International Journal of Scientific Engineering and Technology ISSN : 2277-1581
Volume No. 7, Issue No. 10, PP : 96-99 1 Oct. 2018

Optimized Neural Network Using Genetic Algorithms in Network Intrusion Detection System

Calpephore Nkikabahizi [1], Prof . Cheruiyot w.k, PhD [2] & Dr Ann Kibe[3]
Department of Information Technology, Jomo Kenyata University of Agriculture and Technology
[1]calponkika@gmail.com, [2]wilchery68@gmail.com, [3]annkibe@gmail.com

Abstract*- The growth of datasphere and its usage which is exponentially increased in different areas, have led to the security mechanism. The intrusion detection system is one of security infrastructures that is found on the second line of the security, and is very crucial to confidentiality, integrity and availability. However, the existing intrusion detection systems are claimed to suffer from the issue of the false alarms, time consuming to detect the intruders, and large volume of alerts which is unmanageable and overwhelming to the human analyst. To fit this gap, this paper proposed a Genetic Algorithm for weights optimization of neural network classifier to escape the above mentioned issues, and then the research findings revealed that this classifier performs at rate of 99.68 % at low false positive rate 0.01%, with ROC value of 99.82%*

Key words: Genetic Algorithms , Artificial neural networks, Optimization, Intrusion Detection Systems, Confusion Matrix, ROC

1. Introduction

The growth of datasphere from entertainment data, non-entertainment image/video, productivity data, and embedded data (Reinsel et al. 2017), is expressed in 5'vs dimensions velocity, volume, variety , veracity, and value (Shen & Okyay , 2015) . This data growth and evolution of computing technology based on Bertalanffy's general system theory (Bertalanffy,1968 ; Mahoney, 1988) are resulting from baser sentiments which is the root of both material progress and the miseries and injustice associated with it in human development (Adam and Karl, 2007) . Moreover the data are also marginally increased over network under digitized form (SAN,2016) and require the security as human being need (Maslow ,1943). This complexity of data deluge and information technology tools require both physical and logical security known as cybersecurity (White,2014;Carney,2011),whereby intrusion detection systems are integrated to monitor the normal and abnormal traffic of data on network or host (Harris , 2013; Hutter, 2016).
The Intrusion detection system capable to detect the malicious activities has to fulfill the five important requirements, which are accuracy, completeness, performance, fault tolerance, and scalability (Dacier et al.,1999).The high detection rate at low false alarm rate and minimum error between targets and outputs depend on the algorithms that used in building model for classifying attack versus no attack to alarm the users .This process of minimizing error and maximizing detection rate is in line with optimization quote of *Leonhard Euler (1707-1783). "Since the fabric of the universe is most perfect, and is the work of a most wise Creator, nothing whatsoever takes place in the universe in which some form of maximum or minimum does not appear "*

To optimize the classifier, the researchers use different techniques such as the gradient descent method known as steepest descent or Cauchy's method (Gallero et al. , 2015; Yuchuan et al., 2016) , Newton–Raphson Method that based on the second order Taylor series expansion of the function f around the point x: (Makanjualo et al .,2015) , Nelder–Mead or downhill simplex method (Nelder,1965), a heuristic algorithm for multidimensional unconstrained optimization problems. Simulated annealing, which is stochastic method and its algorithm simulates the evolution of a solid in a heat bath to thermal equilibrium (Rasdi et al.,2015).Other techniques are stochastic approximation and evolutionary methods (Kingma & Lei , 2015; Skraba et al .,2016) includes differential evolutionary (Gonuguntla et al., 2015) ,and genetic algorithms (Nelder & Mead, 1965; Elena , 2015). Particle swarm optimization (Spichakova ,2016; Aljarah & Ludwig, 2013),Ant Colony Optimization and tabu search are also other optimization methods.
Neural network is one of classifiers and has a broad application in real world business problem, and is already successively applied in many industries, and has been carried out in different area such as speech recognition (Lippmann 1988),character recognition (Sharma & Chaudhary , 2013), signature verification (Abdul , 2015;Ashwini & Shalini ,2012) , human face recognition (Nisha & Sandeep ,2015) , Medical diagnosis (Qeethara , 2011), for the purpose of classification, (Zhang , 2000), forecasting (Cocianu & Grigoryan ,2015) , and clustering (Junyuan , Girshick & Farhadi ,2016). However, this classifier suffers from the issues of structure complexity and optimal weight vector parameter that result in minimum error and maximum detection rate.

This paper uses genetic algorithms to optimize neural network classier and find the optimal initial weight vector parameter which is set in the model, and thereafter to compare the results which are get with initial random weights and specific weights.

