International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 3 Issue 1, January - 2014

A Survey of Fuzzy Clustering Techniques for Intrusion Detection System

Richa Sampat, Shilpa Sonawani

Maharashtra Institute of Technology, Pune[1, 2]

Abstract

Internet has become a vital part of any organization. Sensitive and confidential information is being sent over the network. But with the growth of internet, intrusion and attacks have also increased. Thus, there arises a need of robust and powerful intrusion detection systems which can detect the attacks. Recently, many novel methods are experimented to build strong IDSs. A lot of soft computing techniques are used. One such technique is fuzzy logic. This paper gives a survey on fuzzy clustering techniques that are used to build IDS.

1. Introduction

With the growth of network technology, nowadays more and more people are learning various ways to attacks through the network resources and carry out extremely destructive attacks. In recent years, the amount of hackers' attacks is growing 10 times per year. Therefore, network security becomes a vital factor of computer technology.

The concept of Intrusion detection system (IDS) proposed by Denning (1987) is useful to detect, identify and track the intruders [1]. Thus, Intrusion Detection is the process of monitoring and analyzing the events occurring in a computer system in order to detect signs of security problems.

In today's scenario, Intrusion Detection is an important aspect as very large amount of sensitive information is stored and processed in networked systems across the globe. IDS' modeling is a complex task and in recent years, focus has shifted to data mining approaches and soft computing techniques to strengthen network security. These approaches include neural networks, decision trees, genetic algorithms, support vector machines, Naïve Bayes classification, clustering, fuzzy logic, etc. These techniques can detect both known and unknown patterns of attacks, thereby helping in development of smart intrusion detection systems. In this paper, our focus is on fuzzy clustering techniques and its variations that are used for building a robust intrusion detection system.

The rest of the paper is organized as follows. Section 2 gives the details of intrusion detection and its types. Section 3 gives details of fuzzy logic concepts and fuzzy clustering techniques. Section 4 gives the details of different fuzzy clustering techniques and explains how each technique helps in detecting intrusions. Section 5 presents the conclusion.

2. Intrusion detection

Intrusion is defined as a set of actions that attempt to compromise the integrity, confidentiality or availability of any resource on computing platform [2]. IDS are systems that monitor the network looking for malicious or suspicious behavior in users' activity. The main goal of IDS is to alarm the network administrator that any malicious or suspicious activity has happened.

According to the different detection methods, we can say that there are two types of intrusion detection systems.

2.1. Misuse detection

Misuse detection systems are the approach that tries to match user activity to signatures of known attacks that are stored in the database. Such detection systems use a prior defined knowledge to check the new activity happening on the network. It has high speed of detection and low percentage of false alarm. However, it fails in discovering the new attacks that are not defined in the database.

2.2. Anomaly detection

Anomaly Detection approach works on principle “anomalies are not normal”. Such detection approach tries to find whether the change from the normal usage pattern can be called as intrusion or not. Thus, the anomaly detection technique stores the systems normal profile activity and raises an alarm if any abnormal behavior (i.e. intrusive activity) occurs which deviates from normal behavior. Anomaly detection helps in finding new attacks in the network.

This type of IDS can be further divided into two categories: Host-based Intrusion Detection System (HIDS) and Network-based Intrusion Detection System (NIDS). An HIDS resides on a particular host and looks for indications of attack on that host. An NIDS resides on separate machines that look for indications of attack in the whole network. The current trend in intrusion detection is to combine both host-based and network-based information to develop hybrid systems.

3. Fuzzy logic

Fuzzy logic is logic of fuzzy sets. A Fuzzy set can have an infinite range of truth values between one and zero. Fuzzy logic is capable of supporting human type of reasoning in natural form to a considerable extent. It does so by allowing partial membership for data items in fuzzy subsets. Integration of fuzzy logic with data mining techniques has become one of the key aspects of soft computing in handling the challenges posed by the massive collection of natural data.

The fuzzy set is different from a crisp set in that it allows the elements to have a degree of membership. The core of a fuzzy set is its membership function which defines the relationship between a value in the set’s domain and its degree of membership in the fuzzy set. The relationship is functional because it returns a single degree of membership for any value in the domain.

f(s,x) is the fuzzy membership value for the element

s: the fuzzy set

x: the value from the underlying domain.

Fuzzy sets provide a means of defining a series of overlapping concepts for a model variable since it represent degrees of membership.

3.1. Crisp (Hard) clustering techniques

Traditional clustering techniques form uniquely defined cluster within the dataset by grouping related attributes. Each data item in the sample space is assigned to only one cluster. K-means algorithm and its different variations are the most commonly used methods to achieve this. The value k stands for the number of cluster initially provided for the algorithm. This algorithm takes the input k and partitions a set of m objects into k clusters. The technique works by computing the distance between a data item and the cluster center to add an item into one of the clusters so that intra-cluster similarity is high but inter cluster similarity is low. A common method to find the distance is to calculate to sum of the squared difference as follows and it is known as the Euclidian distance. With the definition of the distance of a data point from the cluster centers, the k-means algorithm is quite simple. The cluster centroids are randomly initialized and we assign a data point x_i into a cluster to which it has minimum distance. When all the data points have been assigned to clusters, new cluster centers are calculated by finding the weighted average of all data points in a cluster. The cluster center calculation causes the previous centroid location to move towards the center of the cluster set. This is continued until there is no change in cluster centers.

