

Research Article

Using Harmony Search Algorithm in Neural Networks to Improve Fraud Detection in Banking System

Sajjad Daliri 

Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

Correspondence should be addressed to Sajjad Daliri; sajjad.daliri@srbiau.ac.ir

Received 11 November 2019; Accepted 4 January 2020; Published 8 February 2020

Guest Editor: Octavio Loyola-González

Copyright © 2020 Sajjad Daliri. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Financial fraud is among the main problems undermining the confidence of customers in addition to incurring economic losses to banks and financial institutions. In recent years, along with the proliferation of fraud, financial institutions began looking for ways to find a suitable solution in the fight against fraud. Given the advanced and varied changes in methods of fraud, extensive research has been conducted to detect fraud. In this paper, the Artificial Neural Network technique and Harmony Search Algorithm are used to detect fraud. In the proposed method, hidden patterns between normal and fraudulent customers' information are searched. Given that fraudulent behavior could be detected and stopped before they take place, the results of the proposed system show that it has an acceptable capability in fraud detection.

1. Introduction

Fraud in the financial system is known as abuse of the system to improve profitability of a person or an organization. In a competition-based environment, the occurrence of fraud can cause a critical problem in business. This issue has become critically serious in recent years [1]. Recently, fraud has been dramatically increasing incurring billion-dollar annual losses to the owners of financial institutions and banks. On the other hand, the advancement of technology in various fields has caused generation of data in high volumes. The volume of data is positively and directly related to the complication of interrelationships [2]. With regard to the issue, data mining is used as exploratory data analysis with the aid of other sciences in which exploration of hidden and unknown information out of the bulk of data is under focus. Access to hidden information contained in big data is more difficult and more complicated. The science of data mining with other methods such as machine learning, databases, and artificial intelligence has spread to detect patterns among such data [3, 4].

2. Review of the Related Literature

Since the advent of business activities, some individuals aimed to increase their own profits and incur losses to companies and products of others. In the past, no considerable financial exchanges occurred; however, gradually, the increase in the population and closer communications amplified financial exchanges, and at the same time fraud has also increased [5]. Nevertheless, today, given the new technologies, costs related to fraud have fallen dramatically, yet fraud methods have also progressed. Recently, the issues of telecommunications, e-commerce, and provision of new services have led to the occurrence of fraud in new ways with different dimensions. With the increasing use of the Internet, new types of fraud have appeared, but with the rise of security systems to prevent fraud, there has not been significant progress due to lack of appropriate patterns [6].

Frauds such as technical fraud that exploit weaknesses in the system mostly occur when a new system is introduced, and developers are unaware of those weaknesses. Fraud often occurs when there is unauthorized access to

customer accounts. A subscription fraud occurs when users subscribe to a system without any intention to use it. In this case, other people may use their unused subscriptions [7]. However, much of what happens as fraudulence is the use of social engineering. In this case, swindlers use their skills to obtain detailed information about the system and use it instead of trying to discover the weaknesses of a financial system [8].

In a research study [9], a two-layer system was proposed. In the first layer, general rules and specific rules of each customer are located. The degree of suspiciousness of a transaction can be signified in this layer. In the second layer, techniques of Game theory are used to detect fraud. In another research study [10], a neural network was used to detect fraud. For modeling the neural network, a nonlinear discriminant analysis was used. Moreover, to reduce the volume of calculations, a scoring system was used for transactions. The results revealed that the scoring system had provided better results and higher performance compared with the method in which all transactions are evaluated. However, in another research [11], researchers concluded that, when the financial ratio is used as data, using a neural network has premier performance than other methods. In this research, the dataset of 54 firms, each with 20 attributes were used.

In a study [12], Bank Sealer was proposed. Bank Sealer is a decision support system for online banking fraud analysis and investigation. In the first step, the system quantifies the anomaly of a transaction based on customer historical data. In the second step, it finds the cluster wherein users with similar behavior are included, and finally with the help of information obtained, anomalous or normal transaction is proved.

