

Review of Latest Advancements and Trends in Machine Learning

K. Vinod Kumar^{1*}, P. Santosh Kumar²

^{1,2}Dept. of Computer Science and Engineering, RGUKT-RK Valley, Andhra Pradesh, India

*Corresponding Author: *kethineni.vinod@gmail.com*, Tel.: +91 9490688039

DOI: <https://doi.org/10.26438/ijcse/v7i9.189192> | Available online at: www.ijcseonline.org

Accepted: 13/Sept/2019, Published: 30/Sept/2019

Abstract— In this paper, various machine learning algorithms have been discussed. These algorithms are used for various purposes like data mining, image processing, predictive analytics, etc. to name a few. The main advantage of using machine learning is that, once an algorithm learns what to do with data, it can do its work automatically.

Keywords— Machine Learning, Data Mining, Predictive Analytics, Image Processing, Algorithms

I. INTRODUCTION

Machine learning is used to teach machines how to handle the data more efficiently. Sometimes after viewing the data, we cannot interpret the pattern or extract information from the data. In that case, we apply machine learning [1]. With the abundance of datasets available, the demand for machine learning is in rise. Many industries from medicine to military apply machine learning to extract relevant information. The purpose of machine learning is to learn from the data. Many studies have been done on how to make machines learn by themselves [2] [3]. Many mathematicians and programmers apply several approaches to find the solution of this problem. Some of them are demonstrated in Fig. 1.

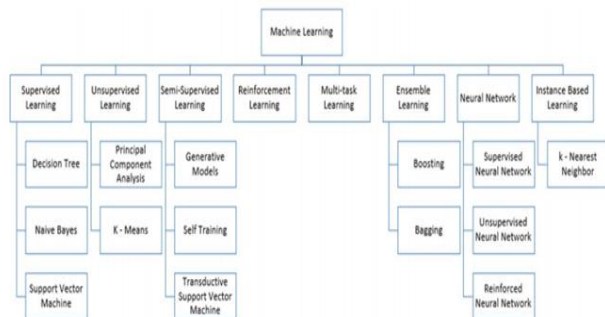


Figure:1 - Types of Learning [2]

II. TYPES OF LEARNING

A. Supervised Learning:

The supervised machine learning algorithms are those algorithms which needs external assistance. The input dataset is divided into train and test dataset. The train dataset has output variable which needs to be predicted or classified. All algorithms learn some kind of patterns from the training

dataset and apply them to the test dataset for prediction or classification [4]. The workflow of supervised machine learning algorithms is given in Fig2.

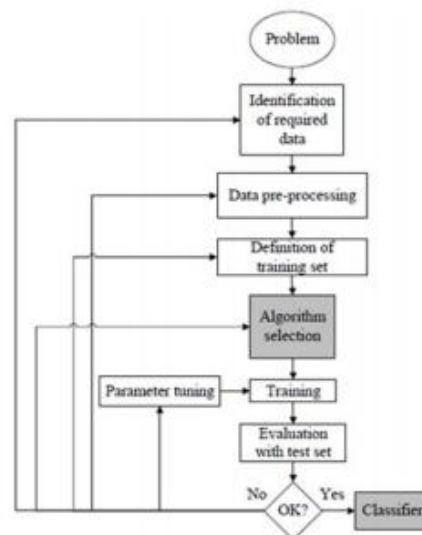


Figure 2: Workflow of supervised machine learning algorithm [4]

Three most famous supervised machine learning algorithms have been discussed here.

1) Decision Tree: Decision trees are those type of trees which groups attributes by sorting them based on their values. Decision tree is used mainly for classification purpose. Each tree consists of nodes and branches. Each nodes represents attributes in a group that is to be classified and each branch represents a value that the node [4]. An example of decision tree, where S, A and y are training set, input attribute and target attribute respectively.

2) Naïve Bayes: Naïve Bayes mainly targets the text classification industry. It is mainly used for clustering and classification purpose [6]. The underlying architecture of Naïve Bayes depends on the conditional probability. It creates trees based on their probability of happening. These trees are also known as Bayesian Network.

