

A Hybrid Simulated Annealing and Back-propagation Algorithm for Feed-forward Neural Network to Detect Credit Card Fraud



Ardalan Husin Awlla

Ministry of Education, Sulaimani 46001, Iraq

ABSTRACT

Due to the ascent and fast development of E-commerce, utilization of credit cards for online buys has significantly expanded, and it brought about a blast in the credit card fraud. As credit card turns into the most prevalent method of installment for both online and also normal buy, cases of fraud associated with it are additionally rising. In actuality, false exchanges are scattered with veritable exchanges, and basic example for coordinating procedures is not frequently adequate to identify those frauds accurately. Usage of effective fraud recognition frameworks has in this manner gotten to be basic for all credit card distributing banks to decrease their losses. Many current systems based on artificial intelligence, Fuzzy logic, machine learning, data mining, sequence alignment, genetic programming, and so on have advanced in distinguishing different credit card fake transactions. A reasonable seeing on all these methodologies will absolutely lead to an efficient credit card fraud detection framework. This paper suggested an anomaly detection model based on a hybrid simulated annealing (SA) and back-propagation algorithm for feed-forward neural network (FFNN), which joined the significant global searching capability of SA with the precise local searching element of back-propagation FFNNs to improve the initial weights of a neural network toward getting a better result for detection fraud.

Index Terms: Artificial Neural Network, Back-propagation, Back-propagation Feed-forward Neural Network, Feed-forward Neural Network, Simulated Annealing, Simulated Annealing-back-propagation Feed-forward Neural Network

1. INTRODUCTION

The convenience of credit cards is common in modern day community. Credit card utilization has expanded among the clients since credit card installment is key one and it is helpful to pay the amount. It is utilized either online or conventional shopping. Due to the expansion and fast advancement in the fields such as E-commerce, the utilization of credit card is also expanded radically [1]. As the use of credit card is

development, the credit card fraud is additionally increments. The fraud is characterized as a restricted movement by a client for whom the record was not anticipated [2]. The clients who are utilizing the credit card not having the associations with the cardholder and has no goal of making the repayments for the obtain they done. At present, commercial fraud is turning into a serious issue, and successful identification of credit card is a troublesome effort for the experts [3].

Identifying credit card fraud is a tough effort when applying traditional methods; therefore, the growth of the credit card fraud discovery model has matured off significance, either in the educational or trades the society recently.

Credit card fraud detection belongs to the classification and identification problem with a large number of non-linear

Access this article online

DOI: 10.21928/uhdjst.v1n2y2017.pp31-36 E-ISSN: 2521-4217
P-ISSN: 2521-4209

Copyright © 2017 Awlla. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: ardalan.husin@gmail.com

Received: 10-03-2017

Accepted: 25-03-2017

Published: 29-08-2017

situations, which cause it significant to consider non-linear integrated ways to explaining the problem [4].

Artificial neural network (ANN) is a mathematical description of the network of neurons in the mind and share relationships functionalities, such as accepting inputs, processing it, and then produces output [5]. It follows a combined graph of nodes, which are twisted by the weighted links related to the biological neurons. There are different models ANN, for example, feed-forward neural network (FFNN), multiple-layered perceptron, and Kohonen network. Adaptive resonance network and the initial two nets work as a classifier, i.e. these can learn from patterns, and the knowledge can be immediately supervised. Although the other nets learn from attention and later update the network weights, through serve unsupervised learning system seen in a case of clustering. In this paper, a FFNN has been improved for classification intention. FFNN allows the information to pass from the input to output layer in a feed-forward path through the hidden layer(s) [6]. All FFNNs, as stated, possibly trained in a supervised way so that it can learn the feature pattern accessible within the data. To achieve the wanted accuracy in class prediction, fit training is compulsory. While training, the purpose is to catch the network learning feature as the best, which is mirrored by reducing the squared error (i.e., the squared variation between the calculated and the wanted output).

