ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/280234928

A hybrid Data mining method for Intrusion and Fraud Detection in E-Banking Systems

Article · January 2014

CITATION
1

READS
415

3 authors, including:

Maryam Khademi
Islamic Azad University, South Tehran Branch
77 PUBLICATIONS **129** CITATIONS
SEE PROFILE

Behrouz Minaei
Iran University of Science and Technology
296 PUBLICATIONS **3,207** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Farsi Knowledge Graph View project

Project Related Article Project View project

All content following this page was uploaded by Maryam Khademi on 22 February 2016.

The user has requested enhancement of the downloaded file.

Copyright © 2014 American Scientific Publishers
All rights reserved
Printed in the United States of America

*Journal of
Computational Intelligence and Electronic Systems*
Vol. 3, 1–6, 2014

A Hybrid Data Mining Method for Intrusion and Fraud Detection in E-Banking Systems

Milad Malekpour[1, *], Maryam Khademi[2], and Behrouz Minae-Bidgoli[3]

[1] *Department of Computer Engineering, Islamic Azad University South Tehran Branch, Tehran, Iran*
[2] *Department of Applied Mathematics, Islamic Azad University South Tehran Branch, Tehran, Iran*
[3] *Department of Computer Engineering, Iran University of Science and Technology, Tehran, Iran*

Today Electronic Banking has become a popular business in the world. Like all other economic sectors, this new business is not totally immune from threats such as fraud and abuse. With online exchange growing, type and the number of these frauds are increasing. Computer system Intrusion and hacking is one of the ways that fraudulent and swindlers apply in their fraudulent activities. The main objective of this study is to propose a hybrid model based on clustering and classification in order to predict network intrusions with a high accuracy. A case study has been applied using KDD Cup 99 data. The advantage of this dataset is various class labels which differ in the number of records. Modeling has been applied on 10% of the records. Principle Component Analysis used for dimension reduction of data, and then the dataset clustered into 14 clusters by k-means algorithm and optimized cluster number selected by Silhouette index. Then 7 different classification algorithms including 3 base classifiers (Logistic regression, Decision Tree and Neural networks) and 2 ensemble methods including Bagging and Boosting applied to every cluster. Results show that the overall performance of boosted Neural Network has better prediction accuracy among other classifiers in all 5 classes. Moreover by comparing the modeling results with other studies it can be concluded that the proposed model in overall accuracy and also predicting 3 classes of attacks namely Dos, U2R and R2l has a better performance and higher accuracy.

Keywords: Fraud Detection, Hybrid Models, Electronic Banking, Ensemble Methods, Clustering.

RESEARCH ARTICLE

1. INTRODUCTION

Development of Information Technology along with tremendous growth in computing power has created a new area and E-Banking is one of its outcomes. These new types of businesses like other traditional ones face with risks and threats such as theft and fraud. In some cases like E-Banking these threats are highly sensitive and require more considerations.

Fraudulent activities in financial institutions and Banks cause heavy and irrecoverable losses to the organizations since financial analysts and investors rely on documents and financial statements in their decision making.[1]

Network Intrusion Detection is one of the most critical frauds in E-Banking systems.[2] Due to online transactions in financial institutions and banks, it is essential to use fraud detection systems to detect fraudulent and destructive activities with sufficient efficiency and accuracy.

An investigation has shown that review of only 2% of transactions lead to decrease in fraud losses to 1% of the total value of transactions.[3] On the other hand examining 30% of transactions can result in reducing fraud losses to %0.06.[3] According to statistics for General and online Fraud the cost of not using anti-fraud softwares has been estimated about $60 Billion in 2005.[4] Fraud detection has been also 1% of the total businesses income which has 1% increase in comparison to 2010.[5] So it's clear that using an anti-fraud solution for online banking will preserve a great amount of financial resources. These systems must have the ability to identify and predict fraudulent activities very quickly and accurately.

Moreover increasing in online data volume,[6, 7] variable types and amounts,[8] financial fraud diversity and data storing velocity[7] have led to further researches on knowledge discovery from databases and fraud detection.

