VOL. 10, NO. 9, MAY 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

ENHANCING QUICK REDUCT ALGORITHM FOR UNSUPERVISED BASED NETWORK INTRUSION DETECTION

V. R. Saraswathy, N. Kasthuri and K. Kavitha
Department of Electronics and Communication Engineering, Kongu Engineering College, Erode, Tamil Nadu, India
E-Mail: vrsaraswathy@kongu.ac.in

ABSTRACT

Network intrusion detection has been identified as one of the most challenging needs of the network security community in recent years. Intrusion detection systems (IDS) can analyze a large amount of data in a reasonable time to detect the attacks. Feature selection is necessary to reduce the time consumption and memory wastage. The dataset may be imprecise, incomplete or uncertain. Rough sets deals with vagueness and uncertainty. Rough set theory (RST) is used as a selection tool to find data dependencies and reduce the number of attributes which are redundant in a dataset. Particle swarm optimization (PSO) is known to effectively solve large-scale nonlinear optimization problems. An unsupervised hybrid feature selection based on PSO and RST for high dimensional network dataset is proposed. Feature selection algorithm namely PSO-quick reduct is applied for the different dimensions of network datasets. The simulation results for the unsupervised learning show that hybridization of PSO with rough set algorithm selects features more effectively than rough set algorithm without hybridization of PSO.

Keywords: network intrusion detection, unsupervised, rough set, quick reduct, particle swarm optimization.

1. INTRODUCTION

a) Network intrusion detection

Intrusion Detection [1, 2] is a software application that continuously monitors the network for policy violations and produces reports to a management station. Network intrusion detection [1] has been identified as one of the most challenging needs of the network security community in recent years. This is because of the inflated number of users and the amount of data exchanged which makes it difficult to distinguish the normal data connections from others that contain attacks. This requires the development of intrusion detection systems (IDS) that can analyze large volume of data in a moderate time in order to take necessary actions against the attacks or violations. IDS [1, 2, 3] can be classified in to two types. (i)Anomaly (behaviour), (ii) Misuse(knowledge).

In this, misuse or signature contains descriptions which are matched against the stream of data looking for an attack. Once an attack is found, or abnormal behaviour is sensed, the alert can be sent to the administrator. An example of an NIDS would be placing it on the subnet where firewalls are situated in order to find if someone is trying to break into the firewall.

b) Feature selection

In recent years, most of the data contains large amount of information in the form of rows and columns. Rows are called as instances and columns are called as features or attributes [15]. Feature selection [14] or attribute selection is known for selecting a subset of relevant features and removes irrelevant and redundant ones which have no useful information. It is an attractive part of data analysis. It increases the efficiency.

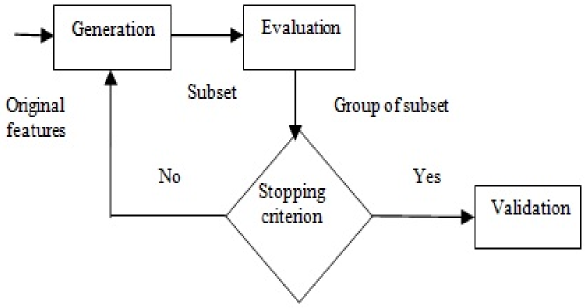


Figure-1. Framework of feature selection.

The different methods of feature selection are

- Filter
- Wrapper and
- Embedded approaches

Rough set theory is one of the effective approaches used for the process of feature selection or feature reduction. It uses the concept of approximations.

c) Rough set theory

Rough Set Theory [12] can be used as a mathematical tool for analyzing the imperfect data. It can be very much useful to find reasoning about knowledge of objects. It is needed in variety of applications like data mining, knowledge discovery etc. The main advantage is that it doesn’t require additional information. It has two basic assumptions.

- Objects are represented by the values of features.
- Objects having same information are considered as indiscernible.

The rough set [13] is the approximation of a vague concept (set) with a pair of definite concepts, namely lower and upper approximations. The basic concepts are (i) approximation space, (ii) information system (iii) decision table.

