International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org ||Volume 7 Issue 2|| February 2018 || PP. 51-57

Fuzzy Rough Set Based Network Intrusion Detection with Wrapper Subset Evaluator

Ashalata Panigrahi[1] Manas Ranjan Patra[2]
Department of Computer Science, Berhampur University, Berhampur, India
Corresponding Author: Ashalata Panigrahi

ABSTRACT :*With the growing penetration of Internet, concerns of cyber security is looming high world-wide. In one hand there is enthusiasm to develop internet based applications for wider access, on the other hand seriousefforts are being made to evolve techniques to deal with the menace of cyber-attacks. One of the research areasfocusses on developing techniques to analyze available network intrusion data and build intrusion detection models which can be used effectively to raise alerts whenever suspicious user behavior is observed in a network. In this work, we propose a soft computing based technique to construct an intrusion detection model using five classifiers, namely, Fuzzy Nearest Neighbour (FNN), Fuzzy-Rough Nearest Neighbour (FRNN), Fuzzy-Rough Ownership Nearest Neighbour (FRONN), Vaguely Quantified Nearest Neighbour (VQNN), and Ordered Weighted Average Nearest Neighbour (OWANN). Further, the most relevant features in the input data have been extracted through a preprocessing stage using the wrapper subset evaluator. Finally, the performance of the model has been evaluated on the NSL-KDD intrusion dataset in terms of accuracy, precision, detection rate, and false alarm rate*

KEYWORDS:*Cyber Security, Feature Reduction, Fuzzy Rough Classifiers, Network Intrusion detection, Wrapper Subset Evaluator.*

Date of Submission: 25-01-2018 Date of acceptance: 17-02-2018

I. Introduction

Overwhelming expansion in network based information processing throughout the world has made computer networks more vulnerable to security threats. There has been increase in number of network intrusions leaving organizations in jeopardy. Traditional intrusion prevention techniques such as data encryption, firewalls, and access control mechanisms have their limitations to completely protect computer networksfrom attacks and malwares which are becoming more and more sophisticated. Therefore, Intrusion Detection Systems (IDS) have become an indispensable component of security infrastructure to detect such threats before they inflict widespread damage.Anomaly based network intrusion detection techniques have the capacity to detect new types of intrusionsand thereby protect target systems and networks from malicious activities. Soft computing is an innovative approach to build a computationally intelligent system which parallels the extraordinary ability of the human mind to reason and learn in an environment of uncertainty and imprecision. [1]. Soft computing consists of several computing paradigms, including genetic algorithms, neural networks, fuzzy sets, rough sets, approximate reasoning, simulated annealing etc. Different soft computing approaches have been tried in the area of intrusion detection.

Ektefa et al. [2] have compared the performance of C4.5 algorithm with Support Vector Machine in detecting intrusions and the results revealed thatC4.5 performed better than SVM in terms of intrusion detection and false alarm rate.Tong et al. [3] have proposed a hybrid IDS based RBF/Elman neural network wherein the RBF neural network is employed as a real time pattern classifier while Elman neural network is employed to restore the memory of past event. Ritu et al. [4] have proposed a combination of SOM and Radial Basis Function (RBF) network and have shown that the combined system offers better results than IDS based on RBF network alone.Sung et al.[5] have proposed an approach for IDS with the use of Rank based featureselection and have shown that Support Vector Machines (SVMs) perform much better Artificial Neural Networks (ANNs) in terms of speed of training, scale and accuracy. Tie-Jun [6] used the Back-Propagation network (BPN) with Genetic algorithm to improve the performance of back propagation technique. He could achieve an overall detection accuracy rate of 91.61% with false alarm rate of 7.35%.Ibrahim and Laheeb[7] focused on self-organization map (SOM) model to compare the detection rate between two data sets: KDD99 and NSL-KDD. Detection rate of SOM with KDD99 is 92.37% while it is 75.49% for NSL-KDD data.Atefi et al. [8] have proposed a hybrid model using SVM and GA (Genetic Algorithm). They compared true negative and false positive rates between SVM and hybrid model SVM+GA. Hybrid model recorded low false negative rate of 0.5013 % and high true negative value of 98.2194 %. The result shows high accuracy of the hybrid model

Following the trend of research so far in the area of network intrusion detection, in this paper, we propose a hybrid intrusion detection approach which combines techniques based on both fuzzy and rough set theories to classify network data as normal and anomalous. Further, in order to improve upon the performance of the model attribute reduction has been carried out through a preprocessing stage using the wrapper approach.

