Applied Soft Computing 13 (2013) 1206–1213

Contents lists available at SciVerse ScienceDirect

Applied Soft Computing

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

Artificial neural network training using a new efficient optimization algorithm

Alireza Askarzadeh*, Alireza Rezazadeh

Faculty of Electrical and Computer Engineering, Shahid Beheshti University, G.C., Evin 1983963113, Tehran, Iran

ARTICLE INFO

Article history:
Received 12 March 2012
Received in revised form 9 September 2012
Accepted 23 October 2012
Available online 12 November 2012

Keywords:
Artificial neural network
Weight training
Bird mating optimizer
Fuel cell

ABSTRACT

Because search space in artificial neural networks (ANNs) is high dimensional and multimodal which is usually polluted by noises and missing data, the process of weight training is a complex continuous optimization problem. This paper deals with the application of a recently invented metaheuristic optimization algorithm, bird mating optimizer (BMO), for training feed-forward ANNs. BMO is a population-based search method which tries to imitate the mating ways of bird species for designing optimum searching techniques. In order to study the usefulness of the proposed algorithm, BMO is applied to weight training of ANNs for solving three real-world classification problems, namely, Iris flower, Wisconsin breast cancer, and Pima Indian diabetes. The performance of BMO is compared with those of the other classifiers. Simulation results indicate the superior capability of BMO to tackle the problem of ANN weight training. BMO is also applied to model fuel cell system which has been addressed as an open and demanding problem in electrical engineering. The promising results verify the potential of BMO algorithm.

1. Introduction

Artificial neural networks (ANNs) are computational modeling tools that are defined as structures comprised of densely interconnected adaptive simple processing elements. They are able to perform massive parallel computations for data processing and knowledge representation [1]. ANN training process is an optimization task with the aim of finding a set of weights to minimize an error measure. Owing to this fact that search space is high dimensional and multimodal which is usually polluted by noises and missing data, the problem of ANN training needs powerful optimization techniques. Most often, some conventional gradient descent algorithms, such as backpropagation (BP) [2], are considered for solving the problem. The gradient-based algorithms are susceptible to be converged at local optima, because they are local search methods that the final result depends strongly on the initial weights. If the initial weights are located near local optima, the algorithm would be stuck at them.

To tackle the complexity of AAN training problem, metaheuristic optimization algorithms such as genetic algorithm (GA) [3], particle swarm optimization (PSO) [4] and ant colony optimization (ACO) [5,6] have been highly proposed to search for the optimal weights of the network. Study of the literature indicates that these algorithms have been used to train the networks, design their architecture, and feature subsets [7–9]. In contrast with conventional methods, metaheuristic algorithms do not use any gradient information, and have more chance to avoid local optima by sampling simultaneously multiple regions of search space.

Recently, a novel metaheuristic optimization algorithm, trying to simulate the evolution process of bird species, has been devised by the authors [10]. This algorithm, named bird mating optimizer (BMO), has been applied to an engineering optimization problem and superior results have been obtained in comparison with the other algorithms. Simple concept and good efficiency are the major advantages of BMO algorithm. The adequate efficiency of BMO originates from using distinct moving patterns to explore the search space. Using distinct moving patterns increases the flexibility of the algorithm to provide good balance between exploration and exploitation. The main goal of this paper is to deal with the application of BMO algorithm for finding ANN weights.

Birds are the most speciose[1] class of tetrapod[2] vertebrates having around 10,000 living species [11]. Mating process in bird's society has many similarities with an optimization process in which each bird breeds or attempts to breed a brood with high quality genes, because a bird with better genes has more chance to live. Similarly, an optimization process searches to discover the global solution in which the quality of each solution is determined by a criterion named objective (fitness) function. In engineering optimization, decision variables are given values in the search space and a solution vector is made. If a good solution is made, that experience

* Corresponding author. Tel.: +98 21 29904178; fax: +98 21 22431804.
E-mail addresses: a_askarzadeh@sbu.ac.ir, askarzadeh_a@yahoo.com (A. Askarzadeh), a-rezazade@sbu.ac.ir (A. Rezazadeh).

[1] In biology means rich in species.
[2] Animals with backbones and spinal columns.

1568-4946/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.asoc.2012.10.023

is memorized and the possibility of making a better one increases at the next time.

During mating season, birds employ a variety of intelligent behaviors such as singing, tail drumming or dancing to attract potential mates. Some courtship rituals are quite elaborate and serve to form a bond between the potential mates. The quality of each bird is specified by its features such as beak, tail, wing, etc. The related gene of each feature determines the quality of that feature, together making the overall quality of the bird. A gene is a hereditary unit that can be passed on through breeding to the next generations. Imagine a bird which has good genes among the species. This bird can fly adeptly and get more food. Hence, it is healthier than the other birds, lives longer and breeds more. The bird passes these genes for better ones on to its broods by selecting a superior mate. They also live longer and have more broods and the gene continues to be inherited generation after generation.

The ultimate success of a bird to raise a brood with superior features depends on the used strategy. Different ways result in broods with diverse features. Study of bird's society reveals that they employ different strategies to perform mating process. In general, there are four strategies: monogamy, polygyny, polyandry, and promiscuity [11–13]. According to its species each bird makes use of one of these ways to breed. Most birds are monogamous, meaning that a male bird only mates with a female one. In polygynous species a male tends to mate with several females while in polyandrous a female tends to mate with several males. In the bird's society, polygyny is much more common than polyandry. Promiscuity is another mating strategy employed by a few bird species, meaning mating systems with no stable relationships in which mating between two birds is a one-time event. This type of mating indicates a rather chaotic social structure in which the male will almost certainly never see his brood or the nest, and most likely will not see the female for another brief visit.

