Machine Learning: An artificial intelligence methodology

Anish Talwar[#1], Yogesh Kumar[*2]

[#1]UG, CSE

Dronacharya Group of Institutions,

Greater Noida

talwar.anish@yahoo.in

[*2]UG, CSE

Dronacharya Group of Institutions,

Greater Noida

M18nikhilkumar@gmail.com

Abstract— **The problem of learning and decision making is at the core level of argument in biological as well as artificial aspects. So scientist introduced Machine Learning as widely used concept in Artificial Intelligence. It is the concept which teaches machines to detect different patterns and to adapt to new circumstances. Machine Learning can be both experience and explanation based learning. In the field of robotics machine learning plays a vital role, 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.**

Keywords— **machine learning, supervised learning, unsupervised learning, algorithms**

I. INTRODUCTION

Learning is considered as a parameter for intelligent machines. Deep understanding would help in taking decisions in a more optimized form and also help then to work in most efficient method. As seeing is intelligence, so learning is also becoming a key to the study of biological and artificial vision. 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.

Pattern recognizing is also a concept in machine learning. Most algorithms use the concept of pattern recognition to make optimized decisions. As a consequence of this new interest in learning we are experiencing a new era in statistical and functional approximation techniques and their applications to domain such as computer visions.

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. Machine Learning algorithms do not totally dependent on nature's bounty for both inspiration and mechanisms. Fundamentally and scientifically these algorithms depends on the data structures used as well as theories of learning cognitive and genetic structures. But still natural procedure for learning gives great exposures for understanding and good scope for

variety of different types of circumstances. Many machine learning algorithm are generally being borrowed from current thinking in cognitive science and neural networks. Overall we can say that learning is defined in terms of improving performance based on some measure. 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 consider the two main types of algorithms i.e. supervised & unsupervised learning.
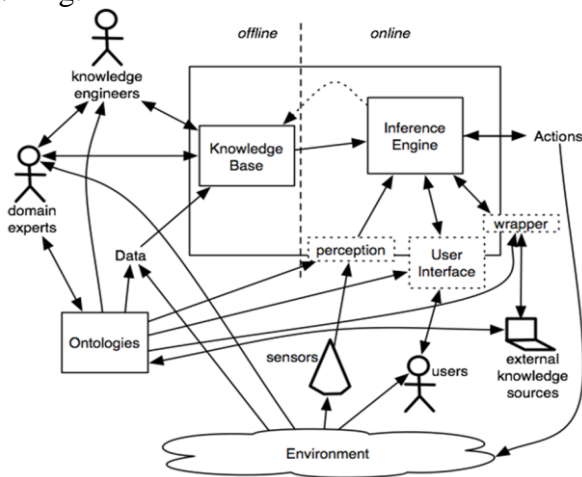

. Fig. 1 Diagram representing Machine Learning Mechanism

II. RELATED WORKS

Sally Goldman et.al [1] proposed the practical learning scenarios where we have small amount of labeled data along with a large pool of unlabeled data and presented a "co-training" strategy for using the unlabeled data to improve the standard supervised learning algorithms. She assumed that there are two different supervised learning algorithms which both output a hypothesis that defines a partition of instance space for e.g. a decision tree partitions the instance space with one equivalent class defined per tree. She finally concluded that two supervised learning algorithms can be used successfully label data for each other.

Zoubin Ghahramani et.al[2] gave a brief overview of unsupervised learning from the perspective of statistical modelling. According to him unsupervised learning can be motivated from information theoretic and Bayesian principles. He also reviewed the models in unsupervised learning. He further concluded that statistics provides a coherent framework for learning from data and for reasoning under uncertainty and also he mentioned the types of models like Graphical model which played an important role in learning systems for variety of different kinds of data.

