Association nationale des Auditeurs jeunes
de l'Institut des Hautes Etudes de Défense nationale

Première association française de jeunes sur les problématiques de défense et de sécurité

THE USE OF AI IN CYBERSECURITY

Cyber Week – 7th-15th February 2019

COMMITTEE : Cyberdefense – ANAJ-IHEDN

This article exclusively reflects the author's opinions. The ideas or opinions expressed cannot in any case be considered as the expression of an official position.

Securisation of Information Systems (IS), also referred to as cybersecurity, is getting more and more crucial for all types of companies. Indeed, over the past decade, several major attacks have shown the sensitivity and dependance of our institutions regarding information technologies (IT) : the Estonian administration and media were blocked during three weeks in 2007, Iran's nuclear plants stopped in late 2009, TV5monde programs inaccessible for two days in 2015, WannaCry and NotPetya blocked companies and public services worldwide in 2017, and so on.

Hence, all institutions (both public and private sectors, as well as individuals) must treat cybersecurity as seriously as possible, especially as more than 55% of the global population (over 4 billion people) are connected to the Internet in 2018[1].
To do so, tools and methods have kept evolving to continuously upgrade the protection level of entities and individuals : antiviruses, firewalls, intrusion detection and protection systems (IDS and IPS), file and message crypting are a few of them.

Nowadays, and it is what will interest us to study in this series of articles, the great technical evolution brought by artificial intelligence (AI) is meeting cybersecurity. This series of articles will present AI mechanisms through the explanation of the main different types of algorithms (I), the added values it can bring to cybersecurity (II), as well as the risks ensued by its use from malicious actors (III).

[AI mechanisms introduction]

When talking about Artificial Intelligence, most people actually refer to machine learning (ML) or even deep learning (DL), which consist in programing a machine so that it will learn by itself from either data or rules, and take its environment into account to make decisions. AI consists in creating and studying intelligent agents: « any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals »[2], often with the objective to solve problems of high computational complexity (involving huge energy, memory, and computational resources to be solved).

Nowadays, this technology is taking more and more ground as it is being integrated to all our electronic devices : smartphones and watches, tablets, computers, cars, homes... AI is used for facial and voice recognition, self-driving cars, personal or home assistants, and preference settings among others. For instance, Siri and Alexa are AI based assistants, whereas Netflix uses AI to recommend movies and series by learning your preferences.

---

[1] Internet Growth Statistics published on Internet World Stats.
[online], URL : https://www.internetworldstats.com/stats.htm

[2] Poole, Mackworth and Goebel, *Computational Intelligence* p. 1, 1998. [online], URL : https://www.cs.ubc.ca/~poole/ci.html – in which AI is referred to as computational intelligence.

In a great number of cases, AI is used to personalize and adapt a service to each user with a lower need in resources. The common property of these tools is that they take contextual information into account before making decisions. As this series of article aims at studying the use of AI in cybersecurity, we are going to focus on machine learning, the most widely used subfield in our context.

Machine learning actually refers to programs enabling computers to learn on their own. ML is usually based on great sets of data and scenarios. A few years ago, the term of *big data* was taking a great importance, especially in the media. Nowadays, it is what comes from big data that allows most of the machine learning algorithms to train. ML enables the machine to see logs and their dynamic at a global scale in near real-time conditions.

SUPERVISED ALGORITHMS

Supervised algorithms enable to work with clean datasets. This means that all data has been labeled and clustered (marked as belonging to a group of similar data). The machine's objective is to learn the *mapping function*, ie. the function which enabled to get from the input (the unlabeled raw data) to the output (result to predict).
There are two main categories :

- classification : sorting items into different categories (shapes, colors, etc.),
- regression : drawing a predictive curve and foresee a comportment of real values (currency, weight, temperature, etc.).

Once this mapping function is well enough approached, the algorithm will be able to predict the function's output with a new given input. Thus, the algorithm will learn by analysing how the groups were formed, searching for similarities within subgroups and differences in between them.
This type of algorithms are said to learn through examples, and several datasets are necessary in order for the algorithm to mimic the original function.

The following image illustrates this process : the raw data is submitted (apple photographs here), then an operator gives indication as to what is the expected output. Thereafter processing, the intelligent agent is able to recognise whether the submitted input corresponds to what it learnt to recognise, or not (apple or not, in our case).

