more transparent, conceptually closer to human life, but it does not add intelligence. The basis for automation already existed, i.e. the TV was previously prepared to be turned on and off remotely. Being honest, it's really hard to create the integration between the TV manufacturer automation API and the virtual voice assistant platform, but it's just hard work. What would be truly smart would be for the television to turn itself on, at the right time, without the user having to do anything.

Oncken and Wass, in their seminal article in Harvard Business Review [7], described five levels of initiative that a collaborator can have concerning their leadership:

1. Wait until told (lowest initiative);
2. Ask what to do;
3. Recommend, then take resulting action;
4. Act, but advise at once;
5. Act on own, then routinely report (highest initiative).

Except for the firmest supporter of micro-management, the remaining bosses and most of us within our own homes would love our employees (or robots) to take a little more initiative. So, this same scale of initiative or proactivity could also be applied to the smart devices or smart home solutions.

The problem is that most home systems, even the most integrated and even those with voice interfaces, are at the initiative level 1. They are always waiting to be told what to do, sometimes reaching level 2, but not going beyond that.

In other contexts, nonetheless, automation has already reached additional levels of initiative. For example, when digital video recorders (DVR) where introduced, the user had to tell the system which and every program he wanted the DVR to record. It did that by going to the electronic program guide and scheduling the recordings (yes, before that there were much more archaic systems involving VHS cassette recorders, but this was on another century and practically no one could use them anyway). The problem with scheduling recordings is that the user had to remember to plan for things he wanted to see. If there was something he didn't know was airing, he would never have the chance to schedule the recording. However, using technology and cost improvements in storage, it was possible to create a solution to the scheduling problem. With the introduction of full line-up catch-up TV (in Portugal called Automatic Recordings), it was achievable to go from level 1, where the user had to say precisely what he wanted to record, to level 2, where the system records everything and asks the user which program he wants to watch.

Yet, with the increasing number of channels and programs available on catch-up TV, choosing what to watch has also become a problem. We are now taking the first steps towards moving to initiative level 3. Based on the TV consumption of the user, and using ML techniques and predictive analytics, the system makes a recommendation of potentially interesting things for the user to see. It is not perfect yet, but since it is constantly evolving, it may even be possible to reach level 4 with these techniques. When this level is attained, the system will automatically select the television program that the user wants to see and, also automatically, begins to play it when the user sits in front of the TV.

Going back to the beginning of the example: smart is not the user saying, "turn on TV", smart is the user sitting on the couch and the television turn on automatically and start playing the most appropriate content to the moment and the assistance.

But this kind of cognitive approach, with increased levels of proactivity, can also apply to a host of other smart home situations, like the one already presented for Nest thermostats. Inside of the home, advanced Wi-Fi signal processing, powered by ML can recognize if someone is at home, if it is moving, or even identify him and decide that it could be an intruder. This cognitive layer can also be used to evaluate sets of indicators and parameters in order do a lot of things at the 3rd initiative level, for instance, to remember the user to buy milk or other groceries. In time, when we reach the 4th or 5th level, the system will buy the milk automatically and just report that to the user. In the same way, the smart home can predict equipment failures and plan maintenance activities. Some of these solutions are already being implemented at Altice Labs, to predict precisely this kind of situation for telecom operators, but this is what will be expected from a proactive smart home system (see **Figure 2** for several examples of the cognitive layer applied to the smart house).

Security and privacy concerns

As we can see, ML is the key factor for improving the proactivity level, and to reduce how much tweaking the user needs to apply to his smart home devices. However, all of this intelligence doesn't materialize out of thin air. For ML to work, it needs to collect a massive amount of information. More than this recollection, the data needs to be sent to a centralized place ("the cloud") to be processed and to create the needed models. This brings a significant amount of security and privacy concerns.

When people do notice that they are being monitored, it's clear that there is an increase in care about data privacy. A recent study from the University of Oxford shows that "*the overwhelming majority of Americans (82%) believe that robots and/or AI should be carefully managed. This figure is comparable to with survey results from EU respondents.*", and that the

FIGURE 2 – Cognitive layer applied to the smart house

topmost AI governance challenge was "*preventing AI-assisted surveillance from violating privacy and civil liberties*" [8].

A recent article, in the front page of The New York Times, raised awareness for the location tracking some smartphone apps are doing, and the companies that are profiting from it [9]. Although this data is anonymously collected, it's easy to connect it to a single person, like the article showed, because most of our routes are quite individual. From simple things like seeing where the smartphone spent the night and where it spent the day is possible to pinpoint with a high probability the correct person. Other apps, like this summer's hit FaceApp, have very murky privacy policies or openly sell the user data for targeted advertising [10].

Taking some more hints from additional public awareness on data breaches and privacy violations, like Cambridge Analytica case [11], and complementing them with breakthroughs in cryptography and deep learning, there's a perception that it's time to invest in privacy-preserving ML. Legislation like GDPR or, for instance, its Chinese equivalent, is a step in the direction of making all this data collection more transparent for the end-user [12], and also promote solutions where the data doesn't need to be transferred to 3rd parties for the systems to learn and personalize their behaviour to the end-user. This could be a window of opportunity to leverage, not only privacy-preserving algorithms, but also the edge computing aspect built-in 5G proposals so that the smart house can keep evolving in the proactive direction.

But when systems reach initiative levels 4 and 5, we will also need some sort of higher-level control of these smart houses, like we already need for autonomous vehicles. That is, if systems start to perform actions by themselves, automatically, they need to present that to the user in a way that does not appear to be intrusive, and that can be easily counteracted. Not implementing this kind of failsafe solutions could mean that one morning we will wake with the house full of milk or, much worse, burning in a sauna.

Conclusion

In this article, we addressed the resurgence of the smart house, this time based on highly available wireless internet connectivity and supported mostly in smartphone apps. This reappearance was also driven by the rising need for energy-savings, low carbon emission solutions and the growing concerns about safety, security and convenience among young people.

These existing solutions, a lot better than some proposals in the past, are, however, still quite disconnected, with each manufacturer providing completely separated platforms and experiences. But the increasing introduction and integration of these systems, with virtual assistants driven by voice, is giving a unification layer to all these disparate solutions.

The challenge now is to add, to the smart house, a cognitive layer that could enable it to go beyond the reactive functions and give a proactive approach to our residential experience, by using the power of cloud computing and machine learning. Altice Labs products and services could be leveraged in designing an offer for this domain, for example, the Wi-Fi 6 Fiber Gateway, the BotSchool, the DataPlaxe and a host of cognitive solutions being infused in all of our products.

All of this smartness and proactivity can't forget some needed controls at the privacy and security level. With our increased reliance on automated systems, we are putting our lives in the hands of machines and on the companies that create and develop software for them. Some legislation like GDPR, although representing increased work and constrains for the product design, epitomise some of the controls that will be needed in the future and there's a lot of essential work yet to be done in this area.

References

[1] Markets and Markets, "Smart Home Market by Product, Software & Services, and Region - Global Forecast to 2024," 2019.

[2] Kotsanas Museum, "The automatics of Philon of Byzandium." [Online]. Available: http://kotsanas.com/gb/cat.php?category=04.

[3] The Ambient, "Smart home visions through the ages: The history of home automation," 2019. [Online]. Available: https://www.the-ambient.com/features/visions-through-the-ages-history-of-home-automation-178.

[4] IFTTT, "How does IFTTT work?," 2019. [Online]. Available: https://help.ifttt.com/hc/en-us/articles/115010158167-How-does-IFTTT-work-.

[5] Amazon Alexa, "Make Money and Reach More Customers with Alexa Skills," *Amazon Developer*, 2019.

[6] CHIC, "Cooperative Holistic view on Internet and Content," 2019. [Online]. Available: https://chic.mog-technologies.com/.

[7] W. Oncken and D. Wass, "Management Time: Who's Got the Monkey?," *Harvard Business Review*, 1999.

[8] B. Zhang and A. Dafoe, "Artificial Intelligence: American Attitudes and Trends," *Future of Humanity Institute, Center for the Governance of AI*, 2019.

