Application of Fuzzy Association Rules-Based Feature Selection and Fuzzy ARTMAP to Intrusion Detection

Mansour Sheikhan[1], Maryam Sharifi Rad[1], Hossein M. Shirazi[2]
1- Department of Electrical and Computer Engineering, Islamic Azad University, South Tehran Branch, Tehran, Iran.
Email: msheikhn@azad.ac.ir, taravat.sharifi@gmail.com
2- Faculty of ICT, Malek-Ashtar University of Technology, Tehran, Iran
Email: shirazi@mut.ac.ir

Received: September 2010 Revised: January 2011 Accepted: April 2011

ABSTRACT:
Intrusion Detection System (IDS) deals with a very large amount of data that includes redundant and irrelevant features. Therefore, feature selection is a necessary data pre-processing step to design IDSs that are lightweight. In this paper, a novel feature selection method based on data mining techniques is proposed, which uses fuzzy association rules to obtain the optimum feature subset. In this research, the fuzzy ARTMAP neural network is used as the classifier to evaluate the goodness of the obtained feature subset. The effectiveness of proposed method is evaluated by experiments on KDD Cup99 dataset. According to the performance comparisons with some other machine learning methods that have used the same dataset, the proposed method is the most efficient on detection rate, false alarm rate and cost per example.

KEYWORDS: intrusion detection, feature selection, fuzzy association rules, fuzzy ARTMAP.

1. INTRODUCTION

In the recent decade, some of the activities over the Internet such as internet banking, online shopping and electronic commerce have been developed. Hence, the number of intrusions into computer networks has been grown extensively. The reason is that automated intrusive tools are emerging every day. Therefore, intrusion detection system (IDS) plays a momentous role in detecting different kinds of attacks. In general, IDSs fall into two categories according to the detection methods that they employ, namely 1) misuse detection and 2) anomaly detection. Misuse detection identifies intrusions by matching observed data with pre-defined descriptions of intrusive behavior. While anomaly detection hypothesizes that abnormal behavior is rare and different from normal behavior. Hence, it builds models for normal behavior and detects anomaly in observed data by noticing deviations from these models [1].

Up to now, several research and methods of intrusion detection have been developed. However, there is a growing interest in intrusion detection community toward the application of machine learning techniques in this field. Considering this trend and the extensive amount of data involved in intrusion detection problem, data mining approaches seem to be appropriate for this purpose [2]-[4].

In general, IDS deals with the huge amount of data which contains irrelevant and redundant features causing slow training and testing process, higher resource consumption as well as poor detection rate [5]. The significance of feature selection can be viewed in two aspects: 1) for filtering out the noise and removing redundant and irrelevant features, 2) as an optimization procedure of search for an optimal subset of features that better satisfy a desired measure [6]. Although using feature selection is not a very popular procedure in intrusion detection. However, some of the studies use different feature selection methods for their experiments. This implies that feature selection could improve some certain level of classification accuracy in intrusion detection [2]. For example, Chebrolu et al. in [7] have combined the genetic algorithm with decision tree classifiers to find an optimal subset of features for decision tree classifiers. In [8] neural networks and support vector machine have been applied for feature selection in the intrusion detection system. In [9], sequential backward floating search has been proposed to find an optimal subset of features in intrusion detection.

Until now, to our knowledge, using the fuzzy association rules as a feature selection method has not been tried for intrusion detection problem. In this study, the feasibility of applying fuzzy association rules for feature selection in the intrusion detection systems will be demonstrated. To do this, a feature selection engine based on fuzzy association rules (FSE-FAR) is developed and a fuzzy ARTMAP neural network is

used for classification, as well.

The rest of paper is organized as follows. In section 2, the main concepts related to the methodology used in this work are described. In section 3, the proposed framework for intrusion detection is introduced in details. In section 4, the results of the experiments, carried out on knowledge discovery and data mining group (KDD) dataset, are presented and compared with some recent works in literature using the same dataset. Finally, section 5 draws conclusions.

2. METHODOLOGY

The objective of data mining is to obtain useful and non-explicit information from data stored in large repositories. One important topic in data mining research is concerned with the discovery of interesting association rules. Association rules determine interesting relationships between the large sets of data items. This technique was initially applied to the so-called market basket analysis, which aims at finding regularities in shopping behavior of customers of supermarkets [10].

