How Artificial Intelligence, Machine Learning and Deep Learning are Radically Different?

Article *in* International Journal of Advanced Research in Computer Science and Software Engineering · March 2018

DOI: 10.23956/ijarcsse.v8i2.569

Sanjay Tiwari
Pandit Ravishankar Shukla University

International Journals of Advanced Research in Computer Science and Software Engineering
ISSN: 2277-128X (Volume-8, Issue-2)

Research Article **February 2018**

How Artificial Intelligence, Machine Learning and Deep Learning are Radically Different?

Tanya Tiwari
Samsung India Electronics Pvt. Ltd, Plot No 2A, Sector 126 Noida-201304 Uttar Pradesh, India
tt14ryp@gmail.com

Tanuj Tiwari
Department of Computer Science, University of Texas Dallas Richardson TX 75080 USA
tanuj98@gmail.com

Sanjay Tiwari
SoS in Electronics, Pt. Ravishankar Shukla University, Raipur, Chhattisgarh, India
drsanjaytiwari@gmail.com

Abstract: There is a lot of confusion these days about Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL). A computer system able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages. Artificial Intelligence has made it possible. Deep learning is a subset of machine learning, and machine learning is a subset of AI, which is an umbrella term for any computer program that does something smart. In other words, all machine learning is AI, but not all AI is machine learning, and so forth. Machine Learning represents a key evolution in the fields of computer science, data analysis, software engineering, and artificial intelligence. Machine learning (ML)is a vibrant field of research, with a range of exciting areas for further development across different methods and applications. These areas include algorithmic interpretability, robustness, privacy, fairness, inference of causality, human-machine interaction, and security. The goal of ML is never to make "perfect" guesses, because ML deals in domains where there is no such thing. The goal is to make guesses that are good enough to be useful. Deep learning is a particular kind of machine learning that achieves great power and flexibility by learning to represent the world as nested hierarchy of concepts, with each concept defined in relation to simpler concepts, and more abstract representations computed in terms of less abstract ones. This paper gives an overview of artificial intelligence, machine learning & deep learning techniques and compare these techniques.

Key words: Artificial Intelligence, Machine Learning, Deep Learning

I. INTRODUCTION

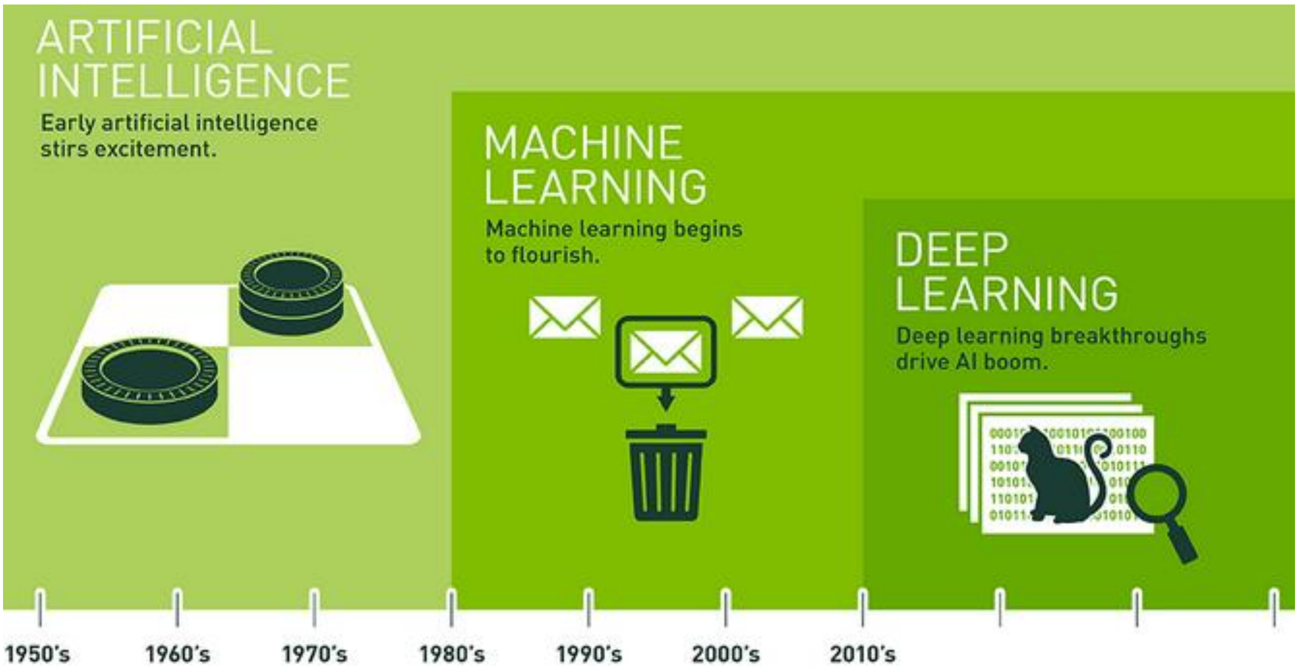
John McCarthy, widely recognized as one of the godfathers of Artificial Intelligence (AI), defined it as "the science and engineering of making intelligent machines. AI is also defined as a branch of computer science dealing with the simulation of intelligent behavior in computers.

