Petroleum and Coal

Article **Open Access**

PREDICTION OF GAS HYDRATE FORMATION TEMPERATURE IN THE PRESENCE OF METHANOL INHIBITOR IN GAS PIPELINE USING ARTIFICIAL INTELLIGENCE (AI) AND GENETIC ALGORITHM (GA)

Niloufar Fatourehchi[1], Zahra Mashayekhi[2], Saeed Sadeghpour Galooyak[3], Majid Masoumi[3]*

[1] *Gas Research division, Research Institute of Petroleum Industry, Tehran, Iran*
[2] *Environment and Biotechnology division, Research Institute of Petroleum Industry, Tehran, Iran*
[3] *Petroleum Refining and Processing Technology Development Division, Research Institute of Petroleum Industry, Tehran*

Received August 18, 2019; Accepted October 31, 2019

Abstract

The phenomenon of gas hydrate formation in different oil and gas industries in downstream processes hinders or decreases productions. Artificial intelligence techniques such as neural network may efficiently help in solving gas hydrate formation issues and also avoid extra costs by predicting hydrate formation conditions. In this study, a hybrid multilayer perceptron neural network with genetic algorithm has been used to model the conditions of gas hydrate formation. The variables of natural gas pressure, specific density, and mass percentage of methanol in aqueous solution are three input parameters of the neural network, and the output parameter of the neural network is the gas hydrate formation temperature. The optimal number of hidden layer neurons was calculated as 7 using the genetic algorithm. The total number of data pieces used in modeling was 156, 70% of which is for training, 15% for monitoring, and the rest for network testing. The small numerical error values obtained for the test data (0.29) and the high correlation coefficient (close to 1) for the test data indicate that the hybrid neural network using a genetic algorithm successfully performs in predicting hydrate formation temperature. At the end of the modeling, a closed formula has been presented to calculate the gas hydrate formation temperature.

Keywords: *Hydrate formation temperature, Methanol inhibitor, Modeling, Hybrid neural network, Genetic algorithm.*

1. Introduction

Gas hydrate is a combination of light gases such as methane, ethane or carbon dioxide (CO_2), which, under a certain temperature and pressure, combine with water molecules to form ice-like substances. Gas hydrates are ice-like solid crystalline compounds that are composed of water molecules as hosts and gas molecules as guests at low temperatures (close to zero degree) or at high pressures (usually > 2.6 MPa). Many gas components such as methane, ethane, and carbon dioxide can be encapsulated by water molecules through hydrogen bonding and form hydrates under appropriate conditions.

Hydrate crystals have a complex 3D structure in which the water molecules form cages in which the guest molecules are trapped. The stability of the guest molecule is attributed to the vander Waals forces. There is no bond between the guest and host molecules. In the process of hydrate formation, no bond is formed or broken, but only phase change occurs. The guest molecules are free to rotate and move within the cages made of host molecules.

Low temperature, high pressure, and hydrate components (methane, ethane, and carbon dioxide) contribute to the hydrate formation. To avoid hydrate formation, it is sufficient to remove only one of the above three conditions. Typically, hydrate components cannot be removed from the mixture. In the case of natural gas, the hydrate components are the desirable

products. Therefore, hydrates are counteracted through the two remaining factors, e.g., the combination of temperature and pressure as well as water content. Other factors forming the hydrate include perturbation and turbulence, increased nucleation sites, and free water [1]. Another way of preventing hydrate formation is to inject inhibitors into the system. Inhibitors are substances making the hydrate form at a certain pressure at a lower temperature. Alcohols, glycols, and salts are examples of these substances.

The methods cited above change the thermodynamic equilibrium of hydrate formation and are known as thermodynamic inhibitory methods. Because in these methods, the system becomes thermodynamically unstable by changing the composition, temperature, or operating pressure, and hydrate formation will not occur as long as the system is unstable. Another way to prevent the formation of hydrates is by using kinetic inhibitors. These materials allow the system to remain stable under thermodynamic conditions but prevent the growth of hydrate crystals. Their mechanism is to block the fine particles of the hydrate crystals, thereby prevening their further growth or the hydrate crystal bonding. Inhibitors are widely used in the oil and gas industry to prevent gas hydrate formation. In this paper, methanol has been used as thermodynamic inhibitor. Artificial intelligence tools such as genetic algorithm and neural network are new computational methods for machine learning, displaying knowledge, and applying knowledge to predict the output responses of complex systems. The main idea beyond such tools is partly inspired by the performance of the biological nervous system for data processing and information for learning and knowledge creation. A key element of this idea is to create new structures for the information processing system. This system consists of a large number of highly interconnected processing elements called neurons that work together to solve a problem and transmit information through synapses (electromagnetic communications) [2-4].

