Extreme Learning Machine with sigmoid activation function on large data

R. R. S. Ravi kumar, G. Apparao

Abstract: This paper describes an efficient algorithm for classification in large data set. While many algorithms exist for classification, they are not suitable for larger contents and different data sets. For working with large data sets various ELM algorithms are available in literature. However the existing algorithms using fixed activation function and it may lead deficiency in working with large data. In this paper, we proposed novel ELM comply with sigmoid activation function. The experimental evaluations demonstrate the our ELM-S algorithm is performing better than ELM,SVM and other state of art algorithms on large data sets.

Index Terms: MachineLearning, Activation function,Sigmoid, Classification

I. INTRODUCTION

Machine Learning is a subset of AI is the scientific investigate of algorithms and statistical models turn this way systems use to automatically learn and improve from experience without being explicitly programmed.

ELM is a neural network model used for knowledge discovery and knowledge representation. It was used and designed by Guang-Bin-Huang. Extreme Learning machines are timber move forward neural networks for grouping, fading, clustering, sparse approximation, compression and feature Sophistication with a single layer or multiple layers of guarded nodes, where the parameters of adjacent to nodes (not just the weights connecting inputs to hidden nodes) need not be tuned[1][2]. These hidden nodes can be assigned randomly and not at all updated (i.e. they are projected random but with nonlinear transforms), or can be inherited from their ancestors without being changed. In most outstanding cases, the collect weights of hidden nodes are usually learned in a single step, which essentially amounts to learning a linear model.

Machine Learning algorithms are adjusted for identifying these intrusions in a given input,these algorithms can be categorized into batch and incremental algorithms. The factual algorithms are computatioanally weak and not easily scalable to unstinting datasets for knowledge extraction and representation.

Several distributed algorithms exist for extracting intrusions in large data sets, these algorithms have a disadvantage with their implementation method or execution environment on which they are running. The explanation behind this is the detection of intrusions is an iterative procedure and the current structures are not appropriate for iterative algorithms

In this paper, we introduce a SVM, ELM-S algorithms on large data set.To summarize following contributions:

1.We explore ELM-S framework to support classification tasks in large data sets.

2.We develop a general method, and accordingly present the ELM-S method.

3.Finally we conduct extensive experiments to demonstrate the our ELM-S in classification

II. RELATED WORK

All the algorithms discussed by wang et al[2],kuang et al[4],Aburomman[5],Fernaaz and Jabbar[8],to improve the performance of intrusion detection systems. These intrusion discovery calculations (for example SVM,ELM) which are executed utilizing appropriated structures like Jupyter to process bigger informational collections.The algorithms is executed dependent on wang.et.al[1] proposed an interruption identification structure dependent on SVM and approve their strategy on the NSL-KDD informational dataset. They claimed that their method, which has 99.92% effeciveness rate was superior to other approaches; but,they do not mention used data sets statistics, number of training and testing samples. Furthermore, the SVM show performance decrease when large data are involved and it is not an ultimate choice for analyzing huge network traffic for intrusion detection. Second algorithm for intrusion detection is proposed by kuang et.al. which applied a hybrid model of SVM and KPCA with GA to intrusion detection,and their system showed 96% detection rate .Intrusion detection systems provide assistance in detecting, preventing and resulting unauthorized access. Thus Aburomman and Reaz [5] proposed an ensemble classifier method which is a combination of PSO and SVM. This order classification performed superior to different methodologies with 92.90% precision.

The above algorithms utilized the learning advancement and information mining of intrusion informational dataset.. The above algorithms utilized the learning advancement and information mining of intrusion informational data. Which has number of disadvantages, for example, the SVM isn't suitable for substantial information, for example, observing the high transmission capacity of the network [6][7].

Moreover the Support Vector Machine is anything but a decent decision for huge information investigation since its presentation debases as

information size increments. SVM'S also not preferred for heavy data sets because of the high computation cost and poor performance[10][11].So to overcome all these challenges, we introduced ELM to decrease false alarms and to increase the detection rate.ELM learns faster and attain higher generalization capability as compared with other feed forward hidden layer biases to minimize the training time .Hence we named all algorithm as the Multi Layer Perceptron -Extreme Learning machine in short ELM-S.

III. PROPOSED MODEL

3.1. Data Set Description:

The Performance of the system is based on the correctness of the data set.The more accurate the data the greater the effectiveness of the system.The data sets used are KDDcup, Mushroom, Iris, Wine and Adult. The information about the data sets is depicted in UCI machine learning repository.

3.2. Support Vector Machine(SVM)

In this section, we discuss a detailed an analysis of the support vector machine classifier method. SVMs were at first proposed by Vapnik(1995) for taking care of issues of grouping and relapse analysis.SVM is a managed learning system that is prepared to order various classifications of information from different disciplines[1]. These have been utilized for two-class characterization issues and are material on both direct and non straight information classification.