"Machine learning is a branch of artificial intelligence that allows computer systems to learn directly from examples, data, and experience. Through enabling computers to perform specific tasks intelligently, machine learning systems can carry out complex processes by learning from data, rather than following pre-programmed rules.

AI has been part of our imaginations and simmering in research labs since a handful of computer scientists rallied around the term at the Dartmouth Conferences in 1956 and birthed the field of AI. In the decades since, AI has alternately been heralded as the key to our civilization's brightest future, and tossed on technology's trash heap as a harebrained notion of over-reaching propellerheads. Frankly, until 2012, it was a bit of both.

Over the past few years AI has exploded, and especially since 2015. Much of that has to do with the wide availability of GPUs that make parallel processing ever faster, cheaper, and more powerful. It also has to do with the simultaneous one-two punch of practically infinite storage and a flood of data of every stripe (that whole Big Data movement) – images, text, transactions, mapping data, you name it.

Machine learning is a technology that allows computers to learn directly from examples and experience in the form of data[1-4]. Traditional approaches to programming rely on hardcoded rules, which set out how to solve a problem, step-by-step. In contrast, machine learning systems are set a task, and given a large amount of data to use as examples of how this task can be achieved or from which to detect patterns. The system then learns how best to achieve the desired output. It can be thought of as narrow AI: machine learning supports intelligent systems, which are able to learn a particular function, given a specific set of data to learn from[5-8]. AI is the all-encompassing umbrella that covers everything from Good Old Fashion AI (GOFAI) all the way to connectionist architectures like Deep Learning. ML is a sub-field of AI that covers anything that has to do with the study of learning algorithms by training with data. There are whole swaths (not swatches) of techniques that have been developed over the years like Linear Regression, K-means, Decision Trees, Random Forest, PCA, SVM and finally Artificial Neural Networks (ANN). Artificial Neural Networks is where the field of Deep Learning had its genesis from.



Recent years have seen exciting advances in machine learning, which have raised its capabilities across a suite of applications. Increasing data availability has allowed machine learning systems to be trained on a large pool of examples, while increasing computer processing power has supported the analytical capabilities of these systems. Within the field itself there have also been algorithmic advances, which have given machine learning greater power. As a result of these advances, systems which only a few years ago performed at noticeably below-human levels can now outperform humans at some specific tasks. Many people now interact with systems based on machine learning every day, for example in image recognition systems, such as those used on social media; voice recognition systems, used by virtual personal assistants; and recommender systems, such as those used by online retailers. As the field develops further, machine learning shows promise of supporting potentially transformative advances in a range of areas, and the social and economic opportunities which follow are significant. In healthcare, machine learning is creating systems that can help doctors give more accurate or effective diagnoses for certain conditions. In transport, it is supporting the development of autonomous vehicles, and helping to make existing transport networks more efficient. For public services it has the potential to target support more effectively to those in need, or to tailor services to users. And in science, machine learning is helping to make sense of the vast amount of data available to researchers today, offering new insights into biology, physics, medicine, the social sciences, and more.

- Machine-learning technology powers many aspects of modern society: from web searches to content filtering on social net- works to recommendations on e-commerce websites, and it is increasingly present in consumer products such as cameras and smartphones.
- Machine-learning systems are used to identify objects in images, transcribe speech into text, match news items, posts or products with users' interests, and select relevant results of search. Increasingly, these applications make use of a class of techniques called deep learning.

