International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 12 (2015) pp. 28961-28967

http://www.ripublication.com

Entropy Based Support Vectors And Ant Colony Networks For Intrusion Detection

G.Sravani

M. Tech 2nd year, department of CSE
SVEC, Tirupati, AP, India
giddalur.saisravani@gmail.com

Abstract

Intrusion detection normally generates a large number of false rate alerts particularly in large scale networks. That results in the dispute on efficiency and accuracy of network attack detection. In this paper for the classification of data as normal and abnormal, a new algorithm ESVAC (entropy based support vectors and ant colony networks) is proposed that is employed to network intrusion detection to take rewards of two algorithms keeping off their disadvantages. The algorithm followed out and valuated using KDD99 data set. As a result, experiments show that ESVAC performs better in price of training time and detection accuracy.

Keywords: classification, entropy, similarity, svm, ids, ant colony networks.

1. Introduction

Intrusion demeanors force to great damage of systems. So go-aheads search for intrusion detection systems to protect their systems. Intrusion detection is to notice the actions that are attempted to compromise the confidentiality, integrity or availability in the network. Intrusion Detection Systems are contrived to guard computer systems from several computer viruses and cyber attacks.

IDS figures out effectual classification models to separate abnormal behaviors from normal behaviors.

Data mining supported approaches for IDS:

Number of present IDSs still cannot sufficiently deal with dynamic computing environments, therefore frequent updating is required to the installed IDS. As it is time and energy consuming business to update manually, so employing data mining

approaches to IDS to acquire demeanor of networks automatically by canvassing the data trails of their actions.

The main idea of data mining approaches is to take a data centric point of view and conceive intrusion detection as a data analysis process, that let in four steps as

1. Capturing data
2. Taking out set of features
3. Learning a model that depicts the activities as normal and abnormal
4. Detecting the intrusions

2. Literature review

Yihua Liao, V.Raovemuri[1] has explained about k-nearest neighbor classifier as it is an algorithm that is use to a database where data points are separated to predict classification. In the KNN algorithm points are selected randomly selected from training set and points which are similar and near are taken into groups to train classifier. But reliability, timing, scaling are limitations for this algorithm.

Amir Azimi Alsti Aharbi[2] has explained about self organized maps as it is not a classification algorithm, but also used for the classification purpose. By using som ratio of errors may be high but only TP (true positive) alerts are trained and FP (false positive) alerts are ignored.

Wenying Feng[5] has explained about hybrid algorithm CSVAC (combining support vectors with ant colony) in which points are clustered by the ant colony and trains support vector machine. It may give high accuracy but it takes more time to train all 41 features in a record.

In this paper, we proposed entropy based algorithm as ESCAV (entropy based support vectors and ant colony networks) so that can reduce features to train SVM, in order to reduce training time and to increase detection rate.

3. Problem statement

Classification is the basic task in the detection of attacks in the intrusion detection system algorithms.

In the combined support vectors and ant colony algorithm, the classification has following steps.

Input: data set (KDD99 data set)

Output: classification of data as normal and abnormal

1. collect the data set
2. perform SVM training and AC phase
3. detection the attacks by classification as normal and abnormal
4. calculate the training time , testing time and detection accuracy

In this algorithm, data has several features so it is a time taking process to train by all the features.

4. Proposed Method

We propose an entropy based algorithm to reduce training time with more accuracy in detection. For the problem of classification of data as normal and abnormal in intrusion detection system an ESCAV (entropy based support vectors and ant colony) method is proposed to reduce the parameter training time with accuracy in detection.

In ESVAC algorithm by using entropy metric we can calculate the entropy measure based on the similarity values for the feature selection and selected features are involved in training.

