Neurocomputing 127 (2014) 4–12

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Investigating the use of alternative topologies on performance of the PSO-ELM

Elliackin M.N. Figueiredo, Teresa B. Ludermir *

Centro de Informatica, Universidade Federal de Pernambuco, Brazil

ARTICLE INFO

Article history:
Received 25 January 2013
Received in revised form
24 April 2013
Accepted 31 May 2013
Available online 26 August 2013

Keywords:
Extreme learning machine
Particle swarm optimization
ELM-PSO
PSO topology

ABSTRACT

In recent years, the Extreme Learning Machine (ELM) has been hybridized with the Particle Swarm Optimization (PSO) and such hybridization is called PSO-ELM. In most of these hybridizations, the PSO uses the Global topology. However, other topologies were designed to improve the performance of the PSO. In the literature, it is well known that the performance of the PSO depends on its topology, and there is not a best topology for all problems. Thus, in this paper, we investigate the effect of eight PSO topologies on performance of the PSO-ELM. The results showed empirically that the Global topology was more promising than all other topologies in optimizing the PSO-ELM according to the root mean squared error (RMSE) on the validation set in most of the evaluated datasets. However, no correlation was detected between this good performance on the RMSE and the testing accuracy.

1. Introduction

In 2004, Huang et al. [1] proposed an efficient training algorithm for single-hidden layer feedforward neural network (SLFN) called Extreme Learning Machine (ELM) that overcomes some disadvantages of gradient-based methods such as back-propagation and its variant the Levenberg–Marquardt [2]. In the ELM, first the input weights and hidden layer biases are randomly assigned and then the output weights are analytically determined through solving a linear system by using Moore–Penrose (MP) generalized inverse. Using this strategy, ELM not only can be thousands of times faster than traditional learning algorithms but it can also obtain SLFNs with better generalization performance [3,4]. ELM also avoids many difficulties faced by gradient-based learning methods such as stopping criteria, learning rate, learning epochs, and local minima [3,5].

However, ELM tends to require more hidden neurons than traditional tuning-based algorithms in many cases [3], which may lead ELM to respond slowly to unknown data [4]. Thus, in [3], Zhu et al. used an evolutionary algorithm to select the input weights of a SLFN and the Moore–Penrose generalized inverse to analytically determine the output weights. Using this method, they were able to achieve SLFNs more compact (with less hidden neurons) with good generalization performance. Particle Swarm Optimization (PSO) [6] has also been combined with ELM in order to improve the generalization capacity of the SLFNs. PSO has some advantages with respect to evolutionary algorithms [4]. PSO for example has no complicated operators as evolutionary algorithms and it has less parameters which need to be adjusted [7]. In 2006, Xu and Shu presented an evolutionary ELM based on PSO, PSO-ELM, and applied this algorithm in a prediction task [8]. In 2011, Han et al. [9] proposed an Improved Extreme Learning Machine, IPSO-ELM, that uses an improved PSO with adaptive inertia to select the input weights and hidden biases of the SLFN. However, unlike the PSO-ELM, the IPSO-ELM optimizes the input weights and hidden biases according to the RMSE on the validation set only (instead of the whole training set to avoid overfitting [3]) and the norm of the output weights. Thus, IPSO-ELM algorithm can obtain good performance with more compact and well-conditioned SLFN than other approaches of ELMs.

In all these works, the social topology used in the PSO is the Global. In Global topology, all particles are fulled connected. That is, each particle can communicate with every other particle. In this case, each particle is attracted towards the best solution found by the entire swarm. Besides of the Global topology, other topologies were designed to improve the performance of the PSO. As observed in [10], the performance of the PSO depends strongly on its topology and there is no outright best topology for all problems. An improper choice of the topology may lead the PSO to premature convergence or may lead to low search efficacy [11]. Thus, it is important to investigate the effect of the PSO topologies on the performance of the PSO-ELM in the training of SLFNs.

Studies on the effect of the PSO topology on the performance of the PSO in training neural networks have been carried out in [12]. Van Wyk and Engelbrecht investigated the overfitting in neural networks trained with the PSO based on three topologies: Global, Local and

* Corresponding author. Tel.: +55 8199981038; fax: +55 8121268430.
E-mail address: tbl@cin.ufpe.br (T.B. Ludermir).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.05.047

Von Neumann [13]. Cheng et al. [11] studied the effect of the topologies on the swarm diversity in several benchmark problems.

In a previous work [14] we studied the effect of five topologies (Global, Local, Von Neumann, Wheel and Four Clusters) on the performance of the PSO-ELM. In this paper, we carried out an extension of this previous work including more datasets (more classification datasets and function approximation problem) and two dynamic topologies, Clan and Multi-Ring. These new topologies are appropriate topologies for multi-modal problems [15,16].

The rest of this paper is organized as follows. Section 2 presents the ELM algorithm. Section 3 presents the PSO algorithm and the topologies used in this work. The experimental arrangement is detailed in Section 4. Section 5 discusses the experimental results. Finally, Section 6 summarizes our conclusions.

2. Extreme learning machine

Given N distinct training samples $(\mathbf{x}_i, \mathbf{t}_i) \in \Re^n \times \Re^m$, where $\mathbf{x}_i$ is the input pattern i such that $\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T$ and $\mathbf{t}_i$ is the target i such that $\mathbf{t}_i = [t_{i1}, t_{i2}, \ldots, t_{im}]^T$. A SLFN with K hidden nodes and activation function $g(\cdot)$ can approximate these N samples with zero error. This means that

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T}, \tag{1}$$

where $\mathbf{H} = \{h_{ij}\}$ $(i = 1, \ldots, N$ and $j = 1, \ldots, K)$ is the hidden layer output matrix, $h_{ij} = g(\mathbf{w}_j^T \cdot \mathbf{x}_i + b_j)$ denotes the output of jth hidden node with respect to $\mathbf{x}_i$; $\mathbf{w}_j = [w_{j1}, w_{j2}, \ldots, w_{jn}]^T$ denotes the weight vector connecting the jth hidden node and the input nodes, and b_j is the bias (threshold) of the jth hidden node; $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_K]^T$ is the matrix of output weights and $\boldsymbol{\beta}_j = [\beta_{j1}, \beta_{j2}, \ldots, \beta_{jm}]^T$ $(j = 1, \ldots, K)$ is the weight vector connecting the jth hidden node and the output nodes; $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_N]^T$ denotes the matrix of targets.

ELM works as follows. Initially the ELM generates the input weights and hidden biases randomly. Next, the determination of the output weights, linking the hidden layer to the output layer, consists in finding simply the least-square solution to the linear system given by Eq. (1). This solution is given by the following equation:

$$\hat{\beta} = \mathbf{H}^{\dagger}\mathbf{T}, \tag{2}$$

where $\mathbf{H}^{\dagger}$ is the Moore–Penrose (MP) generalized inverse of matrix $\mathbf{H}$.

