Network Intrusion Detection by Multi-group Mathematical Programming based Classifier

Gang Kou[1,2], Yi Peng[*1], Yong Shi[3,1], And Zhengxin Chen[1]
[1]*College of Information Science & Technology, University of Nebraska at Omaha, Omaha, NE 68182, USA, {gkou, ypeng, yshi, zchen}@mail.unomaha.edu*
[2]*Thomson Legal & Regulatory, R&D, 610 Opperman Drive, Eagan, MN 55123, USA*
[3]*Chinese Academy of Sciences Research Center on Data Technology & Knowledge Economy, Graduate University of the Chinese Academy of Sciences, 100080, China*
** The corresponding author. Tel: ++1-402-4030269.*

Abstract

The growing number of computer network attacks or intrusions has caused huge lost to companies, organizations, and governments during the last decade. Intrusion detection, which aims at identifying and predicting network attacks, is a fast developing area that has attracted attention from both industry and academia. Technologies have been developed to detect network intrusions using theories and methods from statistics, machine learning, soft computing, mathematics, and many other fields. We have previously proposed multiple criteria linear programming (MCLP) and multiple criteria nonlinear programming (MCNP) models for two-group intrusion detection. Although these models achieve good results in two-group classification problems, they perform poorly on multi-group situations. In order to solve the problem, we introduce the kernel concept into multiple criteria models in this paper. Experimental results show that the new model provides both high classification accuracies and low false alarm rates in three-group and four-group intrusion detection.
Keywords: *Network intrusion detection, Security, Multiple criteria mathematical programming, Multi-group classification*

1. Introduction

The growing number of computer network attacks or intrusions has caused huge lost to companies, organizations, and governments during the last decade. Intrusion detection, which aims at identifying and predicting network attacks, is a fast developing area that has attracted attention from both industry and academia. Technologies have been developed to detect network intrusions using theories and methods from statistics, machine learning, soft computing, mathematics, and many other fields. Classification methods are one the major tools in network intrusion detection. A successful network intrusion detection system needs to have high classification accuracies and low false alarm rates. Different types of attacks may need different responses and multi-group classification model is needed. But it is more difficult to classify data into multiple groups than two groups.

In the past decade, researchers have successfully applied mathematical programming methods (See, for example, [4], [5], [6] and [18]) in various data mining problems, such as classification and clustering [3]. As one of such efforts, we have applied multiple criteria linear programming (MCLP) and multiple criteria nonlinear programming (MCNP) to network intrusion detection [10], [11]. However, both of these approaches have been applied to only two-group classifications. In order to solve multi-group classification problems, we re-examine the MCLP and MCNP two-group classification models and develop a kernel-based multi-group multiple criteria mathematical programming (MCMP) model. This paper has two objectives: (1) propose a new multi-group multiple criteria mathematical programming (MCMP) classification model; (2) test this model on two network intrusion datasets and compare the results with decision tree method.

The two network intrusion datasets used in this paper are: NeWT dataset and KDDCUP-99 dataset. NeWT dataset is collected using Tenable NeWT Security Scanner in the local area network.

Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06)
0-7695-2702-7/06 $20.00 © 2006 IEEE

KDDCUP-99 [15] is the network intrusion dataset used in KDD99 cup. These two datasets are different in two aspects. The first difference is the size. NeWT lab dataset has 34929 records, while KDDCUP-99 has more than 1 million records. The second difference is that NeWT lab dataset is used to classify three groups of network records and KDDCUP-99 is used to classify four groups of records in the experiments.

This paper is organized as follows. The next section discusses the formulation of kernel-based multi-group MCMP model. The third section describes the data collection process. The fourth section presents MCMP and see5 experimental results. The last section concludes the paper.

2. Multi-Criteria Mathematical Programming (MCMP) Model

This section describes the revised kernel-based multi-group MCLP model for classification.

