I. J. Computer Network and Information Security, **2021, 4, 26-34**
Published Online August 2021 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijcnis.2021.04.03

Intrusion Detection System Using Ensemble of Rule Learners and First Search Algorithm as Feature Selectors

D P Gaikwad
AISSMS College of Engineering, Pune, Maharashtra, India
E-mail: dp.g@rediffmail.com

Received: 21 January 2021; Revised: 02 March 2021; Accepted: 17 March 2021; Published: 08 August 2021

Abstract: Recently, the use of Internet is increased for digital communication to share a lot of sensitive information between computers and mobile devices. For secure communication, data or information must be protected from adversaries. There are many methods of safeties like encryption, firewalls and access control. Intrusion detection system is mainly used to detect internal attacks in organization. Machine leaning techniques are mostly used to implement intrusion detection system. Ensemble method of machine learning gives high accuracy in which moderately accurate classifiers are combined. Ensemble classifier also provides less false positive rates.

In this paper, a novel ensemble classifier using rule combination method has proposed for intrusion detection system. Ensemble classifier is designed using three rule learners as base classifiers. The benefits and feasibility of the proposed ensemble classifier have demonstrated by means of KDD'98 datasets. The main novelty of the proposed approach is based on three rule learner combination using rule of combination method of ensemble and feature selector. These three base classifiers are separately trained and combined using average probabilities rule combination. Base classifier's accuracies have compared with the proposed ensemble classifier. Best First search algorithm has used to select relevant features from training dataset. This algorithm also helped to reduce dimension of training and testing dataset which benefits in reduction of training time. Several comparative experiments are conducted for evaluating performances of classifiers in term of accuracy and false positive rates. Experimental results show that the proposed ensemble classifier provide significant improvement of accuracy compared to individual classifiers with less positive rates.

Index Terms: Homogeneous Classifier, PART, Jrip, Ensemble, Rule Learner.

1. Introduction

Now a day, computer networks are being widely used in digital communication to share a lot of sensitive information between computer in network and mobile devices. For secure communication, data or information must be protected from enemies. There are many methods of securities, like encryption, firewalls and access control. This securities mechanism protects data in certain cases. In some cases, network security holes rise day by day [1]. To avoid these holes, intrusion detection systems are playing vital role. Many investigators have proposed different intrusion detection systems using various methods and techniques. Researchers have proposed misuse detection method to compare activities in networks. In this method, to represent a specific attack pre-defined pattern or signatures extracted from characteristic features are used. This method is useful to determine known attacks with a low false positive rate. Misuse detection methods do not have capability to detect unknown attacks. Anomaly detection method also has proposed by many researchers to detect novel attacks in networks. In this method, attacks are detected by identifying eccentricities from normal activities in network. It is well known in detecting unknown attacks with high false positive rate [2, 3]. To overcome limitations of both methods, data mining algorithms have proposed by many researchers. Intrusion detection systems require analyzing millions of records in network. Data mining is useful to analyze data to find hidden pattern and launch relationships between data in database. Different approaches of data mining such as clustering, classification, regression, association rules and outlier detection are being used for intrusion detection. These approaches of data mining are also known as machine learning algorithms. Machine learning process consists of three stages; data exploration, model building and deployment. In data exploration stage, data cleaning, data transformation and relevant feature selection are performed. In model building stage, best model is selected based on predictive performance of model. In development stage, selected best model is applied for new data to estimate expected outcome [4]. Data preprocessing plays vital role in intrusion detection system for selecting relevant features from training dataset. It is also help in reducing training time to enhance performance of intrusion detection system. In general, a base single

This work is open access and licensed under the Creative Commons CC BY 4.0 License.

classifier is not proficient to obtain the best representation in the hypothesis space. For insufficient input dataset to train a model, single classifier may lead to a weak or false hypothesis. For producing suitable hypothesis, a single classifier may take more time. For overcoming above limitations, it is essential to combine autonomous classifiers to improve the predictive performance. Ensemble is a machine learning method in which individual classifiers are combined to give best performance. It is a method of machine learning in which several homogeneous or heterogeneous base classifiers can be combined to obtain high accuracy with less false positive rate [5]. Ensemble classifier has higher stability and accuracy. There are many algorithms available for ensemble base classifiers; Bagging and Boosting are widely used ensemble algorithms which produce higher classification accuracy. Voting combination method of ensemble is used to improve performance of classification. Other algorithms such as Bayesian parameter averaging and stacking also used to ensemble base classifiers [6, 7].