There are methods grounded on the genetic algorithm (GA) such as [9], where a combination of a GA and a Support Vector Machine (SVM) is presented to detect anomalies in data. In the proposed method, a GA has been used to select the best subset of attributes showing anomalies in the system. Then, the dataset has been applied to SVM for training. However, there are methods based on data mining techniques such as [13], which use a distributed approach for fraud detection. In the study, a large dataset of labeled data with algorithms such as C4.5, ID3, Ripper, and CART has been used.

3. The Proposed Method

In the present research study, a hybrid system based on the Artificial Neural Network (ANN) technique and Harmony Search algorithm (HAA) is used to detect fraud. The system uses HSA for optimizing the parameters of ANN, while ANN itself is used for fraud detection.

The proposed system is shown as a flowchart in Figure 1. The figure illustrates how ANN is integrated into HSA. Obviously, to calculate the proportion of the solutions, HAS requires building a corresponding ANN. Finally, after building and training of the ANN neural network corresponding to each solution, the performance of ANN is

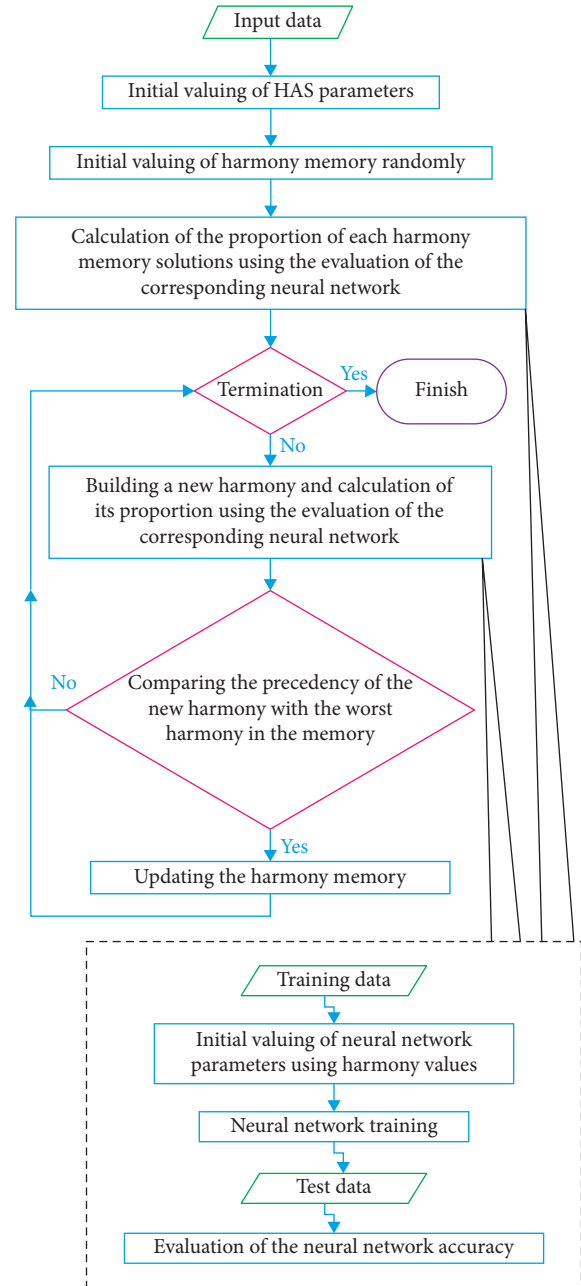


FIGURE 1: Flowchart of NNHS.

recorded as the proportion of HSA solutions. The method proposed in this study is termed NNHS.

In the neural network used in the present research, back propagation is used for learning, so that the network learns the patterns between inputs and outputs. In this way, for each case in the training data, the desired output is prerecorded.

