Robust Feature Selection Technique for Intrusion Detection System

Pullagura Indira Priyadarsini[1], Manikonda Srininivasa Sesha Sai[2],
Akula Suneetha[3] and Munnangi Velengini Bala Teresa Santhi[4]

*[1,2,3] Dept. of CSE, KKR & KSR Institute of Technology & Sciences,
Guntur, AP, India*
[4]Dept. of CSE, Koneru Lakshmaiah University, Guntur, AP, India
indupullagura@gmail.com

Abstract

Intrusion Detection System (IDS) is one of the vital steps in network defense mechanism since the intrusive behavior of network data is a bit bewildering. Besides enormousness of the data, noisy or irrelevant features cannot be ruled out using classic data mining techniques. Feature selection is of substantial prominence in pattern recognition which is potent in augmenting learning efficacy, increasing generalization effect, and achieving data visualization. Feature selection process with the Fuzzy system is strong and effective. A novel method is proposed by exploiting the fuzzy system for feature selection. It can be implemented for reducing the computational complexity and improving the classification accuracy in IDSs. In this paper, a Robust Feature selection (RFS) algorithm is given which was an ensemble of three filtering methods: Euclidean distance, chi-square distance, and correlation coefficient. The ensemble of the three filtering methods is done using fuzzy aggregation operator. The tests made on KDD cup 99 data set, ensured good results and generated a greater proportion of recall and precision when compared to other feature selection methods. The average area under the curves (AUCs) will be given as 0.889 which can be a pretty good fit for the proposed algorithm.

Keywords*: Intrusion Detection System (IDS), Fuzzy system, data mining, Euclidean distance, chi-square distance, correlation coefficient, Support Vector Machine (SVM), KDD cup 99 data set, recall, precision, area under the curves (AUCs)*

1. Introduction

The extent of computer and communication technologies made network security a challenging issue for both the researchers and enterprises. The Intrusion Detection System (IDS) is the one which automatically scans a network activity and is an indispensable enhancement of traditional security defense mechanisms such as firewall and data encryption. The challenges in designing and deploying IDSs are increasing due to the wider reach of the Internet services and non-availability of standard procedure for characterizing the intruders [1]. Therefore, IDSs are noticed as a substantial module in the security group organizations [2]. Intrusion detection techniques can be classified into misuse detection and anomaly detection [3]. It is a watchdog in maintaining security from illegitimate network accesses, evading forbidding attacks and so on [4]. Applying fuzzy methods for the development of IDS is more reliable compared to the classical approaches [5]. Fuzzy logic pattern normally, deals with imprecision and vagueness which suits well for the development of security systems. Fuzzy Intrusion Recognition Engine (FIRE) is developed for detecting security attacks particularly against port scans and probes [6]. It uses simple data mining techniques for processing the network input data and produce

fuzzy sets for every observed feature. An Intrusion Detection System requires a sophisticated decision process which comprises a number of factors, implicating fuzziness and vagueness. A feature selection method named Fuzzy Mutual Information-based Feature Selection with Non-Dominated solution (FMIFS-ND) using a fuzzy mutual information measure is given which chooses features based on feature-class fuzzy mutual information and feature-feature fuzzy mutual information applied on UCI data set using four classifiers was developed [7]. An efficient Intrusion Detection System is designed using Fuzzy_Chi_Euc algorithm where fuzzy inference system is exploited for selecting best features in KDD cup 99 dataset and yielded better accuracy [8].

Many researchers have shown that feature selection is adopted as a tool for building effectual data models [9]. This wider use of data preprocessing techniques results in remodeling of known models for allied schemes or absolutely new proposals. The aspect of feature selection is to depute a subset of the features instead of using all the features which make more complex for classification. However, the latest growth of dimensionality of data produces a major difficulty for numerous current feature selection techniques in accordance with effectiveness and performance. Since the filtering methods have low costs and low accuracy rates, merging two methods possibly improve the accuracy [10]. So in this paper, an ensemble of filter methods is done for choosing reliable features and eliminating irrelevant features thus improving accuracy for intrusion detection process.

Applying fuzzy nature to the feature selection approaches such as Euclidean distance, chi-square distance, and the correlation coefficient is done in this work. Then their attainments are considered using fuzzy operations [11]. For doing this, a model is given which constitutes mainly two phases: feature selector and the classification. Experiments made have shown that this approach has achieved reduced training time and good accuracy rate.