Algorithm:

Step1: SVM makes hyperplane or numerous hyperplanes in a high-dimensional space, and the best hyperplane in them is the one that ideally isolates information into various classes with the biggest detachment between classes. A non-Linear classifier utilizes different bit capacities to boost edges between hyperplanes.

Step2: The kernel function uses squared Euclidian separation between two numeric vectors and maps input information to a high dimensional space to ideally isolate the given information into their individual assault classes. In this classifier method, every one of the recreations have been directed utilizing the openly accessible library package.

Step3:Given that the chosen problem is a multi class classification problem, the algorithm uses the notion of one versus all for attack classification. In this notion, the multiclass problem is divided into a two-class problem. The radial basis function(RBF) bit is utilized in this calculation. Which is spoken to as follows

$$K(a,b)=e^{-x\|a-b\|2},x>0$$

For the given training samples (a_i,b_i),i=1,2,,3,.....n,Where i is the maximum number of samples in the training data, $a_i \sum R^n$ and $y_i \sum\{1,-1\}$,where 1 shows samples from a positive class and -1 represents sequences from the negative class.

Step 4:

The solution is obtained by

$$\min_{w,b} 1/2\ W^TW+C_{i=1}\sum i$$

Step 5:The observed results via the algorithm used in the SVM model are not significantly convincing as the maximum margin classifier sometimes fails to find any separating hyperplane.To overcome this problem soft margin is used soft margin uses positive slack variables in the constraints.

Step 6: The final outcomes created from the above

advances are then delegated the yield classes, into their separate attack classes.

3.3. Extreme Learning Machine: The extreme learning machine (ELM) has attracted much attention over the past decade due to its fast learning speed and convincing generalization performance[1][3]. In this section, we discuss a detailed analysis of the Extreme Learning Machine Classifier technique algorithm.ELM is used to solve various classification, clustering, regression, -and feature engineering problems[12][13].ELM is another name for single or multiple hidden layer and output layer. The Learning algorithm involves input layer, one or multiple layers and the output layer.. ELM with K hidden neurons described as follows:

$$\sum_{a=1}^{K}\beta_a\ g(w_a.x_b+b_a)= O_b,b=1,2,3,..........N$$
Using training data set $\{(\ x_a,t_b)\}_{a=1}^{N}$ with X_b

3.4 Deficiencies in ELM:

The activation function is one of the crucial factors in neural networks. While using random hidden node parameters in ELM brings a fast training speed and easy implementation advantages, using fixed activation function also leads to a deficiency in the ELM[15]. In the ELM, the hidden node parameters are independent of input data and the activation function fixed during learning. Relying merely on solving the output weights through linear equations could lead to poor generalization performance for irregular distributions in the large data scaling[16].

3.5 Extreme Learning Machine with sigmoid(ELM-S):

In this, we proposed ELM-S to overcome the deficiencies of ELM [3]listed above. We presented activation function in ELM-S. Output from every neuron is generated after applying activation function to the values being calculated with set of inputs and their weights. Here we are utilizing sigmoid activation function. The sigmoid activation[17][3] capacity is composed as underneath.

$y = \sigma(w^T+b)$ where $\sigma(z)=1/1+e^{-z}$

If z is very large then e^{-z} is close to zero and $\sigma(z)=1/1+0 =1$

If z is very small then e^{-z} is large and $\sigma(z)=1/1+$large number $=0$

The algorithm of Extreme learning machine with sigmod activation function as below.

Algorithm :

Input: Training data set,Hidden node number,Set of regularization parameters,Activation function

Output:Accuracy

1.Generate hidden node weight vector and bias randomly.

2.Calculate net input matrix (V) of hidden layer.

3.Estimate Sigmoid parameter for zero-near gaussian distribution.

4.Find the matrix output of hidden layer.

   $H=g(s.V+t.l)$

5.Use the algorithm for optimization

5.1. Iteration t

5.2.Compute dw,db

   $V_{dw}=\beta_1\ V_{dw}+(1-\beta_1)dw$
   $V_{db}=\beta_1\ V_{db}+(1-\beta_1)db$
   $S_{dw}=\beta_2\ S_{dw}+(1-\beta_2)dw^2$
   $S_{db}=\beta_2\ S_{db}+(1-\beta_2)db^2$

IV. EVALUATION

In this section, we analyze the experiments we conducted, to prove ELM-S algorithm is performing better than other existing Machine Learning on various data sets.The algorithms compared with ELM-S algorithms are ELM,SVM.

