ISSN 2278-3091
Volume 9, No.5, September - October 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse264952020.pdf
https://doi.org/10.30534/ijatcse/2020/264952020

Exploring the Machine Learning Algorithms and its Classification

[1]Dr. M. Shanmukhi, [2]Kuruva Harinath, [3]M. Srikanth Sagar, [4]Dr. Annaluri Sreenivasa Rao
[1]Department of CSE, Vasavi College of Engineering, Ibrahimbagh, Hyderabad - 500 031
Telangana, India. Email: shanmukhi.marella@staff.vce.ac.in
[2]Department of IT, Mahatma Gandhi Institute of Technology, Gandipet, Hyderabad, India.
[3]Department of CSE, Mahatma Gandhi Institute of technology, Gandipet, Hyderabad, India.
[4]Deparment of IT, VNR Vignana Jyothi Institute of Engineering and Technology, Bachupally, Hyderabad, India-500090.

ABSTRACT

Over previous years, Machine Learning (ML) has developed from endeavor of some computer enthusiasts developing the probability of computers learning, and a portion of Mathematics that seldom deliberated computational methods, to an independent survey discipline, which has not only offered the essential base for statistical-computational principles of learning processes. This manuscript concentrates on describing the perception and evolution of ML, few popular ML algorithms. ML methods might be applied on any broad area. ML will be procedure of self-learning from practices and performances without human interference. The study of ML methods will be displayed here. And provided the success companies see deriving value from massive number of accessible data, everyone wants in. However, whereas the thought of ML might seem overwhelming, and main thoughts are fairly easy. Now we have displayed an establishment for understanding the thoughts behind few popular ML methods and its classification.

Key words: Machine Learning, cyber-physical systems (CPS), supervised learning, unsupervised learning, classification, Internet of things (IoT), machine-to-machine (M2M), Artificial Intelligence (AI), reinforcement learning (RL), supervised learning (SL), unsupervised learning (UL), Semi-supervised learning (SSL)

1. INTRODUCTION

Machine learning (ML) will be the methodology that naturally enhances from survey and behaves without being unequivocally programmed [1–3]. ML might have been making our registering procedures much dependable, efficient & cost effective. ML generates models by examining even much difficult information automatically, fast & much accurately. It will be basically categorized into semi supervised, supervised, unsupervised, &RL. The ML strength lies in their capacity to give comprehensive answers through a design, which might know to enhance its execution. Due to interdisciplinary behavior, it assumes a critical part in different fields incorporating medical, engineering, & computing. Latest progresses in ML are applied to resolve different problems access massive number of information gathered by sensors, & extract useful data from data will be not so simple without ML. It also utilizes to incorporating IoT, CPS, & M2M [4].

2. ML METHODS

In this segment, we investigate different ML strategies &their learning processes, which support to know the next segments. On the basis of the learning types, ML methods are classified into semi supervised, supervised, unsupervised, &RL. The Figure 1 represents nomenclature of ML methods.

2.1 Supervised Learning (SL)

The SL will be the much significant information processing methods in ML. In SL, we give a group of inputs & outputs, & it discovers the association among them whereas training framework. The main tasks of SL methods have to produce method that signifies associations & estimate objective outputs. The SL resolves different issues in WSNs like localization [5]. The supervised learning classified into classification & regression. The classification might be separated into perceptron based (deep learning & ANN), logic based (RF & DT), instance based (k-NN), and statistical learning (SVM & Bayesian) procedures.

2.1.1 Regression

The regression will be a SL model & it will calculate certain value (Y) on the basis of a provided features set (X). The regression method variables are quantitative or continuous. The regression will be much easy ML method & calculates exact outcomes for least errors. The numerical documentation for linear regression [6] is represented in Eq 1.

$$Y = f(x)+\varepsilon \qquad (1)$$

Here, x specifies independent variable(input),Y be dependent variable(output),f be a function, which it creates the association among x & Y, a nd ε signifies probable random error. The operational method of “simple linear regression” will be represented in Figure 2.