In literature review, we found that there are many ensemble classifiers which have been used for intrusion detection system. Jasmeen K. Chahal and Amanjot Kaur [8] have proposed hybrid intrusion detection to improve false positive and false negative rate. Authors have used K-mean algorithm for clustering data and adaptive Support Vector Machine for classification purpose. In this contribution, author has mentioned that the proposed Hybrid classifier offers 98.47 % accuracy on NSL_KDD dataset with false positive rate 0.53%. Author has not mentioned accuracy on test dataset. Our aim of this research work was to increase classification accuracy of ensemble classifier on both training and test dataset. The objectives of this research work to select suitable rule learner as a base classifiers to improve classification accuracy and lesser false positive rates. The main objective of this research work is to increase accuracy of intrusion detection system than above mentioned proposal. In this paper, we have chosen three rule learners which are accurate and efficient for rule making. Jrip, Partial Decision Tree and OneR have selected for intrusion detection and combined using rule combination method of machine learning. Three rule learners are used as base classifiers to take advantages of each base classifier. To reduce dimension of dataset, feature selection method is implemented. Researchers have used many feature selection methods to reduce dimension of training dataset. Search method is widely used for feature selection. The best first search is used to select best features from training dataset. In search methods, the attributes are traversed in the attribute space to find a good subset of attributes. Attribute subset evaluator is used to measure quality of selected attributes [9]. Hill climbing with backtracking capability is used in Best First Search. The accuracy of intrusion detection system have evaluated on test dataset. Paper is organized as follows. In section 2, literature survey on ensemble classifiers is done. Section 3 is used to present the proposed ensemble model. In section 4, experimental results have discussed with analysis. In final conclusion section, paper is concluded with future scope.

2. Literature Survey

Ensemble of classifier is method of machine learning. In this method, base classifiers are combined by using different combination rules. It is found that this method is very useful to reduce false positives rates. The selection of base classifier is very important step in Ensemble method. Many researchers have proposed many ensemble classifiers for intrusion detection system. In ensemble method, homogeneous and heterogeneous weak classifiers can be used. There are many approaches of ensemble base classifiers. We have studied some related research works as below.

Xiaofeng Zhao and Hua Jiang, LiYan Jiao [10] have introduced a data fusion-based intrusion detection model. The system information is abstracted from low-level to high-level for intrusion detection system. Different basic detectors' output has fused to get the detection more accuracy. Authors have introduced D-S evidence theory to model at information layer. The experiments show that the proposed method provides the best detected rate. It is also found that this proposed method provides better scalability and robust system which gives lower false negatives rate and false positive rate. Saeed Khazaee and Karim Faez [11] have used fuzzy clustering and several neural networks for intrusion detection. Fuzzy C-mean algorithm have used for clustering training samples and to detect and move inappropriate data to another dataset. Neural network have trained by new labels regression combination and classification. Authors have used fuzzy ARTMAP neural network to classify outlier. The proposed system has performed improved than the earlier works. Reena Sharma and Gurjot Kaur [12] have proposed RBF neural network technique for spam detection in email. The proposed system has implemented to improve the precision measure of the existing spam detection. The results obtained by using proposed technique are compared with Support vector machine. Experimental result show that proposed methods is better than support vector machine in term of recall, precision and accuracy. Authors have suggested modifying the proposed spam detection system by combining it with other algorithms for making a hybrid spam detection system.

Nabil Moukafih [13] has proposed reliable neural network based intrusion detection system. Combination of simple feed forward neural networks have used as a weak classifier. They have also proposed an approach to create weak classifier using neural network as an expert. The performance of this approach is superior to deep learning models. Smitha Rajagopal et. al., [14] have proposed a stacking ensemble method for intrusion detection system. They have stacked four base classifiers to implement real time intrusion detection system. They have used logistic regression, support vector machine, K-nearest neighbour and random forest tree as a base classifier. Two real time training datasets namely UGR'16 and UNSW NB-15 have used for training and testing their proposal. The proposed system is capable to generate predication accuracy up to 97 % on real dataset. M. Govindarajan [15] has proposed ensemble classifier using

