IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 6 Issue 9, September 2019
ISSN (Online) 2348 – 7968 | Impact Factor (2019) – 6.248
www.ijiset.com

Hybridization of Neural Network-cum-Evolutionary Algorithm variants for Fraud Detection in Credit Cards Online Transaction

Oladimeji W. Ismaila[1], Folasade. M Ismaila[2]

[1] Department of Computer Science and Engineering, Ladoke Akintola University of Technology, Ogbomoso, Nigeria, woismaila@lautech.edu.ng

[2] Department of Computer Science, Osun State Polytechnic, Iree, Osun State, Nigeria. folaismaila@gmail.com

Abstract

Fraud is now regarded as an adaptive crime and it is increasing day by day. So, it needs special methods of intelligent data analysis to detect and prevent it. Existing fraud detection systems may not be so much capable to eliminate or reduce fraud transaction rate. Improvement in fraud prevention/detection practices has become important to maintain existence of payment system. Hence, this work investigated the extent to which Radial Basis Function (RBFN)-cum-Particle Swarm Optimization (PSO) was used to detect frauds in credit card online transactions.

The simulated dataset used contained 100 cards (60 transactions each) with legal transactions sparsely intertwined with malicious types. The work flow of the proposed system consisted of data preparation, implementation and evaluation phases. The metrics used were false alarm rate, recall, precision and accuracy. The results showed that RBFN were 80.8%, 14.2%, 71.4% and 71.0% while RBFN-PSO generated 95.1%, 23.0%, 91.7% and 93.9% for accuracy, false alarm rate, precision and recall respectively.

Keywords: *Fraud detection systems, Simulated dataset, RBFN, PSO, False alarm rate*

.

1. Introduction

The Internet has become the backbone for conducting electronic commerce. Many products, tangible and intangible, are browsed through and sold over the Internet. Although there are a number of possible payment methods, such as electronic cash, electronic cheque, debit/credit card, and electronic wallets but debit and credit cards stands out in usage. Thus, increase in credit cards transactions caused proliferation of online related frauds. E-commerce frauds can be broadly classified into three categories, i.e., traditional card related frauds, merchant related frauds and Internet frauds. The different types of methods for committing credit card frauds are described below [1]. Fraud will continue to be a crime of opportunity and the thieves will continue to steal methodically from companies until those companies go out of business. But there is a need to work to stop them in their tracks and turn their "perfect crime" into a big mistake. Credit card fraud was defined as "Unauthorized account activity by a person for which the account was not intended while Credit card fraud detection system is a computer program that attempts to perform fraud detection by identifying fraud or fraud transaction as quickly as possible once it has been perpetrated. Credit card fraud detection has drawn a lot of research interest and a number of techniques, with special emphasis on data mining and neural networks [2]. Existing fraud detection systems may not be so much capable to reduce fraud transaction rate. Improvement in fraud detection practices has become essential to maintain existence of payment system [3]. The aim of this work was to develop fraud detection system in online transaction using radial basis function modified by particle swarm optimization.

Evolutionary algorithms have been employed over the years to curb online transaction frauds. The authors in [4] presented artificial intelligence techniques, viz MLP neural networks and particle swarm optimization algorithm to detect intrusion and attacks of unauthorized users in cloud computing. . The methods were tested on two databases. The results showed improved accuracy in detecting attacks and intrusions by unauthorized users. [5] researchers proposed a Adaptive Neuro-Fuzzy Inference System (ANFIS) to detect bank credit cards related frauds. By combining evolutionary algorithms with ANFIS, the optimal tuning of ANFIS parameters is achieved by the Teaching-Learning-Based Optimization (TLBO) and the Particle Swarm Optimization (PSO) in order to improve the network performance and to reduce calculation complexities. The proposed algorithm is implemented and evaluated on credit cards data to detect fraud. The results demonstrate superior performance of the designed scheme compared to other intelligent identification methods. [6] authors proposed system that mainly concentrate on finding the optimum membership functions of a fuzzy system using particle swarm optimization (PSO) algorithm. The proposed algorithm was used to optimize the Gaussian membership functions of the fuzzy model system. It was clearly proved that the optimized membership functions (MFs) provided better performance

than a fuzzy model for the same system, when the MFs were heuristically defined.PSO has no evolution operators such as crossover and mutation. The researchers in [7] developed an hybridized system (particle swarm optimization was used to improve the performance of the artificial neural network) in order to recognize the fraud pattern more accurately. In their work,the accuracy, sensitivity, and specificity in bank fraud detection were as much as 90.32%, 88.06%, and 90.12%, respectively.

