International Journal of Engineering Research-Online
A Peer Reviewed International Journal
Email:editorijoer@gmail.com http://www.ijoer.in

Vol.4., Issue.2., 2016
(Mar-Apr)

REVIEW ARTICLE

ISSN: 2321-7758

STUDY ON VARIOUS MACHINE TECHNIQUES AND ITS APPLICATIONS

RAHUL HOODA[1], SANDEEP KUMAR[2], DEEPIKA[3]

[1]Ph.D Student, Computer Science, PEC University, Chandigarh, India
[2]Ph.D Student, Electronics & Communication, PEC University, Chandigarh, India
[3]M.tech Student, Computer Science, PACIFIC University, Udaipur Rajasthan, India

ABSTRACT

Learning and classification in data mining has gained a lot of significance in literature and it has application in different areas from astronomy to medicine and from text classification to banking. So a major focus on machine learning research is given to recognizing complex patterns and making intelligent decisions by automatic learning based on data. Machine learning, like all subjects in artificial intelligence, require cross-disciplinary proficiency in several areas, such as probability theory, statistics, pattern recognition, cognitive science, data mining, adaptive control, computational neuroscience and theoretical computer science. Machine learning is broadly divided into 3 categories: supervised, unsupervised and reinforcement learning. This paper gives a brief overview of machine learning, history, its types and its various approaches and applications used in different fields.

Keywords: Machine learning, supervised learning, data mining, artificial intelligence.

I. INTRODUCTION

Machine learning is the science of making computers to take decisions without being explicitly programmed and can also be defined as the study of computational algorithms for improving performance by making the process of knowledge acquisition from experience automatic. Machine learning focuses on the development of computer programs that can learn to grow and modify themselves when exposed to new data. Although it can be argued that machine learning has been around for long time as part of statistics and artificial intelligence but it really became a separate topic in the 1990's. Machine learning is a combination of various like artificial intelligence, statistics, biology and psychology.

Machine learning can also be defined as the branch of artificial intelligence that is pertained with the aim of developing algorithms that allow computers to change behaviours based on experimental data. It is necessary for the computers to acquire knowledge as intelligence is measured in terms of knowledge.

Machine learning is a method of analysis of data that builds analytical model automatically. Machine learning algorithms perform iterative learning from data that allows machine to find hidden information without being explicitly programmed to do so. The iterative aspect of

machine learning is crucial because models adapt independently as they are exposed to new data. Learning from previous computations produces reliable results using decisions at repeated intervals. Machine learning is so pervasive today that you probably use it various times a day without even knowing it.

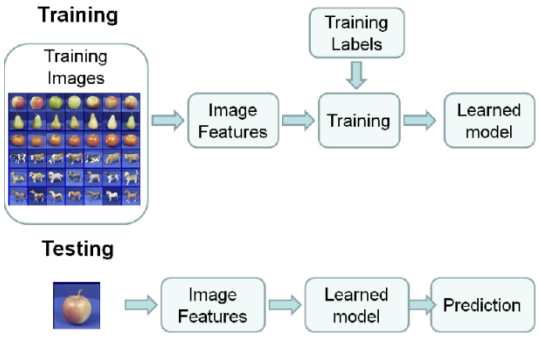

Figure 1: Process of Machine learning

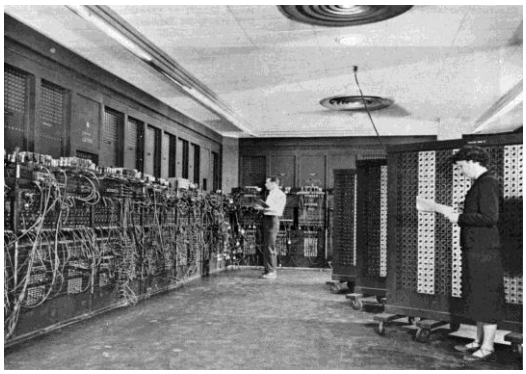

Figure 2: First fully Electronic Computer ENIAC[16]

History of machine learning goes back to 1946 when first fully electronic computer ENIAC (short for Electronic Numerical Integrator and Computer) was built by John Mauchly and John Eckart. The machine was manually operated with the idea human learning and learning could be translated logically in such computer.