heterogeneous and homogeneous base classifiers. They have used Support vector machine and Radial Basis Function as a base classifier. They have used real time and benchmark data sets for training models. They have compared accuracy of each classifier with both ensemble classifiers. It is found that accuracy of heterogeneous ensemble classifier exhibits better accuracy than homogenous ensemble classifier. Habil Damania et.al. [16] have proposed ensemble classifier with three base classifiers. They have used NSL_KDD dataset for training base and ensemble classifiers. They have use Random forest and Support vector machines as a base classifier. Ansam Khraisat et.al. [17] have developed a new intelligent intrusion detection system. They have integrated both signature and anomaly detections methods. C5 classifiers have used to generate signatures which are able to generate a rules pattern. This classifier is also able to detect intrusions using less numbers of signatures. C5 classifier is ensemble with one class support vector machine which provide better detection rate. They found that ensemble classifier has exhibits better results than base classifiers. Mrutyunjaya Panda and Manas Ranjan Patra [18] have proposed an intrusion detection system using voting ensemble classifier. They have used J48 decision tree, Sequential Minimal Optimization) and rule learner as a base classifier. Results show that this voting ensemble classifier yield better results compared to other systems. Waweru Mwangi and Dr. Otieno Calvin [19] have proposed an intrusion detection system using two ensembles classifiers. It includes feature ensemble selecting classifier and data mining classifier. Feature ensemble selecting classifier consists of four classifiers using dissimilar feature sets. Data mining classifier is used to extract normal behaviour from training dataset. For getting final decision, authors have combined these two classifiers. They have concluded that ensemble approach takes more training. Ngoc Tu Pham, Ernest Foo and Suriadi Suriadi [20] have proposed ensemble method for intrusion detection system. They have used feature selection to reduce dimension of training and testing dataset. Authors have used two methods of feature-selection for reducing dimensionality of dataset. In this ensemble classifiers, Bagging and Booting techniques are using tree-based base classifiers. They have used J48, REPTree, Random Tree and Random as a base classifier. They pointed out that proposed model shows the outperformance in comparison with other existing models with some limitations. They have suggested using different datasets for different classifiers to improve classification accuracy. Hariharan Rajadurai and Usha Devi Gandhi [21] have proposed stacked ensemble-based intrusion detection system. In this system, authors have used random forest tree and gradient boost as a base classifier. Gradient boost is an improved version of Adaboost which can boost any differentiable loss function. The Random Forest built set of rules to generate many decision trees. It is found that stacked ensemble learning shows better accuracy compare to other intrusion detection system. Mohammad Reza Parsaei et.al. [22] have proposed intrusion detection system using classification and clustering. In this research, authors have exploited SMOTE created boundary margin for low frequent attacks. Specifically, this system has ability to detect U2R and R2U attacks. They have reduced dataset by using LOO method. The dataset is balance by using CANN to extract single dimension dataset. Experimental results show that proposed method outperforms the base line method concerning detection rate.

3. Methodology of the Proposed Intrusion Detection System

3.1. Feature Selection to Reduce Dimension of Dataset

Many feature selection methods are available in literature. Search method is widely used for feature selection. In search methods, the attributes are traversed in the attribute space to find a good subset of attributes. Attribute subset evaluator is used to measure quality of selected attributes. Best First Search use concept of greedy hill climbing approach. Best first search performs backward and forward search. In this forward search, set of attributes is initializing to zero and searching starts in forward direction. In backward search, set of attributes keep full and start searching in backward direction. Otherwise, it can start at any point and search in both directions in some cases. Finally, selected and evaluated subset is cached for efficiency. There are different evaluators such as correlation-based attribute subset evaluator, class consistency-based evaluator, classifier-based evaluator and coefficient-based evaluator. Coefficient based evaluator ranks attributes according to the magnitude of the coefficients learned by a support vector machine [9]. In this paper, we have used Best First Search algorithm to select relevant features. Following 12 attributes have selected from 41 attributes in training dataset which are listed in table1.