In this study, we try to predict the conditions of gas hydrate formation in a gas pipeline by using artificial intelligence tool via MATLAB software and obtain the optimal conditions. The gas pressure, the mass percentage of methanol in the solution (as inhibitor), the gas density, and the constituents of the gas are all input variables that affect the output variable, e.g. the temperature of gas hydrate formation.

2. Computational procedures and programme

Traditional modeling methods based on mass, energy, and momentum balances require precise process data [5-6]. There are time-consuming analytical methods which have a considerable error when there is insufficient knowledge of the processes. On the other hand, computational intelligence-based methods do not require a precise description of the process mechanism [5, 7-8] and can simulate the system output only using empirical (laboratory) data with reasonable accuracy. These models are mainly applied when the complexities of the process prevent the analytical modeling of them.

Artificial neural networks, as a type of computational intelligence-based models, have been inspired by the parallel structure of neural computing in the human brain. The overall structure of the artificial neural network model is determined by an algorithm or by an operator. The network parameters are adjusted by learning algorithms and experimental data to minimize the output error. Data-driven models such as neural networks estimate the variable based on the input and output conditions of the systems or, in other words, initial and final conditions. In this study, a multilayer perceptron neural network has been used to model the gas hydrate formation conditions. The variables of natural gas pressure, specific density, and mass percentage of methanol in aqueous solution are three input parameters of the neural network, and output parameter of the neural network is the temperature of gas hydrate formation.

After creating the neural network structure, the network should be trained by a robust and reliable database. In fact, relevant laboratory data are used for training, validation, and testing. Here, a database of empirical data related to the process of gas hydrate formation in the presence of methanol inhibitors will be used. In Figure. 1, the input and output variables of modeling the formation temperature of gas hydrate using an artificial neural network are shown.

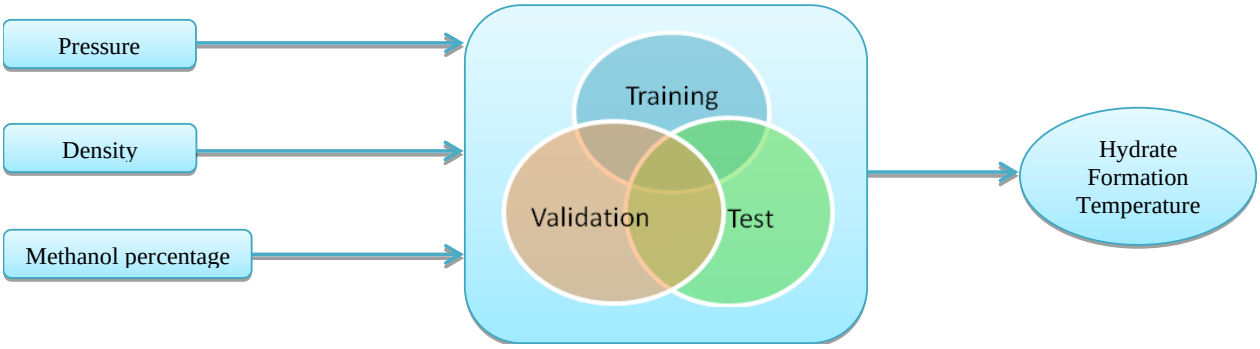

Figure 1. Schematic view of modeling input and output data in this study

In this study, the Levenberg-Marquardt algorithm has been used to train the network [7]. Verifying the capability and reliability of the artificial neural network in process simulation will be examined using test data. A review of the literature indicates that few studies have been performed on the artificial neural networks application for gas hydrate modeling.

In this study, the data obtained from the Hasheminejad Refinery Gas Transmission Line were used for modeling in the presence of methanol inhibitors with different mass percentages. The data were divided randomly into three categories: training, validation, and testing. In the training phase, the free parameters of the network (weights and biases) approach their optimal values. The genetic algorithm is used to find the optimal structure of the neural network. The results, which include a comparison of both real and modeling data, are presented at the end. Error value and regression coefficient are also calculated for modeling and real data. Also, the optimal network structure has been extracted by the genetic algorithm.

2.1. Topology of the neural network used

Neural networks are data-driven models that can determine the relationship between independent and dependent data in a physical system by a network of interconnected nodes. Figure 2 shows the topology of 1-7-3 for the neural network. The structure shown has 3 neurons in the input layer (temperature, pressure, and percentage of methanol), 7 neurons in the hidden layer, and 1 neuron in the output layer (gas hydrate formation temperature).