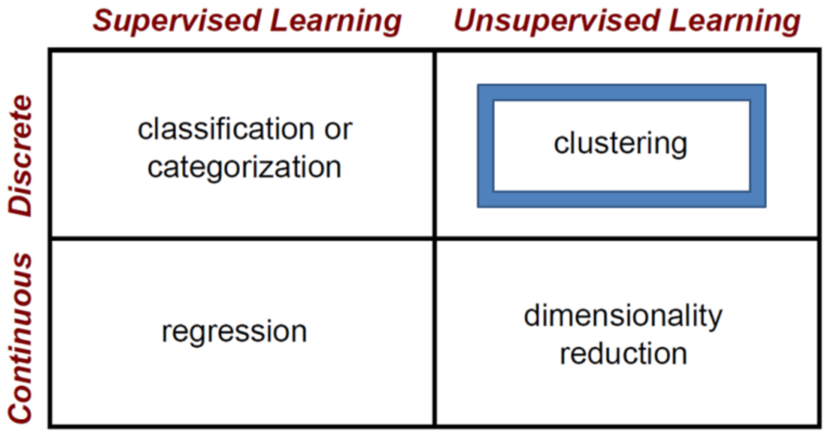
  Machine learning can be used to extract knowledge from data,learn tasks that are difficult to formalise and create software that improves over time
- Conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, constructing a pattern-recognition or machine-learning system required careful engineering and considerable domain expertise to design a feature extractor that transformed the raw data (such as the pixel values of an image) into a suitable internal representation or feature vector from which the learning subsystem, often a classifier, could detect or classify patterns in the input.

II. MACHINE LEARNING METHODS

Learning in MC involves learning general models from data, Data is cheap and abundant, Knowledge is expensive and scarce, Customer transactions to computer behavior and build a model that is a good and useful approximation to the data

In machine learning, tasks are generally classified into broad categories. These categories are based on how learning is received or how feedback on the learning is given to the system developed. Among the different types of Machine learning categories, a crucial distinction is drawn between supervised and unsupervised learning:

- **Supervised machine learning:** The program is "trained" on a pre-defined set of "training examples", which then facilitate its ability to reach an accurate conclusion when given new data.
- **Unsupervised machine learning:** The program is given a bunch of data and must find patterns and relationships therein.



2.1 Supervised Machine Learning

n supervised learning, the computer is provided with example inputs that are labeled with their desired outputs. The purpose of this method is for the algorithm to be able to "learn" by comparing its actual output with the "taught" outputs to find errors, and modify the model accordingly. Supervised learning therefore uses patterns to predict label values on additional unlabeled data. In the majority of supervised learning applications, the ultimate goal is to develop a finely tuned predictor function $h(x)$ (sometimes called the "hypothesis"). "Learning" consists of using sophisticated mathematical algorithms to optimize this function so that, given input data x about a certain domain (say, square footage of a house), it will accurately predict some interesting value $h(x)$ (say, market price for said house).

In practice, x almost always represents multiple data points. So, for example, a housing price predictor might take not only square-footage ($x1$) but also number of bedrooms ($x2$), number of bathrooms ($x3$), number of floors ($x4$), year built ($x5$), zip code ($x6$), and so forth. Determining which inputs to use is an important part of ML design. However, for the sake of explanation, it is easiest to assume a single input value is used.

2.2 Unsupervised Learning

In unsupervised learning, data is unlabeled, so the learning algorithm is left to find commonalities among its input data. As unlabeled data are more abundant than labeled data, machine learning methods that facilitate unsupervised learning are particularly valuable.

The goal of unsupervised learning may be as straightforward as discovering hidden patterns within a dataset, but it may also have a goal of feature learning, which allows the computational machine to automatically discover the representations that are needed to classify raw data.

Unsupervised learning is commonly used for transactional data. You may have a large dataset of customers and their purchases, but as a human you will likely not be able to make sense of what similar attributes can be drawn from customer profiles and their types of purchases. With this data fed into an unsupervised learning algorithm, it may be determined that women of a certain age range who buy unscented soaps are likely to be pregnant, and therefore a marketing campaign related to pregnancy and baby products can be targeted to this audience in order to increase their number of purchases.

Without being told a "correct" answer, unsupervised learning methods can look at complex data that is more expansive and seemingly unrelated in order to organize it in potentially meaningful ways. Unsupervised learning is often used for anomaly detection including for fraudulent credit card purchases, and recommender systems that recommend what products to buy next. In unsupervised learning, untagged photos of dogs can be used as input data for the algorithm to find likenesses and classify dog photos together.

2.3 Representation Learning

It is a set of methods that allows a machine to be fed with raw data and to automatically discover the representations needed for detection or classification. Deep-learning methods are representation-learning methods with multiple levels of representation, obtained by composing simple but non-linear modules that each transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level. With the composition of enough such transformations, very complex functions can be learned. For classification tasks, higher layers of representation amplify aspects of the input that are important for discrimination and suppress irrelevant variations. An image, for example, comes in the form of an array of pixel values, and the learned features in the first layer of representation typically represent the presence or absence of edges at particular orientations and locations in the image. The second layer typically detects motifs by spotting particular arrangements of edges, regardless of small variations in the edge positions. The third layer may assemble motifs into larger combinations that correspond to parts of familiar objects, and subsequent layers would detect objects as combinations of these parts. The key aspect of deep learning is that these layers of features are not designed by human engineers: they are learned from data using a general-purpose learning procedure.

