Network Anomaly Detection via Clustering and Custom Kernel in MSVM

Arvind Mewada
Dept. Of Computer Science
Maulana Azad National Institute of Technology, MANIT
Bhopal, India
mewadatec@yahoo.co.in

Shamaila Khan
Dept. Of Computer Science
RITS, Bhopal, India
shamailabist@gmail.com

Prafful Gedam
Dept. Of Computer Science
Technocrat Institute of Technology, TIT, Bhopal, India
Prafful_it@yahoo.com

***Abstract*—Multiclass Support Vector Machines (MSVM) have been applied to build classifiers, which can help Network Intrusion detection. Beside their high generalization accuracy, the learning time of MSVM classifiers is still a concern when applied into Network intrusion detection systems. This paper speeds up the learning time of MSVM classifiers by reducing the number of support vectors. In this study, we proposed KMSVM method combines the K-means clustering technique with custom kernel in MSVM. Experiments performed on KDD99 dataset using KMSVM method, and the results show that the KMSVM method can speed up the learning time of classifiers by both reducing support vectors and improve the detection rate on testing dataset.**

Keywords-IDS; K-mean; MSVM; RBF; KDD99, Custom Kernel.

I. INTRODUCTION

The intrusion detection system is designed in such a way that any kind of malicious activities in computer network and its resources can be identified and vigilance [1]. *Intrusion Detection Systems* (IDS) are computer programs that tries to perform intrusion detection by comparing observable behavior against suspicious patterns, preferably in real-time. Intrusion is primarily network based activity [2]. The primary aim of Intrusion Detection Systems (IDS) is Monitoring and analyzing both user and system activities, Analyzing system configurations and vulnerabilities, Assessing system and file integrity, Ability to recognize patterns typical of attacks, Analysis of abnormal activity patterns and Tracking user policy violations to protect the availability, confidentiality and integrity of critical networked information systems. IDS can be classified based on which events they monitor, how they collect information and how they deduce from the information that an intrusion has occurred. IDSs that operates on a single workstation are known as host intrusion detection system (HIDS), A HBIDS adds a targeted layer to security to particularly vulnerable or essential systems, it monitors audit trails and system logs for suspicious behaviors [3] while A network-based IDS monitors network traffic for particular network segments or devices and analyzes network, transport, and application protocols to identify suspicious activity. Misuse detection uses the "signatures" of known attacks to identify a matched activity as an attack instance. Misuse detection has low false positive rate, but unable to detect novel attacks. It is more accurate but it lacks the ability to identify the presence of intrusions that do not fit a pre-defined signature, resulting not adaptive [4]. Misuse detection discovers attacks based on patterns extracted from known intrusions [5].

II. SVM/MSVM AND K-MEAN ALGORITHM

A. BINARY CLASS SUPPORT VECTOR MACHINE

The basic principle of SVM is finding the optimal linear hyperplane in the feature space that maximally separates the two target classes. The hyperplane which separates the two classes can be defined as:

$$\omega \cdot x + b = 0$$

Here xk is a group of samples:

$\{(x_1, y_1), (x_2, y_2), \cdots (x_k, y_k)\}, x_k \in R^n, y_k\{-1,1\}$, and

k is the number of styles; n is the input dimension; w and b are nonzero constants [6] [7].

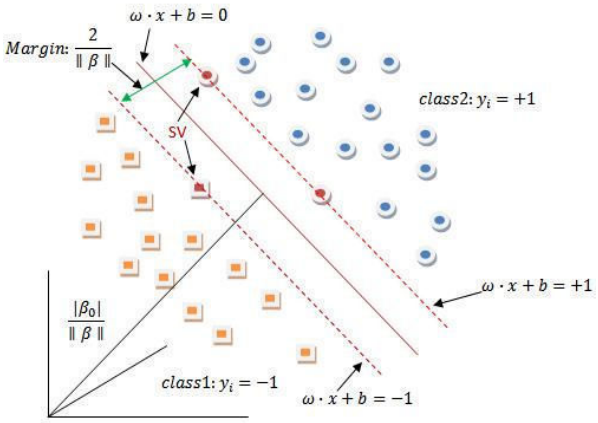

Figure 1. The optimal linear hyperplane (SV=Support vector)

Assume a training set:

$\{(x_1, y_1), (x_2, y_2), \cdots (x_k, y_k)\}, x_k \in R^n, y_k\{-1,1\}$, k is the number of samples. Thus, the problem can be described as:

$$\text{Minimize} \quad \frac{1}{2}\| \omega \|^2 \qquad (1)$$

Subject to $y_i(\omega \cdot x_i + b) \geq 1, i = 1,2,\cdots,k$ .This is a quadratic programming (QP) problem. To solve it, we have to introduce Lagrangian:

$$L(\omega, b, \alpha) = \frac{1}{2}(\omega \cdot \omega) - \sum_{i=1}^{k} \alpha_i \{[(x_i \cdot \omega) + b]y_i - 1\} \qquad (2)$$

