

Prediction of asphaltene precipitation using artificial neural network optimized by imperialist competitive algorithm

Mohammad Ali Ahmadi

Received: 23 August 2011 / Accepted: 17 October 2011 / Published online: 1 November 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract One of the most important phenomena in petroleum industry is the precipitation of heavy organic materials such as asphaltene in oil reservoirs, which can cause diffusivity reduction and wettability alteration in reservoir rock and finally affect oil production and economical efficiency. In this work, the model based on a feed-forward artificial neural network (ANN) optimized by imperialist competitive algorithm (ICA) to predict of asphaltene precipitation is proposed. ICA is used to decide the initial weights of the neural network. The ICA–ANN model is applied to the experimental data reported in the literature. The performance of the ICA–ANN model is compared with Scaling model and conventional ANN model. The results demonstrate the effectiveness of the ICA–ANN model.

Keywords Asphaltene · Precipitation · Artificial neural network · Imperialist competitive algorithm · Prediction

List of symbols

R	Solvent to oil ratio (g/mol)
M	Molecular weight (g/mol)
W	Amount of precipitated asphaltene (weight percent)
Y	Function defined by Eq. 1
X	Function defined by Eq. 2
x	Function defined by Eq. 4
y	Function defined by Eq. 5
ANN	Artificial neural network

PSO Particle swarm optimization

ICA Imperialist competitive algorithm

Introduction

The precipitation and deposition of crude oil polar fractions such as asphaltenes in petroleum reservoirs reduce considerably the rock permeability and the oil recovery. So, many researchers studied this important subject. They introduced experimental procedures or even analytical models, but a fully satisfactory interpretation is still lacking. The available models for description of asphaltene precipitation are divided into two general groups. The first group consists of thermodynamic models, which need asphaltene properties such as density, molecular weight and solubility parameter for prediction of asphaltene phase behavior. All those models consider asphaltene as a pure pseudo-component, but this assumption causes much deviation in the prediction of asphaltene phase behavior (Pedersen et al. 1989); the second group of models is based on the scaling approach which explained separately. In this paper, the ability of the artificial intelligence in establishing and predicting amount of asphaltene precipitation is to be investigated. Artificial intelligence have been widely used and are gaining attention in petroleum engineering because of their ability to solve problems that previously were difficult or even impossible to solve. One example of ability neural network in well-log analysis. This technique has been increasingly applied to predict reservoir properties using well-log data (Doveton and Prenskey 1992; Balan et al. 1995).

A soft sensor is a conceptual device whose output or inferred variable can be modeled in terms of other parameters that are relevant to the same process (Rallo et al. 2002). According to Rallo et al. (2002), artificial

M. A. Ahmadi (✉)
Department of Petroleum Engineering,
Ahwaz Faculty of Petroleum Engineering,
Petroleum University of Technology,
Kut Abdollah, Ahwaz, Iran
e-mail: ahmadi6776@yahoo.com;
Mohammadali.ahmadi@ait.put.ac.ir

neural network (ANN) could be used as soft sensor building approach.

The determination of network structure and parameters is very important; some evolutionary algorithms such as genetic algorithm (GA) (Qu1 et al. 2008), back propagation (BP) (Tang and Xi 2008), pruning algorithm (Reed 1993), simulated annealing (de Souto et al. 2002) can be used for this determination. Recently, a new evolutionary algorithm has been proposed by Atashpaz-Gargari and Lucas (2007) which has inspired from a socio-political evolution, called imperialist competitive algorithm (ICA).

In the present work, we propose ICA for optimizing the weights of feed-forward neural network. Then simulation results demonstrate the effectiveness and potential of the new proposed network for asphaltene precipitation prediction compared with scaling model (Hu and Guo 2001) using the same data.

Scaling model

The three variables involved in the scaling equation are the weight percent of precipitated Asphaltenes, W (based on the weight of feed oil), the dilution ratio, R (defined as the ratio of injected solvent volume to weight of crude oil), and the molecular weight of solvent, M . Rassamdana et al. (1996) combined the three variables into two (X , Y) as follows:

$$X = \frac{R}{M^Z} \quad (1)$$

$$Y = \frac{W}{R^{Z'}} \quad (2)$$

Z and Z' are two adjustable parameters and must be carefully tuned to obtain the best scaling fit of the experimental data. They suggested Z' is a universal constant of -2 and $Z = 0.25$ regardless of oil and precipitant used. The proposed scaling equation is expressed in terms of X and Y through a third-order polynomial function

$$Y = A_1 + A_2X + A_3X^2 + A_4X^3 (X > X_c) \quad (3)$$

where X_c is the value of X at the onset of asphaltene precipitation.