4.1 Execution Environment:

The Experiments are conducted on Python and Jupyter framework. The Algorithm is implemented by using SKLearn. The performance of the algorithm is assessed with extensive experiments conducted using UCI ML repository. The algorithm executed on Mushroom, KDDcup,Iris,wine and German data sets etc. UCI ML Repository by considering various properties Accuracy, Precision, and Recall.

4.2 Results and Analysis:

We present here the results of ELM-S Algorithm. The results mostly focus on the precision, Recall and F1-Score. Here we presented the comparison of the results of both the algorithm shown in fig.In all of the cases the ELM-S performed better than SVM and saves time while processing on the given Input data sets.

To find the performance of an algorithm,we consider AOC/ROC properties .These are precision,Recall,Accuracy and F1-Score.The Accuracy of SVM,ELM-S on 20% testing and 80% training data samples is shown in fig.

1.ELM performing better than SVM.

2.If the data set is small SVM better perform than ELM.

3.If the data set is large ELM perform better than SVM.

To illustrate the capability of the proposed ELM-S framework to handle large data with complex classification, we run several experiments on some real world data sets from UCI ML repository. In detail, we compare the performance of our proposed ELM-S method against a number of state-of-the art classification algorithms, including SVM and ELM .

4.2.1Evaluation Metrics:

Precision: Precision is the ratio between the number of correct positive predictions and the total number of positive predictions.

Recall: Recall is the ratio between the number of correct positive predictions and the total number of positives.

Accuracy: Accuracy is the ratio of the total number of correct prediction, True Positive plus True Negative and the total number of a dataset Positive plus Negative.

F-measure: F-measure is the compound measure of precision and recall combination.

4.3.Experimental Results:

Table 1.demonstrates the precision of both the ELM-S and SVM approach on 20% testing and 80% training data samples. The precision of ELM-S is better than that of SVM Linear and RF on the full data samples, and it also outperforms that of RF.

Data sets	ELM-S	ELM	SVM
KDD Cup	97.50	95.40	94.50
Iris	97.40	94.80	94.40
Wine	97.60	95.50	94.40
Mushroom	97.50	95.30	94.30
Wisconsin	97.80	96.20	94.30

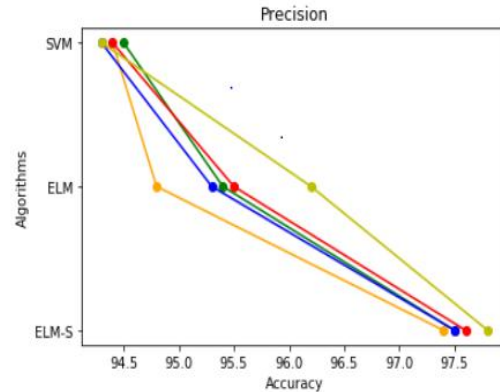Table 1: Comparison by precision with ELM-S and other Classification methods

Fig 1: Precision of Novel ELM,ELM and SVM(80%Training and 20% Testing)

The recall of NOVEL ELM and SVM approach is on on 20% testing and 80% training data samples comparatively shown in Fig.2. On full data samples the recall of NOVEL ELM outperforms ELM and SVM.

Data sets	ELM-S	ELM	SVM
KDD Cup	99.6	98.3	96.4
Iris	98.7	97.5	95.3
Wine	99.55	96.5	95.4
Mushroom	99.5	97	96.5
German	98.4	96	94.3

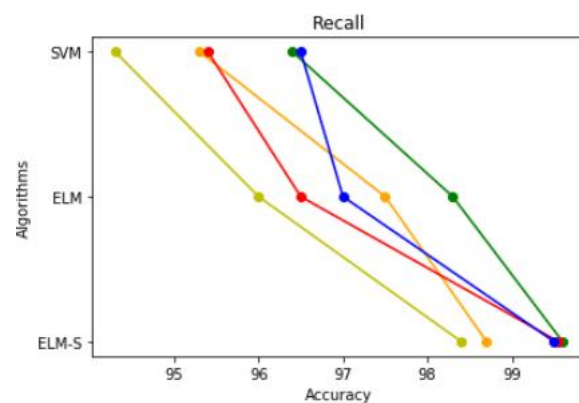Table 2: Comparison by Recall with NOVEL ELM and other Classification methods

Fig 2: Recall of ELM-S,ELM and SVM(80%Training and 20% Testing)

The Accuracy of ELM-S and SVM approach is on on 10% testing and 90% training data samples comparatively shown in Table 3. On full data samples the recall of ELM-S outperforms ELM and SVM.