To tackle these problems, this study proposes a hybrid model combining supervised and unsupervised leaning methods in order to build a model which can predict fraud activities in E-Banking. The rest of this paper is organized as follows:

In Section 2, some definitions used in the paper are formalized. Section 3 presents the clustering-based intrusion

[*] Author to whom correspondence should be addressed.

2326-3008/2014/3/001/006 doi:10.1166/jcies.2014.1068

detecting method. In Section 4, we discuss factors influencing detection results and some methods of selecting parameters. Experimental results are given in Section 5. Finally, Section 6 concludes the paper.

2. RELATED WORKS

The research reports on fraud detection, despite their similarity, are less than those on intrusion detection. Usually due to the security concerns, fraud datasets are not accessible to public.[2] So in this section an attempt has been made to review the researches which published in the field of fraud detection and intrusion detection.

Ngai et al.[9] in 2011 have surveyed application of data mining in financial fraud detection. They simply categorized data mining techniques which have been used for fraud detection as: classification, regression, clustering, prediction, outlier detection. They also refer to regression, artificial neural networks, decision trees, and Bayesian networks as the most used algorithms.

Some studies employed association rules mining for fraud detection[10, 11] and the others simply just used classification methods. Padmaja[12] used naïve bayes, K-nearest neighbor, radial base function, and C4.5 decision tree, Bhattacharyya[13] used logistic regression, support vector machine, and random forest for fraud detection. Panigrahi[14] proposed a model consists of four components, namely:

(1) rule-based filtering,
(2) Dempster–Shafer adder,
(3) transaction history database and
(4) Bayesian learner and achieved true positive rate of 98%.

Also for intrusion detection models, Altwaijry[15] applied Bayesian methods, Naidu[16] conducted a comparative report using C5 decision tree, Riper rule and support vector machines and Limkar in 2012[17] proposed a hidden Markov model.

There are also some researches for fraud and intrusion detection used unsupervised learning methods. Olszewski,[18] Quah,[4] and Zaslavsky[19] used self-organizing maps, Liang[20] applied hierarchical clustering, Jiang[21] developed a new clustering algorithm, and Lei[2] proposed two new clustering algorithms: (1) improved competitive learning network (ICLN) and (2) supervised improved competitive learning network (SICLN), Dokas[22] compared abilities of four clustering methods including Mahalanobis-distance based outlier detection, nearest neighbor approach, Density Based Local Outliers, and Unsupervised Support Vector Machines for fraud and intrusion detection.

Some of the published papers used combination of clustering and classification methods to detect fraud or intrusion. Unsupervised property of these methods helps to uncover new trends and patterns which were unknown previously, they also help to provide summarized and higher-quality training data.

Xiang in 2008[23] combined Bayesian clustering with decision tree, Wang[24] used fuzzy clustering along with artificial neural networks which applied on all cluster members separately. Horng[25] and Khan[26] proposed a two-step model using hierarchical clustering and support vector machine. Ganapathy[27] mixed fuzzy clustering with immune genetic algorithm for intrusion detection modeling.

A few articles mentioned the imbalanced data problem in fraud and intrusion detection datasets. Padmaja in 2007[12] employed sampling techniques such as under sampling and SMOTE to cope with imbalanced data. Others used ensemble methods like random forests,[13] cost sensitive decision tree,[28] and a proposed algorithm named Contrast Miner[29] to deal with this problem.

3. FRAMEWORK

Through a comprehensive literature review which is presented in this paper it can be concluded that there are less researches applied ensemble methods in E-Banking fraud detection. On the other hand clustering techniques along with classification methods which can increase the prediction model accuracy are discussed less in researches. The main objective of this paper is to propose a two steps model. In the first step k-means algorithm has been used in order to cluster data into k clusters. Then different base classification algorithms such as Logistic regression, decision tree, Artificial neural networks (ANN) and ensemble methods like Bagging and Boosting has been applied on records in every cluster. The process model of this study is demonstrated in Figure 1.

