Journal of Engineering and Applied Sciences 12 (16): 4180-4185, 2017
ISSN: 1816-949X

Enhance of Extreme Learning Machine-Genetic Algorithm Hybrid Based on Intrusion Detection System

[1]Mohammed Hasan Ali , [1]Mohamad Fadli Zolkipli, [2, 3]Mohammed Abdulameer Mohammed and [2]Mustafa Musa Jaber
[1]Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang, Pahang, Malaysia
[2]Nabu Research Academy, Selangor, Malaysia
[3]Universiti Utara Malaysia, Kedah, Changlun, Malaysia

Abstract: This study presents a new scheme of the hybrid Extreme Learning Machine-Genetic Algorithm (ELM-GA). ELM has been proved to be exceptionally fast and achieves more generalized performance for learning Single hidden Layer Feedforward Neural networks (SLFN). However, due to the random determination of parameters for hidden nodes and the number of hidden neurons, some un-optimal parameters may be generated to influence the generalization performance and stability. Some of the papers used GA as a hybrid to solve this problem in ELM but ELM-GA still has some limitations where they used the GA to find the optimal weights for the ELM. In this research, we try to let the GA not only find the best weights but find the best classifier (weights and structure). Intrusion Detection System (IDS) facing big challenge in high rate of false alarms. This research proposes a new method in validation of the classifiers to be sure that the classifiers training enough to mitigate the false alarm's rates.

Key words: ELM, SLFN, ANN, IDS, GA, Malaysia

INTRODUCTION

Intrusion-Detection Systems (IDS) is a security tool that, like other measures such as firewalls, access control schemes and antivirus are intended to strengthen the security of information and communication systems as shown by Peyman and Ghorbani (2005). These traditional techniques have failed to be full protect networks and systems from increasingly sophisticated attacks and malwares. As a result, IDS have become an indispensable component of security infrastructure used to detect these threats before they inflict widespread damage. In fact, the process of automatically constructing models from data is not trivial, especially for detection problems. This is because intrusion detection faces such problems as huge network traffic volumes, highly imbalanced attack class distribution, the difficulty to realize decision boundaries between normal and abnormal behavior and requiring continuous adaptation to a constantly changing environment. Artificial intelligence and machine learning have shown limitations to achieving high detection accuracy and fast processing times when confronted with these requirements (Wu and Banzhaf, 2010). Extreme Learning Machine (ELM) was proposed by Huang *et al.* (2004) as novel efficient learning methodology to train Single-hidden Layer Feed forward Neural networks (SLFNs) faster than any other learning approaches. ELM is a tuning free algorithm with b an extreme fast learning speed by randomly generating the inputs (weights, hidden bias) and also random select the numbers of hidden neurons N.

However, due to the random select of learning parameters, some un-optimal input weights and biases may have generated which is negative impact on the performance stability and generalization ability. To alleviate such weakness, some modifications on ELM have been proposed and another researchs proposed optimized the ELM by hybrid it. Zhu *et al.* (2005) proposed Evolutionary Extreme Learning Machine (E-ELM) by adopting a differential evolution algorithm to select optimal input weights and biases. Similarly, Feng *et al.* (2012) proposed an algorithm called ES-ELM in which they use the crossover mechanism derived from the Genetic algorithm to select the optimal hidden nodes for ELM. Although, the parameters could be optimized using the evolutionary algorithms, a single neural network is still not robust enough. Moreover as ELM minimizes the training error with the whole training dataset, it may suffer from overtraining which might also

Corresponding Author: Mohammed Hasan Ali, Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang, Pahang, Malaysia

degrade the generalization performance. Many research tried to alleviate this weakness in the ELM; the hybrid with GA gave better results for many applications that used as ELM-GA. Despite the fact that the works based on ELM-GA gives good results when apply it with many different applications this research try to analysis these works and proposes a new model to enhance the ELM-GA based on IDS.

