A Survey on Machine Learning: Concept, Algorithms and Applications

Hossam Saad & Shady Yasser & Hassan Gaber & Aya Samy & Amany Taha & Ghada Salah & Amira Abdel Salam

Abstract- In the past few decades, the field of computer learning has witnessed a remarkable development, so some have used it in the practice of computer games, some have been interested in mathematics and statistics, and hence the emergence of so-called algorithms, which have developed greatly and which have been used to identify the forms and texts, which led to a separate research interest in itself. In order to extract the data to see the extent of its conformity to the correct.

Keywords- Machine Learning, Algorithm, Data, Training, accuracy.

1. INTRODUCTION

Machine learning is a reference to learning from past experiences and data to get better results in the future. This learning indicates a better picture of the algorithm. Based on previous experiences, the program design has a specific goal, for example the Robert Galbraith series and the Harry Potter series.

The London Sunday Times newspaper reported that the authenticity of the claim that Rowling was the author of the books in question through forensic machine learning.

The training and development of the machine learning algorithm is done through several books, including the Rolling book.

In this paper we will explain the concept of machine learning and identify its algorithms.

A. MACHINE LEARNING: INTERSECTION OF STATISTICS AND COMPUTER SCIENCE

Machine learning is an extension of statistical science and computer science that focuses on the ways to solve problems and find out whether the problem is solvable or not through a specific approach that helps the statistics to know the data and measure the credibility of the conclusions, in addition to computer programming manually.

Additional concerns arise for automated learning through the effectiveness of algorithms to perform data processing.

B. MACHINE LEARNING AND HUMAN LEARNING

Machine learning in the third field of research was closely related to the study of the human and animal mind in neuroscience, psychology and others.

According to the jurisprudence of scientists, the method of learning the machine will not differ greatly from the way of learning of human or animal, however, research in this direction has not been very fruitful to this time, in terms of statistics.

C. DATA MINING, ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Scientific studies have shown the overlap of these disciplines and their similarity very much.

In scientific methods, it is not limited to the sampling of information from many sources and only or the use of analysis and patterns, and therefore the search for data is a way to draw hypotheses useful to the subject and used as a basis for artificial intelligence and learning automation.

What is artificial intelligence ??

He knows artificial intelligence and briefly explains that the machine can solve a particular problem on its own.

Automated learning provides basic data and then identifies basic patterns using statistics. In short, the system cannot be considered intelligent if it lacks the ability to learn from its previous failures.

2. PRESENT RESEARCH QUESTIONS & RELATED WORK

From the above, it is clear to us that significant progress has been made in machine learning algorithms. It is a broad discipline and many scientists and researchers have added their work in this field.

A. USING UNLABELLED DATA IN SUPERVISED LEARNING[10][11][25][26][27]
B. TRANSMITTING THE LEARNING EXPERIENCE

There is a problem that the supervised algorithm may include learning a lot of functions rather than one function such as diagnostic functions for hospitals around the world.

Machine learning algorithms provide hierarchical Bayesian methods one approach that assumes that the learning parameters for both functions are for Kolkata and London, respectively, and the subject becomes more accurate when the transfer between functions is doubled.

C. LINKING DIFFERENT ML ALGORITHMS

Algorithms have been presented and tested in many fields, and research has emerged in the discovery of possible correlations between these algorithms. In general, the theoretical understanding of machine learning algorithms, their properties and their effectiveness remains, to this day, questionable.

D. BEST STRATEGICAL APPROACH FOR LEARNERS WHICH COLLECTS THEIR OWN DATA

The research focuses on learning systems. Instead of using data collected automatically, the research aims to deeply understand the effectiveness of the system in finding the most effective strategy to fully control the learning algorithm. For example, consider the drug testing system and find out how successful the drug is during Monitor your patients with potential side effects and try to minimize them.

E. PRIVACY PRESERVING DATA MINING

This section searches for and access data successfully, in addition to obtaining the results without resorting to the original information. For example, there is a medical examination using data from hospitals around the world, but special concerns lead to the failure to follow this application very much even if this the study of data and its specificity, and recently the emergence of thinking to find a common learning algorithm in order to maintain the privacy of patients using statistics and modern coding techniques.

F. NEVER-ENDING LEARNERS

Among the learning requirements of learning automation is to train learners using a specific database and then take advantage of the outputs.

Despite the tremendous commercial applications of ML algorithms, learning in machines has largely lacked in human or animal learning.

