Advances in Social Science, Education and Humanities Research, volume 218
First International Conference on Science, Mathematics, and Education, (ICoMSE 2017)

The Application of Multi-Class Support Vector Machines on Intrusion Detection System with the Feature Selection using Information Gain

1st Zuherman Rustam
Department of Mathematics
University of Indonesia
Depok, Indonesia
rustamzr@yahoo.com

2nd Jihan Maharani
Department of Mathematics
University of Indonesia
Depok, Indonesia
jihan_m@rocketmail.com

***Abstract*—Nowadays, the intrusions often occur in a network system. One of ways that Intrusions can be prevented or detected is by using Intrusion Detection System. Therefore, IDS (Intrusion Detection System) is indispensable to detect intrusions in a network. In this paper, we will discuss the classification of IDS's data using Multi-class SVM with Information Gain Feature Selection and for the data used KDD Cup Dataset. As a result, we will discuss the accuracy of SVM combined with information gain feature selection.**

Keywords—Information gain, Intrusion detection system, Support vector machine

I. INTRODUCTION

In this modern life, technology has an important role in every aspect. Technology appears on government, education, payment, data storage, daily life, transportation, and others. For data storage, we have a lot of data we want to keep it either secretly or make it public. However, we don't want our data misused by some people who are not responsible. But, currently, there is a lot of crime that attacking our important data for examples like the occurrence of data theft, identity theft, money theft, and others. Those things happened because there are network users who he did not authorize with our network and that users want to get our special information from the network. The person behind that network crime action can be said as a hacker [2]. Therefore, we have to safely keep our data from that kind of people. In government, it's important to safely keep the data because of course they really have important data to be saved or can't make it public for the sake of country or society. So that, because of the importance of saving the data, one of the tools that can prevent undesirable activities in a network is Intrusion Detection System (IDS). Intrusion detection is monitoring the process that happening in a computer system or network and analyzing for the signs of attack or intrusion. The intrusion happened because there is a user who unauthorized with the network due to his activity purpose to get know our special information. Meanwhile, IDS itself is one of software or hardware that automatically monitoring and analyzing the process [1].

In this paper, we proposed an intrusion detection system model. We want to construct a model that has a high accuracy on detecting an intrusion and low positive false alarm. Therefore, we build a model using a classifier combined with feature selection. We choose Multi-Class Support Vector Machine (SVM) as a classifier and Information Gain (IG) as a feature selection. SVM is one of the supervised learning model that works by classifying the data based on an objective function that resulting hyperplane with the largest margin [10]. Besides, IG is one of the filtering techniques that works by selecting and sorting the feature based on entropy value that contained by each feature [5].

The purpose of the combination of SVM and IG is to see the comparison of accuracy level that resulted by IDS Model using SVM Classifier with and without IG Feature Selection as this paper's results.

II. METHODS

Dataset

In this research, we use KDDCup 99 dataset for intrusion detection system data. KDDCup 99 dataset have 41 features that we have to look for and 5 classes as a base of classification. From that 5 classes, there 4 main categories of attacks: DoS (Denial of Service), R2L (Root to Local), U2R (User to Root), and Probes; and the other category is normal.

By using this data, we want to build a predictive model capable of distinguishing between "bad" connections, called intrusions or attacks, and ``good" normal connection [4].

Table 1. List of features in KDDCUP'99 Dataset

Features	
1. duration	22. is_guest_login
2. protocol	23. count
3. service	24. srv_count
4. flag	25. serror_rate
5. src_bytes	26. srv_serror_rate
6. dst_bytes	27. rerror_rate
7. land	28. srv_rerror_rate
8. wrong_fragment	29. same_srv_rate
9. urgent	30. diff_srv_rate
10. hot	31. srv_diff_host_rate
11. num_failed_logins	32. dst_host_count
12. logged _in	33. dst_host_srv_count
13. num_compromised	34. dst_host_same_srv_rate
14. root_shell	35. dst_host_diff_srv_rate
15. su_attemted	36. dst_host_same_src_port_rate

