A BI-OBJECTIVE HYPER-HEURISTIC SUPPORT VECTOR MACHINES FOR BIG DATA CYBER-SECURITY

T.Sasi Vardhan[1], B. Vyshnavi[2], G.Lahari[3], V.S.S.N.Akhila4, Y.Shravya[5]

[1]Assistant Professor, Department of CSE(CS), MallaReddy Engineering College for Women, Hyderabad, TS, India.
Email: sasivardhan.t@gmail.com

[2,3,4,5]UG Students, Department of CSE(CS), MallaReddy Engineering College for Women, Hyderabad, TS, India.

ABSTRACT

Cyber security in the context of big data is known to be a critical problem and presents a great challenge to the research community. Machine learning algorithms have been suggested as candidates for handling big data security problems. Among these algorithms, support vector machines (SVMs) have achieved remarkable success on various classification problems. However, to establish an effective SVM, the user needs to dene the proper SVM configuration in advance, which is a challenging task that requires expert knowledge and a large amount of manual effort for trial and error. 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. We propose 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. The obtained results

demonstrate that the proposed framework is very effective, if not superior, compared with its counterparts and other algorithms.

Keyword : Big Data, Secure, SVM, Malware.

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 problemspecific 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.

Literature survey

SVMs are a class of supervised learning models that have been widely used for classification and regression SVMs are based on statistical learning theory and are better able to avoid local optima than other classification algorithms. An SVM is a kernel-based learning algorithm that seeks the optimal hyper plane. The kernel learning process maps the input patterns into a higher-dimensional feature space in which linear separation is feasible. The existing

kernel functions can be classified as either local or global kernel functions. Local kernel functions have a good learning ability but do not have good generalization ability. By contrast, global kernel functions have good generalization ability but a poor learning ability. For example, the radial kernel function is known to be a local function, whereas the polynomial kernel function is a global kernel function. The main challenge lies in determining which kernel function should be used for the current problem instance or the current decision point. This is because the kernel selection process strongly depends on the distribution of the input vectors and the relationship between the input vector and the output vector (predicted variables). However, the feature space distribution is not known in advance and may change during the course of the solution process, especially in big data cyber security. Consequently, different kernel functions may work well for different instances or in different stages of the solution process and kernel selection may thus have a crucial impact on SVM performance. To address this issue, in this work, we use multiple kernel functions to improve the accuracy of our algorithm and avoid the shortcomings of using a single kernel function.

OBJECTIVES OF SYSTEM

The goal of proposed Bi-objective Hyper-Heuristic system is to maximize the detection accuracy, to minimize false positive rate and detector generation time. The Objective of the proposed application is as follows: To design and implement a Bi-objective Hyper-Heuristic system using SVM and FGA in big data environment. To improve the performance of overall network To detect all types of attacks in online as well as offline environment like NIDS and HIDS. (e,g, DOS, PROBE, U2R, R2L, Unknown) Define the security and privacy in wireless network virtualization over the network

EXISTINGSYSTEM:

SVMs are a class of supervised learning models that have been widely used for classification and regression SVMs are based on statistical learning theory and are better able to avoid local optima than other classification algorithms. An SVM is a kernel-based learning algorithm that seeks the optimal hyper plane. The kernel learning process maps the input patterns into a higher-dimensional

feature space in which linear separation is feasible. The existing kernel functions can be classified as either local or global kernel functions. Local kernel functions have a good learning ability but do not have good generalization ability. By contrast, global kernel functions have good generalization ability but a poor learning ability. For example, the radial kernel function is known to be a local function, whereas the polynomial kernel function is a global kernel function. The main challenge lies in determining which kernel function should be used for the current problem instance or the current decision point. This is because the kernel selection process strongly depends on the distribution of the input vectors and the relationship between the input vector and the output vector (predicted variables). However, the feature space distribution is not known in advance and may change during the course of the solution process, especially in big data cyber security. Consequently, different kernel functions may work well for different instances or in different stages of the solution process and kernel selection may thus have a crucial impact on SVM performance. To address this issue, in this work, we use multiple kernel functions to improve the accuracy of our algorithm and avoid the shortcomings of using a single kernel function.

PROPOSEDSYSTEM

The proposed hyper-heuristic framework for configuration selection is shown in Figure 2. It has two levels: the high- level strategy and the low-level heuristics. 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. To address the bi-objective optimization problem, we propose a population-based hyper-heuristic framework that operates on a population of solutions and uses an archive to save the non-dominated solutions. The proposed framework combines the strengths of decomposition- and Pareto (dominance) - based approaches to effectively approximate the Pareto set of SVM configurations. Our idea is to combine

the diversity ability of the decomposition approach with the convergence power of the dominance approach. The decomposition approach operates on the population of solutions, whereas the dominance approach uses the archive. The hyper heuristic framework generates a new population of solutions using the old population, the archive, or both the old population and the archive. This allows the search to achieve a proper balance between convergence and diversity. It should be noted that seeking good convergence involves minimizing the distances between the solutions and PF, whereas seeking high diversity involves maximizing the distribution of the solutions along PF. The main components of the proposed hyper-heuristic framework are discussed in the following subsections.