According to the Kuhn-Tucher conditions, we obtain

$$\sum_{i=1}^{k} y_i \alpha_i = 0, \omega = \sum_{i=1}^{k} \alpha_i y_i x_i \qquad (3)$$

With the Lagrange multiplier $\alpha \geq 0$ for all i =1, 2… k. So the dual of equation (1) is:

$$\text{maximize} \sum_{i=1}^{k} \alpha_i - \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k} \alpha_i \alpha_j \; y_i y_j (x_i . x_j) \qquad (4)$$

$$\text{subject to} \sum_{i=1}^{k} y_i \alpha_i = 0, \alpha_i \geq 0 (i = 1,2,\cdots,k)$$

For this problem, we also have the complement condition

$$\alpha_i \left(y_i(\omega \cdot x_i + b) - 1\right) = 0$$

So the optimal separating hyperplane is the following indicator function:

$$f(x) = \text{sign}\{(\omega \cdot x) + b\} = \text{sign}\left\{\sum_{i=1}^{i} \alpha_i y_i (x_i \cdot x) + b\right\} \qquad (5)$$

We can obtain the value of vector ω from (3). In the non-linear problem, it can be solved by extending the original set of variables x in a high dimensional feature space with the map Φ. suppose that input vector x Є Rd is transformed to feature vector Φ (x) by a map Φ: Rd→H, then we can find a function K (R', R') → R that satisfies condition K (xi, xj) = Φ(xi).Φ(xj) , so we can replace the inner-product between two vectors (xi , xj)by K (xi, xj) and the QP problem expressed by (4) becomes:

$$\text{maximize} \sum_{i=1}^{k} \alpha_i - \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k} \alpha_i \alpha_j \; y_i y_j (x_i . x_j) \qquad (6)$$

$$\text{subject to} \sum_{i=1}^{k} y_i \alpha_i = 0, \alpha_i \geq 0 (i = 1,2,\cdots,k)$$

The optimal separating hyperplane (5) can be rewritten as:

$$f(x) = \sum_{\text{sup vector}} \alpha_i y_i \Phi(x_i)\Phi(x) + b$$

$$= \sum_{\text{sup vector}} \alpha_i y_i K(x_i, x) + b \qquad (7)$$

B. MULTICLASS SUPPORT VECTOR MACHINE

The multi-class classification problem is commonly solved by decomposition to several binary problems for which the standard SVM can be used. The MSVM can be constructed in two kinds of way: One-Against-All (OAA) and OAO. OAO approach for multi class classification has been shown to perform better than OAA. OAO method constructs k (k −1) / 2 classifiers where each one is trained on data from two classes. For the training data from i-th and j-th classes, we solve the following binary classification problem:

$$\min_{w^{ij}, b^{ij}, \xi^{ij}} \frac{1}{2}(w^{ij})^T w^{ij} + C\sum_{t} \xi_t^{ij} (w^{ij})^T$$

$$s.t. (w^{ij})^T \varphi(x_t) + b^{ij} \geq 1 - \xi_t^{ij}, x_t \text{ belong to i} - \text{th}$$

$$(w^{ij})^T \varphi(x_t) + b^{ij} \leq 1 - \xi_t^{ij}, x_t \text{ belong to j} - \text{th}$$

$$\xi_t^{ij} \geq 0 \qquad (8)$$

After all *k* (*k* −1) / 2 classifiers are constructed, we use the following voting strategy to do future test: if $\text{sign}((w^{ij})^T \varphi(x_t) + b^{ij})$ says ***x*** is in the *i*-th class, then the vote for the *i*-th is added by one. Otherwise, the *j*-th increased by one. Then we predict ***x*** is in the class with the largest vote. In case those two classes have identical votes, we simply select the one with the smaller index. Practically we solve the dual of Eq. (8) whose number of variables is the same as the number of data in two classes. Hence if in average each class has *l* / *k* data points, we have to solve *k* (*k* −1) / 2 quadratic programming problems where each of them has about 2*l* / *k* variables.