1.1 Genetic Algorithms

Genetic Algorithm (GA) is one of the optimization techniques (Mehrdad ,2003) ,and is based on Darwinian's theory of evolution and survival of fitness that make population effective candidates the fact which leads near a predicted fitness, and was invented by John Holland 1960 (Li, 2004).The idea of GA is in the same class with other Evolutionary Computation (EC) techniques such Evolution Strategy (ES) by Rechenberg (1960) in optimizing real valued parameter for a devices, Genetic Programming (GP) by

(Koza,1992), Evolutionary Programming (EP) by Fogel (1966) in evolving finite state machine in evolution of artificial intelligence ,evolution programs which is a particular kind of genetic algorithm by Mitchalewic (1996), and co-evolutionary algorithms that based on continuous interaction between population that are evolving by Poter (1997), Hills (1992).

The history of development search in evolutionary computing is shown in the following timeline EC(Rechernberg,1960)=GP(Koza,1992)+ES(Rechernberg,1965)+EP (Fogel ,1962)+GA(Holland ,1970). Genetic Algorithms are often able to find a good solution relatively and are also suitable for handling large complex population to find global solution which is difficult to find using classical optimization methods. According to Holland (1975), Genetic Algorithms use three operators , which are selective operator (Goldberg and Deb,1991), crossover operator which should be one –point crossover (Reeves and Rome ,2003), Two-point crossover (Kaya ,2011), Intermediate /Blending crossover , Heuristic crossover ($xi\ (t+1) = xi(t) + \alpha\ (yi(t) - xi(t))$) , $\alpha\ \varepsilon < 0, 1 >$ Arithmetic ($xi\ (t+1) = \alpha\ * xi(t) + (1-\alpha) * (yi(t))$,or uniform crossover with probability of 0.5 for exchanging bit (Syswerda 1989);and/or mutation operator which consists on changing chromosomes based on theory of parthenogenesis by Owen (1849) and restrain immature convergence phenomenon of crossover (Jian et al. , 2005).

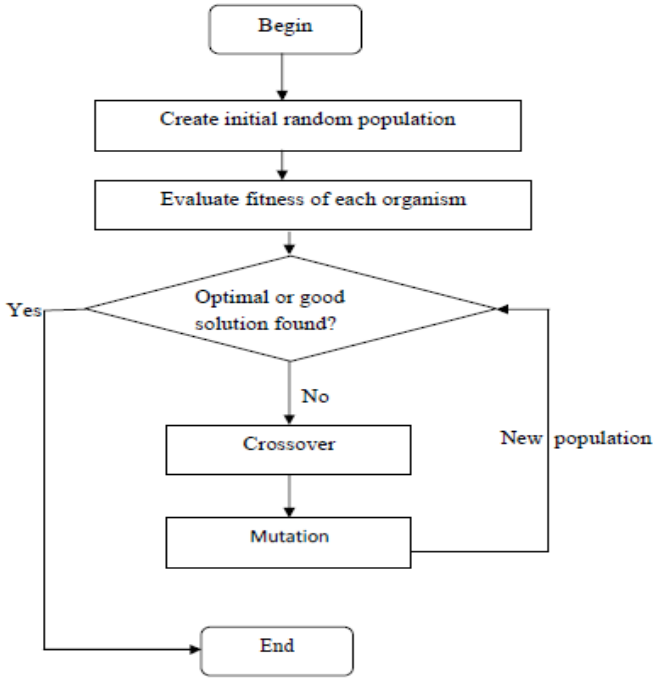

Figure 1.1: Flowchart of Genetic Algorithm

2. Related Work

Badal et al. (2014) proposed multi-objective algorithms to optimize artificial neural network architectures that lead to enhanced results and reduction in computation cost for forecast electrical load .
Zhou and Xiong (2013) conducted an experiment by training the initial weights of backpropagation neural network by genetic algorithms before constructing a new classifier . Using the Iris data , the results shows that backpropagation neural network improved by Genetic Algorithm has a high accuracy classification and greater gradient of convergence than of backpropagation neural network proposed before the experiment.

Montana and Davis (1898) Conducted a set of experiments on a sonar image classification problem, and then the results showed that genetic algorithm is suit for complex pattern classification which find a nearly globally optimal set of weights in considerable short time

Dharmistha (2012) simulated the data to evaluate the performance of Ad hoc On Demand Distance Vector (AODV) routing protocol on Mobile Ad hoc Network (MANET) to determine hello interval , whereby the neural network weights were optimized using genetic algorithms, and thereafter the results revealed that weight and biases are satisfactory trained relatively to traditional neural network training .

