A Survey on Intelligent Machine Learning Algorithms, Concepts and Techniques

[1]Vinodh M R,[2]Dr.P.J.Arul Leena Rose

[1]Assistant Professor,[2]Associate Professor
[1,2]Department of Computer Science, SRM Institute of Science and Technology
Kattankulathur, Tamilnadu, India.

Abstract: The field of study that gives computers the capability to learn without being explicitly programmed is Machine Learning. ML is one of the exciting technologies that everyone is keen to know about. Apparently from the name itself it defines that making the computers to learn that which makes it more similar to humans: ***The knack to learn and drive***. Machine learning aims on the development of programs that can get accessed to data and use the data to learn it for them. This paper focus on summarizing various algorithms of machine learning that is widely used in Artificial Intelligence that helps for prediction in various fields.

Index Terms: **Machine Learning, ML algorithms.**

I. INTRODUCTION

Machine learning (ML) is a type of algorithm that allows software applications to become more precise in predicting result without being explicitly programmed. It lets the computers or the machines to make data-driven decisions rather than being explicitly programmed for carrying out a certain responsibility. These programs or algorithms are designed in a way that they learn and improve over time when are exposed to new data. The basic assertion of machine learning is to build algorithms that can receive input data. And to predict an output, statistical analysis is being used. While updating outputs as new data, it becomes available. The courses of action involved in machine learning are similar to that of data mining and predictive modelling. Currently we are living in the primitive age of machines, while the future of machine is enormous and is beyond our scope of imagination. Machine Learning algorithms are trained using a training data set. These trained dataset then creates a model. Based on this model, when we input new data to the ML algorithm, it makes a prediction based on historical data.

The prediction is examined for accuracy. And if the accuracy is valid, the Machine Learning algorithm is deployed. If the accuracy is not valid, then the Machine Learning algorithm is trained again and again with an augmented training data set.
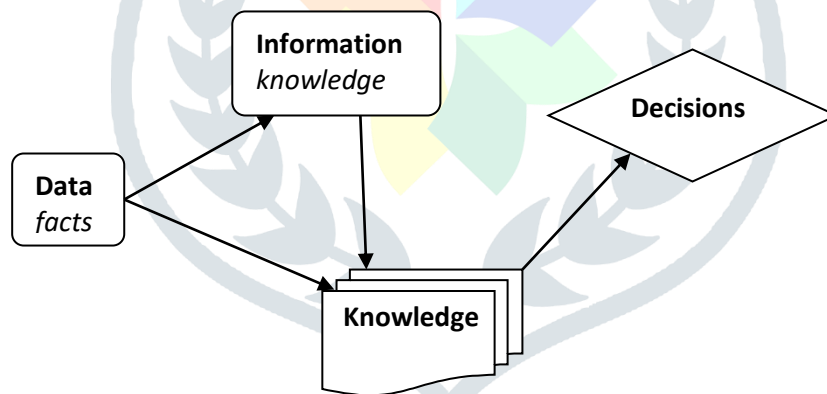
Fig.1 Machine Learning

Data plays a vital role in machine learning. Without data we cannot able to train models. Infer information from data and learn the data that results to support a normal user or even for business.

II. MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

Artificial Intelligence is the emerging field going to redefine the world of software and IT in the near future. AI directs the making of machines, frameworks and different gadgets savvy. It do so by enabling them to think and do errands as all people generally do. Machine Learning is a part of AI that learns from the data that also involves the information gathered from the previous experiences and allows the computer program to change its behaviour accordingly. Machine Learning (ML) manages persuading user's machine to gain from the exterior environment. This exterior environment can be sensors, electronic segments, external storage gadgets and numerous other devices.