Each row of a $n \times r$ matrix $\boldsymbol{A}=(\boldsymbol{A_1},\ldots,\boldsymbol{A_n})^{\mathrm{T}}$ is an vector $\boldsymbol{A_i} = (a_{i1},...,a_{ir}) \in \Re^r$ which corresponds to one of the records in the training dataset of a multi-group classification problem, $i = 1,..., n$; n is the total number of records in the dataset. Suppose k groups, G_1, G_2,…, G_k, are predefined. $G_i \cap G_j = \Phi, i \neq j, 1 \leq i, j \leq k$ and $A_i \in \{G_1 \cup G_2 \cup ... \cup G_k\}$, $i = 1,..., n$. A series of boundary scalars $b_1<b_2<...<b_{k-1}$ is set to separate these groups. The boundary b_j is used to separate G_j and G_{j+1}. Let $\boldsymbol{X} = (x_1,...,x_r)^T \in R^r$ be a vector of real number to be determined. Thus, we can establish the following linear inequations [8], [9]:

$A_i X < b_1$, $\forall A_i \in G_1$; (1)

$b_{j-1} \leq \boldsymbol{A}_i \boldsymbol{X} < b_j$, $\forall \boldsymbol{A}_i \in G_j$; (2)

$A_i X \geq b_{k-1}$, $\forall A_i \in G_k$; (3)

$2 \leq j \leq k-1$, $1 \leq i \leq n$.

Therefore, a single-criterion mathematical programming model for multi-group classification problems can be formulated as following:

(Model 1) *Minimize* $\frac{\| X \|_2^2}{2} + (W_\alpha + W_\zeta) \sum_{j=2}^{k-1}\sum_{i=1}^{n}(\eta_{i,j})^2 + W_\alpha \sum_{j=1,k}\sum_{i=1}^{n}(\eta_{i,j})^2 - W_\zeta \sum_{j=1}^{k}\sum_{i=1}^{n}\eta_{i,j}$ (4)

Subject to: $Y(<A \cdot X> - eb) = e - \eta$

where $e=(1,1,\ldots,1)^{\mathrm{T}}$,

$b = (\underbrace{b_1, b_1, \ldots b_1}_{n \text{ times}}, ..., \underbrace{b_{k-1}, b_{k-1}, \ldots b_{k-1}}_{n \text{ times}})^T$,

$\eta = (\eta_{1,1},\ldots,\eta_{n,1},...,\eta_{1,k-1},\ldots,\eta_{n,k-1},)^T$,

$\eta_{i,j}$ and $\boldsymbol{X}$ are unrestricted, $1 \leq i \leq n$, $1 \leq j \leq k-1$.

According to Wolfe Dual Theorem, $\nabla_X L(X,\eta,\theta) = X - A^T Y\theta = 0$,

$\nabla_\eta L(X,\eta,\theta) = 2(W_\alpha + W_\zeta)\eta - W_\zeta e - \theta = 0, 2 \leq j \leq k-2$

and $\nabla_\eta L(X,\eta,\theta) = 2W_\alpha \eta - W_\zeta e - \theta = 0, j = 1, k-1$.

Introduce the above 3 equations to the constraints of Model 1, we get:

$$Y((A \cdot A^T)Y\theta - eb) + \frac{1}{2(W_\alpha + W_\zeta)}(\theta + W_\zeta e) = e,$$

$$, 2 \leq j \leq k-2$$

$$Y((A \cdot A^T)Y\theta - eb) + \frac{1}{2W_\alpha}(\theta + W_\zeta e) = e,$$

$$, j = 1, k-1$$

$$\Rightarrow \theta = \frac{(1 + Yb + \frac{W_\zeta}{2(W_\alpha + \delta' W_\zeta)})e}{\frac{I}{2(W_\alpha + \delta' W_\zeta)} + Y((A \cdot A^T))Y},$$

$$, \delta' = \begin{cases} 1, & 2 \leq j \leq k-2 \\ 0, & j = 1, k-1 \end{cases} \quad (7)$$

Algorithm 1 summarizes the steps to solve a multi-group classification problem.