There are various algorithms to optimize such learning method. Backpropagation (BP) is one of the standard traditional ANN training algorithms for supervising learning. The weights are adjusted and updated with a statement delta rule to minimize the prediction error during iterations. The weight improvement methodology covers BP the errors from output layer into hidden layer, so obtaining the optimal set of weights [7].

Simulated annealing (SA) is a probabilistic meta-algorithm for global optimization [8]. It is parallel to the physical method where a solid is casually begin cooled till it is construction is in a cold state, which occurs at a minimum energy form [9]. Similarly, BP algorithm, in SA, the weight has to go into some configuration on the rule till it leads the global minimum [10]. There are besides various other optimization methods such as evolutionary algorithm, for example, genetic algorithm, practical swarm optimization, genetic programming (GP), and so on, there are behind scope of this paper.

The principal purpose of this paper is work to experiment the achievement hybrid of SA and BP compare with BP in the FFNN structure for detection credit card fraud.

2. FEATURE SELECTION

The essential step in developing credit card fraud detection is how to extract the key features. They will influence in recognition rate and improved false alarms. By flattering feature, the data reservation will also be enhanced, so the training and time for data set will be more able for classification that runs under constant environment.

The example dataset that we are running was obtained from a data mining blog. This dataset includes the rundown of the transactions of 20,000 dynamic credit card holders recent months. The input fields incorporate credit card ID, authentication type, current balance, average bank balance, book balance, total number credit card used, and 8 distinctive cardholder classifications such as overdraft, average overdraft, number of location usage, and so on. The data set essentially gives the analysis of the cardholders' exchanges without expressing whether the exchanges were legal or fraudulent. Concerning a given cardholder the dataset based on the following critical values, we can identify which exchange is legal or fraud:

1. Based on credit card usage frequency: Frequency can be found as total number card used/credit cardholder age, if the result <0.2 , it implies this property is not relevant for fraud.
2. Based on a number of location credit card usage: Number of locations credit card used per day so far achieved from the dataset, if location is <5 , it means this property is not relevant for fraud.
3. Based on credit card average overdraft: With respect to card used happened so far considers, the average overdraft can be found as number of overdraft/total number of card used, if overdraft with respect to card used is <0.02 , it means this property is not relevant for fraud.
4. Based on credit card book balance: Regular book balance can be found as current book balance/average book balance, if book balance is equal or <0.25 , it implies *that* this property is not relevant for fraud (Table I).

3. FFNN STRUCTURE

According to chosen features from the dataset, we created different networks. The number of hidden layers of every network is restricted to one for active and manageable calculation. The amount of neurons in the hidden layer is changed to test the results [5]. Fig. 1 illustrates the last proper structure of the network achieved it among them.

TABLE I
Sample of Dataset

Transaction No.	1	2	3	4	5
Credit card ID	11111	11112	11113	11114	11115
Authentication type	111	112	113	114	115
Current balance	20000	25000	15000	100000	15000
Average bank balance	80000	55000	70000	60000	61000
Book balance	0.25	0.4545	0.214	1.6666	0.245
Total number card used	13	40	21	90	85
Overdraft	4	20	3	29	17
Average overdraft	0.3076	0.5	0.142	0.3222	0.2
Number of location usage	3	4	2	11	3
Amount of transaction	9000	15000	8500	12000	19000
Card holder age	25	64	50	21	43
Average daily balance	2666	1833	2333	2000	2033
Card frequency	0.52	0.625	0.16	4.2857	0.18
Card holder marital status	0	1	0	1	1

