Machine Learning - A Revolution in Artificial Intelligence

Abhishek Masiwal, Sudhakar Sharma

Abstract— **In the field of machine learning one considers the important question of how to make machines able to "learn". In the future, intelligent machines will replace or enhance human capabilities in many areas. Machine learning is the intelligence exhibited by machines or software. It is the subfield of computer science. Machine learning is becoming a popular field in computer science as it has enhanced the human life in many areas. Machine learning in the last two decades has greatly improved performance of the manufacturing and service systems. , it helps in taking an optimized decision for the machine which eventually increases the efficiency of the machine and more organized way of preforming a particular task. Now-a-days the concept of machine learning is used in many applications and is a core concept for intelligent systems which leads to the introduction innovative technology and more advance concepts of artificial thinking**

Index Terms— **machine learning, supervised learning, unsupervised learning, algorithms**

I. INTRODUCTION

Instead of building heavy machines with explicit programming now different algorithms are being introduce which will help the machine to understand the virtual environment and based on their understanding the machine will take particular decision. This will eventually decrease the number of programming concepts and also machine will become independent and take decisions on their own. Different algorithms are introduced for different types of machines and the decisions taken by them. Designing the algorithm and using it in most appropriate way is the real challenge for the developers and scientists. This research paper emphasizes on different types of machine learning algorithms and their most efficient use to make decisions more efficient and complete the task in more optimized form. Different algorithm gives machine different learning experience and adapting other things from the environment. Based on these algorithms the machine takes the decision and performs the specialized tasks. So it is very important for the algorithms to be optimized and complexity should be reduced because more the efficient algorithm more efficient decisions will the machine makes. To know whether an agent has learned, we must define a measure of success. The measure is usually not how well the agent performs on the training experiences, but how well the agent performs for new experiences. In this research paper we will discuss the two main types of algorithms i.e. supervised and unsupervised learning.

Manuscript received Jan 27, 2017

Abhishek Masiwal, Student, Computer Science and Engineering Department, Inderprastha Engineering College, Ghaziabad- India

Sudhakar Sharma, Student, Computer Science Department, Bhai Parmanand Institute of Business Studies, Delhi- India

II. PROBLEMS FACED IN MACHINE LEARNING

Learning is a complex process as lot of decisions are made and also it depends from machine to machine and from algorithm to algorithm, how to understand a particular problem and on understanding the problem how it responds to it. Some of the issues make a complex situation for the machine to respond and react. These problems not only make problem complex but it also affects the learning process of the machine. Some of them are-

Noise- Noise exists in the data when some features have been assigned the wrong value, there are inadequate features (the features given do not predict the classification), and often there are examples with missing features. One of the important properties of a learning algorithm is its ability to handle noisy data in all of its forms.

Bias- The tendency to prefer one hypothesis over another is called a bias. Consider the agents A and B. Saying that a hypothesis is better than A's or B's data - both A and B accurately predicts all of the data given. Without a bias, an agent will not be able to make any predictions on unseen examples. A learning algorithm with low bias must be "flexible" so that it can fit the data well. But if the learning algorithm is too flexible, it will fit each training data set differently, and hence have high variance. A key aspect of many supervised learning methods is that they are able to adjust this trade-off between bias and variance (either automatically or by providing a bias/variance parameter that the user can adjust).

Function complexity-

The amount of training data available relative to the complexity of the "true" function (classifier or regression function). If the true function is simple, then an "inflexible" learning algorithm with high bias and low variance will be able to learn it from a small amount of data. But if the true function is highly complex (e.g., because it involves complex interactions among many different input features and behaves differently in different parts of the input space), then the function will only be learnable from a very large amount of training data and using a "flexible" learning algorithm with low bias and high variance.

Dimensionality of the input space-

If the input feature vectors have very high dimension, the learning problem can be difficult even if the true function only depends on a small number of those features. This is because the many "extra" dimensions can confuse the learning algorithm and cause it to have high variance. Hence, high input dimensionality typically requires tuning the classifier to have low variance and high bias

III. UNSUPERVISED LEARNING

In unsupervised learning the machine simply receives the input y1, y2… but obtains neither supervised target outputs, nor rewards from its environment. But it is possible to develop a formal framework for unsupervised learning based on the notion that the machine's goal is to build representations of the input that can be used for decision making, predicting future inputs, efficiently communicating the inputs to another machine, etc. Example of unsupervised learning is clustering and dimensionality reduction.