2.4 Supervised Learning.

The most common form of machine learning, deep or not, is supervised learning. Imagine that we want to build a system that can classify images as containing, say, a house, a car, a person or a pet. We first collect a large data set of images of houses, cars, people and pets, each labelled with its category. During training, the machine is shown an image and produces an output in the form of a vector of scores, one for each category. We want the desired category to have the highest score of all categories, but this is unlikely to happen before training. We compute an objective function that measures the error (or distance) between the output scores and the desired pattern of scores. The machine then modifies its internal adjustable parameters to reduce this error. These adjustable parameters, often called weights, are real numbers that can be seen as 'knobs' that define the input–output function of the machine. In a typical deep-learning system, there may be hundreds of millions of these adjustable weights, and hundreds of millions of labelled examples with which to train the machine.

To properly adjust the weight vector, the learning algorithm computes a gradient vector that, for each weight, indicates by what amount the error would increase or decrease if the weight were increased by a tiny amount. The weight vector is then adjusted in the opposite direction to the gradient vector.

Many of the current practical applications of machine learning use linear classifiers on top of hand-engineered features. A two-class linear classifier computes a weighted sum of the feature vector components. If the weighted sum is above a threshold, the input is classified as belonging to a particular category.

Since the 1960s we have known that linear classifiers can only carve their input space into very simple regions, namely half-spaces separated by a hyperplane. But problems such as image and speech recognition require the input–output function to be insensitive to irrelevant variations of the input, such as variations in position, orientation or illumination of an object, or variations in the pitch or accent of speech, while being very sensitive to particular minute variations (for example, the difference between a white wolf and a breed of wolf-like white dog called a Samoyed).

To make classifiers more powerful, one can use generic non-linear features, as with kernel methods, but generic features such as those arising with the Gaussian kernel do not allow the learner to generalize well far from the training examples. The conventional option is to hand design good feature extractors, which requires a considerable amount of engineering skill and domain expertise. But this can all be avoided if good features can be learned automatically using a general-purpose learning procedure. ***This is the key advantage of deep learning***.

We think that deep learning will have many more successes in the near future because it requires very little engineering by hand, so it can easily take advantage of increases in the amount of available computation and data. New learning algorithms and architectures that are currently being developed for deep neural networks will only accelerate this progress.

Deep learning is a very effective supervised and (kind of) unsupervised machine learning method, for certain areas, for example very good at fuzzy recognition of objects/patterns in images. A deep-learning architecture is a multilayer stack of simple modules, all (or most) of which are subject to learning, and many of which compute non-linear input–output mappings. Each module in the stack transforms its input to increase both the selectivity and the invariance of the representation. With multiple non-linear layers, say a depth of 5 to 20, a system can implement extremely intricate functions of its inputs that are simultaneously sensitive to minute details — distinguishing Samoyeds from white wolves — and insensitive to large irrelevant variations such as the background, pose, lighting and surrounding objects.

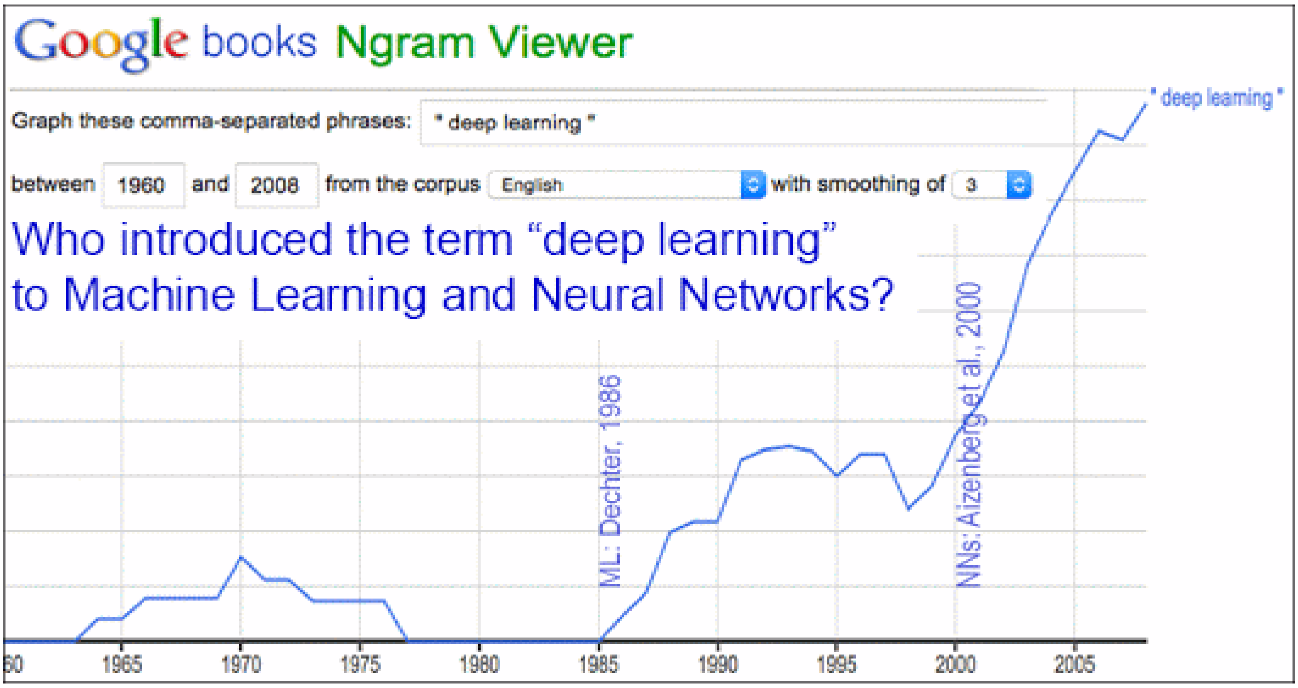
Backpropagation. The backpropagation procedure to compute the gradient of an objective function with respect to the weights of a multilayer stack of modules is nothing more than a practical application of the chain rule for derivatives.

