ORIGINAL PAPER

Earthquake magnitude prediction by adaptive neuro-fuzzy inference system (ANFIS) based on fuzzy C-means algorithm

Masoomeh Mirrashid

Abstract   This paper investigates the prediction of future earthquakes that would occur with magnitude 5.5 or greater using adaptive neuro-fuzzy inference system (ANFIS). For this purpose, the earthquake data between 1950 and 2013 that had been recorded in the region with 2°E longitude and 4°N latitude in Iran has been used. Thereupon, three algorithms including grid partition (GP), subtractive clustering (SC) and fuzzy C-means (FCM) were used to develop models with the structure of ANFIS. Since the earthquake data for the specified region had been reported on different magnitude scales, suitable relationships were determined to convert the magnitude scales into moment magnitude and all records uniformed based on the relationships. The uniform data were used to calculate seismicity indicators, and ANFIS was developed based on considered algorithms. The results showed that ANFIS-FCM with a high accuracy was able to predict earthquake magnitude.

Keywords   Adaptive neuro-fuzzy inference system (ANFIS) · Earthquake prediction · Moment magnitude · Fuzzy C-means algorithm (FCM)

1 Introduction

Earthquake prediction is one of the most difficult problems in the world. Many efforts have been made in earthquake prediction based on artificial intelligence. The studies would be classified into two groups of those based on neural networks and those based on fuzzy systems. Neural networks are much more popular techniques than fuzzy systems in earthquake prediction, and the most efforts in predicting have been made by them. Yudong et al. (1994) investigated the artificial neural network method for prediction of the type of earthquake sequences. The results show that the performance of the artificial neural

M. Mirrashid (✉)
Civil Engineering Department, Gonbad Kavoos Branch, Islamic Azad University, Gonbad Kavoos, Iran
e-mail: masoomeh_mirrashid@hotmail.com

network approach is good (Yudong et al. 1994). BP neural networks are used to mid-term earthquake prediction by Wang et al. (2000). They applied this method to space scanning of North China. The result shows that the mid-term anomalous zone usually appeared obviously around the future epicenter 1–3 years before earthquake (Wang et al. 2000). Liu et al. (2005) proposed a method called constructive ensemble of RBF neural networks (CERNN), in which the number of individuals, the number of hidden nodes and training epoch of each individual are determined automatically. Experiments on datasets demonstrate that CERNN can be applied to earthquake prediction (Liu et al. 2005). Alves (2006) used an earthquake forecasting method that integrated in a neural network several forecasting tools that had been originally developed for financial analysis. This method was tested with the seismicity of the Azores, predicting the July, 1998, and the January, 2004, earthquakes, albeit within very wide time and location windows (Alves 2006). Neural networks are investigated for predicting the magnitude of the largest seismic event in the following month based on the analysis of eight mathematically computed parameters known as seismicity indicators by Panakkat and Adeli (2007). The problem is modeled using three different neural networks: a feed-forward Levenberg–Marquardt back-propagation (LMBP) neural network, a recurrent neural network, and a radial-basis function (RBF) neural network. The models are trained and tested using data for Southern California and the San Francisco bay region. Overall the recurrent neural network model yields the best prediction accuracies compared with LMBP and RBF networks (Panakkat and Adeli 2007). Adeli and Panakkat (2009) investigated the application of a probabilistic neural network for predicting the magnitude of the largest earthquake in a pre-defined future time period in a seismic region using eight mathematically computed parameters. The model is trained and tested using data for the Southern California region. The model yields suitable prediction accuracies for earthquakes of magnitude (Adeli and Panakkat 2009). Sri Lakshmi and Tiwari (2009) studied artificial neural network methods based on the back-propagation algorithm to predict behavior of earthquake dynamics in the crucial tectonic regions of Northeast India (Sri Lakshmi and Tiwari 2009). Panakkat and Adeli (2009) presented a recurrent neural network for predicting the location and time of occurrence of future moderate-to-large earthquakes based on neural network modeling and using eight seismicity indicators as input. Seismicity indicators are computed, and their relationship to the latitude and longitude of the epicentral location and time of occurrence of the following major earthquake is studied (Panakkat and Adeli 2009). Chattopadhyay and Chattopadhyay (2009) estimated the magnitude of earthquake over Indian subcontinent using artificial neural network with back-propagation learning. The day, month, and year of occurrence of the earthquake, latitude of the place, and the longitude of the place are considered as the predictor variables. It has been found that the nonlinear multilayer perceptron (MLP) produces estimates of earthquake magnitude with error <20 % (Chattopadhyay and Chattopadhyay 2009). RBF neural network is used to predict the magnitude of earthquake by Ying et al. (2009). Firstly, RBF neural network is used to learn the data which include the information of earthquake. Then, they use the trained RBF neural network to predict the test samples. The inevitable results indicate that the method has certain application value (Ying et al. 2009). Yi et al. (2010) used a method which momentum term with adaptive momentum factor was introduced into the gradient descent method to modify the parameters of RBF neural network. Combined with the improved RBF neural network model, the seismic time series were predicted (Yi et al. 2010). Sheng-Zhong (2010), investigated the probabilistic neutral network to predict the magnitude of the serious earthquake in future time in a seismic area, depending on mathematically computed parameters as seismicity indicators. The model can be used to