This paper is planned as follows: In Section 2, the preliminary perspectives are discussed in detail, followed by a proposed model in Section 3. Next in the preceding Section 4 experiments conducted and evaluation of the results was discussed. In the last section, conclusions and future work were given.

2. Preliminary Perspectives

2.1. Motivation

In some cases, redundant features can lead to noisy data that distract the learning algorithm and degrade the accuracy of the IDS through which, training and testing processes will be slowed down. Significant features are confessed to have a high significance on the performance of the classifiers. Thus handling appropriate feature selection methods renders the models to make them feasible to construe, reducing the training times and augment the generalization [12]. Filtering approach is used as a robust one in building IDS, a set of features is chosen which are treated as most effective correlating to the classification procedure [13].

2.2. Ensemble Technique

The ensemble organization is beneficial in constructing light weight Intrusion Detection Systems (IDSs). If we set some threshold value and select the features, the remaining features which are not selected are neglected or discarded. However, they might be playing a very important role in classification process which is nearer to a threshold. So to consider the remaining best features in discarded ones, the fuzzy system is used to utilize them for feature selection. Individually each filtering method may have some disadvantages; integrating them using fuzzy logic can be done for giving some degree to the features. Accordingly, the top-ranked ones are retrieved. Former tests

presented that integrating feature selection methods could feasibly improve classification accuracy [14].

Euclidean distance is used as a feature selection method in IDS construction and extracted robust features especially for known patterns using C4.5 and SVM in the work made by Anirut et.al [15]. Correlation between feature and class as well as the correlation between feature and feature is done through two algorithms Correlation-based Feature selection (CFS) and fast correlation-based filter (FCBF) using C4.5 and naive Bayes machine learning algorithms and achieved highest averaged accuracies [16]. Feature selection based on Chi-Square and enhanced C4.5 approach is performed on 10% KDD99 dataset and brought the immediate effects for applications [17]. In order to combine the three filtering methods, fuzzy logic is used which increase robustness and adaption ability to the IDS.

3. The Proposed Model

The proposed IDS model presents an entire structure for selecting the best sub set of KDD cup 99 dataset which will proficiently characterize normal traffic and abnormal traffic through building computationally capable IDS. The main intention of this approach is to use fuzziness to integrate filtering methods. Robust feature selection (RFS) algorithm has been proposed which works by combining three filtering methods for fastening the training process. The proposed Framework of the IDS was given in Figure 1. In essence, it has four steps 1) Pre-processing 2) Applying Filter approach for three methods 3) Ensemble Feature Selection 4) Classification through SVM. While the proposed method is generic and can be pertained to any dataset, for testifying the proposed method KDD Cup 99 dataset has been used.
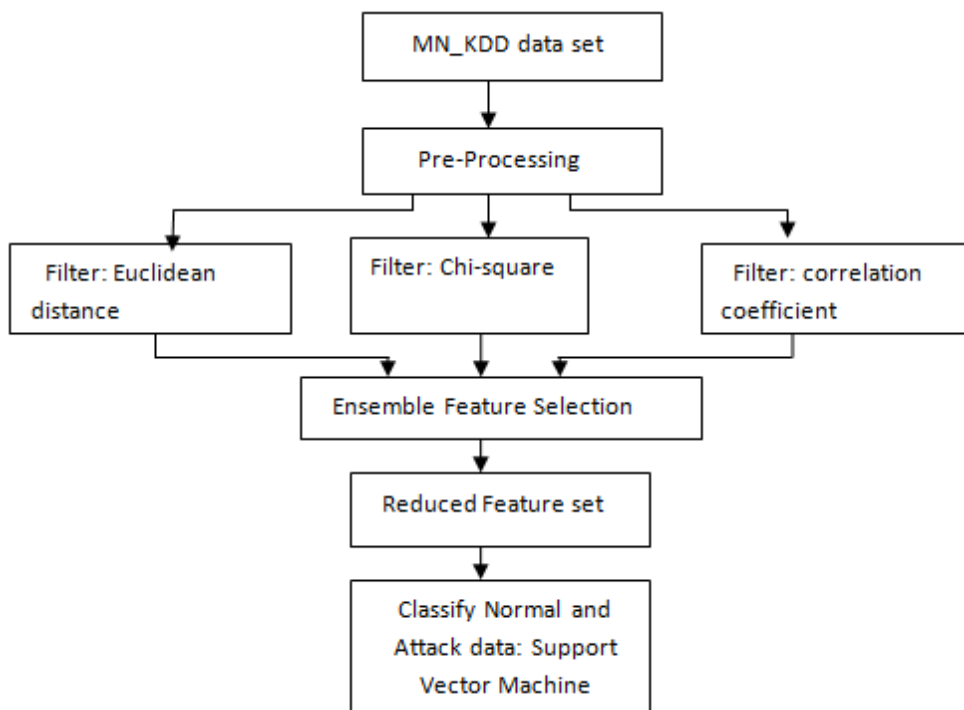

