International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 03 | Mar 2020 www.irjet.net p-ISSN: 2395-0072

A REVIEW ON MACHINE LEARNING ALGORITHMS AND THEIR APPLICATIONS

A. Kavitha

Assistant Professor of Computer Science, Aditanar College of Arts and Science, Tiruchendur, Tamilnadu, India

---

Abstract- *In the field of computing, the concept of Machine Learning is one of the most exciting technologies that make our work easier in our day to day life. Machine Learning is used to solve real-world problems by storing, manipulating, extracting and retrieving data from large sources. It is a very powerful tool. It focuses on the development of computer programs that can access data and use it learn for themselves without human intervention. In this paper, different types of machine learning algorithms and their applications are discussed.*

Keywords: Machine learning, Supervised, Unsupervised, Reinforcement, Regression, Random forest

1. INTRODUCTION

Traditional Programming refers to manually written program that uses input data and runs on a computer to produce the output. In Machine Learning the input data and output are fed to an algorithm to create a program. The following diagram Fig-1 shows the difference between Traditional Programming and Machine learning.

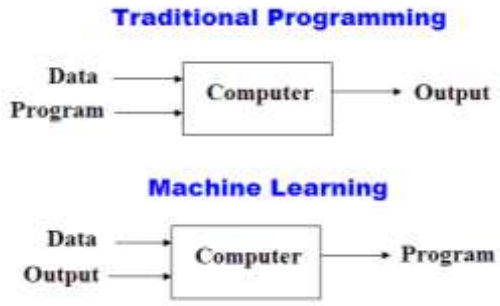

Fig -1 : Difference between Traditional Programming and Machine Learning

Machine learning is an application of AI (Artificial Intelligence) that gives systems the ability to automatically learn and improve from experience without being explicitly programmed. It is used to teach machines the way to handle the data more efficiently. In recent years, machine learning has become more popular in research and is used in large number of applications, including Data mining, Image processing, Speech Recognition, Health care, social network analysis, text mining, video recommendation and so on.

The main objective of Machine learning is development of computer programs that can access data and use it learn for themselves. Learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide [1][2].

2. USES OF MACHINE LEARNING

- **Face Recognition-**The face recognition is one of the immense features that are developed by machine learning only
- **Voice Recognition-**Machine learning helps in voice recognition. It also mentioned as virtual personal assistants. It will assist you to find the information when asked over the voice. After your question, that assistant will look out for the data or the information that has been asked by you and collect the required information to provide you with the best answer.
- **Videos Surveillance-**It helps to detect the crime or any miss happening that is going to happen before it happens. It helps in tracking the unusual behavior of individuals like napping on benches and standing still from a long time, stumbling etc. and it will create an automatic alert to the people who all are posted there and that they can help to avoid any issues or problems.
- **Social Media Platform**-Social Media is used for providing better news feed and advertisement as per the user's interest is especially done through the uses of machine learning only. Some of the examples like friend suggestions, page suggestions for Facebook, songs, and videos suggestion on YouTube.
- **Spam and Malware**-Email clients use variety of spam filtering and these spam filters are continuously getting updated and these are mainly done by the uses of machine learning. Similarly, varieties of malware are detected and these are detected mainly by the system security programs that are mainly helped by machine learning only.
- **Search Engine-** Search engine use variety of machine learning algorithms to response user queries.

- **Applications/Companies**-There are many applications and companies that used machine learning for doing their day to day process because it is being more accurate and precise than manual interventions. These companies are Netflix, facebook, google maps, Gmail, Google search etc [3][4][5].