[9] J. Valentino-DeVries, N. Singer, M. H. Keller, and A. Krolik, "Your Apps Know Where You Were Last Night, and They're Not Keeping It Secret," *The New York Times*, Dec-2018.

[10] H. Denham and D. Harwell, "FaceApp went viral with age-defying photos. Now Democratic leaders are warning campaigns to delete the Russian-created app 'immediately,'" *The Washington Post*, Jul-2019.

[11] I. Lapowsky, "How Cambridge Analytica Sparked the Great Privacy Awakening," WIRED, 2019. [Online]. Available: https://www.wired.com/story/cambridge-analytica-facebook-privacy-awakening/.

[12] W. M. Wenyan, "China is Waking up to Data Protection and Privacy. Here's Why That Matters," World Economic Forum, 2019. [Online]. Available: https://www.weforum.org/agenda/2019/11/china-data-privacy-laws-guideline/.

[13] Google, "Google no Twitter: "Now able to connect with over 5,000 devices for the home #GoogleAssistant is here to help you turn your house into a smart home" Twitter, 2018. [Online]. Available: https://twitter.com/google/status/992126515365187585.

09 HCI boosted by AI: from smart interfaces to immersive cognitive environments

Fausto de Carvalho, Altice Labs
cfausto@alticelabs.com

Hugo Paredes, UTAD & INESC TEC
hparedes@utad.pt

This article addresses the recent evolution of Human-Computer Interaction through the inclusion of technologies and features that envision the rise of immersive cognitive environments with seamless interaction. Exploratory research carried out by Altice Labs looks forward to this paradigmatic change, researching the market for potential exploitation of these technologies.

Keywords

Smart interfaces; Immersive environments; Artificial intelligence; Digital twins; Extended reality; Augmented cognition; Augmented humans

Introduction

The history of interaction is a convergence path, closing the gap between humans and machines. The first steps were taken by man adapting to machine communication constraints, using binary code and switches to establish a communication channel, but we've been witnessing enormous advances in bringing machines closer to common natural forms of interaction. The human interaction paradox is a huge challenge, not only for its richness and complexity, but also for the level of cognition required to understand it: there is a multiplicity of forms of interpersonal communication, ranging from spoken language to body and facial expressions, and multisensory capacity heavily enriches human interaction, nevertheless introducing extreme complexity, with countless possibilities.

Research in the area of human-computer interaction (HCI) has been progressing aligned with this perspective, exploring the context of application of interaction technologies and enriching the existing multidisciplinary knowledge. This multifaceted approach has allowed us to broaden the technological domain of intervention and evolve into topics such as natural language processing, motion capture and body language recognition, which, complemented by the most recent advances in artificial intelligence such as deep learning and machine learning, potentiated the emergence of smart interfaces.

Smart interfaces and immersive environments

Human interaction is continuous, deeply contextual and inherently multimodal. Verbal communication is complemented with body language and references or even interactions with one's own environment [1].

Natural interfaces aim to transpose this conversational richness into human-computer interactions, simplifying communication and exposing the natural spontaneity of interaction. This semantically rich and multichannel user-centred interaction is characterized by its broad-spectrum, combining voice, image and behaviour [2], whose perception is context-dependent. The need for user control of interaction, by making it user-centric, is crucial for the success of the multimodality of the interaction [3].

Natural interaction and the multimodality of interaction have been enriched by the growing interest of the scientific community and industry in conceptualizing innovative interfaces. The multimodality of interaction assumes a particular role in investigating interaction with intelligent software agents, as described in the study by Norouzi et al. [4]. On the other hand, Silva et al. [5], in a systematic literature review, identified 19 articles that present innovative trends in multimodal interaction. In this paper, the vast majority of identified articles fall into the category of natural interaction, including gestures, voice, vision, smell, and cognition. The evolution of natural interfaces includes the use of tactile surfaces that extend interaction to the environment and enable novel manipulation techniques [6].

On the other hand, advances in affordable devices for immersion and visualization, such as Oculus Rift, HTC Vive, Samsung Gear VR and Google Cardboard head-mounted displays (HMDs), leverage new environments and augmented three-dimensional (3D), virtual reality (VR) content or even mixed [7]. These new emerging environments open new challenges in interaction. Being based on traditional interaction techniques, these environments limit the tracking of users' movements and consequently restrict manipulation in the three-dimensional environment, especially in situations of free interaction in space where the accuracy of spatial interactions is a requirement. These limitations further constrain user performance [7]. Nevertheless, this technology has potential benefits in different fields, from education [7] to smart cities [8], for their collaborative and co-creation capabilities [9]. Mixed reality (MR) generates new types of user experience by enhancing a semantic understanding of the environment and consequently presenting itself as a solution to the complex challenges of interaction design [10].

The user experience provided by these environments is largely due to their immersibility. For Dede et al. [11] immersion is the subjective impression that someone is participating in a comprehensive and realistic experience. Schubert et al. [12] associate immersion with presence as a subjective sensation of being in the virtual place. The authors argue that presence is observable when people interact in the virtual world, claiming that presence is perceived as a result of the active interpretation of a virtual environment, arising when the possibilities for bodily action in the virtual world are mentally represented. Slater et al. [13] introduce the perspective of a first-person body transfer and argue that immersive virtual reality is a powerful tool in the study of body representation and experience.

In a recent study, Kang [14] highlights the positive impact of assisted natural interaction in virtual reality environments, increasing the level of empathy, efficiency and emotional effect. In this paper, the author evaluates various technologies that allow three-dimensional body interaction, allowing users to control the contents of the environment in a similar way to the real world by partially or totally moving their body. LaViola [10] refers that the investigation of interactions in 3D space has received considerable attention from the community, presenting several strategies to improve the accuracy of recognition of three-dimensional gestures, in particular broad sets of gestures, which can be associated with their context. The contextualization of interaction in three-dimensional environments leads to emerging concepts presented by Lee et al. [15]. Ubiquitous virtual reality (U-VR) is an emerging paradigm that assures the user applications according to their interaction context [16]. To ensure the pervasiveness of virtual reality environments, this paradigm is based on three pillars: collaborative mediated reality, wearable mediate reality, and context-aware mediated reality. Similarly, the concept of pervasive augmented reality is introduced by Grubert et al. [17] as "*a continuous, ubiquitous and universal augmented interface of information in the physical world*". Practical applications of this kind of applications are commonly presented as "*context-aware augmented reality*" [18], emphasizing the conjugation of augmented reality (AR), intelligent virtual agents, and the Internet of things (IoT) analyzed in the systematic review by Norouzi et al. [19]

In a nutshell, the convergence we see from research in the domain of human-computer interaction goes towards Engelbart's view of machines as enablers of human intelligence [1]: smart interfaces are not intended to automate tasks, but to improve human decision making by interacting with the real world through virtual interactions.

Exploratory research

Mouse and graphical user interface (GUI) paradigm deeply shaped the way we have been interacting with computers for decades (**Figure 1**), so in this world dominated by mice, keyboards and joysticks for games, all based on similar triggering principles for interaction, the release of Microsoft Kinect in 2010 was a truly disruptive innovation.

FIGURE 1 – Engelbart computer mouse prototype (1968) [25]

Our research in immersive 3D environments has made evidence of its remarkable potential in collaborative activities, but the classical GUI is quite limitative for user embodiment (avatar). Natural interaction is required for improved immersivity, ease of use and pervasiveness, and motion recognition provided by Kinect and its integration with virtual worlds in Online Gym project allowed the creation of a low-cost multi-user virtual gym prototype [20]. Tests conducted with real users shown the ease of interaction for the elderly and the improved socialization features, but also highlighted some significant shortcomings, namely in the recognition of fine gestures, the accuracy of recognition, depth and concealment.

