A SURVEY ON MACHINE LEARNING ALGORITHMS, TECHNIQUES AND APPLICATIONS

Dr. C. Thiyagarajan[1], S. Shylaja[2]

[1]*Assistant professor, Department of computer science, PSG College of Arts And Science, Coimbatore, India*
[2]*Research scholar, Department of computer science, PSG College of Arts And Science Coimbatore, India*

Abstract: Machine Learning is an application of artificial intelligence that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. The main idea of any machine learning application is that we have dataset about any topic we try to make prediction for it and apply this data set on machine learning algorithm to get intelligence app. This paper is about machine learning and its Algorithms and Application covering supervised and unsupervised learning as well as reinforcement learning .In supervised learning we present Decision tree, Naïve base, Support vector. While in unsupervised learning we present principal component analysis and k-Means, and other Algorithms types.

Keywords: **Machine Learning, Algorithms, supervised learning, unsupervised learning**

I. INTRODUCTION

Machine Learning is a latest buzzword floating around. It deserves to, as it is one of the most interesting subfield of Computer Science .The term Machine Learning was coined by Arthur Samuel in 1959, an American pioneer in the field of computer gaming and artificial intelligence and stated that "it gives computers the ability to learn without being explicitly programmed" .Machine Learning focuses on the development of computer programs that can access data and use it learn for themselves**.** Many studies have been done on how to make machines learn by themselves .Many mathematicians and programmers apply several approaches to find the solution of this problem. Some of them are demonstrated here.

1.1 Supervised Learning:

The supervised machine learning algorithms are those algorithms which needs external assistance. The inputdataset is divided into train and test dataset.

The train dataset has output variable which needs to be predicted or classified. All algorithms learn some kind of patterns from the training dataset and apply them to the test dataset for prediction or classification.

1.2 Unsupervised Learning:

The unsupervised learning algorithms learns few features from the data. When new data is introduced, it uses the previously learned features to recognize the class of the data. It is mainly used for clustering and feature reduction.

1.3 Semi - Supervised Learning:

Semi – supervised learning algorithms is a technique which combines the power of both supervised and unsupervised learning. It can be fruit- full in those areas of machine learning and data mining where the unlabeled data is already present and getting the labeled data is a tedious process.

1.4 Reinforcement Learning:

Reinforcement learning is a type of learning which makes decisions based on which actions to take such that the outcome is more positive. The learner has no knowledge which actions to take until it's been given a situation. The action which is taken by the learner may affect situations and their actions in the future. Reinforcement learning solely depends on two criteria: trial and error search and delayed outcome.

1.5 Multitask Learning

Multitask learning has a simple goal of helping other learners to perform better. When multitask learning algorithms are applied on a task, it remembers the procedure how it solved the problem or how it reaches to the particular conclusion. The algorithm then uses these steps to find the solution of other similar problem or task. This helping of one algorithm to another can also be termed as inductive transfer mechanism. If the learners share their experience with each other, the learners can learn concurrently rather than individually and can be muchfaster.

1.6 Ensemble Learning

When various individual learners are combined to form only one learner then that particular type of learning is called ensemble learning. The individual learner may be Naïve Bayes, decision tree, neural network, etc. Ensemble learning is a hot topic since 1990s. It has been observed that, a collection of learners is almost always better at doing a particular job rather than individuallearners.

1.7 Neural Network Learning:

The neural network (or artificial neural network or ANN) is derived from the biological concept of neurons. A neuron is a cell like structure in a brain. To understand neural network, one must understand how a neuron works. A neuron has mainly four parts. They are dendrites, nucleus, soma and axon. The dendrites receive electrical signals. Soma processes the electrical signal. The output of the process is carried by the axon to the dendrite terminals where the output is sent to next neuron. The nucleus is the heart of the neuron. The inter- connection of neuron is called neural network where electrical impulses travel around the brain. An artificial neural network behaves the same way. It works on three layers. The input layer takes input (much like dendrites). The hidden layer processes the input (like soma and axon). Finally, the output layer sends the calculated output (like dendrite terminals).

