Caliphate Journal of Science & Technology (CaJoST)

ISSN: 2705-313X (PRINT); 2705-3121 (ONLINE)

Research Article

Article Info

Received: 14th May 2020

Revised: 25th June 2020

Accepted: 29th June 2020

[1]Department of Mathematics, Sokoto State University Sokoto, P.M.B 2134 Sokoto State
[2]Department of Mathematics, Isa Kaita College of Education, P.M.B 5007, Dutsin-Ma, Katsina State
[3]Department of statistics, Abubakar Tatari Ali Polytechnic, Bauchi, Bauchi State

Corresponding author's email:
sas0010@yahoo.com

Cite this: *CaJoST,* 2020, 2, 142-147

Data Mining via 3-Satisfiability Reverse Analysis Method with Hopfield Neural Network for Infertility Detection

Samaila Abdullahi,[1] Nuraddeen Y. Adamu[2] and Sani Musa[3]

Medical diagnosis is quite complicated since every patient sign and symptoms will be screened and when the number of features increases the medical practitioner will not able to be screened appropriately. 3-Satisfiability Reverse Analysis Method (3-SATRA) incorporated with Hopfield neural network is a new approach for the early screening of infertility among humans. 3-SATRA were proposed to extract the best logic rule that will represent the features of a data set since the conventional data extraction techniques focus only on standalone neural network. The 3-Satisfiability Reverse Analysis Method was integrated with logic programming as a data mining instrument. The proposed method is applied to infertility dataset obtained from machine learning repository center restate proposed model. thus, the results of the analysis will promote the early screening stage used by medical practitioners. The simulation was executed using Dev C++ 5.11 as a tool for training, testing and validating the performances of the proposed method. The performance of the method was measured based on Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Sum of Squared Error (SSE) and Computational Time. The performance and accuracy of the results obtained have shown the effectiveness of 3-SATRA in data mining.

Keywords: 3-staisfiability problem; discrete Hopfield neural network; infertility data set; logic programing and data mining

1. Introduction

In humans, infertility is the inability to become pregnant or carry a pregnancy to full term, there are many causes of infertility some of which are hormonal, structural, diminished ovarian reverse. Demographers tend to define infertility as childlessness in a population of women of reproductive age [1]. The medical specialist is unable to screen for infertility properly when the number of instances that attributed to the problem increased. The best and efficient way to deal with uncertainty is to apply data mining techniques incorporated with Hopfield neural network and data mining which are complemented to each other and has the capacity to identify the relationship and compete with the system. To achieve this, 3-STRA model was formulated to serve as a data mining tool. Basically, 3-Satisfiability (3-SAT) problem is a mapping problem from a logic programming in three conjunctive normal form (CNF) to truth values. In theory, 3-SAT can be defined as a formula in a conjunctive normal form with a collection of clauses where each comprises strictly 3 literals per clause [2]. SAT involves definitive task of searching truth assignments that satisfies a Boolean formula. the formula can be in any combination as the number of atoms can be varied except for the literals that are strictly equal to three per clause [3]. Conceptual nowadays recurrent neural network is inspired by human biological brain to mock the computations employed by the human nervous system [3]. The Hopfield neural network (HNN) consists of bipolar threshold unit which is full superimposed in dual direction and the connections are symmetric which cause the network to settle to minimum energy. Furthermore, the Hopfield neural network composed of exceptional features such as better stability, recurrent for faster execution and efficient content addressable memory. The discrete Hopfield neural network computation is performed by collection of activated neurons [4]. Moving on, Sathasivam introduced a powerful Hopfield model based on reverse analysis method in solving Horn-satisfiability problem [5]. Logic programming was justifying consistently in mapping as well as representing the attribute of data set into the symbolic logic form [6]. Thus, the major issues with logic mining is the generating logical rule which will be embedded in the Hopfield network or any machined learning models. Therefore, 3-satisfiability reversed analysis as a logical rule that has been used to generate the best logic rule and pattern that represent the attribute of the real data set [3]. Furthermore, the robust nature of data mining technique can be hybridized with 3-SAT logic mining. However, the new reform reverses analysis method for extracting valuable information among the attribute of a real data set by using the conjunctive normal form (CNF) logical rule.