3. Particle swarm optimization

This section presents the basic algorithm of Particle Swarm Optimization and the information sharing topologies.

3.1. Basic concepts

The Particle Swarm Optimization (PSO) is a population-based stochastic optimization technique developed by Kennedy and Eberhart in 1995 [6]. PSO is inspired by the social behavior of biological organisms such as birds in a flock [4]. In PSO, each particle represents a candidate solution within a n-dimensional search space. The position of a particle i at iteration t is denoted by $\mathbf{x}_i(t) = [x_{i1}, x_{i2}, \ldots, x_{in}]$. At every iteration of the PSO, each particle moves through the search space with a velocity $\mathbf{v}_i(t) = [v_{i1}, v_{i2}, \ldots, v_{in}]$ calculated as follows:

$$v_{ij}(t+1) = wv_{ij}(t) + c_1 r_{j1}[y_{ij}(t) - x_{ij}(t)] + c_2 r_{j2}[\hat{y}_{ij}(t) - x_{ij}(t)], \tag{3}$$

where $j \in [1, 2, \ldots, n]$ is a dimension of the search space, w is the inertia weight, $\mathbf{y}_i(t)$ is the personal best position of the particle i at iteration t and $\hat{\mathbf{y}}(t)$ is the global best position of the swarm iteration t. The personal best position is named *pbest* and it represents the best position found by the particle during the search process until the iteration t. The global best position is named *gbest* and it represents the best position found by the entire swarm until the iteration t. The parameters c_1 and c_2 are acceleration coefficients and the terms r_{j1} and r_{j2} are random numbers sampled from an uniform distribution $U(0, 1)$. The velocity is limited to the range $[\mathbf{v}_{min}, \mathbf{v}_{max}]$. After updating velocity, the new position of the particle i at iteration $t+1$ is calculated using the following equation:

$$\mathbf{x}_i(t+1) = \mathbf{x}_i(t) + \mathbf{v}_i(t+1). \tag{4}$$

In [17], Shi and Eberhart proposed an adaptive inertia in which the parameter w reduces gradually as the iteration increases according to the following equation:

$$w(t) = w_{max} - t \times \frac{(w_{max} - w_{min})}{t_{max}}, \tag{5}$$

where w_{max} is the initial inertia weight, w_{min} is the final inertia weight and t_{max} is the maximum number of iterations.

Algorithm 1 summarizes the PSO.

Algorithm 1. PSO Algorithm.

```
Initialize the swarm S in the search space;
while maximum number of iterations is not reached do
    for all particle i of the swarm do
        Calculate the fitness of the particle i
        Set the personal best position y_i(t);
    end for
    Set the global best position ŷ(t) of the swarm;
    for all particle i of the swarm do
        Update the velocity v_i(t);
        Update the position x_i(t);
    end for
end while
```

3.2. Topologies

The PSO described above is known as gbest PSO. In this PSO, each particle is connected to every other particle in the swarm in a type of topology referred to as Global topology. Different topologies have been developed for PSO and empirically studied [10], each affecting the performance of the PSO in a potentially drastic way [13,18]. Some of the most common topologies are [11,10]:

Global or Star: In Global topology, the particles are fully connected and each particle communicates with every other particle of the swarm as illustrated in Fig. 1a. Thus, each particle is attracted towards the best solution found by the entire swarm. In this topology particles can spread information quickly through the swarm and it has faster convergence than other topologies, but with a susceptibility to be trapped in local minima [10].

Local or Ring: The Local topology was proposed as a way to deal with problems in which the Global topology fails. That is, problems that contain cliffs, variable interactions, and other features that are not typified by smooth gradients [19]. In the Local topology, each particle communicates with its n_N immediate neighbors. Thus, instead of a global best particle, each particle selects the best particle in its neighborhood. When $n_N = 2$, each particle within the swarm communicates with its immediately adjacent neighbors as illustrated in Fig. 1b. The particles are typically included in their own neighborhood. Thus, they may influence themselves, if they have found the best problem solution in the neighborhood so far [19,10]. Using this topology, different regions of the search space can be explored simultaneously [16].

Von Neumann: In Von Neumann topology, each particle is connected to its left, right, top and bottom neighbors on a two dimensional lattice. For example, in a swarm with 20 particles, the

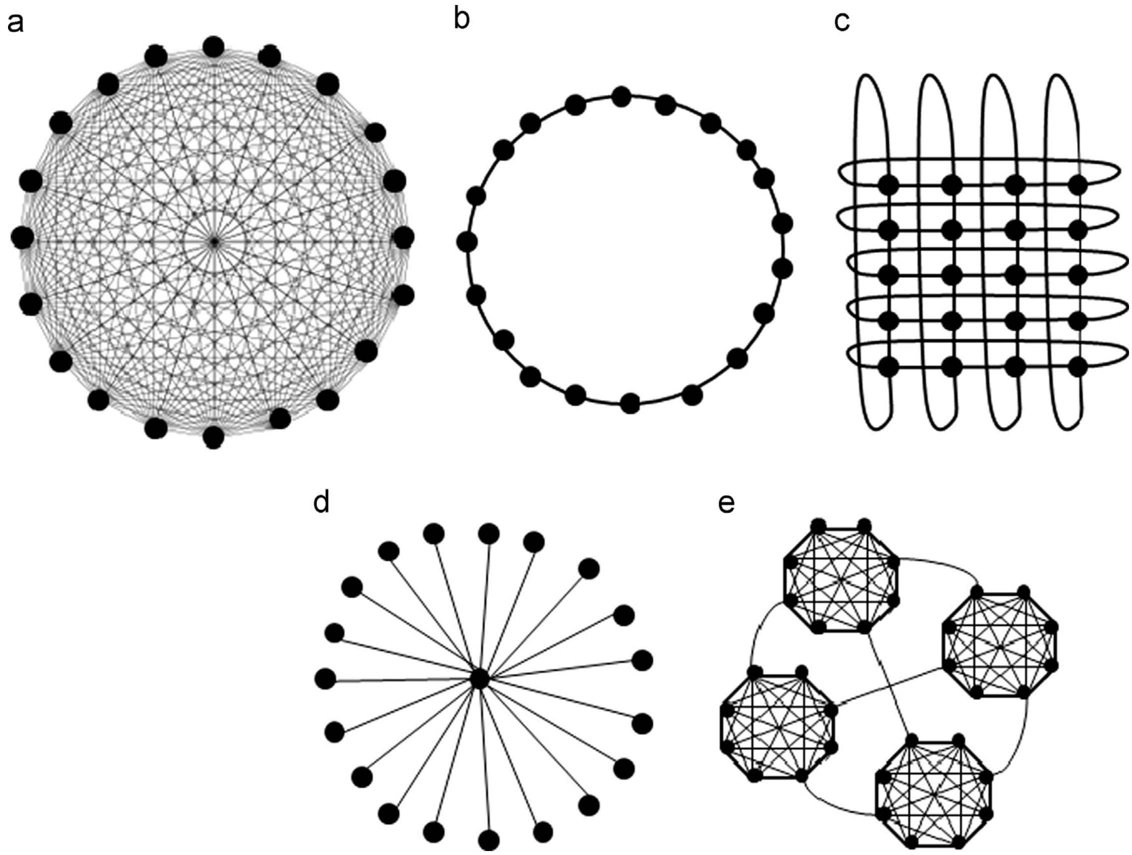


Fig. 1. Topologies of the PSO: (a) Global, (b) Local, (c) Von Neumann, (d) Wheel, (e) Four Cluster.