(i)Approximation space: a pair (U, R) where U is a finite set (Universe) and R is an equivalence relation described on U.
(ii)Information system: a pair S = (U, A) where A is a set of attributes.
(iii) Decision table: S = (U, A=C $\cup \{d\}$) where C are the conditional attributes, d is the conditional attribute.

d) Particle swarm optimization

Particle Swarm Optimization [10] is an evolutionary computation technique developed by Eberhart and Kennedy in 1995 which was based on the social behaviour of bird flocks. It is a population-based search procedure where the particles are grouped into a swarm. It is called population f particles. A candidate solution to the optimization problem can be indicated by each particle. In PSO [8],[9] each particle fly through the multidimensional search space and its position can be adjusted regarding its own experience and that of the neighbouring particles. A particle uses best position faced by itself (pbest) and the best position faced by its neighbours. Each particle's performance is measured with respect to a fitness function which is related to the problem being solved. At each iteration, the velocity and position of the individual particles are stochastically adjusted according to equation 1 and 2.

By considering pbest, gbest and the velocity of each particle the update rule for their position is as the following equations:

$$V_{t+1} = w_t + C_1 * rand() * (pbestx_t) + C_2 * rand() * (gbest - x_t) \quad (1)$$

$$x_{t+1} = x_t + V_{t+1} \quad (2)$$

where w is inertia weight which provides the effect of previous velocity vector (V_t) on the new vector, C_1 and C_2 are acceleration constants and rand () is a random function in the range [0,1] and x_t is current position of the particle.

e) Unsupervised learning

Unsupervised learning algorithms operate on unlabelled examples, i.e., desired output is unknown for the given input. Here the main goal is to find structure in the data, not to map inputs to outputs. It does not focus on predetermined attributes. It finds relation among data.

2. UNSUPERVISED QUICK REDUCT ALGORITHM

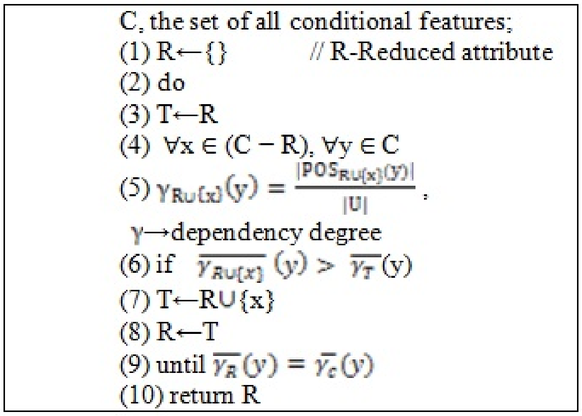
Quick Reduct (QR) [6, 7] algorithm calculates dependency degree based on POS(positive region). The QR algorithm attempts to calculate a reduct without completely creating all possible subsets of attributes.. It begins with a null set (R) and adds subset 'x' with those features that has the outcome of significant growth in the rough set dependency metric using equation (3), until this produces its maximum possible value 1.

a) Dependency degree for unsupervised data

$$\gamma_{R\cup\{x\}}(y) = \frac{|POS_{R\cup\{x\}}(y)|}{|U|} \quad (3)$$

where y - all the conditional attributes of the dataset.
R-selected subset, initially it is an empty set

b) Pseudocode for US-QR



```
C, the set of all conditional features;
(1) $R \leftarrow \{\}$          // R-Reduced attribute
(2) do
(3) $T \leftarrow R$
(4) $\forall x \in (C - R), \forall y \in C$
(5) $\gamma_{R\cup\{x\}}(y) = \frac{|POS_{R\cup\{x\}}(y)|}{|U|}$ ;
 $\gamma \rightarrow$ dependency degree
(6) if $\overline{\gamma_{R\cup\{x\}}}(y) > \overline{\gamma_T}(y)$
(7) $T \leftarrow R \cup \{x\}$
(8) $R \leftarrow T$
(9) until $\overline{\gamma_R}(y) = \overline{\gamma_c}(y)$
(10) return R
```

c) Explanation with worked example

Table-1. Example data set.

