

# Artificial Neural Network Weight Optimization: A Review

Abdirashid Salad Nur<sup>\*1</sup>, Nor Haizan Mohd Radzi<sup>1</sup>, Ashraf Osman Ibrahim<sup>1,2</sup>

<sup>1</sup>Soft Computing Research Group, Faculty of Computing, Universiti Teknologi Malaysia (UTM),  
81310 Skudai, Johor, Malaysia

<sup>2</sup>Faculty of Computer Science and Information Technology, Alzaeim Alazhari University (AAU),  
Khartoum, Sudan

\*Corresponding author, e-mail: salaadnuur@yahoo.com

## Abstract

*Optimizing the weights of Artificial Neural Networks (ANNs) is a great important of a complex task in the research of machine learning due to dependence of its performance to the success of learning process and the training method. This paper reviews the implementation of meta-heuristic algorithms in ANNs' weight optimization by studying their advantages and disadvantages giving consideration to some meta-heuristic members such as Genetic algorithm, Particle Swarm Optimization and recently introduced meta-heuristic algorithm called Harmony Search Algorithm (HSA). Also, the application of local search based algorithms to optimize the ANNs weights and their benefits as well as their limitations are briefly elaborated. Finally, a comparison between local search methods and global optimization methods is carried out to speculate the trends in the progresses of ANNs' weight optimization in the current resresearch.*

**Keywords:** artificial neural networks, local search, global optimization, meta-heuristics, weight optimization

**Copyright © 2014 Institute of Advanced Engineering and Science. All rights reserved.**

## 1. Introduction

Artificial Neural Network (ANN) is a computational model that imitates the human brain system. The most recognised characteristic of ANN is the ability to learn from observing the data sets and improves its performance through learning. ANN also capable to self-organize, self adaptive and conducted real time learnig. These characteristic have been exploited for many years by the researchers to solve classification problems in pattern recognition and attracted the data mining research community [1-2], [6, 10]. Although ANN has been applied in wide spectrum of problems, the researchers still seek for improvement since the performance of ANN is largely attributed by selection of initial parameter values, hidden layer selection, selection of learning patterns (dataset partition) and the convergence of the training algorithm.

For instance, the number of neurons of ANNs have a direct impact to the performance of the ANNs. Too many neurons can cause overfitting problem while too few neurons can contribute in poor approximation [1]. The vector of weight for network is also also crucial since it is contributing to the better fitness function value. These two parameters are amongst the optimization problem occurs in ANN. Meta-Heuristic algorithms have been given the attention in order to optimize the design and ANNs parameters.

Meta-Heuristic techniques are stochastic optimization algorithms that seek for a high quality solutions for an optimization problem in a reasonable computational effort [2, 3]. Mostly, local search methods are heuristic information based techniques since their parameter search is focused on local differences and might be the optimal solution is outside this local area. Meta-Heuristic techniques employ learning strategies in which combine variant intelligent concepts for exploring the search space to arrange information [4]. Genetic algorithm (GA), particle swarm optimization (PSO), simulated annealing (SE), tabu search (TS), ant colony optimization (ACO) and harmony search algorithm (HSA) are examples of meta-heuristic techniques. Majority of these algorithms except HSA are inspired from biological processes. Whilst, HSA is from musical inspiration process. These training algorithms have global searching ability to learn approximate optimal solution lacking the gradient information of the error functions [1, 5].

This paper is organized as follows. Section 2, presents related work. Section 3, provides Harmony Search Algorithm. Finally, Section 4 gives the conclusion.

## 2. Related Work

The ANN training process involves with adjusting the weight of neuron iteratively to reduce the error due to the high dimensionality and multimodal of the exploration search space. Hence, the process of NN learning needs more powerful optimization techniques [6]. In general, a set of dataset known as training set is presented to inputs to begin the network training to determine the correct outputs. After finishing this process, unseen data which is known as testing set is presented to the network's input to test the ability of the classifier to generalize [7].

There are three major learning paradigms for ANN. These paradigms include supervised, unsupervised and reinforced learning. The most common and widely used learning algorithm is Back-Propagation (BP), a supervised learning algorithm. BP is a gradient descent based algorithm. This algorithm calculates the output of the network and reduces the mean square error (MSE) between the actual output and the desired output through adjusting weights accordingly [6-10]. However, many researchers have found that that BP subjected to problem associated with local minima in their learning instances. Furthermore, the performance of BP is also attributed by the selection of appropriate learning parameters.

