Analysis on Machine Learning Algorithms

Dr.L.Thara[1], K.TamilSelvan[2]
[1]Associate Professor, Dept of MCA
[2]Dept of MCA
[1, 2] PSG College of Arts & Science, Coimbatore, Tamilnadu, India

Abstract- *The study of algorithms and statistical models that computer systems employ to carry out a particular task without being explicitly taught is known as machine learning (ML). There are several daily-used programmes that incorporate learning algorithms. One of the reasons an online search engine like Google works so well every time it is used to search the internet is because of a learning algorithm that has mastered the art of ranking web sites. These algorithms are utilised for many different things, including data mining, image processing, predictive analytics, etc. The main benefit of machine learning is that once an algorithm understands how to use data, it can carry out its work autonomously. In this essay, a quick summary and outlook.In this paper, a brief review and future prospect of the vast applications of machine learning algorithms has been made.*

Keywords- Algorithm, Machine Learning, Supervised Learning, Unsupervised Learning, Reinforcement Learning

I. INTRODUCTION

Since the beginning of time, humans have used a variety of instruments to complete various jobs more quickly. Different devices were created thanks to the human brain's ingenuity. These devices made life easier for humans by allowing them to fulfil a variety of demands, such as travel, industry, and computing. And among them, machine learning is one.

According to Arthur Samuel, the scientific subject that enables computers to learn without explicit programming is known as machine learning. For his computer programme that played checkers, Arthur Samuel gained fame. The application of machine learning (ML) teaches computers how to handle data more effectively. Sometimes, even after viewing the data, we are unable to evaluate or extrapolate the information. We then use machine learning in that situation. The availability of a large number of datasets has increased demand for machine learning. Machine learning is used in many industries to retrieve pertinent data. Learning from the data is the goal of machine learning. On how to make robots learn on their own without being explicitly programmed, numerous experiments have been conducted. Many programmers and mathematicians submit applications. There are numerous ways to solve this problem, which has a large amount of data[1].

II. METHODOLOGIES

2.1 Supervised Learning

A function that maps an input to an output is learned through supervised learning using sample input-output pairs. It uses labelled training data made up of a collection of training examples to infer a function. Algorithms that require outside help are known as supervised machine learning algorithms. Train and test datasets are created from the input dataset. The output variable in the train dataset has to be predicted or categorised. Every algorithm Learn certain patterns from the training dataset, then use them to predict or classify data from the test dataset. The diagram below shows the process used by supervised machine learning algorithms. Here, the most well-known supervised machine learning methods have been covered.

2.2 Decision Tree

A decision tree is a graph that displays options and their outcomes as a tree. The edges of the graph indicate the conditions or rules for making decisions, whereas the nodes in the graph represent an event or a choice. Nodes and branches make up eachcharacteristics that needs to be categorised, while each branch indicates a possible value for the node tree. Each node represents a set of[2].

2.3 Navie Bayes

It is a classification method built on the Bayes Theorem and predicated on the idea of predictor independence. A Naive Bayes classifier, to put it simply, believes that the presence of one feature in a class has nothing to do with the presence of any other feature. Nave Bayes is primarily used in the text categorization sector. It is mostly used for classification and clustering purposes based on the conditional likelihood of occurring**.**

2.4 Support Vector Machine

Another most widely used state-of-the-art machine learning technique is Support Vector Machine (SVM). In machine learning, support-vector machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. It basically, draw margins between the classes. The margins are drawn in such a fashion that the distance between the margin and the classes is maximum and hence, minimizing the classification error

2.5 Unsupervised Learning:

These are referred to as unsupervised learning because, in contrast to the supervised learning described above, there are no correct answers and no teachers. The algorithms are left to find and display the intriguing structure in the data on their own. The algorithms for unsupervised learning only extract a small number of features from the data. When new data is introduced, it recognises the class of the data using the previously learned features. It is primarily utilised for feature reduction and grouping.

2.6 Principal Component Analysis

A set of observations of potentially correlated variables are transformed via an orthogonal transformation in the statistical process known as principal component analysis into a set of values of linearly uncorrelated variables known as principal components. This reduces the dimension of the data to speed up and simplify computations. It is used to use linear combinations to explain the variance-covariance structure of a group of variables. It is frequently employed as a dimensionality-reduction method.

2.7 K-Means Clustering

One of the most straightforward unsupervised learning techniques to handle the well-known clustering problem is K-means. The process uses a predetermined number of clusters to classify a given data set in a straightforward and straightforward manner[3]. To define k centres, one for each cluster, is the main notion. These centres should be strategically positioned because different locations yield various effects. Therefore, it would be advisable to arrange the masfar apart from one another as you can.

2.8 Semi Supervise Learning:

Combining supervised and unsupervised machine learning techniques is known as semi-supervised machine learning. It may be useful in machine learning and data mining applications where unlabeled data is already present and obtaining labelled data is a time-consuming procedure. With more widespread supervised machine learning techniques, you train a machine learning algorithm using a dataset that has been "labelled" so that each record contains the result data. Below is a discussion of a few semi-supervised learning algorithms.

2.9 Transductive SVM

Semisupervised learning frequently uses transductive support vector machines (TSVM) to handle partially labelled input. Due of the dearth of knowledge around its generalization-based foundation, mystery has surrounded it. It is used to label the unlabeled data so that the margin between the labelled and unlabeled data is as small as possible. It is NP-hard to find a precise answer using TSVM[4].