x∈ U	a	b	c	d
1	1	1	1	1
2	1	1	1	2
3	2	3	1	1
4	3	2	1	1
5	3	3	2	1
6	3	1	2	2
7	2	3	2	2
8	1	2	2	1
9	1	3	2	1
10	3	2	1	1
11	1	3	2	2
12	2	2	1	2
13	3	3	2	1
14	3	2	3	2

The given example [16] has four attributes or features. It is unsupervised data since it has no decision attribute.

$$\gamma_a(a) = \frac{|POS_a(a)|}{|U|} = \frac{|\{1,2,3,4,5,6,7,8,9,10,11,12,13,14\}|}{|\{1,2,3,4,5,6,7,8,9,10,11,12,13,14\}|} = \mathbf{1} \quad (4)$$

$$\gamma_a(b) = \frac{|POS_a(b)|}{|U|} = \frac{|\{\}|}{|\{1,2,3,4,5,6,7,8,9,10,11,12,13,14\}|} = \mathbf{0} \quad (5)$$

$$\gamma_a(c) = \frac{|POS_a(c)|}{|U|} = \frac{|\{\}|}{|\{1,2,3,4,5,6,7,8,9,10,11,12,13,14\}|} = \mathbf{0} \quad (6)$$

$$\gamma_a(d) = \frac{|POS_a(d)|}{|U|} = \frac{|\{\}|}{|\{1,2,3,4,5,6,7,8,9,10,11,12,13,14\}|} = 0 \tag{7}$$

$$\gamma_a(y) = (\gamma_a(a) + \gamma_a(b) + \gamma_a(c) + \gamma_a(d))/4 \tag{8}$$

Similarly for $\gamma_b(y) = 0.25, \gamma_c(y) = 0.30357$

$$\gamma_a(y) = o.25 \tag{9}$$

$$\gamma_d(y) = 0.25. \tag{10}$$

Among the four dependency degrees, 'c' has the maximum value, hence the combination of 'c' that is {ac},{bc},{cd} are taken as the subsets. Now $\gamma_{ac}(y) = 0.67857$, $\gamma_{bc}(y) = 0.6785$, $\gamma_{cd}(y) = 0.53571$. From these the maximum dependency is taken and its goes on continuously until it reaches 1. For this example data, all the four features are selected to reach the dependency degree of 1.

3. PSO BASED QUICK REDUCT ALGORITHM

Unsupervised PSO based Quick Reduct Algorithm [4, 5, 6] computes a reduct set without generating all possible subsets. It starts with an empty set and it adds one at a time, in turn. A population of particles is created with random positions and velocities on S dimensions. That is each particle's position is represented as binary bits of length N, where N is the total number of features or rows. Therefore, each particle's position is an attribute subset. Fitness function for each particle is evaluated.

For the example data given in Table-1, there are four features. Hence population of four particles is created. It is shown in Table-2.

Table-2. Population of particles.

a	b	c	d
1	0	0	1
0	0	1	1
1	1	0	0
0	1	1	0

Table-2 shows randomly generated position of the particles .In this population of particles, the selected subsets are 1.{ad}, 2.{cd}, 3.{ab}, 4.{bc}. For all these particles, fitness function can be calculated using the equation 3. Position having the best fitness value is taken as the gbest. Particles individual position is pbest.

If any one of the particle has the fitness value '1', then the algorithm stops here with reduced features.

Otherwise, the position and velocity can be updated using the equation 1 and 2.The fitness is calculated for new position. If it is greater than the previous fitness of the particle, then its position is taken as the pbest. The process is repeated until it reaches maximum value 1. Finally, the reduced set is obtained using number of '1's in the gbest.