Some algorithms for unsupervised learning are –

1. **K-means Structuring-**

In this algorithm we have to first select the number of cluster in advance, they might converge to a local minimum. K-means can be seen as a specialization of the expectation maximization (EM) algorithm. It is more efficient (lower computational complexity) than hierarchical clustering

Algorithm:

1. K-means ((X= {d1, . . . ,dn} ⊆ Rm , k): 2R)
2. C: 2R /*μ a set of clusters */
3. d = Rm x Rm -> R /*distance function*/
4. μ: 2R -> R /* μ computes the mean of a cluster */
5. select C with k initial centers f1,….fk
6. while stopping criterion not true do
7. for all clusters cj ∈ C do
8. cj ← {di|∀fld(di, fj) ≤d(di, fl)}
9. done
10. for all means fj do
11. fj <- μ(cj)
12. done
13. done
14. return C

2. **Hierarchical Clustering-**

This algorithm builds a multilevel hierarchy of cluster by creating a cluster tree.

Inputs: objects represented as vectors

Outputs: a hierarchy of associations represented as "dendogram".

Algorithm:

1. hclust(D: set of instances): tree
2. var: C, /* set of clusters */
3. M /* matrix containing distances between pairs of cluster */
4. for each d ∈ D} do
5. Make d a leaf node in C
6. done
7. for each pair a,b ∈ C do
8. Ma,b ← d(a, b)
9. done
10. while(not all instances in one cluster) do
11. Find the most similar pair of clusters in M
12. Merge these two clusters into one cluster.
13. Update M to reflect the merge operation.
14. done
15. return C

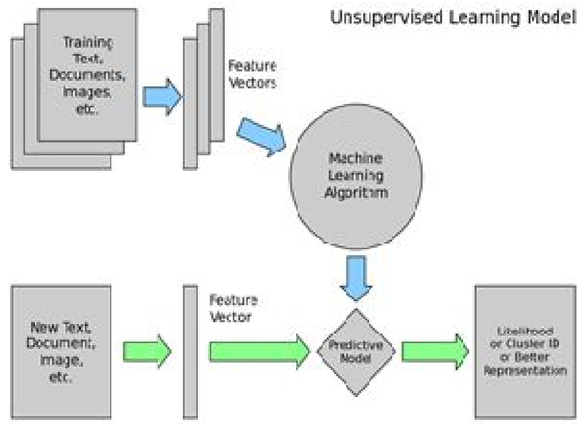

Fig. 1 working mechanism of unsupervised learning

IV. SUPERVISED LEARNING

Supervised learning is an algorithm in which both the inputs and outputs can be perceived. Based on this training data, the algorithm has to generalize such that it is able to correctly respond to all possible inputs. This algorithm is expected to produce correct output for inputs that weren't encountered during training. In supervised learning what has to be learned is specified for each example. Supervised classification occurs when a trainer provides the classification for each example. Supervised learning of actions occurs when the agent is given immediate feedback about the value of each action. In order to solve a give problem using supervised learning algorithm one has to follow some certain

Steps:-

1. Determine the type of training examples.
2. Gather a training set.
3. Determine the input feature representation of learned function.
4. Determine the structure of learning function & corresponding learning algorithm.
5. Complete the design and run the learning algorithm on the gather set of data.
6. Evaluate the accuracy of the learned function also the performance of the learning function should be measured and then the performance should be again measured on the set which is different from the training set.

Supervised Learning can be split into two broad categories:

1. Classification of responses that can have just a few values, such as 'true' or 'false'. Classification algorithm applies to nominal, not ordinal response values.

2. Regression for responses that are a real number, such as miles per gallon of a particular car.

Some algorithms for supervised learning are-

1. **K-Nearest Neighbour Classification-** Arguably the simplest method is the K-Nearest Neighbour classification. Here the k points of the training data closest to the test point are found, and a label is given to the test point by a majority vote between the k points. This method is highly intuitive and attains – given its simplicity – remarkably low classification errors, but it is computationally expensive and requires a large memory to store the training data.