II. Methodology :

2.1 Hybridization of Fuzzy-Rough Set Theory

Fuzzy set and Rough set are two complementary characteristics of imperfect data and knowledge. Fuzzy-Rough set theory [9] is a hybridization of rough sets [10] and fuzzy sets [11] which is capable of dealing with imprecision and uncertainty in discrete and real-valued noisy data or mixture of both without domain information. Fuzzy rough set theory was designed to model imperfect knowledge, i.e., the instances are similar to each other but belonging to different classes and it is more suitable for partially exposed and unbalanced dataset. There are several techniques based on both fuzzy and rough set based theories. In the following sections, we discuss five such techniques which we have experimented on the NSL-KDD data set.

2.2 Fuzzy Nearest Neighbor (FNN) Classification

Fuzzy K-Nearest Neighbor (FNN) Algorithm

FNN (X, C_D, y, K)
Input: X: the training data set;
C_D: the set of decision classes;
y: the objects to be classified;
K: the number of nearest neighbours
begin
$N \leftarrow$ get Nearest Neighbors (y, K)
foreach $C \in C_D$ do
$C'(y) = \sum_{x \in N} R(x,y)C(x)$
end
end
Output: arg max (C' (y))

2.3 Fuzzy-Rough Nearest Neighbor Algorithm (FRNN)

In FRNN algorithm [13] the nearest neighbors are used to construct the fuzzy lower and upper approximations of decision classes, and test instances are classified based on their membership to these approximations. Fuzzy Rough NN is more suitable for partially exposed and unbalanced dataset.

Fuzzy Rough Nearest Neighbor Algorithm

FRNN (X, C, y)
X, the training data set; C_D, the set of decision classes;
y, the objects to be classified;
begin
N ← get Nearest Neighbors (y, K)
τ ← 0, Class ← Ø
foreach $C \in C_D$ do
if $((R\downarrow C)(y) + (R\uparrow C)(y)) / 2 \geq \tau$ then
Class ← C
τ ← $((R\downarrow C)(y) + (R\uparrow C)(y)) / 2$
end
end
end
Output: Class

2.4. Fuzzy Ownership Algorithm

Fuzzy ownership [14] is an attempt to handle both "fuzzy uncertainty" and "rough uncertainty

```
FRONN( X, A, C_D, y )

X, the training data set; A the set of conditional features;
C_D the set of decision classes; y the object to be classified
begin
        for each a Є A do
        K_a = |X| / (2 Σ_{x∈X} ||a(y) − a(x)||2/(m−1))
        end
        N ← | x |
        for each C Є C_D do τ_c(y) = 0
        for each x Є N do
                d = Σ_{a∈A} Ka(a(y) − a(x))²
        for each C Є C_D do
                τ_c(y)+ = C(x).exp(− d^{1/(m−1)} ) / | N |
        end
        end
        end
end
Output: arg max τ_c(y)
        C Є C_D
```

2.5 *Vaguely Quantified Nearest Neighbors (VQNN)*

VQNN [13] depends only on the summation of the similarities of each class. It uses the linguistic quantifiers "most" and "some". Given a couple (Q_u, Q_l) of fuzzy quantifiers that represent "most" and "some" respectively, the lower and upper approximation of C. The resulting VQNN algorithm assigns a class to a target instance y as follows:

- Determine NN, the K nearest neighbors of y.
- Assign y to the class C for which

$(R\downarrow^{Qu} C)(y) + (R\uparrow^{Ql} C)(y)$ is maximal

The upper and lower approximation of Vaguely Quantified rough sets are defined as

$$((R\downarrow^{Qu} C)(y)) = Q_u\left(\frac{\sum_{x\in X} \min (Rx,y),C(x)}{\sum_{x\in X} R(x,y)}\right) \ldots. (1)$$

$$((R\uparrow^{Ql} C)(y)) = Q_l\left(\frac{\sum_{x\in X} \min (Rx,y),C(x)}{\sum_{x\in X} R(x,y)}\right) \ldots. (2)$$

The fuzzy quantifiers Q_u, Q_l are increasing $[0,1] \rightarrow [0,1]$ mapping such that
$Q_u(1) = Q_l(1)=1$ and $Q_u(0) = Q_l(0)=0$

An element y belongs to the lower approximation of X if most of the elements related to y are included in X. An element y belongs to the upper approximation of X if some of the elements related to y are included in X. This classifier based on rough set theory is capable of handling noisy data.

