Australian Journal of Basic and Applied Sciences, 7(7): 480-485, 2013
ISSN 1991-8178

Two Features Selection Algorithmsbased Onensemble of SVM Classifier For Intrusion Detection

[1]IhsanAbodKhalaf, [1]Abdallah M Abualkishik, [2]Abdulla Amin Aburomman, [2]Mamun Bin IbneReaz

[1]College of information technology, UniversitiTenaga Nasiona
[2]Department of Electrical and Electronics Engineering, UniversitiKebangsaan Malaysia

Abstract: With the huge expanding in internet usage and with the existence of many hacking tools and people who like to put their hands on others secret information to get their benefit of it, the existing preventive procedures are not appropriate for avoiding such activities. Network security is very important issue. As long as computers are connected to internet, Intruders or hackers will always try to explore the network security imperfection, they will try different types of attacks and techniques, so protecting the network, and developing a strength measure protection is essential to avoid organizational and personal loss or data damage. In this paper the network intrusion detection system (IDS) that is depending on support vector machines (SVM) classifier and two different features selection algorithms which are (SOM) self-organizing map and (PCA) Principle component analysis is presented. The novelty in this paper is the combination of both feature selection algorithm using voting technique. Combination provides better results comparing to stand alone feature selection based on SVM classifier. For evaluating the proposed system, a random subset of KDD99 data set is chosen. Results of this work showed that different features selection algorithm can affect the classification output in different manner. A comparison study between the SOM and PCA based on binary SVM classifier is presented in respect to their accuracy results, then an ensemble of SVM classifiers and vetoing technique applied so that SVM can select the best features which results in a best accuracy.

Key words: Classifiers, Computer networks, Data mining, KDD99

INTRODUCTION

Among the list of classification techniques problems, intrusion detection still needs researcher's attention to deal with. The main concern in intrusion detection challenged by researchers is the dependability issue. When the invasive routines in the network are not familiar to the current expertise, thus false alarms will tend to be produced(Kok-Chin, Choo-Yee *et al*. 2009). Numerous studies have targeted to boost the recognition rate of IDs by way of recommending new classifiers, but growing the performance of classifiers seriously is not an easy task, though feature selection might be accustomed to optimize the current classifiers. Feature selection techniques are actually brought to getting rid of the trivial features in intrusion detection filed. Feature selection is advantageous to lower the training time (computational complexity), redundancy elimination of the information, accomplish data familiarity, generalization improvements and accuracy enhancing of the classifier. (Amiri, Rezaei Yousefi *et al*. 2011).

Feature selections led to increase overall accuracy, increased low frequency detection of instances in the training data, and reduced the amount of false positives(Dartigue, Hyun Ik *et al*. 2009).Both SOM and PCA are good choice as feature selection to reduce the dimensionality of data, yet it doesn't prove which of them is better in terms of comparison. SVM keeps small error on separate data set records as soon as the training data records are limited. Consequently Support Vector Machines captivated plenty of researchers in the field of intrusion detection (Xiaozhao, Wei *et al*. 2010). This paper present a comparison study between those two feature selection algorithms with regards to their effectiveness, false positive, false negative and overall accuracy based on ensemble of binary SVM classifiers. Then a combination between two feature selections is presented based on SVM using Weighted Majority Voting (WMV).

This paper is structured as follow; section 2 presented an overview of SVM classifier, PCA and SOM features selection. Section 3 presented methodology, dataset, pre-processing and proposed system. Section 4 experiments and results. Section 5 conclusion presented.

Classifierandfeaturesselection:

A. Support Vector Machines:

SVM has become the greatest classification algorithms, and it has just lately captivated plenty of researchers because of its appealing properties such as high generalization performance and globally optimal solution. Support vector machines creates a hyper-plane that separate higher dimensional feature space into two classes(Jiaqi, Ru *et al*. 2011).

In figure 1 a classical example of SVMs linear classification is depicted.