This system continues, as long as the difference between the network outputs and the data outputs is as low as possible. In this network, weights are randomly selected, and then based on how much difference exists between the output from the network and the output recorded in the dataset, weights will be adjusted at each stage. As expressed, HAS is used to determine the parameters of ANN. The

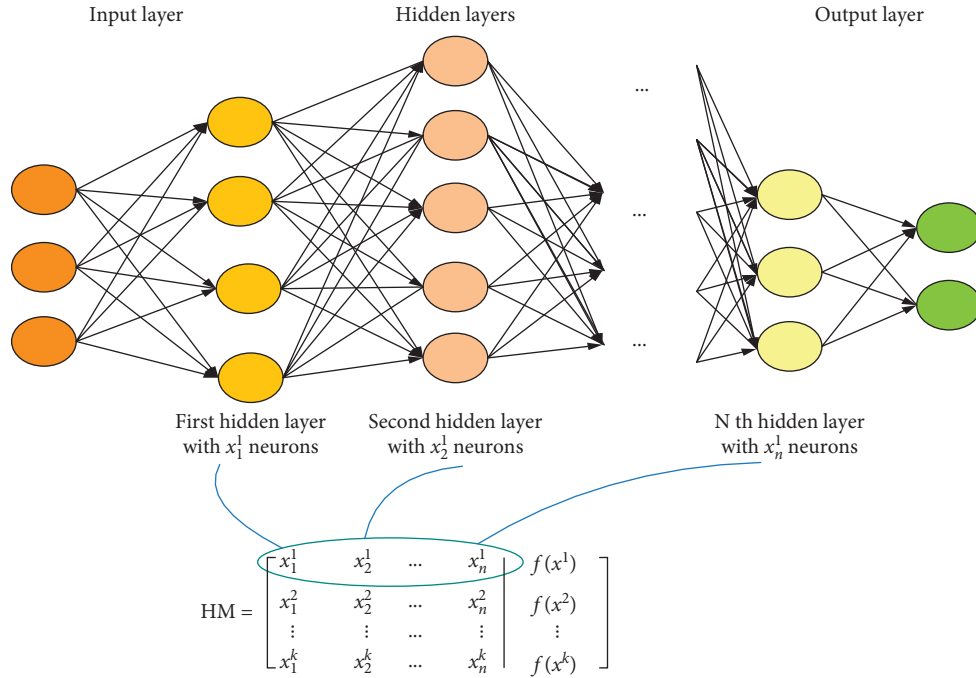


FIGURE 2: Determining the parameters of the neural network using HSA.

structure of the neural network using the harmony memory can be seen in Figure 2.

The HSA task is to find the most suitable structure for the neural network. At the beginning of the execution of the algorithm, parameters such as the size of a harmony memory, the rate of consideration of a harmony memory, the adjustment rate of pitch, and other values are set. The next step is to create the first-generation algorithm randomly. After this step (after each iteration), new harmonies are generated, and accordingly via building, the corresponding neural network is calculated and accuracy is determined. The next step is to update the memory where if the proportion of the New Harmony is lower than the worst proportion recorded in the system, the new one replaces the worst and thus the harmony memory will be updated. In the end, the algorithm stops after performing the specified number of iterations, which is the condition for the termination of the proposed algorithm, and the best harmonies recorded are extracted as the most appropriate solution.

3.1. Research Data. The dataset of the present research study is the German Dataset used to evaluate and test the proposed system. This dataset is available at the UCI website and is used in many studies [14]. This dataset helps provide a better comparison with other methods. It contains 100 records of which 300 cases are abnormal and 700 cases are normal.

