Journal of Engineering Design and Analysis
Volume 3, Issue 2 - 2020, Pg. No. 88-91
Peer Reviewed Journal

Article

A Literature Survey on Machine Learning Algorithms and Applications

Prasanta Kumar Dash

Gandhi Institute for Technological Advancement, Bhubaneswar, Odisha, India.

I N F O

E-mail Id:
prashant_cse@gita.edu.in
Orcid Id:
https://orcid.org/0000-0002-6323-9188
How to cite this article:
Dash PK. A Literature Survey on Machine Learning Algorithms and Applications. *J Engr Desg Anal* 2020; 3(2): 88-91.

Date of Submission: 2020-11-04
Date of Acceptance: 2020-11-18

A B S T R A C T

Machine Learning (ML) is a subset of the Artificial Intelligence, a field of computer science that provides system to automatically learn and improve from past experience without being explicitly programmed. Machine Learning (ML) is a multidisciplinary field, a combination of statistics and computer science algorithms which is widely used in prediction and classification. The second section of the paper focuses on the basic machine learning methods and algorithms. This paper will highlight the various machine learning tools needed to run the machine learning programs. The main concern of concerned paper is to study the main approaches and case studies of using machine learning for forecasting in different areas such stock price forecasting, weather forecasting, tourism demand forecasting, solar irradiation forecasting, supply chain demand and consideration of neural network in machine learning methods.

Keywords: Machine Learning, Supervised Learning, Unsupervised Learning, Classification, Regression, Clustering

Introduction

Over the past decades, Artificial Intelligence (AI) stream has become the broad and exciting field in computer science as it prepare the machines to perform the tasks that human being may do and it aims to train the computers to solve real world problems with the maximum success rate. As perceiving scientific growth and advancement in technology AI systems are now capable to learn and improve through past experiences without explicitly assistance code if they exposed to new data. Eventually it leads to technology of Machine Learning (ML) which uses learning algorithms to learn from the data available.[1] Machine Learning uses data mining techniques to extract the information from the huge size datasets. ML and Data Mining techniques explore data from end to end to find the hidden patterns inside dataset.[2] Machine Learning and data mining algorithms has been deployed in various fields such as Computer networking, travel and tourism industry, finance, forecasting, telecommunication industry and electric load forecasting and many more.[2]

Preliminaries

Machine Learning Methods

Over past years an enormous number of ML algorithms was introduced. Only some of them were able to solve the problem so they replaced by another one.[3] There are three ML algorithms for example unsupervised learning and reinforcement learning, supervised learning, which are displayed.

Supervised Learning

It consists of a given set of input variables (training data) which are pre labeled and target data.[5] Using the input variables it generates a mapping function to map inputs to required outputs. Parameter adjustment procedure continues until the system acquires a suitable accuracy.

Unsupervised Learning

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 without guidance.

In this algorithm we only have training data rather a outcome data. The input data is not previously labeled. It is used in classifiers by recognizing existing patterns or cluster in the input datasets.[4]

Reinforcement Learning

Reinforcement Learning (RL) is an area of machine learning concerned with how software agents ought to take actions in an environment so as to maximize reward in a particular situation. Applying this algorithm machine is trained to map action to a specific decision hence the reward or feedback signals are generated. The machine trained itself to find the most rewarding actions by reward and punishment using past experience.

Algorithm of Machine Learning

There is massive number of algorithms used by machine learning are designed to erect models of machine learning and implemented in it.[4] All algorithms can be grouped by their learning methodology as follows.

Regression Algorithms

In Regression algorithms predictions are made by the model with modeling the relationship between variables using a measure of error. The outcome of a Regression technique is continuous value. The variable can be a price, a temperature. Different regression algorithms are as follows:

- Linear Regression algorithm
- Ordinary Least Squares Regression
- Multivariate Adaptive Regression Splines
- Logistic Regression
- Locally Estimated Scatter plot Smoothing
- Stepwise Regression

Instance Based Learning Algorithm

In machine learning, instance-based learning (sometimes called memory-based learning) is a family of learning algorithms that, instead of performing explicit generalization, compares new problem instances with instances seen in training, which have been stored in memory. Instance based learning refers to a family of teccniques for classification and regression which produce a class label/ prediction based on the similiary of the query to its nearest neighbour in the training set.

In explict contrast to other methods such as decision tress and neural network, instance based learning algorithm do not create an abstraction from specific instances. Rather, they simply store all the data and at query time drive an answer from an examination of the query nearest neighbour.