In 1952, Arthur Samuel, an IBM scientist, developed the first learning program which plays checkers game. Program learned in a supervised manner by playing various game with itself and other human players. The computer program detected which moves were winning strategies and modified its programming to incorporate those strategies. In 1957, Frank Rosenblatt designed the perceptron which was inspired from the neural network. The perceptron is used to connect a web of points and simple decision are made that come together in the large program to solve more complex problem.

Figure 3: Frank Rosenblatt with Perceptron Machine[16]

In 1967 first programs were able to recognize patterns. These programs were based on nearest neighbor algorithm. In 1981, Gerald Dejong introduced explanation based learning(EBL) which uses prior knowledge of the world obtained from training examples. The program examines the training data and removes the irrelevant information to make a generalized rule to follow. In the 1990s machine learning was applied in data mining, web applications, text learning, adaptive software and language learning.

II. TYPES OF MACHINE LEARNING

Machine learning can be broadly classified into three main categories depending on the nature of learning signal available to the learning system. These are supervised learning, unsupervised learning and reinforcement learning.

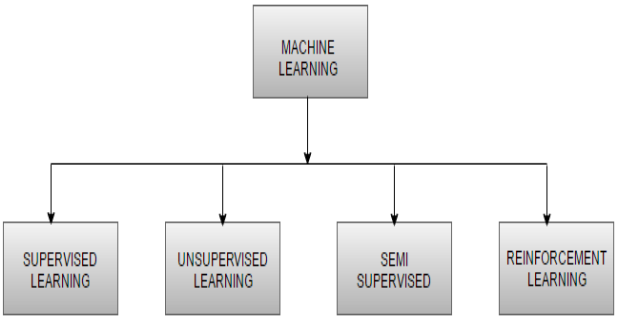

Figure 4: Types of Machine Learning

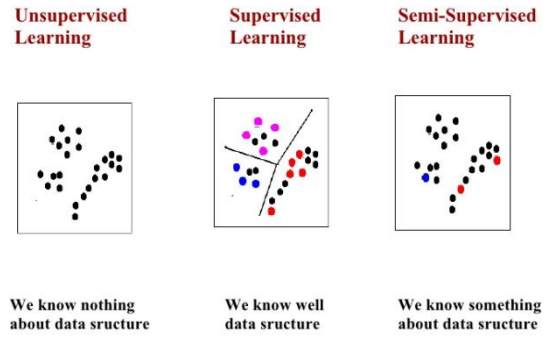

Figure 5: Types of Machine Learning

1. Supervised learning is the machine learning method of deducing a function from labelled training data. The training data contains set of examples in which the pair of input object and desired output is mentioned. This algorithm studies the training dataset and develops an inferred function which can be utilized to map new test data. Supervised learning process consists of usually 2 phases: learning phase and testing phase. In the learning phase, training data is used to learn a model with the help of some extracted features whereas in the testing phase the model is tested using unseen test data to assess the model accuracy.

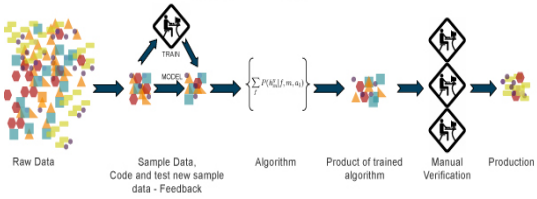

Figure 6: Pictorial representation of Supervised Learning[17]

Some of the areas of application of supervised learning methods are:

- Classification
- Regression
- Ranking

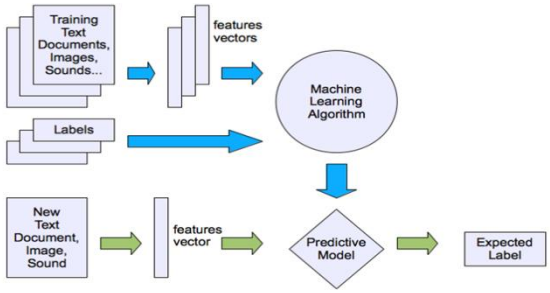

Figure 7: Flowchart of Supervised Learning[18]

2. Unsupervised learning addresses a different type of problem as the data has no labels, and we have to try to find similarities between the objects given in problem. Unsupervised learning categorizes the unlabeled data without having any given dataset for its training. In other words, you can think of unsupervised learning as a means of identifying labels from the data itself. Unsupervised learning is a very powerful tool for dissecting available data and look for trends and patterns. It is most commonly used for clustering similar input data into logical groups.

