

# A Review of Heuristic Global Optimization Based Artificial Neural Network Training Approaches

D. Geraldine Bessie Amali<sup>1</sup>, Dinakaran M.<sup>2</sup>

<sup>1</sup>School of Computing Science and Engineering, VIT University

<sup>2</sup>School of Information Technology, VIT University

---

## Article Info

### Article history:

Received Nov 5, 2016

Revised Jan 9, 2017

Accepted Feb 18, 2017

---

### Keyword:

Artificial Neural Networks  
Feedforward Neural Networks  
Optimization Algorithms  
Recurrent Neural Networks  
Training Algorithms

---

## ABSTRACT

Artificial Neural Networks have earned popularity in recent years because of their ability to approximate nonlinear functions. Training a neural network involves minimizing the mean square error between the target and network output. The error surface is nonconvex and highly multimodal. Finding the minimum of a multimodal function is a NP complete problem and cannot be solved completely. Thus application of heuristic global optimization algorithms that computes a good global minimum to neural network training is of interest. This paper reviews the various heuristic global optimization algorithms used for training feedforward neural networks and recurrent neural networks. The training algorithms are compared in terms of the learning rate, convergence speed and accuracy of the output produced by the neural network. The paper concludes by suggesting directions for novel ANN training algorithms based on recent advances in global optimization.

*Copyright © 2017 Institute of Advanced Engineering and Science.  
All rights reserved.*

---

## Corresponding Author:

D. Geraldine Bessie Amali  
School of Computing Science and Engineering,  
VIT University,  
Near Katpadi Road, Vellore-632014.  
Email: geraldine.amali@vit.ac.in

---

## 1. INTRODUCTION

Artificial Neural Network (ANN) is a mathematical model of the biological nervous system. A neural network consists of sets of adaptive weights, i.e. numerical parameters that are tuned by a learning algorithm, and are capable of approximating non-linear functions [18]. It can also be used to solve various problems including pattern recognition, classification, and function approximation. A neural network is an inter connection of neurons arranged in layers. It consists of an input layer, an output layer and zero or more hidden layers. The input layers and the subsequent layers are connected by links with weights. The strength of the link depends on the weight. The performance of the neural network lies in the number of neurons in each layer and also the weights. The goal of any training algorithm used by the neural network is to determine the weights of the links so that it reduces the error between the output produced by the neural network and the ideal output.

### 1.1. Feedforward Neural Networks

In typical feedforward architecture the neurons are arranged in cascaded layers where all the neurons in one layer are connected to all the neurons in the adjacent layer [18]. However, the neurons are not connected to other neurons in the same layer. The branching between the neurons is unidirectional therefore information can pass only in one direction and there is no feedback. The output of the first layer is presented as input to the second layer. The branches have weights associated with them which can be adjusted by the learning algorithm. A multi-layer perceptron with adequate number of neurons and one hidden layer can

approximate a nonlinear function [18]. Figure 1 shows the architecture of a feedforward neural network with one hidden layer.

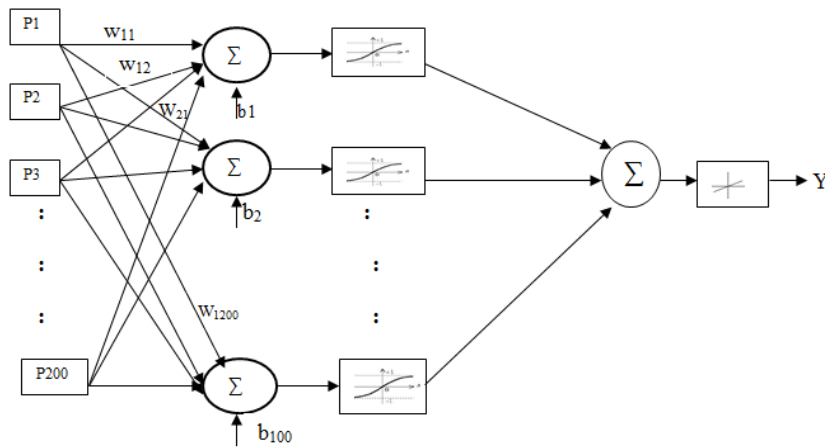


Figure 1. A Feedforward Neural Network with One Hidden Layer.