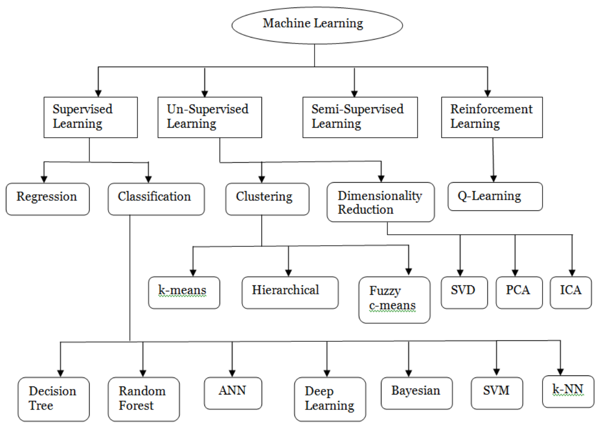

Figure 1: Nomenclature of ML methods

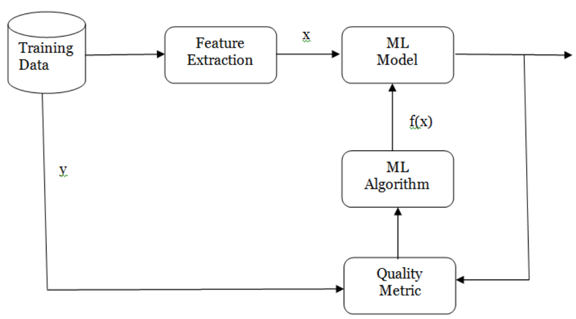

Figure 2: The "simple linear regression" method [6]

2.1.2 Decision trees (DT)

The DTisa category of supervised ML method for categorization relies on group of if then guidelines to improve readability. A DT comprises 2 kinds of nodes named as decision nodes & leaf nodes [7]. The DT utilizes to anticipate a target by making training method built on "decision rules" inferred from training information. A sample diagram representational of DT will be represented in Figure 3. The main benefits of DT are transparent, decrease uncertainty in decision making and permits for complete investigation.

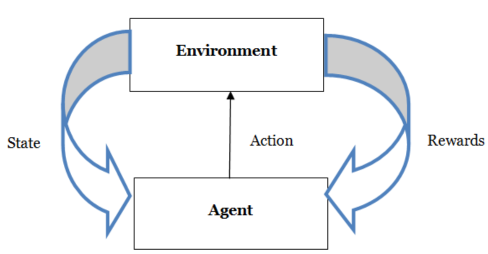

Figure 3: Graphical depiction of a DT [7]

2.1.3 Random forest (RF)

The RF method will be supervised ML method with group of trees & every tree in forest provides a classification. RF method works in 2 phases, RF creation classifier &prediction of outcomes [8]. RF works proficiently for bigger datasets & heterogeneous information. This methodology faultlessly calculates missing values. The effect of choosing a subset of training instances & separating variables at every tree node will generate big amount of DTs. Hence, RF classifier sensitivity level will be low with assessing to other streamline ML classifier due to training instances quality & to over robust DTs. The current classification technologies have facing important issues because of dimensionality. The RF classifier is best proper approach for classifying hyper-spectral information [9].

2.1.4 Artificial neural networks (ANN)

TheANN will be supervised ML method relies on method of human neuron to categorizing the information [10], [11]. The ANN associated with massive amount of neurons, which process data & generate exact effects. ANN commonly works on layers, these layers associated with nodes & every node related with "active function". The Figure 4 indicates essential layer arrangement of ANN. Every ANN holds 3 layers named input, hidden, & output layers. ANN categorizes nonlinear & difficult datasets much simple, & there will be no limitation for inputs such as other categorization approaches. The numerous real time WSN applications are utilizing ANN through it has greater computational necessity.

