A SURVEY ON MACHINE LEARNING : TECHNIQUES, APPLICATIONS

Ch.Keerthika Student, Department of Computer Science,Narayana Engineering College Gudur
D.Nihitha Student, Department of Computer Science,Narayana Engineering College Gudur
Dr.V.Sucharitha Professor, Department of Computer Science,Narayana Engineering College Gudur
P.Madhavi Assistant Professor, Department of Computer Science, Narayana Engineering College Gudur
G.Lakshmi sagarika Student, Department of Computer Science, Narayana Engineering College Gudur

Abstract
In machine Learning we have a large number of algorithms, those algorithms are categorized under supervised machine learning, unsupervised machine learning. We also have another technique called reinforcement learning. In this paper we are going to focus on some algorithms that come under supervised and unsupervised techniques.
Keywords: Machine Learning, Algorithm, Supervised learning, unsupervised learning, Networks.

Introduction

The utilization and improvement of PC frameworks that can learn and adjust without adhering to express directions, by utilizing calculations and factual models to break down and attract deductions from designs information. Machine learning is a model that may imply acquiring from past experience to improve future execution. The sole point of convergence of this field is customized learning strategies. Learning insinuates change or improvement of computation reliant upon past "experiences" subsequently without any external assistance from human. Maybe than arranging estimation to address the issue directly using Machine learning ,a researcher search for a procedure through which the machine ,i.e., the computation will devise its own answer reliant upon the model or planning educational record provided for it from the outset[1][2].

The rule part we are focusing in this paper is AI in which Machine learning are made for right estimate making. The essential idea behind AI is if PC could acquire indeed its support increases. It focuses how to normally sort out some way to make accurate figures subject to past discernments and essential piece of AI is Classification .in which models are gathered into given course of action of characterizations. In some cases the precision of machine expectation are significantly more exact than human – created rules. The Primary objective of present day AI is exceptionally precise expectations on test information .There are a few uses of AI, the fundamental of everything is Data mining .Every case in an informational collection utilized by these calculations is addressed same arrangement of highlights[3].

These highlights might be persistent ,categorical, and parallel .If occurrence are given with marks and right yield then it is called supervised learning in any case if case are not named it is unsupervised learning[4][5] ,In this paper we audit various strategies of machine learning and its diverse algorithms. The learning calculations experienced are sorted as administered or unaided calculations.

Machine learning techniques

Machine learning is one of the applications of artificial intelligence which provides a computer system or a machine the ability to learn automatically and also to get improved from the experience without any explicit programming. Generally Machine learning can be categorized into three main learning techniques they are supervised learning, unsupervised learning, reinforcement learning. In Machine learning algorithms learn by example from historical data to predict outcomes and uncover patterns not easily spotted by humans. Machine learning algorithms are completely dependent on data because it is the most crucial aspect that makes model training possible. This makes data preparation the most important step in ML process. Data preparation may be defined as the procedure that makes our dataset more appropriate for Machine learning process[6][7].