Researchers have adopted several meta-heuristic global optimization algorithms to overcome the local minima problem. Genetic Algorithms (GA), is used in ANN learning by [9], [11-13]. GA is a guided global search algorithm influenced by the natural evolution process of reproduction, mutation, crossover and selection. It behaves as an intelligent process of stochastic searches within a specified search space to solve a problem [14]. GA has been used to optimize ANNs' weights and has resulted in a good performance in ANN training [8,28]. Unlike BP, GA generates a pool of initial solutions in which the selection of the near optimal solution can be made for a choice. However, GA suffers from mutation problem that needs more time to conduct a search process that leads a premature convergence. [14].

Particle Swarm Optimization (PSO) is another so called global search optimization algorithm adopted by the researchers in ANN learning [12], [15-17]. PSO algorithm mimicked from the social behavior of fish schooling or flying birds looking for food sources through collaborative work of the population. Therefore, in PSO algorithm, population and members, known as particles are created, to initialize the search space and worked together to search the best solution through providing the particles random positions and velocities. The important features of PSO such as memory and positive collaboration between its individuals gave it the ability to prevent the mutation problem of GA. Zamani and Sadeghian claimed that PSO able to find global search [17].

However, the exploration of the search space could be slow since PSO calls evaluation function by the number of its population size in each iteration (Soltani et al., 2011). Furthermore, PSO also has potential to be trapped in local optimum if improper values are assigned to the parameters [18, 19]. Apart from GA and PSO, Ant Colony Optimization (ACO) [20, 21] and recently introduced Harmony Search algorithm (HSA) [7, 10, 22] have been widely proposed to seek for the optimal network weights.

Attempt to hybrid BP with metaheuristic algorithm in order to improve BP performance is also made by researchers. For instance, hybridization of BP with PSO and GA [19], [23-27]. In these hybrid BP-GA and BP-PSO, the GA and PSO are applied to initialize and adjust the weights of BP network to escape from the local minima. Both works found the meta-heuristic algorithm has more potential than the local search algorithms for avoiding local minima since it is capable to explore different area of search space simultaneously.

## 3. Harmony Search Algorithm (HSA)

HSA is considered relatively new stochastic global optimization (SGO) algorithm in the field of meta-heuristic algorithms. It has been derived from the process of musical improvisations, where the solution vector is analogous to the harmony in music. HSA can be easily used to solve the optimization problem as the global and local search schemes are analogous to the music requirements [28]. The concept of HSA is similar to other SGO methods in combining the rules of randomness to simulate the musical process. In HSA, a musician attempts to play and test a tone on his instrument to choose an excellence outcome in the harmony with the rest of the band. To produce a new harmony, either a tone from harmony memory (HM) is played with a rather modification or a current tone from the harmony is played

or totally a new tone from a range of acceptable tones is played. Therefore, the best solutions in the HM are selected and stored in the HM till better ones are obtained. Suppose the received harmony is better than the one in the HM then the worst in the HM is swapped by the best solution obtained [10, 22]

HS method is competitive substitute to other SGO techniques and can be used for handling both continuous and discrete variables [7]. The method had shown a noticeable improvement and increased its application in many aspects of IT related problems such as classification problems, robotics and web clustering [10]. HSA has been successfully applied in a variant of optimization problems and presented a number of benefits over other meta-heuristic optimization techniques. The benefits include no extra computational effort required in finding a solution. HSA employs stochastic searching strategies, therefore no derivative information needed to assist the search space. Furthermore, it combines the current solution vector to make a new solution. These characteristics of HSA has made it robust and successful method [29].

