A Review of Machine Learning Algorithm

[1]Dr. A. Sahaya Arthy, [2]Mrs. J.F. Jude Hema
[1,2]Assistant Professor, PG Department of Computer Science
A.P.C. Mahalaxmi College for Women, Thoothukudi

Abstract

Machine learning (ML) is the scientific study of algorithms and statistical models that systems of computers employ to carry out a specific task without being explicitly programmed. Many applications that we use every day use learning algorithms. If a web search engine like Google is used to search the internet, one of the reasons it works so well is because of a learning algorithm that has learned how to rank web pages. These algorithms are used for a variety of purposes including data mining, image processing, predictive analytics, and so on. The main advantage of using machine learning is that once an algorithm learns what to do with data, it can do so automatically. This paper provides a brief overview and outlook on the numerous applications of machine learning algorithms.

Keywords: Machine Learning, Knowledge Acquisition, Classification

1. Introduction

Humans have used a variety of tools to perform various tasks in a more efficient manner since their evolution. The human brain's creativity resulted in the development of various machines. These machines facilitated human life by allowing people to meet a variety of needs, such as travel, industry, and computing. And one of them is machine learning.

Machine learning, as defined by Arthur Samuel, is the field of study that gives computers the ability to learn without being explicitly programmed. Arthur Samuel was well-known for his checkers programme. Machine learning (ML) is a technique for teaching machines how to handle data more efficiently. Sometimes, after viewing the data, we are unable to interpret the extracted information. In that case, machine learning is used. With so many datasets available, the demand for machine learning is increasing. Machine learning is used in many industries to extract relevant data. Machine learning is designed to learn from data. Many studies have been conducted on how to teach machines to learn on their own without being explicitly programmed. Many mathematicians and programmers use various approaches to solve this problem, which has large data sets.

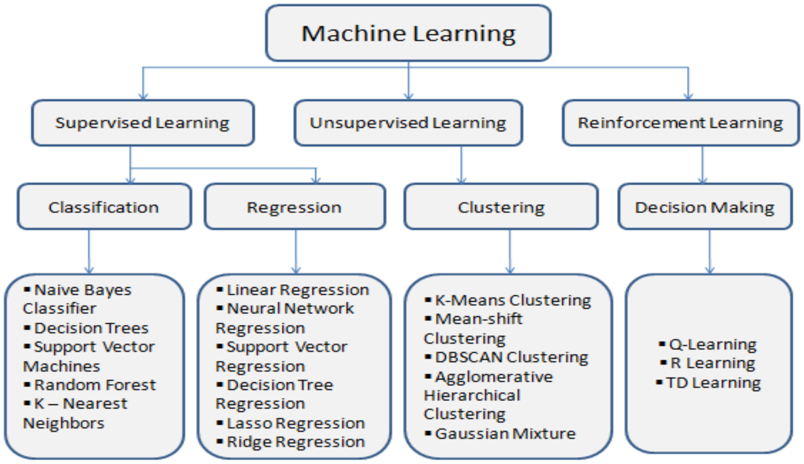

Figure.1

To solve data problems, Machine Learning employs various algorithms. Data scientists like to emphasise that there is no one-size-fits-all type of algorithm that is best for solving a problem. The type of algorithm used is determined by the type of problem to be solved, the number of variables, the best model for it, and so on

1.1 Machine Learning's Purpose

The field of machine learning can be divided into three major research areas:

- **Task-Oriented Studies**: Learning development and analysis. Systems designed to solve a specific set of tasks (also known as the "engineering approach").
- **Cognitive Simulation**: The study and computer simulation of human learning processes (also referred to as the "cognitive modelling approach").
- **Theoretical Analysis:** the theoretical exploration of the space of possible learning methods and algorithms those are independent of the application domain.

1.2 The Aim of Machine Learning

Although many research efforts are primarily aimed at one of these goals, progress in one often leads to progress in another. To investigate the space of possible learning methods, for example, a reasonable starting point may be to consider the only known example of robust learning behaviour, namely humans (and possibly other biological systems). Similarly, psychological studies of human learning may be supported by theoretical analysis that suggests various learning models. In stone task-oriented study, the need to acquire a specific form of knowledge may spawn new theoretical analysis or pose the question: "How do humans acquire this specific skill (or knowledge)?" Expert system research, cognitive simulation, and theoretical studies all contribute to intelligence (cross-fertilization of problems and ideas) (Jaime G. Carbonell, 1983).

