Proceedings of the 4th International Conference on Statistics: Theory and Applications (ICSTA'22)
Prague, Czech Republic – July 28- 30, 2022
Paper No. 129
DOI: 10.11159/cist22.129

Towards Deep Learning: A Comprehensive Overview on PSO with Machine Learning

Yaya Sylla[1], Adama Coulibaly[2], Pierre Morizet[3]
[1]ENSEA
Abidjan, Côte d'Ivoire
Yaya.sylla@ensea.edu.ci
[2]Université d'Abidjan
Abidjan, Côte d'Ivoire
adama.coulibaly@ensea.edu.ci
[3]Université de Technologie de Compiègne
Compiegne, France
Pierre.morizet@utc.fr

Abstract - In this era of deep learning where AI and machine learning models are used in well-known applications such as speech to text, real-time translation, image recognition, requires a large amount of data for the training. To process a large amount the data, model efficiency is a major concern, for this, there are few optimizations' methods being developed by a different researcher. In our research, we briefly discussed the famous optimization method such as PSO and their implementation with different machine learning models and perform experiments with different volumes of textual data and overview the performance with or without PSO algorithm, besides we also utilize a similar approach with neural networks. Experimental results show the proposed overview significance on data volume.

Keywords**:** Optimization; Data integrity; Machine learning; PSO: Big data

1. Introduction

Artificial intelligence makes it possible to automate complex problems using machine learning and deep learning. This state-of-the-art technology enables efficient execution of iterative aspect tasks In an organization, AI provides conveniences freeing up resources for priority tasks. AI-based applications have been widely used and are having a huge impact globally. Through this, scientists have developed different machine and deep learning algorithms to fully meet the requirements. To create an AI-based application, it requires a large amount of data to train the model and get the expected output. Machine learning algorithms at another end are used for this purpose. Among them, a few algorithms are widely used such as logistic regression, KNN, SVM etc… to process data in terms of classification and predictions [3]. Many researchers have worked to create effective methods for the application perspective, in addition to a few algorithms also developed for optimization. To conduct an in-depth study of AI-based methods with efficiency, few approaches are used such as PSO, and we applied it with different machine, deep learning algorithms to validate the performance of different methods with multiple records. In this research, we will study ML and DL algorithms in relation to optimization with different volumes of data. The impact of optimization with or without in ML is also with deep learning.

The motivation of this research is to study the comprehensive analysis of different Machine learning models and deep learning algorithms with PSO and without PSO for data processing and classification purpose. The objectives of this paper are as follows:

- To offer automated and robust data processing mechanism.
- To allow data-based processing of real-time data of different levels.
- To compare PSO-SAW-based data classification mechanism with Machine learning models on textual data.

2. Machine Learning

Machine learning (ML) is a subfield of Artificial Intelligence and can learn from different data and process the concerned task. ML is the most widespread research direction and played a considerable role in different areas, image recognition, voice recognition, real-time translation, and recommended system is examples of a machine learning algorithm.

In recent years machine learning algorithms are used in various applications such as IoT, to solve different problems. In IoT robotics is the best example for ML-based algorithm, in which it learns from the previous task and executes the job in real-time [1]. Moreover, ML algorithms are used to forecast the event by learning from previous data, such as weather prediction, financial operations in banking, or tracking Covid patients on a location-based [2]. There are a different number of algorithms are available for numerous use cases, each algorithm has a distinct feature for a different category of data. Moreover, in artificial intelligence, the computer has the ability to solve complex problems but to give predictions and perform classification where the machine learning comes [4].

As per the time, data is increasing in massive amount and the complexity increased during optimization, the researcher faces more challenges to make it more efficient during training, such optimization algorithm is being introduced to solve respective problems. ML is used to give a solution for a range of problems, nevertheless, applications such as robotics is the best example of AI with the usage of ML base algorithm, Robot vision [5], navigation process used by AI [6], field Robotics [7], Humanoid Robotics, in medical AI is also being used [8]. Due to the adaptation of AI in a different field, it's become popular in the last few years. AI becomes an essential part of human life in current days. Furthermore, machine learning which is a subfield also has categories known as supervised [9], unsupervised [10], and reinforcement learning [11]. In this section, we are going to discuss a different type of machine learning algorithm and we also give a brief overview of different optimization methods.

