ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 2, February 2020

A Bi-objective Hyper-Heuristic Support Vector Machines for Big Data for Network-Security: An overview

Shital Kalokhe[1], Prof. Manoj Wakchaure[2]

P.G. Student, Department of Computer Engineering, Amrutvahini College of Engg, Sangamner, MH India

Professor, Department of Computer Engineering, Amrutvahini College of Engg, Sangamner, MH India

ABSTRACT: Big Data Cyber security Analytics is increasingly becoming an important area of research and practice aimed at protecting networks, computers, and data from unauthorized access by analyzing security event data using big data tools and technologies. Whilst a plethora of Big Data Cyber security Analytic Systems has been reported in the literature, there is a lack of a systematic and comprehensive review of the literature from an architectural perspective. In this paper, we formulate the SVM configuration process as a bi-objective optimization problem in which accuracy and model complexity are considered as two conflicting objectives. System proposes a novel hyper-heuristic framework for bi-objective optimization that is independent of the problem domain. This is the first time that a hyper-heuristic has been developed for this problem. The proposed hyper-heuristic framework consists of a high-level strategy and low-level heuristics. The high-level strategy uses the search performance to control the selection of which low-level heuristic should be used to generate a new SVM configuration. The low-level heuristics each use different rules to effectively explore the SVM configuration search space. To address bi-objective optimization, the proposed framework adaptively integrates the strengths of decomposition- and Pareto based approaches to approximate the Pareto set of SVM configurations. The effectiveness of the proposed framework has been evaluated on two cyber security problems: Microsoft malware big data classification and anomaly intrusion detection.

KEYWORDS: Hyper-heuristics, big data, cyber security, optimization.

I. INTRODUCTION

The high-level strategy operates on the heuristic space instead of the solution space. In each iteration, the high-level strategy selects a heuristic from the existing pool of low-level heuristics, applies it to the current solution to produce a new solution and then decides whether to accept the new solution. The low level heuristics constitute a set of problem-specific heuristics that operate directly on the solution space of a given problem.In proposed work we define three different layers for detection the malicious data or connection using SVM and evolutionary base heuristic approach.
The proposed system carried out multi objective heuristic approach to detect the malicious attack from network environment. Initially system deals with training face where the background knowledge has generated from various network dataset. KDD Cup 99 dataset has used to extract the basic features of network attack and stored those features in train model. In strategic approach system evaluate search network packet using support vector machine (SVM), the system works like supervised learning approach for label classification so, it needs to generate a background knowledge before evaluate the test instances. In this work system first execute data preprocessing as well as data normalization.Once the background knowledge has generated buy system it is purely applicable for testing, interesting phase we have written heuristic kernel function for evaluate to each test object. The background knowledge has used to generate the runtime similarity for each known as well as unknown type of attacks.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 2, February 2020

II. LITARATURE SURVEY

According to Soheily-Khah, Saeid, Pierre-François Marteau[1].Data mining techniques play an increasing role in the intrusion detection by analyzing network data and classifying it as 'normal' or 'intrusion'. In recent years, several data mining techniques such as supervised, semi-supervised and unsupervised learning are widely used to enhance the intrusion detection. This work proposes a hybrid intrusion detection (kM-RF) which outperforms in overall, according to our experimentation, the alternative methods through the accuracy, detection rate and false alarm rate. A benchmark intrusion detection dataset (ISCX) is used to evaluate the efficiency of the kM-RF, and a deep analysis is conducted to study the impact of the importance of each feature defined in the pre-processing step.

According to Alaei, Parisa, and FakhroddinNoorbehbahani[2].With the proliferation of the internet and increased global access to online media, cybercrime is also occurring at an increasing rate. Currently, both personal users and companies are vulnerable to cybercrime. A number of tools including firewalls and Intrusion Detection Systems (IDS) can be used as defense mechanisms. A firewall acts as a checkpoint which allows packets to pass through according to predetermined conditions. In extreme cases, it may even disconnect all network traffic. An IDS, on the other hand, automates the monitoring process in computer networks. The streaming nature of data in computer networks poses a significant challenge in building IDS. In this paper, a method is proposed to overcome this problem by performing online classification on datasets. In doing so, an incremental naive Bayesian classifier is employed. Furthermore, active learning enables solving the problem using a small set of labeled data points which are often very expensive to acquire. The proposed method includes two groups of actions i.e. offline and online. The former involves data preprocessing while the latter introduces the NADAL online method. The proposed method is compared to the incremental naive Bayesian classifier using the NSL-KDD standard dataset. There are three advantages with the proposed method: (1) overcoming the streaming data challenge; (2) reducing the high cost associated with instance labeling; and (3) improved accuracy and Kappa compared to the incremental naive Bayesian approach. Thus, the method is well-suited to IDS applications.