3.2. Test Data. To improve the testing accurateness, accuracy was used. With an initial population of 150, HAS was executed over 100 iterations. To obtain more reliable results than the proposed system, the test was executed over 10 iterations with the same conditions, and the results were

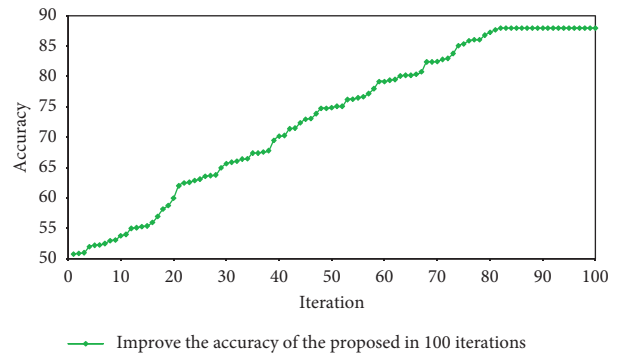


FIGURE 3: The accuracy of NNHS in different generations.

averaged. In Figure 3, accuracy is displayed in 100 generations.

As seen in the Figure 3, after 81 generations, accuracy reaches 88%. In this study, as mentioned earlier, after 10 iterations, the test result averages 86% of accuracy. In the second phase of the test, the scattering matrix (SM) was used. The SM calculation method is given in Table 1, and the numerical comparison progress is shown in Figure 4.

4. Comparison of the Results

First, NNHS is compared with other types of methods. Hayashi et al. evaluated the criteria of accuracy. They used of the recursive-rule extraction algorithm to detect fraud. In Table 2, NNHS is compared with the method of Hayashi et al. [15].

In another study conducted by Hassan et al., the recursive value was evaluated. The results of this method are also compared with NNHS in Table 3 [16].

TABLE 1: The SM calculation process.

TN	FN	Negative predictive value = $TN / (TN + FN)$ False omission rate = $FN / (TN + FN)$
FP	TP	Precision = $TP / (TP + FP)$ False discovery rate = $FP / (TP + FP)$
True negative rate = $TN / (TN + FP)$ False positive rate = $FP / (TN + FP)$	True positive rate = $TP / (TP + FN)$ False negative rate = $FN / (TP + FN)$	Accuracy = $(TP + TN) / \text{total}$ Misclassification rate = $(FP + FN) / (TN + FN + TP + FP)$

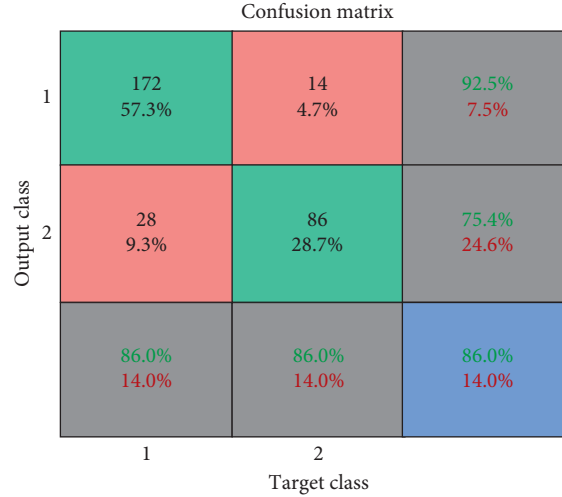


FIGURE 4: The correlation matrix of our proposed method, NNHS.

TABLE 2: Comparison of results of [15].

Method	Training data	Test data	Training accuracy	Test accuracy
Hayashi et al. [15]	700	300	81.32 ± 1.02	80.53 ± 0.88
NNHS	700	300	88	86

TABLE 3: Comparison of results of NNHS and Hassan et al. [16].

Method	Training data	Test data	Recall
Hassan et al. [16]	700	300	76.8
NNHS	700	300	87

Considering other methods, one of the main advantages of NNHS is the improvement of the parameters of ANN using HSA. However, according to the structure of the proposed network, the method is troubled against data with unbalanced information, and detection of the hidden pattern in this type of data has been exposed to many problems.