particles are arranged in a grid 5×4 whose edges are wrapped [19]. In some studies, this topology has shown to be more promising than other topologies [10].

Wheel: In Wheel topology, the particles are isolated from one another and one particle serves as the focal point in which all information flow occurs, as illustrated in Fig. 1d. In this paper, the focal particle was chosen randomly.

Four Clusters: In Four Clusters topology, the swarm is divided into four subgroups in which each subgroup is a star topology. Each subgroup has three particles connected to the three other groups as illustrated in Fig. 1e.

The neighborhood of the particles in all topologies is based on their indexes. For a more detailed description of these topologies, the reader may consult [10].

All topologies aforementioned are static topologies, that is, the neighborhood of each particle does not change during the search process. More recently dynamics topologies have been proposed in the literature. For example, Bastos-Filho et al. [15] proposed a topology named Multi-Ring [15] and Carvalho et al. [16] proposed a topology named Clan. In all these topologies, the neighborhood of the particles varies over the iterations. These topologies are explained hereafter:

Clan Topology: In Clan topology, the swarm is divided into subswarms named clans and each clan has a fully-connected structure (gbest topology). Fig. 2a shows a swarm divided into four clans. In each iteration of the PSO, each clan performs a search and the particle that had reached the best solution of the entire clan is chosen as the leader for its clan. Once leaders are determined, a process named conference occurs. In the conference, the leaders generate a new swarm using a Global topology and adjust their positions using this topology. Fig. 2b illustrates the conference process of the leaders. In Fig. 2b, the leaders (in gray) are placed together in the same swarm and the leaders need to decide which of them is the gbest. The conference process uses only the leaders of each clan to perform a new PSO search. The leaders attending the conference may change in the course of the search process.

The conference can also be performed using the Local topology instead of Global topology. Fig. 3 illustrates a conference among leaders using the Local topology. In this paper, three clans were used as in [16] in the Clan-Global and Clan-Local topologies.

Multi-Ring Topology: The Multi-Ring is inspired in the ring topology. This topology is formed by layers stacked one over the other using the ring topology as illustrated in Fig. 4. Fig. 4 shows an example of this topology with n layers. Each layer has the same number of particles. The Multi-Ring topology is similar to the Von Neumann topology but the first layer does not communicate with the nth layer. Thus, let k_i be ith particle in the layer k then the neighbors of the particle k_i are $\{k_{i-1}, k_{i+1}, (k+1)_i, (k-1)_i\}$ if $1 < k < n$. Otherwise, the neighbors of the particle k_i are $\{k_{i-1}, k_{i+1}, (k+1)_i\}$ if $k=1$ or $\{k_{i-1}, k_{i+1}, (k-1)_i\}$ if $k=n$.

During the search process, the particles of same layer may stagnate in many regions of the search space. This fact in turn may lead the stagnation of the entire swarm [15]. In order to try to solve this problem, the Multi-Ring topology has a strategy in which the layers can rotate and thus to change the neighborhood of the particles. The rotation process occurs as follows. If layer k does not improve its own best solution, this layer is rotated. Thus, each particle in this layer has its index modified to $i=(i+d)\ \mathrm{mod}(nl)$, where d is the rotation distance and nl is the number of particles in a layer. For illustrating the rotation process, consider Fig. 5. Initially, particle E communicates with particles $\{D, F, B, H\}$. After the rotation process, the neighborhood of particle E is modified to $\{D, F, A, G\}$. This process improves the information flow in the swarm and hence it also improves the convergence capacity of the swarm. It is important to note that not only particle E changes its neighborhood, but all

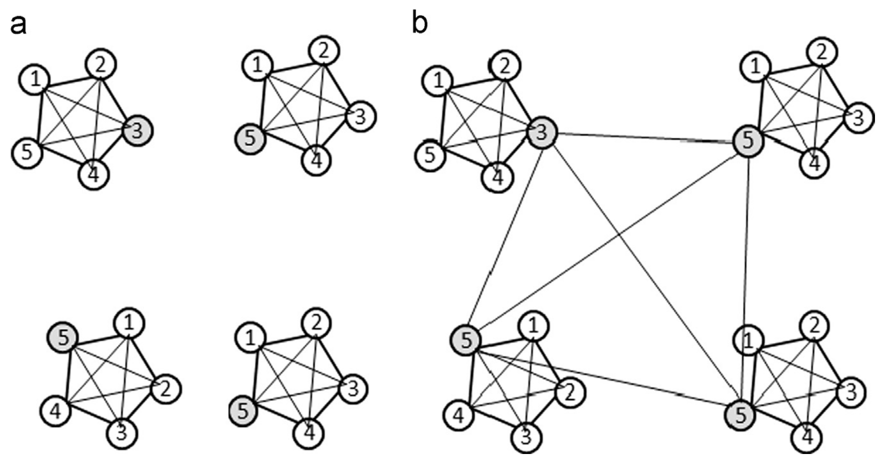


Fig. 2. Clan topology: (a) individual clans, (b) conference of the leaders with Global topology.

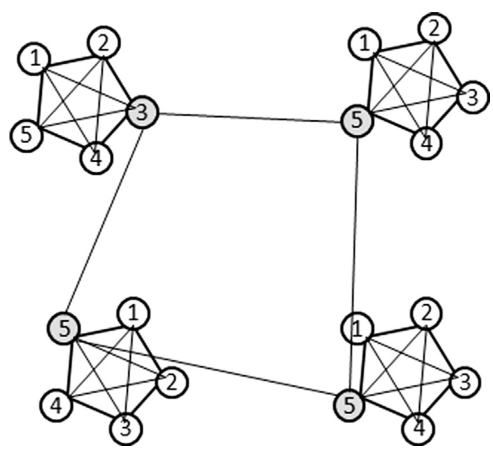


Fig. 3. Conference of the leaders with local topology.