2.1. Association Rules

Given an itemset I and a transaction set T, where each transaction is a subset of I, an association rule is said to be an "implication" of the form $A \Rightarrow C$ denoting the presence of itemsets A and C in some of the T transactions, assuming that $A,C \subset I$, $A \cap C = \varnothing$; and $A,C \neq \varnothing$. The usual measures proposed in [11] for establishing an association rule's fitness are the *support*(Supp($A \Rightarrow C$), the joint probability $p(A \cup C)$), and the *confidence*(Conf($A \Rightarrow C$), the conditional probability $p(C|A)$).

Apriori [11] is the best known basic algorithm to find quickly boolean association rules. In contrast to Boolean association rules, which handle only simple item-based transactions, the next generation of association rules faced quantitative attributes which their values were elements of continuous domains such as a real number domain R. However, the typical Apriori algorithm was not capable of dealing directly with such attributes. Therefore, in [12] an algorithm has been proposed to mine quantitative association rules. This algorithm starts by partitioning the attribute domains and then transforming the problem into a binary one. This method can solve problems introduced by quantitative attributes, but it causes the "sharp boundary" problem. In other words, it either ignores or over-emphasizes the elements near the boundary of intervals in the mining process. As a remedy to the sharp boundary problem, the fuzzy set concept, introduced by Zadeh [13], has been used more frequently in mining quantitative association rules. This approach is better than partitioning method, because fuzzy sets provide a smooth transition between members and non-members of a set and increase the flexibility of systems. In this study, the use of fuzzy association rules is considered as the key component of proposed approach because of the affinity with the human knowledge representation.

2.2. Fuzzy Association Rules

Mining fuzzy association rules is the discovery of association rules, using fuzzy set concepts, such that the quantitative attributes can be handled. Let $I=\{i_1,\ldots,i_m\}$ be an item set and T a fuzzy transaction set, in which each fuzzy transaction is a fuzzy subset of I. Given the transaction $t \in T$, we will use $t(i)$ to denote the membership degree of item I in the transaction t. Various proposals for fuzzy association rules can be found in the literature such as generalization of association rules when initial data are fuzzy [14]-[18]. An interesting in depth study into the extensions to quantitative attribute cases can be found in [18]. In this study, fuzzy grids based rules mining algorithm (FGBRMA) [18] is used to mine fuzzy association rules. In this algorithm, each attribute is viewed as a linguistic variable and the variables are divided into various linguistic terms. FGBRMA is an efficient algorithm since it scans the database only once and applies boolean operations on tables to generate large fuzzy grids and fuzzy association rules.

3. PROPOSED IDS ARCHITECTURE

The proposed framework for intrusion detection has composed of two modules; FSE-FAR module and classification module. Figure 1 shows a schematic view of the proposed intrusion detection system. In the FSE-FAR module, the system uses a fuzzy data mining algorithm to generate fuzzy association rules. The fuzzy association rules can discover relationships between features in dataset. So a subset of the features discovered by the fuzzy data mining algorithm is used as fuzzy ARTMAP inputs.

3.1. FSE-FAR Module

In this study, KDD'99 dataset [19] is used to train and test the proposed intrusion detection framework. The detail of KDD'99 dataset is described in section 4.1. The FSE-FAR module comprises the following three stages:

1. Defining Fuzzy Membership Functions

In the case of KDD dataset there are totally 41 features used to describe each session. To define fuzzy membership functions, each feature value is transformed to three linguistic terms (Low, Medium, and High). In other words, each feature is divided into three sub-features with the linguistic term. A predefined membership function is assigned to each feature, and the linguistic terms can be expressed by the membership function shown in Figure 2. The

parameters α, β, and γ in the fuzzy membership function for feature F_i are set as follows [20]:
β : average value of feature F_i in the dataset;
γ : the largest value of feature F_i in the dataset;
$\alpha = 2\beta - \gamma$.