4.1 Support vector machines

In SVM, data is referred as the vectors and with d dimensional feature space. Each data point in feature space is belongs to either of classes A or B. And those data points are labeled as -1 and 1 by L as follows

$$L=\begin{cases}-1 \; pi \; \epsilon \; classA \\ 1 \; pi \; \epsilon \; class \; B\end{cases}$$

-1 is referred to negative points
1 is referred to positive points

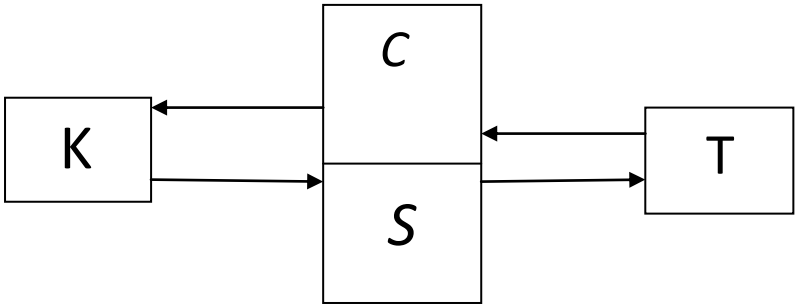
In the case of linear separable, data points are separated as positive and negative by many hyperplanes. From those hyperplanes, an exact margin is defined by calculating the sum of distance from nearest or closest negative and positive points. In case of non linear, kernel function is used to create non linear decision boundaries.

To compute, margin M have to take hyperplanes h1 and h2 which are holding closest positive and negative points.

M: N.Y-d=0 Yϵ R^d
h1: N.Y-d=1 Yϵ R^d
h2: N.Y-d=-1 Yϵ R^d
N is normal to M
d is distance from M to origin
Where M, h1, h2 are parallel

4.2 IDS based SVM

Consider the data set as K and T is the data set for training from K. In the SVM there are two parts as *c* and *s*. *c* is the classifying part and *s* is searching part of new data points from K into T for next training process.



4.3 Ant colony networks

Ant colony network has properties of self organized, robustness, flexibility and decentralization. In the paper ACN is used for the purpose of similarity measure to form clusters.

Similarity of two data points can be described as follows

$$S(x_{i,}\, x_j) = 1 - \frac{d(x_{i,}\, x_j)}{\beta} \qquad \text{eq (1)}$$

β - positive coefficient.
d (x_i, x_j) - distance function among two data points.
S ($x_{i,}$ x_j) - similarity function between two data points.

4.4 IDS based ACN

In the ACN process by adjusting β value in the equation (1), based on that value the algorithm constructs the training data set into q clusters $\{c_{1,}c_{2,}\ldots\ldots,c_q\}$, $c_i=\{x_{ij}\}$ where i=1,2,….,q , j=1,2,…..,n_i x_{ij} are the data points taken into clusters.

4.5 Entropy measure

Entropy value is calculated for each feature based on the similarity value from the similarity phase and then the mean of entropy values is calculated. so that features are selected which are greater than or equal the mean value otherwise the feature is ignored.

Entropy can be calculated as follows

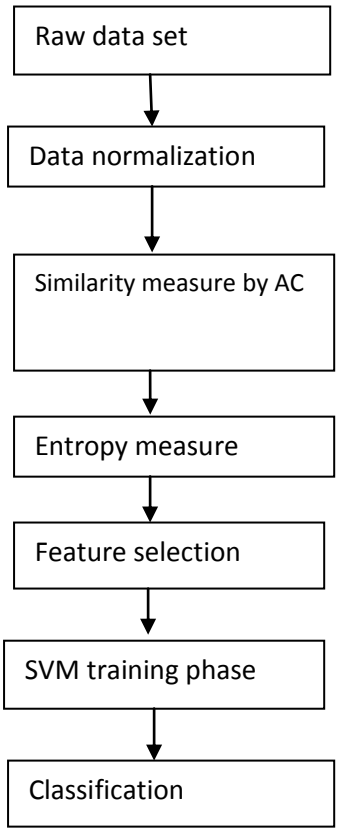
$$E= \sum_{i=1}^{n} (xi \log(xi) + (1 - xi) \log(1 - xi)$$

Where E is the entropy measure
Xi is the similarity value of xi

5. Entropy based support vectors and ant colony networks

Here we introduce a algorithm of combining SVM and ACN with entropy measure as ESVAC which classifies the data into normal and abnormal with reduced training

time.