3.1.1 Limitations of crisp (hard) clustering algorithms. The main limitation of these algorithms comes from its crisp nature in assigning cluster membership to data points. Depending on the minimum distance, a data point always becomes a member of one of the clusters. This works well with highly structured data. The real world data is almost never arranged in clear cut groups. Instead, clusters have ill defined boundaries that smear into the data space often overlapping the perimeters of surrounding clusters. In most of the cases the real world data have apparent extraneous data points which do not clearly belong to any of the clusters and they are called outlier points. The k-means algorithm is not capable of dealing with over- lapping clusters and outlier points since it has to include a data point into exactly one of the existing clusters. Because of this even extreme outlier points will be included in to some cluster based on the minimum distance

3.2. Fuzzy clustering techniques

The central idea in fuzzy clustering is the non-unique partitioning of the data in a collection of clusters. The data points are assigned membership values for each of the clusters. The fuzzy clustering algorithms allow the clusters to grow into their natural shapes. In some cases the membership value may be zero indicating that the data point is not a member of the cluster under consideration. Many crisp clustering techniques have difficulties in handling extreme outliers but fuzzy clustering algorithms tend to give them very small membership degree in surrounding clusters. The non-zero membership values, with a maximum of one, show the degree to which the data point represents a cluster. Thus fuzzy clustering provides a flexible and robust method for handling natural data with vagueness and uncertainty.

4. Fuzzy clustering techniques for Intrusion Detection System

In this section, a study of different fuzzy clustering techniques that is used for intrusion detection is presented.

In [3] authors state that clustering is the best techniques for intrusion detection. Here, k-means clustering is used for intrusion detection because it gives efficient results in case of large datasets. But the author also states that sometimes k-means clustering fails to give best result because of class dominance problem and no class problem.

Researchers have used k means clustering along with the other method for improving the detection rate of intrusion detection system. Some of them are given as follows:

Om H [4] proposed a hybrid intrusion detection system that combines k-Means, and two classifiers: K-nearest neighbor and Naïve Bayes for anomaly detection. Here, first feature selection is done using an entropy based feature selection algorithm which selects the important attributes and removes the redundant attributes. This algorithm operates on the KDD-99 Data set. The next step is clustering phase using k-Means. Their approach gives better results than simple k-means algorithm.

W. Ren [5] has developed a method that applies fuzzy c-means clustering algorithm to detect network intrusion. He carries out fuzzy partition and clustering of data which separates normal data and attack data effectively. His experiment shows the feasibility and validity of fuzzy c-means clustering algorithm.

E. Narayan, *et al* [6] have proposed algorithms on expectation maximization fuzzy c-means clustering (EMFCM). Proposed algorithms provide better result to fuzzy c-means clustering by avoiding the looping problems and saves time. EMFCM clustering algorithm has fast convergence in a few iterations regardless of the initial number of clusters.

H. Wang [7] has proposed a mixed fuzzy clustering algorithm that uses Quantum-behaved Particle Swarm Optimization (QPSO) algorithm and combines with fuzzy c-means (FCM) for abnormally detection. New hybrid algorithm is proposed which is based on the gradient descent of FCM. This technique avoids the local minimum problems of FCM by including in the algorithm a strong global searching capacity. Also, FCM is no longer largely dependent on the initialization values. Their result shows that with FCM, the proposed algorithm not only has the favorable convergent capability of the global optimizing but also has obviously improved the robustness, and has the higher performance in intrusion detection than FCM and K-means algorithm.

F. Guorui [8] in his paper developed a semi-supervised learning algorithm for intrusion detection which is combined with the fuzzy c-Means algorithm. The sensitivity to the initial values and the probability of trapping in local optimum are greatly reduced by using few labeled data to improve the learning ability of the FCM algorithm. The KDD CUP99 data set is adopted as the experimental subject. The result proves that the attack behaviors can be more efficiently found from the network data by the semi-supervised FCM clustering algorithm.

J. Visumathi, *et al* [9] proposed a weighted fuzzy c–means clustering based on immune genetic algorithm for intrusion detection system. A new weighted fuzzy c-means clustering module is designed to make the system more accurate for attack detection, using the immune genetic algorithm which is used to improve the performance of the network. It also solves the high dimensionality problem in the given data set.

T. Fries [10] in his paper presents a fuzzy-genetic approach to intrusion detection that is shown to provide performance superior to other GA-based algorithms. In addition, the method demonstrates improved robustness in comparison to other GA-based techniques.

S. Chittineni, *et al* [11] proposed fuzzy c-means algorithm using neural network algorithm. The proposed work involves two steps. First, an Enhanced K-means Fast Leaning Artificial Neural Network (KFLANN) frame work is used to determine cluster centers. Secondly, fuzzy c-means uses these cluster centers to generate fuzzy membership functions.

