



INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact factor: 4.295

(Volume 4, Issue 3)

Available online at: www.ijarit.com

Machine learning: Trends and applications

Ashwini Dhoke

ashuahirrao@gmail.com

Dr. D. Y. Patil Institute of
Technology, Pune, Maharashtra

Sarika Kadam

sarikaengg.patil3@gmail.com

Dr. D. Y. Patil Institute of
Technology, Pune, Maharashtra

Shital Sungare

shitalsungare@gmail.com

Dr. D. Y. Patil Institute of
Technology, Pune, Maharashtra

ABSTRACT

As in past fifteen years with the ton of data accessible, machine learning has become one in each of the mainstays of data technology and thereupon, a rather central, albeit sometimes hidden, a part of our life. With the ever-increasing amounts of information changing into accessible, there's the sensible reason to believe that good data analysis can become additional pervasive as a necessary ingredient for technological progress. Machine learning is leverage of self-learning behavioral model. Through the past learning experience, the machine is used to extract hidden value from raw data. Data-intensive nature of machine learning makes it useful in the various aspect of life such as healthcare, marketing, automotive industry, information retrieval, intrusion detection, education and so on. This paper takes a review of various machine learning trends and application.

Keywords: Machine learning, Supervised learning, Decision tree, Support vector machine (SVM), Neural network

1. INTRODUCTION

Machine learning is a branch of artificial intelligence. It used to derive desecrate relations from huge and complex data by applying different statistical and optimal techniques automatically by a machines. In machine learning machine are trained using past data, then this expertise is used in various decision driven applications. Machine learning has progressed dramatically over the past fifteen years .It is emerged as a practical software branch of Artificial intelligence for disease prediction, computer vision, semantic analysis, and many other applications. Data-intensive machine learning is adopted across different industry, such as fault detection in complex model, customer support, and supply chain management. Similarly machine learning has diverse effect on empirical sciences also from social science to biomedical to cosmology. Following are various applications of machine learning that we are using our daily life:

1. Prediction while travelling: Online traffic prediction by Google also online transportation services like Uber.
2. Virtual Assistance: Alexa, Siri are worked as personal assistance
3. Social Media Service: Friend recommendation by Facebook, Similar pins are recognized by Pinterest in image by using computer vision.
4. Spam Filtering: Spam and malware filtering from emails.
5. Customer Support: Many website provide chat boats to talk with customer.
6. Search Engine Result Refinement: Google and other search engine uses ML to personalized search.

With this, we can define Machine learning as a process of self-learning based on training provided to improve the performance. For example, in learning to predict cancer tumor, the task is to assign a label of “malignant” or “non-malignant” to any given biopsy sample. The performance metric to be improved might be the accuracy of this tumor classifier, and the training experience might consist of a collection of historical biopsy report, each labeled in retrospect as malignant i.e. cancerous or not. Alternatively, one might also define a different type of training experience—for example, by including unlabeled biopsy report.

2. MACHINE LEARNING TECHNIQUES

A machine learning techniques are classified into two types of learning: supervised learning and unsupervised learning. Supervised learning is used to make predictions. E.g. Decision Tree, Neural Network, Naïve Bayes, Support Vector Machine.

There are two types of supervised learning:

1. Regression: It is used to predict next value of variable eg. Temperature.
2. Classification: It is used to predict category of data eg. Identification of car model using webcam image.

Unsupervised learning: It is used when data is unknown, by exploring that data internal relations are find out eg. Clustering

Decision Trees (DT)

A decision tree is the type of supervised classification. It is used to predict responses. In this data is checked from root to leaf node. At every node, condition is checked, and as per result data is traveled along the branch. This condition derived while training the tree. The number of branches and the values of the condition is decided at the time of training process. It has following features less memory requirement, easy to interpret.

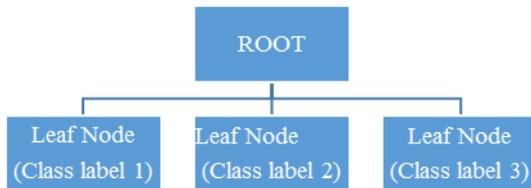


Fig. 1: Decision Tree

Neural Network (NN)

NN are inspired by the human nervous system. It used to predict the no-linear, time volatile data. It is used for incremental learning i.e. as per data availability model will adapt. It is hard to interpret. It consists of a highly connected network of neurons.

