DEPÓSITO LEGAL ppi 201502ZU4666
Esta publicación científica en formato digital es continuidad de la revista impresa
ISSN 0041-8811
DEPÓSITO LEGAL pp 76-654

Revista de la Universidad del Zulia

Fundada en 1947
por el Dr. Jesús Enrique Lossada

Ciencias del Agro Ingeniería y Tecnología

Año 11 N° 29
Enero - Abril 2020
Tercera Época
Maracaibo-Venezuela

Comparison of neural networks and genetic algorithms to determine missing precipitation data (Case study: the city of Sari)

Ali Mahdavi*
Mohsen Najarchi *
Emadoddin Hazaveie *
Seyed Mohammad Mirhosayni Hazave *
Seyed Mohammad Mahdai Najafizadeh *

ABSTRACT

Neural networks and genetic programming in the investigation of new methods for predicting rainfall in the catchment area of the city of Sari. Various methods are used for prediction, such as the time series model, artificial neural networks, fuzzy logic, fuzzy Nero, and genetic programming. Results based on statistical indicators of root mean square error and correlation coefficient were studied. The results of the optimal model of genetic programming were compared, the correlation coefficients and the root mean square error 0.973 and 0.034 respectively for training, and 0.964 and 0.057 respectively for the optimal neural network model. Genetic programming has been more accurate than artificial neural networks and is recommended as a good way to accurately predict.

KEY WORDS: neural networks, genetic algorithms, prognostic data, Sari.

*Department of Technical and Engineering, Arak Branch, Islamic Azad University, Arak, Iran, m-najarchi@iau-arak.ac.ir

Recibido: 13/02/2020 Aceptado: 10/03/2020

Comparación de redes neuronales y algoritmos genéticos para determinar datos faltantes de precipitación (Estudio de caso: la ciudad de Sari)

RESUMEN

Redes neuronales y programación genética en la investigación de nuevos métodos para la predicción de precipitaciones en el área de captación de la ciudad de Sari. Para la predicción se utilizan varios métodos, como el modelo de series temporales, redes neuronales artificiales, lógica difusa, Nero difuso y programación genética. Se estudiaron los resultados basados en los indicadores estadísticos de error cuadrático medio y coeficiente de correlación. Se compararon los resultados del modelo óptimo de programación genética, los coeficientes de correlación y la raíz del error cuadrático medio 0.973 y 0.034 respectivamente para entrenamiento, y 0.964 y 0.057 respectivamente para el modelo de red neuronal óptima. La programación genética ha sido más precisa que las redes neuronales artificiales y se recomienda como una buena forma de predecir con precisión.

PALABRAS CLAVE: redes neuronales, algoritmos genéticos, datos de pronóstico, Sari.

Introduction

Factors affecting water resources systems analysis, including precipitation predicted in one of the major issues in the design, operation and studies relating to these sources is considered. For this purpose, various methods such as artificial neural networks, fuzzy logic, Genetic programming and time series.

Circulating among the genetic programming algorithm is a method that is based on Darwin's theory of evolution is based on. Mentioned algorithm to define an objective function in the form of quantitative criteria and then mentioned function to compare different solutions to solve the problem in a step by step process to apply correct data structure and finally, providing the perfect answer. Genetic Programming is a recent development in the methods of evolutionary algorithms due to sufficient accuracy, the application is more (Feuring and Golubski, 2000).

Khu and colleagues (Khu ST, 2001) in a study on the catchment Avrgval in France, Genetic programming used to predict runoff and results with observed data and calculated values were analyzed by classical methods. The result indicates the accuracy of genetic programming is acceptable.

Lee Jung et al (Liong SY, 2002) studied the relationship between rainfall-runoff that concluded that the use of genetic programming algorithm to predict the behavior of precipitation - runoff in watershed will cause fewer errors. (Aytek and Kisi , 2008), a study on the phenomenon of sediment transport in streams, genetic programming as a suitable approach for modeling of suspended sediment were introduced. According to research (Aytek and Kisi, 2008; Aytek, Asce and Alp, 2008). Genetic Programming a suitable method and the action is in anticipation of rainfall-runoff relationship (Liong et al., 2002). Genetic Programming with incomplete data to estimate the height of waves in the Gulf of Mexico found the method of accuracy in the prediction of time series is very good.