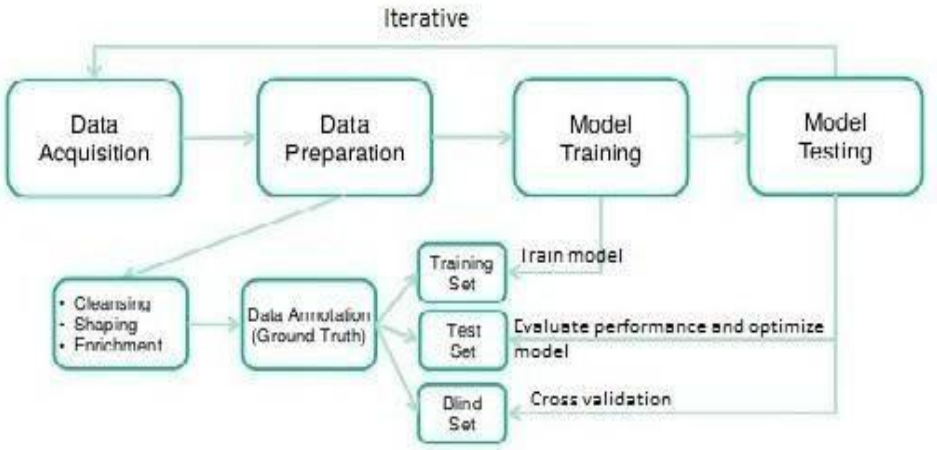

Fig 1: Machine Learning flow diagram

Data Acquisition: It is a part of Machine learning work flow; data acquisition generally takes input as physical conditions and then converts them to digital form for future analysis, storage and evaluation.
Data Preparation: It is also known as data processing, Means Process of Transferring raw data Data preparation also involves three main activities they are Data collection, Data preprocessing, Data transferring.
Model training: Model training includes five activities those are Analysis, Design, Development, Implementation, Evaluation. This last activity Evaluation almost concludes the process.
Model testing: Here the model testing can be referred as the process where performance evaluated from a completely trained model is evaluated in a training set.
As we know we have three learning models in Machine learning let us discuss what actually happens in each learning models
Supervised learning: In supervised learning we first feed output of our algorithm to our system .After feeding the output to the system then preprocessing takes place then the result will be displayed. Also in supervised learning algorithm searches for patterns within the values assigned to the data points. As mentioned in the above diagram (i), from raw data trained data set and desired output are given to the system then after algorithm gets executed it is then takes to next phase of supervised learning i.e., preprocessing, finally we will get the output.

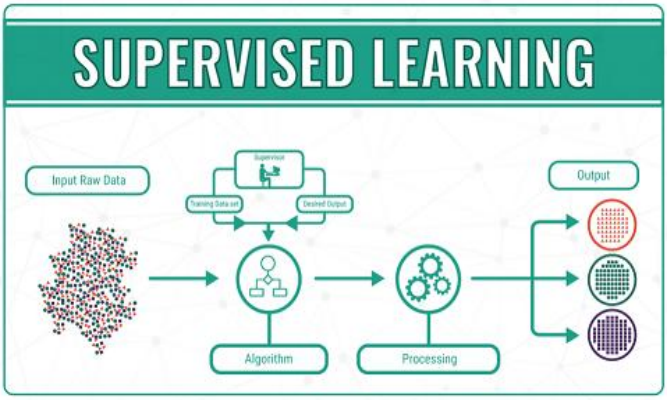

Fig 2.Supervised machine learning

Unsupervised learning: In unsupervised learning, we won't feed output to the system. Also the system does not have any data sets and outputs or results obtained from most of the problems are unknown. The goal of unsupervised learning is to achieve some kind of structure in datasets.

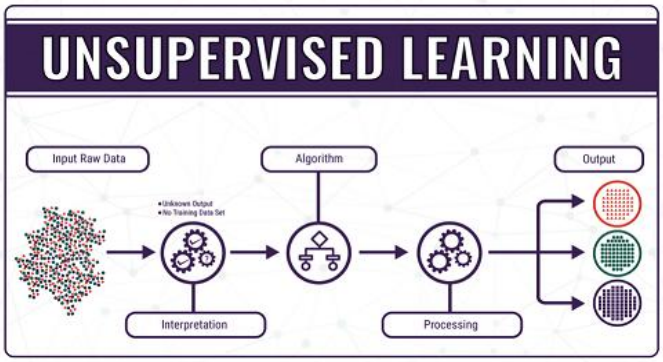

Fig 3.Unsupervised machine learning

Reinforcement learning: By using this algorithm, machine is trained to make certain decisions. To implement reinforcement learning we have three approaches they are value based, policy based, Model based .Goal of this reinforcement learning is maximize the total reward.

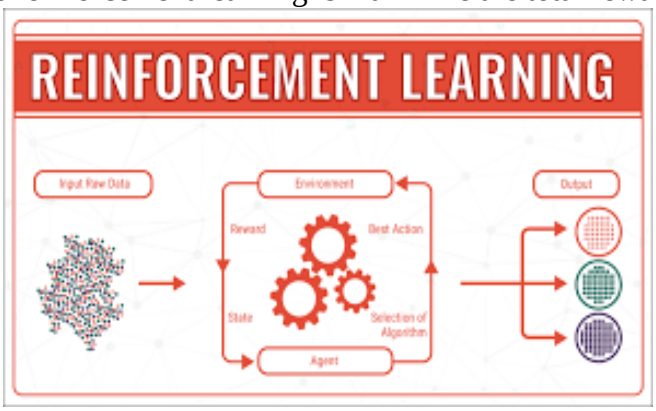

Fig 4.Reinforcement learning.