2.2. Input and output data

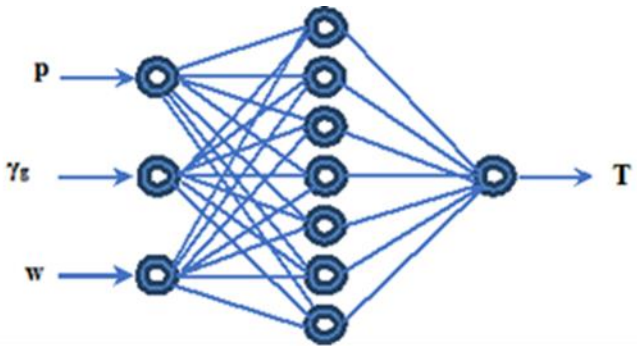

Figure 2. The neural network structure used in this study with three neurons in the input layer (pressure, specific density, molar percentage) and one neuron in the output layer

The amount of input data is fundamental to network efficiency. If the training models do not provide all the components of the problem which have to be solved by the network, the network will not be trained well, and the network performance will be very poor for the test data. The only general formula which can be mentioned is to use a large amount of input data that defines the problem properly. Another aspect that must be kept in mind is the relationship between the total input data and the number of network weights. If the number of input samples is less than the network weights, one can expect to assign a single weight to each sample.

This process reduces network generalization. The number of training samples should be at least twice the weight of the network.

In Figure 3, the schematic display of the type of input and output data of the network is presented along with their upper and lower bounds. Input and output data values were also obtained from the Hashemitejad gas refinery pipeline. The number of samples used in this study was 156.

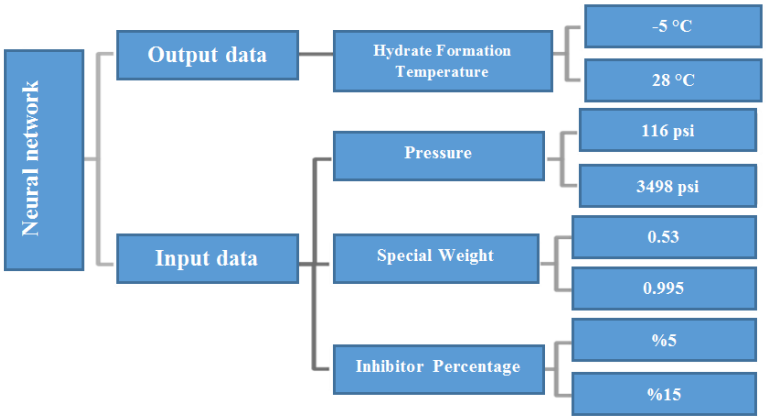

Figure 3. Input and output data (Hasheminejad Gas Refinery in Mashhad) for creating a data-oriented model and their upper and lower bounds using artificial intelligence techniques

2.3. Data segmentation

In this study, according to neural network theory, the data are randomly divided into three categories of training, monitoring and testing network, which includes 70%, 15% and 15% of the total data, respectively.

2.4. The governing equations

The connection between the nodes of the input and the hidden layer is made by the edges. Each of these edges has a specific weight. The number of edges between the input and the hidden layer is (7 × 3), so there are 21 weights that are placed in IW weight matrix, respectively (according to Equation 1).

$$IW_{S\times R} = \begin{bmatrix} IW_{11} & IW_{12} & IW_{13} \\ IW_{21} & IW_{22} & IW_{23} \\ \dots & \dots & IW_{\cdot\cdot} \\ IW_{S1} & IW_{S2} & IW_{S3} \end{bmatrix}_{7\times 3} \quad (1)$$

Matrix 1 is known as the input layer weight matrix. Matrix 2 is the weight matrix of the hidden layer. The entries of both matrices are all unknown and obtained during the network training process.

$$IW_{1\times S} = [IW_{11} \quad IW_{12} \quad \dots \quad IW_{17}]_{1\times 7} \quad (2)$$

In following, the input layer bias matrix (b^1) and the hidden layer bias matrix (b^2) has been presented.

$$b^1 = \begin{bmatrix} b_1^1 \\ b_2^1 \\ \vdots \\ b_S^1 \end{bmatrix}_{7\times 1} \quad \text{and} \quad b^2 = [b_1^2]_{1\times 1} \quad (3)$$

Figure 4 illustrates the relationship between the input and output layers using the IW and LW weight matrices as well as $b^1$and b^2bias matrices. Also, the transfer functions between layers ($f_1$andf_2) are shown in Figure 4. $T_{NN}$is the output variable of the neural network (hydrate formation temperature). In this figure, the neural network performance is simply illustrated by the following equations.

