IJO-SCIENCE
INTERNATIONAL JOURNAL ONLINE OF SCIENCE
ISSN: 2455-0108

Ensemble Classification for Intrusion Detection

Shweta Sharma[1], Prof. Aishwarya Mishra[2]
[1]*Department of Computer Science & Engineering, ICOT, RGPV University, Bhopal, INDIA*
[2]*Associate Professor, Department of Computer Science & Engineering, ICOT, RGPV University, Bhopal, INDIA*
[1]shwetasharma772@gmail.com
[2]mishra.aishy@gmail.com

***Abstract*— Security of network systems is becoming an important issue, as more and sensitive information is being stored and manipulated online. Therefore, it is essential to find an effective way to protect it. Mining approach can play very important role in developing an intrusion detection system. This paper presents various data mining Classification techniques applied on intrusion detection systems for the effective identification of both known and unknown patterns of attacks, to develop secure information systems.**

***Keywords*— Intrusion Detection, Data mining, Classification, Decision tree, KNN, Support vector machine.**

I. INTRODUCTION

Intrusion Detection Systems are divided into two categories [6] according to the detection approaches: Misuse Detection and Anomaly Detection.

A. Misuse detection

Misuse detection, finds intrusions by looking for activity corresponding to known techniques for intrusions i.e. it build pattern for malicious behavior first, and then identify intrusion based on this known pattern. The advantage of misuse detection is that it has higher detection accuracy to all known attack. The weakness of this approach is that it can only detect intrusions that follow predefined patterns.

B. Anomaly detection

In anomaly detection, the system defines the expected behavior of the network or profile in advance. Any significant deviations from such expected behavior are reported as possible attacks. But not all such deviations are attacks. The advantage of anomaly detection is that it can examine unknown and more complicated intrusion. But the weakness is that the normal mode knowledge base is difficult to establish and the false detection rate is high. The shortcoming of anomaly detection is its high false alarm rate. Between these two approaches, only anomaly detection has the ability to detect unknown attacks. To improve the performance an Intrusion Detection System requires high detection rate as well as low false alarm rate.

Classification is used to classify each item in a set of data into one of predefined set of classes or groups. The data analysis task classification is where a model or classifier is constructed to predict categorical labels (the class label attributes). Classification is a data mining function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data. For example, a classification model could be used to identify loan applicants as low, medium, or high credit risks. A classification task begins with a data set in which the class assignments are known. For example, a classification model that predicts credit risk could be developed based on observed data for many loan applicants over a period of time. In addition to the historical credit rating, the data might track employment history, home ownership or rental, years of residence, number and type of investments, and so on. Credit rating would be the target, the other attributes would be the predictors, and the data for each customer would constitute a case. Classifications are discrete and do not imply order. Continuous,

IJO-SCIENCE
INTERNATIONAL JOURNAL ONLINE OF SCIENCE
ISSN: 2455-0108

floating-point values would indicate a numerical, rather than a categorical, target. A predictive model with a numerical target uses a regression algorithm, not a classification algorithm. The simplest type of classification problem is binary classification. In binary classification, the target attribute has only two possible values: for example, high credit rating or low credit rating. Multiclass targets have more than two values: for example, low, medium, high, or unknown credit rating. In the model build (training) process, a classification algorithm finds relationships between the values of the predictors and the values of the target. Different classification algorithms use different techniques for finding relationships. These relationships are summarized in a model, which can then be applied to a different data set in which the class assignments are unknown. Classification has many applications in customer segmentation, business modeling, marketing, credit analysis, and biomedical and drug response modeling. Data classification is defined as two-step process shown in figure 1.

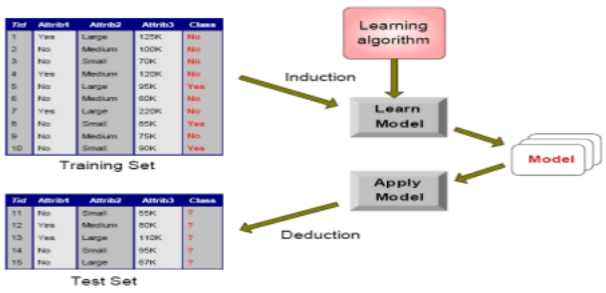

Figure 1: illustrating classification task.

