International Journal of Advance Research in Science and Engineering
Volume No. 11, Issue No. 01, January 2022
www.ijarse.com
IJARSE
ISSN 2319 - 8354

Analysis of Machine Learning Algorithms and Its Applications

V.Valliappan[1], V.Kabilan[2], K.Srivarshini[3], P.Santhiya[4],

[1,2,3] B.E-CSE\I-year, [4]Assistant Professor

Department of CSE, Builders Engineering College, Kangayam, Tiruppur Dt, Tamilnadu, India.

ABSTRACT

For the past 10 years, machine learning (ML) which is seen as a part of artificial intelligence has been ruling the world. Machine learning is also referred to as predictive analytics and it reduces human time and work. Machine learning is highly efficient and it is the need of the hour. For the past few years, machine learning algorithms have been useful in various fields like Cyber Security, Big Data Analytics., Computer Networks., Cloud Computing., etc. These algorithms are very much useful for analyzing and detecting purposes. This paper clearly explains about the machine learning algorithms and its application.

Keywords: Face detection, Machine learning, Spam detection, supervised learning and unsupervised learning.

INTRODUCTION

Machine Learning is defined as the study of computer programs that influence algorithms and models to learn through inference and patterns without being explicitly programmed. It builds a model based on the samples, to make predictions without being explicitly programmed to do so. Machine learning is currently used in multiple fields and industries. It provides information based on the past experience. It is used in medicine, computer vision, image processing etc..Machine learning is important because it allows the user to collect an immense amount of data from the computer and analyze it and make decisions based on the data. It is also used in the development of robotics. We will discuss about it in the machine learning applications briefly.

SUPERVISED LEARNING

Supervised machine learning is used to label the datasets to train algorithms that classify data outcomes accurately. It uses training data to learn mapping function that turns into input variables and output Variables, for example we take (X) as input Variable and (Y) as output variable and it solves f in the equation given below,

Y = f(X) equation

when new inputs are given, it accurately generates outputs. It can be separated into two types,

1. CLASSIFICATION
2. REGRESSION

1. CLASSIFICATION

It uses an algorithm to accurately assign the test data. It is used to predict the outcome of a sample when the output is in the form of Categories. It might look at input data and try to predict things like"sick" or "healthy".We can also use machine learning techniques for classification problems. In classification problems, we classify objects of similar nature into a single group. For example, a set of 100 students say, we may like to group them into three groups based on their heights - short, medium and long. Measuring the height of each student, we will place them in a proper group. Now, when a new student comes in, we will put him in an appropriate group by measuring his height. By following the principles in regression training, we will train the machine to classify a student based on his feature – the height. When the machine learns about how the groups are formed, it will be able to classify any unknown student correctly. Once again, we would use the test data to verify that the machine has learned your technique of classification before putting the developed model in production. Supervised Learning is where Artificial intelligence really began its journey. The following technique was applied successfully in several cases. We have used this model while doing the hand-written recognition on your machine.

2. REGRESSION

Regression is used to understand the connection between dependent and independent variables. It is used to predict the outcome of a Sample when the output is in the form of real values. for example, amount of rainfall can be predicted by processing input data from the regression model. Some of the advantages are Classes representthe quality on the ground, Training data is reusable unless features change. Some of the disadvantages are Classes may not match, Consistency differs in classes, To select training data cost and time are involved. Similarly, in the case of supervised learning, we give known examples to the computer. We say that for given value x1 the output is y1, for x2 it is y2, for x3 it is y3, and so on. Based on this data, we let the computer figure out an empirical relationship between x and y. Once the machine is trained in this way with a sufficient number of data points, now we would ask the machine to predict Y for a given X. Assuming that we know real value of Y for this given X, we will be able to deduce whether the machine's prediction is correct.Thus, we will test whether the machine has learned by using the known test data. Once we are satisfied that the machine is able to do the predictions with a desired level of accuracy (say 80 to 90%) you can stop further training the machine.Now, we can safely use the machine to do the predictions on unknown data points, or ask the machine to predict Yfor a given X for which we do not know the real value of Y. This training comes under the regression that we talked about earlier.

2. UNSUPERVISED LEARNING

Unsupervised machine learning refers to the use of AI algorithms to identify patterns in datasets containing data points that are neither classified nor labelled. They are used when we have only input variable (X) and not the output variable(Y).It uses unlabeled data by modelling the underlying structure of the data in the algorithm. Some of its uses are It is used for customer sections and understanding different customer groups to build

marketing or other business, Anomaly detection, including fraud detection or detecting defective mechanical parts. Some of the advantages are Human error is controlled, It produces unique spectral classes. Some of the disadvantages are It does not necessarily represent the features on the ground in spectral classes, Spatial relationships are notconsidered in the data.