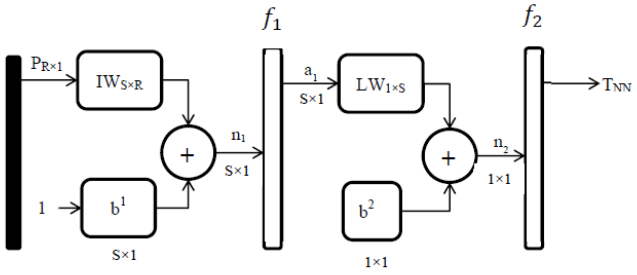

Figure 4. Relationship between input and output data using weight matrix and bias among layers

Matrix 4 shows the neural network input data matrix. As seen, this matrix has three inputs, including gas pipeline pressure, gas specific weight, and methanol weight percentage in aqueous solution.

$$P_R = \begin{bmatrix} p \\ s_g \\ w \end{bmatrix}_{3\times1} \quad (4)$$

After definitions of the different matrices, the method of calculating the network output using the inputs is shown.

$$\eta_1 = IW \times P_R + b^1 \quad (5)$$

$$a_1 = f_1(\eta_1) \quad (6)$$

$$\eta_2 = LW \times a_1 + b^2 \quad (7)$$

$$T_{NN} = f_2(\eta_2) \quad (8)$$

2.5. Transfer functions

The transfer function is a linear or nonlinear function, also called a transformation and activation function. Like all mathematical functions, the transfer function takes one input (x) and yields an output. In this study, the hyperbolic tangent function and linear function are used.

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad \text{Hyperbolic tangent} \quad (9)$$

$$f(x) = x \quad \text{Linear} \quad (10)$$

2.6. Neural network training

As seen in Figure 5, all the parameters are known except for the weight and bias matrix values that are unknown. To obtain these unknowns, the neural network needs to be trained with the help of laboratory data. During training, the weight and bias values change untill the output (T_{NN}) and real output of the lab (T_{Real}) match.

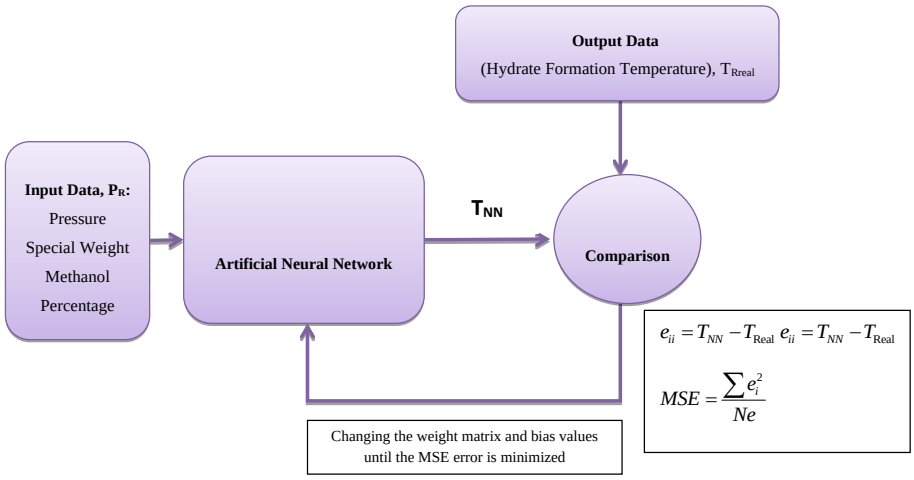

Figure 5. How the neural network works at a glance

2.7. Hybridizing the network with genetic algorithm

Using the trial and error method, one can find the proper size of the network; but in this study, the genetic algorithm has been used to find the optimal structure. In other words, the optimal number of hidden layer neurons will be obtained by the genetic algorithm. It is necessary to note that the objective function of this problem is the mean square error, and the independent variables are the number of neurons in the hidden layer. In the following, results have been presented for the optimal state. At the optimal state, there are 7 neurons in the hidden layer of the neural network.

3. Results and discussion

In this section, a comparison has been made between the experimental and the modeling results. The data (156 pieces) have been randomly divided into three categories of training, validation, and testing. The results have been summarized for each of these three categories as well as for all of them. Figure 6 presents the comparison of results for all data altogether with the corresponding error.