The key insight is that the derivative (or gradient) of the objective with respect to the input of a module can be computed by working backwards from the gradient with respect to the output of that module (or the input of the subsequent module) . The backpropagation equation can be applied repeatedly to propagate gradients through all modules, starting from the output at the top (where the network produces its prediction) all the way to the bottom (where the external input is fed). Once these gradients have been computed, it is straightforward to compute the gradients with respect to the weights of each module.

There was, however, one particular type of deep, feedforward network that was much easier to train and generalized much better than networks with full connectivity between adjacent layers. This was the convolutional neural network (ConvNet). It achieved many practical successes during the period when neural networks were out of favour and it has recently been widely adopted by the computer- vision community.

III. CONVOLUTIONAL NEURAL NETWORKS.

ConvNets are designed to process data that come in the form of multiple arrays, for example a colour image composed of three 2D arrays containing pixel intensities in the three colour channels. Many data modalities are in the form of multiple arrays: 1D for signals and sequences, including language; 2D for images or audio spectrograms; and 3D for video or volumetric images. There are four key ideas behind ConvNets that take advantage of the properties of natural signals: local connections, shared weights, pooling and the use of many layers.



Deep Learning, as the term "deep" specifies is inspired by the human brain and it consists of artificial neural networks (ANN) that are modelled on a similar architecture present in the human brain. In Deep Learning, the learning is performed through a deep and multi-layered "network" of interconnected "neurons". The term "deep" usually refers to the number of hidden layers in the neural network. According to a Mathwork blog, traditional neural networks only contain 2-3 hidden layers, while deep networks can have as many as 150. In 2006, Geoffrey Hinton coined the term "deep learning" to explain new algorithms that allow computers to distinguish objects and text in images and videos. Deep-learning theory shows that deep nets have two different exponential advantages over classic learning algorithms that do not use distributed representations. Both of these advantages arise from the power of composition and depend on the underlying data-generating distribution having an appropriate componential structure. First, learning distributed representations enable generalization to new combinations of the values of learned features beyond those seen during training (for example, 2n combinations are possible with n binary features). Second, composing layers of representation in a deep net brings the potential for another exponential advantage (exponential in the depth).

The recent success of machine learning owes no small part to the explosion of data that is available in some areas, such as image or speech recognition. This data has provided a vast number of examples, which machine learning systems can use to improve their performance. In turn, machine learning can help address the social and economic benefits expected from so-called 'big data', by extracting valuable information through advanced data analytics. Supporting the development of this function for machine learning requires an amenable data environment, based on open standards and frameworks or behaviours to ensure data availability across sectors. As machine learning systems become more ubiquitous, or significant in certain fields, three skills needs follow. Firstly, as daily interactions with machine learning become the norm for most people, a basic understanding of the use of data and these systems will become an important tool required by people of all ages and backgrounds. Introducing key concepts in machine learning at school can help ensure this. Secondly, to ensure that a range of sectors and professions have the absorptive capacity to use machine learning in ways that are useful for them, new mechanisms are needed to create a pool of informed users or practitioners. Thirdly, further support is needed to build advanced skills in machine learning. There is already high demand for people with advanced skills, with specialists in the field being highly sought after, and additional resources to increase this talent pool are critically needed. 'No regrets' steps in building digital literacy and informed users will also help prepare the UK for possible changes in the employment landscape, as the fields of machine learning, artificial intelligence, and robotics develop. There is a vast range of potential benefits from further uptake of machine learning across industry sectors, and the economic effects of this technology could play a central role in helping to address the UK's productivity gap. Businesses of all sizes across sectors need to have access to appropriate support that helps them to understand the value of data and machine learning to their operations. To meet the demand for machine learning across industry sectors, the UK will need to support an active machine learning sector, which capitalises on the UK's strength in this area, and its relative international competitive advantages. The UK's start-up environment has nurtured a number of high-profile success stories in machine learning, and strategic consideration should be given to how to maximise the value of entrepreneurial activity in this space.