The researchers in [9] applied the neural data mining method on online transactions. This model was based on customer's behaviour pattern. Deviation from the usual behaviour pattern was taken as an important task to create this model. The neural network was trained with the data and the confidence value was calculated. The credit card transaction with low confidence value was not accepted by the trained neural network and it was considered as fraudulent. If the confidence value was abnormal, then again it was checked for additional confirmation. The detection performance was based on the setting of threshold. Related work can be found in [2, 8]. [12] authors proposed the combination of the synthetic minority oversampling technique (SMOTE) and the radial basis function (RBF) classifier to deal with classification for imbalanced two-class data. In order to enhance the significance of the small and specific region belonging to the positive class in the decision region, the SMOTE is applied to generate synthetic instances for the positive class to balance the training data set. Based on the over-sampled training data, the RBF classifier is constructed by applying the orthogonal forward selection procedure, in which the classifier structure and the parameters of RBF kernels are determined using a particle swarm optimization algorithm based on the criterion of minimizing the leave-one-out misclassification rate. The experimental results on both simulated and real imbalanced data sets are presented to demonstrate the effectiveness of the proposed algorithm. [13] presented a combined Self-Organizing Maps (SOMs), unsupervised neural network, and Particle Swarm Optimization to introduce a new method for anomaly detection. This method was performed on forest fire detection. The simulated dataset is constructed of 14,987 normal cases and 104 abnormal cases, and the real dataset is constructed of 422 normal cases and 95 abnormal cases. In fact, the results showed that the new method would be a generic algorithm for anomaly detection that may need few changes for implementation in different domains. Also, the authors in [14] presented a novel method of combining the use of Digital Signature of Network Segment (DSNS) with the evolutionary technique called Particle Swarm Optimization (PSO) and neural network training, applied in a real data set. The proposed anomaly detection system uses the Support Vector Machine in order to clusterize the traffic collected by SNMP agents and its respective DSNS. The PSO is combined with the SVM in order to improve performance and quality of the solution in the clusterization and calculation of clusters centroids. Numerical results showed that the obtained detection and false alarm rates are promising.

2. Materials

2.1 The RBF Network

RBF are embedded into a two-layer feed-forward neural network. The set of input can be attributes that are common to various classes needed for classification. While the output units are the various classes that the attributes can resolve to. In between the inputs and the outputs is a layer of processing units called hidden unit and each of them implements a radial basis function. Training of RBF network is accomplished in two stages: parameters of the radial basis functions are set so that they approximate model the unconditional data density of the training set, after that output weights are learned. In pattern classification problems, the Gaussian function demonstrates higher accuracy. [11]

The weights and the center of activation functions are adjusted using the gradient descent method to minimize the Sum of Squared Error (SSE). The hidden layer function is calculated by the equation 1.

$$F(x) = \sum_{i=1}^{N} w_i \emptyset(\|x - u_i\|) \quad (1)$$

where: N is the number of neurons in the hidden layer, u_i is the center vector for neuron i, and w_i is the weight of neuron i in the linear output neuron.

2.2 Particle Swarm Optimization

Particle swarm optimization (PSO) is a stochastically global optimization method that belongs to the family of Swarm Intelligence [16] and Artificial Life. PSO is based on the principles that flock of birds, school of fish, or swarm of bees searches for food sources where at the beginning the perfect location is not known. However, they eventually reach the best location of food source by means of communicating with each other. [15]. PSO has many advantages which includes: effective in nonlinear optimization problems; easy to implement; Only a few input parameters need to be adjusted in PSO and can be efficiently used on large data sets.[10].

2.3 Hybridized RBF-PSO model

This section discussed the optimization of Radial Basis Function (RBF) by Particle Swarm Optimization (PSO). PSO optimization was interposed into this framework in order to optimize the RBF training parameters so that the best particles having optimal parameter settings can be obtained. An initial population size was assumed and fed into RBF framework that performs the classification. Error Function is the average error incurred when RBF classified the large input data. Our objective was to minimize Error Function and obtain the best particle for an optimal set of RBF parameters that was used for prediction purposes. The initial weights were randomly selected.