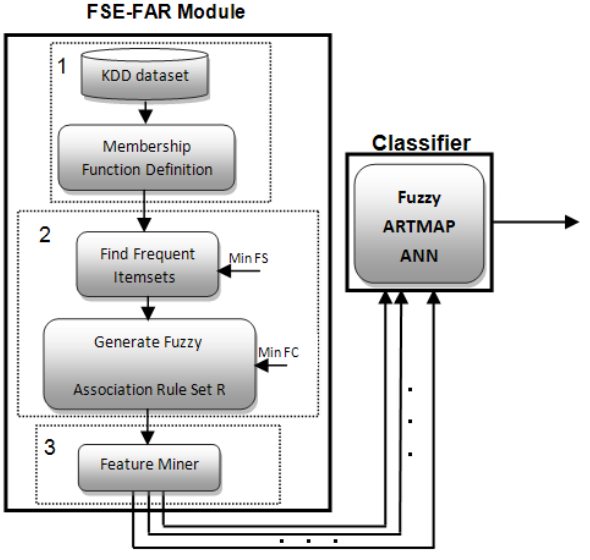

Fig. 1. Block diagram of proposed intrusion detection framework

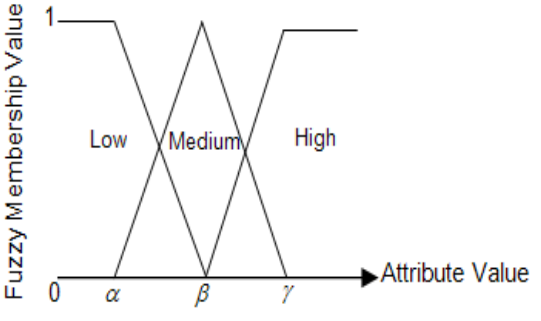

Fig. 2. Definition of fuzzy membership function

Table 1 shows an example of a small database with two features. Figures 3 and 4 show the fuzzy membership functions for features F_1 and F_2, respectively. Table 2 shows the database with fuzzy membership values after the transformation using the membership functions.

Table 1. Example of a small database

Identifier	F_1	F_2
1	10	1000
2	20	800
3	30	600
4	40	400
5	50	200

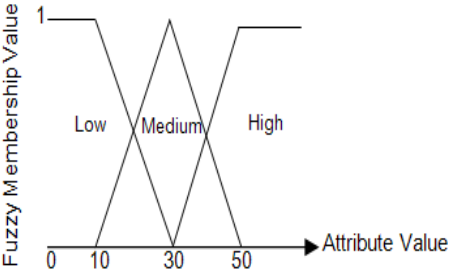

Fig. 3. Membership function for feature F_1

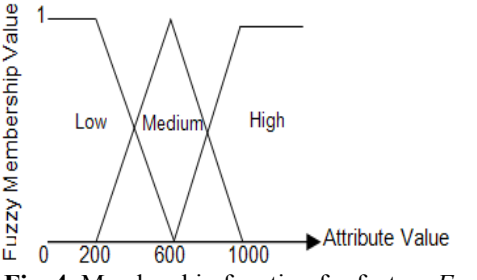

Fig. 4. Membership function for feature F_2

Table 2. Database with fuzzy membership values

Identifier	Feature F_1			Feature F_2		
	Low F'_{11}	Med. F'_{12}	High F'_{13}	Low F'_{21}	Med. F'_{22}	High F'_{23}
1	1.0	0	0	0	0	1.0
2	0.5	0.5	0	0	0.5	0.5
3	0	1.0	0	0	1.0	0
4	0	0.5	0.5	0.5	0.5	0
5	0	0	1.0	1.0	0	0

2. Search for Fuzzy Association Rules

In this study, as we mentioned earlier, FGBRMA algorithm [18] is used to mine fuzzy association rules. In this stage, the frequent item sets are found by computing the fuzzy support counts of candidate itemsets. To check whether each candidate item set is large or not, its fuzzy support is computed. When its fuzzy support is larger than or equal to the pre-determined minimum fuzzy support (Min FS), it can be said that it is a frequent item set. After finding all of the frequent item sets, fuzzy association rules are generated using frequent item sets. To check whether each r rule is acceptable or not, its fuzzy confidence is computed. When its fuzzy confidence is larger than or equal to the pre-determined minimum fuzzy confidence (Min FC), the rule is considered as an acceptable rule. (Note: Min FS and Min FC are the thresholds that are determined by the user).

