# Evolving Machine Learning Paradigm

**Prof. Farhana Kausar[1], Ms. Chinmayi Bandargal[2], Mr.Joseph James[3]**
[1, 2, 3] Dept of Computer Science
[1, 2, 3] Atria Institute of Technology

***Abstract-*** *Machine learning is one of the most promising artificial intelligence tools, conceived to support smart radio terminals. Future smart 5G mobile terminals are expected to autonomously access the most meritorious spectral bands with the aid of sophisticated spectral efficiency learning and inference, in order to control the transmission power, while relying on energy efficiency learning/inference and simultaneously adjusting the transmission protocols with the aid of quality of service learning/inference. Hence, we briefly review the rudimentary concepts of machine learning and propose their employment in the compelling applications of 5G networks, including cognitive radios, massive MIMOs, femto/small cells, heterogeneous networks, smart grid, energy harvesting, device-to device communications, and so on. Our goal is to assist the readers in refining the motivation, problem formulation, and methodology of powerful machine learning algorithms in the context of future networks in order to tap into hitherto unexplored applications and services.*

***Keywords-*** user centric, machine learning system, Heterogenous network , femto cells

## I. INTRODUCTION

Machine learning has found wide-ranging applications in image/audio processing, finance and economics, social behaviour analysis, project management, and so on. Explicitly, a machine learns the execution of a particular task T, with the goal of maintaining a specific performance metric P, based on a particular experience E, where the system aims to reliably improve its performance P while executing task T, again by exploiting its experience E. Depending on how we specify T, P, and E, the learning might also be referred to as data mining, autonomous discovery, database updating, programming by example, and soon.

Machine learning algorithms can be simply categorized as supervised and unsupervised learning, wherethe adjectives “supervised /unsupervised” indicate whether there are labelled samples in the database. Later, reinforcement learning emerged as a new category that was inspired by behavioural psychology. It is concerned with an agent’s certain form of reward/utility, who is connected to its environment via perception and action. The family of machine learning algorithms can also be categorized based on their similarity in terms of their functionality and structure, yielding regression algorithms, instance-based algorithms, regularization algorithms, decision tree algorithms, Bayesian algorithms, clustering algorithms, association rule based learning algorithms, artificial neural networks, deep learning, algorithms dimension reduction algorithms, ensemble algorithms, and so on. In this article, we will introduce the basic concept of machine learning algorithms and the corresponding applications according to the category of supervised, unsupervised, and reinforcementlearning.

## II. MACHINE LEARNING PARADIGMS FOR NEXT-GENERATION WIRELESS NETWORKS

A range of future research ideas on machine learning in 5G networks can be summarized as follows. The family of supervised learning techniques relies on known models and labels that can support the estimation of unknown parameters. They can be utilized for massive MIMO channel estimation and data detection, spectrum sensing and white space detection in cognitive radio, as well as for adaptive filtering in signal processing for 5G communications. They can also be applied in higher-layer applications, such as inferring the mobile users’ locations and behaviours, which can assist the network operators to improve the quality of their services. Unsupervised learning relies on the input data itself in a heuristic manner. It can be utilized for cell clustering in cooperative ultra- dense small-cell networks, for accesspoint association in ubiquitous WIFI networks, for heterogeneous base station clustering in HetNets, and for load-balancing in HetNets. It can also be applied in anomaly/ fault/intrusion detection and for the users’ behaviour- classification.

Reinforcement learning relies on a dynamic iterative learning and decision-making process. It can be utilized for inferring the mobile users’ decision making under unknown network conditions, for example during channel access under unknown channel availability conditions in spectrum sharing, for distributed resource allocation under unknown resource quality conditions in femto/small-cell networks, and base station association under the unknown energy status of the base stations in energy harvesting networks.

## III. APPLICATIONS

In heterogeneous networks constituted by diverse cells, handovers may be frequent, where both the KNN and SVM can be applied to finding the optimal handover solutions. At the application layer, these models can also be used for learning the mobile terminal's specific usage pattern in diverse spatio-temporal and device contexts, as discussed in [5]. This may then be exploited for prediction of the configuration to be used in the location-specific interface. Given a set of context machine learning algorithms are capable of exploiting the user context learned for the sake of dynamically classifying the cues into a system state for the sake of saving energy, while maintaining a high level of user satisfaction. Donohoo et al. also conducted experiments using five real user profiles, including the user-locations and energy consumption, but their data is not accessible to the public. The experiment showed that up to 90 percent successful energy demand prediction is possible with the aid of the KNN algorithms.

