Combining backpropagation and genetic algorithms to train neural networks

George Papakostas*, Yiannis Boutalis and Sofoklis Samartzidis
Department of Electrical and Computer Engineering,
Democritus University of Thrace, Xanthi, Greece
E-mail: gpapakos@ee.duth.gr
E-mail: ybout@ee.duth.gr
E-mail: ssamrt@ee.duth.gr
*Corresponding author

Dimitrios Karras
Department of Automation
Chalkis Institute of Technology, Chalkis, Greece
E-mail: dakarras@teihal.gr

Basil Mertzios
Department of Automation, Laboratory of Control Systems and Comp. Intell.
Thessaloniki Institute of Technology, Thessaloniki, Greece
E-mail: mertzios@uom.gr

Abstract: In the present paper a comparative study of two possible combinations of the Backpropagation (BP) and a Genetic Algorithm (GA), for Neural Networks training is performed. The performance of these approaches is compared to each other and to each algorithm incorporated separately in the training procedure. The construction of hybrid optimization algorithms is originated from the need to manipulate and solve difficult optimization problems by combining their advantages. The locality and globality behaviour of BP and GA is investigated by the presented hybrid structures, by applying them in five popular benchmark problems. It is concluded, that a more sophisticated structure based on the collaboration of two powerful optimization algorithms can be used to train a typical neural network more efficiently.

Keywords: neural networks, backpropagation, genetic algorithms, hybrid algorithms.

Biographical notes: George A. Papakostas received the diploma of Electrical and Computer in Engineering in 1999 and the Msc. Degree in Electrical and Computer Engineering (topic in Feature Extraction and Pattern Recognition) in 2002, from Democritus University of Thrace (DUTH), Greece. Currently he is a Phd student in DUTH, in the field of Neural Networks. His research interests are focused in the field of pattern recognition, neural networks, feature extraction, optimisation, signal and mage processing.

Yiannis Boutalis received the diploma of Electrical Engineer in 1983 from the Democritus University of Thrace (DUTH), Greece and the phD degree in Electrical and Computer Engineering (topic Image Processing) in 1988 from the Computer Science Division of National Technical University of Athens, Greece. Since 1996, he serves as a faculty member, at the Department of Electrical and Computer Engineering, DUTH, Greece, where he is currently an Associate Professor. He served as an assistant visiting professor at University of Thessaly, Greece, and as a visiting professor in Air Defence Academy of General Staff of airforces of Greece. He also served as a researcher in the Institute of Language and Speech Processing (ILSP), Greece, and as a managing director of the R&D SME Ideatech S.A, Greece, specializing in pattern recognition and signal processing applications. His current research interests are focused in the development of Computational Intelligence techniques with applications in Control, Pattern Recognition, Signal and Image Processing Problems.

Dimitrios A. Karras received his Diploma and M.Sc. Degree in Electrical and Electronic Engineering from the National Technical University of Athens, Greece in 1985 and the Ph. Degree in Electrical Engineering, from the National Technical University of Athens, Greece in 1995, with honors. In 1990 he received, also, his Diploma Degree in Mathematics from the University of Athens. Since 2004, after his election, he has been with the Chalkis Institute of Technology, Automation Dept., Greece as full professor in Digital Systems as well as with the Hellenic Open University, Dept. Informatics as a visiting professor. He has published more than 30 research Journal papers in various areas of pattern recognition, image/signal processing and neural networks and more than 70 research papers in International Scientific Conferences. His research interests span the fields of pattern recognition and neural networks, multidimensional digital signal processing, image processing and analysis, communications and security, as well as parallel algorithms and fast processing.

1 INTRODUCTION

Artificial Neural Networks constitute an essential part of a modern decision making system, in which the need of knowledge storing and knowledge based decision has been inspired by the neural networks of a human's brain (Haykin,, 1999). Due to their knowledge storage capability, neural networks are able to be used for pattern recognition tasks and classification problems (Papakostas et al., 2005), while their ability to repeatedly learn their internal representation makes them very useful to real-time image and signal processing applications (Haykin, 2000; Langlet et al., 2001).

The most widely used algorithm for training neural networks, is the backpropagation (BP). Backpropagation is a gradient based algorithm with local behaviour, and thus the probability of converging to a local optimum, increases. In addition, BP has a slow convergence rate, so it needs quite a time in order to find a solution (Fahlman, 1988), especially when the optimal weight set is located in complicated weight spaces.