Hu et al. (2000) performed a detailed study on the application of scaling equation proposed by Rassamdana et al. (1996) for asphaltene precipitation. They examined the universality of exponents Z and Z' and found that Z' is a universal constant ($Z' = -2$) while exponent Z depends on the oil composition and independent of specific precipitant (n -alkane) used. For the experimental data used, they found also that the optimum value of Z is generally within the range of $0.1 < Z < 0.5$.

Despite the simplicity and accuracy of the scaling equation mentioned above, it is restricted to use at a constant

temperature and since temperature is not involved in the scaling equation as a variable, it is not adequate for correlating and predicting the asphaltene precipitation data measured at different temperatures. Due to this issue, Rassamdana et al. modified their scaling equation by implanting temperature parameter in the scaling equation. Based on the previous equation, they defined two new variables x and y :

$$x = X/T^{C_1} \quad (4)$$

$$y = Y/X^{C_2} \quad (5)$$

in which X and Y are variables defined as in Eqs. (1) and (2) and constant C_1 and C_2 are adjustable parameters. They reported that the good fit of their experimental data can be achieved by setting $C_1 = 0.25$ and $C_2 = 1.6$.

Again the new scaling equation is a third-order polynomial in general form of:

$$y = b_1 + b_2x + b_3x^2 + b_4x^3 (x > x_c) \quad (6)$$

Hu et al. (2001) studied the effects of temperature, molecular weight of n -alkane precipitants and dilution ratio on asphaltene precipitation in a Chinese crude oil experimentally. The amounts of asphaltene precipitation at four temperatures in the range of 293–338 K were measured using seven n -alkanes as precipitants. They found that their experimental data could not be well correlated by setting $C_1 = 0.25$ and $C_2 = 1.6$ as recommended by Rassamdana et al. (1996). They reported that their experimental data could be correlated successfully by choosing $C_1 = 0.5$ and $C_2 = 1.6$. Regression plot of predicted asphaltene precipitation using scaling model (Hu and Guo 2001) against experimental data is shown in Fig. 1.

Artificial neural networks

Artificial neural networks are parallel information processing methods which can express complex and nonlinear relationship use number of input–output training patterns from the experimental data. ANNs provides a non-linear mapping between inputs and outputs by its intrinsic ability (Hornik et al. 1990).

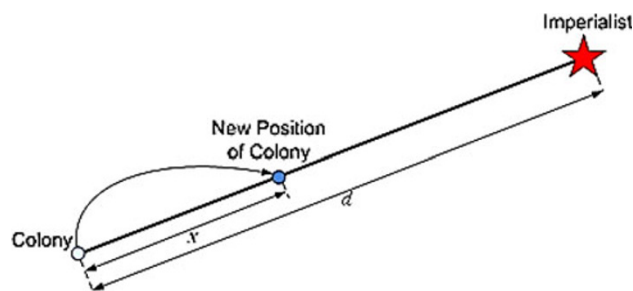


Fig. 1 Movement of colonies toward their relevant imperialist

The most common neural network architecture is the feed-forward neural network. Feed-forward network is the network structure in which the information or signals will propagate only in one direction, from input to output. A three layered feed-forward neural network with back propagation algorithm can approximate any nonlinear continuous function to an arbitrary accuracy (Brown and Harris 1994; Hornick et al. 1989).

The network is trained by performing optimization of weights for each node interconnection and bias terms until the output values at the output layer neurons are as close as possible to the actual outputs. The mean squared error of the network (MSE) is defined as:

$$MSE = \frac{1}{2} \sum_{k=1}^G \sum_{j=1}^m [Y_j(k) - T_j(k)]^2 \tag{7}$$

where m is the number of output nodes, G is the number of training samples, $Y_j(k)$ is the expected output, and $T_j(k)$ is the actual output. The data are split into two sets: a training data set and a validating data set. The model is produced using only the training data. The validating data are used to estimate the accuracy of the model performance.

Imperialist competitive algorithm

The ICA is a new evolutionary algorithm in the evolutionary computation field based on the human’s socio-political evolution (Atashpaz-Gargari and Lucas 2007). Like other evolutionary algorithms, the ICA starts with initial populations called countries. There are two types of countries: colony and imperialist (in optimization terminology, countries with the least cost) which together form empires. In the imperialistic competition process, imperialists try to attempt to achieve more colonies. So during the competition, the powerful imperialists will be increased in the power and the weak ones will be decreased in the power. When an empire loses all of its colonies, it is assumed to be collapsed. At the end, the most powerful imperialist will remain in the world and all the countries are colonies of this unique of this empire. In this stage, imperialist and colonies have the same position and power.

The implementation procedures of our proposed matching strategy based on ICA are described as follows.