### 1.2. Recurrent Neural Networks

In feedforward neural networks the weights associated with the branches are fixed and therefore the state of the neuron solely depends upon the input given to the neuron [18]. This is a static model because it does not depend on the past state of the neurons. Recurrent neural networks on the other hand utilize feedback. Unlike feedforward neural networks this architecture uses nonlinear processing units, is fully connected and is fault tolerant [18]. Due to their dynamic nature and temporal behaviour recurrent neural networks are used in high intelligent systems with applications in symbolic reasoning. Figure 2 shows the architecture of recurrent neural network with four neurons in the input layer, two in the hidden layer and two output neurons.

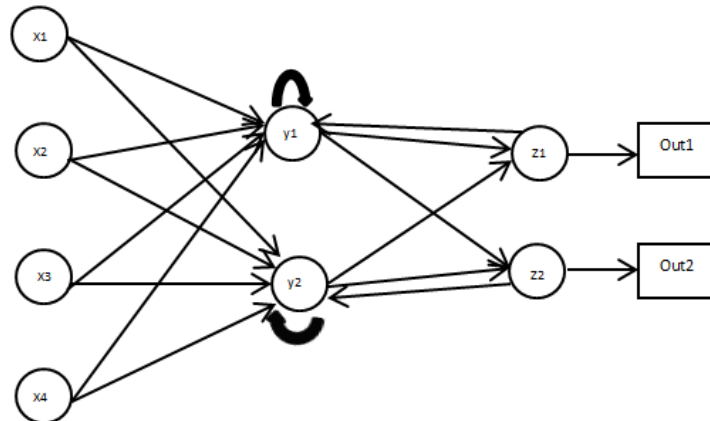


Figure 2. A Recurrent Neural Network with One Hidden Layer

### 1.3. Learning in Artificial Neural Networks

Learning also called as training a neural net is a very crucial process and is done by systematically adjusting the connection weights. The ANN learning can be one of the following types namely supervised, unsupervised or reinforcement learning [21]. In supervised learning the neural net is trained by input and corresponding pairs. The connection weights are adjusted in such a way so as to reduce the error. Unsupervised learning however trains the neural net based on the correlation of data. Reinforcement learning is a special type of learning which employs feedback from the environment. The neural net is trained with positive rewards and negative rewards based on the performance [21].

#### 1.4. Global Optimization

Mathematical optimization is minimization or maximization of a real valued function by selecting the best solution from an available set of feasible solutions [32]. In the case of minimizing a real valued function also called as the cost function the goal is to determine the value of the input which minimizes the function the most. An optimization problem can be formulated as follows

$$\begin{aligned} & \text{minimize } f(x) \\ & \text{Subject to } x \in \Omega \end{aligned}$$

Where  $f: R^n \rightarrow R$  is the function that is to be minimized and  $n$  is the dimensionality of the vector  $x$ . The set  $\Omega$  is a subset of  $R^n$ . When the set  $\Omega$  is the whole of  $R^n$  the optimization is said to be unconstrained. In this paper an unconstrained optimization problem is considered. The optimization problem above can be viewed as finding the vector  $x^*$  from the domain  $\Omega$  such that that  $f(x^*) \leq f(x)$ . A point  $x^* \in \Omega$  is a local minimum of  $f$  over  $\Omega$  if there exists  $\varepsilon > 0$  such that  $f(x^*) \leq f(x)$  for all  $x \in \Omega \setminus \{x^*\}$  and  $\|x - x^*\| < \varepsilon$ . On the other hand a point  $x^* \in \Omega$  is a global minimum of  $f$  over  $\Omega$  if  $f(x^*) \leq f(x)$  for all  $x \in \Omega$ . In general, it is only practical to compute a good local minimum as the problem of nonlinear global optimization is NP-complete.

Optimization problems can either be convex or non-convex. A Convex function has one local minimum which is also the global minimum. Highly efficient algorithms like the interior point algorithm exist to compute the global minimum for convex optimization problems [32]. Non-convex functions on the other hand may contain numerous local minima which makes the problem of locating the global minimum either very difficult or impossible. In many instances optimization algorithms get stuck in local minimum without converging to the global minimum. Thus only good local minimum can be computed in general. However a majority of important problems in engineering like filter design involve non-convex global optimization. Hence development of heuristic random search algorithms inspired by biology like PSO and GA which can compute near optimal solutions are of interest [20]

The rest of the paper is organized as follows. Section 2 presents the evolution of the training algorithms and compares the various optimization algorithms used in training feedforward neural networks. Section 3 reviews the algorithms used in training recurrent neural networks. Conclusion is presented in section 5.

