Brief study on machine learning

Article · October 2020

CITATIONS
0

READS
2,166

1 author:

Simran Gurung
Nepal College of Information Technology
4 PUBLICATIONS **2** CITATIONS

SEE PROFILE

Brief study on machine learning

Simran Gurung
Nepal college of Information and technology (NCIT), Balkumari, Lalitpur, Nepal
simran.192825@ncit.edu.np

***Abstract*— Machine learning is the fastest growing areas of software engineering. Machine learning is the major success factor in the progressing computerized change across businesses. Startups and behemoths alike announce new products that will figure out how to perform errands that beforehand no one but people could do, and play out those assignments better, quicker, and more intelligently. It has the ability to lets the computer to create a program on it's on. It is a subset of Artificial Intelligence (AI), and comprises of the further developed techniques and models that empower the computers to make sense of things out from the data and deliver.**

It is a field of learning and extensively divided into supervised learning, unsupervised learning, and reinforcement learning. There are numerous fields where the Machine learning algorithms are utilized. This article will investigate the use of machine learning and its methodologies with its applications in the various fields from published papers, workshop materials & material collected from books and material gathered from books and material accessible online on the internet.

Additionally, the field of deep learning and ANN will be explained and discussed.

***Index Terms*— Machine Learning, Optimization, Supervised, Unsupervised, Reinforcement, Clustering. Big Data, Deep Learning, Neural Networks, Artificial Neural Network Intelligence.**

Table of Acronyms

ACRONYM	DEFINITION
ML	Machine Learning
MLlib	Machine Learning Library
AI	Artificial Intelligence
SciPy	Scientific Python
NumPy	Numerical Python
TSP	Travelling Salesman problem

I. INTRODUCTION

Machine learning is the subfield of computer science that, according to Arthur Samuel in 1959, gives "computers the capability to learn without being explicitly programmed." [1] Huge amount of data is produced every day from numerous sources called as big data. Big Data is a collection of enormous datasets of aggregate volume, velocity and variety such large amount of data is difficult to manage and process. [2] To deal with such an enormous load of data methods like Hadoop, data Mining, Machine learning etc. are used. Machine learning is various and energizing field to deal with big data and there are numerous approaches to define it. It transforms data into program and automate the automation system. It tells how to perform task with past experience from provided examples.

Optimization is one of the core components of machine learning and artificial Intelligence. [3] [4] The essence of most machine learning algorithms is to construct an optimization model and learn the parameters in the objective function from the given data. Optimization algorithm for instance: Genetic algorithm, Differential algorithm, ant colony optimization, Particle Swarm Optimization, [3] [4] and so on have great impact on different fields of machine learning (ML). In machine learning (ML), genetic algorithms were utilized during the 1980s and 1990s. On the other hand, machine learning techniques have been utilized to improve the performance of genetic and evolutionary algorithms. [5]

The Travelling Salesman problem (TSP) or Hamiltonian tour is an exemplary problem and one of the benchmark in software engineering and Operations Research. Heuristic optimization algorithms for example, genetic algorithm, tabu search and ant colony optimization have been broadly used to discover ideal solution in travelling salesman problem. [5] [6]

II. STEPS OF MACHINE LEARNING ALGORITHM

• **Problem Framing:** frame a machine learning problem in terms of what we need to foresee and what sort of observation data we have to make those predictions.
• **Gathering data:** Input the data. That data can be structured or unstructured.
• **Data Preparation:** Loading data into suitable place and set it up for machine learning training.
• **Choosing Model:** This is significant step to pick a legitimate model to implement and predict the output. There are numerous models that are made by specialists and data scientist throughout the long term. Some are very well appropriate for image data and some are for numeric data.
• **Training data:** Data gradually improves the model's ability to foresee the output.
• **Evaluation**: Evaluation permits 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 solution is obtained. [5]

III. DOMAINS MACHINE LEARNING IS USED FOR

Machine learning is used in different domains. Here are a few examples:

1. Security heuristics: that distill attack patterns to protect, for instance, ports or networks;

2. Image analysis: To detect distinctive forms and shapes, Like that for medical investigations or face and fingerprint recognition;

3. Deep learning: To generate rules for data analytics and huge data handing, likely to be used in marketing and sales promotions;

4. Object recognition and predictions: From collective video streams, data and provided data driven intuitions, decisions, and predictions multisensory fusion for autonomous driving;

5. Pattern recognition: To examine code for weaknesses for example, criticality and code smells (for a related case study, see the sidebar). [7]

IV. MACHINE LEARNING ALGORITHMS

Machine learning employs the following approaches supervised learning, unsupervised learning, semi-Supervised learning and reinforcement learning.