2.1. Supervised Machine Learning

In this type of machine learning where the learning is taking place by using the existing data for the training set and splitting them into for validation known as supervised learning. In supervised the input variable is used as (X) and the output known as (Y) which is being used for classification during training. The function used in this scenario is processed as input to output Y= function(X). Classification, regression is a major part of supervised machine learning [12]. For sequential data regression algorithms is one of the examples whereas linear regression, Decision tree, Arima e.tc for classification logistic regression, SVM, KNN are notably used in supervised machine learning.

2.2. Linear Regression

In supervised machine learning, linear regression is used for continuous data. It is also used for forecasting data such as basic weather forecasting, in real estate price forecasting also used in other industries such as oil firs, financial sector, etc. linear regression takes an input between two or more variables. Whereas X is known as the independent variable and Y is known as a dependent variable. The dependent and independent variables demonstrate the linear correlation. In this machine learning method where the analysis is done based on a linear model called linear regression. In linear regression, it finds the most suitable parameters and creates a straight line using data points, as shown in Figure 1.

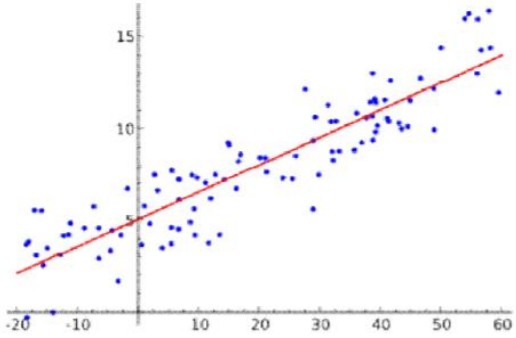


Figure 1 Linear Regression

2.3. Logistic Regression

In this supervised machine learning method, which is mainly used for classification, the difference between linear and logistic regression its data point is not coordinated. The structure of datapoints is settled in a pile, whereas each pile is known for a different category in labelled form as shown in figure 2.

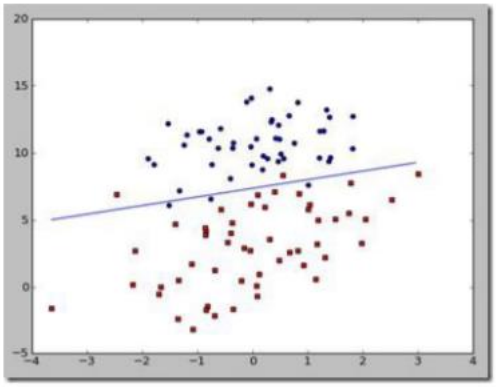


Figure 2 Logistic Regression.

In this supervised algorithm it used for applications such as spam filter, win or lose, classification for cat and dog e.tc. To perform such method data is required for training, and during training it use optimization algorithm to locate the best regression coefficient [13]. The input is obtained by using a function known as;

$$p(x) = \frac{1}{1 + e^{-(x-\mu)/s}}$$

Therefore μ is a position parameter (the midpoint of the curve, where p(μ)=1/2) s is known as scale proportional to the variance. The above function is used to get the range from positive infinity to negative. 0, 1 is the range for dependent values. Moreover, the sigmoid function is extensively used during training. Logistic regression-based classification is given an arbitrary set of inputs, and then the output is obtained by a function, which is the classification of the input data. For example, to simplify processing, when classifying, the function output 0 or 1 in the two classifications represents two classes. According to the actual needs and the above analysis, the above function argument range is from positive infinity to negative infinity. The dependent variable range is 0 or 1. Many functions satisfy the above conditions. The most intuitive one is the 0-1 step function. However, the step function is not steerable at the step point, which is not conducive to mathematical processing. Therefore, the Sigmoid function is now widely used, as shown in fig 3