Data sets	NOVEL ELM	ELM	SVM
KDD Cup	99.80	98.50	97.50
Iris	98.50	97.50	95.50
Wine	98.60	97.50	95.00
Mushroom	98.80	98.00	96.50
German	99.80	98.00	94.50

Table 3: Comparison by accuracy with ELM -S and other Classification methods
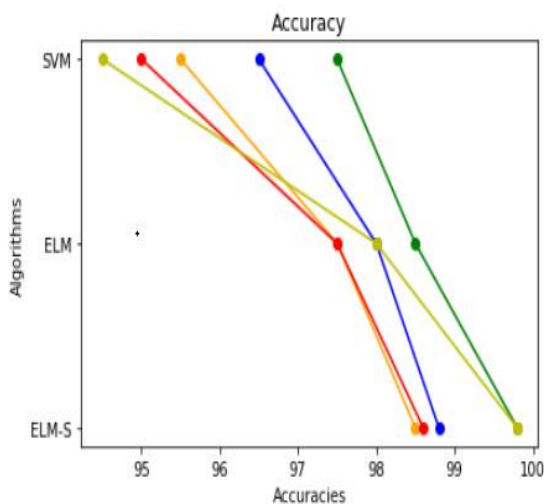

Fig 3: Accuracy of ELM-S,ELM and SVM(80%Training and 20% Testing)

V. CONCLUSION

There are several algorithms proposed for Extreme learning machine,but the approach discussed in this paper for scalable ELM for knowledge discovery in large data sets using jupyter framework. The implementation of ELM-S applied over large data sets.We have presented a detailed analysis of our results to show that ELM-S has performed well in the distributed environment.Experiments conducted on five bench mark data sets also show that the proposed ELM-S outperform several state of the art algorithm.

REFERENCES

1. Iftikhar Ahmad, Mohammad Basheri, Muhammad Javed Iqbal,And Aneel Rahim,"Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection"pp.33789-33795,2018
2. H.Wang, J. Gu, and S.Wang, ``An effective intrusion detection framework based on SVM with feature augmentation," pp. 130-139, Nov. 2017.
3. Jiuwen Cao , Kai Zhang, Hongwei Yong, Xiaoping Lai, Badong Chen , and Zhiping Lin," Extreme Learning Machine With Affine Transformation Inputs in an Activation Function",pp 1-15,2018
4. F. Kuang,W. Xu, and S. Zhang, "A novel hybrid KPCA and SVM with GA model for intrusion detection," pp. 178-184, 2014.
5. A. A. Aburomman and M. B. I. Reaz, "A novel SVM-kNN-PSO ensemble method for intrusion detection system," pp. 360-372, 2016.
6. M. R. G. Raman, N. Somu, K. Kirthivasan, R. Liscano, and V. S. S. Sriram, "An efficient intrusion detection system based on hypergraph Genetic algorithm for parameter optimization and feature selection in support vector machine", pp. 1-12, Oct. 2017.
7. S. Teng, N.Wu, H. Zhu, L. Teng, andW. Zhang, "SVM-DT-based adaptivecand collaborative intrusion detection", pp. 108-118, 2018.
8. N. Farnaaz and M. A. Jabbar, "Random forest modeling for network intrusion detection system," pp. 213-217,2016.
9. R. M. Elbasiony, E. A. Sallam, T. E. Eltobely, and M. M. Fahmy, "A hybrid network intrusion detection framework based on random forests and weighted k-means," pp. 753-762, 2013.
10. I. Ahmad and F. e Amin, "Towards feature subset selection in intrusion detection", 2014, pp. 68-73.
11. J. Jha and L. Ragha, "Intrusion detection system using support vector machine," pp. 25 30,2013.
12. S. M. H. Bamakan, H. Wang, T. Yingjie, and Y. Shi, "An effective intrusion detection framework based on MCLP/SVM optimized by time- varying chaos particle swarm optimization," pp. 90-102, 2016.
13. C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," pp.1-27,2011
14. Y. Liu, Y.Wang, and J. Zhang, "New machine learning algorithm: Random foest,"B. Liu, M. Ma, and J. Chang, Eds. Berlin, pp. 246-252,2012.
15. P. Potocnik and E. Govekar, "Practical considerations in training extreme learning machines,", pp. 1–5, 2015
16. C. M. Bishop, "Pattern Recognition and Machine Learning"a New York, NY, USA: Springer-Verlag, 2006.
17. Sridhar Swaminathan,"leadingindia.ai".

AUTHORS PROFILE

R.R.S.Ravikumar is currently working as Assistant Professor in Computer Science and Engineering Department,VJIT,Hyderabad. He received his B.Tech.in CSE from Pondicherry University,Puducherry.and receivd his M.Tech in CSE from Acharya Nagarjuna University,Guntur.He is pursuing Ph.D at GITAM,Visakhapatnam,India.

Dr.G.Appa Rao is currently working as Professor in Computer Science andEngineering Department, GIT,GITAM,Visakhapatnam. His research includes Software Engineering,Data mining,Machine Learning and Big data.