On the other hand genetic algorithms have proved to be efficient optimization methods, for hard optimization problems. Genetic algorithms have been used successfully in the past to find the optimal set of a neural network's weights (Montana et al., 1989; Harpharm et al., 2004). Due to their parallelism they can provide high convergence rates, while their stochastic behaviour can highly guarantee the globality of the solution founded (Coley, 2001).

Therefore, it worths wondering if the above optimization algorithms BP and GA can collaborate, and combine their advantages. Heading to this way, two possible arrangements of these algorithms may be considered, as presented in the following sections.

2 HYBRID ALGORITHMS

In this section, two possible combinations of the BP and a GA, are presented. In these approaches, the aim is to investigate how these optimization algorithms can collaborate in order to increase the neural network training performance.

2.1 Backpropagation Followed by a Genetic Algorithm

This hybrid algorithm (BPGA) is a direct combination of the BP and a GA. As it can be seen from Figure 1, a Multilayer Perceptron is being trained by using the backpropagation algorithm, and the resulted suboptimal weight set inserts, as a chromosome to the GA's initial population. In the sequence, the GA having the suboptimal weight set in its candidate solutions tries to further optimize it, by finding the optimal network weights.

By using this structure, the BP capability in finding the global optimum can be studied, since it performs a local searching in the weight space, and also the ability of the GA to search an optimum solution globally.

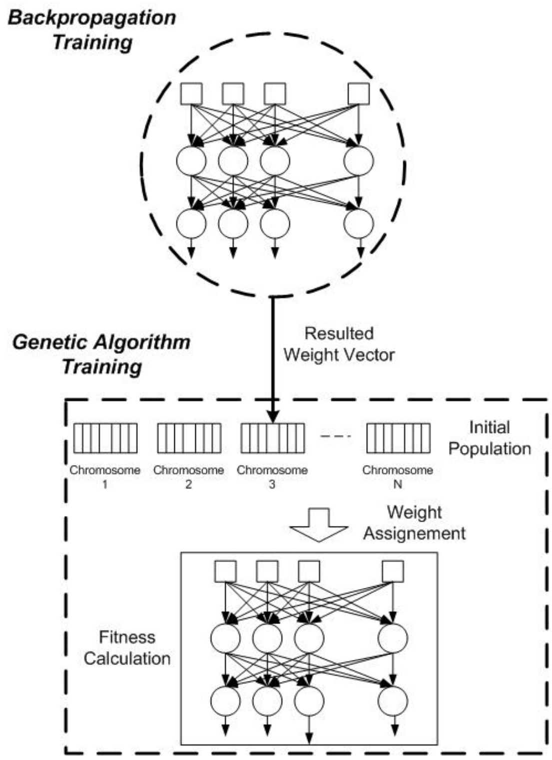


Figure 1 *Training procedure using hybrid algorithm BPGA*

By using this structure, the BP capability in finding the global optimum can be studied, since it performs a local searching in the weight space, and also the ability of the GA to search an optimum solution globally. The fitness value which describes the usefulness of each candidate solution is calculated by measuring the *Mean Squared Error (MSE)*. The *MSE* performance index is defined as follows,

$$MSE = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(y_{ij} - d_{ij}\right)^2$$

where M is the number of training patterns, N the number of network outputs and $(y_{ij} - d_{ij})$ the difference between the i^{th} network output and its corresponding desired value (target), when the j^{th} pattern appears to the network's input.

2.2 Genetic Algorithm Initializes the Backpropagation

Another approach for incorporate BP with GA is to use the GA to initialize the weights of the backpropagation (Liang and Dai, 1998). The BP is then used to train the multilayer neural network and the training *MSE* corresponds to the fitness value of the candidate initial weight set.

The block diagram of this algorithm (GABP) is illustrated in Figure 2, where it can be seen that the GA provides the initial weights of the backpropagation and so it searches an optimal initial weights set that gives minimum MSE.