In [12], Yu-Ping Zhou has proposed a system in which Principal Component Analysis (PCA) neural network is used to reduce the dimensions of the feature space. A modified fuzzy c-means clustering algorithm is used to cluster the learning data to obtain fuzzy rules. Also a hierarchical neuro-fuzzy classifier is developed. The experiments and evaluations of the proposed method were performed with the KDD Cup 99 intrusion detection dataset. Results indicate the high detection accuracy for intrusion attacks and low false alarm rate of the reliable system.

In [13], the authors propose a method of intrusion detection using an evolving fuzzy neural network. This type of learning algorithm combines Artificial Neural Network (ANN) and Fuzzy Inference Systems (FIS), as well as evolutionary algorithms. The algorithm uses fuzzy rules and allows new neurons to be created in order to accomplish this. They use Snort to gather data for training the algorithm and then compare their technique with that of an augmented neural network.

Li Jian-guo, *et al* [14] proposed an improved weighted fuzzy clustering algorithm based on rough set by using the methods of attributes contracted in the rough set theory to improve the FCM algorithm.

B. Thomas, *et al* [15] proposed a new fuzzy clustering method which is more efficient in handling outlier points than conventional fuzzy c-means algorithm. The new method excludes outlier points by giving them extremely small membership values in existing clusters while fuzzy c-means algorithm tends give them outsized membership values. The new algorithm also incorporates the positive aspects of k-means algorithm in calculating the new cluster centers in a more efficient approach than the c-means method.

5. Conclusion

In this paper, a survey of different fuzzy clustering techniques and algorithms is done and a comparison is drawn between them. With the growth and development of network, new attacks are going to happen. Techniques based on simple k-means are easy to implement and promise a highly efficient result when the data is structured. But, in real networks that is hardly the case and we need techniques which can detect attacks on loosely structured data and also improve the detection over time. Fuzzy clustering and classification techniques, though difficult to understand and implement, are capable of achieving this to good extent. Thus, fuzzy clustering techniques and algorithms show a promising way towards the development and enhancement of robust IDS

6. References

[1] D.E. Denning, "*An Intrusion Detection Model*", IEEE Transaction on Software Engineering, vol SE-13, no.2, pp. 222-232, 1987.

[2] K. Labib, "*Computer Security and Intrusion Detection*", from Crosswords, The ACM Students Magazine.

[3] K. Bharti, S. Shukla and S. Jain, "Intrusion Detection using unsupervised learning", International Journal on Computer Science and Engineering. Vol. 02, No. 05, 2010, 1865-1870.

[4] Om H, *"A hybrid system for reducing the false alarm rate of anomaly intrusion detection system"* Recent Advances in Information Technology (RAIT), 2012 1st International IEEE Conference.

[5] W. Ren, *"Application of Network Intrusion Detection Based on Fuzzy C-Means Clustering Algorithm",* Intelligent Information Technology Application, 2009. IITA IEEE 2009.

[6] E. Narayan, P. Singh and G. Tak, *"Intrusion Detection System Using Fuzzy C Means Clustering with Unsupervised Learning via EM Algorithms", VSRD-IJCSIT, Vol. 2 (6), 2012,* 502-510.

[7] H. Wang, *"Network intrusion detection based on hybrid Fuzzy C-mean clustering*", Fuzzy Systems and Knowledge Discovery (FSKD), 2010 Seventh International IEEE Conference.

[8] F. Guorui, *"Intrusion detection based on the semi-supervised Fuzzy C-Means clustering algorithm",* Consumer

Electronics, Communications and Networks (CECNet), 2012 2nd IEEE International Conference.

[9] J. Visumathi, Dr. K. L. Shanmuganathan and Dr. K. A. M. Junaid, *"Misuse and Anomaly-based Network Intrusion Detection System using Fuzzy and Genetic Classification Algorithms"* International Conference on Computing and Control Engineering (ICCCE 2012), 12 & 13 April, 2012.

[10] T. Fries, *"A Fuzzy-Genetic Approach to Network Intrusion Detection",* Department of Computer Science Coastal Carolina University Conway, South Carolina.

[11] S. Chittineni and Dr. R.B. Bhogapathi, *"Neural Network Based Fuzzy C MEANS Clustering Algorithm",* International Journal of Electronics Signals and Systems, Vol. 1, Issue 1.

[12] Yu-Ping Zhou, *"Research on Neuro-fuzzy Inference System in Hierarchical Intrusion Detection",* Information Technology and Computer Science, 2009. ITCS 2009 International IEEE Conference.

[13] M. Panda1 and M. R. Patra, *"Network Intrusion Detection Using Naïve Bayes*". IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.12, December 2007.

[14] LI Jian-guo and G. Jing-Wei, *"Research on Improved Weighted Fuzzy Clustering Algorithm based on Rough Set",* Proceedings of International Conference on Computer Engineering and Technology, pp.98- 102, 2009.

[15] B. Thomas and G. Raju, *"A Novel Fuzzy Clustering Method for Outlier Detection in Data Mining",* International Journal of Recent Trends in Engineering, Vol. 1, No. 2, May 2009.