Algorithm 1

Input: a $n \times r$ matrix $\boldsymbol{A}$ as the training dataset.

Output: classification accuracies for each group in the training dataset, score for every record, and a decision function

Step 1 compute $\theta^* = (\theta_1, \ldots, \theta_n)^T$ by (7). W_ς *and* W_α are chosen by cross-validation.

Step 2 compute $X^* = A^T Y \theta^*$.

Step 3 classify an incoming A_i by using decision function $(X^* \cdot A_i)\{\in(-\infty, b_1], \Rightarrow A_i \in G_1 \in(b_{j-1}, b_j], \Rightarrow A_i \in G_j \in(b_k, +\infty), \Rightarrow A_i \in G_k$.

END

3. Network Intrusion Detection Datasets

The performance of MCMP model developed in the previous section is tested using two network datasets. The first one is called NeWT dataset which is collected and processed following the steps described in Figure 3. A free version of Tenable NeWT Security Scanner is installed in a local area network node as the attacker and Ethereal version 0.10.1 [2] is used as the data capturer in the victim machines. The local network node used to collect the data is provided by STEAL lab [16] at University of Nebraska at Omaha. Tenable NeWT Security Scanner simulates the major network intrusions by generating attacks from one network node to the others and runs the same vulnerability checks using Nessus vulnerability scanner for the Microsoft Windows platform [17]. The attack types are simulated by the sub-catalogs from Tenable NeWT Security Scanner and the normal data records consist of regular operations through networks, such as internet browsing, ftp, and data files transferring. Each file collected from the network intrusion simulation contains the connection records traced from the raw binary data by Ethereal. Each connection record encapsulates the basic TCP/IP characteristics of all the IP traffic during the lifetime of a connection. Each record has 29 fields that are delimited by coma.

Four types of network attacks are collected: denial-of-service (DOS); unauthorized access from a remote machine (R2L); unauthorized access to local root privileges (U2R); and probe. The definition and categorization of DOS, U2R, R2L, and Probe is the same as the KDDCUP-99 data [15]. Because DOS, U2R, and R2L each have a small number of data records, we group them together into one class, named "other attack". Thus, NeWT data has three classes: probe, other attack, and normal records. The total number of data records is 34929, which include 4038 Probe, 1013 other attack and 29878 Normal. In order to apply the data mining technology such as MCMP in the original data set, non-numeric attributes are either dropped or transformed into numerical type. For example, we dropped the attributes contains IP address and time sequence information, such as "Source IP", "Destination IP", "First packet time", "Last packet time" and "Elapsed time" and transformed the "Connection status" from string to numeric. Each record ends up with 23 attributes. The attributes are separated by comma and the target attribute is the last column.

The second dataset is the KDDCUP-99 data set which was provided by DARPA in 1998 for the evaluation of intrusion detection approaches. A version of this dataset was used in 1999 KDD-CUP intrusion detection contest [15]. After the contest, KDDCUP-99 has become a de facto standard dataset for intrusion detection experiments. KDDCUP-99 collects nine weeks of raw TCP dump data for a LAN simulating a typical U.S. Air Force LAN. Multiple attacks were added to the LAN operation. The raw training data was about four gigabytes of compressed binary TCP dump data from seven weeks of network traffic. A connection is a sequence of TCP packets occurred during a specified time period and each connection is labelled as either normal or attack. There are four main categories of attacks: denial-of-service (DOS); unauthorized access from a remote machine (R2L); unauthorized access to local root privileges (U2R); surveillance and other probing. Because U2R has only 52 distinct records, we test KDDCUP-99 as a four-group classification problem. The four groups are DOS (247267 distinct records), R2L (999 distinct records), and Probe (13851 distinct records), and normal activity (812813 distinct records).

4. Experimental Study and Results

4.1 Classify NeWT Lab Data into Three Groups Using MCMP and see5

For the comparison purpose, See5 release 1.19 for Windows [14], a decision tree tool, was applied to both NeWT data and KDDCUP-99 data and the results of See5 were compared with the results of MCMP and Kernel-based MCMP. See5 is chosen because it is the winning tool of KDD 99 cup [13].