In order to study the usefulness of BMO algorithm three real-world classification problems, namely, Iris flower, Wisconsin breast cancer, and Pima Indian diabetes, have been considered here. Iris flower has become a typical test case for many classification techniques in machine learning. Wisconsin breast cancer and Pima Indian diabetes contain missing attribute values and are usually polluted by noises. So, they are some of the most challenging problems in machine learning field. In theses cases, the BMO performance will be compared with the results from some sophisticated classification methods. These classifiers are the best result found by an ANN trained by a subset of features selected by a binary encoded GA (GANet-best) [9], the best result of eight least squares SVM classifiers (SVM-best) [14], decision tree ensembles of CCSS [15] and EDTs [16] and hybrid evolutionary decision tree (OC1-best) [17]. Decision trees (DTs) are popular classifiers, and there are many algorithms induce a tree classifier from a data set [18]. The results are also compared with an evolutionary ANN ensemble evolved by cooperative coevolution (COOP) [19], a constructive algorithm for training cooperative ANN ensembles (CNNE) [20], and an algorithm which evolves ANN structure and connection weights (MGNN) [21].

As a further test, BMO algorithm is used to build an ANN-based model for proton exchange membrane fuel cell (PEMFC) system. As an open and demanding problem, accurate modeling of PEMFC is one of the most challenging issues in electrical engineering which has been the main focus of various researches [22–24]. PEMFC is a nonlinear, complex, time varying, and strongly coupled system that is hard to model by conventional methods. Accompanied with many assumptions and approximations, PEMFC mathematical models are based on the knowledge of electrochemistry, thermodynamics, and fluid mechanics. As a result, they cannot reflect the accurate performance of the system and are difficult to be understood by an electrical engineering person. As an alternative, ANN-based models have received significant attention to represent the PEMFC behavior. An ANN can learn the behavior of PEMFC from a set of input–output data which are experimentally obtained from the system.

This paper is arranged in four sections; Section 2 describes BMO algorithm in detail; in Section 3, the proposed algorithm is first applied to classify three well known problems, and then is used to model a PEMFC system, finally, conclusion is stated in Section 4.

2. Bird mating optimizer

The population of BMO algorithm is called a *society* and each individual in the society is called a *bird*. The society contains four types of birds: monogamous, polygynous, polyandrous, and promiscuous, breeding in a d-dimensional search space, $S \subset R^d$, to find the optimum solution. Assume that we have a set of birds in a society indicated by χ. The birds of the society are categorized based on their fitness values so that $\chi = \mu \cup \kappa \cup \psi \cup \xi$, where μ, κ, ψ, and ξ represent the set of monogamous, polygynous, polyandrous, and promiscuous birds, respectively. Each bird is associated with a predefined number of genes and shown by a vector $\vec{x}\,(\chi) = (x(\chi,1), x(\chi,2), \ldots, x(\chi,d))$. In the society, any bird is a feasible solution of the problem under consideration with a quality represented by $\mathrm{fit}(\vec{x}\chi)$. The birds attempt to pass on better genes to their broods. Consequently, as the algorithm progresses, the quality of the bird's society improves. For convenience of computation, we assume that there is only one brood when a bird mates with other one(s). The society is then updated with the better birds. The breeding among the society continues until a criterion named maximum number of generations, $\mathrm{gen}_{\max}$, is met.

In the proposed algorithm, it is also assumed that the birds of the society can switch their types during generations. At each generation, society birds which have the most promising genes, are chosen as polyandrous birds (females). They have the best fitness values among the society. A predefined percentage of the other birds which have the worst fitness are abandoned from the society and replaced by new ones produced by using a chaotic sequence. The new birds are considered as promiscuous. The remaining birds of the society are regarded as monogamous and polygynous birds. Monogamous birds have better fitness than polygynous ones. Monogamous, polygynous and promiscuous birds make the males of the society. In BMO, the percentage of each type is determined manually. Monogamous and polygynous types have a great portion and polyandrous and promiscuous types have a low percentage of the society.

Monogamous birds are those males that tend to mate with one female. During mating season, a monogamous bird starts to sing and tries to attract female birds. Polyandrous birds receive his song and gather at the vicinity of him. They employ intelligent behaviors such as dancing or tail drumming to catch the attention of the male bird. The ultimate aim of the male bird is to pass on better genes to his brood by combining his genes with the genes of his interesting elite female. Therefore, he evaluates the quality of the females, employs a probabilistic approach to select one of them as his interesting elite female, and mates with her. Female birds with more promising genes have more chance of being selected. Besides, each gene of the brood may be produced by mutation in the bird gene. The probability of mutation is controlled by a factor named mutation control factor, mcf, which varies between 0 and 1. This factor helps the algorithm maintains the diversity and avoids premature convergence. As a result, the resultant brood is produced by Eq. (1):

$$
\begin{aligned}
&\text{for} \quad j = 1 : d \\
&\quad \text{if} \quad r_1 < \text{mcf} \\
&\qquad x(\text{brood}, j) = x(\mu, j) + w \times r_2 \times (x(ef, j) - x(\mu, j)) \\
&\quad \text{else} \\
&\qquad x(\text{brood}, j) = x(\mu, j) + m_w \times ((r_3 - r_4) \times (u(j) - l(j)) \\
&\quad \text{end} \\
&\text{end}
\end{aligned}
\tag{1}
$$

where $\vec{x}(\text{brood})$, $\vec{x}(\mu)$, and $\vec{x}(\text{ef})$ are, respectively, the resultant brood, monogamous bird, and interesting elite female, d denotes the problem dimension, j is the variable index, w is a time-varying weight to adjust the importance of the elite female, r_1, r_2, r_3 and r_4 are normally distributed random numbers between 0 and 1, m_w denotes mutation weight, and $u(j)$ and $l(j)$ are the upper and lower bounds of variable jth, respectively.

