© INNOVATIVE RESEARCH THOUGHTS | Refereed | Peer Reviewed | Indexed
ISSN : 2454 – 308X | Volume : 07 , Issue : 02 | April - June 2021

Feature Selection Based on Hybrid Technique in Intrusion Detection KDDCup's99 dataset

Pavan kaur
M.tech-IT
Research Scholar
GKU, Talwandi Sabo(Bathinda)
Psran35@gmail.com

Dr. Dinesh kumar
Associate Professor
Department of CSE
GKU,Talwandi Sabo(Bathinda)
kdinesh.gku@gmail.com

Abstract : Interruption location has turn into a basic segment of system organization because of the immeasurable number of assaults relentlessly debilitate our PCs. Customary interruption recognition frameworks are restricted and do not give a complete answer for the issue. They hunt down potential noxious exercises on system traffics; they once in a while succeed to discover genuine security assaults and oddities. Nonetheless, much of the time, they neglect to identify noxious practices (false negative) or they fire alerts when nothing incorrectly in the system (false positive). Moreover, they require comprehensive manual preparing and human master obstruction. Applying Data Mining (DM) strategies on system movement information is a promising arrangement that helps grow better interruption identification frameworks. Experimental results on the KDDCup'99 data set have demonstrated that our rare class predictive models are much more efficient in the detection of intrusive behavior than standard classification techniques.

I. INTRODUCTION

During the last few years there is a dramatic increase in growth of computer networks. There are various private as well as government organizations that store valuable data over the network. This tremendous growth has posed challenging issues in network and information security, and detection of security threats, commonly referred to as intrusion, has become a very important and critical issue in network, data and information security. The security attacks can cause severe disruption to data and networks. Therefore, Intrusion Detection System (IDS) becomes an important part of every computer or network system. Intrusion detection (ID) is a mechanism that provides security for both computers and networks. Feature selection and feature reduction is important area of research in intrusion direction system. The size and attribute of intrusion file are very large. Due to large size of attribute the detection and classification

mechanism of intrusion detection technique are compromised in terms of detection rate and alarm generation. For the improvement of intrusion detection process various authors and researchers work together for feature reduction and feature selection for intrusion detection system. In current scenario the feature reduction and selection process focus on entropy based technique[6]. Some authors used neural network model such SOM and RBF neural network model for classification of intrusion data during attacking mode and normal mode of network traffic. On the mechanism of detection intrusion detection divide into two section host based intrusion detection system and network based intrusion detection system. Host based intrusion detection system in generally know as signature based intrusion detection system. Instead signature based intrusion detection system come along with another variant is called anomaly based intrusion detection. In anomaly based intrusion detection various technique are used such as supervised learning and unsupervised learning. In network intrusion Detection, independent and redundancy attributes leads to low detecting rate and speed of classification algorithms. Therefore, how to reduce network attributes to raise performance of classification algorithms by applying optimal algorithm has become a research branch of intrusion Detection [8, 9]. A new approach for network intrusion detection feature selection based on PCNN-SVM attribute selection and reduction is presented in the paper. The available approaches for intrusion detection focus on improving detection accuracy and restraining false alarms, and given enough time, most of them can achieve satisfactory results in terms of these criteria. However, in practice, intrusion detection is a real-time critical mission, that is, intrusions should be detected as soon as possible or at least before the attack eventually succeeds. In addition, there is usually an initial training period for an intrusion detector to characterize the observable object's behavior, and most existing methods are based on the assumption that high quality labeled training data are readily available. Present a new approach; based on pulse Coupled Neural Networks (PCNN) to identify important input features for intrusion detection. Through identifying the important inputs and redundant inputs, a classifier can achieve the reduced problem size, faster training and more accurate results[1,3]. Then, applied modified Gaussian Support Vector Machines (GSVMs) based on training algorithm to, anomaly detection over noisy data. GSVMs effectively address the over-fitting problem introduced by the noise in the training data set. With GSVMs, the incorporation of an averaging technique in the standard support vector machines makes the decision surface smoother and controls the amount of regularization

automatically. Moreover, the training algorithm can significantly reduce training time with better generalization performance and fewer support vectors while maintaining high detection accuracy. They thus require less computational overhead and running time and so are more desirable for real time intrusion detection.