3. Feature Extraction

The feature miner unit is the most important component of the proposed framework. The aim of this unit is to segregate the irrelevant and redundant features from original dataset. It finds the relationships among features in rule set R and then eliminates some unnecessary features. Suppose r rule form $X \Rightarrow Y$; where

X is the antecedent and *Y* is the consequence. In this rule, *Y* item set depends on *X* item set. Thus, all items in
Y item set can be eliminated because they are redundant. Figure 5 shows the details of feature miner algorithm.
This algorithm for each rule *r*, employs *interesting(r)* Boolean function to determine whether the rule *r* is interesting. If *interesting(r)* function (Figure 6) returns "True," then the linguistic variables that are only appeared in the antecedent of the rule *r* are extracted and all of them are added to the S_f. Then, a simple deletion procedure is performed that cancels all of the linguistic variables covered by the rule *r* from the rule set *R*.

```
begin
1)  Sf=∅; //The final feature subset
2)  if (Rule Set R==∅) then break
3)  else
4)     for each rule[r]∈R
5)        if interesting(r) then
6)        Sf=Sf+(extract_linguistic_variables_
          in_antecedent[r]);
7)        Update R by deletion of all linguistic
          variables covered by the rule[r];
8)     end for;
9)  end if;
10) return Sf;
end
```

Fig. 5. Feature miner algorithm

```
bool interesting(r)
1) begin
2)   if(fuzzy confidence[r]≥size-adjustment)
3)     return true:
4)   else
5)     return false;
6) end
```

Fig. 6. Interesting(r) function

At the final step, the feature set S_f will be the result of the feature selection process.

In *interesting(r)* function, the *size-adjustment* parameter controls the size of the feature subset. This parameter is a threshold which is determined by the user. Choosing smaller values for *size-adjustment* parameter leads to larger size feature subsets. Thus, a suitable value of the *size-adjustment* parameters should be determined by the user.

3.2. Classification Module

The fuzzy ARTMAP neural network is used as a classification tool to report the usefulness of the proposed feature subset. This network achieves a synthesis of fuzzy logic and adaptive resonance theory (ART) neural networks by exploiting a close formal similarity between the computations of fuzzy method and ART category choice, resonance and learning [21]. It is composed of two fuzzy ART modules, ART_a and ART_b, interconnected by an inter-ART using an associative memory module as illustrated in Figure 7. The inter-ART module has a self-regulator mechanism, match tracking; whose objective is to maximize the generalization and minimize the network error. The F_2^a layer is connected to the inter-ART module by the weights w_{jk}^{ab}. The steps of fuzzy ARTMAP algorithm are summarized as follows.

1. *Input data*: The input pattern of ART_a is represented by the vector $a = [a_1 \ldots a_{Ma}]$ and the input pattern of ART_b is represented by the vector $b = [b_1 \ldots b_{Mb}]$.
2. *Parameters*: There are three fundamental parameters corresponding to the performance and learning of fuzzy ART network [22].

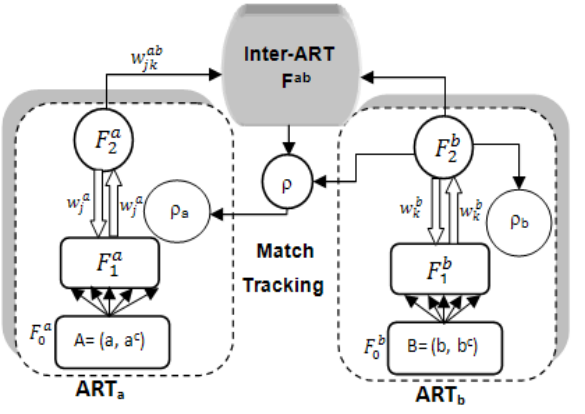

Fig. 7. Structure of fuzzy ARTMAP

- The choice parameter, ($\alpha > 0$): which acts on the category selection.
- Learning rate, ($\beta \in [0,1]$): that controls the velocity of network adaptation.
- Vigilance parameter, ($\rho \in [0,1]$): that controls the network resonance. The vigilance parameter is responsible for the number of formed categories.

