ISSN2394-3777 (Print)
ISSN2394-3785 (Online)
Available online atwww.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 10, March 2018

Performance Analysis of Intrusion Detection System Using Data Mining Neural Classifiers

Sahilpreet Singh[1], Amarvir Singh[2]

Research Scholar, Department of Computer Science, Punjabi University Patiala[1]

Assistant Professor, Department of Computer Science, Punjabi University Patiala[2]

Punjab, India

Abstract:- One of the central areas in network intrusion detection is how to build effective systems that are able to distinguish normal from intrusive traffic. Many researchers using data mining technique to build effective intrusion detection system. In this paper four different algorithms are used namely as Multilayer Perception, Radial Base Function, Logistic Regression and Voted Perception.All these neural based algorithms are implemented in WEKA data mining tool to evaluate the result and compare their relative performance. For experimental work, NSL KDD dataset is used. Based on study, it is concluded that Multilayer Perceptron classifier is most suitable neural algorithm as compare to other with high True positive rate and Low False positive rate.

Keywords:-Data Mining, Intrusion Detection System, Neural network classifier algorithm, NSL KDD dataset

I.INTRODUCTION

Data mining is the process of discovering interesting knowledge from large amounts of data stored in databases, data warehouses, or other information repositories. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use[5].It is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, and database systems.Data mining applications can use a variety of parameters to examine the data.

With the development ofinternet, network security becomes an indispensable factor of computer technology. The concept of Intrusion Detection System (IDS) proposed by Denning (1987) is useful to detect, identify and track the intruders. An intrusion detection system (IDS) is a device or software application that monitors network or system activities for malicious activities or policy violations and produces reports to a management station. The intrusion detection systems are classified as Network based or Host based attacks. The network based attack may be either misuse or anomaly based attacks. The network based attacks are detected from the interconnection of computer systems. The host based attacks are detected only from a single computer system and is easy to prevent the attacks. Data mining can help improve intrusion detection by addinga level of focus to anomaly detection[1].It helps in to classify the attacks to measure the effectiveness of the system. Classification is the process of finding the hidden pattern in data. With the use of classification technique it is easy to estimate the accuracy of the resulting predictive model, and to visualize erroneous predictions. The goal of classification is to accurately predict the target class for each case in the data. In Data Mining, various neural network algorithms are used to classify and analyse the attacks.

II.NEURAL NETWORK ALGORITHM

*A.*Multilayer Perceptron-

A multilayer perceptron (MLP) is a feed forward artificial neural network model that maps sets of input data onto a set of appropriate outputs.The term "multilayer perceptron" does not refer to a single perceptron that has multiple layers. Rather, it contains many perceptrons that are organized into layers An MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Except for the

ISSN2394-3777 (Print)
ISSN2394-3785 (Online)
Available online atwww.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 10, March 2018

input nodes, each node is a neuron (or processing element) with a nonlinear activation function. MLP utilizes a supervised learning technique called backpropagation for training the network. The multilayer perception consists of three or more layers (an input and an output layer with one or more *hidden layers*) of nonlinearly-activating nodes. The main advantages of this method are that they are easy to use, and that they can approximate any input/output map.

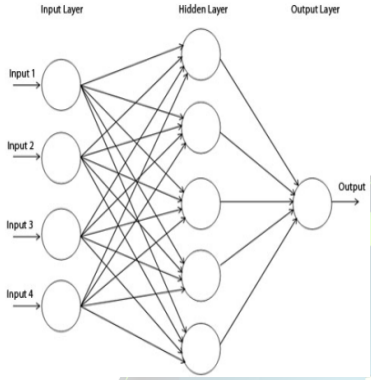

Fig 1. General structure of Multilayer Perceptron Neural Network

***B*.RBF Network-**

A radial basis function network is an artificial neural network that uses radial basis functions as activation functions. The output of the network is a linear combination of radial basis functions of the inputs and neuron parameters. Radial basis function networks have many uses, including function approximation, time series prediction, classification, and system control.It has three layers: an input layer, a hidden layer with a non-linear RBF activation function and a linear output layer. Due to the fact that RBF networks possess only one hidden layer, it simplifies the designing of the network. Also another important property of RBF networks is that while training it builds the network in incremental way, adding one neuron at a time unless a certain MSE (Mean Squared Error) is achieved or training epochs arereached.