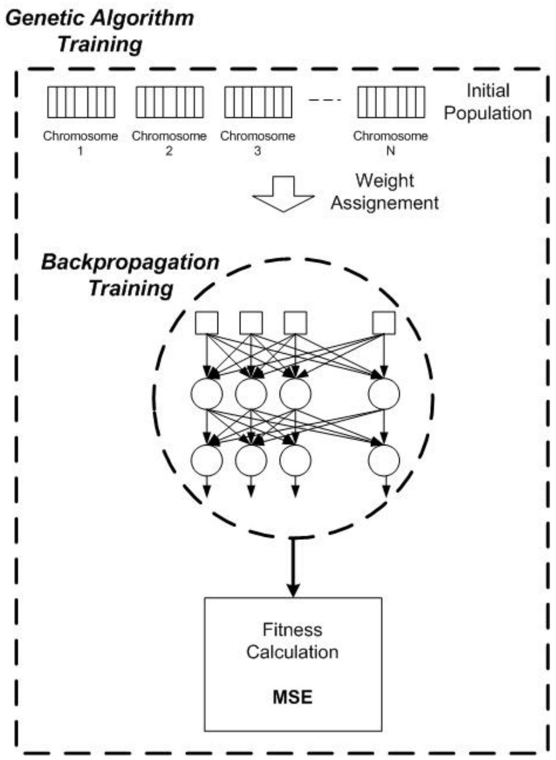


Figure 2 *Training procedure using hybrid algorithm GABP*

The GABP algorithm manages to make a hard tuning in the weight search space, by finding the best initial weights set by performing global searching. In the following the backpropagtion trains the neural network with optimal initial weights by making a fine tuning in a reduced weight space. In this structure it is attempted to exploit the ability of the genetic and backpropagation algorithms to global and local searching respectively.

3 EXPERIMENTAL RESULTS

For the experiments, a Simple Genetic Algorithm (SGA) is selected (Coley, 2001), in order to examine the optimization capabilities of the algorithm without using advanced diversity promotion methods. Additionally, the backpropagation initial weights are initialized with randomly generated values in the range [-1,1]. The calibration of the used SGA is presented in Table 1.

In the following a set of appropriate benchmark problems, is used to examine the training capabilities of each one of the hybrid algorithms, described in the previous sections. The performance of these algorithms is compared with those of the BP and GA which have been achieved, when applied separately in the training process.

Table 1 *Simple Genetic Algorithm settings.*

Population Size	50
Variables Range	[-5,5]
Maximum Generations	350
Elitism	YES, 2 chromosomes
Crossover Points	2 points
Crossover Probability	0.6
Mutation Probability	0.01
Selection Method	Stochastic Universal Approximation (SUS)

The next simulations are made in two different types of problems, classification and function approximation, problems.

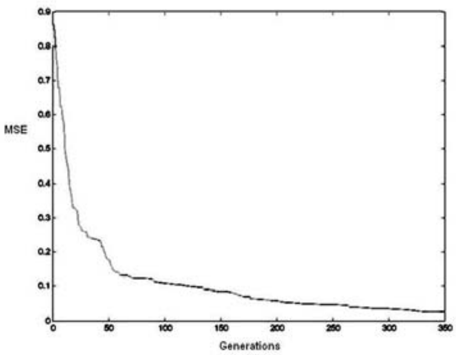
3.1 Classification Benchmark Problems

Next sections present four popular classification benchmark problems, which have been used in order to investigate the performance of the training topologies already discussed. The respective resulted classification rates are summarized in Table 2 for each of the algorithms.

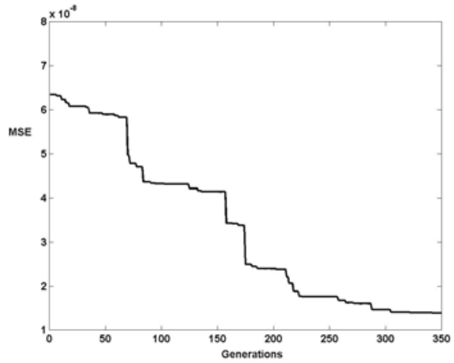
3.1.1 Iris Data

The classification of the Iris data set is a commonly pattern recognition task, for testing the efficiency of the neural structures. In order to realize this pattern classification test, a 4-10-3 multilayer neural network, is used. The four inputs correspond to the measurements that describe each one of the three classes, the tree types of Irish flowers. The Iris data consist of 135 patterns, which 90 randomly selected patterns constitute the training set and the remaining 45 the testing ones.