In order to select the interesting elite, we resort to roulette wheel approach. In this approach, as the quality of a bird increases, the probability of its selection increases, too. In roulette wheel approach, the selection probability of the bird kth from a group including m birds is defined by the following formula:

$$
p_k = \frac{1/\text{fit}(\vec{x})}{\sum_{i=1}^{m} 1/\text{fit}(\vec{x}_i)}
\tag{2}
$$

Based on its selection probability, each candidate bird is devoted a range between 0 and 1. The birds with better qualities have wider range than the others. Then, a random number is uniformly generated between 0 and 1. That range which includes the generated number is specified and the corresponding bird is selected as the interesting elite bird. It is obvious that the birds with better quality have more chance of being selected.

Polygynous birds are those males that have tendency to couple with multiple females. In nature, a polygynous bird mates with several females resulting in a number of broods, but in BMO this behavior is metaphorically adopted in which by mating a polygynous bird with multiple females only one brood is raised which its genes are a combination of the females genes. After mating a polygynous bird with his interesting elite females, the resultant brood is given as follows:

$$
\begin{aligned}
&\text{for} \quad j = 1 : d \\
&\quad \text{if} \quad r_1 < \text{mcf} \\
&\qquad x(\text{brood}, j) = x(\kappa, j) + w \times \textstyle\sum_{i=1}^{n_{ef}} r_i \times (x(ef, j) - x(\kappa, j)) \\
&\quad \text{else} \\
&\qquad x(\text{brood}, j) = x(\kappa, j) + m_w \times ((r_2 - r_3) \times (u(j) - l(j)) \\
&\quad \text{end} \\
&\text{end}
\end{aligned}
\tag{3}
$$

where $\vec{x}(\kappa)$ and $\vec{x}(ef_i)$ are, respectively, the polygynous bird and ith elite female, n_{ef} denotes the number of elite females, and r_i are normally distributed random numbers between 0 and 1. A polygynous bird mixes the information of more candidate solutions into a new one. This behavior may lead to raising a brood with more promising genes. The collaboration of each interesting elite female in each gene of the brood is random because the coefficients r_i are independently generated.

A polygynous bird mates with a female by use of an annealing function with the following probability:

$$
Pr = \exp\left(\frac{-\Delta f}{T}\right)
\tag{4}
$$

where Pr is the probability of mating, Δf denotes the absolute difference between the objective functions of the polygynous bird and female one, and T is an adjustable parameter to control the probability. The probability of mating is high when the quality of the polygynous bird is as good as the favorite female's quality, or when the value of T is high. However, a random number between 0 and 1 is generated and compared with the calculated probability. If it is less than the calculated probability, that female bird is selected for mating. Otherwise, the selection of that female is failed.

In BMO, each polyandrous bird seeks for superior males to breed a brood with high-quality genes. Consequently, she makes aware the best males of the society (monogamous birds) of her conditions for mating. The males receive her signs and participate in her ritual. In order to increase the probability of raising a good brood a predefined number of monogamous birds which have a better quality than the others participate in this ritual. The female bird evaluates the quality of the males, employs a probabilistic approach to select her interesting elite males, and mates with them. Each gene of the resultant brood is obtained as follows:

$$
\begin{aligned}
&\text{for} \quad j = 1 : d \\
&\quad \text{if} \quad r_1 < \text{mcf} \\
&\qquad x(\text{brood}, j) = x(\psi, j) + w \times \textstyle\sum_{i=1}^{n_{em}} r_i \times (x(em, j) - x(\psi, j)) \\
&\quad \text{else} \\
&\qquad x(\text{brood}, j) = x(\psi, j) + m_w \times ((r_2 - r_3) \times (u(j) - l(j)) \\
&\quad \text{end} \\
&\text{end}
\end{aligned}
\tag{5}
$$

where $\vec{x}(\psi)$ is the polyandrous bird, $\vec{x}(em_i)$ is the ith elite male, and n_{em} denotes the number of interesting elite males.

Polyandrous birds use the annealing function to select their males, too. However, in order to increase the probability of raising a good brood a predefined percentage of monogamous birds with better qualities participate in this rituals.

As previously mentioned, promiscuous birds are produced using a chaotic sequence. With different qualities, they attend during each generation and mate with their interesting elite females. The behavior of each promiscuous bird is the same as that of monogamous bird. As a result, each gene of the resultant brood is given as follows:

$$
\begin{aligned}
&\text{for} \quad j = 1 : d \\
&\quad \text{if} \quad r_1 < \text{mcf} \\
&\qquad x(\text{brood}, j) = x(\xi, j) + w \times r_2 \times (x(e, j) - x(\xi, j)) \\
&\quad \text{else} \\
&\qquad x(\text{brood}, j) = x(\xi, j) + m_w \times ((r_3 - r_4) \times (u(j) - l(j)) \\
&\quad \text{end} \\
&\text{end}
\end{aligned}
\tag{6}
$$

where $\vec{x}(\xi)$ denotes the promiscuous bird.

