Improved Intrusion Detection Technique based on Feature Reduction and Classification using Support Vector Machine and Particle of Swarm Optimization

Sunita Patel
M Tech Scholar
NIRT, RGPV, Bhopal

Jyoti Sondhi
Assistant Prof.
NIRT, RGPV, Bhopal

Anand Motvani
Assistant Prof.
NIRT, RGPV, Bhopal

Anurag Shrivastava
Assistant Prof.
NIRT, RGPV, Bhopal

ABSTRACT
Reduces the file size and increase the performances of classification and intrusion detection technique used in current research trend. The reduction of file size and number of attribute used dimension reduction algorithm and optimization algorithm. Various authors used genetic algorithm, ANT colony optimization and neural network. In this paper used particle of swarm optimization technique for feature reduction and feature selection for support vector machine classification process. The proposed algorithm implemented in MATLAB software and used DARPA dataset for evaluation of proposed method. Our empirical result shows that better detection ratio in compression of other exiting technique such as FCMNN, GSVM.

Keyword
IDS, Feature Reduction, SVM, PSO

1. INTRODUCTION
In this paper proposed an improved intrusion data classification technique using support vector machine and particle of swarm optimization. The classification of intrusion data is very challenging job in the field of network security and machine learning. Various authors used classification technique such as neural network, decision tee and some rule based classification technique [1]. In the process of classification, large number of network attribute put confusion status for classifier. Some authors used feature reduction technique for reduction of feature and increased the classification ratio of intrusion detection. In the process of feature reduction used principle of component analysis and LDPA method by other authors. Some another authors used heuristic and meta-heuristic function for feature optimsation and feature reduction process[2].For example silakari and saliendra[16] gives the feature reduction technique based on PCA algorithm and also used ensemble based classification technique for data classification. In consequence of heuristic function Li [17] was used genetic algorithm for feature reduction process. Jain and Upendra [18] gives a technique of feature reduction cum feature selection for classification of KDDCUP99 data. They also used some machine learning algorithm and improved the performance of classification technique such as decision tree and J48 algorithm. Muda et al. [19] used a process of intrusion data classification based on clustering technique and optimsation algorithm. Some another author used support vector machine classification technique along with feature reduction algorithm in [8, 9]. Z. Xue-qin et al. gives a model of intrusion detection based on support vector machine and FLDA. Zhang and M. Zulkernine [20] used random forest tree algorithm for feature reduction cum classification technique. In this paper used support vector machine for classification of intrusion data and particle of swarm optimization used for the process of feature reduction cum feature selection process. The particle of swarm optimization technique is dynamic population based searching technique, its nature is find the most appropriate nature of attribute for the process of classification. In section II we describe feature of reduction and POS. In section III proposed algorithm. In section IV discuss experimental result analysis and finally conclude in section V.

2. FEATURE REDUCTION AND PSO
In this section discuss feature reduction process and types of feature of network file. The process of feature reduction gives a new set of feature attribute. The new feature attribute set process for classification and clustering task. Some authors used some well know algorithm for feature reduction such as PCA and LDPA. The principle of component analysis is mathematical model used in pattern recognition process[5,6]. PCA is an algorithm that checks and converts the data set for all the correlated variables into a set of uncorrelated variables, also known as principal components. Feature reduction process is illustrated in Figure1. On the left there are the features (F0…FN) that are available from the input of network file data. On the right is the output (V0…VN) of the reduction tool [21]. The number of features in the output usually is less than in the input but it might as well be the same. The new features (V0…VN), can be calculated based on a single feature or a combination of multiple features (F0…FN).

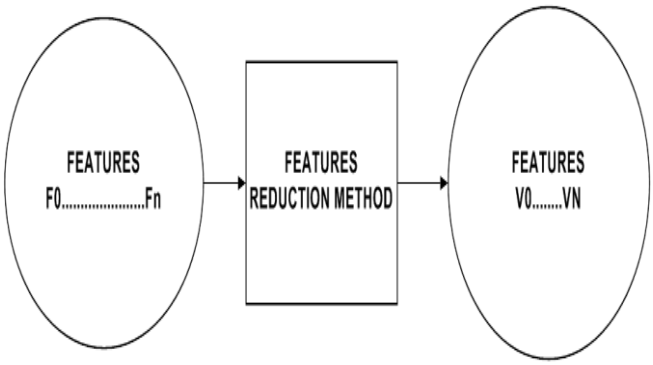

Fig 1: Process of feature reduction

Particle of swarm optimization is dynamic population based optimization technique. The dynamic population based selection process of feature attribute of network traffic data. The network traffic data categories into different section of number of particle. The number of particle process as the

number of attribute are distributed along with rang of path. Some steps are

Step 1: the process of feature attribute in range and distribute and define velocity of particle.