1.2.1 Applied Learning Systems

Now, instructing a computer or a computer-controlled robot to perform a task necessitates the development of a complete and correct algorithm for that task, followed by the laborious programming of that algorithm into a computer. These activities typically entail a time-consuming and labor-intensive effort by specially trained personnel. Computer systems today cannot truly learn to perform a task by using examples or analogy to a similar, previously solved task. They also cannot significantly improve based on previous mistakes or acquire new skills by observing and imitating experts. Machine learning research strives to open the possibility of instructing computers in such novel ways, and thus promises to alleviate the burden of manually programming ever-increasing volumes of increasingly complex information into tomorrow's computers. The rapid expansion of computer applications and availability today makes this possibility even more appealing and desirable.

1.2.2 Knowledge acquisition

When approaching a task-oriented knowledge acquisition task, it is important to remember that the resulting computer system will interact with humans and should thus closely parallel Human capabilities. The conventional wisdom holds that an engineering approach does not have to reflect human or biological performance and thus is not truly applicable to machine learning. Since aeroplanes, a successful result of an almost pure engineering approach, bear little resemblance to their biological counterparts, one could argue that applied knowledge acquisition systems could be equally divorced from any consideration of human capabilities. This argument does not apply in this case because airplanes are not required to interact with, or understand, birds. Learning machines, on the other hand, will have to interact with the people who use them, and thus the concepts and skills they acquire—even if not necessarily through their internal mechanism—and must be understandable to humans.

1.3 Machine Learning as a Science

The learning mechanism, which is the ability to learn facts, skills, and more abstract concepts, is a clear contender for a cognitive invariant in humans. Understanding human learning well enough to reproduce aspects of that learning behaviour in a computer system is thus a worthy scientific goal in and of itself. Furthermore, the computer can be used to test the consistency and completeness of learning theories, as well as to enforce a commitment to the fine-structure process level detail that precludes meaningless tautological or untestable theories (Bishop, 2006)

The study of human learning processes is also very important in practise. Understanding the principles that underpin human learning abilities is likely to result in more effective educational techniques. The goal of machine learning research is to create intelligent computer assistants or computer tutoring systems, and many of these goals are shared by the machine learning fields. According to Jaime et al (Jaime G. Carbonell, 1983), computer tutoring is beginning to include abilities to infer models of student competence from observed performance. Inferring a student's knowledge and skills in a specific area allows for much more effective and personalised tutoring of the student (Sleeman, 1983).

The basic scientific goal of machine learning is to investigate possible learning mechanisms, including the discovery of various induction algorithms. The scope of theoretical limitations of certain methods appears to be the information that must be available to the learner, the problem of dealing with imperfect training data and developing general techniques applicable to a wide range of task domains. There is no reason to believe that human learning methods are the only way to gain knowledge and skills. In fact, common sense suggests that human learning is just one point in an uncharted space of possible learning methods—a point that has evolved to be particularly well suited to coping with the general physical environment in which we exist. The majority of theoretical work in machine learning is focused on the development, characterization, and analysis of general learning methods, with a focus on analysing generality and performance rather than psychological plausibility.

Knowledge Acquisition and Skill Refinement

There are two basic forms of learning:

1) Knowledge Acquisition
2) Skill refinement

When someone says they learned mathematics, it means they acquired mathematical concepts, understood their meaning, and their relationship to one another. as well as to the rest of the world. The importance of learning in this case is knowledge acquisition, which includes the description and modelling of physical systems and their behaviours, incorporating a variety of representations ranging from simple intrusive mental model models, examples, and images to fully testing mathematical equations and physical laws. A person is said to have learned more if this knowledge explains a broader range of situations, is more accurate, and can better predict typical world behaviour (Allix, 2003). this type of learning is typically applied to a wide range of situations and is generally acquired knowledge. As a result, knowledge acquisition is defined as learning a new task and being able to apply the information effectively. Learning by practise is the second type of learning that involves the gradual improvement of motor and cognitive skills. Learning activities such as:

- Learning to drive a car
- Learning to play keyboard
- Learning to ride a bicycle
- Learning to play piano

If one obtains all textbook knowledge on how to perform the activities, this is the first step in developing the necessary skills. As a result, the majority of the learning process is spent taming the acquired skill and improving mental or motor coordination or learning coordination through repetition and correction of deviations from desired behaviour. This type of learning is also known as skill taming. This differs from knowledge acquisition in numerous ways. Whereas knowledge acquisition is a conscious process that results in the creation of new representative knowledge structures and mental models, skill taming is learning by example or repeated practise without concerted conscious effort. According to Jamie (Jaime G. Carbonell, 1983), most human learning appears to be a combination of the two, with intellectual endeavours favouring the former and motor coordination tasks favouring the latter. The majority of current machine learning research focuses on knowledge acquisition, though some studies, particularly those concerned with problem-solving and transforming declarative instructions into effective actions; touch on aspects of both types of learning.