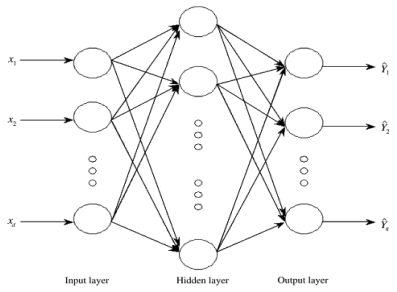

Fig 2. General structure of RBF Neural Network

***C*. Logistic Regression-**

Logistic regression is a type of regression analysis used for predicting the outcome of a categorical dependent variablebased on one or more predictor variables. That is, it is used in estimating empirical values of the parameters in a qualitative response model. Logistic regression measures the relationship between a categorical dependent variable and one or more independent variables. It can be binomial or multinomial. Binomial or binary logistic regression refers to the instance in which the observed outcome can have only two possible types (for example, "dead" vs. "alive"). Multinomial logistic regression refers to cases where the outcome can have three or more possible types (e.g., "better" vs. "no change" vs. "worse").

***D*. Voted Perceptron-**

Voted Perceptron neural network helps in replaces all missing values,and transforms nominal attributes into binary one. It helps in to predicting the outcome in binary one.

III.RELATED WORK

Jimmy Shum and Heidar A. Malki[1] presents a neural network-based intrusion detection method for the internet-based attacks on a computer network. Neural networks are used to identify and predict unusual activities in the system. Training and testing data were obtained from the Defense Advanced Research Projects Agency(DARPA)and showed promising results on detection intrusion systems using neural networks.XiangmeiLi[2] has presented difference between the attack categories, for this he had adjusted 41-dimensional input features of the neural-network-based multiple classifiers intrusion detection system and found that the every adjusted sub-classifier is better in convergence precision, shorter in training time than the 41-features sub-classifier and it also shows that intrusion detection system is higher in the detection rate, and less in the false negativerate.

MahbodTavallaeeet al.[3] has described that researcher hasovercome the weakness of signature-based IDSs in detecting novel attacks, and KDD' 99 cup is the mostly widely used data set for the evaluation of these systems. KDD'99 cup faces two main issue which effects the performance of system. Therefore a new data set, NSL-KDD, has been proposed which does not suffer from any mentioned shortcomings.Usman Ahmed and Asif Masood[4] presented the Intrusion Detection System which uses Radial Base Function Neural network whichprioritize the speed and efficiency of the training phase and also limits the

ISSN2394-3777 (Print)
ISSN2394-3785 (Online)
Available online atwww.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 10, March 2018

false alarm rate and it shows that result shows that the radial Basis Functions Neural Networks provide betterdetection rate and very low training time.

Jorge Blasco et al. [5] havebuild the effective intrusion detection system with the use of Genetic Programming algorithm. They concluded that with the use of Genetic Programming, efficiency and effectiveness of the system has been improved.S.Devaraju and Dr.S.Ramakrishnan [6]has comparedthe performance of intrusion detection withvarious neural network classifiers. For this, the performance of the full featured KDD Cup 1999 dataset is compared with that of the reduced featured KDD Cup 1999 dataset. With the help of MATLAB software the efficiency of the system has been measured and also proved that the reduced dataset is performing better than the full featured dataset**.**

Mohd. JunedulHaque, Khalid.W. Magld, Nisar Hundewale [7] has present focus on intrusion detection system based on data mining**.** The main part of Intrusion Detection Systems (IDSs) is to produce huge volumes of alarms. With the use of data mining they proposed a framework K-mean clustering, which is helpful for detect the intrusion.Purva Adlakha and PritiSubramanium [8] has classified the attacks and measured attacks in intrusion detection system by using Multilayer Perception based artificial neural network. For experimental result they use KDD'99 cup dataset and at last it shows that the proposed system not only detect attacks but also classify them in 6 groups with the accuracy of approximately 83% .

Rohit Arora and Suman [9] has compare the two classification algorithm J48 and Multilayer Perception to analysed better performance..After comparing both algorithm it found that, Multilayer Perceptron is better algorithm in most of the cases

IV.PROPOSED FRAMEWORK

In this section, we elaborate the whole framework of our new approach.

