International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(REVIEW ARTICLE)

Check for updates

A review on machine learning

Shaveta *

Assistant Professor, DCSA, Guru Nanak College, Ferozepur Cantt, Punjab, India.

International Journal of Science and Research Archive, 2023, 09(01), 281–285

Publication history: Received on 18 April 2023; revised on 26 May 2023; accepted on 28 May 2023

Article DOI: https://doi.org/10.30574/ijsra.2023.9.1.0410

Abstract

Machine learning is a particular branch of artificial intelligence that teaches a machine how to learn, whereas artificial intelligence (AI) is the general science that aims to emulate human abilities. An AI method called machine learning teaches computers to learn from their past experiences. Machine learning algorithms don't rely on a predetermined equation as a model, but instead "learn" information directly from data using computational techniques. As the quantity of learning examples increases, the algorithms adaptively get better at what they do. This paper provides an overview of the field as well as a variety of machine learning approaches, including supervised, unsupervised, and reinforcement learning and various languages used for machine learning.

Keywords: Machine learning; learning; Supervised learning; Artificial intelligence; Languages

1. Introduction

A developing technology called machine learning makes it possible for computers to learn autonomously from historical data. Machine learning uses a variety of techniques to create mathematical models and make predictions based on previous information or data. Currently, it is utilised for many different things, including recommender systems, email filtering, Facebook auto-tagging, image identification, and speech recognition [1].

Modern machine learning differs from machine learning in the past due to new computing technologies. Researchers interested in artificial intelligence wanted to discover if computers could learn from data; pattern recognition and the idea that they can learn without being programmed to do certain tasks gave rise to it. Because models may independently adjust as they are exposed to fresh data, machine learning's iterative component is crucial. In order to provide trustworthy, reproducible decisions and results, they learn from earlier calculations. It's an old science, but it has recently garnered new momentum [2]. A subset of artificial intelligence known as "machine learning" is stated to focus on creating algorithms that enable computers to independently learn from data and past experiences. In 1959, Arthur Samuel coined the phrase "machine learning."

Machine learning algorithms create a mathematical model with the aid of historical sample data, or "training data," that aids in making predictions or judgements without being explicitly programmed. Computer science and statistics are used with machine learning to create prediction models. Algorithms that learn from past data are created by machine learning or used in it [3]. The performance will be higher the more information we supply.

Whenever new data is received, a machine learning system predicts the outcome for it by learning from historical data, creating prediction models, and learning from past data. The volume of data used to develop the model influences how well it predicts the output, as a larger volume of data allows for more accurate model construction.

* Corresponding author: Shaveta

Instead of coding code specifically for a difficult scenario where we need to make predictions, let's say we merely need to give the data to generic algorithms, which then use the data to build logic and anticipate the results. Machine learning has altered the way we approach the issue [4].

Machine learning is becoming increasingly necessary as time goes on. Because machine learning can perform activities that are too complex for a person to carry out directly, it is necessary. Since we cannot access the vast quantity of data manually as humans, we need computer systems, and here is where machine learning comes in to simplify our lives.

By giving machine learning algorithms a vast amount of data, we can train them to examine the data, build models, and anticipate the desired output automatically. The cost function can be used to estimate how well the machine learning algorithm performs given the amount of data. We can save time and money by using machine learning. Its use cases make it simple to understand the significance of machine learning. Machine learning is currently employed in self-driving cars, cyber fraud detection, face recognition, and Facebook friend suggestion, among other applications [5]. Several leading businesses, including Netflix and Amazon, have developed machine learning algorithms that make use of a tremendous quantity of data to analyse customer interest and make product recommendations.

Because it can solve issues at a speed and scale that cannot be matched by the human mind alone, ML has been shown to be useful. Machines can be trained to recognise patterns in and relationships between incoming data by putting large amounts of processing power behind a single activity or a number of focused tasks. This allows machines to automate repetitive tasks.

Machine learning can be broadly divided into three categories:

- Supervised learning
- Unsupervised learning
- Reinforcement in learning

2. Supervised learning

One form of machine learning technique is supervised learning, in which we train the machine learning system using sample labelled data and then watch as it predicts the outcome.

In order to anticipate future events, supervised machine learning algorithms use labelled examples to apply what they have learned in the past to fresh data. The learning method creates an inferred function to forecast output values by examining a known training dataset [6]. After adequate training, the system may provide targets for any new input. In order to identify flaws and correct the model as necessary, it can also compare its output with that which is proper and intended.

In order to comprehend the datasets and learn about each one, the system builds a model using labelled data. After training and processing, the model is tested by being given a sample set of data to see if it predicts the output accurately or not [7].