According to Falcón-Cardona, Jesús Guillermo et al. [3]in recent years, Indicator-based Multi-Objective Evolutionary Algorithms (IB-MOEAs) have become a relatively popular alternative for solving multi-objective optimization problems. IB-MOEAs are normally based on the use of a single performance indicator. However, the effect of the combination of multiple performance indicators for selecting solutions is a topic that has rarely been explored. A hyper-heuristic which combines the strengths and compensates for the weaknesses of four density estimators based on R2, IGD, and p. The selection of the indicator to be used at a particular moment during the search is done using online learning and a Markov chain. Additionally, a novel framework that aims to reduce the computational cost involved in the calculation of the indicator contributions. Our experimental results indicate that our proposed approach can outperform state-of-the-art MOEAs based on decomposition (MOEA/D) reference points (NSGA-III) and the R2 indicator (R2-EMOA) for problems with both few and many objectives.

According to Rahul, Vigneswaran K., et al. [4]Intrusion detection system (IDS) has become an essential layer in all the latest ICT system due to an urge towards cyber safety in the day-to-day world. Reasons including uncertainty in finding the types of attacks and increased the complexity of advanced cyber-attacks, IDS calls for the need of integration of Deep Neural Networks (DNNs). In this paper, DNNs have been utilized to predict the attacks on Network Intrusion Detection System (N-IDS). A DNN with 0.1 rate of learning is applied and is run for 1000 number of epochs and KDDCup-`99' dataset has been used for training and benchmarking the network. For comparison purposes, the training is done on the same dataset with several other classical machine learning algorithms and DNN of layers ranging from 1 to 5. The results were compared and concluded that a DNN of 3 layers has superior performance over all the other classical machine learning algorithms.

According to Gaied, Imen, Farah Jemili, et.al [5]there is no standard solution we can use to completely protect against computer network intrusion. Every solution has its advantages and drawbacks. Soft computing is considered as a promising paradigm to cope with the dynamic evolution of networks. In previous works, we presented two soft

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 2, February 2020

computing approaches of intrusion detection. The first one is based on the neuro-fuzzy and the second one is based on the genetic fuzzy one. In this work, we elaborate an empirical comparative study to highlight the benefits of each method in intrusion detection and exploit their complementarities to enhance the detection rate of all types of attacks as well as decrease the false positives rate.

According to Potteti, Sumalatha, and NamitaParati.[6]The describes how Hybrid IDS is used in Fuzzy Genetic algorithm in wireless networks or networks. These days Intrusion Detection System (IDS) which is define d as a solution of system security is employed to identify the abnormal activities in a computer system or network. The IDS is to detect the attacks and generate the proper response. The drawback of the anomaly based intrusion detection in a wireless network is the high rate of false positive. By designing a hybrid intrusion detection system can solve this by connecting a misuse detection module to the anomaly detection module. In this paper, we propose to develop a hybrid intrusion detection system for wireless local area networks based on Fuzzy Genetic logic. The proposed Fuzzy Genetic logic-based system could be able to detect the intrusive activities of the computer networks as the rule base holds a better set of rules.

According to Mukane, Rohit V. et.al [7]Rotating machine is device that supports an important category of manufacturing industry. Machine fault detection is very advantageous for diagnosis of faults before it occurs and to protect the machinery from catastrophic damages. Vibration monitoring is important for running rotating equipment evenly for years. There are diverse reasons that cause vibrations in the rotating machinery like rotor unbalance, machine looseness, permanent bow (bend shaft or warped shaft), bearing damage, misalignment, eccentricity, oil whirl, cracked tooth, etc. There are several methods used to detect machinery faults such as, model based technique and signal based techniques which includes Hilbert-Huang Transform (HHT), Wavelet Transform (WT), Artificial Neural Networks (ANN), Support Vector Machine (SVM), Shock Pulse Monitoring (SPM), Fuzzy Logic. The proposed work shows a novel approach of classification using fuzzy logic for decision making about the machine condition. The experimental results show that, algorithm implemented identifies severity of faults to appropriate condition.

According to Behera, SantiKumari, et al. [8]a suggests a computer vision based system which have ability to identify deformity in the orange fruits and also organize the flaw type appeared on the surface of orange fruit. The symptoms of flaw mark imply the seriousness of the disease and recommend the optimal approach to deal with the disease. It's conjointly required to diagnose the disease properly with prior to great damage by providing proper treatment. Further, estimation of severity of disease is required for applying proper amount of pesticides to avoid the environmental pollution and economic burden. Here we use multi class SVM with K-means clustering for classification of diseases with 90% of accuracy and Fuzzy logic to compute the degree of disease severity.

