An Empirical Comparison and Feature Reduction Performance Analysis of Intrusion Detection

Upendra[1] and Yogendra Kumar Jain[2]

[1]Research Scholar M.Tech, Department of Computer Science & Engineering, Samrat Ashok Technological Institute, Vidisha, M.P., India
upendra.chaurasiya@gmail.com
[2]Head of Department, Computer Science & Engineering, Samrat Ashok Technological Institute, Vidisha, M.P., India

ABSTRACT

This paper reports on the empirical evaluation of five machine learning algorithm such as J48, BayesNet, OneR, NB and ZeroR using ten performance criteria: accuracy, precision, recall, F-Measure, incorrectly classified instances, kappa statistic, mean absolute error, root mean squared error, relative absolute error, root relative squared error. The aim of this paper is to find out which classifier is better in its performance for intrusion detection system. Machine Learning is one of the methods used in the intrusion detection system (IDS).Based on this study, it can be concluded that J48 decision tree is the most suitable associated algorithm than the other four algorithms. In this paper we compared the performance of Intrusion Detection System (IDS) Classifiers using seven feature reduction techniques.

KEYWORDS

Intrusion Detection, Machine Learning, BayesNet, C 4.5, NB, ZeroR, KDD 99

1. INTRODUCTION

Empirical studies indicate that feature reduction technique is capable of reducing the size of dataset [28]. Recently research on machine learning for intrusion detection has standard much attention in the computational intelligence community. In intrusion detection algorithm, immense strengths of audit data must be analyzed in order to conception new detection rules for increasing number of novel attacks in high speed network. Intrusion detection algorithm should consider the composite properties of attack behaviors to improve the detection speed and detection accuracy. Analyze the large volume of network dataset and the better performances of detection accuracy, intrusion detection become an important research field for machine learning. In this work we have presented J48 decision tree algorithm for intrusion detection based on machine learning. The Intrusion Detection System (IDS) is Process of monitoring the events occurring in a computer system or network and analyzing them for signs of possible incidents. IDS was first introduced in 1980 by James. P. Anderson [3] and then improved by D. Denning [4] in 1987.

DOI : 10.5121/ijctcm.2012.2101

Various paradigms namely Support Vector Machine [30], Neural Networks[31], K-means based clustering[32] have been applied to intrusion detection because it has the advantage of discovering useful knowledge that describes a user's or program's behavior. They are two basic approaches for Intrusion Detection techniques, i.e. Anomaly Detection and Misuse Detection (signature-based ID) [17]. Anomaly Detection is basically based on assumption that attacker behavior is different from normal user's behavior [1]. In this paper, we present the application of machine learning to intrusion detection. We analyse five learning algorithms (J48, BayesNet, OneR,NB and ZeroR) for the task of detecting intrusions and compare their relative performances. There is only available data set is KDD data set for the purpose of experiment for intrusion detection.KDD data set [2] contain 42 attributes. The classes in KDD99 [18] dataset can be categorized into five main classes (one normal class and four main intrusion classes: probe, Dos, U2R and R2L).Many feature reduction methods use information theory based metrics to measure the relevance of features.

2. RELATED WORK

Intrusion detection started in 1980's and since then a number of techniques have been introduced to build intrusion detection systems [12], [13], [14]. In 2007, Panda and Patra [10] determined a method using naive Bayes to detect signatures of specific attacks. They used KDD99 dataset for experiment, in the early 1980's, Stanford Research Institute (SRI) developed an Intrusion Detection Expert System (IDES) that monitors user behavior and detects suspicious events. Meng Jianliang [6] used the K Mean algorithm to cluster and analyze the data. He used the unsupervised learning technique for the intrusion detection. Mohammadreza Ektefa et al., [8] in 2010, compared C4.5 with SVM and the results revealed that C4.5 algorithms better than SVM in detecting network intrusions and false alarm rate. Zubair A.Baig et al. (2011) proposed An AODE-based Intrusion Detection System for Computer Networks. They suggested that the Naive Bayes (NB) does not accurately detect network intrusions [7]. In 2010, Hai Nguyen et al. [5] applied C4.5 and BayesNet for intrusion detection on KDD CUP'99 Dataset. Jiong Zhang and Mohammad Zulkernine [9] done the intrusion detection using the random forest algorithms in anomaly based NIDS. Cuixio Zhang, Guobing Zhang, Shanshan Sun [15] used the missed approach for the intrusion detection. He designed the mixed combining the anomaly detection and misuse detection in this model the anomaly detection module is built using unsupervised clustering method and the algorithm is an improved algorithm of K means clustering algorithm. The new algorithm learns the strong points from the k-means and improved relations trilateral triangle theorem. Gary Stein [11] applied the genetic algorithm and the decision tree algorithm for the intrusion detection. He used the genetic algorithm technique for the feature reduction. W.Lee et al. [29] propose a technique to measure the performance of an intrusion detection system by quantifying the benefits and costs of detection rules.