In addition, HSA has been implemented in the training process of ANNs particularly in a weight optimization in which important for the success of the training process [30]. It has capability of significantly converge to the optimal or near optimal solution [18]. Due to that, HSA owe the fastest convergence rate and the possibility of scapping local minima compared to BP, GA and PSO [7, 10]. The variant of HSA versions such standard harmony search algorithm (HSA) improved harmony search algorithm (IHS), modified improved harmony search algorithm (MIHS), global-best harmony search algorithm (GHS) and self-adaptive global-best harmony search algorithm (SGHS) has been developed to improve HSA. Most of these variants have used for weight optimization for the supervised ANNs and they showed remarkable improvements in the learning of feed forward ANNs. The technique increased the convergence speed and reduced the convergence error rate. Since the research of metat-heuristic algorithms especially the HSA variants are at their early stages thus there could be an interesting research area to apply HSA for weight optimization or optimizing the hidden layer selection and their respective neurons or investigate the possibility of HSA could be used for the optimal learning rate.

**3.1. Hybridization of BP with HSA**

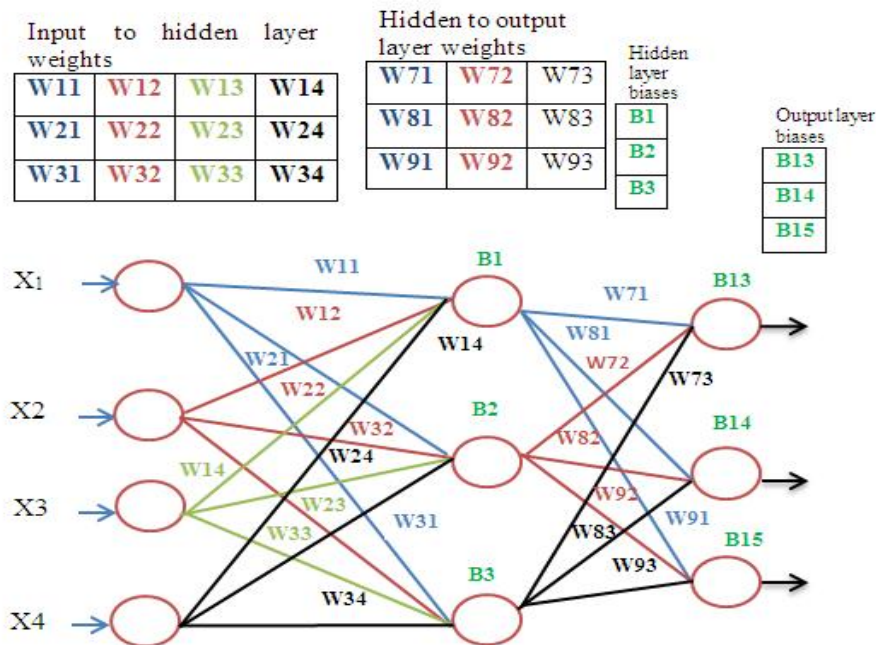


Figure 1. FFANN Sample for Weight Vector Representation

Basically, the training process of ANNs involves with deciding the connection weights between the neurons to minimize the error. In this work, we attempt to exploit the capability of HSA in giving global optima value to improve BP algorithm performance. Although GA has solved the problem of gradient descent methods like BP, it requires long time execution in mutation and also possible to have premature convergence. Attempt to use PSO in ANN learning able to reduce ANNs error function by providing near optimal weights to ANNs network and minimizes the GA's long execution in mutation. But, PSO is also subjected to local optimal problem whenever the parameter values are assigned improperly [19]. PSO reduces the convergence speed and increase the execution time compared to HSA since the algorithm calls the fitness function in each iteration. [18]. Furthermore, the increasing number of hidden layers and their respective neuron can add more dimensions to the particles. On the other hand, making changes to the parameters can affect the performance of ANNs and vice versa though training of ANNs with PSO is very competitive associated to other ANNs training algorithms [17]. Due to the above mentioned causes, researchers turned their attention to the newly introduced in the field of meta-heuristic algorithms known as harmony search algorithm (HSA). It has been successfully applied in a variant of optimization problems and has presented a number of benefits over other meta-heuristic optimization algorithms.

In applying the HSA for optimizing the weights of ANNs, HS generates a pool of solution vector and keeps them in the HM to be taken as weights of ANNs since it produces a near optimal weight value in which supports the success of the training process. the harmony memory represents the weights and bias of input layer to hidden layer and from hidden layer to output layer [12] as shown in Figure 1.