Throughout recent years, advances in VR and AR technology catalyzed the emergence of a series of devices based on new acquisition paradigms, including computer vision, electromyography and gyroscopes. To address previously identified problems, project InMERSE explored the fusion and contextualization of interaction in immersive environments: fine gestures using Myo and LeapMotion devices with greater precision (**Figure 2**), complementing Kinect motion capture, plus context synchronization between immersive display devices, namely (by then state of the art) Oculus Rift and Google Glass. Furthermore, to achieve the technological objectives of the project, a multi-device framework was created allowing to abstract motion recognition and its application semantics, thus also mitigating technological obsolescence in an area where devices are coming and going at a fast pace. The software library for natural interaction was made available in open-source [21] and allowed the development of two demonstrators: a digital signage prototype and an interactive, immersive multi-user installation [22]. Although being a big advance relatively to most of the limitations found in previous projects, this approach is not yet a solution for widespread use cases because of the complexity and cost of the required mix of the devices.

More recently, a small revolution has been happening in the software side, with the

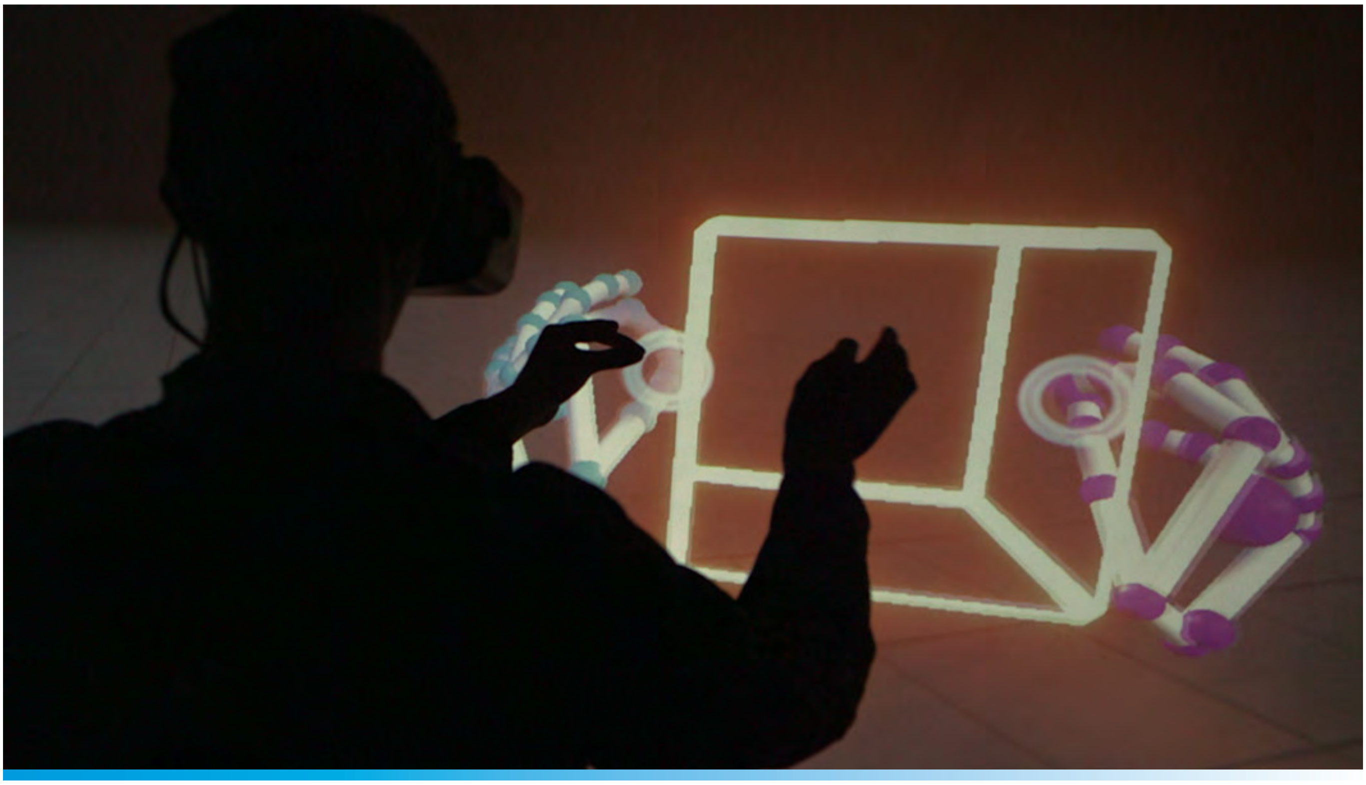


FIGURE 2 – LeapMotion tracking hands VR demo [26]

introduction of new toolkits for gestures recognition based on state of the art machine learning/deep learning techniques. ARaNI project explored these emerging technologies in AR/MR scenarios.

The SmartMirror prototype took advantage of the latest AI techniques for motion recognition using very low-cost devices and modest processing capabilities. The use of the OpenPose framework [23], allowed the creation of a multimodal "how to make a tie knot" demonstration without any special interaction device besides a simple cheap webcam (**Figure 3, top**). Gestures aside, user can also interact in plain voice, to guide the prototype flow while having his hands busy.

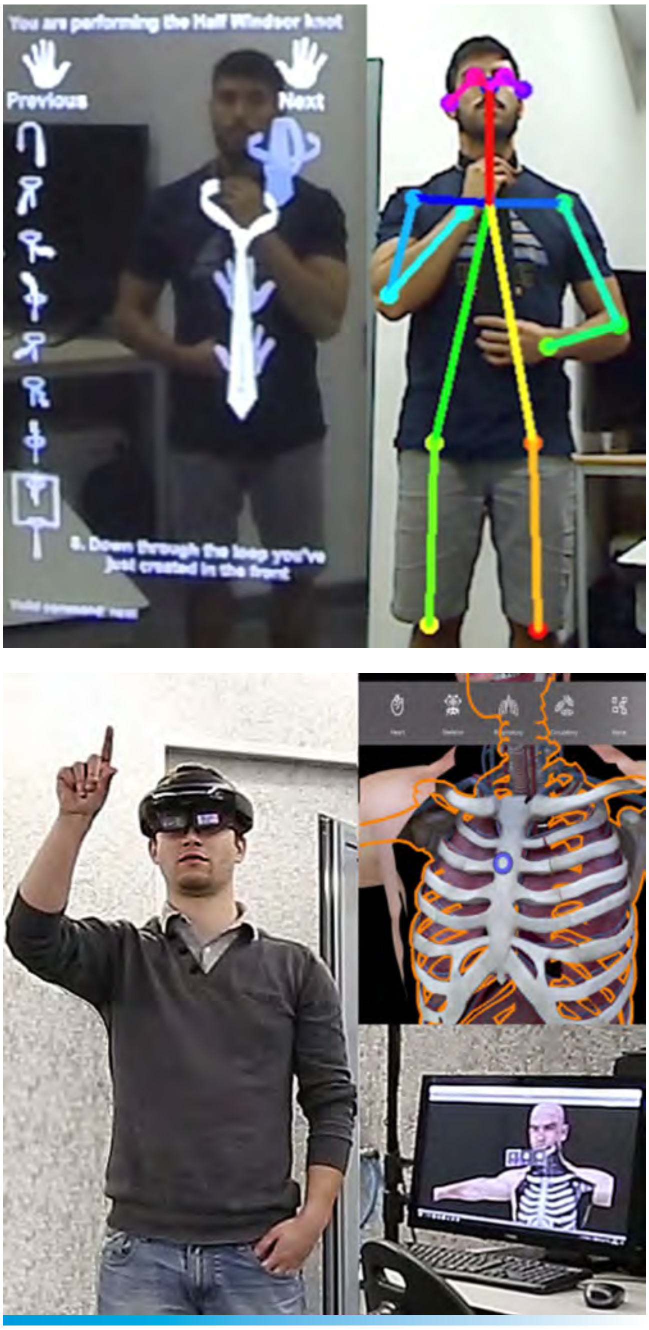


FIGURE 3 – AraNI prototypes

In another prototype, we explored Microsoft HoloLens mixed reality headset applied to a human body 3D model exploration demonstrator (**Figure 3, bottom**). In addition to using the device capabilities for multimodal interaction using gestures and voice, we also addressed several aspects of the collaborative interaction within the MR environment, allowing multiple users to interact with the virtual 3D model simultaneously (via HoloLens or a web interface, easily extendable to other devices).