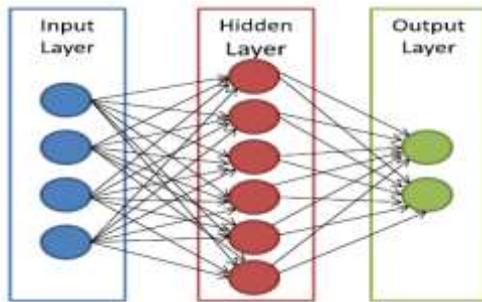


Fig. 2: Neural Network

Support Vector Machine (SVM)

It classifies the data by finding linear decision boundary that separates all data points of one class from all other classes. The best decision boundary is chosen based margin between two classes. It mostly used for the data that exactly has two classes. It is easy to interpret and most accurate one.

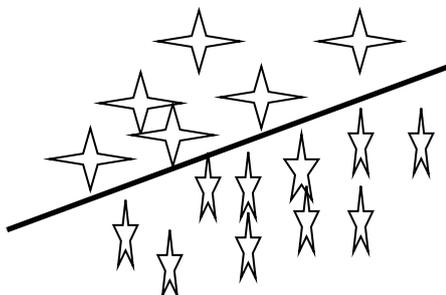


Fig. 3: Support Vector Machine

Naive Bayes Classifier

It classifies the data based on the probability. It assumes that presence of one feature in a class is independent of other feature. It is mostly used when the data set is small, easily interpret and having more number of parameter. It is mostly used in financial and medical data set to train the model when some unknown scenario encountered.

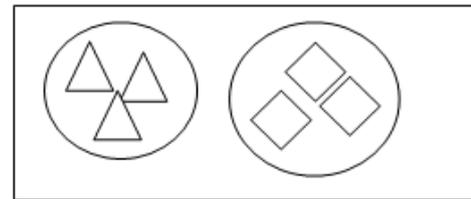


Fig-3: Naïve Bayes Classifier

3. LITERATURE REVIEW

Machine learning techniques are used in the different file. In this chapter, we are going to discuss about various techniques and applications of machine learning used till now.

Sajjad Ahmad et.al [1] used SVM a regression-based approach to measuring soil moisture using remote sensing data. This project carried out in the USA at 6 different sites in the year 1998 to 2005. Through this study, they concluded that SVM model performed batter as compared to ANN model.

Prashant K. Shrivastva et.al. [2] Used Artificial neural network (ANN), support vector machine (SVM), relevance vector machine and generalized linear models the SMOS derived soil moisture. Through their study, they concluded that down scaling gives a better result as compared to continuous time series data to measure soil moisture. ANN model performed better as compared to other models.

Lukas Povoda et al. [3] proposed semantic analysis based on SVM. Semantic analysis is analysis performed on text messages, emoji present on social media like twitter, facebook. In their work, text relevance classifier is proposed for four languages using different machine learning method. As per the resulting classifier work better for the English language.Result is increased by 11% to as compared to traditional method.

J.R. Otukei et.al [4] proposed land cover change assessment over remotely sensed data using different machine learning methods. The pixel base algorithm for the maximum likelihood classifier (MLC), support vector machines (SVMs) and the decision trees (DTs) are used. DTS is checked over dataset the 1986 and 2001 Landsat TM. The results were compared with those obtained using SVMs, and MLC. DTs are performed better as compared to other two methods.

Andre Luckow et.al [5] Used deep learning method i.e sub set of machine learning method in the automotive industry. In this work, they have used machine learning for computer vision inspection in manufacturing and social media analytics. For this k-means, SVM, logistic regression techniques are used.

Min Chen et.al. [6] Proposed CNN based disease risk prediction algorithm for structured and unstructured data. For this data set created in China based hospital chain. For structured data, they have used Naive Bayes classifier, Decision trees, and KNN methods to predict the risk of cerebral infraction diseases.

Dursun Delen[7] develop the model based on machine learning techniques to predict the retention of a student in an institute. In prediction model, he used four machine learning techniques: decision tree (DT), neural network (NN), SVM, and Logical Regression. Also, he used 10 fold cross

validation for model building. Using this model institute came who is going to drop out in the second year. Comparatively, SVM gives better result over the others.