A parallel processing system using artificial neural networks mimic the human brain are very simple biological nervous system. This emulation is based on a mathematical configuration, so that consists of several layers of neurons in each layer. The performance of these networks is thus an input layer, and acting inappropriately. And layers (layers) between the data processing and finally, the output layer to output resulting from application of its models.

(Abrahart and See, 2000) in a research area Wi-Vavs a comparative study of artificial neural networks method and ARMA model to predict the flow of the river is dealt. Comparative benchmark results show that the neural network capable of simple results than ARIMA model with the same input data generated. Kisi (2004) in a research Gvsvdr basin located in the State Issaquah Japan of artificial neural network and multiple regression to predict the monthly flow. The results showed higher accuracy than regression models were artificial neural networks (Kisi, 2005) in an investigation into Blackwater and Gila basin using artificial neural network models to predict the flow of the river's Atoregrsion. The results showed higher accuracy than the artificial neural network approach is Atoregrsion. Dogan et al (2006) in a research on Sakarya basin of two artificial neural networks and Atoregrsion used to predict the flow of daily revolt. The results show high accuracy recurrent neural network model to predict the course of daily revolt. (Firat M, 2007) in a study on watershed in Turkey's Cihan neuro-fuzzy methods, artificial neural networks feedforward neural network generalized multiple regression to predict daily river flow data can be used. Results showed superiority over the other two methods are neuro-fuzzy approach.

The aim of this study was to predict rainfall data at the basin city of Sari using genetic programming and artificial neural network model and compare the results of one of the most accurate method is common as is.

1. Methods

1.1. Genetic Algorithm

In genetic algorithm, first available block that includes the purpose and function of the input variables and connecting them, defined and reasonable structure model and its coefficients are determined. It involves a link between input variables and output equation and thus able to automatically select the appropriate variable and remove the variable is unrelated This will reduce the size of input variables. Select the appropriate inputs one of the most important things that need to be addressed in this way. It is also used in the absence of secondary input data, it will be more important; Because they provide unbiased input data, reduces the accuracy of the model and create more complex models That interpretation is faced with more difficult.

In engineering applications, the genetic programming is widely used in modeling to determine the structure of the phenomenon comes into action.

Genetic Programming for process step by step the following steps (Willis et al, 1997):

1. An initial population of combined functions of predictive models, considered to be random.

2. Any member of the population using the fitting functions, are evaluated.
3. In each production, the following steps will be taken to select a new population:

a) A cross operator, mutation and selection can be copied.
b) A good number of people in the crowd are selected.
c) The selection operator is used to produce offspring.
d) Son is mentioned in a new population.
e) The model is evaluated using different brushes.

4. The third step is to achieve the maximum number of publications, will be repeated.

In this way at the start of any relationship is not intended function and this method is able to optimize the structure of the model and its components.

Genetic Programming for Computational step by step flowchart shown in Figure 1.