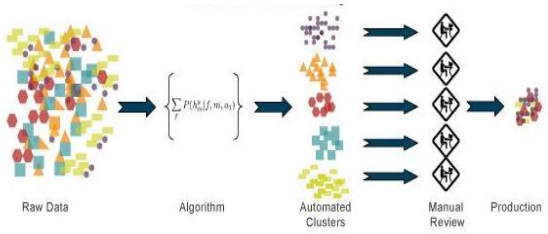

Figure 8: Pictorial view of Unsupervised Machine Learning[17]

Some of the areas of application of unsupervised learning methods are:

- Clustering
- Association Mining
- Segmentation
- Density estimation
- Dimension Reduction

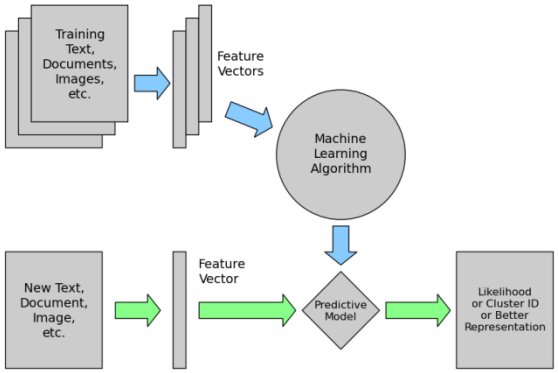

Figure 9: Flowchart of Unsupervised Machine learning[18]

3. Reinforcement learning is a category of machine learning which allows computers and software agents to automatically find out the ideal behaviour within a predefined context so that the performance is maximized. Software agent learns its behaviour using simple reward feedback and this

signal is known as the reinforcement signal. In the Reinforcement Learning, the machine or software agent learns its behaviour based on feedback provided by the environment. This behaviour can be learnt at one time only, or keep on adapting as time passes by. The ideal behaviour that maximises the reward for Reinforcement Learning algorithms is to reach the global optimum which it can attain if problem is modelled with care. Reinforcement learning is different from supervised learning as correct input/output pairs are never given and sub-optimal actions are not explicitly corrected. Also, in reinforcement learning focus is on on-line performance, which is all about finding the correct balance between exploitation of current knowledge and exploration of uncharted territory.

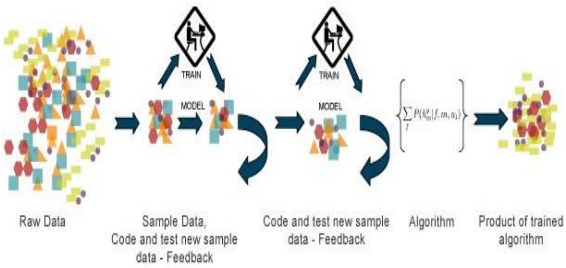

Figure 10: Pictorial View of Reinforcement Learning[17]

Some of the areas of application of reinforcement learning methods are:

- Decision Process
- Reward System
- Recommendation
- Systems

4. Semi-supervised learning is a learning process in which input data is a mixture of labelled as well as unlabelled examples. This process is a problem based on desired prediction but the model must learn the structures to organize the data and subsequently make predictions. Example problems are regression and classification. Algorithms based on this method are versions of other flexible methods that make assumptions about the way in which the unlabelled data will be modelled.

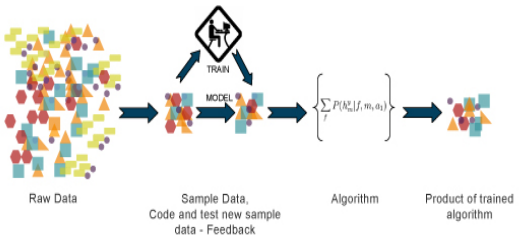

Figure 11: Pictorial View of Semi Supervised Learning [17]

III. Various type of Algorithm Used in Machine Learning:

1. Decision trees: 14 classifiers.
2. Rule-based methods: 12 classifiers.
3. Neural networks: 21 classifiers
4. Deep Learning
5. Inductive logic programming: 3 classifiers.
6. Support vector machines: 10 classifiers
7. Clustering
8. Bayesian approaches: 6 classifiers
9. Discriminant analysis: 20 classifiers
10. Boosting: 20 classifiers
11. Bagging: 24 classifiers
12. Stacking: 2 classifiers.
13. Random Forests: 8 classifiers.
14. Other ensembles: 11 classifiers.
15. Nearest neighbor methods: 5 classifiers.
16. Genetic Algorithms: 4 classifiers
17. Other Methods: 10 classifiers.