The alternative approach is defined as the most pluralistic approach and the accumulated awareness of human learning. It is the never-ending science. For example, the learner (NELL) [8] is the learner of his job learning to read web pages and has reported that he reads the World Wide Web every hour Since January 2010, NELL has acquired more than 80 million registered shares.

3. CATEGORISATION OF ML ALGORITHMS

Over the past years, a large number of automated learning algorithms have been designed. Of course, we do not know them all very well. Some of them did not solve the problem.

A. ALGORITHMS GROUPED BY SIMILARITY

1-linear regression algorithms

Before knowing what is linear regression, let us get ourselves accustomed to regression. Regression is a method of modeling a target value based on independent predictors. This method is mostly used for forecasting and finding out cause and effect relationship between variables. Regression techniques mostly differ based on the number of independent variables and the type of relationship between the independent and dependent variables.

Simple linear regression is a type of regression analysis where the number of independent variables is one and there is a linear relationship between the independent(x) and dependent(y) variable. The red line in the above graph is referred to as the best fit straight
line. Based on the given data points, we try to plot a line that models the points the best. The line can be modeled based on the linear equation shown below.

\[ y = a_0 + a_1 \times x \]  

## Linear Equation

The motive of the linear regression algorithm is to find the best values for \( a_0 \) and \( a_1 \). Before moving on to the algorithm, let’s have a look at two important concepts you must know to better understand linear regression.

### 2-Instance-based Algorithms

Also known as Memory based learning, Instance based learning is a supervised classification learning algorithm that performs operation after comparing the current instances with the previously trained instances, which have been stored in memory. Its name is derived from the fact that it creates assumption from the training data instances.

Time complexity of Instance based learning algorithm depends upon the size of training data. Time complexity of this algorithm in worst case is \( O(n) \), where \( n \) is the number of training items to be used to classify a single new instance.

To improve the efficiency of instance based learning approach, preprocessing phase is required. Preprocessing phase is a data structure that enables efficient usage of run time modeling of test instance.

Various instance based learning algorithms:

- K Nearest Neighbor (KNN)
- Learning Vector Quantization (LVQ)
- Self-Organizing Map (SOM)

Example using KNN neighbor:
1. Splitting data into test and training datasets.
2. Building and training model with training data.

### 3-Regularization Algorithm

Have you seen this image before? As we move towards the right in this image, our model tries to learn too well the details and the noise from the training data, which ultimately results in poor performance on the unseen data.

In other words, while going towards the right, the complexity of the model increases such that the training error reduces but the testing error doesn’t. This is shown in the image below.

![Figure 2: underfitting, just right, overfitting](image)

### 4-Decision Tree Algorithms

Our daily life is full of decisions. In many cases, we face positions where we need to make a decision and there are many options that can be close or similar, so choosing a decision will not be easy.
In this case, the person is in a state of great confusion because of the inability to make a decision which may cause disruption of the daily work of the individual, and can have negative side effects that harm the human. So it was necessary to find a way to analyze the options and to prefer a trend from another direction and this method or method is what is known as the tree of decisions.

It is a mathematical method that helps the decision maker to know all the possible alternatives and the probability of occurrence and is usually used to determine the choice of the optimal solution from future cases. This method depends on the calculation of the expected financial value.

It is known that a person estimates the estimated future financial value of each option (decision), due to the choice of the decision with the highest financial value.

More clearly, each branch of the decision tree represents one of the alternatives and each branch can branch out to other branches of the future. In general, the decision tree is a decision analysis tool.

5-Bayesian Algorithms
Bayes’ theorem finds many uses in the probability theory and statistics. There’s a micro chance that you have never heard about this theorem in your life. Turns out that this theorem has found its way into the world of machine learning, to form one of the highly decorated algorithms.

As you might have guessed, this requires us to view things from a probabilistic point of view. Just as in machine learning, we have attributes, response variables and predictions or classifications. Using this algorithm, we will be dealing with the probability distributions of the variables in the dataset and predicting the probability of the response variable belonging to a particular value, given the attributes of a new instance. Let's start by reviewing the Bayes’ theorem:

Bayes theorem : This lets us examine the probability of an event based on the prior knowledge of any event that related to the former event. So for example, the probability that price of a house is high, can be better assessed if we know the facilities around it, compared to the assessment made without the knowledge of location of the house. Bayes’ theorem does exactly that.

\[
P(A | B) = \frac{P(B | A) P(A)}{P(B)}
\]

Now, let's see how this suits well to the purpose of machine learning: Take a simple machine learning problem, where we need to learn our model from a given set of attributes (in training examples) and then form a hypothesis or a relation to a response variable. Then we use this relation to predict a response, given attributes of a new instance. Using the Bayes’ theorem, it's possible to build a learner that predicts the probability of the response variable belonging to some class, given a new set of attributes. Consider the previous equation again. Now, assume that A is the response variable and B is the input attribute. So according to the equation, we have(A|B) : conditional probability of response variable belonging to a particular value, given the input attributes. This is also known as the posterior probability's(A) : The prior probability of the response variable(B) : The probability of training data or the evidence's(B|A) : This is known as the likelihood of the training data.