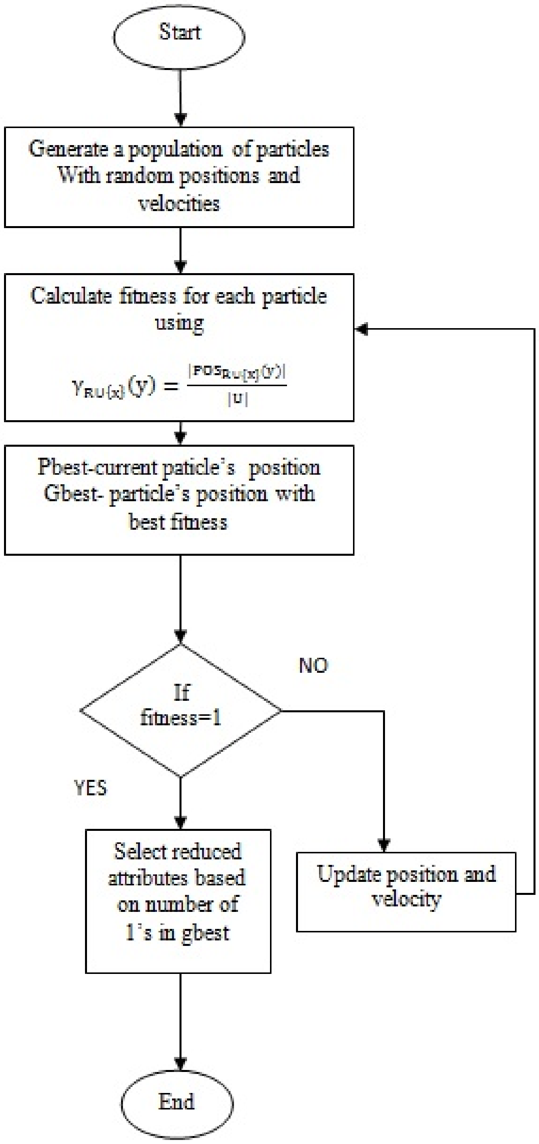

Start
Generate a population of particles With random positions and velocities
Calculate fitness for each particle using
$\gamma_{RU(x)}(y) = \frac{|POS_{RU(x)}(y)|}{|U|}$
Pbest-current paticle's position Gbest- particle's position with best fitness
If fitness=1
NO
YES
Select reduced attributes based on number of 1's in gbest
Update position and velocity
End

Figure-2. Flowchart for US-PSO-QR algorithm.

4. DATASET DESCRIPTION

NSL KDD dataset overcomes the several problems of KDD cup 99. NSL KDD cup 99 is one of effective dataset to help the researches in the field of IDS. It has dimension of 11850x39 including decision attribute. The features include smurf, guess_password, snmp_guess, mscan, normal etc. The dataset can be obtained by using the link [11].

5. RESULTS AND ANALYSIS

Table-3. Results of US-QR & US-PSO-QR

Dimension	Algorithm			
	US-QR		US-PSO-QR	
	Features selected	Time (s)	Features selected	Time (s)
50x38	3	462.2	20	330.8
100x38	6	2047.8	22	821.6
150x38	6	9261.8	23	1872.4
200x38	6	13570.2	23	1889.2
300x38	7	22293.6	34	10062.2

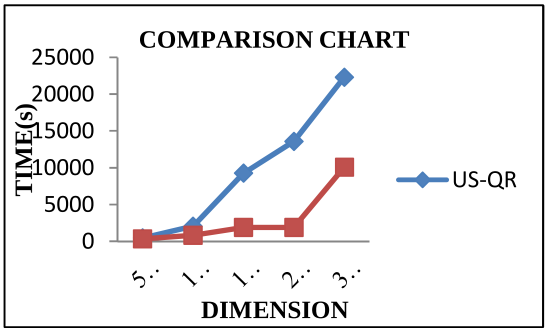


Figure-3. Comparison of US-QR & US-PSO-QR in terms of time.

MATLAB R2012b tool is used to simulate these algorithms. From the results obtained, it is clear that US-PSO-QR selects the attributes efficiently compared to US-QR. US-PSO-QR takes lesser computational time. In all the cases of dimension shown in Table-3, US-PSO-QR selects more features than US-QR.

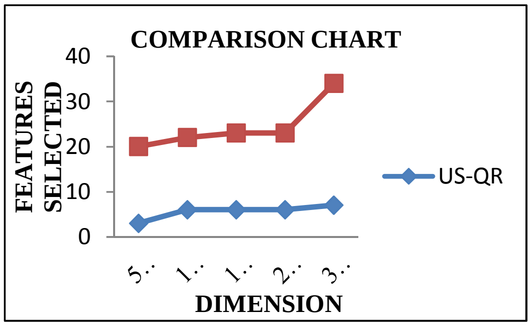


Figure-4. Comparison of US-QR & US-PSO-QR in terms of features selection.

