A Survey on Various Machine Learning Algorithms

Abhishek Singh, Dhwani Agrawal

Assistant Professor, Dept. of Computer Science & Engineering, GNIOT, Gr. Noida.

Abstract - *Apparently, we are living in the most defining and developing period of human history. This is the period where computing generation reached from large mainframes to PCs to cloud. But what makes it defining is not what happened, but what is coming our ways in future. There is no doubt that machine learning/ artificial intelligence has rapidly gained more vogue in the previous couple of years. As the hottest mania in the tech industry at present, machine learning extremely powerful to make predictions and calculated suggestions which is generally based on the very large amount of data. This paper tells about how the machine learning algorithms adaptively enhance their performances as the inputs available for learning increases.*

Keywords: *Machine learning algorithms.*

1. INTRODUCTION

In the era, where almost all manual works are being automated, the definition of manual is reshaping. Machine learning algorithms can help computers to play games, perform surgeries and get smarter and more private.

We are living in the world where technology is changing very rapidly like day-by-day. One of the main features of these transformations is how computing techniques and tools have been democratized. In the past few years, data scientist has assembled evolutionary data-crunching machines by seamlessly executing advanced techniques. The results are amazing.

Machine learning is a data analytics technique through which computers learns to do what comes naturally to humans and animals i.e. learn from experience. Machine learning algorithms use computational methods to remember information which is directly from the data without depending on a predetermined equation. The machine learning algorithms adaptively enhance their performances as the inputs available for learning increases.

There is no doubt that artificial intelligence/ machine learning has rapidly gained more popularity in the previous couple of years. As Big Data is the trending mania in the tech industry at present moment, machine learning is very strong for calculated suggestions and make predictions which is based on the large amount of data. Some of the very common and famous examples of machine learning are Netflix's algorithms to make movie suggestions based on movies you have watched in past or Amazon's algorithms that recommend books based on the books you have bought or searched before.

2. EVOLUATION OF MACHINE LEARNING

Machine learning was born from pattern recognition and the theory that computers can learn without being programmed to perform some specific tasks. But researchers who are concerned in artificial intelligence wanted to see that if computers could learn from data. The repetitive side of machine learning is important because models learn from computations to generate reliable, repeatable decisions and results.

While there are many machine learning algorithms have been around for a very long time, the capability to automatically apply complex mathematical calculations to big data is a new development.

3. TYPES OF MACHINE LEARNING ALGORITHMS

Machine learning algorithms can be broadly categorized into 3 categories: - supervised learning, unsupervised learning, and reinforcement learning. Supervised learning is very handy in the cases where a label is available for a certain training set, but needs to be predicted for other objects. Unsupervised learning is useful when the challenge is to find implicit relationships in a given unlabeled datasets. Reinforcement learning lies between these 2 categories.

3.1 SUPERVISED LEARNING

This algorithm consists of an outcome variable or dependent variable. Using these set of variables, it generates a function that maps inputs to their desired outputs. This process will go on until the model reaches a desired level of accuracy on the training data. Some examples of supervised learning are Decision Tree, Regression, Logistic Regression etc.

3.2 UNSUPERVISED LEARNING

In this type of algorithm, there is no target or outcome variable to estimate. This algorithm is used for gathering population in different groups, which is widely used for segmenting customers in different groups for specific intervention.

3.3 REINFORCEMENT LEARNING

In this algorithm, the machine is coach to make specific decision. Machine is exposed to the environment where it trains itself continually using trial and error. Then the machine learns its past experience and then tries to catch the possible best knowledge to make accurate decisions.

4. SOME MACHINE LEARNING ALGORITHMS

1. **LINEAR REGRESSION:** In this process, a relationship is established between dependent and independent variables by fitting them in a line. This line is known as regression line and represented by a linear equation.

2. **LOGISTIC REGRESSION:** Logistic regression is generally used to calculate discrete values from a set of independent variables. This helps to predict the probability of an event. It is also called logit regression.

   Here are some methods that are often used to improve logistic regression models:

   - Include interaction terms
   - Eliminate features
   - Regularize techniques
   - Use a non-linear model

3. **DECISION TREE:** This is the most famous machine learning algorithm that is being used in present scenario. This is one kind of supervised algorithm that is used for classified algorithms. It works well for classifying both categorical and continuous dependent variables.