1) Supervised learning

The supervised learning comprises classification algorithms, which take as input a dataset and the class of each piece of data so that the computer can learn how to classify new data.

Classification can utilize logic regression, classification trees, support vector machines, random forests, artificial neural networks (ANNs), or other algorithms.

Supervised learning is straightforward and easy to understand. It depends on earlier data basic like a mother instructs to her kid. Given data are labeled data for instance weather data, time, holidays, route select, house cost and so on learning algorithm used these labeled data and attempt to estimate the given labeled data with the input and trained the machine.

Once the machine is trained it begins to predict and give decision when new data is given to it. Regression algorithms foresee an estimation of an entity's attribute ("regression" here has an extensive sense than only statistical regression). Regression algorithms include linear regression, decision trees, Bayesian networks, fuzzy classification, and ANNs. Here are few types of regression techniques. [5] [7]

1. Decision tree:

Decision tree algorithm is simple supervised machine learning algorithm. It comprises a root node, branches, and leaf nodes. Each internal node signifies a test on an attribute, each branch denotes the result of a test, and each leaf node holds a class label. [5] [8]

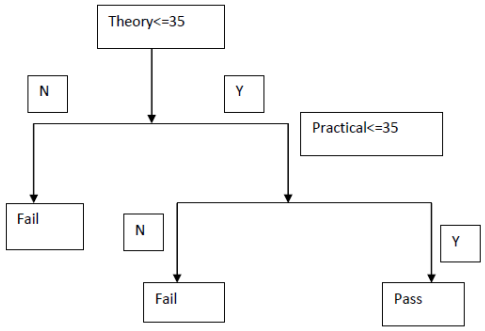

Figure 1: example of decision tree

2. Random forest:

Random forests are the most adaptable and simple to utilize supervised learning algorithm. It can be utilized both for classification and regression. However, it is generally utilized for the classification. A forest consist of trees.

Random forests have a range of uses, for example, recommendation engines, image classification and feature selection. [5] [8]

3. Logistic Regression:

Logistic regression method used in estimation of the probability of the target value. Target value is discrete value which means it is in binary form having data coded in 1 for progress/yes and 0 for disappointment/no. [8]

4. Naïve Bayes:

Naïve Bayes is one of the probabilistic machine learning (ML) algorithms based on Bayes theorem. It is beneficial for text data. Some of the application of this method includes filtering spam, classifying documents, sentiment prediction etc. [5]

5. Support vector machine:

Support vector machines (SVMs) are a set of related supervised learning methods utilized for classification and regression. However, they are generally used in classification problems. [3] SVMs have their distinctive mode of implementation in contrast to other machine learning (ML) algorithms. L Recently, they are tremendously popular because of their ability to deal with multiple continuous and categorical variables. [5]

Challenges in Supervised machine learning

Here, are difficulties faced in supervised machine learning:

- Irrelevant input feature present training data could give mistaken outcomes.
- Data preparation and pre-processing is consistently a test.
- Accuracy suffers when unmanageable, doubtful, and inadequate values have been inputted as training data.
- If the concerned expert is not existing, then the other method is "brute-force." It implies you need to think that the right features (input variables) to train the machine on. It can be imprecise. [9]

Advantages of Supervised Learning

- Supervised learning permits you to gather information or produces a data output from the past experience.
- Helps you to enhance performance models utilizing experience.
- Supervised learning encourages you to solve different kinds of real-world computation issues.
- The main advantage of supervised learning algorithm is clearness of information and simplicity of training.
- Supervised learning can be extremely useful in classification problems. [9]

Disadvantages of Supervised Learning

- Supervised learning is restricted in a sense that it can't deal with a portion of the complex assignments in machine learning.
- Supervised learning cannot give you obscure information from the training data like unsupervised learning do.
- It cannot cluster or classify data by learning its features on its own, unlike unsupervised learning.
- As well as many shortcomings such as the incapability to learn by itself.
- Classifying vast data can be a real challenge.
- Training for supervised learning needs a great deal of calculation time. [9]

2) Unsupervised Learning Algorithms

It comprises of clustering algorithms, which take as input a dataset covering different dimensions and segment it into groups fulfilling certain standards.