The most well known algorithms based on instance learning algorithms are:

- Learning Vector Quantization
- Self-Organizing Map
- k-Nearest Neighbor
- Locally Weighted Learning

Decision Tree Based Learning Algorithm

Decision trees are a type of supervised machine learning (that is you explain what the input is and what the corresponding output is in the training data) where the data is continuously split according to certain parameter. The tree can be explained by two entities, namely decision nodes and leaves. The leaves are the decisions or the final outcomes and the decision nodes are whrere the data is split.

An example of a decision tree can be explained usinh above binary tree. Let's say you want to predict whether a person is fit given their information like age, eating, habit and physical activity etc. The decision nodes here are questions like, What's the age?, Does he exercise?, Does he eat a lot of Pizza? And the leaves, which are outcomes like either fir or unfit. In this case was a binary classification problem (a yes no type problem).

There are two main types of Decision Trees:

Classification Trees (Yes/ No types)

What we have seen above is an example of classification tree, where the outcome was a variable like fit or unfit. Here the decision varriable is Categorical.

Reggration Trees (Continuous data types)

Here the decision or the outcome variable is continuous example like 123.

The most well known algorithm using decision tree are:

- Iterative Dichotomized 3
- M5
- Chi squared Automatic Interaction Detection
- C5.0 and C4.5 (different versions of a powerful approach)
- Decision Stump
- Classification and Regression Tree
- Conditional Decision Tree

Baysian Algorithms

Machine Learning is multidisciplinary field of Computer Science like Statistics and algorithm. Statistics manages and quantifies the uncertainty and are represented by bayesian algorithms based on probablity theory and Bayes Theorem. The most famous Bayesian algorithms are:

ISSN: 2582-5607

- Bayesian Belief Network (BBN)
- Multinomial Naive Bayes Bayesian Network (BN)
- Gaussian Naive Bayes
- Naive Bayes

Data Clustering Algorithms

This algorithm split items into different types of batches. It groups the item set into clusters in which each subset share some similarity. It is unsupervised learning method and its methods are categorized as hierarchical or network clustering and partitioned clustering. The most well known algorithms for clustering are:

- K Means
- Expectation Maximization(EM)
- K Medians
- Hierarchical-Clustering

Learning Algorithms Using Association Rule

Learning algorithms using Association rule are generally utilized by the organization commercially when multidimensional datasets are huge in size. They are used as extraction methods that can explore observed relationships between variables and data. The most well known learning algorithms using association rule are:

- ECLAT algorithm
- Apriori algorithm

Algorithms Using Artificial Neural Network

Artificial neural networks models are based on the biological neuron structure and uses supervised learning. It consists of artificial neurons which have weighted interconnections among units. They are also well known by parallel distributed processing networks. The most famous or well known algorithms for artificial neural network are:

- Radial Basis Function Network (RBFN)
- Back-Propagation
- Perceptron
- Hopfield Network

Deep Learning

Deep learning is an artificial intelligence function that imitates the workings of the human brain in processing data and creating patterns for use in decision making. Deep learning is a subset of machine learning in artificial intelligence (AI) that has networks capable of learning unsupervised from data that is unstructured or unlabeled.

Deep learning has evolved hand-in-hand with the digital era, which has brought about an explosion of data in all forms and from every region of the world. This data, known simply as big data, is drawn from sources like social media, internet search engines, e-commerce platforms and online cinemas, among others. This enormous amount of data is readily accessible and can be shared through fintech applications like cloud computing. Famous algorithms for deep learning are:

- Stacked Auto Encoders
- Deep Boltzmann Machine (DBM)
- Convolution Neural Network (CNN)

Algorithms using Dimensionality Reduction

Dimensionality reduction method is widely used in case of large number of dimensions, large volume of space concerned. Then that problem requires a statistical significance. Dimensionality reduction methods used for minimizing the number of dimensions outlined the item and removes unrelated and unessential data which lessen the computational cost. Some of these methods are used in classifying and regression. The algorithms using reduction in dimensionality are as follows:

- Partial Least Squares Regression
- Multidimensional Scaling
- Principal Component Analysis
- Flexible Discriminant Analysis
- Mixture Discriminant Analysis
- Sammon Mapping
- Projection Pursuit
- Linear Discriminant Analysis
- Principal Component Regression
- Quadratic Discriminant Analysis

Literature Review

Rob Law (1998)[7] applies neural networks to forecasts occupancy rates for the rooms of Hong Kong hotels and finds that neural networks outperforms naïve extrapolation model and also superior to multiple regression. This research studied the feasibility incorporating the neural network to predict the rate of occupancy of rooms in Hong Kong hotel industry.