3.2. Introduction to Base Classifiers

In this section, base classifiers are introduced in detail as follows. **Jrip Rule Base Classifier:** Jrip is proposed by W William and it is based on Repeated Incremental Pruning to Produce Error Reduction induction process. In Ripple induction mechanism, three stages are used to process rules. In first stage, growing and pruning sets are formed from training dataset. Using pruning sets, total rules are reduced. In optimization stage, all reduced rules are reviewed to reduce the error on the entire training dataset. In cleaning phase, if Description Length rule increases rules sets then it is deleted V. Veeralakshmi [23]. Jrip rule classifier is guessing and governs RIPPLE rule classifiers which develop proportional rules. Jrip is improved version of IREP (incremental reduced error pruning) algorithm which was developed by Furnkranz and Widmer. William W. Cohen modified IREP algorithm to improve it. Three modifications have implemented. In first modification, a different metric is used for assessing the value of rules in the pruning phase. A new heuristic is used to stop adding rules to rule set. In optimization step, the rules produced by IREP algorithm are post processed. Repruning each rule helped minimize the error of the complete rule set [24]. General Algorithm of Jrip

is given below. Algorithm is divided in three subroutines. FORMRULESET procedure creates rule set from training dataset. Procedures RULESETOPTIMZE create optimize rule set using growing and pruning rules. Procedure JRIP call two procedures for creating set of rules and optimizing rule set.

Table 1. Selected attributes using Best First Search.

Sr. No.	Attribute Name
1	Service
2	Flag
3	Src_bytes
4	dst_bytes
5	logged_in
6	serror_rate
7	srv_serror_rate
8	same_srv_rate
9	diff_srv_rate
10	dst_host_srv_diff_host_rate
11	dst_host_serror_rate
12	dst_host_srv_serror_rate

Algorithm 1: Jrip Algorithm
Procedure JRIP (P, N, k)
RuleArray: = FORMRULESET (P, N)
For k time RuleArray: = RULESETOPTIMZE (RuleArray, P, N)
return (RuleArray)
end JRIP

Procedure FORMRULESET (P, N)
P=positive Candidates, N=negative Candidates.
RuleArray = {}
DL=DescriptionLength (RuleArray, P, N)
while P not equal to {}
// Grow and prune a new rule
split (P, N) into (GrowPositive, GrowNegative) and (PrunePositive, PruneNegative)
SingleRule: = GrowRule (GrowPositive, GrowNegative)
SingleRule: = PruneRule (SingleRule, PrunePositive, PruneNegative)
add SingleRule to RuleArray
if DescriptionLength (RuleArray, P, N) > DL+64 then
// Prune the whole rule set and exit
for each rule R in RuleArray
if DescriptionLength (RuleArray -> R}, P, N) < DL then
delete R from RuleSet
DL: = DescriptionLength (RuleArray, P, N)
endif
end for
return (RuleSet)
endif
DL: = DescriptionLength (RuleArray, P, N)
delete from P and N all examples covered by Rule
end while
end FORMRULESET

Procedure RULESETOPTIMIZE (RuleArray, P, N)
for each rule R in RuleArray
delete R from RuleArray
UpdatePositive: = examples in P not covered by RuleArray
UpdateNegative: = examples in N not covered by RuleArray
Spilt (UpdatePositive, UpdateNegative) into (GrowPositive, GrowNegative) and (PrunePositive, PruneNegative)
RepeteRule: = GrowRule (GrowPositive, GrowNegative)

RepeteRule: = PruneRule (RepeteRule, PrunePositive, PruneNegative)
ReviseRule: = GrowRule (GrowPositive, GrowNegative, R)
ReviseRule: = PruneRule (ReviseRule, PrunePositive, PruneNegative)
choose better of RepeteRule and ReviseRule and add to RuleArray
end for
end RULESETOPTIMZE

PART base classifier: Supervised Rule learner is a machine learning algorithm which tries to make a set of rules. C4.5 and RIPPER are two good rule learners which perform two steps to induce rules. Initially, rules are determined and then rules are discarded according by global optimization technique. C4.5 produces an unpruned decision tree and converts this tree into a set of rules. After rule ranking, each rule is simplified separately. It deletes rules from the rule set as long as the rule set's error rate on the training instances decreases. For rule generation in RIPPER, a divide-and-conquer method is used. Only one rule is generated at a time. The instances are deleted which are covered by this rule. For remaining examples in training dataset, RIPPER iteratively creates new rules. Frank and Witten developed a rule generation wherein no global optimization strategy is used to generate appropriate rules. Partial Decision Trees generates rules using divide-and-conquer approach of RIPPER and organizations it with C4.5 decision tree. PART produces a set of rules according to the divide-and-conquer strategy. It eliminates all examples from the training dataset that are covered by this rule. It precedes recursively until no instance remains. PART forms a partial decision tree for the existing set of instances to produce a single rule. It selects the leaf with the largest coverage as the new rule. For avoiding early generalization, the partial decision tree is discarded [25].