In few years, we have definitely walked away from the limited and somewhat clumsy natural interaction in 3D virtual worlds requiring special devices. The convergence of several rapidly expanding technological areas is creating the opportunity for smooth shared eXtended Reality (XR, AR/VR/MR) collaborative interactive scenarios that we may anticipate being soon widely available, enriched with even more interesting capabilities.

Raising up augmented humans: a roadmap through seamless AI for HCI

Exploratory research carried on by Altice Labs in collaboration with University de Trás-os-Montes e Alto Douro yielded a set of outcomes that allow us to anticipate a roadmap for the potential of smart interfaces and some of the building blocks for their combination with extended reality environments.

Strictly at the interaction level, the use of AI is already happening, with the integration of deep learning techniques associated with image processing and gesture recognition, such as the powerful OpenPose library explored in the ARaNI project. This open-source real-time multi-person system allows very complex use cases, by jointly detecting human body, hand, facial and foot key points, up to a total of 135 key points on single images. Besides, it's based on TensorFlow [24], an end-to-end open-source platform for machine learning, originally developed by researchers and engineers working on the Google Brain team, which became a popular de facto standard with a comprehensive and flexible ecosystem of tools, libraries and community resources.

The potential for using AI techniques in interaction goes far beyond their application in computer vision. We can foresee AI facilitating interaction, with the ability to predict user actions and adapting to their habits, based on a comprehensive set of personalization aspects, usage information, and also the context of the interaction, using human-inspired techniques such as neural networks embedded into the interactive systems. In this way, it will be possible to make tasks simpler and consequently improving the performance and quality of the interaction, opening new horizons for assisting complex tasks, more prone to human error, but which require human cognition in the decision process.

This role of interaction facilitator reverses a more conservative view of technology as an obstacle for general access to knowledge and the information society development, becoming, in fact, a catalyst for technology's ability to scale up, due to the broadening potential of personalization, fitting each and every user, and as such facilitating education and contributing to a more inclusive (information) society. At this turning point, technology takes on its responsibility to act as a bridge to overcome communication and social interaction hurdles.

This process of interaction transformation is not complete without the mutual involvement of all actors in the course of action, and literature makes evidence of the role of immersion as a stimulus for human involvement. Virtual reality presents an opportunity for reinforcing this immersion, with the new real-world representation potential brought by the IoT sensorization.

Digital twins are a particular example of this merger, being accurate dynamic virtual representations of the physical entities, which are being applied in multiple areas, notably Industry 4.0. In these scenarios, virtual replicas allow to improve operations, increase efficiency or even predict problems. Furthermore, whole lifecycles can be adressed as early as in the design phase, and simulation of new functionalities or brand new products, before a physical prototype is built. Lessons learned in the virtual environment can then be applied in a real environment, already knowing its consequences (**Figure 4**).

Digital twins concept also means an evolution of virtual reality, bridging over the traditional real/virtual segregation, which offers total immersion on the virtual side but involves a clear boundary to the real environment. In this transformation, virtual reality is combined seamlessly with the real world, often giving way to augmented reality or even mixed reality realistic fusion, with enrichment coming from information formerly scattered among several computer applications or in purely virtual environments. The concept of smart everything anywhere, as a new dimension of the Internet of things, is a two-way portal between real and virtual that enables this new augmenting reality.

Looking ahead, we may expect a convergence of interaction technologies that have been following autonomous research and development streams, towards an integration in a unified interface. The materialization of this vision of pervasive, personalized and customizable mixed reality environments is strongly conditioned by the computing capabilities of mobile and wearable devices. On the other hand, moving the processing to high availability cloud architectures is conditioned by latency, with a major impact on the user experience. The deployment of 5G and technologies such as edge and fog computing, ensuring low latency of the network infrastructure and all the needed processing capacity, will

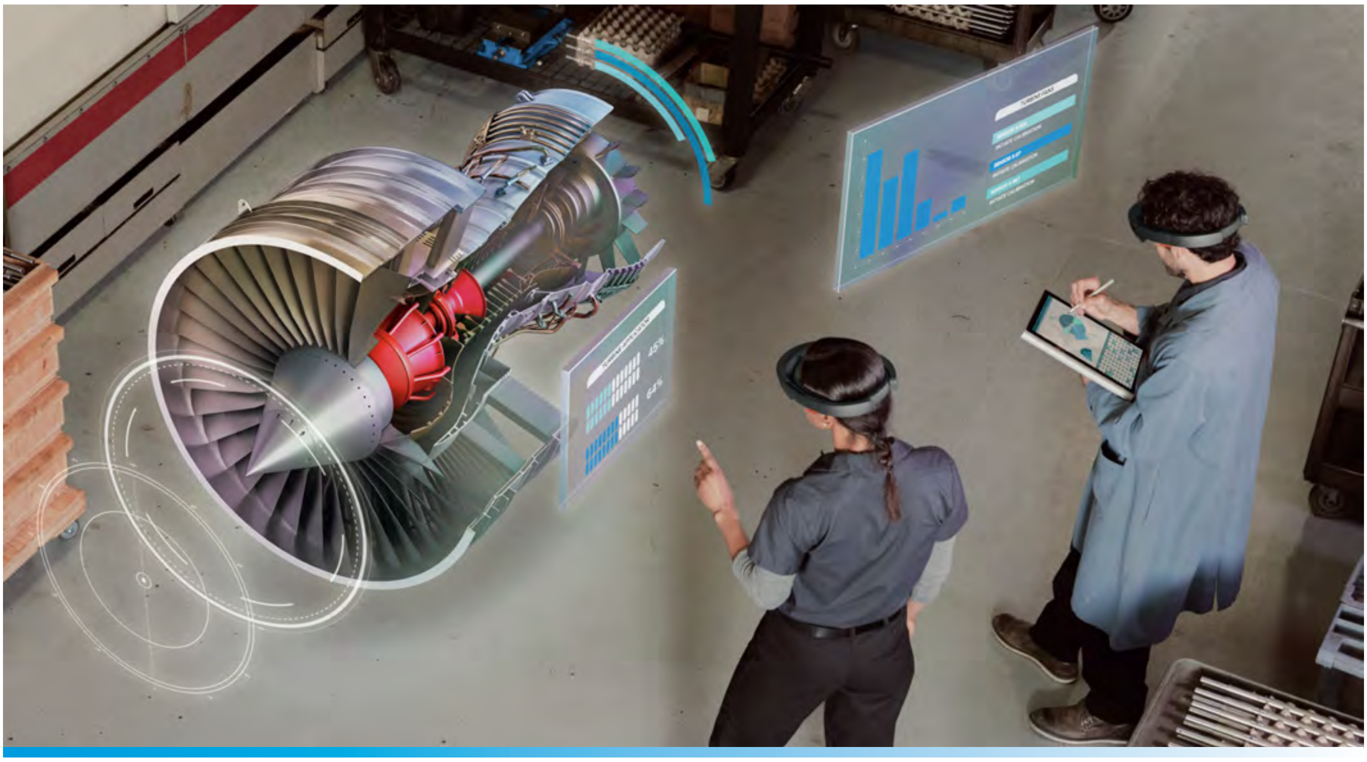


FIGURE 4 – XR in Industry 4.0 context [27]

definitely accelerate the introduction and contribute to market widespread of interesting and powerful mixed reality solutions.

Ultimately, technological conditions for experiencing rich mixed reality in pervasive environments will be met, and we will witness symbiotic man-machine cooperation, with high computational power augmenting human cognition and facilitating/enabling people to accomplish a wide spectrum of tasks. Engelbart's vision of technology as a catalyst for human capacity can be materialized in this hybrid interaction environment, with the imperceptibility between real and virtual allowing the genesis of Augmented Humans who, acting naturally and endowed with technology, are empowered in their full cognitive potential and decision-making power.