3. METHODOLOGICAL APPROACH

Decision tree technology is a common, intuitionist and fast classification method [21]. Its construction process is top-down, divide-and-rule. Essentially it is a greedy algorithm. Starting from root node, for each non-leaf node, firstly choose an attribute to test the sample set; Secondly divide training sample set into several sub-sample sets according to testing results, each sub-sample set constitutes a new leaf node; Thirdly repeat the above division process, until having reached specific end conditions. In the process of constructing decision tree, selecting testing

attribute and how to divide sample set are very crucial. Different decision tree algorithm uses different technology. In practice, because the size of training sample set is usually large, the branches and layers of generated tree are also more. In addition, abnormity and noise existed in training sample set will also cause some abnormal branches, so we need to prune decision tree. One of the greatest advantages of decision tree classification algorithm is that: It does not require users to know a lot of background knowledge in the learning process. As long as training samples can be expressed as the form of attribute-conclusion, you can use this algorithm to study. But decision tree technology also has a lot of deficiency, such as: When there are too many categories, classification accuracy is significantly reduced; It is difficult to find rules based on the combination of several variables. At present, there are a lot of decision algorithms, such as: ID3, SLIQ, CART, CHAID and so on. But J48 algorithm is the most representative and widely used. It is proposed by Quinlan in 1993.

A Naive Bayes classifier [19] is a simple probabilistic classifier based on applying Bayes' theorem (from Bayesian statistics) with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model". In simple terms, a naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. For example, a fruit may be considered to be an apple if it is red, round, and about 4" in diameter. Even if these features depend on each other or upon the existence of the other features, a naive Bayes classifier considers all of these properties to independently contribute to the probability that this fruit is an apple Depending on the precise nature of the probability model; naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without believing in Bayesian probability or using any Bayesian methods.

3.1 INFORMATION GAIN BY AN EXAMPLE DATA SET

The proposed feature reduction technique can be easily understood by the following example. To demonstrate efficiency of the proposed technique, we have used weather database [20] to calculate information gain.

TABLE I. WEATHER EXAMPLE DATASET

day	Outlook	Temperature	Humidity	Windy	Class:Play
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rainy	Mild	High	Weak	Yes
D5	Rainy	Cool	Normal	Weak	Yes
D6	Rainy	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes

D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rainy	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rainy	Mild	High	Strong	No

Table I. presents a training set, D, of class-labelled tuples randomly selected from the All Electronics weather database. In this example, each attribute is discrete-valued. The class label attribute, play compute, has two distinct values (namely, yes, no); therefore, there are two distinct classes (that is, m = 2).Let class C1 correspond to yes and class C2 correspond to no. There are nine tuples of class yes and five tuples of class no. A (root) node N is created for the tuples in D. We compute the information gain of each attribute. We first compute the expected information needed to classify a tuple in D:

Info (D)= - 9/14 log2 (9/14) –5/14 log2 (5/14)=0.940 bits....(1)

Next, we need to compute the expected information requirement for each attribute. Let's start with the attribute outlook.We need to look at the distribution of yes and no tuples for each category of outlook. For the outlook category sunny, there are two yes tuples and three no tuples. For the category overcast, there are four yes tuples and zero no tuples. For the category rainy, there are three yes tuples and two no tuples. Now we calculate the Info for an attribute Outlook. The expected information needed to classify a tuple in *D* if the tuples are partitioned according to outlook is:

Info outlook (D) = 5/14 × (- 2/5 log2 2/5 – 3/5 log2 3/5) + 4/14 ×

(-4/4 log2 4/4 -0/4 log2 0/4) + 5/14 × (-3/5 log2

3/5 – 2/5 log2 2/5)

= 0.694 bits..... (2)

Hence, the gain in information from such a partitioning would be equation (1) – (2)

