Intrusion Detection System using Support Vector Machine and Decision Tree

Snehal A. Mulay
Bharati Vidyapeeth University,
Pune, Maharashtra
India

P.R. Devale
Bharati Vidyapeeth University
Maharashtra
India

G.V. Garje
Pune University
Maharashtra
India

ABSTRACT

Support Vector Machines (SVM) are the classifiers which were originally designed for binary classification. The classification applications can solve multi-class problems. Decision-tree-based support vector machine which combines support vector machines and decision tree can be an effective way for solving multi-class problems. This method can decrease the training and testing time, increasing the efficiency of the system. The different ways to construct the binary trees divides the data set into two subsets from root to the leaf until every subset consists of only one class. The construction order of binary tree has great influence on the classification performance. In this paper we are studying an algorithm, Tree structured multiclass SVM, which has been used for classifying data. This paper proposes the decision tree based algorithm to construct multiclass intrusion detection system.

Keywords

Intrusion detection system, support vector machine, decision tree.

1. INTRODUCTION

Security is becoming a critical issue as the Internet applications are growing. The current security technologies are focusing on encryption, ID, firewall and access control. But all these technologies cannot assure flawless security. The system security can be enhanced by Intrusion detection. The ability of an IDS to classify a large variety of intrusions in real time with accurate results is important. The patterns of user activities and audit records are examined and the intrusions are located.

IDSs are classified, based on their functionality, as misuse detectors and anomaly detectors. Misuse detection system uses well defined patterns of attack which are matched against user behavior to detect intrusions. Usually, misuse detection is simpler than anomaly detection as it uses rule based or signature comparison methods. Anomaly detection requires storage of normal usage behavior and operates upon audit data generated by the operating system.

Support vector machines(SVM) are the classifiers which were originally designed for binary classification[5][6], can be used to classify the attacks. If binary SVMs are combined with decision trees, we can have multiclass SVMs, which can classify the four types of attacks, Probing, DoS, U2R, R2L attacks and Normal data, and can prepare five classes for anomaly detection. Our aim is to improve the training time, testing time and accuracy of IDS using the hybrid approach.

The paper is organized as follows. Section 2 describes the basic principles of SVM. Section 3 discusses the Multiclass SVM. In section 4 we will study the decision tree method to implement multiclass SVM. And the section 5 describes the proposed method to implement the IDS .

2. PRINCIPLES OF SUPPORT VECTOR MACHINE

Binary classification problems can be solved using SVM [4]. An SVM maps linear algorithms into non-linear space. It uses a feature called, kernel function, for this mapping. Kernel functions like polynomial, radial basis function are used to divide the feature space by constructing a hyperplane. The kernel functions can be used at the time of training of the classifiers which selects support vectors along the surface of this function. SVM classify data by using these support vectors that outline the hyperplane in the feature space.

This process will involve a quadratic programming problem, and this will get a global optimal solution. Suppose we have N training data points $\{(x_1, y_1), (x_2,y_2), (x_3, y_3), ..., (x_N , y_N)\}$, where $x_i \in R_d$ and $y_i \in \{+1,-1\}$. Consider a hyper-plane defined by (w, b), where w is a weight vector and b is a bias. The classification of a new object x is done with

$$f(x) = \text{sign}(w.x + b) = \text{sign}(\sum_{i}^{N} \alpha_i y_i (x_i \cdot x) + b)$$

The training vectors x_i occur only in the form of a dot product. For each training point, there is a Lagrangian multiplier α_i. The Lagrangian multiplier values α_i reflect the importance of each data point. When the maximal margin hyper-plane is found, only points that lie closest to the hyper-plane will have $\alpha_i > 0$ and these points are called support vectors. All other points will have $\alpha_i = 0$. That means only those points that lie closest to the hyper-plane, give the representation of the hypothesis/classifier. These data points serve as support vectors. Their values can be used to give an independent boundary with regard to the reliability of the hypothesis/classifier.

3. MULTICLASS SUPPORT VECTOR MACHINE

Multiclass SVM constructs k different classes at the training phase of IDS. Some typical methods [2][10] , for construction of multiclass SVM are one-versus-rest(OVR), one-versus-one(OVO) and method based on Directed Acyclic Graph(DAG).

