IJCSE International Journal of Computer Sciences and Engineering Open Access

Review Paper Vol.-6, Issue-12, Dec 2018 E-ISSN: 2347-2693

A Review on effect of SVM in Intrusion Detection System

C. Amali Pushpam[1*], J. Gnana Jayanthi[2]

[1]Research Scholar, Rajah Serfoji College, Tamil Nadu, India
[2]Dept. of Computer Science, Rajah Serfoji College, Tamil Nadu, India

Corresponding Author: joemarycap@gmail.com Tel.: +91 9600627074

Available online at: www.ijcseonline.org

Accepted: 10/Dec/2018, Published: 31/Dec/2018

***Abstract*-**Intrusion detection system is a system combining both software and hardware that monitors and analysis huge volume of network traffic and detects malicious activities. The role of IDS in system security is significant but not sufficient. Data analysis is a part of the IDS process. Data mining is a data analytic tool. If it is integrated with IDS, performance of IDS will be elevated. One of the data mining classification algorithms is SVM. It is widely applied in IDS. In this paper a methodical study on SVM in IDS was done. This paper reports the effect of SVM in IDS. It is observed that SVM increases the performance of IDS and also it has some limitations. This review provides new ways for further research to overcome these limitations.

***Keywords*-** Data mining, Intrusion Detection System, SVM

I. INTRODUCTION

If we take any field, security is an important and unavoidable concept and it should be searched extremely well. Science always has two sides i.e. merits and demerits. It should be handled properly. Whenever the society is enjoying the merits of it, it should be ready to face the challenges of it also. Internet is the one of the important benefits enjoyed by the present world. Intrusion and its detection are the very big challenges of it. Though much advancement in this field, still a lot of researches are being carried out as it is a major issue.
Communication without network is unbelievable. In internet, growth and usage of data is unpredictable. Every Yoctosecond, enormous amount of data are being transmitted through channels and hosts. These valuable resources are continuously subjected to attacks by hackers. The very big responsibility of research community is to protect these resources. Though intrusion detection system is one of the solutions to this problem, it is not fulfilling the requirements at satisfactory level. Analysing huge volume of data is a major issue in IDS. In order to simply this task, data mining is integrated with IDS. So many data mining techniques are applied in IDS. Among them SVM, ANN, DT and K-Means data mining techniques are frequently used in this field. Because of its characteristics, SVM is a focal point to researchers.

This paper is summarized with six sections explaining Support Vector Machine and kernel parameters in Section-II, Section-III discuss about applications of SVM, Section- IV concise the related work of different researchers, Section- V discuss about limitations of SVM and finally Section-VI is about conclusion.

II. SUPPORT VECTOR MACHINE

Support Vector Machine is a popular machine learning algorithm coming under classification category of data mining techniques. Support Vector Machine classifies data in such a way that data are plotted in n-dimensional space where n indicates the number of features / attributes. In [1], authors illustrate how data are plotted in different SVM classifiers. It is shown in figure 2. Two input variables (columns) would form a two-dimensional space. By finding proper hyper plane, it does the classification. Optimization techniques solve the problem of finding the optimal hyper plane. In geometry, a hyperplane is a subspace. Its dimension is one less than that of its ambient space. In SVM a hyperplane categories input variables by splitting them. In two-dimensions, hyperplane is a line and it can be visualized. When dimensions increases (linear to non-linear), it is difficult to define and visualize the hyperplane. In that situations, kernels are supporting to define hyperplane. Kernels are mathematical functions used by SVM for pattern analysis.

According to Jiapu Zhang [2], nearly 21 kernel functions are used in SVM. Linear, polynomial, radial basis function (RBF), and sigmoid are well-known kernel functions. They are given below

Kernel Functions

$$K(\mathbf{X}_i, \mathbf{X}_j) = \left\{ \begin{array}{ll} \mathbf{X}_i \cdot \mathbf{X}_j & \text{Linear} \\ (\gamma \mathbf{X}_i \cdot \mathbf{X}_j + C)^d & \text{Polynomial} \\ \exp(-\gamma |\mathbf{X}_i - \mathbf{X}_j|^2) & \text{RBF} \\ \tanh(\gamma \mathbf{X}_i \cdot \mathbf{X}_j + C) & \text{Sigmoid} \end{array} \right\}$$

where

$$K(\mathbf{X}_i, \mathbf{X}_j) = \phi(\mathbf{X}_i) \bullet \phi(\mathbf{X}_j)$$

That is, the kernel function represents a dot product of input data points mapped into the higher dimensional feature space by transformation ϕ

In pattern analysis, we find and study general types of relations in datasets like classifications, clusters, rankings, correlations, principal components..etc. Linear, nonlinear, polynomial, radial basis function (RBF), and sigmoid are kernel functions. RBF is the most used type of kernel function.