The issue of representation lies at the heart of the debate between the logic-inspired and the neural-network-inspired paradigms for cognition. In the logic-inspired paradigm, an instance of a symbol is something for which the only property is that it is either identical or non-identical to other symbol instances. It has no internal structure that is relevant to its use; and to reason with symbols, they must be bound to the variables in judiciously chosen rules of inference. ***By contrast, neural networks just use big activity vectors, big weight matrices and scalar non-linearities to perform the type of fast 'intuitive' inference that underpins effortless commonsense reasonin***g.

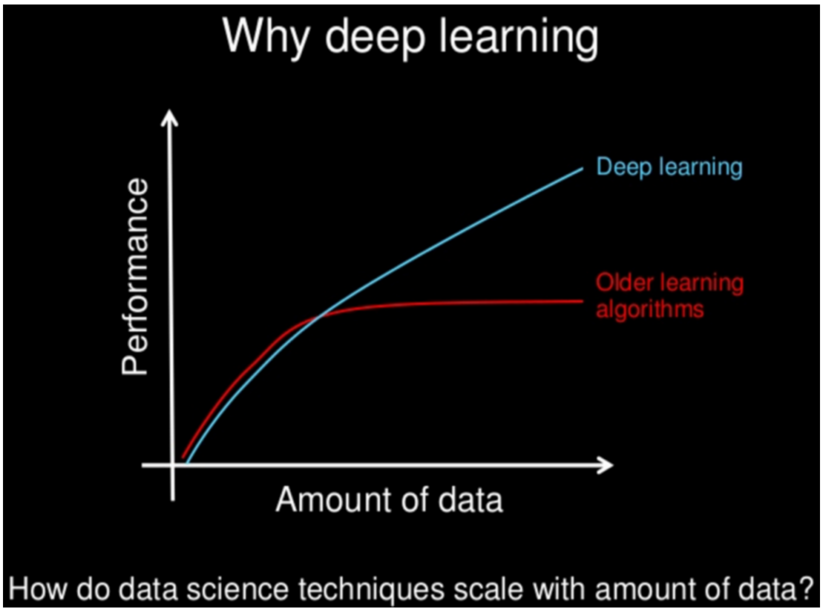
- Before the introduction of neural language models, the standard approach to statistical modelling of language did not exploit distributed representations: it was based on counting frequencies of occurrences of short symbol sequences of length up to N (called N-grams). The number of possible N-grams is on the order of VN, where V is the vocabulary size, so taking into account a context of more than a handful of words would require very large training corpora.
- **Recurrent neural networks.** When backpropagation was first introduced, its most exciting use was for training recurrent neural networks (RNNs). For tasks that involve sequential inputs, such as speech and language, it is often better to use RNNs. ***RNNs process an input sequence one element at a time, maintaining in their hidden units a 'state vector' that implicitly contains information about the history of all the past elements of the sequence***. When we consider the outputs of the hidden units at different discrete time steps as if they were the outputs of different neurons in a deep multilayer network, it becomes clear how we can apply backpropagation to train RNNs.
- RNNs are very powerful dynamic systems, but training them has proved to be problematic because the backpropagated gradients either grow or shrink at each time step, so over many time steps they typically explode or vanish.

IV. COMPARISON BETWEEN MACHINE LEARNING AND DEEP LEARNING

Over the years, machine learning has evolved in its ability to crunch huge volumes of data and is widely used in everyday applications technology. ML powers many aspects of our daily interaction –from spam filtering to content filtering on social networks, recommendations on e-commerce sites, and it is increasingly present in consumer products such as cameras and smartphones, the speech recognition (as in Siri, Apple's voice assistant) and handwriting recognition (Optical Character Recognition). The main differences between Machine Learning and Deep Learning is illustrated in following paragraph:

1. Data dependencies

The performance of deep learning and traditional machine learning varies drastically as the scale of data increases. Deep learning algorithms does not perform that well when the data is small as its algorithms need a large amount of data to understand it perfectly. In contrary, traditional machine learning algorithms with their handcrafted rules works well even when data is small as shown in Fig



2. Hardware dependencies

Machine learning algorithms, can work efficiently on low-end machines as they include GPUs which are an integral part of its working and inherently do a large amount of matrix multiplication operations whereas Deep learning algorithms heavily depend on high-end machines.