3.1 One-versus- Rest

It constructs *k* two-class SVMs. The *i*th SVM (*i* = 1, 2, . . . , *k*) is trained with all training patterns. The *i*th class patterns are labeled by 1 and the rest patterns are labeled by −1. The class of an unknown pattern $\boldsymbol{x}$ is determined by argument max*i*=1,2,...,*k* $f_i(\boldsymbol{x})$, where $f_i(\boldsymbol{x})$ is the decision function of the *i*th two-class SVM. In short, a binary classifier is constructed to separate instances of class yi from the rest of the classes. The training and test phase of this method are usually very slow.

3.2 One-versus-one

It constructs all possible two-class SVM classifiers. There are $k(k-1)/2$ classifiers by training each classifier on only two out of *k* classes. A Max Wins algorithm is used in test phase,: each classifier casts one vote for its favored class, and finally the class with most votes wins. The number of the classifiers is increased super linearly when *k* grows. OVO becomes slow on the problem where the number of classes is large.

3.3 Directed Acyclic Graph

The training of DAG is same as OVO. DAG uses tree-structured two-class SVMs to determine which class an unknown pattern belongs to. DAG is constructed in test phase. So the classification of DAG is usually faster than OVO.

4. DECISION TREE BASED SVM

Decision tree based SVM [1] is also a good way for solving multiclass problems. It combines the SVM and the decision tree approaches for preparing decision making models. Integrating different models gives better performance than the individual learning or decision making models. Integration reduces the limitations of individual model.

The problem of existence of unclassifiable regions can be resolved by decision tree based SVM for multi-classification. Our proposed system, as shown in Figure 1, prepares five SVM models for five types of labeled data. This data include four types of attacks and normal data. The five types of patterns are then organized into a binary tree.

We are using KDD CUP'99 intrusion detection data set (TCP dump data)[8][9], which is most commonly used for evaluation . The data has 41 attributes for each connection record plus one class label. The data set contains 24 attack types, which are categorized into four types as follows:

1. **Denial Of Service (DOS)** : In this type of attack legitimate user is denied to access a machine by making some computing resources or memory full. For example TCP SYN, Back etc.
2. **Remote to User (R2L)** : In this type of attack remote user tries to gain local access as the user of the machine. Foe example FTP_write, Guest etc.

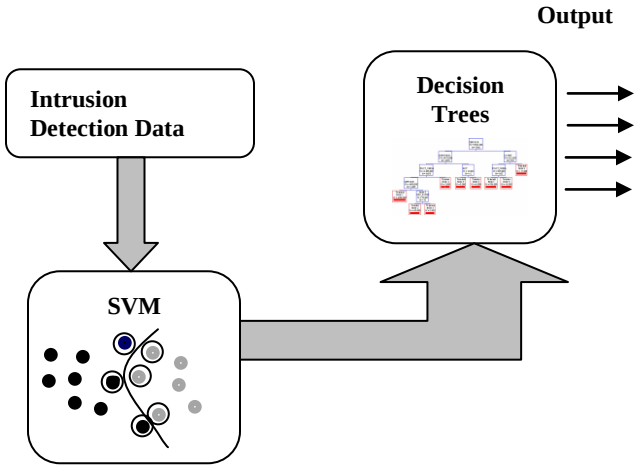

Figure 1. The proposed IDS

3. **User to Root (U2R)** : In this type of attack the attacker tries to gain root access to the system. For example Eject, Fdformat etc.
4. **Probing** : In this type of attack attacker tries to scan a network of computer to fine known vulnerabilities or to gather information. For example Ipsweep, Mscan.

The IDS prepares a binary decision tree of the SVMs which at the first stage partitions the data into two classes "normal and "attack". At later stages, we repeat the process for the four types of attacks.

5. THE PROPOSED METHOD

In [2], the algorithm for multiclass support vector machines is used for classification of US Postal Service data set. We are proposing the algorithm for our IDS [11], as it gives better performance results for multiclass classification.

