Asst. Prof. Manish Mishra, 2021, 9:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

A Survey on Machine Learning Features and Techniques

Asst. Prof. Manish Mishra, Professor Mahesh Motwani
Department of Computer Science and Engineering
UIT RGPV, MP,India

Abstract- Machine learning algorithm are help to generate unique pattern from raw dataset. Number of researcher has proposed different techniques for the processing of raw dataset to extract information. In these algorithms input is user data output is predicted result in form of pattern, cluster, class, etc. This paper introduce some algorithms developed by researcher to drive information from data. The paper throw light on implementation area of machine learning technique, type of features and there requirement in different dataset and evaluation parameter for the analysis of prediction, classification, clustering algorithm.

***Keywords*- Machine Learning, Genetic algorithm, Neural Network.**

I. INTRODUCTION

Machine learning is about using computers for calculations and data retrieval [1, 2, 3]. Machine learning can be accomplished in a supervised or an unsupervised way. In supervised learning, the system receives a dataset with different example parameter values and decisions/classification, from which it infers a mathematical function, which automatically maps an input signal to an output signal [4].

Unsupervised learning, on the other hand, means that the system acts and observes the consequences of its actions, without referring to any predefined type cases other than those previously observed. This is pure `learning by doing' or trial-and error. Compared to supervised learning, unsupervised methods perform poorly in the beginning, when they are un-tuned, but as they tune themselves, performance increases [4].

Rest of paper was arrange in few section where second section has summarized different machine learning methods adopt by researcher to extract information. Third gives a brief explanation of different features used in different type of dataset, while fourth is collection of machine learning techniques. In fifth section different evaluation parameters were explained with there formulas.

II. LITERATURE SURVEY

Hamideh et. al. [5], gave ensemble methods of the core function of neural networks. The paper proposed a function that is a combination of linear single losses and integrated it into the proposed function into the machine learning network. By this, the weights that were associated with combination losses were learned by using backpropagation during the stage of training. Thus, at last their impact was studied under art of neural networks for classification of text.

Fan et. al. [6] built a superior model of regression for predicting the ratings of users in neural networks. Based on received item and user data item feature matrix and user feature matrix was constructed. It was constructed by using 4 types of neural networks pooling layer or PL and fully connected layer or FCL, embedding layer or EL, convolution layer or CL. Using this matrix further user feature matrix was created using FCL. Thus this regression model was trained to generate a recommendation user rating list.

Jin Z.et. al. [7] gave a DDAL or discriminative deep association of learning framework for the recognition of facial expression. In this, the unlabeled data is prefered to train the DDNs with the help of labeled data at the same time in a multi-loss class of network

based association. Discrimination loss is used to confirm intra-class group of clusters and separation.
CHUANLONG et. al. [8] Given a model on conceptual learning named intrusion detection system(RNN-IDS). Authors have to study the outcome of this model in multiclass and binary classification, and also how different learning rates and a number of neurons affect the performance of the model. Finally, the authors have compared this with J48, , support vector machine random forest, artificial neural network and several other machine learning methods given by researchers earlier.

V. Anitha et. al. [9] In this system at the first stage self-organizing map trains the features that were extracted from the different wavelet transformation and mix wavelets. Then the obtained filter vector is trained by the nearest neighbor and then finally the verification process was done in two stages. This proposed two-level classification was found more effective than the traditional training process.
Yuki et. al. [10] gives a watermarking technology in digital data to claim the ownership of machine learning networks. In this, the first paper contains the problems related embedding of watermarks in deep machine learning networks. The second paper gives the common framework to embed watermarking in model parameters by using a parameter form of regularizer. The proposed way does not give the output performance of networks after watermarking as watermark was embedded while the training of a network(host).

E. Jadon et al. (2017) [11] gave classification of multiclass documents for text documents. To solve text classification problem Naïve Bayes classifier was used. The experiment was done first with first linear and second in a hierarchical way to get effective results. Finally, there was a conclusion that the hierarchical way is more effective than the linear approach and it improved the efficiency and accuracy of such classifier.