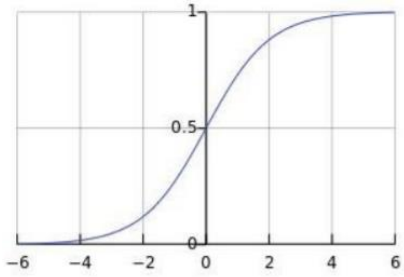


Figure 1 Sigmoid function

2.4. KNN

In supervised learning KNN which is known as K-Nearest Neighbour are commonly used in machine learning for classification and extensively used in many applications [14]. In KNN each sample belongs to category. For the classification there are few factors take place during training such as setting up the K- Value distance volume, and decision rule. Whereas "the minority obeys the majority" [15] used for the optimization. Value for k is selected on the bases of distance to perform training in optimized manner. There is different procedure take place to perform optimized training by various researchers.

PSO is being used by Chen et-al [16] to optimized weights of KNN. GWO is also used by Xu et-al [17]. TS [18] is being used by Tahir et-al for the selection of K value. Similarly other optimized algorithms are also being used such as evolution DE [19], whale algorithm [20]

2.5. Naïve Base

The Naïve Bayes (NB) classifier is a supervised probabilistic learning technique that is regularly used for classification purposes, it is simple but very accurate. It is based on Bayes' theorem where the conditional probability of an event X occurring, for an index Y, is determined by Bayes' rule:

$$P(X/Y) = P(X)* P(Y/X) / P(Y)$$

The N.B. has a disadvantage of being sensitive in feature selection, which may affect its accuracy. The feature-based classification methods developed in these studies produce good accuracy [27]. However, this classification has a limitation which is unable to fulfill the assumption of independence between attributes. This can have an effect on the level of precision [28].

2.6. SVM

In 1992 -1995 Boser et al. introduced a method which is known as Support vector machines (abbreviated SVM) Linear support vector machines (SVMs) are the origin form of binary classification. SVM is a method whose main objective is to find linear separators that separate between different classes. The SVM finds a hyperplane with the largest possible margin [25]. The SVM takes as input a large number of examples and assigns them to different classes. The formed data which are examples considered as points in space. The point we obtained belongs to one of the classes which are pre-separated, and the distance obtained is called hyperplane. When we have a new input data, we put it in the SVM, it will be assigned to one of the classes after a separation by the margin. The larger the margin, the lower the generalization error of the classifier [26].

3. Optimization

In artificial intelligence to process data and provide prediction in after training in efficient way is the biggest challenge, to overcome such problem different optimization algorithm produced by various researchers, in which it uses to maximize and minimize the function to reduce the error. It varies on model learning parameters for the computation of the target value (Y) from predictor value (x) being used. In the neural network, there are some weights (w) and the bias (b) used as learning parameters during computations and gives output updates, in optimal solution after minimizing the loss function in the training process of the neural network. The main purposed of using optimization algorithm is to minimize or maximize a loss function using gradients value w.r.t parameters, one of the most useful algorithms is Gradient Descent, SGD, ADAM, PSO etc.

3.1. Particle Swarm Optimization (PSO)

Particle swarm optimization (PSO) is an evolutionary algorithm that uses a population of possible solutions to develop an optimal solution to the problem. The first version of PSO was proposed by Russel Eberhart (electrical engineer) and James Kennedy (socio-psychologist) in 1995 [Kenn 95]. The particle test is a set of simple agents, known as particles. Each of these particles is considered as a solution to the problem, whereas it has a position (the solution vector) and a velocity. In addition, each particle has a memory allowing it to memorize its best performances (in position and in value) and the best performance obtained by the "neighbouring" (informant) particles: each particle has a group informant, historically called his neighbourhood.

Related Work

A study based on the optimization of the parameters of Deep Learning models using Particles swarm optimization, was targeted to build a Deep Learning model to predict the number of occupants at a given location in 15, 30 and 60 minutes. Three parameter sizes were used, namely: size 10, 25, 50 particles. In this study, the dataset used is 6 weeks in 7 parts; each part corresponds to a day of the week. Each data set has the following features: access point ID (APID), date, time, user MAC address and building number. It should be recalled that the result of this study once again showed the influence of PSO in Deep Learning algorithms. [21]