Generating initial empire

A country formed as an array of variable values to be optimized. In a N_{var} dimensional optimization problem, this array defined by:

$$\text{Country} = [P_1, P_2, P_3, \dots, P_{N_{var}}] \tag{8}$$

The cost of a country is found by evaluating the cost function f :

$$\text{Cost} = f(\text{country}) = f([P_1, P_2, P_3, \dots, P_{N_{var}}]) \tag{9}$$

The algorithm starts with the number of initial countries ($N_{country}$), number of imperialist (N_{imp}) and number of the remaining country are colonies that each belongs to an empire (N_{col}) the initial number of colonies of an empire in convenience with their powers. To divide the colonies among imperialists proportionally, the normalized cost of an imperialist is defined by:

$$C_n = c_n - \max_i \{c_i\} \tag{10}$$

where c_n is the cost of n th imperialist and C_n is its normalized cost. Having the normalized cost of all imperialist, the power of each imperialist is calculated by:

$$P_n = \left| \frac{C_n}{\sum_{i=1}^{N_{imp}} C_i} \right| \tag{11}$$

In the other hand, the normalized power of an imperialist is determined by its colonies. Then, the initial number of an imperialist will be:

$$NC_n = \text{round}\{P_n \cdot N_{col}\} \tag{12}$$

where NC_n is the initial number of colonies of n th empire and N_{col} is the number of all colonies. To divide the colonies among imperialists, NC_n of the colonies is selected randomly and assigned them to each imperialist. The colonies together with the imperialist form the n th empire.

Moving colonies of an empire toward the imperialist

The imperialist countries try to improve their colonies and make them a part of themselves. This fact is modeled by moving all colonies toward their relevant imperialist. Figure 1 (Atashpaz-Gargari and Lucas 2007) shows this movement. In this figure, the colony moves toward the imperialist by x (is a random variable with uniform distribution) units.

$$x \sim U(0, \beta \times d) \tag{13}$$

where β is a number greater than 1 and d is the distance between a colony and an imperialist. In the moving process, a colony may reach a position with lower cost than that of its imperialist. In this case, the imperialist and the colony change their positions. Then, the algorithm will continue by the imperialist in the new position and then colonies start moving toward this position.

The total power of an empire

The total power of an empire depends on both the power of the imperialist country and the power of its colonies. This fact is modelled by defining the total cost by:

Table 1 Compositions (mol%) and properties of the degassed Cao-qiao crude oil and separator gas

Component	Degassed oil	Separator gas
CO ₂	0.0	2.96
N ₂	0.0	1.18
C ₁	0.0	89.37
C ₂	0.0	3.34
C ₃	0.0	2.10
i-C ₄	0.0	0.32
n-C ₄	0.0	0.26
i-C ₅	0.16	0.22
n-C ₅	0.58	0.15
n-C ₆	1.2	0.12
C ₇ ⁺	98.06	
C ₁₁ ⁺	87.16	
C ₇ ⁺ molecular weight (g/mol)	503.6	
C ₇ ⁺ density (at 293 K)	0.9526	
Reservoir temperature	343	
Bubble point pressure at 343 K (MPa)	9.8	
Gas oil ratio (GOR, m ³ /m ³)	30.2	
Saturates (wt%)	38.0	
Aromatics (wt%)	47.6	
n-C ₅ asphaltenes (wt%)	7.26	
Resins (wt%)	18.6	

Table 2 Comparison between the performances of ICA-ANN and scaling model

	ICA-ANN	ANN	Scaling
MSE	0.0032749	0.83759	0.69396
R ²	0.99367	0.95586	0.96413

$$TC_n = \text{Cost}(\text{imperialist}_n) + \xi \text{mean}\{\text{cost}(\text{colonies of impire}_n)\} \quad (14)$$

where TC_n is the total cost of the n th empire, and ξ is a positive number which is considered to be less than 1. A small value for ξ implies that the total power of an empire to be determined by just the imperialist and increasing it will increase the role of the colonies in determining the total power of an empire. The value of 0.1 for ξ is a proper value in most of the implementations.