2.1.5 Deep learning (DL)

The DL will be a supervised ML method utilized for categorization, & it will be sub category of ANN. The deep learning methods have deep learning representation models with multilayer demonstrations. It comprises with basic non-linear modules, which convert representational from "lower to higher layer" to attain better solution [12]. It will be motivated by association designs & data processing in "human nerve frameworks" [13]. The main advantages of DL have extracting "high level features" from data, work without or with labels, & it might be trained to achieve numerous objects. It might be advantageous in different domains like business intelligence, speech recognition, bioinformatics, hand writing recognition, social network analysis, benefits of the business intelligence, medicinal image processing, hand writing recognition, speech recognition, & bioinformatics. The benefits of DL have attracted investigators of WSNs.

2.1.6 Support vector machine (SVM)

The SVM is supervised ML classifier that discovers an "optimal hyper-plane" to classify the information. The SVM executes better categorization utilizing coordinate individual & hyper plane observation [14]. A large amount of training information is redundant once a limit created & a group of points supports to examine the limit. The focuses which have utilized to discover the limit named as SVM.

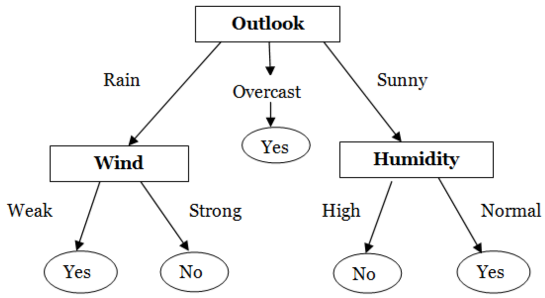

Figure 4: ANN architecture with diverse layers

The SVM gives the better categorization from a provided set of information. Hence, the method difficulty of SVM will be unaesthetic by amount of features run into in training information. For this cause, SVMs have well appropriate to deal with learning responsibilities, whereas amount of features will be big w.r.t the amount of training samples.

2.1.7 Bayesian

Bayesian will be supervised ML method relies on statistical learning methods. The Bayesian learning discovers associations between datasets by "learning conditional independence" utilizing numerous statistical models (Chi-square test). The Bayesian learning permits diverse probability tasks for diverse class node variables.

2.1.8 K-Nearest Neighbor (k-NN)

K-NN will be very lazy, "instance-based learning" model in classification & regression. The k-NN usually categorizes on basis of distance among identified training & testing instances. The k-NN model utilizes numerous distance functions like Canberra, Euclidean, Minkowski, Hamming, Chebychev, &Manhattan distance. The unpredictability of k - NN method relies on input dataset size & optimal presentation, whether the similar scale of information. This method discovers the probable missing values from feature space & decrease dimensionality [15], [16].

The correlations of categorization methods w.r.t different factors have represented in Table 1. The number 1 demonstrates poor presentation, 2 might be deliberated as satisfactory, 3 signify the very good, and 4 demonstrates the better presentation.

2.2 Unsupervised learning (UL)

In UL, there will be no output connected with inputs; even method attempt to extract association from information. The UL method utilized as classifying group of same patterns into clusters, anomaly recognition, &dimensionality decrease, from information. The unsupervised learning further classified into dimensionality reduction (ICA, PCA, & SVD) & clustering (fuzzy c-means, k-.means, & hierarchical).

Terms	DT	ANN	DL	RF	Bayesian	SVM	k-NN
Accuracy	2	3	3	2	1	4	2
Handling noise	2	2	3	3	3	2	1
Speed of classification	4	4	4	4	4	4	1
Managing attributes that are independent highly	2	3	3	2	1	3	1
Managing of aspects	3	1	2	3	4	1	3
Managing variables that are unrequired	2	2	2	2	1	3	2
speed of Learning	3	1	1	2	4	1	4
Fitting	2	1	1	2	3	2	3
Managing irrelevant variables	3	1	2	3	2	4	2
Managing values that are missing	3	1	2	2	4	2	2

Table 1: Comparisons of ML methods with numerous specifications [17]

2.2.1 K-means clustering

The k -means method simply creates specific amount of clusters from provided dataset [18]. Primarily, k-amount of unplannedplaces will be deliberated and remaining focuses connected with closest focuses. Once clusters have created by covering whole points from dataset, a novel centroid from every cluster will be re-estimated. The cluster centroid variation in every iteration, & repeat method until no much variation in centroid of whole clusters. The time complexity of k-means method is *O(n k i d)*, whereas n signifies amount of points, k demonstrates amount of centroids, d signifies the amount of attributes, & i be amount of iterations.

