Journal of Theoretical and Applied Information Technology
10th November 2015. Vol.81. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

FUZZY KERNEL C-MEANS ALGORITHM FOR INTRUSION DETECTION SYSTEMS

[1]ZUHERMAN RUSTAM, [2]AINI SURI TALITA

[1] Senior Lecturer, Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Indonesia, DEPOK, INDONESIA

[2] Associate Lecturer, Faculty of Computer Science and Information Technology, Gunadarma University, DEPOK, INDONESIA

E-mail: [1]rustam@ui.ac.id, [2]ainisuri@staff.gunadarma.ac.id

ABSTRACT

Intrusion Detection Systems (IDS) are used as security management systems. There are two approaches of IDS, Misuse Detection (knowledge-based intrusion detection) and Anomaly Detection (behavior-based intrusion detection). Misuse detection is performed by monitoring activities which is suspected as an intrusion based on prior information about specific attacks. While anomaly detection is based on the observation of the activity that is incompatible with the acceptable behaviors in normal conditions and makes it possible to determine new type of attacks in the system. Some Computational Intelligence models have been developed to solve Intrusion Detection Systems problems such as Neural Network and Neuro-Fuzzy methods. They are chosen because IDS involves large data sets with several different features that can bring out negative effects on IDS accuracy and its computational time. Naïve Bayes, Decision Tree (C4.5) and Kernel Matrix Methods can be used to reduce the number of features at data sets. We propose Fuzzy Kernel C-Means Algorithm as another method to solve IDS problems that we claim provides better results while combined with Kernel Matrix method to reduce the number of selected data features.

Keywords: *Data Features*, *Fuzzy C-Means*, *Intrusion Detection Systems*, *Kernel Matrix*, *Kernel Method*

1. INTRODUCTION

The importance of computers in modern life leads them as a potential target of attacks for personal gain. The system must be protected to prevent infiltration by unauthorized parties that intend to discover its weakness or classified information it contains. Intrusion Detection Systems (IDS) can be used to monitor and analyze user activities, maintain data and systems integrity, recognize the pattern of activity that indicates an attack, give respond to activity that detected as an attack automatically, make a report of activity detection and IDS also capable of detecting the unknown type of attack.

There are several ways to prevent attacker infiltration into the system, such as, using authentication methods (password, biometric authentication), avoiding programming errors at the system, and use information protection (encrypted information). One of the most important goals of IDS is to protect the target of the attack: user password, file systems, and kernel system.

IDS technique generally can be divided into two approaches, anomaly detection that based on acceptable behavior rules, and misuse detection that based on activity patterns and signatures. On anomaly detection, when a certain behavior diverge from the learned normal behavior then it will be recognized as an attack. Misuse detection use prior information about attack and system vulnerability to categorize an attack.

The attack itself can be divided into two categories, host-based attacks [1] and network-based attacks [2]. Host-based attacks use inner source system to infiltrate it. Network-based attacks prevent authorized users to use system source by ensure that network traffic and host become overload, one of the example is mail bomb.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

One of the research that related to IDS is Chou et al. [3] that use decision tree (C4.5) and Naïve Bayes method for data classification and feature extraction. At this paper, we will classify IDS data by using Fuzzy Kernel C-means (FKCM) and Kernel Matrix [4] for feature extraction.

2. CLASSIFICATION AND INTRUSION DETECTION SYSTEMS

As explained before, IDS is a security management system that in general, aim to protect system from attacks. One problem that rise on this field is related with classification, which is how to classify input data as an attack or normal condition event. And more, how to decide which type of attacks that occur, since there are different types of system defensive reactions for different types of attacks. The goal of this research is to classify KDD99 [5] data set that is the Benchmark data for IDS researchers. It consists of more than 500.000 data from 24 different types of attacks that can be categorized into 4 groups:

1) Denial of Service Attacks (DoS): Legitimate user access to the system is denied because the network or machine resource unavailable. Examples for this type of attacks are Apache2, Back, Land, Mail Bomb, SYN Flood, Ping of Death, Smurf, Teardrop, etc.

2) User to Super user or User to Root Attacks (U2R): Attacker access the system by using normal user account and find the system weakness to get into root system, as in Eject, Ffbconfig, Fdformat, Loadmodule, Perl, Ps, and Xterm.

3) Remote to Local Attacks (R2L): Attackers send the systems some packages through the network and look into system weakness so that they can act as local users, as in Dictionary, Ftp Write, Guest, Imap, Named, Phf, Sendmail, Xlock, and Xsnoop.