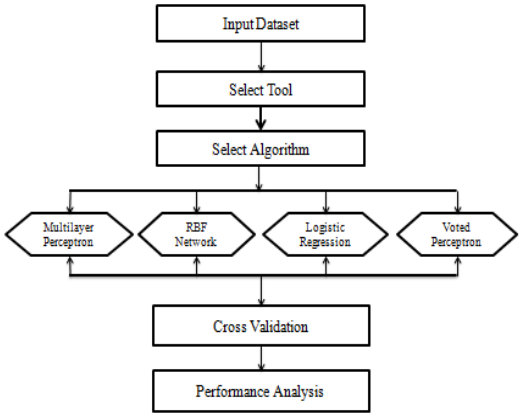

Fig. 1 Block diagram of proposed technique

a)Input and Read the Dataset- Firstly, NSL KDD dataset has been selected for classify purpose. There are 25192 classified instances and 41 attributes in this dataset, out of which 12 attributes are selected to evaluate performance.

b)Select Tool- After selecting the dataset, the next step is to load the dataset in proposed WEKA machine learning tool.

c)Select Algorithm- When the dataset has been loaded in the WEKA tool, next step is to test the dataset to measure the detection rate by selecting particular algorithm at a time.

d)Cross Validation- In this step, the dataset has been tested with the help of selected neural algorithm separately, for classifying the attacks or to measure the accuracy.

e)Performance Analysis- The last step, where the performance of each algorithm is evaluated and compared with respect to each other.

V. EXPERIMENTAL SETUP& RESULT

To evaluate the performance of our approach, a series of experiments were conducted. We carried out these experiments by implementing proposed intrusion detection system in machine learning tool.

TYPE	NAME	SOURCE
Tool	WEKA	www.cs.waikato.ac
Dataset	NSL KDD	UCI 99
Algorithm	MPNN, RBFNN, Logistic Regression, Voted Perception	Data mining Algorithm
Result	Microsoft Power Point	Microsoft Office

Table 1 Building blocks of IDS system

ISSN2394-3777 (Print)
ISSN2394-3785 (Online)
Available online atwww.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 10, March 2018

a)WEKA TOOL-

We have used WEKA toolkit to analyze the dataset with data mining algorithm. WEKA is open source application formally called Waikato Environment for Knowledge Learning, is a computer program that was developed at the University of Waikato in New Zealand for the purpose of identifying information from raw data gathered from agricultural domains. It supports many different standard data mining tasks such as data pre-processing, classification, clustering, regression, visualization and feature selection.

b)NSL KDD Dataset-

NSL-KDD is a data is used to solve some of the inherent problems of the KDD'99 data set.In NSL-KDD dataset there is no duplicate records in the proposed test sets; therefore, the performance of the learners are not biased by the methods. This dataset contains number of attributes, which are supportive for measure the attacks.

c)Algorithm-

The selected dataset is test with data mining algorithms. There are four types of data mining algorithm namely as Multilayer Perception, Radial Base Function, Logistic and Voted Perception are used to evaluate the performance. All these algorithms individually tested on a dataset.

d)Results-

The performance of each tested algorithm on a selected dataset is graphicallyexposed with the help of Microsoft Power Point. It makes easy to understand.

VI. PERFORMANCE EVALUATION

After applying cross validation on all four neural network algorithms, the performance of each algorithm on a conducted dataset is evaluated and the following results are obtained.

ALGORITHM	CCI (Correctly Classified Instances)	ICI (Incorrectly Classified Instances)	TP Rate (True Positive Rate)	FP Rate (False Positive Rate)
Voted Perceptron	82.37	17.62	0.78	0.2
RBF Network	90.12	9.87	0.9	0.11
Logistic	93.66	6.03	0.93	0.06
Multilayer Perceptron	94.94	5.05	0.94	0.05

Table 2. Performance Measure Result

From above table II, the graphical representation of evaluated parameter result by individually tested neural network algorithm is shown below.

ISSN2394-3777 (Print)
ISSN2394-3785 (Online)
Available online atwww.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 10, March 2018

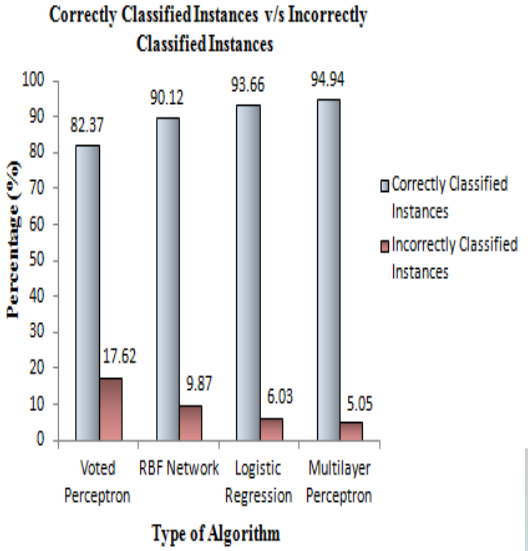

Fig 2.