3. APPLICATIONS OF MACHINE LEARNING

- **Marketing**-Machine learning plays an important role in personalized digital marketing. Ads click predictions, showing relevant Ads to customers, identifying target customers, churn analysis, etc. are some of the important applications of machine learning in the marketing sector.
- **Healthcare**-ML is becoming a fast-growing trend in healthcare. In today's world, machine learning enables us to form data-driven decisions that can prevent diseases, helps in better patient diagnosis, faster root cause detection, etc.
- **Digital Media and Entertainment-**Machine learning has tremendous applications in digital media, social media and entertainment. Personalized recommendation (i.e Youtube video recommendation), user behavior analysis, spam filtering, social media analysis, and monitoring are some of the most important applications of machine learning.
- **E-commerce**-Advancements in machine learning is also a key stakeholder in today's e-commerce transformation. When we are browsing an e-commerce site, we can see personalized recommendations, which are achieved through content-based or collaborative filtering. Perhaps the availability of large scale user data is what keeps e-commerce giants ahead in the race than retailers. Machine learning is also used in fashion designing. Indian E-Commerce giant Myntra has multiple brands that are designed by deep learning systems.
- **Banking and Financial**-In a digital economy, machine learning helps banks and other financial organizations to safeguard from frauds, money laundering, illegal financial detection, identifying valuable customers, etc. It also helps financial organizations with stock market predictions, demand forecasting, offering personalized banking solutions to the customers, etc.
- **Automobile**-Almost every automobile manufacturer is using AI for optimizing fuel consumption, breakdown prediction and even for self-driving.
- **Governance and Surveillance**-Machine learning is reshaping modern Governance and defense systems. With the help of the state of the art deep learning algorithms and infrastructures, security agencies are now enabled with real-time image detection, drone surveillance, automated social network monitoring, etc.
- **Transportation**-Machine learning technology is used in Transportation firms and delivery organizations to carry out data analysis and data modeling to make informed decisions and help their customers make smart decisions when they travel.
- **Oil and Gas**-Machine learning technology is used in oil & gas industry from analyzing underground minerals and finding new energy sources to streaming oil distribution. [3][4][5].

4. TYPES OF MACHINE LEARNING

Machine learning tasks are typically classified into three broad categories: Supervised learning, Unsupervised learning and Reinforcement learning as shown in Fig- 2.

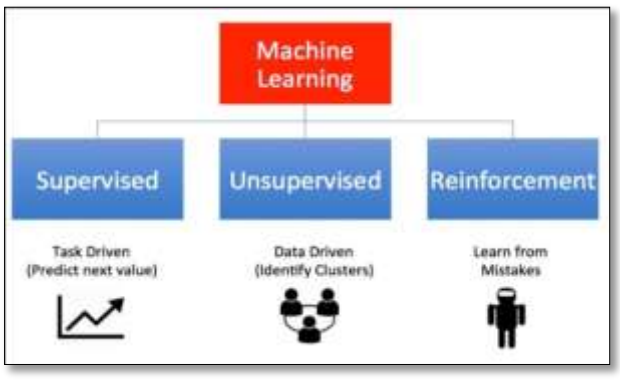

Fig –2: Types of Machine Learning

- **Supervised Learning-** Used when we know the correct answers from past data, but need to predict future outcomes. For example, using past house prices to predict the present and future value. Effectively using a trial and error based statistical improvement process, the machine gradually improves accuracy by testing results against a group of values provided by a supervisor.

- **Unsupervised Learning-** Where there is no distinct correct answer, but we would like to get something new from the data. Most often used to classify or group data, for example, to classify music on Spotify, to help recommend which albums you might listen to. It will then classify the listeners, to see if they're more likely to listen to Radiohead or Justin Bieber [6].
- **Reinforcement Learning-**Reinforcement learning is all about making decisions sequentially. In simple words we will say that the output depends on the state of the current input and the next input depends on the output of the previous input. In Reinforcement learning decision is dependent, so we give labels to sequences of dependent decisions. Example: Chess [7].

5. MACHINE LEARNING ALGORITHMS

Machine Learning algorithm is an evolution of the regular algorithm. It makes our programs "smarter", by allowing them to automatically learn from the data we offer. Machine learning algorithms are classifies as follows

5.1. Supervised Learning Algorithm

This algorithm consists of a target / outcome variable (or dependent variable) which is to be predicted from a given set of predictors (independent variables). Using these set of variables, we generate a function that map inputs to desired outputs. The training process continues until the model achieves a desired level of accuracy on the training data. Examples of Supervised Learning: Linear Regression, Decision Tree, Random Forest, KNN, Logistic Regression etc.