1.8 Instance-Based Learning:

In instance-based learning, the learner learns a particular type of pattern. It tries to apply the same pattern to the newly fed data. Hence the name instance-based. It is a type of lazy learner which waits for the test data to arrive and then act on it together with training data. The complexity of the learning algorithm increases with the size of the data.

2. ALGORITHMS GROUPED BY SIMILARITY:

2.1 Regression Algorithms:

Regression analysis is part of predictive analytics and exploits the co-relation between dependent (target) and independent variables. The notable regression models are Linear Regression, Logistic Regression, Stepwise Regression, Ordinary Least Squares Regression (OLSR), Multivariate Adaptive Regression Splines (MARS) , Locally Estimated Scatterplot Smoothing (LOESS) etc.

2.2 Instance-based Algorithms:

Instance-based or memory-based learning model stores instances of training data instead of developing an precise definition of target function. Whenever a new problem or example is encountered, it is examined in accordance with the stored instances in order to determine or predict the target function value.It can simply replace a stored instance by a new one if that is a better fit than the former. Due to this, they are also known as winner-take-all method. Examples: K-Nearest Neighbour (KNN), Learning Vector Quantization (LVQ), Self-Organising Map (SOM), Locally Weighted Learning (LWL) etc.

2.3 Regularization Algorithm:

Regularization is simply the process of counteracting over fittingor abate the outliers. Regularization is just a simple yet powerful modification that is augmented with other existing ML models typically Regressive Models. It smoothes up the regression line by castigatingany bent of the curve that tries to match the outliers. Examples: Ridge Regression, Least Absolute Shrinkage and Selection Operator (LASSO) , Elastic Net, Least-Angle Regression (LARS) etc.

2.4 Decision Tree Algorithms:

A decision tree constructs atree like structure involving of possible solutions to a problem based on certain constraints. It is so namedfor it begins with a single simple decision or root, which then forks off into a number of branches until a decision or prediction is made, forming a tree. They are favored for its ability to formalize the problem in hand process that in turn helps identifying potential solutions faster and more accurately than others. Examples: Classification and Regression Tree (CART), Iterative Dichotomiser 3 (ID3), C4.5 and C5.0, Chi-squared Automatic Interaction Detection (CHAID) , Decision Stump, M5, Conditional Decision Trees etc.

2.5 Bayesian Algorithms:

A group of ML algorithms employ Bayes' Theorem to solve classification and regression problems. Examples: Naive Bayes, Gaussian Naive Bayes, Multinomial Naive Bayes, Averaged One- Dependence Estimators (AODE), Bayesian Belief Network (BBN), Bayesian Network (BN) etc.

2.6 Support Vector Machine (SVM):

SVM is so popular a ML technique that it can be a group of its own. Ituses a separating hyperplane or a decision plane todemarcate decision boundaries among a set of data pointsclassified with different labels. It is a strictly supervised classification algorithm. In other words, the algorithm develops an optimal hyperplane utilizing input data or training data and this decision plane in turnscategories new examples. Based on the kernel in use, SVM can perform both linear and nonlinear classification.

2.7 Clustering Algorithms:

Clustering is concerned with using ingrained pattern in datasets to classify and label the data accordingly. Examples: K-Means, K-Medians, Affinity Propagation, Spectral Clustering, Ward hierarchical clustering, Agglomerative clustering. DBSCAN, Gaussian Mixtures, Birch, Mean Shift, Expectation Maximization (EM)etc.

2.8 Association Rule Learning Algorithms:

Association rules help discover correlation between apparently unassociated data. They are widely used by ecommerce websites to predict customer behaviors and future needs to promote certain appealing products to him. Examples: Apriori algorithm, Eclat algorithmetc.

2.9 Artificial Neural Network (ANN) Algorithms:

A model based on the built and operations of actual neural networks of humans or animals. ANNs are regarded as non-linear modelsas it tries to discover complex associations between input and output data. But it draws sample from data rather than considering the entire set and thereby reducing cost and time. Examples: Perceptron, BackPropagation, Hop-field Network, Radial Basis Function Network (RBFN)etc.