OneR Base classifier: OneR is rule learner algorithm and known as "One Rule". It is a simple classification algorithm that produces a one-level decision tree. OneR is able to deduce clearly easy and correct classification rules from a set of instances. OneR is also able to grip missing values and numeric attributes showing flexibility in spite of simplicity. The OneR algorithm produces one rule for each attribute in the training data. It selects the rule with the minimum error rate [26].

4. Proposed Ensemble Classifier for Intrusion Detection System

In this section, the proposed ensemble classifier is explained in detail. The proposed system is rule based which is faster in training as compared with other existing intrusion detection system. The system is implemented by combing three best rule learners. The ensemble classifier offers highest accuracy with less false positive rates. Initially, three base classifiers are trained separately. These three base classifiers have tested on training, test dataset and on cross validation. To take advantages of these classifiers, they have combined together using "Average Probabilities" combination rule. Figure 1 represents system architecture of suggested Ensemble classifiers. In this planned architecture, both the training and testing dataset were applied for feature selection using Best First Search algorithm. This process eliminates irrelevant feature from training dataset which help in reduction of dimensionality. The selected relevant features are given to each base classifier for training. In final stage, ensemble classifier has trained and tested using reduced training dataset. It is found that "average Probabilities" rule is sturdiest non-trainable combination types. In this scheme, an average of confidence score of a certain class through individual base classifier is calculated to obtain the final score of the class using Equation 1.

$$r_{n=3} = \frac{1}{M}\sum_{m=1}^{M} P_m^{n=3} \quad (1)$$

Where, P_m^n is the posterior probability score of class n gained from classifier m for any data instances. Where, r_n be the combined probability score of class n. In algorithm 2, steps are given to calculate final score of classifiers. This final score is used to take final decision to decide class of test sample.

Algorithm 2: Ensemble class for Intrusion Detection system
Input: M sample in Training Dataset
Assumptions: C is class $r_{n=3}$ final score.
Train PART rule learner C_1 //posterior score of class C_1
Calculate $P_m^{n=1}$
Train Jrip rule learner C_2 // posterior score of class C_2
Calculate $P_m^{n=2}$
Train OneR rule learner C_3 // posterior score of class C_3
Calculate $P_m^{n=3}$
Calculate combined score using $r_{n=3} = \frac{1}{M}\sum_{m=1}^{M} P_m^{n=3}$
Output: Combined Decision of Classifiers.

5. Experimental Results and Analysis

In this section, we discussed experimental process and analysis of results. The proposed ensemble classifier and base classifiers have trained and tested using KDD'99 dataset. Various experiments have conducted for performance evaluations of base classifiers and ensemble classifier. Initially, we have conducted experiments on each base classifier on training and test dataset. The performances of each base classifier have noted and listed in table for further comparisons with ensemble classifier. Then, three classifiers have been combined together using voting rule combination method of machine learning. Average of Probabilities rule combination is used to combine base classifiers. The base classifiers and Ensemble classifier are trained and tested on Intel (R) Core (TM) i5, 3210M CPU @ 2.50 GHz ,8 GB RAM and coding is done in WEKA programming. Following metrics are used to evaluate classification accuracies of classifiers. Performances of classifiers are measured in term of precision, accuracy and recall using followings equations [2-4].

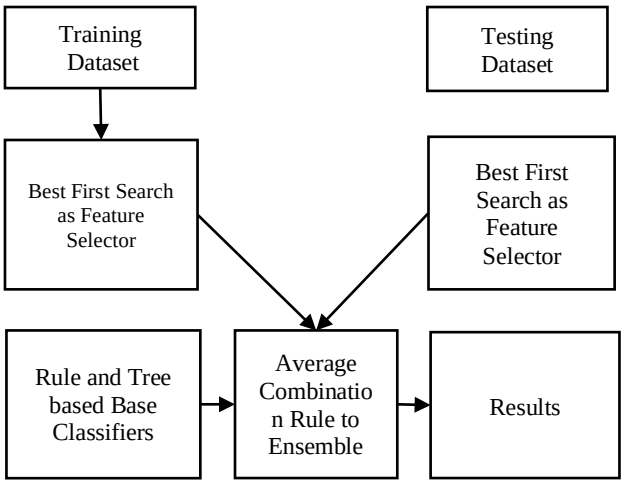

Fig.1. The Proposed Architecture of Ensemble Classifier.