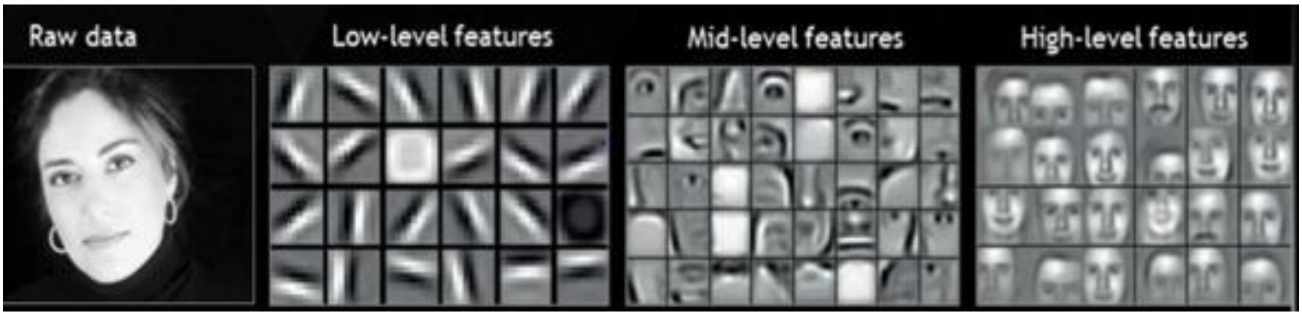
3. Feature extraction

Feature extraction is a process of putting domain knowledge into the creation of feature extractors to reduce the complexity of the data and make patterns more visible to learning algorithms to work. This process is difficult and expensive in terms of time and expertise.

Most of the applied features need to be identified by an expert and then hand-coded as per the domain and data type in Machine learning. The features can be pixel values, shape, textures, position and orientation. The performance of most of the Machine Learning algorithm depends on how accurately the features are identified and extracted.

Deep learning algorithms extract high-level features from data. This is major feature of Deep Learning which is major advantage over traditional Machine Learning. As a result of deep learning reduces the task of developing new feature extractor for every problem.

Many programmers point out that feature extraction is a painstaking process and relies a lot on how insightful the developer is. Hence, for complex problems such as object recognition or handwriting recognition, feature extraction in traditional ML becomes a huge challenge. On the other hand, in Deep Learning, raw data can be fed via neural networks and extract high-level features from the raw data.



4. Problem Solving approach

In machine learning algorithm, the problem solving requires to break the problem down into different parts, solve them individually and combine them to get the result. Deep learning in contrast solves the problem end-to-end. Let us have a task of multiple object detection where it is to be identified that what is the object and where is it present in the image.

In a typical machine learning approach, you would divide the problem into two steps, object detection and object recognition. First, you would use a bounding box detection algorithm like grab cut, to skim through the image and find all the possible objects. Then of all the recognized objects, you would then use object recognition algorithm like SVM with HOG to recognize relevant objectswhereas in deep learning approach, end-to-end process is done. For example, in a YOLO net which is a type of deep learning algorithm, when an image is passed, it would give out the location along with the name of object.

5. Execution Time

In deep learning algorithm, there are large number of parameters so takes a long time to train. State of the art deep learning algorithm ResNet takes about two weeks to train completely for beginners. In contrast, machine learning comparatively takes much less time to train, ranging from a few seconds to a few hours.

The deep learning algorithm takes much less time to run on testing time. If compare it with a type of machine learning algorithm as k-nearest neighbors , test time increases on increasing the size of data. This is not true for all machine learning algorithms, as some of them have small testing times too.

