Tuning FCM Parameters with AMOSA

Seyed Mahmood Hashemi[1] Jingsha He[2]
Hashemi2138@yahoo.com Jhe@bjuit.edu.cn
[1,2]: Beijing University of Technology, School of Software Engineering
Beijing Engineering Research Center for IoT Software and Systems

Abstract:- Clustering of attack patterns is used in intrusion detection and discussed in this paper. Fuzzy C-Means (FCM) is a popular method for clustering; the performance of FCM is depends to its parameters radically. In this paper, performance of FCM is evaluated with some Validity Indices and represented in a multi-objective optimization problem. Multi-Objective Simulated Annealing (AMOSA) has high power to solve multi-objective optimization problems, so it is used in this paper to tune the parameters of FCM.

Key-Words:- Intrusion Detection, Clustering, FCM, Multi-Objective Optimization, AMOSA

1 Introduction

Intrusion is the act of wrongfully entering upon, seizing, or taking possession of the property of another. In Information Technology (IT), intrusion is defined as an unauthorized access to or a malicious activity on a computer or an information system [4]. Intrusion detection provides the real-time protection for internal attacks, external attacks and disoperation [1]. Intrusion detection technologies can be divided into two categories: misuse detection and anomaly detection. Misuse intrusion detection firstly groups attack patterns, then manually set up the corresponding detection rules and patterns to construct the intrusion detection system. Anomaly intrusion detection is mainly dependent on the intuition and experience to select the statistic feature to construct the intrusion detection system. Since diversion in network is abruptly, misuse intrusion detection is used more usable. Firstly, there is need to powerful approach to group attack patterns. Since there is no pre-knowledge about structures of attack patterns, clustering is more useable.

The work in this paper has been supported by National Natural Science Foundation of China (61272500), National High-tech R&D Program (863 Program) (2015AA017204) and Beijing Natural Science Foundation (4142008).

A common approach which can be used in intrusion detection is clustering [2]. Fuzzy C-Means Algorithm (FCM) is a popular approach for clustering. Some necessary basics of FCM (such as convergence) are represented in papers [3]. Fuzzy C-Means has already been widely used in bioscience, medicine, agriculture, target recognition, geography and so on [5]. FCM performance is radically depends to tune of two parameters: number of clusters; fuzzification. There are some researches that try to optimize them. Modification of FCM is represented with *Validity Index*. Since there are number of validity indices (and each of them focus on one aspect of FCM), we can optimum them simultaneously as a *Multi-Objective Optimization* problem.

The aim of this paper is clustering of patterns with FCM. *Mutli-Objective Simulated Annealing* is used to optimize more validity indices and find optimum values for cluster number and fuzzification.

The rest of this paper is organized as follow: related works are discussed in section 2. Section 3 is about definition of FCM and its validity indices. Section 4 is about multi-objective optimization and multi-objective imperialist competitive algorithm. Proposed algorithm and its result are represented in section 5. Finally, section 6 is about conclusion.

2 Related Works

To evaluate the intrusion detection system, two major indicators of performance were of interest [13]: the detection rate and false positive rate. The detection rate (DR) is defined as the number of intrusion instances detected by the system divided by the total number of intrusion instances present in the test set. The false positive rate (FPR) is defined as the total number of normal instances that were (incorrectly) classified as intrusions by the total number of normal instances.

With the development of internet the quantitative risk is assessed and automatic decision-making of network security has been especially important. Clustering algorithm is used to mine network security evaluation rules which are utilized to build an algorithm model of network security model to assess decision-making [14]. First of all, network security is based on two basic assumptions. User behavior and procedures are visible and normal behavior and acts of intrusion can be distinct essentially. Under normal network environment behavior is the mainstream and the invaded is individual. The testing data set may be considered to divide into the normal behavior of the cluster groups and various non-normal mode or small group of cluster. Namely, only are the normal behavior patterns classified into a concentrated clustering set and various type of invaded behavior is not further clustered. Therefore, a normal pattern of cluster groups are established which keeps the normal behavior patterns of behavior away from the invaded as soon as possible.

The proposed algorithm in [15] can speed up the overall computation time and reduce the total number of calculation. The main idea of the proposed algorithm is to refine the initial cluster centers. The algorithm does not note to the value of fuzzification, so the shape of the clusters are not optimize.

