An Incremental-Learning Method for Supervised Anomaly Detection by Cascading Service Classifier and ITI Decision Tree Methods

Wei-Yi Yu and Hahn-Ming Lee

Department of Computer Science and Information Engineering National Taiwan University of Science and Technology Taipei, 106, Taiwan, R.O.C
{M9315912,hmlee}@mail.ntust.edu.tw

Abstract. In this paper, the incremental learning method to cascade Service Classifier and ITI (incremental tree inducer) methods for supervised anomaly detection, called "SC+ITI", is proposed for classifying anomalous and normal instances in a computer network. Since the ITI method can not handle new instances with new service value, the SC+ITI cascading method is proposed to avoid this. Two steps are in SC+ITI cascading methods. First, the Service Classifier method partitions the training instances into n service clusters according to different service value. Second, in order to avoid handling instances with new service value, the ITI method is trained with instances with the same service value in the cluster. In 2007, Gaddam et al. showed KMeans+ID3 cascading method which mitigates two problems 1) the Forced Assignment problem and 2) the Class Dominance problem. His method with Nearest Neighbor (NN) combination rule outperforms the other three methods (i.e., K-Means, ID3 and KMeans+ID3 with Nearest Consensus rule) over the 1998 MIT-DARPA data set. Since the KDD'99 data set was also extracted from the 1998 MIT-DARPA data set, Nearest Neighbor combination rule within K-Means+ITI and SOM+ITI cascading methods is used in our experiments. We compare the performance of SC+ITI with the K-Means, SOM, ITI, K-Means+ITI and SOM+ITI methods in terms of the Detection Rate and False Positive Rate (FPR) over the KDD'99 data set. The results show that the ITI method have better performance than the K-Means, SOM, K-Means+ITI and SOM+ITI methods in terms of the overall Detection Rate. Our method, the Service Classifier and ITI cascading method outperforms the ITI method in terms of the Detection Rate and FPR and shows better Detection Rate as compared to other methods. Like the ITI method, our method also provides the additional options of handling missing values data and incremental learning.

Keywords: anomaly detection system (ADS), K-Means clustering, Kohonens' self-organizing maps (SOM), ITI (incremental tree inducer), KDD'99.

1 Introduction

The intrusion detection systems (IDS) can be commonly classified into two categories according to the modeling methods used. One is misuse detection or rule-based

H. Chen et al. (Eds.): PAISI 2009, LNCS 5477, pp. 155–160, 2009.

method that uses stored signatures of known intrusion instances to detect a malicious attack with low false-positive error. However this technique is hard to detect novel attacks and variants of known attacks whose rules are not stored. The other one is anomaly detection method that analyzes large amount of data to model a normal profile and attempts to identify patterns of activity that deviate from the defined profile. Although it remedies the problem of detecting novel attacks, the drawback of this technique is that normal behavior deviating from the defined profile may be labeled as an intrusion, resulting in high false-positive error.

In this paper, the incremental learning Service Classifier and ITI (SC+ITI) cascading method is proposed for classifying anomalous and normal instances. Since the ITI method can not handle new instances with new service value, the SC+ITI cascading method guarantees the ITI method is trained with instances with same service value. The SC+ITI cascading method has three phases which are described in section 2.2.

In 2007, Gaddam et al. [3] presented the novel method cascading the clustering method (K-Means) [4] with the decision tree (ID3) learning method [8] called "KMeans+ID3" which alleviates two problems in the cluster: 1) the Forced Assignment problem and 2) the Class Dominance problem. The first problem, Forced Assignment arises when similar anomaly and normal instances are assigned to the cluster. The second problem, Class Dominance arises in the cluster when subset of training data in the cluster contains an amount of instances from one particular class and few instances from the remaining classes. Since Gaddam et al. presented KMeans+ID3 cascading method which mitigates two problems: 1) the Forced Assignment problem and 2) the Class Dominance problem, SC+ITI cascading method is evaluated with the performance of K-Means, SOM , ITI, K-Menas+ITI, SOM+ITI methods using two measures (Detection Rate and False Positive Rate) in this paper.