MATERIALS AND METHODS

Preliminaries:

Extreme learning machine: ELM reduces the learning time of SLFNs by finding the weights using a simple inverse generalization calculation. Huang proved that the basic structure in Fig. 1 is capable of forming classification decisions for any number of disjoint regions as long as the number of hidden neurons was allowed to grow as large as required (Huang*et al*., 2000). The SLFN depicted in equation has the following output function:

$$(x)^{\rightarrow} = t_j = \sum G\,(Li = lx^{\rightarrow}, a_i, b_i)\beta_i \quad (1)$$

Where:

$x^{\rightarrow}$ = Represents the set of input data

L = The number of nodes in the hidden layer

The a_i, b_i values represent the weights or strength of connections between the ith hidden node and each of the inputs while the b_i sets the threshold of the ith hidden node (Haykin, 2000). G is the activation function which for additive nodes is expressed as (a. x_j+b_i) where x_j are the inner product.

For classification problems such a system would map N examples in the input data set $x^{\rightarrow}$ through transformation G into classes t_j, substituting h(x) for G Eq. 1 can be rewritten as $fL(x)^{\rightarrow} = t_j = h(x)\,\beta_i$. In matrix notation this becomes:

$$H\beta = T \quad (2)$$

where, H is considered to be the output matrix for the hidden layer between x and T as explained (Ding *et al*., 2015). In this case:

$$H = \begin{bmatrix} g(a_1 \cdot x_1 + b_1) & \cdots & g(a_L \cdot x_1 + b_L) \\ \vdots & \vdots & \vdots \\ g(a_1 \cdot x_N + b_1) & \cdots & g(a_L \cdot x_N + b_L) \end{bmatrix}_{N\times L}, \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L\times m}$$

$$T = \begin{bmatrix} \beta_{1m}^T \\ \vdots \\ \beta_{Nm}^T \end{bmatrix}_{N\times m} \quad (3)$$

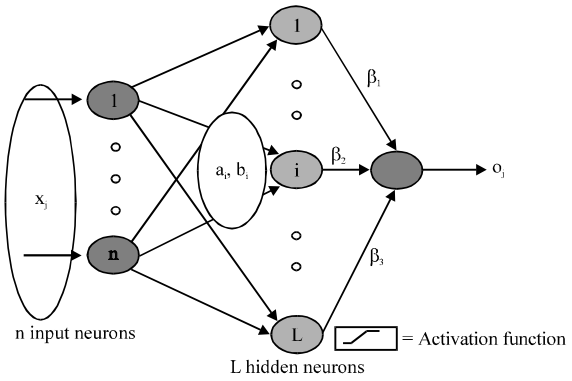

Fig. 1: Basic structure of single hidden layer feedforward neural network adapted from (Huang *et al*., 2012)

Given (Eq. 2 and 3) above, the learning objective is to obtain an optimal solution β such that the training error is minimized and the system can accurately classify new examples of data with similar properties to the training dataset. That is the ability to generalize when testing the system with new data.

For the basic ELM, Huang discovered that instead of using an algorithm to iteratively tune (i.e., train) the SLFN using a supervised learning technique such as Back Propagation (Haykin, 2000) in order to find a_i, b_i values that meet some maximum error target by just simply randomizing the a_i, b_i values and solving for b in Eq. 2 using Moore-Penrose inverse yields a result that meets the learning objective in many cases (Huang *et al*., 2004). In mathematical form this is stated as follows:

$$\beta = H^{*}T \quad (4)$$

where, * is the Moore-Penrose generalized inverse of H (Huang *et al*., 2004).

Genetic algorithm: Optimization is at the heart of many real-world problem solving processes. However, finding the optimal solution for such problems is often tedious, especially in the presence of non-linearity, high dimensionality and multi-modality. Over the last few decades, Evolutionary Algorithms (EAs) have shown tremendous success in solving complex optimization problems. The Genetic Algorithm (GA) is the most popular and widely used in practice (Elsayed*et al*., 2014). Figure 2 shown the general genetic algorithm. Basically, populations of strings (called chromosomes) which encode candidate solutions to an optimization problem evolve toward better solution and to yield a good result in many practical problems is composed of three operators reproduction, crossover and mutation.