`Supervised machine learning techniques

Support-vector machines, Linear regression ,Logistic regression ,Naive Bayes ., Linear discriminant analysis, Decision trees, K-nearest neighbor algorithm, Neural networks are some of the supervised machine learning techniques .we mainly focus on Support vector machine, Neural network ,Decision trees , Navie Bayes techniques. All these techniques learn high level concept from low level image features[8].

Support vector machine: Support vector is supervised machine learning also called as SVM. It is one of the most popular supervised learning algorithms, SVM is used for both classification and regression problems[9]. The goal of support vector machine is to make the best line or choice limit that can isolate n-dimensional space into classes so we can undoubtedly put the new information point in the right classification later on. This best line is known as Hyper lane.

SVM picks the limit focuses/vectors that assistance in making the hyperlane. These limit cases are called as help vectors, and consequently calculation is named as Support Vector Machine. Consider the beneath outline in which there are two unique classifications that are ordered utilizing a choice limit or hyperlane.SVM algorithm can be used in different applications like Face detection, Image classification, text categorization.

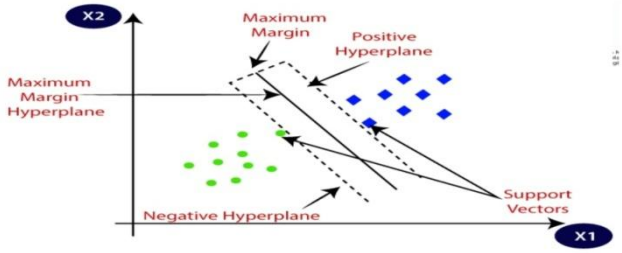

Fig5.Support vector machine

SVM types

SVM can be classified into two types

a.Linear**:** Linear SVM is utilized for linearly distinguishable information, which implies if a dataset can be characterized into two classes by utilizing a single straight line, at that point such information is named as linearly detachable information, and classifier is utilized called as Linear SVM classifier**.**

b.Non-Linear**:** Non-Linear SVM is utilized for non- linearly distinguishable information, which implies if a dataset cannot be characterized into two classes by utilizing a single straight line, at that point such information is named as non-linearly detachable information, and classifier is utilized called as non-linear SVM classifier**.**

Advantages of using SVM

- SVM functions admirably when there is a reasonable edge of partition between classes.
- SVM is more compelling in high dimensional spaces.
- SVM is compelling in situations where the quantity of measurements is more prominent than the quantity of tests.

Disadvantages of Using SVM

- SVM is not suitable for huge data sets
- SVM is not that much efficient with data sets where target classes are overlapping.
- SVM underperforms whenever data point crosses the number of samples in training data.

Applications of SVM

SVM can be applied in Detecting Facial expression, Image classification, text and hypertext categorization

Neural Network

An Artificial Neural Network in the field of Artificial intelligence where it endeavors to mimic the network of neurons makes up a human mind so PCs will have a choice to get things and settle on choices in a human-like way. The artificial neural network is planned by programming PCs to act just like interconnected braincells. To understand the idea of the design of a artificial network, we need to understand what a neural network comprises of. To characterize a neural organization that comprises of an enormous number of artificial neurons, which are named units arranged in a sequence of layers.

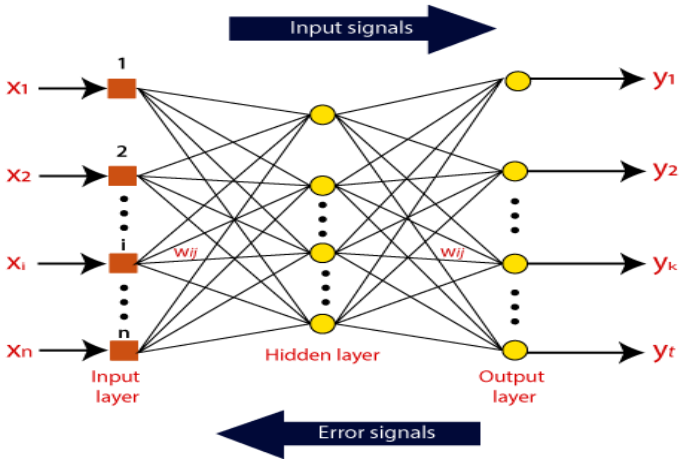

Fig6.Neural Network

Layers of Neural Network

A neural network consists of three layers namely input layer, hidden layer, output layer.Each layer corresponds to particular action.