3. *Algorithm structure*: After the resonance is confirmed in each network, *J* is the active category for the ART_a network, and *K* is the active category for the ART_b network. The next step is match-tracking to verify, if the active category on ART_a corresponds to the desired output vector presented to ART_b. The vigilance criterion is given by [22]:

$$\rho_{ab} = \frac{\left| y^b \wedge w_{JK}^{ab} \right|}{\left| y_b \right|} \qquad (1)$$

4. *Learning*: After the input has completed the resonance state by vigilance criterion, the weight

adaptation is implemented. The adaptation of the ART_a and ART_b module weights is given by [22]:

$$w_J^{new} = \beta\left(I \wedge w_J^{old}\right) + \left(1-\beta\right)w_J^{old} \quad (2)$$

4. EXPERIMENTAL RESULTS

4.1. Dataset

As mentioned before, KDD dataset [19] is used to evaluate the proposed framework for intrusion detection. This dataset is a common benchmark for evaluation of intrusion detection techniques. KDD'99 consists of several components, that two of them are used in this work. This dataset contains a number of connection records where each connection is a sequence of packets containing values of 41 features. Furthermore, attack types in this dataset fall into four main categories: denial of service (DoS), probe, user to root (U2R), and remote to local (R2L). In our experiments, '10% KDD' dataset has been used for training and 'Corrected KDD' dataset has been used as a test set. Several new (never-before-seen) attacks have been used in 'Corrected KDD' in order to assess the generalization ability of IDS. Statistical details of the two mentioned KDD components are summarized in Table 3.

Table 3. Characteristics of KDD'99 components used for training and test

KDD Dataset	No. of Attack Patterns	No. of Normal Patterns	No. of Total Patterns
10%	396,743	97,278	494,021
Corrected	250,436	60,593	311,029

Table 4. Cost matrix values for KDD'99

Actual \ Predicted	Normal	Probe	DoS	U2R	R2L
Normal	0	1	2	2	2
Probe	1	0	2	2	2
DoS	2	1	0	2	2
U2R	3	2	2	0	2
R2L	4	2	2	2	0

4.2. Evaluation Criteria

Before discussing about the results of experiments, it seems necessary to mention the standard metrics that had been developed for evaluating IDS. Detection rate (DR) and false alarm rate (FAR) are the two the most common metrics. DR is computed as the ratio between the number of correctly detected attacks and the total number of attacks, while FAR is computed as the ratio between the number of normal connections that are incorrectly misclassified as attacks and the total number of normal connections. Another metric that is used here is the classification rate (CR). Classification rate for each class of data is computed as the ratio between the number of test instances correctly classified and the total number of test instances of this class. For classifier algorithm evaluation, another comparative measure is defined, which is cost per example (CPE) [23]. CPE is calculated using the following formula:

$$CPE = \frac{1}{N_T}\sum_{i=1}^{m}\sum_{j=1}^{m} CM(i,j).C(i,j) \quad (3)$$

where CM and C are confusion matrix and cost matrix, respectively. N_T represents the total number of test instances and m is the number of classes in classification. CM is a square matrix in which each column corresponds to the predicted class, while rows correspond to the actual classes. An entry at row i and column j, $CM(i,j)$, represents the number of misclassified instances that originally belong to class i, although incorrectly identified as a number of class j. The entries of the primary diagonal, $CM(i,i)$, stand for the number of properly detected instances. Cost matrix is similarly defined, as well and entry $C(i,j)$, represents the cost penalty for misclassifying an instance belonging to class i into class j. Cost matrix values employed for the KDD'99 classifier learning contest are shown in Table 4 [19].

4.3. Experiments Setup and Results

In this work, the simulations have been run on a PC powered by a Pentium IV, 3.6 GHz of CPU, and 2 GB of RAM. After implementation of the proposed approach with minimum fuzzy support of 40% and minimum fuzzy confidence of 75%, the total of 3437 rules had been discovered. From these, there were 578 rules with two elements, 1267 rules with three elements, 973 rules with four elements and 619 rules with five elements.

In this study, the value of *a size-adjustment* parameters has been set to 0.8. So, by applying a feature miner algorithm, the linguistic variables that are only appeared in the antecedent of each rule have been extracted and all of them are added to the final feature subset. By using the proposed algorithm, the dimension of input feature space has been reduced, and the most important features are selected for classification. As it was mentioned in the section 4.1, each network connection record in KDD'99 dataset consists of 41 features (Appendix). This algorithm results in the approximate 25% reduction of the features, as the dimension of input feature space is reduced from 41 to 31. The selected features are reported in Table 5.