In PSO algorithms, the particle swarm is initialized randomly in searching space and each particle has initial speed and position. So the searching quality and the speed have randomness. The path of particle is updated through individual best position and the path of swarm is updated via global best location, which is found by the entire population (population size of 100 is used). This makes particles move to the optimal solution. The Stepwise procedure of the RBF-PSO algorithm are presented below:

Step 1: Generate random population of N, Set parameter ω_{min}, ω_{max}, c_1 and c_2 of PSO

Step 2: Initialize population of particles having positions x_j and velocities v_j

Step 3: Set iteration k = 1

Step 4: Calculate fitness of particles $F_{ij} = f(\widehat{H}_{LS})$ and find the index of the best particle *b*

The fitness function for parameters selection is as follow

The length of the particles is determined by the number of input factors, the number of layers, and the number of nodes in each layer. The total length of the particle L was calculated as

$$\widehat{H}_{LS} = n_{input} * n_1 + \sum_{i=1}^{k-1}(n_j * n_{i+1}) + n_k \qquad 5$$

Where " n" input is the number of input attributes for the RBF, k is the number of internal layers, and n_i is the number of nodes in layer. The last term in the equation is for the weights between the last internal layer and the output layer, which consists of a single node.

Step 5: Select $Gbest_{ij} = \widehat{H}_{LS}$ and $Pbest_{ij} = H$

Step 6: $\omega = \omega_{max} - k \times (\omega_{max} - \omega_{min})/Max\_no$

Step 7: Update velocity and position of particles

$$\vec{v}_{ij} = \omega\vec{v}_{ij} + c_1 r_1\left(P_{best} - \widehat{H}_{LS}\right) + c_2 r_2\left(P_{best} - \widehat{H}_{LS}\right) + c_3 r_3\left(G_{best} - \widehat{H}_{LS}\right)$$

$$\vec{x}_{ij} = \vec{x}_{ij} + \vec{v}_{ij}$$

Step 8: Evaluate fitness $F_{bj} = f(\widehat{H}_{LS})$ and find the index of the best particle b_1

Step 9: Update $Pbest$ of population

If $F_{ij} < F_{bj}$ then $Pbest_{bj} = H$ else

$$Pbest_{ij} = Pbest_{bj}$$

Step 10: Update $Gbest$ of population

If $F_{ij} < F_{bj}$ then $Gbest_j = Pbest_{bj}$ and set $b = b_1$ else

$$Gbest_{bj} = Gbest_j$$

Step 11: If $k < Max\_no$ then $k = k + 1$ and goto step f else goto step l

Step 12: Output optimum solution as $Gbest_{bj}$.

$$Gbest_{bj} = H_{LS}$$

3. METHODOLOGY

The proposed credit card transaction system mainly includes the two phases: data preparation phase and implementation phase. The workflow is shown in figure 1.

Data Preparation

A simulator was used to generate a mix of genuine and fraudulent transactions. This involves preparing the accumulated data for training. It can be regarded as the data mapping phase which has to do with matching the parameters of RBF-PSO to selected variables of cardholder's transaction data. The accumulated data was presented in the form acceptable by the model with respect to its parameter.

The dataset was encoded to include several indicator variables to make it suitable for the RBF algorithm and the categorical response variables anticipated. Several attributes that are ordered categorical have been coded as integer, for instance the predicted response class label y, was dichotomously defined as follows:

$$y = W(x) = \begin{cases} 0, & if\ a\ transaction\ is\ genuine \\ 1, & if\ s\ transaction\ is\ malicious \end{cases} \quad (2)$$

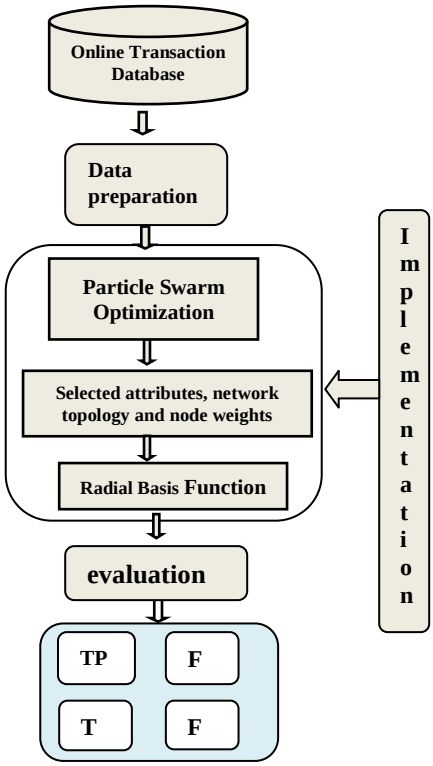

Figure 1: Workflow of the proposed system

Implementation phase

The implementation phase encompassed the training phase, prediction phase and detection engine. The detection phase employed the parameters from trained RBF-PSO model to identify the new transactions as fraudulent or not.

