International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2020 IJSRCSEIT | Volume 6 | Issue 3 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT2062111

Machine Learning – Learning Techniques, CNN, Languages and APIs

Krishna Kumar Joshi*[1], Neelam Joshi[2], Ravi Ray Chaudhari[3]
*[1]Computer Science & Applications, ITM University, Gwalior, Madhya Pradesh, India
[2]Computer Science, ITM Group of Institutions, Gwalior, Madhya Pradesh, India
[3]Computer Science & Applications, ITM University, Gwalior, Madhya Pradesh, India

ABSTRACT

Nowadays, Artificial intelligence is an important part in everyone's life. It can be derived in two categories named as Machine learning and deep learning. Machine learning is the emerging field of the current era. With the help of the machine learning, we can develop the computers in such a way so that they can learn themselves. There are various types of leaning algorithms used for machine learning. With the help of these algorithms, machines can learn various things and they can behave almost like the human beings. Nowadays, the role of the machine is not limited in some defined fields only; it is playing an important role in almost every field such as education, entertainment, medical diagnosis etc. In this research paper, the basics about machine learning is discussed we have discussed about various learning techniques such as supervised learning, unsupervised learning and reinforcement learning in detail. A small portion is also used to cover some basics about the Convolutional Neural Networks (CNN). Some information about the various languages and APIs, designed and mostly used for Machine Learning and its applications are also provided in this paper.

Keywords : Machine Learning, Supervised Learning, Unsupervised Learning, Reinforcement Learning, Convolutional Networks.

I. INTRODUCTION

Machine learning helps computer systems in such a way to provide them the ability to learn automatically via various learning algorithms. **The primary aim is to train the computers so that they can work without human assistance and take the decisions themselves [1].**

Nowadays, **Machine learning is showing the importance in various fields such as education, entertainment, health etc. The machine learning process starts with the data observations. Data observation can be through experience (direct or indirect), examples or through instructions [2].**

Machine learning algorithms are categorized into four types named as supervised, unsupervised, semi supervised and Reinforcement (As shown in the figure 1).

In the following section a detailed description is provided on three most popular learning algorithms named as Supervised Learning, Unsupervised Learning and reinforcement Learning. No detail description about semi supervised learning is provided in this paper.

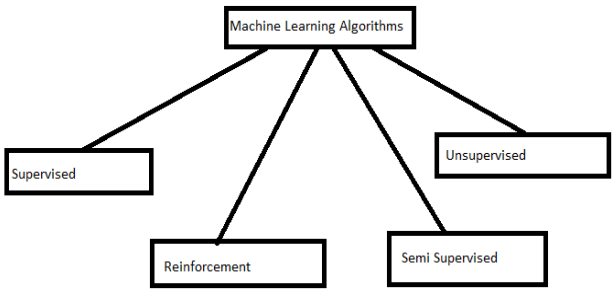

Fig. 1. Types of Machine Learning Algorithms

II. SUPERVISED LEARNIN ALGORITHMS

Supervised learning is one of the popular and important learning methods used in machine learning. In this, input data is pre classified with unique concept labels. For every data instance, we can have the input a and the corresponding output b. And, from this the machine learning system will build a model so that given a new observation a will try to find out what is the corresponding b[3]. The goal of supervised learning is to, based on example input-output pairs, learn a concept definition which best approximates the relationship between input data and the concept labels. An optimal scenario will allow for the algorithm to correctly determine the concept labels for unseen data-items. Fig 2 is showing the concept of supervised learning:

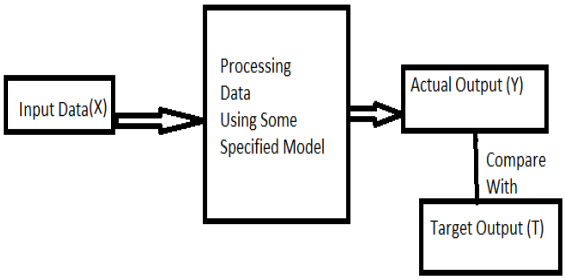

Fig. 2. Supervised Machine Learning Algorithms

It is classified into two types: Regression and Classification (as shown in the figure 3 below)