References

[1] J. Landay, "Smart Interfaces for Human-Centered AI," *The Stanford Institute for Human-Centered Artificial Intelligence*, 2019. [Online]. Available: https://hai.stanford.edu/news/smart-interfaces-human-centered-ai.

[2] W. Liu, "Natural user interface- next mainstream product user interface," *2010 IEEE 11th Int. Conf. Comput. Ind. Des. Concept. Des. 1*, vol. 1, pp. 203–205, 2010.

[3] D. Roscher, M. Blumendorf, and S. Albayrak, "A meta user interface to control multimodal interaction in smart environments," in *Proceedings of the 14th international conference on Intelligent user interfaces (IUI '09)*, 2009, pp. 481–482.

[4] N. Norouzi et al., "A Systematic Survey of 15 Years of User Studies Published in the Intelligent Virtual Agents Conference," in *Proceedings of the 18th International Conference on Intelligent Virtual Agents (IVA '18)*, 2018, pp. 17--22.

[5] T. Silva, P. Almeida, J. Abreu, and E. Oliveira, "Interaction paradigms on iTV: A survey towards the future of television," *IMCIC 2018 - 9th Int. Multi-Conference Complexity, Informatics Cybern. Proc.*, vol. 2, no. Imcic, pp. 18–23, 2018.

[6] D. Mendes, F. M. Caputo, A. Giachetti, A. Ferreira, and J. A. Jorge, "A Survey on 3D Virtual Object Manipulation: From the Desktop to Immersive Virtual Environments," *Comput. Graph. Forum*, vol. 38, pp. 21–45, 2019.

[7] R. R. Divekar et al., "Interaction challenges in AI equipped environments built to teach foreign languages through dialogue and task-completion," *DIS 2018 - Proc. 2018 Des. Interact. Syst. Conf.*, pp. 597–610, 2018.

[8] L. Morgado et al., "Cities in Citizens' Hands," *Procedia Comput. Sci.*, vol. 67, no. September, pp. 430–438, 2015.

[9] T. Kohler, J. Fueller, K. Matzler, and D. Stieger, "CO-creation in virtual worlds: The design of the user experience," *MIS Q. Manag. Inf. Syst.*, vol. 35, no. 3, pp. 773–788, 2011.

[10] L. J. Joseph J., "Context aware 3D gesture recognition for games and virtual reality," in *ACM SIGGRAPH 2015 Courses (SIGGRAPH '15)*, 2015, pp. 10:1--10:61.

[11] C. Dede, "Immersive interfaces for engagement and learning," *Science (80-.).*, vol. 323, no. 5910, pp. 66–69, 2009.

[12] T. Schubert, F. Friedmann, and H. Regenbrecht, "Embodied Presence in Virtual Environments," in *Visual Representations and Interpretations*, 1999, pp. 269--278.

[13] M. Slater, B. Spanlang, M. V. Sanchez-Vives, and O. Blanke, "First person experience of body transfer in virtual reality," *PLoS One*, vol. 5, no. 5, 2010.

[14] J. Kang, "Effect of Interaction Based on Augmented Context in Immersive Virtual Reality Environment," *Wirel. Pers. Commun.*, vol. 98, no. 2, p. pp 1931–194, 2018.

[15] Y. Lee, S. Oh, C. Shin, and N. Woo, "Recent trends in ubiquitous virtual reality," *Proc. - Int. Symp. Ubiquitous Virtual Reality, ISUVR 2008*, no. May 2014, pp. 33–36, 2008.

[16] S. Jang, S. Kim, and W. Woo, "When VR meets UbiComp," *Virtual Real.*, pp. 1–2, 2005.

[17] J. Grubert, T. Langlotz, S. Zollmann, and H. Regenbrecht, "Towards Pervasive Augmented Reality: Context-Awareness in Augmented Reality," *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 6, pp. 1706–1724, 2017.

[18] A. H. Basori, A. M. A.-G. Al-Sharif, A. O. F. AL-Oufi, A. O. Almagrabi, and O. M. Barukab, "Intelligence Context Aware Mobile Navigation using Augmented Reality Technology," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 4, no. 1, p. 65, 2018.

[19] N. Norouzi, G. Bruder, B. Belna, S. Mutter, D. Turgut, and G. Welch, "A Systematic Review of the Convergence of Augmented Reality, Intelligent Virtual Agents, and the Internet of Things," in *Artificial Intelligence in IoT. Transactions on Computational Science and Computational Intelligence*, 2019, pp. 1–24.

[20] F. Cassola, L. Morgado, F. de Carvalho, H. Paredes, B. Fonseca, and P. Martins, "Online-Gym: A 3D Virtual Gymnasium Using Kinect Interaction," *Procedia Technol.*, vol. 13, no. December, pp. 130–138, 2014.

[21] L. Fernandes et al., "InMERSE Framework," 2015. [Online]. Available: https://bitbucket.org/Apidcloud/inmerse-framework/src/master/.

[22] L. M. Alves Fernandes et al., "Exploring educational immersive videogames: an empirical study with a 3D multimodal interaction prototype," *Behav. Inf. Technol.*, vol. 35, no. 11, pp. 907–918, 2016.

[23] G. Hidalgo, Z. Cao, T. Simon, S.-E. Wei, H. Joo, and Y. Sheikh, "OpenPose: Real-time multi-person keypoint detection library for body, face, hands, and foot estimation," 2019. [Online]. Available: https://github.com/CMU-Perceptual-Computing-Lab/openpose.

[24] "TensorFlow," 2019. [Online]. Available: https://www.tensorflow.org/.

[25] Wikimedia Commons, "Wikimedia Commons," 9 December 2018. [Online]. Available: https://commons.wikimedia.org/wiki/File:SRI_Computer_Mouse.jpg

[26] Ultraleap Ltd, "LEAP MOTION" 2019. [Online]. Available: https://www.leapmotion.com/press/

[27] Microsoft, "Mixed Reality," 2019. [Online]. Available: https://news.microsoft.com/presskits/windows-mixed-reality/

10 ENTERing the future through open innovation

João Fernandes, Odysai
joao.fernandes@docdigitizer.com

Luis Reis, Altice Labs
luis-m-reis@alticelabs.com

Nuno Silva, Altice Labs
nunos@alticelabs.com

Working with startups and scaleups paves the way to go after innovative ideas. Open Innovation, as well, strategically leverages corporate knowledge and resources along with the startup's flow of new ideas and niche knowledge. Aware of the need to take a strong approach when it comes to innovation, Altice Portugal has launched the ENTER program, an open innovation and business corporate venture program, that acts as a bridge between innovative startups and the Altice Portugal business units.

Keywords

ENTER; Corporate venture; Startups; Artificial intelligence; Data capture; Optical character recognition; Cognitive services

Introduction

The business landscape is changing at great speed with tech innovations and disruptive business models fundamentally transforming the traditional TELCO space. Innovative technologies such as cloud computing and storage, blockchain and, especially, artificial intelligence (AI) and machine learning (ML) may bring a significant contribution to Altice Group's digital business transformation, promoting and optimising the evolution of its processes, products and services.

But to maintain leadership in a global digital market, no company today can resort only to its own internally generated knowledge and creativity. Many companies, Altice Labs included, are now working regularly with universities, academy-industry interface units, startup incubators and accelerators, thus getting fast access to the best in the breed, both in scientific research, creativity and entrepreneurship areas.

Under this scope, Altice Portugal has launched ENTER, an open innovation and business corporate venture program that acts as a bridge between innovative startups and scaleups and its business and innovation units. ENTER aims to strategically leverage the Altice Group knowledge and resources along with the startup's agility, the flow of new ideas and deep niche knowledge.