Figure 1. Proposed Structure

3.1. Fuzzy Feature Selection

The key idea behind using fuzzy nature in feature selection lies in permitting the second best features which may be important for the classification process. For the data set chosen, Euclidean distance [18], correlation coefficient [19] and chi-square distance [20] computations are made. They are further transformed into three fuzzy sets. The α-cut operation of fuzzy aggregation operations is applied for all the three fuzzy sets. From the corresponding features of three α-cut values obtained, the features are selected. For a given pattern {M, N} where M= no. of instances and N= no. of attributes or features in the data set which is 41. F= {F1, F2, ---F41} denotes the feature set of pattern input M.

RFS Algorithm
> **Input:** MN_KDD data set,
>   N= {F1, F2, -------F41},
>   Class, a,b,c,D[],E[],F[],D_α , E_α , F_α ,s1,s2,s3,x, Ω , Ψ, X
> **Output:** {Feature Selector, classification measures: accuracy, precision, recall, ROC}

> **Start:**
> 1. Use Proper pre-processing techniques to the MN_KDD data set
> 2. for each feature in the N
> 3.   Do
> Calculate a =Euclidean_distance (Feature, class);
> Calculate b =Chi_square (Feature, class);
>   Calculate c=Correlation (Feature, class);
> 4.   End
> 5.   Convert 'a' to fuzzy values
> 6.   Set as D [41];
> 7.   Convert 'b' to fuzzy values
> 8.   Set as E [41];
> 9.   Convert 'c' to fuzzy values
> 10. Set as F [41];
> 11. s1= D_α ;
> 12. s2= E_α ;
> 13.   s3= F_α ;
> 14. Ω =All the corresponding features of s1= {Feature | D(x) = Feature for   some x $\in$ s1};
> 15. Ψ = All the corresponding features of s2= {Feature | E(x) =   Feature for some x $\in$ s2};
> 16. X= All the corresponding features of s3= {Feature | F(x) = Feature for some x $\in$ s3};
> 17.   Feature Selector= Ω U Ψ U X;
> 18.   Pass the reduced Feature Set into classification algorithm (SVM);
> **End**

Figure 2. The Proposed RFS Algorithm

For any single feature F1 $\in$ F the Euclidean distance is calculated as

$$d_j (F_j, C) = \sqrt{\sum_{i=1}^{n} (x_{1j} - c_i)^2} \qquad \text{-------- (1)}$$

Where F_j is individual feature and j=1 to 41 and C is a class label, n=number of instances. This is computed for all the 41 features. c_i =1 to 5.By applying the equation 1 above, we get some score to each feature. It is converted to fuzzy value by standardizing

them. Then $d_j \epsilon$ {real values between 0 and 1 including both}. In the same manner, chi-square distance and correlation coefficient are calculated for each feature and class label.

The data set taken is called MN_KDD data set with 41 attributes and five classes. The Robust Feature Selection algorithm (RFS) is given in the Figure 2 above. The input is MN_ KDD data set with features F1, F2, ---- F41 and a class. Then necessary pre-processing techniques are applied to the data set. With RFS algorithm, choosing the finest feature subset of the 41 features is done. As mentioned in the algorithm above, we perform three filtering methods to each feature in the data set. Then for each feature and class label calculating 1) Euclidean distance 2) chi-square distance 3) correlation coefficient is done. This was indicated in line 3 of the RFS algorithm above. Then by converting them to fuzzy values, by therefore we get three fuzzy sets as shown in lines 6, 8, and 10 of the Figure 2 above. It means that we get membership functions μ_D, μ_E, μ_F in fuzzy sets D[], E[], F[] of the given algorithm. D_α, E_α, F_α mentioned in the algorithm means that Aggregation operator α-cut is calculated for all the three fuzzy sets. The α-cut returns all the values whose $\mu_D(x) \geq \alpha$. Here we have taken $\alpha=0.5$. The definition of α-cut is illustrated in table 1. Therefore, Ω is the one which gives corresponding features of the α-cut values obtained in D_α as given in line 14. In the same way we got Ψ, X as mentioned in line 15 and line 16 of the algorithm above. Then union of all the features obtained is placed in *Feature selector* as given in line 17 of the above mentioned algorithm. Now the features selected are passed to the classification using SVM which returns attack category or normal category.