III. FEATURES OF MACHINE LEARNING

As per type of data text, image, number features were extract so based on three type of data features are fetch for learning.

1. Image Features Set

In the transform domain, the host image is segmented into multiple frequency chains using several transformations such as DWT(Discrete wavlelet Transform) or DCT (Discrete Cosine transform) and many more [12]. Then, the inverse transform is applied to obtain the watermarked image. DWT and DCT are most widely used transform for image watermarking.

The frequency transformation in the DCT domain divides image into different frequency bands, so they facilitate embedding the watermarking information in a specific frequency band [13]. It has been found that the middle-frequency bands are most suitable for embedding the watermark because the low-frequency band carries the most visual essential parts of the image.

While the high-frequency band is exposed to removal through compression and noise attacks on the image. Therefore, embedding the watermark in the middle frequency band neither affects the visible essential parts of the image (low frequency) nor overexposes them to removal through attacks when high-frequency components are targeted [14]. The DWT transform divides into four different parts namely: LL, LH, HL and HH sub bands. Majorly, LL sub band are utilized for watermark because LL sub band conatins Low frequency components that attains resistance against different attacks**.**

2.Text Feature

Term Frequency: The TF is the count of category-of-words of every category in each document. So the documents term frequency for a category is the occurrence of the words in single document or article [15].

2.1 Document Term Frequency: It is the number of documents in the collection that contain a term. IDF: Inverse Document Frequency, is a measure of how much information the word provides, i.e., if it's common or rare across all documents. It is the logarithmically scaled inverse fraction of the documents that contain the word.

$$IDF(t) = log\left(\frac{N}{n}\right)$$

N represents the total number of documents in the dataset, n represents the number of documents that term t appears

2.2 TF-IDF: TF-IDF [16] (Term Frequency-inverse Document Frequency), puts weighting to a term based on its inverse document frequency. It means that if the more documents a term appears, the less

important that term will be, and the weighting will be less.

$$TFIDF(t) = TF_t * log\left(\frac{N}{n_t}\right)$$

2.3 TF-IDF-CF: As per the Shortcomings of TF-IDF has, introduce a new parameter to represent the in-class characteristics, and authors have call this class frequency, which calculates the term frequency in documents within one class.

$$TFIDFCF(t) = \log(TF_t + 1) * log\left(\frac{N+1}{n_t}\right) * \frac{n_{c,t}}{N_c}$$

the number of documents where term t appears within the same class c document. Nc represents the number of documents within the same class c document.

IV.NUMERIC FEATURE EXTRACTION

1.Markov Model: The Kth order markov models were develop from a series of numbers. These are patterns obtain from the numeric dataset like weblog page visiting sequence [17].

2.Regression: As per requirement different type of regression (linear / logistic) features were extract from the numeric data [18]. Finding a feature from temporal data is done by this regression.

3.Machine Learning Techniques

3.1 Regression Algorithms

The Regression analysis deals to exploit the relation between independent variables [18] and dependent (target) and is a part of prediction analytics. The fanous regression models are:, Ordinary Least Squares Regression or OLSR, Stepwise Regression Linear

3.2 Regression

Multivariate Adaptive Regression Splines or MARS, Logistic Regression , Locally Estimated Scatter plot Smoothing or LOESS, etc.

3.2.1 Linear Regression- It comes under the category of checking learning of machine learning related algorithms. Regression is utilized to predict the continuous value and the forecasting method is used to predict the past and future values. In the field of statistics, linear regression is commonly used to find out the relation in input and output numerical data variables. Now it has been equipped with machine learning to accomplish Model, finding of the mean and variance value, covariance, and related Assess coefficients [25], and Simple Linear Regression. The result is a type of linear equation that takes the input value(x) and predicts the output(y) on type of the input provided. The pair of input v/s output should be of numerical values.

