Introduction to Machine Learning and Its Applications: A Survey

Vanita Tonge Buradkar, Manisha More
Assistant Professor
Department of Computer Science & Engineering
Rajiv Gandhi College of Engineering Research & Technology, Chandrapur
Corresponding Author's email id: *vsburadkar@gmail.com, more.manisha3030@gmail.com*
DOI: *http://doi.org/10.5281/zenodo.3775092*

Abstract

Machine learning is the fastest growing areas of computer science. It has the ability to lets the computer to create the program. It is a subset of Artificial Intelligence (AI), and consists of the more advanced techniques and models that enable computers to figure things out from the data and deliver. It is a field of learning and broadly divided into supervised learning, unsupervised learning, and reinforcement learning. There are many fields where the Machine learning algorithms are used. The objective of the paper is to represent the ML objectives, explore the various ML techniques and algorithms with its applications in the various fields from published papers, workshop materials & material collected from books and material available online on the World Wide Web.

Keywords*: - Machine Learning (ML), Artificial Intelligence (AI), Optimization, Supervised, Unsupervised, Reinforcement, Clustering.*

INTRODUCTION

The term Machine Learning was coined by Arthur Samuel in 1959, an American pioneer in the field of computer gaming and artificial intelligence and stated that, "It gives computers the ability to learn without being explicitly programmed".

The traditional programming is differs significantly from machine learning algorithm as follows [15]

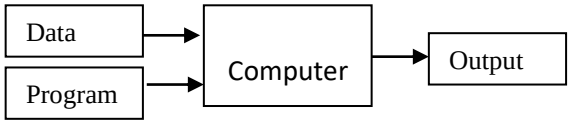

Fig.1 Traditional Program

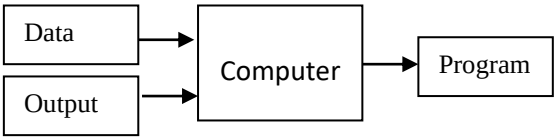

Fig.2 Machine learning

Huge amount of data is generated daily from many sources called as big data. Big Data is a collection of large datasets of increasing volume, velocity and variety such large amount of data is hard to manage and process.[1] To handle such a massive load of data techniques like Hadoop, data Mining, Machine learning etc. are used . Machine learning is diverse and exciting field to handle big data and there are multiple ways to define it. It converts data into program and automate the automation system. Using past experience it teaches computer how to perform task by providing examples.

Optimization is one of the core components of machine learning and artificial Intelligence [18][19]. The essence of most machine learning algorithms is to build an optimization model and learn the parameters in the objective function from the given data[13]. Optimization algorithm for example Genetic algorithm, Differential algorithm,ant colony optimization, Particle Swarm Optimization[18][19] etc. have great influence on various fields of machine learning. In machine learning, genetic algorithms were used in the 1980s and 1990s. Conversely, machine learning techniques have been used to improve the performance of genetic and evolutionary algorithms[7].Scheduling is a decision making process. To solve this problem improved differential evolutionary algorithm, differential evolution algorithm using classical NEH and iterated local search with enhanced swap operator and DE with position based crossover operator with the objectives of minimizing makespan is used[14].

The Travelling Salesman problem (TSP) or Hamiltonian tour is a type of classic old problem and one of the benchmark in Computer Science and Operations Research. Heuristic optimization algorithms such as genetic algorithm, tabu search and ant colony optimization have been widely used to find optimal solution in travelling salesman problem. Standard GA have a good performance for finding the promising regions of the search space, they are not so successful at determining local minimum in terms of convergence speed. In order to overcome these disadvantages of GA in numeric optimization problem, Differential

evolution algorithm has been introduced by Storn and Price [16].

Steps of machine learning Algorithm.

- **Problem Framing:** frame a machine learning problem in terms of what we want to predict and what kind of observation data we have to make those predictions.

- **Gathering data:** Input the data it can be structured or unstructured.

- **Data Preparation:** Loading Data into suitable place and prepare it for machine learning training.

- **Choosing Model:** This is important step to choose a proper model to implement and predict the output. There are many models that are created by researchers and data scientist over the years. Some are very well suited for image data and some are for numeric data.

- **Training data:** Data incrementally improves the model's ability to predict the output.