Using a chaotic sequence to produce new feasible solutions in the search space increases the capability of the algorithm to discover potential solutions in as yet untested regions of the space. Chaos has some good properties such as ergodicity, stochastic properties, and regularity. A chaotic sequence can go through every state in a certain area according to its own regularity, and every state in experienced only once. Therefore, BMO can more easily escape from local optima by using chaotic movement. At the initial generation, each promiscuous bird is produced using Eq. (7) where z is chaos variable and its initial value is a random number between 0 and 1 (not the points of 0.25, 0.50 and 0.75). At the next generation, the parameter z is firstly updated by the well-known Logistic map using Eq. (8) and then, the new promiscuous bird is produced.

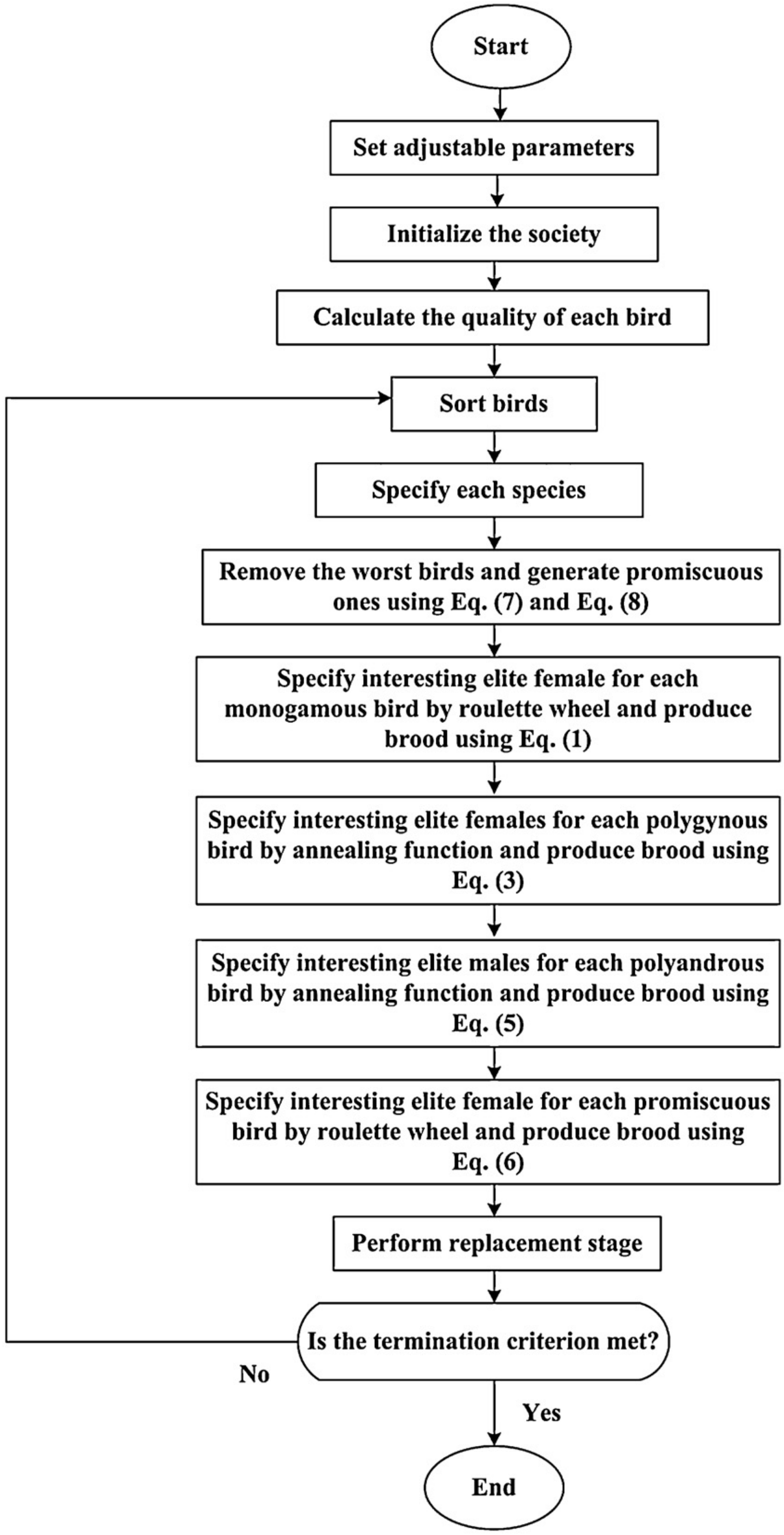

Fig. 1. Flowchart of BMO algorithm used for ANN training.

for $j = 1 : d$

$$x(\xi, j) = l(j) + z^{\text{gen}} \times (u(j) - l(j)) \tag{7}$$

end

$$z^{\text{gen}+1} = 4z^{\text{gen}}(1 - z^{\text{gen}}) \tag{8}$$

At the end of each generation, associated with each bird one brood has been raised. Replacement is the next stage. At this stage, any bird makes a decision to replace its brood instead of itself or not. The bird evaluates the quality of its brood. If the brood has better genes than the bird, the bird abandons the society and the brood attaches to it, otherwise, the brood is abandoned and the bird stays

```
Initialization:
    Determine the society size, percentage of monogamous, polygynous, promiscuous, and
  polyandrous birds, maximum number of generations, and the other parameters
Do
    Compute objective function of the birds
    Sort birds based on their objective function
    Partition the society into males and females
    Specify monogamous, polygynous, and polyandrous birds
    Remove the worst birds and generate promiscuous birds based on the chaotic sequence
    Compute objective function of the promiscuous birds
    For i = 1 to number of monogamous birds
      Select interesting elite bird
      Produce the brood based on Eq. (1)
    Next i
    For i = 1 to number of polygynous birds
      Select interesting elite birds
      Produce the brood based on Eq. (3)
    Next i
    For i = 1 to number of polyandrous birds
      Select interesting elite birds
      Produce the brood based on Eq. (5)
    Next i
    For i = 1 to number of promiscuous birds
      Select interesting elite bird
      Produce the brood based on Eq. (6)
    Next i
    Compute objective function of the broods
    Perform replacement stage
    Update the parameters
 Until termination criterion is met
```

Fig. 2. Pseudocode of BMO algorithm.

in the society. The flowchart and pseudocode of BMO algorithm have been represented in Figs. 1 and 2, respectively.

