ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Special Issue 1, August 2021

HYPER PARAMETER OPTIMIZATION IN STACKED DEEP NEURAL NETWORK FOR MEDICAL DIAGNOSIS

Ambily Merlin Kuruvilla
Research Scholar
Department of Computer Applications, Karpagam Academy of Higher Education, Coimbatore, India.
ambilykuruvilla@gmail.com

Dr. N V Balaji
Dean, Faculty of Arts, Science and Humanities, Karpagam Academy of Higher Education
Coimbatore, India.
balajinv@karpagam.com

Abstract— Diabetes is a metabolic disease where the blood sugar rate of an individual is consistently above normal. Due to the modern lifestyle and work culture, diabetics is widespread and affects the productivity and quality of life for an individual. A diabetics patient is at a very high risk of various health issues like organ failure and even it can even result in loss of life. An early prediction of this chronic disease can avoid health issues and save many lives. The aim of this article is to develop a better predictive model for diabetics using an automated hyper parameter optimization (HPO) approach in Multilayer Perceptron (MLP). This article provides an efficient way to increase the accuracy of Neural Network to a substantial level through the HPO process using Grid Search Optimization (GSO) through the stacking ensemble model. In order to run the ensemble model at an optimal level and to minimize errors, appropriate hyperparameters must be calculated. Three GSO methods are utilized to tune the hyper parameters. To build the stacking ensemble model, the PIMA data set was used.

Keywords: HPO, MLP, GSO: Hyper Parameter Optimization, Multilayer Perceptron, Grid Search Optimization.

INTRODUCTION

A study carried out by the WHO recently revealed that in 2016, diabetics was one of the leading causes of death worldwide. Diabetes has resulted in 1.6 million fatalities in 2016 and this statistic replaces HIV / AIDS with diabetes as one of the most frequent cause of death [4]. The burden of diabetes disease grew from 108 million in 1980 to 422 million [5] in 2014, and the percentage of diabetic patients amongst adults over 18 years of age rose from 4.7% in 1980 to 8.5% in 2014[5]. 642 million people i.e. (1 in 10 people) are expected to contract diabetes by 2040. 46.5% of people with diabetes have not been diagnosed officially [6]. This makes it necessary to develop techniques and procedures to assist in the early detection of diabetes in order to reduce the number of deaths related to diabetes, as late diagnosis is responsible for a majority of deaths linked to diabetes [7].

There is a need to implement sophisticated information processing to develop cutting-edge strategies for the early detection of diabetes. Data mining tools can also be effectively applied. The ability to remove and uncover previously unseen, secret, yet important patterns from a large database repository is given by data mining [7]. These tools can assist medical evaluation and decision making.

LITERATURE SURVEY

Roshan Birjais [1] conducted a research in many classification algorithms for diabetes prediction and his team found out that the Gradient Boosting algorithm outperform other classifiers and

obtained 86% accuracy. They have analysed various prime factors for the cause of diabetes disease.

Md. Maniruzzaman [2] wrote an article in which he used LR for identifying the prime features for the diagnosis of diabetes disease. Accuracy and Area Under curve are used as the performance measures for the classifiers. It revealed that LR along with the Random forest generates the maximum accuracy.

Atik Mahabub[3] used an ensemble voting classifier for forecasting diabetes disease. The Out of eleven classifiers, the best performing three will be used for the ensemble classifier. Accuracy, Precision, F-Measure and Recall [3] are used as the parameters for evaluating the classifiers. The outcome clearly shows that the ensemble method outperform the base classifiers.

BACKGROUND STUDY

Neural network is used for the prediction. Parameter tuning mechanism in neural network is applied for improving the accuracy.

ARTIFICIAL NEURAL NETWORK

Artificial Neural Network [17] is built by multiple nodes that reproduce the human brain's biochemical neurons. The neurons are linked and are communicating with each other. The nodes are capable of taking input data and executing the operations. The outcome of these operations is transferred to other neurons. The output is referred to as its activation or node value for a node. A technique known as Gradient Descent in the Artificial Neural Network, which takes places in the backpropagation period whereby it intends to regularly resample the gradient of the model parameter in the reverse direction based on the weight 'W', periodically updating till the global minimum of G(W) function is reached.

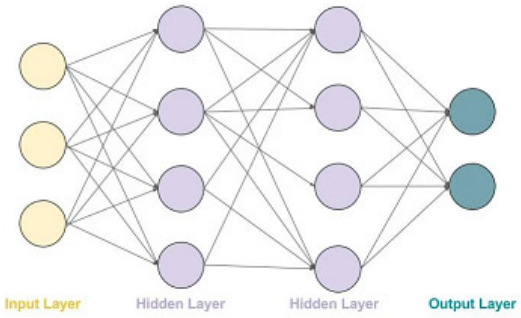

Figure 1: Hidden Layers in Artificial Neural Network

A loss represents the prediction error found in the Artificial Neural Network. In deep learning, this is estimated as a loss function. The Loss Function describes the model's operating efficiency. For the estimation of the loss function, stochastic gradient descent was used. For each iteration, weights are updated, and the model is trying to reach the global minimum point.

