A Comparative Analysis on Hybrid SVM for Network Intrusion Detection System

Gaddam Venugopal[1], Dr. Gatram Ramamohan babu[2]

[1]Research Scholar, Dr. Y.S. Rajasekhar Reddy University College of Engineering & Technology, Acharya Nagarjuna University, Nagarjuna Nagar, Guntur.
[2]Professor, Department of Information Technology,RVR & JC College of Engineering, Chowdavaram, Guntur.
[1]venugopal.gaddam@gmail.com, [2]rmbgatram@gmail.com

Article History: Received: 11 January 2021; Accepted: 27 February 2021; Published online: 5 April 2021

Abstract: Rapid growth in technology, not only makes smoother the life style, but also reveals a lot of security issues. Day by day changing of attack types distractsnot only organizations, companies but also the people who are using network services for their daily needs.Intrusion Detection Systems (IDS) have been developed to avoid financial losses caused by network attacks. KDD CUP 99, NSL-KDD, KYOTO 2006+, CIDDS-01 etc., some of the Intrusion Datasets available for researchers to test and develop their IDS models. In this paper, an attempt is made to compare the effect of various SVM Kernel based models and Hybrid kernel based models etc., on CIDDS-01 dataset. Results were drawn.

Keywords:IDS, Machine Learning, Support Vector Model, Kyoto 2006+ dataset, CIDDS-01, Intrusion Detection System.

1. Introduction

The growth in the technology leads to the wide spread usage of Internet. Internet is one of the major source for wide spread services in various sectors like business, medical, education, banking etc. which facilitate the customers as well as vendors. These online services are facing major security issues which become a greater threat to the network users for their valuable data; money etc. Network Intrusion detection system (NIDS) is an effective mechanism that provides security for the network users.

There are several machine learning techniques that help in developing Intrusion detection systems. Most of the cases, supervised learning techniques were applied by the various authors due to the availability of the class labels in the intrusion datasets.Classification models like Nearest Neighbor classifiers, Support Vector Machine (SVM), Convolutional Neural Networks, Decision Tree Induction etc., can be adopted on the intrusion datasets and can achieve good prediction rates for attacks.

There are numerous benchmark datasets readily available in the web in various formats that can be easily downloaded and can be applied any NIDS model. Among them KDD CUP 99, NSL KDD, Kyoto 2006+, CIDDS-01 [7][15] are a few datasets. All these datasets are collected through network traffic and contain several parameters such as Source IP Address, Destination IP Address, Source Port, Destination Port, Duration etc., through these features the patterns of a specific network activity is observed and detected whether it is either an attack or a benign.

Some of these datasets are generated by constructing testbeds and some of them are collected from real-time network traffic.

The objective of this paper is to study the behavior of the Hybrid Kernel based SVM algorithm on the dataset CIDDS-01 and compare with the results of other datasets.

i. Data Set Description

CIDDS-001(Coburg Network Intrusion Detection Dataset) is a labeled unidirectional flow based dataset generated by emulating small business environment in cloud for the evaluation of Network Intrusion Detection System (NIDS). It consists of real traffic data from an internal server with open stack environment (Web, E-Mail servers etc.) and external server (file synchronization, web server). Python scripts emulate normal user behavior on the clients.

The dataset contains 14 attribute, the first attributes 1 to 11 are default NetFlow attributes whereas the attributes 12 to 14 are additional attributes described the attacks. Table I provides the description of CIDDS-001 dataset attributes [16Ring, M., Wunderlich].

Table 1: Features and their Description of CIDDS-001 data set

Feature Name	Description
Date first seen	flow first seen at particular Start time
Duration	Duration of the flow
Proto_type	Transport Protocol (e.g. ICMP, TCP, or UDP)
Src_IP_Addr	IP Address of Source
Src_Pt	Source Port number
Dst_IP_Addr	IP Address of Destination
Dst_Pt	Destination Port number
Packets	Number of packets transmitted
Bytes	Number of bytes transmitted
Flags	OR concatenation of all TCP Flags
ToS	Type of Service
Class	Class label (Normal, Attacker, Victim, Suspicious and Unknown)
Attack Type	Type of Attack (Port Scan, DoS, Brute force, Ping Scan)
Attacked	Unique Attack id. Allows attacks which belong to the same class carry the same attack id
Attack Description	additional information about the set attack parameters Provided (e.g. the number of attempted password guesses for SSH-Brute-Force attacks)

The remainder of this paper is as follows: In Section II researchers' work in intrusion detection system is discussed. A detailed description of Hybrid Kernel based SVM (HKSVM) and its feature selection approaches is given in section III. Section IV provides methodology adopted for testing the dataset, results are discussed in section V and finally conclusion is provided in Section VI.