predict earthquake magnitude (Sheng-Zhong 2010). Artificial neural networks have been applied to learn the cyclic behavior of seismicity in the independent seism genic sources to predict their future trends by Sharma and Tyagi (2010). The outcome of the ANN is used to interpret the future seismicity of the future seismicity cycles (Sharma and Tyagi 2010). Mohsin and Azam (2011) used both deterministic and un-deterministic optimized algorithms to determine the future earthquake in Pakistan. Shah et al. (2011) studied the application of ABC algorithm that simulates the behavior of a honey bee swarm. MLP trained with the standard back-propagation algorithm (Shah et al. 2011). Xie et al. (2011) used the seismicity variation rate as the input and create a neural network model for prediction based on the earthquake data during the past period of time in East China. Yaghmaei-Sabegh (2012) used an approach for ranking of ground motion prediction equations named generalized regression neural networks (GRNN) as a one-pass learning algorithm, and an effective type of radial-basis neural network was chosen. The results showed potential of designed GRNN for ranking of ground motion prediction equations (Yaghmaei-Sabegh 2012).

Although neural networks enjoy great popularity, the solution that network has learned cannot be expressed explicitly and for the user it is a black box. On the other hand, in fuzzy systems against neural networks, it is possible to formulate rules which are based on linguistic expressions and to apply them for learning processes. There are few researches on earthquake prediction that have been done based on fuzzy models. Wang et al. (1997) investigated the application of fuzzy associative memory (FAM) network model for earthquake prediction. This system has functioned for knowledge learning without disadvantages of neural network, which the learned knowledge implied in network is difficult to be interpreted by expert system (Wang et al. 1997). A neural network model named impulse force based on ART neural network (IFART) is presented, and IFART network is applied to predict magnitude of earthquake by Liu et al. (2004). The principle, learning and rules of fuzzy artificial neural network model, is introduced to predict seismic trend after main shock by Cong et al. (2006). Andalib et al. (2009) used a methodology for the development of a fuzzy expert system with application of earthquake prediction. The rules were created by an expert to generate a fuzzy rule base. The model was employed to attain the performance of an expert who used to predict earthquakes in the Zagros area in Iran based on the idea of coupled earthquakes (Andalib et al. 2009). Zhong and Zhang (2010) obtained the risk class through fuzzy theory by one candidate assessment unit into consideration. A fuzzy mathematical prediction model of three-gorge reservoir in China has been created. It is concluded that compared with other mathematical prediction models, the reservoir-induced earthquake prediction model based on the fuzzy theory can analyze data with fuzzy processing (Zhong and Zhang 2010).

It is sometimes difficult to specify parameters of a fuzzy system for earthquake prediction. Several factors have an impact on this phenomenon, and determining the fuzzy rules and function parameters has not been quite possible by an expert human. The combination of neural networks and fuzzy systems which is called neuro-fuzzy can help to solve this problem and can be enhance the performance of the model. In this paper, the application of one of the most powerful neuro-fuzzy systems with the purpose of predicting magnitude of the future earthquakes of equal to or greater than 5.5 is investigated. A seismic region in Iran was chosen as the study zone, and the data of earthquakes occurring between 1950 and 2013 were collected to develop a predictive ANFIS model.

2 Dataset

To evaluate the ability of each model by artificial intelligence a set of data that considered to train and test the system is needed. In this paper, the information of recorded earthquakes in International Institute of Earthquake Engineering and Seismology (IIEES) has been used. This database is available for free download at www.iiees.ac.ir. The collected data consists of 1958 earthquakes which are recorded in the region with 2°E longitude and 4°N latitude between 1950 and July 10, 2013.