Open Innovation and ENTER

In the scenario above described, open innovation is becoming crucial! In one hand, because an exclusively internal innovation model has proven to be insufficient to cope with the pace the market and the industry require; in the other hand, is seen as a promising source of innovative opportunities for R&D and product development teams. Since it brings novelty and agility from the outside into the organisation, open innovation assumes established organisations:

- **may use** external ideas to reduce the lack of knowledge in that area;
- **may search** for core products to expand their business scope;
- **may look** for new solutions to improve the features of their core product;
- **may want to find** adjacent technologies to become more efficient at a specific stage of the value chain.

Moreover, corporate venture offers a way to work between established corporations and innovative startups (see **Figure 1**), allowing corporates to innovate at the speed of startups, thus leading to a rapid influx of new products and opportunities. On the startup side, it gives access to resources and markets, which is particularly important in sectors where a small number of large players dominate the market, so partnering with them enables instant access to a customer base, the corporate partner included.

However, the collaboration, even when wanted by an organisation, is not always possible since there are often barriers which inhibit effective collaboration, namely, issues related with strategy, structure, organisational culture or internal processes from the corporate side. By far, the biggest challenges reported by startups when trying to collaborate with corporates are the long cycle times and slow decision-making process. To fight the unsuccessful approaches and to make sure the co-working scenario is the best for both parties, there are different means for corporates to engage with startups, namely:

- **organising one-off events**, with competitions such as hackathons;
- **sharing resources**, for example, having co-working spaces and using common tools;
- **providing business support**, with accelerators and incubators;

	Corporations	Startups
Disadvantages	a) Slow organization b) Lack of creativity c) Standardized inflexible processes d) Limited motivation e) Slow-pace growth f) Aversion to risk	1) Difficulties in accessing new markets 2) Small workforce 3) Lack of resources 4) Tight budget 5) Small number of partners 6) Narrow visibility
Corporations	a) Knowledge and access to market b) Experienced workforce c) Resources, experience and power d) Available capital e) Wide network of partners f) Visibility and quality assurance	1) Organizational agility 2) Flow of new ideas and niche knowledge 3) Desire to challenge the status quo 4) Highly motivated teams 5) Potentially rapid growth 6) Little impact if it fails

FIGURE 1 – Opportunities for collaboration between large corporations and startups [1]

- **promoting partnerships**, such as product co-development or procurement from startups;
- **making investments** (corporate venture capital).

Aware of the need to co-innovate with startups, Altice Portugal has launched the ENTER program, an open innovation and corporate venture program, promoting faster and less risky access to innovative products and services, based on collaboration with technological and digital startups. Aimed at Altice Portugal's business, technological and business support units were designed to reinforce both Altice Portugal's competitive advantage and its internal operational efficiency.

The ENTER program's strategy is mostly supported by mechanisms or venturing tools like scouting, product pilot access, networking opportunities, prizes (Altice International Innovation Award; IoT Challenge) and services in kind, such as free office facilities (ENTER premises). In summary, ENTER program "*offers a collaboration framework that acts as a bridge between innovative and disruptive startups and an established corporation*" [2].

Shaping the future: AI transforming the World

AI is set to be the key source of transformation, disruption and competitive advantage in today's fast-changing economy. Some of its main technologies [3] are described in **Figure 2** (on next page) .

AI is also emerging as an important tool to evolve business process efficiency, drive new revenue streams and enhance the Customer experience. Some of the areas where AI will be the key to leverage digital service providers' (DSP) operations and innovations are [4]:

- **Customer engagement:**
  - The emerging of touchpoints such as virtual assistants, chatbots, conversational machine voice attendants, among others, are reinventing the relationship with Customers;

FIGURE 2 – Main technologies of AI [3]

- The use of ML algorithms allows for the discovery of patterns, themes and opportunities in the largely unstructured communications data set owned by DSP;
- Using an ML, AI-driven, omnichannel approach holds great promise for lowering customer support costs. It will also improve the support experience.

• **Smart devices** (smartphones, smartwatches, etc.): these devices will soon incorporate AI technologies at large. As image processing and natural language processing are two major AI-enabled technologies, there is a clear technical and functional rationale to incorporate AI features behind smartphones cameras, micros and speakers;

- **In service and business assurance**, where the human experience of recurring faults and incidents correlation is a key element for fast and efficient resolution of trouble tickets, ML techniques can play a major role in its automation, potentiating 'zero-touch assurance' goals;
- **In network management**, by implementing and leveraging cognitive technologies in the traffic management area to categorise services automatically based on traffic characteristics analysis. Also, in the automation of software-defined networks (SDN) and network functions virtualisation (NFV) for complex events correlation, root cause analysis, preventive maintenance, preventive assurance and incident resolution.

From the personal assistants in our mobile phones to the profiling, customisation, and cyber protection that lie behind more and more commercial interactions, AI now touches almost every aspect of our lives, transforming the world where we live!

Startups bringing AI/ML to open new opportunities and revenue stream

Many innovative and disruptive startups are playing a key role in AI's new digital frontier. According to Roland Berger & Asgard, "*AI innovation now stems largely from research laboratories, big tech digital platforms and startups. These are the players creating algorithms and developing use cases; they are the brains behind innovations in AI functional applications like image recognition, natural language processing and automated driving.*" [5]

Furthermore, startups host entrepreneurs and innovators, researchers and business people. Not seldom these profiles are embodied by one single person! Quite often, this person is the researcher or post-doctoral student that turn to the idea of creating a startup. AI startups champions tend to be "*something of a hybrid between companies and research labs in terms of their areas of focus, team makeup and duration of product development*" [5]. That may explain why - taking the United States (US) as an example - active US AI startups increased 113% from 2015 to 2018 [6], and venture capital funding for US AI startups increased 750% from 2013 to 2018, accounting for USD 9.3B during 2018 [7]. At its scale, the Portuguese startup ecosystem is also thriving and a new breed of Portuguese AI-related startups is increasingly driving relevant developments in technology and business, namely, in the area of cognitive services.

AI in the form of cognitive services is a key enabler to boost innovation and to enable a whole new era for the digital industry, being not only recognized as the driving force behind the 4th industrial revolution, but also a promise to enhance the way Altice Group positions itself as a DSP. Therefore, through the ENTER program, Altice Portugal is working with Odysai, a startup providing an AI cognitive application [8] for transforming human-readable content into actionable data.

A cognitive application by Odysai

Today, information is not only being exchanged on more channels, but it is also getting less structured in nature (e-mails, chatbot logs, etc.). If, on the one hand, the improvement of Customer's experience imposes tearing down the barriers in the way people communicate, on the other hand, breaking down these barriers also leads to more significant organisation inefficiencies in the active and timely treatment of data. In most cases, leading with less structured content requires a significant human intervention for data structure and validation, which limits data delivery time, as well as influences quality. To solve these issues, solutions that attempt to mimic human behaviour, such as robot process automation (RPA), are emerging to orchestrate and execute repetitive actions, emulating human behaviour on the software systems.

Nevertheless, if a robot is highly efficient at transferring information from a spreadsheet to any other system, the same efficiency is no longer obtained in the analysis of an e-mail or any other content with a format not previously recognised by the robot. Thus, RPA technologies are frequently integrated with data capture and optical character recognition (OCR) tools, being their goal to allow to structure information in a way that the robot can make it actionable. These data capture tools mostly have two approaches to interpret and structure information, namely:

- **Templating**: which is based on the idea of previous knowledge over the format and location of key information within a document;
- **ML and NLP:** which is based on the idea of pattern interpretations through which it is possible to induce where key information is present.

The main goal of these technologies is to ensure that the automation processes are fed with actionable, structured and standardised information. Technologies that entail the use of templates or strict configurations on the information format have a reduced scope of application, in addition to presenting very high implementation and maintenance costs. The challenge must then be addressed by new platforms supporting data processing that rely on the scale and speed of the most modern data capture systems based on AI – see **Figure 3**.

In what human-readable content is concerned, one of the main challenges resides in supporting efficient data capture over documents that are unpredictable in format or nature. Just like humans are highly efficient in information recognition and recovery, by being able to learn new methods empirically, as well as to adapt prior knowledge to new situations, so should be a data capture system. In other words, if, in some cases, a more template approach is the most appropriate, in others, the preferable approach is the text interpretation.