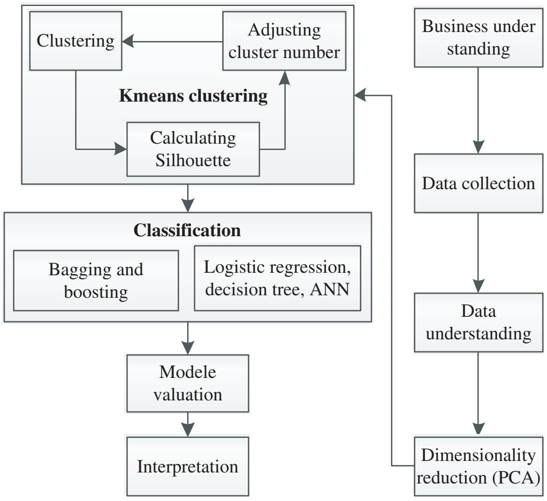

Fig. 1. Proposed model's framework.

3.1. Ensemble Learning

Most of intrusion and fraud detection datasets are suffering from class imbalance problem.[30] While the number of examples that represent one class is much lower than the ones in other classes, neither of base classifiers can solve the class imbalance problem and consequently cannot reach high prediction accuracy. So Ensemble methods can be applied in order to increase the accuracy of these models.[30] Two most used ensemble algorithms already used in this paper are Bagging and Boosting.

Bagging: Bootstrap aggregating abbreviated as bagging is based on voting. It trains different classifiers with bootstrapped replicas of the original training data-set in order to reach the diversity needed.[30] Bagging is more compatible with algorithms which are highly depended on the datasets. (ANN and Decision Tree are examples of these algorithms while KNN is not).

Boosting: due to the training error of base classifiers, this learning algorithm uses different weighted training data in each iteration.[30] Boosting attempts to generate weak learners (while weak learner is slightly better than random guessing) and turn them into a strong learner with a higher accuracy in the sense of probably approximately correct (PAC) learning framework.

4. EXPERIMENTAL DETAILS

To evaluate the performance of the proposed model a series of experiments on KDD CUP 1999 dataset were conducted. In these experiments, we implemented and evaluated the proposed methods in IBM modeler on a Windows PC with Dual-Core 1.83 GHz CPU and 2 GB RAM.

KDD CUP 1999 dataset is a version of the original 1998 DARPA intrusion detection evaluation program that is prepared and managed by the MIT Lincoln Laboratory. The dataset contains about five million connection records and a set of 41 features derived from each connection which specifies the status of connection records as either normal or 4 specific attack type namely

(1) Denial of Service (DoS),
(2) Remote to Local (R2L),
(3) User to Root (U2R) and
(4) Probing (PRB).

Table I shows detailed information about the number of all records.

Table I. Adjuster parameters.

Methods	Adjusted parameters
Logistic regression	Method: Integer (no feature selection) Maximum iteration: 20
Decision tree	Splitting criterion: Gini Minimum records in parent branch: 2% Minimum records in child branch: 1%
Artificial neural networks	Multi-layer perceptron Stopping rule: after 15 minutes
Boosting	10 iteration
Bagging	10 component models

It is important to mention that in this study, like some other researches, 10% of the original dataset samples is used to build the model to reduce the computational complexity and time.[31–33] KDD Test data is also used in order to test the proposed model.

4.1. Data Preparation

Although the KDD Cup1999 dataset presented 41 features for each network connection, not all features are needed in the design the system. It is critical to identify important features of network traffic data, in order to achieve maximal performance. In this study, features are selected based on Principal Analysis Component (PCA) which converts a set of observations of possibly correlated variables into linearly uncorrelated variables using covariance matrix and eigenvector.

In this regard the software determined to extract factors whose eigenvectors are more than 1/2 from dataset. Maximum amount of iteration of 100 was also adjusted to achieve required convergence. Finally 6 factors extracted from 41 variables in dataset which had a total explained variance of 61%.

