ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 1, January 2017

Machine Learning Algorithms

[1] Mayank Kumar,

[1] Department of Electronics and Communication Engineering, Galgotias University, Yamuna Expressway Greater Noida, Uttar Pradesh

[1] mayank.kumar@Galgotiasuniversity.edu.in

ABSTRACT: Machine learning is used to teach machines how to more effectively handle the data. Machine learning is predominantly an area of Artificial Intelligence, a key component of digitalization solutions that has attracted a great deal of attention in the digital arena. In this paper author intends to make a brief review of various machine learning algorithms that are used most frequently and are therefore the most popular. Through their application perspective, the author aims to highlight the merits and demerits of the machine learning algorithms to assist in informed decision-making to choose the correct learning algorithm to fulfill the application's specific requirements Specific machine learning algorithms were discussed in this article. For various purposes such as data mining, image processing, predictive analytics, etc., these algorithms are used to name a few. The main advantage of using machine learning is that it can do its work automatically once an algorithm knows what to do with data.

KEYWORDS: Algorithm, Decision Tree, Regression, Machine Learning, Support Vector Machine.

INTRODUCTION

Starting with the fundamental concept of machine learning[1] will be a good starting point for this paper. A computer program is assigned to perform certain tasks in Machine Learning and it is said that the machine has benefited from its experience if its observable output in these tasks is improving as it gains more and more experience in performing these tasks. Therefore the computer takes decisions and makes data-based predictions / projections. Take the example of a computer program that learns from a patient's medical research reports to detect / predict cancer. It will improve performance as it gathers more experience by analyzing reports of medical research from a wider patient population this success will be calculated by an experienced oncologist's count of correct predictions and detections of cancer cases. Machine learning is implemented in a wide range of fields: robotics, automated personal assistants (such as Google), computer games, pattern recognition[2], natural language processing, data mining[3], prediction of traffic, Online transport network (e.g. peak hour estimation by Uber app), product recommendation, market share prediction, medical diagnosis, online fraud prediction, farm advisory, search engine refining (e.g. Google search engine), BoTs (online customer support Chabot's), e-mail spam filtering, video surveillance device crime prediction, social media services (face recognition in Facebook).

Machine Learning generally discusses three types of issues: classification, regression, and clustering. Depending on the availability of training data types and categories, it may be appropriate to choose from the "supervised learning" techniques available, to implement the correct machine learning algorithm, "unsupervised learning," "semi-supervised learning" and "reinforcement learning." Some of the most widely used machine learning algorithms will be discussed in the next few pages.

Machine learning is about learning from the results. There have been several studies on how to make machines learn on their own. Many mathematicians and programmers use several methods to solve this problem. Some of these are shown in Figure 1.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 1, January 2017

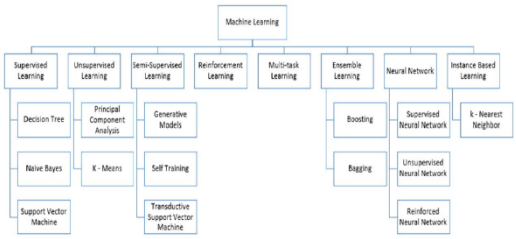

Fig.1: Types of Learning

1. Supervised Learning:

The supervised algorithms of machine learning are those algorithms that require external help. The data set for input is separated into the data set for train and test. The train data set has a variable output that must be projected or categorized. Both algorithms learn from the training dataset some patterns and add them to the prediction or classification test dataset. Here was discussed three of the most popular supervised machine learning algorithms.

1.1 Decision Tree:

Decision trees[4] are the types of trees that group attributes based on their values by sorting them. The decision tree is used mainly for the purpose of classification. Every tree is made up of branches and nodes. In a group to be categorized, each node represents attributes and each branch represents a value the node will take. Figure 2 shows an example of a decision tree.

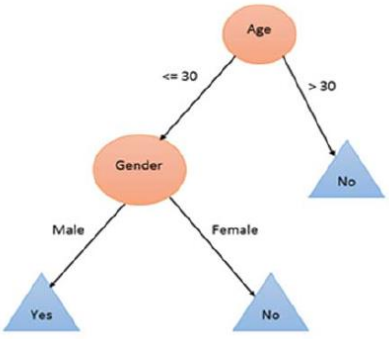

Fig.2: Decision Tree

1.2 Naïve Bayes:

Naïve Bayes[5] is mainly aimed at the industry of text classification. It is used mainly for the purpose of clustering and classification. Naïve Bayes ' underlying design relies on the conditional probability. This produces trees based on their probability of occurrence. Often known as the Bayesian Network, such trees.