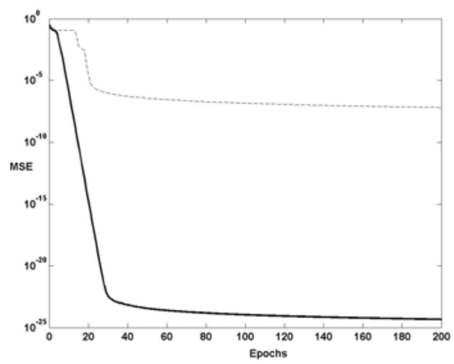
The following diagrams, describe the MSE variation through the iterations of each one of the algorithms, compared. As can be seen from these diagrams, the most efficient algorithm seems to be the GABP, since it outperforms even the standard backpropagation with random initial weight values. The simple genetic algorithm (SGA) achieves high MSE equal to 0.0063, the BP (6.3456×10^{-8}), the BPGA (1.4362×10^{-8}), and GABP (4.6026×10^{-25}).



(a)



(b)



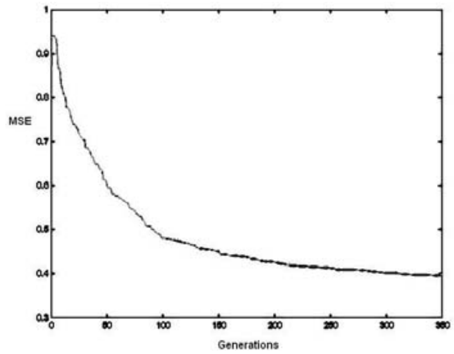
(c)

Figure 3 *Training process for the case of (a) SGA, (b) BPGA and (c) BP (dashed line) and GABP (solid line).*

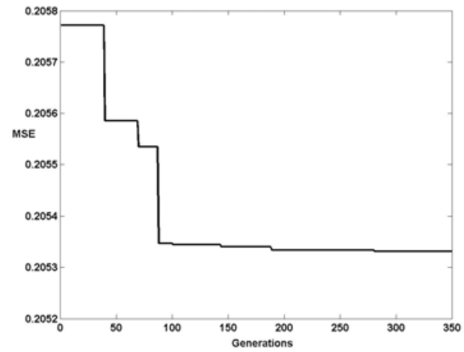
3.1.2 The Two Spirals Problem

The two spirals problem is an extremely hard problem for algorithms of the BP family to solve. In order to realize this pattern classification test, a 2-10-1 multilayer neural network, is used. The goal of this problem is to learn to discriminate between two sets of training points which lie on two distinct spirals in the x-y plane (Langlet et al., 2001).

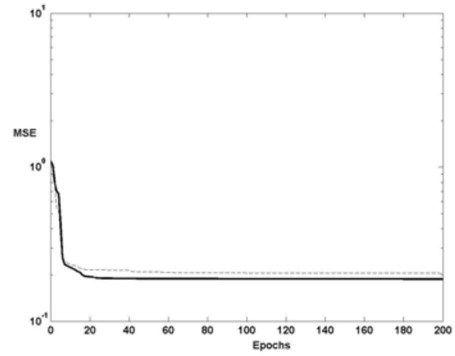
As can be seen from Figure 4, the GABP algorithm converges to a minimum MSE of 0.18793, very close to this of BP (0.20578), while the MSE of SGA (0.39560) and of BPGA (0.20533) are of the same orders.



(a)



(b)

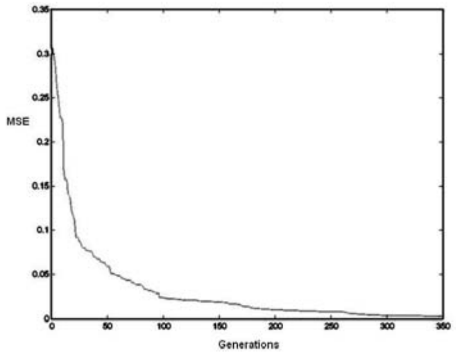


(c)