Complex network offers a bran-new view for studying complexity [16]. Degree k_i of node *i* is defined as number of other odes connected by the node. Aggregation coefficient of node *i* is defined as follow: $C_i = \frac{2E_i}{(k_i(k_i-1))}$. in characteristic view the formula for aggregation coefficient is changed to

$$C_i = \frac{\textit{the number of traingles connected to node } i}{\textit{the number pf triples connected to node } i}$$

.

The weights between nodes would influence to difference between the networks. In this study of network security evaluation model, whether the nodes are connected must be not only considered, but also the risk conductivity should be considered in the transmission of the risk between two nodes.

3 FCM

Fuzzy C-Means (FCM) is proposed by Bezdek [2]. FCM clustering algorithm determines whether the ting belongs to a certain category on the basis of its membership. Let $X = \{X_1, X_2, \cdots, X_n\}$ is data set and each of data vectors belong to d-dimensional space i.e. $X_j \in \mathbf{R}^d$. FCM algorithm minimizes the following objective function with respect to membership degree of each data and cluster centroid [7].

$$J_m = \sum_{i=1}^{c}\sum_{j=1}^{n}(\mu_{ij})^m d^2(X_j, V_i) \qquad (1)$$

$$\textit{Such That: } \sum_{i=1}^{c}\mu_{ij} = 1$$

Where

$$d^2(X_j, V_i) = (X_j - V_i)^T A(X_j - V_i) \qquad (2)$$

A is a definite matrix, c number of clusters and V_i is centroid of each cluster, n is number of data vectors (or data points), μ_{ij} is membership of specific data vector in specific cluster and $m > 1$ is fuzzification.

FCM is an iterative algorithm. In the first iteration, values of cluster centroids are selected randomly, but in the next iterations they are determined according to the following formula:

$$V_i = \frac{\sum_{j=1}^{n}(\mu_{ij})^m X_j}{\sum_{j=1}^{n}(\mu_{ij})^m} \qquad (3)$$

Membership degrees are recomputed in each epoch.

$$\mu_{ij} = \frac{\left(\frac{1}{d^2(X_j, V_i)}\right)^{\frac{1}{m-1}}}{\sum_{i=1}^{c}\left(\frac{1}{d^2(X_j, V_i)}\right)^{\frac{1}{m-1}}}$$

(4)

Iterations of algorithm repeat until the value of J_m do not change.

The probability of selecting i[th] cluster given the j[th] data vector is:

$$h(i|X_j) = \frac{1/d^2(X_j, V_i)}{\sum_{k=1}^{c} 1/d^2(X_j, V_k)}$$

(5)

Fuzzy Covariance Matrix of the i[th] cluster is calculated with:

$$FCM_i = \frac{\sum_{j=1}^{N} h(i|X_j)(X_j - V_i)(X_j - V_i)^T}{\sum_{j=1}^{N} h(i|X_j)}$$

(6)

There are two parameters influence on performance of FCM: value of fuzzification ($m > 1$) and number of clusters (n). To get the best performance of FCM, important parameters of FCM must be optimized simultaneously according to *Validity Index*. Validity indices are some measures which determine suitable clustering. Validity functions typically suggest finding trade-off between intra-cluster and inter-cluster variability [9]. There are numbers of validity criteria for fuzzy clustering [10, 11], but four of them cover other ones. Bezdek designed *Partition Coefficient* to measure the amount of "overlap" between clusters [7].

$$F = \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n}(\mu_{ij})^2 \quad (7)$$

Where *n* is number data vectors and *c* is number of clusters.

Compactness and Separation validity is computed by the following formula:

$$S = \frac{\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^2 \|V_i - X_j\|^2}{n \min_{i,j}\|V_i - V_j\|^2}$$

(8)

Fuzzy Hyper-Volume [8] is calculated with (9).

$$F_{HV} = \sum_{i=1}^{c}[\det FCM_i]^{1/2} \quad (9)$$

Where FCM_i is (6). *Partition Density* is:

$$D_p = \frac{A}{F_{HV}}$$

(10)

Where $A = \sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}$

4 Mutli-Objective Optimization

Multi-Objective Optimization (MOO) are necessary when multiple cost functions are considered in the same problem. The aim of MOO is tuning the decision variables to satisfy all objective functions to optimum value. This class of problems is modeled by (11).

$$Optimize\ [F_1(X), \cdots, F_k(X)]$$

(11)

$$Subject\ to:\ g_i(X) \le 0, h_j(X) = 0; i = 1, \cdots, m; j = 1, \cdots\cdots, p$$

Where k is the number of objective functions, X is the decision vector, m is the number of inequality constraints and p is the number of equality constraints.

