

An Insight on Machine Learning Algorithms and its Applications

R. P. Ram Kumar, Sanjeeva Polepaka, Lazarus S F, Dasari Vamsi Krishna

Abstract— Machine Learning (ML) furnishes the ability of insights on automatic recognizing patterns and determining the prediction models for the structured and unstructured data even in the absence of explicit programming instructions. Today, the impact of Artificial Intelligence (AI) has grown up to several heights, ranging from Life sciences to the Management techniques. The integration of ML led to reduce or eliminate the errors in the prediction, classification and simulation models. The objective of the paper is to represent the ML objectives, explore the various ML techniques and algorithms with its applications in the various fields.

Keywords: Machine learning, Supervised learning, Unsupervised learning, Semi-supervised learning, Reinforcement learning,

I. INTRODUCTION

The ML is a process of making the system to learn automatically based on the earlier experimental data. The objective of ML is to determine the predictions based on the existing data. The following shows the procedure for ML adopted by Gartner [16].

Step 1: Gather the input data (may be structured or unstructured)

Step 2: Apply the ML algorithms (such as Supervised, Unsupervised, Semi-supervised, and Reinforcement Learning methods)

Step 3: Determine the Output (such as predictive, exploratory, classification)

After the data collection process, knowledge extraction or pattern determination process is imitated from the observations. ML has the following subdivisions based on the observations summarized from [17]. They are;

- Supervised learning,
- Unsupervised learning,
- Semi-supervised learning and
- Reinforcement learning.

The ML process is viewed as two phases, namely, (a) learning phase and (b) prediction phase. The training data is fed to the learning phase for predictions. In the prediction phase, the predicted results are obtained for the new data regarding the concerned trained model.

Revised Manuscript Received on August 05, 2019.

R.P. Ram Kumar, Professor, Department of CSE, Malla Reddy Engineering College (A), Secunderabad, Telangana, India.

Sanjeeva Polepaka, Associate Professor, Department of CSE, Malla Reddy Engineering College (A), Secunderabad, Telangana, India.

Lazarus S F, B. Tech., Department of CSE, Malla Reddy Engineering College (A), Secunderabad, Telangana, India.

Dasari Vamsi Krishna, B. Tech., Department of CSE, Malla Reddy Engineering College (A), Secunderabad, Telangana, India.

II. MACHINE LEARNING APPROACHES

The following section deals with the machine learning approaches.

Supervised Machine Learning (SML): SML approach's aim is to learn the mapping function 'f' from the input variable (X) to the output variable (Y) as represented in the equation (1) summarized from [18].

$$Y = f(X) \quad (1)$$

The mapping function tries to estimate during the arrival of novel input data (X) and forecast the corresponding output variable (Y). This algorithm learns from the training set and stops learning when the acceptable performance level is reached. The subdivisions of SML are classification (output variable needs categorization) and regression (output is real value). Some of the examples of SML algorithms are linear regression, random forest and Support Vector Machine (SVM).

UnSupervised Machine Learning (USML): The objective of the USML is learning, understanding and exploring the unlabeled input data (X) without historical data [19] and [17]. The USML is subdivided into association (discover rules to describe the data) and clustering (discover inherent group in data). Examples of USML include apriori and k-means methods for association and clustering problems, respectively.

Semi-Supervised Machine Learning (SSML): The SSML is considered as a mixture of SML and USML, and arises when huge unlabeled input data (X) with few labelled data (Y) [18]. That is, the mixture of very few labelled data and huge unlabeled data. The SML and USML algorithms can be applied to get knowledge and discover the data.

Reinforcement Learning (RL): The objective of RL is to map the actions according to situations to yield the maximum reward. During the mapping process, the algorithm should take in to account the immediate next and future rewards too. Some of the examples include Temporal Difference, Markov decision, Q-learning and Monte-Carlo methods [20].