- **Evaluation:** Evaluation allows us to test our model against data that has never been used for training.

- **Parameter Tuning:** To improve the further training some parameters are assumed and try other values.

- **Prediction:** This is the step where we got the answer.[17]

MACHINE LEARNING ALGORITHMS

Machine learning is classified as supervised learning, unsupervised learning, Semi-Supervised learning and reinforcement learning.

A. Supervised Learning Algorithms

Supervised learning is simple and easy to understand. It is based on prior information simple like a mother teaches to her child. Given data are labeled data for example weather data, time of the day, Holidays, Route chosen, House price etc. learning algorithm used these labeled data and try to approximate the given labeled data with the input and trained the machine.

Once the machine is trained it starts to predict and give decision when new data is

given to it. Following are the different supervised algorithms Regression technique predicts a single real valued output using training data. Here are few types of regression techniques.

1. **Decision tree:** it is widely used technique for classification. Classification is a tree each node is the features of an instance and every branch represents a value which the node can assume. This method of classification is known as a decision tree.Decision tree algorithm: Decision tree algorithm is simple supervised machine learning algorithm. It includes a root node, branches, and leaf nodes. Each internal node denotes a test on an attribute, each branch denotes the outcome of a test, and each leaf node holds a class label.

The topmost node in the tree is the root node.To understand the concept of Decision Tree consider an example. Let's say you want to predict whether a student succeeded in exam or not. Information is given like marks scored in theory exam and practical exam etc. The decision nodes are the questions like 'What's the score in theory?', 'score in practical?', 'Does he score more than 35% in both? And the leaves represent outcomes like either 'pass', or 'fail'.

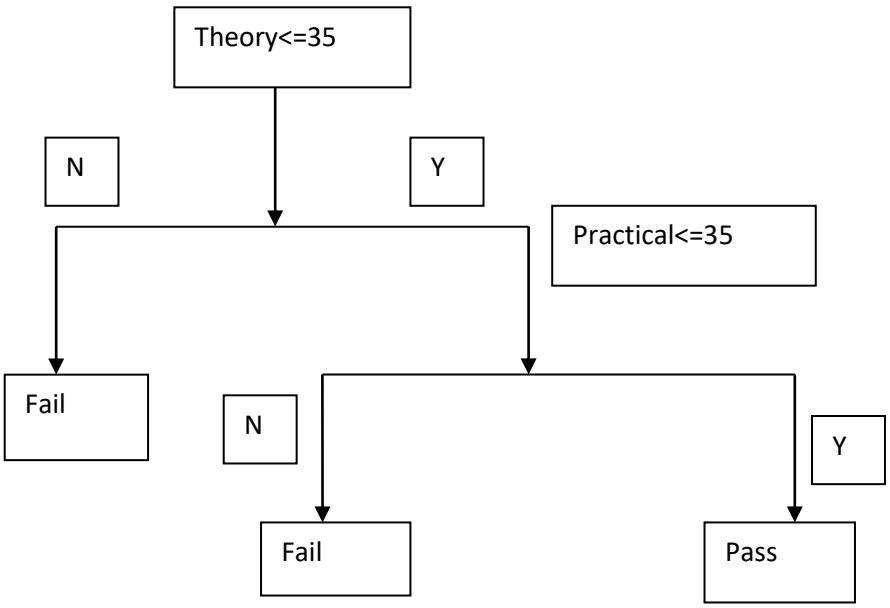

Fig. 3 Example of Decision Tree

2. **Random forest:** Random forests are the most flexible and easy to use supervised learning algorithm. It can be used both for classification and regression. But it is mostly used for the classification. A forest is comprised of trees. As it consist more trees, the more robust forest it has. Random forest algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting. Random forests have a variety of applications, such as recommendation engines, image classification and feature selection. Further classification algorithms are KNN, Trees.

3. **Logistic Regression:** Logistic regression method used to estimate probability of the target value. Target value is discrete value which means it is in binary form having data coded in 1 for success/yes and 0 for failure/no.