Imperialistic competition

All empires try to take the possession of colonies of other empires and control them. The imperialistic competition gradually brings about a decrease in the power of weaker empires and an increase in the power of more powerful ones. This competition is modelled by just picking some

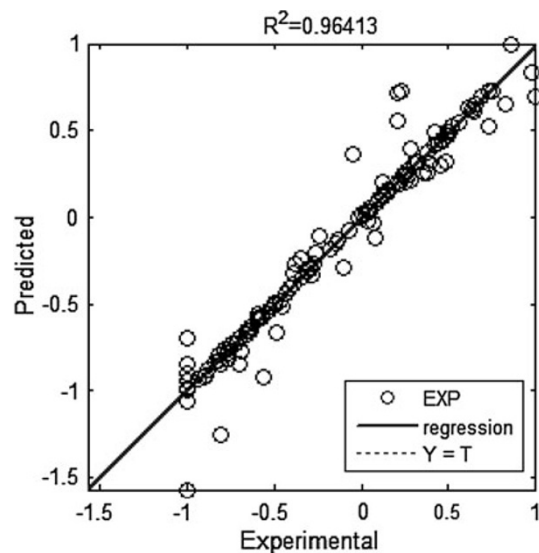
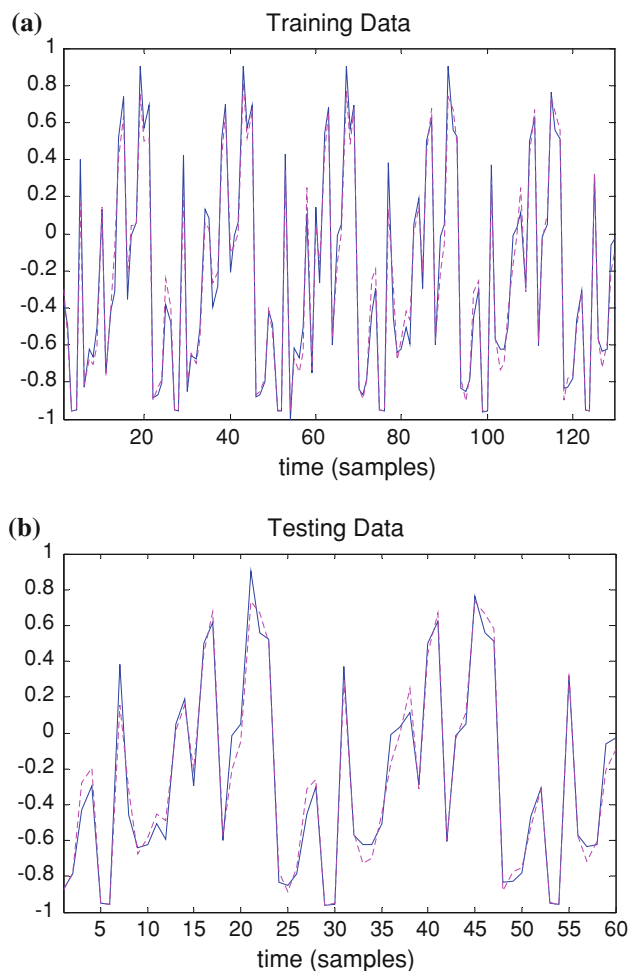
**Fig. 2** Regression plot of prediction by scaling equation (Hu and Guo 2001)**Fig. 3** Comparison between measured and predicted asphaltene precipitation (ICA-ANN): **a** training, **b** test