The understanding of data types also referred as measurement scales in ML becomes mandatory for exploring the data analysis. Understanding the data type helps in dealing the right visualization (ML) method. The data types are broadly classified into (1) Categorical and (2) Numerical values. The former type represents the characteristics and the later represents the countable data. Categorical data is sub-divided into (a) Nominal type, to label the discrete variables without order, and (b) Ordinal type, to label the discrete variables with ordering. Numerical

data is further sub-divided into (a) Discrete data, i.e., the countable data and (b) Continuous Data (represents the measurable data) is subdivided into interval and ratio data [21] and [22]. Table 1 shows the summary of applications of ML algorithms on the corresponding data types summarized from [23].

Table 1 Categorization of ML Algorithms

ML algorithm types Data types	SML Algorithm	USML Algorithm
Discrete values	Classification or Categorization	Clustering
Continuous values	Regression	Dimensionality Reduction

III. MACHINE LEARNING TECHNIQUES

The following section deals with the introduction to ML techniques summarized from [1], [2], [4], [5] and [6]. They are;

A. Supervised Learning:

It is the common method adopted in industry use cases. Here both inputs, as well as desired outputs, are provided to a system with the aim to classify the future data. It is a routine process when an algorithm learns from a new training dataset. Self-driving cars, Chabot’s, face recognition methods adopts Supervised Learning approaches.

B. Unsupervised Learning:

It is a complicated process and prone to use in smaller applications. “Teaching themselves” one of Machine Learning’s property implies Unsupervised Learning. There are no explicit target outputs available in this method. In this approach, the classification is carried without the prior knowledge, that is, there is no prior knowledge about the data being classified. For example, clustering and association related problems fall into this learning approaches.

C. Semi-Supervised Learning

It is the study of how computers learn in the presence of labeled and unlabeled dataset. This technique put forths the protocol that generation of labelled data is costlier than the unlabelled data. Examples of this learning approaches are speech analysis and web content classification.

IV. MACHINE LEARNING ALGORITHMS

The following section deals with the description regarding the traditional ML algorithms summarized from [5], [7] and [8]. They are;

1. Naïve Bayes Classifier (NBC) Algorithm

NBC algorithm helps to classify a webpage, a document, a lengthy text note or an email, it is not easy to do it manually. Its popular application is spam filtering. The classifier function assigns a population’s element value as one of the accessible categories. Sentiment Analysis is also its popular application. It performs well when we have categorical input variables.

2. K-Means Clustering Algorithm

When a user searches for a keyword “jaguar” in the World Wide Web (WWW), results will be the aggregation of Jaguar animal, Jaguar car and Jaguar Mac Operating System. The similar characteristic featured contents (may be text, images, videos, etc) are grouped and displayed to the user. The grouping (also referred as clustering) of similar features is done by K-means clustering Algorithm. Further, this algorithm is adopted by many search engines (namely, Google and Yahoo) to group the web pages based on the similarity and identifies ‘relevance rate’ of concerned search results.

3. Support Vector Machine (SVM) Algorithm

SVM, a supervised algorithm, is used for classification problems. SVM learns about classes of input dataset which are differentiated by a line in the dataset to classify the new data. It is categorized as Linear and Non-linear SVM. Linear SVM can classify using just a hyperplane but it’s not possible for non-linear SVM. This technique does not make any substantial assumptions on the data. It is used to forecast stock market values.

4. Apriori Algorithm

Apriori algorithm, an unsupervised algorithm, depicts association rules from the input data set.

Association rule entails that on occurrence of item ‘X’, item ‘Y’ happens definitely, similar to “if-then” rule. For instance, **IF** a person purchases bike, **THEN** certainly they will purchase helmet. Critical observations on such purchases are examined to ensure the optimum decisions. Detecting unfavourable drug reactions and auto-complete applications are its applications.

5. Linear Regression (LR) Algorithm

LR algorithm depicts the association between the two variables and how they influence each other. It shows the impact on the dependent variable (termed as factor of interest) when the independent variable (termed as explanatory variables) is changed. It is a wide spread ML technique with rapid speed. Estimating sales and risk assessment is its applications.