2. 3-Satisfiability Problem

Theoretically, satisfiability (SAT) can be defined as the task of searching a truth assignment that creates an arbitrary Boolean expression true [7]. Basically, the satisfiability problem can be treated as the combinatorial optimization problem according to logic programming perspective. In other words, satisfiability problem refers to decision making problems based on constraints. Consequently, 3-SAT can be defined as conjunctive normal form formula with an array of clauses connected with logic operator where each comprises strictly 3 literals per clause [8]. three satisfiability (3-SAT) will defined as a mapping problem to assign a Boolean logic in conjunctive normal form (CNF) to truth values. The 3-SAT can be summarized as follows:

1. The 3-SAT IN Boolean logic form consist a set of n variable $R_1, R_2, ..., R_n$ for each clause, $n=3$.
2. The set of literals can be either positive or negative and the set of d clause in Boolean logic formula W. $\exists W : W = C_1 \wedge C_2 \wedge, ..., C_d$.
3. The clause is combined with operator **OR** ($\vee$) and connected with operator **AND** ($\wedge$).

The recommended formula for 3-satisfiability is given by:

$$R_{3-SAT} = \bigwedge_{i=1}^{n} c_i \tag{1}$$

where c_i, $i = 1, 2, ...n$ is given by:

$$c_i = \bigvee_{i=1}^{n} c_i(p_i, q_i, r_i) \tag{2}$$

The 3-satisfiability formula can be denoted as:

$$R = (P \vee Q \vee R) \wedge (S \vee T \vee U) \wedge (V \vee W \vee X) \tag{3}$$

The equation (3) prescribed the 3-Satisfiability problem in conjunction normal forma (CNF) and it can be formulated in various combination when the number of literals is assigned at random and simplify the output of the given problem.

3. Discrete Hopefield Neural Network

The data mining techniques can be model by integrating logic mining and the Hopfield neural network. Theoretically, Hopfield Neural Network is a fully interconnected neural network, in such a way that each neuron is connected to every other neuron. The architecture of HNN requires a set of neurons.

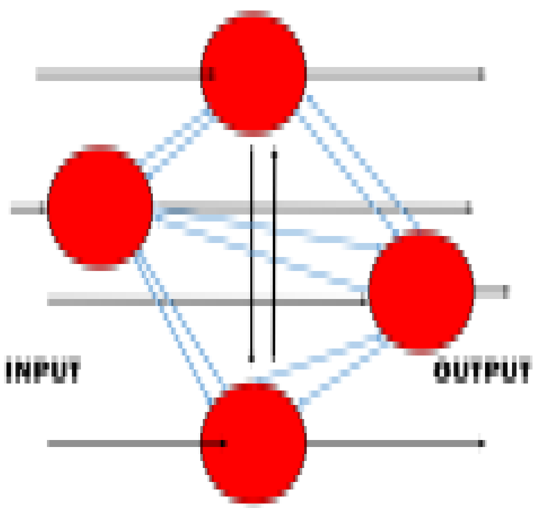

Figure 1. Simple Topology of Hopfield Neural Network

The HNN consists of good remarkable properties which include parallel execution, faster computation, better stability, and higher memory capacity [8]. The neurons in discrete HNN are termed bipolar which strictly considers values of 1 and -1. The principal description for neuron state activation is shown:

$$s_i = \begin{cases} 1, if \sum_j W_{ij} S_j > \xi_i \\ -1, otherwise \end{cases} \tag{4}$$

where W_{ij} is the connection strength from unit j to i. S_j is the state of unit j. Next, ξ_i is the threshold of unit i. The network's architecture comprises of N recognized neurons, each was described by an Ising spin variable model [9]. The connection in Hopfield neural network is symmetric (i.e. no connection within itself) and zero-diagonal $w_{ii} = w_{jj} = 0$. This gave Hopfield the symmetrical features in terms of architecture. This causes each neuron to flip until the equilibrium is reached. Thus, it follows the dynamics $s_i \to \operatorname{sgn}\left[h_i(t)\right]$ where h_i is the local field of the neuron's connection. When dealing with higher order neurons connection, the sum of the field induced by each neuron is given in Equation (3.2) formulated by [10].