2.2.2 Hierarchical clustering (HC)

The HC scheme integrates similar objects into the clusters that were top-down & bottom-up pre-organized order. Here, the HC that is bottom-up termed as clustering agglomerative in this model, each observation assigns its cluster on the basis of density operations [21] [22]. Moreover, the hierarchical clustering of top-down also termed to be clustering divisive; further, in this model, the huge single split division repeatedly until 1 cluster aimed at every notice. Moreover, in HC model, none of the former data pre-requisite is regarding quantity of clusters, where it would be performed easily.

2.2.3 FCM(Fuzzy-c-means clustering)

The FCM clustering is established by Bezdek in 1981 utilizing "fuzzy set theory" that allocates perception to 1 or more clusters [23]. In this method, clusters have signified on basis of comparison measurements like distance, intensity. Based on datasets or applications, methods might deliberate 1 or many comparison measures. The method iterates on clusters to discover "optimal cluster centers". Such as "k-means clustering", it also needs past information about amount numerous clusters. The FCM time complexity will be higher than other clustering methods, & it mostly based on numerous clusters, iterations, dimensions, & data points. This clustering method utilized I numerous industries like image segmentation, business intelligence, pattern recognition, & bioinformatics, & so on.

2.2.4 Singular value decomposition (SVD)

The SVD will be matrix factorization model that will be utilized to decrease dimensionality. The matrix factorization signifying a matrix into product of matrices [24].

In Eq. (2), a m × n matrix M's SVD is signified as.

$M = U \sum V (2)$

2.2.5 Principle component analysis (PCA)

The PCA will be multivariate investigation feature extraction model for dimensionality decrease [25]. The PCA mixes whole data & drops less priority data from feature space to decrease dimensionality. The PCA output will be linear mixture of detected variables, few time utilized to identify anomalies from regression & data. In WSNs, sensors constantly collect data from environments & transmitting to base station.

2.2.6 Independent component analysis (ICA)

The ICA discovers a novel base for depiction of data & decomposes multivariate perceptions under additive sub-modules. Here, sub-modules are non-Gaussian perceptions [26]. ICA will be a much capable method over PCA, in different words; it will be prolonged version of PCA. The ICA is eliminating the higher sequence dependencies, while PCA might have been incapable to do. ICA investigated information from different application industries like digital images, social networking, psychometric measurements, web content, & business intelligence, & so on.

Specifications	***k* -means**	**Hierarchical clustering**	**Fuzzy-*c*-means**
Accuracy	Low	High	High
Clustering speed	Fast	Fast	Slow
Outcomes of randomness in datasets	Moderate	Good	Moderate
Presentation with minorexplanations in datasets	High	Moderate	Moderate
Average prediction accuracy	High	Low	Low
Quality with massive datasets	High	Moderate	Moderate
Noise datasensitivity	High	Low	Low

Table 2: Comparisons of numerous clustering processes

2.3 Semi-supervised learning (SSL)

The majority of real-life application's information may be mixture of unlabeled & labeled. The supervised learning methods effectively work on labeled data, &UL effectively works on unlabeled information. The SSL presented to work on information with mixture of unlabeled & labeled. In includes "semi supervised classification" to execute categorization on incompletely labeled information, reserved clustering to executes clustering with unlabeled & labeled information, dimensionality diminishment for labeled information & regression with unlabeled information [27], [28].