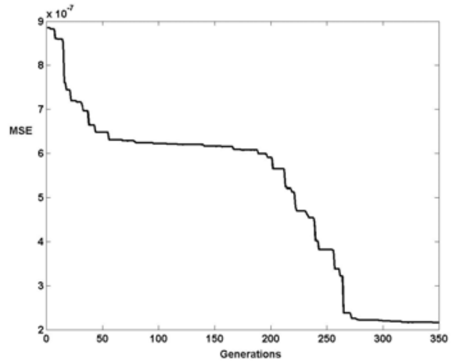
Figure 4 *Training process for the case of (a) SGA, (b) BPGA and (c) BP (dashed line) and GABP (solid line).*

3.1.3 4-parity Problem

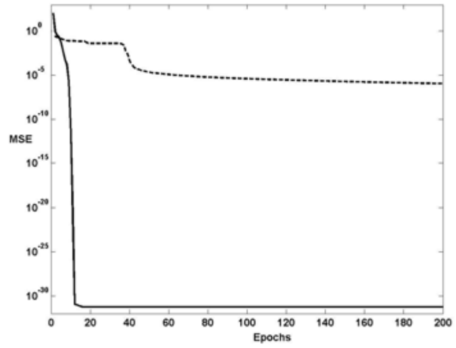
The *n*-parity problems are widely used as test data, for determining the performance of a training algorithm. These problems generates 2^n combinations of *n* bits as input data, while the output is a single bit equal to 1 if there is an odd number of high bits or 0 if there is an even number of high bits, in the input pattern (Looney, 1997). In order to realize this pattern classification test, a 4-10-1 multilayer neural network, is used. In this paper the 4-parity problem is selected as representative example.



(a)



(b)



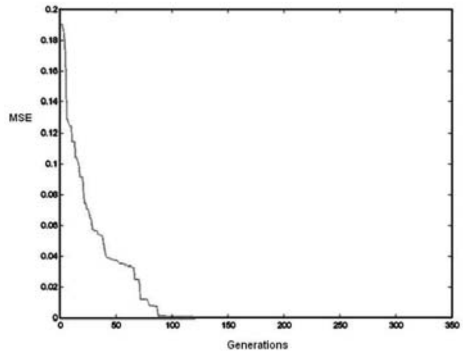
(c)

Figure 5 *Training process for the case of (a) SGA, (b) BPGA and (c) BP (dashed line) and GABP (solid line).*

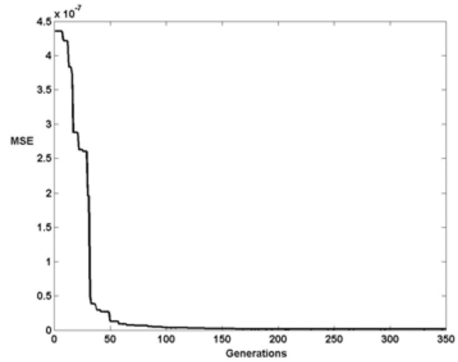
For the case of 4-parity problem, all the algorithms work well, SGA ($2.8879x10^{-3}$), BP ($8.8235x10^{-7}$), BPGA ($2.2557x10^{-7}$), where in the case of GABP ($5.6237x10^{-32}$) the MSE converges to the smallest level.

3.1.4 XOR Problem

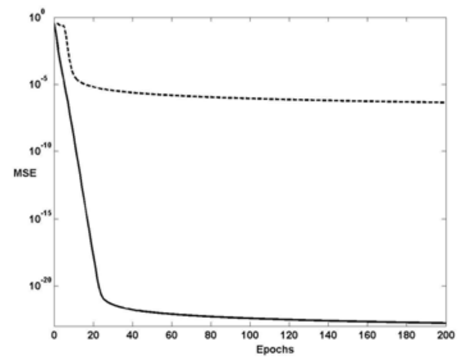
This problem corresponds to the 2-parity problem, and it is the most widely used to evaluate the performance of pattern recognition systems. In order to realize this pattern classification test, a 2-10-1 multilayer neural network, is used. It is an easy problem, for the neural network structure used to our simulations, since the MSE achieves small values for all the algorithms SGA ($1.30017x10^{-5}$), BP ($4.3530x10^{-7}$), BPGA ($2.3783x10^{-9}$) especially for the case of GABP ($1.6432x10^{-23}$), as depicted in Figure 6.