Corresponding Author: IhsanAbodKhalaf, College of information technology, UniversitiTenaga Nasiona
E-mail: reoroman@yahoo.com

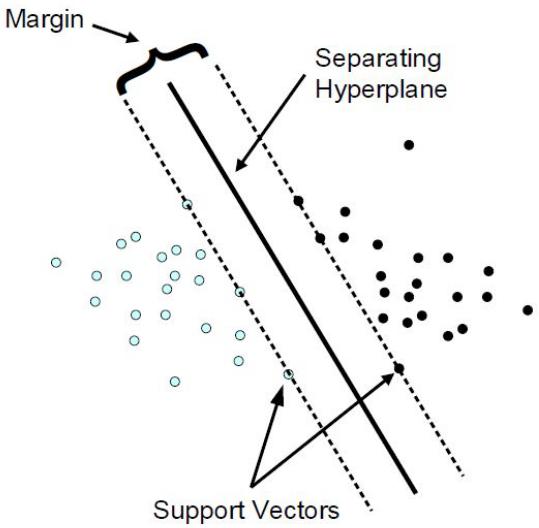

Fig. 1: Classical example of SVMs linear classifier

Assume we have now N training points of data$\{(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_N, y_N)\}$, where$x_i \in R^d$and $y_i \in \{+1, -1\}$. Think about a hyper plane determined by(w, b), exactly where w is really a weight vector and also b is really a base. A brand new object x might be classified using the subsequent functionality:

$$f(x) = sign\,(w.x + b) = sign\,(\sum_{i=1}^{N} \propto_i y_i(x, x_i) + b) \quad (1)$$

Where $\propto_i$ is called Lagrange multipliers.

When data are not capable of being divided or dissociated (linearly separated), the kernel trick function can be implemented to convert data into dimension of higher space, then it can be possible to implement a linear model. Below is the SVM nonlinear decision function:

$$f(x) = sign\left(\sum_{i=1}^{N} \propto_i y_i k(x, x_i)) + b\right) \quad (2)$$

Where K (x, x_i) is the Kernel Function.

The RBF kernel has a lesser amount of mathematical complications comparing to polynomial kernel, because its value are lying in between 0 and 1 whilst the polynomial might have to go to 0 or infinity(Eid, Darwish *et al*. 2010),for that reason RBF is used in this field of study.

B. PCA feature selection:

To improve the classification performance, feature selection is presented as a key factor for this purpose by searching for features subset which is best suited to the classifier algorithm. It is important to remove the redundant and irrelevant features which contribute toward better classification accuracy, which also important in real time detection. PCA is a vital algorithm for features selection and data reduction in intrusion detection (Eid, Darwish *et al*. 2010)

Suppose that $\{X_t\}$in which $t = 1, 2 \ldots, N$ random n-dimensional data input tend to be report using mean (μ). Can be described by this equation:

$$\mu = \frac{1}{N}\sum_{t=1}^{N} x_t \quad (3)$$

Covariance matrix of x_t described by

$$c = \frac{1}{N}\sum_{t=1}^{N}(x_t - \mu).(x_t - \mu)^T \quad (4)$$

Eigenvalue problem of covariance matrix C Can be solved by PCA as follow:

$$Cv_i = \lambda_i v_i \quad (5)$$

Where eigenvalues are $\lambda_i (i = 1, 2, \dots, n)$ and the corresponding eigenvectors are $v_i (i = 1, 2, \dots, n)$. Computing the m eigenvectors to the m largest eigenvectors (m < n) can represent data samples with low dimensional vectors.

Let

$$\emptyset = [v1, v2, \dots, v_m], \Lambda = diag[\lambda_1, \lambda_2, \dots, \lambda_m] \quad (6)$$

Then

$$C\emptyset = \emptyset\Lambda \quad (7)$$

Relation holds can be represented by the following equation when estimation accuracy of m biggest eigenvectors announced by v argument.

$$\frac{\sum_{i=1}^{m} \lambda_i}{\sum_{i=1}^{n} \lambda_i} \geqslant v \quad (8)$$

Based on last two equations, eigenvectors numbers could be chosen so accuracy parameter v is offered, the reduced dimensional feature vector of the brand new input data x solved by:

$$x_f = \emptyset^t{}_x \quad (9)$$

In the experiment, in PCA algorithm, Kaiser's rule has been used to find the numbers of best features by applying this formula

$$\lambda_i > \frac{\sum_{n=1}^{N} \lambda_n}{N} \quad (10)$$

C. SOM feature selection:

SOM or self-organizing map is an algorithm which belongs to neural networks. There is essential process during the development of self-organizing map. The first one is called competition, which means, for each and every input pattern, the neurons within the output layer they will determine the value of a function named discriminate function, so each neuron will compute a discriminate function, and this function provides the basic of the competition, so the actual neuron using the largest discriminate functionality is declared the winner. The mathematical module of the competitive process with considering m-dimensional input as follow

$$\vec{X} = [x_1, x_2, \dots, x_m]^t \quad (11)$$

$$\overrightarrow{w_j} = \left[w_{j1}, w_{j2}, \dots, w_{jm}\right]^t \; where \; j = 1,2 \dots, L \quad (12)$$

Where L would be the count of output neurons within the network.