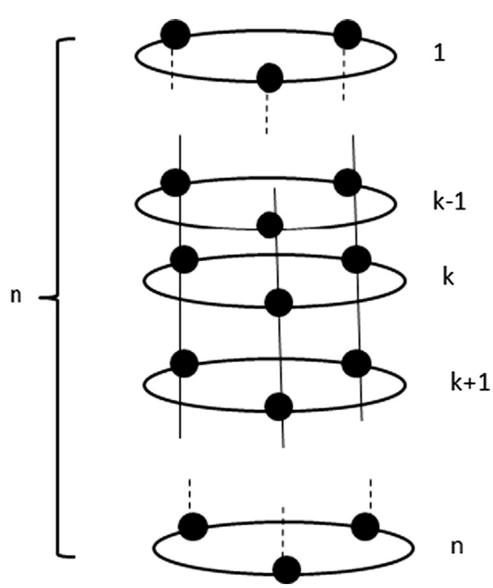


Fig. 4. Multi-Ring topology.

other particles of the same layer change too. The rotation process starts when the best solution of the layer does not change during t_r iterations. In this paper, we used nl=4 (i.e., the number of rings was 5), t_r=20 and $d=nl/2$. The value $d=nl/2$ was chosen according to the recommended by [15].

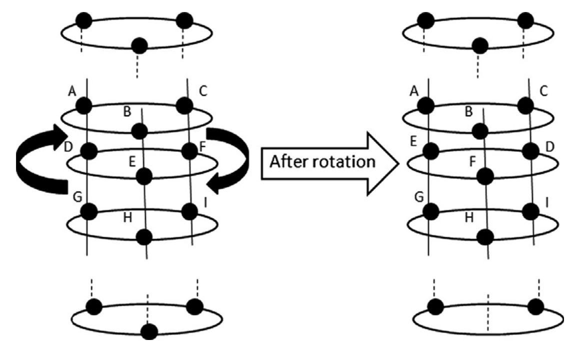


Fig. 5. Rotation skill example.

4. Experimental arrangement

4.1. PSO-ELM description

The combination of the PSO with the ELM used in this paper was defined as follows. The PSO was used to select the input weights and hidden biases of the SLFN, and the MP generalized inverse is used to analytically calculate the output weights using a training set. Thus, each particle in the swarm represents the input weights and the hidden biases. For example, the particle i is given by $\mathbf{x}_i=[w_{11}, w_{12}, \ldots w_{1n}, \ldots w_{21}, w_{22}, \ldots w_{2n}, \ldots, w_{K1}, w_{K2}, \ldots w_{Kn}, b_1, b_2, \ldots, b_K]$. The fitness of each particle is adopted as the root mean squared error (RMSE) on the validation set. We used the RMSE because it is a continuous metric thus the search process can be conducted smoothly. The RMSE on the validation set was also used in [4] both to prediction and classification problems. In the SLFN trained by ELM, the activation function was the sigmoid function given by the following equation:

$$g(v)=\frac{1}{1+e^{-v}}. \tag{6}$$

4.2. Performance metrics

In order to evaluate the performance of PSO-ELM, two metrics are used: the root mean squared error (RMSE) over the validation set referred to as $RMSE_V$ and the testing accuracy denoted by TA. The testing accuracy TA refers to the percentage of correct classifications produced by the trained SLFNs on the testing set and it is given by the following equation:

$$TA=100\times\frac{n_c}{n_T}, \tag{7}$$

where n_T is the size of the testing set and n_c is the number of correct classifications in this set.

The RMSE on the validation set of size N is calculated using the following equation:

$$RMSE_V = \sqrt{\frac{\sum_{i=1}^{N}\sum_{j=1}^{m}(t_{i,j}-o_{i,j})^2}{N \times m}}, \tag{8}$$

where m is the number of output units in the SLFN, $t_{i,j}$ is the target to the pattern i in the output j, $o_{i,j}$ is the output obtained by the network to the pattern i in the output j and N is the number of samples.

Other metric used in this work is the global swarm diversity D_S [20] that is calculated as follows:

$$D_S = \frac{1}{|S|}\sum_{i=1}^{|S|}\sqrt{\sum_{k=1}^{I}(x_{ik}-\overline{x}_k)^2}, \tag{9}$$

where $|S|$ is the swarm size, I is the dimensionality of the search space, $\mathbf{x}_i$ and $\overline{\mathbf{x}}$ are particle position i and the swarm center, respectively. This metric is used to measure the convergence of the PSO. In this metric, a small value indicates swarm convergence around the swarm center while a large value indicates a higher dispersion of particles away from the center.

4.3. Problem sets

In order to evaluate the effect of the several topologies on the performance of the PSO-ELM, we used nine benchmark classification problems and one regression problem. The classification datasets are obtained from UCI Machine Learning Repository [21] and they present different degrees of difficulties and different number of classes. The classification problems and chosen sizes of the hidden layer are given in Table 1. The sizes of the hidden layer used were the same as [22] for the problems Cancer, Diabetes, Glass, Heart, Iris, Ecoli, and as [23] for the problems Ionosphere, Sonar, Vehicle. For all the datasets, the data was treated as presented in [22], that is, the attributes have been normalized into the range [0,1], while the outputs (targets) have been normalized into [−1,1]. Each dataset was divided into training, validation and testing sets, as specified in Table 2. This division was based on [22,23] depending on the dataset. For all topologies, 30 independent executions were done for each dataset. The training, validation and testing sets were randomly generated at each trial of simulations. All topologies were executed on the same sets generated.

4.4. PSO configuration

In this paper the parameters used in the PSO were defined as follows. The coefficients c_1 and c_2 were both set to 2.0 and the adaptive inertia was used where the initial inertia is 0.9 and the end inertia is 0.4. All components are limited within the range [−1, 1] and we used $v_{max}=1$ and $v_{min}=-1$ for all dimensions of the search space. The swarm size used was 20. A small number of particles was used because we wanted to observe the effect of topology on information sharing. If we use a very large number of particles as in [9,4,3], the effect of the topology on the communication of the particles could be hidden. The PSO was executed for 1000 iterations. We use many iterations because we wanted to provide enough time for the exchange of information between the particles for all topologies.

Table 1
Specification of the classification problems used in the experiments.

Problem	Attributes	Classes	Hidden nodes
Cancer	9	2	5
Pima Indians Diabetes	8	2	10
Glass Identification	9	6	10
Statlog Heart	13	2	5
Iris	4	3	10
Ecoli	7	8	20
Ionosphere	34	2	20
Sonar	60	2	20
Vehicle	18	4	20

Table 2
Partition of the datasets used in the experiments.