Leung et al. (2003) presented tuning of the structure and parameters of a neural network using an improved genetic algorithm, and then their experiment results showed that the improved GA perform better than the standard GA on sunspot forecasting and associative memory.

Yu-Tzu et al. (2012) conducted a study on patients' sample to find the optimal set of initial weights to enhance the accuracy of artificial neural network by using genetic algorithm for Hip bone fracture prediction, and then the results revealed that GA has been proved to be effective for improving the accuracy of artificial neural networks.

3. Methodology and material

3.1 Study Sample

The research uses a sample size of 57,279 records that is selected using a stratified and purposive methods .This sample size satisfies the two important criterions ,one hand is greater than the minimum sample size found using Power laws, and on another hand it converge on the same point as the entire population , $\Pr(acc(N) - acc(n) > \varepsilon) \leq \delta$. The structure of optimized neural network has 34 inputs neurons , 28 hidden neurons and 5 outputs .To initialize the vector parameter of weights and biases , each layer is connected with a specified weights and biases through which information are propaged from input neuron to all hidden neurons , and then thereafter information processes use transfer function sigmoid at each hidden neuron, and the optimizer is mean square error . The chain rule is still used to update the weights until the stopping criteria are met or minimum error obtained.

3.2 Genetic Algorithms Parameters

In the current study, genetic algorithms parameters are determined as follows , creation function is constraint dependent, roulette as selective function , scattered mutation function , heuristic crossover method , number of generation

ISSN : 2277-1581
1 Oct. 2018

=10, population size =200, and then the remain parameters are used by default .

The fitness function for this research to be optimized is

$$\min imize = \begin{cases} E(w) = \dfrac{1}{1length(Inputs) * \sum_{i}^{5} c_i} \left[\sum_{j=1}^{s} \sum_{i=1}^{s} (net(Inputs)_i - t\arg ets_i)^2 \right] \\ or \\ Acc(w) = -\left[\dfrac{\sum_{i=1}^{5} cm(i,i)}{\sum_{i=1}^{5} \sum_{i=1}^{5} cm(:,i)} \right] \end{cases}$$

The model purpose is to minimize error, E(w), and maximize the probability that algorithm can correctly predict positive and negative examples, whereby w represents the weights. This probability or accuracy of model is represented by Acc (w) ,and Cm (i,j) is confusion matrix entry on i^{th} row and j^{th} column.

The number of variables is 1125 and are represented in one vector parameter, Wb_{1x1125} ,and thereafter is separated into four vectors , input bias weight , b_{28x1} ,input weight matrix , IW_{28x34} , output bias weight, b_{5x1} ,and then output weight matrix, Lw_{5X28} .

These sub- vectors have been set into the model , , trained, tested , evaluated and then compared to its results using initial specific weights, Wb_{1x1125} , and initial random weights and biases. To test this experiment, the research has also used other weight vector parameters which are selected from a pool of offspring's of population.

4. Result and discussion

Table 4. 1: Comparative results from BP and GA

Training Algorithm	Classification Accuracy (%)	Performance	Time (seconds)	epoch (iteration)
Backpropagation	99.66	0.00475	36	249
$net.\vec{Wb}\left[\vec{IW},\vec{b1},\vec{b2},\vec{LW}\right]$	99.66	0.00482	0.01	6
$net.\vec{\lvert Wb\rvert}\left[\lvert\vec{IW}\rvert,\lvert\vec{b1}\rvert,\lvert\vec{LW}\rvert,\lvert\vec{b2}\rvert\right]$	99.50	0.00637	31	213
$net.\vec{Wb}\left[\vec{IW},\lvert\vec{b1}\rvert,\lvert\vec{b2}\rvert,\lvert\vec{LW}\rvert\right]$	99.68	0.00477	17	115

The results from table 4.1 reveal that classification accuracy of neural network 34I/28H/5O has increased from rate of ninety nine point sixty six (99.66 %) percent by random initial weights to rate of ninety nine point sixty eight (99.68 %) percent by initializing an optimal weight vector $\vec{Wb}_{1x1125}$ in the model . Every component of $\vec{Wb}'_{1x1125}$ is normalized using minmax technique except the input weight matrix, as follows

$$\vec{Wb}'_{1x1125}\left[\vec{IW}_{28\times34},\lvert\vec{b1}\rvert_{28\times1(\min\max)},\lvert\vec{LW}\rvert_{5\times28(\min\max)},\lvert\vec{b2}\rvert_{5\times1(\min\max)}\right]$$

The time which was spent to build the model was decreased from 36 seconds to 17 seconds, and number of iteration was also decreased from 249 to 115,and the small changes has been noticed on performance results from 0.00475 to 0.00477.