Logistic Regression: Gives the maximum-like hood estimation based on the provided dataset. The function form of logistic regression is different from artificial neural sets. The logistic regression function is called a parametric method while the artificial neural nets are called semi-parametric or completely non-parametric [26]. Their difference is important as it gives parameters (coefficients and intercept) in logical regression can be interpreted while in the other it is not always the matter with neural network of weights parameter.

4. Instance-based Algorithms

These are memory-based learning models that store instances of the training facts instead of making precision target functions. Whenever it encounters any new problem it checks it with the already stored instances to predict the function value of target [19]. It will also store and replace any new instance with the older one if it thinks that it fits better than the other. For this reason, it is also called the winner takes all method. Examples are, K-Nearest Neighbor or KNN [34], Learning Vector Quantization or LVQ Self-Organizing Map or SOM, Locally Weighted Learning or LWL etc.

5. Regularization Algorithm

Regularization [39] is a process in machine learning which provides the values with a reduced error coefficient. Overfitting is avoided in this technique so that the function is fixed appropriately. In the case of machine learning reducing such values of errors is called decay of weight. In the case of neural networks and error function, this regularization gives an excess penalty term.

6. Decision Tree Algorithms

It is a tree-like structure that gives many hopeful solutions to a problem which depends on constraints. The beginning of the tree is from the root and then spreads into several branches and reaches the under the prediction of decision is made. It tends to provide the potential solution to a problem faster and with accuracy than others.

Examples of the decision tree are, Classification and Regression Tree or CART, Conditional Decision Trees Decision Stump, Iterative Dichotomiser 3 or ID3,C4.5 and C5.0, Chi-squared Automatic Interaction Detection or CHAID, M5, etc.

7. Bayesian Algorithms

It is a collection of ML algorithms to solve regression and classification problems. Examples are, Averaged One-Dependence Estimators or AODE, Multinomial Naive Bayes Naive Bayes, Bayesian Belief Network or BBN, Gaussian Naive Bayes, Bayesian Network or BN, etc.

8. Support Vector Machine (SVM)

It is quite a famous technique that has a group of itself. To demarcate the decision boundaries in the data set with different labels it uses a dividing hyper plane or a decision type plane. In other words, this algorithm makes an perfect hyperplane by using input data or data of training into categories. Examples are SVM which can perform both like linearity and non-linearity classification results.

9. Association Rule Learning Algorithms

It gives a relation between dissimilar data. They are mainly used in e-commerce sites to know the behavior of the customer and to promote their products accordingly. Examples are Eclat algorithm and Apriori algorithm,and many more.

10. Artificial Neural Network (ANN) Algorithms

It is a model which is the exact replica of the neural networks of animals or humans. ANN is considered as a non-linear model that gives the complex relationship between input and output data[8]. But is reduces both cost and time[21] as it compares only the data and not the complete data set. Examples are Hop-related Network, CNN, Recurrent machine learning, Perceptron, Back- Propagation, associative type memory networks, ART, counter propagation networks.

11. Clustering

Clustering was used to reduce the size of the data to manage the large dataset. Here cluster centers were identified and each cluster tends to select data units from other clusters. It is a tough task to select good bunch centers in large units of data. Many researchers used algorithms such as K-means, Clara, divisive, k-medoid, FCM, etc [22] related to clustering of data. Out of which some were considered unsupervised while some were partially supervised in which steps were taken to improve the accuracy of cluster selection. [35]. Once a cluster identifies other elements present in the group similarity value is obtained.

12. Genetic Algorithm

It needs plenty of time because the combination increases exponentially with the increase of the sets of data and a solution was needed for this. So, random choosing of solution was done to reduce the execution time by using genetic type of algorithms [23]. These algorithms worked on the concept of environmental and biological activity of the surroundings. Research implemented this concept to solve many problems like clustering, load balancing, shortest path identification, feature selection, classification, etc [24]. Butterfly, Bee colony, Ant Colony, PSO, etc are some of the well-known genetic algorithms. It depends on the nature of the problem that which genetic algorithm needs to apply.