$$Accuracy = \frac{TPc+TNc}{TPc+TNc+FPc+FNc} \quad (2)$$

$$Pc = \frac{TPc}{TPc+FPc} \quad (3)$$

$$Rc = \frac{TPc}{TPc+FNc} \quad (4)$$

Where, TP_c, a number of test instances correctly classified as belong to class c (True positives) and TN_c, a number of test instances correctly classified as not belong to class c (True Negative). FP_c is a number of test instances incorrectly classified as belonging to class c (False positives).). FN_c is a number of test instances incorrectly classified as not belonging to class c (False negative).

The experimental results have summarized in Tables 1-3. Classification accuracies of the proposed ensemble classifiers and base classifiers on training and test datasets have listed in Table 1. Essential potency of intrusion detection system is depending on identification of new or unknown attack. The strength of any intrusion detection system is measured in term of accuracy on test dataset because this accuracy is measured on unknown patterns. According to Table 2 and Figure 2, classification accuracy of the proposed ensemble classifier is higher than its base classifiers. It can be also concluded that PART provides more accuracy on training dataset and cross validation. It is also can be seen that the proposed ensemble classifier offers higher classification accuracy on cross validation using KDD dataset than existing system referred in reference number 8.

Table 2. Accuracies of base and ensemble Classifiers.

Classifier Name	Accuracy on Training Dataset	Accuracy on Test Data	Accuracy on Cross Validation
Jrip	99.85%	81.50 %	99.74 %
PART	99.89%	81.08 %	99.82 %
OneR	99.80%	81.37 %	96.29 %
Hybrid of K-Mean and SVM [ref-8]	NA	NA	98.47%
The proposed Ensemble classifier	99.87%	82.97 %	99.80 %

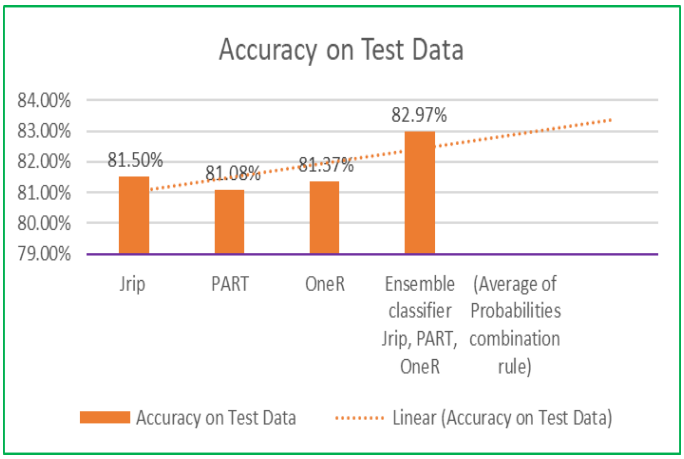

Fig.2. Accuracies of Classifiers on Test Dataset.

For any intrusion detection system, reducing false positive rate is very difficult task. The objective of this research was to reduce false positive and increase true positive rates. Table 3 shows the values of true positive, false positive, precision and recall values offered by ensemble and base classifiers on test dataset. According to Table 3, we can see that proposed ensemble classifier produced lowest false positive and highest true positive rates. It can be also seen that both precisions and recall values of the proposed classifier are higher than its base classifiers.

Table 3. True Postive, False Postive, Precision and Recals of Classifiers.

Classifier Name	TP Rate	FP Rate	Precision	Recall
Jrip classifier on Test Dataset	0.815	0.148	0.855	0.815
OneR on Test Dataset Jrip classifier on Test Dataset	0.814	0.151	0.85	0.814
PART classifier on Test Dataset	0.811	0.165	0.833	0.811
The proposed Ensemble classifier.	0.83	0.137	0.863	0.830
Hybrid of K-Mean and SVM [ref-8]	NA	0.530	NA	NA

Finally, Training time is also very important feature of any classifier. In table 4, training timings of base classifiers and ensemble classifier have listed. According to Table 4, it can be concluded that the base classifiers take less time for training than ensemble classifier.

Table 4. Number of Rules Generated by Classifiers and Model Building Timings.