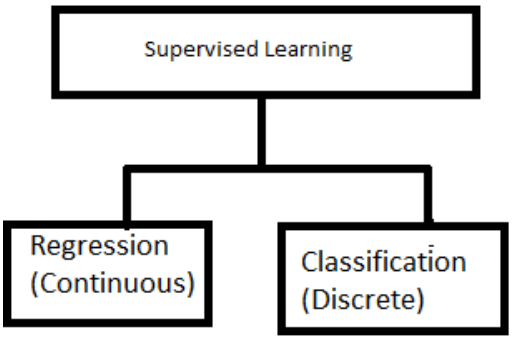

Fig. 3. Classification of Supervised Machine Learning Algorithms

A. Regression

Regression is a technique from statistics that is used to predict values of a desired target quantity when the target quantity is in continuous form. Here, the target quantity can be complex and typically multi dimensional.[4]

B. Classification

Classification is different from Regression. Classification predicts discrete number of values. Here, the data is categorized under different labels according to some parameters and then the labels are predicted for the data. Here, the data can be multi dimensional.

C. Decision Tree

An important method or technique used in supervised learning helps for taking decisions is **Decision Tree**.
In decision analysis, a decision tree can be used to visually and explicitly represent decision alternatives and decision-making options.
Some properties of decision tree are:
(i) A decision tree is drawn upside down with its root at the top.
(ii) Nodes in the decision trees represent features.
(iii) Edges represent feature values or feature intervals (decision options) and
(iv) leaves represent either discrete or continousoutcomes (classification and regression).
A simple view of decision tree is shown in the figure 4 below:

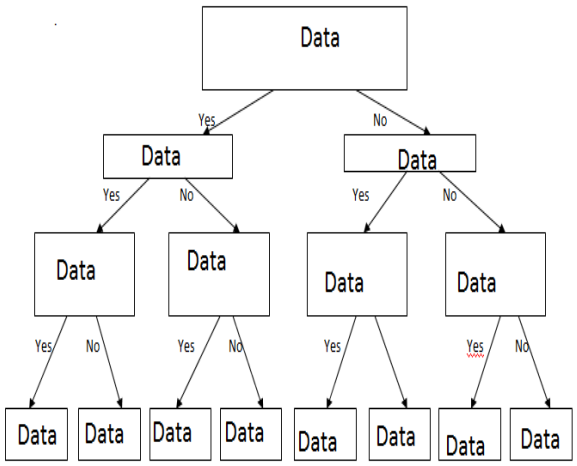

Fig. 4. Decision Tree

Classification of Supervised Machine Learning Algorithms Decision trees are typically used in a purely normative or prescriptive mode in case of the decisionmaking scenario.

In Machine learning applications, in general, we want to build up classification or regression decision trees from data-sets collected in some domain. For such cases, we need to construct an optimal tree design so that the tree optimizes the benefits of considered data-items and the predictive performance for still unseen data-items.

In the Decision tree, evaluation starts from the top and continue with the values of features in the given order eventually ending up with a unique outcome.

Decision trees analysis can be Classification tree analysis or Regression Tree analysis. The only difference between these two approaches is that in Classification tree analysis the leaves are the classes, while in Regression tree analysis the leaves are real numbers or intervals.

D. Support Vector Machine (SVM)

Support Vector Machine is a kind of machine learning algorithm operates in supervised mode with pre-classified examples but it can also operate in an incremental model. SVM is an instance of a non-probabilistic binary linear classifier system, where binary means that it classifies instances into two classes and linear means that the instance space have to be linearly separable. It shows very much significance and popularity because of it is very easy to use and has good generalization performance and the most important thing is that the same algorithm solves a variety of problems with little tuning [7].

III. UNSUPERVISED LEARNIN ALGORITHMS

In Unsupervised learning, input data are not classified i.e. contains only input data and lack concept labels. Unsupervised learning algorithms therefore have to identify commonalities and structures in the data-set and to group the input based on similarity.[8]

Basic concept of Unsupervised Learning is shown in the figure 5, below:

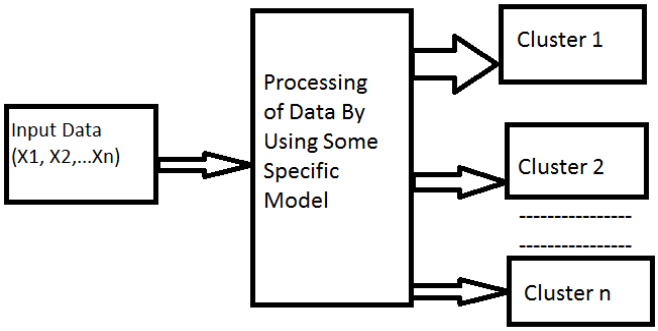

Fig. 5. Unsupervised Machine Learning Algorithms

Unsupervised learning algorithms have to decide on a optimal portfolio of concepts that best matches the data-set and arrange groupings of subsets of the data-set so that it matches the portfolio of concepts. these algorithms therefore have to identify commonalities and structures in the data-set and to group the input based on similarity. Here, the data sets are divided into clusters.