Evaluation phase

In this work, confusion matrix was employed for evaluating the results of the proposed system. The confusion matrix was used estimate these metrics viz; accuracy, precision, recall and false alarm rate.

Confusion matrix		Predicted class	
		C	NC
Actual class	C	FP	TN
	NC	TP	FN

Normal class – NC ***Novelty - C***
FP-false positive TP-true positive
FN-false negative TN-true negative

$$FAR = \frac{FP}{FP+TN} * 100 \quad (3)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}}\ x\ 100 \quad (4)$$

$$Precision = \frac{\text{FP}}{\text{TP+FP}}\ x\ 100 \quad (5)$$

$$Accuracy = \frac{TP+TN}{TP+FN+TN+FP} x\ 100 \quad (6)$$

4. Implementation and Results

The interface of developed RBN-PSO model was designed using C-sharp language, as shown in figure 2. The dataset used contain 100 cards which consist of 60 transactions each with fraudulent type and non-fraudulent type, 30 transactions of each card were used to trained the system and 30 transactions were used to test the system. The total number of transactions used for training amount to 1500 all together for 50 cards and 1500 in total for 50 cards was used for testing while 3000 transactions for 50 cards were left untrained.

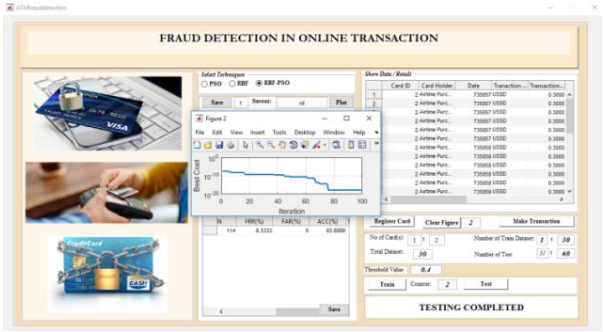

Figure 2: RBFN-PSO fraud detection system interface

The results, as shown in table 1, showed that PSO produced overall accuracy of 78.8%, precision and recall were 68.3% and 68.0% and false alarm rate had 15.8%. The RBFN produced overall accuracy of 80.8%, false alarm rate of 14.2%, overall precision and recall were 71.4% and 71.0%. While the hybridized RBFN-PSO system produced overall accuracy of 95.1% , false alarm rate of 4.2%, precision of 91.7 % and recall of 93.9%.

Table 1: Table showing results generated with PSO, RBF and RBF-PSO

Technique	TP	FN	FP	TN	FAR (%)	RECALL (%)	PREC (%)	ACC (%)	Total Time(sec)
RBFPSO	1408	92	127	2873	4.23	93.87	91.73	95.13	118.35
RBF	1065	435	427	2573	14.23	71.00	71.38	80.84	1591.12
PSO	1020	480	474	2526	15.80	68.00	68.27	78.80	1960.78

5. Conclusions

PSO and RBFN demonstrated powerful problem-solving ability. They were based on quite simple principles, but took advantage of their mathematical nature: non-linear iteration. Particle Swarm Optimizations are global search methods that are based on principles like personal learning coefficient and global learning coefficient. This study examined how Particle Swarm Optimizations can be used to optimize the network topology of RBFN, to enhance RBFN predictive capability. The RBFN-PSO system produced low false alarm rate and fastest matching time which is required for effective online fraud detection. Creating and implementing comprehensive credit fraud detection techniques that include prevention, detection and resolution components not only will limit losses but also will strengthen businesses.

Acknowledgments

I acknowledge the Tertiary Trust Fund (TETFUND) Nigeria for sponsoring this research work and publication.

References

[1] R. Patidar,. and L. Sharma, (2011).Credit Card Fraud Detection Using Neural Network. International Journal of Soft Computing and Engineering (IJSCE), India: Jaipur, Vol 1, pp. 32-38.