## 2. REVIEW OF OPTIMIZATION ALGORITHMS USED FOR TRAINING FEEDFORWARD NEURAL NETWORKS

Training ANNs using back propagation algorithm had limitations in terms of overfitting, increase in learning time with the size of the training data, and most importantly the risk of getting stuck in the flat regions of the search space thereby converging to a local minimum and not finding the global minimum[1]. Therefore, biologically inspired optimization algorithms like Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) were used to train ANNs.

GA is inspired by natural evolution and adopts the principles of selection, crossover and mutation [24]. It is stochastic and derivative free and therefore can be applied to both continuous and discrete optimization problems. In [2] [23] the authors have used GA with crossover to calculate the weights of the feedforward neural network (FNN). He has demonstrated in his work [3] that GA outperforms back propagation algorithm. Training neural networks with distributed GA reinforced by perceptron learning rule was proposed and applied by Olikier et al in [4]. A version of GA known as soft algorithm is combined with backpropagation and soft-bp is applied to train ANNs in [19] by Adawy et.al. This algorithm obtains a good weight vector thereby reducing the error of the output. The parallel version of GA has been used in [22] for time series prediction.

PSO is inspired by the swarm behaviour of flock of birds or school of fishes. It was proposed by Kennedy and Eberhart in [5] and has been used for training neural networks by Gudise and Venayagamoorthy in [6]. The performance of the PSO algorithm is compared with the backpropagation algorithm by training the neural network to learn a nonlinear function. It was found the PSO was faster between the two algorithms to learn the nonlinear function as it required less number of computations than BP to attain the same error goal [6]. PSO algorithm itself has certain limits in terms of convergence, precision and parameter selection. It was slower during the final stages of evolution and had lower precision.

Therefore, Chen et al in [31] proposed an algorithm called Artificial Fish Swarm Algorithm (AFSA)- PSO parallel hybrid evolutionary algorithm (APPHE) for training FNNs. This algorithm divides the PSO population into two sub populations. PSO is executed in one sub population and AFSA in the other in parallel. The best solution of both the sub population is given back to the swarm and PSO is now executed in both sub populations. The algorithm terminates when a termination criterion is satisfied. The authors tested

the performance of the proposed APPHE algorithm with the Levenberg- Marquardt Back propagation (LMBP) algorithm by training the neural network in iris data classification. The neural network trained using the APPHE algorithm did better than LMBP in terms of faster convergence to the global minimum and accuracy of the result.

Hybrid algorithms that combine global optimization algorithm and local search algorithms were used to train ANNs. Hybrid Artificial Bee Colony Algorithm that combines the Artificial Bee Colony algorithm (ABC) and the Levenberg- Marquardt (LM) is used for training ANN in [7]. ABC is a global optimization algorithm and finds the global minimum whereas LM is used to exploit the local minimum. Therefore a hybrid algorithm that combines the exploration ability of the ABC and exploitation ability of the LM has been proposed by Ozturk in [7]. The hybrid algorithm performs better than the algorithms by themselves. A modified LM algorithm which addressed the demand of memory for large jacobians and the need for inverting the large matrices was proposed by Wilamowski and Chen in [11]. Their proposed algorithm used a new performing index which reduces the size of the matrix that is to be inverted thereby increasing the computation speed.

Simulated annealing is a global search heuristic that is inspired by annealing in metallurgy. Metallurgy is a physical process of heating metal to very high temperatures and then cooling it very slowly. This helps remove the defects in the crystals formed. Simulated annealing is combined with local gradient search algorithm (Rprop) in [8] and in [9] with tabu search to train ANNs. A composite squared error algorithm was proposed and applied to train ANNs by Gonzaga et al. in [11]. In this algorithm the first part of training uses the linear error while the second part uses the nonlinear. By doing so the algorithm escapes the suboptimal solutions and converges to the optimal solution faster than the backpropagation algorithm.