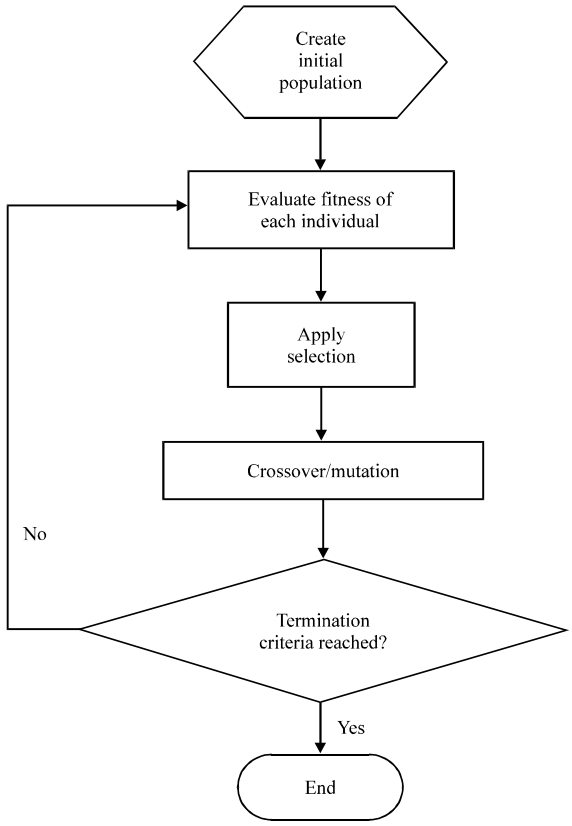

Fig. 2: The schematic flowchart of general genetic algorithm (Pereira *et al*., 2013)

RESUTS AND DISCUSSION

Extreme Learning Machine-Genetic Algorithm (ELM-GA): Since, ELM just randomly chooses the input weights and hidden biases (Zhu *et al*., 2005) much of learning time traditionally spent in tuning these parameters is saved. However as the output weights are computed based on the prefixed input weights and hidden biases, there may exist a set of non-optimal or unnecessary input weights and hidden biases.

GA is applied in many works to find the optimal weights between input and hidden layers and bias values in many works. Initially, a population is generated with a set of random numbers and these values are fed to ELM and find the classification accuracy as fitness for each chromosome. Alexandre *et al*. (2015) to reducing traffic noise and improving intelligent transportation systems, they proposed ELM-Ga. ELM played the key role of providing fitness of candidate solution in each generation for the GA by special part called the Probability of correct classification (PCC) is used as a measure of classifier works. ELM accuracy plays the key role in fitness and evaluated as the aim of maximizing of PCC. They used social data set to evaluate the model based three kinds of sound (cars, motor bikes, trucks). They did the training and validation 50 times that is the mean at the end will fit 50 ELM classifiers so the final value of the PCC is the mean value and standard deviation of the 50 probabilities (Cao *et al*., 2015; Xue *et al*., 2014). All of these works they proposed ELM-GA in different fields but with same strategy on hybrid ELM and GA. ELM plays the key to find the fitness function for the GA based on the Root Mean Square Error (RMSE) and the parameter's connections between them are only (w_i, b_i). Xue *et al*. (2014) proposed the same hybrid but to solve the random select in GA during a crossover and mutation. They divided the ELM into two steps first one called downward-climbing to find the relationship between the parent and child population during crossover and second one named upward-climbing to find the relationship between the parent and child population during mutation. Also, used special parameters based on the data set, data take for the water inflow from two hydro plants.