4) Probing Attacks (Probe): Attackers are looking for information at the system or its weakness by scanning the computer network, as in Ipsweep, Mscan, Nmap, Saint, and Satan.

Each KDD99 data consists of 41 features that given at Table 1.

Table 1: KDD99 Data Features

No	Feature Name
1	duration
2	protocol_type
3	Service
4	flag
5	src_bytes
6	dst_bytes
7	land
8	wrong_fragment
9	urgent
10	hot
11	num_failed_logins
12	logged_in
13	num_compromised
14	root_shell
15	su_attemted
16	num_root
17	nu_file_creations
18	num_shells
19	num_access_file
20	num_outbond_cmds
21	is_host_login
22	is_guest_login
23	count
24	srv_count
25	serror_rate
26	srv_serror_rate
27	rerror_rate
28	srv_rerror_rate
29	same_srv_rate
30	diff_srv_rate
31	srv_diff_host_rate
32	dst_host_count
33	dst_host_srv_count
34	dst_host_same_srv_rate
35	dst_host_diff_srv_rate
36	dst_host_same_src_port_rate
37	dst_host_srv_diff_host_rate
38	dst_host_serror_rate
39	dst_host_srv_serror_rate
40	dst_host_rerror_rate
41	dst_host_srv_rerror_rate
42	attack_type

3. FUZZY C-MEANS AND KERNEL METHOD

Fuzzy C-Means method (FCM) is one of fuzzy clustering method. It was found by Bezdek [6]. For a data set $X = \{x_1, x_2, \dots, x_m\} \subseteq \mathbb{R}^d$, we define $n \times c$ Membership Matrix $U = [u_{ij}]$, $1 \le i \le n$, $1 \le j \le c$, and Cluster Center $V = \{v_1, v_2, \dots, v_c\}$ where each object in V is an element of d-dimensional Euclidean Space.

Mathematical model of Fuzzy C-Means method is given by Equation (1).

Journal of Theoretical and Applied Information Technology
10th November 2015. Vol.81. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$J(U,V) = \min \sum_{i=1}^{n} \sum_{j=1}^{c} (u_{ij})^m d^2(x_i, v_j) \quad (1)$$

with constraints:

$$\begin{aligned} &\sum_{j=1}^{c} u_{ij} = 1, \ \ i = 1, 2, \dots, n \\ &\sum_{i=1}^{n} u_{ij} > 0, \ \ j = 1, 2, \dots, c \\ &u_{ij} \in [0,1], \ j = 1, 2, \dots, c \end{aligned} \quad (2)$$

where d is distance or dissimilarity function. And $m \in [1, \infty)$ is the fuzziness degree for cluster partition.

Cluster center and membership values are updated by using:

$$v_j = \frac{\sum_{i=1}^{n} u_{ij}^m x_i}{\sum_{i=1}^{n} u_{ij}^m}, \ j = 1, 2, \dots, c \quad (3)$$

And

$$u_{ij} = \left(\sum_{j=1}^{c} \left(\frac{d(x_i, v_i^t)}{d(x_i, v_j^t)} \right)^{\frac{2}{m-1}} \right)^{-1}, 1 \le i \le n \quad (4)$$

The accuracy of a FCM classification is dependent on the types of data. For a non-linearly separable data, its convergence is slow and inaccurate. To solve this problem, the data set are transformed into another space (feature space) that the dimension is much higher than the data space. It is expected that the transformed data behavior close with linearly separable data so that classification accuracy can be improved.

While working directly on feature space (high dimensional space) results in expensive cost (memory use, computational time), we need another tool as a "connector" between data space and feature space, so we can have a better accuracy without directly working at feature space. This concept is called Kernel method that was given by Vapnik [7] and elaborate further by Platt [8], Cristianini and Taylor [9], and Scholkopf et al. [10].

Set a nonlinear mapping φ from input data space $\mathbb{R}^d$ into feature space F. The clustering process take place at F other than $\mathbb{R}^d$. We need to find a way to measure distance between transformed data $\varphi(x)$ and $\varphi(y)$, x, y are objects at data space without knowing explicit form of φ. To solve this problem we use kernel function K as in [7]. By using kernel function, distance between $\varphi(x)$ and $\varphi(y)$ can be measure by:

$$\begin{aligned} d^2(\varphi(x), \varphi(y)) &= \|\varphi(x) - \varphi(y)\|^2 \\ &= \varphi(x)^t \varphi(x) - 2\varphi(x)^t \varphi(y) + \varphi(y)^t \varphi(y) \\ &= K(x,x) - 2K(x,y) + K(y,y) \end{aligned} \quad (5)$$

The success of kernel method on classification problem [9, 10] had inspired other researchers to apply kernel method on classic classification method, as in [11] that combine Fuzzy K-Medoids algorithm and kernel method to solve multiclass multidimensional data classification problem.