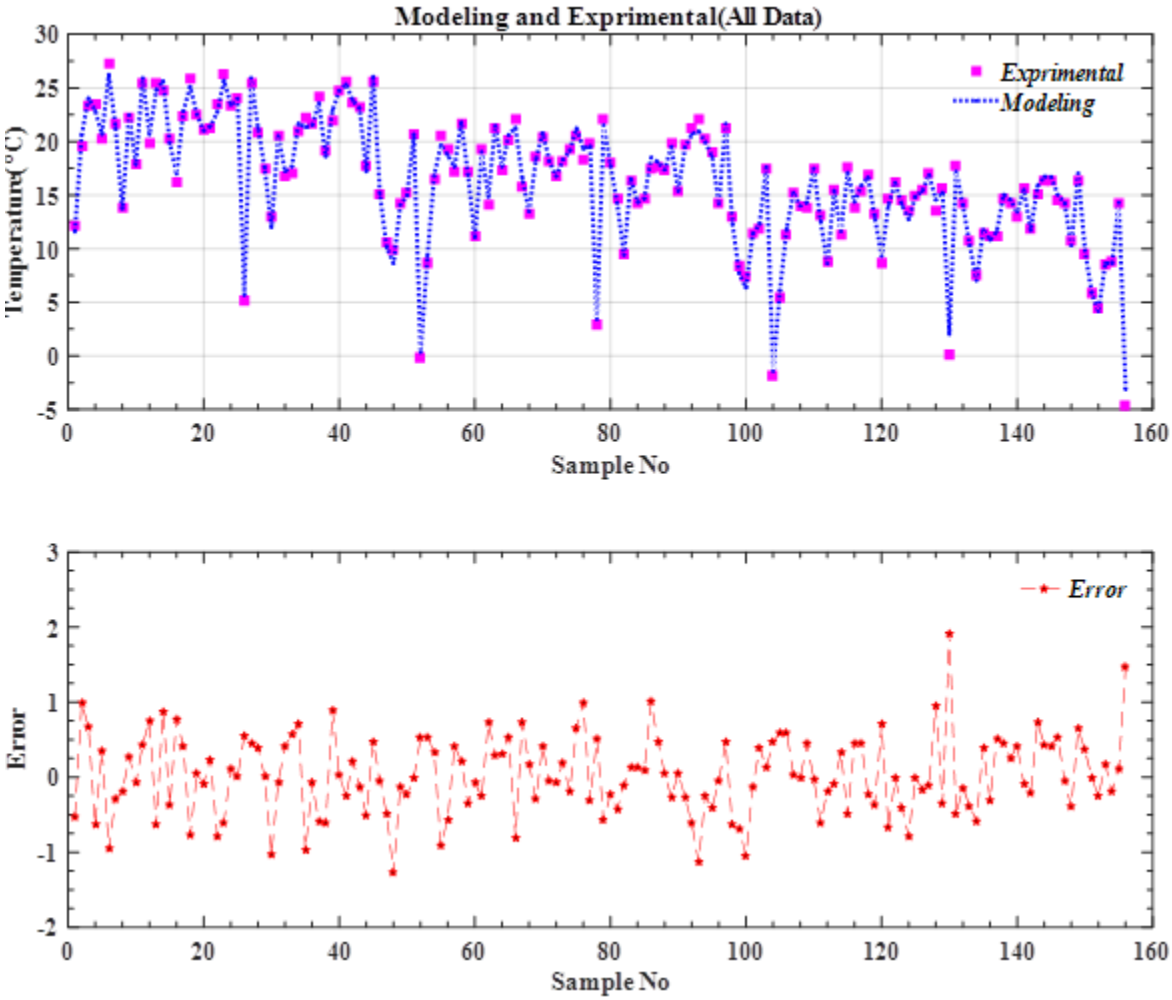

Figure 6. Total data (comparing neural network prediction values and experimental data) including training, validation, and test data along with their error

As can be seen, the error of prediction using the neural network is mainly in the range of +1 to -1°C. In the following figures, the diagram of comparison of experimental data and modeling is given separately for all three categories. It is obvious that the results related to

the comparison of the test data indicate the accuracy and actual performance of modeling using the neural network.

The neural network complements the learning process using the training data and assesses the quality of learning of the network by validation. Finally, the network can be actually tested using the test data to evaluate its accuracy in prediction. As can be seen, the neural network performs very successfully in the test phase and is able to predict the experimental values with high accuracy, and the prediction error has been small (Figure 7).