The number of the hidden neurons in ELM is a necessary parameter. Zhu *et al*. (2005) mention that ELM may require more hidden neurons than the conventional tuning-based learning algorithm in some application which may make ELM response slowly to unknown testing data. This one may expect the more compact network in the applications which requires a faster response of the training network. Furthermore, another work (Alexandre *et al*., 2015) it's noted that the number of hidden nodes is a free parameter of ELM training and must be estimated for obtaining good results. And previous works did not propose something for problem of the random select the number of hidden nodes of neurons. We propose to let Ga find not only best (w_i, b_i) but to find also the optimal number of hidden nodes neural. So, in end GA will provide best ELM classifier. Table 1 shown related work based on the hybrid between ELM and GA in other fields.

The problem with the random select is not just about the hidden number of nodes neurons but also about the activation functions of the ELM classifier. It is used to activation level as a unit (neuron) into an output. In (Karlik and Olgac, 2011) the aim of this study is to analyze the performance of generalized Multilayer Perception (MLP) use different activation functions (Radial Basis Function (RBF), sigmoid, tanh) and in the end, the results showed that tanh function performs better accuracy than other functions. Also, Lin and Lin (2003) explained the sigmoid kernel is not better than RBF kernel in general (Ozkan and Erbek, 2003). This research comparison some

activation functions for multispectral Landsat TM image classification with one and two of hidden layered network. The results shown for one-hidden layer network the tangent hyperbolic activation function was better than sigmoid function for classification problem. And for two-hidden layer network structure, the sigmoid gave the highest test data overall accuracy. I proposed a primitive but without giving details and analysis for the current solutions that provide same hybrid in other fields (Ali and Zolkipli, 2016). For all Works based ELM-GA hybrid didn't mention about this difference to the activation function's performance and they had also chosen randomly the functions. We propose to check the new hybrid of ELM-GA with several of activations functions to find the best structure of this hybrid.

Intrusion-detection systems based on ELM-GA: Many works used artificial intelligence and machine learning to design the IDS because of the problems, we mentioned before the benefits of these algorithms. Xiang *et al.* (2014) proposed ELM to work based on IDS because they considered traditional supervised learning methods are too slow to work as IDS and ELM method is still fairly novel in the Intrusion IDS field in the sense that it is still not mentioned in recent surveys of the methods employed (Bhuyan *et al.*, 2014).

They showed that ELM based intrusion detection can extend its applicability to significantly larger datasets than datasets currently used in most papers and its possible without increasing training time drastically. In the end this research not propose any solved of the random select problem of ELM same like Cheng *et al.* (2012) work they used basic kernel ELM to work as IDS but without any solve propose for the random select problem. In this research, we propose to mitigate the disadvantage of ELM by hybrid it with GA in a new model based on IDS. Table 2 shown some related works based on IDS.

Table 1: Some related works based hybrid ELMGA

Researchers and years	Activation function	Encoding chromosome	Fitness function	Data set
Xue *et al.* (2014)	Sigmoid	W_i, B_i	RMSE	Abalone, Boston housing, California housing
Alexandre *et al.* (2015)	(RBF)	wave describing parameters	RMSE with only 60 generations	Wave data set
Alexandre *et al.* (2015)	(RBF)	Special for GA and normal ELM	Probability of Correct Classification (PCC)	Sound features
Matias	Linear	W_i, B_i	Using special equation based on RMSE	Experimental results in 16 Benchmark data sets based classification
Yang	Used linear nonlinear	Special parameters based the work	Find fitness based ELMs. They divided to two ELM one with a crossover and one with mutation	Data take for the water inflow from two hydro plants
Raajamanickam and Mahalakshmi	RBF	W_i, B_i	Particle Swarm Optimization (PSO)	Used multi datasets based water marking
Cao *et al.* (2015)	Sigmoid	W_i, B_i	RMSE	Used apical data set based Finite Element Model (FEM) simulation
Qingfeng	RBF	W_i, B_i	RMSE	Test data from UCI machine learning repository