Decision tree learning: It is defined as a predictive model which uses decision tree and helps in mapping observations about an item to the findings about the item's target value.

Rule-based learning: It is a method to store and manipulate knowledge in large databases to discover interesting relations between variables.

Artificial neural networks (ANN): An ANN learning algorithm, which is also known as neural network, is an algorithm that is inspired by the functional as well as structural aspects of biological neural networks. Modern neural networks are data modeling tools which uses non-linear data statistics. They are basically used to model complex relationships between inputs and outputs by finding patterns in the data or by capture the statistical structure between known variables.

Deep Learning: Development of the concept of deep learning can be attributed to the consistently falling hardware prices and the increase in usage of GPUs in computers in the last few years. Deep learning can be defined as an ANN which consists of multiple hidden layers. This technique tries to model the way the human brain works to process light and sound and uses them for vision and hearing. Deep learning concept is being successfully used in areas of computer vision and speech recognition.

Inductive logic programming (ILP): ILP is an approach which uses logic programming to represent input examples, background knowledge, and hypotheses and to develop predicate description. This approach can be used with any programming language to derive hypotheses and not only logic programming.

Support vector machines (SVM): SVMs are supervised learning models with related algorithms that can analyze the data used for classification and regression. In the training par with the help of set of examples, SVM builds a prediction model that concludes whether the testing object falls into one category or the other.

Clustering: Cluster analysis is done to assign a group of observations into different subsets (also called clusters) so that observations within the same cluster satisfy a pre-designated criterion or criteria, while observations drawn from different clusters are doesn't meet those criteria. Cluster analysis is not a single algorithm and it can be solved using different algorithm suitable for that particular problem. Clustering is a unsupervised learning method and a common technique used for statistical data analysis.

Bayesian networks: A Bayesian network, which is also called belief network or directed acyclic graphical model belong to the family of probabilistic graphical model that represents a set of random variables and their conditional independencies using a directed acyclic graph (DAG). These networks perform inference and learning using some existing efficient algorithms.

Discriminant Analysis: It is a technique to find linear combination of features that separates two or more classes of objects. It is is a supervised technique of feature extraction used to find a linear combination of the available features which separate the classes. The main objective is to reduce the dimension of data, in order to reduce the computational cost of classification.

Boosting: Boosting refers to a family of algorithms used for primarily reducing bias and variance in supervised learning, and which convert weak learners to strong ones. A weak learner is a classifier which is correlated with the true classification but only slightly whereas a strong learner is a classifier that is well-correlated with the true classification.

Bagging: Bagging stands for Bootstrap Aggregation. It is the way to improve the stability and accuracy of your prediction by generating additional data for training. Additional training data is generated from the original dataset by producing multiset data of same cardinality as original data using combinations and repetitions of data. By increasing the size of the training set you can decrease the variance but can't improve predictive force of the model, and hence tune expected outcome prediction.

Stacking: Stacking, also called stacked generalization, is similar to boosting and in this several models are applied to the original data. In this method prediction of several other learning methods are combined and combined learning algorithm is trained on the above data. Then, combined algorithm makes a final prediction. Stacking performs better than all other single algorithms used for its training. It can be used on supervised as well as unsupervised learning tasks.

Random forests: A random forest is a group of decision trees which will create different decision tree at training time and output a mean prediction value of individual trees. Each decision tree is constructed by using a random subset of the training data. After training of the forest is done, each test row can be passed to output a prediction.

Nearest Neighbor: In this method whenever a new point needs to be classified then a predefined number of training samples closest in distance to new point are located and the label is predicted from the label of those samples. The number of samples which are tested can be a user-defined constant as in case of k-nearest neighbor learning,

or vary based on the local density of points as in radius-based neighbor learning.

Genetic algorithms (GA): A genetic algorithm (GA) is a search heuristic that mimics the process of natural selection. In machine learning, genetic algorithms belong to evolutionary algorithm class. This technique uses techniques like inheritance, selection, mutation and cross-over which are inspired by natural evolution and generate solutions to optimization problems.

