Machine learning and its algorithms: A Research

Virendra Singh Kushwah, ArunaBajpai

Abstract—The calculations which are actualized on the machines and which are additionally used to make machines clever are called as AI calculations, likewise they can make sense of how to perform significant assignments by summing up from precedents. Therefore, AI is broadly utilized in software engineering, man-made reasoning and different fields. Be that as it may, creating fruitful AI applications requires comprehension of keen frameworks and calculations required to build it. This paper totals and condenses the kinds of AI types and calculation that are required to develop a shrewd or a specialist framework. These calculations are utilized for different purposes like information mining, picture preparing, prescient examination, and so forth to give some examples. The principle bit of leeway of utilizing AI is that, when a calculation realizes how to manage information, it can do its work consequently.

Index Terms— Machine learning, Supervised learning, Unsupervised learning, Algorithms, Reinforcement learning.

I. INTRODUCTION

Machine Learning is an idea to learn from examples and experience, without being explicitly programmed. Instead of writing code, you feed data to the generic algorithm, and it builds logic based on the data given. Machine learning methods are beginning to be used for various aspects of survey research including responsive/adaptive designs, data processing and nonresponse adjustments and weighting. Machine learning (ML) is a branch of Artificial Intelligence that pushes forward the idea that, by giving access to the right data, machines can learn by themselves how to solve a specific problem [1]. By leveraging complex mathematical and statistical tools, ML renders machines capable of performing independently intellectual tasks that have been traditionally solved by human beings. This idea of automating complex tasks has generated high interest in the networking field, on the expectation that several activities involved in the design and operation of communication networks can be offloaded to machines. Some applications of ML in different networking areas have already matched these expectations in areas such as intrusion detection [2], traffic classification [3], and cognitive radios [4].

II. NEED OF MACHINE LEARNING

Machine learning is needed for tasks that are too complex for humans to code directly. Some tasks are so complex that it is impractical, if not impossible, for humans to work out all of the nuances and code for them explicitly. So instead, we provide a large amount of data to a machine-learning algorithm and let the algorithm work it out by exploring that data and searching for a model that will achieve what the programmers have set it out to achieve.

Resurging interest in machine learning is due to the same factors that have made data mining and Bayesian analysis more popular than ever. Things like growing volumes and varieties of available data, computational processing that is cheaper and more powerful, and affordable data storage.

All of these things mean it's possible to quickly and automatically produce models that can analyze bigger, more complex data and deliver faster, more accurate results – even on a very large scale. And by building precise models, an organization has a better chance of identifying profitable opportunities – or avoiding unknown risks.

III. APPLICATIONS OF MACHINE LEARNING

Most industries working with large amounts of data have recognized the value of machine learning technology.

A. Finance Services

Banks and other businesses in the financial industry use machine-learning technology for two key purposes: to identify important insights in data, and prevent fraud. The insights can identify investment opportunities, or help investors know when to trade. Data mining can also identify clients with high-risk profiles, or use cyber surveillance to pinpoint warning signs of fraud.

B. Health Care

Machine learning is a fast-growing trend in the health care industry, thanks to the advent of wearable devices and sensors that can use data to assess a patient's health in real time. The technology can also help medical experts analyze data to identify trends or red flags that may lead to improved diagnoses and treatment.

C. Oil & Gases

Finding new energy sources.Analyzing minerals in the ground.Predicting refinery sensor failure.Streamlining oil distribution to make it more efficient and cost-effective. The number of machine learning use cases for this industry is vast – and still expanding.

D. Government

Government agencies such as public safety and utilities have a particular need for machine learning since they have multiple sources of data that can be mined for insights. Analyzing sensor data, for example, identifies ways to increase efficiency and save money. Machine learning can also help detect fraud and minimize identity theft.

E. Retails

Websites recommending items you might like based on previous purchases are using machine learning to analyze your buying history. Retailers rely on machine learning to

capture data, analyze it and use it to personalize a shopping experience, implement a marketing campaign, price optimization, merchandise supply planning, and for customer insights.