The 10-fold cross validation results of MCMP, See5, and MCMP with kernel on NeWT data are

summarized in Table 1. Looking at Table 1 we see that all three methods achieve almost perfect results for Probe and excellent classifications for other attack. The difference is their performance on Normal class. The classification accuracies of See5, MCMP, and MCMP with kernel for Normal are 96.13%, 94.66%, and 98.52%, respectively. Since the classification accuracy of Normal class indicates the false alarm rates, a classifier with low classification accuracy for Normal class will generate a lot of false alarms.

4.2 Classify KDDCUP-99 Data into Four Groups Using MCMP and see5

KDDCUP-99 data is thirty times larger than the NeWT data and has four groups, which increases the classification difficulties. Table 2 shows the 10-fold cross validation results of MCMP, See5, and MCMP with kernel on KDDCUP-99 data. The overall classification accuracy, which is defined as the average of each class's accuracy, of See5, MCMP, and MCMP with kernel is 93.08%, 95.71%, and 97.2%, respectively. MCMP with kernel outperforms See5 and MCMP in every class, especially for Normal class. The high classification accuracy of MCMP with kernel leads to low false alarm rates.

5. Conclusion and Future Remark

In this paper, we have developed a highly efficient kernel-based MCMP model for multi-group classification problems and tested its applicability in network intrusion detection. The computation requires only simple matrix computation ($O(n^2r^3)$ for dense matrix and $O(nr^3)$ for sparse matrix). Furthermore, the definition of e-support vector is introduced to reduce the complexity (e.g. only 1% to 5% of the data need to be computed for the model in large scale problems). The experimental results showed that the proposed MCMP with kernel achieves better classification accuracies and lower false alarm rates than See5 and MCMP model without kernel on KDDCUP-99 dataset and a self-collected network intrusion dataset.

In network intrusion detection, both high classification accuracy and low false alarm rate are important performance criteria. Classifiers that produce high false alarm rates are distracting and annoying [1] and impede intrusion detection systems achieving efficiencies. The reported results of NeWT and KDDCUP-99 data demonstrated that the proposed model is capable of achieving both high classification accuracy and low false alarm rate, which makes it a potential choice for network intrusion detection systems

Acknowledgment

This research has been partially supported by Key Project #70531040,#70472074, National Natural Science Foundation of China; 973 Project #2004CB720103, Ministry of Science and Technology, China; and BHP Billiton Co., Australia.

References

[1] Allen J., Christie A., Fithen W., McHugh J., Pickel J., Storner E. (2000) *State of the Practice of Intrusion Detection Technologies*, Tech. Rep. CMU/SEI-99-TR-028, Software Engineering Institute, Carnegie Mellon University.

[2] Ethereal, available at: http://www.ethereal.com/.

[3] Fayyad U. M. G. Piatetsky-Shapiro, and P. Smith, From Data Mining to Knowledge Discovery: an Overview, Advance in Knowledge Discovery and Data Mining, AAAI Press/The MIT Press, Cambridge, 1996

[4] Fung , G. "Machine learning and data mining via mathematical programming-based support vector machines" Ph.D thesis, The University of Wisconsin - Madison. 2003

[5] Fung, G. and Mangasarian, O. L. Multicategory Proximal Support Vector Machine Classifiers, Machine Learning 59, 2005, 77-97.

[6] Hsu, C. W. and Lin, C. J. (2002) A comparison of methods for multi-class support vector machines, *IEEE Transactions on Neural Networks*, 13(2), 415-425.

[7] LINDO Systems Inc., *An overview of LINGO 8.0*, http://www.lindo.com/cgi/frameset.cgi?leftlingo.html;lingof.html.