Table 2: Relate works based on IDS

Researchers	Single	Hybrid	Anomaly	Signature	Algorithm	Data set	Limitations
Alomari and	-	✓	✓	-	Bees Algorithm (BA)+SVM	KDD cup 99	ELM lower computational requirements than SVMs, ELMs have shorter training time requirements than SVMs, ELMs work directly on multi-class classification problems
Shivhare and Chaturvedi	-	✓	✓	-	Back propagation+DBSCAN Algorithm+featuere selection method	KDD cup99	The computational cost using ELM is very small in comparison to back propagation Another problem of the conventional back propagation clearing algorithms is slow coverage rate
Senthilnayaki	-	✓	✓	-	GA+ Decision tree algorithm	KDD cup99	It is difficult to precisely model all behaviors since anomaly based detection can detect only known attacks
Deshmukh	-	✓	✓	-	Naive Bayes, decision tree	NSL-KDD	Bayes needs large data sets to work because the classes are assumed to be independent, and also it is difficult to estimate the actual probabilities of the network traffic
Fossaceca	-	✓	-	✓	Multiple Kernel-ELM	KDD cup99	The researcher during testing mode didn't depend on the data set testing mode to evaluate the results. This research evaluated based on KDD99 and we mentioned already the problems with this data set

Table 2: Continue

Researchers	Single	Hybrid	Anomaly	Signature	Algorithm	Data set	Limitations
Gholipour Goodarzi	-	✓	✓	-	SVM, ABC	KDD cup99	ELMs have shorter training time requirements than SVMs, ELMs work directly on multi-class classification problems. This research evaluated based on KDD99 and we mentioned already the problems with this data set
Deshmukh,	✓	-	✓	-	Naïve Bayes	NSL-KDD 99	Bayes needs large data sets to work because the classes are assumed to be independent, and also it is difficult to estimate the actual probabilities of the network traffic
Cheng *et al.*	✓	-	✓	-	ELM	KDD cup99	This research used normal ELM with the (2012) random select problem. This research evaluated based on KDD99 and we mentioned already the problems with this data set
Sangkatsanee,	✓	-	_	✓	Decision tree	RLD09 (Reliability lab data 2009)	Decision Tree uses a recursive approach to learning to divide data into specific classes, however decision tree algorithms canbe unstable or very complex
Lee	✓	-	✓	-	ID3 (Decision tree various)	KDD cup99	Most of their anomaly detection rates were lower than the detection rates with known/trained data. This research evaluated based on KDD99 and we mentioned already the problems with this data set
Peddabachigri,	-	✓	✓	-	Decision trees (DT), SVM	KDD cup99	SVM suffer from long training time and require parameters turning or don't perform well in multi-class classification. This research evaluated based on KDD99 and we mentioned already the problems with this data set

CONCLUSION

In this study, we have proposed the use of both the basic of ELM and GA for intrusion detection in a computer network. One of the advantages is that parameters during the learning process cannot be randomly assigned. This greatly increases the accuracy of ELM classifier and provides good scalability. Many works tried to apply ELM with IDS but this is the work proposed the ELM-GA hybrid based on IDS and this hybrid showed a good result when applied in another field. In this case, detection accuracy matters <speed to detect. Therefore, much research should be done in the future about this study in order to provide more results.

REFERENCES

Alexandre, E., L. Cuadra, S. Salcedo-Sanz, A. Pastor-Sanchez and C. Casanova-Mateo, 2015. Hybridizing extreme learning machines and genetic algorithms to select acoustic features in vehicle classification applications. Neurocomputing, 152: 58-68.

Ali, M.H. and M.F. Zolkipli, 2016. Review on hybrid extreme learning machine and genetic algorithm to work as intrusion detection system in cloud computing. ARPN. J. Eng. Appl. Sci., 11: 460-464.