1. CLUSTERING

Clustering is a type of unsupervised learning method. Drawing references from datasets consisting of input data without labelled responses is an unsupervised learning method. Clustering is a way of grouping the data points into clusters consisting of similar data points. The goal is to group or cluster observations that have similar characteristics. It does not use output information for training, but the algorithm defines the output.We use only visualization to inspect the quality of the solution. We can cluster almost anything, and the more similar the items are in the cluster, the better the clusters are. It is called k-means clustering because it finds 'k' unique clusters, and the center of each cluster is the mean of the values in that cluster. For example, The 2000 and 2004 Presidential elections in the United States were closevery close. The largest percentage of popular vote that any candidate received was 50.7% and the lowest was 47.9%. If a percentage of the voters were to have switched their sides, the outcome of the election would have been different. There are small groups of voters who properly appealed to switch their sides. These groups may not be big, but with such close races, they may be able to change the outcome of the election. How do you find these groups of people? How we can appeal to them with a limited budget? The answer might be clustering without any doubt.

1. First, we collect information on people either with or without their consent: any sort of information that might give some clue about what is important to them and what will influence how they vote in the election.
2. Then we put this information into some sort of clustering algorithm.
3. Next, for each cluster we craft a message that will appeal to these voters.
4. Finally, we deliver the campaign and measure to see if it's working.

2. ASSOCIATION

It is a type of unsupervised learning technique that checks for the dependency of one data item on another data item and maps accordingly so that it can benefit more to the users. It tries to find interesting relations or associations among the variables of the dataset.The Association rule algorithm tries to learn without a teacher as data are not labelled. Association rule mining is commonly used for Market basket analysis, customer clustering in details, price bundling, etc. Some of the examples are The Cancer patients grouped by their gene expression measurement, Groups of shoppers based on their browsing and purchasing histories, The rating is given by movie viewers.

TYPES OF UNSUPERVISED LEARNING

1. LINEAR CLASSIFICATION BEHAVIOUR

It is a statistical classification that is used to identify the objects, which classes it belongs to. This can be achieved by making a classification behaviour based on the value of the linear combination of Characteristics. It

is mostly used for finding the relationship between variables and forecasting. It makes predictions for continuous, real or numerical values or variables by this type. The result is predicted by Known parameters which are correlated with output. It predicts the values within a Continuous range rather than classifying them.

2. NON-LINEAR CLASSIFICATION BEHAVIOUR

It is used to separate non-linear objects. Data cannot be separated by a simple threshold sometimes. For More Complex data we can use non-linear classification.

Adding permutations is the simplest way for non-linear Classification. Its class boundaries cannot be approximated with linear classifiers, So, non-linear classifiers are more accurate than linear classifiers. It Maps data into high dimensional space to classify. linear lines Cannot be separated easily.

3. ANGLE BASED BEHAVIOUR

It defines angles of a data point with other data points, which define two angles. Then it measures the variance of those angles; anomalies result in very small variance.

4. NEURAL NETWORK

A Neural Network is a series of algorithms that endeavours to recognise underlying relationships in a set of data through a process that the way the human brain operates with. In this sense, it refers to a system ofneurons, either organic or artificial. It is the quality and ability of human beings that is given to machines. If the machine needs to work as humanwe need to give a specific program to do the job. AI was introduced in the early 1950s. For example, The children cannot know about the danger of fire, it will burn their hands if they keep near the fire. Likewise, if we create a new machine it gets damaged in a particular place. Next time it can occur in the same place because it has no thinking capacity like humans. It will do the sameuntil the program was changed by the programmer. Scientists got an idea to give thinking capacity to machines to reduce the work. Neuro is related to the human body. In our body a nervous system transfers sense to the human brain with the help of neurons. Input is used to sense the data. Hidden layers can have many layers in the single neural network, it is mainly because of complexity. The machine cannot learn without the input given by the user, so we need to give many inputs to generate a suitable output.

First, the neural network will classify the input connected to hidden layers by parameters. Parameters mean connecting lines, in other words, it is called weights. An example for single neuron is, It has three inputs x1,x2,x3 and corresponding weights w1,w2,w3.We take input and weights and add them with bias.

Then take the result to apply it to the sigmoid. Sigmoid is a function.

$Z=(X*W)+B$

sigmoid(Z)

Some of its applications are Character recognition, Electronicnose, Image compression, Neural network applications, Medical applications, Security applications.