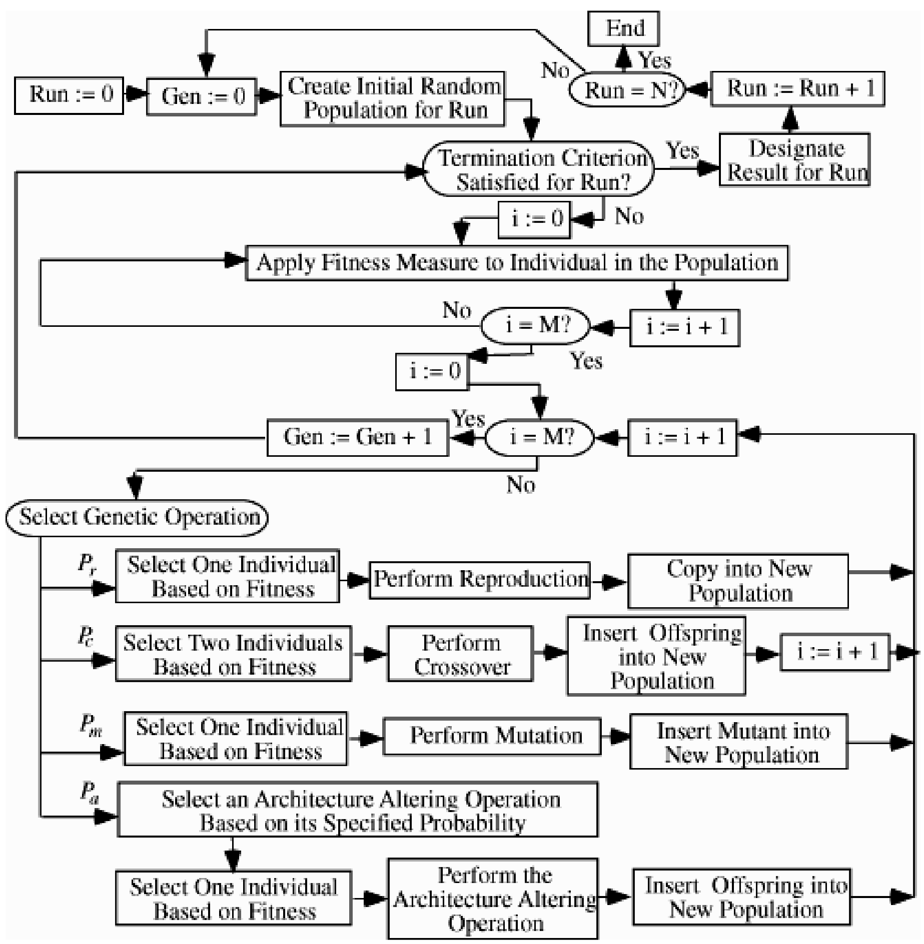

Figure 1. Flowchart genetic programming (Kozo, 1992)

1.2. Artificial Neural Networks

These networks mimic the neural networks in organisms and using a large number of artificial neurons are interconnected to perform the necessary calculations. A neural network consists of several nodes or compute nodes and In cases where the weighted inputs is used, The nodes are able to produce appropriate outputs will be using functions. Each layer may be formed from multiple neurons and neural network also includes one or more layers will be connected together. A three-layered structure of a network that consists of an input layer I, a hidden layer H and an output layer O. The above description is given in Figure 2.

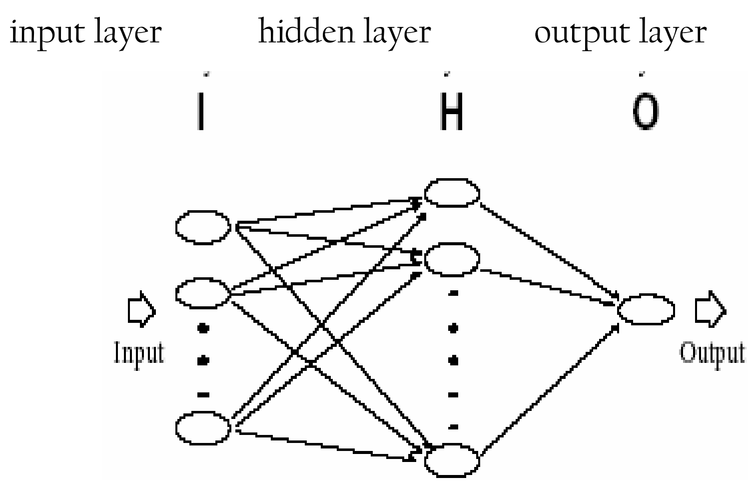

Figure 2. Structure of an artificial neural network (Alvisi et al., 2005)

The main elements of a neural network, or artificial neurons are nerve. Input pattern to a node dendritic cells is biologically similar That it can be a vector with N elements as indicated whit $X = (x_1, x_2, ..., x_n)$. Sum of the weights of fellow inputs can be scalar quantity S displayed.

$$s = \sum_{n=1}^{N} w_n x_n = W^T X \qquad (1)$$

Where $W = (w_1, w_2, ..., w_n)$ is the weight vector of the neuron. Quantity s Then f is a nonlinear function to output result:

$$y = f(s) \quad (2)$$

Non-linear transfer function is usually in the form of sigmoid function is defined as follows:

$$f(s) = (1 + \exp(-s))^{-1} \quad (3)$$

The output can be the result medal, or input is the next layer multi-layer network.Various algorithms to calculate the optimal weights of the algorithms presented in the "propagation" which is the most widely used. Networks that were used in this study process modeling is a Multilayer Perceptron network. The leading network for action and the Structural optimal design as much as possible try to take a middle layer. Multilayer Perceptron network training using back propagation algorithm was used.