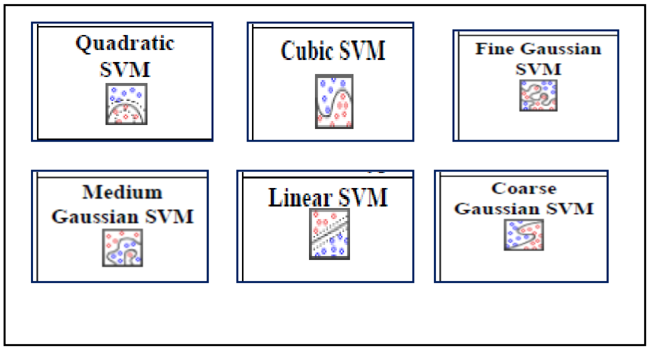

Figure 1: Plotting of data in different SVM classifiers

A. *Kernel parameters*

Cost and gamma are kernel parameters. Optimal selection of kernel function and its parameters has a very great impact on classification accuracy of SVM. For higher value of cost and gamma, model will classify more data points correctly and model will capture the shape of the dataset well. By tuning these kernel parameters, SVM will take different kernel functions “Linear”, “Poly”, “rbf”. Kernel functions are applied to map low dimensional data to high dimensional space. Also it is called similarity function as it knows how to compare two objects (integer, vectors, trees, real value). With the support of kernel, hyper planes are defined.

In SVM a hyperplane categories input variables by splitting them. In two-dimensions hyperplane is a line and split the data into two classes. As it classifies the data into two classes, it is called binary classifier. Through proper kernel selection, it can act as multi classifier. Support vector machine gets training through labelled data and develop models based on that and predict the value of target variable. Hence it is called supervised learning technique. The SVM, originally a type of pattern classifier based on a statistical learning technique for classification and regression with a variety of kernel functions, has been successfully applied to a number of pattern recognition applications [3].

In SVM, linear hyper plane is simple and easy to have it. In 2- dimensional space, linear hyper plane classifies data into two classes. By applying different types of kernel functions like linear, polynomial, rbf and sigma, SVM takes low dimensional input space and transform it to a higher dimensional space. This technique is called kernel trick.

III. APPLICATIONS OF SVM

Aim of SVM is to classify unlabelled data correctly. It has a number of applications used in various fields.

- ✓ Face detection: In this SVM classifies images into 2 classes i.e part of face or non- face.
- ✓ Text and hypertext categorization: SVM categorizes documents on the basis of score generated and then compares with the threshold value.
- ✓ Classification of images: SVM classifies images very effectively and produces better accuracy than traditional techniques
- ✓ Bioinformatics: SVM performs different types of classifications like protein classification, cancer classification..etc. It classifies genes, patients based on biological problems.
- ✓ Handwriting recognition: By applying SVM in handwriting recognition, widely used hand written characters are recognized.
- ✓ Intrusion detection system: In intrusion detection, data are classified into two categories i.e. normal and malicious. So Intrusion detection system is formulated as a binary classification problem. An effective binary classification technique called support Vector Machine is applied to detect intrusion.

IV. LITERATURE REVIEW

In IDS, SVM has become one of the popular techniques. Many authors have applied SVM in their work. Hybrid method produces better result than single method. In hybrid, SVM is mostly combined with other techniques to increase efficiency of IDS.

In 2015, Minakshi Bisen, Amit Dubey, have applied SVM , Genetic Algorithm (GA) and Hierarchical Clustering for Network Intrusion Detection System (NIDS). By applying GA on KDD Cup 1999 dataset along with Hierarchical Clustering, fewer, efficient instances of dataset are provided to SVM. Hence SVM classify the data more accurately [4].

In 2017, Zhenlong Li, Qingzhou Zhang and Xiaohua Zhao have analyzed the performance of K-nearest neighbor, support vector machine, and Artificial Neural Network (ANN) classifiers for driver drowsiness detection with different road geometries and concluded that the support vector machine achieved the fastest classification time and the highest accuracy (80.84%) [5].

In 2016, Sandeep Ranode has combined Support Vector Machine(SVM) and Clustering based on Self-Organized Ant Colony Network(CSOACN) to increase both the classification rate and runtime effectiveness [6].

In 2016, Liliya Demidova ,Evgeny Nikulchev and Yulia Sokolova have introduced SVM Classifiers with the Modified Particle Swarm Optimization and the SVM Ensembles to classify Big Data and they have proved that the SVM classifiers on the base of the modified PSO algorithm classify data with the high classification accuracy[7].

In 2012, Lei Shi1, Qiguo Duan2, Xinming Ma1, and Mei Weng1 have applied SVM for agriculture data classification and proved that it outperforms two popular algorithms, i.e., naive bayes and Artificial Neural Network (ANN) in terms of the $F1$ measure. They have concluded from the experimental results that the SVM is an effective method for classification of agricultural data [8].

In 2014, Vitthal Manekar and Kalyani Waghmare have proposed a new method for intrusion detection system using Support Vector Machine(SVM) with Particle Swarm Optimization (PSO). In this method, first PSO performed parameter optimization using SVM and then performed feature optimization to get optimized feature. Then these optimized parameters and features are given to SVM to get higher accuracy.
NSL-KDD dataset was used for experiment[9].