a.Input layer: As the name it self says, this layer accepts various types of inputs with different formats given by programmer.

b.Hidden layer: This layer is placed in between input and output layers ,it is named as hidden layer because it performs all the calculations in order to find features and patterns which are hidden .

c.Output layer**:** Using this hidden layer the input goes through a series of transformations. This finally results as output which is conveyed using output layer.

The neural network takes input and figures the weighted amount of the information sources and includes an bias. This computation is represented as a form of a transfer function.

$$\sum_{i=1}^{n} Wi*Xi+b$$

It decides weighted complete is passed as a contribution to an actuation capacity to deliver the yield. Enactment capacities pick if a hub should fire. Just the individuals who are terminated make it to the yield layer. There are particular initiation capacities accessible that can be applied upon the kind of undertaking we are performing.

Table 1: Relationship between Biological neural network and artificial neural network:

Biological Neural Network	Artificial Neural Network
Dendrites	**Inputs**
Cell nucleus	**Nodes**
Synapse	**Weights**
Axon	**Output**

Advantages of using neural network

Store information on the entire network, The ability to work with insufficient knowledge, Good fault tolerance, Distributed memory, Gradual Corruption, Ability to train machine, The ability of parallel processing.

Disadvantages of using neural network

Hardware Dependence, Unexplained functioning of the network, Assurance of proper network structure, The difficulty of showing the problem to the network, The duration of the network is unknown.

Applications of neural network

Human Face Recognition, Speech detection , handwriting detection, and also in driving.

Unsupervised machine learning techniques

In this learning technique input information isn't la-belled and doesn't have a known outcome. A model is set up by deriving structures present in the info information and concentrate general principles. It might go through a numerical interaction to deliberately decrease repetition or may put together information through comparability. The primary objective of the model is to decide Data designs/gathering. The information have no objective property and we need investigate the information to track down some natural construction in them. Bunching is a method for discovering similitude bunches in information called groups. Means it bunches information cases that are like each other in one group and information in positions that are altogether different from one another into various bunches. Bunching is regularly called as solo learning as no class esteems signifying and a logical gathering of the information occasions are given. It alludes to an issue of discovering covered up structures in the named information. It has no estimations of result, to control the learning cycle. Picture bunching is commonly unaided learning procedure. It groups the arrangement of pictures in such a manner, that the comparability between various groups should be limited. The Algorithm utilized is Apriori calculation and K-mean calculation.Here we are going to have a look on k-means clustering and principle component analysis.

K-Means Clutsering

It intends to segment the given n perceptions into K clusters. The mean of each cluster is found and the picture is put in a cluster, whose mean has the most least Euclidean distance with the picture include vector. Because of the complex dispersion of the picture information, the k-mean clustering regularly can't separate pictures with various ideas all around ok. clustering like relapse depicts the class of issue and the class of strategies. Bunching techniques are regularly coordinated into two demonstrating approaches as Centroid-based and Hierarchical. The most mainstream among everything is K-mean which essentially goes under the class of grouping in unaided learning. K-mean is a sort of unaided calculation which takes care of the bunching issue. Its methodology follows a basic and simple approach to characterize a given informational index through a specific number of bunches (take as K clusters).Data focuses inside a bunch are homogeneous and heterogeneous to peer gatherings.

Let the set data points be y_1; y_2; :::y_n where y_{i1}; y_{i2}; ::y_{ir} is a vector in a re-valued space Y R^r and here r is the number of attributes in the data. This algorithm partitions the input data into clusters. Each cluster with its centroid .Here k is specified by user.