To figure out the very best match up between $\vec{X}$ and $\overrightarrow{w_j}$ we have to compute $\overrightarrow{w_j}^t\vec{X}$ for $j = 1, 2, \dots, L$ and we have to choose the biggest among that. So j gives us the largest value is the winner. In this case we maximizing $\overrightarrow{w_j}^t\vec{X}$where this effect is about to minimizing the Euclidian distance between the$\left\|\vec{X} - \overrightarrow{w_j}\right\|$. So using the index $i(\vec{X})$where i is the index and it is indexes based on some input vector x, and the index $i(\vec{X})$= arg $\min_j \left\|\vec{X} - \overrightarrow{w_j}\right\|$ and the corresponding weight vector v1 $i(\vec{X})$ is the closest weight vector.

The second one is the cooperation; here the succeeding neuron determines the unique place of topological neighborhood of excited (thrilled) neurons. So excitation is actually cooperation because it also strength the neurons closer to it. And the neurons which is far will be eliminated by the winner takes all mechanism. This can be described by the following formula:

$$hci\,(t) = h\,(\|rc - ri\|); t) \quad (13)$$

Where rc a representative of succeeding neuron and c, ri a representative in the position of excitatory neurons I. the most used in the analysis of the data is bubble function. When time $\|rc - ri\| < R(t), hci\,(t) = \alpha(t)$, otherwise, there is $hci\,(t) = 0$; where in$\alpha(t)$is performance of learning, $\alpha(t)$ scaled-down the convergence much more slower, the mandatory learning training more longer, $\alpha(t)$is commonly shown by formulation$\alpha(t) =$

$A_2E^{-\frac{1}{T_2}}$ where $A_2$is the foremost learning speed in the training from the outset, T2 signifies the rate of decay.

The 3rd step could be the synaptic adaptation; it allows the excited neurons to improve their own particular person values of discriminant function with regards to the input pattern. So the excited neurons will have their discriminant value increased, and can show by this formula (Li and Wang 2009).

$$w_i(t+1) += w_i(t) + hci\ (t)\ [x(t) - w_i(t)] \tag{14}$$

Methodology:

A. Data preprocessing and grouping:

- Conversion of symbolic features to numeric.

As it can be seen in the KDD99 data set in features 2, 3 and 4, they're a symbolic characters for which we need to convert it to numeric as well as the classifier requires a numeric form. So for example, a UDP can be replaced by 1 and TCP replaced by 2, etc…

- Normalization in where all normalized values of each feature fall between range 0 and 1. The normalization equation used in this paper is as follow:

$$N_i(x) = \frac{(i(x) - V_{min}(x))}{V_{max}(x) - V_{min}(x)} \tag{15}$$

- Discretization (continues to discrete conversion).Conversion of continuous features to discrete using Equal Frequency Discretization (EFD) method.
- Feature selections, using the above stated two methods, PCA and SOM.
- Binary classification for every class-pair using SVM.
- Voting of all SVM and ensemble of classifiers on testing dataset. Classifier outputs combined by Majority Voting, a method named: Weighted Majority Voting (WMV).

B. Dataset

For evaluating the performance intrusion detection system, the KDD CUP99 datasets have been selected.KDD99 is one of the common used data set for intrusion detection experiments. It is considered the only one for which supplies labels to training and testing data sets (Koc, Mazzuchi *et al*. 2012).

The original data have both training and testing data set, where every connection has 41 different features. In the experiment a random subset of training records has been taken from training file"kddcup.data_10_percent.gz"and a random subset of testing records has been taken from the testing file "corrected.gz", samples are as bellow:

	normal	Probe	DoS	U2R	R2L
Training	2000	2000	1000	52	600
Testing	2000	2000	1000	52	800

C. Proposed system

Figure 2 depict the presented system framework, where all data being grouped and pre-processed in the first stage, followed by feature selection and classification then ensemble stage.

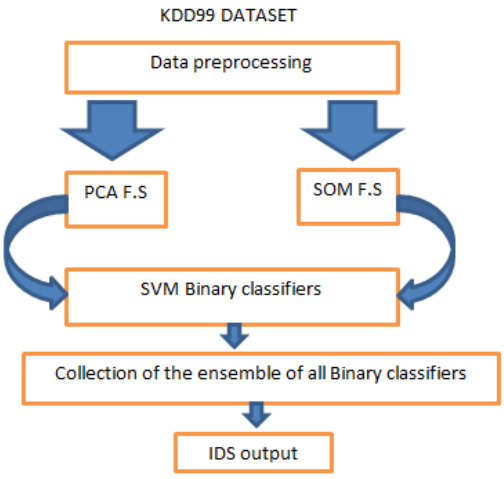

Fig. 2: Proposed system

To improve the classification performance, 20 SVMs has used in this experiments, so in the training phase the feature selection algorithm has fed the 10 classifiers with 10 pairs of classes, so that every classifier become an expert on different training samples. In the testing phase all 20 classifiers will combined by "majority vote" so that the overall performance can be effectively improved.