Clustering and Cluster Analysis

Cluster analysis is the assignment of a set of observations into subsets, called clusters. The observations within the same cluster are similar according to one or more pre-designated criteria, while observations drawn from different clusters are dissimilar. Clustering is shown in the figure below:

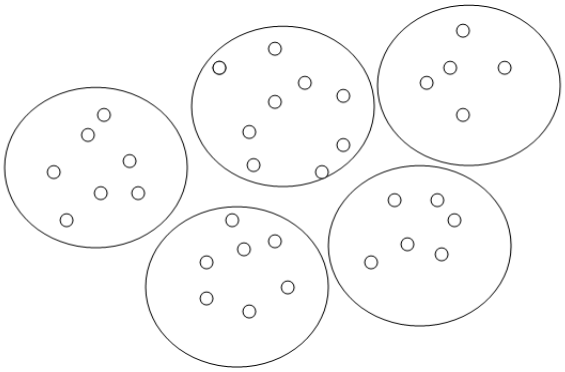

Fig. 6. Clustering

Cluster analysis is an important element in unsupervised concept learning, i.e. learning of multiple concepts from unsorted examples . It can be used as a stand alone technique for particular categorization purposes. As instances are not classified in the unsupervised scenario, algorithms have to identify commonalities and structures in the data-set and to group the instances based on similarity.

Cluster analysis is the task of grouping a set of objects in such a way that objects in the same group called a cluster are more similar (in some sense) to each other than to those in other groups (clusters).[9]

Cluster analysis can be achieved by various algorithms that differ significantly in their understanding of what constitutes a cluster and how to efficiently find them. In General, Clustering algorithms are dependent on several hyper parameter settings.

Some important parameters are:

(i) Number of Clusters to establish
(ii) Number of features used to describe instances
(iii) Type of distance measure to employ
(iv) Treshold value
(v) Alternative density threshold measures
(vi) Number of sessions for inspection of the data-set

Following clustering techniques are used in Machine Learning

(a) Hierarchical clustering
(b) Partitioning–based clustering
© Density-based clustering
(d) Grid-based clustering
(e) Model-based clustering

IV. REINFORCEMENT LEARNIN ALGORITHMS

Reinforcement learning is another learning technique used in Machine Learning. In Reinforcement learning there are three important terms: **Agents, Actions** and **Environment.** Reinforcement learning is concerned with how agents take *actions* in an *environment* and changing state so as to maximize some notion of cumulative *reward*. In general, the agent can concisely manage the handling of the reward so that components of the agent can be modified to produce a new state. Also, It is assumed that the typical scenario for re-inforcement learning, is a strong-theory-based hardware and software system and that adaption happens on the margin of the systems behaviour[10].

The mapping from states to possible actions is called a **Policy**. The achievement of goals is defined by **Rewards** or reward signals being the feedback on actions from the environment.

The **Return** is defined as the cumulative over an **Episode** = action sequence, leading to a **Terminal state**. The main goal of Reinforcement Learning algorithm is to establish a policy that maximizes the **Returns**.

Concept of Reinforcement learning is shown in the figure below:

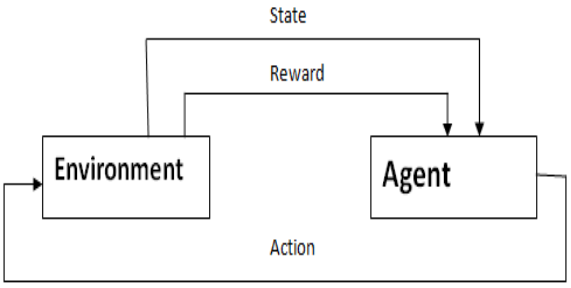

Fig. 7. Reinforcement Machine Learning Algorithms

V. STEPS INVOLVED IN MACHINE LEARNING PROCESS

Machine Learning Process can be classified into seven steps:

(i) Data Collection
(ii) Data Preparation
(iii) Choose a Model
(iv) Train the Model
(v)Evaluate the Mode
(vi) Parameter Tuning
(vii) Make Predictions.

Phases of Machine learning process are shown in the figure 8.