The decision tree model in this method consists of a series of two class SVMs. The structure of the tree is determined by distance between two class patterns and the number of each class patterns. Let n_i denote the number of i^{th} class patterns $\boldsymbol{x}_i$, where *i* = 1, 2, . . . , *k*. The center point of i^{th} class patterns is calculated using following equation:

$$c_i = \frac{\sum_{m=1}^{n_t} x_m^i}{n_i} \qquad (1)$$

The Euclidian distance between i^{th} class and j^{th} class patterns is calculated as follows:

$$Ed_{ij} = \left\| c_i - c_j \right\|$$

The separability or the distribution of two class patterns is given by a distance equation as follows:

$$d_{ij} = \frac{Ed_{ij}}{\gamma_i + \gamma_j} \quad (2)$$

where,

$$\gamma_i = \frac{\sum_{m=1}^{n_i} \left\| x_m^i - c_i \right\|}{n_i}$$

obviously d_{ij} equals to d_{ji}. Find a pair (i^*, j^*) by calculating the distances of all pairwise classes, where the distance between $\boldsymbol{x}^{i*}$ and $\boldsymbol{x}^{j*}$ is extreme. To explain how to construct the tree, an example of five-class pattern recognition problem is illustrated as shown in Figure 2. The five classes are denoted by a set

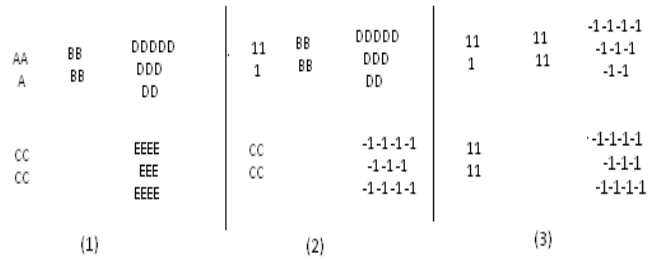

Figure 2 How to separate five class patterns into two classes

{A,B,C,D,E}. From Figure 2.1, the distance between Class *A* and E, *dAE* is biggest. In the pair (*A*,*H*), let *A* be Class 1 and *E* be Class −1. Next, separate the rest three class patterns *{B,C,D}* into two classes (Class 1 or −1). Compare the distances *dBA* with *dBE*. The distance *dBA* < *dBE*, so Class *B* is put into Class 1. Similarly, calculate the distances for C and D, where *dCA* < *dCE*, *dDA* > *dDE*.The result is as shown in Figure 2.3. At last, compare the number of patterns in Class 1and −1. *n*1(*n*−1) denote the number of patterns in Class1(−1).

$$\mu = \frac{\min(n_1, n_{-1})}{\max(n_1, n_{-1})} \quad (3)$$

Parameter μ is the balance of distribution of five-class patterns.

$$v_e = \frac{d_{eA}}{d_{eH}} \quad (4)$$

$$\mu' = \frac{\min(n_1 + n_F, n_{-1} - n_F)}{\max(n_1 + n_F, n_{-1} - n_F)} \quad (5)$$

Supposition holds, if μ' is larger than the original parameter μ, and if not, the supposition is canceled. Repeat the operation until the parameter μ is not increased

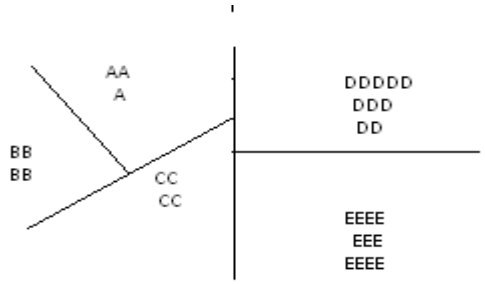

Figure 3.1) An intuitive division of the five class patterns.

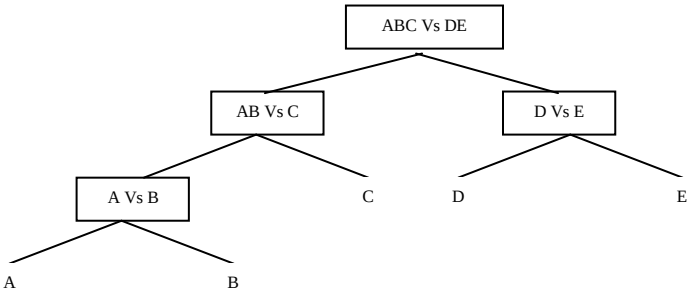

Figure 3.2) A decision tree for classifying the five classes

any more. At the end the five class patterns are divided into two subsets: *{A,B,,C}* and *{D,E}*. These two class patterns give us a classifier as (1). Now construct a decision tree by separating the subsets and by constructing the corresponding classifiers recursively until each subset remains with only one element. Figure 3 shows the tree-structured SVM obtained using this method. Figure 3.1 is the division of the five class patterns and Figure 3.2 is a decision tree. An unknown pattern will start testing from the root, and will end at any of the leaves by traveling the decision tree. This leaf node will be the resulting class of the data.