C. CUSTOM KERNEL AND SUPPORT VECTOR MACHINE

D. K-MEAN ALGORITHAM[21]

K-means is a centroid-based clustering with low time complexity and fast convergence, which is very important in intrusion detection due to the large size of the network traffic audit dataset. Each cluster in profile can be simply expressed as a centroid and an effect influence radius. So a profile record can be represented as the following format

(Centroid, radius, type)

Centroid is a centric vector of the cluster, radius refers to influence range of a data point (represented as the Euclidean

distance from the centroid), and type refers to the cluster's category, e.g. normal or attack. We can determine whether a vector is in the cluster or not only by computing the distance between the vector and the centroid and comparing the distance with the radius. If the distance is less than radius, we consider that the vector belongs to the cluster. And then we can label the vector as the cluster's type. Therefore, the whole search in the profile only includes several simple distance calculations, which means we can deal with the data rapidly. Of course, not all clusters can serve as the profile. Some maybe include both normal and attack examples and not fit for the profile apparently. It is necessary to select some clusters according to a strategy. A majority example is an example that belongs to the most frequent class in the cluster. The higher the purity is, the better the cluster is served as a profile. A cluster with small purity means that there are many attacks with different types in the cluster, so we don't select such cluster as our profile. Instead, we use them as the training set for classifier. After the clusters are selected for the profile, we put them into the profile repository. The basic contents include centroid, radius and type. Here, we use the type of majority examples in one cluster as the whole cluster's type regardless of the minority examples.

III. PRAPOSED KMSVM MODEL

To separate attacks from legitimate activities, all of the machine learning based intrusion detection technologies will have two main phases, training procedure and detection procedure. As shown in Fig. 2, in the training procedure of KMSVM, K-mean is used to extract the optimal discriminate support vectors of the whole training data. In MSVM for making decision function needs support vectors other vectors far from decision boundary useless for MSVM

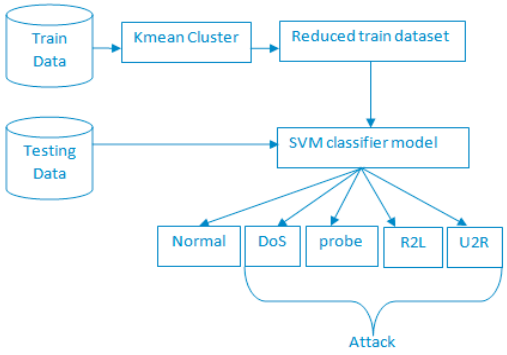

Figure 2. KMSVM model

K-mean clustering algorithm is used to cluster the projected training datasets and remove unused data by using the information provided by the K-mean clustering results, to setup the multiclass SVM detection model. The detection model consists of many binary detection models. Each binary detection model includes two main types of information, the optimal discriminate support vectors extracted from the sub training data by using K-mean and the MSVM classifier based on the projected sub training data. In the detecting procedure, we project the test dataset according to the detection model and then classify the data as normal or malicious by using the detection.

A. KMSVM ALGORITHM

Step 1: three input parameters are selected: the kernel parameter γ, the penalty factor C, and the compression rate CR

Step 2: the K-means clustering algorithm is run on the original data and all cluster centers are regarded as the compressed data for building classifiers

Step 3: SVM classifiers are built on the compressed data

Step 4: three input parameters are adjusted by the heuristic searching strategy proposed in this paper according to a tradeoff between the testing accuracy and the response time

Step 5: return to Step 1 to test the new combination of input parameters and stop if the combination is acceptable according to testing accuracy and response time

Step 6: KMSVM classifiers are represented as the formula in equation (8)

IV. DATASET AND EXPERIMENTS

The KDD Cup 1999 uses a version of the data on which the 1998 DARPA Intrusion Detection Evaluation Program was performed. Each instance in the KDD Cup 1999 datasets contains 41 features that describe a connection. Features 1-9 stands for the basic features of a packet, 10-22 for content features, 23-31 for traffic features and 32-41 for host based features. There are 38 different attack types in training and test data together and these attack types fall into five classes: normal, probe, denial of service (DoS), remote to local (R2L) and user to root (U2R) [14]. In this experiment we use Pentium (IV 3GH) processor, 512 MB RAM, running window XP (SP2) based SVM multiclass [15].The experiment using RBF [16] [17] [18], polynomial and Custom kernel function for intrusion detection (multiclass classification) with parameters as g=0.001, c=0.01, q=50, n=40. Results are shown in below table.

Table1: RBF kernel (confusion matrix)

	normal	Probe	DoS	U2R	R2L
normal	172175	106	0	0	0
Probe	2953	806	301	0	0
DoS	228	0	57721	0	0
U2R	16176	13	0	0	0
R2L	60198	277	75	0	0

Table2: Polynomial kernel (confusion matrix)

	normal	Probe	DoS	U2R	R2L
Normal	156	72	0	0	0
Probe	3529	441	196	0	0
DoS	180309	8	49536	0	0
U2R	16183	6	0	0	0
R2L	60455	102	36	0	11374

Table3: Custom kernel (confusion matrix)

	Normal	Probe	DoS	U2R	R2L
Normal	69	0	159	0	0
Probe	705	0	3461	0	0
DoS	116802	0	113051	0	0
U2R	15361	0	828	0	0
R2L	58637	0	1956	0	0

Above give table 1, table 2 and table 3 show that RBF kernel give good result for class normal, probe and DoS, polynomial kernel gives the best result for R2L and DoS attack and Custom kernel gives best result for DoS class. We used complete dataset for training and testing and got accuracy of multi class support vector machine 73%for whole testing dataset.