5. NEIGHBOUR BASED BEHAVIOUR

The k-nearest neighbours (KNN) algorithm is a simple, easy-to-implement supervised machine learning algorithm that can be used to solve classification. Some of the advantages are The algorithm is easy to implement, There is no need to build a model, It can be used for classification and regression,It works very well in low dimensions for complex decision surfaces. One of the disadvantage is The algorithm gets significantly slower as the number of examples increases. An example algorithm,

1. Load the data
2. Apply genetic search mechanism on the loaded data
3. Based on the Values attributes are ranked
4. Subset is selected from the higher ranked
attribute
5. Apply Genetically and KNN Algorithm to Maximize
Classification Accuracy
6. Calculate the accuracy of the given classifier.

K-NEAREST NEIGHBORS

It is an effective classification method. For a new instance to be classified, k nearest neighbours of the instances are selected, and then the major class of the k neighbours is assigned to the new instance class. The Euclidean distance is used in kNN.A window of NN values is used up to size k. It is a simple but effective classification method.

ANOMALY DETECTION

Anomaly detection refers to the problem of finding patterns in data that do not conform to expected behaviour. These patterns are also known as outliers, exceptions,
aberrations, or contaminants in different application domains.
Anomalies occur far from their closest neighbours and require a distance or similarity measure defined between two data instances.

DISTANCE/SIMILARITY MEASURES

For continuous attributes in similarity measure, Euclidean distance is the correct choice.
For example, the square of the Euclidean distance is the distance metric, since it involves fewer and less expensive computations.Some of the advantages are It need not know the data distribution model,An anomaly has a close neighbourhood is very low,For different data types, it defines an appropriate distance measure.Some of the disadvantages are In the testing phase, the computational complexity is challenging, Defining distance measures between instances is also challenging when the data is graphs.

6. DENSITY-BASED BEHAVIOUR

A density-based clustering algorithm has played a role in finding the non-linear structures based on density. The most used density-based algorithm is DBSCAN (Density-Based Spatial Clustering of Applications with Noise). It uses two concepts: density reachability, density connectivity.

Algorithm for DBSCAN clustering

x={x1, x2, x3,……., xn} is a set of data points.

1)Start with an arbitrary starting point that has not been visited.

2)Using E, extract the neighbourhood of this point.

3)Around this point, if there are sufficient neighbourhoods then the clustering process starts. Otherwise, the point is marked as noise.

4)If the point is part of clusters, then its E neighbourhood is also part of clusters. For all E neighbourhood points, the procedure from step 2 is repeated. Until all points in clusters are determined, it is repeated.

5)A new unvisited point is found, then it's leading to find clusters or noise.

6)Until all the points are marked as visited, this process will continue.

Some of the advantages are While clustering, it can be able to identify the noise data, It does not require a specification of clusters, DBSCAN algorithm can find arbitrary size and shaped structures. Some of the disadvantages are It fails in varying the density clusters, It doesn't work well in high dimensional data.

7. DIMENSIONALITY REDUCTION

Dimensionality is the number of variables, characteristics or features present in the dataset. Sometimes most of the features are redundant and correlated. Then the dimensionality reduction algorithm was used. Dimensionality reduction is the process that reduces the number of variables by obtaining a set of principal variables under consideration.There are two components such as Feature selection and Feature extraction

Feature selection: find a subset of the original set of variables or features, to get a smaller subset that is used to model the program. It has three ways such as filter, wrapper, and embedded. Feature extraction: It reduces the data from high dimensional space to low dimensional space. Some of the advantages are Helps in data compression, Reduces storage space and computation time, Also helps in removal of redundant. some of the disadvantages areLeads to few amount of data loss, We may not know how many principal components to keep - in practice. Some thumb rules are applied, PCA fails in where covariance and mean are not enough to define datasets.

REINFORCEMENT LEARNING

It is a type of machine learning algorithm that allows an agent to decide the best action based on the Current state by learning behaviour that maximize the reward. It usually learns optimal actions through error and trial. For example, a videogame in which a player move from Certain places at Certain times to earn points. An algorithm will be playing by moving randomly through trial and error, so it would learn where and when it is needed to move in the game to maximize the character portal.

International Journal of Advance Research in Science and Engineering
Volume No. 11, Issue No. 01, January 2022
www.ijarse.com
IJARSE
ISSN 2319 - 8354

MACHINE LEARNING APPLICATIONS

1. FACE DETECTION

It is one of the most common applications of machine learning. It is used to identify a face, person, etc. To identify a face from the given number of photos, it will tag the photo automatically when the same photo is given next time. It is used in mobile phones, which is giving safety to our mobile phones. It is used in NEET, JEE examination to identify the right person who is writing the exam in the hall.