3. Application of BMO to ANN training

3.1. Using BMO in classification problems

The ANN tuned by our BMO algorithm is a three-layer feed-forward network. The nodes of the input layer are passive, meaning that they do not modify the features, they only receive them. The inputs are connected to all the hidden units, which in turn all connected to all the outputs. All neurons are connected to a bias unit, with constant output of 1. The units calculate their net activation as $\text{net} = \sum_{i=1}^{p} u_i v_i + v_0$, where p is the number of inputs to the neuron, u_i denotes an input, v_i is the corresponding weight, and v_0 is the weight corresponding to the bias unit. Hidden units employ hyperbolic tangent as their activation function, while output units make use of step function. Each hidden unit emits an output according to $f(\text{net}) = \tan h\,(\beta \text{net})$, we set $\beta = 1$ in all the experiments. Connection weights are adjusted by our BMO algorithm as represented in Fig. 3. In order to evaluate the performance of BMO-trained ANN (BMOANN), several well-studied machine learning benchmark problems from the UCI machine learning repository which are investigated by human experts in practice, have been considered: Iris flower, Wisconsin breast cancer, and Pima Indian diabetes.

In the experiments, parameter setting of BMO algorithm is as follows: The society size is set to 100; Monogamous, polygynous,

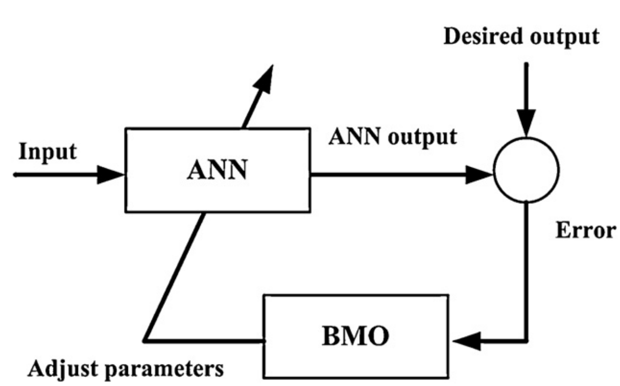

Fig. 3. Schematic diagram of BMO-based ANN.

Table 1
Error rate (%) of BMOANN for Iris flower problem. The results have been averaged over 50 runs.

Algorithm	Training set				Test set			
	Avg.	Std.	Min.	Max.	Avg.	Std.	Min.	Max.
BMOANN	1.50	0.78	1.00	4.00	3.88	1.72	0.00	8.00

Table 2
Comparison of BMOANN and the other ANNs in terms of average testing error rate (%) on the Iris flower problem.

Algorithm	BMOANN	GANet-best	SVM-best	CCSS	OC1-best
Test error rate (%)	3.88	6.40	**1.40**	4.40	3.70

polyandrous, and promiscuous birds make 50%, 40%, 5%, and 5% of the society, respectively; T, w, and m_w are defined as decreasing linear functions where $T_{\max} = 300$, $T_{\min} = 50$, $w_{\max} = 2.5$, $w_{\min} = 0.5$, $m_{w,\max} = 0.01$, and $m_{w,\min} = 0.0001$; mcf is selected 0.9 and number of monogamous birds participating in rituals of the polyandrous birds is set to 5; Maximum number of generations (epochs) is set to 100 for Iris flower and Pima Indian diabetes, and is set to 50 for Wisconsin breast cancer; To limit the search rage l and u are set to −15 and 15, respectively. It should be noted that the parameter setting is based on trial and no attempt has made to optimize it.

Matlab environment is implemented to code our BMOANN. Due to the fact that the nature of metaheuristic algorithms is stochastic, the results obtained in one attempt will differ from the results obtained in another attempt. Therefore, the performance analysis must be statistically based. As a result, in each case, 50 independent runs are performed to get an average result of the algorithm. To compare the training capability of our BMO, the results will be compared with those of the other classifiers from the literature. The results have been adopted from the literature directly for comparison.

3.1.1. Iris flower

This dataset consists of 150 samples from three species of Iris flowers (Setosa, Virginica, and Versicolor). From each species, four features including the length and the width of sepal and petal have been measured. In order to develop an ANN to classify this multiclass problem, the dataset is randomly partitioned as a training set including 100 instances and a test set including the rest 50 instances.

The architecture of BMOANN is experimentally selected 4-5-2, meaning that 5 hyperbolic tangent and 2 step functions are located in the hidden and output layers, respectively. Due to the face that Iris flower is a multiclass problem with three classes, in the output layer two nodes are used. The computer program has been written so that the outputs of 00, 01, and 11 denote the first, second, and third classes, respectively.

Table 1 reports the performance of BMOANN statistically in solving Iris flower problem over 50 runs. As can be seen, BMOANN even solves this problem with 0% of error rate for the test set. To evaluate the performance of BMOANN in solving the Iris flower problem, the results of the other methods are tabulated in Table 2 in comparison with the BMOANN result. As Table 2 indicates, BMOANN produces better results than GANet-best and CCSS and is outperformed by SVM-best and OC1-best methods. However, the error rate is slightly worse than that of the OC1-best. The performance of BMOANN in comparison with CCSS and OC1-best is promising; because they are state-of-the-art decision trees classifiers. It is worth to mention that GANet-best, SVM-best, and CCSS classifiers use k-fold cross-validation, which leads to generating more optimistic results.