The mapping of input and output data is the aim of supervised learning. A pupil learning under the teacher's supervision is the same as supervised learning because it is based on supervision. Spam filtering serves as an illustration of supervised learning. Supervised learning can be grouped further in two categories of algorithms:

2.1. Classification

When the output variable is categorical, meaning there are two classes, such as Yes-No, Male-Female, True-False, spam detection, etc., classification methods are used. Several Classification algorithms are: Random Forest, Decision Trees, Logistic Regression, Support Vector Machines

2.2. Regression

Regression procedures are applied if there is a correlation between the input and output variables. It is used to forecast continuous variables like weather, market trends, and other things. Several well-liked regression algorithms that fall under supervised learning are listed below: Linear Regression, Regression Trees, Non-Linear Regression, Bayesian Linear Regression, Polynomial Regression

3. Unsupervised learning

Unsupervised learning is a type of learning where a computer picks up information without any human intervention. The machine is trained using a set of unlabeled, unclassified, or uncategorized data, and the algorithm is required to respond independently to that data [8]. Unsupervised learning's objective is to reorganize the incoming data into fresh features or a collection of objects with related patterns. There is no predefined outcome in unsupervised learning.

When training data is neither categorised nor labelled, unsupervised machine learning techniques are utilised. Unsupervised learning investigates how systems might extrapolate a function from unlabeled data to describe a hidden structure. The system can never be guaranteed that the output is correct. Instead, it infers what the result should be based on datasets. The machine searches through the vast volume of data for helpful insights. It can also be divided into two types of algorithms:

3.1. Clustering

Using the clustering technique, items are grouped into clusters so that those who share the most similarities stay in one group and share little to none with those in another. The data objects are classified based on the existence or lack of commonalities discovered by cluster analysis.

3.2. Association

An unsupervised learning technique called an association rule is used to uncover the connections among the variables in a sizable database. It establishes the group of items that co-occur in the collection. Marketing strategy is more effective because to the association rule. People who buy X (let's say, bread) also frequently buy Y (let's say, butter or jam). Market Basket Analysis is an illustration of an association rule in action.

4. Reinforcement learning

A learning agent in a reinforcement learning system receives a reward for each correct action and receives a penalty for each incorrect activity. With the help of this feedback, the agent automatically learns and performs better [9]. The agent explores and engages with the environment during reinforcement learning. An agent performs better since its objective is to accrue the most reward points.

Algorithms for reinforcement learning interact with their surroundings by taking actions and identifying successes or failures. Trial-and-error learning and delayed rewards are two of reinforcement learning's most important features. With the help of this technique, machines and software agents may automatically select the best course of action in a given situation to enhance performance. The reinforcement signal, which is straightforward reward feedback, is necessary for the agent to figure out which action is better [10]. Reinforcement learning is demonstrated by the robotic dog, which automatically learns how to move its arms [11].

5. Languages used for Machine Learning

5.1. Python

A fantastic programming language is Python. Although it is easy to read, it is also capable of a wide range of tasks. Unlike other languages, it enables quick iteration. If you only need to alter one or two lines of your code, you can immediately run the changed version. For minor modifications, there's no need to update numerous locations. No need to wait minutes or hours for your code to be recompiled before running it. Python has one of the most straightforward syntaxes and one of the most natural languages. Learning, reading, and correcting errors are therefore simple. Python is an open-source language, which is the best part [12]. Hence, both accessing and distributing it are free. It consequently features flexible libraries. The burgeoning fields of data science, AI, and machine learning can benefit from this language. Python is anticipated to continue to be in high demand in the future because of this use. It is often used in web-based applications as well. Additionally, you can use it to create desktop and mobile applications [13].

A library is a group of programs or modules created to carry out particular tasks. Among other things, some modules can assist with the graphical user interface, data analysis, and graphics. You can import the code from the library rather than having to write it each time you need it.

R is a statistics-related open-source programming language. This one is the finest substitute for conventional languages like Stata, SAS, and SPSS. R uses graphics to assist in the visualisation of statistical data. It is utilised by big businesses like Facebook and Google [14]. Being open-source software, it is available for free download and usage.

Data mining, time series analysis, regression analysis, stock-market modelling, risk assessment, and simulation are a few applications of R. R is also useful for businesses who need to gather and analyse client data. These include financial services, social networking, e-commerce, and the healthcare industry. R has a significant advantage over Python when it comes to advanced statistics. Pythons is a multipurpose programming language that uses a broad statistical methodology. R, however, can perform in-depth statistical analysis. R, however, is a lot more difficult to learn than Python.

5.2. Java and JavaScript

Java is becoming more popular among machine learning engineers who have backgrounds in Java development, despite Python and R still being the preferred programming languages of machine learning enthusiasts. This is because these engineers don't need to learn a new programming language like Python or R in order to implement machine learning. Huge Java code bases already exist in many organisations, and Java is the primary programming language used in the majority of open-source big data processing tools like Hadoop and Spark. It is simpler for machine learning engineers to integrate with existing code repositories when using Java for machine learning applications. Its ease of use, package services, improved user interaction, simple debugging, and graphical data representation make it a preferred machine-learning language.