Problem	Training	Validation	Testing	Total
Cancer	350	175	174	699
Pima Indians Diabetes	252	258	258	768
Glass Identification	114	50	50	214
Statlog Heart	130	70	70	270
Iris	70	40	40	150
Ecoli	180	78	78	336
Ionosphere	175	88	88	351
Sonar	104	52	52	208
Vehicle	423	212	212	846

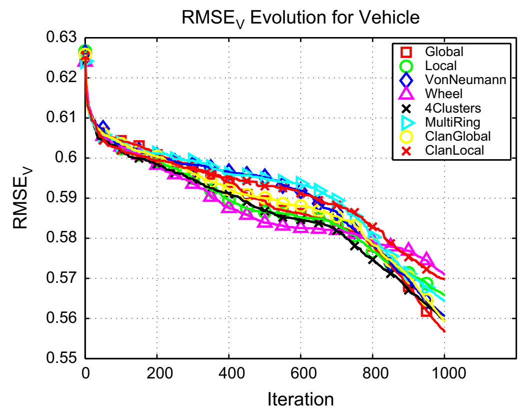


Fig. 6. $RMSE_V$ over time on the Vehicle dataset. Graph shows mean over 30 runs.

5. Experiment results and discussion

This section presents the experimental results and analyses of the effect of the PSO topology on performance of the PSO-ELM. First we analyzed the evolution of the $RMSE_V$ along the iterations in order to observe the behavior of the PSO (Section 5.1). Due to space limitations, we restrict this analysis only to the Vehicle dataset. Vehicle is considered one of the most difficult dataset of the UCI repository. Then we show the performance of the eight topologies on the nine datasets using the three metrics: $RMSE_V$, TA and D_S (Section 5.2). However, it is necessary to assess whether the performance of the topologies is statistically different on all datasets. Thus, we performed the one-way analysis of variance (ANOVA) with 5% of significance on all experiments (Section 5.3).

5.1. Evolution of the $RMSE_V$ for Vehicle dataset

In this section, we analyzed the evolution of the $RMSE_V$ along the iterations in order to observe the behavior of the PSO. The objective of this section is to verify if the PSO works as expected. Fig. 6 shows the evolution of the $RMSE_V$ on the eight topologies investigated. This graph corresponds to the mean over 30 samples. As can be seen from the graph, all topologies had similar behavior up to approximately 200th iteration. After this iteration, the decrease in the $RMSE_V$ occurred differently for each topology up to 700th iteration. In particular, the Wheel topology reached lower $RMSE_V$ in this stage than with other topologies. After the 700th iteration, the PSO begins the exploitation process. Again, this happens in a different way for each topology. The Wheel topology, for example, starts the exploitation stage later than other topologies. On the other hand, the other topologies start the exploitation stage approximately at the same time. By the end of the search process, the Wheel topology reached a worse $RMSE_V$ than other topologies. On the other hand, the Global topology reached the best $RMSE_V$ value for this problem.

Table 3
Performance of the eight topologies on datasets Vehicle, Ionosphere and Sonar.

Topology	$RMSE_V$	TA (%)	D_S
Vehicle			
Global	**0.5565** ± 0.0174	77.49 ± 2.22	**3.5563** ± 0.9151
Local	0.5659 ± 0.0118	77.68 ± 2.64	7.7824 ± 0.2122
Von Neumann	0.5606 ± 0.0153	77.60 ± 2.12	6.8379 ± 0.8367
Wheel	0.5709 ± 0.0163	77.57 ± 2.80	5.9935 ± 0.8497
Four Clusters	0.5596 ± 0.0136	77.88 ± 2.49	7.2259 ± 0.3611
Multi-Ring	0.5642 ± 0.0125	77.95 ± 2.28	7.4822 ± 0.5445
Clan Global	0.5592 ± 0.0149	**78.18** ± 2.54	7.8553 ± 0.3691
Clan Local	0.5697 ± 0.0137	77.36 ± 2.83	8.2825 ± 0.3247
Ionosphere			
Global	**0.4196** ± 0.0526	88.45 ± 4.08	**4.8252** ± 1.0642
Local	0.4493 ± 0.0495	**88.90** ± 3.53	10.7470 ± 0.2648
Von Neumann	0.4285 ± 0.0464	88.03 ± 3.97	9.3085 ± 1.0479
Wheel	0.4787 ± 0.0633	88.67 ± 3.67	8.2832 ± 0.8108
Four Clusters	0.4370 ± 0.0575	88.83 ± 3.07	9.9415 ± 0.4645
Multi-Ring	0.4381 ± 0.0548	88.75 ± 4.21	10.3991 ± 0.4644
Clan Global	0.4239 ± 0.0486	89.32 ± 3.45	11.0203 ± 0.4039
Clan Local	0.4689 ± 0.0498	88.64 ± 3.24	11.3703 ± 0.4413
Sonar			
Global	**0.3242** ± 0.0588	76.99 ± 6.03	**4.2517** ± 1.2101
Local	0.4067 ± 0.0591	76.03 ± 6.64	13.8073 ± 0.2837
Von Neumann	0.3780 ± 0.0462	75.90 ± 5.57	10.9651 ± 1.5757
Wheel	0.4408 ± 0.0926	75.90 ± 5.64	9.7381 ± 1.9339
Four Clusters	0.3548 ± 0.0546	74.87 ± 5.67	12.2801 ± 0.7902
Multi-Ring	0.3935 ± 0.0523	**77.69** ± 6.01	12.8096 ± 0.8124
Clan Global	0.3431 ± 0.0435	76.73 ± 7.66	13.8341 ± 0.4714
Clan Local	0.4524 ± 0.0403	75.06 ± 5.59	14.6731 ± 0.3839

Table 4
Performance of the eight topologies on datasets Cancer, Diabetes and Ecoli.