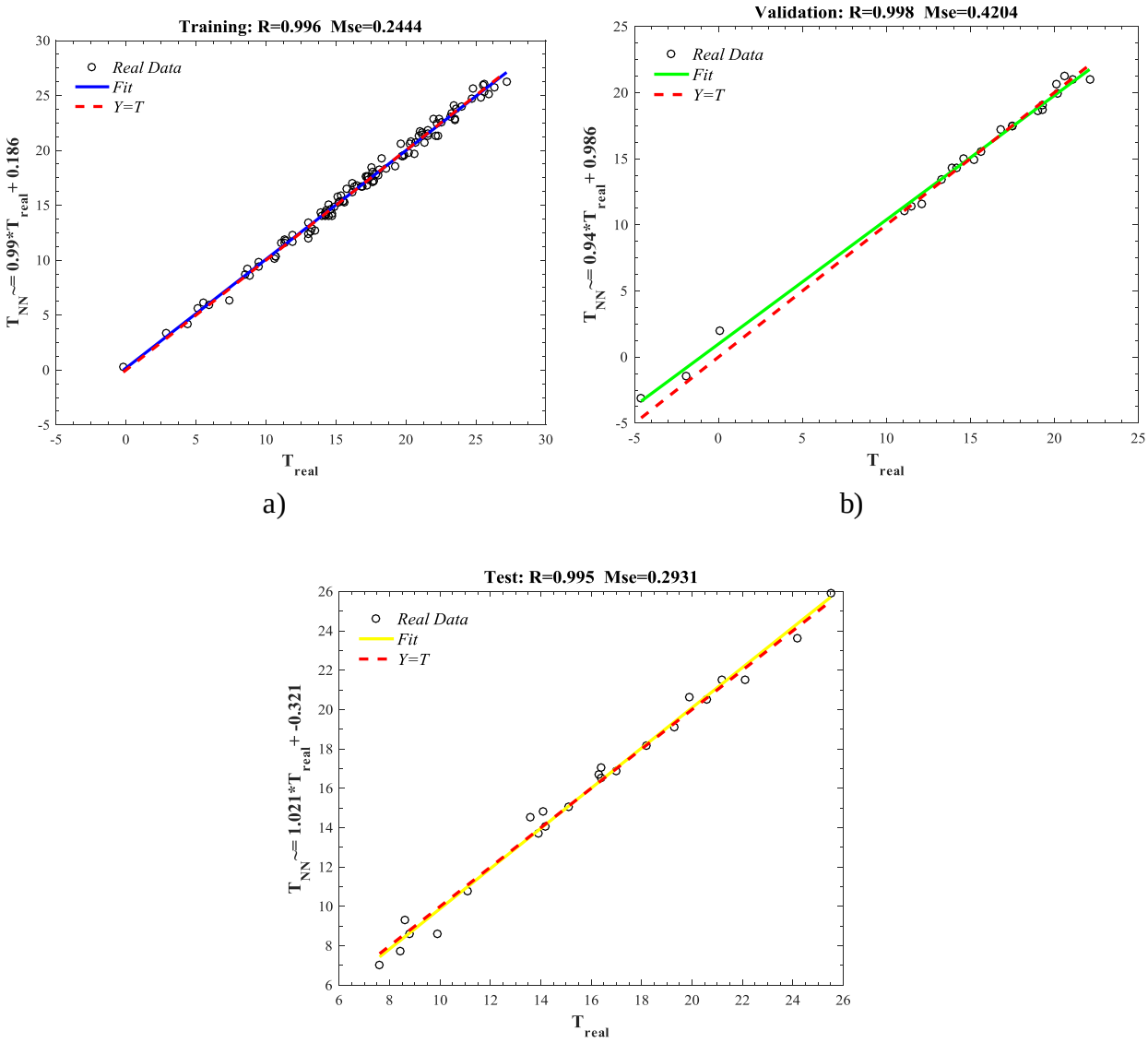

Figure 7. Comparison of the results of neural network modeling and real (experimental) results for (a) training data; b) monitoring data, and c) test data

Table 1. Mean squared error and correlation coefficient for all three categories of training, validation and testing and total data

	Training (70%)	Validation (15%)	Testing (15%)	Total (100%)
Mean squared error (MSE)	0.24	0.42	0.29	0.27
Regression coefficient	0.996	0.998	0.995	0.996

As expected, the low error value and the high correlation coefficient indicate the proper performance of the neural network in predicting the hydrate formation temperature. In the following curve, the real hydrate formation temperature and temperature obtained from the modeling has been indicated. Ideally, when the network error is zero, all points are placed on the line. In practice, there is a small error value that causes the points to be scattered above and below this line. The equation of the best trend line from the points of this diagram, along with its correlation coefficient, is given below.

$$T_{NN} = 0.99 \times T_{\text{Real}} + 0.186 R = 0.996 \quad (11)$$

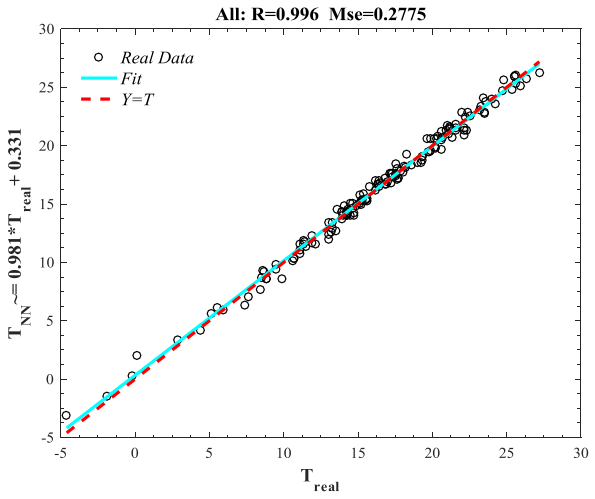

Figure 8. Regression diagram of all real data and modeling data (neural network)