Import of raw data to reduce network speed and accuracy. For authentication of such circumstances and in order to equalize the value data for the network, the normal practice was that this would shrink the excessive weight and early saturation is inhibiting the neurons. Frequently, data normalization in the range (0.9 and 0.1) takes place. In this study, all data input before applying to the network were normalized using the following equation:

$$X_{normal} = 0.1 + 0.8 * \left(\frac{X_0 - X_{\min}}{X_{\max} - X_{\min}} \right) \quad (4)$$

In this relation X_{normal} is the normal amount, X_0 is real amount, $X_{\max}$ and $X_{\min}$ The data are the maximum and minimum values respectively. The artificial neural network model data for training and testing are divided into two categories, In this study divided into two categories: training and testing patterns, respectively, 75 percent and 20 percent considered.

The performance of these networks is thus an input layer, action to accept data and layers (layers) between the data processing and finally, The output layer to output resulting from the application of the model. During the modeling coefficients related to errors in nodes for trial and error corrected which in most cases mean error profile data is used. It does so by comparing output model the performed with data input observations.

$$R = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}} \tag{5}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - x_i)}{N}} \tag{6}$$

In the last equation, R is the correlation coefficient, $RMSE$ is Root Mean Square Error, x_i is The observed value at time step, y_i is The amount calculated in the same time, N is The number of time steps, the mean of observations and $\bar{y}$ The mean values is calculated.

1.3. Case study

Sari in Mazandaran province in northern Iran, the largest and most populous city in Mazandaran province. Sari, located in the foothills of the Alborz mountain range has two parts is mountainous and plain In longitude and latitude 36 degrees 53 degrees 5 minutes and 4 minutes. Tajan River passes through the City of Surrey. The river flows in East Surrey. Zaremrod of 3 main branch of the river in the north and the main branch of Tajan in central and southern branch of the White River basin is formed and, moreover, they are on their way tributaries such as the Valley Babrcheshmeh, Pelaroy, ShirinRood, Salar Dareh, Tirjari water and one of them receives. The study and forecast data on precipitation is very important.

Figure 3. Location city Sari

2. Results

According to various studies done on the effectiveness of artificial neural networks and genetic programming models predict precipitation data, Clearly, the ability of these techniques according to different patterns and structures and The nature of the problem we solve is different.

Surrey city station rainfall data from 1971 to 2014 and was used for this study. In Table 1, comparable statistical indicators related to the results of the application of genetic algorithm to predict rainfall data is given.

Table 1. Statistical analysis accuracy capability

Test			Training			
RMSE	R^2		RMSE	R^2	Run number	Input pattern
0/043	0/925		0/045	0/937	200	1
0/035	0/969		0/038	0/971	200	2
0/033	0/971		0/034	0/973	200	3

From the above results it can be concluded that the results are gradually improving and is a convergent answer.

In the present study to investigate the accuracy of genetic programming, artificial neural network model with the same input patterns with input data were normalized. For each input pattern by changing the number of hidden layers and the number of hidden layer neurons, Different neural networks were trained and finally a structure that has the fewest errors, was chosen as the most suitable structure.

Table 2 shows the most appropriate structure for each input pattern. For each input pattern, Structure with five neurons in a hidden layer and an output layer neurons in the best structure has been sigmoid function of the type of stimuli used. In between these structures, such as genetic algorithms to improve with fewer errors and higher correlation coefficient is the fourth structure.

Table 2. Statistical analysis of the accuracy of artificial neural networks

Training		Test			
RMSE	R^2	RMSE	R^2	Structure	Input pattern
0/048	0/945	0/049	0/948	2-6-1	1
0/051	0/958	0/055	0/956	3-6-1	2
0/055	0/960	0/057	0/964	4-6-1	3

Following the results of the two methods is shown as a graph.