5.1.1. Linear Regression

The methodology for measuring the relationship between the two continuous variables is known as Linear regression. It comprises of two variables such as

Independent Variable – "x"
Dependent Variable – "y"

In a simple linear regression, the predictor value is an independent value that does not have any underlying dependency on any variable. The relationship between x and y is described as follows

$$y = mx + c$$

Here, m is the slope and c is the intercept. Based on this equation, we can calculate the output that will be through the relationship exhibited between the dependent and the independent variable.

5.1.2. Decision Tree

It is a kind of supervised learning algorithm that is mostly used for classification problems. A decision tree makes a tree like structure contains possible solutions to a problem based on certain constraints. It is so named for it begins with a single simple decision or root, which then split into a number of branches until a decision or prediction is made, forming a tree.

For example if you want to go to the market to purchase a shampoo. First, you will analyze if you really do require shampoo. If you run out of it, then you will have to buy it from the market. Furthermore, you will look outside and assess the weather. That is, if it is raining, then you will not go and if it is not, you will. We can visualize this scenario intuitively with the following diagram Fig-3.

Fig -3: Decision Tree Example

With the same principle, we can construct a hierarchical tree to obtain our output through several decisions. There are two procedures towards building a decision tree – Induction and Pruning. In Induction, we build the decision tree and in pruning, we simplify the tree by removing several complexities [8][9][10][11][13].

5.1.3. Random forest

Random forest is a supervised learning algorithm which is used for both classification as well as regression. But however, it is mainly used for classification problems. As we know that a forest is made up of trees and more trees means more robust forest. Similarly, random forest algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting. It is an ensemble method which is better than a single decision tree because it reduces the over-fitting by averaging the result.

We can understand the working of Random Forest algorithm with the help of following steps –

- First, start with the selection of random samples from a given dataset.
- Next, this algorithm will construct a decision tree for every sample.
- Then it will get the prediction result from every decision tree.
- Next voting will be performed for every predicted result.
- At last, select the most voted prediction result as the final prediction result.

5.1.4. kNN (k- Nearest Neighbors)

kNN is one of the many supervised machine learning algorithms that we use for data mining as well as machine learning. Based on the similar data, this classifier then learns the patterns present within. It is a non-parametric and a lazy learning algorithm. By non-parametric, we mean that the assumption for underlying data distribution does not hold valid. In lazy loading, there is no requirement for training data points for generating models. The training data is utilized in testing phase causing the testing phase slower and costlier as compared with the training phase.

5.1.5. Logistic Regression

It is a classification not a regression algorithm. It is used to estimate discrete values (Binary values like 0/1, yes/no, true/false) based on given set of independent variable(s). In simple words, it predicts the probability of occurrence of an event by fitting data to a logistic function. Hence, it is also known as logistic regression. Since, it predicts the probability, its output values lies between 0 and 1 (as expected).

5.2. Unsupervised Learning Algorithm

In this algorithm, we do not have any target or outcome variable to predict / estimate. It is used for clustering population in different groups, which is widely used for segmenting customers in different groups for specific intervention. Examples of Unsupervised Learning: Apriori algorithm, K-means.

5.2.1. Apriori Algorithm

Apriori is especially used for sorting large amounts of data. Sorting data often occurs because of association rules. Rules help show what aspects of data different sets have in common. Categories can then be built around those association rules. With data in categories, algorithms and users can spot new trends and structure data sets. They may have a better ability to point out trends over time. The algorithm can also be used to track how relationships develop and categories are built.

Apriori can be used as a basis for an Artificial Neural Network. It can help the network make sense of large reams of data and sort data into categories by frequency almost instantaneously. An artificial neural network using Apriori can also tweak the weighting on different categories to expand or diminish the importance of those categories. As a result, an artificial neural network can process data, identify trends, and elaborate on patterns that would otherwise be missed. Apriori is incredibly helpful for data analysts in numerous fields. Its importance will only continue to grow as more and more fields use artificial intelligence to make sense of massive data sets. Apriori will continue to be an essential tool in the growth of machine learning and artificial intelligence for years to come.