4.4. Neural Net Structure and Specifications

Before evaluating the system, we have determined the best values of important parameters for neural net. For this purpose, some primary experiments have been carried out and the values of Table 6 are achieved. A noticeable point realized from these primary experiments is the influence of the vigilance

parameters on detection performance of the system. Neither small nor a big amount of this parameter is suitable for this purpose. This is due to the fact that the vigilance parameter determines the similarity degree of patterns that are placed on the same class. Hence, the low value of this parameter causes dissimilar patterns to be placed in the same class and so the neural net is unable to precisely distinguish some of the patterns from each other.

Table 5. Selected features based on *size-adjustment* value

***Size-adjustment* value**	**Selected Features**	**Size of feature subset**
0.8	$F_1,F_3,F_4,F_5,F_6,F_{10},F_{11},F_{13},F_{14},F_{16},F_{17},F_{18},F_{19},F_{22},F_{23},F_{24},$ $F_{25},F_{26},F_{27},F_{28},F_{29},F_{30},F_{31},F_{32},F_{35},F_{36},F_{37},F_{38},F_{39},F_{40},F_{41}$	31

Table 6. Specifications of fuzzy ARTMAP

Number of output layer units	400
Number of epochs	100
Choice parameter (α)	0.01
Learning rate (β_a)	0.5
Learning rate (β_b)	0.5
Vigilance parameter (ρ_a)	0.97
Vigilance parameter (ρ_b)	0.99

The high value for the parameter also causes increasing the sensitivity of the network and reduces its flexibility in placing a new pattern in the previously formed classes (of normal and known attacks). As a result, by finding out a proper value of the vigilance parameter, it is possible to determine the optimum sensitivity level of the system, during the training phase. After determining appropriate structure and parameter values for fuzzy ARTMAP (Table 6), the performance of the proposed system is evaluated in terms of detection rate (DR), false alarm rate (FAR) and cost per example (CPE). The confusion matrix when using fuzzy ARTMAP classifier without applying FSE-FAR module is reported in Table 7.

Table 7. Confusion matrix of fuzzy ARTMAP classifier

Predicted / Actual	Normal	Probe	DoS	U2R	R2L
Normal	6037	8	13	0	1
Probe	28	328	61	0	0
DoS	577	39	22366	3	0
U2R	2	1	2	2	0
R2L	690	4	3	0	938

Table 8. Confusion matrix of FSE-FAR+Fuzzy ARTMAP classifier

Predicted / Actual	Normal	Probe	DoS	U2R	R2L
Normal	6048	8	2	0	1
Probe	28	354	33	0	2
DoS	14	10	22920	0	41
U2R	2	1	2	1	1
R2L	664	0	2	0	969

The confusion matrix for the hybrid structure of FSE-FAR+Fuzzy ARTMAP is shown in Table 8, as well. The duration of the training process for fuzzy ARTMAP with 31 features is approximately 178.83 seconds. Using the same machine, the training takes 224.02 seconds for 41 features. It can be seen that the reduced set of features decreases the computation time more than 20 percent. The performance of the proposed framework in terms of CR, DR, FAR, and CPE along with the performance of some other machine learning methods have been shown in Table 9.

As shown in Table 9, the proposed system has higher or equal classification rate for all the classes, as compared to the systems reported in [23]-[26]. This system performs better in terms of DR, FAR, and CPE, as well. So, it can be inferred that the proposed approach improves the detection rate and decreases the false alarm rate and the cost per example, effectively. It is interesting to note that, as expected, capabilities of this prototype IDS reveals the effectiveness of data mining techniques.

Table 9. Performance of proposed IDS framework as compared to other machine learning models

Model	**Classification rate**					**DR**	**FAR**	**CPE**
	Normal	Probe	DoS	U2R	R2L			
Proposed model with feature selection	99.82	84.93	99.72	17.52	59.28	96.81	0.18	0.0934
Proposed model without feature selection	99.65	78.83	97.31	18.90	57.37	94.37	0.36	0.1341
PNrule [23]	99.5	73.2	96.9	6.6	10.7	91.1	0.4	0.2371
Winner of KDD in 2000 [24]	99.5	83.3	97.1	13.2	8.4	91.8	0.6	0.2331
Runner up of KDD in 2000 [25]	99.4	84.5	97.5	11.8	7.3	91.5	0.6	0.2356
ESC-IDS [26]	98.2	84.1	99.5	14.1	31.5	95.3	1.9	0.1579
MLP with 38 selected features [27]	99.6	75.5	99.7	14.3	32.7	94.9	0.36	0.1517