4.2. Data Preparation

In the paper, *k*-means algorithm is employed on 6 variables extracted from PCA in the previous section. Since this method does not have the ability to determine the number of clusters, clustering phase ran with different number of

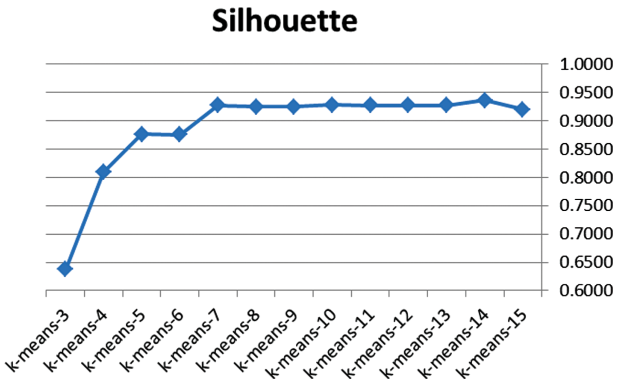

Fig. 2. Silhouette index for 13 clustering model.

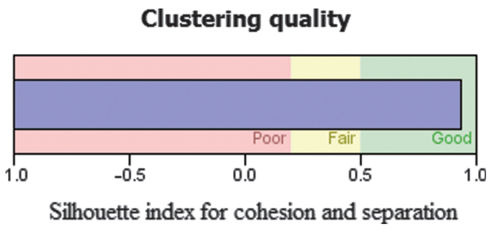

Fig. 3. Silhouette index for cohesion and separation.

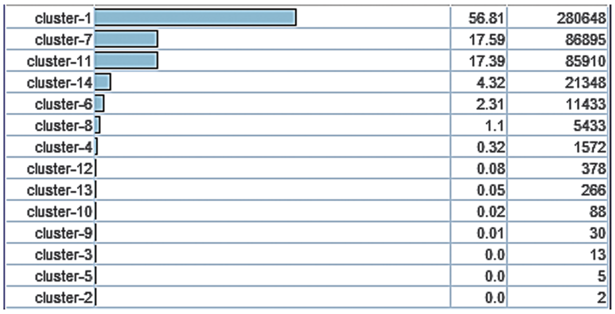

Fig. 4. Proportion of examples in each cluster.

Table II. Number of each class.

Class	Quantity
Dos	391458
Normal	97235
Probe	4032
R2L	1106
U2R	52
Total	493883

clusters (3 to 15) and a clustering validity measure is calculated to identify the optimal k. Many criteria have been developed for determining cluster validity, all of which have a common goal to find the result is a set of clustering that are as compact and well separated. In this paper we employed silhouette index which is one of the most popular measures for clustering Figure 1 indicate different values of silhouette index for each k. The *result* is a set of *clusters* that are as *compact* and *well-separated*.

The results shown in Figure 1, it can be inferred that silhouette index achieve its optimal value when k is equal to 14. Figure 2 shows when K is equal to 14, silhouette index is in good interval.

By clustering examples with 14 clusters, distribution of records in each cluster can be shown in below in Figure 3. As depicted in Figure 1. Some clusters include fewer members than other clusters. So in this step these clusters are identified as outliers and have been eliminated from training data set. So five clusters each of which has less than 0.01% of examples have been eliminated from the main dataset.

4.3. Classification

In order to classify cluster' members, seven classification methods have been used including tree base classifier: logistic regression, Classification and Regression Tree (CART), artificial neural networks and two ensemble methods: bagging and boosting which are applied to last two mentioned base classifier to increase their prediction accuracy. The lists of parameters which have been adjusted for each of models are demonstrated in Table I.

While five out of fourteen clusters have been eliminated, there were only nine clusters that form different training subsets. Table II shows the proportion of each of classes in datasets after noise elimination. In this section all of the classification methods employed on different training subsets. Due to stopping rules, decision tree could not built the model using some of training subsets so the result of this method is not listed here.

Some measurements such as true positives, true negatives, false positives and false negatives are often proposed to evaluate the detection precision of IDS. In this study we used Accuracy in order to evaluate our model. Table III contains the total quantity of correct prediction of each class. Accuracy of each method can be calculated using total correctly predicted records divided by total number of records.

According to Table III it is clear that Boosting ANN and Bagging ANN are the most accurate methods for intrusion

Table III. Accuracy of all methods.