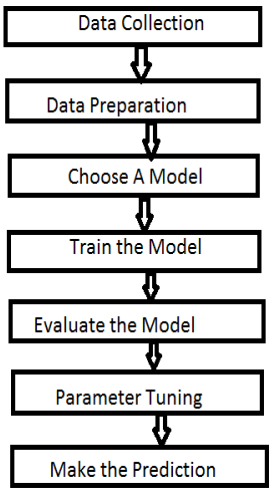

Fig. 8. Phases of Machine Learning Process

In Data Collection phase the required data is gathered by using some requirement gathering steps [13]. In Data Preparation phase the collected data is prepared for the processing. Cleaning the data is an important step in this phase. In Data Collection phase the required data is gathered by using some requirement gathering steps. The collected data should be accurate[4]. In Data Preparation phase the collected data is prepared for the processing. Cleaning the data is an important step in this phase. By using cleaning the data various errors, redundancy, false values, empty or missing values can be removed or corrected. [14]. In the third phase the particular model is selected that is relevant and most efficient for the given dataset and tasks.

Now, train the model selected in the previous phase. The main goal of the training is to find the correct prediction as expected for the particular inputs.. To evaluate the trained model some measurement matrices are used. The main motive of the evaluation of the trained model is to check the performance of the model.[3] The next step is Parameter Tuning. This step is used to enhance the performance of the performance parameters. The Last and the final step is Make Prediction. In this phase some additional data which is not used previously, used in this step to check the validity of this model.

VI. CONVOLUTIONAL NEURAL NETWORKS (CNN)

Convolutional Neural Networks (CNN) was developed primarily for image recognition problems. Its architectures are strongly influenced by human and animal visual perception model. CNNs use relatively little pre-processing compared to other image classification algorithms. This means that the network learns the filters that in traditional algorithms were hand-engineered. This independence from prior knowledge and human effort in feature design is a major advantage[5].

Convolution is a mathematical operation on two functions (f and g) to produce a third function that expresses how the shape of one is modified by the other. The term convolution refers to both the result function and to the process of computing it. Convolution is similar to cross-correlation and related to autocorrelation.

It can be expressed as:

$$f(x) * g(x) = \int_{-\infty}^{\infty} f(a)g(x-a)da$$

Here, we express each function in terms of a dummy variable (let, a). then, Reflect one of the functions: g(a) → g(-a) and after that, Add a time-offset, x, which allows g to slide along the a-axis from $-\infty$ to $+\infty$.

Wherever the two functions intersect, find the integral of their product.

VII. PROGRAMMING LANGUAGES/API USED FOR MACHINE LEARNING

Here, are some important programming languages which are typically used for Machine Learning:

(i) Python

It is an open source, interpreted, high-level, general-purpose programming language with a dynamic type system and supports multiple programming paradigms, including object-oriented, functional and procedural.

(ii) R

R is an open source programming language and software environment for statistical computing and graphics. It is widely used among statisticians and data miners for developing statistical software and data analysis.

(iii) Julia

Julia is an open source high-level general-purpose dynamic programming language developed mainly for high-performance numerical analysis and computational science.

(iv) Scala

Scala is an open source, general-purpose programming language providing support for object oriented and functional programming with a strong static type system. [11]

(v) Lisp

Lisp is an open source programming language which is commonly used for Artificial Intelligence.

(vi) Java

Java is a high-level general-purpose programming language used for multiple paradigms.

Some important APIs used for Machine learning are:

(i) Anaconda

Anaconda is one of the most popular open sources API used for ML. In general, it is a distribution of the Python and R programming languages. Mostly used for scientific computing application such as data science and machine learning applications

(ii) TensorFlow

Tendsorflow is one of the most popular open source API used for implementation of ML application. it was developed by Google in 2015.

(iii) Microsoft Cognitive Toolkit

Microsoft Cognitive Toolkit is an open-source toolkit for commercial-grade distributed deep learning. It was developed by Microsoft in 2016. It describes neural networks as a series of computational steps via a directed graph.[11]

(iv) Theano

Theano is a popular Python library that helps to define, optimize, and evaluate mathematical expressions.

(v) Jupyter Notebook

Jupyter Notebook is an open source, which is a product of the Project Jupyter that supports execution environments in several dozen languages. Project Jupyter's name is a reference to the three core programming languages supported by Jupyter, which are Julia, Python and R.