2. Related work

Idhammadet al., [4] suggested detection system of DDoS attacks in a cloud environment based on information theoretical entropy and random forest classifier. Time-based sliding window algorithm is used to estimate the entropy of network header characteristics of incoming traffic. When estimated entropy exceeds its normal range then incoming traffic is preprocessed and then random forest classifier is applied. The significant improvement of the accuracy of 2.5% is noticed here compared to the accuracy of Random forest tested directly on the CIDDS-001 which is 97%.

Raneel kumar, Lal and Sharma (2015) proposed an Intrusion Detection system (IDS) to detect DoS attacks emanating from one or more Virtual machines to another in cloud environment which has got multiple VM's as multi-tenanted set up. The Intrusion Detection system composed of a packet sniffer, a function extractor, and one – class Support Vector Machine classifier. The proposed Intrusion Detection System showed promising results to detect seven different types of DoS attacks.

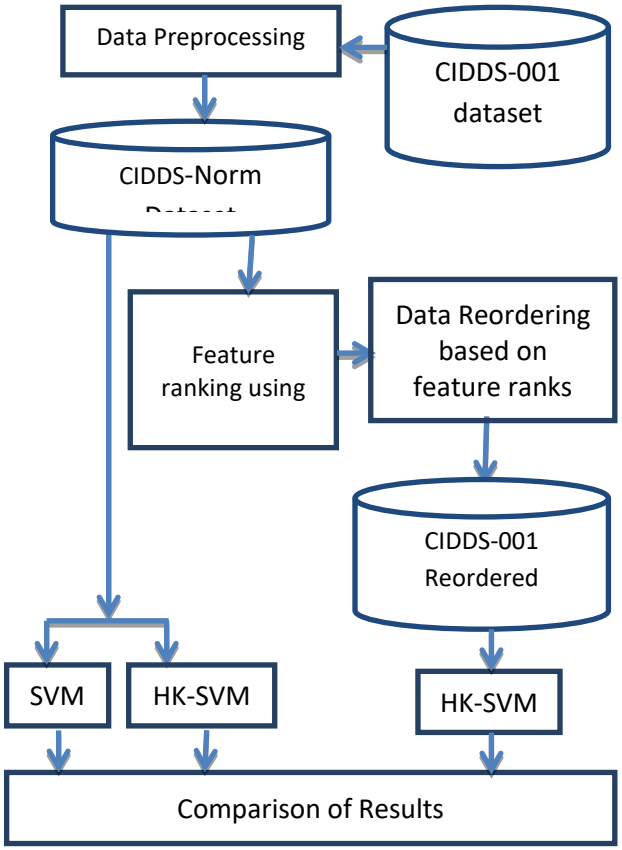
Ertam et al., 2014 proposed a method to arbitrate whether data captured on the internet was normal or malicious. Theclassifiers in the proposed work were analyzed with recall, precision, F measure metrics, falserate and accuracy rate values.Ye and Yu, 2015combined binary ELM classifiers of each class into an ensemble classifiers using one to all strategyto classify network intrusion and evaluated the accuracy of different approach. ELM is used with least human intervene. The experiment was performed on NSL-KDD dataset.

Wang et al. 2017, proposed support vector machine (SVM) based intrusion detection framework with feature augmentation. The framework supports feature augmented technique for providing immense quality, concise data for training SVM classifiers. The proposed system improved detection along with reduction in training time. Proposed model used NSL-KDD dataset for finding out performance of classifiers. The performance was found to be superior for the metric viz. false alarm rate, accuracy, detection rate.

3. Methodology

This section will discuss about the proposed methodology to implement a Hybrid Kernel based SVM (HKSVM) [1]and an Ensemble Hybrid Kernel based SVM (EHK-SVM) a feature selection approach [2]on network intrusion detection datasets. For the purpose of experimental test two benchmark datasets namely Kyoto2006+ and CIDDS-001 are used.