Table 1. Fuzzy Aggregation Operations

S.no	Label of the operation	Explanation
1.	**Support**	It returns all the values whose $\mu_A(x) > 0$.
2.	**Crossover point**	It returns all the values whose $\mu_A(x) = 0.5$.
3.	**Height**	It returns the largest membership value attained by any point.
4.	**α-cut**	It returns all the values whose $\mu_A(x) \geq \alpha$ where α can be between 0 to 1.
5.	**Strong α-cut**	It returns all the values whose $\mu_A(x)$ is strictly greater than α where α can be between 0 to 1.

3.2. Classification through Support Vector Machine

Support Vector Machines (SVMs) are excellent for classification in several areas.SVM is a classification process which divides the data into two classes by constructing Maximal marginal hyper plane (MMH) through identifying support vectors. The data point can belong to the positive class and negative class in case of binary classification. They are put forward by Vapnik [21]. SVMs exploit Statistical Learning Theory (SLT). They are specific on training samples and have better generalization ability on testing samples. SVMs are built with linear and nonlinear decision boundaries which can be solved with optimization problem. One of the benefits of SVM is that it exhibits good accuracy in high dimensional space.

4. Experiments and Results

The KDD cup 99 data set [22] is used for testing the proposed algorithm. It is the benchmark data set widely used for IDS evaluation. This is the data set taken from only 10 percent of the original data set. From this data set a portion of it is taken containing 14207 instances and it is named as "MN_KDD" data set with a relative size of records as in KDD cup99 data set. It contains 3000 Normal instances, 10000 DoS instances, 574 probe instances, 401 R2L and 52 U2R instances. MN_KDD data set contain discrete and continuous features. Then Feature rescaling is done for every feature individually. The five classes are mapped with 1, 2,3,4,5 for U2R, R2L, Probe, DoS, Normal respectively. Features are taken as $\{F_1, F_2....F_{41}\}$. Here totally three experiments were conducted on MN_KDD data set 1) The proposed RFS algorithm 2) With filter method Euclidean distance for feature selection and SVM 3) With no Feature Selection and SVM.

In this robust feature selection approach, three filtering methods namely Euclidean distance, chi-square, correlation are applied to the MN_KDD dataset. Then obtained values are sorted and converted to fuzzy values. This can be done by 'x' which is an actual value, then the fuzzy value becomes $f(x) = \dfrac{x}{max - min}$ where max is the maximum of all the 14207 values obtained and min is the minimum of all 14207 values obtained. Therefore for the Euclidean distance, chi-square distance and correlation coefficient, their fuzzy sets are taken each containing 41 values. They are obtained as D [41], E [41], and F [41]. Then by considering their membership values, we get D_α , E_α , F_α of fuzzy sets D,E,F respectively. Now considering all the corresponding features of D_α , E_α , F_α values we get Ω, Ψ, X. As a result, Ω is equal to F23, F24, F29, F32, and F33. Ψ is equal to F2, F24, F29, F30, F32, F33, F36, and F38. Then X is equal to F23, F24, and F36. Then union of Ω, Ψ, X will be F2,F23,F24,F29,F30,F32,F33,F36,F38 which is the final feature set.

Euclidean distance is calculated for feature and class label values. This is done for all the 41 features. For instance, the Euclidean distance computation and fuzzy values obtained for all the 41 features in the MN_KDD data set is shown table II below. Here considering the α-cut of the fuzzy set D(from Euclidean distance obtained) as mentioned above we take F23,F24,F29,F32,F33 since their μ values are greater than or equal to 0.5.

