International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue V, May 2019- Available at www.ijraset.com

Machine Learning Techniques: A Review

Chitra T. Wasnik[1], Kanchan Mankar[2]
[1, 2]*Department of Computer Engineering, Lokmanya Tilak College of Engineering, Navi Mumbai, Maharashtra-400709, India*

Abstract: Machine learning (ML) is one of the core elements of artificial intelligence, which studies from previous understandings to convalesce the functioning of intelligent programs. ML develops a real and efficient learning model that ascertains how to guesstimate from a specified instance's training data. Information technology denotes collection of themes concerned about algorithm conception and estimation that expediate pattern differentiation, categorization, and likelihood on the basis of models stemmed from prevailing data. Currently, ML aims to validate the promise of generating precise valuations. This paper presents a review of various ML techniques along with their advantages and disadvantages and elaborates on the operational model of ML.
Keywords: Machine Learning, Artificial Intelligence, Classification, Differentiation

I. INTRODUCTION

Machine learning (ML) can be defined as an architype that implies discovering from bygone incidence (i.e., previous data) in order to recuperate forthcoming operation. Computerised learning techniques is one of the solitary purposes of ML. Learning means algorithm alteration on the basis of older knowhows in an automated manner deprived of any human support. ML instructs machineries how to proficiently manage data.

From time to time after observing data, understanding the pattern of that data or extricating intelligence becomes problematic. Herein, applying ML becomes decisive [1]. Nowadays, ML is gaining utmost importance because of abundant readiness of datasets. For instance, medical industries and armed forces use ML in order to excerpt pertinent data. Learning from existing data is one of the sole reasons of ML.

Several studies have been carried out that explain how to make machines learn by themselves [2, 3]. ML is a multidisciplinary field in AI, probability, mathematics and statistics, and information theory. It unravels real-world problems by developing a suitable model for approximating data. The investigation of ML has expanded from the attempts of searching whether computers could learn to impersonate human brain and a field of statistics to a wide-ranging discipline that has produced fundamental statistical computational theories of learning processes [4]. Recently, adaptive programming is reconnoitred, which uses ML, wherein programs can recognize patterns, ascertain from knowledge, hypothesise new information, and optimize adeptness and exactness of its processing and output.

On account of new computing technologies, ML today isn't like ML of the past. Nonetheless, voluminous ML algorithms have been developed, contemporary development in ML is the competence to instinctively use multifaceted mathematical calculations to big data. Today, ML gains popularity because of rising volumes and varieties of available data, inexpensive and robust computational processing, and inexpensive storage of data.

This indicates that it is feasible to spontaneously create models that not only can examine larger and more complex data but also provide speedier and precise results on a large scale. The models of ML produce high-value predictions that assist in taking improved decisions in real time without human beings. ML not only responds to present-day demand but also tries to predict real-time demand. Because computation gets low-priced, ML makes unwieldly things manageable, which helps in making and intelligent infrastructure.

There exists a lot of scope to develop new algorithms to increase the discipline of ML and extensive work needs to be conducted to substitute existing algorithms from new ones.

II. MODEL OF MACHINE LEARNING

In ML, how the process of learning takes place can be explained broadly in two steps, i.e., training and testing processes.

A. Operational Model of ML

Figure 1 shows the operational model of ML. In the training process, certain samples are selected from training data that act as input where features are learned by a learning algorithm and then a learning model is built.

However, in the testing process, learning model uses execution engine to make prediction for test or production data. The final output of learning model that gives final prediction or classified data is called labeled data.