Gain (outlook) = Info (D) – Info outlook (D)

= 0.940 – 0.694

= 0.246 bits

Similarly, we can compute Gain (temperature) = 0.029 bits,

Gain(humidity) = 0.151bits, and Gain (windy) = 0.048 bits

Using the method above for calculation of information gain, we calculate the info gain of the all the attribute of the KDD99 data set. The info gain of the all the attribute is given below in table I. In our proposed technique we are using the KDD99 dataset with these selected features and train and test the algorithm. For the testing we are using the 10 fold cross validation. Features selection techniques have been employed by Researchers. In other domain to extract important features.Skurichina and Duin [16] suggested that predictive accuracy can be improved by combining feature sets.

TABLE II. A SAMPLE CONFUSION MATRIX

	Predicted Class Positive	**Predicted Class Negative**
Actual Class Positive	a	b
Actual Class Negative	c	d

In this confusion matrix, the value a is called a true positive and the value d is called a true negative. The value b is referred to as a false negative and c is known as false positive.

3.2 True Positive Rate,Fale Positive Rate

In the context of intrusion detection, a true positive is an instance which is normal and is also classified as normal by the intrusion detector. For a good IDS TP rate should be high. False positive means no attack but IDS detect the attack. For a good IDS FP should be low.

3.3 Accuracy

This is the most basic measure of the performance of a learning method. This measure determines the percentage of correctly classified instances. From the confusion matrix, we can say that:

$$\text{Accuracy} = \frac{a+d}{a+b+c+d}$$

This metric gives the number of instances from the dataset which are classified correctly i.e. the ratio of true positives and true negatives to the total number of instances.

1 **Mean Absolute Error:** In statistics, the mean absolute error (MAE) is a quantity used to measure how close forecasts or predictions are to the eventual outcomes. The mean absolute error (MAE) is given by

$$\frac{1}{n}\sum_{i=1}^{n} |f_i - y_i| = \frac{1}{n}\sum_{i=1}^{n} |e_i|$$

The mean absolute error is an average of the absolute errors $e_i = f_i - y_i$, where f_i is the prediction and y_i the true value.

2 **Root Mean Squared Error (RMSE):** or Root-Mean-Square deviation (RMSD). It is a frequently-used measure of the differences between values predicted by a model or an estimator and the values actually observed from the thing being modeled or estimated

$$\sqrt{\frac{\sum \left(f_{(x_i)} - y_i\right)2}{n}}$$

3 **Relative Absolute Error:** The relative absolute error Ei of an individual program i is evaluated by the equation:

$$E_{i=} \frac{\sum_{j=1}^{n} \left| P_{(ij)} - T_j \right|}{\sum_{j=1}^{n} \left| T_j - \overline{T} \right|}$$

where P(ij) is the value predicted by the individual program i for sample case j (out of n sample cases); Tj is the target value for sample case j; and $\overline{T}$ is given by the formula:

$$\overline{T} = \frac{1}{n} \sum_{j=1}^{n} T_j$$

For a perfect fit, the numerator is equal to 0 and Ei = 0. So, the Ei index ranges from 0 to infinity, with 0 corresponding to the ideal. Root Relative Squared error: The root relative squared error Ei of an individual program i is evaluated by the equation:

$$\mathbf{E_i} = \sqrt{\frac{\sum_{j=1}^{n} \left(P_{(ij)} - T_j\right) 2}{\sum_{j=1}^{n} \left(T_j - \overline{T}\right) 2}}$$

3.4 J48

Decision tree J48 developed by Johan Ross Quinlan [25]. C4.5 is an extension of Quinlan's earlier the Interactive Dichotomizer3 (ID3) Algorithm. J48 builds decision trees from a set of labelled training data using the concept of information entropy.The Decision tree is a classifier expressed as a recursive partition of the instance space, consists of nodes that form a rooted tree, meaning it is a directed tree with a node called a root that has no incoming edges referred to as an internal or test node. All other nodes are called leaves (also known as terminal or decision nodes). Decision trees [22], [23], [24] are one of the most commonly classification methods used in supervised learning approaches.