2.6 *Ordered Weighted Average Nearest Neighbors*

In Fuzzy-rough hybridization based on ordered weighted average (OWA) operator [14],themembership degrees to be approximated are computed by an aggregation process. The OWA based approach has the following benefits:

i. It is monotonous with respect to the fuzzy indiscernibility relation.
ii. The traditional fuzzy-rough approximation can be recovered by a particular choice of the OWA weight vectors.

III. The Proposed Model

3.1 Model Description

The proposed model aims at building an intrusion detection system with low false alarm rate and high detection rate by applying hybridization of fuzzy rough techniques viz. Fuzzy Nearest Neighbour, Fuzzy-Rough Nearest Neighbour, Fuzzy-Rough Ownership NN,Vaguely Quantified Nearest Neighbours, and Ordered Weighted Average Nearest Neighbours.Further, the Wrapper subset evaluator is applied on the dataset to select the most relevant features as described below. The purpose is to improve the performance of the intrusion detection model in terms of time and accuracy.

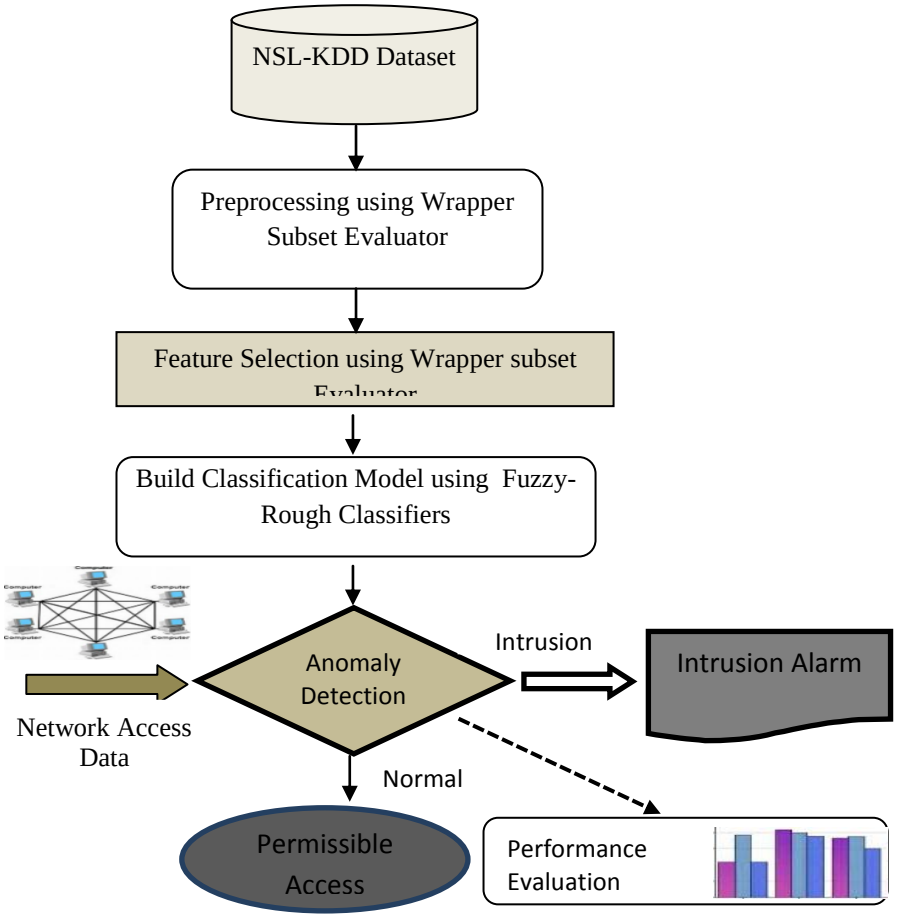

Fig. 1. Proposed Model

3.2 Wrapper Subset Evaluator for Feature Selection

A major challenge while constructing a high performance intrusion detection system is to deal with data that contain large number of features, all of which may not be relevant to a particular processing requirement. The NSL-KDD dataset which is used for our experiments has many irrelevant features. Most of the existing detection models use all the 41 features to evaluate the performance of a model. This makes the detection process consume more time and degrades the performance of the IDS. In this paper, we attempt to identify the most suitable subset of features by employing a wrapper based feature selection algorithm. In the wrapper approach, search for a good subset is conducted using an induction technique.A search requires a state space, an initial state, a termination condition and a search engine. The goal of the search is to find the state with the highest evaluation using a heuristic function. The wrapper approach conducts a search in the space of possible parameters and the accuracy of the induced classifiers is estimated using accuracy estimation techniques.

IV. Experimental Setup

4.1 NSL-KDD Dataset

The NSL- KDD dataset [12] is a reduced version of the original KDD CUP 99 dataset. NSL-KDD dataset consists of same features as KDD CUP 99 train dataset but has some advantages over the original KDD'99 dataset. The data set consists of 41 feature attributes for each connection record plus one class label. Out of 41 attributes 38 are numeric and 3 are symbolic. Symbolic attributes are protocol type, service, and flag. The total number of records in the data set is 125973 after removal of the redundant data; out of which 67343 are normal and 58630 are attacks. The dataset contains 24 different attack types which can be classified into four main categories, namely, Denial of Service (DOS), Remote to Local (R2L), User to Root (U2R), and Probing.