16. num_root	37. dst_host_srv_diff_host_rate
17. nu_file_creations	38. dst_host_serror_rate
18. num_shells	39. dst_host_srv_serror_rate
19. num_access_file	40. dst_host_rerror_rate
20. num_outbond_cmds	41. dst_host_srv_rerror_rate
21. is_host_login	

Feature Selection

Feature selection is a technique that we want to add to the IDS model. By Feature selection, we can reduce the dimensionality of IDS data, removing irrelevant and redundant features. Also, Feature Selection is an important pre-processing step in machine learning and data mining due to the rapid accumulation of high-dimensional data [11]. Feature selection has 2 types of method, there are wrapper and filter methods. The difference between them, filter method is working based on the class label (supervised) and the other one is can work without the class label. We hope after we add this feature selection can make our data accurately classified and obtain a higher accurate result.

Information Gain Feature Selection

Information gain (IG) is one of many feature selection techniques. IG can be categorized as a filter feature selection. Information gain evaluates the features by the value that gained by each feature. The value of information gain feature selection based on entropy concept. As a result, IG will rank the features based on the value of each feature.

Let m be the number of classes in data set. Let D be a data set has n features and a set of classes in the other name, the dataset has n+1 attributes. Entropy (D) shows by Equation 1.

$$E(D) = -\sum_{i=1}^{m} p_i * log_2(p_i) \quad (1)$$

where p_i is a probability that each instance in D belongs to class c_i.

However, KDDCup 99 has multi label data or multi class data ($m > 2 = l$). So that, The Equation 1 modified to be a new entropy calculation, shows by Equation 2 and IG's name adapted as Multi Label Information Gain (MLIG) [5].

$$E(D) = -\sum_{i=1}^{l} p_i * log_2(p_i) + q_i * log_2 q_i \quad (2)$$

where a new p_i is a probability of each instance in D belongs to class c_i, q_i is a probability of each instance in D doesn't belong to class c_i or $q_i = 1 - p_i$, and l is the number of classes in data set.

Support Vector Machine (SVM)

Support Vector Machine (SVM) is one of binary classifiers and kind of supervised machine learning. The problem of assigning labels to record where the labels are assigned from a finite element set can be solved by SVM. Because of SVM is the binary classifier, the class labels contain only 2 values, there are +1 and -1 for each. SVM maximizes the margin when constructing the hyperplane to separate 2 classes. As a visualization. Look at the Fig. 1. Given a set of N training points of data (x_i, y_i) for $i = 1,2,\dots,N$ where x_i is the vector of input and y_i is related classes +1 and -1 for each input. In Equation 3, u is zero for separating hyperplane, w is normal vector to separate hyperplane, and x is the vector of input.

$$u = \vec{w}.x - b \quad (3)$$

For the parallel hyperplane, in Equation 3 the value of $u = \pm 1$ and the margin $m = \frac{1}{\|w\|^2}$. In equation 3, $\vec{w}$ and b show in Equation 4 & 5 where α is called as Lagrange Multiplier.

$$\vec{w} = \sum_{i=1}^{N} y_i \alpha_i \vec{x}_i, for\ \alpha_i > 0 \quad (4)$$

$$b = \vec{w}.\vec{x}_k - y_k \quad (5)$$

On the other hand, we use *kernel* for non-linier SVM. So that, $\vec{x}_i$ in Equation 4 as a kernel function so that Equation 3 becomes

$$u = \sum_{j=1}^{N} y_j \alpha_j K(\vec{x}_j, \vec{x}) - b \quad (6)$$

Kernel's function mapping input space to features space. There are 3 types of known kernel:

a. RBF kernel function :
$$\boldsymbol{k(x, x_i) = \exp\left(\frac{\|x - x_i\|^2}{\sigma^2}\right)}$$

b. Polynomial kernel function :
$$\boldsymbol{k(x, x_i) = (x.x_i + 1)^d}$$

c. Linear kernel function :
$$\boldsymbol{k(x, x_i) = x.x_i}$$

However, on our intrusion detection data which is KDDCup 99 Dataset has a multi-class type of data. So that, in this paper we use Multi-Class SVM for IDS model.