1.3 Support Vector Machine:

Support Vector Machine (SVM)[6] is another state-of - the-art machine learning technique that is commonly used. It is mainly used to classify. SVM operates on the measurement of margins principle. This essentially draws the classes ' margins. The margins are drawn in such a way as to minimize the classification error by minimizing the distance between the margin and groups. Figure 3 provides an example of how SVM works.

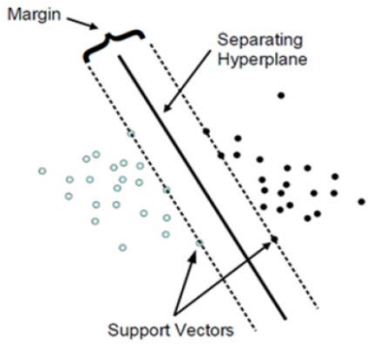

Fig.3: Working of Support Vector Machine

2. Unsupervised Learning:

The unattended learning algorithms learn from the data a few features. By introducing new data, it uses the features previously learned to understand the data type. It is used mainly for clustering and reduction of apps. An example of unsupervised learning workflow is shown in Figure 4.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 1, January 2017

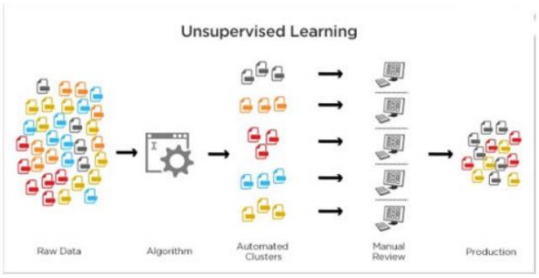

Fig.4: Example of Unsupervised Learning

3. Semi-Supervised Learning:

Semi–supervised learning algorithms is a technique that combines both supervised and unsupervised learning capabilities. In those areas of machine learning and data mining where unlabeled data is already present, it can be fruitful and getting the labeled data is a tedious process. There are a lot of semi-supervised learning categories. Some of them will be discussed below:

3.1 Generative Models:

Generative models are one of the oldest semi-supervised learning methods that assumes a structure such as $p(x, y) = p(y)\ p(x|y)$ where $p(x|y)$ is a mixed distribution, for example. Models of [7]. The mixed components can be represented in the unlabeled data. Once example labeled per part is sufficient to validate the distribution of the mixture.

3.2 Self-Training:

A classifier is trained in self-training with a portion of labeled data. The classifier is then fed with data that is not labelled. In the training set, the unlabeled points and the projected labels are added. Once again, this procedure is repeated. Hence the term self-training since the classifier is studying itself.

3.3 Transduced SVM:

A SVM extension is a transduced vector support system or TSVM. TSVM considers both the labeled and unlabeled data. It is used to mark the unlabeled data to optimize the difference between the labeled and unlabeled data. It is a NP-hard problem to find an exact solution by TSVM.

4. Reinforcement Learning:

Learning to strengthen is a type of learning that makes decisions based on what actions to take to make the outcome more positive. The learner has no knowledge of what actions to take until a situation has been given. The learner's behavior will affect future situations and their behavior.

5. Multitask Learning:

Learning multitask has a simple goal to help other learners do better. When multitask learning algorithms are applied to a task, it recalls how it solved the problem or how it reaches the specific conclusion. Then the algorithm uses these steps to find some similar problem or task solution. It can also be called this aid from one algorithm to another as an inductive transfer mechanism. If the learners share their experiences with each other, they can learn at the same time rather than individually, and they can be much faster.

6. Ensemble Learning:

If different individual learners are combined to form only one learner, the particular type of learning is called learning of the ensemble. Naïve Bayes, decision tree, neural network, etc. can be the person learner. Training for the ensemble has been a hot topic since the 1990s. It has been found that, in doing a particular job, a group of learners is almost always better than individual students.

7. Neural Network Learning

The neural network[8] (or neural artificial network or ANN) derives from the neuronal biological concept. A neuron in a brain is a cell-like structure. It is necessary to understand how a neuron functions to understand the neural network. There are primarily four sections of a neuron[9] (see Fig. 5). Its dendrites, nuclei, soma and axon.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 1, January 2017

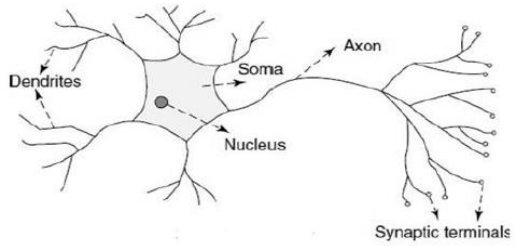

Fig.5: A Neuron

Electric signals are transmitted by the dendrites. The electrical signal[10] is interpreted by Soma. The process output is transmitted by the axon to the dendrite terminals where the output is transmitted to the next neuron. The nucleus is the neuron's heart. Neuron interconnection is called the neural network in which electrical impulses pass around the brain.

