ISSUE
02
February
2020

Royal
Academy of
Engineering

Digitalization Transformation Guides

In this issue:

- What is Artificial Intelligence?
- How to bring benefit from Machine Learning in industry?
- Dangers and ethical considerations of AI

Solutions for a New Economy

There are numerous applications in which Machine Learning is the right solution and can bring benefit to the industry.

Industrial Implementation of Artificial Intelligent Systems in Industry

.

The prominence of AI is demonstrated in popular culture with television and movies. Artificial Intelligence refers to that of human intelligence demonstrated by machines which are programmed to think like humans and replicate decisions, behaviours and actions. The application of AI is widely associated to machines having the capability to learn and to solve problems as humans would, or to a better degree.

Machine Learning (ML), and AI in general, are still relatively new research fields and are not necessarily the solution to every technological and scientific problem. While it is expected that the presence of AI in the industry will largely increase in the near future, there is a broad spectrum of problems for which AI may not be yet a satisfactory solution.

Loughborough
University

Contents

What is Artificial Intelligence?

Brief History

Although not generally reported, formal research on Artificial Intelligence can be said to have originated at the Macy conferences on Cybernetics (1945-1953). The transactions from those events include landmark outputs, such as the first artificial neural networks by McCulloch and Pitts – capable of performing basic arithmetic – and the information entropy measure, which constitutes the basic loss function for training most classification models today. The Macy conferences were shortly followed by the well-known Dartmouth Workshop (1956), where the moniker Artificial Intelligence (AI) was coined. The Dartmouth Workshop established AI as an independent scientific domain. It may seem surprising that many of the so-called 'state-of-the-art' AI algorithms are improved versions of the ideas of that time.

From the beginning, AI could be categorized into two sub-domains, each with an opposing view on how human cognition arises: *symbolic AI* and c*onnectionism.* Symbolic AI is based on the premise that human thinking can be represented (or even replicated) by symbol manipulation (a sort of symbolic calculus or logics). Symbolic AI methods, include search trees, expert systems, fuzzy systems and heuristics. Game playing and path planning are among the first successful applications within the symbolic framework. Connectionists, on the other hand, believe that cognitive processes result from a massive network or interconnected computational units (neurons), linked to each other through channels of different intensity (synapses). Thus for the connectionist, cognition is an emergent process resulting from a complex system (a neural network).

Both approaches gradually drifted from the original aim of achieving human cognition and/or intelligence. That goal belongs today to a different domain known as Artificial General Intelligence (AGI). Research in AGI is very limited today, as the driving force of modern AI is based on practical applications. Nevertheless, serious AGI research still exists today.

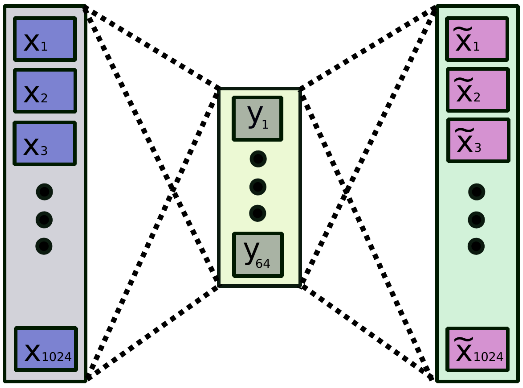
Since the 1980's, and especially during the last ten years, the connectionist approach has monopolized most of the interest, but this has not been always de case; AI has gone through various hype cycles and disillusionment periods through its history. In the last ten years, the connectionist approach has yielded impressive results in computer vision, game playing, recommender systems and natural language processing (NLP). Most of these applications are within the AI techniques known as *machine learning*.

Difference between AI and Machine Learning (ML)

ML corresponds to a subclass of AI algorithms that are able to learn autonomously. These algorithms extract their knowledge from a dataset or from sensor data collected in real time. This is done through a process called *training*, in which the algorithm iteratively browses through the available data while adjusting its own parameters. These algorithms range from the simple linear regression model to connectionist structures – i.e., artificial neural networks

(ANNs). Some ANNs display extremely complex architectures, although in general all these algorithms are trained through similar principles.