Let’s take a problem, where the number of attributes is equal to n and the response is a Boolean value, i.e. it can be in one of the two classes. Also, the attributes are categorical(2 categories for our case). Now, to train the classifier, we will need to calculate P(B|A), for all the values in the instance and response space. This means, we will need to calculate 2*(2^n - 1), parameters for learning this model. This is clearly unrealistic in most practical learning domains. For example, if there are 30 Boolean attributes, then we will need to estimate more than 3 billion parameters.

6-Support Vector Machine (SVM)
Support Vector Machine” (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well (look at the below snapshot).

Support Vectors are simply the co-ordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes (hyper-plane/ line).

How does it work?

1. How does it work? Identify the right hyper-plane: we have three hyper-planes (A, B and C) Now, identify the right hyper-plane to classify star and circle.
2. Identify the right hyper-plane: the distances between nearest data point (either class) and hyper-plane will help us to decide the right hyper-plane. This distance is called as Margin.
3. Identify the right hyper-plane.
4. Can we classify two classes.
5. Find the hyper-plane to segregate to classes.
How to tune Parameters of SVM?

Tuning parameters value for machine learning algorithms effectively improves the model performance.

**Pros:**
- It works really well with clear margin of separation.
- It is effective in high dimensional spaces.
- It is effective in cases where number of dimensions is greater than the number of samples.
- It uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.

**Cons:**
- It doesn’t perform well, when we have large data set because the required training time is higher.
- It also doesn’t perform very well, when the data set has more noise i.e. target classes are overlapping.
- SVM doesn’t directly provide probability estimates, these are calculated using an expensive five-fold cross-validation. It is related SVC method of Python scikit-learn library.

7-Association Algorithms

Association rule learning is a rule based method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using some measure of interestingness.

Based on the concept of strong rules, Rakish, Tomasz and Arum Swam introduced association rules for discovering regularities between products in large-scale transaction data recorded by point of sale (POS) systems in supermarkets.

Association rules are employed today in many application areas including web usage mining, intrusion, detection continuous production and bioinformatics. In contrast with sequence mining, association rule learning typically does not consider the order of items either within a transaction or across transactions.

Many algorithms for generating association rules have been proposed. Some well-known algorithms are Apriori, Éclat and FP-Growth, but they only do half the job, since they are algorithms for mining frequent item sets. Another step needs to be done after to generate rules from frequent item sets found in a database.

Apriority algorithm: a breadth-first search strategy to count the support of item sets and uses a candidate generation function which exploits the downward closure property of support.

![Image of associations between selected items](image_url)

**Figure 4:** Associations between selected items

Éclat algorithm: Éclat (alt. ECLAT, stands for Equivalence Class Transformation) is a depth-first search algorithm based on set intersection. It is suitable for both sequential as well as parallel execution with locality-enhancing properties.

FP-growth algorithm: In the first pass, the algorithm counts the occurrences of items (attribute-value pairs) in the dataset of transactions, and stores these counts in a 'header table'. In the second pass, it builds the FP-tree structure by inserting transactions into a Items in each transaction have to be sorted by descending order of their frequency in the dataset before being inserted so that the tree can be processed quickly. Items in each transaction that do not meet the minimum support requirement are discarded. If many transactions share most frequent items, the FP-tree provides high compression close to tree root.

8-Artificial Neural Network (ANN) Algorithms

Artificial Neural Network algorithms are inspired by the human brain. The artificial neurons are interconnected and communicate with each other. Each connection is weighted by previous learning events and with each new input of data more learning takes place. A lot of different algorithms are associated with Artificial Neural Networks and one of the most important is Deep learning. An example of Deep Learning. It is especially concerned with building much larger complex neural networks.

www.jcsis.org/
ANN is rarely used for predictive modeling. The reason being that Artificial Neural Networks (ANN) usually tries to over-fit the relationship. ANN is generally used in cases where what has happened in past is repeated almost exactly in same way. Hence, it is a kind of machine learning technique which has enormous memory. But it does not work well in case where scoring population is significantly different compared to training sample. For instance, if I plan to target customer for a campaign using their past response by an ANN. I will probably be using a wrong technique as it might have over-fitted the relationship between the response and other predictors. For same reason, it works very well in cases of image recognition and voice recognition.