In the study on the decision tree combined with the selection of PSO features for sentiment analysis, the dataset uses the review of online transport users in Indonesia. The algorithm used is the decision tree and the PSO algorithm implemented in Python. Several scenarios were realized using several PSO parameters. At the end of all the scenarios, the PSO gives good results when applied to the decision tree, it is up to the engineer to select the right parameters in order to obtain a satisfactory result. [22]

The study on the prediction and classification of heart disease using machine learning algorithms optimized by particle swarm optimization and ant colony optimization. In this study, the FCBF (Fast Correlation-Based Feature Selection) method was exploited to filter out redundant features to improve the quality of heart disease classification. Subsequently, the classification algorithms were used and optimized by Particle Swarm Optimization (PSO) combined with Ant Colony Optimization (ACO) approaches. The dataset used is a heart disease dataset; the results show the efficiency and robustness of the proposed hybrid method. We obtain a classification of 99.65% using the optimized model proposed by FCBF, PSO and ACO. [23]

The study on improving the accuracy of the C4.5 algorithm using the characteristic selection and the bagging technique of a particle swarm (PSO) in the diagnosis of breast cancer. The aim of this study was to select the functionalities of the Particle Swarm Optimization algorithm in order to obtain the best result. To do this, the data used was the Breast Cancer Wisconsin (Original) Data Set (1992) obtained from the UCI machine learning repository. At the end of this study of the application of PSO as a selection and application function to improve the performance of the C4.5 algorithm by increasing the precision of 5.11% with an initial precision of 93.43% at 98.54%. [24]

3.1. Methodology

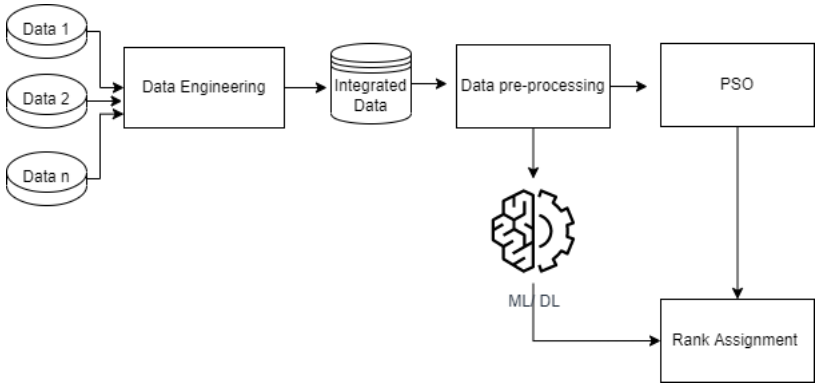


Figure 2 Proposed framework for the size and content-based classification

In our proposed method we are more focused on distinguishing the efficiency of different machine learning algorithm by using PSO. In this mechanism, it automates the classification on big dataset by using different supervised machine learning and deep learning algorithms. The motivation of this research is to validate the accuracy of text-based data, of different levels. We have divided our dataset into three categories on the bases of data quantity, such as low medium and large. We managed to collect real time data for sentiment analysis using tweets and IMBD reviews. In our case, there are several features, along with data variability. Then, by using the feature selection method, we are more concern to validate the PSO performance on different level of data. The overall approach of our model is given below in Fig.4 it shows the core part of our methodology and shows how we manage to validate the performance of different models. In the beginning of data pre-processing, we collect data and perform the analysis. After that integrate data and perform the data engineering for the feature selection. After the data pre-processing part, we get the final weight of our data. The data is being processed into two parts training and testing. Data is being send to the various machine learning and deep learning algorithms, for training purpose. In ML and Deep learning model, it initializes the input layer for NN and added few hidden layers for maximum performance during training use the best weight from pooling layer and get the predicted value. We use four NN by using our own text data for classification. Our proposed model also managed to get the highest model output and gives the best model regarding their training and testing performance. Similarly, we also use machine learning algorithms such for classification such as KNN, Random Forest, Logistic Regression, SVM for training and testing. We use KNN because it uses a multiclass

classification method, and this algorithm works best on textual data for more than two categories. In this scenario our model also cross validates the model performance and select the best model output. We are also using the PSO in parallel for same data to get the comprehensive analysis of model output. In figure 5 the overview working of proposed algorithm by using PSO in machine learning algorithms the structure is given below:

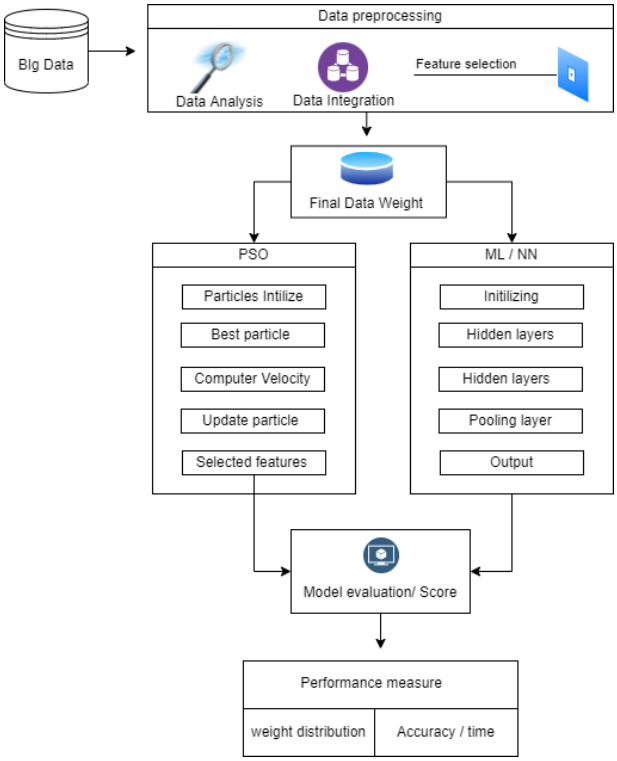


Figure 3 System architecture for the applied approach for data handling and classifying

Pseudocode

Step 1: $x_1 - X_{n-1}$ weights have been counted as the population of particles.

Step 2: The below equation is used for velocity of different particles based on different weight size.

$$V_i(t+1) = w_i(t) + c_1\, r_1\left[x_i'\ \ (t) - x_i(t)\right] + \ c_2 r_2\,[g(t) - \ x_i(t)]$$

In which the *vi (t)* denotes the particle velocity at time t.

xi (t) Denotes the particle position at time t.

x̂ i (t): Denotes the individual best of particle at time t.

g(t): Denotes the swarm best at time t that is global best.

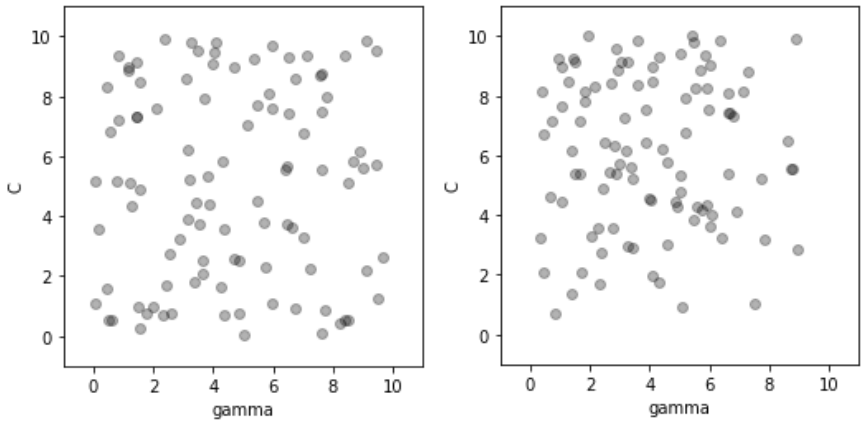
w: Inertia coefficient. c1 and c2: Cognitive coefficient

Step 3: In this step the position of current weight is associated with the updated weights and gives the best position as per requirements. All the weights are also updated. The range is for *w* is 0.8 to 1.2 and the value for c_1 and c_2 is given as 2.

Step 4: in this process the following procedure will take place:

$$x_i(t+1) = x_i(t) + v_i(t+1)$$

Step 5: the position is being set and assigned to each weight and classification is determined.



In above figures it shows the particle displacement in order to use PSO with machine learning during training.