On contrast, symbolic AI techniques involve a human explicitly programming the rules of the algorithm for a particular task.



ML Basics

Hype headlines in the media may mislead us to believe that the autonomy in learning displayed by ML is roughly equivalent to human cognition. In reality it is not so. It is important to remember that these algorithms are trained by people to perform specific tasks (e.g., image classification), and that specific data must be given to the algorithm according to the task at hand.

One can think of a ML algorithm as a device that records a certain aspect of human intelligence, and then is able to reproduce it, within the same restricted environment where the algorithm was trained. ML algorithms act by borrowing an aspect of human knowledge and don't generalize well beyond that particular application domain.

Taxonomy of ML techniques

ML can be subdivided into three main subcategories (*supervised learning*, *unsupervised learning* and *reinforcement learning*).

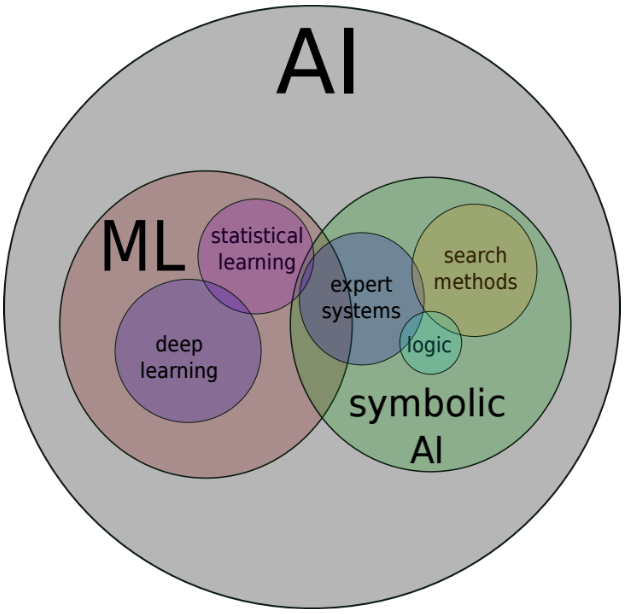
i. *Supervised learning:* Supervised algorithms are trained from pairs of inputs and correct outputs (labels). Thus the trained model is able to predict the outputs from new data, for which the outputs are not known. These predictions can be a regression task – when the prediction is a real value (e.g., stock price, temperature) – or a classification task – when the output is an integer denoting a specific category or class. The input varies according to the task (image data, time series, etc.). Supervised learning is the most common approach, but it requires a labelled data set. An example of supervised learning is face recognition.

ii. *Unsupervised learning:* Unsupervised models are trained only with inputs without labels. A canonical example is *cluster analysis*, that chunks data measurements into different groups, for which a human will attempt to assign a meaning. A novel unsupervised technique is that of Generative Adversarial Networks (GANs), partly responsible of the deepfake craze (Sec. 6). Because humans are able to learn in an unsupervised manner, there is increasing interest in academia towards unsupervised learning.

iii. *Reinforcement learning:* This method is also inspired in the way in which humans learn. In reinforcement learning, an agent interacts with the environment (simulated or otherwise) by collecting sensor data. The algorithm makes a series of choices which result in better or worse outcomes. Good outcomes give a higher reward and reinforce certain types of behaviour. Bad outcomes discourage the actions that led to them. Reinforcement learning gained large attention in 2016, when it was successfully employed to defeat the 18-time Go world champion by 4-1. Go is a more complex game than chess. Indeed, the computer that defeated Garry Kasparov in 1997 in chess employed mostly symbolic AI search methods, not suitable for the larger number of choices in each turn of Go match.

Further nomenclature includes the term *deep learning*, which originally refereed to ANNs with two or more hidden layers. As most ANNs satisfy that criteria nowadays, *deep learning* has become a synonym of ANNs in ML, independently of the training method being supervised, unsupervised o through a reward system.