According to Theresa, W. Gracy, and S. Sakthivelet.al [9]MANETs are potential wireless network, where mobile nodes are connected dynamically in ad hoc basis. This unique characteristic attracted many promising applications and the one among is battlefield communication. The war troops at the edge of the network do not have any computing infrastructure for communication. Therefore MANET plays a vital role in battlefield communication. Since MANET is openness to eavesdropping, routing of information causes vulnerabilities and degrades the performance of network. This necessitated developing a MANET with a novel intrusion detection system to provide reliability and security to the battlefield communication. A Di-Fuzzy logic technique which provides two phase detection for intrusion detection in the network. It works in a cluster based routing environment to co-ordinate and control the entire network. The selection of cluster head is based on the node with the maximum energy to improve the life time of the network. The proposed technique is simulated in network simulator NS 2.8 and the performance is evaluated by comparing with the existing Fuzzy based IDS (F-IDS) & intrusion detection and adaptive response (IDAR) system. From comparison the proposed Di-Fuzzy logic technique improves its performance in all metrics and thus provides a safe environment for battlefield communication.

According to Alqahtani, Saeed M., and Robert John [10]Intrusion detection system (IDS) as one of huge research problem in network security is the most effective tool of protection. It is a method of parsing network traffic data to

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 2, February 2020

detect security abuses. Data mining can play a very significant role in evolving an IDS. The dataset of IDSs or soft computing techniques based IDS can be classified into normal and abnormal traffic in order for generated alerts to detect threats. In this paper, we utilized the most common classification algorithms: Decision Tree (J48), Naive Bayes, OneR, and K-Nearest Neighbor (K-NN). These algorithms were chosen after investigating the most effective classification algorithms that are widely used. The aim of this study is to present a comparative study for the performance of each system that was gained from our previous experiments: Snort IDS, Suricata IDS, FL-Snort IDS, and FL-Suricata IDS in order to test which classifier algorithm is the best for our systems results, and investigate which system presents significant results. The performance of these classification algorithms was evaluated using 10-fold cross validation. Experiments and assessments of these methods were performed in the WEKA environment using the ISCX dataset.

III. DATASET DETAILS

The inherent flaws in the KDD cup 99 dataset [9] has been publicized by several statistical analyses has affected the detection accuracy of many IDS modeled by researchers. NSL-KDD data set [3] is a refined version of its precursor. It contains vital records of the complete KDD data set. There are a group of downloadable files at the disposal for the researchers. They are listed in the table 1.

Table 1 : List of NSL-KDD Dataset Files and Their Description

S. No.	File Name	Description
1	KDDTrain+.ARFF	The complete NSL-KDD train set with binary labels in ARFF format
2	KDDTrain+.TXT	The complete NSL-KDD train set including attack-type labels and difficulty level in CSV format
3	KDDTrain+_20Perce nt.ARFF	A 20% subset of the KDDTrain+.arff file
4	KDDTrain+_20Perce nt.TXT	A 20% subset of the KDDTrain+.txt file
5	KDDTest+.ARFF	The complete NSL-KDD test set with binary labels in ARFF format
6	KDDTest+.TXT	The complete NSL-KDD test set including attack-type labels and difficulty level in CSV format
7	KDDTest-21.ARFF	A subset of the KDDTest+.arff file which does not include records with difficulty level of 21 out of 21
8	KDDTest-21.TXT	A subset of the KDDTest+.txt file which does not include records with difficulty level of 21 out of 21

In each record there are 41 attributes clarifying different features of the flow and a label allocated to each either as an attack type or as normal. The essentials of the characteristics namely the attribute name, their description and sample data are listed in the tables. Table 2 encompasses type information of all the 41 attributes present in the NSL-KDD data set. The 42nd attribute holds data about the several 5 classes of network connection vectors and they are categorized as one normal class and four attack class. The 4 attack classes are further grouped as DoS, Probe, R2L and U2R. The explanation of the attack classes.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 2, February 2020