This goal causes different between these algorithms and their ancestor Single-Objective Optimization, which is base on concept of best, while the multi-objective optimization uses the concept of dominance. Dominance is defined in [17]:

$$\vec{U} = (u_1, \cdots, u_n) \prec \vec{V} = (v_1, \cdots, v_n)\ iff\ \forall i \in \{1, \cdots, n\}: u_i \le v_i$$

(12)

In words, a vector $\vec{U} = (u_1, \cdots, u_n)$ dominates another vector $\vec{V} = (v_1, \cdots, v_n)$ if and only if $\vec{U}$ can reach to optimal value for some criteria without causing a simultaneous non-optimal value for at least one criterion. If two vectors cannot dominate each other, they are called as non-dominated vectors.

4.1. Multi-Objective Simulated Annealing

Basic concept in Simulated Annealing is evolution of the solution by simulating the decreasing temperature (*tmp*) in a material, where higher the temperature meaning that higher the modification of the solution at a generation. If temperature of a hot material

decreases very fast its internal structure may divers and material becomes hard and fragile. Decreasing temperature slowly yields higher homogeneity and less fragile material. Evolution of the solution is carried at specific temperature profiles. At the first iterations a diverse set of initial solutions for the problem is produced at higher temperature. And, these solutions are evolved while the temperature decreases to get their local optimums. In multi-objective situation, there are non-dominated solutions which must be kept in the archive, as a candidate of optimal solution.

Along the run of AMOSA algorithm, there are two solutions: current-so and new-so [18]. They can have three states compared to each other: i- current-so dominates new-so, ii- current-so and new-so are non-dominated each other and iii- new-so dominates current-so.

If new-so is dominated by current-so, there may be solutions in archive which dominates new-so. New-so is accepted to the archive by the probability:

$$p = \frac{1}{1 + exp(\Delta . tmp)} \quad (13)$$

Where Δ is differencing between new-so and other solutions which dominated new-so. If there are *A* solutions in the archive,

$$\Delta = \frac{\sum_{i=1}^{A} \Delta_i + \Delta}{A + 1} \quad (14)$$

Solutions can escape from local-optima and reach to the neighborhood of the global-optima by this probable acceptance.

If new-so is dominated by some solutions in the archive, (14) is modified to:

$$\Delta = \frac{\sum_{i=1}^{A} \Delta_i}{A} \quad (15)$$

When new-so is non-dominated with all members in archive, then new-so is set as current-so and it is added to the archive.

If new-so dominates some solutions in the archive, then new-so is set as current-so and it is added to the archive and solutions in the archive which are dominated by new-so are removed.

If new-so is dominated by by some solutions in the archive, then (13) is changed to:

$$p = \frac{1}{1 + exp(-\Delta)} \quad (16)$$

Where Δ is the minimum of the difference between new-so and dominating solutions in the archive. New-so is set as current-so with the probability (13). If new-so is non-dominated by all solutions in the archive it is set as current-so and added to the archive. If new-so dominates some solutions in the archive, it is set as current-so; it is added to the archive; and all dominated solutions are removed from the archive.

5 Proposed Algorithm

The performance of FCM depends to their parameters radically. In other words, attack patterns are clustered very well if we tune the number of clusters and the value of fuzzification. The aim of this paper is optimization of FCM parameters with AMOSA according to validity indices. There are number of validity indices. In [19, 20] some features for suitable clusters are represented. Clusters must have optimum *Adherence* and *Coherence* simultaneously. There is need to clusters which have maximum *Adherence* (means patterns in cluster have the most similarity to each other) and minimum *Coherence* (means minimum similarity between patterns in different clusters). Since formulas (7), (8) and (10) represent adherence and coherence of clusters, they must be optimum simultaneously. [18] represents the ability of AMOSA in multi-objective optimization, so AMOSA is used in this paper. Therefore problem is declared as follow:

$$Optimize\ [F_1(X), F_2(X), F_3(X)] \quad (17)$$

Where $F_1(X)$ is Partition Coefficient (7) is, $F_2(X)$ is Compactness and Separation Validity (8) and $F_3(X)$ is Partition Density (10).