Conclusion

This paper has discussed the genetic algorithms that optimize neural network classifier, and then the results revealed that the optimal weight vector parameter has been proved to be effective in improving majority of metrics such as performance, classification accuracy, ROC,R-squared, number of iteration and time cost for model training . The experiment has led the research to deduce the following assumption : Most of $wb_i \in P$,where wb_i and p are respectively weights and biases vectors , and population intervene in optimizing classification accuracy , Acc(w) ,and classification error , E(w) . If given two vectors $wb_1 \in \vec{Wb}\left[\vec{IW},\vec{b_1},\vec{b_2},\vec{Lw}\right]$ and $wb_2 \in \vec{Wb}\left[\vec{IW},\lvert\vec{b_1}\rvert,\lvert\vec{b_2}\rvert,\lvert\vec{LW}\rvert\right]$

$then\ \exists!\ \delta \in \square_+, \sigma \in \square_-$ such that $\delta = Acc(wb_2) - Acc(wb_1), and \sigma = E(wb_2) - E(wb_1)$ i.e the vector wb_2 increases the accuracy and decreases the classification errors.

REFERENCES

i. *Adam, S. & Karl , M. (2007) . Human development. The culture mandala: The Bulletin of the Centre for East-West Cultural and Economic Studies ,7(2), 1-12.*

ii. *Aljarah, I., & Ludwig, S. (2013). MapReduce Intrusion Detection System based on A Particle Swarm Optimization Clustering Algorithm: Evolutionary Computation, 955 - 962.*

iii. *Ashwini .P. & Shalini B.(2012). Handwritten Signature Verification using Neural Network. International Journal of Applied Information Systems 1(2), 44-49*

iv. *Badal ,U. I., Zuhairi,B., Muhammad,Q. R., & Perumal , N. (2014). Optimization of neural network architecture using genetic algorithm for load forecasting. Intelligent and advanced System, 5th International conference.*

v. *Carney, J. (2011).Why integrating physical and logical security .White paper , 1-8.*

vi. *Cocianu , C. & Grigoryan, H. (2015). An Artificial Neural Network for Data Forecasting Purposes. Informatica Economică , 19 (2) , 34-45.*

vii. *Dacier , M., Debar, H. & Wespi, A.(1999. Towards a taxonomy of intrusion detection systems. Computer Networks 31, 805–822.*

viii. *Dharmistha,V. D. (2012). Genetic Algorithm based Weights Optimization of Artifical Neural Network. International Journal of advanced research in electrical , electronics and instrumentation engineering ,1(3),206-2011.*

ix. Gallero, F.,A. , Quntero, J.&Riano J.C. (2015).Convergence of the steepest descent method ith the line search and uniformly convex objective in reflexing banach spaces. Mathematical communication 3(2), 47-50

x. Goldberg, D. ,E.& Deb, K. (1991).A comparative analysis of selection schemes used in genetic algorithms Foundation of genetic algorithms . San Mateo, Morgan Kaufmann.

xi. Gonuguntla ,V.Mallipeddi ,R. & Kalyana, C.V.(2015). Differnetial evolutional with population and strategy parameter adaptation mathematical problems in engineering .Hindawi, 1-10

xii. Harris ,S.(2013).Physical and environmental security In CISSP Exam Guide 6th ed,427-502. USA MacGraw –Hill.

xiii. Hillis,W.D.(1992).Co-evolving parasites improve simulated evolution as an optimization procedure. Addison-Wesley, Redwood City

xiv. Holland,J. H.(1975).Adaption in natural and artificial systems. The University of Michigan Press.Ann Arbor , MIT.

xv. Hutter,D.(2016). Physical Security and why it important. The NANS institute. InfoSec reading Room.1-31.

xvi. Jian-Cong ,B., Hui-you ,C., & Yang ,Y (2005). A partheno-genetic algorithm for multidimensional knapsack problem. In Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference, 5, 2962–2965.

xvii. Junyuan , X. , Girshick ,R. & Farhadi ,A. (2016). Unsupervised deep embedding for clustering analysis.ICML'16 Proceedings of the 33rd International Conference on International Conference on Machine Learning 48, 478-487,New York USA

xviii. Kaya, M. (2011). The effects of two new crossover operators on genetic algorithm performance. Applied Soft Computing, 11, 881–890.

xix. Leung ,F.H., Lam, H. K., Ling, S.H. & Tam, P.K.. (2003).Tuning of the structure and Parameters of a neural network using an improved genetic algorithms. IEEE Transaction on ,14(1),79-88.