We have also utilized Audio base data for audio classifier/ recognizer and give accuracy analysis w.r.t. of different Machine Learning models.

Experiment Results

To perform our experiments we consider different dataset, firstly we get a real-time data from tweeter feeds, using trending hashtags. For data-segmentation we perform different analysis. Besides we also collect some data from IMDB movie review. We perform data engineering and successfully made test and training data sample. Our framework has been developed by using Jupyter notebook with python 3.8. the empirical solution was performed by using as Intel® Core™ i7–7200 U CPU running at 2.8 GHz with 16 GB RAM. System type is 64-bit operating system and×64-based processor. For more effective results we use cross-validation with the dataset, the parameters are considered such as data size, content-based attribute weight, average attribute weight. We did a comprehensive analysis for different data range by using different ML and NN.

Table 1 shows the training accuracy with PSO of Machine Learning model

	Normal Data		Medium Data		Big Data	
	Training Accuracy	With PSO	Training Accuracy	With PSO	Training Accuracy	With PSO
Decision Tree	53%	56%	44%	46%	36%	36%
Naïve Base	49%	51%	48%	48%	52%	52%
Logistic Regression	55%	56%	33%	34%	33%	33%
KNN	52%	69%	55%	61%	58%	59%

Table 2 Shows the training accuracy of different size of data of Deep Learning model

	Normal Data		Medium Data		Big Data	
	Training Accuracy	With PSO	Training Accuracy	With PSO	Training Accuracy	With PSO
Simple ANN	99%	99%	55%	46%	54%	36%
GRU	99%	99%	77%	48%	71%	52%
LSTM	97%	96%	76%	64%	71%	33%
CNN	99%	99%	52%	61%	49%	59%

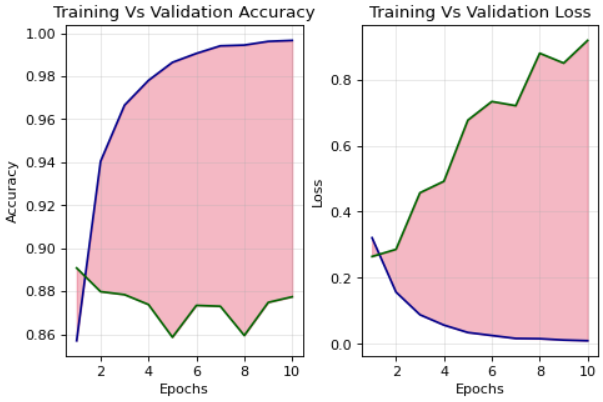
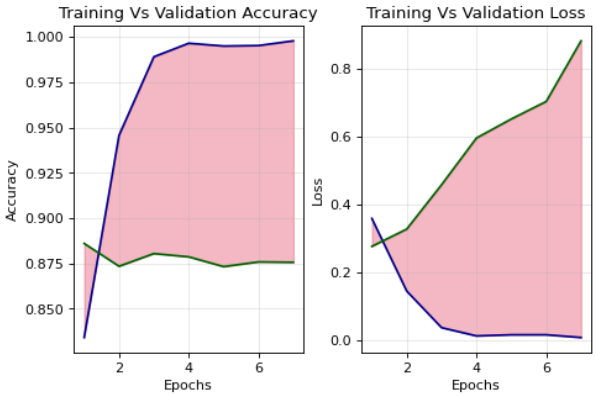
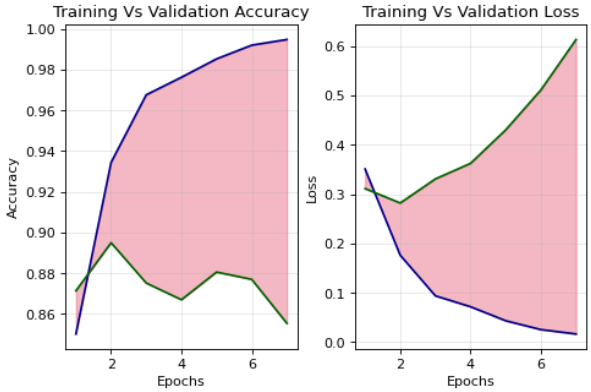


