International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 06 | June 2018 www.irjet.net p-ISSN: 2395-0072

Comparison between Supervised Learning and Unsupervised Learning

Bhagyashree Kolge[1], Lavina Jadhav[2]

[1]Student, Institute of Computer Science, Mumbai Education Trust, Maharashtra, India
[2]Professor, Institute of Computer Science, Mumbai Education Trust, Maharashtra, India

---

Abstract - *In the past couple of years, Machine Learning has gained huge popularity. Machine learning allows a computer system to learn with the given data and regressively improve its performance without being programmed explicitly. This paper presents a comparison between Supervised Learning Model and Unsupervised Learning Model with regards to their pattern evaluation. It has been observed that though Supervised Learning Model has proved to be very productive for solving number of real-time non-linear problems, Unsupervised Learning Model also offers efficient ways to solve and classify present study.*

Key Words: **Supervised learning; Unsupervised learning; KSOM; ANN; MLP; Simple K-Mean**

1.INTRODUCTION

An artificial neural network (ANN) learning algorithm, usually called "neural network" (NN), is an efficient information processing learning algorithm which resembles its characteristics with a biological neural network. ANNs possess large number of highly interconnected processing elements called nodes or units or neurons which usually operate in parallel. Each neuron is connected with the other by a connection link. Each connection link is associated with weights which contain information about the input signal. This information is used by the neuron net to solve a particular problem. ANN is based on three types of parameters; (a) the learning rule (supervised/ unsupervised /reinforcement etc.,); (b) its application function (like Classification model, Optimization model, Association model) and (c) its interconnection property (as feed forward network and recurrent network) [1].

These models are unique in nature and have their own importance. ANN has varied applications. Pattern classification has been the main focus of ANN. The structural design and learning methods of ANN allows it to perform classification tasks very efficiently. Degree of inference and learning ability allow to differentiate amongst different learning algorithms. Hence there is no definite way to design ANN model.

Organization of the paper is as follows:

Section 2 describes the ANN learning paradigm which includes more details of Supervised Learning and Unsupervised Learning. Section 3 describes the difference between Supervised and Unsupervised Learning based on it's type. Section 4 includes an educational experiment and its output. Section 5 describes the end result observations of the experiment. Section 6 concludes the research paper.

2. ANN LEARNING PARADIGM

Learning refers to gaining knowledge. Machine Learning allows a computer system to 'learn' with data without being explicitly programmed and denotes changes in the system that are adaptive in nature.
The main property of ANN is its capability to learn. ANN learning paradigms can be classified into 3 categories: supervised, unsupervised and reinforcement learning. In Supervised Learning model the learning task is performed with the help of a teacher, whereas, in Unsupervised learning model the task of learning is performed without the help of a teacher and Reinforcement learning learns through trial and error interactions with its environment (reward/penalty assignment).
Learning is dependent on the space of interconnection of neurons, although learning is addressed in different ways in these models. That is, unsupervised learning learns using information associated with a group of neurons, supervised learning learns by adjusting its interconnection eight combinations with the help of error signals whereas reinforcement learning modifies local weight parameters by using reinforcement function.
Learning or training is a task by which a neural network adapts itself to a stimulus by making proper parameter adjustments, resulting in the production of desired response[1]. This parameter adjustment plays important role in differentiating the learning algorithm as supervised or unsupervised models.[4]

2.1 Supervised Learning

In ANN's Supervised learning model, each input vector requires a corresponding target vector, which represents the desired output. Input vector and the target vector are together called as training pair. A precise information about what should be given as output is known to the network.

Fig. 1 gives a simple depiction of supervised learning model.

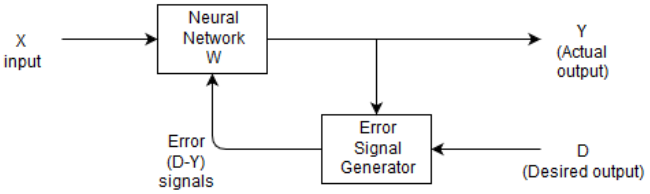

Fig-1 : Supervised Learning Model [2]

Training Process:

- The input vector presented to the network produces an output vector which is the actual output vector.
- This actual output vector is compared with the target output vector for discrepancies.
- The network generates an error signal if there exist a difference between the two output vectors.
- The network uses the error signal for adjustment of weights until the actual output matches the desired output. [2].

2.2 Unsupervised Learning

In ANN's Unsupervised learning model, the input vector of similar type is grouped together without the use of training data to specify how a member of each group looks or to which group a number belongs to.
Fig 2 depicts a simple Unsupervised learning model.