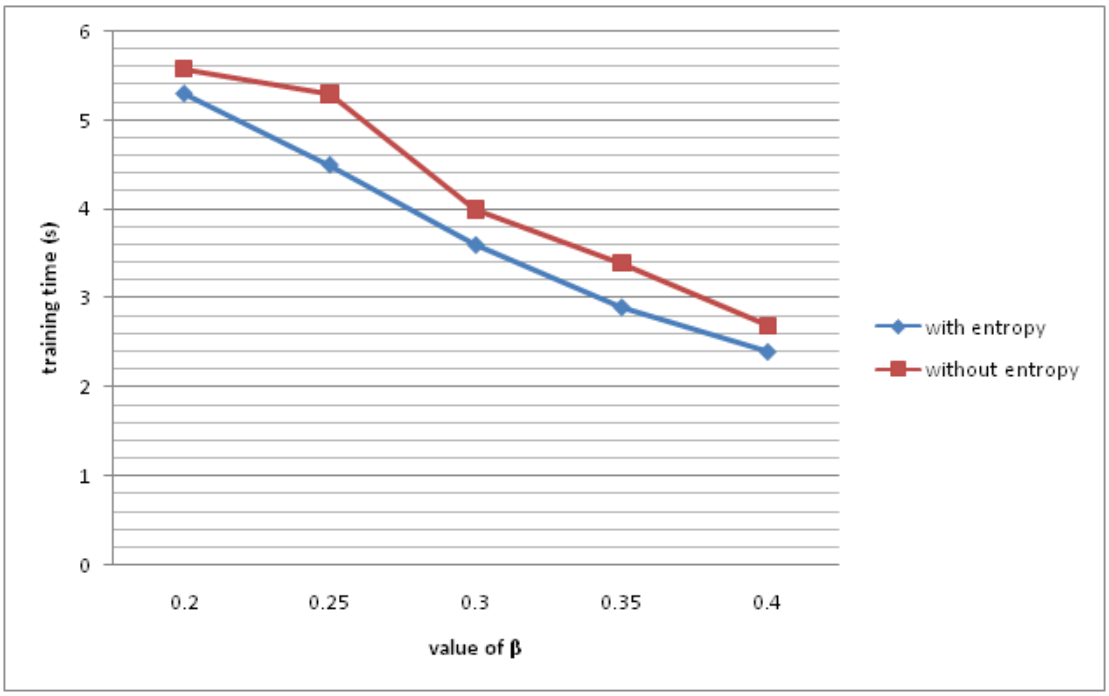
In the entire data set, randomly some data points are picked out for the SVM training phase. However those data points are separated by the hyperplanes based on the support vectors, each time by the adding new data points hyperplanes may be modified and obviously data points around the support vectors are probably to be support vectors for the next training phase. Based on those support vectors the similarity will be calculated.

Let V be the number of data points in the cluster and SIM be the similarity value treated as threshold value based on these two values clusters will formed. If the SIM is taken less than V value of cluster will be lower, SIM is taken high the V value will be more.