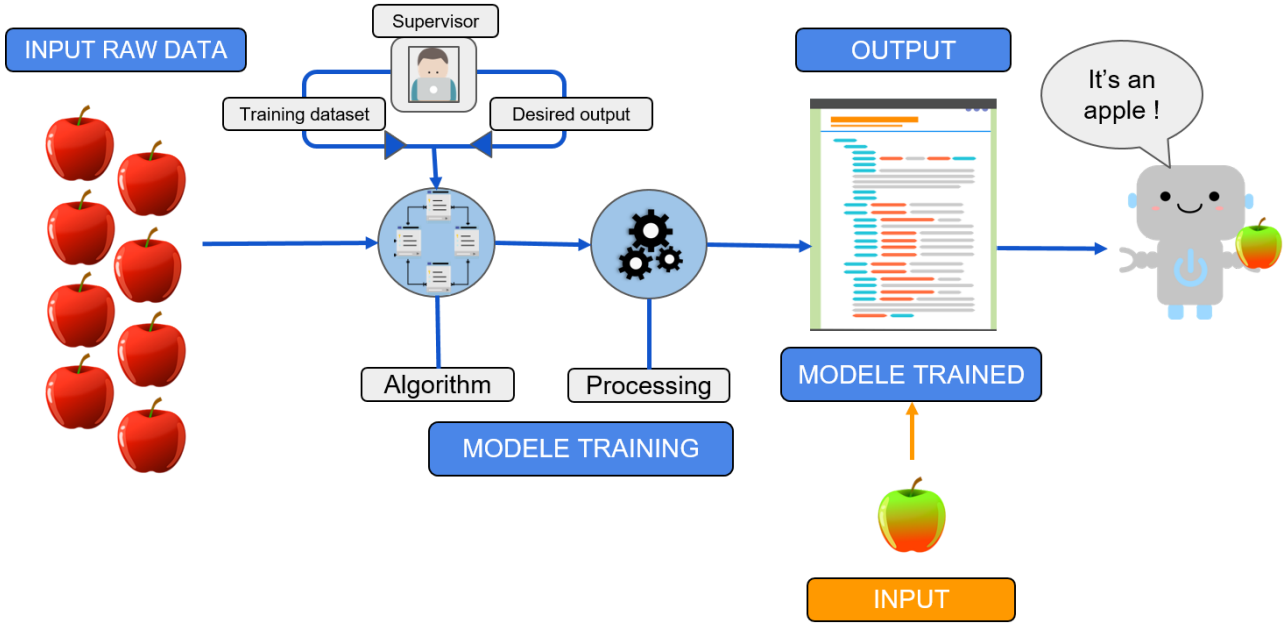

Illustration 1 - Supervised machine learning process

REINFORCEMENT LEARNING

To treat datasets that have been poorly labeled (as in most real-life examples) we can use semi-supervised algorithms. Indeed, if a dataset provides both an input and the corresponding output (the label) for only 10% of the data, it might be too expensive to label every datapoint and use a supervised algorithm.

Thus, two options are available : using unsupervised algorithm (see below) and flout the provided output, or use a four-steps process sometimes referred to as *reinforcement learning* (see Illustration 2):

1. Use supervised algorithm to learn the mapping function thanks to the 10% of labeled data,
2. Train on a similarly sized data set (another 10%),
3. Check the algorithm's output,
    b. correct it if necessary,
4. Feed it to the algorithm as a new training set.

In the reinforcement learning context, the intelligent agent will react to its environment (dataset and feedbacks) to maximise its positive rewards.

Illustration 2 - Reinforcement learning process

UNSUPERVISED ALGORITHMS

The other main ML algorithms are unsupervised algorithms. These take datasets which have not been labeled at all as an input. They let the machine completely responsible for the detection of similarities in the dataset because it has no reference to identify the "right" answer.
Here again, we can split this type of algorithms into two main categories : classification and association. Classification algorithms aim at discovering the underlying structure of a dataset to determine coherent subgroups (subgroups such as client having the same buying comportments in a store); whereas association algorithms will rather try to identify underlying rules to understand data more widely (rules such as "client buying milk also tend to buy diapers" for instance). The way the algorithms choose to group data can be studied through community detection[3] in particular.

Here is an example of classification algorithm objectives : instead of being able to say if a given object is an apple or not, the algorithm is able to classify the entry as being an apple, or a banana, or a mango.

[3] Applied mathematical field closely linked to graph theory.