Above Fig-II describes the correctly classified instances and the incorrectly classified instances of each algorithm. After classification of NSL KDD test dataset it is clearly shown that the Multilayer Perceptron algorithm shows the higher detection accuracy among all other algorithms.

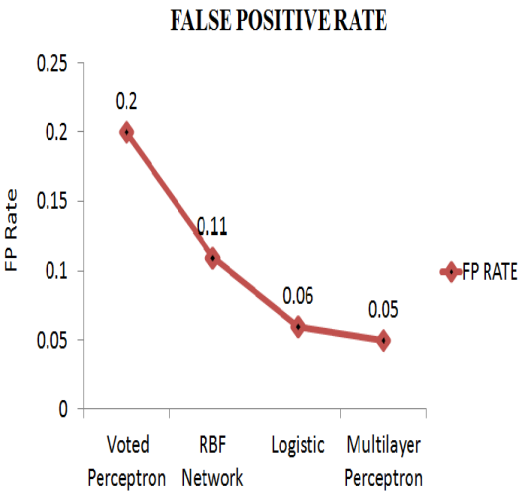

Fig 4.

For a good Intrusion Detection System, FP Rate i.e False Positive Rate should be low. Above figure shows that FP Rate of the Multilayer Perceptron, RBF Network,Logistic, Voted Perceptron when run with all the attributes. From fig IV, it is clear that Multilayer Perceptron algorithm has lowest TP Rate remaining than the other classifiers algorithms which is desirable.

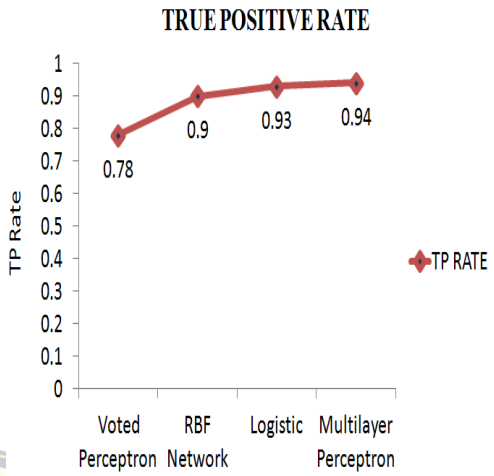

Fig 3.

For a good Intrusion Detection System, TP Rate i.e True Positive Rate should be high. Above figure shows that TP Rate of the Multilayer Perceptron, RBF Network, Logistic, Voted Perceptron, when run with all the attributes. From fig III, it is clear that Multilayer Perceptron algorithm has highest TP Rate remaining than the other classifiers algorithms which is desirable.

VII. CONCLUSION

This study approached the best classification algorithm application for machine learning to intrusion detection system. For this, we have presented different neural based data mining classifier algorithms to classify attacks in an efficient manner. After doing experimental work, it is clear that Mulitilayer Perceptron has the highest classification accuracy with the lowest error rate. To enhance the results the tree based classifier algorithms are applied to KDD test dataset and implemented using WEKA machine learning tool. We showed that machine learning methodology can be used in the field of Intrusion detection system.

REFERENCES

[1]Jimmy Shum and Heidar A. Malki,"Network Intrusion Detection System Using Neural Networks" Fourth International Conference on Natural Computation in IEEE 2008.

[2] Xiangmei Li, "Optimization of the Neural-Network-Based Multiple Classifiers Intrusion Detection System" in IEEE 2010.