The rest of the paper is organized as follow: In Section 2, we briefly discuss the K-Means, SOM, ITI, K-Means+ITI, SOM+ITI and SC+ITI learning-based anomaly detection methods. In Section 3, we discuss experiments, data sets and measures. In Section 4, we discuss the results of above six methods. We conclude our work and propose future work in section 5.

2 Methodologies for Anomaly Detection

Since anomaly detection with the K-Means [4], SOM [5], K-Means+ITI, SOM+ITI methods were quite similarly discussed in [3], these methods for anomaly detection will not be described in this section. In the section, we only briefly discuss the ITI 10 and SC+ITI methods for supervised anomaly detection. Nearest Neighbor combination rule within the K-Means+ITI and SOM+ITI cascading methods is adopted here instead of Nearest-Consensus rule because of two reasons: 1) Gaddam's K-Means+ID3 cascading method with Nearest Neighbor (NN) combination rule [3] outperform the other proposed methods over the 1998 MIT-DARPA data set 2) Nearest Consensus of the K-Means and ID3 cascading method probably doesn't exist if user defined parameter f is too small.

2.1 Anomaly Detection with the ITI Decision Tree

After trained with instances, the ITI method will build the binary decision tree. For detecting anomalies, ITI method outputs binary classification of "0" to indicate normal and "1" to indicate anomaly class. This is quite similarly to Gaddam's anomaly detection with ID3 which is described in [3]. We choose the ITI method because of four reasons: 1) inconsistent training instances 2) missing values data 3) incremental learning 4) numeric variables. Inconsistent training instances, missing values and incremental learning had been discussed in [10]. Numeric variables and limitations are discussed as follows:

2.1.1 Numeric Variables

For handling numeric variables, there are differences between the Gaddam's ID3 and ITI decision tree methods. In Gaddam's ID3 method, the training space was discretized into n equal-width intervals where n is predefined. Fayyad's splitting point selection [2] for numeric variables is adopted here in the ITI algorithm to deal with this problem.

2.1.2 Limitations

Two limitations were discussed in [6]. First, the ITI method needs to have "sufficient" training instances that cover as much variation of the normal behavior as possible. Second, this method can not handle new instances with new (i.e., unseen service) class labels. During the incremental-learning phase, this method can not incorporate instances with new service value into the binary decision tree and update the rules incrementally. However, at the testing phase, this method can be tested with these instances. Instances with new service value for the test at the decision node of service attribute will be passed down the false branch. Details were described in [6].

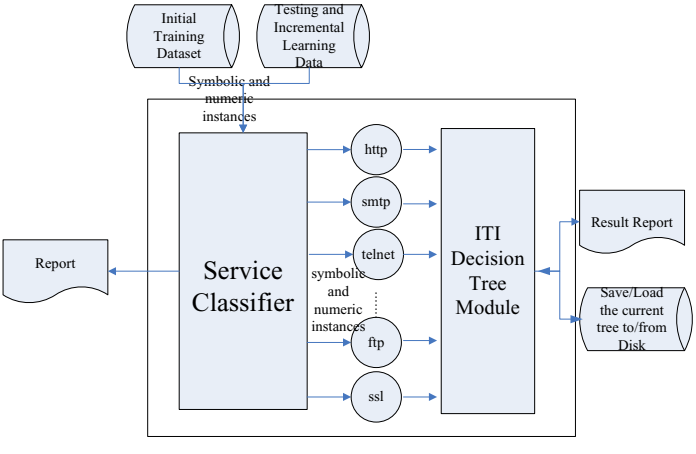

Fig. 1. The Service Classifier and ITI cascading method for Anomaly Detection

2.2 Anomaly Detection Using Service Classifier (SC) and ITI

Since the ITI method can not handle new instances with (i.e., unseen service) class labels, the service classifier and ITI cascading method is proposed to avoid this. Since Service Classifier ensures the ITI method is trained with instances with only one service value, the ITI method is trained with instances incrementally without the attribute service. The cascading method has three phases: 1) training, 2) testing and 3) incremental learning. During first training phase, the service classifier method is first applied to partition the training data set into m disjoint clusters according to different service. Then, the ITI method is trained with the instances in each cluster. The Service Classifier method ensures that each training instance is associated with only cluster. The testing phase, we will find the cluster to which the test instances are belong. Then, the ITI method is tested with the instances. At the last incremental learning phase, the Service Classifier and ITI cascading method will not be re-trained again and uses incremental learning data to train the existing ITI binary tree. The architecture of the Service Classifier and ITI cascading method is described in Fig. 1.