Another three closely related applications may be found in cognitive radio networks. In a cooperative wideband spectrum sensing scheme based on the EM algorithm was proposed for the detection of a primary user (PU) supported by a multi-antenna assisted cognitive radio network Clustering is a common problem in 5G networks, especially in heterogeneous scenarios associated with diverse cell sizes as well as WIFI and D2D networks. For example, the small cells have to be carefully clustered to avoid interference using coordinated multi-point transmission (CoMP), while the mobile users are clustered to obey an optimal offloading policy, the devices are clustered in D2D networks to achieve high energy efficiency, the WIFI users are clustered to maintain an optimal access point association, and so on.

Both the PCA and ICA constitute powerful statistical signal processing techniques devised to recover statistically independent source signals from their linear mixtures. One of their major applications may be found in the area of anomaly-detection, fault-detection, and intrusion-detection problems of wireless networks, which rely on traffic monitoring. Furthermore, similar problems may also be solved in sensor networks, mesh networks, and so on.

The family of MDP/POMDP models constitute ideal tools for supporting decision making in 5G networks, where the users may be regarded as agents and the network constitutes the environment. Q-learning has also been extensively applied in heterogeneous networks, usually in conjunction with the aforementioned MDPmodels. In the authors presented a heterogeneous fully distributed multi-objective strategy based on a reinforcement learning model con structed for the self-configuration/optimization offemtocells.

The MAB and MP-MAB models, as a family of emerging signal processing tools, are capable of solving challenging resource allocation problems in wireless scenarios, where either the channel conditions or some other wireless environment parameters have to be "explored," while the known channels also have to be "exploited" by a group ofusers.

## IV. A USER-CENTRIC MACHINE LEARNING FRAMEWORK FOR CYBER SECURITY OPERATIONS CENTRE

Cyber security incidents will cause significant financial and reputation impacts on enterprise. In order to detect malicious activities, the SIEM (Security Information and Event Management) system is built in companies or government. The system correlates event logs from endpoint, firewalls, IDS/IPS, DNS (Domain Name System), DHCP (Dynamic Host Configuration Protocol), Windows/Unix security events, VPN logs etc. The security events can be grouped into different categories. The logs have terabytes of data each day.

From the security event logs, SOC (Security Operation Centre) team develops so-called use cases with a pre-determined severity based on the analysts' experiences. They are typically rule based correlating one or more indicators from different logs. These rules can be network/host based or time/frequency based. If any pre-defined use case is triggered, SIEM system will generate an alert in real time. SOC analysts will then investigate the alerts to decide whether the user related to the alert is risky (a true positive) or not (false positive). If they find the alerts to be suspicious from the analysis, SOC analysts will create OTRS (Open Source Ticket Request System) tickets. After initial investigation, certain OTRS tickets will be escalated to tier 2 investigation system as severe security incidents for further investigation and remediation by Incident ResponseTeam.

However, SIEM typically generates a lot of the alerts, but with a very high false positive rate. The number of alerts per day can be hundreds of thousands, much more than the capacity for the SOC to investigate all of them. Because of this, SOC may choose to investigate only the alerts with high severity or suppress the same type of alerts. This could potentially miss some severe attacks. Consequently, a more intelligent and automatic system is required to identify risky users.

## V. APPLICATIONS

The machine learning system sits in the middle of SOC work flow, incorporates different event logs, SIEM alerts and SOC analysis results and generates comprehensive user risk score for security operation centre. Instead of directly digging into large amount of SIEM alerts and trying to find needle in a haystack, SOC analysts can use the risk scores from machine learning system to prioritize their investigations, starting from the users with highest risks. This will greatly improve their efficiency, optimize their job queue management, and ultimately enhance the enterprise's security.