Table 2. Euclidean Distance for Each Feature and its Transformed Fuzzy Values

Feature Name	Feature	Euclidean distance	μ values
Duration	F1	489.261	0.017
Protocol_type	F2	272.6004	0.009
Service	F3	2221.573	0.08
flag	F4	701.0228	0.02
Src_bytes	F5	481.1676	0.02
Dst_bytes	F6	1299.587	0.04
Dos	F7	491.754	0.02
Wrong fragment	F8	491.7082	0.02
Urgent	F9	491.7469	0.02
hot	F10	529.103	0.02
Num_failed_logins	F11	491.754	0.02
Logged_in	F12	405.4257	0.01
Inum_compromised	F13	488.4475	0.02

Iroot_shell	F14	491.6899	0.02
Isu_attempted	F15	491.754	0.02
Inum_root	F16	491.4591	0.02
Inum_file_creations	F17	492.2124	0.02
Inum_shells	F18	491.7449	0.02
Inum_access_files	F19	491.2962	0.02
Inum_outbound_cmds	F20	491.754	0.02
Is_host_login	F21	491.754	0.02
Is_guest_login	F22	490.7362	0.02
count	**F23**	**20414.29**	**0.7**
Srv_count	**F24**	**20446.24**	**0.7**
Serror_rate	F25	491.6011	0.02
Srv_serror_rate	F26	491.5809	0.02
Rerror_rate	F27	491.5221	0.02
Srv_rerror_rate	F28	491.1651	0.02
Same_srv_rate	**F29**	**29178.8**	**1.0**
Diff_srv_rate	F30	490.9823	0.017
Srv_diff_host_rate	F31	482.4524	0.017
Dst_host_count	**F32**	**21835**	**0.76**
Dst_host_srv_count	**F33**	**28030**	**0.97**
Dst_host_same_srv_rate	F34	383.8136	0.01
Dst_host_diff_srv_rate	F35	488.4898	0.02
Dst_host_same_src_port_rate	F36	467.7594	0.02
Dst_host_srv_diff_host_rate	F37	489.7643	0.02
Dst_host_serror_rate	F38	491.5383	0.02
Dst_host_srv_serror_rate	F39	491.6404	0.02
Dst_host_rerror_rate	F40	491.2453	0.02
Dst_host_srv_rerror_rate	F41	491.0838	0.02

From the proposed RFS algorithm given in section 3.1, the features selected are nine. With this reduced number of features, the rapidity of the learning process is increased. At this point, one vs. rest methodology has been employed for classification and thus five classifications are made. Training is conducted using Radial Basis Function (RBF) kernel and 10 fold cross-validation. Generally, intrusion detection is a true or false investigative problem, and the training of two-class classifiers appears much proficient and precise. At the implementation phase, classification part of SVM is a facile weighted sum and aimed at one vs. rest method *n* SVMs are required to be built if there are *n* classes. John C. Platt's sequential minimal optimization algorithm (SMO) for training a support vector classifier using RBF kernels has been used [23]. The precision obtained for the proposed algorithm is 80.2. The Results obtained is illustrated with graphs. The overall Accuracy rate and the average time taken to build all the models are depicted in Figures 3 & 4. The Recall for proposed model is 78.96%. Table 3 shows the comparison of average Precision, average recall and overall accuracy rate of all the three experiments done.

Table 3. Comparison of RFS algorithm with Other Experiments using Euclidean + SVM, with no FS

S.no	Evaluation metrics/Algorithm	RFS Algorithm	Euc+SVM	No FS
1.	Recall	78.96%	72.4%	56.8%
2.	Precision	80.2%	76.4%	52.3%
3.	Accuracy Rate	96.25%	88.45%	80.2%