Topology	$RMSE_V$	TA (%)	D_S
Cancer			
Global	**0.2898** ± 0.0418	96.38 ± 1.28	**0.7787** ± 0.1867
Local	0.3008 ± 0.0391	96.40 ± 1.32	3.0384 ± 0.1396
Von Neumann	0.2954 ± 0.0417	96.36 ± 1.34	1.9405 ± 0.4816
Wheel	0.3016 ± 0.0407	96.44 ± 1.40	1.3158 ± 0.4965
Four Clusters	0.2980 ± 0.0400	96.40 ± 1.46	2.6360 ± 0.2826
Multi-Ring	0.2963 ± 0.0402	96.30 ± 1.45	2.3937 ± 0.3064
Clan Global	0.2903 ± 0.0394	96.34 ± 1.33	3.1545 ± 0.2891
Clan Local	0.2973 ± 0.0389	**96.48** ± 1.30	2.9577 ± 0.4175
Diabetes			
Global	0.7600 ± 0.0261	76.65 ± 2.19	**2.0740** ± 0.4341
Local	0.7620 ± 0.0225	76.65 ± 1.92	3.9642 ± 0.1577
Von Neumann	0.7576 ± 0.0225	76.45 ± 1.79	3.5954 ± 0.4541
Wheel	0.7649 ± 0.0259	76.49 ± 2.07	2.8312 ± 0.4015
Four Clusters	0.7572 ± 0.0243	76.67 ± 1.95	3.7435 ± 0.2525
Multi-Ring	0.7594 ± 0.0232	**77.12** ± 2.18	3.8673 ± 0.2629
Clan Global	**0.7571** ± 0.0224	76.58 ± 1.96	4.0092 ± 0.2314
Clan Local	0.7593 ± 0.0237	76.41 ± 2.10	4.1290 ± 0.2733
Ecoli			
Global	0.3606 ± 0.0580	84.23 ± 3.97	**3.2578** ± 0.4549
Local	**0.3506** ± 0.0404	**84.62** ± 4.31	5.5475 ± 0.2585
Von Neumann	0.3508 ± 0.0451	84.19 ± 4.32	5.0891 ± 0.3584
Wheel	0.3559 ± 0.0416	83.97 ± 4.49	4.3296 ± 0.6231
Four Clusters	0.3617 ± 0.0569	83.93 ± 4.22	5.3475 ± 0.3268
Multi-Ring	0.3592 ± 0.0575	84.23 ± 3.90	5.3875 ± 0.3091
Clan Global	0.3593 ± 0.0551	84.57 ± 3.92	5.4900 ± 0.2691
Clan Local	0.3592 ± 0.0471	83.97 ± 4.88	5.5709 ± 0.2576

5.2. Performance of topologies on classification datasets

The main objective of the experiments is to investigate the influence of different topologies on the performance of a PSO-ELM. The results of the experiments are organized in tables for the datasets. The $RMSE_V$, TA and D_S are emphasized in bold to indicate the best value obtained among the topologies and the results are for 30 independent executions. Table 3 shows the results for Vehicle, Ionosphere and Sonar datasets in which the Global topology reached the smallest $RMSE_V$ and D_S on the three datasets.

Tables 4 and 5 show the results for the other six datasets. As can be seen from these three tables, the Global topology reached the best $RMSE_V$ in five datasets (Vehicle, Ionosphere, Sonar, Cancer and Heart). For the other datasets the best topologies were Local, Von Neumann, Four Clusters and Clan-Global. In general, we observed that in problems with more than four classes (Glass and Ecoli), the topologies less connected (Von Neumann and Local) presented better $RMSE_V$ than other topologies. On the other hand, in problems with few classes (≤ 4), the topologies more connected (Global, Clan-Global, Four Clusters) presented better $RMSE_V$ than other topologies. In terms of diversity, the Global topology reached lower values than other topologies on all datasets, this means that the particles, in the end of the algorithm, are near each other around the center of swarm what may indicate convergence of the swarm. This result is according to what we expected from this topology and a similar result was obtained by [13] in the case in which neural networks are fully trained by PSO. From an empirical analysis of the experimental results, we may conclude that the performance of the PSO with respect to $RMSE_V$ (that is the fitness function) depends on its topology. This analysis is consistent with the literature [10,11,15,16,19] that says that the topologies have different performances depending on the problem. However, it is important to say that the $RMSE_V$ is statistically similar for the topologies Global, Von Neumann, Four Clusters and Clan-Global as we are going to see in the next section.

5.3. Statistical analysis of experimental results on classification datasets

So far our analysis on the $RMSE_V$ and TA was empirical. However, it is necessary to assess whether the performances of the topologies are statistically different on all datasets. Thus, we performed the one-way analysis of variance (ANOVA). To use ANOVA is necessary to know if the $RMSE_V$ samples came from a normal distribution for all topologies for each dataset. Thus, we used the normality test known as Lilliefors test with 1% of significance for all topologies in each dataset. We conclude that all topologies come from a normal distribution for all datasets except to the Ecoli dataset. Moreover, for each dataset, it is necessary to verify whether the $RMSE_V$ samples of the different topologies have the same variance. We used Bartlett's test (Matlab function vartestn) with 5% of significance to verify this hypothesis. Thus we conclude that the variances of the distributions of all topologies are the same for all datasets except to Sonar dataset. Therefore we performed the ANOVA analyses in all datasets except to the datasets Ecoli and Sonar because in these datasets the two hypotheses for using the ANOVA were not satisfied. We used the ANOVA with 5% of significance (Matlab function anova). We verify that the means of all topologies are equal in each dataset except to the datasets Vehicle and Ionosphere.

In order to find out what are the different topologies, we performed a multiple comparison analysis with 5% of significance (Matlab function multcompare). Fig. 7 shows the confidence intervals with 95% of confidence of the topologies for Vehicle dataset. As can be seen from this figure, there are a significance difference between the Global topology and the Wheel topology with respect to $RMSE_V$.

Fig. 8 shows the confidence intervals with 95% of confidence of the topologies for the Ionosphere dataset. In this dataset, the Global, Von Neumann and Clan Global topologies are better than the Wheel topology.

Table 5
Performance of the eight topologies on datasets Glass, Heart and Iris.

Topology	$RMSE_V$	TA (%)	D_S
Glass			
Global	0.5460 ± 0.0324	**64.27** ± 5.50	**2.2269** ± 0.4757
Local	0.5539 ± 0.0298	63.33 ± 5.74	4.2742 ± 0.1323
Von Neumann	**0.5441** ± 0.0266	63.67 ± 6.10	3.6571 ± 0.3624
Wheel	0.5573 ± 0.0275	62.07 ± 6.07	3.1868 ± 0.5424
Four Clusters	0.5473 ± 0.0315	62.40 ± 5.79	4.0281 ± 0.1995
Multi-Ring	0.5449 ± 0.0297	62.53 ± 6.66	3.9938 ± 0.2639
Clan Global	0.5444 ± 0.0290	63.60 ± 6.44	4.3108 ± 0.2347
Clan Local	0.5514 ± 0.0290	62.33 ± 6.08	4.4233 ± 0.2285
Heart			
Global	**0.5725** ± 0.0410	81.90 ± 4.81	**0.9267** ± 0.2427
Local	0.6045 ± 0.0470	81.62 ± 4.80	3.5632 ± 0.1663
Von Neumann	0.5933 ± 0.0437	81.81 ± 4.71	2.5175 ± 0.5661
Wheel	0.6062 ± 0.0506	**82.48** ± 4.55	1.8858 ± 0.5036
Four Clusters	0.5898 ± 0.0436	82.00 ± 5.21	3.0651 ± 0.4207
Multi-Ring	0.5938 ± 0.0445	82.19 ± 5.32	2.9421 ± 0.4551
Clan Global	0.5870 ± 0.0470	81.67 ± 5.57	3.6413 ± 0.2493
Clan Local	0.5970 ± 0.0464	82.38 ± 4.32	3.6674 ± 0.2733
Iris			
Global	0.2931 ± 0.0479	95.67 ± 2.70	**1.6205** ± 0.3058
Local	0.2968 ± 0.0441	**96.42** ± 2.68	3.1497 ± 0.1504
Von Neumann	0.2893 ± 0.0486	96.33 ± 2.60	2.7647 ± 0.2156
Wheel	0.3034 ± 0.0422	95.50 ± 2.66	2.2210 ± 0.3820
Four Clusters	**0.2887** ± 0.0447	95.58 ± 2.68	2.8807 ± 0.1909
Multi-Ring	0.2927 ± 0.0453	96.25 ± 3.06	2.8546 ± 0.2098
Clan Global	0.2941 ± 0.0477	95.67 ± 2.78	3.0957 ± 0.1607
Clan Local	0.2942 ± 0.0472	95.92 ± 2.58	3.1395 ± 0.2255