13. k-NN

Also called as K nearest neighbors [34, 35] and is a supervised learning machine used to solve regression and classification problems. Based on resemblance like Euclidian distance it gives new data points. So this algorithm is used to classify data points like Euclidian distance, it differentiates the data points that are similar.

14. K-Means The Clustering

IT divides the data into set-off disjoint group. Each item can be a member of the group if it's similar. K-mean [34,35] is commonly used in the partitioned clustering algorithm. It clusters the n number of data points into k groups. It gives k centroids that are randomly selected which is single for each cluster. After this, it sends each point of data in the nearest centroid.

15. Random forest Tree

It constructs a collection of random independent and non-independent identical tresss of decision based on the randomization method. Each decision tree randomly selects vector parameter, feature samples, and a subset of sample data as its training set [28]. The developing algorithm of the random forest is like k gives the number of trees of decision based from any random forest, n gives the number of samples in the training data set corresponding to each decision

tree, and thus segmentation is carried out on a isolate node in the decision tree perfectly.

16. Dimension Reduction Technique

It is the processing of obtaining a group of important variables [29] by considering various random variables. Feature extraction and future selection are the two major categories that this dimension reduction approach follows.

16.1 Principal Component Analysis

It is a trait extraction process that gives new features which were in the singular combination of the starting features. In each PCA maps instance k<d where d is dimensional data space and k is a dimensional subspace. All the newly generated k dimensions are called Principal components and each one of them is bound for towards the maximum difference except such variance which has already been counted in previous components. [29]. So the first component covers maximum variance followed by lesser to least value of variance.

17. Singular Value Decomposition:

The SVD is considered an important algorithm as it leads the data from a high dimension to a lesser dimension through a matrix. As it chooses less number of dimensionless accuracy is achieved[30]. So SVD gives a fair illustration of any matrix and eliminates a compact amount of the matrix parts and gives a good approximation in any preferred number of the dimensions. Let M denotes an mxn matrix and its rank is r . In this, the rank of the matrix is known to be the max number of rows or any column in the matrix and authors chooses the linear combination of rows to vector 0(zero).

18. Reinforcement-Learning Model

An agent is linked via action and perception in the normal reinforcement learning model. On every step, the agent receives input and some indication from the current state and chooses an act to give the output. A Scalar reinforcement signal[31] is used to communicate the value of state transition. It's the behavior of the agent to choose any action that helps in achieving long-run values of the environmental signal. It will learn this in due course by trial and error that will be examined by it in several algorithms together.

19. Independent Component Analysis

The all-purpose model of ICA is that it generates a source through singular basis transformation in which noise is present. For example if a source have n independent signal si(t),i = 1,2,3,4,...,N. The authors assume that these sources cannot directly be observed and each si(t) depends on probability distribution corresponding to time. T.

If authors have observed the signals using N number of the sensors then authors obtained N signals of observation xi(t), i = 1,..., N which are a mixture of these resources. Sensors should be spatially separated to obtain a different combination of resources which is the basic aspect of the combination process. With this concept, the authors can get the matrix multiplication of spatial separation as-

$$x(t) = As(t)$$

Where the A= unknown or mixing matrix and the two x(t), s(t) are the vectors that represent the experiential signals and source signal[32]. The justification of this description is blind as the author is unfamiliar and does not have any knowledge on the mix matrix or even of the source. The aim is to recover the original signal si(t) from the experiential vector xi(t).

20. Transfer Learning

Transfer learning (also known as knowledge transfer, learning to learn) refers to a subfield of machine learning. The aim of transfer learning is to learn an objective predictive function for a target task with help of not only the target domain but also from other source domain and source tasks. Tasks in traditional machine learning process are independent of each other, and every task should be learned from scratch [33]. However, in transfer learning previous knowledge extracted from other source tasks can be transferred to the learning process of a new task. Transfer learning can be applied in various scenarios, such as web document classification, indoor WiFi localization problem and sentiment classification etc

IV. COMPARATIVE EVALUATION PARAMETERS

Recall and Precision [36] is used to evaluate the quality of clustering that is done using the clustering algorithm. Recall is defined as the number of data points that are correctly kept into a cluster by the

clustering algorithm used divided by the correct points that should have been present in that cluster. Precision is defined as the number of data points of the given data set that have been correctly kept into a cluster divided by all the points that are kept into that cluster by the clustering algorithm.