II. INTRUSION DETECTION SYSTEM

An intrusion is an attempt to compromise the integrity, confidentiality, availability of a resource, or to bypass the security mechanisms of a computer system or network. James Anderson introduced the concept of intrusion detection in 1980 [1].It monitors computer or network traffic and identify malicious activities that alerts the system or network administrator against malicious attacks. Dorothy Denning proposed several models for IDS in 1987 [2]. Approaches of IDS based on detection are anomaly based and misuse based intrusion detection. In anomaly based intrusion detection approach [3], the system first learns the normal behavior or activity of the system or network to detect the intrusion. If the system deviates from its normal behavior then an alarm is produced. In misuse based intrusion detection approach [4], IDS monitors packets in the network and compares with stored attack patterns known as signatures. The main drawback is that there will be difference between the new threat discovered and signature being used in IDS for detecting the threat. Approaches of IDS based on location of monitoring are Network based intrusion detection system (NIDS) [5] and Host-based intrusion detection system (HIDS) [6]. NIDS detects intrusion by monitoring network traffic in terms of IP packet. HIDS are installed locally on host machines and detects intrusions by examining system calls, application logs, file system modification and other host activities made by each user on a particular machine.

III. GOAL FOR WORK

In today's era detection of security threats that are commonly referred to as intrusion, has become a very important and critical issue in network, data and information security. Highly confidential data of various organizations are present over the network so in order to preserve that data from unauthorized users or attackers a strong security framework is required. Intrusion detection system plays a major role in providing security to computer networks.

Our Objectives are as follows:

- The goal is to detect the Intrusion from network from different dataset using KNN , SVM and GA in Weka tool

- A comparative analysis of different feature selection methods based on KDDCUP'99 benchmark dataset.
- To evaluate the performance are evaluated in terms of detection rate, root mean square error and computational time.

IV. METHODOLOGY

The research work is to detect the intrusion from network. It is based upon weka tool. Their are the programmable files containing the information about the dataset. The Intrusion detection system deals with large amount of data which contains various irrelevant and redundant features resulting in increased processing time and low detection rate. Therefore feature selection plays an important role in intrusion detection. There are various feature selection methods proposed in literature by different authors. In this a comparative analysis of different feature selection methods are presented on KDDCUP'99 benchmark dataset and their performance are evaluated in terms of detection rate, root mean square error and computational time.

The proposed step for work is :

Step 1: Start the weka tool.

Step 2: Browse the dataset for preprocessing.

Step 3: Select the attribute with different attribute selection.

Step 4: keep the selected attribute and remove the unselected attribute.

Step5: Classify the selected attribute with different classifier.

Step 6: Analyze the different values after the classification.

Step 7: visualize the resulted graph with different values.

Step 8: repeat the step 3 to step 7 for different classifiers.

Step 9 : Stop

RESULT

The following figures display the results of work:

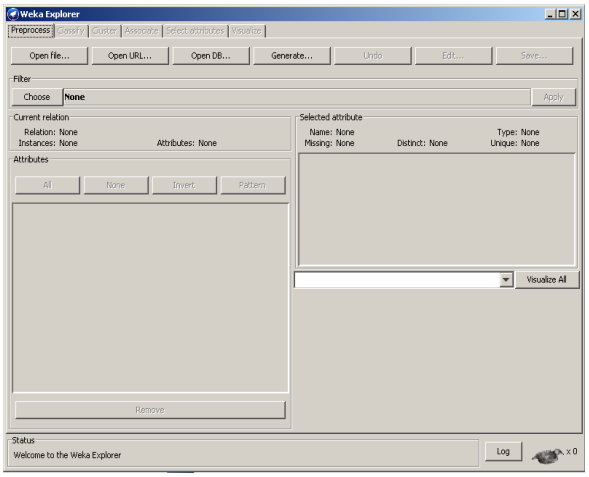

Figure 1: Browsing dataset window for weka tool

© INNOVATIVE RESEARCH THOUGHTS | Refereed | Peer Reviewed | Indexed
ISSN : 2454 – 308X | Volume : 07 , Issue : 02 | April - June 2021