6. Decision Tree Algorithm

It uses the branching method to represent the possible consequences for a decision with respect to specific scenario. The decisions in a tree are inversely proportional to the expected precision. Two sub-categories of decision trees are Classification and Regression trees. The classification and regression trees are used during the classification of nature and continuous/numerical variables, respectively. The decision tree classification is mainly used in finance and banking sectors to classify the loan applicants based on the probability of payments in a specific periodic intervals.

7. Random Forest Algorithm

Random Forest algorithm works by creating a cluster of decision tree with a subset of randomly chosen data. The prediction performance is improved by repeated training of a model with random samples. The final prediction is made from the integration of decision trees output followed by the polling procedure.

8. Logistic Regression Algorithm

Mainly this type of classification is applied to classify the discrete classes. For instance, to detect the mail as spam or not, online transaction made as a genuine or fraud, tumor classification as benign and malignant, etc. This type of regression transforms the desired output using sigmoidal function to determine the probability value. If the resultant classification is with two outputs (for instance, benign or malignant), then it is referred as Binary Logistic Regression. On the other hand, if the resultant output is of categorical values (for instance, dogs, cats, horses, etc), then it is referred as Multi-Linear Logistic Regression method.

V. MACHINE LEARNING APPLICATIONS

The following section deals with the insight on ML applications summarized from [9] and [10].

Spam Detection: Received emails are identified whether they are spam if yes then they are not shown to the user with regular emails in their inboxes. All these spam emails are maintained in a separate folder. This is done by learning how to recognize a spam mail by identifying the characteristics from newly received ones.

Face Detection: Identifying faces from a given number of photos and tagging them automatically if next time the same face is detected for a new photo. Currently, Facebook is popularly known for using this technique where when we upload a picture some suggestions of friends are automatically detected and shown to the user whose friends are in the picture. For example, Google photos separate all the pictures according to people in it.

Credit Card Fraud Detection: According to customers past transactions, if there are any inappropriate purchases made then the customer is warned immediately about the condition.

Digit Recognition: Machine's camera detects postal codes that are handwritten and arranges all the letters according to the geographical locations they have to reach. The machine is trained to learn handwritten numbers and transforms into digital signs.

Speech Understanding: Deals with listening to a speech by the user, understand user intentions process what machine has understood. Machine is expected to follow instructions from the users accurately. "Cortana" by Windows, "Siri" by Apple and "Okay Google" by Google are popular and successfully implemented applications of this technique.

Product Recommendation: With all the data of a customer's past purchases or interests online, the machine will recommend some products which would attract customers to view them and maybe even purchased. Flipkart, Amazon and many other e-commerce sites have been implanting this technique where we receive only recommended products as advertisements.

Medical Diagnosis: For detecting diseases more accurately, hospitals these days are using machines that could decide whether a person is affected with any diseases for symptoms he/her has, with the help of complete data about all diseases and symptoms. IBM designed a system with 95% precision in predicting the cancerous images in contrast to 75%-84% precision by doctors.

Stock Trading: Given the current and past price fluctuations of a stock, the machine decides to hold or sell the stock for better service of the customer(s).

Customer Segmentation: With the help of past behavior patterns, the ML algorithms tries to predict the number of users who will choose the paid versions from the trial versions. Amazon Prime has implemented this technique.

Chat smarter with Allo: Allo is a messenger application by Google which will learn from its users, how they are responding to what kind of messages and recommend the user with a response that he /she has thought of typing.

Financial Service: Companies can identify the company insights of financial sector data and can overcome the occurrence of financial fraud. It is used to identify the opportunities for investment and trade. We can also prevent the financial risks prone institutions by using cyber surveillance and take necessary actions to prevent fraud.

