Shallow Learning Vs Deep Learning

Dr. Amirthalingam Ramanan *reconnoiters major aspects of machine learning with profound comparison on how they are practically applied in AI.*

Machine Learning (ML) is a sub-field of Artificial Intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Deep Learning (DL) is a subfield of ML concerned with algorithms inspired by the structure and function of the brain called artificial neural networks (ANNs). The term Shallow learning is generally used with those techniques of ML that are not deep. Although ML is a field within Computer Science, it differs from traditional computational approaches. In traditional computing, algorithms are sets of explicitly programmed instructions used by computers to perform calculations or problem-solving. ML algorithms instead facilitate computers in building models from sample data to automate decision-making processes based on data inputs.

The building block of ML has the ability to weigh the features of the data fed to that algorithm to make the most accurate prediction. Computing the weights of these features is a big part of what ML is. A feature is an individual measurable property or characteristic of a phenomenon being observed. The feature extraction in shallow learning is a manual process that requires domain knowledge of the data that we are learning from. In other words, it is a type of ML where we learn from the data described by the pre-defined features. On the other hand, in deep learning the feature extraction is algorithmically computed without manual human intervention, i.e., a DL algorithm automatically learns these features along with their weights from raw data with little or no preprocessing.

Shallow learning algorithms vary depending on the nature of prediction that they are trying to make. These algorithms can be grouped as follows:

1. Supervised learning (e.g., classification, regression),

2. Unsupervised learning (e.g., clustering, dimension reduction), and

3. Reinforcement learning (e.g., model-based, model-free).

In supervised learning, a predictive model is used for tasks that involve the prediction of a given output (target/concept) using other variables (features/attributes) in the dataset. i.e., the learning algorithm in a predictive model attempts to discover and model the relationships among the target variable and the other variables. Classification and regression are types of supervised ML algorithms. In classification, the task is to approximate the mapping function of the predictive model from input variables to discrete output variables (e.g. labels). The main goal is to identify which class/category the new data will fall into. The different types of classification algorithms include: Nearest neighbour, Naïve Bayes, Support Vector Machine (SVM), Multilayer Perceptron (MLP), and random decision forest. In regression, algorithms predict a continuous value based on the input variables. The different types of regression algorithms include linear, logistic, and generalised regression.

In unsupervised learning, the understanding of a given data is performed without a target variable. i.e., the learning is concerned with identifying groups in the dataset. The groups may be defined by the rows (i.e., clustering) or the columns (i.e., dimension reduction). However, the

The writer is a Senior Lecturer attached to the Department of Computer Science, University of Jaffna. He earned his PhD in Computer Science from the University of Southampton, UK. His major areas of expertise include Machine Learning, Deep Learning, and Computer Vision.

motive in each case is quite different. Clustering and dimension reduction are types of unsupervised ML algorithms. In clustering, the task is to divide the data points into a number of groups such that data points in the same groups are more similar to other data points in the same group than those in other groups. The different types of clustering algorithms include K-means, mean-shift, hierarchical clustering, affinity propagation, spectral clustering, Gaussian Mixture Models (GMMs), and Latent Dirichlet Allocation (LDA). In dimension reduction, the task is to reduce the higher dimensional data by projecting it to a lower-dimensional subspace which captures the 'essence' of the data. The different types of dimension reduction algorithms include Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Canonical Correlation Analysis (CCA).

In reinforcement learning (RL), an AI agent learns in an interactive environment by trial-and-error using feedback from its own actions and experiences, i.e., it refers to goal-oriented algorithms, which learn how to attain a complex goal over many steps. There are three approaches to implement a reinforcement learning algorithm: value-based, policy-based, and model-based. Generally, model-based learning attempts to model the environment then choose the optimal policy based on its learned model, whereas in model-free learning the agent relies on trial-and-error experience for setting up the optimal policy.

In addition, Semi-supervised learning (SSL) falls between supervised learning and unsupervised learning. i.e., it combines a small amount of labeled data with a large amount of unlabeled data during training which is particularly useful when extracting relevant features from the dataset is difficult, and labeling examples is a time-intensive task for experts. S3VM and Graph-based SSL are example of SSL algorithms.

Deep learning (DL) is a subfield of ML. DL has outperformed ML techniques in many domains, e.g., cybersecurity, natural language processing, bioinformatics, robotics, and medical information processing, among many others. Remarkably, DL has achieved outstanding results on several complex cognitive tasks, matching or even beating those provided by human performance. The reason is that though ML models do become progressively better at whatever their function is, they still need some guidance. With a DL model, an algorithm can determine on its own if a prediction is accurate or not through its own neural network. Since DL involves multiple levels of representation and multiple layers of non-linear processing units (or neurons), it has seemed appropriate to describe them as 'deep'. DL techniques are classified into the following major categories:

1. Deep Supervised learning (e.g., Recurrent Neural Networks [RNNs], Convolutional Neural Networks [CNNs], and Deep Neural Networks [DNNs]),

2. Deep Unsupervised learning (e.g., Generative Adversarial Networks (GANs), restricted Boltzmann machines, and auto-encoders), and

3. Deep Reinforcement learning.

Among the many implementations of DL models, CNNs are particularly suited for several domain of applications. In image classification, CNN performs by discovering low-level features (such as edges and curves) and then building up to more abstract representations through a series of convolutional layers. A commonly used type of CNN consists of numerous convolution layers preceding sub-sampling (pooling) layers, while the ending layers are fully connected layers. Some of the well-known CNN architectures are AlexNet, VGG, GoogleNet, ResNet, Xception, Residual attention neural network, DenseNet, MobileNet, CapsuleNet, and HRNetV2.

In contrast, ML models can be trained with lesser training data which usually takes less time for training and the training can be performed using CPUs, whereas DL models take longer time for training and the proper training is performed using GPUs. DL enables a machine to efficiently analyse problems through its hidden layer architecture which are otherwise far more complex to be programmed manually. Example applications of DL includes Virtual Assistants (e.g., Amazon Alexa, Siri, and Microsoft's Cortana), Advanced driver-assistance systems (e.g., Mobileye) Chatbots, Speech-to-speech translation (e.g., Google's Translatotron), Entertainment (e.g., Netflix, YouTube), and Robotics (e.g., RoboWatch).