ISSN No. 0976-5697

Volume 8, No. 3, March – April 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Network Intrusion Anomaly Detection Using Radial Basis Function Networks

Dr.R Ravinder Reddy
Department of Computer Science and Engineering
Chaitanya Bharathi Institute of Technology
Hyderabad, India

Abstract: The network intrusion detection is big threat to the current generation. In these days the usage of the Internet is part of our lives. The enormous growth of the computational intelligence makes us to use many devices, which are connected to the internet. This makes the attacker penetrates into the network to get the unauthorized access of the resources, to alter or modify them. By compromising the security mechanism and steal the valuable information. This makes the network intrusion detection needs to upgrade in every moment. The machine learning techniques enhances the detection rate by learning the new computational models. In this work we presented the Radial Basis Function (RBF) Networks are used to analyse the intrusion detection.

Keywords: Radial Basis Function (RBF) Networks, Intrusion detection, Machine learning, artificial neural network (ANN).

INTRODUCTION

Network Intrusion detection is very critical in these days, the dynamic behavior of the intruder in the network is unable identify accurately. Because of this dynamic nature the existing methodologies are not sufficient. The enormous growth of the machine learning techniques will able to solve this problem. In spite of huge work in the machine learning based intrusion detection, still there are some lapses to accelerate the intrusion detection model. The radial basis function (RBF) network is a mathematical modeling for artificial neural network. It works on the principle of radial basis function.

The conventional techniques of the intrusion models are suffer with the generalization problem. It decides the models performance. The Generalization ability of the classifier is increased by feedback neural network. Basically, the Generalization is affected by 3 factors in back propagation.

1. The size of the training set.
2. The architecture of the network.
3. Physical complexity of the problem at hand.

To overcome these problems and enhance the ability of the generalization, in this work we proposed the radial basis network for intrusion detection. This work has produces the better ability for representation of the model. An artificial neural network represents incredibly existing and powerful machine learning based techniques used to solve many real world problems. The network intrusion detection is determined by the ANN approach for enhanced detection rate.

The main motivation of using Artificial Intelligence for Intrusion Detection system is to overcome the drawbacks of conventional techniques. The major limitations of the conventional system are constant update of database with new signatures and false alarm rate. These drawbacks are overcome with the AI based techniques. The main advantages are flexibility, adaptability and learning abilities [1]. The Capability of the neural network largely depends on the learning algorithm and the network architecture used.

The intrusion detection can be treated as the classification problem. The error rate of the model is depends on the hypothesis learning technique [2]. In machine learning techniques the learning model must be generalized to reduce the error rate in the given hypothesis space. In this approach, for enhancing the generalization ability of the model, the huge datasets has considered for experimentation. But, its ability is affected by the free parameters.

The size of the dataset N,

$$N = O\left(\frac{W}{\in}\right)$$

Number of hidden units W is a free parameter, which is adjusted to get best predictive performance of the model. $\in$-fraction of permissible class error. The size of the dataset is crucial in the process of training the model, which can used to determine the best predictive model. The dynamic nature of the intrusion and the new types of threats are penetrating into the network. Learning of these novel threats and updating itself and detecting the intrusions are typical.

The remaining portion of the paper is organized as follows. In section II discussed the related work, sections III outlines the radial basis function network. Methodology is discussed in section IV, results and analysis is presented in section V. conclusion and future enhancements are presented in section VI.

RELATED WORK

A. *Intrusion detection system (IDS)*

The concepts of intrusions are identified in the seminal report of Anderson, in the evaluation of the log records in the early 80s [10]. With the exponential growth of the computers the signs of intrusions are also increased. Basically intrusion detection is divided into aspects.

1. **Misuse detection**: it is static nature, consists of fixed signature database for finding the intrusions. The main problem is needs to update the database when the new signs of intrusion.
2. **Anomaly detection:** it is dynamic nature can find the novel intrusions based on the normal profile. But the high false alarm rate is the main disadvantage of the model.