Step 1: A classifier is built describing a predetermined set of data classes or concepts. (This is also known as supervised learning).

Step 2: Here, the model is used for classification. First, the predictive accuracy of the classifier is estimated. (This is also known as unsupervised learning).

The commonly used methods for data mining classification tasks can be classified into the following groups.1.Decision tree induction methods, 2. Neural networks, 3. Bayesian network, 4. Support vector machines. Rest of the paper is arranged in the same fashion.

II. DECISION TREE INDUCTION

Decision tree learning is a method commonly used in data mining. The goal is to create a model that predicts the value of a target variable based on several input variables. Each interior node corresponds to one of the input variables; there are edges to children for each of the possible values of that input variable. This is illustrating in Figure 1. Each leaf represents a value of the target variable given the values of the input variables represented by the path from the root to the leaf.

Decision tree induction algorithms are function recursively. First, an attribute must be selected as the root node. In order to create the most efficient (i.e., smallest) tree, the root node must effectively split the data. Each split attempts to pare down a set of instances (the actual data) until they all have the same classification. The best split is the one that provides what is termed the most information gain.

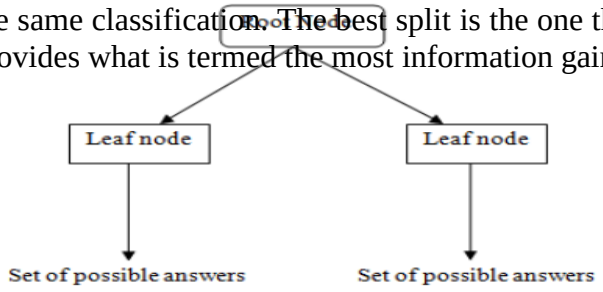

Figure 2: Decision tree induction

The basic algorithm for decision tree induction is a greedy algorithm that constructs decision trees in a top-down recursive, divide-and-conquer manner. The Tree induction algorithm, summarized as follows.

IJO-SCIENCE
INTERNATIONAL JOURNAL ONLINE OF SCIENCE
ISSN: 2455-0108

Step1: The algorithm operates over a set of training instances, C.

Step2: If all instances in C are in class P, create a node P and stop, otherwise select a feature or attribute F and create a decision node.

Step3: Partition the training instances in C into subsets according to the values of V.

Step4: Apply the algorithm recursively to each of the subsets C.

These algorithms usually employ a greedy strategy that grows a decision tree by making a series of locally optimum decisions about which attribute to use for partitioning the data. For example, Hunt's algorithm, id3, c4.5, cart, sprint are greedy decision tree induction algorithms.

A. Hunt's Algorithm

Hunt's algorithm grows a decision tree recursively by partitioning a training data set into smaller, purer subsets. This algorithm contains two steps in order to construct a decision tree.

Step 1: which is the terminating step for the recursive algorithm, checks if every record in a node is of the same class? If so, the node is labelled as a leaf node with its classification the class name of all the records within.

Step 2: If a node is not pure then selects/creates an attribute test condition to partition the data into two purer data sets. From here a child node is created for each subset.

The algorithm recurses until the all leaf nodes are found. A common means of deciding which attribute test condition should be used is the notion of the "best split." This concept boils down to choosing the test condition that results in subsets that are purer (where purity is richer when the set contains records of the same class). Three common formulas for calculating impurity is entropy, gini, and classification error.

B. (Iterative Dichotomiser 3) Algorithm

ID3 (Iterative Dichotomiser 3) is an algorithm invented by Ross Quinlan used to generate a decision tree. ID3 is the precursor to the C4.5 algorithm.

The ID3 algorithm can be summarized as follows:

- Take all unused attributes and count their entropy concerning test samples
- Choose attribute for which entropy is minimum (or, equivalently, information gain is maximum)
- Make node containing that attribute

C. C4.5 Algorithm

C4.5 is an algorithm used to generate a decision tree developed by Ross Quinlan.[2] C4.5 is an extension of Quinlan's earlier ID3 algorithm. The decision trees generated by C4.5 can be used for classification, and for this reason, C4.5 is often referred to as a statistical classifier.