As from the above analysis, RFS algorithm has the worth of high classification precision, rapid training velocity, and good generalization capability. The feature selection algorithm can choose subset with lesser features; however, the subset acquires extremely high average classification precision. Therefore, the RFS algorithm is firm and powerful.
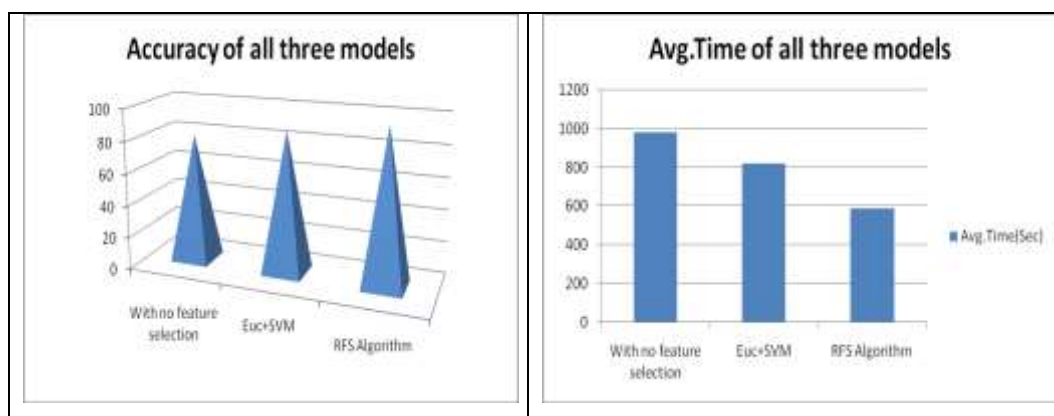
Figure 3. The Overall Accuracy Rate for Three Models

Figure 4. The Average Time is taken to build the Three Models

The ROC curves for four classifications namely DoS-Normal, Probe-Normal, R2L-Normal and U2R-Normal for the proposed approach are shown separately in the below figures. The average area under the curves (AUCs) will be given as 0.889 which can be a pretty good fit for the proposed algorithm.
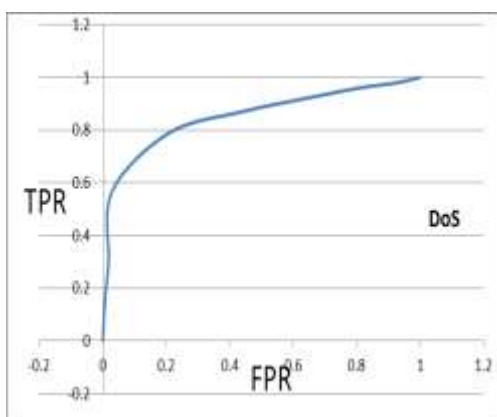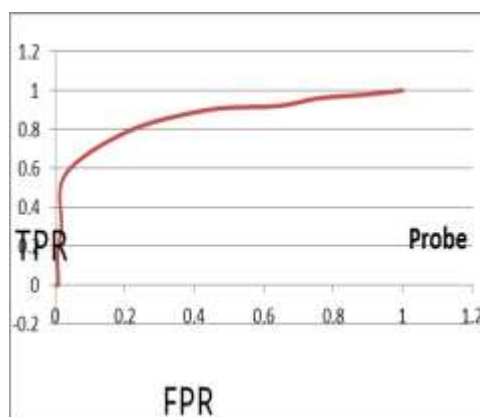

Figure 5. ROC Curve for DoS

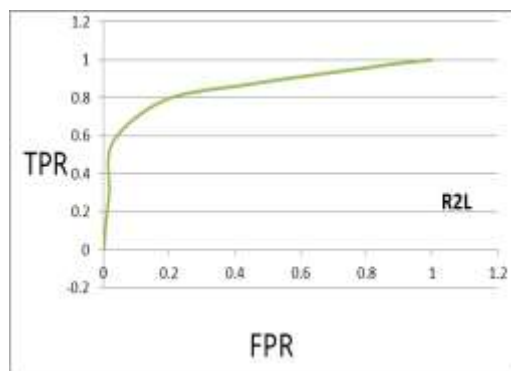**Figure 6. ROC Curve for Probe**
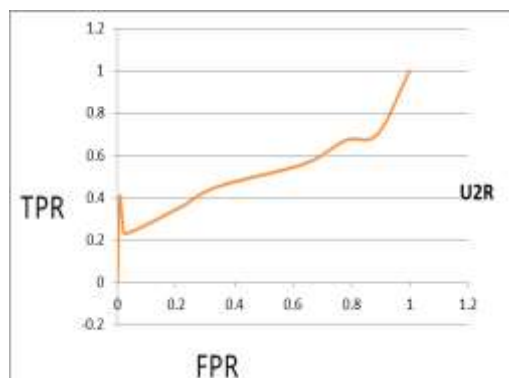
Figure 7. ROC Curve for R2L   Figure 8. ROC Curve for U2R