LOGISTIC REGRESSION

Logistic Regression is a well-known classification algorithm used to estimate the probability of a target variable. The design of the goal or dependent variable is binary, indicating that only two possible groups are available. Mathematically, f(Xi), a LR model predicts P(Yi=1).

b=wt0+wt1 x1+wt2x2+…+wtnxn.

A sigmoid function can be used to map the predicted values to the probabilities.

S(y)=1/(1+e(-y))

S(y) is the estimated output; y is the value inputted to the function and e is the base of natural log.

NAVIE BAYES

The Navie Bayes algorithm follows the Bayasian principle which belongs to a category of conditional probabilities (CP). The CP is the probability that an event B will happen, provided A has already occurred. Though Bayes Theorem provides a principled way for calculating conditional probability, in practice its computationally expensive and thus using some assumptions, bayes theorem is simplified by making some assumptions and turning it into an effective classification model referred to as Naive Bayes. Conditional probability can yield the probability of an event using prior information.

P(H/E) = (P(E/H) * P(H))/(P(E))

PROPOSED MODEL

Stacking Ensembles

To obtain better performance, ensemble methods allow for combining the results of many methods. More the models, better will be the performance of the ensemble strategies. An ensemble method where the no. of models is stacked in a way that their observations act as input to a new model is called Stacking.

Hyper Parameters in Neural Network

The variables that determines how the network is trained are called the Hyperparameters. Hyper Parameters also determine network structure. These variables are set before the training phase, i.e. before the weight and bias is optimized.

Network Weight Initialization

Based on different activation functions applied to each layer, it is preferred to use separate weight initialization schemes.

Activation function

It is used to apply nonlinearity to frameworks that will permit nonlinear prediction restrictions to be learned by deep learning models. The most popular among the activation functions is the rectifier activation function. While making predictions, for binary, Sigmoid is used while for multi-class predictions, softmax is deployed in the output layer.

Gradient Descent

The learning rate determines how fast its parameters are modified by a network. The learning process is slowed down by a low learning rate, but converges efficiently. The higher learning rate accelerates learning, but does not converge. A decreasing learning rate is usually prioritised.

Momentum

Momentum is used to eliminate oscillations and to know the course of the next step with knowledge of the previous phase. For momentum, usually, a value between 0.5 and 0.9 is used.

Number of epochs

The number of epochs measures the number of times the entire training data is given to the model during the training process. The number of epochs is raised until the accuracy begins to decline, while the accuracy of the training is improved due to overfitting.

Batch size

Refers to the no. of sub samples provided to network before the parameter is updated.

METHODOLOGY

Diabetes prediction using MLP Model is described in the following steps. In this model these steps are used to predict diabetes more accurately.

Data Collection

PIMA Indian Data set [17], from the UCI repository is used for the analysis. The data set consists of features including age, number of pregnancies, Body Mass index, Diabetes Pedigree Function, Glucose Level [17] etc…

HYPER PARAMETER TUNING USING GRID SEARCH.

It is one of the traditional hyper parameter tuning method [13]. Before the learning process starts, the value of the hyper parameter needs to be calculated. Grid Search is also known as a comprehensive search [13], with each mixture of hyper parameters explored by Grid Search. This means that each variation of the hyper parameter values listed would be tried. There can be several parameters for models, and it can be viewed as a search problem to find the right combination values to the parameters. The purpose of algorithm tuning is to find the best values of the parameter corresponding to the particular problem. Grid Search can be expanded to provide the highest results by using automatic approaches for finding optimum values to the parameters.

HYPER PARAMETER TUNING USING GRID SEARCH.

Also known as a comprehensive search, it's a traditional hyper parameter tuning method. The value of the hyper parameter needs to be measured before the start of the learning process. Here, each variation of the hyper parameter values is explored which means that each variation of the hyper parameter values that are listed is tested. On further expanding, grid search can provide the highest results using automatic approaches that would optimum values for the parameters.