Based on the source of the detection the IDS divided into:

1. **Host based IDS**: this is at server or host to analyze the system calls.
2. **Network based IDS**: This is used analyze the network packets which are captured in the network. Used these packets to analyze for the signs of intrusions.

In the research of intrusion detection many techniques were adopted to detect intrusions like statistical, data mining, neural network and machine learning approaches. All these approaches are used to detect the accurate intrusion detection.

B. *Rough set theory*

Rough set theory (RST) is a conventional set theory proposed pawlak [6], which is used to represent the knowledge. Mainly RST is used to select appropriate feature from the dataset. It is based on the lower and upper boundary approximations. The RST reduces the dimensionality of the feature vector without affecting the performance of the model. This is works on the principle of indiscernibility, which is calculates based on the equivalence classes of the relation. In this work we used the quick QUICKREDUCT [5] feature selection algorithm based on RST for obtaining the minimal feature vector.

The QUICKREDUCT is based on the dependency function approach. It attempts to calculate a reduct, initially it starts with empty set and adds attributes, one at a time, those attributes that result in the greatest increase in the rough set dependency metric, until this produces its maximum possible value for the dataset. Quickreduct algorithm calculates the dependency of each attribute, and selects the best attributes. The dependency of attributes is used to form the patterns for identifying the intrusion. Use of dependency is good sign in the process of evaluation IDS.

C. *Datasets*

Evaluation of the model requires proper datasets, in the evaluation process of the intrusion detection there three datasets which are used to evaluate the model. In this work we consider these dataset for evaluation of the model.

1. KDDCUP99 Dataset: this is the first standard dataset for evaluation of the intrusion model, introduced in the KDD conference in the year 1999. This simulates many new intrusions. In this dataset, it contains the derived features which helps to find the new types of intrusion [11].
2. UNB ISCX dataset: it is prepared on the real network captured data. It is available from the 2012 [13].
3. HTTP CSIC dataset: this dataset is for web traffic analysis. Which contains the different web users data for the normal and intrusion analysis [11].

Table 1: The quick reduct for the IDS datasets

Reduct algorithm Dataset	Total number of conditional features	No. of features for Quick Reduct
KDDCUP99	41	24
HTTP CSIC	17	8
ISCX	20	13

These datasets are used for the model evaluation. These datasets contain both the normal and intrusive records. The number of feature contained in each dataset and the optimal feature obtained by using the quickreduct are shown in the table 1.

RADIAL BASIS FUNCTION NETWORK

The Radial Basis Function (RBF) Networks are the type of artificial neural networks. It is used to solve supervised learning problems. Radial basis function are used to increase (or decrease) the distance monotonically from the centre point. The typical radial basis function is the Gaussian function, which can be represented as.

$$g(x) = exp\left(-\frac{(x-c)^2}{r^2}\right) \qquad \text{Eq1}$$

Where r is the radius and c is the centre.

The RBF networks are introduced by Broomhead and Lowe's [8] in their seminal report in the 1988. Based on the radial function movement and the number of hidden layers the RBF Network is nonlinear. Typically used for function approximation, pattern classification, etc. It consists of two layer feed-forward structure with each hidden unit implementing radial activated function. The training involves updating centre of network for hidden neuron and output layer weights adjusted according to the desired objective.

The functionality of the RBF is divides into two steps.

1. Radial is only depends on the distance between x and center.
2. Approximating the value based on back propagation.

The most important aspect of back propagation is its computational efficiency. The Model selection is depends on the different parameters.

1. The curse of dimensionality
2. Minimizing the misclassification rate

As the dimensionality data is affecting the RBF networks training time. The misclassification is also an important criterion in the training process of the model. Needs to supply a bias to the RBF networks to reduce the classification error. RBF Networks is the linear aggregation of radial hypotheses [4]. In the neural network and RBF networks are differs in no of hidden layers, in the neural network it calculates based on the inner product and tanh while in RBF it is distance and Gaussian. RBF Networks is looks like structurally similar to the Multi Layer Perceptron (MLP). But the vice versa is not true. The back propagation is the process of making a stochastic approximation of the process. It means learning with the example. Basic form of RBF consists of 3 layers.