Conclusions

AI has the potential to fundamentally disrupt the market through the creation of innovative new services and entirely new business models. With the eruption of AI, some of the market leaders in ten, even five years may be unknown companies nowadays. In turn, some of today's biggest

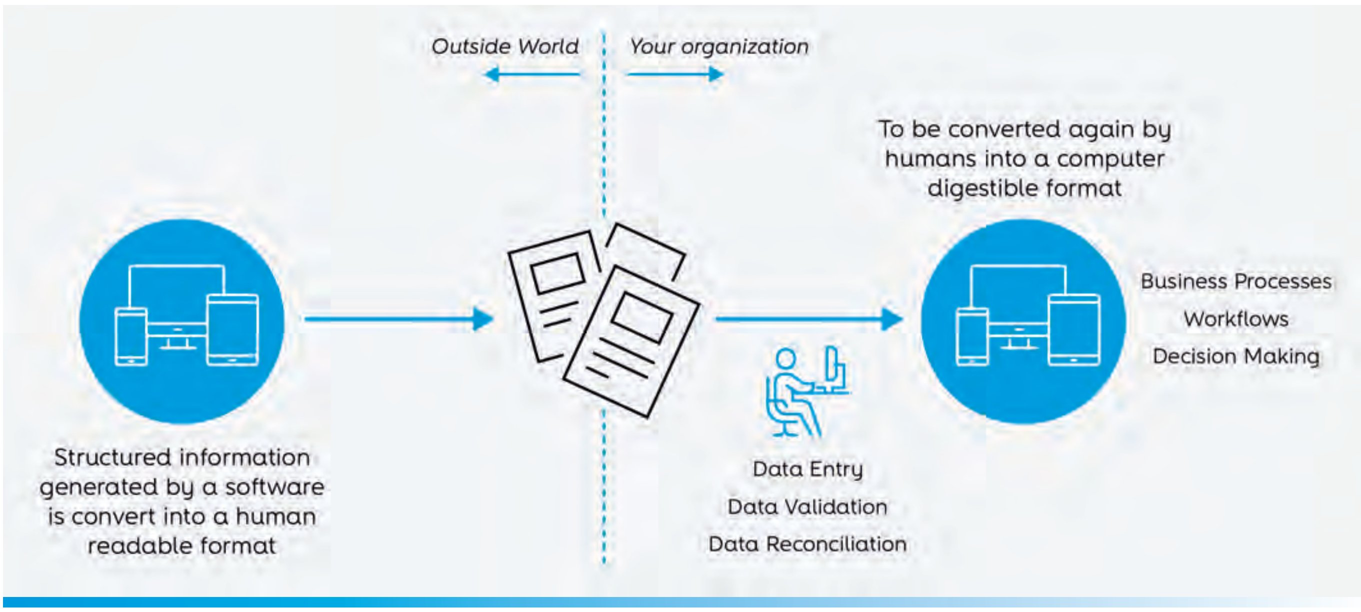


FIGURE 3 – DocDigitizer Platform-As-A-Service, from Odysai [8]

commercial names could be struggling to sustain relevance, if their response has been too little or too late.

Corporates have been putting in place ambitious venture strategies to establish symbiotic relationships with AI-enabled startups. Believing it is possible to achieve a cross-fertilisation both in terms of ideas and business opportunities, and paving the way for more innovation and increased revenue generation for the whole group, Altice Portugal has launched the ENTER programme. ENTER addresses the challenges for the successful implementation of open innovation, with AI specialised startups being one of the main focus, since they are inexorable sources of fresh knowledge relating to AI technologies, tools and techniques.

References

[1] M. J. Prats, J. Siota, T. Canonici and X. Contijoch, “Open Innovation: Building, Scaling and Consolidating Your Firm’s Corporate Venturing Unit,” IESE Business School; Opinno, 2018.

[2] M. J. Prats, P. Amigó, X. Ametller and A. Batlle, “Corporate Venturing: Achieving Profitable Growth Through Startups,” IESE Business School; mVenturesBcn, 2017.

[3] A. Rao, G. Verweij and E. Cameron, “Sizing the prize: What’s the real value of AI for your business and how can you capitalise?,” PwC, 2017.

[4] P. Blanchard, “Leveraging AI for a quantum leap in digital transformation,” TMForum, 2018.

[5] A. LEMAIRE, R. LUCAZEAU, T. RAPPERS, F. WESTERHEIDE and C. E. HOWARD, “Artificial Intelligence – A strategy for European startups,” ROLAND BERGER GMBH, ASGARD CAPITAL VERWALTUNG GMBH, 2018.

[6] Y. Shoham, R. Perrault, E. Brynjolfsson, J. Clark, J. Manyika, J. C. Niebles, T. Lyons, J. Etchemendy, B. Grosz and Z. Bauer, “The AI Index 2018 Annual Report,” Stanford University, 2019.

[7] “MoneyTree™ Report,” PwC; CB Insights, 2018.

[8] DocDigitizer. (2019). *DocDigitizer*. Retrieved from DocDigitizer: https://www.docdigitizer.com/

Acronyms & Terms

2

2G/3G/4G/5G Second/third/fourth/fifth generation mobile networks

3

3D Three-dimensional
3GPP Third Generation Partnership Project, a collaboration between groups of telecommunications standards associations

5

5GAA 5G Automotive Association
5G-ACIA 5G Alliance for Connected Industries and Automation, a global forum for collaboration between automation, engineering, and process industries, and telecom operators and suppliers
5G-PPP 5G Infrastructure Public Private Partnership

A

AC Air Conditioning
ADSL2+ Asymmetric Digital Subscriber Line for 2MHz
AI Artificial Intelligence
API Application Programming Interface
App Application
AR Augmented Reality
ARaNI Altice Labs and Universidade de Trás-os-Montes e Alto Douro project for augmented reality and natural interaction for smart living
ASOP Autonomous Service Operations Platform

B

B/OSS Business Support Systems/Operations Support Systems
B2B Business-to-business
B2C Business-to-consumer
BBF BroadBand Forum
BC Before Christ
BI Business Intelligence
BOM Bill of Materials
BOSA Bi-directional Optical Sub-Assembly
BotSchool Altice Labs' cognitive virtual assistant programming solution
BPON Broadband PON
BSS Business Support System

C

CAPEX Capital Expenditures
CATV Cable TV
CCO Cloud Central Office
CD Continuous Delivery
CeX Customer Satisfaction
CHIC Cooperative Holistic View on Internet and Content, a P2020 project
CI Continuous Integration
CNCF Cloud Native Computing Foundation
CPU Central Processing Unit
CSFB Circuit Switched FallBack
CSP Communications Service Providers
CUPS Control and User Plane Separation
C-V2X Cellular-V2X

D

D2D Device-to-Device
DataPlaxe Altice Labs' platform to collect, orchestrate, process, transact, perform analytical calculations over large dataset volumes, its visualisation and reporting
DC Data Centre
DL Deep Learning
DNN Deep Neural Networks
DNPW Do Not Pass Warning
DPU Distribution Point Unit
DSP Digital Services Provider
DSRC Dedicated Short-Range Communication
DVR Digital Video Recorder

E

eMBB enhanced Mobile BroadBand
eMMC embedded MultiMedia Card
eNB enhanced Node B
ETSI European Telecommunications Standards Institute
EU European Union
E-UTRAN Evolved Universal Terrestrial Radio Access Network

F

FaceApp A mobile application for iOS and Android developed by Russian company Wireless Lab
FCW Forward Collision Warning