Implementation

Input values are converted to vectorised format using Logistic regression. The output of Logistic Regression is saved as a collection in Python and this will be given as the input for the MLP. Grid Search is used for doing the parameter optimization. The model optimizes the following hyper parameters

- Epoch
- Batch Optimization.
- Gradient descent

By using Keras Classifier Grid parameters such as Batch size and Epochs are optimized. In this model stochastic Gradient is introduced for training the Artificial Neural Network. The internal parameters such as epoch and Batch values are optimized and the model is automized so the best values for Epoch and Batch optimization is find out and it is assigned. This leads to an increase in accuracy and reduced the loss and MSR (Mean Square Error). The Vanishing Gradient Problem is overcome by using ReLU (Rectified Linear unit) activation function. The formulae for ReLU is

R= {0, z}

{1 for z > 0}

{0 for z < 0}

The input of each neuron is passed to the activation function and there it is processed. ReLU overcoming the problem of vanishing gradient by keeping the derivative value as positive. So always there is a difference between Wold and Wnew in ReLU. Feedback connections were introducing LSTM and the best score for grid search is found out. By using standard Grid Search Value of Epoch and Batch size is automated and it is found out that for this dataset the best result was obtained for Batch size =20 and Epoch=20. Forward propagation is improved by adding the features of LSTM [12]. So a new memory part was introduced to store [12] the activation details of the hidden layers. Various steps involved in this model are described below:

- Choose the data set
- Convert all the input values to Vectorised format for standardization Find out the training values
- Initialise the model using Keras Classifier
- Three sub models are introduced, and, in each model, Optimal parameters are finding out using hyper parameter tuning in Grid Search Model.
- In model1 the best values for Epoch and Batch size was find out using Grid search in Keras Classifier. The model is tested with three batch sizes (10,20 and 30) and three epochs (10,20 and 30). All combinations were tested and the best accuracy will be find out for the combinations of batch size 20 and epoch 20.
- In Model2 Forward propagation is improved by introducing LSTM and Adam Optimizer. By automating the model best values for epoch and batch size was predicted. Learning rate and dropout rate is find out and best classifying accuracy is calculated.
- In Model3 ReLU activation function is introduced along with Softmax function. Best predicted result was observed.
- Stacked the three Models using tensor flow method and the accuracy matrix was printed. Logistic Regression is used for stacking the Models.

FINDING OPTIMAL PARAMETERS USING HYPER PARAMETER TUNING IN GRID SEARCH MODEL.

- In model1 the best values for Epoch and Batch size was find out using Grid search in Keras Classifier. The model is tested with three batch sizes (10,20 and 30) and three epochs (10,20 and 30). All combinations were tested and the best accuracy will be findout for the combinations of batch size 20 and epoch 20.
- In Model2 Forward propagation is improved by introducing LSTM and Adam Optimizer. By automating the model best values for epoch and batch size was predicted. Learning rate and dropout rate is find out and best classifying accuracy is calculated.
- In Model3 ReLU activation function is introduced along with Softmax function. Best predicted result was observed.
- Stacked the three Models using tensor flow method and the accuracy matrix was printed. Logistic Regression is used for stacking the Models.

RESULT AND DISCUSSION

From the observations it is found out that the accuracy is incremented in the new model comparing with the base classifiers. Here three base classifiers are used for the accuracy comparison. Naive Bayes, Logistic Regression, Multilayer Perceptron. From the observations it is found out that the automated hyperparameter optimization with stacked ensembled model increases the accuracy level.

Algorithm	Accuracy	Precision	Recall
Navie Bayes	76.3	75.9	76.3
Logistic	77.2	76.7	77.2

Regression			
MLP	78.3	75.0	74.0
Stacked Model	90.7	94.0	82.0

Table 1: Performance Comparison of Classifiers

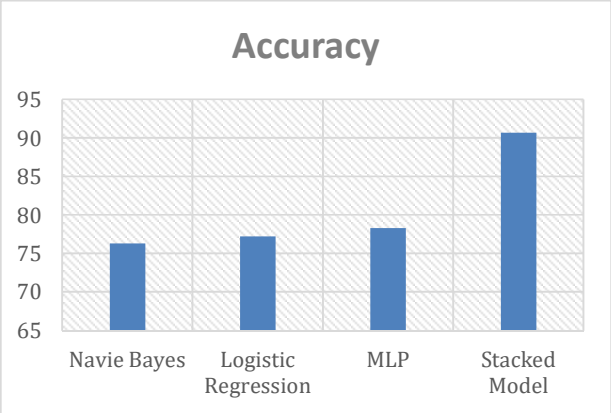

Figure 2: Classifier Performance based on Accuracy

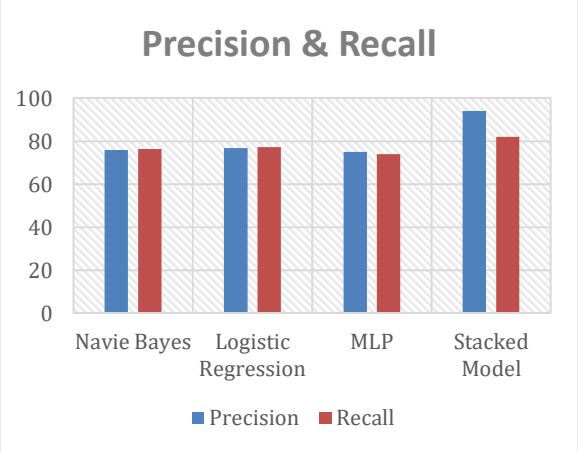

Figure 3: Classifier Performance based on Precision and Recall.

