ISSN No:-2456-2165

Improved Intrusion Detection using Feature Selection

[1]C. Sudhakar Reddy, [2]Dr K. venugopal Rao
Associate Professor, Professor
[1]Dept of IT, Dept of CSE
G. Narayanamma Institute of Technology & science, Hyderabad, India

Abstract:- Intrusion Detection is one of the most important technique used in the context of security over network. Many tools are available for intrusion detection which use classification for intrusion detection. Accuracy is the most important characteristics to assess the usefulness of the tool. Accuracy of the classifier can be improved by applying feature selection methods. Many of the existing studies illustrate the application of feature selection methods improve accuracy of the model generated by classifier. In order to improve the accuracy of the classification model further in this paper we have proposed hybrid approach for feature selection. Hybrid approach uses the combination of wrapper filter and mutual information measure is used to identify the redundant attributes and remove them before applying classification algorithm.

I. INTRODUCTION

Growth of the access to internet and technologies leading to many security violations by the users of the system. In order to provide security for the system in a network firewall is most widely used solution, but many firewalls can detect only known attack for which they are configured so we need a system which can detect new possible attacks. Intrusion detection helps to detect new possible attack by performing the analysis on the meta data collected from various resources. Many of the existing researchers have used classification techniques for intrusion detection Accuracy of the model generated by classification algorithm depends on features used for selection of the relevant attributes. Many feature selection techniques are used for selecting relevant attributes, but most of the feature selection methods uses exhaustive search which is time consuming. In order to improve the performance in this paper first we have sued relative dependency to find subset of given set of features Using the reduced set we have applied wrapper methods for reducing the attributes further which will reduce the time required for exhaustive search.

II. PROPOSED METHOD

In this paper we propose an integrated method for feature selection. An integrated approach uses the combination of wrapper method and the measure called relative dependency together to identify the features relevant for the given classification. The procedure for integrated approach is shown in the figure 1

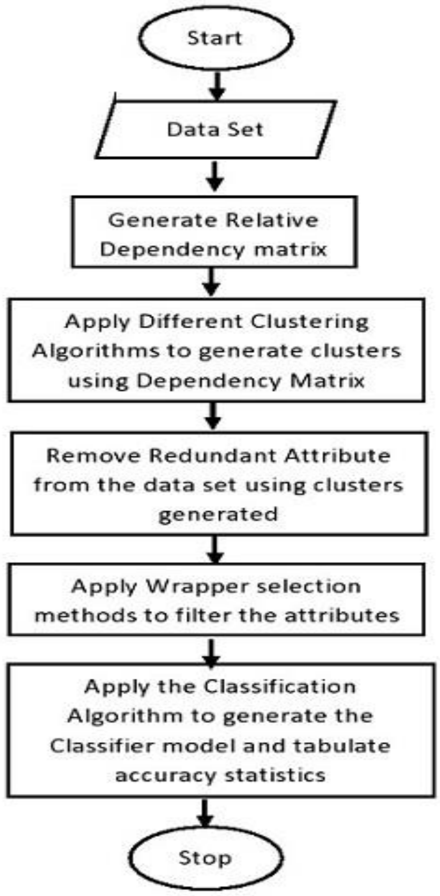

Fig 1:- Flow chart to generate model

Dependency between any two attributes namely A_i and A_j is calculated using attribute dissimilarity.

- *Attribute Dissimilarity*

Dissimilarity between the two attributes is calculated as fallows

Given two attributes A_i and A_j then dependency between A_i and A_j is represented as Dep (A_i, A_j)

$$\mathrm{Dep}(A_i, A_j) = \frac{\pi_{A_i}(R)}{\pi_{A_i, A_j}(R)}$$

$\Pi_A(R)$ indicates projection of attribute A over the relation R.

As the $Dep(A_i,A_j)$ is not symmetric we calculate the the Dependency as average of $Dep(A_i,A_j)$ and $Dep(A_j,A_i)$

The distance (dissimilarity) measure for the pair of attributes Ai and Aj is thus proposed as follows

$$Dep(A_i,A_j) = \frac{1}{Avg(Dep(A_i,A_j),Dep(A_j,A_i))}$$

Dissimilarity between various attributes is calculated and represented as a matrix.

After generating the dissimilarity matrix clusters are constructed using simple k means. Membership of the attributes is used to reduce the attributes further in order to increase the accuracy and decrease the learning time.

- *Example*

	inter	btech	tec ev	nteecev	comm	Placed
A	IB	EB	TOK	NOK	GOOD	YES
B	IA	EB	TOK	NBEST	BAD	YES
a	IA	EB	TOK	NOK	OK	YES
A	Dist	EA	TGOOD	NBEST	OK	NO
A	Dist	EA	TGOOD	NGOOD	OK	YES
A	Dist	EA	TGOOD	NGOOD	BAD	YES
A	Dist	EB	TOK	NBEST	OK	YES
A	Dist	EB	TGOOD	NBEST	OK	YES
A	Dist	EB	TOK	NOK	GOOD	YES
A	Dist	EC	TOK	NGOOD	GOOD	NO
B	IA	EC	TOK	NGOOD	BAD	NO
A	Dist	EC	TOK	NBEST	OK	NO

Table 1 (Relation R)

$\Pi_{Placed}(R)=2$
$\Pi_{Btech,}(R)=3$
$\Pi_{Btech,Palced}(R)=4$
So Dependency (B,Tech, Placed)=3/4=0.75

III. EXPERIMENTAL RESULTS

To carry out the experiment we collected data sets from UCI repositories and kaggle. The data sets size is varying from few hundreds to thousands. Comparative performance of J48 and Random Forest on various data sets is shown in the table.

REFERENCES

[1]. Mario Beraha, Alberto Maria Metelli, Matteo Papini, Andrea Tirinzoni, Marcello Restelli, "Feature Selection via Mutual Information: New Theoretical Insights" , arXiv:1907.07384v1 Jul 2019.
[2]. Bhavesh Patankar, Dr. Vijay Chavda, "Improving Classifier Performance Using Feature Selection with Ensemble Learning", 2016 IJSRCSEIT | Volume 1 | Issue 1 | ISSN : 2456-3307.
[3]. Tzung-Pei Hong ·,Yan-Liang Liou , Shyue-Liang Wang , Bay Vo " Feature selection and replacement by clustering attributes" , Vietnam J Computer Science (2014) , DOI 10.1007/s40595.
[4]. Samuel H. Huang, "Supervised feature selection: A tutorial" Artificial Intelligence Research, 2015, Vol. 4, No. 2, ISSN 1927-6974.
[5]. N. Hoque, D. K. Bhattacharyya, J. K. Kalita, "A Mutual Information-based Feature Selection Method", Expert systems with applications (Elsevier) April 21, 2014.
[6]. Jianchao Han, Ricardo Sanchez, Xiaohua Hu, T.Y. Lin, "Selection Based on Relative Attribute Dependency: An Experimental Study", 10th International Conference, RSFDGrC 2005, DOI: 10.1007/11548669_23.
[7]. Saurabh Pal , "Classification Model of Prediction for Placement of Students", DOI: 10.5815/ijmecs.2013.11.07.
[8]. Dianhong Wang, Liangxiao Jiang , "An Improved Attribute Selection Measure for Decision Tree Induction" IEEE 2007 (FSKD 2009).
[9]. https://archive.ics.uci.edu/ml/index.php.
[10]. https://www.kaggle.com/data.