xx. Li, W. (2004). A Genetic algorithm approach to Intrusion Detection System . SANS institute, USA.

xxi. Kingma,D.,P.& Lei,B.,J.(2015).Adam: A methods for stochastic optimization .ICLR 1-15.

xxii. Koza, J.R. (1992).Genetic programming :on the programming of computers by means of natural selection . The MIT press ,Cambridge, Massachusetts. USA.

xxiii. Mahoney ,S.M. (1988). The History of Computing in the History of Technology. Annals of the History of Computing 10(1988), 113-125).

xxiv. Makanjualo,F.A. ,Kayode J.A.,&Adejoke,O.D (2015).On Quasi-newton mewthon method for solving unconstrained optimization problems .American journal of applied mathematics3(2) ,47-50.

xxv. Maslow ,A. H. (1943). A theory of human motivation .Psychological review ,50,370—396.

xxvi. Mehrdad Dianati,(2003).An introduction to Genetic algorithms and Evolution strategies .Insop Song and Mark Treiber.

xxvii. Montana ,I.D. & Davis, L.(1898).Training feed forward neural networks using Genetic Algorithm. IJCI'89 proceedings of the 11th international conference on artificial intelligence 1,762-767. Morgan Kaufmann , San Francisco, USA.

xxviii. Nelder, J.A. & Mead , R. (1965). A simplex method for function minimization. Computer Journal 7, 308–313

xxix. Nisha & Sandeep, D. (2015).Face Detection and Expression Recognition using Neural Network Approaches. Global Journal of Computer Science and Technology: F Graphics & Vision 15(3),21-24.

xxx. LIPPMANN ,R.P. (1988).Neural Network Classifiers for Speech Recognition .The Lincoln Laboratory Journal, 1, (1), 107-124

xxxi. Qeethara,K.A.(2011). Artificial Neural Networks in Medical Diagnosis. IJCSI International Journal of Computer Science Issues ,8 (2) , 150-154

xxxii. Rasdi,L. M. , Fanany M.I.& Aniati, M.A.(2015).Simulated annealing algorithms for deep learning . The 3rd Information systems international conference . Procedia computer science 72,137-144.

xxxiii. Reeves, C. R. & Rome, J. E. (2003). Genetic Algorithms Principles and Perspectives, Kluwer Academic Publishers. Dordrecht, 2003.

xxxiv. Reinsel , D., Gantz, J. & Ryding (2017).Data age 2025:The evolution of data to life critical. Do not focus on data, focus on the data that's big. An IDC white paper ,1-25.

xxxv. Sharma,A., Chaudhary,D.R..(2013) Character Recognition Using Neural Network. International Journal of Engineering Trends and Technology (IJETT) 4(4)-662-667

xxxvi. Skraba,A.,Stanovov ,V.Semenkin,E.,&Kofjac , D. (2016).Hybridazationof stochastic localsearch and genetic algorithms for human resource planning management. Organizacija 49,42-54.

xxxvii. Shen,Y., & Okyay , K.. (2015).Big data for modern industry :Challenges and trends. Proceedings of IEEE, 103(2), 143-146.

xxxviii. Spichakova ,M.(2016).Modified partical swarm optimization algorithms based on gravitational field interaction .Proceedings of Estonia academy of science 65(1) 15-27.

xxxix. Sukumaran, S. & Kesavaraj,G. (2013). A study on classification techniques in data mining. IEEE 4th international Conference on computing , communication and network technologies

xl. Syswerda, G. (1989). Uniform crossover in genetic algorithms. Proceeding 3rd International conference on genetic algorithms, Morgan –Kaufmann, USA .

xli. Zhou,W.& Xiong, S. (2013).Optimization of BP neural network classifier using Genetic algorithm Intelligent computation and evolutionary computation , 599-605.

xlii. Yuchuan,Q., Baldur ,V.L., Lelievetdt,B.P.F.,& Staring, M. (2016).Fast automatic step size estimation for gradient decent optimization of image registration .IEEE Transaction on medical imaging, 35(2),391-403.

xliii. Yu-Tzu,C., Jinn, L., Jiann-Shing, S. & Maysam, F. A. (2012).Optimization the Initial Weights of Artificial Neural Networks via Genetic Algorithm Applied to Hip Bone Fracture Prediction . Advanced in Fuzzy Systems, 1-10.

xliv. Whitley D. (1995). Genetic Algorithms and Neural Networks. Genetic Algorithms in Engineering and Computer Science. John Wiley.