4. **Naïve Bayes:** Naïve Bayes is one of the probabilistic machine learning algorithms based on Bayes theorem. It is advantageous for text data. Application includes filtering spam, classifying documents, sentiment prediction etc. Bayes theorem named after Rev. Thomas Bayes. It works on conditional probability. Conditional probability is the probability that something will happen, given that something else has already occurred. Using the conditional probability, we can calculate the probability of an event using its prior knowledge. Bayes' theorem is stated mathematically as the following equation:

   $$P(A/B)=(P(B/A)P(A))/P(B)$$

   where A and B are events.

5. **Support vector machine:** Support Vector Machine (SVM) was first heard in 1992, introduced by Boser, Guyon and Vapnik in COLT-92. Support vector machines (SVMs) are a set of related supervised learning methods used for classification and regression [7].But generally, they are used in classification problems [18]. SVMs have their unique way of implementation as compared to other machine learning algorithms. Lately, they are extremely popular because of their ability to handle multiple continuous and categorical variables [8].

Challenges in Supervised machine learning

Here, are challenges faced in supervised machine learning:

- Irrelevant input feature present training data could give inaccurate results.
- Data preparation and pre-processing is always a challenge.
- Accuracy suffers when impossible, unlikely, and incomplete values have been inputted as training data.
- If the concerned expert is not available, then the other approach is "brute-force." It means you need to think that the right features (input variables) to train the machine on. It could be inaccurate.[9]

Advantages of Supervised Learning

- Supervised learning allows you to collect data or produce a data output from the previous experience.
- Helps you to optimize performance criteria using experience
- Supervised machine learning helps you to solve various types of real-world computation problems.[9]
- The main advantage of supervised learning algorithm is clarity of data and ease of training.
- Supervised learning can be very helpful in classification problems.[10]

Disadvantages of Supervised Learning

- Supervised learning is limited in a variety of sense so that it can’t handle some of the complex tasks in machine learning.
- Supervised learning cannot give you unknown information from the training data like unsupervised learning do.
- It cannot cluster or classify data by discovering its features on its own, unlike unsupervised learning.
- As well as many disadvantages such as the inability to learn by itself.[10]
- Classifying big data can be a real challenge.
- Training for supervised learning needs a lot of computation time.[9]

B. Unsupervised Learning Algorithms

Unsupervised learning is opposite of supervised learning. Instead of using label data unsupervised learning is fed with lot of data and a tool to understand the properties of the data. Unsupervised learning algorithm helps you to finds all kinds of pattern in data. Output of the unsupervised learning is group or cluster of data having similar characteristics. Types of unsupervised learning are clustering and association.

Clustering: It is an important concept when it comes to unsupervised learning. It mainly deals with finding a structure or pattern in a collection of uncategorized data. [9]

Following are some clustering algorithms.

K-means- K-means is the clustering algorithm used to determine the natural spectral grouping present in a data set. [6] K-Means comes under unsupervised clustering method. Data will be partitioned into k clusters, based on their features. Each cluster is represents by its centroid, defined as the center of the points in the cluster. K-Means is simple and fast but it doesn't yield to the same result with each run. [12] Single value decomposition and principal Component Analysis are also examples of Clustering algorithms.

Association: Association rules allow you to establish associations amongst data objects inside large databases. This unsupervised technique is about discovering interesting relationships between variables in large databases. For example, people that buy a new home most likely to buy new furniture. [9]Analysis like Apriori, FP-growth.

C. Semisupervised Learning Algorithm

Semi-supervised learning falls between unsupervised learning and supervised learning. Many machine-learning researchers have found that unlabeled data, when used in conjunction with a small amount of labeled data, can produce a considerable improvement in learning accuracy. [7]

D. Reinforcement Learning Algorithm

Reinforcement algorithm is different from supervised and unsupervised algorithm. Supervised algorithm trains the data with answer key whereas reinforcement learning algorithm trains data without correct answer key. Reinforcement learning can be understood using the concepts of agents, environments, states,

actions and rewards. The reinforcement learning agent decides what to do in order to perform the given task. In absence of training data set reinforcement learning agent uses the experience. To connect the agent to the environment, we give it a set of actions that it can take that affect the environment. To connect the environment to the agent, we have it continually issue two signals to the agent: an updated state and a reward.