III. WORKING PRINCIPLE OF MACHINE LEARNING

Train the machine Learning algorithm using a training dataset for creating the model. Whenever we produce new input or instance to the algorithm, it makes a prediction based on the model. The prediction is checked for accuracy and if it is acceptable, the Machine Learning algorithm used for the model is deployed. If the evaluated accuracy is not acceptable, then the Machine Learning algorithm is trained again and again with an augmented training data set till we get the accuracy. Also it predicts its class based on previous experiences end also with existing records.
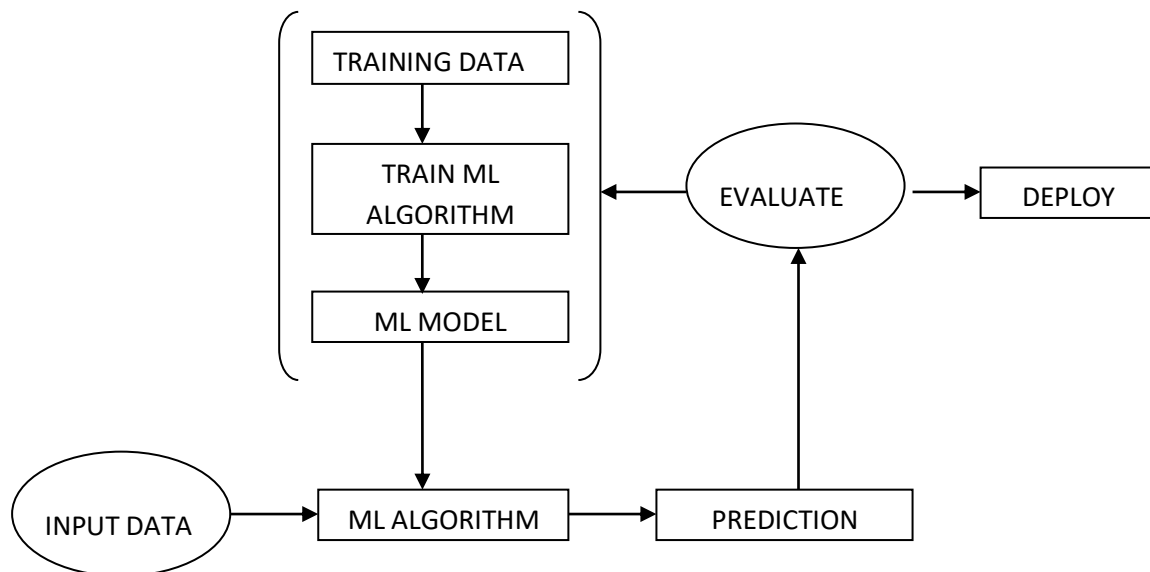

Fig.2 Working of Machine Learning

IV. MACHINE LEARNING ALGORITHMS

Learning generally means learning from past experiences. Machine learning is also based on same principle that learns from the historical data with respect to the task and checks for its improvement in performance. Taiwo Oladipupo Ayodele in his "Types of machine learning algorithms" described types of ML algorithms with their details, advantages, and disadvantages [1]. Learning is classified into various types: Supervised learning, Unsupervised learning and Reinforcement learning.

4.1 Supervised Learning

Learning with someone's help is supervised learning. We should train the model or a machine with an dataset. After the model get trained, it can start predicting for new data. V. N. Vapnik in his springer edition of NewYork mentioned that, in order to solve a give problem using supervised learning algorithm one has to follow some certain steps [2].

i)Determine the various type of training examples.
ii)Collect and organize a training set.
iii)Identify the input feature representation of learned function.
iv)Identify the structure of learning function & corresponding learning algorithm.
v)Complete the design and run the learning algorithm on the collected set of data.
vi)Evaluate the accuracy of the learned function also the performance of the learning function. The performance should be measured again on the set which is different from the training set.
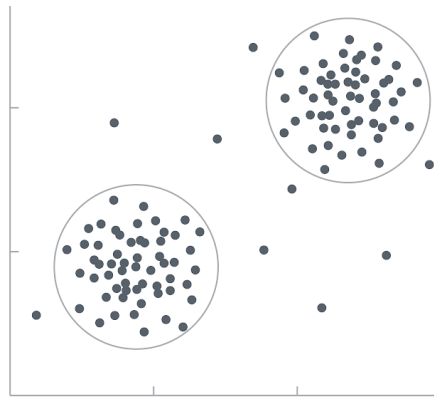

Fig.2 Supervised Learning

4.2 Unsupervised Learning

Learning through observations and finding structures in the data is unsupervised learning. When the model is given to a dataset, it automatically searches for patterns and relationships in the dataset by creating clusters in it. Impossible thing in unsupervised learning is adding labels to the cluster. Usually this type of training will generally fit into the decision problem framework. Because the goal is not to only produce a classification model but also to make decisions that maximize rewards. I.

Wittenet.al in 2011 elaborated that this approach nicely generalizes to the real world, where agents might be rewarded for doing certain actions and punished for doing others[3].