(a)



(b)



(c)

Figure 6 *Training process for the case of (a) SGA, (b) BPGA and (c) BP (dashed line) and GABP (solid line).*

Table 2 *Classification results.*

Algorithm	Benchmark Problems			
	Iris	Two Spiral	4-Parity	XOR
BP	100%	66%	100%	100%
SGA	100%	54%	100%	100%
BPGA	100%	58%	100%	100%
GABP	100%	66.8%	100%	100%

3.2 Function Approximation

It has been proved that Neural Networks are universal approximators (Cybenco, 1989; Hornik et al., 1989). This means that any arbitrary multivariable function can be approximated by an appropriate neural network. In this section, a one variable function is used as approximation target, in order to study the approximation capabilities of a neural network structure, trained by each one of the hybrid algorithms, already presented. For this reason, a 1-10-1 multilayer perceptron is being used.

3.2.1 Benveniste Function

A common function that is being used for function approximation tests is the so-called Benveniste function (Zhang and Benveniste, 1992). This single variable piecewise function is defined in [-10,10] as follows,

$$f(x)=\begin{cases} -2.186x-12.864 & , \quad -10\le x<-2 \\ 4.246x & , \quad -2\le x<0 \\ 10e^{-0.05x-0.5}\cdot\sin[(0.03x+0.7)x] & , \quad 0\le x\le 10 \end{cases}$$

The following diagrams figure that the MSE for each one of the algorithms were, SGA (3.2577), BP (3.5975×10^{-4}), BPGA (1.9856×10^{-4}) and GABP (1.1378×10^{-4}).