Classifier Name	Number of Rules	Model Building Time in Sec
Jrip	46	964.35
PART	112	270.44
OneR	One Rule on Src_bytes	31.06 seconds
The proposed Ensemble classifier	Combination of Rules	1089.67 seconds

6. Conclusions and Future Scope

Intrusion detection system is very important system to identify malicious activities in Network. Machine leaning techniques are mostly used to implement intrusion detection system to offer high accuracy with less false positive rate. Recently, Ensemble method of machine leaning is widely used to take advantages of different classifiers. In this paper, ensemble method of machine leaning for intrusion detection has proposed. Ensemble method is very advanced technique of machine learning which provide highest accuracy than its base classifiers. For obtaining higher classification accuracy, ensemble method of machine learning has been proposed for Intrusion detection system. Initially, three rule learners have trained separately using KDD99 dataset. These three base classifiers are combined using average probabilities rule combination. Best First search algorithm has used to select relevant features from training dataset. This algorithm also helped to reduce dimension of training and test dataset which benefits in reduction of training time. For robust intrusion detection system, high accuracy on test dataset is very essential. Classification accuracy on test dataset is very vital for any classifier because this accuracy is offers on new samples.

Experimental results show that the proposed Ensemble classifier using three rule learner based base classifiers offer highest accuracy on test dataset than its base classifiers. It is also found that PART provides more accuracy on training dataset and cross validation. For any intrusion detection system, reducing false positive rate is very challenging task. In this research work, we found that false positive rate is reduced by the proposed Ensemble classifier on test dataset as compared with reference paper 8. False positives rate on test dataset of Ensembles is less than other classifiers. False positives rate on training dataset and cross validation is same for PART and Ensemble classifier. Recall and Precision call value of Ensemble classifier on test dataset is also higher than all base classifiers. The performance of Ensemble classifier using homogeneous base classifiers have studied and analyzed. In future, the system will be implemented using ensemble of heterogeneous base classifiers to study and compare performances of this and new system.

References

[1] Reda M. et.al, “A Hybrid Network Intrusion Detection Framework Based on Random Forests and Weighted K-means”, Ain Shams Engineering Journal (2013) 4, 753–762, 2013.

[2] Barbara D at.el., “DAM: detecting intrusions by data mining”, In: Proc 2nd annu IEEE workshop in assure secure, New York; 2001. pp. 11–6, 2001.

[3] Zhang J et.al. “Random forest-based network intrusion detection systems”, IEEE Transactions on Systems, Man, and Cybernetics – Part C. Applications and Reviews 2008; 38(5):648–58, 2008.

[4] Snehlata S., Kapil Wankhade and Dongre,” Intrusion Detection System Using New Ensemble Boosting Approach”, International Journal of Modeling and Optimization, Vol. 2, No. 4, August 2012.

[5] Gaikwad D. P. and R. C. Thool, “Intrusion detection system using bagging ensemble method of machine learning”, In International Conference on Computing Communication Control and Automation, IEEE. pp. 291–295. doi:10.1109/ICCUBEA.2015.

[6] Yuyang Zhou, “Building an Efficient Intrusion Detection System Based on Feature Selection and Ensemble Classifier”, arXiv:1904.01352v4 [cs.CR] 2 Apr 2020.

[7] Breiman, L., “*Bagging predictors. Machine learning”*, *24*, 123–140. Doi: 10.1007/BF00058655.

[8] Jasmeen K. Chahal, Amanjot Kaur,"A Hybrid Approach based on Classification and Clustering for Intrusion Detection System", International Journal of Mathematical Sciences and Computing, Vol.2, No.4, pp.34-40, 2016.

[9] Ian H. Witten, Eibe Frank and Mark A. Hall, “Data Mining Practical Machine Learning Tools and Techniques”, Third Edition, Morgan Kaufmann Publishers, 2011.

[10] Xiaofeng Zhao,Hua Jiang,LiYan Jiao,"A Data-Fusion-Based Method for Intrusion Detection System in Networks", International Journal of Information Engineering and Electronic Business, vol.1, no.1, pp.32-40, 2009.

[11] Saeed Khazaee, Karim Faez,"A Novel Classification Method Using Hybridization of Fuzzy Clustering and Neural Networks for Intrusion Detection", International Journal of Modern Education and Computer Science, vol.6, no.11, pp.11-24, 2014.