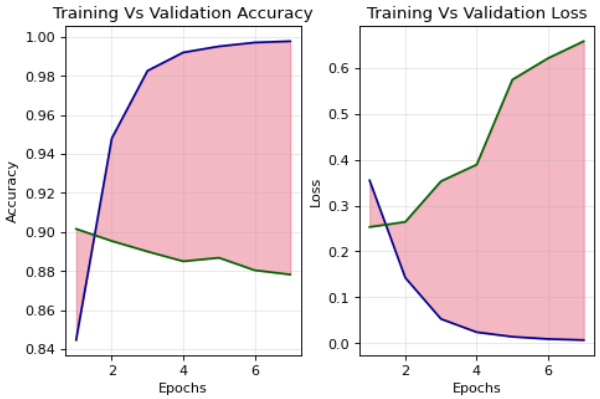
Fig 4: ANN Accuracy for Normal Data (a)



(b) CNN Accuracy for Normal Data



(c) LSTM Accuracy for Normal Data



(d) GRU Accuracy for Normal Data

Conclusion

In this research we give an overview of different machine learning algorithm on real-time collected data and perform the classification for sentiment analysis. presented an approach to prepare a data to select best model. In addition, we solved a real-world problem with the help of Machine Learning and neural network. We performed data analytics to find meaningful relationships, then cleaned the data and trained models on it. We give a comprehensive overview of different machine learning algorithm w.r.t to PSO. We also applied PSO on NN to overview the significance on similar dataset. We apply NN and ML on different size of dataset to verify the significance by using PSO. As per our analysis PSO works fine with only few algorithms, as per the data increase the effect of PSO doesn't make empirical effect on model performance. Although a lot of research are implemented in the field of Optimization by using different way for different data. But many challenges are still needed to address. One of the core challenges is system heterogeneity and statistical heterogeneity. Our future approach will be using huge amount of data (image and textual) and use different resource for computation such as Massive Parallel Processing systems which are solely use for data pre-processing, and analysed the accuracy w.r.t efficiency of data. Beside we will also apply optimization algorithm in MPP and validate the performance.

Acknowledgements

A short acknowledgement section can be written between the conclusion and the references. Sponsorship and financial support acknowledgments should be included here. Acknowledging the contributions of other colleagues who are not included in the authorship of this paper is also added in this section. If no acknowledgement is necessary, this section should not appear in the paper.

References

The reference of this research paper is following:

[1] Knox, W.B., Glass, B.D., Love, B.C., Maddox, W.T., Stone, P.: How humans teach agents. Int. J. Social Robot. 4(4), 409–421 (2012)

[2] Yin, Y.; Zhao, Y.; Li, C.; Zhang, B. Improving Multi-Instance Multi-Label Learning by Extreme Learning Machine. Appl. Sci. 2016, 6, 160, doi:10.3390/app6060160.

[3] S. Rongheng, Applied Mathematical Statistics (3rd Edition) ,CA: Science Press, 2014.

[4] Michalski, R.S., Carbonell, JG., Mitchell, TM. (eds.): Machine Learning: an Artificial Intelligence Approach. Springer Science & Business Media (2013)

[5] Rosten, E., Drummond, T.: Machine learning for high-speed corner detection. In: Computer Vision–ECCV, pp. 430–443. Springer, Berlin (2006)

[6] Sofman, B., et al.: Improving robot navigation through self-supervised online learning. J. Field Robotics. 23, 59–75 (2006)

[7] Yang, S.Y., Jin, S.M., Kwon, S.K.: Remote control system of industrial field robot. In: 6th IEEE International Conference on Industrial Informatics, pp. 442–447 (2008)

[8] Taylor, R.H., Menciassi, A., Fichtinger, G., Dario, P.: Medical robotics and computerintegrated surgery. In: Handbook of Robotics, pp. 1199–1222. Springer, Berlin (2008)

[9] Battiti. R., Brunato. M.: Reactive search optimization: learning while optimizing. In: Handbook of Metaheuristics, pp. 543–571. Springer, US (2010)

[10] Stavens, D., Thrun, S. A.:self-supervised terrain roughness estimator for off-road autonomous driving. arXiv preprint arXiv:1206.6872 (2012)

[11] Kober, J., Bagnell, J.A., Peters, J.: Reinforcement learning in robotics: a survey. Int. J. Robot. Res. 32(11), 1238–1274 (2005)

[12] Wright, Raymond E. "Logistic regression." (1995)

[13] Z. Qian, Meng. Deyu, Xu. Zongben, "L_(1/2) Regularized Logistic Regression," Pattern Recognition and Artificial Intelligence, vol. 25(05), pp. 721-728, 2012.