9-Deep Learning Algorithms
The field of artificial intelligence is when machines can perform tasks that require human intelligence. It includes machine learning, where machines can learn through experience and acquire skills without human intervention. Deep learning is a subset of automated learning where artificial neural networks, algorithms inspired by the human brain itself, learn from large amounts of data. Similar to how we learn from experience, the deep learning algorithm will perform a task frequently every time it is slightly modified to improve the result. We refer to "deep learning" because neural networks have multiple (deep) layers that enable learning. Any problem that requires "thinking" is a problem that deep learning can learn to solve.

The amount of data we produce every day is amazing - currently estimated at 2.6 billion bytes - a resource that makes deep learning possible. Because deep learning algorithms require a lot of data to learn from, this increase in data creation is one of the reasons that has led to increased learning capabilities in recent years. In addition to creating more data, deep learning algorithms benefit from the power of today's powerful computing as well as the spread of artificial intelligence. Artificial Intelligence was harnessed as a service to smaller organizations and provided access to artificial intelligence technology and artificial intelligence algorithms required for deep learning without significant initial investment.

10-Ensemble Algorithms
Ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithm alone. Unlike a statistical ensemble in statistical machine which is usually infinite machine learning ensemble consists of only a concrete finite set of alternative models, but typically allows for much more flexible structure to exist among those alternative.

Ensemble size edit: while the number of component classifiers of an ensemble has a great impact on the accuracy of prediction, there is a limited number of studies addressing this problem.

Ensemble learning applications:
Change detection is an image analysis problem, consisting of the identification of places where the land cover has changed over time. Change detection is widely used in fields such as urban growth, forest and vegetation dynamics, land use and disaster monitoring.

4. MEASURING AND COMPARING PERFORMANCES OF POPULAR ML ALGORITHMS
A. NAIVE BAYES CLASSIFIER
This is the way every pair of feature is mutually independent, its performance quite well in many practical situations, like in text classification and spam detection. It is one of the best types advanced classification.
B. SUPPORT VECTOR MACHINE
SVM supervised classification algorithm we perform classification by finding the hyper-plane that differentiate the two classes very well.
C. DECISION TREE
One of the most successful learning aids is manageable. It constructs a graph to show each result a probable decision
The efficiency of an algorithm somewhat depends on the data set and the domain it is applied to Under certain conditions.

5. APPLICATION
Machine learning is research area that have attached a lot of mind and has the potential divulge further.
A. SPEECH RECOGNITION
All current speech recognition systems available in the market use machine learning approaches to train the system for better accuracy.
In practice, most of such system implement learning.
B. COMPUTER VISION
Majority of resent vision systems, facial recognition software, systems capable of automatic classification microscopic images of cells.
C. BIO-SURVEILLANCE
Several government initiatives to track probable outbreaks of diseases uses ML algorithm. Consider the RODS project in western Pennsylvania, this project collects admissions reports to emergency rooms in the hospitals there.
D. ROBOT OR AUTOMATION CONTROL
ML method are largely used in robot and automated system. For example, consider the use of ML to obtain control.
E. EMPIRICAL SCIENCE EXPERIMENTS
A large group data-intensive science disciplines use ML methods in several of it researches. For example, ML is being implemented in genetics to identify unusual celestial objects in astronomy.

6. FUTURE SCOPE

A. PROGRAMMING LANGUAGES CONTAINING MACHINE LEARNING PRIMITIVES

In majority of applications, ML algorithms are incorporated with manually coded programs as part of an application software. The need of a new programming language that is self-sufficient to support manually written subroutines as well as those defined as “to be learned.” It could enable the coder to define set of inputs-outputs of every “to be learned” program and opt for an algorithm from the group of basic learning methods already imparted in the language. Programming languages like Python, JavaScript, C++.

B. PERCEPTION

A generalized concept of computer perception that can link ML algorithms which are used in numerous forms of computer perception today including but not limited to highly advanced vision, speech recognition etc., is another potential research area. One thought-provoking problem is the integration of different senses (e.g., sight, hear, touch etc) to prepare a system which employ self-supervised learning to estimate one sensory knowledge using the others.

7. CONCLUSION

This is we know about the machine learning, types and their application. We now know as we study that the machine learning have important rule in Machine Learning is the science of getting computers to learn and act like humans do, and improve their learning over time in autonomous fashion.