Vector based representation and adjacency matrix based representation are two common strategies used for feed-forward artificial neural network weight representation. The vector based representation is more appropriate for HS algorithm. However, the method must be accepted the common ranges specified by vectors  $[x^L \ x^U]$  as NN weights have common value ranges and they are not discrete variables for optimization problems. In such case, the HM representation can be regarded as being of different musicians that are exploiting similar musical instrument. Doing so; means the musical instrument will indicate that they will have a common pitch range. Hence each component value of  $[x^L \ x^U]$  is identical to all decision variables. Each harmony vector in HM is represented using a vector representation as shown in Figure 1. The vector comprises a complete set of NN weights and biases. In this work, HSA is only used for optimizing the weights of ANNs whenever BP's error is unchanged. For example, the error remain the same after six consecutive iterations. We suspect, BP has trapped into local minimum or overfitting problems. This condition is used as indicator of BP's unsuccessful learning process. Hence, HSA is employed to generate a fresh near optimal weights to continue the training process.

#### 4. Conclusion

In this paper, several methods in the field of meta-heuristic algorithms and gradient information based algorithms have been reviewed and discussed in their application of ANNs weight optimization. The research of hybridizing ANNs with meta-heuristic algorithms is its early stage and therefore the need for solving different research aspects such as theories and methods require to be enhanced and standardized. The meta-heuristic algorithms have provided a solution to the problem of local search based techniques and some of them have introduced some problems such as GA and PSO. However, HSA has shown a good promising characteristic in which can support it to be a good candidate for training ANNs by providing a near optimal weights.

#### References

- [1] Abdull Hamed HN, Shamsuddin SM, Salim N. Particle Swarm Optimization For Neural Network Learning Enhancement. *Jurnal Teknologi*. 2012; 49(1): 13–26.
- [2] Askarzadeh A, Rezazadeh A. Artificial neural network training using a new efficient optimization algorithm. *Applied Soft Computing*. 2012.
- [3] Carvalho M, Ludermitr TB. *Particle swarm optimization of neural network architectures and weights*. Proceedings of the 2007 Hybrid Intelligent Systems. 7th International Conference, IEEE. 2007; 336-339.