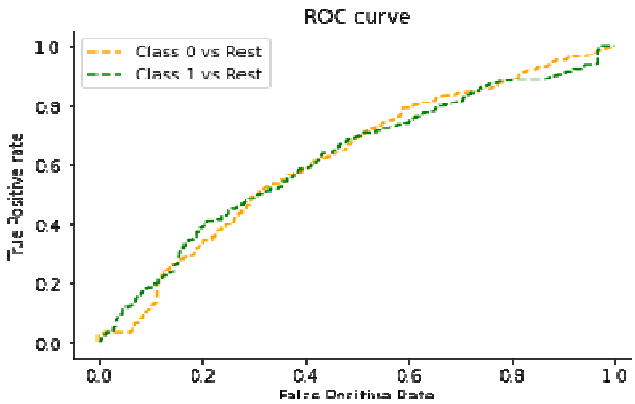

Figure 4: ROC Curve

ROC curve is a graphical plot that shows the system's diagnostic potential as its discriminating threshold is varied. The ROC curve is generated by plotting the true positive rate at different threshold settings against the false positive rate.

CONCLUSION

In this article we are introducing an ensembled hyper parameter tuning mechanism to tackle the deficiencies and improving the accuracy. For this purpose, we have used an automated stacked ensemble method which combines various hyper parameters. Grid Search Optimisation method is used, and three different models were created as base learners using Neural Network by combining various activation functions. The optimum value for each parameter is calculated and stored into an external file by each model. Three output files are created, and these output files are inputted to a logistic regression model which is a learning model. We have used LR Model as the learning model. It is found out that this model improves the accuracy to good extend.

REFERENCES

1.Roshan Birjais, Ashish Kumar Mourya,Ritu Chauhan, Harleen Kaur " Prediction and diagnosis of future diabetes risk : a machine learning approach". Springer.

2.Md.Maniruzzaman,JahanurRahman, BenojirAhammed, Md.Menhazul bedin "Classification and prediction of diabetes disease using machine learning paradigm", Springer

3. Atik Mahabub "A robust voting approach for diabetes prediction using traditional machine learning techniques", Springer.

4.www.geeksforgeeks.org/ml-chi-square-test-for-feature-selection

5.www.niddk.nih.gov/health-information/diabetes/overview/what-is-diabetes Retrieved.

6.www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death, Accessed 27th Jul 2018.

7.www.who.int/news-room/fact-sheets/detail/diabetes retrieved 27/07/2018.

8.www.diabetesdaily.com/learn-about-diabetes/what-is-diabetes/how-many-people-have-diabetes.

9.Changsheng Zhu, Christian Uwa Idemudia "Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques, ScienceDirect.
10.https://towardsdatascience.com/ways-to-detect-and-remove-the-outliers-404d16608dba
11.https://mc.ai/deep-learning-in-the-real-world-how-to-deal-with-non-differentiable-loss-functions/
12.Merijn Beeskma, Suzan Verberne, Antal van den Bosch Predicting life expectancy with a long short-term memory recurrent neural network using electronic medical records.
13.Sampath Kumar Palaniswamy, Hyperparameters tuning of ensemble model for software effort estimation, Springer.
14.Aramesh Rezaeian, Marzieh Rezaeian "Prediction of mortality of premature neonates using neural network and logistic regression", Journal of Ambient Intelligence and Humanized Computing, Springer.
15. AH. Habbi and M. Zelmat "Fuzzy Logic Based Gradient Descent Method with Application to a P1-type Fuzzy Controller Tuning: New Results", IEEE 2017.
16. Huaping Zhou, Raushan Myrzashova and Rui Zheng "Diabetes prediction model based on an enhanced deep neural network" EURASIP Journal on Wireless Communications and Networking, Springer.
17. Kalaiselvi, G. M. Nasira, "A New Approach for Diagnosis of Diabetes and Prediction of Cancer Using ANFIS", IEEE
18. Pahulpreet Singh Kohli, Shriya Arora "Application of Machine Learning in Disease Prediction", IEEE.
19. Abdulhakim Salum Hassan, I. Malaserene, A. Anny Leema , "Diabetes Mellitus Prediction using Classification Techniques", International Journal of Innovative Technology and Exploring Engineering (IJITEE)
20. Sai Poojitha Nimmagadda, Sagar Yeruva, Rakesh Siempu, "Improved Diabetes Prediction Model for Predicting Type-II Diabetes", International Journal of Innovative Technology and Exploring Engineering (IJITEE).