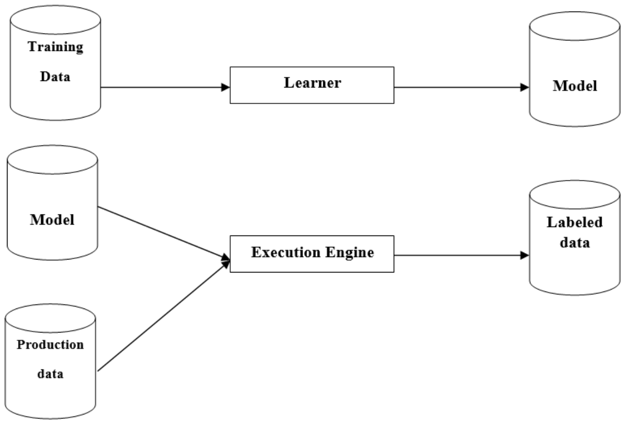

Fig. 1 Operational model of ML

III. TECHNIQUES OF MACHINE LEARNING

Various techniques of machine learning are as follows:

A. *Supervised or Directed Learning*

The machine is designed by exploiting the *a priori* known information in the form of 'direct' training examples consisting of observed values of system states (input vectors): $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}$, and the response (output) to each state: $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(N)}$.
The 'supervisor' has thus provided the following data:

$$Ɗ = \{s^{(i)}, y^{(i)}\};\ i = 1, \ldots, N$$

$$s^{(i)} = \mathbf{x}^{(i)}: \{x_1^{(i)}, x_2^{(i)}, \ldots, x_n^{(i)}\} \qquad \textbf{(1)}$$

The dataset Ɗ is used for inferring a model of the system.
If the dataset Ɗ lies in the region **X** of the state space Ř^n (X ⊂ Ř^n), then X must be fully representative of situations. Choice of features/attributes x_j; $j = 1, \ldots, n$, significantly affects the output.
Figure 2 shows the flow of supervised or directed learning. Here, an algorithm makes distinction between raw observed data (i.e., training data) [e.g., text, document, or image], and some label is given to the model during training. Supervised or directed learning algorithm builds prediction model in the process of training.
After the model is trained, the fitted model attempts to predict most likely labels for a new set of samples in test data. On the basis of the nature of the target, supervised or directed learning can be classified as follows:

1) If the target has values in a fixed set of categorical outcomes, then the task to predict the target is called classification.
2) If the target has floating point values, then the task to predict the target is called regression.

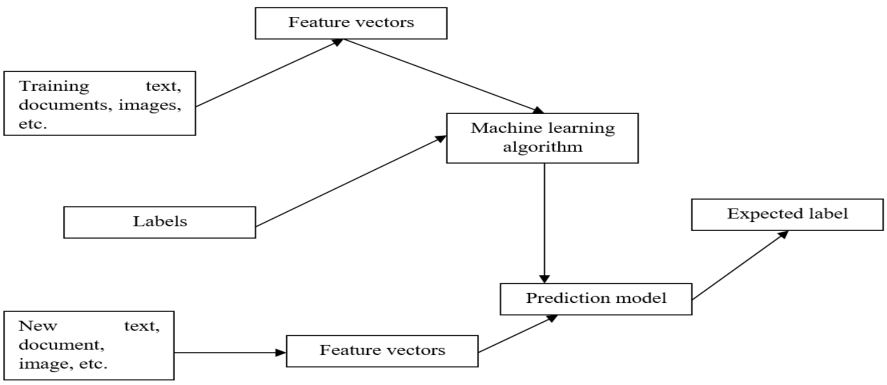

Fig. 2 Flow of supervised or directed learning

Moreover, supervised or directed learning technique is the one that needs external assistance. The input dataset is divided into two sets, i.e., training and testing datasets. The output variable that needs to be classified is possessed by training dataset, which in turn is supplied to testing dataset for classification [5].

There are two types of tasks for which supervised/directed learning is used, i.e., pattern recognition and numeric prediction.

B. Unsupervised or Undirected Learning

Another form of ML task is when output $y^{(i)}$ is not available in training data. In this type of problem, a set of feature vectors $\mathbf{x}^{(i)}$ is given, and the goal is to unravel underlying similarities.

Figure 3 shows the flow of unsupervised or undirected learning. From the figure, it is evident that unsupervised or undirected learning algorithm uses only a single set of observations with *n* samples and *n* features. Moreover, unsupervised algorithms do not used any kind of label.

During the process of training, unsupervised or undirected learning algorithm builds prediction model that tries to fit its parameters to best summarize regularities found in data.