Table 4
Comparison of BMOANN and the other ANNs in terms of average testing error rate (%) on the Wisconsin breast cancer problem.

Algorithm	BMOANN	GANet-best	COOP	CNNE	EDTs
Test error rate (%)	1.16	1.06	1.23	1.20	2.63
Algorithm	MGNN	SVM-best	CCSS	OC1-best	
Test error rate (%)	3.05	3.1	2.72	3.9	

3.1.2. Wisconsin breast cancer

The purpose of this dataset is to classify a tumor as either benign or malignant. This dataset has 9 integer attributes with 699 instances of which 458 instances are benign and 241 instances are malignant. In order to train the ANN for classification, the data set is divided into three sets: a training set with the first 349 instances is used to adjust the ANN weights, a validation set with the following 175 instances is used to minimize overfitting, and a test set with the final 175 instances is used only for final solution in order to verify the ANN predictive power. As previously mentioned, the validation set is used to minimize overfitting. The network weights are not adjusted with this data set. This set is only used to verify that any increase in the accuracy over the training set actually yields an increase over the accuracy of a data set that has not been shown to the network before, or at least the network has not been trained by it (i.e. validation data set). If the accuracy over the training data set increases while the accuracy over the validation set decreases or stays the same, to avoid overfitting the training process must be stopped. The architecture of BMOANN is selected 9-5-1.

We have tabulated the performance of BMOANN over 50 runs in solving Wisconsin breast cancer problem in Table 3. It is obvious that BMOANN at its best performance reaches to 0% of error rate for the test set. The performance of the other eight techniques has been summarized in Table 4 in comparison with the BMOANN result. As results reveal, BMOANN produces better results than COOP, CNNE, EDTs, MGNN, SVM-best, CCSS, and OC1-best. In this case, it is only outperformed by GANet-best technique. From the table, we can see that BMOANN markedly outperforms SVM-best and OC1-best algorithms.

3.1.3. Pima Indian diabetes

This problem is one of the most difficult ones because the dataset is relatively small and is heavily polluted by noise. This dataset has 768 instances of patients of which 500 patients have signs of diabetes and there are no signs of diabetes for the other 268 patients. From each patient, eight features have been measured. The data set is divided into three sets: a training set with the first 384 instances, a validation set with the following 192 instances, and a test set with the final 192 instances. The architecture of BMOANN is experimentally selected 8-5-1.

Table 3
Error rate (%) of BMOANN for Wisconsin breast cancer. The results have been averaged over 50 runs.

Algorithm	Training set				Validation set				Test set			
	Avg.	Std.	Min	Max	Avg.	Std.	Min	Max	Avg.	Std.	Min	Max
BMOANN	4.07	0.94	3.15	6.88	2.89	0.82	1.71	4.57	1.16	0.59	0.00	2.29

Table 5
Error rate (%) of BMOANN for Pima Indian diabetes. The results have been averaged over 50 runs.

Algorithm	Training set				Validation set				Test set			
	Avg.	Std.	Min	Max	Avg.	Std.	Min	Max	Avg.	Std.	Min	Max
BMOANN	23.57	1.44	20.83	28.13	19.45	1.25	16.67	22.92	22.55	1.89	18.23	27.08

Table 6
Comparison of BMOANN and the other ANNs in terms of average testing error rate (%) on the Pima diabetes disease.

Algorithm	BMOANN	GANet-best	COOP	CNNE	SVM-best
Test error rate (%)	22.55	24.70	19.69	19.60	22.7

Algorithm	CCSS	OC1-best
Test error rate (%)	24.02	26.0

The statistical performance of BMOANN over 50 runs in solving Pima Indian diabetes problem has been summarized in Table 5. The results obtained by the other classifiers are shown in Table 6. From the table, the best performance belongs to CNNE classifier. The results found by COOP and BMOANN classifiers are in the next orders. In this case, BMOANN outperforms GANet-best, SVM-best, CCSS, and OC1-best classifiers.

In order to make a conclusion, the rank of each classifier to solve the benchmark problems has been indicated in Table 7. Rank 1, Rank 2, and Rank 3 denote the rank of each classifier to solve Iris flower, Wisconsin breast cancer, and Pima Indian diabetes problems, respectively. In summary, as the last column of Table 7 indicates, the capability of the classifiers tested here can be ordered as CNNE > BMOANN > COOP > GANet-best > SVM-best > CCSS = EDTs > OC1-best > MGNN. It can be found that BMOANN has been ranked the second and outperformed by CNNE classifier. It should be noted that the most of the classifiers are sophisticated ones using different techniques to improve their performance, while BMOANN has simple structure which only uses the search power of BMO to tune the ANNs parameters.

3.2. Using BMO for modeling of PEMFC

As one of the most promising renewable energy resources, PEMFC has significantly attracted the attention of industrial owners. PEMFC is an electrochemical device which converts the stored chemical energy of hydrogen and oxygen into electricity. Polarization curve, representing the fuel cell voltage vs. current (*V*–*I*), is one of the most important characteristics of fuel cells. It is an important tool for researchers, because optimization of fuel cell operating points and design of the power conditioning units, simulators for fuel cell stack systems, as well as the system controllers depend on such characteristic [25]. Therefore, accurate modeling of *V*–*I* characteristics is necessary.