[12] Reena Sharma, Gurjot Kaur,"E-Mail Spam Detection Using SVM and RBF", International Journal of Modern Education and Computer Science, Vol.8, No.4, pp.57-63, 2016.

[13] Nabil, Ghizlane and Said El Hajji, “Neural Network-Based Voting System with High Capacity and Low Computation for Intrusion Detection in SIEM/IDS Systems”, In Security and Communication Networks Volume 2020, Article ID 3512737, pages-15,2020.

[14] Smitha Rajagopal et. al., “A Stacking Ensemble for Network Intrusion Detection Using Heterogeneous Datasets”, Hindawi, Security and Communication Networks, 2020, Article ID 4586875, pages 9, https://doi.org/10.1155/2020/4586875. 2020.

[15] M. Govindarajan, “Evaluation of Ensemble Classifiers for Intrusion Detection,” World Academy of Science, International Journal of Computer and Information Engineering, Vol: 10, Issue No: 6, 2016.

[16] Habil Damania et.al. “MAIDEn: A Machine Learning Approach for Intrusion Detection using Ensemble Technique”, *International Journal of Computer Applications, Volume 179 – No.13.2018.*

[17] Ansam Khraisat et.al., “Hybrid Intrusion Detection System Based on the Stacking Ensemble of C5 Decision Tree Classifier and One Class Support Vector Machin”, MDPI, Electronics 2020, Vol. 9, No. 173, 2020.

[18] Mrutyunjaya Panda and Manas Ranjan Patra, “Ensemble Voting System for Anomaly Based Network Intrusion Detection”,

International Journal of Recent Trends in Engineering, Vol 2, No. 5, November 2016.
[19] Waweru Mwangi and Dr. Otieno Calvin, "Ensemble Network Intrusion Detection Model Based on Classification and Clustering for Dynamic Environment", International Journal of Engineering Research and Technology, ISSN: 2278-0181, Vol. 7 Issue 02, February-2018.
[20] Ngoc Tu, Ernest Foo and Suriadi, "Improving Performance of Intrusion Detection System Using Ensemble Methods and Feature Selection", Australasian Computer Science Week, ACSW, Brisbane, QLD, Australia. ACM, New York, NY, USA, 2018.
[21] Hariharan Rajadurai and Usha Devi Gandhi, "A stacked Ensemble Learning Model for Intrusion Detection in Wireless Network," in Neural Computing and Applications, Springer-Verlag London Ltd., 2020.
[22] Mohammad Reza Parsaei, Samaneh Miri Rostami, Reza Javidan, "A Hybrid Data Mining Approach for Intrusion Detection on Imbalanced NSL-KDD Dataset", *International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 7, No. 6, 2016.*
[23] V. Veeralakshmi, "Ripple down Rule learner (RIDOR) Classifier for IRIS Dataset," In international Journal of Computer Science Engineering, ISSN: 2319-7323 Vol. 4 No.03, 2015.
[24] William Cohen, "Fast Effective Rule Induction Machine Learning," in the Proceedings of the 12th International Conference, 2015.
[25] Helmut, Dieter and Michael, "Exploiting Partial Decision Trees for Feature Subset Selection in e-Mail Categorization", in SAC'06, April 23-27, 2006, Dijon, France, 2006.
[26] Vaishali S. Parsania et.al., "Applying Naïve Bayes, BayesNet, PART, JRip and OneR Algorithmson Hypothyroid Database for Comparative Analysis", in International Journal of Darshan institute of Engineering research and Emerging Technologies, Vol. 3, No. 1, 2014.

Authors' Profiles

D P Gaikwad received the B.E. degree in Computer Science and Engineering from SGGS Institute of Engineering Technology and Maharashtra State, India in 1995. He has completed his M.Tech in Computer Science and Engineering from College of Engineering Pune in 2006. He has been awarded Ph.D degree in Computer Science and Engineering in 2017 from SGGSIOET, Nanded, India. Currently, he is working as Associate Professor and Head of Computer Engineering Department. He has published more than 40 papers in international journal and conferences. He received best researcher award in International Scientist Award Conference held at Vishakhapatnam, India.

How to cite this paper: D P Gaikwad, "Intrusion Detection System Using Ensemble of Rule Learners and First Search Algorithm as Feature Selectors", International Journal of Computer Network and Information Security(IJCNIS), Vol.13, No.4, pp.26-34, 2021. DOI: 10.5815/ijcnis.2021.04.03