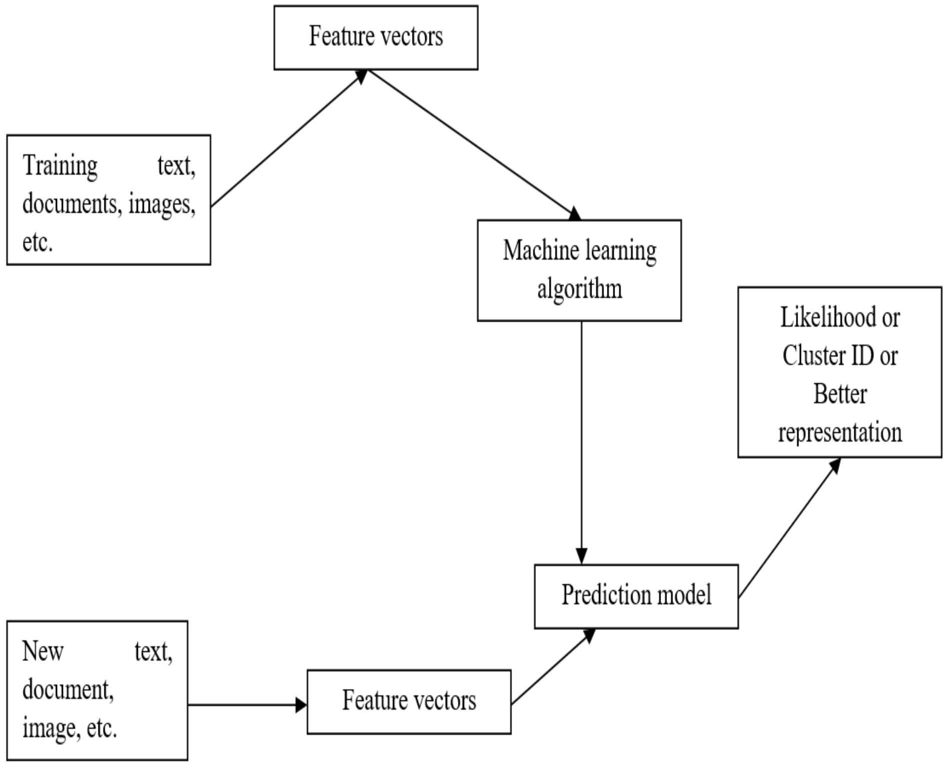

Fig. 3 Flow of unsupervised or undirected learning

Association and cluster analyses are the two different types of learning tasks that frequently appear in real-world applications of unsupervised or undirected learning.

1) *Association analysis:* It uses unsupervised learning to discover patterns in the data where no target is specified earlier. It is up to human interpretation to make sense of the patterns.
2) *Cluster analysis:* It creates clusters of similar records based on measurements made for these records. A primary issue in clustering is that of defining 'similarity' between feature vectors $\mathbf{x}^{(i)}$; $i = 1, 2, \ldots, N$, representing the records. The other issue is algorithm scheme selection that will cluster vectors based on accepted similarity measure.

C. Semi-Supervised Learning

It is a technique that merges supervised learning as well as unsupervised learning. This technique is only used when unlabeled data is previously present and obtaining labeled data becomes uninteresting [6]. Generative models, self-training, transductive support vector machine, etc. are some of the categories of semi-supervised learning.

D. Reinforcement Learning

It is based on the concept that if an action is followed by a satisfactory state of affairs, or by an improved state of affairs, then the inclination to produce that action becomes stronger, i.e., reinforced. The idea can be extended to permit action choices to be dependent on state information, which then brings in the aspect of feedback. Thus, a reinforcement learning system is the one that enhances its performance by obtaining feedback in the form of a scalar reward—a reinforcement signal, that is indicative of the suitability of the response. The learning system is not instructed with regard to what action has to be taken. Instead it is expected to find out which actions produce maximum reward by trying them. The actions may influence not only immediate reward but also next situation, and through that all subsequent rewards. Trial-and-error search and cumulative reward are the two significant distinguishing attributes of reinforcement learning. In other words, it is a technique that makes decisions on the basis of which actions to take such that a positive outcome is obtained. A learner has no knowledge about what type of actions to be undertaken until a particular situation is provided. In addition, the action that a learner takes might influence situations in the future [7].