A novel ant algorithm proposed Dorigo in [28] is used by Li and Liu in [27] and applied to train feedforward neural net for call admission control. Ant colony optimization (ACO) is a global optimization algorithm that is inspired by the swarm behaviour of ants following a path seeking food from their colonies [28]. The ants have to perform two tasks. First they have to select the path which they want to follow and secondly adjust their pheromone level along the chosen path. A version of the aforementioned ACO is used for training a neural net in [27]. This ACO trained neural network performed well when compared to BP however its performance degrades with the increase in the number of inputs due to the communication overhead.

Quantum Computing with optimization algorithms started to evolve. Quantum inspired GA [13] a quantum inspired parallel GA was proposed in [14]. A Quantum Shuffled Frog Leaping Algorithm (QSFLA) was proposed and used for training ANNs by Liu and Zhang in [15]. This algorithm efficiently solved continuous optimization problem in high dimensional space and did better than the BP algorithm in terms of convergence and accuracy.

### 3. REVIEW OF OPTIMIZATION ALGORITHMS USED FOR TRAINING RECURRENT ARTIFICIAL NEURAL NETWORKS

Breeding swarm algorithm is a hybrid of GA and PSO. This was proposed by Matthew et al in [10] and used it to training of ANNs. Their algorithm uses the classical PSO formula for updating the velocity and the positions of the particles, and uses the selection, mutation and crossover principles from GA. In addition the authors have also introduced a parameter called the breeding parameter, which determines the population size that should undergo breeding. Since breeding swarms algorithm was a general population based algorithm, when it was used to train recurrent neural networks it was found that the algorithm was able to scale better. In [16] a hybrid Bayesian learning method which combines Markov chain Monte Carlo methods with fuzzy membership functions and GA is used by Kocadagli for training Bayesian neural networks. The author has addressed the problems of complexity in choosing the parameters of the model, the training time associated with the Bayesian neural networks. He argues that the proposed hybrid model can overcome the problems faced with normal training algorithms. A hybrid model that combines the gradient descent and metaheuristics is proposed and used in [17]. An improved version of PSO with time varying parameter and constriction helps in improving the search ability and convergence. In order to prevent overfitting a cross validation method is also included in the algorithm. A variant of PSO called the modified binary Particle Swarm Optimization (MPSO) was proposed by Eberhart for binary problems [21]. This version of PSO was used to train recurrent neural networks in [29] for decoding of 1/n rate convolutional codes. This approach provided low latency and converged to a global minimum thereby making it more practicable.

#### 4. DISCUSSION

A review of the literature shows that optimization algorithms approximate nonlinear functions and provide near accurate solutions. Biologically inspired optimization algorithms like GA, PSO, ACO, ABC, AFSA are stochastic in nature. This helps these algorithms explore the search space and escape from local minima. Backpropagation however uses gradient descent and might get stuck in the local minimum. Literature shows that variants of the optimization algorithms especially hybrid algorithms that combine global and local search heuristic perform better than the algorithms by themselves. These hybrid algorithms performed better than the algorithms by themselves in terms of the learning rate, accuracy and convergence. Because these hybrid algorithms combine the exploratory ability of the global optimization algorithms and the exploitation ability of the local search algorithms they provide better results. Literature also confirms that hybrid algorithms outperform the algorithms by themselves in training recurrent neural networks. Shown in Table 1.

Table 1 Comparison of Global Optimization Algorithms

ANN Training Algorithm	Successes	Challenges
Genetic Algorithm	Explores large and complex search space	Slower convergence
Particle Swarm Optimization	Fewer number of computation required to learn	Slower convergence
Hybrid Artificial Bee Colony	Accuracy of results	Not practicable for high dimensional classification problems
Hybrid PSO with stop criteria	Faster convergence	Less exploration of the search space
Ant Colony Optimization	Accuracy of result	Performance degrades with the increase in the number of inputs.
Artificial Fish Swarm Optimization	Faster convergence	Limited applications
Hybrid Simulated Annealing	Accuracy of results	Overfitting

In recent years novel heuristic global optimization algorithms that outperform current state of the art optimization algorithms have been proposed [32]. In the approach proposed in [32] alternating cycles of exploration and exploitation are used to achieve a compromise between exploration of new solutions and exploitation of existing solutions and avoid premature convergence to local minima. Gradient based algorithms like Backpropagation have a tendency to get stuck in local minima leading to poor performance of the ANN. Thus application of novel heuristic global optimization algorithms like the Galactic Swarm Optimization (GSO) algorithm [32] to ANN training is of interest. Comparison of different heuristic global optimization algorithms such as [32] and [33] on benchmark ANN training problems can be considered for future work.