Transportation: By the travel history and pattern of travelling in various routes, machine learning helps in Transportation Company to predict the problem in routers and advice their customers to choose a different route. Transportation firmly uses machine language to carry out data analysis and data modelling.

Few other applications of ML are as follows:

Computational Intelligence: For many years computational intelligence is being developed actively. Constant improvements and improvements are being carried out on classical methods like machine learning algorithms. These days computational intelligence has been for many applications directly or indirectly.

Natural language processing: It is a field that which involves both computer understanding and manipulation of human language and its good in gathering new possibilities. It is mostly seen in a large pool of legislation or other document sets, trying to discover new patterns or to root out corruption. It is a better way to analyze, understand and find the meaning of human language easily and smartly. By using NLP developer can perform tasks such as speech recognition, entity recognition, automatic translation, and summarization.

Discrete Event Simulation is a technique where patients are modeled as an independent event associating each with some attribute information like age, weight and problematic scenarios, etc. Natural Language Processing is a technique used for system to read the physician's notes and convert it to digital data. Proprietary Predictive model is used to make predictions such as admissions can be predicted by hospitals to spread expertise which is in short supply. Disease prediction and diagnosis is achieved by helping radiologists to make intellectual decisions with radiology data (for

example, CT, Radiographs and MRI).

Sentiment Classification: It is difficult to classify documents based on sentiment rather on topic, like separation of comments into positive and the negative categories. Hence, the following ML techniques are adopted to classify the sentiments and subjected to further analysis.

- Naïve Bayes method is a trendy algorithm for text classification process; it is given some texts and its results to learn on how to decide about new text.
- Maximum Entropy classification it is an alternative technique to work on natural language processing although it fails to perform sometimes.
- Support Vector Machine is highly effective and can perform better than the other two techniques but they are large margin classifiers rather than probabilistic classifiers. It is useful very much because classifying texts according to sentiments is not easy for human to carry out, whereas machine learning has made it simple.

Mobile Devices: ML techniques when applied on portable devices like Sensors, Smartcards, Smartphones, computing handheld and automotive systems had proved efficient. As mobile terminals are improving a lot these days mobile technology has encouraged technological advancements. The techniques like Naïve Bayesian, Decision trees and C4.5 are supportive for mobile devices. These techniques major categorised as supervised, unsupervised and reinforcement these are used to represent learning methods.

Here, machine learning approach is provided with some training examples for given function. Support Vector Machines (SVM), K-Nearest Neighbour (KNN), Random Forests and Decision trees are some of supervised algorithms. There is no supervisor available and learning in the Unsupervised ML category. Training samples and testing samples uses these techniques. Each ML technique has few strengths and weakness. It gives brief about techniques for application in mobile devices. It presents performance measures for machine learning algorithms.

Data Mining: There are two styles of data mining algorithms summarized from [13], [14] and [15]. They are (1) Predictive, and (2) Descriptive

Predicting the value of special attribute using the other attribute is called predictive whereas deriving the patterns that summarize the relationship among data points is called descriptive.

Supervised Data Mining: It is a pre specified target variable. Its goal is to specify the relationship between predictors and target. Model is provided with many training data where the target is known. It is applied to the data where the target is unknown

Unsupervised Data Mining: It is mostly used to determine whether the classes or clusters exist in data and also for exploratory analysis. All variables are treated in same way.

Pattern Recognition: Pattern recognition is the other main problem in machine learning. Pattern recognising algorithm generally aims to provide a reasonable answer for all possible inputs and to perform closest to matching of input. It uses two learning methods, namely (1) Supervised learning and (2) Unsupervised learning method.

VI. CONCLUSION

The paper presented an overview and insight on the Machine Learning approaches and techniques applied to various applications. Apart from the applications, the paper also put forth an idea of applying the algorithms and its performances to various scenarios. For instance, whenever there exists a sample, the supervised machine learning algorithm comes in to the picture, whereas on the other hand, analyzing the data without existing data, the clustering technique from unsupervised machine learning algorithm is very useful in providing the solutions. As a future enhancement, authors would like to work on intelligent decision support systems integrated with machine learning techniques for predicting the cardiac diseases.