At this paper we use Fuzzy Kernel C-Means (FKCM) that comes by applying kernel method into FCM method to solve IDS problem, which is to classify KDD99 Benchmark data of IDS researchers [5].

4. FUZZY KERNEL C-MEANS ALGORITHM

At Karayiannis and Bezdek research about Fuzzy LVQ [12], for each iteration, a different fuzziness degree m are used, $m = m_i + \frac{t}{T}(m_f - m_i)$, where m_i and m_f are initial value and end value of m, respectively. When the value of m_f is small and m_i is quite big then it is expected that m will decreasing and vice versa.

Fuzzy Kernel C-Means Algorithm (FKCM) was build by applying kernel method into FCM, while using [12] concept to choose fuzziness degree.

Input : $X, c, m_i, m_f, \varepsilon, T$
Output : U and V
1. Initial condition: $V^0 = [v_1, v_2, \dots, v_c]$, $v_j \in C_j$
2. For $t = 1$ to T
3. $m = m_i + \frac{t(m_f - m_i)}{T}$
4. $b = -\frac{1}{m-1}$
5. Calculate membership $U^1 = [u_{ij}]$, $1 \le i \le n$, $1 \le j \le c$ by using $u_{ij} = \frac{d^b(x_i, v_j)}{\sum_{k=1}^{c} d^b(x_i, v_k)}, 1 \le i \le n, 1 \le j \le c$
6. Update cluster center $V^t = [v_1, v_2, \dots, v_c]$, where $v_j = \frac{\sum_{i=1}^{n} u_{ij}^m x_i}{\sum_{i=1}^{n} u_{ij}^m}, j = 1, 2, \dots, c$
7. If $E = \sum_{j=1}^{c} k^2(v_{jt}, v_{jt-1}) \le \varepsilon$, *STOP*.
8. $t = t + 1$

Algorithm 1: Fuzzy Kernel C-Means Algorithm

Journal of Theoretical and Applied Information Technology
10th November 2015. Vol.81. No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

According to Bezdek [6], sequences $\{U^t, V^t\}$ will converge to minimum value of $J(U,V)$.

5. EXPERIMENT RESULTS

By using Algorithm 1, data as in Section 2 of this paper and normal behavior data will be classified. Table 2 consists of explanation about KDD99 41 features data that categorized by the types of attack.

Table 2: KDD99 Data [5]

Data	Number of data	Class
Normal - DoS	488.735	2
Normal - Probe	101.384	2
Normal – U2R	97.329	2
Normal – R2L	98.403	2

At this paper, we perform two types of experiment. The first type is classification by using all data features, and the second one by selecting several features only. The selected features are chosen by using Kernel Matrix concept that was introduced by Rustam [4]. The selected features are given at Table 3.

Table 3: Selected Features by using Kernel Matrix

Classification	Selected features
Normal - DoS	38, 26, 39, 25, 5, 33
Normal - Probe	5, 6, 33, 32, 1
Normal – U2R	5, 6, 33, 1, 3, 23
Normal – R2L	5, 6, 33, 32

As a comparison for our experiment results, we use results by Chou et al. [3], with selected features are given at Table 4 that were selected by using C4.5 decision tree and Naïve Bayes methods.

Table 4: Selected Features by Chou [3]

Classification	Selected features
Normal - DoS	1-6, 12, 23, 24, 31, 32, 37
Normal - Probe	1-4, 12, 16, 25, 27-29, 30, 40
Normal – U2R	1-3, 10, 16
Normal – R2L	1-5, 10, 22

Our experiment results are given at Table 5, where the results that using all of 41 data features are given at the second column. And the third column is for the results with selected features as in Table 3. On each experiment, we use p% of data as training data and $(100-p)\%$ as the testing data to validate the algorithm, where $p = 10, 20, \ldots, 90$. The experiment was repeated 10 times and at every experiment the data were chosen randomly. At this experiment, we use Gaussian kernel $K(x,y) = \exp\left(-\frac{\|x-y\|^2}{\sigma^2}\right)$ with parameter $\sigma^2 = 0.05$. The values of Detection Rate (DR) and False Positive Rate (FPR) are counted by:

$$DR = \frac{TP}{TP+FN},\ FPR = \frac{FP}{FP+TN} \quad (6)$$

where True Positive (TP) is the number of malicious records that are classified correctly, True Negative (TN) is the number of legitimate records that are classified correctly, False Positive (FP) is the number of malicious records that are classified as legitimates one, and False Negative (FN) is the number of legitimate records that are classified as malicious records.