Other ML learning techniques – not based on ANNs but on probability theory and statistics – are generally referred to as *statistical learning*.



Limitations of ML

ML, and AI in general, are still relatively new research fields and are not necessarily the solution to every technological and scientific problem. While it is expected that the presence of AI in the industry will largely increase in the near future, there is a broad spectrum of problems for which AI may not be yet a satisfactory solution.

In particular ML is deficient at problems which cannot be accurately described into a series of well understood rules, in scenarios with a high degree of uncertainty and when it does not exist a rich dataset representative of the problem at hand. As a rule of thumb, it is good to remember that ML is largely based on optimization methods. Thus in most cases a well-defined loss function – specifying the objective of the learning algorithm – is required. While for some tasks clear objectives are more easily delineated (chess playing, image classification, data compression, etc.), many real life scenarios involve a large number of conflictive objectives. Indeed, a part of the human brain is dedicated to handle conflicts resulting from perception cues: the *anterior cingulate cortex*. Current scientific development is not even close to replicate such brain functionality with an algorithmic counterpart. In addition, ML is not good at handling low probability situations. After all, these algorithms only learn from data, and low probability outcomes are typically under-represented in a dataset. On the contrary, humans are very efficient at learning from exceptions and from low probability events. The brain is adapted to detect novelty.

For these reasons, it is highly unlikely that humans are going to be replaced by AI systems in the near future, at least in most of the tasks they perform. It has long been regarded that for most critical tasks, the human cannot be removed from the control loop, and that the design of machine intelligence has to be directed towards unburdening the human operator, instead of replacing them. As an example, path planning with accurate geographical information is a solved problem (rich datasets, clear goals and low uncertainty). On the other hand, driving a car is a high uncertainty task consisting in multiple goals, some of them contradicting each other (e.g., keeping a vehicle within the lane and at the same time exiting a motorway). Further, it is difficult to create a representative dataset for every driving scenario and to replicate every low probability accident. Thus, while there has been a lot of progress in driving automation, fully autonomous cars won't pervade the roads in the near future.

The ML Landscape Beyond the Hype

Beyond its limitations, there are numerous applications in which ML is the right solution and can bring benefit to the industry. A recent report by the McKinsey Global Institute predicted that by 2030 AI will deliver $13 trillion in economic output worldwide. AI (mostly ML) is already yielding large profits through recommender systems, speech recognition, prediction algorithms, search engines and financial services to name a few. Not in vain, the last Turing Award (2018) was endowed to three eminent architects of the deep learning framework: Geoffrey Hinton, Yann Lecun and Yoshua Bengio. Further, the education industry all over the word is incorporating programs to teach ML skills. Copious research projects are funded by governments in Europe, the U.S.A, Japan and China. This is not only changing higher education institutions, but has also resulted in profitable online education programs and in many companies building in-house AI teams.

The near future of AI appears to be bright too. There are high hopes in new AI developments in simulation oriented design, algorithms that generalize better (or that learn from exceptions), unsupervised learning, and on the development or more ethical ML algorithms. Opening frontiers are drug discovery and AI designed biorobots.

Likewise, there has been recent debate on the lack of attention that symbolic approaches have received during the last years. Many researchers believe that the field will benefit of the development of effective hybrid symbolic-connectionist methods. Although present research trends tend to exclude symbolic methods, it is important to remember than many early implementations of AI into the real world belong to symbolic approaches. Consider for example the Sendai subway system in Japan, which uses a fuzzy system (a type of expert system) to control its acceleration, resulting in smoother and more energy efficient speed transitions.

How to bring benefit from ML in industry?

AI is a field that has won prominence very rapidly. Hence industry managing staff may feel baffled by a technology that was barely taught when they were students. Due to this fast development, a considerable amount of new jargon has emerged (including job roles) which may add more confusion to people responsible of recruiting and organizing teams. To mitigate this gap, a number of reports and online quality content have been recently released.