Experiments And Results:

Experiments done on Matlab 2011b-64bit.KDD Cup (1999) data was used in this experiment to compare between two feature selection algorithms which are SOM and PCA based on SVM. After data pre-processing stage, the PCA and SOM has employed to reduce the dimensionality. In testing of PCA base SVM, results can be seen in table 1 below.

Table 1: PCA based SVM testing results confusion matrix

normal	Probe	DoS	U2R	R2L
0.981	0.003	0.01	0.0005	0.0055
0.251	0.7475	0.001	0.0005	0
0.038	0.002	0.96	0	0
0.6731	0.0192	0	0.25	0.0577
0.8833	0.0133	0	0.0067	0.0967

Overall accuracy = 0.819
FP (False positive) = 0.019
FN (False negative) = 0.2457

In table 2, SOM based on SVMconfusion matrix results.

Table 2: SOM based SVM testing results confusion matrix

normal	Probe	DoS	U2R	R2L
0.9465	0.0095	0.0115	0.026	0.0065
0.058	0.9255	0.001	0.002	0.0135
0.037	0.146	0.817	0	0
0.4423	0.0385	0	0.4038	0.1154
0.8817	0.0083	0	0.0167	0.0933

Overall accuracy = 0.82005
FP = 0.0535
FN = 0.1595

Then the experiment of ensemble of best pairs can be seen in table 3 which proves the effectiveness of ensemble technique.

Table 3: Results confusion matrix of Ensemble PCA and SOM

normal	Probe	DoS	U2R	R2L
0.9835	0.007	0	0	0.0095
0.1075	0.8815	0.001	0.009	0.001
0.0385	0.0045	0.957	0	0
0.6154	0	0	0.1731	0.2115
0.8883	0.0033	0	0.0067	0.1017

Overall accuracy = 0.85899
FP = 0.0165
FN = 0.18422

Conclusion:

In this paper a new method of combining two features selection algorithms has been implemented using PCA and SOM based on 20 binary Support Vector machines, then ensemble techniques using majority vote applied to determine the best pairs. It is noticed that the overall accuracy of ensemble technique has been enhanced comparing to one feature selection algorithm, so combining both SOM and PCA feature selection could achieve better detection. Everyfeatureselection algorithm has its advantages and disadvantages for selecting the best features which will affectthe classification output, so instead of them competing to each other, they will work together. Empowering them together could prove better classification with the help of SVM and it is seen in the experiments that it could enhance the overall performance of IDS

REFERENCES

Amiri, F., M. Rezaei Yousefi, *et al*. 2011. Mutual information-based feature selection for intrusion detection systems. Journal of Network and Computer Applications., 34(4): 1184-1199.

Dartigue, C., J. Hyun Ik, *et al*. 2009. A New Data-Mining Based Approach for Network Intrusion Detection. Communication Networks and Services Research Conference, 2009. CNSR '09. Seventh Annual.

Eid, H.F., A. Darwish, *et al*. 2010. Principle components analysis and Support Vector Machine based Intrusion Detection System. Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on.

Jiaqi, J., L. Ru, *et al*., 2011. A New Intrusion Detection System Using Class and Sample Weighted C-support Vector Machine. Communications and Mobile Computing (CMC), 2011 Third International Conference on.

Koc, L., T. A. Mazzuchi, *et al*. 2012. A network intrusion detection system based on a Hidden Naïve Bayes multiclass classifier. Expert Systems with Applications, 39(18): 13492-13500.

Kok-Chin, K., T. Choo-Yee, *et al*., 2009. A Feature Selection Approach for Network Intrusion Detection. Information Management and Engineering, 2009. ICIME '09. International Conference on.

Li, M. and D. Wang, 2009. Anormaly Intrusion Detection Based on SOM. Information Engineering, 2009. ICIE '09. WASE International Conference on.

Xiaozhao, F., Z. Wei, *et al*., 2010. A Research on Intrusion Detection Based on Support Vector Machines. Communications and Intelligence Information Security (ICCIIS), 2010 International Conference on.