Table 2 : Mapping Attack Class with Attack Type

Attack Class	Attack Type
DoS	Back, Land, Neptune, Pod, Smurf,Teardrop,Apache2, Udpstorm, Processtable, Worm (10)
Probe	Satan, Ipsweep, Nmap, Portsweep, Mscan, Saint (6)
R2L	Guess_Password, Ftp_write, Imap, Phf, Multihop, Warezmaster, Warezclient, Spy, Xlock, Xsnoop, Snmpguess, Snmpgetattack, Httptunnel, Sendmail, Named (16)
U2R	Buffer_overflow, Loadmodule, Rootkit, Perl, Sqlattack, Xterm, Ps (7)

IV. RESULTS AND DISCUSSION

The existing survey basically focus on soft computing and classification based detection approach, basically both methods having the good detection rate but at times it generates more false positive ratio. Some systems are also not applicable in real time environment and some can't be focus on misclassified anomalies. As observed, most applications still miss the mark as there is no system that at present gives a 100% discovery rate and the sky is the limit. The below figure 2 show existing systems of the overall attack found ratio in ensemble approach with the help of table 4. In result, the X-axis presents number of packets (network flow size) provided for finding results and Y-axis presents how many packets correctly classified in particular attack types.

Table 1 : Overall attack found ratio with ensemble approach

Network flow Size (Connections)	Attacks Found			
	DOS	Probe	U2R	R2L
2500	1520	265	8	523
5000	2987	498	17	1029
7500	4562	765	22	1602
10000	5960	936	35	2188

V. CONCLUSION

Since the study of intrusion detection began to gain drive in the security community coarsely ten years before, a number of diverse ideas have emerged for challenging this problem. Intrusion detection systems vary in the sources they use to get data and in the exact techniques they rendezvous to examine this data. Most systems nowadays categorize data either by abuse detection or anomaly detection. Each method has its relative advantages and is accompanied by a set of limitations. It is likely not realistic to expect that an intrusion detection system be capable of correctly categorizing each event that happens on a given system. Perfect detection, like perfect security, is simply not an attainable goal given the complexity and rapid evolution of modern systems. After the completion of this survey we can conclude there are different techniques that can used for detection, some soft computing as well as some classification approaches are effective for detect the different attacks. Some system has work on signature base anomaly detection with creation of different rules. KDD cup dataset has used for training and testing purposed. Finally every system shows the maximum accuracy for attack detection, but none of these are has focused on unknown attack detection or misuse detection.

.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 2, February 2020

REFERENCES

[1]Soheily-Khah, Saeid, Pierre-François Marteau, and Nicolas Béchet. "Intrusion Detection in Network Systems Through Hybrid Supervised and Unsupervised Machine Learning Process: A Case Study on the ISCX Dataset." Data Intelligence and Security (ICDIS), 2018 1st International Conference on.IEEE, 2018.

[2] Alaei, Parisa, and FakhroddinNoorbehbahani. "Incremental anomaly-based intrusion detection system using limited labeled data."Web Research (ICWR), 2017 3th International Conference on. IEEE, 2017.

[3] Falcón-Cardona, Jesús Guillermo, and Carlos A. CoelloCoello. "A multi-objective evolutionary hyper-heuristic based on multiple indicator-based density estimators." Proceedings of the Genetic and Evolutionary Computation Conference.ACM, 2018.

[4] Rahul, Vigneswaran K., et al. "Evaluating Shallow and Deep Neural Networks for Network Intrusion Detection Systems in Cyber Security." 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT).IEEE, 2018.

[5] Gaied, Imen, Farah Jemili, and OuajdiKorbaa. "Neuro-fuzzy and genetic-fuzzy based approaches in intrusion detection: Comparative study." Software, Telecommunications and Computer Networks (SoftCOM), 2017 25th International Conference on.IEEE, 2017.

[6] Potteti, Sumalatha, and NamitaParati. "Intrusion detection system using hybrid Fuzzy Genetic algorithm."Trends in Electronics and Informatics (ICEI), 2017 International Conference on.IEEE, 2017.

[7] Mukane, Rohit V., et al. "LabVIEW Based Implementation of Fuzzy Logic for Vibration Analysis to Identify Machinery Faults." 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA).IEEE, 2017.

[8] Behera, SantiKumari, et al. "Disease Classification and Grading of Orange Using Machine Learning and Fuzzy Logic." 2018 International Conference on Communication and Signal Processing (ICCSP).IEEE, 2018.

[9] Theresa, W. Gracy, and S. Sakthivel. "Fuzzy based intrusion detection for cluster based battlefield MANET." Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), 2017 IEEE International Conference on.IEEE, 2017.

[10] Alqahtani, Saeed M., and Robert John. "A comparative analysis of different classification techniques for cloud intrusion detection systems' alerts and fuzzy classifiers."Computing Conference, 2017.IEEE, 2017.