2.4 Reinforcement learning

The RL method constantly knows by interacting with nature & collects data to take specific actions. The RL improves the presentation by defining the optimal outcome from environment. Figure 5 indicates the purpose of RL. The Q-learning methods are the model free RL methods [29]. In Q-learning, every mediator cooperates with nature& produces an arrangement [30]. The reward matrix R(S, A), whereas S be the set of states & A signify the set of set of actions correspondingly. Learning factor γ, multiplied by extreme value of Q for whole probable actions in subsequent stage as represents in (Eq 3).

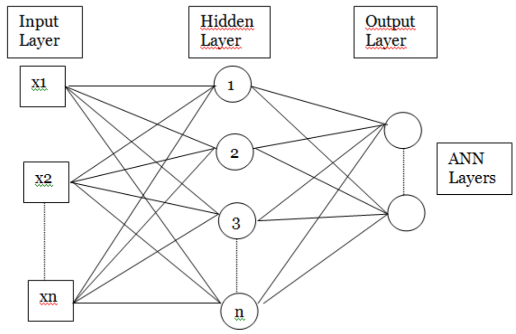

Figure 5: Visualization of RL

$$Q(s_i,a_i)=R(s_i,a_i)+\gamma * Max[Q(nextstate,A)] \quad (3)$$

2.4.1 Evolutionary computation

The evolutionary computation will be issue solving method, which utilizes the computational methods motivated from biological & nature evolution. The evolutional computing will be sub category of AI & it utilizes numerous “combinatorial optimization” methods. In “evolutionary computation”, result of specific issue produces over iteration by iteration [31]. Primarily, it produces a casual group of answers, in each iteration, it eliminates low fit answers as per objective function by trial and error basis to attain optimal outcomes. Figure 6 indicates the common plan of evolutionary method in type of figure. The evolutionary dialects or environment motivated methods incorporate genetic programming, evolutionary programming, “artificial bee colony, artificial immune schemes, memetic procedures”, genetic algorithms, firefly algorithm, evolutionary procedures, differential evolution, and so on[32].

3. Conclusion

This manuscript discovers the numerous ML methods& its classification. The primary target of ML investigators will be to plan many proficient & practical learning approaches, which might execute best over an extensive domain. In ML context, the proficiency with model uses data resources, which are also a significant presentation model along with space &time complexity. The humanly interpretable prediction & higher accuracy of prediction rules have also of high. The investigators might follow the stated algorithms and they might plan their individual thoughts.