Unlike that of supervised learning rather than utilizing label data unsupervised learning is taken care of with data and a tool to comprehend the properties of the data. Unsupervised learning algorithm encourages you to discover a wide range of pattern in data. Output of the unsupervised learning is group or cluster of data having alike features. [7]

Types of this algorithm are: clustering and association.

1. Clustering:

It basically manages finding a structure or pattern in a collection of uncategorized data. [10]

Following are some clustering algorithms.

- K-means- K-means is the clustering algorithm utilized to decide the natural spectral grouping present in a data set. Data will be segregated into k clusters, in view of their features. Each cluster is signified by its centroid, defined as the focus of the points in the cluster. K-Means is basic and quick however it doesn't respect a similar result with each run. Examples: Single value decomposition, principal Component Analysis. [10]
- Association: Association rules permit you to build up relationship among data objects inside large data sets. This unsupervised technique is about learning interesting connections between variables in large datasets.

3) Semi-supervised Learning Algorithm

Semi-supervised learning falls between unsupervised learning and supervised learning. Many machine-learning researchers have discovered that unlabeled data, when used in combination with a small amount of labeled data, can create an impressive improvement in learning precision. It utilizes both labeled and unlabeled data for training – normally a lesser amount of labeled data with a larger amount of unlabeled data (because unlabeled data is less expensive and takes less effort to acquire). This type of learning can be utilized with methods, for example: classification, regression and prediction. Semi-supervised learning is valuable when the expense related with labeling is too high to even consider allowing for a fully labeled training process. Early instances of this include distinguishing an individual's face on a web cam.

- Reinforcement Learning Algorithm

Reinforcement algorithm is unlike than that of supervised and unsupervised algorithm. Supervised algorithm trains the data with answer key however reinforcement learning algorithm trains data without correct answer key. Reinforcement learning can be understood with the concepts of agents, environments, states, actions and rewards. The reinforcement learning agent chooses what to do so as to play out the given task. In lack of training data set reinforcement learning agent utilizes the

experience. Reinforcement learning varies from other types of learning methods, because the system isn't trained with the sample data set. Or maybe, the system learns through trial and error. In this way, a grouping of fruitful choices will bring in the process being reinforced, because it takes care of the current issue. [3] [5]

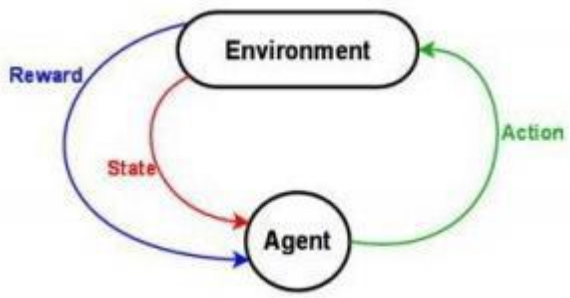

Figure 2: Reinforcement Learning Algorithm

V. ESSENTIAL TOOLS

Machine learning's popularity has brought along an abundance of tools. The vast majority of them are open source, so users can easily experiment with them and learn how to utilize them.

For a field as wide as machine learning, no single tool will do. The best a software engineer can do is to get familiar with a wide range of tools and realize which one is the most fitting for a given circumstance.

All things considered, R is more mainstream with individuals with a somewhat stronger statistical background. It has a superb collection of machine-learning and statistical-inference libraries. The numerical and statistical communities are separated into 2 camps: one that that lean towards R and one that favors Python. Python is more popular with people with a computer science background.

Python has broad libraries for numerical computing (NumPy), scientific computing (SciPy), statistics (StatsModels), and machine learning (scikit-learn). These are largely wrappers of C code, so you get Python's convenience with C's speed. In spite of the fact that there are less machine-learning libraries for Python than there are for R, numerous developers discover working with Python simpler. They may definitely know the language or think that it's simpler to learn than R. They additionally discover Python helpful for preprocessing data: reading it from different sources, cleaning it, and bringing it to the required formats.

For visualization, Python depends on matplotlib. You can do pretty much everything on matplotlib, however you may find you need to invest some effort. In general, R and Python work when the dataset fits in the computer's main memory. If that's not possible, you must use a distributed platform. The most well-known is Hadoop, yet Hadoop isn't the most advantageous for machine learning. Spark has a machine-learning library that executes key algorithms, so for many reasons you don't need to implement anything yourself.