2.1 Uniformization of magnitude quantities

Lack of uniformity in the reported magnitude scales in the IIEES led to determining relationships to uniformize numbers of magnitude. In order to convert magnitude scales, reliable database was used for the earthquakes occurred between 1950 and 2013 in the specified region and those records were taken that the moment magnitude scale and one of the other magnitude scales were reported. The International Seismological Center (ISC) was reference for mb and Ms. The last reported event that has been recorded in two scales was on February 6, 2010. Therefore, for earthquakes that occurred after that date, reference of National Earthquake Information Center (NEIC) was used. For the local magnitude scale, IIEES as a local reference in Iran was considered. Harvard Seismological Center (HRVD) was used for moment magnitude, and for other records that HRVD was not reported Mw, the Global Centroid Moment Tensor (GCMT) was taken into consideration.

2.2 mb to Mw conversion

Additionally, there are 20 records that both scale mb and Mw are reported by ISC. The range of recorded magnitude for mb was 4.6–5.9 and for Mw was between 4.8 and 6.4. Table 1 shows the information of the record.

Orthogonal and least squares regressions and based on distribution of the available values, quadratic regression was performed to determine the best relation between mb and Mw (Table 2). Parameters R^2, MAE, and RMSE, respectively, are correlation coefficient, mean absolute errors, and root mean squared errors. The results showed that the least squares regression has better performance than orthogonal regression. Among these, the strongest relation which has acceptable correlation coefficient of 83 % is the quadratic regression. The relation of this regression has the minimum error RMSE and MAE in comparison with the results of other regressions. That is also visible from the distribution of points in Fig. 1. However, for smaller values of the interval mb whose records are less than 4.6, the Eq. (1) cannot be used. The obtained results for lower values of the range mb from the equation of the curve 1 will be far from reality. The changes of value in the range of <4.6 are also unclear. Therefore, for the amounts of upper than 4.6, the Eq. (1) was

Table 1 The number of taken records for converting mb to Mw

Scale	Author			
	ISC	NEIC	GCMT	HRVD
mb	18	2	–	–
Mw	–	–	7	13

Table 2 The obtained equations and their results from the regressions for converting mb to Mw

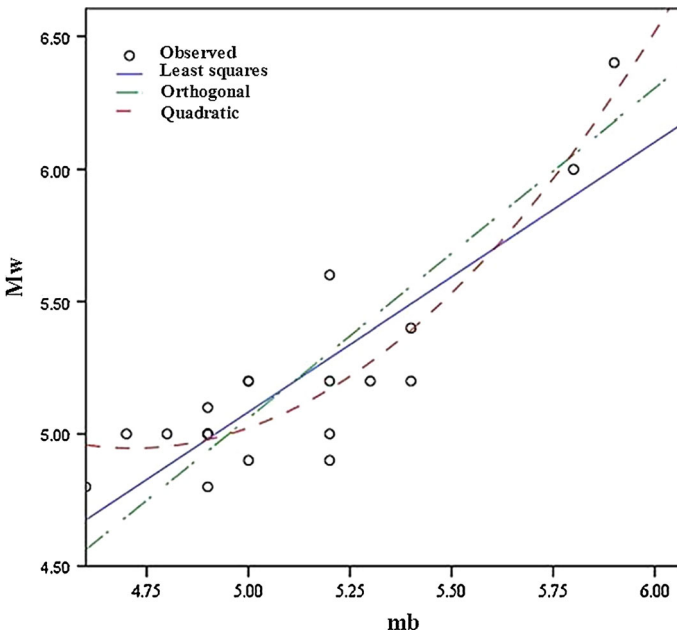| Number of equation	Regression type	Relation	R^2	RMSE	MAE
Eq. 1	Quadratic	Mw = 26.111 − 8.977 mb + 0.952 mb^2	0.83	0.16	0.12
Eq. 2	Least squares	Mw = 1.02 mb − 0.0187	0.73	0.20	0.17
Eq. 3	Orthogonal	Mw = −1.160 + 1.244 mb	0.69	0.22	0.19

Fig. 1 The applied regressions for converting mb to Mw. The *circles* in the figure show the amounts of mb and Mw for 20 available records

selected, and for the range of <4.6 it was assumed that the Eq. (2) is proportional to the slope of the changes and provided better results than Eq. (3) was used.