(vi) Caffe

Caffe is a deep learning framework made with expression, speed, and modularity in mind. It is developed by Berkeley AI Research and by community contributors.[12]

VIII. APPLICATIONS OF MACHINE LEARNING

With the elevate in immensely colossal data, machine learning has become a key technique for solving quandaries in areas, such as:

- Image processing and computer vision, for face apperception, kinetics detection, and object detection
- Biological applications such as for cancer detection, drug revelation, and DNA sequencing [5]
- Aerospace for predictive maintenance
- Natural language processing for voice controlled applications
- Trading and investments in financial applications.

- In Marketing and Sales to analysis of the purchase history of the customers and make personalized product recommendations for their next purchase.
- Government projects to identify the methods to minimize costs and increment in efficiency.

IX. CONCLUSION AND FUTURE SCOPE

Machine learning is emerging technology in the modern era. It is playing an important role in almost every field of human life. Because of the accuracy in the result and the processing speed for computation it now became the helping hand for the everyone. In this paper we have tries to reveal every information related with machine learning in this paper a detailed description on machine learning algorithms (Supervised, Unsupervised and Reinforcement Learning) is provided including the important terms used in these algorithms. Machine Learning is also entered into medical fields. It is used for medical diagnosis technologies such as MRI, CT SCAN and many diagnosis procedures. In future we want to design an algorithm based on the application of machine learning in medical diagnosis.

X. REFERENCES

[1] Jose A. Quesada, Adriana Lopez-Pineda, Vicente F. Gil-Guillén, Ramón Durazo-Arvizu, Domingo Orozco-Beltrán, Angela López-Domenech, Concepción Carratalá-Munuera "Machine learning to predict cardiovascular risk " in The International Journal of Clinical Practice, 2019, DOI: 10.1111/ijcp.13389 .

[2] Sayan Chatterjee, Saruk Chand Sk, Manish Kumar Singh, Judhajit Sanyal,"Machine Learning Based Prediction of Energy Consumption", International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering ol. 7, Issue 5, May 2019, DOI 10.17148/IJIREEICE.2019.7506 19

[3] Purushottama, Kanak Saxenab, Richa Sharma, "Efficient Heart Disease Prediction System" ScienceDirect Procedia Computer Science 85 (2016) 962 – 969, doi: 10.1016/j.procs.2016.05.288

[4] Zhuang Yu, Haijiao Lu, Hongzong Si, Shihai Liu, Xianchao Li, Caihong Gao,Lianhua Cui, Chuan Li, Xue Yang, Xiaojun Yao, "A Highly Efficient Gene Expression Programming (GEP) Model for Auxiliary Diagnosis of Small Cell Lung Cancer" PLOS ONE | DOI:10.1371/journal.pone.0125517 May 21, 2015

[5] Phan Duy Hung, Tran Duc Hanh, Vu Thu Diep, "Breast Cancer Prediction Using Spark MLlib and ML Packages"*ICBRA '18*, December 27–29, 2018, Hong Kong, Hong Kong © 2018 Association for Computing Machinery, DOI: https://doi.org/10.1145/3309129.3309133

[6] Madhuri Gupta, Bharat Gupta, "An Ensemble Model for Breast Cancer Prediction Using Sequential Least Squares Programming Method (SLSQP)" Proceedings of 2018 Eleventh International Conference on Contemporary Computing (IC3), 2-4 August, 2018, Noida, India

[7] Purushottam, Kanak Saxena, Richa Sharma, "Efficient Heart Disease Prediction System using Decision Tree" International Conference on Computing, Communication and Automation (ICCCA2015)

[8] R. Sathya and Annamma Abraham, "Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification", International Journal of Advanced Research in Artificial Intelligence, Vol. 2, No. 2, 2013

[9] Ayon Dey" Machine Learning Algorithms: A Review", International Journal of Computer Science and Information Technologies, Vol. 7 (3) , 2016, 1174-1179

[10] Iqbal Muhammad and Zhu Yan, "SUPERVISED MACHINE LEARNING APPROACHES: A

SURVEY", ICTACT JOURNAL ON SOFT COMPUTING, APRIL 2015, VOLUME: 05, ISSUE: 03

[11] Pulkit Kalia, "Stacking Supervised and Unsupervised Learning Models for Better Performance", International Research Journal of Engineering and Technology (IRJET) Volume: 05 Issue: 09 | Sep 2018

[12] Kajaree Das , Rabi Narayan Behera, "A Survey on Machine Learning: Concept, Algorithms and Applications", International Journal of Innovative Research in Computer and Communication Engineering Vol. 5, Issue 2, February 2017

Cite this article as :

Sh