E. Multi-task Learning

The chief goal of this technique is to assist other learners to work better. When a task is applied with multi-task learning technique, it recollects the procedure as to how a problem was solved and how did it reach to a specific conclusion. Next, this technique uses the abovementioned steps to determine solution of other tasks or problems that are similar in nature, which is indirectly known as inductive transfer mechanism. In other words, if learners share their experience with one another, they can then learn not only concomitantly but also faster [8].

F. Ensemble Learning

When numerous individual learners are coalesced to form a single learner, then such learning is known as ensemble learning.
For given amount and quality of training data, the output of one hypothesis function may be inappropriate for the problem at hand. The ideal model to make more reliable decisions is to create a combination of outputs of different hypotheses. Many ML algorithms do this by learning an ensemble of hypothesis and employing them in a combined form.
Bagging and boosting are the most frequently used techniques under ensemble learning.

1) *Bagging:* In this technique, individual approaches are constructed separately.
2) *Boosting:* Each new model is impacted by the performance of those models that are built earlier. In this technique, a model with accuracy on training set greater than average is created and then a new component called classifier is added to make an ensemble whose joint decision rule possesses a high level of accuracy on training set.

Note that a set of learners is always better at performing a specific task instead of individual learners [9].

G. Neural Network

Neural network learning technique provides a robust approach to approximating real-valued, discrete-valued, and vector-valued target functions. For certain types of problems, such as learning to interpret complex real-world sensor data, artificial neural networks (ANNs) are among the most effective learning methods currently known. Additionally, ANNs are derived from the biological concept of neurons. ANN learning is robust to errors in training data and has been successfully applied to various problems like interpreting visual scenes, speech recognition, and learning robot control strategies.

H. Instance-Based Learning

Methods, such as nearest neighbour and locally weighted regression, under this technique are conceptually straightforward approaches to approximating real-valued or discrete-valued target functions.
In this type of ML technique, a learner studies a specific kind of pattern and attempts to apply that pattern to recently fed data. It is a type of lazy learner that anticipates test data to arrive, and when test data arrives, then the data act on it together with training data. As the size of the data increases, learning algorithm complexity also increases.
Instance-based learning technique uses complex, symbolic representations for instances. One of the main disadvantages of this technique is the cost of classifying new instances can be high because of the fact that nearly all computation occurs at classification time rather than when training examples are first encountered.

IV. CONCLUSION

This paper provided an overview of various techniques of ML. At present, every single person uses ML either intentionally or unintentionally. Moreover, the paper describes the operational model of ML, which gives an overview of the process of ML.

REFERENCES

[1] W. Richert, L. P. Coelho, "Building Machine Learning Systems with Python", Packt Publishing Ltd., ISBN 978-1-78216-140-0
[2] M. Welling, "A First Encounter with Machine Learning", Irvine, CA.: University of California, 12
[3] M. Bowles, "Machine Learning in Python: Essential Techniques for Predictive Analytics", John Wiley & Sons Inc., ISBN: 978-1-118-96174-2
[4] Tom M. Mitchell, "The Discipline of Machine Learning", [Online] Available: http://reports-archive.adm.cs.cmu.edu/anon/ml/abstracts/06-108.html [Accessed on May 08, 2018]
[5] S. B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques", Informatica 31 (2007), pp. 249-268
[6] X. Zhu, A. B. Goldberg, "Introduction to Semi-Supervised Learning", Synthesis Lectures on Artificial Intelligence and Machine Learning, 2009, vol. 3, no. 1, pp. 1-130
[7] R. S. Sutton, "Introduction: The Challenge of Reinforcement Learning", Kluwer Academic Publishers, Boston, Machine Learning, 8, pp. 225-227, 1992
[8] R. Caruana, "Multitask Learning", Machine Learning, Kluwer Academic Publishers, 28, pp. 41-75,1997
[9] D. Opitz, R. Maclin, "Popular Ensemble Methods: An Empirical Study", Journal of Artificial Intelligence Research, 11, pp. 169-198, 1999