Fig.2 UnSupervised Learning

4.3 Reinforcement Learning

It is the ability of an agent to interact with the environment and find out what is the best outcome. The method followed in reinforcement learning is hit and trial method. The agent is rewarded or penalized with a point for a correct or a wrong answer. And the model trains itself based on the positive reward points gained. And it is trained once again so that it gets ready to predict the new data presented to it.

4.4 Naive Bayes Classifier Algorithm

A classifier is a function that issues a population's element value from one of the available classification. One of the popular Naïve Bayes algorithm is Spam Filtering. To all emails, spam filter acts as a classifier that assigns a label "Spam" or "Not Spam" to all the emails. Naïve Bayes Classifier works on the popular Bayes Theorem of Probability. Main purpose of Naïve Bayes is to build machine learning models particularly for disease prediction and document classification. This algorithm works well when the input variables are categorical and also for many complex real world problems.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{1}$$

In Bayes theorem, for two events, A and B, it allows you to figure out p(A|B) from p(B|A).p(A|B) is the probability that event A happened, given that test B was positive and p(B|A) is the probability that test B happened, given that event A happened. S.L. Ting, in 2011 highlighted in his paper about the performance of employing Naïve Bayes in document classification and proven that naive Bayes is best classifier based on his results [4].

4.5 K Means Clustering Algorithm

K-means is a commonly used unsupervised machine learning algorithm for cluster analysis. K-Means is a non-deterministic and iterative method. The algorithm operates on a given data set through pre-defined number of clusters, k. The generated output of K Means algorithm is k clusters. It consists of input data partitioned among the clusters. K-Means produces tighter clusters than hierarchical clustering. Shi Na in 2010, elaborated that K-means is a non-deterministic, numerical, unsupervised, , iterative method which is simple and very fast in many practical applications [5]

4.5 Support Vector Machine

One of the supervised machine learning algorithm is Support Vector Machine, which can be used for both classification and regression problems. For labelled training data, the support vector machine algorithm outputs an optimal hyperplane which categorizes new examples. Zhi-Qiang Zeng in 2008, explained about how to overcome the limitation of support vector machine applications in large sample problems [6].

V. CONCLUSION

The researchers using Machine Learning will always focus on designing efficient implementation of the algorithms and to produce high accuracy of prediction. Machine learning gives high importance in the prediction of many areas especially in the field of healthcare, business solutions. The best part of machine learning algorithms is of using same algorithms to train different models. Supervised Learning adopts many techniques which are very expensive when the systems are implemented over wide range of data. This is due to the certainty that considerable amount of effort and cost is involved because of obtaining large labelled data sets. Thus active learning provides a way to lessen the labelling costs by labelling only the most useful instances for learning.

REFERENCES

[1] Types of Machine Learning Algorithms, Taiwo Oladipupo Ayodele, University of Portsmouth, United Kingdom.

[2] V. N. Vapnik. The Nature of Statistical Learning Theory. Springer, New York, NY, 1995.

[3] I.Witten, E. Frank, and M. Hall. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, San Mateo, CA, 3rd edition, 2011

[4] S.L. Ting, W.H. Ip, Albert H.C. Tsang, "Is Naïve Bayes a Good Classifier for Document Classification?", International Journal of Software Engineering and Its Applications, Vol 5, July 2011.

[5] Shi Na, Liu Xumin, Guan yong, "Research on k-means Clustering Algorithm", Third International Symposium on Intelligent Information Technology and Security Informatics, 2010

[6] Zhi-Qiang Zeng, Hong-Bin Yu, Hua-Rong Xu, Yan-Qi Xie, Ji Gao, "Fast Training Support Vector Machines Using Parallel Sequential Minimal Optimization", International Conference on Intelligent System and Knowledge Engineering, 2008.

[7] Cour, T. and Sapp, B. and Taskar, B. Learning from partial labels, Journal of Machine Learning Research, Volume 12, 1501-1536 2012

[8] A. Carlson, J. Betteridge, B.Kisiel, B.Settles,E. R.Hruschka Jr,and T. M. Mitchell, Toward an architecture for never-ending language learning,
AAAI, volume 5, 3, 2010

[9] H. Hlynsson. Transfer learning using the minimum description length principle with a decision tree application. Master's thesis, University of
Amsterdam, 2007.

[10] Taiwo Oladipupo Ayodele, Types of Machine Learning Algorithms, New Advances in Machine Learning, Yagang Zhang (Ed.), InTech, 2010