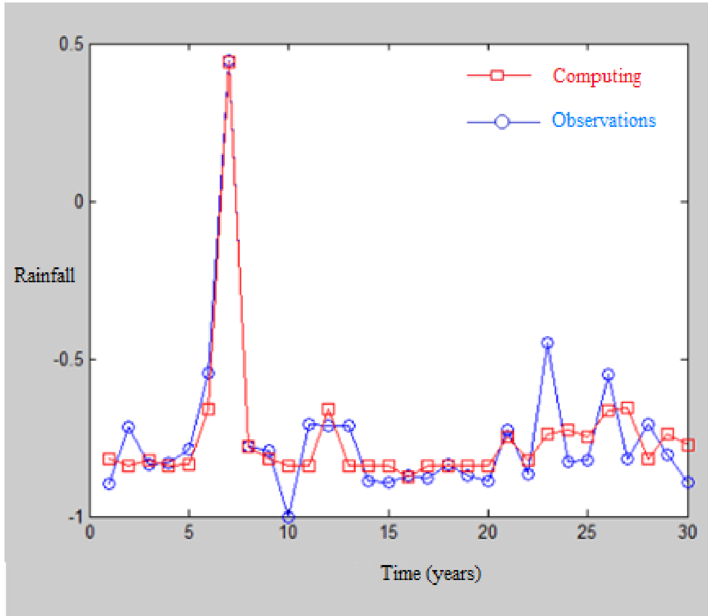

Figure 4. Results related to neural networks (Training)