Step 2: the process of velocity update of particle according to their iteration of each particle agent set.

$$v_i = v_i + c_1 R_1 (p_{i,best} - p_i) + c_2 R_2 (g_{i,best} - p_i) \quad (1)$$

where pi and vi are the position and velocity of particle i, respectively; pibest and gibest is the position with the 'best' objective value found so far by particle i and the entire population respectively; w is a parameter controlling the dynamics of flying; R1 and R2 are random variables in the range [0,1]; c1 and c2 are factors controlling the related weighting of corresponding terms. The random variables support the PSO with the ability of stochastic searching.

Step 3: Position updating – The positions of all particles are updated according to,

$$p_i = p_i + v_i \quad (2)$$

After updating, pi should be tested and limited to the allowed range.

Step 4: Memory updating – Update pi,best and gi,bestwhen condition is met,

$$p_{i,best} = p_i \qquad if\ f(p_i) > f(p_{i,best})$$
$$g_{i,best} = g_i \qquad if\ f(g_i) > f(g_{i,best}) \quad (3)$$

wheref(x) is the objective function to be optimized.

Step 5: Stopping Condition–The algorithm repeats steps 2 to 4 until certain stopping conditions are met, such as a predefined number of iterations. Just the once stopped, the algorithm reports the values of gbest and f(gbest) as its solution.

PSO utilizes several searching points and the searching points gradually get close to the global optimal point using its pbest and gbest. Preliminary positions of pbest and gbest are different. However, using thee different direction of pbest and gbest, all agents progressively get close to the global optimum.

3. PROPOSED METHODOLOGY

In this section discuss the process of proposed methodology working process. The proposed method work in two phase in first phase process of feature reduction and second phase process of classification of data into support vector machine.

3.1 Processing of feature reduction

1. Input the process of KDDCUP99 dataset.
2. The input dataset going on the process of transformation using min- max algorithm
3. After the transformation feature of data are map in particle of swarm optimization.
4. The total number of feature transformed into particle.
5. The velocity of particle updated according to reduction process.
6. After the reduction process found new traffic feature map data is called reduces feature set.

3.2 Processing of support vector machine

Step1. The reduces feature set input in from of vector in support vector machine.

Step2. Here show steps of processing of SVM [14]

1) Initialize Gaussian hyper plane margin.
2) Select a random vector from training data and present it to the SVM.
3) Every plane is examined to find the support vector (SV).
4) Estimate the support vector point according to their plane data.
1) The hyper plane data collects and validate data according to their construct model.

Step 3. Finally gets Intrusion data classified and calculate the value of TP, TN, FP and FN.

4. EXPERIMENTAL RESULT AND ANALYSIS

In this section discuss the performance evaluation of proposed algorithm for intrusion detection. The proposed algorithm work on the base of feature reduction cum classification. The process of result analysis used MATLAB 7.8.0 software and KDDCUP99 dataset [18]. The KDDCUP99 dataset is well known and recognized dataset for analysis of intrusion detection. Total recode in KDDCUP99 have approx. seven lac but we used only 10% KDDCUP99 dataset. The total number of categories in KDDCUP99 is five such as NORMAL, DOS, PROB, U2R, and R2L. The total number of reduces attribute is 20. Now the process of classification only used 22 attribute in collection of all categories. For the measuring performance calculate precision, recall and accuracy.

Table 1: Shows that the comparison of all method on given evaluation parameters such as Precision, Recall and Accuracy

Value	Method	%Accuracy	%Precision	%Recall
0.1	FCMNN	89.7999	86.2731	83.8146
	RSVM	95.3094	88.7004	84.8546
	SVM-PSO	96.3094	89.8004	88.0657
0.2	FCMNN	91.4417	87.9149	88.4565
	RSVM	96.3422	90.3422	86.4964
	SVM-PSO	97.9512	91.4422	89.7075
0.3	FCMNN	89.7999	86.2731	83.8146
	RSVM	95.3095	88.7004	88.8546
	SVM-PSO	96.3094	89.8004	88.0657
0.4	FCMNN	91.499	87.9722	85.5138
	RSVM	97.0085	90.3995	86.5537
	SVM-PSO	98.0085	91.4995	89.7648
0.5	FCMNN	91.559	88.0322	85.5737
	RSVM	97.0685	90.4595	86.6136
	SVM-PSO	98.0685	91.5595	89.8248
0.6	FCMNN	89.7999	86.2731	83.8146
	RSVM	95.3094	88.7004	84.8546
	SVM-PSO	96.3094	89.8004	88.0657

For the visualization of different type of categories of attack and normal data used MATLAB figure window.