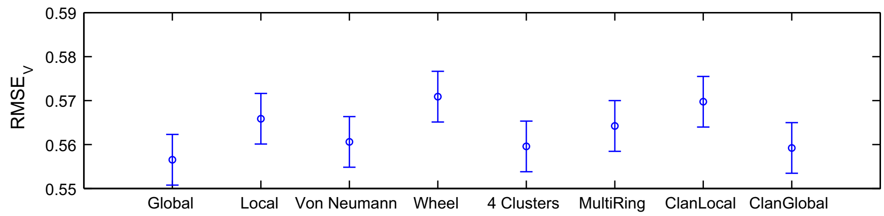


Fig. 7. Confidence intervals of the $RMSE_V$ with 5% of confidence for the Vehicle dataset.

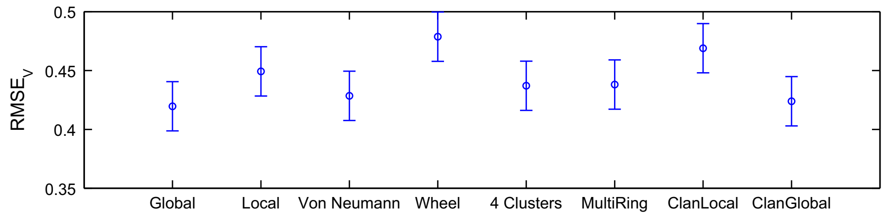


Fig. 8. Confidence intervals of the $RMSE_V$ with 5% of confidence for the Ionosphere dataset.

Table 6
SinC results after 1000 iterations. Performance of the eight topologies over 30 executions is reported with standard deviations.

Topology	$RMSE_V$	$RMSE_T$ (%)	Diversity
Global	0.1162 ± 0.0021	0.0175 ± 0.0031	1.0233 ± 0.1871
Local	0.1162 ± 0.0021	0.0178 ± 0.0031	1.8460 ± 0.2376
Von Neumann	0.1162 ± 0.0020	0.0180 ± 0.0032	1.6772 ± 0.2605
Wheel	0.1162 ± 0.0020	0.0177 ± 0.0034	1.3063 ± 0.3074
Four Clusters	0.1162 ± 0.0020	0.0181 ± 0.0031	1.7370 ± 0.2487
Multi-Ring	0.1162 ± 0.0020	0.0184 ± 0.0033	1.7634 ± 0.1883
Clan Global	0.1162 ± 0.0020	0.0173 ± 0.0028	1.7513 ± 0.1558
Clan Local	0.1162 ± 0.0021	0.0178 ± 0.0036	1.7896 ± 0.2036

We performed a similar analysis for the metric *TA*. We applied the ANOVA method for all datasets except to the datasets Heart, Ionosphere and Iris because these datasets do not satisfy the assumptions of the ANOVA. We conclude that all topologies have the same performance in terms of *TA* for the datasets Cancer, Diabetes, Glass, Ecoli, Vehicle and Sonar.

5.4. Evaluating the PSO-ELM in a regression problem

In this section, we carried out an experiment to evaluate all topologies in an approximation function problem. The function used was the 'SinC' function:

$$y = \begin{cases} \sin(x)/x & \text{if } x \neq 0, \\ 1 & \text{if } x = 0. \end{cases} \tag{10}$$

For this experiment, we created a training set (x_i, y_i) and a testing set (x_j, y_j) with 1000 data, respectively, where x_i's and x_j's are randomly distributed on the interval $(-10, 10)$ according to a uniform distribution. In order to make the regression problem 'real', large uniform noise distributed in $[-0.2, 0.2]$ has been added to all the training samples while the testing samples remain noise-free. In each trial of simulation, we created a new training set and a new testing set. In each simulation, the whole training set was divided into two sets (training and validation) with equal sizes (500 samples to each one). In this experiment, we used the same parameters used in the classification problems in the PSO-ELM but we executed it with 10 hidden nodes.

The results are presented in Table 6, where $RMSE_T$ represents the root mean squared error on testing set. All topologies reached similar values for the metric $RMSE_V$ but they reached empirically different values for the metric $RMSE_T$. However, all topologies have the same performance in terms of $RMSE_T$ (using a similar statistical analysis as done for the classification problems).

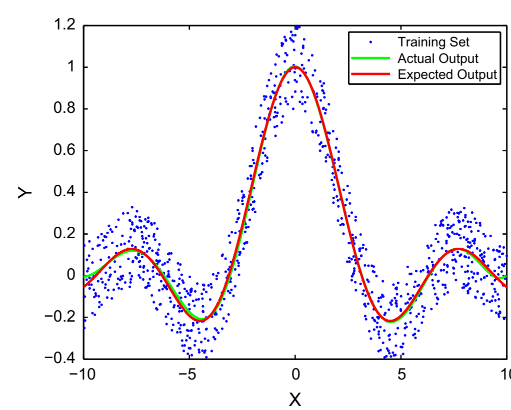


Fig. 9. Outputs of the PSO-ELM learning algorithm using the Global topology.

Fig. 9 shows the true and the approximated function of the PSO-ELM learning algorithm using the Global topology.