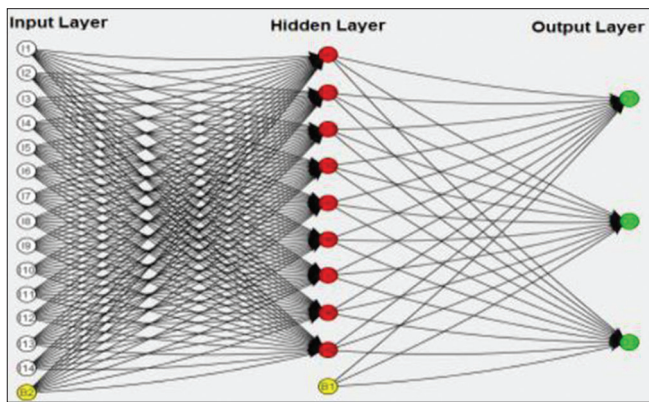


Fig. 1. Description of the FFNN developed

The log sigmoid function in equation 1 is applied as the transfer function connected by the neurons in hidden and output layers to achieve the outputs.

$$f(x) = (1 + e^{-x})^{-1} \tag{1}$$

4. ALGORITHMS

A. Back-propagation Algorithm

BP is a common method for training ANN. The algorithm operates in two forms. First, a training input pattern is given to the input layer, which is forward to the hidden layer then into output layer to produce the network output. Mean square error (MSE) is next calculated by analyzing the estimated output and the target output for all inputs as we explained in equation 3, where “N” indicates the number of entire instances.

$$MS = \frac{1}{N} \sum_{i=0}^N (T0 - E0)^2 \tag{2}$$

In the next step, besides the MSE, the network information back propagates from the output layer to the input layer, and specific connector weights are updated utilizing a “generalized Δ rule” that is held of learning rate (η) and momentum constant (α) [11]. Equations 3 and 4 display the rule of weight updating. In particular equations, the characters “w” means the weights between the connectors “I” and “j” and “t” is the position iteration. An excellent style manual for science writers is [7].

$$\Delta w_{i,j}^t = -\eta \cdot \frac{\partial MSE}{\partial w_{i,j}^t} + \alpha \cdot w_{i,j}^t \tag{3}$$

$$w_{i,j}^{t+1} = \Delta w_{i,j}^t + w_{i,j}^t \tag{4}$$

For achieving the average η which gives the least MSE, rigorous parametric research has been conducted in this research [11]. The η at which the error is minimal is determined for the association with the SA algorithm. In this research, the momentum constant (α) is fixed equal to 0.9 for all states to speed up the learning method. The epoch size is fixed to as 1500.

B. Simulated Annealing Algorithm

The critical parameter for SA is a temperature (T) which is the similarity of the T in physical system. Beginning at a high T, the algorithm ends the minimum T with continuous decrease with attaining of a thermal equilibrium status at each T [8]. At any T, the weights are randomized. A recent set of weights is accepted as the new optimized set if the MSE with this set is under than the prior set or with a possibility that the present set of weights will reach to the global minimum. As estimated in BP, cost function and transfer functions are utilized. This

research is expected if the number of adjustment in the weight set is more than 10 either the number of iterations is more than 1500 then the equilibrium state at a critical T is supposed to be done. The primary T is determined as 10°C , and ultimate T is 1°C randomly. In addition, the T is reduced with by a determinant of 0.95 random because it is challenging to obtain the accurate values of initial, ultimate, and more the threatening of T . The implementation algorithm is as follows ($E(s)$ is an actual function).

SA Algorithm

- 1) *set initial solution in S*
- 2) *Set initial solution T*
- 3) *While not terminate do*
- 4) *Repeat k times*
- 5) *Chose S' a randomly element from $N(s)$*
- 6) $\Delta E = E(S') - E(S)$
- 7) *If ($\Delta E \leq 0$) then*
- 8) $S_{i+1} = S'$
- 9) *else if $S_{i+1} = S'$ with probability $e^{-\Delta E/T}$*
- 10) $S_{i+1} = S'$
- 11) *end if*
- 12) *end repeat*
- 13) *decreased T*
- 14) *end do*

C. Hybrid Algorithms

To defeat the local minimum issue of BP because of initial random weight parameters of the network, various optimization algorithms have been attempted by numerous researchers, which enhance the execution of the classification at the cost of more impalement time. In this paper I, hybridized two algorithms, joining global search SA algorithm, and local search gradient algorithm that defeats the local minimum issue with high speculation and quick union speed.