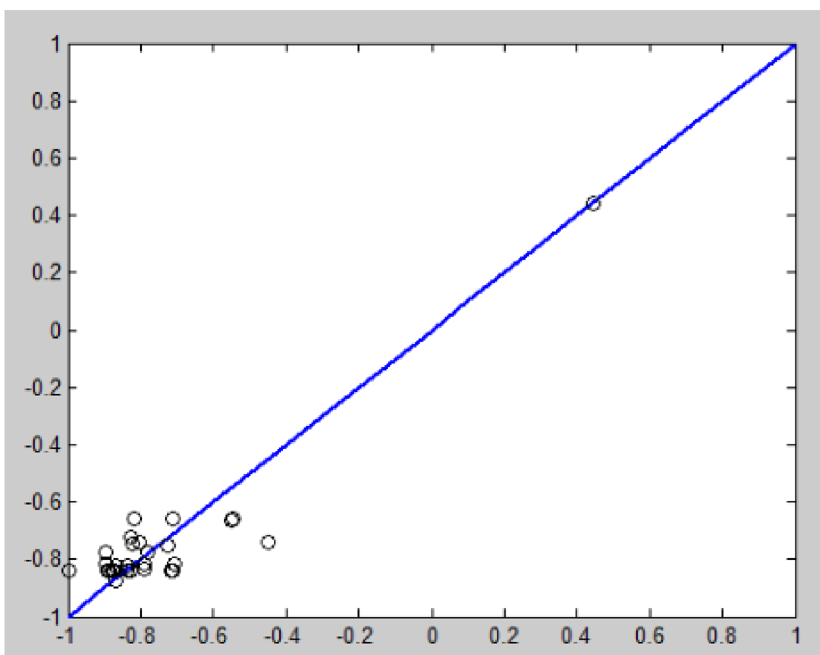

Figure 5. The regression line for Training mode in the Neural Network

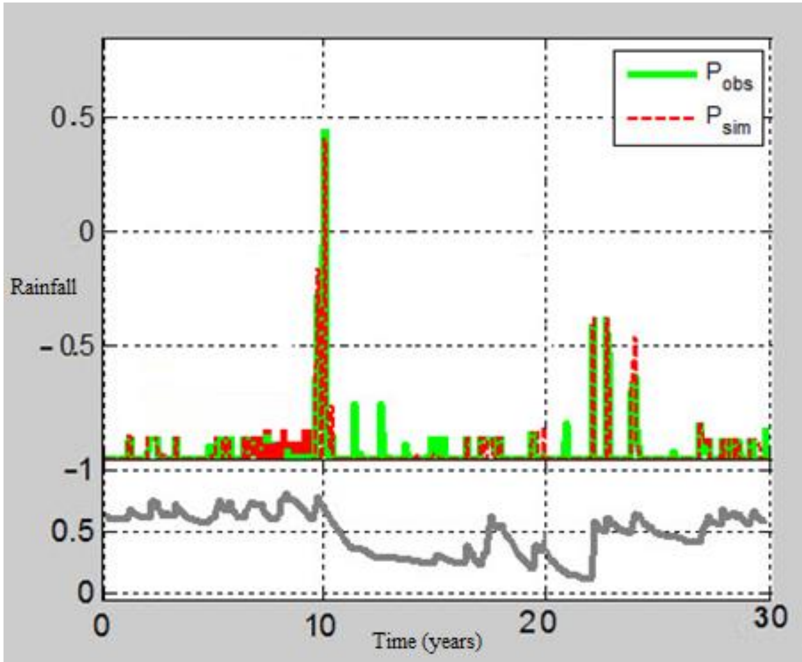

Figure 6. Results of genetic algorithm

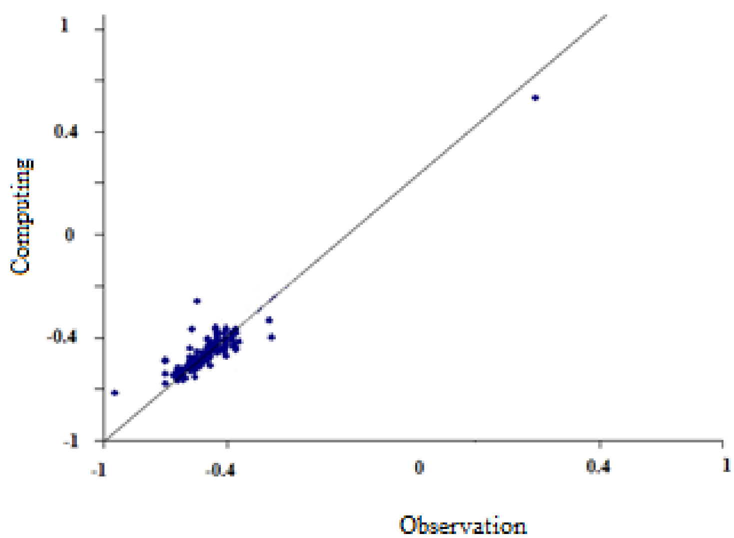

Figure 7. The regression line for the training in Genetic Algorithm

In Figure 8 Comparison of Two Methods used in the study are shown in the diagram.

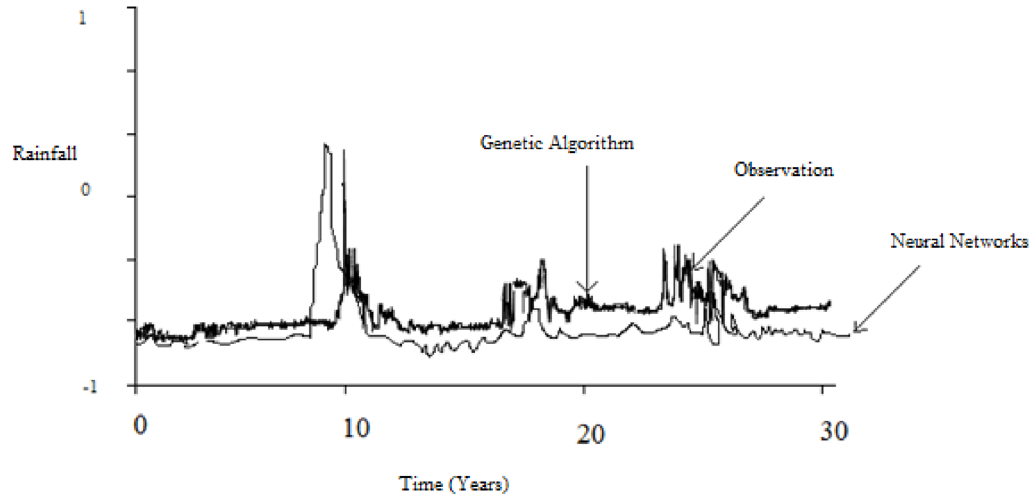

Figure 8. Compare computational methods with the observed data

The results showed that in terms of the values in Table 1 and 2 respectively genetic algorithm is more accurate than using artificial neural networks. In Figure 9 correlation coefficient for the two methods that reflect the above explanation is given for comparison.