Fig. 4 Training state plot of ICA-ANN

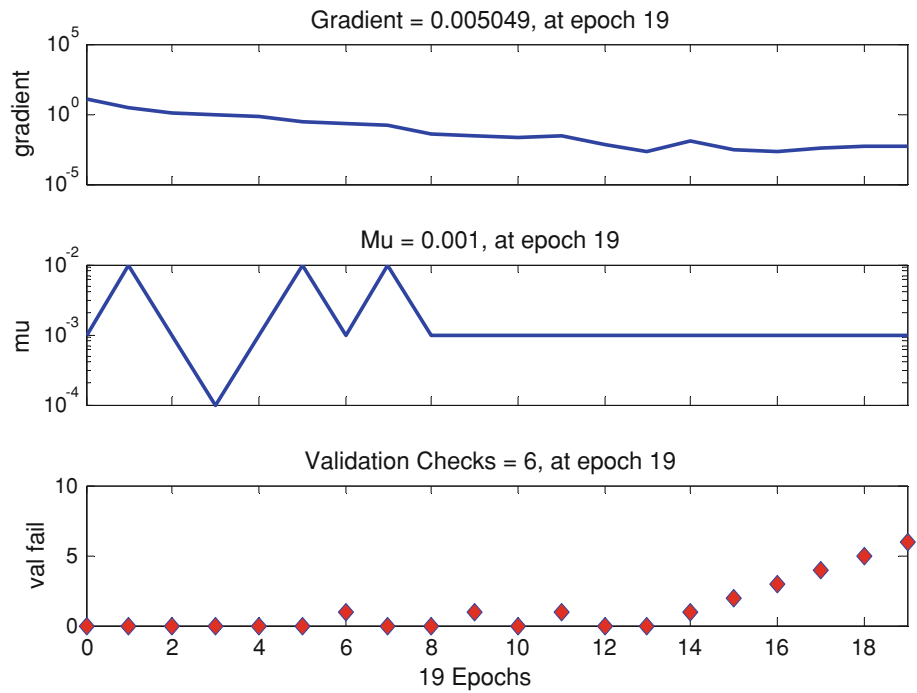
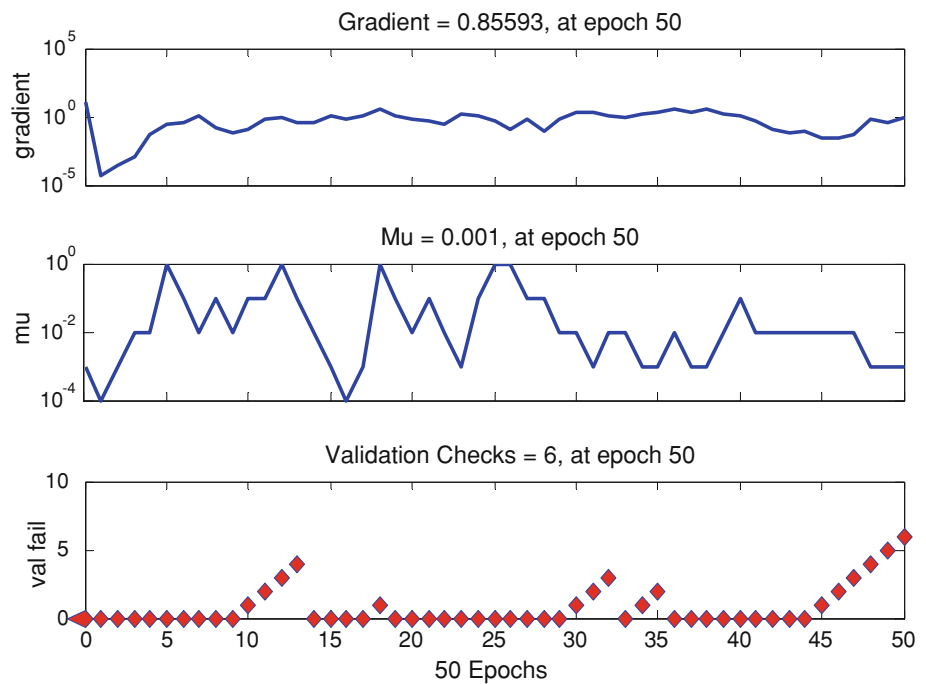


Fig. 5 Training state plot of ANN



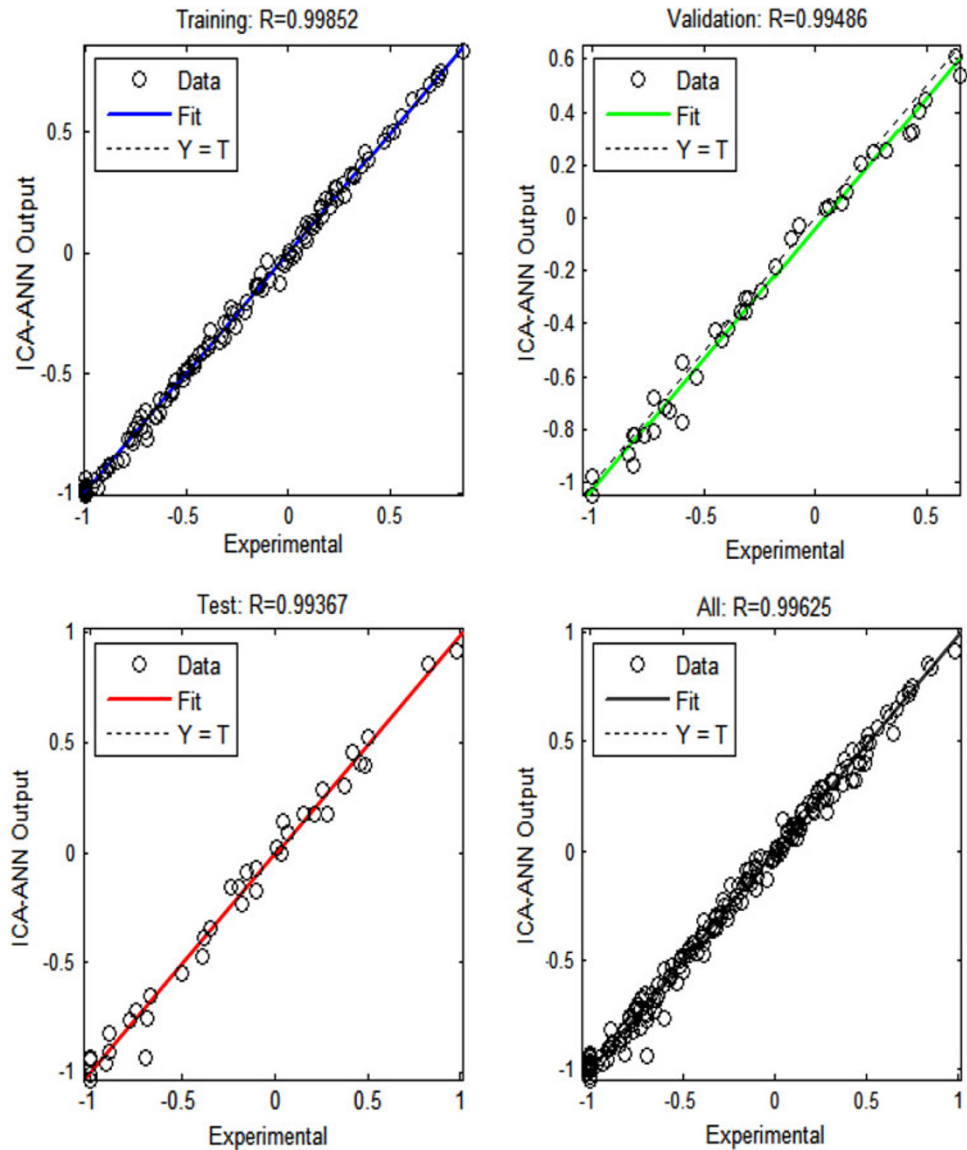
(usually one) of the weakest colonies of the weakest empires and making a competition among all empires to possess this colonies.