8. Instance-Based Learning

The learner learns a particular type of pattern through case-based learning. It is attempting to apply the same pattern to the newly fed results. That is why the name is based on an instance. It's a kind of lazy learner waiting for the test data to arrive and then acting on it along with training data. The learning algorithm's complexity increases with the data size. An example of instance-based learning that is k-nearest neighbor is given below.

CONCLUSION

In this paper, the most commonly used machine learning algorithms were attempted to solve classification, regression and clustering problems. The advantages, drawbacks of these algorithms were discussed in terms of efficiency, learning rate etc., along with comparison of different algorithms (wherever possible). In addition, examples of these algorithms ' practical applications have been discussed. Types of machine learning techniques have been discussed, including supervised learning, unattended learning, and semi-supervised learning. It is expected to give readers insight to make an informed decision to define the available options for machine learning algorithms and then choose the correct machine learning algorithm in the sense of specific problem solving.

REFERNCES

[1] K.Deepika, P.Andrew, R.Santhya, S.Balamurugan, S.Charanyaa, "Investigations on Methods Evolved for Protecting Sensitive Data", International Advanced Research Journal in Science, Engineering and Technology Vol 1, Issue 4, Decermber 2014.

[2] K.Deepika, P.Andrew, R.Santhya, S.Balamurugan, S.Charanyaa, "A Survey on Approaches Developed for Data Anonymization", International Advanced Research Journal in Science, Engineering and Technology Vol 1, Issue 4, December 2014.

[3] Vishal Jain, Dr. Mayank Singh, “Ontology Development and Query Retrieval using Protégé Tool”, International Journal of Intelligent Systems and Applications (IJISA), Hongkong, Vol. 5, No. 9, August 2013, page no. 67-75, having ISSN No. 2074-9058, DOI: 10.5815/ijisa.2013.09.08

[4]. Vishal Jain, Dr. S. V. A. V. Prasad, “Ontology Based Information Retrieval Model in Semantic Web: A Review”, International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE), Volume 4, Issue 8, August 2014, page no. 837 to 842 having ISSN No. 2277-128X

[5] Vishal Jain, Dr. S. V. A. V. Prasad, “Role of Ontology with Multi-Agent System in Cloud Computing”, International Journal of Sciences: Basic and Applied Research (IJSBAR), Jordan, Volume 15, No. 2, page no. 41 - 46, having ISSN No. 2307-4531

[6] R Santhya, S Balamurugan, “A Survey on Privacy Preserving Data Publishing of Numerical Sensitive Data”, International Journal of Innovative Research in Computer and Communication Engineering , Vol. 2, Issue 10, October 2014

[7] S. ichi Amari, “Machine Learning,” in *Applied Mathematical Sciences (Switzerland)*, 2016.

[8] J. O’Rourke and G. T. Toussaint, “Pattern recognition,” in *Handbook of Discrete and Computational Geometry, Third Edition*, 2017.

[9] "Introduction to data mining," *Intell. Syst. Ref. Libr.*, 2011, doi: 10.1007/978-3-642-19721-5_1.

[10] C. Bulac and A. Bulac, "Decision Trees," in *Advanced Solutions in Power Systems: HVDC, FACTS, and AI Techniques*, 2016.

[11] P. Cichosz, "Naïve Bayes classifier," in *Data Mining Algorithms*, 2015.

[12] "Tutorial on Support Vector Machine," *Appl. Comput. Math.*, 2016, doi: 10.11648/j.acm.s.2017060401.11.

[13] T. Marwala, "Gaussian Mixture Models," in *Handbook of Machine Learning*, 2018.

[14] Y. Goldberg, "A primer on neural network models for natural language processing," *J. Artif. Intell. Res.*, 2016, doi: 10.1613/jair.4992.

[15] C. Hammond, "Neurons," in *Cellular and Molecular Neurophysiology: Fourth Edition*, 2015.

[16] M. Pavlovic and B. Balint, *SpringerBriefs in Electrical and Computer Engineering*. 2012.