4.2 Cross Validation

Cross validation calculates the accuracy of the model by separating the data into two different populations, a training set and a testing set. In n-fold cross-validation the dataset is randomly partitioned into n mutually exclusive folds, In 10-fold cross validation, a given dataset is partitioned into 10 subsets, of these 10 subsets 9 subsets are used to perform a training fold and a single subset is retained as the testing data. This cross-validation process is then repeated 10 times (the number of folds) by rotating the folds. The 10 sets of results are then aggregated via averaging to produce a single model estimation. The advantage of 10-fold cross validation over random sub-sampling is that all objects are used for both training and testing, and each object is used for testing only once per fold.

4.3 Confusion Matrix

An intrusion detection model can be evaluated by its ability to make accurate prediction of attacks. Intrusion detection systems mainly discriminate between two classes, attack class and normal class. The confusion matrix reports the number of False Positives (FP), False Negatives (FN), True Positives (TP), and True Negatives (TN). Based on these values the following performance measurements can be made:

$$\text{Accuracy} = \frac{TP+TN}{TN+TP+FN+FP}$$

$$\text{Detection Rate or Recall} = \frac{TP}{TP+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{False Alarm Rate} = \frac{FP}{TN+FP}$$

5. Results and Discussions

Five different fuzzy rough classifiers, namely, Fuzzy Nearest Neighbour, Fuzzy-Rough Nearest Neighbour, Fuzzy-Rough Ownership NN, Vaguely Quantified Nearest Neighbours, and Ordered Weighted Average Nearest Neighbours with wrapper subset evaluator were applied on the NSL-KDD intrusion dataset. The performance of different classifiers are evaluated on the basis of accuracy, detection rate, precision, and false alarm rate.

Table I: Comparison of Fuzzy-Rough classifiers with Wrapper Approach feature subset selection

Classifier Techniques	Evaluation Criteria			
	Accuracy in %	Recall or Detection Rate in %	Precision in %	False Alarm Rate in %
Fuzzy NN	96.2841	97.2966	94.8521	4.5973
Fuzzy Rough NN	97.513	95.139	**99.4952**	**0.4202**
Fuzzy OwnershipNN	**98.8132**	98.3336	99.1283	0.7529
VQNN	98.664	98.0795	99.0407	0.8271
OWANN	98.6584	98.1221	98.6584	0.8746

It is observed that Fuzzy ownership nearest neighbour classification technique yields better accuracy than other classification techniques. Fuzzy rough NN gives low false alarm rate(see Table I). The performance measures such as accuracy, detection rate, precision, and false alarm rate can be seen in figures 2, 3, 4, and 5.