5. CONCLUSIONS

In this research, an intrusion detection framework based on fuzzy association rules, and fuzzy ARTMAP neural network has been proposed. Fuzzy association rules mining is able to sufficiently handle large amounts of data and it can discover important relationships between large set of data items. In the proposed model, fuzzy grids based rules mining algorithm (FGBRMA) has been used for finding fuzzy association rules to discover the most important features. In this way, the dimension of input feature space has been reduced from 41 to 31. Experimental results have shown that the proposed hybrid model performs better in terms of classification rate, DR, FAR, and CPE as compared to some other machine learning methods.

REFERENCES

[1] S.X. Wu, and W. Banzhaf, **"The use of computational intelligence in intrusion detection systems: A review"**, *Applied Soft Computing,* Volume 10, Issue 1, pp. 1-35, 2010.

[2] C.F. Tsai, Y.F. Hsu, C.Y. Lin, and W.Y. Lin, **"Intrusion detection by machine learning: A review"**, *Expert Systems with Applications*, Vol. 36, Issue 10, pp. 11994-12000, 2009.

[3] S.Y. Wu, and E. Yen, **"Data mining-based intrusion detectors"**, *Expert Systems with Applications*, Vol. 36, Issue 3, Part 1, pp. 5605-5612, 2009.

[4] 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*, Vol. 37, Issue 9, pp. 6225-6232, 2010.

[5] W. Li, J.L. Wang, Z.H. Tian, T.B. Lu, and C. Young, **"Building lightweight intrusion detection system using wrapper-based feature selection mechanisms"**, *Computers & Security*, Vol. 28, Issue 6, pp. 466-475, 2009.

[6] A. Zainal, M.A. Maarof, and S.M. Shamsuddin, **"Feature selection using Rough-DPSO in anomaly intrusion detection"**, *Lecture Notes in Computer Science and its Applications*, Vol. 4705, pp. 512-524, 2007.

[7] S. Chebrolu, A. Abraham, and J.P. Thomas, **"Feature deduction and ensemble design of intrusion detection system"**, *Computers & Security*, Vol. 24, Issue 4, pp. 295-307 , 2005.

[8] S. Mukkamala, and A.H. Sung, **"Feature selection for intrusion detection using NN and SVM"**, *Journal of Transport Research Board National Acada, Transport Research Record* No. 18822, pp. 33-39, 2003.

[9] C.B Vilakazi, and T. Marwala, **"Application of feature selection and fuzzy ARTMAP to intrusion detection"**, *In the Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, pp. 4880-4885, 2006.

[10] A. Tajbakhsh, M. Rahmati, and A. Mirzaei, **"Intrusion detection using fuzzy association rules"**, *Applied Soft Computing*, Vol. 9, Issue 2, pp. 462-469, 2009.

[11] R. Agrawal, T. Imieliński, and A. Swami, **"Mining association rules between set of items in large databases"**, *In Proceedings of the 1993 ACM SIGMOD Conference*, New York, NY, USA, pp. 207-216, 1993.

[12] R. Srikant, and R. Agrawal, **"Mining quantitative association rules in large relational tables"**, *In Proceedings of the ACM SIGMOD International Conference on Management of Data, Montreal*, Canada, pp. 1-12, 1996.

[13] L. Zadeh, **"Fuzzy sets"**, *In Proceeding of Information and Control*, Vol. 8, Issue 3, pp. 338-353, 1965.

[14] F. Berzal, I. Blanco, D. Sánchez, and M.A. Vila Miranda, **"A new framework to assess association rules",** *In Proceedings of the 4th International Conference on Advances in Intelligent Data Analysis, Springer-Verlag***,** Vol. 2189, pp. 95-104, 2001.

[15] M. Delgado, N. Marín, D. Sánchez, and M.-A. Vila, **"Fuzzy association rules: General model and applications"**, *IEEE Transactions on Fuzzy Systems*, Vol. 11, Issue 2, pp. 214-225, 2003.