To start the competition, first, the possession probability of each empire is found based on its total power. The normalized total cost is obtained by:

$$NTC_n = TC_n - \max_i\{TC_i\} \tag{15}$$

where, TC_n and NTC_n are the total cost and the normalized total cost of n th empire, respectively. Having the normalized total cost, the possession probability of each empire is given by:

Fig. 6 Regression plot of ICA-ANN



$$P_{P_n} = \left| \frac{NTC_n}{\sum_{i=1}^{N_{imp}} NTC_i} \right| \tag{16}$$

To divide the mentioned colonies among empires, vector **P** is formed as

$$\mathbf{P} = [P_{P_1}, P_{P_2}, P_{P_3}, \dots, P_{P_{N_{imp}}}] \tag{17}$$

Then the vector **R** with the same size as **P** whose elements are uniformly distributed random numbers is created,

$$\mathbf{R} = [r_1, r_2, r_3, \dots, r_{N_{imp}}] \tag{18}$$

Then vector **D** is formed by subtracting **R** from **P**

$$\mathbf{D} = \mathbf{P} - \mathbf{R} = [D_1, D_2, D_3, \dots, D_{N_{imp}}] \tag{19}$$

Referring to vector **D**, the mentioned colony (colonies) is handed to an empire whose relevant index in **D** is maximized.

Powerless empire will collapse in the imperialistic competition and their colonies will be divided among other empires. At the end, all the empires except the most powerful one will collapse and all the colonies will be under the control of this unique empire. In this stage, imperialist and colonies have the same position and power.

Results and discussion

In this study, an ANN was used to build a model to predict asphaltene precipitation using the data reported in literature