John O. Awoyemi [14] et.al used machine learning techniques to detect credit card fraud. For this, they sourced data from European card holder. It is highly skewed data. After data preprocessing fraud is detected using 3 machine learning method naïve bays, k-nearest neighbor, and logistic regression. Naive bays and k-nearest neighbor gives very good accuracy nearly 98%.

4. PERFORMANCE MEASURE

All Machine learning techniques are evaluated using common formulas used in this field. Following four measures are used to calculate accuracy, efficiency, and reliability of machine learning method.

1. Accuracy = $\frac{TP+TN}{P+N}$
2. Recall = $\frac{TP}{P}$
3. Precision = $\frac{TP}{TP+FP}$
4. F1score = $2 * \frac{Precision * Recall}{Precision + Recall}$

Where TP, FP, TN, FN are truly positive, false positive, true negative, false negative respectively.

Following table gives a rough comparison of machine learning method in a different domain.

Table 1: Rough Comparison of Machine Learning Techniques based on Accuracy

Area	Machine Learning Techniques		
	SVM	DT	ANN
Soil Moisture	High	-	Moderate
Land Coverage	Moderate	High	-
Automotive	High	Low	Moderate
Health care	High	Low	High
Education	High	Low	Moderate
Semantic analysis	Low	High	High

5. CONCLUSION

We have discussed various areas and applications where machine learning can be used. We discussed about techniques used in machine learning application.

We have discussed the use of machine learning in daily life, in health care, fraud detection, education, automotive industry. This gives a brief idea about machine learning. Still, there are many unexplored opportunities. The main issue is how to make a model to learn for lifelong like human beings.

6. REFERENCES

[1] Ahmad, S., Kalra, A., Stephen, H. (2009). "Estimating soil moisture using remote sensing data: A machine learning approach. *Advances in Water Resources*", 33(2010), 69-80. Elsevier.

[2] P.k.Shrivastva Machine Learning Techniques for Downscaling SMOS Satellite Soil Moisture Using MODIS Land Surface Temperature for Hydrological Application *Water Resour Manage* (2013) 27:3127–3144 Elsevier

[3] Povoda, Lukas, R. Burget, and M. K. Dutta. "Sentiment analysis based on Support Vector Machine and Big Data." *International Conference on Telecommunications and Signal Processing* 2016.

[4] Otukei, J.R. and Blaschke, T. (2010) Land Cover Change Assessment Using Decision Trees, Support Vector Machines and Maximum Likelihood Classification Algorithms. *International Journal of Applied Earth Observation and Geoinformation*, 12, S27-S31.

[5] Andre Luckow_, Matthew Cook_, Nathan Ashcraftz, Edwin Weilly, Emil Djerekarov_, Bennie Vorste” *Deep Learning in the Automotive Industry: Applications and tools* “International conference on Big Data 2016.

[6] hen, Min & Yixue, Hao & Hwang, Kai & Wang, Lu & Wang, Lin. (2017). Disease Prediction by Machine Learning Over Big Data From Healthcare communities. *IEEE Access*. PP. 1-1. 0.1109/ACCESS.2017.2694446.

[7] Dursun Delen (2010) “A comparative analysis of machine learning techniques for student retention management” *Journal of Decision Support system* Vol 49.

[8] Chen, Long Sheng, C. H. Liu, and H. J. Chiu. "A neural network based approach for sentiment classification in the blogosphere." *Journal of Informetrics* 5.2(2011):313-322.

[9] Xu X, Zhang Y, Zou L, Wang M, Li A. A gene signature for breast cancer prognosis 896using support vector machine. *IEEE* 2012:928–31.

[10] H. Huang, H. Liu, Big data machine learning and graph analytics: current state and future challenges, in: 2014 IEEE International Conference on Big Data (Big Data), Oct. 2014, pp.16–17.

[11] Y. Lv, Y. Duan, W. Kang, Z. Li, F.-Y. Wang, Traffic flow prediction with big data: a deep learning approach, *IEEE Trans. Intell. Transp. Syst.* PP (99) (2014) 1–9

[12] X. Chen, A. Shrivastava, and A. Gupta, Neil: Extracting visual knowledge from web data, In *Proceedings of ICCV*, 2013

[13] Gregory, P. A. and Gail, A. C. Self-supervised ARTMAP Neural Networks, Volume 23, 265-282, 2010

[14] J. O. Awoyemi, A. O. Adetunmbi and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," 2017 International Conference on Computing Networking and Informatics (ICCNI), Lagos, 2017, pp. 1-9.