F. Transportation

Analyzing data to identify patterns and trends is key to the transportation industry, which relies on making routes more efficient and predicting potential problems to increase profitability. The data analysis and modeling aspects of machine learning are important tools to delivery companies, public transportation and other transportation organizations.

Fig.1: Applications of Machine learning

IV. CLASSIFICATIONS OF MACHINE LEARNING ALGORITHM& RESULTS

Machine Learning algorithms can be classified into:

A. Supervised Algorithms - Linear Regression, Logistic Regression, Support Vector Machine (SVM), Decision Trees, Random Forest

B. Unsupervised Algorithms - K Means Clustering.

C. Reinforcement Algorithm

A. Supervised learning algorithms

In this type of algorithm, the data set on which the machine is trained consists of labelled data or simply said, consists both the input parameters as well as the required output [5]. For example, classifying whether a person is a male or a female. Here male and female will be our labels and our training dataset will already be classified into the given labels based on certain parameters through which the machine will learn these features and patterns and classify some new input data based on the learning from this training data [6].

1. Classification algorithm

Just as the name suggests, these algorithms are used to classify data into predefined classes or labels. We will discuss one of the most used classification algorithm known as the K-Nearest Neighbor (KNN) Classification Algorithm [7].

KNN Classification Algorithm

This algorithm is used to classify a set of data points into specific groups or classes based on the similarities between the data points. Let's consider an example where we need to check whether a person is fit or not based on the height and weight of a person. Suppose we give the following table as the training data set:

Table: Data for KNN algorithm

Height (cms)	Weight (kg)	Fitness
150	50	Fit
187	75	Fit
156	80	Not Fit
163	60	Fit
170	49	Not Fit
179	70	Fit

Now consider a new person needs to be classified as fit/not fit. Let us consider the value of K=3, which means will consider 3 nearest neighbours. The nearest neighbours can be found out by determining the Euclidean difference between the height and weight of one person and the height and weight of the persons given in the table. The persons with the 3 least differences will be considered as the nearest neighbours. Now we will check how many out of these 3 are fit. If 2 or more out of the 3 are fit, then we will classify the new person as fit and vice versa. In case, we get an equal number of neighbours with different outcomes, then we can increase the value of K and check again.

2. Regression algorithm

These algorithms are used to determine the mathematical relationship between two or more variables and the level of dependency between variables. These can be used for predicting an output based on the interdependency of two or more variables. For example, an increase in the price of a product will decrease its consumption, which means, in this case, the amount of consumption will depend on the price of the product. Here, the amount of consumption will be called as the dependent variable and price of the product will be called the independent variable. The level of dependency of the amount of consumption on the price of a product will help us predict the future value of the amount of consumption based on the change in prices of the product.

We have two types of regression algorithms: Linear Regression and Logistic Regression

(1) Linear Regression

Linear regression is used with continuously valued variables, like the previous example in which the price of the product and amount of consumption are continuous variables, which means that they can have an infinite number of possible values. Linear regression can also be represented as a graph known as scatter plot, where all the data points of the dependent and independent variables are plotted and a straight line is drawn through them such that the maximum number of points will lie on the line or at a smaller distance from the line. This line - also called the regression line, will then help us determine the relationship between the dependent and independent variables along with which the linear regression equation is formed.

(2) Logistic Regression

The difference between linear and logistic regression is that logistic regression is used with categorical dependent variables (eg: Yes/No, Male/Female, Sunny/Rainy/Cloudy,

Red/Blue etc.), unlike the continuous valued variables used in linear regression. Logistic regression helps determine the probability of a certain variable to be in a certain group like whether it is night or day, or whether the colour is red or blue etc. The graph of logistic regression consists of a non-linear sigmoid function which demonstrates the probabilities of the variables.

If you wish to implement Logistic Regression in Python, then you can refer this article which includes the codes to do so. Apart from these, there are more Supervised algorithms like Support Vector Machine (SVM), Decision Trees, Random Forests.