2.3 Ms to Mw conversion

To convert Ms to Mw, 19 records including magnitudes between minimum 3.9 and maximum 6.4 magnitudes were reported. Scale of Mw was also between 4.8 and 6.4. Table 3 shows information of the records.

From the applied regression analysis for converting Ms to Mw, least squares regression and growth function form regression provided the best results. Table 4 shows the results of the correlation coefficient, RMSE, and MAE for these regressions.

The presented equations in Table 4 have very similar results. It seems that Eqs. (5) and (6) are more suitable equations due to lack of complexity of the data distribution and are easier than Eq. (4). Artificial intelligence models produce results based on values that have been given to them. The closer the data are to reality, the stronger the results are. The

Table 3 The number of taken records for converting Ms to Mw

Scale	Author			
	ISC	NEIC	GCMT	HRVD
Ms	19	–	–	–
Mw	–	–	5	14

Table 4 The obtained equations and their results from the regressions for converting Ms to Mw

Number of equation	Regression type	Relation	R^2	RMSE	MAE
Eq. 4	Growth	Mw = EXP (1.134 + 0.11 Ms)	0.98	0.062	0.050
Eq. 5	Least squares	Mw = 2.369 + 0.609 Ms	0.97	0.072	0.055
Eq. 6	Orthogonal	Mw = 2.287 + 0.627 Ms	0.97	0.068	0.053

selected equation in this section will be applied directly to calculate the seismic indicators which are used in modeling and because of closer obtained results of the regressions, additional criteria for the selection proper equation between the presented relationships are determined. If it is specified that the linear Eqs. (5) and (6) have less error than Eq. (4) then Eq. (6) is chosen because of its better results than Eq. (5). For this purpose, analysis of variance and standard deviation has been done. The results showed that Eq. (4) was also better than linear equations. According to Fig. 2 distribution of the observed points was in a linear form. Moreover, according to Table 5, the standard deviation coefficients Eq. (4) is less than linear equation. Therefore, to convert the scale Ms to Mw, Eq. (4) was used.

2.4 ML to Mw conversion

There were only 11 records to convert the recorded values of the scale ML to Mw (Table 6). Range of magnitude for ML was reported between 4.8 and 5.1 and between 4.8 and 6 for Mw. It should be noted that IIEES that is an Iranian reference is used to take data with local magnitude scale ML.

Due to the large number of records that have been recorded with ML scale in recent years for the chosen region, determining of the relationship between ML and Mw was very important. Selecting an inappropriate relationship would have bad effects on the results of the model. Several regressions were used to determine the best relationship. The results are shown in Table 7. Due to the distribution of points in Fig. 3 and the values for the correlation coefficient, MAE and RMSE, Eq. (7) seems more appropriate. However, because the range of variation ML for 11 reviewed records was more than 4.8 and the most recorded earthquakes had ML value less than this amount, Eq. (7) was selected to convert ML to Mw only for amounts greater than 4.8. The obtained slope changes in the Eq. (7) show that if the goal was to determine the corresponding value of Mw for magnitudes <4.8, the amounts of Mw would change in a little closer to 5 and it could not be acceptable for all of the considered range, and therefore, Eq. (7) cannot provide a good prediction for small amounts. It is also true for the quadratic regression in a way that using this equation for small amounts will show large amount that is visible from the shape and slope curve of
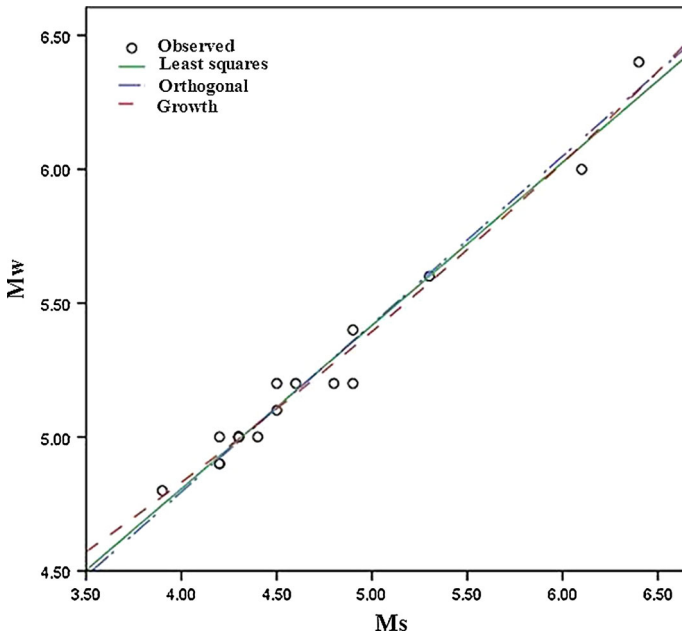
Fig. 2 The applied regressions for converting Ms to Mw. The *circles* in the figure show the amounts of Ms and Mw for 19 available records