Methods	Normal	Dos	U2R	R2L	Probe	Corrected classified	Total accuracy (%)
Logistic regression	92106	376612	12	316	1930	470976	95.36
Boosting decision tree	96671	391300	26	788	3922	492707	99.76
Bagging decision tree	96798	390826	0	341	3585	491550	99.53
Artificial neural networks (ANN)	96923	391232	0	877	3842	492874	99.80
Boosting ANN	96962	391254	20	920	3926	493082	99.84
Bagging ANN	97020	391312	1	826	3933	493092	99.84

Table IV. Accuracy of all methods.

Methods	Normal (%)	Dos (%)	U2R (%)	R2L (%)	Probe (%)
Logistic regression	94.73	96.21	23.08	28.57	47.87
Boosting decision tree	99.42	**99.96**	**50.00**	71.25	97.27
Bagging decision tree	99.55	99.84	0.00	30.83	88.91
Artificial neural networks (ANN)	99.68	99.94	0.00	**79.29**	95.29
Boosting ANN	**99.72**	99.95	**38.46**	**83.18**	**97.37**
Bagging ANN	**99.78**	**99.96**	1.92	74.68	**97.54**

Table V. Testing the proposed model.

Methods	Normal (%)	Dos (%)	U2R (%)	R2L (%)	Probe (%)	Total accuracy (%)
Logistic regression	96.33	74.89	58.97	3.12	68.87	77.81
Boosting decision tree	3.31	99.80	28.21	6.94	82.33	77.74
Bagging decision tree	98.47	99.80	25.64	1.95	78.08	97.13
Artificial neural networks (ANN)	3.16	99.87	0.00	5.24	87.21	77.77
Boosting ANN	3.22	99.74	0.00	1.57	88.77	77.62
Bagging ANN	98.22	99.65	25.64	4.64	89.31	97.3

Table VI. Comparison of results with similar studies.[34–36]

Methods	Year	Normal (%)	Dos (%)	U2R (%)	R2L (%)	Probe (%)	Total accuracy (%)
Proposed model		98.22	**99.65**	**25.64**	4.64	89.31	**97.31**
Horng (hierarchical clustering + SVM)	2011	99.29	99.53	19.73	**28.81**	**97.55**	95.72
Toosi and kahani (ESC-IDS)	2007	98.2	99.5	14.1	31.5	84.1	95.3
Xuren (association rule)	2006	**99.5**	96.8	3.8	7.9	74.9	–
Pfahringer (KDD 99 winner)	2000	**99.5**	97.1	12.3	8.4	83.3	91.8

detection prediction in this study. Among all the presented methods, artificial neural network's base methods are more accurate. On the other hand logistic regression is the less accurate one.

However it should be noted that the total accuracy alone is not a good indicator for evaluating the Strengths of methods due to this fact that the best methods should also correctly classified each of separate classes. The percentages of correctly classified samples are given in Table IV.

In Table IV, methods with high accuracy are highlighted in bold face. As it is demonstrated in the table, ensemble methods achieve higher accuracies. The methods based on artificial neural networks also perform better than decision tree and logistic regression. Bagging ANN is better in predicting the majority class (Prob, Dos, Normal) while boosting ANN is more successful in prediction of minority classes (L2R and U2R). Low accuracy in predicting U2R attacks is the main weakness of bagging ANN.

5. RESULTS AND DISCUSSION

In this section further investigations have been conducted in order to analyze the validity of proposed model using KDD cup test set. Table V contains the result of applying different methods on the test set. This dataset include some new attacks which do not exist in training set. In this study according to unsupervised property of proposed model this new attacks are removed from test set.

As shown in Table IV, bagging ANN is the best method among all the base and ensemble methods. Boosting ANN is also well performed in predicting individual classes.

We can compare our results with the results of existing studies in the literature. Table VI demonstrates results of accuracy obtained from other studies in comparison to ours which reveals that to some extent the proposed model achieves the higher accuracy.

According to Table VI it is clear that proposed model perform better than other researches in the prediction of Dos and U2R attacks. It can also be noted that the total accuracy of proposed model is significantly higher among other studies.