V.CONCLUSION

As large amount of data available in different platform, so information extraction depends on machines learning algorithms. This paper has summarize techniques of machine learning used by researcher in various field like IDS (intrusion Detection System), Web Mining, Text Mining, Image Processing, etc. As training of machine learning algorithm is done by extracted features from the dataset, so features of data as per type of dataset were also detailed in this paper. Evaluation parameters were also showed in the paper for comparison of machine learning algorithm. In future one can develop a generalize technique which work in all set of datasets.

REFERENCES

[1]. Welling, "A First Encounter with Machine Learning". University of California Irvine 2011.

[2] S.B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques", Informatica 31 (2007) 249-268

[3] Gupta, N.A. Literature Survey on Artificial Intelligence. 2017. IJERT Volume 5 Issue 1 7 January 2020.

[4] Carvalho, T.P.; Soares, F.A.A.M.N.; Vita, R.; da Francisco, P.R.; Basto, J.P.; Alcalá, S.G.S. A systematic literature review of machine learning methods applied to predictive maintenance. Comput. Ind. Eng. 2019.

[5] Hajiabadi, H., Molla-Aliod, D., Monsefi, R. et al. Combination of loss functions for deep text classification. Int. J. Mach. Learn. & Cyber. 11, 751–761 (2020).

[6] Bi, J., Liu, Y. & Fan, Z. A deep neural networks based recommendation algorithm using user and item basic data. Int. J. Mach. Learn. & Cyber. 11, 763–777 (2020).

[7] Jin, X., Sun, W. & Jin, Z. A discriminative deep association learning for facial expression recognition. Int. J. Mach. Learn. & Cyber. 11, 779–793 (2020).

[8] Chuanlong Yin , Yuefei Zhu, Jinlong Fei, And Xinzheng He. "A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks". IEEE Access November 7, 2017.

[9] V. Anitha, S. Murugavalli. "Brain tumour classification using two-tier classifier with adaptive segmentation technique". ISSN 1751-9632 24th July 2014.

[10] Yuki Nagai, Yusuke Uchida, Shigeyuki Sakazawa, Shin'ichi Satoh. "Digital Watermarking for Deep Neural Networks". Computer Vision and Pattern Recognition, 6 Feb 2018.

[11] E Jadon, R Sharma et al. "Data Mining: Document Classification using Naive Bayes Classifier" International Journal of Computer Applications Volume 167 – No.6, June 2017.

[12] P. Dabas and K. Khanna, "A study on spatial and transform domain watermarking techniques," International Journal of Computer Applications, vol. 71, no. 14, pp. 38–41, 2013.

[13] A. Al-Ataby, W. Al-Nuaimy, and M. A. M. Abdullah, Wavelet Transform-Multidisciplinary Applications. InTech, 2012, pp. 255–250.

[14] G. C. Langelaar, I. Setyawan, and R. L. Lagendijk, "Watermarking digital image and video data. a state-of-the-art overview," IEEE Signal Processing Magazine, vol. 17, no. 5, pp. 20–46, 2000.

[15] Qaiser, Shahzad & Ali, Ramsha. Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. International Journal of Computer Applications. IJCA 2018.

[16] Kim, S., Gil, J. Research paper classification systems based on TF-IDF and LDA schemes. Hum. Cent. Comput. Inf. Sci. 9, 30 (2019).

[17] Mamoun A. Awad and Issa Khalil. "Prediction of User's Web-Browsing Behavior: Application of Markov Model". Ieee Transactions On Systems, Man, And Cybernetics—Part B: Cybernetics, Vol. 42, No. 4, August 2012.