Specifically, our approach constructs a framework of user- centric machine learning system to evaluate user risk based on alert information. This approach can provide security analyst a comprehensive risk score of a user and security analyst can focus on those users with high risk scores.

The main contribution of this paper is as follows:

- An advanced user-centric machine learning system is proposed and evaluated by real industry data to evaluate user risks. The system can effectively reduce the resources to analyse alerts manually while at the same time enhance enterprise security.
- A novel data engineering process is offered which integrates alert information, security logs, and SOC analysts' investigation notes to generate features and propagate labels for machine learning models.

## VI. CONCLUSION

Furthermore, computational intelligence paradigms, such as neural networks and neuro- fuzzy methods, swarm intelligence algorithms such as ant colony optimization, and evolutionary algorithms such as the competitive imperialist algorithm, may also be applied to improve the performance of 5G networks. Among those compelling techniques, neural networks and deep learning have recently become particularly popular.

Generally, a neural network consists of a number of neurons and weighted connections among them, where the neurons can beregarded as variables and the weights can be viewed as parameters. The network should be appropriately configured with the aid of learning techniques to ensure that the application of a set of inputs produces the desired set of outputs. Explicitly, this can be achieved by iteratively adjusting the weights of the existing connections among all neuron pairs with the aid of learning based on the labelled data for supervised learning or unlabelled data for unsupervised learning. Neural networks have been widely utilized for spectral white state estimation [17], prediction [18], and handoff decisions [19] in cognitive radio networks. Note that the algorithms introduced in this article are only limited samples of the machine learning field. There are many other algorithms that can also be applied to the next-generation networks. For example, the family of evolutionary algorithms, such as genetic algorithm scan solve optimization problems by mimicking a natural selection process, which can be utilized to solve resource allocation problems in HetNets [20]. By contrast, machine learning relies on two phases, the training phase and the testing phase, where the training phase imposes a much higher complexity than the testing phase. Due to the energy constraints and computational complexity constraints of mobile terminals, it is recommended to only implement the testing phase on shirt-pocket-sized mobile terminals.

This article reviewed the benefits of artificial intelligence aided wireless systems equipped with machine learning. We introduced the major families of machine learning algorithms and discussed their applications in the context of next- generation networks, including massive MIMOs, the smart grid, cognitive radios, heterogeneous networks, femto/small cells, D2D networks, and so on. The classes of supervised, unsupervised, and reinforcement learning tools were investigated, along with the corresponding modelling methodology and possible future applications in 5G networks. In a nutshell, machine learning is an exciting area for artificial intelligence aided networking research.

In the research paper, we present a user-centric machine learning system which leverages big data of various security logs, alert information, and analyst insights to the identification of risky user. This system provides a complete framework and solution to risky user detection for enterprise security operation centre. We describe briefly how to generate labels from SOC investigation notes, to correlate IP, host, and users to generate user- centric features, to select machine learning algorithms and evaluate performances, as well as how to such amachine learning system in SOC production environment. We also demonstrate that the learning system is able to learn more insights from the data with highly unbalanced and limited labels, even with simple machine learning algorithms. The average lift on top 20% predictions for multi neural network model is over 5 times better than current rule-based system. The whole machine learning system is implemented in production environment and fully automated from data acquisition, daily model refreshing, to real time scoring, which greatly improve efficiency and enhance enterprise risk detection and management. As to the future work, we will research other learning algorithms to further improve the detection accuracy.

## REFERENCES

[1] SANS Technology Institute. “The 6 Categories of Critical Log Information.” 2013.
[2] X. Li and B. Liu. “Learning to classify text using positive and unlabeled data”, Proceedings of the 18th international joint conferenceonArtificialintelligence,2003
[3] A. L. Buczak and E. Guven. “A survey of data mining and machine learning methods for cyber security intrusion detection”, IEEE Communications Surveys & Tutorials 18.2 (2015): 1153- 1176.
[4] S. Choudhury and A. Bhowal. “Comparative analysis of machine learning algorithms along with classifiers for network intrusion detection”, Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM),2015.
[5] N. Chand et al. “A comparative analysis of SVM and its stacking with other classification algorithm for intrusion detection”, Advances in Computing, Communication, & Automation (ICACCA),2016.