4. **SUPPORT VECTOR MACHINE (SVM):** It is the method for classification in which points are plotted as raw data in an n-dimensional space. Then the value of each feature is tied to a particular coordinate, making it very

easy to classify the data. Lines that are used to spilt the data are called classifiers and they can used to plot the graph on them.

5. **NAÏVE BAYES:** This algorithm assumes that the presence of particular feature in a class is not related to the presence of any other feature.

   Naive Bayes classifier consider these variables independent even if they are related to each other when calculating the probability of a particular outcome.

   A Naïve Bayes model is very useful for massive datasets and easy to build. It is a simple model which is known to outperform highly sophisticated classification methods.

6. **K-NEAREST NEIGHBORS (KNN) :** This algorithm can be applied to both classification and regression problems. It is widely used in solving classification problems. It is very simple algorithm that stores all available cases and classifies any new case by taking a majority vote of its k neighbors.

   It is very easy algorithm that can be understood by comparing it to the real life. Here are some things that must be considered before selecting KNN:

   • KNN is computationally expensive.
   • Variables should be normalized otherwise higher range variables can bias the algorithm.
   • Data still needs to be pre-processed.

7. **K-MEANS:** K- Means is an unsupervised algorithm through which one can solve clustering problems. In this algorithm, data sets being classified into a particular number of clusters in such a way that within a cluster, all the data points are homogeneous and heterogeneous from the data in other clusters.
   Here are some points through which one can know how K-Means form clusters:

   - The K-Means algorithm picks k number of points for each cluster. These points are known as centroids.
   - Each data point forms a cluster with the closest centroids.
   - Then it creates new centroids, based on the existing cluster members.
   - Now with these new centroids, determine closest distance for each data points. This process is being repeated until the centroids do not change.

8. **RANDOM FOREST:** A collective of decision tree is called random forest. In this, each tree is classified and tree votes for that class and this process is done to classify a new object based on its attributes. The forest chooses the classification which has the most votes.

   Each tree is planted and grown as follows:

   - If the number of cases in the training set is N, then a sample of N cases is taken at random. This sample will be the training set for the growing tree.
   - Each tree is growing to the largest extent possible. There is no pruning.

9. **DIMENSIONALITY REDUCTION ALGORITHMS:** Vast amounts of data is being stored and analyzed by corporates in today's world.

   Dimensionality reduction algorithms remove the challenges coming in the identifying significant patterns and variables.

10. **GRADIENT BOOSTING AND ADABOOST:** These are the boosting algorithms used when enormous loads of data have to be handled in order to make predictions with high accuracy. Boosting is an ensemble learning algorithm that combines the predictive power of several base estimators to improve robustness.

    To summarize, it combines multiple weak or average predictors to build a strong predictor. These boosting algorithms always works well in data science competitions like Kaggle, AV Hackathon etc. these are the most preferred machine learning algorithms today.

REFERENCES

[1] Bateson, G. (1972). *Steps to an ecology of mind.* New York: Ballantine P. Gupta and P. R. Kumar. The capacity of Wireless Networks.

[2] Davis, L., & Coombs, S. (1987). Genetic algorithms and communication link speed design: Theoretical considerations. *Genetic Algorithms and Their Applications: Proceedings of the Second International Conference on Genetic Algorithms* (pp. 252-256). Cambridge, MA: Lawrence Erlbaum.

[3] Edelman, G. M. (1987). *Neural Darwinism: The theory of neuronal group selection.* New York: Basic Books.

[4] Fourman, M. P. (1985). Compaction of symbolic layout using genetic algorithms. *Proceedings of the First International Conference on Genetic Algorithms and Their Applications* (pp. 141-152). Pittsburgh, PA: Lawrence Erlbaum.

[5] Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning.* Reading, MA: Addison-Wesley.

[6] Grefenstette, J. J. (Ed.). (1985). *Proceedings of the First International Conference on Genetic Algorithms and Their Applications.* Pittsburgh, PA: Lawrence Erlbaum.

[7] Grefenstette, J. J. (Ed.). (1987). *Genetic Algorithms and Their Applications: Proceedings of the Second International Conference on Genetic Algorithms.* Cambridge, MA: Lawrence Erlbaum.

[8] Holland, J. H. (1962). Outline for a logical theory of adaptive systems. *Journal of the Association for Computing Machinery, 3,* 297-314.

[9] Holland, J. H. (1975). *Adaptation in natural and artificial systems.* Ann Arbor, MI: University of Michigan Press.

[10] Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction: Processes of inference, learning, and discovery.* Cambridge, MA: MIT Press.