5. Conclusion

In the present study, a fraud-detection model termed NNHS was proposed through the integration of ANN and HSA. The proposed system offers a solution based on HAS succeeding in predicting the best structure for ANN and detecting the hidden algorithm in the mass of data. The results of the comparisons have shown that the best accuracy obtained from the German dataset for the proposed system is 86. In addition, the best value obtained for the same recall criteria is 87. However, the values obtained in [15] and [16] were,

respectively, 81.53 and 76.8. Therefore, the results of the evaluation show that having a relatively high performance, NNHS has been able to detect dishonest customers with comparatively high accuracy.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] E. Caldeira, B. Gabriel, and A. Pereira, "Fraud analysis and prevention in e-commerce transactions," in *Proceedings of the 2014 9th Latin American Web Congress*, pp. 42–49, Ouro Preto, Brazil, October 2014.

- [2] A. Sharma and P. Kumar Panigrahi, "A review of financial accounting fraud detection based on data mining techniques," *International Journal of Computer Applications*, vol. 39, no. 1, pp. 37–47, 2013.
- [3] H. L. Sithic and T. Balasubramanian, "Survey of insurance fraud detection using data mining techniques," *International Journal of Innovative Technology and Exploring Engineering*, vol. 2, no. 3, pp. 989–994, 2013.
- [4] P. Mishra, N. Padhy, and R. Panigrahi, "The survey of data mining applications and feature scope," *Asian Journal of Computer Science & Information Technology*, vol. 2, no. 4, pp. 68–77, 2013.
- [5] N. Jain and V. Srivastava, "Data mining techniques: a survey paper," *IJRET: International Journal of Research in Engineering and Technology*, vol. 2, no. 11, pp. 116–119, 2013.
- [6] J. A. Parejo, A. Ruiz-Cortés, S. Lozano, and P. Fernandez, "Metaheuristic optimization frameworks: a survey and benchmarking," *Soft Computing*, vol. 16, no. 3.3, pp. 527–561, 2012.
- [7] D. Manjarres, "A survey on applications of the harmony search algorithm," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 8, pp. 1818–1831, 2013.
- [8] M. T. Hagan, *Neural Network Design*, PWS Publishing Company, Boston, MA, USA, 1996.
- [9] V. Vatsa, S. Sural, and A. K. Majumdar, *A Game-Theoretic Approach to Credit Card Fraud Detection*, vol. 3803, pp. 263–276, Springer, Berlin, Germany, 2005.
- [10] J. R. Dorronsoro, F. Ginel, C. Sgnchez, and C. S. Cruz, "Neural fraud detection in credit card operations," *IEEE Transactions on Neural Networks*, vol. 8, no. 4, pp. 827–834, 1997.
- [11] K. M. Fanning, K. O. Cogger, and K. O. Cogger, "Neural network detection of management fraud using published financial data," *International Journal of Intelligent Systems in Accounting, Finance & Management*, vol. 7, no. 1, pp. 21–41, 1998.
- [12] M. Carminati, R. Caron, F. Maggi, I. Epifani, and S. Zanero, "BankSealer: a decision support system for online banking fraud analysis and investigation," *Computers & Security*, vol. 53, pp. 175–186, 2015.
- [13] P. K. Chan, W. Fan, A. L. Prodromidis, and S. J. Stolfo, "Distributed data mining in credit card fraud detection," *IEEE Intelligent Systems*, vol. 14, no. 6, pp. 67–74, 1999.
- [14] UCI Machine Learning Repository (Center of Machine Learning and Intelligent Systems), <https://archive.ics.uci.edu/ml/datasets.html>.
- [15] Y. Hayashi, S. Nakano, and S. Fujisawa, "Use of the recursive-rule extraction algorithm with continuous attributes to improve diagnostic accuracy in thyroid disease," *Informatics in Medicine Unlocked*, vol. 1, pp. 1–8, 2015.
- [16] R. Hassan, S. M. Arafat, and K. Begg, "Fuzzy-genetic model of the identification of falls risk gait," in *Proceedings of the Symposium on Data Mining Applications, SDMA2016*, Riyadh, Saudi Arabia, March 2016.