The main difference between these types is that supervised ML algorithm learn from training data, unsupervised ML algorithms doesn't required any training data whereas Reinforcement ML algorithm learn on its own.[10][11][12]. Reinforcement learning differs from other types of supervised learning, because the system isn't trained with the sample data set. Rather, the system learns through trial and error. Therefore, a sequence of successful decisions will result in the process being reinforced, because it best solves the problem at hand [18].

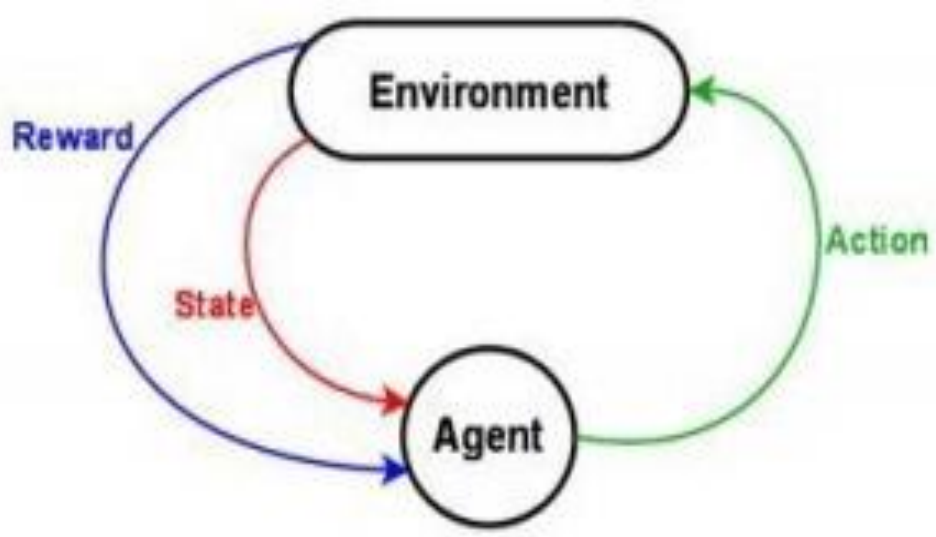

Fig. 4 Flow of reinforcement algorithm

Comparison between Supervised and Unsupervised Learning Algorithm

Basis for Comparison	Supervised Algorithm	Unsupervised Algorithm
Input Data	Uses Known and Labeled Data as input	Uses Unknown Data as input
Method	Both Input and Output data is given	Only Input data is given
Output	Predicts the output	Finds the hidden patterns in data
Computational Complexity	Very Complex	Less Computational Complexity
Real Time	Uses off-line analysis	Uses Real Time Analysis of Data
Number of Classes	Number of Classes are known	Number of Classes are not known
Sub-Domain	Classification and Regression	Clustering and Association
Accuracy of Results	Accurate and Reliable Results	Moderate Accurate and Reliable Results

APPLICATION OF ML

Machine learning algorithms are used in wide area of research such as:

Financial Services: Machine learning rapidly earns popularity in financial services such bank to prevent fraud and also to find patterns in data.

Sales and Marketing: Massive data is generated from marketing sites. With the boom of data, marketing department relies on machine learning to optimize the relationship between customer and product campaign.

Government: The government makes use of ML to manage public safety and utilities.

Healthcare: Machine learning is playing an important role in healthcare. Today, machine learning is helping to streamline administrative processes in hospitals, map and treat infectious diseases and personalize medical treatments.

Transportation: Based on the travel history and pattern of traveling across various routes, machine learning can help transportation companies predict potential problems that could arise on certain routes, and accordingly advise their customers to opt for a different route.

LITERATURE SURVEY

"A Combination of Machine Learning and Image Processing Technologies for the Classification of Image Regions" In this paper authors discuss about an approach to applying machine learning (ML) technologies for the automated creation of classification rules for image regions is introduced [2].

"DECISION TREES", Decision Trees are considered to be one of the most popular approaches for representing classifiers. In this paper authors gives an updated survey of current methods for constructing decision tree classifiers in a top-down manner also suggested a unified algorithmic framework for presenting these algorithms and describes various splitting criteria and pruning methodologies [3].

"Image Segmentation using K-means Clustering Algorithm and Subtractive Clustering Algorithm", Nowadays image segmentation becomes one of important tool in medical area where it is used to extract or region of interest from the background. One of most used clustering algorithm is k-means clustering. It is simple and computationally faster and it can also work for large number of variable [4].