A promising development is the Julia programming language for technical computing, which aims at top performance. Since Julia is new, it doesn't have close to the same number of libraries as Python or R. However, because of its noteworthy speed, its popularity might grow. Strong commercial players include Matlab and SAS, which both have a distinguished history. Matlab has since quite a while ago offered strong tools for numerical computation, to which it has added machine-learning algorithms and implementations. For engineers acquainted with Matlab, it might be a natural fit. SAS is a software suite for advanced statistical analysis; it also has added machine-learning capabilities and is widespread for business intelligence tasks. [7]

VI. APPLICATION OF ML

Machine learning algorithms are utilized in wide territory of exploration, for example:

1. Financial Services: Machine learning acquires prominence in monetary administrations such bank to forestall extortion and also to find patterns in data.
2. Sales and Marketing: Huge information is created from marketing sites. With the boom of data, marketing department relies on machine learning to optimize the relationship between customer and product campaign.
3. Government: The government administration can use ML to manage public safety and utilities.
4. Healthcare: Machine learning is playing a significant part in medical care. Today, machine learning is assisting with smoothing out regulatory cycles in emergency clinics, guide and treat infectious diseases and personalize medical treatments.
5. Transportation: Based on the travel history and pattern of traveling across various routes, machine learning help transportation organizations foresee potential issues that could emerge on specific courses, and in like manner encourage their clients to decide on an alternate course. [5]

VII. ANNS AND DEEP LEARNING

The Concept of Artificial neural networks are inspired by the biological model proposed by Nobel laureates David H. Hubel & Torsten Wiesel in 1959.

ANNs' fundamental concept have been around for about 50 years. Though, it's now that we can architect them in new ways. ANNs can be utilized across a range of machine learning: classification, regression, clustering, and dimensionality reduction.

Innovations in ANN architectures and the accessibility of modest computing resources to run ANNs has brought about the expansion of deep learning concepts—utilizing big ANNs to perform machine learning. Throughout the most recent couple of years, deep learning has chalked up headline-grabbing successes by beating humans in Jeopardy! Also Go, learning to play arcade games, showing an uncanny capability to recognize images, performing automatic translation, and so on.

Deep learning is especially good at common tasks requiring the elicitation of more significant level, theoretical ideas from

the input data, which is what the numerous layers of an ANN excel at.

As of late, Google delivered an open source TensorFlow library for working with ANNs. You can connect with TensorFlow through a Python API. A C++ API is likewise accessible; despite the fact that not as simple to use, it may give some performance benefits.

The use of the expression "Deep Learning" in the context of Artificial Neural Networks was introduced by Igor Aizenberg and colleagues in 2000. [11] Deep learning is the study of artificial neural networks and related machine learning algorithms that contain more than one hidden layer. [12] Deep learning is part of a larger family of machine learning methods based on learning representations of data. An observation (e.g., a picture) can be represented in multiple points of view, for example, a vector intensity values per pixel, or in a more theoretical path as a lot of edges, areas of specific shape, and so forth. Some representations are better than others at improving the learning task (e.g., face recognition or expression recognition). One of the guarantees of deep learning is replacing handcrafted features with efficient algorithms for unsupervised or semi-supervised learning and hierarchical feature extraction. [13]

Although Deep learning has been described as a trendy expression, or a rebranding of neural networks, deep neural nets have shown an ability to out-perform other machine learning algorithms on tasks for example, object recognition in the field of computer vision.

VIII. CONCLUSIONS

This paper summed up some of machine learning (ML) algorithms and its applications. Aside from its application, the paper additionally concise about distinction between supervised, Unsupervised and Reinforcement algorithms. [5] Artificial Intelligence, machine learning and deep learning are fundamentally machine perception. It is the ability to decipher sensory data. Two primary ways we decipher things are by naming what we sense; for example we hear a sound as we state ourselves "That's my daughter's voice." Or we see a haze of photons and we state "That's my mother's face." If we don't have names for things, we can still identify resemblances and dissimilarities. You may see two faces and realize that they were mother and little girl, without knowing their names; or you may hear two voices and realize that they originated from a similar town or state by their intonation. Algorithms train to name things through supervised learning, and to bunch things through unsupervised learning. The dissimilarity between supervised and unsupervised learning is whether there is a labeled training set to work with or not. The labels you apply to data are just the results you care about. Perhaps you care about distinguishing irate or nasty messages, which are on the whole simply unstructured masses of text. Or maybe it is about identifying individuals in images. Or you're taking a gander at time arrangement information - a flood of numbers - and you care about whether the following occasions in the time arrangement will be sequential. So by applying deep learning, working with other algorithms, we can classify, cluster and predict information. It does as such by figuring out how to peruse the signs, or structure, in data automatically. At the point when deep learning algorithms train, they make deductions about the data, measure the error of those deductions against the training set, and afterwards they correct the way they make prediction in order to become more precise. That is optimization. [14]