Cluster formation in K-Mean

- It picks k number of focuses for each group known as centroid.
- Each information point shapes a group with the nearest centroids implies k clusters.
- Find the centroid of each bunch dependent on existing group members.Here we have new centroids.
- Now we have new centroids ,rehash 2 and 3 steps.Find the nearest distance for every information point from new centroid and get related with new k-clusters.Continue rehashing the interaction until we arrive at assembly for example centroids doesn't change.

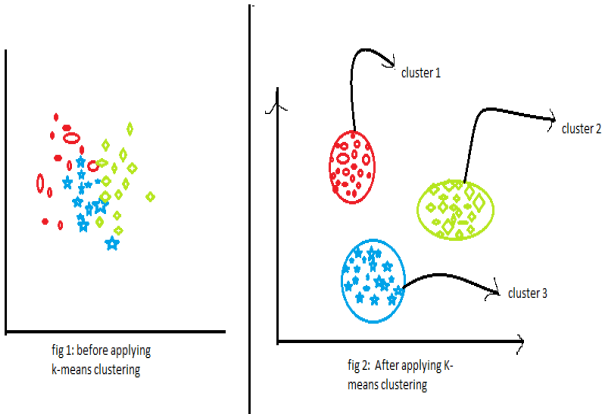

Fig.7.K-means clustering

Advantages of using K-means clustering

Guarantees convergence. Can warm-start the positions of centroids. Easily adapts to new examples. Generalizes to clusters of different shapes and sizes, such as elliptical clusters.

Disadvantages of using K-means clustering

- The client needs to indicate k (the quantity of bunches) at the outset.
- k-means can just deal with mathematical information.
- k-implies expects that we oversee round gatherings and that each bunch has commonly comparable amounts of insights.

Applications of K-means clustering

k-means can be applied to information that has fewer measurements, is numeric, and is persistent. for example, record grouping, distinguishing crime inclined regions, client division, protection misrepresentation recognition, public vehicle information investigation, clustering of IT cautions… and so forth

Conclusion

In this survey various machine learning techniques and their advantages, disadvantages, and their applications were discussed. Machine learning provides software adaptability and flexibility when ever required. From the references and this survey machine learning algorithms are providing the systems to have an ability to learn by itself and improve the experience without any involvement of explicit program.

References

1. Sunpreet Kaur, "A survey on Machine learning Algorithms"International Journal of Innovative Research in Advanced engineering(IJIRAE)issue 11.Volume 3(November 2016).
2. Rabi Narayan Behr,"A Survey on Machine learning:concept,Algorithm,and Applications (IJIRAE) Volume 5,Issue 2,February 2017.
3. P. Venkateswara Rao , A. Ramamohan Reddy , V. Sucharita, Computer Aided Shrimp Disease Diagnosis in Aquaculture. International Journal for Research in Applied Science & Engineering Technology Volume 5 Issue II, February 2017 ISSN: 2321-9653
4. Taiwo Oladipupo Ayodele ,Types of machine learning Algoritms ,new advances I machine learning ,yagang Zhang(Ed),In Tech 2010.
5. Chapelle,O.and Sindhwani,V.and keerthi,s.s.Optimization Techniques for Semi-Supervised Support Vector machines ,Journal of machine learning Research, volume9,203-233,2013.
6. V. Sucharita, S. Jyothi ,D.M. MamathaA Comparative Study on Various Edge Detection Techniques used for the Identification of Penaeid Prawn Species ,International Journal of Computer Applications (0975 – 8887) Volume 78 – No.6, September 2013
7. A Survey on Computer Vision and Image Analysis based Techniques in Aquaculture S. Jyothi, V. Sucharita, D.M. Mamatha- CIIT International Journal of Digital Image Processing, 2013
8. Texture Feature Extraction for the Classsification of Penaeid Prawn Species using Gabor Filter V. Sucharita1, S. Jyothi2 and D. M. Mamatha,Indian Journal of Science and Technology, Vol 8(17), 69538, August 2015
9. Venkateswara Rao, P., Ramamohan Reddy, A., Sucharita, V.An approach of detecting white spot syndrome of peaneid SHRIMP using improved FCM with hybrid back propagation neural network, International Journal of Pharmacy and Technology, 2016, 8(4), pp. 22351–22363