Table 5 The results of variance and standard deviation analysis for choosing Ms to Mw

Type of regression	Residual of regression	Mean square	Std. error of the estimation	Std. error of coefficients	
				Constant	Ms
Growth	0.002	0.091	0.012	0.021	0.004
Least squares	0.081	2.784	0.071	2.369	0.609

Table 6 The number of taken records for converting ML to Mw

Scale	Author		
	IIEES	GCMT	HRVD
ML	11	–	–
Mw	–	6	5

the Eq. (7). Therefore, if we assume that the slope of the linear regression is approximately true for earthquakes that were recorded with a magnitude value <4.8, Eq. (8) can be used to convert ML to Mw. It is worth mentioning that according to Table 7, Eq. (8) has more correlation and less error than Eq. (9); therefore, it is more suitable for converting values of <4.8.

Table 7 The obtained equations and their results from the regressions for converting ML to Mw

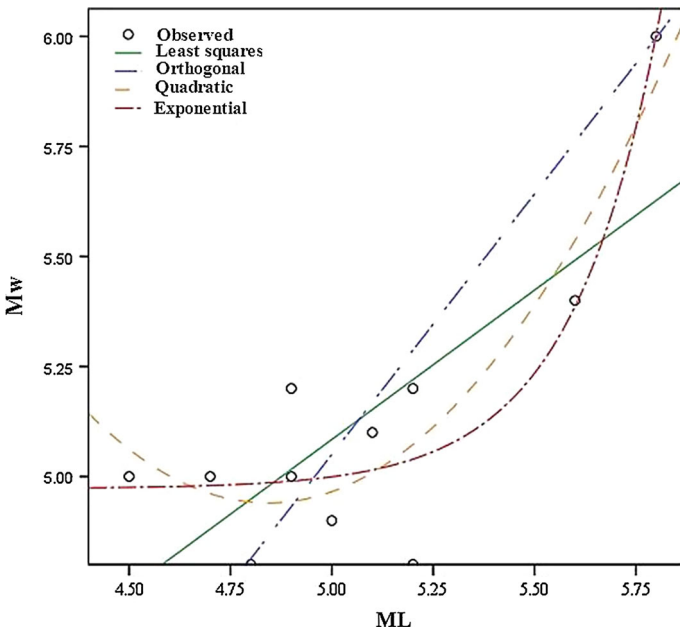| Number of equation	Regression type	Relation	R^2	RMSE	MAE
Eq. 7	Exponential	Mw = 4.972 + 3.331e−012 EXP(4.561 ML)	0.84	0.13	0.09
Eq. 8	Least squares	Mw = 1.691 + 0.678 ML	0.57	0.21	0.17
Eq. 9	Orthogonal	Mw = −0.87 + 1.184 ML	0.25	0.28	0.21
Eq. 10	Quadratic	Mw = 29.32 − 10.071 ML + 1.04 ML^2	0.80	0.14	0.12

Fig. 3 The applied regressions for converting mb to Mw. The *circles* in the figure show the amounts of ML and Mw for 11 available records

2.5 Scale conversion

After the relationship was determined, all 1958 data taken from IIEES were converted to moment magnitude. Table 8 shows the scale conversion results and the number of values in each of the studied decades. It is clear that the number of recorded earthquakes which had been recorded before 1980 consisted of a small number of events due to the lack of seismic instruments recorded in different parts of the region. According to Table 8, the area that has been studied in this paper and had five earthquakes with magnitude upper than 6.5 and one earthquake with magnitude upper than 7 has been a very high seismicity region in the recent 63 years. Therefore, subject of earthquake prediction has great importance for this region.