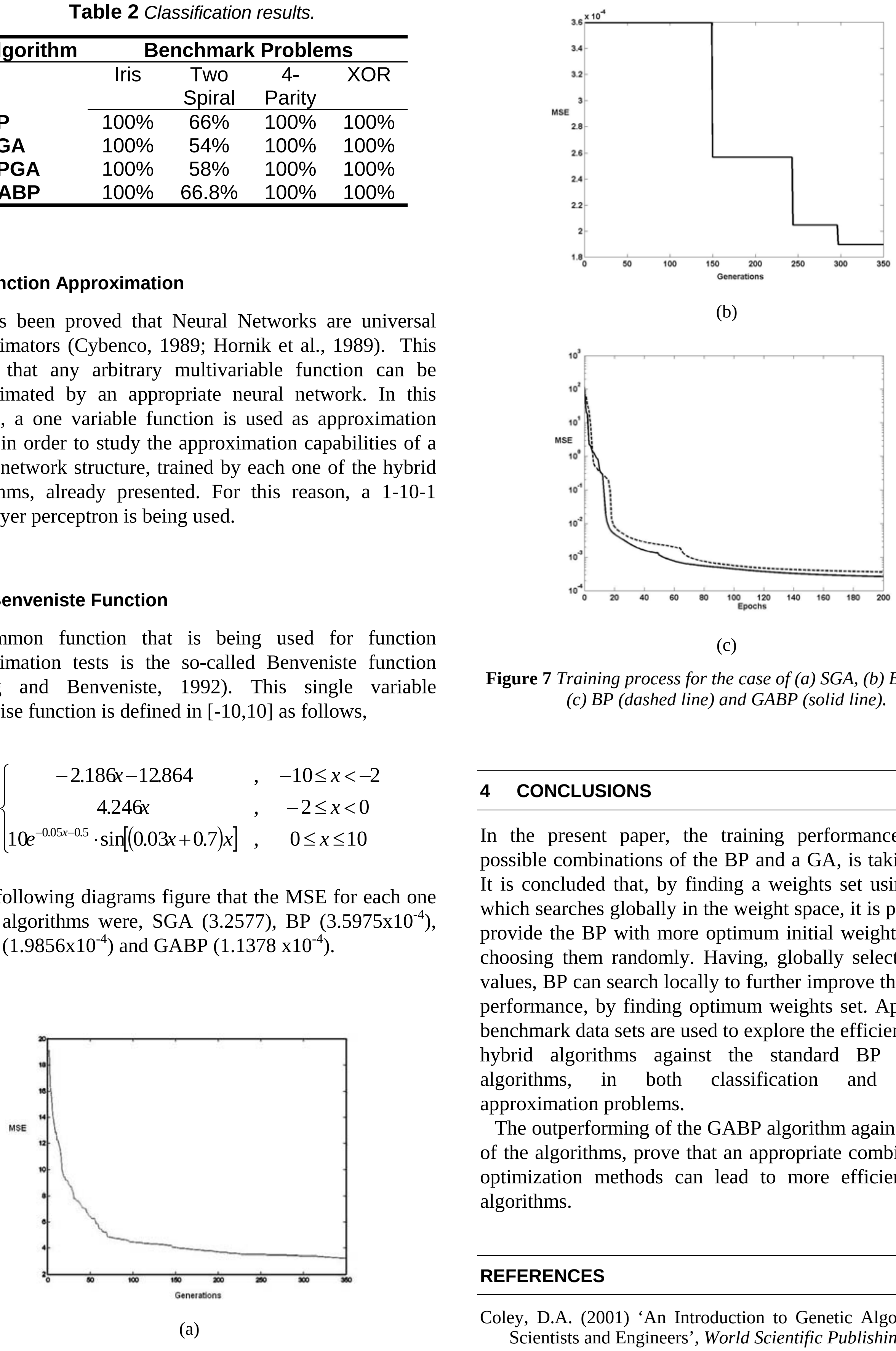


Figure 7 *Training process for the case of (a) SGA, (b) BPGA and (c) BP (dashed line) and GABP (solid line).*

4 CONCLUSIONS

In the present paper, the training performance of two possible combinations of the BP and a GA, is taking place. It is concluded that, by finding a weights set using a GA, which searches globally in the weight space, it is possible to provide the BP with more optimum initial weights than by choosing them randomly. Having, globally selected initial values, BP can search locally to further improve the training performance, by finding optimum weights set. Appropriate benchmark data sets are used to explore the efficiency of the hybrid algorithms against the standard BP and GA algorithms, in both classification and function approximation problems.

The outperforming of the GABP algorithm against the rest of the algorithms, prove that an appropriate combination of optimization methods can lead to more efficient hybrid algorithms.

REFERENCES

Coley, D.A. (2001) ‘An Introduction to Genetic Algorithms for Scientists and Engineers’, *World Scientific Publishing.*

Cybenco, G. (1989) ‘Approximation by Superpositions of a Sigmoidal Function’, *Mathematics of Control, Signals and System*, Vol.2, pp. 303-314.

Fahlman, S.E., (1988), 'An empirical study of learning speed in back-propagation networks', *Technical Report,* CMU-CS-88-162.

Harpham, C., Dawson, C.W. and Brown, M.R. (2004) 'A review of genetic algorithms applied to training radial basis function networks', *Neural Comput. & Applic.* Vol. 13, pp. 193-201.

Haykin, S. (1999) 'Neural Networks – A Comprehensive Foundation', *Prentice Hall, 2nd Edition*.

Haykin, S. (2000) 'Adaptive Digital Communication Receivers', *IEEE Communications Magazine*, December, pp. 106-114.

Hornik, K., Stinchcombe, M. and White, H., (1989) 'Multilayer Feedforward Networks are Universal Approximators', *Neural Networks*, Vol.2, pp. 359-366.

Langlet, F., Abdulkader, H., Roviras, D., Mallet, A. and Castanie, F. (2001), 'Comparison of Neural Network Adaptive Predistorsion Techniques for Satellite Down Links', *Proceedings of International Joint Conference on Neural Networks,* pp. 709-714.

Liang, H. and Dai, G. (1998), 'Improvement of Cascade Correlation Learning Algorithm with an Evolutionary Initialization', *Information Sciences* Vol. 112, pp. 1-6.

Looney, C.G. (1997), 'Pattern Recognition Using Neural Networks, Theory and Algorithms for Engineers and Scientists', *Oxford University Press.*

D.J. Montana, D.J. and Davis, L. (1989) 'Training Feedforward Neural Networks Using Genetic Algorithms', *In Proceedings of International Conference on Artificial Intelligence*, pp. 762-767.

Papakostas, G.A., Karras, D.A., Mertzios, B.G. and Boutalis, Y.S. (2005), 'An Efficient Feature Extraction Methodology for Computer Vision Applications using Wavelet Compressed Zernike Moments', *ICGST International Journal on Graphics, Vision and Image Processing, Special Issue: Wavelets and Their Applications,* Vol. SI1, pp.5-15.

Zhang Q. and Benveniste, A. (1992), 'Wavelet Networks', *IEEE Transactions on Neural Networks,* Vol. 3, No. 6, pp.889-898.