2. SPAM DETECTION

Emails received by the unknown, if it is identified as spam then it is not shown to the users in the regular inboxes. All the spam mails are maintained in a separate folder. We always receive important mail in our inbox with the important symbol and spam emails in our spam box. Some spam filters used by gmail are Content Filter, Header filter, General blacklists filter, Rules-based filters, Permission filters. This program is used to block unreal or machine-based messages and e-mails.

3. CREDIT CARD FRAUD DETECTION

Machine learning is making our online transactions safe and secure by detecting fraud transactions. According to past transactions by customers, any unwanted purchase made by someone immediately the customer is warned about the condition. It can take place fake ids, fake accounts and there may be a chance to make money in the middle of the transaction. For each transaction, the output is converted into some hash values, and these values become the input for the next round. For each transaction, there is a specific pattern that gets changed for the fraud transaction.

4. MEDICAL DIAGNOSIS

For detecting diseases, Hospitals are using machines to identify the disease affected to his/her, with the help of complete data about diseases. IBM designed a system with 95% precision in predicting the cancerous images in contrast to 75%-84% precision by doctors. It analyzes the medical data for detecting regularities in data, It is used to handle inappropriate data. It explains about data generated by medical units, It is also used for effective monitoring of patients. This technology is growing very fast and is able to produce 3D models.

5. COMPUTATIONAL INTELLIGENCE

Computational intelligence has been developed actively for many years. Constant improvements and improvements are carried out on machine learning algorithms. It is to facilitate the spreading of theoretical, experimental and applied research. It is used to provide professionals, students, academics, and scholars free access to the latest and most advanced research outcomes, It is helpful in findings, and studies. It is being used in the field of Artificial Intelligence & Machine Learning. The fact that both Computational Intelligence and

Machine Learning is an open-access journal means that anyone from any part of the world can gain access to entire issues at any time they can.

6. MOBILE DEVICES

When machine learning techniques core applied. On portable devices like smartphones, automotive systems, sensors, etc., the ML approach is provided with some training examples such as Vector Machines, Random Forests etc... Each technique has its strengths and weaknesses. It presents performance measures for the machine learning algorithms.

7. PATTERN RECOGNITION

It is another main problem in machine learning. It aims to provide reasonable answers for all possible inputs. They use computer algorithms to recognize data regularities and patterns. This type of recognition can be done on various input types, such as biometric recognition, colours, image recognition, and facial recognition. It has been applied in various fields such as image analysis, computer vision, and healthcare.

CONCLUSION

The paper presented an overview of Machine learning and its techniques in various applications. Supervised machine learning algorithm and Unsupervised machine learning algorithm are the most commonly used machine learning algorithms. Clustering in the unsupervised machine learning algorithm is the very useful in providing the solutions. Machine learning techniques are helpful in many fields like Medical, Mobile devices etc..This will be used in our future projects.

REFERENCES

1. https://en.wikipedia.org/wiki/Machine_learning
2. https://bigdata-madesimple.com/top-10-real-life-examples-of-machine-learning/
3. https://www.netapp.com/artificial-intelligence/what-is-machine-learning/
4. https://machinelearningmastery.com/machine-learning-algorithms-mini-course/
5. https://winder.ai/404-nonlinear-linear-classification/
6. https://www.aitude.com/svm-difference-between-linear-and-non-linear-models/
7. https://machinelearningmastery.com/types-of-classification-in-machine-learning/
8. https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/
9. https://www.datasciencecentral.com/unsupervised-learning-an-angle-for-unlabelled-data-world/
10. https://towardsdatascience.com/unsupervised-machine-learning-clustering-analysis-d40f2b34ae7e
11. https://en.wikipedia.org/wiki/Unsupervised_learning
12. https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/
13. https://en.wikipedia.org/wiki/Machine_learning

14. https://pythonistaplanet.com/
15. https://medium.com/
16. https://link.springer.com/
17. https://www.javatpoint.com/
18. https://www.salesforce.com/blog
19. https://www.asquero.com/
20. https://www.ibm.com/
21. https://www.hindawi.com/
22. https://www.sciencedirect.com/
23. https://pathmind.com/
24. https://www.researchgate.net/
25. http://www.javatpoint.com/clustering-in-machine-learning
26. https://www.geeksforgeeks.org/clustering-in-machine-learning/
27. http://www.javatpoint.com/density-based
28. https://sites.google.com/site/dataclusteringalgorithms/density-based-clustering-algorithm
29. https://towardsdatascience.com/unsupervised-learning-dimensionality-reduction
30. https://machinelearningmastery.com/dimensionality-reduction-for-machine-learning/
31. https://en.m.wikipedia.org/wiki/DBSCAN
32. https://en.m.wikipedia.org/wiki/Cluster_analysis