2.10 Deep Learning Algorithms:

These are more modernized versions of ANNs that capitalizeon the profuse supply of data today. They are utiliseslarger neural networks to solve semi-supervised problems where major portion of an abound data is unlabelled or notclassified. Examples: Deep Boltzmann Machine (DBM), Deep Belief Networks (DBN), Convolutional Neural Network (CNN), Stacked Auto-Encoders etc.

2.11 Dimensionality Reduction Algorithm:

Dimensionality reduction is typically employed to reduce a larger data set to its most discriminative components to contain relevant information and describe it with fewer features. This gives a proper visualization for data with numerous features or of high dimensionality and helps in implementing supervised classification more efficiently. Examples: Principal Component Analysis (PCA), Principal Component Regression (PCR), Partial Least Squares Regression (PLSR), Sammon Mapping, Multidimensional Scaling (MDS), Projection Pursuit, Linear Discriminant Analysis (LDA), Mixture Discriminant Analysis (MDA), Quadratic Discriminant Analysis (QDA), Flexible Discriminant Analysis (FDA) etc.

2.12 Ensemble Algorithms:

The main purpose of an ensemble method is to integrate the projections of several weaker estimators that are singly trained in order to boost up or enhance generalizability or robustness over a single estimator. The type of learners and the means to incorporate them is carefully chosen as to maximize the accuracy. Examples: Boosting, Bootstrapped Aggregation (Bagging), AdaBoost, Stacked Generalization (blending), Gradient Boosting Machines (GBM), Gradient Boosted Regression Trees (GBRT), Random Forest, Extremely Randomized Trees etc.

3. MEASURING AND COMPARING PERFORMANCES OF POPULARML ALGORITHMS

Though various researchers have contributed to ML and numerous algorithms and techniques have been introduced as mentioned earlier, if it is closely studied most of the practical ML approach includes three main supervised algorithm or their variant. These three are namely, Naive Bayes, Support Vector Machine and Decision Tree. Majority of researchers have utilized the concept of these three, be it directly or with a boosting algorithm to enhance the efficiency further. These three algorithms are discussed briefly in the following section.

Naive Bayes Classifier:

It is a supervised classification method developed using Bayes' Theorem of conditional probability with a 'Naive' assumption that every pair of feature is mutually independent. That is, in simpler words, presence of a feature is not effected by presence of another by any means. Irrespective of this over-simplified assumption, NB classifiers performed quite well in many practical situations, like in text classification and spam detection. Only a small amount of training data is need to estimate certain parameters. Beside, NB classifiers have considerably outperformed even highly advanced classificationtechniques.

Support Vector Machine:

SVM, another supervised classification algorithm proposed by Vapnik in 1960s have recently attracted an major attention of researchers. The simple geometrical explanation of this approach involves determining an optimal separating plane or hyperplane that separates the two classes or clusters of data points justly and is equidistant from both of them. SVM was defined at first for linear distribution of data points. Later, the kernel function was introduced to tackle nonlinear datas as well.

Decision Tree:

A classification tree, popularly known as decision tree is one of the most successful supervised learning algorithm. It constructs a graph or tree that employs branching technique to demonstrate every probable result of a decision. In a decision tree representation, every internal node tests a feature, each branch corresponds to outcome of the parent node and every leaf finally assigns the class label. To classify an instance, a top-down approach is applied starting at the root of the tree. For a certain feature or node, the branch concurring to the value of the data point for that attribute is considered till a leaf is reached or a label is decided. Now, the performances of these three were roughly compared using a set of tweets with labels positive, negative and neutral. The raw tweets were taken from Sentiment140 data set. Then those are pre-processed and labeled using a python program. Each of these classifier were exposed to same data. Same algorithm of feature selection, dimensionality reduction and k-fold validation were employed in each cases. The algorithms were compared based on the training time, prediction time and accuracy of the prediction.