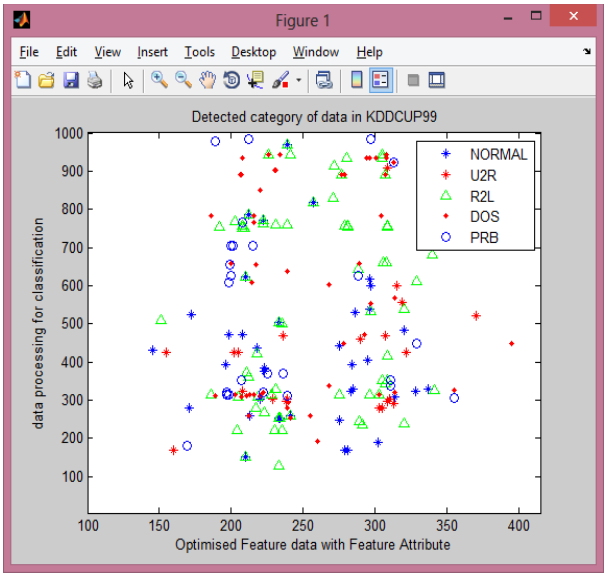

Figure 1: gives the information about the classification map, in this map show that five categories of data classified in different region with different colour map.

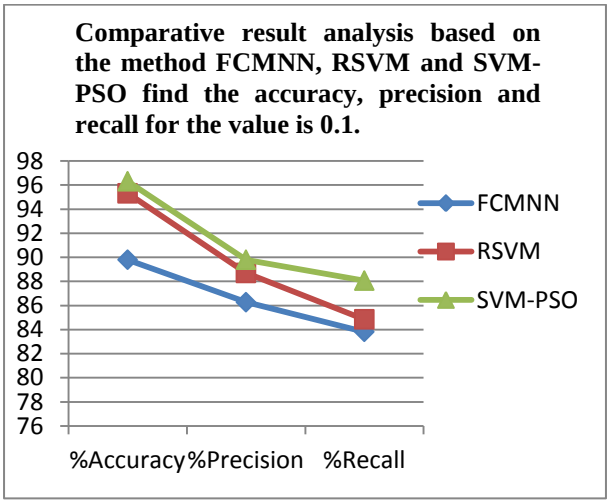

Figure 2: Shows that the performance evaluation of Accuracy, Precision and Recall for the FCMNN, RSVM, SVM-PSO and the given input value is 0.1. The 0.1 value shows that the data generation point for classification

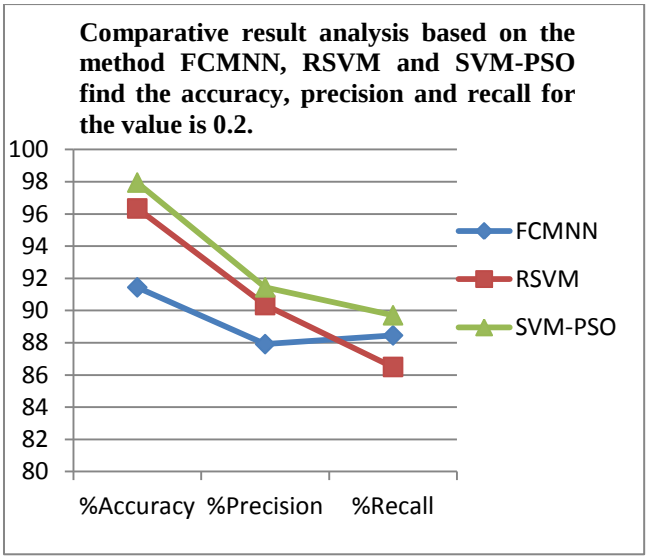

Figure 3: Shows that the performance evaluation of Accuracy, Precision and Recall for the FCMNN, RSVM, SVM-PSO and the given input value is 0.2. The 0.2 value shows that the data generation point for classification

5. CONCLUSION AND FUTURE SCOPE

In this Paper proposed a reduces feature set for the classification of intrusion detection system. The process of feature reduction perform by particle of swarm optimization function. The particle of swarm optimization function dynamically reduces the number of unused feature attribute of traffic data. For the process of classification used support vector machine classifier. The kernel function of suppor vector machine is redial. The proposed algorithm is a combination of feature selection and feature reduction for intrusion detection system. The feature selection and reduction both improved the performance of classification algorithm, but it not achieved the classification ratio 100%. The process of data sampling improved the reduction process and improved the classification ratio up to 100%. The sampling process design as mixed sampler corresponding to the nature of network traffic data, the network traffic data is mixed data type some are continuous and discrete

6. REFERENCES

[1]. A M Chandrasekhar, K raghuveer"Intrusion detection technique by using k-means, fuzzy neural network and SVM classifier" International conference on computer communication and informatics, IEEE, 2013. Pp 1-7.