$\vec{X}$ is a vector which has two fields. One field is number of clusters and another field is value of fuzzification. Values of $\vec{X}$ are used in FCM, then values of FCM are used in (17). After iterations optimum value for $\vec{X}$ includes optimum number of clusters and optimum value of fuzzification can be reached.

In this paper, problem in (17) is tested with a subset of KDD-99 data [21]. KDD-99 is the abbreviation of International Conference on

Knowledge Discovery and Data Mining 1999. KDD-99 data includes a wide variety of intrusions simulated in a network environment. There is need to measures to evaluate the proposed algorithm.

$\vec{X} = (5,1.392)$ is an individual, where 5 in number of clusters and 1.392 is value of fuzzifcation. Thus, according to (7), (8) and (10) respectively:

$F_1 = 13.7525, F_2 = 9.6428, F_3 = 22.351$

AMOSA, same as other evolution optimization algorithms, starts with a random initial individual. We use (5,1.392) as initial. AMOSA runs for 10 epochs. Table 1, represents the values of $\vec{X}$ in each iteration.

Number of Clusters	Fuzzification
5	1.392
7	1.6465
10	1.7302
11	1.7416
11	1.9384
12	1.9506
15	2.001
14	2.1025
15	2.1104
16	2.2138

Table 1. results in each iteration

Since the values in initialization step are set as random, the results of algorithm may change for different runs. Thus, there is need an approach to measure the usability of algorithm in real environment.

A common measure to evaluate the systems is K-Fold [22, 23]. K-Fold technique ensures the quality of simulated system. In other words, K-Fold ensures that the proposed algorithm can be worked in real systems. In K-Fold approach, the set of data divided into K subsets. Approach has K stages. In each stage of approach, K-1 subsets are used as test data and the rest is used as train data. For example, if K=5;

5-Fold approach has 5 stages. In the first stage, subsets: 1, 2, 3, 4 are used as test data and subset 5 is used train data. In the second stage, subsets: 1, 2, 3, 5 are used as test data and subset 4 is used as train data and so on. Each stage has specific Hit Ratio. Average of Hit Ratio represents the power of proposed algorithm in real environment. Hit ratios of 5-fold are 93%, 88%, 87%, 91%, 88%. Therefore, average hit ration is 89.4%.

6 Conclusion

Misuse intrusion detection approaches need to grouping the attack patterns. Because there is no pre-knowledge about patterns, clustering methods are used. Fuzzy C-Means (FCM) is a popular method for clustering. Actually FCM needs to tune the parameters. Tuning the parameters of FCM deeply influence on its performance. There are Validity Indices, which measure the clustering method. Motivation of this paper is tuning of FCM parameters according to partition coefficient, compactness and separation validity and partition density. Validity indices can be seen as objectives, so problem a multi-objective optimization problem. Multi-objective simulated annealing (AMOSA) is used to solve the recent problem. Results of 5-fold cross validation approach ensure that proposed system can be adapt to real environment.

References

[1]. Wuling Ren, Jinzhu Cao, Xianjie Wu, “Application of Network Intrusion Detection Based on Fuzzy C-Means Clustering Algorithm”, IEEE, 2009 Third International Symposium on Intelligent Information Technology Application, DOI 10.1109/IITA.2009.269

[2]. Wei Jiang, Min YAO, Jun YAN, “Intrusion Detection Based on Improved Fuzzy C-Means Algorithm”, IEEE, 2008 International Symposium on Information Science and Engineering, DOI 10.1109/ISISE.2008.17

[3]. Lawrence O. Hall, Dmitry Goldgof, “Convergence of the Single-Pass and Online Fuzzy C-Means Algorithm”, IEEE TRANSACTIONS ON FUZZY SYSTEMS, Vol. 19, No. 4, 2011

[4]. Yu Guan, Ali A. Gorbani, Nabil Belacel, “Y-MENAS: A CLUSTERING METHOD FOR INTRUCTION DETECTION”, CCECE 2003 CCGEI 2003, Montr´eal, May/mai 2003 0-7803-7781-8/03/$17.00 c 2003 IEEE