4. Comparative study

Author	Description	Pros and cons
L. Collingwood, T. Jurka, A.E. Boydstun, E. Grossman,	The process of algorithm learning from the training dataset can be thought of as a teacher upervising the learning process.	Supervised learning can be very helpful in classification problems. Supervised learning cannot give you unknown information from the training data like unsupervised learning do.

Kajaree Das1, Rabi Narayan Behera2	Unsupervised learning is the training of an artificial intelligence (ai) algorithm using information that is neither classified nor labeled and allowing the algorithm to act on that information withoutguidance.	In **unsupervised learning**, no one is required to understand and then to label the data inputs. We cannot get precise information regarding data sorting, and the output as data used in unsupervised learning is labeled
Ayon Dey	Machine learning focus on the development of computer programs that can access data and use it learn for themselves.	A key machine learning benefit concerns this technology's ability to review large volumes of data and identify patterns and trends that might not be apparent to a human. **High Level of Error Susceptibility.** An error can cause havoc within a machine learning interface, as all events subsequent to the error may be flawed, skewed or just plain undesirable
Paras Lakhani , BaskaranSundaram	Deep learning is a subset of machine learning that relies on neural networks that are modeled on the intricacies of the human brain and how it operates.	Reduces the need for feature engineering, one of the most time- consuming parts of machine learning practice. Expensive and intensive training

5. CONCLUSION

The fore most target of ML researchers is to design more efficient (in terms of both time and space)and practical general purpose learning methods that can perform better over a widespread domain. In the context of ML, the efficiency with which a method utilises data resources that is also an important performance paradigm along with time and space complexity. Higher accuracy of prediction and humanly interpretable prediction rules are also of high importance. Being completely data-driven and having the ability to examine a large amount of data in smaller intervals of time, ML algorithms has an edge over manual or direct programming. Also they are often more accurate and not prone to human bias. Consider, speech recognition software that has to be customised according to the needs of the customer. Like e-commerce sites that customises the products displayed according to customers or email reader that enables spam detection as per user preferences. Direct programming lacks the ability to adapt when exposed to different environment.

ML provides a software the flexibility and adaptability when necessary. In spite of some application (e.g., to write matrix multiplication programs) where ML may fail to be beneficial, with increase of data resources and increasing demand in personalised customisable software, ML will thrive in near future. Besides software development, MLwill probably but help reform the general outlook of Computer Science. By changing the defining question from "how to program a computer" to "how to empower it to program itself," ML priories the development of devices that are self- monitoring, self-diagnosing and self-repairing, and the utilises of the data flow available within the program rather than just processing it. Likewise, it will help reform Statistical rules, by providing more computational stance. Obviously, both Statistics and Computer Science will also embellish ML as they develop and contribute more advanced theories to modify the way oflearning.

REFERENCES:

[1].Thiyagarajan C, K. Anandha Kumar and , A. Bharathi "A Survey on Diabetes Mellitus Prediction Using Machine Learning Techniques" International Journal of Applied Engineering Research ISSN 0973- 4562 Volume 11, Number 3 (2016) pp1810-1814.

[2] .PD Stachour and B M Thuraisingham Design of LDV A multilevel secure relational database management system, IEEETrans. Knowledge and Data Eng., Volume 2, Issue 2, 190 - 209, 1990.

[3]. T. M. Mitchell, Machine Learning, McGraw-Hill International, 1997.

[4]. Kalaiselvi, C., and G. M. Nasira. "Classification and Prediction of Heart Disease from Diabetes Patients using Hybrid Particle Swarm Optimization and Library Support Vector MachineAlgorithm."

[5]. Rakesh Agrawal, Ramakrishnan Srikant, Privacy Preserving Data Mining, SIGMOD '00 Proceedings of the 2000 ACM SIGMOD international conference on Management of data, Volume 29 Issue 2,Pages 439-450,2000.

[6]. D. Michie, D. J. (1994). Machine Learning, Neural and Statistical Classification. Prentice Hall Inc. Fausett, L. (19994). Fundamentals of Neural Networks. New York: PrenticeHall.