Bhuyan, M.H., D.K. Bhattacharyya and J.K. Kalita, 2014. Network anomaly detection: methods, systems and tools. IEEE. Commun. Surv. Tutorials, 16: 303-336.

Cao, Z., J. Xia, M. Zhang, J. Jin and L. Deng*et al.*, 2015. Optimization of gear blank preforms based on a new R-GPLVM model utilizing GA-ELM. Knowl. Based Syst., 83: 66-80.

Cheng, C., W.P. Tay and G.B. Huang, 2012. Extreme learning machines for intrusion detection. Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN), June 10-15, 2012, IEEE, Brisbane, Queensland, ISBN:978-1-4673-1488-6, pp: 1-8.

Ding, S., H. Zhao, Y. Zhang, X. Xu and R. Nie, 2015. Extreme learning machine: Algorithm, theory and applications. Artif. Intell. Rev., 44: 103-115.

Elsayed, S.M., R.A. Sarker and D.L. Essam, 2014. A new genetic algorithm for solving optimization problems. Eng. Appl. Artif. Intell., 27: 57-69.

Feng, G., Z. Qian and X. Zhang, 2012. Evolutionary selection extreme learning machine optimization for regression. Soft Comput., 16: 1485-1491.

Haykin, S., 2000. Unsupervised Adaptive Filtering. John Wiley, New York, USA., ISBN: 978-0-471-37941-6, Pages: 200.

Huang, G.B., H. Zhou, X. Ding and R. Zhang, 2012. Extreme learning machine for regression and multiclass classification. IEEE Trans. Syst. Man Cybern. Part B (Cybern.), 42: 513-529.

Huang, G.B., Q.Y. Zhu and C.K. Siew, 2004. Extreme learning machine: A new learning scheme of feedforward neural networks. Proceedings of the IEEE International Joint Conference on Neural Networks, July 25-29, 2004, IEEE, Budapest, Hungary, ISBN:0-7803-8359-1, pp: 985-990.

Huang, G.B., Y.Q. Chen and H.A. Babri, 2000. Classification ability of single hidden layer feedforward neural networks. IEEE. Trans. Neural Netw., 11: 799-801.

Karlik, B. and A.V. Olgac, 2011. Performance analysis of various activation functions in generalized MLP architectures of neural networks. Intl. J. Artif. Intell. Expert Syst., 1: 111-122.

Lin, H.T. and C.J. Lin, 2003. A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods. Masters Thesis, National Taiwan University, Taipei, Taiwan.

Ozkan, C. and F.S. Erbek, 2003. The comparison of activation functions for multispectral landsat TM image classification. Photogramm. Eng. Remote Sens., 69: 1225-1234.

Pereira, G.C., M.M.D. Oliveira and N.F. Ebecken, 2013. Genetic optimization of artificial neural networks to forecast virioplankton abundance from cytometric data. J. Intell. Learn. Syst. Appl., 5: 57-66.

Peyman, K. and A.A. Ghorbani, 2005. Research on intrusion detection and response: A survey. Intl. J. Netw. Secur., 1: 84-102.

Wu, S.X. and W. Banzhaf, 2010. The use of computational intelligence in intrusion detection systems: A review. Appl. oft Comput., 10: 1-35.

Xiang, J., M. Westerlund, D. Sovilj and G. Pulkkis, 2014. Using extreme learning machine for intrusion detection in a big data environment. Proceedings of the ACM Workshop on Artificial Intelligent and Security Workshop, November 07-07, 2014, ACM, Scottsdale, Arizona, ISBN:978-1-4503-3153-1, pp: 73-82.

Xue, X., M. Yao, Z. Wu and J. Yang, 2014. Genetic ensemble of extreme learning machine. Neurocomputing, 129: 175-184.

Zhu, Q.Y., A.K. Qin, P.N. Suganthan and G.B. Huang, 2005. Evolutionary extreme learning machine. Pattern Recognit., 38: 1759-1763.