SVM is slow for large size problems. To solve this problem many decomposition methods can be used which decomposes a large QP problem into small sub-problems of size two. The Sequential Minimal Optimization (SMO) algorithm [7] has been used for decomposition of two class SVMs, in this algorithm. The LIBSVM is adopted to train and test every SVM.

In short, the steps of execution will be as shown in Figure 4.

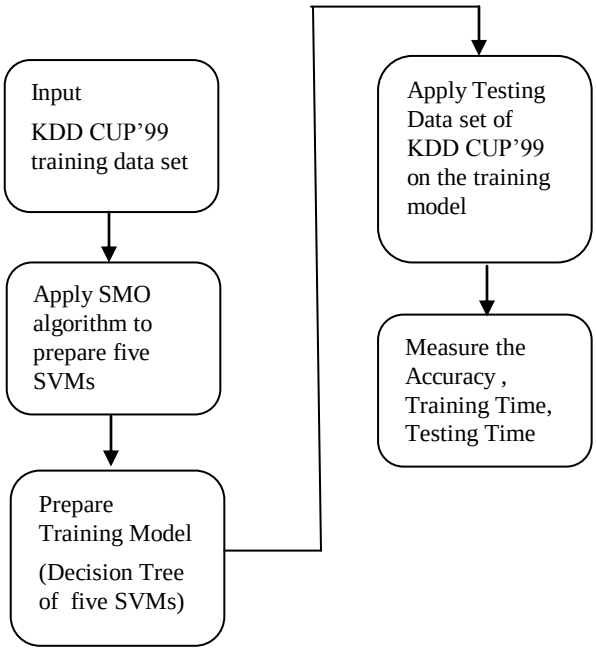

Figure 4. Steps for experimentation

6. CONCLUSION

The paper proposes the above novel multiclass SVM algorithm for implementation of Intrusion Detection System. The integration of Decision tree model and SVM model gives better results than the individual models. The final results for the proposed system are not available yet, but it seems that multi-class pattern recognition problems can be solved using the tree-structured binary SVMs and the resulting intrusion detection system could be faster than other methods.

7. REFERENCES

[1] Sandya Peddabachigari, Ajith Abraham, Crina Grosan, Johanson Thomas. Modeling Intrusion Detection Systems Using Hybrid Intelligent Systems. Journal of Network and Computer Applications-2005.

[2] Jun GUO , Norikazu Takahashi, Wenxin Hu . An Efficient Algorithm for Multi-class Support Vector Machines. IEEE-2008.

[3] Latifur Khan, Mamoun Awad, Bhavani Thuraisingham. A new intrusion detection system using support vector machines and hierarchical clustering. The VLDB Journal DOI 10.1007/s00778-006-0002 , 2007.

[4] V. N. Vapnik. The nature of statistical learning theory. Springer-Verlag,New York. NY, 1995.

[5] Xiaodan Wang, Zhaohui Shi, Chongming Wu and Wei Wang. An Improved Algorithm for Decision-Tree-Based SVM. IEEE-2006.

[6] Pang-Ning Tan, Michael Steinbach, Vipin Kumar. Introduction to data mining. Pearson Education.

[7] K. Crammer and Y. Singer. On the algorithmic implementation of multiclass kernel-based vector machines. Journal of Machine Learning Research, 2:265–292, 2001.

[8] YMahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani. A detailed analysis of KDD CUP'99 data set. IEEE-2009.

[9] http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[10] C. W. Hsu, C. J. Lin. A comparison of methods for multiclass support vector machines. IEEE Trans. On Neural Networks, vol. 13, no. 2, pp.415-425, 2002.

[11] Snehal Mulay, P.R. Devale, G.V. Garje. Decision Tree based Support Vector Machine for Intrusion Detection. ICNIT-2010, unpublished.

[12] Lili Cheng, Jianpei Zhang, Jing Yang, Jun Ma. An improved Hierarchical Multi-Class Support Vector Machine with Binary Tree Architecture**"** 978-0-7695-3112- 0/08 2008 IEEE DOI 10.1109/ICICSE.2008