If there is a difference between the performance of the network in the training and testing phase, we realize that network training has not been performed well. According to Table 1, the small error of the MSE for the test data (0.29) and their high correlation coefficient (close to 1) indicate successful neural network prediction. Figure 8 shows the regression diagram of the total experimental and modeling data.

End-of-training is based on the mean squared error control according to criteria. The MSE curve is called as a function of the iteration of the training algorithm. Another method is to monitor the MSE of the test and validation data. Training should stop when test data error begins to increase. At this point, the most generalization occurs.

To apply this method, the network must first be trained several times. Then, the weights should be fixed, and the network is evaluated with test data. Then, it is time to go back to the training dataset and continue training. This method is slightly overwhelming because, for each iteration of the training, it is necessary to calculate the performance of the network test data.

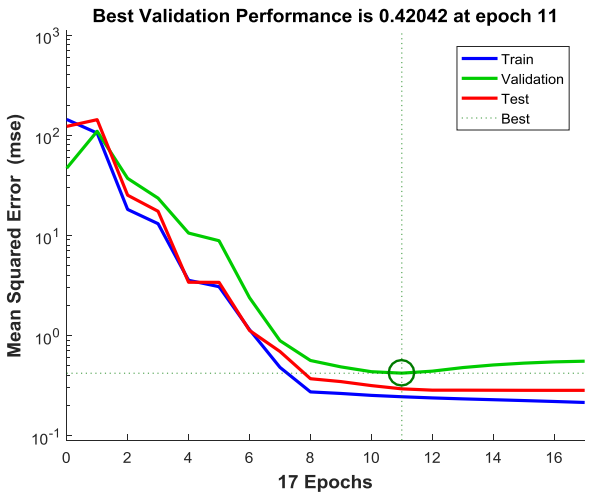

Figure 9. Diagram of the mean squared error (MSE) in different iterations of the training process

In Figure 9, the mean squared error during the training process has been shown. It is obvious that by increasing iteration, the error value is decreasing for all three categories. As the algorithm progresses at each iteration step, the mean squared error is calculated for the validation data. When the validation error is reduced, the algorithm will not stop, and the training will continue. Training is stopped when the validation error is not decreasing in 6 consecutive iterations. Increasing the number of iterations (more than 6) drives the network to overfit and increases test data error and reduces network generalization.

Finally, the point where the verification error has been minimized is considered as the output of the problem. In Figure 9, the validation error decreases before iteration No. 11 and then increases, while the training data error continues to decrease. From iteration No. 11 to

No. 17 (6 consecutive repetitions), the validation error is ascending, so the training algorithm terminates its operation, and the iteration No. 11 is considered as output. We should note that the algorithm calculates the mean squared error as follows:

$$MSE = \sum_{i=1}^{n} \frac{(T_{NN} - T_{\text{Real}})^2}{Ne} \quad (12)$$

In the above equation, T_{NN} is the hydrate formation temperature predicted using the neural network, T_{Real} is the real hydrate formation temperature, and Ne is the total number of samples. As indicated, Figure 10 shows the structure of the network used in which the input layer takes 3 types of data and predicts the output layer of data. According to the figure below, the number of hidden layer neurons is 7, the weight matrix and bias, as well as the activation function, are shown. As shown schematically, the activation function of the first layer is the hyperbolic tangent, and the second activation function is linear.

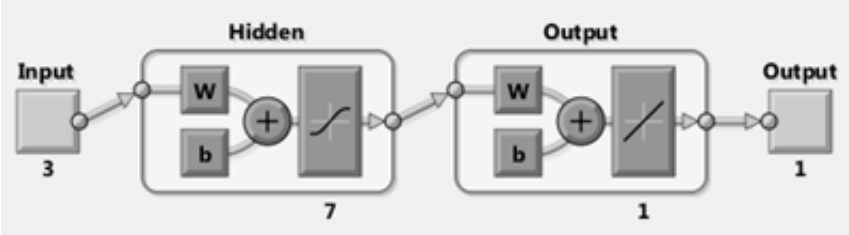

Figure 10. The neural network structure used in this study

4. Conclusion

An exact mathematical model has been presented based on the theory of artificial intelligence and neural network. An accurate model was developed by programming the neural network algorithm and feeding it by the laboratory data. The high correlation coefficient and low prediction error for the test data indicate that the obtained mathematical model can well predict the gas hydrate formation conditions. The present model predicts accurate results and is reliable in the temperature range from 268 to 301 K, the pressure between 116- 3500 psi, inhibition percentage between 5% to 15%, and specific weight between 0.530 to 0.995. The present model has better prediction accuracy than the existing experimental and quasi-experimental models with the mean absolute error percentage of MSE = 0.29.