In pseudo code, the general algorithm for building decision trees is:

1. Check for base cases
2. For each attribute a
3. Find the normalized information gain from splitting on a
4. Let a_best be the attribute with the highest normalized information gain
5. Create a decision node that splits on a_best
6. Recurse on the sub lists obtained by splitting on a_best, and add those nodes as children of node.

D. RndTree (Random Forest)

Random Forests grows many classification trees. To classify a new object from an input vector, put the input vector down each of the trees in the forest. Each tree gives a classification, and we say the tree "votes" for that class. The forest chooses the classification having the most votes (over all the trees in the forest).

Each tree is grown as follows:

If the number of cases in the training set is N, sample N cases at random - but with replacement, from the original data. This same t split on this m is used to split the node. The value of m is held constant during the forest growing.

III. NEURAL NETWORKS

IJO-SCIENCE
INTERNATIONAL JOURNAL ONLINE OF SCIENCE
ISSN: 2455-0108

In more practical terms neural networks are nonlinear statistical data modeling tools. They can be used to model complex relationships between inputs and outputs or to find patterns in Data [1]. Using neural networks as a tool, data warehousing firms are harvesting information from datasets in the process known as data mining. The difference between these data warehouses and an ordinary database is that there is actual manipulation and cross-fertilization of the data helping users makes more informed decisions.

A. Feed forward Neural Network

One of the simplest feed forward neural networks (FFNN), such as in Figure 4, consists of three layers: an input layer, hidden layer and output layer. In each layer there are one or more Processing Elements (PEs). PEs is meant to simulate the neurons in the brain and this is why they are often referred to as neurons or nodes. A PE receives inputs from either the outside world or the previous layer. There are connections between the PEs in each layer that have a weight (parameter) associated with them. This weight is adjusted during training. Information only travels in the forward direction through the network - there are no feedback loops optimization techniques.

IV. Fuzzy Logic

One of the control structures for resolving the issues is the Fuzzy Logic which provides the facility to be installed into the system of different range that is multichannel personal computer or may be at work station or the control-systems also this deals with both hardware and software. Based on the noisy, inaccurate, ambiguous information fuzzy-logic may provide the way to make the proper decisions [11].

Fuzzy logic attempts to systematically and mathematically emulate human reasoning and decision making [12]. It provides an intuitive way to implement control systems, decision making and diagnostic systems in various branches of industry. Fuzzy logic represents an excellent concept to close the gap between human reasoning and computational logic. Variables like intelligence, credibility, trustworthiness and reputation employ subjectivity as well as uncertainty. They cannot be represented as crisp values, however their estimation is highly desirable. Fuzzy systems are emerging technologies targeting industrial applications and added a promising new dimension to the existing domain of conventional control systems. Fuzzy logic allows engineers to exploit their empirical knowledge and heuristics represented in the IF-THEN rules and transfer it to a functional block. Fuzzy logic systems can be used for advanced engineering applications such as intelligent control systems, process diagnostics, fault detection, decision making and expert systems.

V. Classification and Regression Trees (CART)

CART is one of the popular methods of building decision tree in the machine learning community. CART builds a binary decision tree by splitting the record at each node, according to a function of a single attribute. CART uses the gini index for determining the best split. The initial split produces the nodes, each of which we now attempt to split in the same manner as the root node. Once again, we examine the entire input field to find the candidate splitters. If no split can be found then significantly decreases the diversity of a given node, we label it as a leaf node [13].

Eventually, only leaf nodes remain and we have grown the full decision tree. The full tree may generally not be the tree that does the best job of classifying a new set of records, because of over fitting. At the end of the tree-growing process, every record of the training set has been assigned to some leaf of the full decision tree. Each leaf can now be assigned a class

and error rate. The error rate of a leaf node is the percentage of the incorrect classification at that node. The error rate of an entire decision

IJO-SCIENCE
INTERNATIONAL JOURNAL ONLINE OF SCIENCE
ISSN: 2455-0108

tree is a weighted sum of the error rates of all the leaves.