#### 5. CONCLUSION

This paper reviews the global optimization algorithms used for training feedforward and recurrent neural networks. ANNs and most of the global optimization algorithms are biologically inspired and they borrow ideas from the social behaviour and biological structure of the individuals. Thereby it can be positively stated that training neural nets with biologically inspired optimization algorithms will provide a more complete learning. A review of the literature proves that due to the stochastic nature of these algorithms, hybrid algorithms which combine the global and local optimization algorithms outperform the algorithms by themselves in terms of faster convergence and accuracy of the output. In recent years novel heuristic global optimization algorithms that outperform current state of the art optimization algorithms have been proposed. These algorithms can be considered for ANN training in the future.

#### REFERENCES

- [1] M Gori, A Tesi. On the problem of local minima in backpropagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992; 14: 76-86.
- [2] D Whiteley. *Applying Genetic Algorithms to Neural Networks Learning*. Proceedings of 7th Conference of the Society of Artificial Intelligence and Simulation of Behavior. Sussex, England: Pitman Publishing, 1989; 137-144.
- [3] D Whiteley, T Starkweather, C Bogart. Genetic Algorithms and neural Networks: Optimizing Connections and Connectivity. *Parallel Computing*, 1990; 14: 347-361.

- [4] Olikier, S, et al. *Design architectures and training of neural networks with a distributed genetic algorithm*. IEEE International Conference on Neural Networks. 1993; 1: 199-202.
- [5] James Kennedy, R Eberhart. *Particle swarm optimization*. IEEE International Conf. on Neural Networks, Perth, Australia. 1995; 4: 1942-1948.
- [6] Gudise, Venayagamoorthy GK. *Comparison of particle swarm optimization and backpropagation as training algorithms for neural networks*. Proceedings of the 2003 IEEE in Swarm Intelligence Symposium, 2003: 110-117.
- [7] Ozturk, Karaboga. *Hybrid Artificial Bee Colony algorithm for neural network training*. 2011 IEEE Congress on Evolutionary Computation (CEC), 2011: 84-88.
- [8] N Treadgold, T Gedeon. Simulated annealing and weight decay in adaptive learning: the sarprop algorithm. *IEEE Transactions on Neural Networks*, 1998; 9: 662-668.
- [9] T Ludermir et al. An optimization methodology for neural network weights and architectures. *IEEE Transactions on Neural Networks*, 2006; 17(5): 1452-1460.
- [10] Settles M et al. *Breeding swarms: a new approach to recurrent neural network training*. In Proceedings of the 7th annual conference on Genetic and evolutionary computation ACM, 2005; 185-192.
- [11] Gonzaga D et al. *Composite squared-error algorithm for training feedforward neural networks*. Advances in Digital Filtering and Signal Processing IEEE Symposium, 1998; 116-120.
- [12] Wilamowski et al. *Efficient algorithm for training neural networks with one hidden layer*. In Proc. IJCNN Jul 10. 1999; 3: 1725-728.
- [13] Narayanan A. Moore M. *Quantum-inspired genetic algorithms*. In Evolutionary Computation, Proceedings of IEEE International Conference, 1996: 61-66.
- [14] Han KH et al. *Parallel quantum-inspired genetic algorithm for combinatorial optimization problem*. In Evolutionary Computation, 2001 Proceedings of the 2001 Congress, 2001; 2: 1422-1429.
- [15] Liu L, Zhang Q. *Training and application of process neural network based on quantum shuffled frog leaping algorithm*. In Computer Science and Network Technology (ICCSNT), 2013 3rd International Conference. 2013: 829-833.
- [16] Kocadağlı O. A novel hybrid learning algorithm for full Bayesian approach of artificial neural networks. *Applied Soft Computing*, 2015: 52-65.
- [17] Yaghini M et al., A hybrid algorithm for artificial neural network training. *Engineering Applications of Artificial Intelligence*. 2013; 26(1): 293-301.