3.5 KDD Cup 1999 Intrusion Detection Data

the KDD 99 intrusion detection benchmark in the International Knowledge Discovery and Data Mining.The data used in this paper are those proposed in the KDD'99 for intrusion detection [2] which are generally used for benchmarking intrusion detection problems and subversion of DARPA (Defense Advanced Research Projects Agency) 1998 dataset. The 1999 KDDCup data set [18] contains a set of records that represent connections to a military computer network where there have been multiple intrusions and attacks. KDD dataset contains symbolic as well as continuous features. attacks fall into four main categories DoS (Denial of Service), R2L (Remote

to Local), U2R (User to Root) and Probe. This data set was obtained from the UCI KDD archive [26]. The training data set has 65535 data instances with 32 continuous attributes and 9 categorical attributes and dataset includes a set of 41 features. The testing data set is smaller and contains several new intrusions that were not present in the training data set. KDD dataset is divided into training and testing record sets.

4. PERFORMANCE EVALUATION AND RESULT

The Tables III, IV and V Shows the performance of five classification methods based on correctly classified Instances, incorrectly classified Instances ,Kappa statistic, Mean absolute error, Root Mean Squared Error, Relative Absolute Error,Root Relative Squared error,Coverage of cases(0.95 level) and Time taken to build the models respectively. The comparison is performed for 41 and 7 attributes. The five classifier models on the dataset are built and tested by means of 10-fold cross-validation. The Java Heap size was set to 1024 MB for WEKA 3.6.2, the simulation platform is an Intel™ Core i3-2100 processor system with 3 GB RAM under Microsoft Windows XP™ Service Pack-2 operating system, 3.10 GHz with 500 GB memory. the mapped & normalized dataset is further discretized to obtain discrete values for continuous features using WEKA [27].

TABLE III. COMPARISION OF THE RESULTS FOR J48, BAYESNET, ONER,NB AND ZEROR WITH ALL ATTRIBUTE

Parameter	Classifier				
	J48	**BayesNet**	**OneR**	**NB**	**ZeroR**
Correctly Classified Instances	99.5594%	96.5624%	96.18%	89.591%	53.3%
Incorrectly Classified Instances	0.4406 %	3.4376 %	3.810%	10.408%	46.6%
Kappa statistic	0.9911	0.9307	0.923	0.7906	0
Mean absolute error	0.0064	0.0378	0.038	0.1034	0.49
Root mean squared error	0.0651	0.175	0.195	0.3152	0.49
Relative absolute error	1.2854 %	7.6037 %	7.656%	20.781%	100%
Root relative squared error	13.059 %	35.0792%	39.13%	63.189%	100%
Coverage of cases(0.95 level)	99.6229%	97.781 %	96.18%	90.9654 %	54.1%

TABLE IV. COMPARISION OF THE RESULTS FOR J48, BAYESNET, ONER,NB AND ZEROR WITH 7 ATTRIBUTE

Parameter	Classifier				
	J48	**BayesNet**	**OneR**	**NB**	**ZeroR**
Correctly Classified Instances	99.890%	99.243%	97.6761 %	93.569%	59.9649 %
Incorrectly Classified Instances	0.1099 %	0.7568 %	2.3239 %	6.4302 %	40.0351 %
Kappa statistic	0.9978	0.9846	0.9529	0.8708	0
Mean absolute error	0.0007	0.0032	0.0093	0.0266	0.1963
Root mean squared error	0.0206	0.0488	0.0964	0.1538	0.3133
Relative absolute error	0.3358 %	1.6253 %	4.7346 %	13.556%	100 %
Root relative squared error	6.5596 %	15.5687%	30.7728%	49.080%	100 %
Coverage of cases (0.95 level)	99.91 %	99.6414%	97.6761%	94.700%	89.6221 %

From table III and IV.It is clear that The J48 gave the best performance.

Now we compare the result of the J48, BayesNet, OneR, NB and ZeroR algorithms. Firstly we compare the result after run the algorithm with all attribute. Secondly we compare the result after run the algorithm with reduced 7 attribute than only we conclude that which one algorithm is good best for the intrusion detection.