Rich Caruana et.al [3] has studied various supervised learning methods which were introduced in last decade and provide a large-scale empirical comparison between ten supervised learning methods. These methods include: SVMs, neural nets,

logistic regression, naive bayes, memory-based learning, random forests, decision trees, bagged trees, boosted trees and boosted stumps. They also studied and examine the effect that calibrating the models through Platt Scaling and Isotonic Regression has on their performance. They had used various performance based criteria to evaluate the learning methods.

Niklas lavesson et.al [4] answered the fundamental question that how to evaluate and analyse supervised learning algorithms and classifiers. One conclusion of the analysis is that performance is often only measured in terms of accuracy, e.g., through cross-validation tests. However, some researchers have questioned the validity of using accuracy as the only performance metric. They have given a different approach for evaluation of supervised learning, i.e. Measure functions, a limitation of current measure functions is that they can only handle two-dimensional instance spaces. They present the design and implementation of a generalized multi-dimensional measure function and demonstrate its use through a set of experiments. The results indicate that there are cases for which measure functions may be able to capture aspects of performance that cannot be captured by cross-validation tests. Finally, they investigate the impact of learning algorithm parameter tuning.

Yugowati Praharsi et.al[5] had taken three supervised learning methods such as k-nearest neighbour (k-NN), support vector data description (SVDD) and support vector machine (SVM) , as they do not suffer from the problem of introducing a new class, and used them for Data description and Classification. The results show that feature selection based on mean information gain and a standard deviation threshold can be considered as a substitute for forward selection. This indicates that data variation using information gain is an important factor that must be considered in selecting feature subset. Finally, among eight candidate features, glucose level is the most prominent feature for diabetes detection in all classifiers and feature selection methods under consideration. Relevancy measurement in information gain can sort out the most important feature to the least significant one. It can be very useful in medical applications such as defining feature prioritisation for symptom recognition. In this way the analyze the accuracy and working of all the three methods.

III. PROBLEMS FACED IN 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 understands 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 it also affects the learning process of the machine. As the machine is dependent on what it perceives, the perception module of the machine should also focus on different types of challenges and environment which it will face, as different input can produce different outputs and the most appropriate and optimize output should be considered by the machine. Some of the common

problems faced during the learning process are as follows:-

Bias- the tendency to prefer one hypothesis over another is called a bias. Consider the agents *N* and *P*. Saying that a hypothesis is better than *N*'s or *P*'s hypothesis is not something that is obtained from the data - both *N* and *P* accurately predicts all of the data given - but is something external to the data. Without a bias, an agent will not be able to make any predictions on unseen examples. The hypotheses adopted by *P* and *N* disagree on all further examples, and, if a learning agent cannot choose some hypotheses as better, the agent will not be able to resolve this disagreement. To have any inductive process make predictions on unseen data, an agent requires a bias. What constitutes a good bias is an empirical question about which biases work best in practice; we do not imagine that either *P*'s or *N*'s biases work well in practice.

Noise-In most real-world situations, the data are not perfect. Noise exists in the data (some of the 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.

Pattern Recognition-This is another type of problem faced in machine learning process. Pattern recognition algorithms generally aim to provide a reasonable answer for all possible inputs and to perform "closest to" matching of the inputs, taking into account their statistical variations. This is different from pattern matching algorithms which match the exact values and dimensions. As algorithms have well-defined values like for mathematical models and shapes like different values for rectangle, square, circle etc.

It becomes different for machine to process those inputs which have different values e.g. Consider a ball the shape and pattern can be recognized by the machine, but now when we keep an inflated ball then the pattern would be entirely different and the machine will face problem in recognizing the pattern and the entire process comes to halt. This is the major problem faced by most of the machine learning process and algorithms..

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.