The hybrid SA-BP is a training algorithm joining the SA algorithm with the BP algorithm. SA is global optimization algorithm, which has a powerful capability to investigate the whole search space. This algorithm has a drawback that the search over the global optimum solution is slow. In opposite, the BP has exact and quick local searching capacity to investigate locally the optimum result, but it gets stuck to discover global optimum result in complex pursuit space.

By joining the SA and the gradient-based BP algorithm, another algorithm alluded to as hybrid SA-BP algorithm as shown in Fig. 2. The suggested hybrid algorithm has two stages: Initial one a global search stage, the FFNN is trained utilizing the SA algorithm for few pre-characterized

temperature or training error is less than some predefined value, then training mechanism changed to the second stage for searching locally utilizing a deterministic technique the BP algorithm. In this paper, it achieved SA-BP hybrid training algorithm as a strong option way to deal with BP algorithm.

Following steps is the pseudo code for the hybrid SA-BP algorithm:

1. Randomly initialize the weights of the FFNN system appeared in Fig. 1
2. Evaluate weights using SA used in the neural network follow a temperature annealing schedule with the algorithm
3. While first temperature value is under or equal to minimum error then select the best solution for MLP then go to step 7
4. Select a moving method with some probability
5. Try a new solution
6. Evaluate
7. Select the best solution
8. Initialize parameters of BP learning algorithm
9. Initialize weights of the MLP utilizing best solution of SA
10. While new epoch is under or equal to maximum epoch or error converges to minimum error do
11. Using BP update weights to minimize error with training data
12. End while
13. Assess execution of classification with test data
14. End while

5. EXPERIMENTAL STUDY

The network packets that are obtained are separated into two sections. The first section about eight hundred records is used to train SA and BP neural network module. The second section is about two hundred records applied to test the credit card fraud detection. The efficiency of the neural network relies on the number, type, and amount of features and learning algorithm applied to train the neural network. Hence, as to evaluate the execution of a credit card fraud recognition strategy; we have to display a quantitative estimate. In our credit card fraud detection system, we mostly classify the network traffic into two categories, which they are normal and abnormal network traffic. Hence, we need to realize the true positive, true negative, false positive, and finally false negative to define true-positive rate (TPR) and false-negative rate (FNR). TPR and FNR can be calculated using the following mathematical equations [5], [6].