V. Conclusion and Future Work

There are many kernel functions which can be used for intrusion detection purpose. Among those we have conducted experiment using RBF, Polynomial and custom kernel function over MSVM. And found that the RBF kernel function's performance is better for intrusion detection. We can improve over all performance of the MSVM for four types of attack and normal (five classes) by combining above three kernels into one.

References

[1] Jai Sunder balasubamaniyan, Jose Dmar Garcia-Fernandez, David Isacoffet.al, "An Architecture for Intrusion Detection using Autonomous Agents," COAST Laboratory, Purdue University, COAST Technical Report june 1998.

[2] S. Axelsson, "Research in intrusion-detection systems: A survey," Department of Computer Engineering, in Chalmers University of Technology, December15, 1998.

[3] Y. Liu, D. Tian, and B. Li, 2006, "A Wireless Intrusion Detection Method Based on Dynamic Growing Neural Network," *presented at Proceedings of the First International Multi-Symposiums on Computer and Computational Sciences (IMSCCS'06)*.

[4] K. Ilgun, R. A. Kemmerer, and P. A. Porras, March 1995,"State transition analysis: A rule-based intrusion detection approach," *IEEE Transactions on Software Engineering,* 21(3):181–199.

[5] D. Barbara and S. Jajodia, 2002,"Applications of Data Mining in Computer Security," *Norwell, MA: Kluwer*.

[6] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*

[7] Hui Zhao, Yong Yao, and Zhijing Liu Liu, 2007, "A Classification Method Based on Non-linear SVM Decision Tree", *Fourth International Conference on Fuzzy Systems and Knowledge Discovery*.

[8] Kuan-Ming Lin and Chih-Jen Lin, NOV 2003, "A Study on Reduced Support Vector Machines", *IEEE Transactions On Neural Networks*, VOL. 14, NO. 6.

[9] R.Debnath, H.Takahashi, July 16, 2006, "SVM Training: Second-Order Cone Programming versus Quadratic programming", *IEEE International Joint Conference on Neural Networks, Canada*, pp.1162-1168.

[10] K. Goh, E. Chang, K. Cheng, November 10, 2001,"SVM Binary Classifier Ensembles for Image Classification," A*tlanta, Georgia, USA,CIKM'*01.

[11] Bernhard Pfahringer, January 2000, "Winning the KDD99 Classification Cup: Bagged Boosting", *ACM SIGKDD Explorations Newsletter*, Volume 1, Issue 2, p. 65-66.

[12] "http://kdd.ics.uci.edu/databases/kddcup99/task.html", KDD Cup1999.

[13] T. G. Dietterich and G. Bakiri, 1995, "Solving multiclass learning problems via error-correcting output codes", *Journal of Artificial Intelligence Research*, 2:263-286.

[14] Thorsten Joachims Cornell University, Department of Computer Science, "http://svmlight.joachims.org".

[15] M. Bianchini, P. Frasconi, and M. Gori, May 1995, "Learning without local minima in radial basis function networks*," IEEE Tranaction. Neural Network.,* vol 6, no. 3, pp. 749–756.

[16] C. M. Bishop, 1991 ,"Improving the generalization properties of radial basis function neural networks," *Neural Computat.*, vol. 3, no. 4, pp. 579–588,.

[17] M. J. L. Orr, 1996, "Introduction to radial basis function networks," *Center Cognitive Sci., Univ. Edinburgh, Edinburgh, U.K.*

[18] Andrea Passerini, Massimiliano Pontil, and Paolo Frasconi, Member, IEEE, January 2004,"New Results on Error Correcting Output Codes of Kernel Machines" *IEEE Transactions on Neural Networks*, Vol. 15.

[19] Srinivas Mukkamala, Andrew H. Sunga, Ajith Abraham, 7 January 2004, "Intrusion detection using an ensemble of intelligent paradigms", *Elsevier, Science Direct.*

[20] Jiaqi Wang, Xindong Wu, Chengqi Zhang, 2005, "Support vector machines based on K-means clustering for real-time business intelligence systems", Int. J. Business Intelligence and Data Mining, Vol. 1, No. 1.

[21] Hongyu Yang1,2, Feng Xie3, and Yi Lu4," Clustering and Classification Based Anomaly Detection", FSKD 2006, LNAI 4223, pp. 1082–1091, Springer-Verlag Berlin Heidelberg 2006.