1. **Input layer**: which is used to supply the input data to the model

2. **Hidden layers**: these are used to approximation functions based on the centre distance. It is updated based on their distances. It is radial basis function updating. Basically it uses the basis function based the nearest neighbors.
3. **Output**: The output of the model

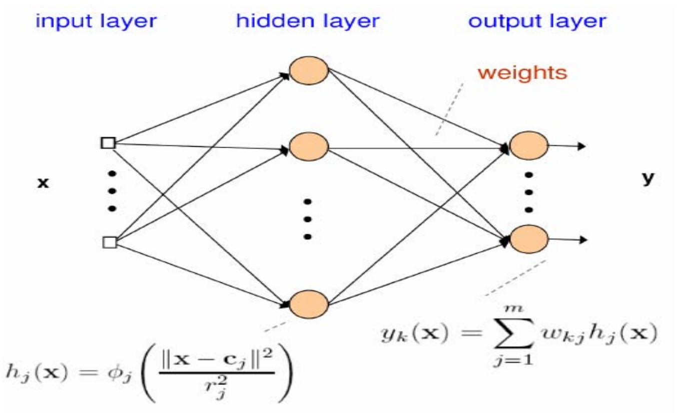
input layer
hidden layer
output layer
weights
x
y
$y_k(\mathbf{x}) = \sum_{j=1}^{m} w_{kj} h_j(\mathbf{x})$
$h_j(\mathbf{x}) = \phi_j\left(\frac{\|\mathbf{x}-\mathbf{c}_j\|^2}{r_j^2}\right)$

Figure 1: Representation of RBF Network

The general representation of the RBF network is shown in the figure 1. The hidden layer nodes are represented by the radial basis function. Input source nodes connected to the environment. Hidden layer only one layer which does the non linear transformation from input space to the hidden space. The dimension of the hidden layer is higher. Output layer supplies the response. According to the cover's theorem a pattern classification problem in high dimensional space is more likely to be linearly separable than in a low dimension. Based on this principle's many of the classifier's transform the data into the higher dimensions, to separate the data linearly. Gaussian SVM achieves large margin in infinite dimensional space. The RBF network is the linear aggregation of the radial hypothesis. Radial is the only depends on distance between x and centre. Basis function is used to combine.

The Support vector machines (SVMs), which are maximal margin classifiers. RBFs and SVMs differ in that at the output layer, a SVM employs convex optimization to find an optimal linear classifier, whereas the output weights of RBF network are typically estimated by a linear least squares algorithm, such as the LMS or recursive least squares (RLS) algorithm [7].

METHODOLOGY

The implementation of the RBF network for the intrusion model is divided into two stages. In the first stage, selection of the appropriate features from the dataset using the RST approach. Second, train the model with RBF network for evaluation of the IDS. RBF network works on the distance similarity to centers as feature transformed into higher dimensions. The combined approach of RST-RBF Networks is performed well on the intrusion model [3]. The transformed feature are evaluated based on the radial function from the centre, it is updated based the distances at the hidden layers. In the evaluation process the basis is uses the KNN approach for evaluation of the model.

Algorithm: RBF network with RST for IDS
Input: IDS dataset
Output: Accuracy of the Model

1. Preprocess the dataset
2. Apply the rough set theory based quick reduct algorithm to reduce the features.
3. Prepare the new dataset with the Derived features.
4. Train the RBF network with the data.
    a. It works with KNN approach, it calculates the center.
    b. Radial basis function will approximate the hidden layer values.
5. Evaluate the learned model.

RESULT ANALYSIS

The experiments are conducted for the network intrusion datasets. In this work we consider the three different datasets for evaluating the model. For the evaluation purpose of the model, different measures are used.