5.3. Julia

A possible rival to Python and R, Julia is a high-performance, general-purpose dynamic programming language that has many key features designed specifically for machine learning. Although it is a general-purpose programming language and can be used to create various applications, it performs best when used for computational science and high-performance numerical analysis. Large organisations like Apple, Disney, Oracle, and NASA are using Julia to power their machine learning applications because it supports all forms of hardware, including TPUs and GPUs on every cloud.

5.4. Lisp

As it adapts to the answer a programmer is coding for, LISP is regarded as the most effective and flexible machine-learning language for tackling specifics. LISP differs from other machine-learning languages due to this. These days, machine learning and inductive logic problems are their main applications. LISP was used to produce the first AI chatbot, ELIZA, and machine learning experts can still use it to build chatbots for eCommerce. Given that developers still use LISP for artificial intelligence projects that heavily rely on machine learning, it is deserving of a spot on the list of the best languages for machine learning. Although LISP is adaptable for machine learning, it does not have the backing of well-known machine learning libraries [15]. Because it is challenging to learn, LISP is neither a machine learning language for beginners nor does it have a sizable user base like Python or R do.

6. Future scope

Machine learning is still essential to the success of AI despite all of its flaws. A different approach to AI that addresses its flaws, such as the "black box" problem that arises when robots learn unsupervised, is necessary for this achievement, though. That strategy is symbolic AI, or a rule-based approach to data processing. A knowledge graph, an open box used in the symbolic approach, is used to define concepts and semantic relationships.

Hybrid AI, which combines ML and symbolic AI, enables AI to comprehend language as well as data. This effective method is changing how data is used throughout the company by providing a deeper understanding of what was discovered and why.

7. Conclusion

In conclusion, machine learning has revolutionized various industries and continues to shape the future of technology. Its ability to analyze vast amounts of data and extract meaningful patterns has led to advancements in areas such as healthcare, finance, transportation, and more. Machine learning algorithms have demonstrated remarkable performance in tasks such as image and speech recognition, natural language processing, and predictive modeling. The benefits of machine learning are evident, as it enables businesses to make data-driven decisions, automate processes,

and optimize operations. By leveraging machine learning techniques, organizations can uncover valuable insights, improve efficiency, and enhance customer experiences.

Compliance with ethical standards

Acknowledgments

I would like to express my heartfelt gratitude to all those who contributed to the completion of this research paper titled "A Review on Machine Learning." I would like to thank our colleagues and peers for their stimulating discussions and valuable insights, which significantly enriched the content and scope of this review. Their inputs have been instrumental in broadening our understanding of the subject matter and refining our analysis.

References

[1] Carbonell, Jaime G., Ryszard S. Michalski, and Tom M. Mitchell. "An overview of machine learning." Machine learning (1983): 3-23.

[2] Alzubi, Jafar, Anand Nayyar, and Akshi Kumar. "Machine learning from theory to algorithms: an overview." Journal of physics: conference series. Vol. 1142. No. 1. IOP Publishing, 2018.

[3] Maleki, Farhad, et al. "Overview of machine learning part 1: fundamentals and classic approaches." Neuroimaging Clinics 30.4 (2020): e17-e32.

[4] Ayodele, Taiwo Oladipupo. "Machine learning overview." New Advances in Machine Learning 2 (2010): 9-18.

[5] Jiang, Tammy, Jaimie L. Gradus, and Anthony J. Rosellini. "Supervised machine learning: a brief primer." Behavior Therapy 51.5 (2020): 675-687.

[6] Löning, Markus, et al. "sktime: A unified interface for machine learning with time series." arXiv preprint arXiv:1909.07872 (2019).

[7] Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. Informatica,31, 249–268.

[8] Alpaydin, E. (2010). Introduction to machine learning (2nd ed.). Cambridge, MA: MIT Press.

[9] Bishop, C. M. (2006). Pattern recognition and machine learning. New York, NY: Springer.

[10] Dietterich, T. G. (2000). Ensemble methods in machine learning. Proceedings of Multiple Classifier Systems: First International Workshop (MCS) (pp. 1–15). Berlin: Springer. doi:http://dx.doi.org/10.1007/3-540-45014-9_1

[11] Singh, Amanpreet, Narina Thakur, and Aakanksha Sharma. "A review of supervised machine learning algorithms." 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom). Ieee, 2016.

[12] Burkart, Nadia, and Marco F. Huber. "A survey on the explainability of supervised machine learning." Journal of Artificial Intelligence Research 70 (2021): 245-317.

[13] Jiang, Tammy, Jaimie L. Gradus, and Anthony J. Rosellini. "Supervised machine learning: a brief primer." Behavior Therapy 51.5 (2020): 675-687.

[14] Sen, Pratap Chandra, Mahimarnab Hajra, and Mitadru Ghosh. "Supervised classification algorithms in machine learning: A survey and review." Emerging technology in modelling and graphics. Springer, Singapore, 2020. 99-111.

[15] Mahesh, Batta. "Machine learning algorithms-a review." International Journal of Science and Research (IJSR).[Internet] 9 (2020): 381-386