$$h_i = \sum_j \sum_k W_{ijk} S_j S_k \sum_j W_{ij} S_j + W_i \tag{5}$$

Therefore, since the synaptic weight in Hopfield neural network is maintained symmetrical, the updating rule maintains as follows:

$$S_i\left(t+1\right) = \operatorname{sgn}\left[h_i\left(t\right)\right] \tag{6}$$

whereby sgn is a sign function of the local field or state of the neuron. According to Equation (6) the updating rule decreases monotonically with the dynamics. In the original HNN model, the state of the system evolves from any initial state to a final state where it is called local minimum of the Lyapunov function [11]

$$E_{Lyapunov} = -\frac{1}{2}\sum_{i=0}^{n}\sum_{j=0}^{n} W_{ij} S_i S_j - \sum_{j=0}^{n} W_i S_j \tag{7}$$

whereby W_{ij} is the synaptic weight, S_i and S_j refer to the neuron state. Thus, the Equation (7) is then recrafted into generalized Lyapunov energy that accommodates neurons with higher complexity as shown in Equation (8):

$$E_{Lyapunov} = -\frac{1}{3}\sum_{i=0,}^{n}\sum_{j\neq 0}^{n}\sum_{k\neq 0}^{n} W_{ij} S_i S_j S_k - \frac{1}{2}\sum_{i=0}^{n}\sum_{j\neq 0}^{n} W_{ij} S_i S_j - \sum_{i=0}^{n} W_i S_i \tag{8}$$

This energy function is significant because it establishes the degree of convergence of the network. The energy value obtained from the equation will be verified as global or local minimum energy. For our case, the network will produce the correct solution when the induced neurons state reached global minimum energy. The energy function is vital as the refine mechanism for determined global minima solution or local minima solution.

4. Implementation

The 3-satisfiability reverse analysis (3-SATRA) method will be integrated with Hopfield neural network and 3-SAT for doing data mining. Thus, the execution of the proposed method will be validated during the training and testing of infertility data set. The simulations of the designed method will be carried out infertility dataset obtained from UCI machine learning repository. restate as follows:

Step 1:

Select the data set to be trained and tested by the proposed method. In this paper, we consider 60% for the training and 40% portion for testing.

Step 2:

Convert attribute of the date set into bipolar and values to the neuron state that will generate the best logic, P_{best}.

Step 3:

Derive a cost function of 3-SAT logic, $E_{3-SAT.}$ by considering $P = \frac{1}{2}(1+S_P)$ and $\neg P = \frac{1}{2}(1-S_P)$. Hence, the neuron state can be verified as $S_P = 1$ or $S_P = -1$.

Step 4:

Check the clauses satisfaction of $E_{3-SAT.}$ the model will be trained by exhaustive search (ES) method and best pattern that represent the data set will be store in CAM of HNN.

Step 5:

Compute the synaptic weight of P_{best} using [7] standard weight management techniques in HNN by comparing the energy function and cost function.

Step 6:

Compute the local field, h_i and apply hyperbolic tangent activation function to determine the final state neurons.

Step 7:

Generate the induced logic, $P_{induced}$ to compare the output of the network and the actual target output from the data set which will verified as Success or Failure.

Step 8:

Compute the corresponding performance evaluation metrics namely, RMSE, MAE, SSE, ACCURACY and Computational time (CPU Time) of the model.

5. Results and Discussion

The platform used in simulating the medical data set is Dev C++ 5.11 for training and testing the proposed method in order to validate the models.

Table 1. Experimental setup

Model P_{best}	
Operating System	Windows 10
Processor	AMD RYZEN 3, 2.5 Ghz
Memory	4GB RAM
Termination of Criterion	0.001
Number of Trial	100
NN	1-8

The best logic for 3-SATRA that represent the attribute of the infertility data set is generated as:

$$R=(\neg P\vee\neg Q\vee R)\wedge(S\vee T\vee U)\wedge(V\vee\neg W\vee\neg X) \quad (9)$$

The induced logic for 3-SATRA that represent the attribute of the infertility data set that will test the outcome of the target values is generated as:

$$R=(P\vee\neg Q\vee R)\wedge(\neg S\vee T\vee U)\wedge(V\vee W\vee\neg X) \quad (10)$$