- [4] Castro PA, Von Zuben FJ. *Training multilayer perceptrons with a Gaussian artificial immune system*. Proceedings of the 2011 Evolutionary Computation (CEC). IEEE Congress. 2011; 1250-1257.
- [5] Correa B, Gonzalez AM. *Evolutionary algorithms for selecting the architecture of a MLP neural network: a credit scoring case*. Proceedings of the 2011 Data Mining Workshops (ICDMW), IEEE 11th International Conference. 2011; 725-732.
- [6] Ding S, Li H, Su C, Yu J, Jin F. *Evolutionary artificial neural networks: a review*. *Artificial Intelligence Review*. 2013; 1-10.
- [7] El-Henawy I, Kamal A, Abdelbary H, Abas A. *Predicting stock index using neural network combined with evolutionary computation methods*. Proceedings of the 2010 Informatics and Systems (INFOS), 2010 The 7th International Conference. 2010; 1-6.
- [8] Er MJ, Liu F. *Genetic algorithms for MLP neural network parameters optimization*. Proceedings of the 2009 Control and Decision Conference. CCDC'09. Chinese: IEEE, 2009; 3653-3658.
- [9] Gao W. *Evolutionary neural network based on new ant colony algorithm*. Proceedings of the 2008 Computational Intelligence and Design. ISCID'08. International Symposium: 2008; 318-321.
- [10] Kattan A, Abdullah R. *Training Feed-Forward Artificial Neural Networks For Pattern-Classification Using The Harmony Search Algorithm*. Proceedings of the 2013 The Second International Conference on Digital Enterprise and Information Systems (DEIS2013): The Society of Digital Information and Wireless Communication. 2013: 84-97.
- [11] Kattan A, Abdullah R, Salam RA. *Harmony search based supervised training of artificial neural networks*. Proceedings of the 2010 Intelligent Systems, Modelling and Simulation (ISMS), International Conference on. 2010; 105-110.
- [12] Kulluk S, Ozbakir L, Baykasoglu,A. *Training neural networks with harmony search algorithms for classification problems*. *Engineering Applications of Artificial Intelligence*. 2012; 25(1): 11-19.
- [13] Li L, Lu Sb, Chu Xs, Yu Gm. *A Combinatorial Search Method Based on Harmony Search Algorithm and Particle Swarm Optimization in Slope Stability Analysis*. Proceedings of the 2009 Computational Intelligence and Software Engineering. CiSE 2009. International Conference. 2009, 1-4.
- [14] Lu C, Shi B, Chen L. *Hybrid BP-GA for multilayer feedforward neural networks*. Proceedings of the 2000 Electronics, Circuits and Systems. ICECS 2000. The 7th IEEE International Conference: 2000: 958-961.
- [15] Osman IH. *Metaheuristics: A General Framework*. 数理解析研究所講究録. 2001: 169-170.
- [16] Pan QK, Suganthan PN, Tasgetiren MF, Liang JJ. *A self-adaptive global best harmony search algorithm for continuous optimization problems*. *Applied Mathematics and Computation*. 2010; 216(3): 830-848.
- [17] Ramos-Pollán R, Guevara López MA, Naimy González de Posada N. *Optimizing the area under the ROC curve in multilayer perceptron-based classifiers*. Proceedings of the 2011 FUTURE COMPUTING, The Third International Conference on Future Computational Technologies and Applications. 2011; 75-81.
- [18] Romero VIM, Tomes LL, Yusiong JPT. *Tetris Agent Optimization Using Harmony Search Algorithm*. *International Journal of Computer Science Issues*. 2011; 8(1): 22-31.
- [19] Shayeghi H, Shayanfar H, Azimi,G. *A Hybrid Particle Swarm Optimization Back Propagation Algorithm for Short Term Load Forecasting*. *International Journal on Technical and Physical Problems of Engineering*. 2010; 1(3): 12-22.
- [20] Soltani MAZ, Haghghat AT, Rashidi H, Chegini TG. *A Couple of Algorithms for K-Coverage Problem in Visual Sensor Networks*. *IACSIT*. 2011: 96-101.
- [21] Sun X, Liu Q, Zhang L. *A BP neural network model based on genetic algorithm for comprehensive evaluation*. Proceedings of the 2011 Circuits, Communications and System (PACCS), 2011 Third Pacific-Asia Conference on: IEEE. 2011: 1-5.
- [22] Tavakoli S, Valian E, Mohanna,S. *Feedforward neural network training using intelligent global harmony search*. *Evolving Systems*. 2012; 3(2): 125-131.
- [23] Wang S, Yin S, Jiang M. *Hybrid neural network based on GA-BP for personal credit scoring*. Proceedings of the 2008 Natural Computation. ICNC'08. Fourth International Conference on: IEEE. 2008: 209-214.
- [24] Yang XS. *Harmony search as a metaheuristic algorithm Music-inspired harmony search algorithm*. Springer. 2009: 1-14.
- [25] Yao X. *Evolving artificial neural networks*. Proceedings of the IEEE. 1999; 87(9): 1423-1447.
- [26] Yoshikawa M, Otani K. *Ant Colony Optimization Routing Algorithm with Tabu Search*. Proceedings of the International Multiconference of Engineers and Computer Scientists. 2010: 17-19.
- [27] Zamani M, Sadeghian,A. *A variation of particle swarm optimization for training of artificial neural networks*. *Computational Intelligence and Modern Heuristics*. Al-Dahoud Ali. 2010.
- [28] Zanchettin C, Ludermir TB, Almeida LM. *Hybrid training method for mlp: optimization of architecture and training*. Systems, Man, and Cybernetics, Part B: Cybernetics, *IEEE Transactions on*. 2012; 41(4): 1097-1109.

- 
- [29] Zhang JR, Zhang J, Lok TM, Lyu MR. A hybrid particle swarm optimization–back-propagation algorithm for feedforward neural network training. *Applied Mathematics and Computation*. 2007; 185(2): 1026-1037.
- [30] Zhu M, Wang L. *Intelligent trading using support vector regression and multilayer perceptrons optimized with genetic algorithms*. Proceedings of the 2010 Neural Networks (IJCNN), The 2010 International Joint Conference on: IEEE, 2010: 1-5.
- [31] Zinati RF, Razfar MR. Constrained optimum surface roughness prediction in turning of X20Cr13 by coupling novel modified harmony search-based neural network and modified harmony search algorithm. *The International Journal of Advanced Manufacturing Technology*. 202; 58(4): 93-107.