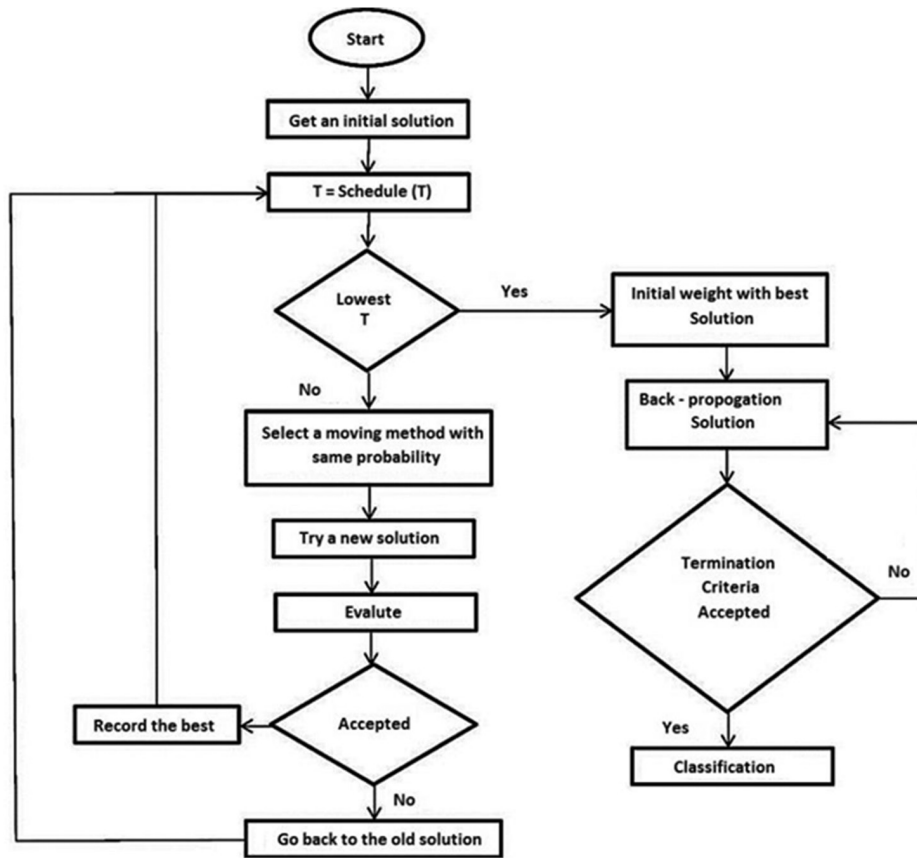


Fig. 2. Structure of hybrid algorithm for classification

$$TPR = TP / (TP + FN) \quad (5)$$

The TPR measures the performance of credit card fraud detection technique concerning the possibility of a suspect data reported correctly as abnormal data. Then, again the FPR measures the performance of credit card fraud detection technique as far as the possibility of a normal traffic reported as abnormal data.

As introduced, the length of parameter temperature has been taken from 10°C to 1°C. The balanced state for it includes of each 10 changes made in set of weights or 1500 iterations, the momentum is 0.9, learning rate is 0.7, maximum error to reaches is 0.01, and weight and threshold values are randomly initialized before training. Consequently, Figs. 2 and 3 show the result of training and test case for the detection rate and FPR of BP and SA-BP (Table II).

Experimental result in Fig. 3. clearly show that SA-BPFFNN more secure in detection credit card farud in comparison to BPFNN. Furthermore, from the Fig. 4. SA-BPFFNN

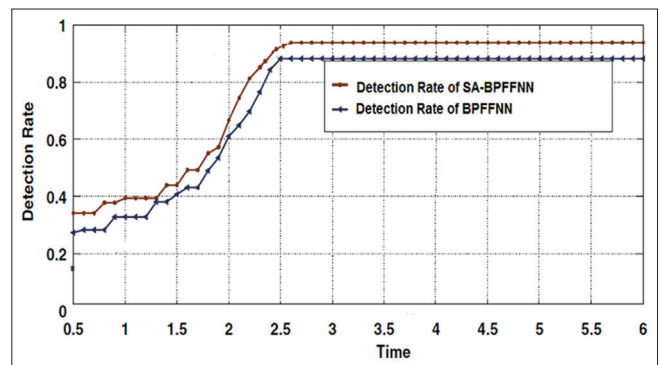


Fig. 3. Detection Rate of SA-BPFFNN and BPFNN

significantly reduce the false-positive rate compare to BPFNN.

6. CONCLUSION

As utilization of credit cards turn out to be increasingly regular in each field of the everyday life, master card or

TABLE II
Summaries the Result During Training and Testing

Desired output SA-BP	Yes	0	1	0	1	0
	No	0	0	1	0	1
	Maybe	1	0	0	0	0
Actual output SA_BP	Yes	0.033	0.9723	0.023	0.9865	0.0146
	No	0.3687	0.0173	0.969	0.0075	0.867
	Maybe	0.5983	0.0104	0.008	0.006	0.1184
Desired output BP	Yes	1	0	1	0	0
	No	0	1	0	1	0
	Maybe	0	0	0	0	1
Actual output BP	Yes	0.0035	0.9148	0.0253	0.9517	0.1277
	No	0.4019	0.073	0.8770	0.0303	0.870
	Maybe	0.5946	0.0122	0.0977	0.0180	0.0023