6. Final considerations

This paper investigated the use of eight topologies on the performance of a PSO-ELM. The empirical results clearly showed that the topologies have different performance. We emphasize the difference on performance mainly to metric $RMSE_V$ which is the function being optimized and to metric D_S that measures the convergence of the PSO. In our experiments, the Global topology reached empirically the best $RMSE_V$ on most of the datasets (Vehicle, Ionosphere, Sonar, Cancer and Heart). For the other datasets, the best topologies were Local, Von Neumann, Four Clusters and Clan-Global. In particular, we noted that, in the evaluated datasets, the topologies more connected (Global, Clan-Global, Four Clusters) presented better $RMSE_V$ than other topologies on datasets with few classes (≤ 4). On the other hand, the topologies less connected (Von Neumann and Local) presented better $RMSE_V$ than other topologies on datasets with many classes. The experiments show that the performance, measured by testing accuracy (TA), is statistically equal using all topologies. In terms of diversity, the Global topology reached lower values than other topologies on all datasets, this means that the particles, in the end of the algorithm, are near each other around the center of swarm. This fact indicates that when the PSO uses this topology the particles converge more fast than when using other topologies. Future work includes investigating the effect of velocity clamping as well as the other parameters of the PSO (w,c_1,c_2) on the performance of the PSO-ELM.

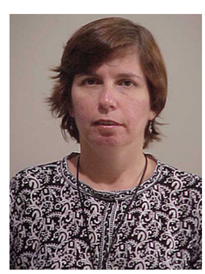
Acknowledgments

The authors would like to thank CNPq, FACEPE, CAPES (Brazilian Research Agencies) for their financial support.

References

[1] G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme learning machine: a new learning scheme of feedforward neural networks, in: Proceedings of the International Joint Conference on Neural Networks, vol. 2, 2004, pp. 985–990.
[2] G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (1–3) (2006) 489–501.
[3] Q.Y. Zhu, A. Qin, P. Suganthan, G.B. Huang, Evolutionary extreme learning machine, Pattern Recognition 38 (10) (2005) 1759–1763.
[4] F. Han, H.F. Yao, Q.H. Ling, An improved evolutionary extreme learning machine based on particle swarm optimization, Neurocomputing 116 (20) (2013) 87–93.
[5] J. Cao, Z. Lin, G.B. Huang, Composite function wavelet neural networks with extreme learning machine, Neurocomputing 73 (7–9) (2010) 1405–1416.
[6] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proceedings of the IEEE International Conference on Neural Networks, 1995, vol. 4, 1995, pp. 1942–1948.
[7] T.B. Ludermir, W.R. de Oliveira, Particle swarm optimization of MLP for the identification of factors related to common mental disorders, Expert Systems with Applications 40 (2013) 4648–4652.
[8] Y. Xu, Y. Shu, Evolutionary extreme learning machine – based on particle swarm optimization, in: Lecture Notes in Computer Science, vol. 3971, 2006, pp. 644–652.
[9] F. Han, H. Yao, Q. Ling, An improved extreme learning machine based on particle swarm optimization, in: Lecture Notes in Computer Science, vol. 6840, 2011, pp. 699–704.
[10] A.P. Engelbrecht, Computational Intelligence: An Introduction, John Wiley & Sons Ltd, 2007.
[11] S. Cheng, Y. Shi, Q. Qin, Population diversity based study on search information propagation in particle swarm optimization, in: 2012 IEEE Congress on Evolutionary Computation (CEC), 2012, pp. 1–8.
[12] R. Mendes, P. Cortez, M. Rocha, J. Neves, Particle swarms for feedforward neural network training, in: Proceedings of the International Joint Conference on Neural Networks, IJCNN '02, vol. 2, 2002, pp. 1895–1899.
[13] A.B. van Wyk, A.P. Engelbrecht, Overfitting by PSO trained feedforward neural networks, in: IEEE Congress on Evolutionary Computation, 2010, pp. 1–8.
[14] E.M. Figueiredo, T.B. Ludermir, Effect of the PSO topologies on the performance of the PSO-ELM, in: 2012 Brazilian Symposium on Neural Networks (SBRN), 2012, pp. 178–183.
[15] C. Bastos-Filho, M. Caraciolo, P. Miranda, D. Carvalho, Multi-ring particle swarm optimization, in: 10th Brazilian Symposium on Neural Networks, 2008, SBRN '08, 2008, pp. 111–116.
[16] D. Carvalho, C. Bastos-Filho, Clan particle swarm optimization, in: IEEE Congress on Evolutionary Computation, 2008, pp. 3044–3051.
[17] Y. Shi, R. Eberhart, A modified particle swarm optimizer, in: The 1998 IEEE International Conference on IEEE World Congress on Computational Intelligence, 1998, pp. 69–73.
[18] J. Kennedy, R. Mendes, Population structure and particle swarm performance, in: Proceedings of the 2002 Congress on Evolutionary Computation, CEC '02, vol. 2, 2002, pp. 1671–1676.
[19] J. Kennedy, R. Mendes, Neighborhood topologies in fully informed and best-of-neighborhood particle swarms, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 36 (4) (2006) 515–519.
[20] O. Olorunda, A. Engelbrecht, Measuring exploration/exploitation in particle swarms using swarm diversity, in: IEEE Congress on Evolutionary Computation, 2008, pp. 1128–1134.
[21] A. Frank, A. Asuncion, UCI Machine Learning Repository, URL: ⟨http://archive.ics.uci.edu/ml⟩, 2010.
[22] D.N.G. Silva, L.D.S. Pacifico, T.B. Ludermir, An evolutionary extreme learning machine based on group search optimization, in: Congress on Evolutionary Computation (CEC), 2011, pp. 574–580.
[23] J.F.L. de Oliveira, T.B. Ludermir, Homogeneous ensemble selection through hierarchical clustering with a modified artificial fish swarm algorithm, in: IEEE International Conference on Tools with Artificial Intelligence (ICTAI), 2011, pp. 177–180.

Elliackin M.N. Figueiredo received the Bachelor's degree in Computer Engineering from University of Pernambuco (UPE) in 2010. Thereafter, he received the Master's degree in Computer Science from Federal University of Pernambuco (UFPE) in 2013. Currently, he is a PhD candidate in Computer Science by Federal University of Pernambuco (UFPE). His research interests include evolutionary computation, particle swarm optimization, and applications these algorithms in engineering optimization problems as well as in computer science and other areas of science and technology.



Teresa B. Ludermir received the BSc and MSc degrees in Computing from the Universidade Federal de Pernambuco, Brazil, in 1983 and 1986, respectively, and the PhD degree in Artificial Neural Networks in 1990 from Imperial College, University of London, UK. From 1991 to 1992, she was a lecturer at Kings College London. She joined the Centro de Informática at Universidade Federal de Pernambuco, Brazil, in September 1992, where she is currently a professor and head of the Computational Intelligence Group. She has published over a 300 articles in scientific journals and conferences, three books in Neural Networks and organized two of the Brazilian Symposium on Neural Networks. Her research interests include weightless neural networks, hybrid neural systems and applications of neural networks.