MultiClass-SVM

As we know that SVM is a binary classifier so that SVM has 2 values of class labels, there are +1 and -1. In multiclass SVM that has multi-class labels, it uses a combination of several binary SVM classifiers [6]. There are 3 methods in this classifier: *one vs one, one vs all,* and, *pairwise coupling.* The difference between them, *one vs one* analysis each pair of classes on generated SVM. *one vs all*: test the data is classified a determined value based on the biggest value if there are N number classes, there will be N decisions function which generated. *Pairwise coupling* works on all output that has gained by *one vs one* and combines them.

III. RESULTS AND DISCUSSION

On Table 1, The result of IDS Model using Multi SVM Classifier and Information Gain Feature Selection (IGFS)

with kernel using RBF, $\sigma = 0.05$ and a number of features: 25.

Table 1. Result of IDS Model using Multi-Class SVM combined with IGFS and with number of features: 25

Experiment	Number of Features	%Data Training	%Accuracy	Running Time
Normal vs DoS	25	20	100	5.86
Normal vs Probe	25	20	92.54	2.39
Normal vs U2R	25	80	98.75	10.39
Normal vs R2L	25	70	93.33	10.70
All: Normal, DoS, Probe, U2R, R2L	25	90	90.59	52.25

On Table 2, The result of IDS Model using Multi SVM Classifier without Information Gain Feature Selection (IGFS).

Table 2. Result of IDS Model using MultiClass SVM without IGFS.

Experiment	Number of Features	%Data Training	%Accuracy	Running Time
Normal vs DoS	All	20	100	5.89
Normal vs Probe	All	90	84.00	4.30
Normal vs U2R	All	90	100	9.25
Normal vs R2L	All	70	97.50	11.42
All: Normal, DoS, Probe, U2R, R2L	All	90	90.59	52.13

IV. CONCLUSION

As we can see at the table result, IDS model has a high accuracy more than 90% with the number of features 25 and less data training when using feature selection into the model. So that, The IDS model when using feature selection has a better accuracy than without feature selection, and also this purpose model has less running time and data training that more or less can represent our IDS' data. In this model, we got what we want at first, that is a high accuracy and less number of features to reduce the dimensionality and running time. We can say from the result, it's a good model when adding Information Gain Feature Selection into the IDS model.

V. REFERENCES

[1] Bace, R and mell, P., "Intrusion detection systems, National Institute of Standards and Technology (NIST)," Technical Report, pp. 800-31, 2001.
[2] H. Dalziel and W. David, Cyber Security Awareness for CEO sand Management. Chapter 3, 2016 pp. 25-29.
[3] Ikram, S. Thaseen., Cherukuri, A. Kumar., "Itrusion Detection Model Using Fusion of Chi-Square Feature Selection and Multi class SVM", 2016.
[4] KDD Cup 1999 Data, http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html
[5] Pereira, Rafael B., Plastino, Alexandre., Zadrozny, Bianca., Merschmann, Luiz H. C., Information Gain Feature Selection for Multi-Label Classification, 2015
[6] Prakash, J.Suriya., Vignesh. K. Annamalai., Ashok, C., Adithyan, T., Multi Class Support Vector Machine Classifier for Machine Vision Application.
[7] Rachman, A.A, Rustam, Z., Cancer classification using Fuzzy C-means with feature selection, 2017.
[8] Rustam, Z., Panca, V., Application of Machine Learning on Brain Cancer Multiclass Classification, 2017.
[9] Shang, Changxing., Li, Min., Feng, Shengzhong., Jiang, Qingshan., Fan, Jianping., Feature Selection Via Maximizing Global Information Gain for Text Classification. 2013.
[10] Vapnik, V., The Nature of Statistical Learning Theory. New-York: SpringerVerlag, 1995.
[11] Zheng, Zhao., Liu, Huan., Spectral Feature Selection for Supervised and Unsupervised Learning, 2007.