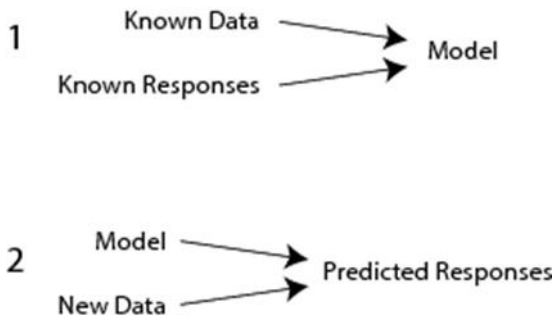| Algorithm	Predictive Accuracy	Fitting Speed	Prediction Speed	Memory Usage	Easy to Interpret	Handles Categorical Predictors
Trees	Low	Fast	Fast	Low	Yes	Yes
SVM	High	Medium	*	*	*	No
Naïve Bayes	Low	**	**	**	Yes	Yes
Nearest Neighbor	***	Fast***	Medium	High	No	Yes***
Discriminant Analysis	****	Fast	Fast	Low	Yes	No

Fig. 2 Diagram representing Supervised Learning Algorithm

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.

Characteristics of Algorithm

This table shows typical characteristics of the various supervised learning algorithms. The characteristics in any particular case can vary from the listed ones. Use the table as a guide for your initial choice of algorithms, but be aware that the table can be inaccurate for some problems.

Fig. 3 Table showing the characteristics of Supervised Learning Algorithms

Discriminant
Analysis
****Fast   Fast   Low   Yes   No

* — SVM prediction speed and memory usage are good if there are few support vectors, but can be poor if there are many support vectors. When you use a kernel function, it can be difficult to interpret how SVM classifies data, though the default linear scheme is easy to interpret.

** — Naive Bayes speed and memory usage are good for simple distributions, but can be poor for kernel distributions and large data sets.

*** — Nearest Neighbor usually has good predictions in low dimensions, but can have poor predictions in high dimensions. For linear search, Nearest Neighbor does not perform any fitting. For kd-trees, Nearest Neighbor does perform fitting. Nearest Neighbor can have either continuous or categorical predictors, but not both.

**** — Discriminant Analysis is accurate when the modeling assumptions are satisfied (multivariate normal by class). Otherwise, the predictive accuracy varies.
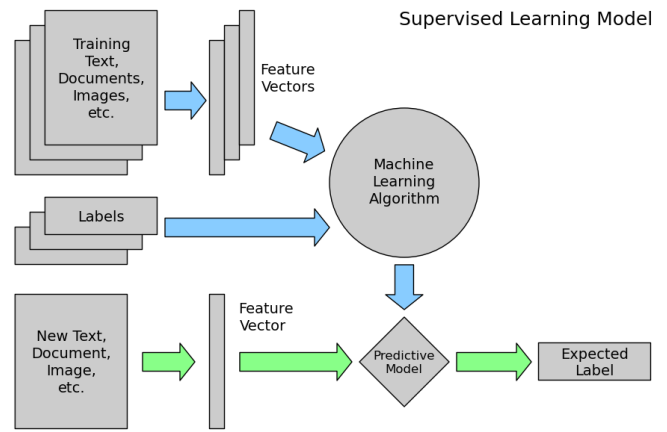

Fig. 4 This block-diagram shows the working mechanism of Supervised Learning.

V. SUPERVISED LEARNING

In unsupervised learning the machine simply receives the input x1, x2… 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 as follows:

1.   **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 a "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*

2.   **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
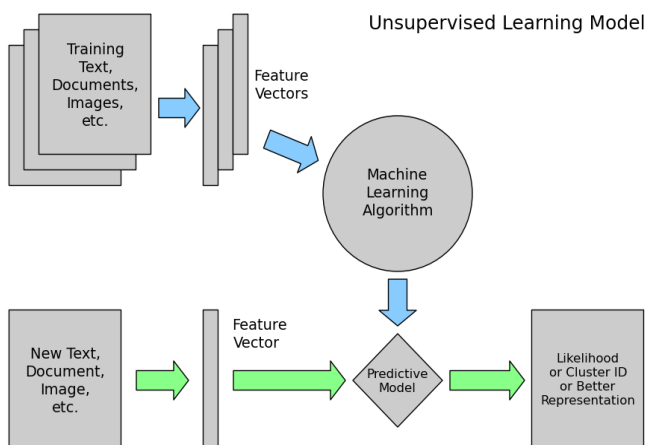13. done
14. return C