3) Support Vector Machine: Another most widely used state-of-the-art machine learning technique is Support Vector Machine (SVM). It is mainly used for classification. SVM works on the principle of margin calculation. 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.

The two main algorithms for clustering and dimensionality reduction techniques are discussed below.

a) K-Means Clustering: Clustering or grouping is a type of unsupervised learning technique that when initiates, creates groups automatically. The items which possesses similar characteristics are put in the same cluster. This algorithm is called k-means because it creates k distinct clusters. The mean of the values in a particular cluster is the centre of that cluster [9].

b) Principal Component Analysis: In Principal Component Analysis or PCA, the dimension of the data is reduced to make the computations faster and easier. To understand how PCA works, let's take an example of 2D data. When the data is being plot in a graph, it will take up two axes. PCA is applied on the data, the data then will be 1D.

B. Semi - Supervised Learning These algorithms combine the power of both supervised and unsupervised learning. It can be fruit-full in those areas of machine learning and data mining where the unlabelled data is already present and getting the labelled data is a tedious process [15]. There are many categories of semi-supervised learning [16]. Some of which are discussed below:

- 1) Generative Models: Generative models are one of the oldest semi-supervised learning method assumes a structure like $p(x,y) = p(y)p(x|y)$ where $p(x|y)$ is a mixed distribution e.g. Gaussian mixture models. Within the unlabelled data, the mixed components can be identifiable. One labelled example per component is enough to confirm the mixture distribution.
- 2) Self-Training: In self-training, a classifier is trained with a portion of labeled data. The classifier is then fed with unlabeled data. The unlabeled points and the predicted labels are added together in the training set. This procedure is then repeated further. Since the classifier is learning itself, hence the name self-training.

3) Transductive SVM: Transductive support vector machine or TSVM is an extension of SVM. In TSVM, the labeled and unlabeled data both are considered. It is used to label the unlabeled data in such a way that the margin is maximum between the labeled and unlabeled data.

C. Reinforcement Learning: It is a type of learning which makes decisions based on which actions to take such that the outcome is more positive. The learner has no knowledge which actions to take until it's been given a situation. The action which is taken by the learner may affect situations and their actions in the future. Reinforcement learning solely depends on two criteria: trial and error search and delayed outcome [17]. The general model [18] for reinforcement learning is depicted in Fig. 3.

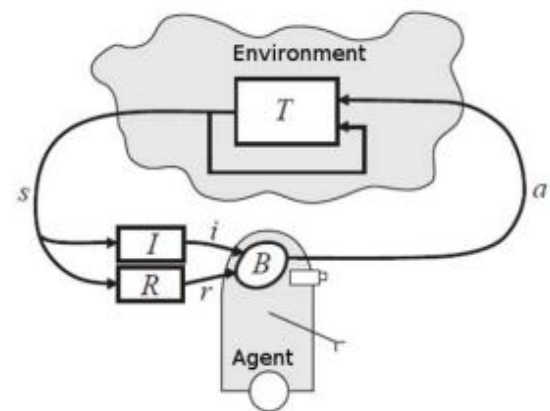


Figure 3: Reinforcement Learning Model [18]

In the figure, the agent receives an input i , current states, state transition r and input function I from the environment. Based on these inputs, the agent generates a behavior B and takes an action a which generates an outcome.

D. Multitask Learning: It has a simple goal of helping other learners to perform better. When multitask learning algorithms are applied on a task, it remembers the procedure how it solved the problem or how it reaches to the particular conclusion. The algorithm then uses these steps to find the solution of other similar problem or task. This helping of one algorithm to another can also be termed as inductive transfer mechanism. If the learners share their experience with each other, the learners can learn concurrently rather than individually and can be much faster [19].

E. Ensemble Learning: When various individual learners are combined to form only one learner then that particular type of learning is called ensemble learning. The individual learner may be Naïve Bayes, decision tree, neural network, etc. Ensemble learning is a hot topic since 1990s. It has been

observed that, a collection of learners is almost always better at doing a particular job rather than individual learners [20].