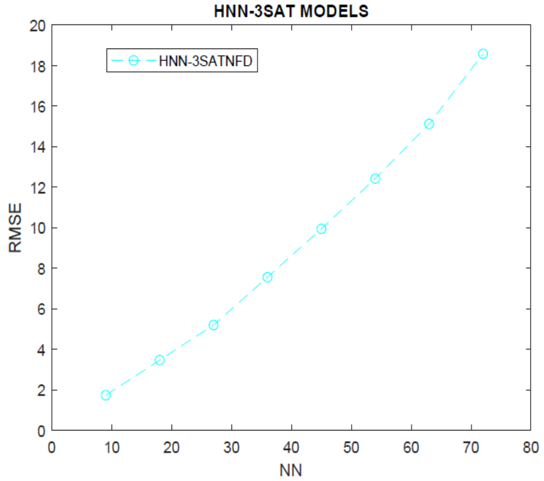

Figure 2. RMSE for HNN-3SATNFD

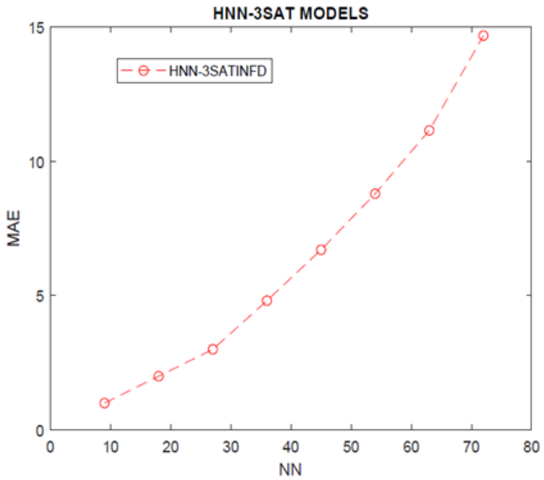

Figure 3. MAE for HNN-3-SATNFD

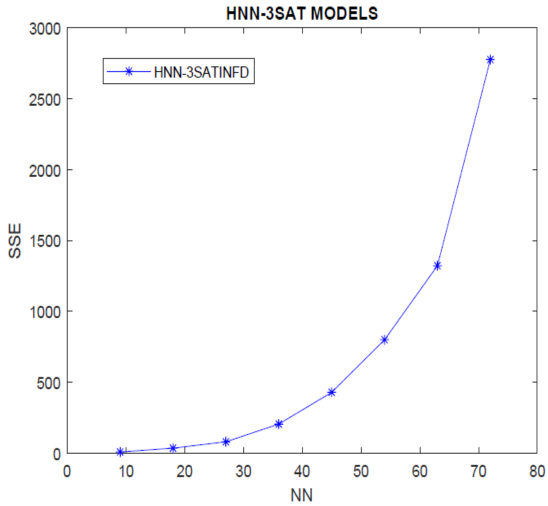

Figure 4. SSE for HNN-3SATNFD

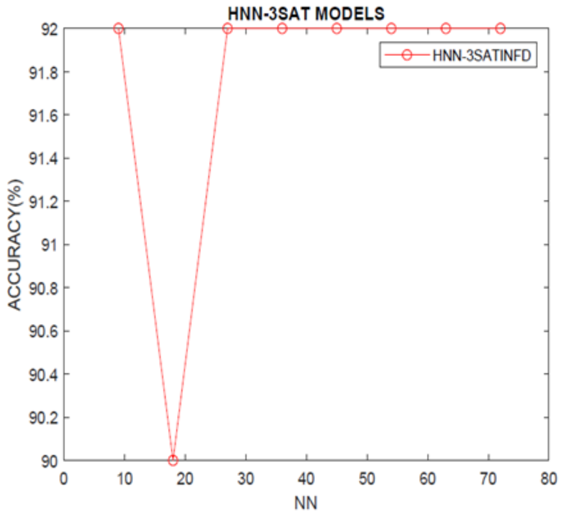

Figure 5. Accuracy for HNN-3-SATNFD

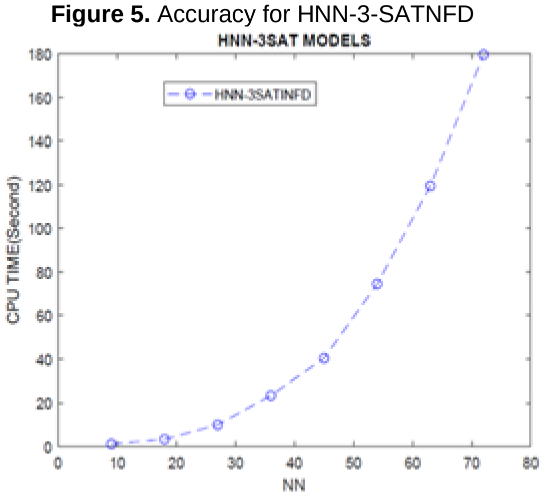

Figure 6. Accuracy for HNN-3-SATNFD

The results revealed the performances of HNN-3SATES based on RMSE, MAE, SSE and CPU Time calibration measures in training infertility data set. In this research, 60% of the data is allocated for training data set and the remaining 40% for testing data. The main task is to extract logical rule for infertility data set. Basically, the proposed network will assist the medical practitioners to identify the presence or absence of the disease during diagnosis. The simulations are carried out until $NC = 8$ for simplicity to check the complexity. The HNN-3SAT model with 3-SATRA has successfully extracted the best logical rule of data set. Hence, the comparison of the RMSE, MAE, SSE, Accuracy and CPU Time for HNN-3SATNFD are as follows. Figure 2: depicts the RMSE values obtained by HNN-3SATNFD for different complexities. Hence, the proposed method was performed consistently better in training the infertility data set. However, the MAE values for HNN-3SATNFD is shown in Figure 3: there is positive correlation since the MAE values are always positive. The lower the MAE values indicate the solution is quite near to the optimal solution. In Figure 4: the SSE recorded for HNN-3SATNFD with different array of number of clause (NC) parameters is demonstrated. The error accumulation can be described through the SSE value as shown in figure 4. Therefore, the magnitude of the SSE will increase as the number of clause increase. The robustness of our 3-SATRA algorithms can be approximately demonstrated base on the effectiveness of the entire computational process. As the number of neurons increases, the CPU Time also increase for all the hybrid networks. Hence, overall CPU Time when $NC = 8$ the proposed paradigms was generally below 150 second as indicated in figure 6. The ACCURACY recorded for HNN-3SATNFD were 92% as shown in figure 5. The HNN-3SAT with 3-SATRA approach to the infertility data set is proven the performance of the proposed model and it can be implemented by medical practitioners as an early diagnosis tool.