[2] S. O. Falaki, B. K. Alese, O. S. Adewale, Ayeni J. O., Aderounmu G. A. and Ismaila W. O. (2012) "Probabilistic Credit Card Fraud Detection System in Online Transactions" In: International Journal of Software Engineering and Its Applications, Vol. 6, No. 4
[3] M. Z. Khan, J. D. Pathan, A. H. Ahmed (2014) "Credit Card Fraud Detection System Using Hidden Markov Model and K-Clustering" In: International Journal of Advanced Research in Computer and Communication Engineering, 3(2), 5458-5461.
[4] S. Ahmad, M. Mehrdad, M. Hamid (2017). Attacks and Intrusion Detection in Cloud Computing Using Neural Networks and Particle Swarm Optimization Algorithms, Emerging Science Journal, Vol. 1, No 4.
[5] M. Ghodsi,, M. Abadeh, (2017). Fraud Detection of Credit Cards Using Neuro-fuzzy Approach Based on TLBO and PSO Algorithms, Journal of Computer & Robotics 10 (2), 2017 57-68
[6] M.C. Kavitha, S. R. Kumari (2013).Particle Swarm Optimization For Adaptive Anomaly-Based Intrusion Detection System Using Fuzzy Controller, International Journal of Computer Trends and Technology (IJCTT) – volume 4 Issue10.
[7] N. Safaria, T. Banirostam, (2018). A New Solution to Reduce Bank Fraud Detection Fault with Particle Swarm Optimization Algorithms, International Journal of Research in Science and Engineering, Vol. 6, No. 3, 2018, pp. 81-89.
[8] Z . Valdimir and A. Strizhak, (2006) "Credit Card Fraud Detection using Self Organizing Maps", Information and Security: An International Journal, 18: 48-63.
[9] G. Tao and L. Gui-Yang (2008), "Neural Data Mining for Credit Card Fraud Detection", International Conference on Machine Learning and Cybernetics, 7; 3630-3634.
[10] X. Xiang, R.D. Ernst, E. Russell, B.M. Zina, and J.O. Robert, (2003).''Gene clustering using self-organizing maps and particle swarm optimization'', Ipdps, International Parallel and Distributed Processing Symposium IPDPS'03, pp. 154b.
[11] E.O. Oyebode, S.G. Fashoto, O.A. Ojesanmi and O.E. Makinde (2011). Intrusion Detection System for Computer Network Security, Australian Journal of Basic and Applied Sciences, 5(12): 1317-1320, 2011.
[12] G. Ming, H. Xia, C. Sheng and J. H. Chris (2011).On Combination of SMOTE and Particle Swarm Optimization Based Radial Basis Function Classifier for Imbalanced Problems, Proceedings of the International Joint Conference on Neural Networks, USA.
[13] M., S. Lotfi, D. Moazzamia, B. Moshiri, M.R. Delavar, (2011). Anomaly detection using a self-organizing Map and particle swarm optimization, Scientia Iranica D (2011) 18 (6), 1460–1468.
[14] D. Dutta and K. Choudhury (2013). Network Anomaly Detection using PSO-ANN, International Journal of Computer Applications (0975 – 8887) Volume 77– No.2.
[15] J. Kennedy,; R. Eberhart, (1995). "Particle Swarm Optimization". Proceedings of IEEE International Conference on Neural Networks.
[16] N. Nedjah and L. M. Mourelle, Swarm Intelligent Systems. Springer- Verlag Berlin Heidelberg: Springer, 2006.
[17] I. O Alabi. and R. G. Jimoh (2018) Financial Fraud Detection using Radial Basis Network, Circulation in Computer Science, Vol.3, No.1, pp: (10-21).

Ismaila W. Oladimeji hasfirst degree in Computer science from Department of Mathematics and Computer Science, Federal University of Technology, Minna in 1998. He got his Master of Technology and PhD in Computer science from Department of Computer Science, Federal University of Technology, Akure in 2004 and 2012. He currently lectures as a Senior lecturer at Department of Computer Science and Engineering, Ladoke Akintola University of Technology, Ogbomoso, Nigeria. His area of research are Soft Computing, Modeling and Simulation and Computer Networks and Cyber-security. He has supervised about eighty-five graduates and twenty postgraduate students. He has published over thirty papers in both national and international journals. He is an associate member of National Computer Society and Computer Professionals of Nigeria.

Ismaila Folasade Muibat got her bachelor of technology in computer science from Department of Computer Science and Engineering, Ladoke Akintola University of Technology, Ogbomoso, in 2008. She has master in Information Technology in 2015 at national open university, Nigeria. She currently lectures at Department of Computer Science, Osun State Polytechnic, Iree. She specialized in research areas as soft computing, Computer Networks and Cyber-security. She has about twelve publications. And has attended about six conferences. She has supervised over forty students in projects.