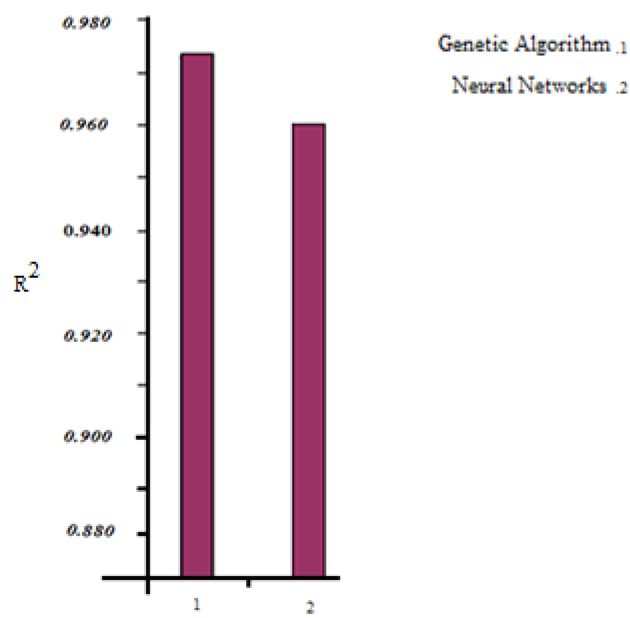

Figure 9. The correlation coefficient for the methods used

Conclusion

Given that in this study, two methods were used to predict after calculating the correlation coefficients and root mean square error of the results of both methods are acceptable. For genetic algorithm correlation coefficient and root mean square error 0.973 and 0.034 respectively for training and artificial neural network model for optimal 0.960 and 0.055 respectively were obtained. According to the results we can say that genetic algorithm is more accurate than using neural networks.

In a study of (Aytek and Kisi, 2008) conducted a genetic programming in modeling sediment extremely high accuracy relative to the regression model showed and Khu et al. (2001) took advantage of genetic programming for runoff prediction and results with observed data and calculated values were compared by classical methods. The result is an expression of ultra-high precision genetic programming. The results with respect to the accuracy of model genetic programming is consistent with the results of the above studies.

Reference

Abrahart R.J. and See L. (2000). Neural network vs ARMA modeling : Constructing benchmark case study of river flow prediction, Pp. 1021-1028. 3rd International Conference on Hydroinformatics. Copenhagen, Denmark.

Alvisi S., Mascellani G., Franchini M. and Bardossy A. (2005). Water level forecasting through fuzzy logic and artificial neural network approaches. J Hydrol Earth Syst Sci 2: 1107-1145.

Aytek A., Asce M. and Alp M., (2008). An application of artificial intelligence for rainfall runoff modeling. J Hydrol Earth Syst Sci 117(2): 145-155.

Aytek, A. and Kisi, O. (2008). A genetic programming approach to suspended sediment modeling. J Hydrol Eng 351: 288-298.

Dogan E., Isik S., Toluk T. and Sandalci M. (2007). Daily streamflow forecasting using artificial neural networks. Pp. 448-459. International Congress River Flood Management. Ankara, Turkey.

Feuring, T.; Golubski, W. (2000). Evolving Neural Network Structures. M. Mohammadian (Ed.). New Frontiers Computational Intelligence and its Aplications. IOS Press, 2000.

Firat M. (2007). Artificial intelligence techniques for river flow forecasting in the Seyhan river catchment, Turkey. J Hydrol Earth Syst Sci 4: 1369-1406.

Kisi O. (2004). River flow modeling using artificial neural networks. J Hydrol Eng 9(1): 60- 63.

Kisi O. (2005). Daily river flow forecasting using artificial neural networks and autoregressive models. J Eng Env Sci 29: 9-20.

Khu S.T., Liong S.Y., Babovic V., Madsen H. and Muttil N. (2001). Genetic programming and its application in real- time runoff forming. J AmWater Res Assoc 37(2): 439-451.

Kozo J. (1992). Genetic programming on the programming of computers by natural selection. MIT Press, Cambridge, MA.

Liong S.Y., Gautam T.R., Khu S.T., Babovic V., Keijzer M. and Muttil N. (2002). Genetic programming: A new paradigm in rainfall runoff modeling. J Am Water Res Assoc 38(3): 705-718.

Willis, Mark & Hiden, Hugo & Marenbach, P. & McKay, Ben & Montague, Gary. (1997). Genetic programming: An introduction and survey of applications. 314 - 319. 10.1049/cp:19971199.