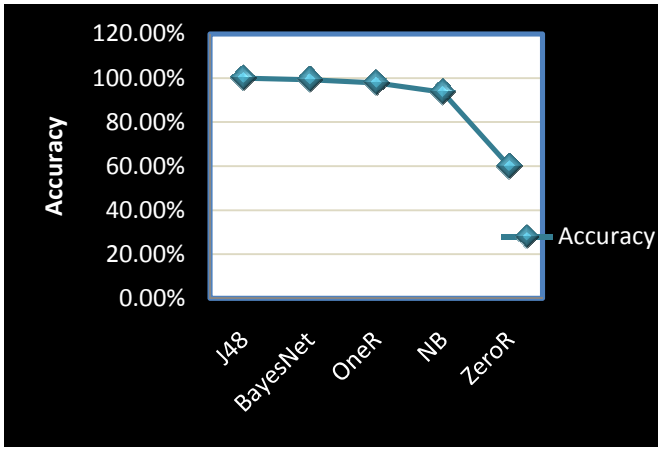

Figure 1. Comparison of accuracy for J48, BayesNet, OneR, NB and ZeroR.

From above figure 1. It is clear that information gain feature reduction method gives the better accuracy which is desirable for good Intrusion Detection System. Especially in the case of J48 accuracy is 99.9%.

Now we compare the TPR of the J48, BayesNet, OneR, NB and ZeroR algorithm with all attribute and with selected 7 attributes.

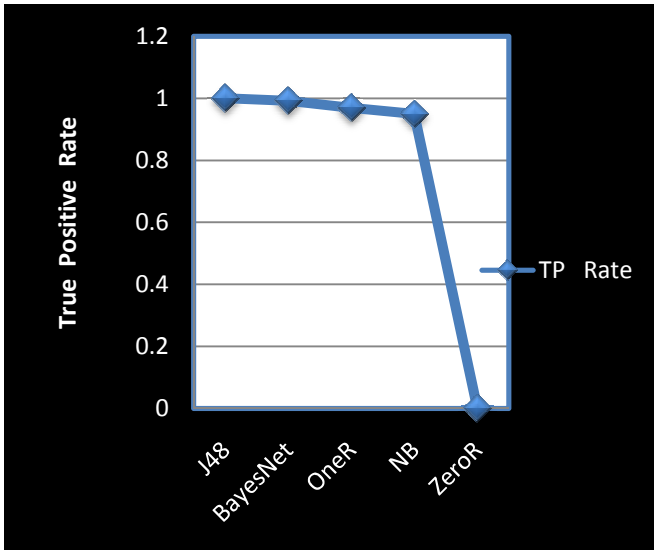

Figure 2. TPR comparison of J48, BayesNet, OneR,NB and ZeroR

For a good IDS TP Rate should be high. Above figure 2. Shows that TP Rate of the J48 algorithm is higher when we reduce the feature of the data set using information gain. Especially in the case of J48 TPR is 1

Figure 2 and Figure 3 above shows the TPR (True Positive Rate) and FPR (False Positive Rate) of the J48, BayesNet, OneR,NB and ZeroR algorithm when run with the all attributes of the data set. Figure 2. Shows that TPR of the J48 is higher than the remaining four algorithms which is desirable. Figure 3. Also shows that FPR of the J48 is almost zero which is desirable for a good intrusion detection algorithm.

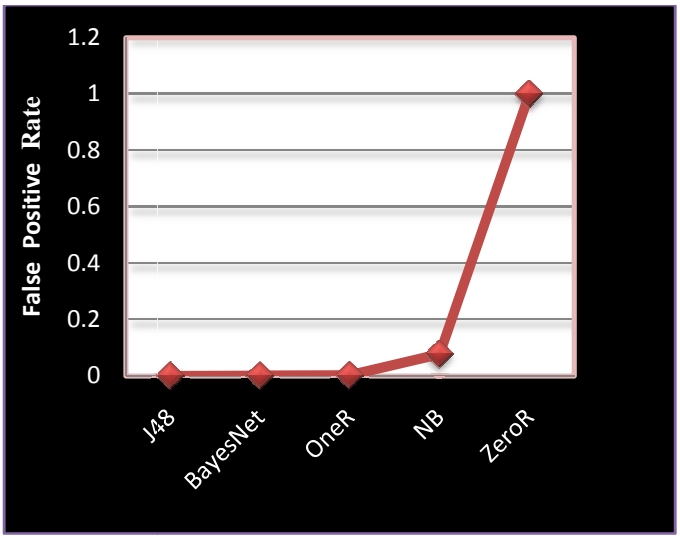

Figure 3. FPR comparison of J48, BayesNet, OneR,NB and ZeroR

For a good IDS FPR should be low. Above figure 3. Shows that FPR of the J48 algorithm is lower when we reduce the feature of the data set using information gain. Especially in the case of J48 FPR is 0.In the case of BayesNet, OneR,NB and ZeroR algorithm FPR of the greater than 0. From above figures 1, 2 and 3 it is clear that J48 algorithm Accuracy, TPR and FPR is better than other four algorithms. So we can say that reduction of the feature using information gain is better technique.