Two limitations are discussed. First, the cascading method needs to have "sufficient" training instances that cover as much variation of the normal behavior as possible. Second, the cascading method can not be tested with instances with new service value at the testing phase.

3 Experiments, Data Set and Measures

3.1 Experimental Setup and Data Set

SOM_PAK [11] and ITI [10] packages are adopted to evaluate their performance here. In our experiments, the k value of the K-Means method was set to 10, m*n value of the SOM method were set to 5*5 and 41 features of KDD'99 data set were all selected here. Neighborhood Parameters of SOM are Gaussian and Hexagonal.

Although a critique of the 1998 and 1999 DARPA IDS evaluations was discussed in [7], the KDD'99 data set [12] is commonly used for comparing the performance of IDSs. Four reasons to choose the KDD'99 data set were discussed in [9]. Other reason we choose this data set is noisy instances (ex: inconsistent training instances, error service value) occur in the KDD'99 data set. That represents real data in the reality. In our experiments, we simulate Sarasamma's training data set which was described in [9]. 169,000 instances from the "10% KDD" data set and 311,029 instances from "Corrected KDD" (Test Set) data set were used for training and testing respectively. The training and test set consist of 22 and 39 attack types respectively which fall into four main categories: Denial of Service (DOS), Probe, Remote to User (R2L), and User to Root (U2R).

3.2 Performance Measures

Anomaly intrusion detection is a two-class classification problem. For each single prediction, there are four possible outcomes. The true-positives and true-negatives are correct classifications. A false-positive occurs when IDS/ADS classifies an instance as an anomaly when it is a normal instance. Measures such as False Positive Rate,

Detection Rate, Accuracy, Precision and F-measure are defined in [1]. Other measures to compute the Area under an ROC (Receiving Operating Characteristic) curve, called AUC, mentioned in [1]. In our experiments, two measures (Detection Rate and FPR) are used.

4 Results

In this section, we present the results of K-Means, SOM, ITI, K-Means+ITI, SOM+ITI and SC+ITI methods over the KDD'99 data set. Table 1 summarizes the Detection Rate of these methods for five categories. The last two rows in Table 1 represent the Overall Detection Rate and FPR of these methods individually. The ITI and SC+ITI methods have better performance than K-Means, SOM, K-Means+ITI and SOM+ITI methods in terms of Detection Rate on U2R, R2L and PROBE attacks. The overall Detection Rate of the ITI and SC+ITI methods is better than other four methods, but the overall FPR of these methods is less than that of the ITI, and SC+ITI methods. In Table 1, we notice that:

- The overall Detection Rate of the K-Means, SOM and cascading K-Means+ITI and SOM+ITI methods is 89.95%, 85.97%, 91.31%, 91.07% respectively. The overall FPR of the K-Means, SOM and cascading K-Means+ITI and SOM+ITI methods is 1.29%, 1.49%, 0.81%, 0.73% individually. Since cascading methods mitigate the Class Dominance and Forced Assignment problems, cascading methods outperform the individual clustering methods in terms of the overall Detection Rate and FPR.