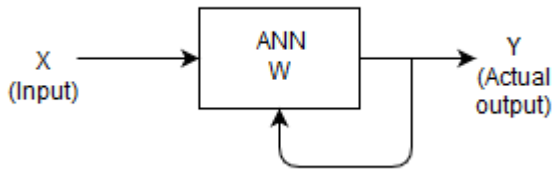

Fig- 2 : Unsupervised Learning Model [2]

Training process:

- The network receives the input patterns and organizes these patterns to form clusters.
- The neural network gives an output response when a new input pattern is applied indicating the class to which the input pattern belongs.
- If a pattern class cannot be found for an input, a new class is generated. [2]

3. TYPES

3.1 Supervised Learning Types

Classification:

Classification is the problem considering new observations and identifying to which of a set of categories (sub-populations) they belong. This is done on the basis of a training set of data containing observations (or instances) whose category membership is known.
Classification is an example of pattern recognition.
Classification is considered a type of supervised learning, i.e. a training set of correctly identified observations is present for learning.

Regression:

Regression analysis is a set of statistical processes used for estimating the relationships among variables. When the focus is on the relationship between a dependent variable and one or more independent variables (or 'predictors'), regression analysis includes many techniques for analyzing and modelling of variables.
Regression analysis is widely used for forecasting and prediction purposes. Its use has been fundamentally overlapped with the field of machine learning. Regression analysis is also used to understand which of the independent variables are related to the dependent variable, and to evaluate the forms of these relationships.

3.2 Unsupervised Learning

Clustering:

Clustering is the function of forming a group of a set of objects in such a way that objects in the same group (called a cluster) have more resemblance (in some characteristics) to each other than to those in other groups (clusters). It is mainly used for exploring data mining, and a common technique for statistical data analysis. It is also used in many fields, including pattern recognition, machine learning, image analysis, information retrieval, data compression and computer graphics.
Cluster analysis is an iterative process of interactive multi-objective optimization or knowledge discovery that involves trial and failure and not an automatic task. Until the result achieves the desired properties, it often becomes necessary to modify data pre-processing and model parameters.

Density Estimation:

Density estimation is the process of forming an estimate, by observing data, of an unobservable

underlying probability density function. The unobservable density function is thought of as the density according to which a large population is distributed; the data are usually thought of as a random sample from that population.

There are many approaches to density estimation, including a range of data clustering techniques, such as vector quantization. Rescaled histogram is the most basic form of density estimation.

4. EXPERIMENTAL OBSERVATIONS

4.1 Supervised Method

An MLP (Multilayer Perceptron) network was designed using error back propagation model. A dataset of 286 was used to create the ANN. A pattern was randomly chosen and provided to the input layer along with the bias and output at the output layer. The learning rate was set to 0.3 and the Number of Epochs to 500. Table I demonstrates the different experimentation process that was done to display the ANN design.

Table -1: Supervised Learning Observation

No. of hidden neurons	No. of Epochs	Mean squared error	Incorrectly Classified Instances	Correctly Classified Instances
3	500	0.4737	27.6224 %	72.3776 %
4	500	0.5182	31.1189 %	68.8811%

4.2 Unsupervised Method

A Simple K-Mean algorithm, which is a type of unsupervised learning, is used for working with unlabeled data. A regulated model of dataset is used to prepare the system. Euclidean distance function was used for instances comparison. The seed number was kept 10. Table 2 the experimental processes done using Simple K-Mean method.

Table – 2 Unsupervised Learning Observations

Number of cluster	Number of iterations	Sum of squared error	Incorrectly clustered instances	Correctly clustered instances
2	9	12.1	33.3333%	66.6667 %
3	6	6.9	11.3333%	88.6667%

5. RESULTS AND DISCUSSIONS

The Unsupervised learning algorithm is favored than the supervised learning algorithm since the accuracy rate of Unsupervised learning algorithm is more than Supervised learning algorithm.

The addition of hidden layers in Supervised learning would have resulted in expansion of difference in the correctness of algorithm. More over the time taken to perform the assemble the network would have increased with the increase in the number of hidden layers.

The Simple K-Mean algorithm in unsupervised learning uses single-pass learning method which makes it quick and exact rather than the MLP learning algorithm in Supervised learning with uses multiple-pass learning method.

A drawback of MLP faced during the performance of the system was that the size of the network increased with the addition of hidden layer which adds to the time required for execution.

As classification is a standout amongst most decision-making tasks of human, in our instruction circumstance, this classification may help the organization to guide the understudies and enhance their execution by appropriate consideration and training.

6. CONCLUSIONS

The designing of a classification network of a provided pattern is a form of learning from observation. This paper provides a comparison between two learning models i.e. supervised and unsupervised based on ANN. We found out supervised learning algorithm is very efficient for many non- linear real time problems and performing on datasets which contain labelled data. Simple K-Mean– the unsupervised model performs efficiently than the supervised learning algorithm.

REFERENCES

[1] L. Fu., Neural Networks in Computer Intelligence, Tata McGraw-Hill, 2003.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Principles of Soft Computing Second Edition, S.N.Sivanandam and S.N.Deepa, Wiley-INDIA.

[3] http://www.ise.bgu.ac.il/faculty/liorr/hbchap8.pdf

[4] 'Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification' by R. Sathya and Annamma Abraham.