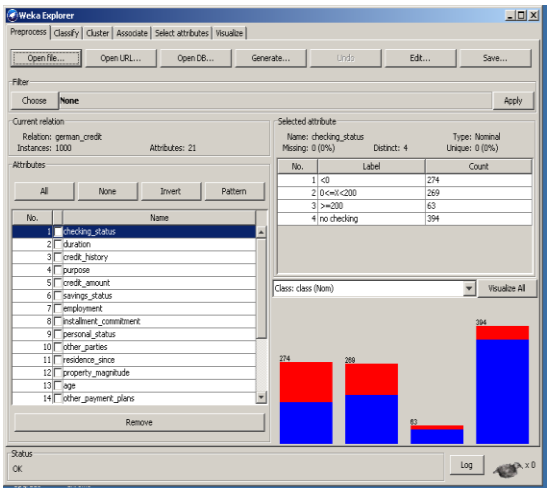

Figure 2 Displaying dataset features

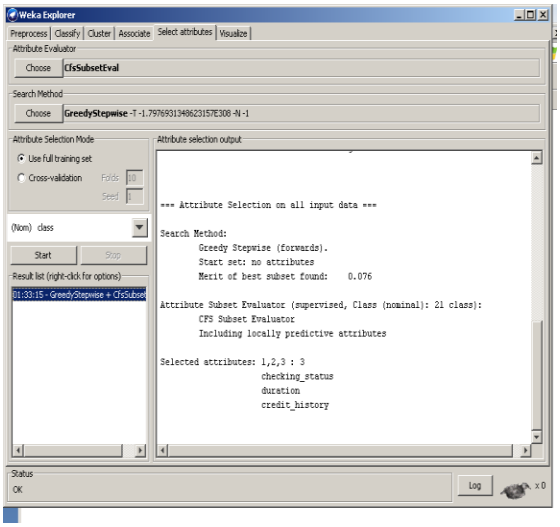

Figure 3 selecting dataset attributes

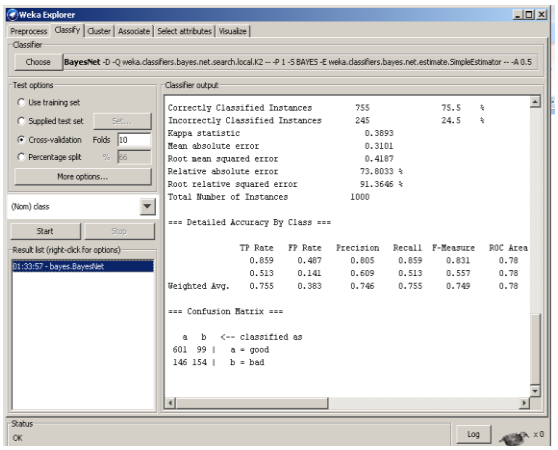

Figure 4 classifying dataset attributes with Bayesnet

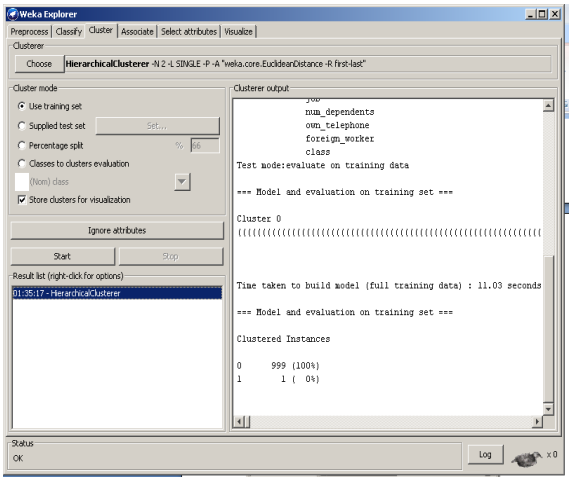

Figure 5 clustering dataset attributes with Hierarchical cluater

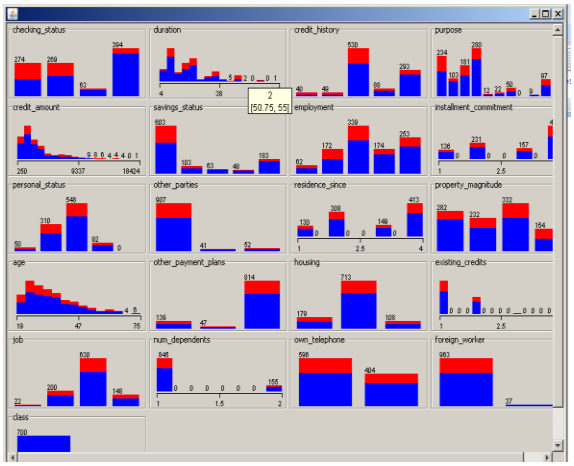

Figure 6 Visualization of the complete dataset

Table no 1. Using bayes net classifier

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
Weighted Average	0.589	0.487	0.805	0.859	0.831	0.78	1
	0.513	0.141	0.609	0.513	0.557	0.78	0
	07.55	0.383	0.746	0.755	0.749	0.78	

Table no 2.using naive bayes classifier

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
Weighted Average	0.873	0.436	0.815	0.873	0.843	0.809	1
	0.573	0.127	0.644	0.537	0.585	0.809	0
	0.772	0.362	0.763	0.772	0.766	0.809	

Table no 3.using k –star classifier

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
Weighted Average	0.813	0.583	0.765	0.813	0.788	0.689	1
	0.417	0.187	0.488	0.417	0.45	0.689	0
	0.694	0.464	0.464	0.694	0.687	0.689	

V. CONCLUSION

An intrusion is an attempt to compromise the integrity, confidentiality, availability of a

resource, or to bypass security mechanisms of a computer system or network. There is multiclass problem during the classification of data. Intrusion detection is a problem of transportation infrastructure protection owing to the fact that computer networks are at the core of the operational control of much of the nation's transport day' era detection of security threats that are commonly referred to as intrusion, has become a very important and critical issue in network, data and information security. Intrusion detection system plays a major role in providing security to computer networks. In this work KDDCups 99 data set is used to detect the intrusion .

References:

[1]. Effective approach toward Intrusion Detection System using datamining techniques by G.V. Nadiammai, M. Hemalatha in EgyptianInformatics Journal (2014) 15, 37-50

[2]. Data Mining in Education for Students Academic Performance: ASystematic Review by Er.Anurag Jindal, Er. Williamjeet Singh in ISSN 2277-3061

[3]. Mining With Noise Knowledge: Error-Aware Data Mining byXindong Wu and Xingquan Zhu in IEEE TRANSACTIONS ONSYSTEMS, MAN, AND CYBERNETICS—PART A: SYSTEMSAND HUMANS,VOL. 38, NO. 4, JULY 2008

[4]. Combined Mining: Discovering Informative Knowledge in ComplexData by Longbing Cao, Senior Member, IEEE, Huaifeng Zhang,Member, IEEE, Yanchang Zhao, Member, IEEE, Dan Luo, andChengqi Zhang, Senior Member, IEEE in IEEETRANSACTIONS ON SYSTEMS, MAN, ANDCYBERNETICS—PART B: CYBERNETICS, VOL.41, NO. 3, JUNE 2011

[5]. Intrusion Detection System Using Data Mining Technique: SupportVector Machine by Yogita B. Bhavsar, Kalyani C.Waghmare in International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 3,March 2013)

[6].]Huy Anh Nguyen and Deokjai Choi"Application of Data Mining to Network Intrusion Detection: Classifer Selection Model" ChonnamNational University, Computer Science Department. Gwangju. Korea.

[7]. Wenke Lee, Salvatore J.Stolfo, Philip K.Chan " Real Time Data Mining-Based Intrusion Detection" Computer science

Department, FloridaInstitute of Technology, Melbourne

[8]. paul Dokas, Levent Ertoz,Vipin kumar, Aleksandar Lazarevic,Jaideep Srivastava, Pang-Nig Tan " Data Mining for Network Intrusion Detection" University of Minnesota, Minneapolis. USA

[9]. Theodoros Lappas and Konstantinos Pelechrinis "Data Mining for (Network) Intrusion Detection system" Department of Computer Scienceand Engineering. Riverside.

[10]. Subaira A.S. Anitha P. "An Efficient Classification Mechanism for Network Intrusion Detection System Based on Data Mining Techniques:

[11]. A survey" in International Journal of Computer Science and Business Informatics. Coimbatore, October 2013 India.