While knowledge acquisition clearly belongs in the domain of artificial intelligence research, one could argue that skill refinement is more closely related to non-symbolic processes like those studied in adaptative control systems. As a result, it is possible that both types of learning (knowledge-based and refinement learning) can be captured in artificial intelligence models.

1.4 Classification of Machine Learning

There are several areas of machine learning that could be used to solve email management problems, and our approach used an unsupervised machine learning method.

Unsupervised learning is a machine learning method in which the algorithm is given only examples from the input space and a model is fitted to these observations.

A clustering algorithm, for example, is an example of unsupervised learning. "Unsupervised learning is a machine learning method in which a model is fitted to observations. It differs from supervised learning in that there is no a priori output. Unsupervised learning begins with gathering a data set of input objects. Unsupervised learning, on the other hand, typically treats input objects as a set of random variables. A joint density model is then constructed for the data set. According to Russell, "the problem of unsupervised learning entailed learning patterns in the input when no specific output values were supplied" (Russell, 2003).

In the unsupervised learning problem, we only observe the features and have no measurements of the outcome. Our task is to describe how the data are organised or clustered". According to Trevor Hastie (Trevor Hastie, 2001), "There is no explicit teacher in unsupervised learning or clustering, and the system forms clusters or "natural groupings" of input patterns. Natural is always defined explicitly or implicitly in the clustering system itself, and given a specific set of patterns or cost function, different clustering algorithms lead to different clusters. The user will frequently set the hypothesised number of different clusters ahead of time, but how should this be done? How can we avoid inappropriate representations?" according to Duda (Richard O. Duda, 2000). Artificial intelligence is classified into several categories. Machine learning systems are classified as follows:

Supervised Machine Learning

Supervised learning is a machine learning technique that uses training data to learn a function. The training data is made up of pairs of input objects (usually vectors) and desired outputs. The function's output can be a continuous value (called regression) or it can predict the class label of the input object (called classification). After viewing a number of training examples, the supervised learner must predict the value of the function for any valid input object (i.e. pairs of input and target output). To accomplish this, the learner must "reasonably" generalise from the presented data to previously unseen situations (see inductive bias). (Consider unsupervised learning.)

Supervised learning is a machine learning technique in which the algorithm is first presented with training data that includes both the inputs and the desired outputs, allowing it to learn a function. The learner should then be able to generalise from the presented data to previously unseen examples" (Mitchell, 2006). A "teacher" may also provide us with a training set of (X, Y) pairs as part of supervised learning. We know the values of f for the m samples in the training set, and we assume that if we can find a hypothesis, h, that closely agrees with f for the members of, then this hypothesis will be a good guess for, especially if is large.

Curve fitting is a simple example of supervised function learning. Consider the values of a two-dimensional function. f, at the four sample points indicated by Figure 9 shows solid circles.

We want to find a function, h, that fits these four points and is drawn from the set, H, of second-degree functions. We show a two-dimensional parabolic surface that fits the points above the 1 x. 2 x, plane. This parabolic function, h, represents our hypothesis about the function f that generated the four samples. In this case, h f at the four samples, but exact matches are not required.

Unsupervised Machine Learning

Unsupervised learning1 is a type of machine learning in which inputs are not labelled manually. It differs from supervised learning approaches, which use a set of human-prepared examples to learn how to perform a task, such as classification or regression. We are only given the Xs and some (ultimate) feedback function on our performance in unsupervised learning. We simply have a training set of vectors that lack function values. Typically, the problem in this case is to divide the training set into appropriate subsets, 1.... R.

1.4.1 Classification of Machine Learning

Classification of machine learning system could be implemented along many different dimensions and we have chosen these two dimensions:

Inference Learning

This is a type of classification based on the underlying strategy involved. These strategies will be determined by the amount of inference performed by the learning system on the information provided to it.

The amount of inference the learner performs on the information provided now distinguishes learning strategies. So, for example, if a computer system performs email classification, its knowledge increases, but it may not perform any inference on the new information, implying that all cognitive efforts are on the part of the analyst or programmer.

However, if the machine learning classifier discovers or adopts new theories on its own, it will perform a significant inference. This is known as deriving knowledge from example, experiments, or observation. For example, if a student wants to solve statistical problems in a text book, this is a process that involves inference, but the solution is not to discover a brand-new formula without the assistance of a teacher or text book.

According to Jaime (Jaime G. Carbonell, 1983) and (Anil Mathur, 1999), it is much more difficult to teach a person by explaining each step in a complex task than by showing them. demonstrating how similar tasks are typically completed. It is even more difficult to programme a computer to perform a complex task than to instruct a person to perform the task, because programming necessitates explicit specification of all prerequisite details, whereas a person receiving instruction can fill in most mundane details using prior knowledge and common sense.