Table 7
Comparison of BMOANN and the other ANNs in terms of their rank in solving the classification problems.

Algorithm	Rank 1	Rank 2	Rank 3	Average rank	Final rank
BMOANN	3	2	3	2.67	2
GANet-best	5	1	6	4	4
SVM-best	1	8	4	4.33	5
CCSS	4	6	5	5	6
OC1-best	2	9	7	6	8
COOP	–	4	2	3	3
CNNE	–	3	1	2	1
EDTs	–	5	–	5	6
MGNN	–	7	–	7	9

Two main modeling can be found in the literature for PEMFC system: mathematical modeling and ANN-based models. Mathematical models describe fuel cell behavior by a variety of equations. The nonlinearity and complexity of PEMFC result in considering many assumptions and approximations during modeling. Besides, PEMFC is a time-varying system whose parameters are extremely related to the operating conditions, and a given set of operating conditions requires a corresponding set of parameters. To obtain accurate results with mathematical models, parameter identification must be performed in each operating condition. These drawbacks greatly limit mathematical model's application. To provide a better approach, ANN can be an efficient candidate to model PEMFC behavior.

The relation of fuel cell voltage, V, and current density, I, is influenced by many operating parameters, such as cell temperature, T_c, humidity, λ, hydrogen pressure, p_{h2}, oxygen pressure, p_{o2}, hydrogen flow rate, q_{h2}, oxygen flow rate, q_{o2}, etc. It is defined by the following equation.

$$V = f(I, T_c, \lambda, p_{h2}, p_{o2}, q_{h2}, q_{o2}, \ldots) \tag{9}$$

A model considering all the operating parameters has not been developed, so far. Our ANN-based model is not exception. Here, current density and cell temperature are taken as variable parameters while the others are constant. So, Eq. (9) is simplified and expressed by the following equation.

$$V = f(I, T_c) \tag{10}$$

A three-layer feed-forward ANN with two inputs and one output is constructed to model polarization curve of the Ballard MK5-E PEMFC system [26]. In this case, activation function used in the output layer is linear and gen_{max} is set to 200. The other parameters are same as those of the previous investigations. Five sets of data are used to train and test the ANN model as Fig. 4. In order to train the ANN, the experimental data obtained at 24, 31, 39, and 72 °C is used, and the data obtained at 56 °C is employed to test the ANN-based PEMFC model.

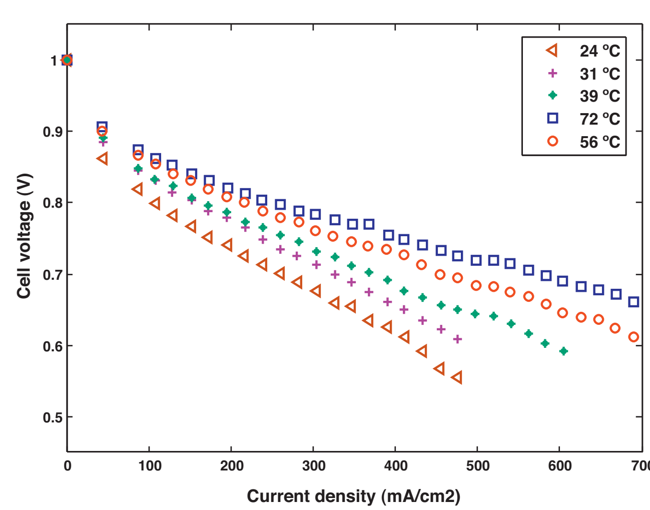

Fig. 4. Experimental data obtained from the Ballard MK5-E PEMFC system.

Table 8
Error rate (%) of BMOANN for modeling of the Ballard MK5-E PEMFC system. The results have been averaged over 50 runs.

Algorithm	Training set				Test set			
	Avg.	Std.	Min	Max	Avg.	Std.	Min	Max
BMOANN	0.11	6.18e−2	0.02	0.29	0.11	7.52e−2	0.01	0.35

Table 9
Comparison of BMOANN and the other ANNs in terms of the best testing error rate (%) on the Ballard MK5-E PEMFC system.

Algorithm	BMOANN	PSOANN	BPANN	ANN-new
Test error rate (%)	**0.01**	0.0247	0.14	0.024

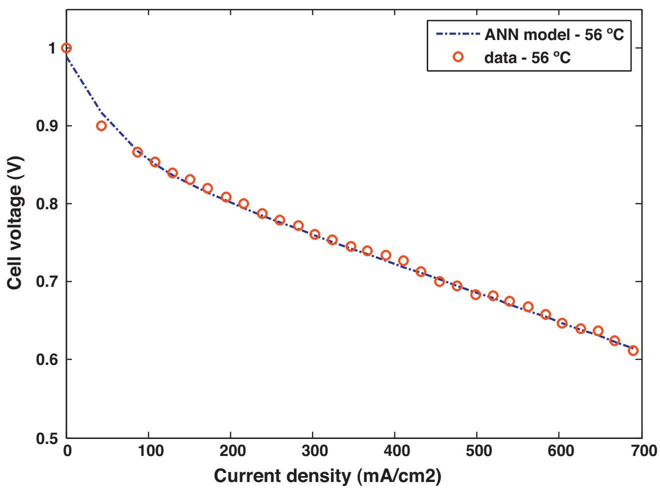

Fig. 5. Comparison between experimental data and ANN model for the test set.