Authors Hua et al., (2006)[8] described support vector machines approach to predict occurrences of non zero demand or load time demand of spare parts which used in petrochemical enterprise in china for inventory management. They used a integrated procedure for establishing a correlation of explanatory variables and autocorrelation of time series of demand with demand of spare parts. On performing the comparison the performance of SVM method with this LRSVM model, Croston's model exponential smoothing model, IFM method and Markov bootstrapping procedure, it performs best across others. Authors Vahidov et al., (2008)[8] compares the methods of predicting demand in the last of a supply chain, the naive forecasting and linear regression and trend moving average with advanced machine learning methods such as neural networks and support vector machines, recurrent neural networks finds that recurrent neural networks and support vector machines show the best performance.

Wang (2007)[8] describes the machine learning method with genetic algorithm (GA)-SVR with real value Gas. The experimental findings investigates this, SVR outshines the ARIMA models and BPNN regarding the base the normalized mean square error and mean absolute percentage error.

Wang et al., (2016)[4] proposed a novel forecasting method CMCSGM based Markov chain grey model which used algorithm of Cuckoo search optimization to make better the performance of the Markov chain grey model. The resultant study indicates that the given model is systematic and fine than the traditional MCGM models.

Barzegar et al., (2017)[3] demonstrates model predict multi-step ahead electrical conductivity i.e. indicator of water quality which is needed for estimating the mineralization, purification and salinity of water based on wavelet extreme learning machine hybrid or WAELM models and extreme learning machine which exploiting the boosting ensemble method. The findings showed that upgrading multi WA ELM and multi WAANFIS ensemble models outshines the individual WAELM and WA ANFIS constructions.

Zhang et al., (2018)[2] suggests a design of multi kernel ELM or MKELM method for segregation of motor imagery electroencephalogram or EEG and investigate performance of kernel ELM and impacts of two different functions of kernel such as polynomial and Gaussian kernel Compares MKELM method gives greater segregation accuracy than other algorithms indicates betterment of the suggested MKELM based.

Applications of Machine Learning

In the research paper we studied various Machine- learning techniques such as supervised and unsupervised learning. Supervised learning is applied in classification problems like face recognition, medical diagnosis, pattern recognition, character recognition, web advertizing.

Unsupervised learning can be applied in clustering, association analysis, CRM, summarization, image compression, bioinformatics. Reinforcement learning is widely applied in game playing and robot control.[3]

Tools used in Machine Learning

Tools make machine learning swift and rapid. Machine learning tools provides interface to the machine learning programming language. They provide best practices for process and implementation. Machine learning tools contains platforms which provides capabilities to run a module or project. Examples of platforms of machine learning are:

- Python SciPy subparts such as scikit-learn, Panda
- R Platform
- WEKA Machine Learning Workbench
- Machine learning tools contain various libraries which provide all capabilities to complete a project and libraries provide various algorithms. Some of libraries are
- JSAT in Java
- Scikit-learn in Python
- Accord Framework in .NET

Conclusion

Machine learning methods and algorithms have been reviewed in this paper. This paper also reviewed algorithms describing the various types of machine learning techniques, algorithms and methodology. Various applications of Machine learning and many tools needed for processing are also being reviewed. In the Literature review section we studied various machine learning algorithms implemented in past years in different areas in combination with the tradition methods and studied how they outperformed the previous models.

References

1. Awad M, Khanna R. Efficient learning machines: Concepts and Applications. Aspress Publications, 2015.
2. Xiuyi T, Yuxial G. Research on Application of machine learning in data mining. IOP Conf. Series: Materials Science and Engineering. 2018
3. Praveen M, Jaignesh V. Litrature review on Supervised Machine Learning Algorithms and Boosting Process. *International Journal of Computer Applications.*
4. Das K, Narayan RB. A survey on Machine learning: concept, algorithm and applications. *International Journal of Innovative Research in Computer and Communication Engineering* 2017; 5.
5. Kotsiants SB. Supervised Machine Learning: A Review of classification tyechniques. *Informatica* 2007; 249-268.
6. Law R. Room occupancy rate forecasting: a neural network approach. *International Journal of Contemporary Hospitality Management* 1998; 10(6): 234-239.
7. Hua Z, Bin Z. A hybrid support vector machines and logistic regression approach for forecasting intermitten demand of spare parts. *Applied Mathematics and Computation* 2006; 181: 1035-1048.
8. Carbonneu R, Kevin L, Vahidov R. Application of Machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research* 2008; 184: 1140-1154.
9. Kuan-Yu C, Cheng HW. Support vector regression with genetic algorithms in forecasting tourism demand *Tourism Management* 2007; 28: 215-226.
10. Wei CH, Dong Y, Chen LY et al. SVR with hybrid chaotic genetic algorithm for tourism demand forecasting. *Applied soft Computing* 2011; 11: 1881-1890.

ISSN: 2582-5607