Table 8 The number of magnitudes recorded at investigated decade in considered region

Year	Mw							Total
	<4.5	4.5–4.9	5–5.4	5.4–5.9	6–6.4	6.5–6.9	7–7.5	
1950–1960	0	1	4	1	1	0	0	7
1961–1970	3	26	8	5	0	0	0	42
1971–1980	37	103	26	8	3	1	0	178
1981–1990	62	114	14	4	3	1	1	199
1991–2000	143	48	15	4	0	2	0	212
2001–2010	963	96	17	2	1	0	0	1079
2011–2013	199	35	6	1	0	0	0	241
Total	1407	423	90	25	8	4	1	1958

3 Determining and evaluating seismicity indicators

Collection of data is required as inputs and output to create a model. Based on the entered information, the system is trained to provide good results. Large amounts of the entered data into the model will increase the accuracy in the learning process. For creating ANFIS, a region with the longitude 56–58 and latitude 27–31 in Iran has been selected, and information of the occurred earthquakes between 1950 and 2013 in the region has been collected. This dataset was used to train and test the model.

There are 38 records out of 1958 that have magnitude of 5.5 and greater. Each of these records is considered as a base and will be named basis earthquake. The main idea in this paper is that magnitude of next basis earthquake can be predicted by earthquakes that occurred after basis earthquake in the past. For this purpose, three input variables which are calculated for each record were used. The first of them is the logarithm of the seismic moment of earthquake (LM0). To determine the seismic moment the Eq. (11) which was calculated by Hanks was used (Hanks and Kanamori 1979):

$$\text{Log } M0 = 1.5 \text{ Mw} + 16.1 \tag{11}$$

To determine the two input remained variables, the region was considered in a grid network so that the distance between each of the lines was 0.1. The point at 27 latitude and 56 longitude is considered as the origin of the grid network with coordinates 0 and 0. In this network, e.g., point with 4 width and 0 length is corresponding to latitude 56 and 31 latitude. By rounding amounts for longitude and latitude of each event and considering it in one decimal, each record will be defined on lines of the grid network. Using the network causes the simulation of earthquakes location for the model to be much simpler. The repetition rates of the points are higher in mesh mode; this issue will increase the accuracy of the model to find the best result. Two seismic indicators along with other input LM0 are the length XO and width YO positions of each record in the grid network. Figure 4 shows position of all 1958 data in the network.

The main goal of ANFIS model is the amount prediction of moment magnitude for the future earthquakes which would occur with magnitude 5.5 or higher (basis earthquake). For this purpose, the difference between magnitude of each earthquake record and the basis future earthquake was determined. This value was calculated for all records and used as output DMw and was applied to train and to test the model.
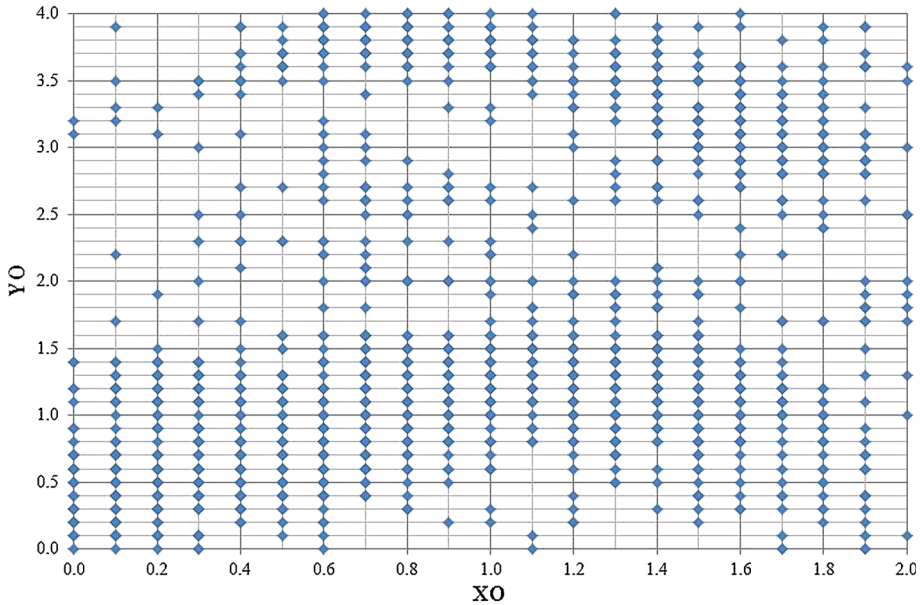
Fig. 4 The grid network for the specified region in Iran for all 1958 data that occurred between 1950 and 2013. In this network, point with coordinates 0 and 0 is the location of point with 27 latitude and 56 longitude

The number of last basis earthquake in the region is 1768. Therefore, calculating DMw values for records that have occurred after this record may not be possible. It is also not feasible to determine DMw for the record number 1768 which is a basis earthquake. Finally, there were 1767 available records, and for each of them four indicators were calculated. Table 9 shows the calculated values for some records. The data in this table related to earthquakes that occurred between two basis earthquake with magnitudes of 5.5 and 5.8.