REFERENCES

1. A Tour of The Most Popular Machine Learning Algorithms, <https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/>
2. Supervised Learning, <http://whatis.techtarget.com/definition/supervised-learning>
3. Supervised V Unsupervised Machine Learning -- What's The Difference?, <https://www.forbes.com/sites/bernardmarr/2017/03/16/supervised-v-unsupervised-machine-learning-whats-the-difference/#66901d12485d>
4. Peter Dayan, "Unsupervised Learning", The MIT Encyclopedia of the Cognitive Sciences, Wilson, RA & Keil, F, Editors, pp. 1-7, 1999.
5. Top 10 Machine Learning Algorithms, <https://www.dezyre.com/article/top-10-machine-learning-algorithms/202>
6. Understanding Semi-supervised Learning, <https://medium.com/@jrodthoughts/understanding-semi-supervised-learning-a6437c070c87>
7. Xin-She Yang, Su Fong Chien, and Tiew On Ting, "Computational Intelligence and Metaheuristic Algorithms with Applications", The Scientific World Journal, Vol. 2014, pp. 1-4, 2014.
8. Introduction to Natural Language Processing (NLP): What is NLP?, <https://blog.algorithmia.com/introduction-natural-language-processing-nlp/>
9. What are some interesting possible applications of machine learning? <https://www.quora.com/What-are-some-interesting-possible-applications-of-machine-learning>
10. What are some real-world examples of applications of machine learning in the field? <https://www.quora.com/What-are-some-real-world-examples-of-applications-of-machine-learning-in-the-field>
11. The Top 10 AI And Machine Learning Use Cases Everyone Should Know About, <https://www.forbes.com/sites/bernardmarr/2016/09/30/what-are-the-top-10-use-cases-for-machine-learning-and-ai/>
12. Bo Pang, Lillian Lee, Shivakumar Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques", Association for Computational Linguistics, Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002), pp. 79-86, 2002.



13. Chaudhar , S. Kolhe, Rajkamal, “Machine Learning Techniques for Mobile Devices: A Review”, International Journal of Engineering Research and Applications, Vol. 3, Issue 6, pp.913-917, Nov-Dec 2013.
14. Introduction to Data Mining and Machine Learning Techniques- Iza Moise, Evangelos Pournaras, Dirk Helbing, Lecture Notes, 2018.
15. Anish Talwar, Yogesh Kumar, “Machine Learning: An artificial intelligence methodology”, International Journal Of Engineering And Computer Science, Vol. 2, No. 12, pp. 3400-3404, Dec.2013.
16. Gartner, Preparing and Architecting for Machine Learning, Technical Professional Advice, Analyst(s): Carlton E. Sapp, 2017, https://www.gartner.com/binaries/content/assets/events/keywords/catalyst/catus8/preparing_and_architecting_for_machine_learning.pdf
17. Machine Learning: What it is and Why it Matters, <https://www.simplilearn.com/what-is-machine-learning-and-why-it-matters-article>
18. Jason Brownlee, Master Machine Learning Algorithms: Discover how they work and implement them from Scratch, 2016, <http://MachineLearningMastery.com>
19. Gavin H.: Mastering Machine Learning with scikit-learn, Packt Publishing, (2014)
20. Swamynathan, M.: Mastering Machine Learning with Python in Six Steps: A Practical Implementation Guide to Predictive Data Analytics Using Python, Apress (2017)
21. Muller, A.: Introduction to Machine Learning with Python: A Guide for Data Scientists, Shroff O'Reilly, 1st edition (2016)
22. Data Types in Statistics, <https://towardsdatascience.com/data-types-in-statistics-347e152e8bee>
23. Supervised vs. Unsupervised Learning, <https://towardsdatascience.com/supervised-vs-unsupervised-learning-14f68e32ea8d>