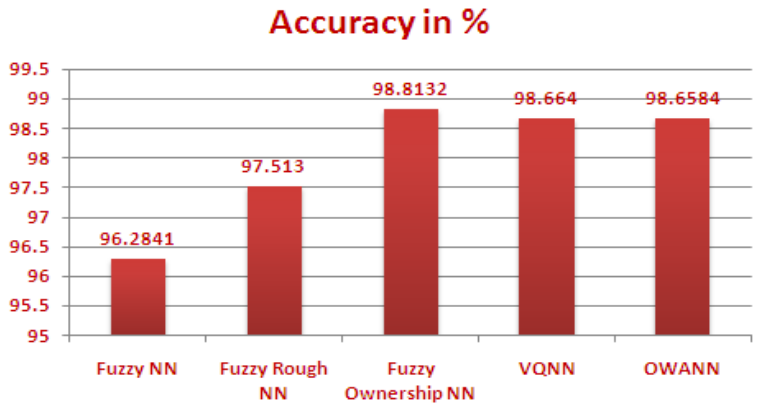

Fig. 2. Comparison of accuracy among the classifiers

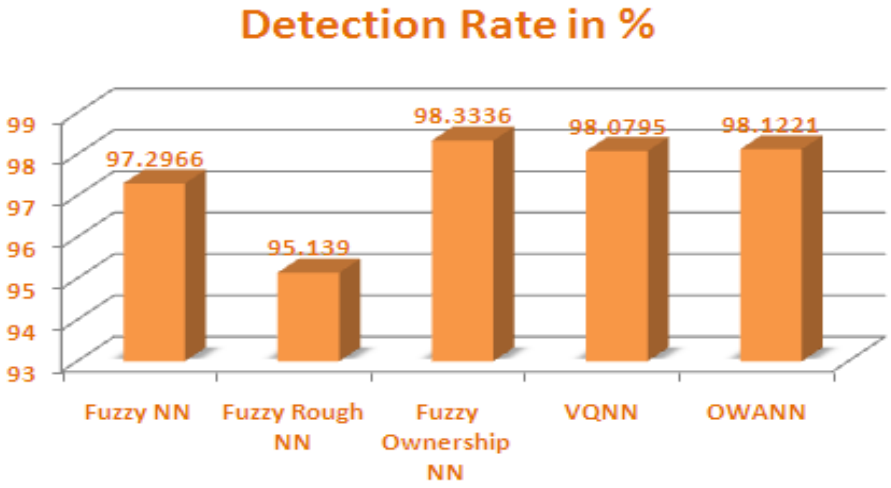

Fig. 3. Comparison of detection rate among th classifiers

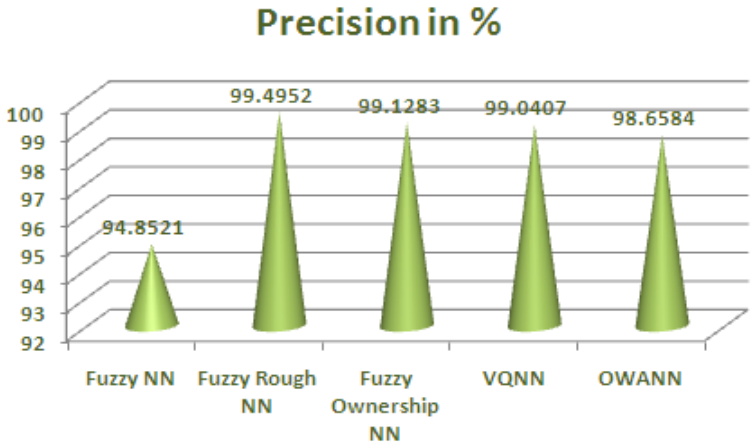

Fig. 4. Comparison of precision among the classifiers

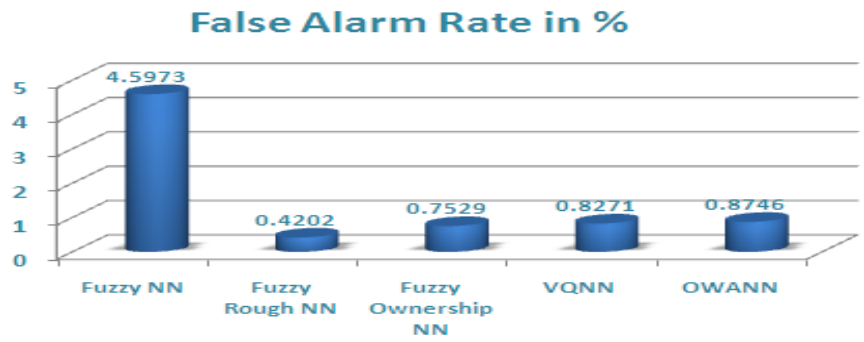

Fig. 5. Comparison of false alarm rates among the classifiers

References

[1]. L.A.Zadeh, Role of soft computing and fuzzy logicin the conception, design and development of information / intelligent systems, Lecture Notes in Computer Science , 1998, 1-9.
[2]. MohammatrezaEktefa, Sara Memar, Fatimah Sidi, Lilly SurianiAffendey,"Intrusion detection using Data Mining Techniques",Proceedings of IEEE International Conference on InformationRetrieval & Knowledge Management, Exploring Invisible World,CAMP' 10, 2010, 200-203.
[3]. X.Tong, Z. Wang and H. Yu, *A research using hybrid RBF/Elman neural network for intrusion detection system secure model*, Computer Physics Communications, *180 (10)* ,2009, 1795-1801.
[4]. R.S.Ritu, , N. Gupta and S.Kumar , *To reduce the false alarm in intrusion detection system using Self Organizing map*. International Journal of soft computing and Engineering, 2011, 27-32.
[5]. Sung, and S.Mukkamala. *The feature selection and intrusion detection problems*, Advances in Computer Science-ASIAN 2004, Higher-Level Decision Making, 2005, 3192-3193.
[6]. Z.Z.Tie-Jun, The research of intrusion detection based on genetic neural network. Proceedings of the 2008 International Conference on Wavelet Analysis and Pattern Recognition, IEEE Xplore Press, Hong Kong, 2008, 276-281.
[7]. Ibrahim, M. Laheeb, *A comparison study for intrusion (KDD99, NSL-KDD) based on self- organization map (SOM) artificial database neural network*, Journal of Engineering Science and Technology, 2013, pp. 107 – 119.