That is, it takes all the necessary features and removes the attributes without having the useful information. It is more suitable when dataset has high dimension. Comparison of two algorithms is shown in Figure 3 and Figure 4.

6. CONCLUSIONS AND FUTURE WORK

In Network security, Intrusion detection plays a major role to detect unauthorized access or attacks. It takes long time for analyzing the data. Feature selection is necessary to reduce redundancy in dataset and reduces analyzing time. The simulation results shows that the RST algorithms namely QR hybridized with PSO algorithm provides better results in terms of time consumption and performs attribute reduction efficiently for different dimensions of network dataset.

In future, QR algorithm can be modified and hybridized with advance swarm optimization techniques and compared with PSO for the high dimensional Network data. It can also be tested for dynamically varying data using Group Incremental approach that utilizes the information entropy.

REFERENCES

[1] Amrita Anand and Brajesh Patel. 2012. “An Overview on Intrusion Detection System and Types of Attacks It Can Detect Considering Different Protocols”, International Journal of Advanced Research in Computer Science and Software Engineering Vol. 2, No. 8, 2012.

[2] Kapil Wankhade and Sadia Patka. 2013. “An Efficient Approach for Intrusion Detection Using Data Mining Methods”, International Conference on Advances in Computing, Communications and Informatics (ICACCI).

[3] Ming Xue and Changjun Zhu. 2009. “Applied Research on Data Mining Algorithm in Network Intrusion Detection”, International Joint conference on Artificial intelligence.

[4] Hannah Inbarani H., Nizar Banu P. K. and Andrews S. 2012. **“**Unsupervised Hybrid PSO – Quick Reduct Approach for Feature Reduction”, Proceedings of IEEE International Conference on Recent Trends in Information Technology, Vol. 24, pp. 11-16.

[5] Hannah Inbarani H., Nizar Banu P. K. 2012. **“**Performance Evaluation of Hybridized Rough Set based Unsupervised Approaches for Gene Selection”, International Journal of Computational Intelligence and Informatics, Vol. 25, No 3-4, pp. 793-806.

[6] Jothi G. and Hannah Inbarani H. 2012. “Soft set based quick reduct approach for unsupervised feature selection”, IEEE –International Conference on

Advanced Communication Control and Computing Technologies.

[7] Velayutham C. and Thangavel K. 2011. “Unsupervised quick reduct algorithm using rough set theory”, Journal of Electronic Science and Technology, Vol. 9, No.3, pp. 193-201.

[8] Eberhart R. C. and Shi Y. 2001. ”Particle swarm optimization: developments, applications and resources”, in: Proceedings of IEEE International Conference on Evolutionary Computation, Vol. 1, pp. 81-86.

[9] Wang X., Yang J. X. T., Xia W. and Jensen R. 2007. “Feature selection based on rough sets and particle swarm optimization”, Pattern Recognition Letters.

[10] Kennedy J. and Eberhart R. C. 1995. “Particle swarm optimization”, Proceedings of IEEE International Conference on Neural Networks.

[11] Nsl-kdd data set for network-based intrusion detection systems, Available : http://nsl.cs.unb.ca/NSL-KDD/

[12] Pawlak Z. 2002. “Rough Set Theory and its applications”, Journal of Telecommunications and Information Technology.

[13] Pawlak Z. 1982. “Rough Sets”, International Journal of Computer and Information Sciences, Vol.11, pp. 341-356.

[14] Dash M. and Liu H. 1997. “Feature Selection for Classification”, Intelligent Data Analysis, Vol.1, pp. 131-156.

[15] Lei Yu and Huan Liu. 2003. “Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution”, Proceedings of the Twentieth International Conference on Machine Learning (ICML).

[16] Velayutham C. and Thangavel K. 2011. ”Improved Rough Set Algorithms for Optimal Attribute Reduct”, Journal of Electronic Science And Technology, Vol. 9, No. 2, June.