6. Conclusion

The proposed algorithm, HNN-3SATES has demonstrated a good performance based on RMSE, MAE, SSE and CPU time obtained during training and testing processed. The hybrid HNN model with 3-SATRA is proven as a robust paradigm according to the experimental results. This again recommended the prospective implementation of the proposed method in developing a faster classification paradigm that will aid the medical practitioners in their decision-making process.

Conflict of interest

The authors declare no conflict of interest.

References

[1] M. Mustafa, J. Hadi, E. M. Iiizam, and S. Aliya, “Male and Female Infertility: Causes , And Management Male and Female Infertility : Causes , And Management,” no. October, 2019.

[2] K. Kutzkov, “New upper bound for the #3-SAT problem,” *Inf. Process. Lett.*, vol. 105, no. 1, pp. 1–5, 2007.

[3] M. A. Mansor, “Accelerating Activation Function for 3- Satisfiability Logic Programming,” no. October, pp. 44–50, 2016.

[4] M. Shareduwan and S. Sathasivam, “Accelerating activation function in higher order logic programming Accelerating Activation Function in Higher Order Logic Programming,” vol. 030006, 2016.

[5] S. Sathasivam, “Upgrading Logic Programming in Hopfield Network,” *Sains Malaysiana*, vol. 39, no. 1, pp. 115–118, 2010.

[6] M. A. Mansor, S. Sathasivam, and M. Shareduwan, “3-Satisfiability Logic Programming Approach for Cardiovascular Diseases Diagnosis,” vol. 020022, 2018.

[7] W. Ahmad Tajuddin Wan Abdullah, “The logic of neural networks,” *Phys. Lett. A*, vol. 176, no. 3–4, pp. 202–206, 1993.

[8] S. Sathasivam, “APPLICATION OF NEURAL NETWORKS IN PREDICTIVE,” pp. 371–376, 2011.

[9] M. A. Mansor, M. S. M. Kasihmuddin, and S. Sathasivam, “VLSI Circuit configuration using satisfiability logic in hopfield network,” *Int. J. Intell. Syst. Appl.*, vol. 8, no. 9, pp. 22–29, 2016.

[10] S. Sathasivam and N. P. Fen, “Developing agent based modelling for doing logic programming in hopfield network,” *Appl. Math. Sci.*, vol. 7, no. 1–4, pp. 23–35, 2013.

[11] S. Sathasivam, “Logic Mining in Neural Networks,” 2006.