[5]. Peiyu LIU, Linshan DUAN, Xuezhi chi, Zhenfang ZHU, “An Improved Fuzzy C-Means Clustering Algorithm Based on Simulated

Annealing", 2013 10th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)
[6]. J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Functions algorithms", Plenum, New York, 1981
[7]. Xuanli Lisa Xie, Gerardo Beni, "A Validity Measure for Fuzzy Clustering", IEEE Log Number 9142752, IEEE, TRANSCATIONS ON PATTERN ANALISYS AND MACHINE INTELLIGENCE, Vol. 13, No. 8, 1991
[8]. I. Gath, B. Geva, "Unsupervised Optimal Fuzzy Clustering", IEEE, TRANSACTION ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, Vol. 11, No. 7, 1989
[9]. Jurgen Beringer, Eyke Hullermeier, "Adaptive Optimization of the Number of Clusters in Fuzzy Clustering", 1-4244-1210-2/07/$25.00 C 2007 IEEE.
[10]. M. Halkidi, Y. Batistakis, M. Vazigiannis, "Clustering algorithms and validity measures", 0-7695-1218-6/01 $10.00 *0* 2001 IEEE
[11]. Hyun-Sook Rhee, Kyung-Whan Oh, "A Validity Measure for Fuzzy Clustering and Its Use in Selecting Optimal Number of Clusters", 0-7803-3645-3/96 $5.00©1996 IEEE
[12]. G. V. Nadiamma, M. Hemalatha, "An Evaluation of Clustering Techniques over Intrusion Detection System", *ICACCI'12*, August 3-5, 2012, Chennai, T Nadu, India. Copyright 2012 ACM 978-1-4503-1196-0/12/08…$10.00.
[13]. Fangfei Wei, Qingshan Jiang, Liang Shi, Nannan Wu, "An Intrusion Detection System Based on the Clustering Ensemble", 1-4244-1035-5/07/$25.00 .2007 IEEE
[14]. Qu Zhiming, Wang Xiaoli, Study of Rough Set and Clustering Algorithm in Network Security Management", IEEE, 2009 International Conference on Networks Security, Wireless Communication and Trusted Computing, DOI 10.1109/NSWCTC.2009.47
[15]. Ming-Chuan Hung, Don-Lin Yang, "An Efficient Fuzzy C-Means Clustering Algorithm", 0-7695-1 119-8/01 $17.00 *0* 2001 IEEE
[16]. Shou Zhi-qin, Tao Jian-ping, Zhou Jian, "Appling Fuzzy C-Means Clustering Algorithm to Campus Network Security Assessment Based on Characteristics of Weighted Complex Network", IEEE, 978-1-4244-5326-9/10/$26.00 ©2010 IEEE
[17]. Carlos A. Coello Coello, David A. Van Veldhuizen, Gary B. Lamont, "Evolutionary Algorithms for Solving Multi-Objective Problems", Speringer, 2nd Ed., 2007, pp. 74
[18]. Sanghamitra Bandyopadhyay, Sriparna Saha, Ujjwal Maulik, Kalynmoy Deb, "A Simulated Annealing-Based Multi-Objective Optimization Algorithm: AMOSA", IEEE, Vol. 12, Issue 3, ISSN 1089-778x, 2008
[19]. R. Babuska, H.B. verbuggen, "Constructing Fuzzy Model By Product Space Clustering", Delft University of Technology, P.O. Box 5031, NL_2600 GA Delft, The Netherlands
[20]. Michio Sugeno, Takahiro Yasukawa, "A Fuzzy-Logic-Based Approach to Qualitative Modeling", IEEE Transaction On Fuzzy Systems, VOL. 1, NO. 1, Feb 1993
[21]. KDD Cup 1999 Data, University of California, Irvine, http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html,
1999.
[22]. Jin-Ho Chung, Kyeongcheol Yang, "k-Fold Cyclotomy and Its Application to Frequency-Hopping Sequences", IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 57, NO. 4, APRIL 2011
[23]. Juan Diego Rodriguez, Aritz Perez, Jose Antonio Lozano, "Sensitivity Analysis of k-fold Cross Validation in Prediction Error Estimation", IEEE TRANSACTION ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 32, NO. 3, MARCH 2010