Table 1. Detection rate of attack catgory, overall detection rate and false positive rate of the methods

Method / Category(Count)	KMeans	SOM	ITI	KMeans +ITI	SOM +ITI	SC+ITI
U2R(228)	55.70%	20.61%	75.44%	27.63%	42.98%	61.40%
R2L(16,189)	0.10%	0.03%	20.61%	3.96%	4.60%	21.22%
PROBE(4,166)	73.91%	71.80%	92.70%	85.14%	84.52%	95.15%
NORMAL(60,593)	98.71%	98.50%	98.14%	99.19%	99.26%	98.20%
DOS(229,853)	96.60%	92.35%	97.44%	97.64%	97.33%	97.65%
Overall Detection Rate	89.95%	85.97%	92.38%	91.31%	91.07%	92.63%
False Positive Rate	1.29%	1.49%	1.86%	0.81%	0.73%	1.80%

5 Conclusion and Future Work

In the Table 1, K-Means+ITI and SOM+ITI cascading methods outperform the individual K-Means and SOM clustering methods in terms of the overall Detection Rate and FPR respectively because they alleviate two problems: 1) the Forced Assignment problem and 2) the Class Dominance problem. The ITI method has better performance than these above four methods in terms of the overall Detection Rate. The SC+ITI cascading method shows better overall Detection Rate and FPR as compared

to the ITI method. We conclude our work. For detecting anomalies, the incremental-learning SC+ITI cascading method shows better Detection Rate as compared to other methods and provides the additional options of handling missing values data and incremental learning.

Our future work includes 1) comparing performance of Gaddam's K-Means+ID3, 2) results of the methods (ex: ITI) tested with instances with new service value in terms of the Detection Rate and FPR, 3) statistical evaluation, 4) comparing performance of the different versions of ID3 or other decision trees, and 5) comparing performance of the incremental learning or multi-level classifier and decision tree (ex: ID4) cascading methods.

Acknowledgments. This work was supported in part by the National Digital Archive Program-Research & Development of Technology Division (NDAP-R&DTD), the National Science Council of Taiwan under grants NSC 95-2422-H-001-007, NSC 95-2422-H-001-024, and also by the Taiwan Information Security Center (TWISC), the National Science Council under grants NSC 95-2218-E-001-001, and NSC 95-2218-E-011-015.

References

1. Fawcett, T.: An introduction to ROC analysis. Pattern Recognition Letters 27(8), 861–874 (2006)
2. Fayyad, U.M., Irani, K.B.: On the handling of continuous-valued attributes in decision tree generation. Machine Learning 8(1), 87–102 (1992)
3. Gaddam, S.R., Phoha, V.V., Balagani, K.S.: K-Means+ID3: A novel method for supervised anomaly detection by cascading k-Means clustering and ID3 decision tree learning methods. IEEE Transactions on Knowledge and Data Engineering 19(3), 345–354
4. Hartigan, J.A., Wong, M.A.: A K-Means clustering algorithm. Applied Statistics 28(1), 100–108 (1979)
5. Kohonen, T.: The self-organizing map. Neurocomputing 21(1-3), 1–6 (1998)
6. Lee, W., Stolfo, S.J., Mok, K.W.: Adaptive Intrusion Detection: A Data Mining Approach. Artificial Intelligence Review 14(6), 533–567 (2000)
7. McHugh, J.: Testing intrusion detection systems: A critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by Lincoln laboratory. ACM Transactions on Information and System Security 3(4), 262–294 (2000)
8. Quinlan, J.R.: Induction of decision trees. Machine Learning 1(1), 81–106 (1986)
9. Sarasamma, S.T., Zhu, Q.A.: Min-Max Hyperellipsoidal Clustering for Anomaly Detection in Network Security. IEEE Transactions on systems, man, and cybernetics-part B: Cybernetics 36(4), 887–901 (2006)
10. Utgoff, P.E., Berkman, N.C., Clouse, J.A.: Decision Tree Induction Based on Efficient Tree Restructuring. Machine Learning 29, 5–44 (1997)
11. Kohonen, T., Hynninen, J., Kangas, J., Laaksonen, J.: SOM_PAK: The Self-Organizing Map Porgram Package, `http://www.cis.hut.fi/research/som_lvq_pak.shtml`
12. Stolfo, S., et al.: The Third International Knowledge Discovery and Data Mining Tools Competition (2002), `http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html`