We have tabulated the results generated by BMOANN over 50 independent runs in Table 8. Table 9 shows the comparison between the BMO result and the best performance of an ANN trained by BP (BPANN), an ANN trained with PSO algorithm (PSOANN), and an ANN trained by a new learning algorithm (ANN-new) [27]. It is obvious that BMOANN yields better result than the other ANNs. To observe the similarity between the experimental data and those predicted by the ANN, Fig. 5 illustrates the cell voltage vs. current density for the test data. From the figure, we can see that the BMOANN has successfully modeled the PEMFC system so that the experimental data and model results are close to each other.

4. Conclusion

The main goal of this paper is to develop a BMO-based learning algorithm to train ANNs. BMO is a recently devised population-based optimization algorithm which imitates the mating behavior of bird species for breeding superior broods and provides different strategies to effectively seek the search space. The main merit of the algorithm is to employ distinct patterns to move through the search space which leads to avoiding premature convergence and maintaining population diversity. BMOANN is firstly used to classify three real-world problems. The results of the BMOANN are promising when we compare its performance with those of the other classifiers. Although BMO algorithm is not the best classifier, has a satisfying performance and can compete the already known classifiers. BMO is then applied to build an ANN model for polarization curve of the Ballard MK5-E PEMFC system. Simulation results disclose that the model results are in good agreement with the experimental data and BMOANN yields better result than the other ANNs. Therefore, BMO algorithm can be an efficient candidate for training ANNs.

References

[1] R.J. Schalkoff, Artificial Neural Networks, McGraw-Hill, New York, 1997.
[2] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, Nature 323 (1986) 533–536.
[3] D.J. Montana, L. Davis, Training feedforward neural networks using genetic algorithms, Machine Learning (1989) 762–767.
[4] V.G. Gudise, G.K. Venayagamoorthy, Comparison of particle swarm optimization and backpropagation as training algorithms for neural networks, in: Proceedings of the 2003 IEEE, Swarm Intelligence Symposium, 2003. SIS '03, 2003, pp. 110–117.
[5] K. Socha, C. Blum, An ant colony optimization algorithm for continuous optimization: application to feed-forward neural network training, Neural Computing & Applications 16 (2007) 235–247.
[6] C. Blum, K. Socha, Training feed-forward neural networks with ant colony optimization: an application to pattern classification, in: Hybrid Intelligent Systems, 2005, HIS '05, Fifth International Conference on, 2005.
[7] Y. Xin, Evolving artificial neural networks, Proceedings of the IEEE 87 (1999) 1423–1447.
[8] X. Yao, A review of evolutionary artificial neural networks, International Journal of Intelligent Systems 8 (1993) 539–567.
[9] E. Cantú-Paz, C. Kamath, An empirical comparison of combinations of evolutionary algorithms and neural networks for classification problems, IEEE Transactions on Systems, Man, and Cybernetics, Part-B 35 (2005) 915–927.
[10] A. Askarzadeh, A. Rezazadeh, A new heuristic optimization algorithm for modeling of proton exchange membrane fuel cell: bird mating optimizer, International Journal of Energy Research, in press
[11] http://en.wikipedia.org/wiki/Bird. in press.
[12] http://www.stanford.edu/group/stanfordbirds/text/essays/Monogamy.html. in press.
[13] http://www.paulnoll.com/Oregon/Birds/courtship-systems.html. in press.
[14] T. Van Gestel, Benchmarking least squares support vector machine classifiers, Machine Learning 54 (2004) 5–32.
[15] S. Džeroski, B. Ženko, Is combining classifiers with stacking better than selecting the best one? Machine Learning 54 (2004) 255–273.
[16] T.G. Dietterich, An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization, Machine Learning 40 (2000) 139–157.
[17] E. Cantu-Paz, C. Kamath, Inducing oblique decision trees with evolutionary algorithms, IEEE Abbreviations for Transactions, Journals, Letters, and Magazines 7 (2003) 54–68.
[18] S.K. Murthy, On Growing Better Decision Trees from Data, Doctoral Dissertation, University of Maryland,1997.
[19] N. García-Pedrajas, C. Hervas-Martinez, D. Ortiz-Boyer, Cooperative coevolution of artificial neural network ensembles for pattern classification, IEEE Abbreviations for Transactions, Journals, Letters, and Magazines 9 (2005) 271–302.
[20] M. Islam, X. Yao, K. Murase, A constructive algorithm for training cooperative neural network ensembles, IEEE Transactions on Neural Networks 14 (2003) 820–834.
[21] P.P. Palmes, T. Hayasaka, S. Usui, Mutation-based genetic neural network, IEEE Transactions on Neural Networks 16 (2005) 587–600.
[22] A. Askarzadeh, A. Rezazadeh, A grouping-based global harmony search algorithm for modeling of proton exchange membrane fuel cell, International Journal of Hydrogen Energy 36 (2011) 5047–5053.
[23] A. Askarzadeh, A. Rezazadeh, Artificial immune system-based parameter extraction of proton exchange membrane fuel cell, International Journal of Electrical Power & Energy Systems 33 (2011) 933–938.
[24] M. Ye, X. Wang, Y. Xu, Parameter identification for proton exchange membrane fuel cell model using particle swarm optimization, International Journal of Hydrogen Energy 34 (2009) 981–989.
[25] X.D. Xue, et al., Unified mathematical modelling of steady-state and dynamic voltage–current characteristics for PEM fuel cells, Electrochimica Acta 52 (2006) 1135–1144.
[26] Z.-D. Zhong, X.-J. Zhu, G.-Y. Cao, Modeling a PEMFC by a support vector machine, Journal of Power Sources 160 (2006) 293–298.
[27] A. Rezazadeh, A. Askarzadeh, M. sedighizadeh, ANN-based PEMFC modeling by a new learning algorithm, The International Review on Modelling and Simulations 3 (2010) 187–193.