V. LIMITATIONS

Basically SVM is binary classifier. It can perform binary classification only. But Intrusion detection requires big data analytics. Multi class classification is required to handle big data. In big data, number of dimension is higher and it affects the performance of SVM classifier.

SVM is supervised machine learning method. It is learning from labelled data. Based on this gained information, it performs classification. This pre existing knowledge may not be available at all the time. As well as it is time consuming process.

SVM treats all the features of data equally. Practically some features are less important and redundant. Hence it requires feature selection process.

SVM also doesn't perform very well, when the data set is larger and has more noise.

VI. CONCLUSION

In this Hi-Tech world, a number of technologies are available to support security. But still there is a demand for updating existing methods and inventing new methods to challenge the challengers. Intrusion detection system integrated with data mining performs well in this field. As Data mining is a very popular data analytic tool, it analysis a huge volume of data which are handled by IDS. A number of data mining techniques are available in market and they are categorized into three main classes i.e classification, clustering and association rule. SVM is becoming a very popular and frequently used classification technique in IDS to enhance its performance. This paper gives overall idea of SVM and its limitations. This field needs further research to overcome the limitations and reach maximum level of classification accuracy.

ACKNOWLEDGMENT

I sincerely thank my Guide Professor Dr. J.Gnana Jayanthi for her guidance and support given.

REFERENCES

[1]. Jamal Hussain, Aishwarya Mishra "An Effective Intrusion Detection Framework Based On Support Vector Machine Using Nsl - KDD Dataset", in Indian Journal of Computer Science and Engineering (IJCSE), e-ISSN : 0976-5166, Vol. 8 No. 6 PP: 703 -713, Dec 2017-Jan 2018

[2]. Jiapu Zhang, "A Complete List of Kernels Used in Support Vector Machines" in Biochemistry & Pharmacology: Open Access, DOI: 10.4172/2167-0501.1000195, PP: 4-5. 2015

[3]. Jayshree Jha, Leena Ragha, " Intrusion Detection System using Support Vector Machine", in *International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868,PP:25-30,2013.*

[4]. Minakshi Bisen1, Amit Dubey2, "An Intrusion Detection System Based On Support Vector Machine Using Hierarchical Clustering And Genetic Algorithm" in International Journal Of Engineering And Computer Science ISSN:2319-7242, Volume 4 Issue 1, PP: 10062-10064, January 2015.

[5]. Zhenlong Li, Qingzhou Zhang and Xiaohua Zhao, "Performance analysis of K-nearest neighbor, support vector machine, and artificial neural network classifiers for driver drowsiness detection with different road geometries" in International Journal of Distributed Sensor Networks 2017, Vol. 13(9) PP:1-9, 2017

[6]. Sandeep Ranode, "Intrusion Detection System Using SVM Classification" in International Journal of Innovative Research in Computer and Communication Engineering, ISSN(Online): 2320-9801, Vol. 4, Issue 6, PP: 12180- 12184, June 2016.

[7]. Liliya Demidova ,Evgeny Nikulchev ,Yulia Sokolova , " Big Data Classification Using the SVM Classifiers with the Modified Particle Swarm Optimization and the SVM Ensembles", in International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 7, No. 5, PP: 294-31, 2016

[8]. Lei Shi1, Qiguo Duan2, Xinming Ma1, and Mei Weng1, "The Research of Support Vector Machine in Agricultural Data Classification" in IFIP International Federation for Information Processing 2012, CCTA 2011, Part III, IFIP AICT 370, PP: 265–269, 2012.

[9]. Vitthal Manekar1, Kalyani Waghmare2, "Intrusion Detection System using Support Vector Machine and Particle Swarm Optimization (PSO)" in International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Volume-4 Number-3 Issue-PP: 808-812, 16 September-2014.

Authors Profile

C.Amali Pushpam, MCA., M.Phil., is currently pursuing Ph.D. and servicing as an Assistant Professor in the Department of Information Technology, Bon Secours College for Women, Thanjavur, affiliated to Bharathidasan University, Tiruchirappalli, India since 2006. During her service, she has organized many international and nationsl conferences, seminars and symposium. Her main research work focuses on Data Mining, Big Data Analytics and Network Security.

J. Gnana Jayanthi, M.C.A., M.Phil., Ph.D., is presently servicing as an Assistant Professor in the Department of Computer Science, Rajah Serfoji Government College, Thanjavur, India. She has published more than 30 research papers in International and National conferences and Technical Journals and are cited in popular refereed publishers, IEEE, ACM and Springer. She is a life member of Computer Society of India (CSI), Member of the World Scientific and Engineering Academy and Society (WSEAS), International Association of Computer Science and Information Technology (IACSIT) and member of International Association of Engineers (IAENG). Her research interests include Distributed DBMS, Big Data Analytics and IoT.