B. Unsupervised learning algorithms

Unlike supervised learning algorithms, where we deal with labelled data for training, the training data will be unlabelled for Unsupervised Machine Learning Algorithms. The clustering of data into a specific group will be done based on the similarities between the variables.

Some of the unsupervised machine learning algorithms are K-means clustering, neural networks. In this article, we will talk about the k-means clustering algorithm.

K-means clustering

Before we understand the working of K-means clustering algorithm, let us first break down the word K-means clustering to understand what it means.

Clustering: In this algorithm, we form clusters, which are a collection of data points grouped together due to their similarities.

K refers to the number of centroids, which will be considered for a specific problem whereas 'means' refers to a centroid which is considered as the central point of any cluster.

Working of K-means Clustering Algorithm

1. Define the value of K. For eg: if K= 2, then we will have two centroids.
2. Randomly select K data points as centroids.
3. Check the distance of each data point with the centroids.
4. Assign the data point to the centroid with which it has a minimum distance, thus forming a cluster of similar data points.
5. Recalculate the centroid of each newly formed cluster and reassign the data points to the cluster whose centroid is at a minimum distance from the data point.

You can decide the number of iterations for repeating step 5 to optimize the algorithm. When the centroid stops changing or remains same after some amount of iterations then that will be our stopping point and the algorithm will be fully optimized.

Another machine-learning concept which is extensively used in the field is Neural Networks.

C. Reinforcement learning algorithms

Reinforcement Learning is a type of Machine Learning in which the machine is required to determine the ideal behavior within a specific context, in order to maximize its rewards. It works on the rewards and punishment principle which means that for any decision which a machine takes, it will be either be rewarded or punished due to which it will understand whether or not the decision was correct. This is how the machine will learn to take the correct decisions to maximize the reward in the end.

For reinforcement algorithm, a machine can be adjusted and programmed to focus more on either the long-term rewards or the short-term rewards. When the machine is in a particular state and has to be the action for the next state in order to achieve the reward, this process is called the Markov Decision Process.

V. CONCLUSION

Many current ML researches suffer from a growing detachment and divide from those real problems. Many investigators and stakeholders withdraw into their private studies with a copy of the data set and work in isolation to solidify algorithmic performance. ML provides a software the flexibility and adaptability when necessary. In spite of some application (e.g., to write matrix multiplication programs) where ML may fail to be beneficial, with increase of data resources and increasing demand in personalized customizable software, ML will thrive in near future. Besides software development, ML will probably but help reform the general outlook of Computer Science.

REFERENCES

1. S. Marsland, Machine learning: an algorithmic perspective. CRC press, 2015.
2. A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 1153–1176, Oct. 2015.
3. T. T. Nguyen and G. Armitage, "A survey of techniques for internet traffic classification using machine learning," IEEE Communications Surveys & Tutorials, vol. 10, no. 4, pp. 56–76, 4th Q 2008.
4. M. Bkassiny, Y. Li, and S. K. Jayaweera, "A survey on machine learning techniques in cognitive radios," IEEE Communications Surveys & Tutorials, vol. 15, no. 3, pp. 1136–1159, Oct. 2012
5. Xu, X., Yang, G. Robust manifold classification based on semi supervised learning (2013) International Journal of Advancements in Computing Technology, 5 (8), pp. 174-183. DOI: 10.4156/ijact.vol5.issue6.21
6. Alajlan, N., Bazi, Y., Melgani, F., Yager, R.R. Fusion of supervised and unsupervised learning for improved classification of hyperspectral images (2012) Information Sciences, 217, pp. 39-55. DOI: 10.1016/j.ins.2012.06.031
7. Silva, J.D.A., Hruschka, E.R. An experimental study on the use of nearest neighbor based imputation algorithms for classification tasks (2013) Data and Knowledge Engineering, 84, pp. 47-58. DOI: 10.1016/j.datak.2012.12.006