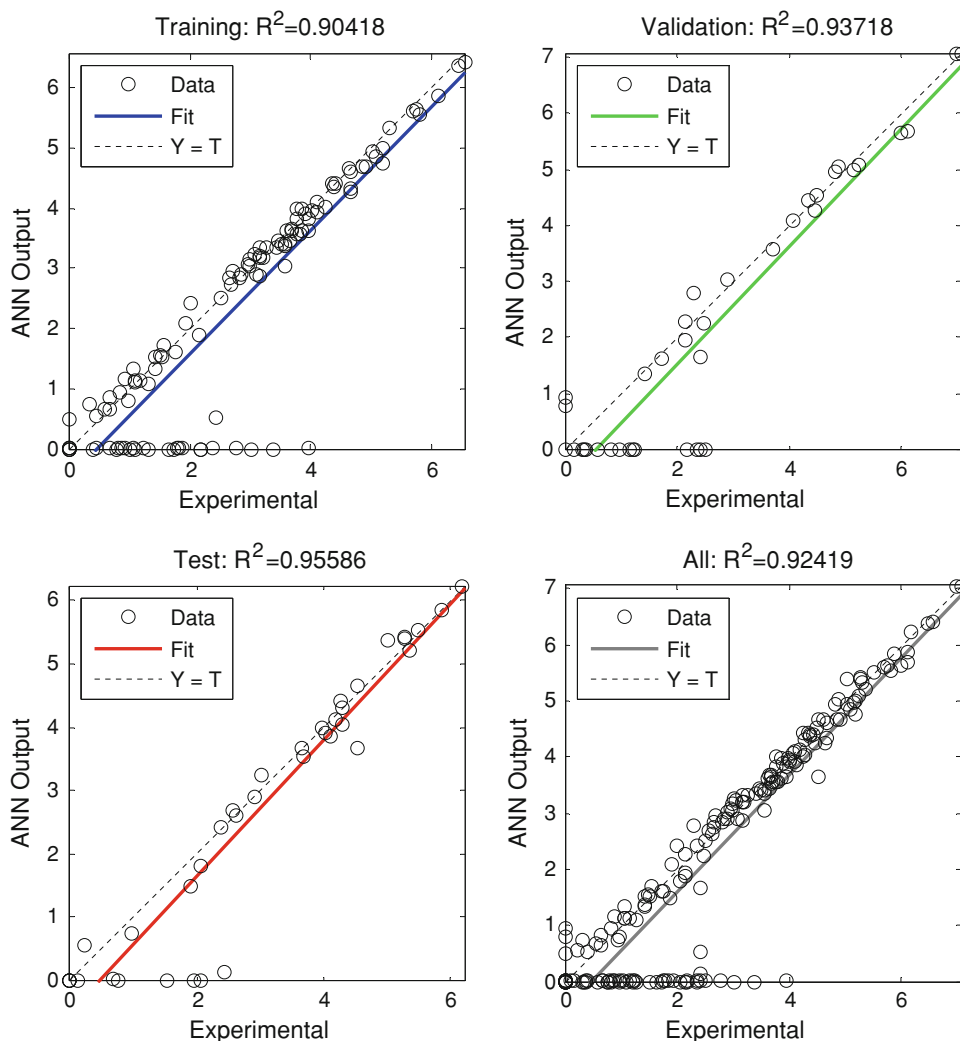


Fig. 7 Regression plot of ANN

(Hu and Guo 2001). The best ANN architecture was: 3-4-10-1 (3 input units, 4 hidden neurons in first layer, 10 hidden neurons in second layer, 1 output neuron). ANN model trained with back propagation network was trained by Levenberg–Marquardt using three parameters: (1) molecular weight, (2) dilution ratio, and (3) temperature as inputs. The transfer functions in hidden and output layer are sigmoid and linear, respectively. Physical and thermodynamic properties of oil used for generating experimental data by Hu and Guo (2001) are shown in Table 1.

ICA is used as neural network optimization algorithm and the MSE used as a cost function in this algorithm. The goal in proposed algorithm is minimizing this cost function. Every weight in the network is initially set in the range of $[-1, 1]$. In these simulations, the number of imperialists and the colonies is considered 4 and 40, respectively; parameter β is set to 2. The number of training and testing data is 130 and 60, respectively.