[16] T.-P. Hong, C.-S. Kuo, and S.-C. Chi, **"Mining association rules from quantitative data"**, *Intelligent Data Analysis*, Vol. 3, Issue 5, pp. 363-376, 1999.

[17] C.M. Kuok, A. Fu, and M.H. Wong, **"Mining fuzzy association rules in databases"**, ACM SIGMOD Record, Vol. 27, Issue 1, pp. 41–46, 1998.

[18] Y.-C. Hu, R.-S. Chen, and G.-H. Tzeng, **"Discovering fuzzy association rules using fuzzy partition methods"**, *Knowledge-Based Systems*, Vol. 16, Issue 3, pp. 137-147, 2003.

[19] 1999 KDD Cup Competition (Available on http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html).

[20] S. Mabu, C. Chen, N. Lu, K. Shimada, and K. Hirasawa, **"An intrusion-detection model based on fuzzy class-association-rule mining using genetic network programming"**, *IEEE Transaction on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, Vol. 41, Issue 1, pp. 130-139, 2011.

[21] G.A. Carpenter, S. Grossberg, N. Markuzon, J.H. Reynolds, and D.B. Rosen, **"Fuzzy ARTMAP: a neural network for incremental supervised learning of analog multidimensional maps"**, *IEEE Transactions on Neural Network*, Vol. 3, Issue 5, pp. 689-713, 1992.

[22] G.A. Carpenter, **"Default ARTMAP"**, *In Proceedings of the International Joint Conference on Neural Networks*, Vol. 2, pp. 1396–1401, 2003.

[23] R. Agrawal, and M.V. Joshi, **"PNrule: A new framework for learning classifier models in data mining (a case-study in network intrusion detection)"**, *IBM Research Division, Technical Report TR 00-015, Report No. RC-21719, Department of Computer Science*, University of Minnesota, 2000.

[24] B. Pfahringer, **"Winning the KDD99 classification cup: bagged boosting"**, *ACM SIGKDD Explorations Newsletter*, Vol. 1, Issue 2, pp. 65-66, 2000.

[25] I. Levin, **"KDD-99 classifier learning contest: LLSoft's results overview"**, *ACM SIGKDD*

Explorations Newsletter, Vol. 1, Issue 2, pp. 67-75, 2000.

[26] A.N. Toosi, and M. Kahani, **"A novel soft computing model using adaptive neuro-fuzzy inference system for intrusion detection"**, *In Proceedings of the IEEE International Conference on Networking, Sensing and Control*, pp. 834-839, 2007.

[27] M. Sheikhan, Z. Jadidi, and M. Beheshti, **"Effects of feature reduction on the performance of attack recognition by static and dynamic neural networks"**, *World Applied Sciences Journal*, Vol. 8, Issue 3, pp. 302-308, 2010.

APPENDIX

Name and type of 41 attributes in KDD dataset

Attribute number	Name	Type
1	duration	continuous
2	protocol_type	discrete
3	service	discrete
4	flag	discrete
5	src_bytes	continuous
6	dst_bytes	continuous
7	land	discrete
8	wrong_fragment	continuous
9	urgent	continuous
10	hot	continuous
11	num_faild_logins	continuous
12	logged_in	discrete
13	num_compromised	continuous
14	root_shell	continuous
15	su_attempted	continuous
16	num_root	continuous
17	num_file_creations	continuous
18	num_shells	continuous
19	num_access_files	continuous
20	num_outbound_cmds	continuous
21	is_host_login	discrete
22	is_guest_login	discrete
23	count	continuous
24	srv_count	continuous
25	serror_rate	continuous
26	srv_serror_rate	continuous
27	rerror_rate	continuous
28	srv_rerror_rate	continuous
29	same_srv_rate	continuous
30	diff_srv_rate	continuous
31	srv_diff_host_rate	continuous
32	dst_host_count	continuous
33	dst_host_srv_count	continuous
34	dst_host_same_srv_rate	continuous
35	dst_host_diff_srv_rate	continuous
36	dst_host_same_src_port_rate	continuous
37	dst_host_srv_diff_host_rate	continuous
38	dst_host_serror_rate	continuous
39	dst_host_srv_serror_rate	continuous
40	dst_host_rerror_rate	continuous
41	dst_host_srv_rerror_rate	continuous