REFERENCES

1. T.M.Mitchell,MachineLearning,1stEdition,McGraw-Hill,Inc.,NewYork,NY,USA,1997.
2. T.O.Ayodele,Introductiontomachinelearning,1stEdition,InTech,2010.
3. P.Langley,H.A.Simon,Applicationsofmachinelearningandruleinduction,Communicationsofthe ACM 38 (11) (1995)54–64.
4. M. A. Alsheikh, S. Lin, D. Niyato, H.-P. Tan, Machine learning in wireless sensor networks: Algo- rithms, strategies, and applications, IEEE Communications Surveys & Tutorials 16 (4) (2014)1996– 2018.
5. S. S. Banihashemian, F. Adibnia, M. A. Sarram, A new range-free and storage-efficient localization algorithmusingneuralnetworksinwirelesssensornetworks,WirelessPersonalCommunications98(1) (2018)1547–1568.
6. C. G. Emily Fox, The simple linear regression model, in: Machine learning : Regression, Coursera, 2018.
7. J.R.Quinlan,Inductionofdecisiontrees,Machinelearning1(1)(1986)81–106.
8. L.Breiman,Randomforests,Machinelearning45(1)(2001)5–32.
9. M.Belgiu,L.Drăguţ,Randomforestinremotesensing:Areviewofapplicationsandfuturedirections, ISPRSJournalofPhotogrammetryandRemoteSensing114(2016)24–31.
10. S.S.Haykin,S.S.Haykin,S.S.Haykin,S.S.Haykin,Neuralnetworksandlearningmachines,Vol.3, PearsonUpperSaddleRiver,NJ,USA:,2009.
11. H. White, Learning in artificial neural networks: A statistical perspective, Neural computation 1 (4) (1989)425–464.
12. Y.LeCun,Y.Bengio,G.Hinton,Deeplearning,nature521(7553)(2015)436–444.
13. A.H.Marblestone,G.Wayne,K.P.Kording,Towardanintegrationofdeeplearningandneuroscience, Frontiers in computational neuroscience 10 (2016)1–41.
14. V. N. Vapnik, An overview of statistical learning theory, IEEE transactions on neural networks 10(5) (1999)988–999.
15. S. A. Dudani, The distance-weighted *k*-nearest-neighbor rule, IEEE Transactions on Systems, Man, andCyberneticsSMC-6(4)(1976)325–327.
16. L.E.Peterson,K-nearestneighbor,Scholarpedia4(2)(2009)1883.
17. . B. Kotsiantis, I. Zaharakis, P. Pintelas, Supervised machine learning: A review of classification techniques, Emerging artificial intelligence applications in computer engineering 160 (2007) 3–24.
18. T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, A. Y. Wu, An efficient k-means clustering algorithm: Analysis and implementation, IEEE transactions on pattern analysis and machine intelligence 24 (7) (2002) 881–892.
19. J. P. Ortega, M. Del, R. B. Rojas, M. J. Somodevilla, Research issues on k-means algorithm: An exper- imental trial using matlab, in: CEUR Workshop Proceedings: Semantic Web and New Technologies, 2009, pp. 83–96.
20. K. Almi’ani, A. Viglas, L. Libman, Energy-efficient data gathering with tour length-constrained mobile elements in wireless

sensor networks, in: Local Computer Networks (LCN), 2010 IEEE 35th Conference on, IEEE, 2010, pp. 582–589.
21. S. C. Johnson, Hierarchical clustering schemes, Psychometrika 32 (3) (1967) 241–254.
22. R. G. D'Andrade, U-statistic hierarchical clustering, Psychometrika 43 (1) (1978) 59–67.
23. W. Peizhuang, Pattern recognition with fuzzy objective function algorithms (james c. bezdek), SIAM Review 25 (3) (1983) 442.
24. V. Klema, A. Laub, The singular value decomposition: Its computation and some applications, IEEE Transactions on automatic control 25 (2) (1980) 164–176.
25. S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemometrics and intelligent labora- tory systems 2 (1-3) (1987) 37–52.
26. T.-W. Lee, Independent component analysis, in: Independent component analysis, Springer, 1998, pp. 27–66.
27. X. Zhu, A. B. Goldberg, Introduction to semi-supervised learning, Synthesis lectures on artificial intelligence and machine learning 3 (1) (2009) 1–130.
28. M. F. A. Hady, F. Schwenker, Semi-supervised learning, in: Handbook on Neural Information Pro- cessing, Springer, 2013, pp. 215–239.
29. C. J. Watkins, P. Dayan, Q-learning, Machine learning 8 (3-4) (1992) 279–292.
30. D. L. Poole, A. K. Mackworth, Artificial Intelligence: Foundations of computational agents, Cambridge.
31. Randhir Singh, Saurabh Pal, "Machine Learning Algorithms and Ensemble Technique to Improve Prediction of Students Performance", International Journal of Advanced Trends in Computer Science and Engineering Volume 9, No.3, May - June 2020.https://doi.org/10.30534/ijatcse/2020/221932020
32. Rhagini A , Dhanush Narayanan T , Santhosh M , Sharmila C , Swetha G, "An Unsupervised approach for Predicting the Breast Cancer using K-Means with Compound Feature Generation", International Journal of Advanced Trends in Computer Science and Engineering Volume 9, No.3, March - April 2020.https://doi.org/10.30534/ijatcse/2020/04922020