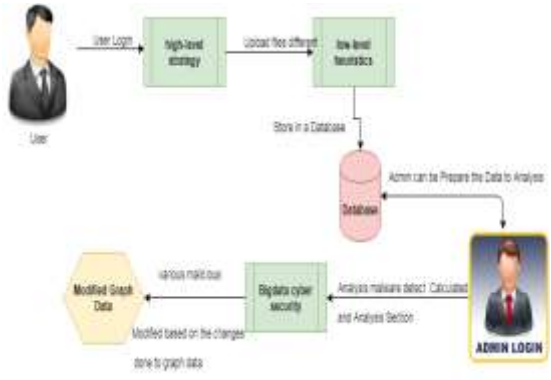
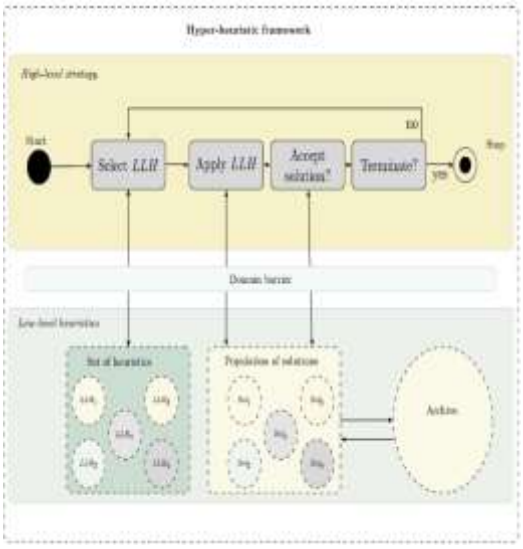


Fig : Architecture

METHODLOGY

In machine learning, support vector machines (SVMs, also support vector networks) are learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier(although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.



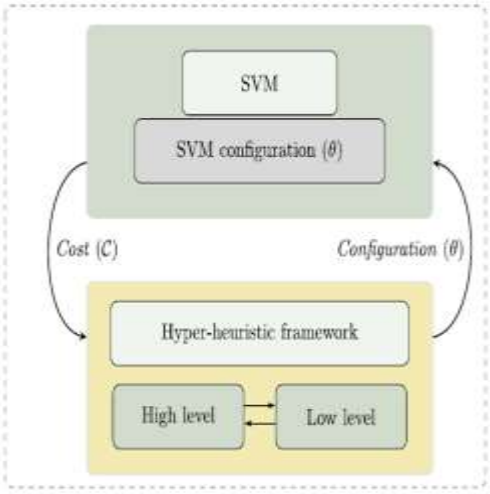


Fig 2 proposed methodology

The project is carried out based on the following modules listed:

Approved Users

In this system users are not allowed to access resources simply. User need verify their information's with admin. Admin are the authorized and trustworthy to the network. User need to send the request to administrator that they are interested to add the community. Admin views the user request and respond with the pass code to access users account through trusted sources like SSL (Gmail).

1. Security Steps and Upload

This is where the proposed algorithm is going to be effective. The admin can be upload the files with proposed classification algorithm and cryptography in order to classify and upload the encrypted details to network with its tag in the mark of understand to user about the resource.

Resource Access

The permissions to access the resource can be sent by users to admin. The requests have been updated by admin with the response to access the resource. Users can decrypt the resource and access the details. The important part is access the resource with the decryption. The passkey to access the details are limited. If the limit of wrong attempts over the threshold value means pass key expires.

Graphical Representation

This is graphical notation of the data given by the system. This phase of implementation will shows the effectiveness of the proposed system through pictorially in the order to better understand of proposed system.

SUPPORT VECTOR MACHINE:

In machine learning, **support vector machines** (**SVMs**, also **support vector networks**) are learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training

examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier(although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

Results:

Fig : User Home

Fig : User Register

Fig : Dataset

Fig : Big Data

Fig : Malware Anylisis

Fig : Data Add

Fig : Malware Analysis

Fig Bar Graph

Fig: Grapgh to Anylisis

Conclusion :

In this work, we proposed a hyper-heuristic SVM optimization framework for big data cyber security problems. We formulated the SVM configuration process as a bi-objective optimization problem in which accuracy and model complexity are treated as two conflicting objectives. This bi-objective optimization problem can be solved using the proposed hyper-heuristic framework. The framework integrates the strengths of decomposition- and Pareto-based approaches to approximate the Pareto set of configurations.

Reference:

[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