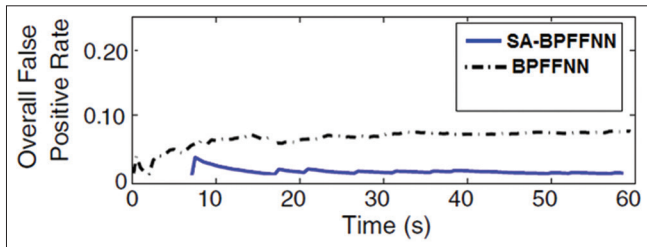


Fig. 4. False positive rate of SA-BPFFNN and BPFFNN

credit card fraud has turned out to be much more rampant. To enhance security of the financial transaction frameworks in an automated and successful way, constructing an accurate and effective credit card fraud detection framework is one of the key efforts for the financial institutions. Credit card fraud detection refers to the classification and recognition issues. The paper hybrids the SA algorithm with BPFNN for fraud detection where the simulated neural network can learn knowledge from a large number of a dataset for training and examining the result of detection. The analysis result illustrated that the using BP is a simple local minimum algorithm, and SA is a good global search algorithm or optimization algorithm based on the analysis, the experimental results indicate that the accuracy of BPFNN is under than applied SA to BPFNN algorithm.

REFERENCES

- [1] N. S. Halvaie and M. K. Akbari. "A novel model for credit card fraud detection using artificial immune systems." *Applied Soft Computing*, vol. 24, pp. 40-49, Nov. 2014.
- [2] C. Yin, A. H. Awlla, Z. Yin and J. Wang. "Botnet detection based on genetic neural network." *International Journal of Security and Its Applications*, vol. 9, pp. 97-104, Nov. 2015.
- [3] V. Van Vlasselaer, C. Bravo, O. Caelen and B. Baesens. "A novel approach for automated credit card transaction fraud detection using network-based extensions." *Decision Support Systems*, vol. 75, pp. 38-48, Jul. 2015.
- [4] D. Sanchez, M. A. Vila, L. Cerda and J. M. Serrano. "Association rules applied to credit card fraud detection." *Expert Systems with Applications*, vol. 36, pp. 3630-3640, 2009.
- [5] S. Suganya and N. Kamalraj. "A survey on credit card fraud detection." *International Journal of Computer Science and Mobile Computing*, vol. 4, pp. 241-244, Nov. 2015.
- [6] J. Bernal and J. Torres-Jimenez. "SAGRAD: A program for neural network training with simulated annealing and the conjugate gradient method." *Journal of Research of the National Institute of Standards and Technology*, vol. 120, pp. 113-128, 2015.
- [7] S. J. Subavathi and T. Kathirvalavakumar, "Adaptive modified backpropagation algorithm based on differential errors." *International Journal of Computer Science, Engineering and Applications*, vol. 1, no. 5, pp. 21-33, Oct. 2011.
- [8] A. T. Kalai. "Simulated annealing for convex optimization." *Mathematics of Operations Research*, vol. 31, pp. 253-266, 2006.
- [9] C. M. Tan, Ed. *Simulated Annealing*. Vienna, Austria: In-Teh is Croatian Branch of I-Tech Education and Publishing KG, Sep. 2008.
- [10] S. H. Zhan, J. Lin, Z. J. Zhang and Y. W. Zhong. "List-based simulated annealing algorithm for traveling salesman problem." *Computational Intelligence and Neuroscience*, vol. 2016, pp. 12, Mar. 2016.
- [11] N. A. Hamid, N. M. Nawi, R. Ghazali and M. N. M. Salleh. "Solving local minima problem in back propagation algorithm using adaptive gain, adaptive momentum and adaptive learning rate on classification problems," *International Conference Mathematical and Computational Biology*. Malacca, Malaysia, pp. 448-455, Apr. 2011.