4 ANFIS

One of the first hybrid neuro-fuzzy systems for function approximation was Jang's model (Jang 1993). Adaptive neuro-fuzzy inference system (ANFIS) is a fuzzy inference system implemented in the framework of adaptive networks. The model proposed can construct an input–output mapping based on both human knowledge in the form of fuzzy rules and stipulated input–output data pairs. It presented a Sugeno-type fuzzy system in five-layer network (the input layer not counted by Jang).

Figure 5 shows ANFIS with two inputs x and y and one output z. Suppose that the rule base contains two fuzzy if–then rules of Takagi and Sugeno's type:

$$\text{Rule 1}: \text{ If } x \text{ is } A_1 \text{ and } y \text{ is } B_1, \text{ then } \quad f_1 = p_1 x + q_1 y + r_1$$

$$\text{Rule 2}: \text{ If } x \text{ is } A_2 \text{ and } y \text{ is } B_2, \text{ then } \quad f_2 = p_2 x + q_2 y + r_2$$

Then, the corresponding equivalent ANFIS architecture is shown in Fig. 5.

Table 9 The number of magnitudes recorded at investigated decade in considered region

| Information of records					Seismicity indicators			
Date	Time	Latitude	Longitude	Mw	LM0	XO	YO	DMw
9/29/1962	6:54:00	28.29	57.48	5.5	24.28625	1.5	1.3	0.3
10/16/1962	11:58:41	30.63	57.36	4.6	22.95695	1.4	3.6	1.2
10/16/1963	7:02:26	28.15	57.79	4.9	23.44100	1.8	1.2	0.9
3/11/1964	11:34:23	27.90	57.61	4.9	23.43425	1.6	0.9	0.9
5/11/1964	6:07:41	28.13	57.38	5.2	23.90225	1.4	1.1	0.6
7/21/1964	11:46:54	27.60	56.46	4.9	23.43425	0.5	0.6	0.9
8/10/1964	6:18:40	30.10	57.67	4.9	23.43425	1.7	3.1	0.9
12/19/1964	11:31:55	27.50	56.88	5.2	23.90225	0.9	0.5	0.6
12/22/1964	4:36:34	28.16	56.90	5.8	24.78425	0.9	1.2	−0.3
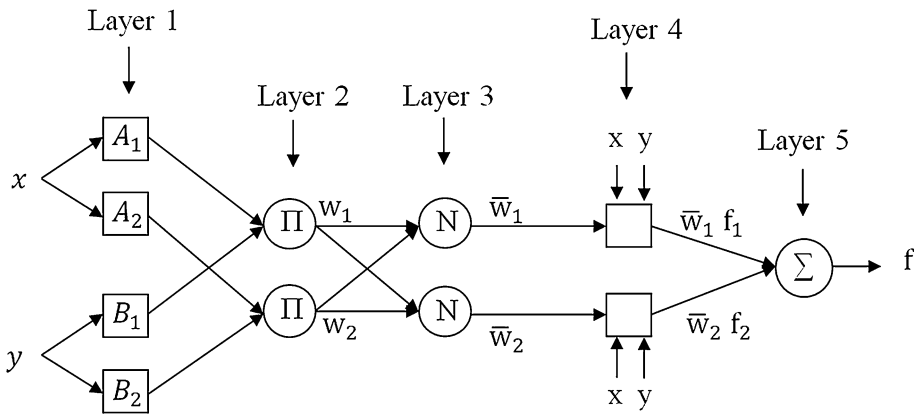
Fig. 5 Structure of ANFIS with two inputs and two rules. A *square node* (adaptive node) has parameters while a *circle node* (fixed node) has none (Jang 1993)

The node functions in the same layer are of the same function family as described below (Jang 1993):

Layer 1:   Every node i in this layer is a square node with a node function

$$O_i^1 = \mu_{A_i}(x) \tag{12}$$

where x is the input to node i and A_i is the linguistic label (such as "small" or "large") associated with this node function. In other words, O_i^1 is the membership function of A_i and it specifies the degree to which the given x satisfies the quantifier A_i. Any continuous and piecewise differentiable function, such as commonly used bell-shaped, trapezoidal or tri-angular-shaped membership functions (MF) are qualified candidates for node function in this layer.