The multilayer perceptron neural network has been used to model the gas hydrate formation conditions. The variables of natural gas pressure, specific density, the mass percentage of methanol in aqueous solution are the three input parameters of the neural network. The output parameter of the neural network is the gas hydrate formation temperature. After creating the neural network structure by a powerful and reliable database, the network has been trained. The results show that the hybrid artificial neural network can well predict the temperature of gas hydrate formation with high accuracy. The optimal number of hidden layer neurons (7 neurons) has been calculated by the genetic algorithm. The total number of data pieces used in the modeling is 156, 70% of which has been allocated to training, 15% for monitoring, and the rest for network testing. The numerical error values (MSE) for training, validation, and test data are 0.24, 0.42, and 0.29, respectively. Also, correlation coefficients for training, validation, and test data were 0.996, 0.998 and 0.995, respectively. As expected, the MSE error for the training data is minimum, and the regression coefficient is maximum. Due to the large size of the variable of the hydrate formation temperature, this error value is very small, and the modeling accuracy is acceptable. The small MSE error (0.29) and the high correlation coefficient (approximately close to 1) for the test data indicate that the neural network is successful in predicting the hydrate formation temperature.

In the following, the main relation of calculating the hydrate formation temperature has been presented based on the input variables. As you can see below, we can obtain the hydrate formation temperature as follows:

Multiply the weight matrix by the input matrix, add it to the first layer bias matrix, and then apply the hyperbolic tangent function to it, and multiply it by the second layer weight matrix, and finally, add a constant value.

$$T_{NN} = [-0.886 \quad +0.076 \quad +0.326 \quad -0.0096 \quad -0.0087 \quad 1.501 \quad -0.0105]$$
$$\times \tanh\left(\begin{bmatrix} -0.210 & -0.292 & -0.067 \\ +2.718 & 0.107 & +4.491 \\ +0.206 & -0.034 & -2.579 \\ +3.429 & -1.735 & +4.147 \\ -3.139 & +1.413 & -1.693 \\ +2.401 & +0.058 & +0.074 \\ +2.441 & +3.700 & +2.797 \end{bmatrix}\right) \times \begin{bmatrix} P \\ \gamma_g \\ w \end{bmatrix} + \begin{bmatrix} -0.216 \\ +4.533 \\ -0.094 \\ +0.436 \\ -0.958 \\ +2.831 \\ +1.691 \end{bmatrix} - 1.3654 \qquad (14)$$

Reference

[1] Suleimaniynazar A. M.Sc. Thesis, Isfahan University, Iran, 2009.

[2] Andersson GG, Kaufmann PJ. Development of a generalized neural network. Chemom. Intell. Lab. Sys., 2000; 50: 101-105.

[3] Balakin BV, Ph.D. Thesis, University of Bergen, Norway, 2016.

[4] Chapoy A, Mohammadi AH, Richon D. Predicting the Hydrate Stability Zones of Natural Gases Using Artificial Neural Networks. Oil Gas Sci., Technol., 2007; 62 (5): 701–706.

[5] Yarveicy H, and Ghiasi MM. Modeling of gas hydrate phase equilibria: Extremely randomized trees and LSSVM approaches. J. Mol. Liq., 2017; 243: 533-541.

[6] Kim AR, Kim HS, Cho GC, and Lee JY. Estimation of model parameters and properties for numerical simulation on geomechanical stability of gas hydrate production in the Ulleung Basin, East Sea, Korea. Quatern Int. J. 2017; 459: 55-68.

[7] Ghavipour M, Ghavipour M, Chitsazan M, Najibi SH, and Ghidary SS. Experimental study of natural gas hydrates and a novel use of neural network to predict hydrate formation conditions. Chem. Eng. Res. Des., 2013; 91 (2): 264-273.

[8] Jeong T, Byun J, Choi H, and Yoo D. Estimation of gas hydrate saturation in the Ulleung basin using seismic attributes and a neural network. J. Appl. Geophys., 2014; 106: 37-49.

To whom correspondence should be addressed: Dr. Niloufar Fatourehchi, Gas Research division, Research Institute of Petroleum Industry, Tehran, Iran, E-mail: fatourehchin@ripi.ir