III. GENERATIVE MODELS

A model that can produce data is called a generative model. It models the entire set of data, including the characteristics and the class. If we model P(x,y), I can create data points using this probability distribution, hence any algorithms that model P(x,y) are generative. For each component, a single labelled sample is sufficient to verify the mixture distribution.

3.1 Self-Training

A classifier is learned using a part of labelled data during self-training. Unlabeled data are then supplied into the classifier. In the training set, the predicted labels and the unlabeled points are combined. The process is then carried out once again. Self-training is a term used since the classifier is always learning.

3.2 Reinforcement Learning

Reinforcement learning is an area of machine learning concerned with how software agents ought to take actions in
an environment in order to maximize some notion of cumulative reward. Reinforcement learning is one of three basic machine learning paradigms, alongside supervised learning and unsupervised learning.

3.3 Multitask Learning

A kind of machine learning called "multi-task learning" uses the similarities between several tasks to attempt to handle many distinct problems at once. This may increase learning effectiveness and serve as a regularizer. Formally, Multi-Task Learning (MTL) will assist in improving the learning of a particular model by using the knowledge contained in all of the n tasks if there are n tasks (conventional deep learning approaches aim to solve just 1 task using 1 particular model), where these n tasks or a subset of them are related to each other but not exactly identical.

3.4 Ensemble Learning

To address a specific computational intelligence issue, many models, such as classifiers or experts, are deliberately developed and mergedin an ensemblelearning process. The main purpose of ensemble learning is to enhance a model's performance or lessen the possibility of making a poor model choice. Other uses for ensemble learning include giving the model's choice a level of confidence, choosing the best features, fusing data, incremental learning, non-stationary learning, and error-correcting.

IV. NEURAL NETWORK

A neural network is a collection of algoithms that aims to identify underlying connections in a piece of data using a method that imitates how the human brain works. In this context, neural networks are systems of neurons that can be either organic or synthetic in origin. Since neural networks are capable of adapting to changing input, the network can produce the optimal outcome without having to change the output criterion. In the creation of trading systems, the artificial intelligence-based idea of neural networks is quickly gaining favour.

4.1 Supervised Neural Network

The output of the input in a supervised neural network is already known. The neural network's anticipated output is contrasted with the actual output. The parameters are modified in response to the mistake, and the neural network is then fed again. The feed forward neural network employs supervised neural networks.

4.2 Unsupervised Neural Network

The neural network has no prior knowledge of the input or output. The network's primary responsibility is to group data into categories based on shared characteristics. The neural network clusters the inputs based on how closely they are correlated[5].

4.3 Reinforced Neural Network

Reinforcement learning is a term used to describe goal-oriented algorithms that learn how to maximise along a given dimension over a number of steps, such as maximising the number of points scored during a game. They are capable of beginning with nothing, and given the correct circumstances, they are capable of superhuman performance. These algorithms are punished when they make the incorrect judgements and rewarded when they make the right ones, just like a toddler who is encouraged by spankings and chocolates.

V. INSTANCE-BASED LEARNIN

A series of approaches for classification and regression known as instance-based learning provide a class label or prediction based on how similar the query is to its closest neighbours in the training set. Instance-based learning algorithms do not abstract from particular cases, in contrast to other approaches like decision trees and neural networks. Instead, they merely store all the data, and when a query is made, they determine the response by looking at the query's nearest neighbour (s)[6].

5.1 K-Nearest Neighbor

The supervised machine learning technique known as the k-nearest neighbours (KNN) can be used to tackle classification and regression issues. It is simple to use and comprehend, but it has the important problem of becoming noticeably slower as the amount of data in use increases[7].

VI. CONCLUSION

Both supervised and unsupervised machine learning are possible. Choose supervised learning if you have fewer data points with well marked training data. For huge data sets, unsupervised learning would often perform and produce superior outcomes. Consider using deep learning techniques if we have a sizable data collection that is easily accessible. This page provides a better understanding of neural networks, their uses, and their drawbacks. In this work, numerous machine learning methods are surveyed. Today, whether intentionally or not, everyone uses machine learning. from posting images on social networking sites to receiving product recommendations when buying online. This article presented various machine learning algorithms and thein deployments were discussed by suggesting the appropriate selection of such algorithm so as to generate the optimal outcome[8].

REFERENCES

[1] W. Richert, L. P. Coelho, "Building Machine Learning Systems with Python", Packt Publishing Ltd., ISBN 978-1-78216-140-0

[2] J. M. Keller, M. R. Gray, J. A. Givens Jr., "A Fuzzy K-Nearest Neighbor Algorithm", IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-15, No. 4, August 1985

[3] https://www.geeksforgeeks.org/machine-learning/

[4] S. Marsland, Machine learning: an algorithmic perspective. CRC press, 2015.

[5] M. Bkassiny, Y. Li, and S. K. Jayaweera, "A survey on machine learning techniques in cognitive radios," IEEE Communications Surveys & Tutorials, vol. 15, no. 3, pp. 1136–1159, Oct. 2012.

[6] https://en.wikipedia.org/wiki/Instance-based_learning

[7] R. S. Sutton, "Introduction: The Challenge of Reinforcement Learning", Machine Learning, 8, Page 225-227, Kluwer Academic Publishers, Boston, 1992

[8] P. Harrington, "Machine Learning in action", Manning Publications Co., Shelter Island, New York, 2012