6. CONCLUSIONS

In this paper, we mainly investigated ensemble methods for fraud and intrusion detection and proposed a hybrid model using clustering and classification methods, applied it to a real dataset.

Compared to other researches by the simulation experiments, the test results indicate that the hybrid model combined with ensemble methods can significantly improve the overall accuracy and also it cause better performance in predicting some of classes.

Several future research directions also emerge. Firstly, other types of feature reduction of Feature Subset Selection (FSS) techniques should be applied to reduce the complexity of models and increase the accuracy. In this case FSS techniques which are based on optimizing the solution space like genetic algorithm are preferable. Secondly, further analyses are encouraged to use other robust classification methods like support vector machines (SVM). Thirdly, for further clarification of model performance, the proposed model should apply on other real world data sets.

References

1. H. A. Ata and I. H. Seyrek, The use of data mining techniques in detecting fraudulent financial statements: An application on manufacturing firms. *Journal of Faculty of Economics and Administrative Sciences* 14, 157 (**2009**).
2. J. Z. Lei and A. A. Ghorbani, Improved competitive learning neural networks for network intrusion and fraud detection. *Neurocomputing* 75, 135 (**2012**).
3. T. P. Bhatla, V. Prabhu, and A. Dua, Understanding credit card frauds. *Cards Business Review* 1 (**2003**).
4. J. T. Quah and M. Sriganesh, Real-time credit card fraud detection using computational intelligence. *Expert Systems with Applications* 35 (**2008**).

5. CyberSource, http://www.cybersource.com/ **(2012)**.

6. N. A. K. Le, S. Markos, M. O'Neill, A. Brabazon, and M. T. Kechadi, An investigation into data mining approaches for anti money laundering, *Proceedings of International Conference on Computer Engineering and Applications (ICCEA 2009)* **(2009)**.

7. T. Pang-Ning, M. Steinbach, and V. Kumar, Introduction to data mining. *In Library of Congress* **(2006)**.

8. R. J. Bolton and D. J. Hand, Statistical fraud detection: A review. *Statistical Science* 17, 235 **(2002)**.

9. E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems* 50, 569 **(2011)**.

10. D. Sánchez, M. A. Vila, L. Cerda, and J. M. Serrano, Association rules applied to credit card fraud detection. *Expert Systems with Applications* 36, 3630 **(2009)**.

11. D. Fu, S. Zhou, and P. Guo, The design and implementation of a distributed network intrusion detection system based on data mining, *WRI World Congress on Software Engineering, 2009 WCSE'09*, IEEE, May **(2009)**, Vol. 3, pp. 446–450.

12. T. M. Padmaja, N. Dhulipalla, P. R. Krishna, R. S. Bapi, and A. Laha, An unbalanced data classification model using hybrid sampling technique for fraud detection, *Pattern Recognition and Machine Intelligence*, Springer, Berlin, Heidelberg **(2007)**, pp. 341–348.

13. S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, Data mining for credit card fraud: A comparative study. *Decision Support Systems* 50, 613 **(2011)**.

14. S. Panigrahi, A. Kundu, S. Sural, and A. K. Majumdar, Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning. *Information Fusion* 10, 354 **(2009)**.

15. H. Altwaijry, Bayesian based intrusion detection system, *IAENG Transactions on Engineering Technologies*, Springer, Netherlands **(2013)**, pp. 29–44.

16. R. N. C. Appala and P. S. Avadhani, A comparison of data mining techniques for intrusion detection, *2012 IEEE International Conference on, Advanced Communication Control and Computing Technologies* (*ICACCCT*) IEEE **(2012)**, pp. 41–44.

17. S. Limkar and R. K. Jha, An effective defence mechanism for detection of DDoS attack on application layer based on hidden markov model, *Proceedings of the International Conference on Information Systems Design and Intelligent Applications*, (INDIA 2012) held in Visakhapatnam, India, Springer, Berlin, Heidelberg, January **(2012)**, pp. 943–950.

18. D. Olszewski, J. Kacprzyk, and S. Zadrożny, Employing self-organizing map for fraud detection, *Artificial Intelligence and Soft Computing*, Springer, Berlin, Heidelberg, January **(2013)**, pp. 150–161.