Two popular Ensemble learning techniques are given below [21]:

1) **Boosting:** Boosting is a technique in ensemble learning which is used to decrease bias and variance. Boosting creates a collection of weak learners and convert them to one strong learner. A weak learner is a classifier which is barely correlated with true classification. On the other hand, a strong learner is a type of classifier which is strongly correlated with true classification [21].

2) **Bagging:** Bagging or bootstrap aggregating is applied where the accuracy and stability of a machine learning algorithm needs to be increased. It is applicable in classification and regression. Bagging also decreases variance and helps in handling overfitting [23].

F. Neural Network Learning: It (or artificial neural network or ANN) is derived from the biological concept of neurons. A neuron is a cell like structure in a brain. To understand neural network, one must understand how a neuron works. A neuron has mainly four parts. They are dendrites, nucleus, soma and axon. The dendrites receive electrical signals. Soma processes the electrical signal. The output of the process is carried by the axon to the dendrite terminals where the output is sent to next neuron. The nucleus is the heart of the neuron. The inter-connection of neuron is called neural network where electrical impulses travel around the brain.

G. An artificial neural network behaves the same way. It works on three layers. The input layer takes input (much like dendrites). The hidden layer processes the input (like soma and axon). Finally, the output layer sends the calculated output (like dendrite terminals) [24].

There are basically three types of artificial neural network: supervised, unsupervised and reinforcement [25].

1) **Supervised Neural Network:** In the supervised neural network, the output of the input is already known. The predicted output of the neural network is compared with the actual output. Based on the error, the parameters are changed, and then fed into the neural network again. Fig. 19 will summarize the process. Supervised neural network is used in feed forward neural network.

2) **Unsupervised Neural Network:** Here, the neural network has no prior clue about the output the input. The main job of the network is to categorize the data according to some similarities. The neural network checks the correlation between various inputs and groups them. The schematic diagram is shown in Fig. 4.

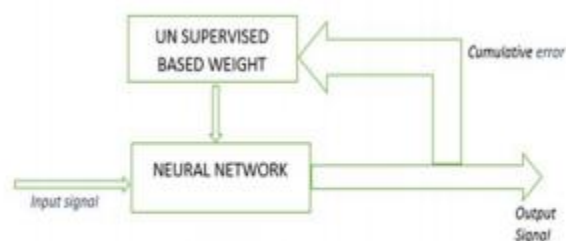


Figure 4: Unsupervised Neural Network

H. Instance-Based Learning: In instance-based learning, the learner learns a particular type of pattern. It tries to apply the same pattern to the newly fed data. Hence the name instance-based. It is a type of lazy learner which waits for the test data to arrive and then act on it together with training data. The complexity of the learning algorithm increases with the size of the data. Given below is a well-known example of instance-based learning which is k-nearest neighbour [26].

1) **K-Nearest Neighbour:** In k-nearest neighbour (or KNN), the training data (which is well-labelled) is fed into the learner. When the test data is introduced to the learner, it compares both the data. k most correlated data is taken from training set. The majority of k is taken which serves as the new class for the test data [27].

III. CONCLUSION

This paper surveys various machine learning algorithms. Today each and every person is using machine learning knowingly or unknowingly. From getting a recommended product in online shopping to updating photos in social networking sites. This paper gives an introduction to most of the popular machine learning algorithms.