Figure 1 presents the methodology for the implementation of the proposed model. The dataset is preprocessed to apply transformation and normalizations. The resultant preprocessed normalized dataset is now ready for the mining. This normalized dataset is now undergone to SVM classification with different kernels like RBF-kernel, Polynomial kernel, Gaussian Kernel and Hybrid Kernel based SVM[1] and their corresponding accuracies are observed. On the other hand a feature selection approach namely Relief is applied on the CIDDS-001 dataset and 6 features are extracted having highest ranking to extract good accuracy. Table 2 presents the list of features that were selected after Relief feature selection method.



4. Results

The experimental study is implemented on JDK 1.7 on Windows 7 environment on Intel Core i5 processor. Table 3 and Table 4 contains tabulated values of accuracies for various kernel based SVM methods on Kyoto 2006+ dataset and CIDDS-001 dataset respectively.

The Figure 2 presents a bar graph for the accuracies of the models on both the datasets.

Table3: Accuracies of the SVM methods on CIDDS-001 dataset.

SVM Model	Accuracy

adopted	
Polynomial kernel	93.31 %
Gaussian Kernel	93.26%
RBF Kernel	92.13%
HKSVM	96.22%
EHK-SVM	97.93%

Table4: Accuracies of the SVM methods on Kyoto2006+ dataset.

SVM Model adopted	Accuracy
Polynomial kernel	83.87%
Gaussian Kernel	67.16%
RBF Kernel	82.99%
HKSVM	92.51%
EHK-SVM(11)	99.08%

Figure 2: Accuracies of the SVM methods on CIDDS-001 dataset

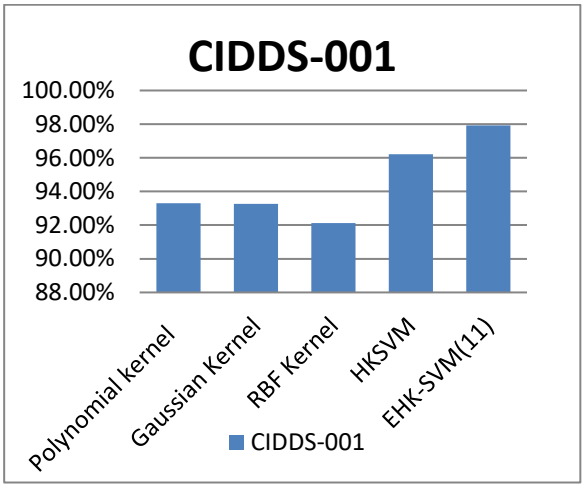
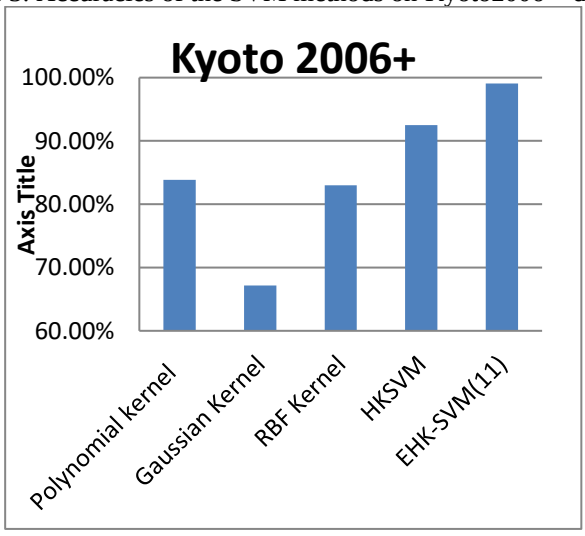


Figure 3: Accuracies of the SVM methods on Kyoto2006+ dataset.



From Table 2 and 3 and Figure 2 and 3 it is observed that HKSVM and EHKSVM are giving highest accuracies for the both the datasets.

5. Conclusion

In this paper, an attempt is made to find the accuracies of various SVM models on CIDDS-001 and Kyoto 2006+ datasets and compared the results. It is concluded that the Hybrid Kernel based SVM is gained good accuracy when compared to other kernel based approaches. It is also observed that an ensemble hybrid kernel based SVM (EHK-SVM) is also yielding good accuracy with 11 features from Kyoto 2006+ dataset and with 6 features from CIDDS-001 dataset. As a future work of this paper, the proposed model can be tested on the real time network traffic.