	Fiber Gateway	Altice Labs' terminal equipment for GPON fiber access integrating the fiber component
	FSAN	Full Service Access Network
	FTTb	Fiber-to-the-business
	FTTB	Fiber-to-the-Building
	FTTc	Fiber-to-the-cell
	FTTC	Fiber-to-the-Cabinet
	FTTdp	Fiber-to-the-distribution-point
	FTTH	Fiber-to-the-Home
	FTTx	Fiber-to-the-x
G	**G.fast**	ITU-T digital subscriber line protocol standard
	GDPR	General Data Protection Regulation
	GIS	Geographic Information System
	Google Cardboard	A virtual reality platform developed by Google for use with a head mount for a smartphone
	Google Glass	An optical head-mounted display designed in the shape of a pair of eyeglasses from Google
	GPON	Gigabit Passive Optical Network
	GPU	Graphical Processing Unit
	GUI	Graphic Users Interfaces
H	**HCI**	Human-Computer Interaction
	HMD	Head-Mounted Display
	HoloLens	A pair of mixed reality smartglasses developed and manufactured by Microsoft
	HSI	High Speed Internet
	HTC Vive	A virtual reality headset developed by HTC and Valve
	HVAC	Heating, Ventilation and Air Conditioning
I	**IEEE**	Institute of Electrical and Electronics Engineers, a professional association for electronic engineering and electrical engineering
	IFTTT	if this then that
	IMS	IP Multimedia Subsystem
	InMERSE	Altice Labs' exploratory study on the use of techniques for natural interaction and immersion devices in specific scenarios
	IoT	Internet of Things
	IP	Internet Protocol
	IPTV	Internet Protocol Television
	ISM	Industrial, Scientific and Medical
	IT	Information Technology
	ITS	Intelligent Transport Systems
	ITU	International Telecommunication Union
	ITU-T	International Telecommunication Union, Telecommunication Standardization Sector
	IVR	Interactive Voice Response
K	**Kinect**	A Microsoft's motion sensor add-on for the Xbox 360 gaming console
	KPI	Key Perform Indicators
L	**Layer 2**	The Data Link layer of the Open Systems Interconnection Reference Model
	LeapMotion	American company that manufactures and markets a computer hardware sensor device that supports hand and finger motions as input
	LED	Light Emitting Diode
	LTE	Long Term Evolution
M	**MAC**	Media Access Control
	MAE	Mean Absolute Error
	MEC	Multi-access Edge Computing
	ML	Machine Learning
	MME	Mobility Management Entity
	mMTC	massive Machine Type Communications
	MPLS	Multiprotocol Label Switching
	MR	Mixed Reality
	MRR	Mean Reciprocal Rank
	MTTR	Mean Time to Repair
	MU-MIMO	Multi-user Multiple-input and Multiple-output
	MUX	Multiplexer
	Myo	United States-based medical device company specializing in myoelectric orthotics for people with neurological disorders

N	**N&S**	Network and Services
	N/SOC	Network/Service Operation Centres
	NAND	inverted AND flash memory type
	NE	Network Elements
	NFV	Network Function Virtualisation
	NG-PON2	Next Generation PON 2
	NLP	Natural Language Processing
	NMS	Network Management System
	NOC	Network Operations Centre
	NOR	inverted OR flash memory type
	NR	New Radio
	NSA	Non-Standalone Architecture
O	**OCP**	Open Compute Project, an organization that shares designs of data centre products among companies
	OCR	Optical Character Recognition
	Oculus Rift	A lineup of virtual reality headsets developed and manufactured by Oculus VR
	ODN	Optical Distribution Network
	OLT	Optical Line Termination
	OMCI	ONU Management and Control Interface
	Online Gym	Exploratory project based on collaborative networked virtual reality that consists of a 3D platform that encourages people to exercise in groups on the Internet
	ONT	Optical Network Terminal
	ONU	Optical Network Unit
	OpenPose	Open-source realtime system for multi-person 2D pose detection, including body, foot, hand, and facial keypoints, from Carnegie Mellon University
	OPEX	Operational Expenditures
	OSF	OpenStack Foundation
	OSS	Operation Support System
P	**P2MP**	Point-to-Multipoint
	P2P	Point-to-Point
	PCA	Principal Component Analysis
	PDCP	Packet Data Convergence Protocol
	PhD	Doctor of Philosophy
	PHY	Physical Layer
	PON	Passive Optical Network
	PSU	Power Supply Unit
Q	**QoS**	Quality of Service
R	**R&D**	Research and Development
	RAN	Radio Access Network
	RCA	Root-Cause Analysis
	RF Overlay	Radio Frequency signals broadcasted over an additional wavelength
	RFoPON	Radio Frequency over PON
	RMSE	Root Mean Squared Error
	RPA	Robot Process Automation
	RRC	Radio Resource Control
	RSSI	Received Signal Strength Indicator
S	**SA**	Standalone Architecture
	Samsung Gear VR	A virtual reality headset manufactured by Samsung
	SAT	Satellite
	SDN	Software Defined Networks
	SLA	Service Level Agreement
	Smart Mirror	Altice Labs' voice and gesture interactive mirror project
	SMS	Short Message Service
	SOC	Service Operation Centres
	SSD	Solid-State Drive
	SVD	Singular Value Decomposition
	SVM	Support Vector Machine
T	**TDM**	Time Division Multiplexing
	TELCO/ TELCOS	Telecommunication Operators
	TensorFlow	Open source artificial intelligence library, using data flow graphs to build models
	TIP	Telecom Infra Project
	TM Forum	A non-proft industry association for service providers and their suppliers in the telecommunications industry
	TSU	Tensor Processing Unit
	TTK	Trouble Tickets
	TV	Television

	TWDM-PON	Time- and Wavelength-Division Multiplexing PON
	TX	Transmission
U	**UAV**	Unmanned Aerial Vehicle
	UC	Use Case
	uMSAN	micro Multi-Service Access Node
	URLLC	Ultra Reliable and Low Latency Communications
	US	United States
	USA	United States of America
	USD	United States Dollar
	U-VR	Ubiquitous Virtual Reality
	UX	User Experience
V	**V2I**	Vehicle-to-Infrastructure
	V2N	Vehicle-to-Network
	V2P	Vehicle-to-Pedestrian
	V2V	Vehicle-to-Vehicle
	V2X	Vehicle-to-everything
	VDSL	Very-high-speed Digital Subscriber Line
	VHS	Video Home System, is a widely-adopted videocassette recording technology developed by JVC
	VIM	Virtual Infrastructure Manager
	VNF	Virtual Network Functions
	VOD	Video On Demand
	VoIP	Voice over IP
	VR	Virtual Reality
	VRU	Vulnerable Road User
	VTF	Vertical Engagement Task Force
W	**WAN**	Wide Area Network
	WAVE	Wireless Access in Vehicular Environment
	Wi-Fi	IEEE 802.11x - Wireless Network (Wi-Fi Alliance)
	Wi-Fi 6	IEEE 802.11ax - the next generation standard in Wi-Fi technology, builds and improves on the current 802.11ac Wi-Fi standard
	WLAN	Wireless LAN
	WM	Wavelength Modulator
X	**XaaS**	Everything-as-a-Service
	XG-PON	10-Gigabit-capable (asymmetrical) PON
	XGS-PON	10-Gigabit-capable Symmetrical PON
	xPON	Designation for several PON technologies
	XR	eXtended Reality

Acknowledgements

It is essential to leave a note of gratitude and acknowledgement to all those who helped, once again, to build this publication. To all authors, for their contributions and for sharing the knowledge resulting from their research. To all technical and editorial reviewers that did a relevant and meticulous job on improving the quality and excellence of all articles and of the entire magazine. To our designer for the excellent job of granting an appealing and coherent design across the entire magazine.

Strategic team

Ana Patrícia Monteiro, Fausto de Carvalho,
Mário Rui Costa, Paula Cravo,
Paulo Pereira and Pedro Carvalho

Editorial team

Ana Patrícia Monteiro and Paula Cravo

Art coordinator and graphic edition

Cátia Santos Pinto

Technical revision team

Ana Patrícia Monteiro, Carlos Parada,
Fausto de Carvalho, Fernando Morgado,
Isilda Costa, Jorge Carapinha,
Lourenço Moura, Luís Cortesão,
Manuel Dinis, Nuno Alexandre Seixas,
Paula Cravo, Pedro Carvalho,
Ricardo Ferreira and Rui Calé

www.alticelabs.com