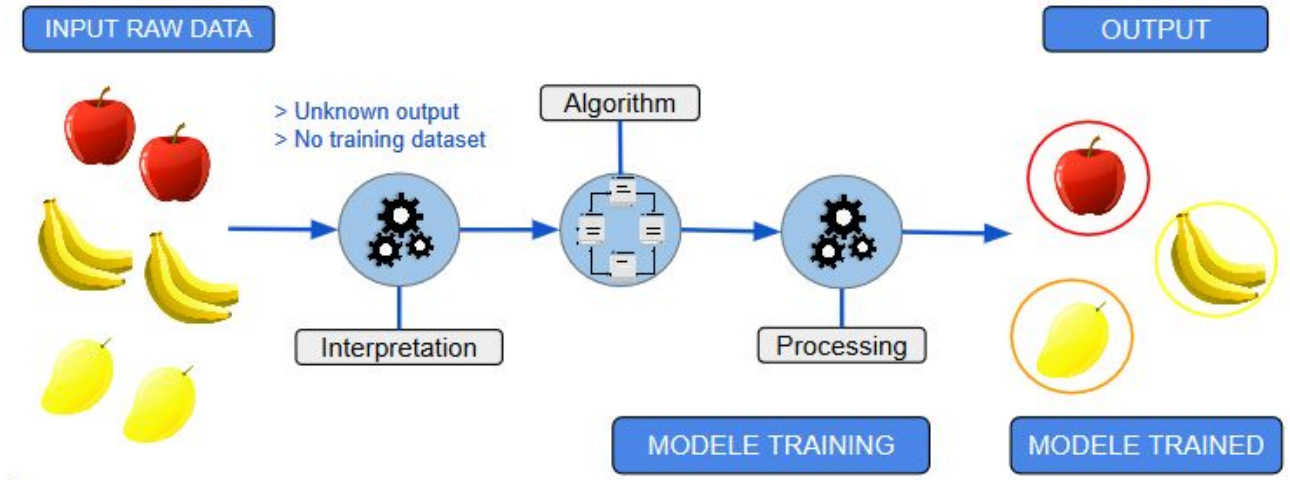

Illustration 3 - Unsupervised machine learning process

DEEP LEARNING

Deep learning is a more technically advanced part of machine learning which operates through several hidden layers. This technology is notably used for complex data structure recognition, such as facial and vocal recognition as its objective is to model a high level of abstraction through non-linear transformation, and thanks to metadata[4].
As shown below, after being trained, a deep learning based agent will decompose the data submitted in order to recognize features by itself and in the end can have more precise outputs.

Deep learning models are thought to approach as much as possible human neuronal networks, which are not fully understood and hold secrets as the way intermediar neuronal layers cooperates and divide (or not) the incoming to-do tasks. Similarly, the *deep neural network* of a deep learning algorithm is made of hidden layers which collaborate to classify features composing objects (see illustration below). There are several types of hidden layers which can apply about any function to its income from previous layer elements (usually linear transformation followed by inequality flattening).
For example, with the input of a human face, a DL algorithm would try to recognise the person, while another less sophisticated algorithm would only recognise that the picture is one of a human being.
As illustrated below, to do so, DL hidden layers treat information at a different scale and base themselves on a mix of the output from one or several previous layers (according to their position in the network). Here, the first hidden layer will identify pixels drawing edges, the second hidden layer will group these edges to identify shapes, and finally the third hidden layer will recompose pieces of faces to allow recognition in the output.

[4] Metadata give information about the data's origins (who created the main data, what for, what it includes), the original data structure (what elements compose it, and internal organisation), the origin of the resources (types and access rights).