"Support Vector Machine Classification of Drunk Driving Behavior", Alcohol is the root cause of numerous traffic accidents due to its pharmacological action on the human central nervous system. This study conducted a detection process to distinguish drunk driving from normal driving under simulated driving conditions. The classification was performed by a support vector machine (SVM) classifier trained to distinguish between these two classes by integrating both driving performance and physiological measurements [5].

In this paper k-means clustering algorithm is used to group similar images into clusters using color which is one of the most widely used features for image similarity retrieval [6].

CONCLUSION

This paper summarized some of machine learning algorithms and its applications. Apart from its application, the paper also brief about difference between supervised, Unsupervised and Reinforcement algorithms. There can be many ways to enhance its features and functions. There are innumerable possibilities of using various other techniques to achieve the same results but with different parameters and results. In future author would like to explore advanced ML algorithm and its applications in detail.

REFERENCES

I. Tonge V.G., More M., Nagrale V.(2017) Introduction to Hadoop, HDFS and Map Reduce for Big Data Application. TECH-CHRONICLE an International e-journal on Emerging trends in science, Technology and Management ISSN NO: 2454-1958 Special Issue Tech-Ed 2017,pp 104-108.

II. Andreas D. Lattner, Andrea Miene, and Otthein Herzog, A Combination of Machine Learning and Image Processing Technologies for the Classification of Image Regions, A. N¨urnberger and M. Detyniecki (Eds.): AMR 2003, LNCS 3094, pp. 185–199, 2004. © Springer-Verlag Berlin Heidelberg 2004

III. Lior Rokach, Oded Maimon, Department of Industrial Engineering Tel-Aviv University liorr@eng.tau.ac.il," DECISION TREES", DATA MINING AND KNOWLEDGE DISCOVERY HANDBOOK.

IV. Nameirakpam Dhanachandra∗, Khumanthem Manglem and Yambem Jina ChanuNational Institute of Technology, Manipur 795 001, India, "Image Segmentation using K-means Clustering Algorithm and Subtractive Clustering Algorithm" 1877-0509 © 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

V. Chen h,Chen l," "Support Vector Machine Classification of Drunk Driving Behaviour",

VI. Tonge V. G.," Content Based Image Retrieval by K-Means Clustering Algorithm", (2011) International Journal of Engineering Science and Technology (IJEST)

VII. Wikipedia Online. Http://en.wikipedia.org/wiki
VIII. https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_support_machine.htm
IX. https://www.guru99.com
X. https://pythonistaplanet.com
XI. https://www.geeksforgeeks.org
XII. https://www.educba.com
XIII. Shiliang Sun, Zehui Cao, Han Zhu, and Jing Zhao," A Survey of Optimization Methods from a Machine Learning Perspective"
XIV. Tonge, V.G., and Kulkarni, October-2013. Solving Permutation Flowshop Scheduling Problem Using Improved Differential Evolutionary Algorithm. In International Journal of Engineering Research and Technology. 2(10).
XV. https://medium.com/swlh/introduction-to-machine-learning-different-types-of-machine-learning-algorithms-f710bb81f5ce
XVI. Vanita G.Tonge," Travelling Salesman Problem using Differential Evolutionary Algorithm", IOSR Journal of Engineering (IOSRJEN) www.iosrjen.org ISSN (e): 2250-3021, ISSN (p): 2278-8719 PP 63-67.
XVII. https://towardsdatascience.com
XVIII. Pravin Kshirsagar, Dr. Sudhir Akojwar, "Classification Human Motion using EEG Signals" ,International Journal of Computer Application, Volume 146-No, 7, July 2016.
XIX. Pravin Kshirsagar, Dr. Sudhir Akojwar, "Hybrid Heuristic Optimization for Benchmark Datasets", International Journal of Computer Application, Volume 146-No, 7, July 2016

Cite this Article as

Vanita Tonge Buradkar, Manisha More (2020) "**Introduction to Machine Learning and Its Applications: A Survey**" Journal of Artificial Intelligence, Machine Learning and Soft Computing, 5 (1), 1- 12

http://doi.org/10.5281/zenodo.3775092