Reference

1. Intrusion Detection with Comparative Analysis of Supervised Learning Techniques and Fisher Score Feature Selection Algorithm. Available from: https://www.researchgate.net/publication/327700381_Intrusion_Detection_with_Comparative_Analysis_of_Supervised_Learning_Techniques_and_Fisher_Score_Feature_Selection_Algorithm.
2. GaddamVenu Gopal,Dr. Gatram Rama Mohan Babu" A Hybrid Kernel-based Support Vector Machine Algorithm for Intrusion Detection System", Jour of Adv Research in Dynamical& Control Systems, Vol. 11(8), 2019: 143-150.
3. MarwaneZekri et al., DDoS Attack Detection using Machine Learning in Cloud Computing Environment, 978-1-5386-1115-9/17@2017 IEEE.
4. Idhammad, Mohamed, Karim Afdel, and Mustapha Belouch. "Distributed intrusion detection system for cloud environments based on data mining techniques." *Procedia Computer Science* 127 (2018): 35-41.
5. SandipHingane et al, Intrusion Detection Techniques: A Review, International Journal of Scientific Research in Computer Science, Engineering and Information Technology © 2018 IJSRCSEIT | Volume 3 | Issue 1 | ISSN : 2456-3307.
6. Pradeep Pundir, Classification and Prediction techniques using Machine Learning for Anomaly Detection, International Journal of Engineering Research and Applications (IJERA) ,ISSN: 2248-9622.
7. Abhishek verma et al., Statistical analysis of CIDDS-001 dataset for Network Intrusion Detection Systems using Distance-based Machine Learning, 6th International Conference on Smart Computing and Communications, ICSCC 2017,7-8 December 2017,Kurukshetra, India.
8. Mohamed Idhammada, Karim Afdel, and Mustapha Belouch, Detection System of HTTP DDoS Attacks in a Cloud Environment Based on Information Theoretic Entropy and Random Forest, Security and Communication Networks Volume 2018, Article ID 1263123, 13 pages , https://doi.org/10.1155/2018/1263123.
9. Yili Ren, Fuxiang Hu, Hongping Miao, The optimization of kernel function and its parameters for SVM in well-logging, 2016 13th International Conference on Service Systems and Service Management (ICSSSM).
10. Raneelkumar , Sunil Pranit Lal &Alok Sharma (2015) , Detecting Denial of Service Attacks in the Cloud, 2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress.
11. ChunhuaGu and Xueqin Zhang,A Rough Set and SVM Based Intrusion Detection Classifier, Second International Workshop on Computer Science and Engineering, 2009.
12. Yong-Xiang Xia, Zhi-Cai Shi, Zhi-Hua Hu, An Incremental SVM for Intrusion Detection Based on Key Feature Selection, 2009 Third International Symposium on Intelligent Information Technology Application.
13. Hetalbhavsar, A Comparative study of training algorithms for supervised machine learning, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-4, September 2012.
14. Yogita B. Bhavsar, KalyaniC.Waghmare, Intrusion Detection System Using Data Mining Technique: Support Vector Machine, 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).
15. CIDDS-001 dataset. (2017, Aug.) [Online] Available: https://www.hs-coburg.de/forschung kooperation/forschungsprojekteoentlich/ingenieurwissenschaften/ cidds-coburg-intrusion-detection-data-sets.html.
16. M Ring, S Wunderlich, D Grüdl, D Landes and A Hotho, Creation of Flow-Based Data Sets for Intrusion Detection,Journal of Information Warfare Vol. 16, No. 4 (Fall 2017), pp. 41-54. (ip address)
17. Chih-Wei Hsu et al., A Practical Guide to Support Vector Classification, National Taiwan University, Taipei 106, Taiwan http://www.csie.ntu.edu.tw/~cjlin Initial version: 2003 Last updated: April 15, 2010.

18. T. Nathiya, G. Suseendran, An effective way of cloud intrusion detection system using decision tree, support vector machine and naïve bayes algorithm, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7 Issue-4S2, December 2018.