REFERENCES

- [1] T. M. Mitchell, Machine Learning, McGraw-Hill International, 1997.
- [2] T.M. Mitchel, The Discipline of Machine Learning, CMU-ML-06-108, 2006
- [3] N. Cristianini and J. Shawe-Taylor. An Introduction to Support Vector Machines. Cambridge University Press, 2000.
- [4] E. Osuna, R. Freund, and F. Girosi. Support vector machines: training and applications. AI Memo 1602, MIT, May 1997.
- [5] V. Vapnik. Statistical Learning Theory. John Wiley & Sons, 1998.
- [6] C.J.C. Burges. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 2(2):1-47, 1998.
- [7] Taiwo Oladipupo Ayodele, Types of Machine Learning Algorithms, New Advances in Machine Learning, Yagang Zhang (Ed.), InTech, 2010
- [8] T. Mitchell, W. Cohen, E. Hruschka, P. Talukdar, J. Betteridge, A. Carlson, B. Dalvi, M. Gardner, B. Kisiel, J. Krishnamurthy, N. Lao, K. Mazaitis, T. Mohamed, N. Nakashole, E. Platanios, A. Ritter, M. Samadi, B. Settles, R. Wang, D. Wijaya, A. Gupta, X. Chen, A. Saparov, M. Greaves, J. Welling, Never-Ending

- Learning, Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2014
- [9] Pedregosa et al., Scikit-learn: Machine Learning in Python, JMLR 12, pp. 2825-2830, 2011.
- [10] Wang, J. and Jebara, T. and Chang, S.-F. Semi-supervised learning using greedy max-cut. *Journal of Machine Learning Research*, Volume 14(1), 771-800 2013
- [11] Chapelle, O. and Sindhvani, V. and Keerthi, S. S. Optimization Techniques for Semi-Supervised Support Vector Machines, *Journal of Machine Learning Research*, Volume 9, 203-233, 2013
- [12] J. Baxter. A model of inductive bias learning. *Journal of Artificial Intelligence Research*, 12:149-198, 2000.
- [13] S. Ben-David and R. Schuller. Exploiting task relatedness for multiple task learning. In *Conference on Learning Theory*, 2003.
- [14] W. Dai, G. Xue, Q. Yang, and Y. Yu, Transferring Naive Bayes classifiers for text classification. AAAI Conference on Artificial Intelligence, 2007.
- [15] H. Hlynsson. Transfer learning using the minimum description length principle with a decision tree application. Master's thesis, University of Amsterdam, 2007.
- [16] Z. Marx, M. Rosenstein, L. Kaelbling, and T. Dietterich. Transfer learning with an ensemble of background tasks. In *NIPS Workshop on Transfer Learning*, 2005.
- [17] R Conway and D Strip, Selective partial access to a database, In *Proceedings of ACM Annual Conference*, 85 - 89, 1976
- [18] P D Stachour and B M Thuraisingham Design of LDV A multilevel secure relational database management system, *IEEE Trans. Knowledge and Data Eng.*, Volume 2, Issue 2, 190 - 209, 1990
- [19] R Oppliger, Internet security: Firewalls and beyond, *Comm. ACM*, Volume 40, Issue 5, 92 -102, 1997
- [20] Rakesh Agrawal, Ramakrishnan Srikant, Privacy Preserving Data Mining, *SIGMOD '00 Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, Volume 29 Issue 2, Pages 439-450, 2000
- [21] 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
- [22] X. Chen, A. Shrivastava, and A. Gupta, Neil: Extracting visual knowledge from web data, In *Proceedings of ICCV*, 2013.
- [23] P. Donmez and J. G. Carbonell, Proactive learning: cost-sensitive active learning with multiple imperfect oracles. In *Proceedings of the 17th ACM conference on Information and knowledge management*, 619-628. ACM, 2008
- [24] T. M. Mitchell, J. Allen, P. Chalasani, J. Cheng, O. Etzioni, M. N. Ringuette and J. C. Schlimmer, Theo: A framework for self-improving systems, *Arch. for Intelligence* 323-356, 1991
- [25] Gregory, P. A. and Gail, A. C. Self-supervised ARTMAP Neural Networks, *Volume 23*, 265-282, 2010
- [26] Cour, T. and Sapp, B. and Taskar, B. Learning from partial labels, *Journal of Machine Learning Research*, Volume 12, 1501-1536 2012
- [27] Adankon, M. and Cheriet, M. Genetic algorithm-based training for semi-supervised SVM, *Neural Computing and Applications* Volume 19(8), 1197-1206, 2010