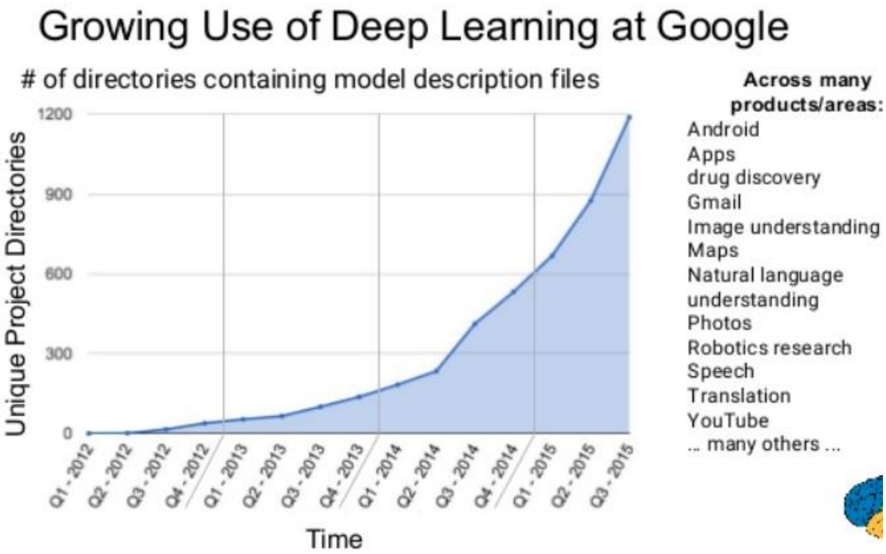
6. Interpretability

Interpretability is an important factor for comparison of machine learning and deep learning and due to this factor deep learning is still thought seriously before its use in industry.

Let us explain by an example of using deep learning to give automated scoring to essays. The performance it gives in scoring is quite excellent and is near human performance. But It does not reveal why it has given that score. Indeed mathematically you can find out which nodes of a deep neural network were activated, but we don't know what there neurons were supposed to model and what these layers of neurons were doing collectivelyand fail to interpret the results.

In contrast, machine learning algorithms like decision trees give us crisp rules as to why it chose what it chose, so it is particularly easy to interpret the reasoning behind it. Therefore, algorithms like decision trees and linear/logistic regression are primarily used in industry for interpretability.

Google is applying machine learning in its various products. Applications of Machine Learning/Deep Learning are endless



V. CONCLUSION

- Machine learning models hit a ground in analytics performance due to data processing limitations, whereas DL algorithms can work at scale and the recent hardware innovations have proven how deep learning training can be reduced dramatically to minutes. There is a lot more going on in deep learning research and DL is poised to make a major impact in areas such as driverless technology, retail, healthcare and its breakthroughs will significantly impact diagnostics. Machine Learning is a method of teaching computers how to perform complex tasks that cannot be easily described or processed by humans and to make predictions. It is a combination of Mathematical Optimization and Statics. In the other hand, Deep Learning is the subset of ML that focus even more narrowly like neuron level to solve any problem.
- Ultimately, major progress in artificial intelligence will come about through systems that combine representation learning with complex reasoning. Although deep learning and simple reasoning have been used for speech and handwriting recognition for a long time, new paradigms are needed to replace rule-based manipulation of symbolic expressions by operations on large vectors.
- Deep learning is part of a broader family of machine learning methods based on learning representations of data. Deep learning is surprising us each and every day, and will continue to do so in the near future. This is because Deep Learning is proving to be one of the best technique to be discovered with state-of-the-art performances.
- Research is continuous in Machine Learning and Deep Learning. But unlike in previous years, where research was limited to academia, research in Machine Learning and Deep Learning is exploding in both industry and academia. And with more funds available than ever before, it is more likely to be a keynote in human development overall.

General AI is currently impossible and Narrow AI is very difficult .Machine Learning is a way to solve some Narrow AI problems, albeit with hand-coding involved. Deep Learning is an advancement on ML, which again is still Narrow AI .Therefore, in the future we could have General AI!

REFERENCES

[1] Bridge, James P., Sean B. Holden, and Lawrence C. Paulson. "Machine learning for first-order theorem proving." Journal of automated reasoning 53.2 (2014): 141-172.

[2] Loos, Sarah, et al. "Deep Network Guided Proof Search." arXiv preprint arXiv:1701.06972 (2017).

[3] Finnsson, Hilmar, and YngviBjörnsson. "Simulation-Based Approach to General Game Playing." AAAI. Vol. 8. 2008.

[4] Sarikaya, Ruhi, Geoffrey E. Hinton, and AnoopDeoras. "Application of deep belief networks for natural language understanding." IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP) 22.4 (2014): 778-784.

[5] "AI-based translation to soon reach human levels: industry officials". Yonhap news agency. Retrieved 4 Mar 2017.Mohri, Mehryar; Rostamizadeh, Afshin; Talwalkar, Ameet (2012). Foundations of Machine Learning. USA, Massachusetts: MIT Press. ISBN 9780262018258.

[6] Zhong, Sheng-hua; Liu, Yan; Liu, Yang (2011). "Bilinear Deep Learning for Image Classification". Proceedings of the 19th ACM International Conference on Multimedia. MM '11. New York, NY, USA: ACM: 343–352. doi:10.1145/2072298.2072344. ISBN 9781450306164.

[7] Szegedy, Christian; Toshev, Alexander; Erhan, Dumitru (2013). "Deep neural networks for object detection". Advances in Neural Information Processing Systems.

[8] Deng, L.; Yu, D. (2014). "Deep Learning: Methods and Applications" (PDF). Foundations and Trends in Signal Processing. **7** (3–4): 1–199. doi:10.1561/2000000039.