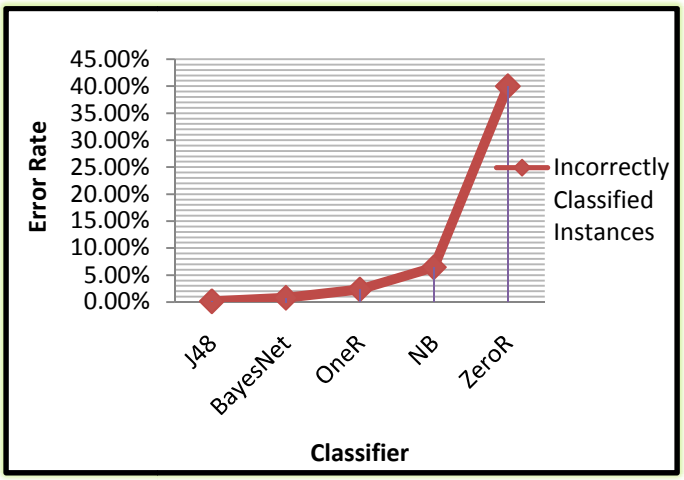

Figure 4. Error rate comparison of J48, BayesNet, OneR,NB and ZeroR

The experimental results shows that Performance Evaluation of five classification models, J48 have much better performance than other four methods and it is also observed that the overall performance of J48 classification has increased their performance using feature reduction method a notable improvement in their classification, means the classification accuracy increases better after feature selection.

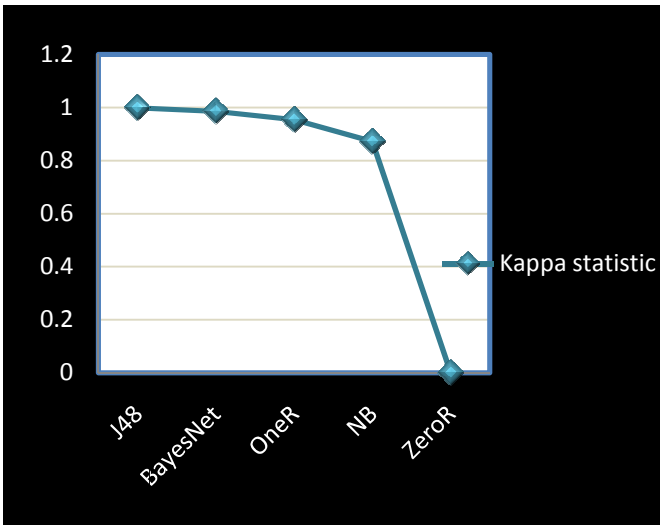

Figure 5. Kapa statistic comparison of J48, BayesNet, OneR,NB and ZeroR

In this paper, the performance of four well known data mining classifier algorithms namely J48, BayesNet, OneR,Naïve Bayes and ZeroR are evaluated based on the 10-fold cross validation test, Experimental results using the KDD CUP99 IDS data set demonstrate that while J48 is one of the most effective inductive learning algorithms, decision trees are more interesting as far as the detection of new attacks is concerned.

From above figure 4 and 5, it is clear from our evaluation that J48 has relatively detection rates and Kappa statistic; Incorrectly Classified Instances are compared and found that J48 is excellent in performance than other classifiers.

TABLE V. COMPARISION OF THE RESULTS FOR J48, BAYESNET, ONER, NB AND ZEROR WITH 7 ATTRIBUTE

Feature Used	Classifier	Accuracy	normal		dos		probe		r 2 l		u 2 r	
			TP Rate	FP Rate	TP Rate	FP Rate	TP Rate	FP Rate	TP Rate	FP Rate	TP Rate	FP Rate
7	J48	99.8901%	1	0.002	0.999	0	0.966	0	0.779	0	0.2	0
7	BayesNet	99.2432%	0.992	0.002	0.999	0.005	0.845	0.003	0.824	0	0.4	0
7	OneR	97.6761%	0.969	0.002	0.999	0.035	0.682	0.001	0.794	0	0	0
7	NB	93.5698%	0.949	0.077	0.921	0.021	0.777	0.008	0.765	0.006	0.6	0.007
7	ZeroR	59.9649%	1	1	0	0	0	0	0	0	0	0