The table below displays a hierarchy of the different ML roles. Further details can be found in a report released by Workera in 2019.

role	tasks	
Data analyst	Data analysts perform tasks related to data preparation. They also analyse data with a business oriented perspective.	
Data scientist	Besides the tasks of a data analyst, a data scientist is skilled at statistical inference and statistical learning methods.	
ML engineer	A ML engineer performs data preparation, modelling – through statistical learning and deep learning – and deploys the trained algorithms into actual products (e.g., a recommender system in an online platform, a face recognition system into a smartphone...).	
ML researcher	The above categories are typically based on the industry. ML researchers, on the other hand, may also be based in academia. Depending on where they are based, their responsibility and role vary:	
	Industry	**Academia**
	Industry ML researchers perform modelling (statistical learning, reinforcement learning and deep learning). Their work involves finding the most suitable models for a particular problem, adapting those models to a particular task and even developing new algorithms.	In the academia, an ML researcher performs similar tasks than those of an ML researcher in the industry. The difference is that their main objective is to find or test new techniques, and to study the scientific and mathematical underpinnings of AI.

Another report, published by Landing AI, offers practical directions on how to reorient business strategy to address the emergence of ML. These include how to create an in-house ML team. In addition, there exist quality online programs that introduce AI and ML to people without the required mathematical and technical skills, such as 'AI for Everyone' and 'Elements of AI'.

Dangers and ethical considerations.

AI must be used wisely, i.e., with a human always in the control loop. In general, it seems more pragmatic to direct research efforts towards collaborative human-machine systems, instead of trying to tackle the problem of full automation directly. Unrealistic expectations often result in disappointment, and may give raise to a new *AI winter.* The high institutionalization of today's scientific community, in combination with the ever larger journalist hunger to produce catchy slogans, repeatedly generates exaggerated claims. Researchers usually feel pressurized to magnify their claims when writing research proposals, so that they are able to compete with other researchers doing exactly the same thing. These claims are always believed and echoed in the media, and sometimes in the industry. Most of the times, they constitute the only information that reaches to the lay person. Further, in the public imagination still pervades the science fiction picture of an AI turning against humans and ending civilization. At the same time, more realistic dangers resulting from the usage of this technology are partially being ignored. Some of these risks are:

i. *Algorithmic bias*: ML learns from data, thus it may incorporate any discriminatory biases that are present in the dataset. For example, search engines may display different search results to different ethnic groups. Search engines also tend to bubble individuals into their own opinions by suggesting results according to their search history. Alternative search engines exist to address these shortcomings.

ii. *Adversarial attacks*: ML algorithms can be cheated more easily than we tend to think. Adversarial attacks consist in using an optimization algorithm to find specific inputs which make an ANN err. Examples include: putting stickers intro traffic signal that completely confuse a computer vision system, and fooling face detection AI by wearing accessories with certain patterns.

iii. *Deepfakes*: D*eepfakes* are fake images or videos produced by generative ANNs. The results can be very impressive. There has been concern that deepfakes could even be used to manipulate elections by spreading false claims through fake but realistic video segments.

iv. *Fake News*: New and powerful NLP algorithms, such as GPT-2 and BERT, have brought concern about the possibility of using these tools to generate realistic and credible fake news articles. At the same time, this technology can be used to detect the presence of fake news in social media platforms too.

v. *Privacy concerns*: London, among other cities, is implementing AI face recognition systems in already existing CCTV networks. Although the right use of this technology in a democratic country may only result in increased security, it is not known how this technology will actually be used in an unseen future with a potentially different political framework.

These challenges exemplify how necessary is to assess ethical issues derived from new technologies in an ever-changing world.

Useful information sources

- This Guide has been funded under the RAEng Regional Engagement Award.

Guide written by:
Miguel Martínez García, Demetrios Joannou, Roy S. Kalawsky,
Loughborough University
Contact details:
M.Martinez-Garcia@lboro.ac.uk
d.joannou@lboro.ac.uk
r.s.kalawsky@lboro.ac.uk

Loughborough University