- [18] Ham FM, Kostanic I. Principles of neurocomputing for science and engineering. *McGraw-Hill Higher Education* 2000 Sep 1.
- [19] El Adawy et al. *A SOFT-backpropagation algorithm for training neural networks*. In Radio Science Conference, 2002 (NRSC 2002). Proceedings of the Nineteenth National 2002: 397-404.
- [20] Noel MM. A new gradient based particle swarm optimization algorithm for accurate computation of global minimum. *Applied Soft Computing*. 2012; 12: 353-9.
- [21] Kennedy J, Eberhart RC. *A Discrete Binary Version of the Particle Swarm Algorithm*. In Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation. 1997 IEEE International Conference, 1997; 5: 4104-4108.
- [22] Mendivil SG et al. Optimization of artificial neural network architectures for time series prediction using parallel genetic algorithms. In *Soft Computing for Hybrid Intelligent System*. 2008: 387-399.
- [23] Zi-wu RE, Ye SA, Improvement of Real-valued Genetic Algorithm and Performance Study [J]. *Acta Electronica Sinica.*, 2007; 2: 017.
- [24] Guo Z, Uhrig RE. *Using genetic algorithms to select inputs for neural networks*. In Combinations of Genetic Algorithms and Neural Networks, 1992, COGANN-92. International Workshop on 1992: 223-234.
- [25] Kitano H. Designing neural networks using genetic algorithms with graph generation system. *Complex Systems Journal*, 1990; 4: 461-76.
- [26] Angeline PJ et al. An evolutionary algorithm that constructs recurrent neural networks. *IEEE transactions on Neural Networks*, 1994; 1: 54-65.
- [27] Li S, Liu Z. *A general CAC approach using novel ant algorithm training based neural network*. In Neural Networks, 1999. IJCNN'99. International Joint Conference on 1999; 3: 1885-1888.
- [28] Dorigo M, Gambardella LM. Ant colony system: a cooperative learning approach to the traveling salesman problem. *IEEE Transactions on Evolutionary Computation*, 1997; 1: 53-66.
- [29] Asvadi R, Ahmadian M. *Modified binary PSO Training of Recurrent neural Network for 1/n Rate Convolutional Decoders*. In Performance Evaluation of Computer and Telecommunication Systems, 2008. SPECTS 2008. International Symposium, 2008: 30-35.
- [30] Luitel B, Venayagamoorthy GK. *A PSO with quantum Infusion Algorithm for Training Simultaneous Recurrent Neural Networks*. In Neural Networks, 2009. IJCNN 2009. International Joint Conference, 2009: 1923-1930.

- [31] Chen X et al. *A novel hybrid Evolutionary Algorithm Based on PSO and AFSA for Feedforward neural Network Training in Wireless Communications*. Networking and Mobile Computing, 2008. WiCOM'08. 4th International Conference on 2008. 2012: 1-5
- [32] Muthiah-Nakarajan V, Noel MM. Galactic Swarm Optimization: A new global optimization metaheuristic inspired by galactic motion. *Applied Soft Computing*, 2016; 38: 771-87.
- [33] Javidy et al. Ions motion algorithm for solving optimization problem. *Applied Soft Computing*. 2015; 32: 72-79.

## BIOGRAPHIES OF AUTHORS



Geraldine Bessie Amali received her M.Tech. in Computer Science and Engineering from VIT university in 2014 where she received gold medal for graduating at the top of her batch. She also has a university rank in Master of Computer Applications from Bharathidasan University. She is currently working as an assistant professor at VIT University and has four years experience teaching computer science. Her research interests include machine learning and biologically inspired optimization algorithms.



Dr. Dinakaran M has completed his B.Tech (IT), M.Tech (IT-Networking) in Vellore Institute of Technology and Ph.D in Anna University, Chennai, Tamil Nadu, India. He worked in TATA Consultancy Services as Assistant System Manager from September 2006 to July 2009. Currently he is working as Associate Professor and Head of the Department in School of Information Technology and Engineering, VIT University, Vellore. He has published 25 articles in various International Conferences and Journals. His research interests are mobile networks, image retrieval and machine learning.