Each leaf's contribution to the total is the error rate at that leaf multiplied by the probability that a record will end up in there.[14] Classification and Regression Trees is a classification method which uses historical data to construct so-called decision trees. Decision trees

are then used to classify new data. In order to use CART we need to know number of classes a priori.

CART methodology consists of three parts:

1. Construction of maximum tree
2. Choice of the right tree size
3. Classification of new data using constructed tree

VI. BAYESIAN NETWORK

A Bayesian network, Bayesian networks are directed acyclic graphs(DAG) whose nodes represent random variables in the Bayesian sense: they may be observable quantities, latent variables, unknown parameters or hypotheses. Edges represent conditional dependencies; nodes which are not connected represent variables which are conditionally independent of each other. Each node is associated with a probability function that takes as input a particular set of values for the node's parent variables and gives the probability of the variable represented by the node. For example, if the parents are m Boolean variables then the probability function could be represented by a table 2^m of entries, one entry for each of the possible 2^m combinations of its parents being true or false.

Similar ideas may be applied to undirected, and possibly cyclic, graphs; such are called Markov networks. Bayesian approaches are a fundamentally important DM technique. Given the probability distribution, Bayes classifier can provably achieve the optimal result. Bayesian method is based on the probability theory. Bayes Rule is applied here to calculate the posterior from the prior and the likelihood, because the later two is generally easier to be calculated from a probability model One.

VII. SUPPORT VECTOR MACHINES

Support Vector Machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data and recognize patterns. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier.

SVM models have similar functional form to neural networks and radial basis functions, both popular data mining techniques. However, neither of these algorithms has the well-founded theoretical approach to regularization that forms the basis of SVM. The quality of generalization and ease of training of SVM is far beyond the capacities of these more traditional methods. The support vector machine (SVM) is a training algorithm for learning classification and regression rules from data, for example the SVM can be used to learn polynomial, radial basis function (RBF) and multi-layer perceptron (MLP)classifiers.

In the classification task, data characteristics tend to discriminate better the algorithms [3] . However, in the regression task, things tend to be more similar referring to the error rate. On the basis of the analysis of the experimental results, some conclusions can be drawn for the classification task, as follows: The deeper is the C5.0 tree the higher is the relative error of the NN to the error of the C5.0; the lower is the number of nominal variables the higher is the relative error of DA to the error of C5.0 tree; the higher is the number of classes the higher is the relative error of LR to the error of C5.0 tree.

The experimental analysis also shows that SPRINT and C4.5 algorithms have a good classification accuracy compared to other algorithms used in the study [4]. The variation of data sets class size, number of attributes and volume of data records is used to determine

IJO-SCIENCE
INTERNATIONAL JOURNAL ONLINE OF SCIENCE
ISSN: 2455-0108

which algorithm has better classification accuracy between IDE3 and CART algorithms.

[5] presents a "comparative study of various data mining classification algorithms "on the dataset "social side of the internet" for two set. For sub set 1, the features selected by Feature ranking and for sub set 2 relief F filtering. The features selected by feature reduction techniques are chosen as input attributes with necessary class variables as target attribute and various classifications algorithms were executed for all selected features one by one and there error rates are tabled below. In this research their conclusion was Rnd Tree performed well for their taken dataset.

VIII. DISCUSSION

Boosted decision tree approach [7] for IDS is an ensemble approach, its detection rate is good but it has moderate false alarm rate. Since it combines several decision trees, it becomes complex, requires more time and space.

Back-propagation artificial neural network model [8] requires a very large amount of data and also considerable time to ensure that the results are accurate. Also, there is some kind of compromise between increasing the percentage of detection and the classification levels.

Hybrid learning approach [9] by using a combination of some classification and naive bayes classification overcomes the drawback of moderate detection rate and high false alarm rate of existing methods. A hybrid IDS [10] that combine k-means, and two classifiers: k-nearest neighbor and naive bayes overcome the drawback of very high false alarm rate in existing method.