Now visualize that, with deep learning, we can categorize, cluster or predict anything you have data about: images, sound, video, text and DNA, time series (touch, financial exchanges, monetary tables, and the climate). That is, whatever people can detect and that our technology can digitize. You have increased your capacity to break down what's going in the world by many times. With deep learning, we are fundamentally enabling society to carry on significantly more wisely, by precisely deciphering what's going on in our general surroundings with software.

Prediction alone is a tremendous power, and the applications are genuinely self-evident. Classification sounds dull, however by naming something, you can conclude how to react. In an event that an email is spam, you send it to the spam envelope and spare the reader time. In an event that the face captured by your entryway door camera is your mother, possibly you advise the smart lock to open the door. In an event an X-ray shows a tumorous pattern, you flag it for deeper assessment by clinical specialists. [14]

There are innumerable prospects of utilizing different strategies to achieve similar outcomes however with different parameters and results. In future I would like to discover further advanced ML algorithm and its applications in detail.

This work has therefore achieved its last goal by furnishing the academic network with likely bearings for future work and will ideally fill in as preparation for extraordinary enhancements in the field of machine learning with Big Data.

XI. REFERENCES

[1] Munoz and Andres, "Machine Learning and Optimization," [Online]. Available: https://www.cims.nyu.edu/~munoz/files/ml_optimization.pdf. [Accessed 1 June 2017].

[2] T. V.G., M. M. and N. V, "Introduction to Hadoop, HDFS and Map Reduce for Big Data Application," *TECH-CHRONICLE an International e-journal on Emerging trends in science*, vol. Technology and Management , no. 2454-1958 Special Issue Tech-Ed 2017, 2017.

[3] P. Kshirsagar and D. S. Akojwar, "Classification Human Motion using EEG Signals," *International Journal of Computer Application*, Vols. 146-No, 7, 2016.

[4] P. Kshirsagar and D. S. Akojwar, "Hybrid Heuristic Optimization for Benchmark Datasets," *International Journal of Computer Application*, Vols. 146-No, 7, 2016.

[5] M. M. Vanita Tonge Buradkar, "Introduction to Machine Learning and Its Applications: A Survey," *Journal of Artificial*

Intelligence, Machine Learning and Soft Computing, vol. 5, no. 1, p. 12, 2020.

[6] V. G.Tonge, "Travelling Salesman Problem using Differential Evolutionary Algorithm," *IOSR Journal of Engineering (IOSRJEN)*, no. (e): 2250-3021, ISSN (p): 2278-8719 PP 63-67..

[7] P. Louridas and C. Ebert, Machine Learning, Paris: THE IEEE COMPUTER SOCIETY, 2016.

[8] A. Joy, "Supervised Learning Algorithms: An Illustrated Guide," Pythonista Planet, [Online]. Available: https://www.pythonistaplanet.com/supervised-learning-algorithms/. [Accessed 10 septemeber 2020].

[9] guru99, "Supervised Machine Learning: What is, Algorithms, Example," guru99, [Online]. Available: https://www.guru99.com/supervised-machine-learning.html. [Accessed 10 September 2020].

[10] A. Joy, "Real-World Applications Of Unsupervised Learning," Pythonista Planet, [Online]. Available: https://www.pythonistaplanet.com/applications-of-unsupervised-learning/. [Accessed 20 September 2020].

[11] I. Aizenberg, N. N. Aizenberg and J. P. Vandewalle, Multi-Valued and Universal Binary Neurons, Springer US, 2000.

[12] L. Deng and D. Yu, "Deep Learning: Methods and Applications," 2014. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.691.3679&rep=rep1&type=pdf. [Accessed 10 Novemeber 2020].

[13] H. A. Song and S.-Y. Lee, Hierarchical Representation Using NMF, Vols. Lecture Notes in Computer Science, vol 8226, KAIST, Daejeon 305-701, Republic of Korea: Springer-Verlag Berlin Heidelberg, 2013.

[14] P. Ongsulee, "Artificial Intelligence, Machine Learning and Deep Learning," in *2017 Fifteenth International Conference on ICT and Knowledge Engineering*, Thailand, 2017.