[8] Kou, G., X. Liu, Y. Peng, Y. Shi, M. Wise and W. Xu, "Multiple Criteria Linear Programming to Data Mining: Models, Algorithm Designs and Software Developments" *Optimization Methods and Software* 18 (4): 453-473, Part 2 AUG 2003

[9] Kou, G., Y. Peng, Y. Shi, M. Wise and W. Xu, "Discovering Credit Cardholders' Behavior by Multiple Criteria Linear Programming" *Annals of Operations Research* 135 (1): 261-274, JAN 2005

[10] Kou, G., Peng, Y., Yan, N., Shi, Y., Chen, Z., Zhu, Q., Huff, J. and McCartney, S. (2004a) "Network Intrusion Detection by Using Multiple-Criteria Linear Programming" *2004 International Conference on Service Systems and Service Management*, July 19 to 21, Beijing, China.

[11] Kou, G., Peng, Y., Shi, Y., Chen, Z. and Chen X. (2004b) "A Multiple-Criteria Quadratic Programming Approach to Network Intrusion Detection" in Y. Shi, et al (Eds.): CASDMKM 2004, LNAI 3327, Springer-Verlag Berlin Heidelberg, pp. 145–153.

[12] MATLAB. User's Guide. The MathWorks, Inc., Natick, MA 01760, 1994-2005. http://www.mathworks.com.

[13] Pfahringer, B. (2000) Winning the KDD99 Classification Cup: Bagged Boosting, *SIGKDD Explorations* 1(2): 65-66.

[14] Quinlan, J. See5.0. (2004) [available at:http://www.rulequest.com/see5-info.html].

[15] Stolfo, S.J., W. Fan, W. Lee, A. Prodromidis, and P.K. Chan. 2000. Cost-based Modeling and Evaluation for Data Mining With Application to Fraud and Intrusion Detection: Results from the JAM Project. *DARPA Information Survivability Conference*.

[16] STEAL (Security Technology Education and Analysis Laboratory), Nebraska University Consortium on Information Assurance (NUCIA), http://nucia.ist.unomaha.edu/steal/labs.php.

[17] Tenable Network Security, available at: http://www.tenablesecurity.com/.

[18] Vapnik, V. N. (1995), *The Nature of Statistical Learning Theory,* Springer, New York.

[19] Vapnik, V. N. The Nature of Statistical Learning Theory. Springer, New York, second edition, 2000.

[20] Zheng, J., Zhuang, W., Yan, N., Kou, G., Peng, H., McNally, C., Erichsen, D., Cheloha, A., Herek, S., Shi, C. and Shi, Y., "Classification of HIV-1 Mediated Neuronal Dendritic and Synaptic Damage Using Multiple Criteria Linear Programming" *Neuroinformatics* 2 (3): 303-326 Fall 2004.

[21] Zhu, D., Premkumar, G., Zhang, X. and Chu, C.H. (2001) *Data Mining for Network Intrusion Detection: A comparison of Alternativest Methods*, Decision Sciences, Volume 32 No. 4, Fall 2001.

Table 1. NeWT Data Classification Results (Confusion Matrix)

Classified As ->	Probe	Other Attack	Normal
	See5		
Probe	4033	5	0
Other Attack	14	994	5
Normal	663	492	28723
	MCMP		
Probe	4036	1	1
Other Attack	5	995	13
Normal	1181	413	28284
	MCMP with kernel		
Probe	4037	1	0
Other Attack	13	998	2
Normal	40	403	29435

Table 2. KDDCUP-99 Classification Results (Confusion Matrix)

Classified As ->	Probe	DOS	R2L	Normal
See5				
Probe	13621	77	12	141
DOS	2656	241743	431	2437
R2L	12	2	856	129
Normal	27295	43	49747	735728
MCMP				
Probe	13466	216	145	24
DOS	1084	244967	1202	14
R2L	1	4	895	99
Normal	16313	59	7623	788818
MCMP with kernel				
Probe	13745	7	73	26
DOS	502	245720	141	904
R2L	40	0	912	47
Normal	4516	430	4092	803775