Knowledge Representation

This is a skill that the learner acquires based on the type of knowledge representation.

1.4.2 Existing Learning Systems

There are numerous other existing learning systems that use multiple strategies and knowledge representations, some of which have been applied to more than one. No inference is used in the knowledge-based machine learning method, but the learner displays the transformation of knowledge in a variety of ways:

- **Learning by programming**: When an algorithm or code is written to perform a specific task. For example, a code could be written as a guessing game for the type of animal. An external entity could alter such a programme.
- **Learning by memorization: Memorization** is the process of memorizing facts or data without drawing any conclusions from the incoming information or data.
- **Learning from examples:** This is an example of inductive learning. Given a set of positive and negative examples and counterexamples for a concept, the learner infers a general concept description that describes all of the positive examples but none of the negative examples. Learning from examples is a method that has received a lot of attention in the field of artificial intelligence. The amount of inference performed by the learner is much greater than in instruction-based learning (Anil Mathur, 1999). (Jaime G. Carbonell, 1983).
- **Learning from Observation:** This is an unsupervised learning approach and a very general form of inductive learning that includes discovery systems, theory formation tasks, the creation of classification criteria to form taxonomic hierarchies, and similar tasks that must be completed without the assistance of an external teacher. As previously stated, unsupervised learning requires the learner to perform more inference than any other approach. The learner is not given a set of data or an example of a specific concept. The above classification of learning strategies should aid in comparing various learning systems in terms of their underlying mechanisms, available external sources of information, and the extent to which they respond on reorganized knowledge.

1.5 Machine Learning Applications

Another consideration for classifying learning systems is the application domain, which adds a new dimension to machine learning. The areas below are where various existing learning systems have been applied. They are as follows:

1) Computer coding
2) Game play (chess, poker, and so on)
3) Image recognition, Speech recognition
4) Medical evaluation
5) Agriculture, Physics
6) Email administration, Robotics
7) Mathematics
8) Music
9) Natural Language processing, among many other things.

2. References

1. Allix, Nicholas M. "Epistemology and knowledge management concepts and practices." *Journal of Knowledge Management Practice* 4.1 (2003): 1-24.
2. Alpaydin, Ethem. *Introduction to machine learning*. MIT press, 2020.
3. Anderson, J. R. (1995). *Learning and Memory*. Wiley, New York, USA.
4. Anil Mathur, G. P. (1999). Socialization influences on preparation for later life. *Journal of Marketing Practice: Applied Marketing Science , 5* (6,7,8), 163 - 176.
5. Ashby, W. R. (1960). *Design of a Brain, The Origin of Adaptive Behaviour*. John Wiley and Son.
6. Batista, G. &. (2003). An Analysis of Four Missing Data Treatment Methods for Suppervised Learning. *Applied Artificial Intelligence , 17*, 519-533.
7. Bishop, C. M. (1995). *Neural Networks for Pattern Recognition.* Oxford, England: Oxford University Press.
8. Block H, D. (1961). The Perceptron: A Model of Brian Functioning. *34* (1), 123-135.
9. Carling, A. (1992). *Introducing Neural Networks* . Wilmslow, UK: Sigma Press.
10. D. Michie, D. J. (1994). *Machine Learning, Neural and Statistical Classification.* Prentice Hall Inc.
11. Fausett, L. (19994). *Fundamentals of Neural Networks.* New York: Prentice Hall.

12. Forsyth, R. S. (1990). The strange story of the Perceptron. *Artificial Intelligence Review , 4* (2), 147-155.
13. Friedberg, R. M. (1958). A learning machine: Part, 1. *IBM Journal* , 2-13.
14. Ghahramani, Z. (2008). Unsupervised learning algorithms are designed to extract structure from data. *178*, pp. 1-8. IOS Press.
15. Gillies, D. (1996). *Artificial Intelligence and Scientific Method.* OUP Oxford.
16. Haykin, S. (19994). *Neural Networks: A Comprehensive Foundation.* New York: Macmillan Publishing.
17. Hodge, V. A. (2004). A Survey of Outlier Detection Methodologies. *Artificial Intelligence Review 22* (2), 85-126.
18. Holland, J. (1980). Adaptive Algorithms for Discovering and Using General Patterns in Growing Knowledge Bases Policy Analysis and Information Systems. *4* (3).
19. Ian H. Witten, E. F. (2005). *Data Mining Practical Machine Learning and Techniques* (Second edition ed.). Morgan Kaufmann.
20. Jaime G. Carbonell, R. S. (1983). Machine Learning: A Historical and Methodological Analysis. *Association for the Advancement of Artificial Intelligence , 4* (3), 1-10.