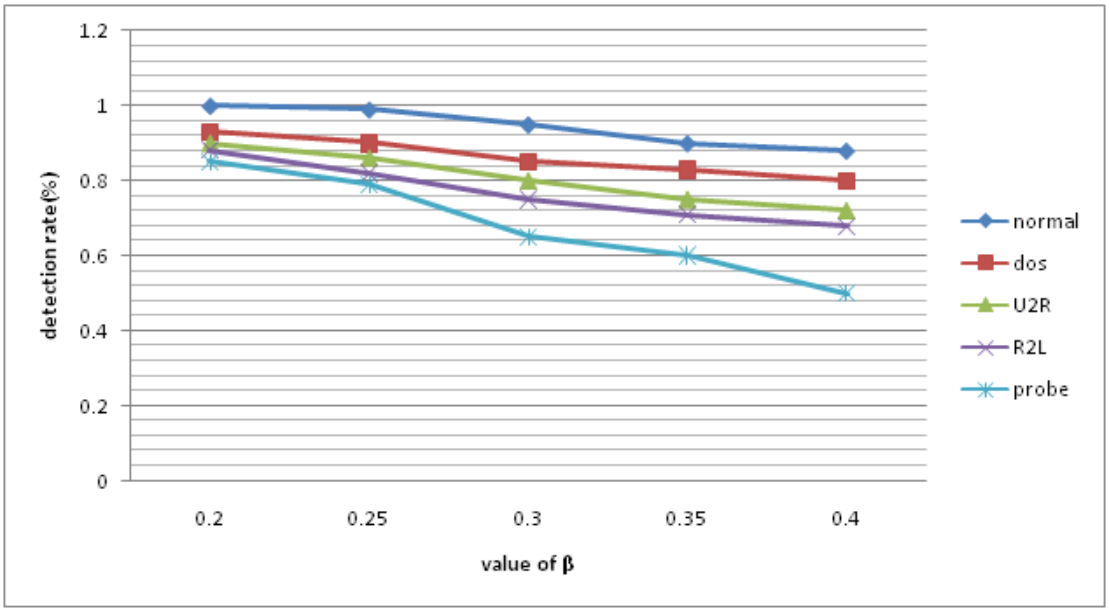
After the similarity measure phase an entropy measurement is to be taken for the feature selection phase based on the maximum similarity value of feature, entropy is to be calculated for each feature. Then mean of the total features of entropy measure is to be calculated, if the feature posses less than mean value can be ignored and which are equal and more than mean can be considered for the training phase. Where training time will be reduced for the selected features rather than entire features training time.

6. Results

We have taken the input data KDD99 data set as 10000 records and performed the algorithm by adjusting the values of β and V. Results are as follows.



(a) Training time



(b) Detection rate

The example detection of attacks as follows

Number	Attributes	Attack type
1	normal	normal
2	nmap	probe
3	back	DOS
4	neptune	DOS
5	satan	probe

7. Conclusion

In this paper, proposed a new algorithm ESVAC to classify the data by applying similarity measure and entropy measure to attain better performance in training time and detection rate to the intrusion detection. This approach aggregates both algorithms SVM and ACN where SVM is for training purpose and ACN is for similarity measure and entropy measure is calculated using similarity values of features, so that those entropy values are used for the feature selection phase. Hence numbers of features for training phase are reduced and therefore training time reduced with efficient detection rate.

8. References

[1] Yihua Lioa, V.Rao Vemuri, Using K-Nearest Neighbor classifier for intrusion detection, *science direct, computer & security,* vol.2,439-448,2002.

[2] Amir Azimi Alsti Ahaerbi, Ahmad Habibizad Navin, Hadi Bahrbegi, Mir Kamal Mirnia, Mehdi Bahrbegi, Elnaz Safarzadeh, Ali Ebrahimi, A New system for clustering and classification of IDS alerts using self organized maps, *International journal of computer science and security*, vol.4,589-597,2012.

[3] Ting Liu, Zhiwen WANG, Haijum Wang, Ke Lu, An Entrop-based method for attack detection in large scale network, *INT J comput commun*, vol.7,509-517,2012.

[4] Jayanti Yasmah Sari, Mutmainnah Muchtar, Mohammad Zarkari, Agus Zainal Arifin, similarity based Entropy on feature selection for the high dimensional data classification, *journal of computer science & information*, vol.7,104-110,2014.

[5] Wenying Feng, Qinglei Zhang, Gongzhu Hu, Jimmy Xiangji Huang, Mining network data for intrusion detection through combining SVMs with ant colony networks, *Elsevier, future generation computer systems*, vol.37, 127-140,2014.