5. CONCLUSIONS

In this paper we compared the performance measure of five machine learning classifiers such as Decision tree J48,BayesNet,OneR,Naive Bayes and ZeroR.The results are compared and found that J48 is excellent in performance than other classifiers with respect to accuracy.

we reduced the features of the data set using information gain of the attributes. This study is approached to discover the best classification algorithm for the applications of machine learning to intrusion detection. Our simulation results show that, in general, the J48 has the highest classification accuracy performance with the lowest error rate. On the other hand, we also found that drastically decreased in learning time of the algorithm and increase in accuracy and TPR. Comparison shows that reduction of the feature using information gain technique is suitable for the feature reduction. Using Weka, we analysed five algorithms towards their suitability for detecting intrusions from KDD99 dataset. We showed that machine learning can be effectively applied to detect novel intrusions and focused on anomaly detection. The five learning algorithms J48, BayesNet, OneR, Naïve Bayes and ZeroR were compared at the task of detecting intrusions. J48 with an accuracy rate of approximately 99% was found to perform much better at detecting intrusions than BayesNet, OneR, NB and ZeroR Based on the experiments done in the paper and their Corresponding results, we can state the following: J48 classifier shows better performance for all the classes (Normal, DOS, R2L, U2R, Prob)

REFERENCES

[1] Lida Rashidi,Sattar Hashem and Ali Hamzeh, "Anomaly detection in categorical datasets using bayesian networks," AICI'11 Proceedings of the Third International Conference on Artificial Intelligence and Computational Intelligence, Volume Part II, Springer-Verlag, Berlin ,Heidelberg, 2011, pp.610–619.

[2] Knowledge Discovery in Databases DARPA archive. Task Description ,KDDCUP 1999 DataSet, http://www.kdd.ics.uci.edu/databases/kddcup99/task.html

[3] James P. Anderson, "Computer Security Threat Monitoring and Surveillance," Technical Report, James P.Anderson Co.,Fort Washington, Pennsylvania, USA , pp.98–17, April 1980.

[4] Dorothy E. Denning,"An Intrusion Detection Model," IEEE Transaction on Software Engineering (TSE), volume–13, No.2, pp.222–232,February 1987.

[5] Hai Nguyen, Katrin Franke and Slobodan Petrovi'c, "Improving Effectiveness of Intrusion Detection by Correlation Feature Selection," International Conference on Availability, Reliability and Security, pp. 17–24, IEEE 2010.

[6] Meng Jianliang, Shang Haikun, "The application on intrusion detection based on K-Means cluster algorithm," International Forum on Information Technology and Application, 2009.

[7] Zubair A. Baig, Abdulrhman S. Shaheen, and Radwan AbdelAal, "An AODE-based Intrusion Detection System for Computer Networks," pp. 28–35, IEEE 2011.

[8] Mohammadreza Ektefa, Sara Memar, Fatimah Sidi, Lilly Suriani Affendey,"Intrusion Detection Using Data Mining Techniques," Proceedings Of IEEE International Conference on Information Retrieval & Knowledge Management,Exploring Invisible World, CAMP'10,2010, pp.200-203.

[9] Jiong Zhang and Mohhammad Zulkernine, "Anomaly based Network Intrusion detection with unsupervised outlier detection," School of Computing Queen's University, Kingston, Ontario, Canada. IEEE International Conference ICC 2006, Volume-9, pp. 2388-2393, 11-15 June 2006.

[10] M. Panda, and M. R. Patra, "Network intrusion detection using naive Bayes," International Journal of Computer Science and Network Security (IJCSNS), Volume -7, No. 12, December 2007, pp. 258–263.

[11] Gary Stein, Bing Chen," Decision Tree Classifier for network intrusion detection with GA based feature selection," University of Central Florida. ACM-SE 43, Proceedings of 43rd annual Southeast regional Conference. Volume-2,2005.ACM,New York,USA.

[12] Shai Rubin, Somesh Jha, and Barton P. Miller, "Protomatching Network Traffic for High Throughput Network Intrusion Detection," In Proceedings of the Proceedings of the 13th ACM conference on Computer and Communications Security, pages 47–58. ACM, 2006.