We use a personal computer with i-7 processor, 1 TB hard disk, 8 GB RAM, and Matlab 2012a as the computation software.

Table 5: Classification Results by FKCM

Data Set	All Features	Selected Features	
	DR	DR	FPR
Normal - DoS	96.55	100	25.72
Normal - Probe	98.03	100	2.08
Normal – U2R	52.58	87.64	2.08
Normal – R2L	97.26	98.66	0.56

Chou et al. [3] classification results are given at Table 6. Chou et al. [3] used Decision Tree C4.5 and Naïve Bayes methods, with selected features are given at Table 4.

Table 6: Classification Results by using C4.5 and Naïve Bayes [3]

Data Set	C4.5		Naïve Bayes	
	DR	FPR	DR	FPR
Normal - DoS	99.98	0.03	99.16	0.01
Normal - Probe	97.98	0.38	96.94	0.87
Normal – U2R	48.08	0	69.23	0.5
Normal – R2L	97.69	0.01	93.25	0.49

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

As we can see at Table 5, by using FKCM algorithm that use all of data features the results even though quite good for three classes of attacks, still can be improved for Normal-U2R. While by using kernel matrix method to select the features, the FKCM results are very satisfying and give a better results compare than Chou et al. [3] (Table 6) when we compare cases by cases for all 4 type of attacks. Kernel method that we combine with FCM and kernel matrix method for features selection gives a better result for classification IDS data problem.

6. CONCLUSION AND FUTURE WORKS

At this paper we propose Fuzzy Kernel C-Means algorithm as another classification method for solving IDS data classification problem. FKCM algorithm is a combination result of FCM method and Kernel Method to overcome possibilities of nonlinearly separable input data. We also use Kernel Matrix concept for feature selection that give better results than using all of the data features. Based on experiments, FKCM algorithm with Kernel Matrix selected features give a better result than using Decision Tree C4.5 and Naïve Bayes methods.

For future works, this research can be continue to find a better method to solve IDS data classification problem, and it is possible to find a nonparametric kernel that can be used to maximize kernel method superiority.

REFERENCES

[1] S. Axelsson, Research in Intrusion Detection Systems: A Survey, Technical Report TR 98-17 (revised in 1999), Chalmers University of Technology, Goteborg, Sweden, 1999.

[2] D. Marchette, "A Statistical Method for Profiling Network Traffic", *Proceedings of the First USENIX Workshop on Intrusion Detection and Network Monitoring*, Santa Clara, CA, 1999, pp. 119-128.

[3] T. S. Chou, K. K. Yen, and J. Luo, "Network Intrusion Detection Design Using Feature Selection of Soft Computing Paradigms", *International Scholarly and Scientific Research & Innovation*, Vol. 2, No. 11, 2008, pp. 458-470.

[4] Z. Rustam, "Feature Ordering Menggunakan Matriks Kernel", Proceedings of Seminar Nasional Matematika, FMIPA UI, Depok, 6 February, 2010, pp. 449-454.

[5] KDD Cup 1999 Data, http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[6] J. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, New York, Plenum Press, 1981

[7] V. N. Vapnik, The Nature of Statistical Learning Theory, New York, Springer-Verlag, 2000.

[8] J. C. Platt, Fast Training of Support Vector Machines Using Sequential Minimal Optimization. In: Advances in Kernel Methods – Support Vector Learning. MIT Press Cambridge, 1999.

[9] N. Cristianini and J. S. Taylor, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, Cambridge University Press, 2000.

[10] B. Scholkopf, A. Smola, and K. R. Muller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem", *Neural Computation*, Vol. 10, No. 5, 1998, pp. 1299-1319.

[11] Z. Rustam and A. S. Talita, "Fuzzy Kernel K-Medoids Algorithm for Multiclass Multidimensional Data Classification", *Journal of Theoretical and Applied Information Technology*, Vol. 80, Issue 1, October 2015, preprint.

[12] N. B. Karayiannis and J. C. Bezdek, "An Integrated Approach to Fuzzy Learning Vector Quantization and Fuzzy C-Means Clustering", *IEEE Trans. Fuzzy Systems*, Vol. 5, No. 4, 1997, pp. 622-628.