19. V. Zaslavsky and A. Strizhak, Credit card fraud detection using self-organizing maps. *Information and Security* 18, 48 **(2006)**.

20. H. Liang, R. Wei-Wu, and R. Fei, Anomaly detection using improved hierarchy clustering, *AICI'09. International Conference on Artificial Intelligence and Computational Intelligence*, IEEE **(2009)**, Vol. 1, pp. 319–323.

21. S. Jiang, X. Song, H. Wang, J. J. Han, and Q. H. Li, A clustering-based method for unsupervised intrusion detections. *Pattern Recognition Letters* 27, 802 **(2006)**.

22. D. Paul, L. Ertoz, V. Kumar, A. Lazarevic, J. Srivastava, and P.-N. Tan, *Proc. Nsf Workshop on Next Generation Data Mining*, Baltimore **(2004)**.

23. C. Xiang, P. C. Yong, and L. S. Meng, Design of multiple-level hybrid classifier for intrusion detection system using Bayesian clustering and decision trees. *Pattern Recognition Letters* 29, 918 **(2008)**.

24. G. Wang, J. Hao, J. Ma, and L. Huang, A new approach to intrusion detection using artificial neural networks and fuzzy clustering. *Expert Systems with Applications* 37, 6225 **(2010)**.

25. S. J. Horng, M. Y. Su, Y. H. Chen, T. W. Kao, R. J. Chen, J. L. Lai, and C. D. Perkasa, A novel intrusion detection system based on hierarchical clustering and support vector machines. *Expert systems with Applications* 38, 306 **(2011)**.

26. L. Khan, M. Awad, and B. Thuraisingham, A new intrusion detection system using support vector machines and hierarchical clustering. *The VLDB Journal—The International Journal on Very Large Data Bases* 16, 507 **(2007)**.

27. S. Ganapathy, K. Kulothungan, P. Yogesh, and A. Kannan, A novel weighted fuzzy C–means clustering based on immune genetic algorithm for intrusion detection. *Procedia Engineering* 38, 1750 **(2012)**.

28. Y. Sahin, S. Bulkan, and E. Duman, A cost-sensitive decision tree approach for fraud detection. *Expert Systems with Applications* 40, 5916 **(2013)**.

29. W. Wei, J. Li, L. Cao, Y. Ou, and J. Chen, Effective detection of sophisticated online banking fraud on extremely imbalanced data. *World Wide Web* 27 **(2013)**.

30. M. Re and G. Valentini, Ensemble Methods: A Review **(2011)**.

31. Y. Yi, J. Wu, and W. Xu, Incremental SVM based on reserved set for network intrusion detection. *Expert Systems with Applications* 38, 7698 **(2011)**.

32. W. Huai-Bin, Y. Hong-Liang, X. Zhi-Jian, and Y. Zheng, A clustering algorithm use SOM and K-means in intrusion detection, *International Conference on E-Business and E-Government* (*ICEE*), IEEE, May **(2010)**, pp. 1281–1284.

33. S. K. Sharma, P. Pandey, S. K. Tiwari, and M. S. Sisodia, An improved network intrusion detection technique based on k-means clustering via Naïve bayes classification, *International Conference on Advances in Engineering, Science and Management* (*ICAESM*), IEEE, March **(2012)**, pp. 417–422.

34. Pfahringer and B. Winning, The KDD99 classification cup: Bagged boosting. *ACM SIGKDD Explorations Newsletter* 1, 65 **(2000)**.

35. A. N. Toosi and M. Kahani, A new approach to intrusion detection based on an evolutionary soft computing model using neuro-fuzzy classifiers. *Computer Communications* 30, 2201 **(2007)**.

36. W. Xuren, H. Famei, and X. Rongsheng, Modeling intrusion detection system by discovering association rule in rough set theory framework, *International Conference on Computational Intelligence for Modelling, Control and Automation, 2006 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce*, IEEE, November **(2006)**, pp. 24–24.

Received: xx Xxxx xxxx. Accepted: xx Xxxx xxxx.

View publication stats