1. KDDCUP99 dataset: which
2. ISCX data set
3. The HTTP CSIC dataset.

The classifier evaluation is based on the generalization of the model. The accuracy of the model is not sufficient to judge the intrusion detection. In this aspect F-measure, precision and recall are used.

Table .1 Confusion Matrix

	Positive	Negative
Positive	Positive Identified as Positive (TP)	Negative Identified as Positive (FP)
Negative	Positive Identified as Negative (FN)	Negative Identified as Negative (TN)

As shown in table 1 the classifier produce the result either positive or negative samples. These values have the significant meaning in the evaluation of the intrusion model. The significance can be measured in the form of the accuracy, precision, recall and f-measure. The equations from 2-5 are used to calculate these measures.

$$\text{Accuracy} = \frac{Tp + Tn}{Tp + Tn + Fp + Fn} \qquad Eq2$$

$$\text{Recall} = \frac{\text{Tp}}{\text{Tp} + \text{Fn}} \qquad \text{Eq3}$$

$$\text{Precision} = \frac{Tp}{Tp+Fp} \qquad \text{Eq4}$$

$$\text{F} - \text{measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \qquad \text{Eq5}$$

Table: The RBF Network measures

	Accuracy	Precision	Recall	F-Measure
KDD	93.05	0.931	0.931	0.931
ISCX	99.96	1	1	1
HTTP	96.39	0.964	0.964	0.964

VI. CONLUSION

The radial basis function network with the rough set theory combinely addresses the intrusion problem. The RST approach has reduced the dimension for decreasing the training time of the RBF network [9]. The trained RBF network is used to detect the new signs of intrusions. The model is tested for the three datasets and obtains the good results. The RBF network model adjusted the center based on the radial basis function is the crucial for accurately evaluation of the model. As the future research the model has to test directly on the real network data with unturned parameters. The RBF network has to design for the detection of the raw network data.

VII. ACKNOWLEDGMENT

I am very thankful to Dr.Y Ramadevi, Head of the department for her support and encouragement in writing this paper, I also than Dr.B Chennakesava Rao principal, CBIT and the management for allowing me to do this work.

VIII. REFERENCES

[1] Yichun, Peng, Niu Yi, and Hu Qiwei. “Research on Intrusion Detection System Based on IRBF”, Computational Intelligence and Security (CIS), 2012 Eighth International Conference on. IEEE, 2012.

[2] Huang, D. & Chow, T.W.S. “Improving the effectiveness of RBF classifier based on a hybrid cost function” Neural Computing & Applications 16(4), pp: 395-405, 2007.

[3] Ding, S., Ma, G. & Shi, Z. “A Rough RBF Neural Network Based on Weighted Regularized Extreme Learning Machine”, Neural Processing Letter, 40(3): 245-260.2014.

[4] H Yu, PD Reiner, T Xie, T Bartczak, BM Wilamowski, An incremental design of radial basis function networks. IEEE Trans Neural Netw and Learning Systems **2**(10), 1793–1803 (2014).

[5] Chouchoulas A, Shen Q. “Rough Set-Aided Keyword Reduction for Text Categorization”, Applied Artificial Intelligence 15(9), 843–873 2001.

[6] Z. Pawlak, “Rough Sets”, Internation Journal of Computer and Information Sciences11, pp.341-356.1982.

[7] Wen, H., Xie, W., Pei, J. et al. “An incremental learning algorithm for the hybrid RBF-BP network classifier” EURASIP Journal on Advances in Signal Processing. (2016).

[8] D.S. Broomhead and D. Lowe. “Multivariate functional interpolation and adaptive networks”. Complex Systems.2:321-355, 1988.

[9] Chiang J-H, Ho S-H “A combination of rough-based feature selection and RBF neural network for classification using gene expression data”. IEEE Transaction on Nano bioscience 7(1):91–99. 2008

[10] J. P. Anderson, “Computer Security Threat Monitoring and Surveillance”. Technical Report, Fort Washington, PA (1980).

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

[12] http:// iec.csic.es/dataset/

[13] www.unb.ca/cic/research/datasets/ids.html