[18] G.Tripepi, K.J.Jager, F.W.Dekker, C.Zoccali. "Linear and logistic regression analysis". Kidney International, Volume 73, Issue 7, 1 April 2008.

[19] Behera, Rabi & Das, Kajaree. "A Survey on Machine Learning: Concept, Algorithms and Applications". International Journal of Innovative Research in Computer and Communication Engineering, (2017).

[20] D. Meyer, "Support Vector Machines – The Interface to Libsvm in Package e1071", November 25, 2019.

[21] Hongyu Yang And Fengyan Wang. "Wireless Network Intrusion Detection Based on Improved Convolutional Neural Network". IEEE Access Volume 7, May 30, 2019.

[22] Hong Huang, Fanzhi Meng, Shaohua Zhou, Feng Jiang, And Gunasekaran Manogaran. "Brain Image Segmentation Based on FCM Clustering Algorithm and Rough Set". Ieee Access, New Trends In Brain Signal Processing And Analysis Volume 7, Feb 6, 2019.

[23] Vinod Sharma, Dr. Shiv Sakti Shrivastava. "Document Class Identification Using Fire-Fly Genetic Algorithm and Normalized Text Features". International Journal of Scientific Research & Engineering Trends, IJSRET Volume 6 Issue 1 Jan, 2020.

[24] Alan Díaz-Manríquez, Ana Bertha Ríos-Alvarado, José Hugo Barrón-Zambrano, Tania Yukary Guerrero-Melendez, And Juan Carlos Elizondo-Leal. "An Automatic Document Classifier System Based on Genetic Algorithm and Taxonomy". accepted March 9, 2018, date of publication March 15, 2018, date of current version May 9, 2018.

[25] Sankranti Srinivasa Rao. "Stock Prediction Analysis by using Linear Regression Machine Learning Algorithm". (IJITEE), ISSN: 2278-3075, Volume-9 Issue-4, February 2020.

[26] Stephan Dreiseitla and Lucila Ohno-Machado. "Logistic regression and artificial neural network classification models: a methodology review". Journal of Biomedical Informatics 35 (2002) 352–359.

[27] Mittal, K., Aggarwal, G. & Mahajan, P. Performance study of K-nearest neighbor classifier and K-means clustering for predicting the diagnostic accuracy. Int. j. inf. tecnol. 11, 535–540 (2019).

[28] Gerard Biau. "Analysis of a Random Forests Model". Journal of Machine Learning Research 13 (2012) 1063-1095.

[29] K.Keerthi Vasan, B.Surendiran. "Dimensionality reduction using Principal Component Analysis for network intrusion detection". Perspectives in Science Volume 8, September 2016, Pages 510-512.

[30] YongchangWang and L. Zhu, "Research and implementation of SVD in machine learning," 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan, 2017, pp. 471-475,

[31] Matthew Botvinick, Sam Ritter, Jane X.Wang, Zeb Kurth Nelson, Charles Blundell, Demis Hassabis. "Reinforcement Learning, Fast and Slow". Trends in Cognitive Sciences, Volume 23, Issue 5, May 2019, Pages 408-422.

[32] Ganesh R. Naik and Dinesh K Kumar. "An Overview of Independent Component Analysis and Its Applications". School of Electrical and Computer Engineering, Informatica 35 (2011) 63–81.

[33] Nazih Sendia, Nadi Abchiche-Mimouni, Farida Zehraoui. "A new Transparent Ensemble Method based on Deep learning". Procedia Computer Science Volume 159, 2019, Pages 271-280..

[34]. Gourav Rahangdale, Manish Ahirwar and Mahesh Motwani. "Application of k-NN and Naive Bayes Algorithm in Banking and Insurance Domain". International Journal of Computer Science Issues (IJCSI) Vol. 13 (No.5) Sept. 2016.

[35] Niti Arora and Mahesh Motwani. "A Distance Based Clustering Algorithm" International Journal of Computer Engineering & Technology (IJCET) Vol.5(No.5) May 2014.