5.2.2. K-Means

It is a type of unsupervised algorithm which solves the clustering problem. Its procedure follows a simple and easy way to perform partitioning of the data consisting of n values into subsequent k subgroups. Each of the n values with the nearest mean belongs to the k cluster.

Given a group of objects, we perform partitioning of the group into several sub-groups. The sub-groups have a similar basis where the distance of each data point in the sub-group has a meaning related to their centroids. It is the most popular form of unsupervised machine learning algorithm as it is quite easy to comprehend and implement.

The main objective of a K-means clustering algorithm is to reduce the Euclidean Distance to its minimum. This distance is the intra-cluster variance which we minimize using the squared error function.

5.3. Reinforcement Learning Algorithm

Using this algorithm, the machine is trained to make specific decisions. It works this way: the machine is exposed to an environment where it trains itself continually using trial and error. This machine learns from past experience and tries to capture the best possible knowledge to make accurate business decisions. Example of Reinforcement Learning: Markov Decision Process, Evolutionary algorithm [8][9][10][11][13].

5.3.1. Markov Decision Process.

The Markov decision process, better known as MDP, is an approach in reinforcement learning to take decisions in a gridworld environment. A gridworld environment consists of states in the form of grids. The MDP tries to capture a world in the form of a grid by dividing it into states, actions, models/transition models, and rewards.

Markov decision process problems (MDPs) assume a finite number of states and actions. At each time the agent observes a state and executes an action, which incurs intermediate costs to be minimized. The cost and

the successor state depend only on the current state and the chosen action. Successor generation may be probabilistic, based on the uncertainty we have on the environment in which the search takes place. For example, an action might sometimes fail to result in the desired target state, instead staying in the current state with a small probability [12].

5.3.2. Evolutionary algorithm

Evolutionary algorithm starts with code generation entirely at random. Each of these codes is tested to see whether it achieves the required goal. In this way, the code evolves. Overtime, it becomes better, and after many generations, if conditions are right, it can become better than any human coder can design. It consists of maintaining a distribution over network weight values, and having a large number of agents act in parallel using parameters sampled from this distribution. Each agent acts in its own environment, and once it finishes a set number of episodes, or steps of an episode, cumulative reward is returned to the algorithm as a fitness score. With this score, the parameter distribution can be moved toward that of the more successful agents, and away from that of the unsuccessful ones. By repeating this approach millions of times, with hundreds of agents, the weight distribution moves to a space that provides the agents with a good policy for solving the task at hand [13].

6. CONCLUSION

Machine learning is a concept which allows the machine to learn from examples and experience, and that too without being explicitly programmed. Today each and every person is using machine learning knowingly or unknowingly. From purchasing a product in online shopping to updating photos in social networking sites. In this paper, we went through what is machine learning and the variety of machine learning algorithms that are essential for the implementation of machine learning models in the data science industry.

REFERENCES

[1]https://expertsystem.com/machine-learning-definition/
[2] Ayon Dey, Machine Learning Algorithms: A Review, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (3) , 2016, 1174-1179
[3]https://www.educba.com/applications-of-machine-learning/
[4]https://www.educba.com/uses-of-machine-learning/
[5]https://www.outsource2india.com/software/articles/machine-learning-applications-how-it-works-who-uses-it.asp
[6] https://dzone.com/articles/machine-learning-in-plain-english
[7]https://www.geeksforgeeks.org/what-is-reinforcement-learning/
[8]https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/
[9]https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/
[10]https://data-flair.training/blogs/machine-learning-algorithms/
[11]https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_algorithms.htm
[12]https://www.sciencedirect.com/topics/ computer-science/markov-decision-process
[13] Anushree Raj , Rio D. Souza , A Review on Machine Learning Algorithms , International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 7 Issue VI, June 2019, ISSN: 2321-9653

BIOGRAPHY

A.KAVITHA [M.C.A] is presently working as Head & Assistant Professor in the department of Computer Science in Aditanar College Of Arts And Science, Tiruchendur, Tuticorin District, Tamilnadu. She has qualified in Tamilnadu State Eligibility Test (SET) Lectureship in 2016.