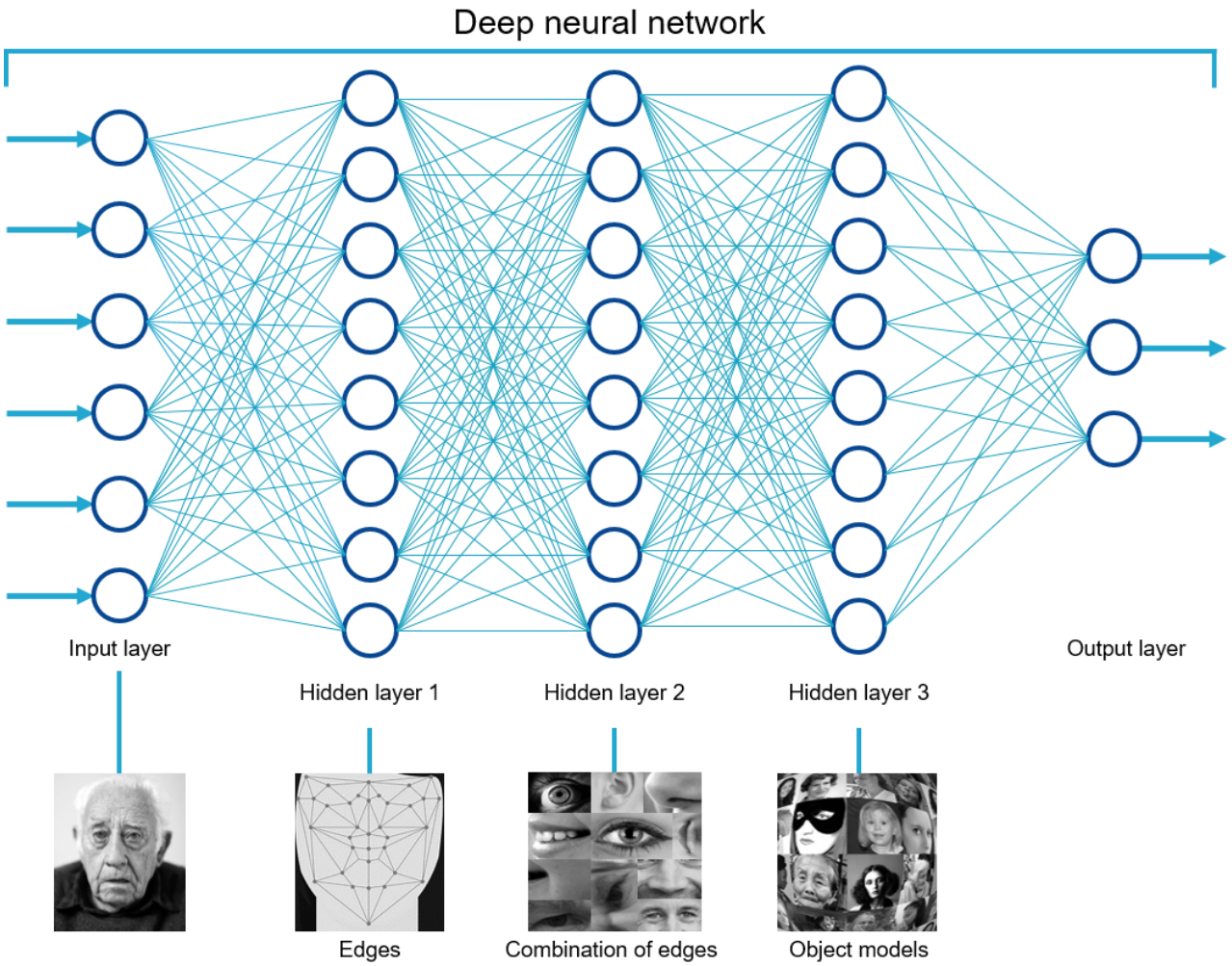

Illustration 4 - Deep learning process to recognize a face

However, this characteristic has quickly become an issue because the users of this technology are not able to access and understand how the machine has made its decision all the way through. This situation could drive to legal issues regarding responsibility attribution, especially when human lives are at stake, as in self driving car accident situations.

To summarize, four types of algorithms can be applied in machine learning, presented in the table below.

Algorithm	Characteristic	Given input	Given output	Objective	Type of operations	Limits
Supervised	Task driven	Data set	Fully labeled dataset	Imitate mapping function linking input and output	Classification and regression	Need of fully labeled set Need of rather detailed instructions
Unsupervised	Data driven	Data set	None	Determine underlying structure of the input	Clustering and association	Need of rather detailed instructions
Reinforcement	Environment adaptive	Data set	Feedback, rewards	Maximise positive rewards	Preferences prediction	Need of feedbacks
Deep learning	Hidden layers	Metadata set	Fully labeled dataset	Extract features to model complex objects	All of above	Lack of understanding Potential legal issues Need of fully labeled set

Table 1 - Machine learning algorithms

To treat data through machine learning, as we have seen above we must have great amounts of it. The condition of the registered data is crucial : the fields which give sense to the ML models must be completed and reliable. For example, data generated when a photograph is taken without a timestamp will be found useless to someone (or some machine) trying to make use of the camera's activity timeline.

We will see in following article that, as cybersecurity currently relies on detection of preset attack scenarios, the advent of AI could allow, with much less effort (in time, reflexion and anticipation) to distinguish normal comportments from malicious ones. Thanks to weak signals automatic analysis, this detection and prevention could help experts focus on crisis resolution rather than its detection.
However, as a result, AI datasets and algorithms get even more sensitive to attacks by compromising training data. Moreover, as it can help institutions to detect weak signals, AI can also be used by attackers to use these signals for the purpose of sneaking into IT networks without being noticed.

_____

Mathilde DELFOSSE-LEGAT
Membre du Comité Cyberdéfense de l'ANAJ-IHEDN
104ème Séminaire Jeunes, Dijon, 2017