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  
rustamr@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

<table>
<thead>
<tr>
<th>Features</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. duration</td>
<td>22. is_guest_login</td>
</tr>
<tr>
<td>2. protocol</td>
<td>23. count</td>
</tr>
<tr>
<td>3. service</td>
<td>24. srv_count</td>
</tr>
<tr>
<td>4. flag</td>
<td>25. serror_rate</td>
</tr>
<tr>
<td>5. src_bytes</td>
<td>26. srv_serror_rate</td>
</tr>
<tr>
<td>6. dst_bytes</td>
<td>27. rerror_rate</td>
</tr>
<tr>
<td>7. land</td>
<td>28. srv_rerror_rate</td>
</tr>
<tr>
<td>8. wrong_fragment</td>
<td>29. same_srv_rate</td>
</tr>
<tr>
<td>9. urgent</td>
<td>30. diff_srv_rate</td>
</tr>
<tr>
<td>10. hot</td>
<td>31. srv_diff_host_rate</td>
</tr>
<tr>
<td>11. num_failed_logins</td>
<td>32. dst_host_count</td>
</tr>
<tr>
<td>12. logged_in</td>
<td>33. dst_host_srv_count</td>
</tr>
<tr>
<td>13. num_compromised</td>
<td>34. dst_host_same_srv_rate</td>
</tr>
<tr>
<td>14. root_shell</td>
<td>35. dst_host_diff_srv_rate</td>
</tr>
<tr>
<td>15. su_attempted</td>
<td>36. dst_host_same_src_port_rate</td>
</tr>
</tbody>
</table>
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 (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 \( n \) 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.

\[
E(D) = - \sum_{i=1}^{m} p_i \log_2(p_i)
\]

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].

\[
E(D) = - \sum_{i=1}^{l} p_i \log_2(p_i) + q_i \log_2(q_i)
\]

where a new \( p_i \) is a probability of each instance in D belongs 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 \( i = 1,2, \ldots, N \) where \( x_i \) is the vector of input and \( y_i \) is related classes +1 and -1 for each input. In Equation 3, \( \mathbf{w} \) is zero for separating hyperplane, \( \mathbf{w} \) is normal vector to separate hyperplane, and \( x \) is the vector of input.

\[
\mathbf{w} \cdot x - b = 0
\]

For the parallel hyperplane, in Equation 3 the value of \( \mathbf{w} = \pm 1 \) and the margin \( m = \frac{1}{|\mathbf{w}|} \). In equation 3, \( \mathbf{w} \) and \( b \) show in Equation 4 & 5 where \( \alpha \) is called as Lagrange Multiplier.

\[
\mathbf{w} = \sum_{i=1}^{N} y_i \alpha_i \mathbf{x}_i \quad \text{for} \quad \alpha_i > 0
\]

\[
b = \mathbf{w} \cdot \mathbf{x}_k - y_k
\]

On the other hand, we use kernel for non-liner SVM. So that, \( \mathbf{x}_j \) in Equation 4 as a kernel function so that Equation 3 becomes

\[
\mathbf{w} = \sum_{j=1}^{N} y_j \alpha_j K(\mathbf{x}_j, \mathbf{x}_i) - b
\]

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

a. RBF kernel function:

\[
k(\mathbf{x}, \mathbf{x}_i) = \exp\left(\frac{||\mathbf{x} - \mathbf{x}_i||}{\sigma^2}\right)
\]

b. Polynomial kernel function:

\[
k(\mathbf{x}, \mathbf{x}_i) = (\mathbf{x} \cdot \mathbf{x}_i + 1)^d
\]

c. Linear kernel function:

\[
k(\mathbf{x}, \mathbf{x}_i) = \mathbf{x} \cdot \mathbf{x}_i
\]

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

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Number of Features</th>
<th>%Data Training</th>
<th>%Accuracy</th>
<th>Running Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal vs DoS</td>
<td>25</td>
<td>20</td>
<td>100</td>
<td>5.86</td>
</tr>
<tr>
<td>Normal vs Probe</td>
<td>25</td>
<td>20</td>
<td>92.54</td>
<td>2.39</td>
</tr>
<tr>
<td>Normal vs U2R</td>
<td>25</td>
<td>80</td>
<td>98.75</td>
<td>10.39</td>
</tr>
<tr>
<td>Normal vs R2L</td>
<td>25</td>
<td>70</td>
<td>93.33</td>
<td>10.70</td>
</tr>
<tr>
<td>All: Normal, DoS, Probe, U2R, R2L</td>
<td>25</td>
<td>90</td>
<td>90.59</td>
<td>52.25</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Number of Features</th>
<th>%Data Training</th>
<th>%Accuracy</th>
<th>Running Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal vs DoS</td>
<td>All</td>
<td>20</td>
<td>100</td>
<td>5.89</td>
</tr>
<tr>
<td>Normal vs Probe</td>
<td>All</td>
<td>90</td>
<td>84.00</td>
<td>4.30</td>
</tr>
<tr>
<td>Normal vs U2R</td>
<td>All</td>
<td>90</td>
<td>100</td>
<td>9.25</td>
</tr>
<tr>
<td>Normal vs R2L</td>
<td>All</td>
<td>70</td>
<td>97.50</td>
<td>11.42</td>
</tr>
<tr>
<td>All: Normal, DoS, Probe, U2R, R2L</td>
<td>All</td>
<td>90</td>
<td>90.59</td>
<td>52.13</td>
</tr>
</tbody>
</table>

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