5. Conclusions and Future Work

The ensemble of filtering methods such as Euclidean distance, chi-square distance, and the correlation coefficient is made through fuzzy aggregation operation α–cut and has yielded good results. Well efficient Intrusion Detection Systems are built by retrieving most important features from the proposed RFS Algorithm. With this algorithm, the top nine features are selected. Since SVMs are good with generalization ability and robustness, these are employed for classification of attack and normal data. From the experiments done, the proposed algorithm has shown considerable recall and precision. The average area under the curves (AUCs) is 0.889 which can be a pretty good fit for the proposed algorithm. Detection rate has also increased by this approach. Since the 41 feature set contains $2^{41}-1$ possible subsets in which one can be a very optimistic subset. As the future work, exploring the optimistic subset of building lightweight Intrusion Detection System models can be done.

References

[1] H.J. Sehumacher and S. Ghosh, "A fundamental framework for network security", Journal of Network and Computer Applications, vol. 20, (1997), pp. 305-322.
[2] H. Debar, M. Dacier and A. Wespi, "Towards a Taxonomy of Intrusion Detection Systems", IBM Technical Paper, Computer Networks, vol.39, issue 9, (1999), pp. 805-822.
[3] M.A.M. Hasan, M. Nasser and B. Pal, "On the KDD'99 Dataset: Support Vector Machine Based Intrusion Detection System (IDS) with Different Kernels", IJECCE, vol. 4, (2013), pp. 1164-1170.
[4] R. H. Gong, M. Zulkernine and P. Abolmaesumi, "A Software Implementation of a Genetic Algorithm Based Approach to Network Intrusion Detection", Proceedings of the Sixth International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing and First ACIS International Workshop on Self-Assembling Wireless Networks, (SNPD/SAWN'05). DOI: 10.1109/SNPD-SAWN.2005, (2005).
[5] M.G. Ouyang, W.N. Wang and Y.T. Zhang, "A fuzzy comprehensive evaluation based distributed intrusion detection", Proceedings First International Conference on Machine Learning and cybernetics,), China, Beijing, (2002), pp. 281-285.
[6] I.E. Dickenon, J. Juslin, 0. Koukousoula and J.A. Dickenon, "Fuzzy intrusion detection", Proceedings IFSA World Congress and ZGih NAFIPS lnternational Conference, vol. 3, (2001), pp. 1506-1510.
[7] N. Hoque, H.A. Ahmed, D.K. Bhattacharyya and J.K. Kalita, "A Fuzzy Mutual Information-based Feature Selection Method for Classification", Fuzzy Information and Engineering, https://doi.org/10.1016/j.fiae.2016.09.004,vol.8, iss.3, (2016), pp. 355-384.
[8] P. Indira priyadarsini, Ch.Anuradha and P.S.R. Murty, "Fuzzy Based Feature Selection for Intrusion Detection System", in International Journal of Engineering Research in Computer Science and Engineering (IJERCSE),vol. 4, issue 11, (2017), pp.42-49.
[9] X. Jin, A. Xu, R. Bie1 and P. Guo, "Machine Learning Techniques and Chi-Square Feature Selection", Springer-Verlag Berlin Heidelberg LNBI 3916, (2006), pp. 106 – 115.
[10] M. A. Ambusaidi, X. He, P. Nanda and Z. Tan, "Building an Intrusion Detection System Using a Filter Based Feature Selection Algorithm", IEEE Transactions on Computers, DOI: 10.1109/TC.2016.2519914, vol. 65, no. I (10), (2016).
[11] T. J. Ross, "Fuzzy Logic with Engineering Applications (3rd Edition)", John Wiley & Sons, New Jersey, USA.

[12] X. Jin, A. Xu, R. Bie1 and P. Guo, "Machine Learning Techniques and Chi-Square Feature Selection", Springer- Verlag Berlin Heidelberg LNBI 3916, **(2006)**, pp. 106 – 115.

[13] S. Khalid, T. Khalil and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning", Science and Information Conference (SAI), **(2014)**, pp. 372–378.

[14] Z. Karimi and A. Harounabadi, "Feature Ranking in Intrusion Detection Dataset using Combination of Filtering Methods", International Journal of Computer Applications (0975 – 8887), vol. 78, Iss (4), **(2013)**, pp. 21–27.

[15] A. Suebsing and N. Hiransakolwong, "Euclidean-based Feature Selection for Network Intrusion Detection", International Conference on Machine Learning and Computing IPCSIT vol.3 **(2011)** ,IACSIT Press, Singapore, pp.222-229.