[2]. V. Bapuji, R. Naveen Kumar, A. Govardhan, S.S.V.N. Sarma "Soft Computing and Artificial Intelligence Techniques for Intrusion Detection System" Network and Complex Systems, Vol-2, 2012. Pp 24-33.

[3]. IftikharAhmad ,Azween Abdullah ,Abdullah Alghamdi , Muhammad Hussain "Optimized intrusion detection mechanism using soft computing Techniques" Springer 2012. Pp 1-9.

[4]. Iftikhar Ahmad, Azween Abdullah, Abdullah Alghamdi "Towards the Selection of Best Neural Network System for Intrusion Detection" 2011. Pp 1-8.

[5]. Bibi Masoomeh Aslahi Shahri, I. R .Adeyami ,Bahareh Maleki Alavi " Intrusion Detection System Using Hybrid

Gsa-K-Means" Proceedings of Global Engineering, Science and Technology Conference, 2013. Pp 1-14.

[6]. ShaohuaTeng, Hongle Du, Naiqi Wu, Wei Zhang, Jiangyi Su "A Cooperative Network Intrusion Detection Based on Fuzzy SVMs" JOURNAL OF NETWORKS, VOL. 5, 2010. Pp 475-484.

[7]. Krishna Kant Tiwari ,Susheel Tiwari , Sriram Yadav " Analyze the Different Kernel Function in SVM for IDS" International Journal of Advanced Research in Computer Science and Electronics Engineering, Vol-2, 2013. Pp 623-632.

[8]. SannasiGanapathy, KanagasabaiKulothungan,"Intelligent feature selection and classification techniques for intrusion detection in networks: a survey" EURASIP Journal on Wireless Communications and Networking , Springer 2013. Pp 1-16.

[9]. Gang Wang, Jinxing Hao, Jian Ma, Lihua Huang "A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering" Expert Systems with Applications, Elsevier 2010. Pp 1-8.

[10]. RituRanjani Singh, Neetesh Gupta "To Reduce the False Alarm in Intrusion DetectionSystem using self Organizing Map" International journal of Computer Science and its Applications, 2010. Pp 65-72.

[11]. AnshulChaturvedi and Prof.VineetRichharia "A Novel Method for Intrusion Detection Based on SARSA and Radial Bias Feed Forward Network (RBFFN)" in international journal of computers & technology vol 7, no 3.

[12]. Mohammad Behdad, Luigi Barone, Mohammed Bennamoun and Tim French "Nature-Inspired Techniques in the Context of Fraud Detection" in ieee transactions on systems, man, and cybernetics—part c: applications and reviews, vol. 42, no. 6, november 2012.

[13]. Alberto Fernandez, Maria Jose del Jesus and Francisco Herrera "On the influence of an adaptive inference system in fuzzy rule based classification system for imbalanced data-sets" in Elsevier Ltd. All rights reserved 2009.

[14]. P. Garcia-Teodoro, J. Diaz-Verdejo, G. Macia-Fernandez and E.Vazquez "Anomaly-based network intrusion detection: Techniques, Systems and challenges" in Elsevier Ltd. All rights reserved 2008.

[15]. Terrence P. Fries "A Fuzzy-Genetic Approach to Network Intrusion Detection" in GECCO 08, July12–16, 2008, Atlanta, Georgia, USA.

[16]. Shailendra Singh, Sanjay Silakari "An Ensemble Approach for Cyber Attack Detection System: A Generic Framework" 14th ACIS, IEEE 2013. Pp 79-85

[17]. X. Li et al., "Smart Community: An Internet of Things Application," IEEE Commun. Mag., vol. 49, no. 11, 2011, pp. 68–75.

[18]. Jain and Upendra "An Efficient intrusion detection based onDecision Tree Classifier using feature Reduction", InternationalJournal of scientific and research Publications , Vol. 2, Jan. 2012.

[19]. Muda, Y. Yassin, M.N. Sulaiman and N.I. Udzir, " A K-Means andNaive Bayes Learning Aproach for Better Information Detection",Information Technology journal, Asian Network For scientificInformation publisher, Vol. 10 , 2011.

[20]. Zhang and M. Zulkernine, "Network Intrusion Detection usingRandom Forests", School of Computing Queen's University,Kingston Ontario, 2006.

[21]. Sunita Patel and JyotiSondhi "A Review of Intrusion Detection Technique using Various Technique of Machine Learning and Feature Optimization Technique" in International Journal of Computer Applications, Volume 93 – No 14, May 2014