[8]. A.Atefi, K. Atefi, K., A.Y.Dak, S. Yahya, A hybrid intrusion detection system based on different machine learning algorithms. In: Proceedings of the 4th International Conference on Computing and Informatics, ICOCI 2013, Sarawak, Malaysia, 2013, 312–320.
[9]. D.Dubois, H.Prade,Putting rough sets and fuzzy sets together. In: Huang S (ed) Intelligent Decision Support, Springer, Netherlands, 1992, 203-232
[10]. Z.Pawlak *Rough sets: Theoretical Aspects of Reasoning About Data.* Kluwer Academic Publishing. Springer, Netherlands, 1991.
[11]. L.Zadeh. Fuzzy sets. Information and Control. doi: 10.1016/S0019-9958(65)90241-X, 1965. 338-353.
[12]. J.M.Killer, M.R.Gray, J.A.Givens, A Fuzzy K-Nearest Neighbour Algorithm. Systems Man and Cybernet. doi: 10.1109/TSMC.1985, 580-585
[13]. R.Jesen, C.Cornelis, , *A New Approach to Fuzzy-Rough Nearest Neighbour Classification*. Rough sets and current trends of computing, . doi: 10.1007/978-3-540-88425-5_32, 2008, 310-319.
[14]. R.R.Yager, On ordered weighted averaging aggregation operators in multicriteria decision making, Systems, Man and Cybernetics. doi: 10.1109/21.87068, 1988, 183-190.
[15]. M.Tavallaee, E.Bagheri, W.Lu, A.Ghorbani, A detailed analysis of the KDD CUP 99 data set. Proceedings of the 2009 IEEE Symposium on Computational Intelligence in Security and Defence Applications doi: 10.1109/CISDA.2009.5356528, 2009, 1-6.

International Journal of Engineering Science Invention (IJESI) is UGC approved Journal with Sl. No. 3822, Journal no. 43302.

Ashalata Panigrahi "Fuzzy Rough Set Based Network Intrusion Detection with Wrapper Subset Evaluator" International Journal of Engineering Science Invention (IJESI), vol. 07, no. 02, 2018, pp. 51–57.