ISSN2394-3777 (Print)
ISSN2394-3785 (Online)
Available online atwww.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 10, March 2018

[3] MahbodTavallaee, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set" proceeding of the 2009 IEEE Symposium on computational intelligence in Security and Defence Application.
[4] Usman Ahmed, Asif Masood, "Host Based Intrusion Detection Using RBF Neural Networks" in IEEE 2009 International Conference on Emerging Technologies.
[5] Jorge Blasco, Agustin Orfila, Arturo Ribagorda, "Improving Network Intrusion Detection by Means of Domain-Aware Genetic Programming" in IEEE 2010 International Conference on Availability, Reliability and Security.
[6]S.Devaraju, Dr.S.Ramakrishnan, "Performance analysis of Intrusion Detection System using various Neural Network Classifiers"in IEEE International Conference on Recent Trends in Information Technology, ICRTIT 2011.
[7] Mohd. JunedulHaque, Khalid.W. Magld, Nisar Hundewale "An Intelligent Approach for Intrusion Detection Based on Data Mining Techniques" in IEEE 2012.
[8] PurvaAdlakha,PritiSubramanium, "Detecting and Classifying Attacks in Network Intrusion Detection System Using Multi-layer Perceptron Based on Artificial Neural Network" in International Journal of Advanced Research in Computer Science and Software Engineering in Volume 3, Issue 6, June 2013.
[9] Rohit Arora, Suman, "Comparative Analysis of Classification Algorithms on Different Datasets using WEKA" inInternational Journal of Computer Applications (0975 – 8887) Volume 54– No.13, September 2012.
[10]Mrs. Sneha Kumari, Dr. Maneesh Shrivastava, "A Study Paper on IDS Attack Classification Using Various Data Mining Techniques" in International Journal of Advanced Computer Research in Volume-2 Number-3 Issue-5 September-2012.
[11]Chunlin Zhang, Ju Jiang, Mohamed Kamel, "Intrusion detection using hierarchical neural networks" in 2004 Elsevier.
[12] GuangqunZhai, Chunyan Liu "Research and Improvement on ID3 Algorithm in Intrusion Detection System" in IEEE 2010 Sixth International Conference on Natural Computation.
[13] Fan Li, "Hybrid Neural Network Intrusion Detection System using Genetic Algorithm" in IEEE 2010.
[14] Gang Wang, Jinxing Hao, Lihua Huang "A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering" in ELSEVIER 2010.
[15] S.A.Joshi, Varsha S.Pimprale "Network Intrusion Detection System (NIDS) based on Data Mining"International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 2, Issue 1, January 2013.
[16] N.S.Chandolikar, V.D.Nandavadekar "Comparative analysis of two algorithm for Intrusion attack classification using dataset" in International Journal of Computer Science and Engineering (IJCSE) in 2012.
[17] Shaik Akbar, Dr.K.Nageswara Rao, Dr.J.A.Chandulal "Intrusion Detection System Methodologies Based on Data Analysis" International Journal of Computer Applications (0975 – 8887) Volume 5– No.2, August 2010
[18]YacineBouzida, Frederic Cuppens "Neural networks vs. decision trees for intrusion detection" in 2011.
[19] Mark Hall, Eibe Frank "The WEKA Data Mining Software: An Update" Department of Computer Science University of Waikato Hamilton, New Zealand.
[20] Anazida Zainal, MohdAizainiMaarof, Siti Mariyam Shamsuddin "Data Reduction and Ensemble Classifiers in Intrusion Detection" in IEEE 2008 Second Asia International Conference on Modelling & Simulation.
[21] L Dhanabal, SP Shantharajah, "A study on nsl-kdd dataset for intrusion detection system based on c1assification algorithms", International Journal of Advanced Research in Computer and Communication Engineering, vol. 4, no. 6, pp. 446-452, 2015.
[22]Dikshant Gupta, Suhani Singhal, Shamita Malik, Archana Singh, Network intrusion detection system using various data mining techniques, pp. 1-6, 2016.
[23] Sasanka Potluri, Christian Diedrich, Accelerated deep neural networks for enhanced intrusion detection system, pp. 1-8, 2016.
[24] Ian H Witten, Eibe Frank, Mark A Hall, Christopher J Pal, Data Mining: Practical machine learning tools and techniques, Morgan Kaufmann, 2016.
[25] Manish Kulariya, Priyanka Saraf, Raushan Ranjan, Govind P. Gupta, "Performance analysis of network intrusion detection schemes using Apache Spark", Communication and Signal Processing (ICCSP) 2016 International Conference on, pp. 1973-1977, 2016
[26] Tuan A Tang, LotfiMhamdi, Des McLernon, Syed Ali Raza Zaidi, Mounir Ghogho, Deep learning approach for network intrusion detection in software defined networking, pp. 258-263, 2016.

ISSN2394-3777 (Print)
ISSN2394-3785 (Online)
Available online atwww.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 10, March 2018