IV. FIELD OF APPLICATIONS

1. Affective computing
2. Bioinformatics
3. Brain-machine interfaces
4. Chem-informatics
5. Classifying DNA sequences
6. Computational finance
7. Computer vision, including object recognition
8. Detecting credit card fraud
9. Game playing
10. Information retrieval
11. Internet fraud detection
12. Machine perception
13. Medical diagnosis
14. Marketing
15. Natural language processing
16. Online advertising
17. Optimization and metaheuristic
18. Recommender systems
19. Robot locomotion
20. Search engines
21. Sentiment analysis
22. Sequence mining
23. Software engineering
24. Speech and handwriting recognition
25. Stock market analysis
26. Structural health monitoring
27. Syntactic pattern recognition
28. Economics and Finance

V. Conclusion

This paper provides a review of machine learning approaches and presents different techniques used. This paper describes best known machine learning types and techniques in relative detail. Machine learning offers a plethora of useful ways to approach problems that otherwise defy manual solution. It should be noted that our paper lists only the publications representing the key ideas related to machine learning and not all the publications on that topic. The paper shows that the key question when dealing with machine learning classification is not whether a learning algorithm is superior to others, but under which conditions a particular method can significantly outperform others on a given application problem.

References

[1]. Menaka, S., and N. Radha. "Text classification using keyword extraction technique." *International Journal of Advanced Research in Computer Science and Software Engineering* 3.12 (2013).

[2]. Ikonomakis, M., S. Kotsiantis, and V. Tampakas. "Text classification using machine learning techniques." *WSEAS Transactions on Computers* 4.8 (2005): 966-974.

[3]. Baharudin, Baharum, Lam Hong Lee, and Khairullah Khan. "A review of machine learning algorithms for text-documents classification." *Journal of advances in information technology* 1.1 (2010): 4-20.

[4]. Rambhajani, Madhura, Wyomesh Deepanker, and Neelam Pathak. "A SURVEY ON IMPLEMENTATION OF MACHINE LEARNING TECHNIQUES FOR DERMATOLOGY DISEASES CLASSIFICATION." *International Journal of Advances in Engineering & Technology* 8.2 (2015): 194.

[5]. Hormozi, Hadi, Elham Hormozi, and Hamed Rahimi Nohooji. "The classification of the applicable machine learning methods in robot manipulators." *International Journal of Machine Learning and Computing* 2.5 (2012): 560.

[6]. Pennacchiotti, Marco, and Ana-Maria Popescu. "A Machine Learning Approach to Twitter User Classification." *ICWSM* 11.1 (2011): 281-288.

[7]. kumar Sen, Sanjaya, and Subhendu kumar Pani. "Application of Machine Learning Algorithm in Indian Stock Market Data."

[8]. Soni, Sneha, and Shailendra Shrivastava. "Classification of Indian Stock Market Data Using Machine Learning Algorithms." *International Journal on Computer Science and Engineering, ISSN* (2010): 0975-3397.

[9]. Kalaivani, P., and K. L. Shunmuganathan. "Sentiment classification of movie reviews by supervised machine learning approaches." *Indian Journal of Computer Science and Engineering* 4.4 (2013): 285-292.

[10]. Danso, Samuel, Eric Atwell, and Owen Johnson. "A comparative study of machine learning methods for verbal autopsy text classification." *arXiv preprint arXiv:1402.4380* (2014).

[11]. Bhavsar, Hetal, and Amit Ganatra. "A comparative study of training algorithms for supervised machine learning." *International Journal of Soft Computing and Engineering (IJSCE)* 2.4 (2012): 2231-2307.

[12]. Singh, Yogesh, Pradeep Kumar Bhatia, and Omprakash Sangwan. "A review of studies on machine learning techniques." *International Journal of Computer Science and Security* 1.1 (2007): 70-84.

[13]. Chaudhary, Archana, Savita Kolhe, and Raj Kamal. "Machine Learning Classification Techniques: A Comparative Study." *International Journal on Advanced Computer Theory and Engineering (IJACTE) ISSN ISSN (Print)*(2013): 2319-2526.

[14]. Trivedi, Deepika, Seema Bissa Harsh, and Rajesh Kumar. "A Brief STUDIES on Machine Learning Techniques."

[15]. Anto, S. "Supervised Machine Learning Approaches for Medical Data Set Classification-A Review 1." (2011).

[16]. https://en.wikipedia.org.

[17]. http://us.hudson.com/blog/tag/predictive-analytics

[18]. http://www.astroml.org/sklearn_tutorial/general_concepts.html”