[13] Marco Cova, Davide Balzarotti, Viktoria Felmetsger, and Giovanni Vigna. Swaddler, "An Approach for the Anomaly-Based Detection," Symposium on Recent Advances in Intrusion Detection(RAID), pages 63–86. Springer, 2007.

[14] Pavel Kachurka, Vladimir Golovko, "Neural Network Approach to Real-Time Network Intrusion Detection and Recognition," The 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Application,15-17 September 2011, pp. 393–397, IEEE 2011.

[15] Cuixiao Zhang, Guobing Zhang, Shanshan Sen., "A mixed unsupervised clustering based Intrusion detection model," Third International Conference on Genetic and Evolutionary Computing, 2009.

[16] M. Skurichina and R.P.W. Duin, "Combining feature subsets in feature selection," Lecture Notes in Computer Science, Vol. 3541,pp-165-175,Springer Verlag, Berlin, 2005.

[17] LI Min and Wang Dongliang, "Anomaly Intrusion Detection Based on SOM," IEEE WASE International Conference on Information Engineering, IEEE Computer Society, 2009, pp. 40-44.

[18] Mahbod Tavallaee,Ebrahim Bagheri,Wei Lu, and Ali A.Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," Proceedings of the 2009 IEEE Symposium on Computational Intelligence in Security and Defense Application(CISDA 2009),IEEE 2009.

[19] R.Dogaru,"A modified Naive Bayes classifier for efficient implementations in embedded systems," Signals Circuits and Systems (ISSCS), IEEE 10th International Symposium on Lasi, June 30,2011-July 1, 2011, pp.1–4.

[20] Jiawei Han and Micheline kamber, "Data Mining Concepts and Techniques,"Second Edition,University of Illinois at Urbana-Champaign The Morgan Kaufmann Series in Data Management Systems,Elsevier 2007.

[21] Juan Wang, Qiren Yang,Dasen Ren, "An intrusion detection algorithm based on decision tree Technology,"Asia-Pacific Conference on Information Processing,APCIP 2009 ,Shenzhen,IEEE 18-19 July 2009. pp. 333–335.

[22] John Ross Quinlan, (1992) "Learning with Continuous Classes",5th Australian Joint Conference on Artificial Intelligence, Singapore, pp.343-348.

[23] Kamarulrifin Abd Jalil and Mohamad Noorman Masrek,"Comparison of Machine learning Algorithms Performance in Detecting Network Intrusion", IEEE 2010 International Conference on Networking and Information Technology,pp.221-226.

[24] G.Meera Gandhi,Kumaravel Appavoo,S.K. Srivatsa, "Effective Network Intrusion Detection using Classifiers Decision Trees and Decision rules", International Journal of Advanced Networking and Applications, Volume: 2, Issue: 3, pp: 686-692, 2010.

[25] John Ross Quinlan, (1993) "C4.5: Programs for Machine Learning", Morgan Kaufmann Publishers, San Mateo,CA.1993.

[26] C. Blake, E. Keogh and C. Merz, UCI repository of machine learning databases,1998.

[27] H. Witten, E. Frank, Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, San Francisco, Second Edition, 2005.

[28] A.K. Jain, D. Zongker, "Feature Selection: Evaluation, Application, and Small Sample Performance", IEEE Trans. Pattern Analysis and Machine Intelligence, 19(2) pp-153-158, 1997.

[29] W. Lee, J. Cabrera, A. Thomas, N. Balwalli, S. Saluja, and Y. Zhang, "Performance adaptation in real-time intrusion detection systems," in Recent Advances in Intrusion Detection. Springer, RAID, 2002.

[30] Jiaqi Jiang, Ru Li,Tianhong Zheng,Feiqin Su, Haicheng Li, "A new intrusion detection system using Class and Sample Weighted C-Support Vector Machine", Third International Conference on Communications and Mobile Computing, IEEE Computer Society,2011,pp-51-54.

[31] E. T. Ferreira, G. A. Carrijo, R. de Oliveira and N. V. S. Araujo, "Intrusion Detection System with Wavelet and Neural Artifical Network Approach for Network Computers," IEEE Latin America Transactions, Vol. 9, No. 5, September 2011,pp-832-837.

[32] Yang Zhong, Hirohumi Yamaki, Hiroki Takakura, "A Grid-Based Clustering for Low-Overhead Anomaly Intrusion Detection," IEEE 2011, pp-17-24.