Fig. 5 This block-diagram shows the working mechanism of Supervised Learning

ANALYTICS: SUPERVISED VS UNSUPERVISED LEARNING

Machine learning algorithms are described as either 'supervised' or 'unsupervised'. The distinction is drawn from how the learner classifies data. In supervised algorithms, the classes are predetermined. These classes can be conceived of as a finite set, previously arrived at by a human. In practice, a certain segment of data will be labeled with these classifications. The machine learner's task is to search for patterns and construct mathematical models. These models then are evaluated on the basis of their predictive capacity in relation to measures of variance in the data itself. Many of the methods referenced in the documentation (decision tree induction, naive Bayes, etc) are examples of supervised learning techniques.

Unsupervised learners are not provided with classifications. In fact, the basic task of unsupervised learning is to develop classification labels automatically. Unsupervised algorithms seek out similarity between pieces of data in order to determine whether they can be characterized as forming a group. These groups are termed clusters, and there are whole families of clustering machine learning techniques.

In unsupervised classification, often known as 'cluster analysis' the machine is not told how the texts are grouped. Its task is to arrive at some grouping of the data. In a very common of cluster analysis (K-means), the machine is told in advance how many clusters it should form -- a potentially difficult and arbitrary decision to make.

It is apparent from this minimal account that the machine has much less to go on in unsupervised classification. It has to start somewhere, and its algorithms try in iterative ways to reach a stable configuration that makes sense. The results vary widely and may be completely off if the first steps are wrong. On the other hand, cluster analysis has a much greater potential for surprising you. And it has considerable corroborative power if its internal comparisons of low-level linguistic phenomena lead to groupings that make sense at a higher interpretative level or that you had suspected but deliberately withheld from the machine. Thus cluster analysis is a very promising tool for the exploration of relationships among many texts.

VI. CONCLUSIONS & FUTURE SCOPE

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. The thesis resolves some issues, e.g., by analyzing current evaluation methods and the metrics by which they measure performance, and by defining a formal framework used to describe the methods in a uniform and structured way. 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 analyze 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. Since time is often of essence when evaluating, e.g., if the evaluation method is used as a fitness function for a genetic algorithm, we have analyzed measure-based evaluation in terms of the time consumed to evaluate different classifiers. The conclusion is that the evaluation of lazy learners is slower than for eager learners, as opposed to cross-validation tests. Additionally, we have presented a method for measuring the impact that learning algorithm parameter tuning has on classifier performance using quality attributes. The results indicate that parameter tuning is often more important than the choice of algorithm. Quantitative support is provided to the assertion that some algorithms are more robust than others with respect to parameter configuration.

VII. REFERENCES

[1] Sally Goldman; Yan Zhou, "Enhancing Supervised Learning with Unlabeled Data", Department of Computer Science, Washington University, St.Louis, MO 63130 USA.

[2] Zoubin Ghahramani,"Unsupervised Learning",Gatsby Computational Neuroscience Unit, University College London, UK.

[3] Rich Caruana; Alexandru Niculescu- Mizil,"An Empirical Comparison of Supervised Learning Algorithms", Department of Computer Science, Cornell University, Ithaca, NY 14853 USA

[4] 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

[5] 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

[6] 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.

[7] Bing Liu, "Supervised Learning", Department of Computer Science,University of Illinois at Chicago (UIC), 851 S. Morgan Street, Chicago

[8]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.

[9]Rich Caruana; Alexandru Niculescu- Mizil,"An Empirical Comparison of Supervised Learning Algorithms", Department of Computer Science, Cornell University, Ithaca, NY 14853 USA
[10] Peter Norvig; Stuart Russell,"Artificial Intelligence: A Modern Approach".