2. **Linear Discriminant Analysis-** Computes a hyper plane in the input space that minimizes the within-class variance and maximizes the between class distance. It can be efficiently computed in the linear case even with large data sets.

3. **Decision Trees-** Intuitive class of classification algorithms are decision trees. These algorithms solve the classification problem by repeatedly partitioning the input space, so as to build a tree whose nodes are as pure as possible (that is, they contain points of a single class). Classification of a new test point is achieved by moving from top to bottom along the branches of the tree, starting from the root node, until a terminal node is reached.

4. **Neural Networks-** They are perhaps one of the most commonly used approaches to classification. Neural networks are a computational model inspired by the connectivity of neurons in animate nervous systems. A further boost to their popularity came with the proof that they can approximate any function mapping via the Universal Approximation Theorem.

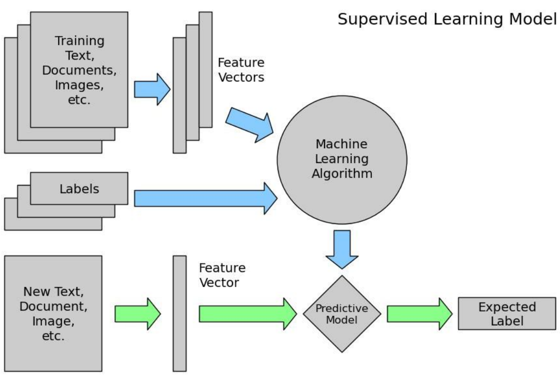

Fig. 2 working mechanism of supervised learning

CONCLUSION

Machine Learning research has been extremely active the last few years. The result is a large number of very accurate and efficient algorithms that are quite easy to use for a practitioner. It seems rewarding and almost mandatory for (computer) scientist and engineers to learn how and where Machine Learning can help to automate tasks or provide predictions where humans have difficulties to comprehend large amounts of data. The question of how to measure the performance of learning algorithms and classifiers has been investigated. This is a complex question with many aspects to consider. One conclusion of the analysis is that classifier performance is often measured in terms of classification accuracy, e.g., with cross-validation tests. Some methods were found to be general in the way that they can be used to evaluate any classifier (regardless of which algorithm was used to generate it) or any algorithm (regardless of the structure or representation of the classifiers it generates), while other methods only are applicable to a certain algorithm or representation of the classifier. One out of ten evaluation methods was graphical, i.e., the method does not work like a function returning a performance score as output, but rather the user has to analyse a visualization of classifier performance. The applicability of measure-based evaluation for measuring classifier performance has also been investigated and we provide empirical experiment results that strengthen earlier published theoretical arguments for using measure-based evaluation. For instance, the measure-based function implemented for the experiments, was able to distinguish between two classifiers that were similar in terms of accuracy but different in terms of classifier complexity.

REFERENCES

1. Rich Caruana; Alexandru Niculescu- Mizil,"An Empirical Comparison of Supervised Learning Algorithms", Department of Computer Science, Cornell University, Ithaca, NY 14853 USA
2. Niklas Lavesson,"Evaluation and Analysis of Supervised Learning Algorithms and Classifiers", Blekinge Institute of Technology Licentiate Dissertation Series No 2006:04,ISSN 1650-2140,ISBN 91-7295-083-8
3. Yugowati Praharsi; Shaou-Gang Miaou; Hui-Ming Wee," Supervised learning approaches and feature selection - a case study in diabetes", International Journal of Data Analysis Techniques and Strategies 2013 - Vol. 5, No.3 pp. 323 - 337
4. Andrew Ng,"Deep Learning And Unsupervised", "Genetic Learning Algorithms", "Reinforcement Learning and Control", Department of Computer Science, Stanford University, 450 Serra Mall, CA 94305, USA.
5. Bing Liu, "Supervised Learning", Department of Computer Science, University of Illinois at Chicago (UIC), 851 S. Morgan Street, Chicago
6. Niklas Lavesson,"Evaluation and Analysis of Supervised Learning Algorithms and Classifiers", Blekinge Institute of Technology Licentiate Dissertation Series No 2006:04, ISSN 1650-2140, ISBN 91-7295-083-8