IX. CONCLUSION

The goal of classification algorithms is to generate more certain, precise and accurate system results. Numerous methods have been suggested for the creation of ensemble of classifiers. Classification methods are typically strong in modeling interactions. Several of the classification methods produce a set of interacting logic that best predict the phenotype. However, a straightforward application of classification methods to large numbers of markers has a potential risk picking up randomly associated markers. But still it is difficult to recommend any one technique as superior to others as the choice of a dataset. Finally, there is no single classification algorithms is best for all kind of dataset. Classification algorithms are specific in their problem domain.

This paper describes different data mining techniques applied for detecting intrusions. Data mining helps to understand normal behavior inside the data and use this knowledge for detecting unknown intrusions. This paper describes different data mining techniques like classification. Classification is a supervised learning technique that can handle only labelled data i.e. it can detect only known attacks.

REFERENCES

[1] Sundar. C, M. Chitradevi and Dr. G. Geetharamani "Classification of Cardiotocogram Data using Neural Network based Machine Learning Technique" International Journal of Computer Applications (0975 – 888) Volume 47– No.14, June 2012.

[2] Smitha .T, V. Sundaram "Comparative Study Of Data Mining Algorithms For High Dimensional Data Analysis" International Journal Of Advances In Engineering & Technology, Sept 2012.IJAET ISSN: 2231-1963

[3] Koliastasis C and D.K. Despotis "Rules for Comparing Predictive Data Mining Algorithms by Error Rate" OPSEARCH, VOL. 41, No. 3, 2004 Operational Research Society of India.

[4] Matthew N. Anyanwu, Sajjan G. Shiva "Comparative Analysis of Serial Decision Tree Classification Algorithms" International Journal of Computer Science and Security, (IJCSS) Volume (3) : Issue (3) page number 230.

[5] Mrs.P.Nancy, R. Geetha Ramani and Shomona Gracia Jacob "A Comparison on Performance of Data Mining Algorithms in Classification of Social Network Data" International Journal of Computer Applications (0975 –8887) Volume 32–No.8, October 2011.

[6] Deepthy K Denatious, Anita John, "Survey on Data Mining Techniques to Enhance Intrusion Detection" In Proceedings of International Conference on Computer Communication and Informatics (ICCCI -2012), Jan. 10 – 12, 2012, Coimbatore, INDIA, IEEE, 2012.

[7] Mrudula Gudadhe, Prakash Prasad, Kapil Wankhade, "A New Data Mining Based Network Intrusion Detection Model", In Proceedings of International Conference on Computer and

ISSN: 2455-0108

Communication Technology (ICCCT 2010), IEEE, 2010, pp.731-735.

[8] Sufyan T. Faraj Al-Janabi, Hadeel Amjed Saeed, "A Neural Network Based Anomaly Intrusion Detection System", 2011 Developments in Esystems Engineering, IEEE, 2011, pp.221-226.

[9] Z. Muda, W. Yassin, M.N. Sulaiman, N.I. Udzir, "Intrusion Detection based on K-Means Clustering and Naïve Bayes Classification", In Proceedings of 7th International Conference on IT in Asia (CITA), IEEE, 2011.

[10] Hari Om, Aritra Kundu, "A Hybrid System for Reducing the False Alarm Rate of Anomaly Intrusion Detection System", In Proceedings of 1st Int'l Conf. on Recent Advances in Information Technology (RAIT-2012), IEEE, 2012.

[11] A.M. Chandrasekhar, "Intrusion Detection Technique By Using K-Means, Fuzzy Neural And SVM Classifier ", 2013 International Conference on Computer Communication and Informatics (ICCCI - 2013), Jan04-06, 2013 Coimbatore, India.

[12] Araki, S., Yamaguchi, Y., Shimada, H., & Takakura, H. (2014), "Unknown Attack Detection by Multistage One-Class SVM Focusing on Communication Interval", In Neural Information Processing (pp. 325-332). Springer International Publishing.

[13] JAYSHRI R. PATEL, " Performance Evaluation of Decision Tree Classifiers for Ranked Features of Intrusion Detection", NOV 12 TO OCT 13 | volum – 02, ISSUE – 02.

[14] Tiwari Nitin, S. R. Singh and P. G. Singh, Intrusion Detection and Prevention System (IDPS) Technology- Network Behavior Analysis System (NBAS), International Science Congress Association , 51-56, July (2012).