[16] T.-S. Chou, K. K. Yen and J. Luo, "Correlation-based feature selection for intrusion detection design", IEEE, DOI: 10.1109/ MIL COM. 2007. 4454806, **(2007)**.

[17] D. Hong and L. Haibo, "A Lightweight Network Intrusion Detection Model Based on Feature Selection" ,IEEE, DOI 10.1109/PRDC.2009.34, **(2009)**.

[18] http://www.econ.upf.edu/~michael/stanford/maeb4.pdf.

[19] L. Yu, and H. Liu, "Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution", In Proceeding of the Twentieth International Conference on Machine Learning (ICML), **(2003)**, pp. 856-863.

[20] http://web.pdx.edu/~newsomj/da1/ho_chisq.pdf.

[21] C. Cortes and V. Vapnik, Support Vector Networks, Machine Learning, vol. 20, iss. 3, **(1995)**, pp. 273-297.

[22] M. Tavallaee, E. Bagheri, W.Lu and A.A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set", Proceedings of the (2009) IEEE Symposium on Computational Intelligence in Security and Defense Applications.

[23] J. Platt, B. Schölkopf, C. Burges and A. Smola, "Fast Training of Support Vector Machines using Sequential Minimal Optimization", Advances in Kernel Methods - Support Vector Learning, MIT Press, **(1998)**.

Authors

Pullagura Indira Priyadarsini, she is an Associate Professor in Dept. of Computer Science & Engineering, KKR & KSR Institute of Technology & Sciences, Guntur, Andhra Pradesh, India. Earlier she worked as Head of the Department of Computer Science & Engineering in Chalapathi Institute of Technology, Mothadaka, AP, India. She has conducted various conferences and workshops and acted as a convener. She has guided several projects at M.Tech level. She has attended several Faculty Development Programmes at NITs and IITs in India. She published and presented articles in several reputed journals and conferences. She completed her Doctor of Philosophy from Acharya Nagarjuna University, Guntur in the area of machine learning in the year 2015.Her research interests are machine learning, data mining and network security. She completed her M.Tech (CSE) from Koneru Lakshmaiah college of Engineering, Guntur. She stood first in the Post Graduate level. She completed her B.Tech from RVR & JC College of Engineering, Guntur. She is hard working and sincere in her deeds.

Manikonda Srinivasa Sesha Sai, he received B.Sc. degree from Acharya Nagarjuna University, India in 1994, Master of Computer Applications from Acharya Nagarjuna University, India in 1997, Master of Philosophy in Computer Science from Alagappa University, India in 2004, Master of Technology in Computer Science & Engineering from Dr. M.G.R. University, India in 2007 and Ph.D. in Computer Science & Engineering from Acharya Nagarjuna University, India in 2011. He is currently working as Professor in the Department of Computer Science & Engineering, in KKR & KSR Institute of Technology & Sciences, Guntur, India.

He is a DEL-EMC certified Data Associate in 'R' Language. His area of interests are Data Science, Data Analytics and Internet of Things & its applications.

Akula Suneetha, she received her M.Tech(Computer Science & Engineering)Degree from Jawaharlal Nehru Technological University, Hyderabad, Andhra Pradesh, India in the year 2009 and B.Tech (Computer Science and Engineering) from Sir C R Reddy College of Engineering; Andhra University in the year 2005.She is currently pursuing the Ph.D Degree.she completed pre-ph.d, She is currently working as Assistant professor in the Department of Computer Science & Engineering, KKR &KSR Institute of science & technology,Guntur, Andhra Pradesh affiliated to jntukUniversity. Her research area is Fuzzy Image processing. She is having 12 years of Teaching Experience.

Munnangi Velengini Bala Teresa Santhi, she is working as Associate Professor in the Dept. of Computer Science & Engineering in Koneru Lakshmaiah University since 2006. She has completed her M.Tech from Acharya Nagarjuna University, Guntur. She received B-Tech degree in Computer Science and Engineering from Jawaharlal Nehru Technological University, Hyderabad, in 2003, M-Tech degree in Computer Science from Acharya Nagarjuna University, in 2010. Her areas of interest are Data Warehousing, Design and Analysis of Algorithms, DBMS, Big Data and Business Intelligence.