[14] Iliyasu, A.M.; Fatichah, C. A Quantum Hybrid PSO Combined with Fuzzy K-NN Approach to Feature Selection and Cell Classification in Cervical Cancer Detection. Sensors 2017, 17, 2935.

[15] Liu, Gaoyuan, et al. "An Enhanced Intrusion Detection Model Based on Improved kNN in WSNs." Sensors 22.4 (2022): 1407.

[16] hen, M.; Guo, J.; Wang, C.; Wu, F. PSO-based adaptively normalized weighted kNN classifier. J. Comput. Inf. Syst. 2015, 11, 1407–1415.

[17] Xu, H.; Fang, C.; Cao, Q.; Fu, C.; Yan, L.; Wei, S. Application of a Distance-Weighted KNN Algorithm Improved by Moth-Flame Optimization in Network Intrusion Detection. In Proceedings of the 2018 IEEE 4th International Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS), Lviv, Ukraine, 20–21 September 2018; pp. 166–170.

[18] Glover, F.; Laguna, M. Tabu Search, Handbook of Combinatorial Optimization; Springer: Boston, MA, USA, 1998; pp. 2093–2229.

[19] Storn, R.; Price, K. Differential Evolution—A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. J. Glob. Optim. 1997, 11, 341–359.

[20] Mirjalili, S.; Lewis, A. The Whale Optimization Algorithm. Adv. Eng. Softw. 2016, 95, 51–67.

[21] Qolomany, B., Maabreh, M., Al-Fuqaha, A., Gupta, A., & Benhaddou, D. (2017, June). Parameters optimization of deep learning models using particle swarm optimization. In 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC) (pp. 1285-1290). IEEE.

[22] Primartha, R., Tama, B. A., Arliansyah, A., & Miraswan, K. J. (2019, March). Decision tree combined with PSO-based feature selection for sentiment analysis. In Journal of Physics: Conference Series (Vol. 1196, No. 1, p. 012018). IOP Publishing.

[23] Khourdifi, Y., & Bahaj, M. (2019). Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization. International Journal of Intelligent Engineering and Systems, 12(1), 242-252.

[24] Saputra, R. H., & Prasetyo, B. (2020). Improve the accuracy of c4. 5 algorithm using particle swarm optimization (pso) feature selection and bagging technique in breast cancer diagnosis. Journal of Soft Computing Exploration, 1(1), 47-55.
[25] The Vanishing Gradient Problem During Learning Recurrent Neural Nets And Problem Solutions. Sepp Hochreiter. International Journal Of Uncertainty, Fuzziness And Knowledge-Based Systems
[26] Classification of Sentimental Reviews Using Machine Learning Techniques. Abinash
[27] K. Schoefegger, T. Tammet and M G, 2013 A survey on socio-semantic information retrieval Comput. Sci. 8 p. 25–46.
[28] Wisnu H Afif M and Ruldevyani Y, 2020 Sentiment analysis on customer satisfaction of digital payment in Indonesia: A comparative study using KNN and Na¨ıve Bayes J. Phys. Conf. Ser. 1444, 1.
[29] Sheraz Ali, Muhammad Maaz Irfan, Abubakar Bomai and Chuan Zhao "Towards Privacy-Preserving Deep Learning: Opportunities and Challenges." 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA). IEEE, 2020.
[30] Irfan Muhammad Maaz, Sheraz Ali, Irfan Yaqoob and Numan Zafar "Towards Deep Learning: A Review On Adversarial Attacks." 2021 International Conference on Artificial Intelligence (ICAI). IEEE, 2021.