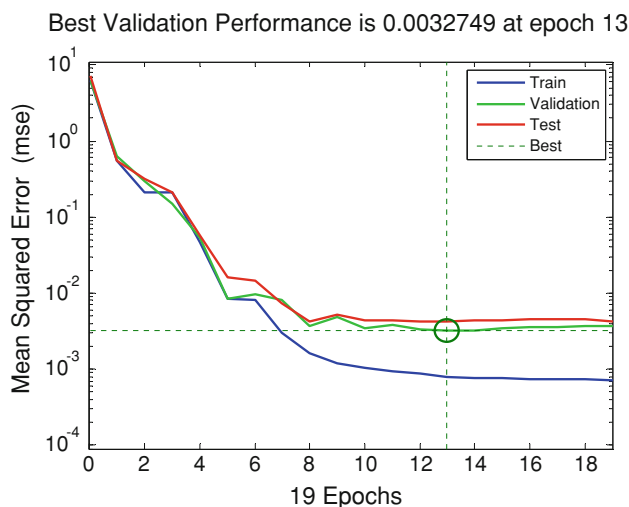


Fig. 8 Performance plot of ICA-ANN

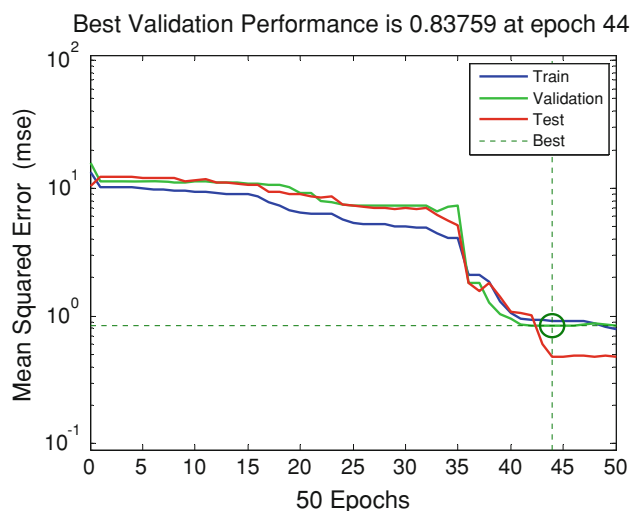


Fig. 9 Performance plot of ANN

The simulation performance of the ICA–ANN model and ANN model were evaluated on the basis of MSE and efficiency coefficient R^2 . Table 2 gives the MSE and R^2 values for the three different models of the validation phases. Prediction of asphaltene precipitation by scaling model is shown in Fig. 2 and prediction of asphaltene precipitation in the training and test phase is shown in Fig. 3. The simulation performance of the ICA–ANN model and ANN model were evaluated on the basis of MSE and efficiency coefficient R^2 . Table 2 gives the MSE and R^2 values for three different models of the validation phases. Training state and regression plot and performance of ICA–ANN and ANN models are shown in Figs. 4, 5, 6, 7, 8 and 9, respectively. It can be observed that the performance of ICA–ANN model is better than scaling model and ANN model.

Conclusions

The idea of ICA algorithm is that each initial point of the neural network is selected by ICA and the fitness of the ICA is determined by a neural network. The experiment with experimental data reported in literature (Hu and Guo 2001) has showed that the ICA–ANN model is successfully demonstrated on prediction of asphaltene precipitation also predictive performance of the proposed model is better than that of scaling model (Hu and Guo 2001) and conventional ANN model. One problem when considering the combination of neural network and ICA for prediction of asphaltene precipitation is the determination of the optimal neural network structure. Proposed neural network structure described in this work is determined manually.

A substitute method is to apply the ICA or another evolutionary algorithm for neural network structure optimization, which will be a part of our future work. The proposed asphaltene precipitation prediction model may be combined with existing asphaltene precipitation modeling softwares to speed up their performance, reduce the uncertainty and increase their prediction and modeling capabilities.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

- Atashpaz-Gargari E, Lucas C (2007) Imperialist competitive algorithm: an algorithm for optimization inspired by imperialistic competition. In: IEEE Congress on Evolutionary Computation, pp 4661–4667
- Balan B, Mohaghegh S, Ameri S (1995) SPE Eastern Regional Conference and Exhibition, West Virginia, pp. 17–21
- Brown M, Harris C (1994) Neural fuzzy adaptive modeling and control. Prentice-Hall, Englewood Cliffs
- de Souto MCP, Yamazaki A, Ludemir TB (2002) Optimization of neural network weights and architecture for odor recognition using simulated annealing. In: Proceedings of the international joint conference on neural networks, vol 1, pp 547–552
- Doveton JH, Prensky SE (1992) Geological applications of wireline logs: a synopsis of developments and trends. Log Analyst 33(3):286–303
- Hornik K, Stinchcombe M, White H (1989) Multilayer feed forward networks are universal approximators. Neural Netw 2:359–366
- Hornik K, Stinchcombe M, White H (1990) Universal approximation of an unknown mapping and its derivatives using multilayer feed forward networks. Neural Netw 3(5):551–560
- Hu Y-F, Guo T-M (2001) Effect of temperature and molecular weight of n-alkane precipitation on asphaltene precipitation. Fluid Phase Equilib J 192:13–25
- Hu Y-F, Chen GJ, Yang JT, Guo TM (2000) Fluid phase Equilib J 171:181–195
- Pedersen KS, Fredenslund A, Thomassen P (1989) Properties of oils and natural gases. Gulf Publishing, Houston
- Qu1 X, Feng J, Sun W (2008) Parallel genetic algorithm model based on AHP and neural networks for enterprise comprehensive business. In: IEEE International conference on intelligent information hiding and multimedia signal processing, pp 897–900
- Rallo R, Ferre-Gin J, Arenas A, Giralt F (2002) Neural virtual sensor for the inferential prediction of product quality from process variables. Comput Chem Eng 26:1735–1754
- Rassamdana H, Dabir B, Nematy M, Farhani M, Sahimi M (1996) Asphalt flocculation and deposition: I. the onset of precipitation. AIChE J 42:10–22
- Reed R (1993) Pruning algorithms—a survey. IEEE Trans Neural Networks 4:740–747
- Tang P, Xi Z (2008) The research on BP neural network model based on guaranteed convergence particle swarm optimization. In: Second international symposium on intelligent information technology application, IITA '08, vol 2, pp 13–16