Layer 2:   Every node in this layer is a circle node labeled Π which multiples the incoming signals and sends the product out. For instance,

$$w_i = \mu_{A_i}(x) \times \mu_{B_i}(y), \quad i = 1, 2. \tag{13}$$

Each node output represents the T-norm operators that combine the possible input membership grades in order to compute the firing strength of a rule.

Layer 3: Every node in this layer is a circle node labeled N. The ith node calculates the ratio of the ith rule's firing strength to the sum of all rules' firing strengths:

$$\overline{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1, 2. \tag{14}$$

For convenience, outputs of this layer will be called normalized firing strengths.

Layer 4: Every node i in this layer is a square node with a node function

$$O_i^4 = \overline{w}_i f_i = \overline{w}_i(p_i x + q_i y + r_i) \tag{15}$$

where $\overline{w_i}$ is the output of layer 3, and $\{p_i, q_i, r_i\}$ is the parameter set. Parameters in this layer will be referred to as consequent parameters that are adjustable.

Layer 5: The single node in this layer is a circle node (adaptive node) labeled $\sum$ that computes the overall output as the summation of all incoming signals, i.e.,

$$O_i^5 = \text{overall output} = \sum_i \overline{w}_i f_i = \frac{\sum_i \overline{w}_i f_i}{\sum_i w_i} \tag{16}$$

It is not adjustable.

For learning of ANFIS, a combination of two methods of back-propagation (gradient descent) and least squares estimate (LSE) are used. First, parameters of the introduction section are supposed steady, and result parameters are estimated using least squares method. Then, result parameters are supposed steady and error back-propagation is used to correct the parameters of introduction. This process is repeated in each learning cycle (Bezdek and Pal 1992).

5 Methods

Two methods are commonly used to generate ANFIS: grid partition (GP) and subtractive clustering (SC). ANFIS with GP algorithm uses a hybrid learning algorithm to identify parameters of inference system. It applies a combination of the least squares method and the back-propagation gradient descent method for training ANFIS membership function parameters.

Grid partition divides the data space into rectangular sub-spaces using axis-paralleled partition based on pre-defined number of MF and their types in each dimension. The number of rules depends on the number of input variables and on the number of MF used per variable, and this partition strategy needs a small number of membership function for each input. It encounters problems when we have a moderately large number of inputs (Jang and Sun 1996).

Clustering is a task of assigning a set of data into groups called clusters to discover structures and patterns in a dataset, and the radius of a cluster is the maximum distance

between all the points and the centroid. There are two main clustering methods: the hard clustering and the fuzzy clustering. The hard clustering is based on classify each point of the dataset just to one cluster. In fuzzy clustering, objects on the boundaries between several clusters are not forced to fully belong to one of them. The subtractive clustering method (SC) as a hard clustering was proposed by Chiu (1994). The SC method assumes that each data point is a potential cluster center and calculates the potential for each data point based on the density of surrounding data points. The measure of potential for a data point is a function of its distances to all other data points. A data point with many neighboring data points will have a high potential value. The data point with highest potential is selected as the first cluster center, and the potential of data points near the first cluster center is destroyed. Then data points with the highest remaining potential as the next cluster center and the potential of data points near the new cluster center are destroyed. It is notable that the influential radius of cluster is critical for determining the number of clusters and data points outside this radius has little influence on the potential. Also, a smaller radius leads to many smaller clusters in the data space, which results in more rules (Chiu 1994).

In this paper, GP, SC, and another technique that is called FCM are used to create the ANFIS model. FCM is a powerful unsupervised algorithm. Fuzzy c-means (FCM) clustering was first reported by Dunn (1973). It can be extended by Bezdek (1981). FCM is an algorithm where each data point has a membership degree between 0 and 1 to each fuzzy subset. In other words, each data in FCM can be belonged to all groups with different membership grades. The algorithm produces an optimal c partition by minimizing the weighted within group sum of squared error function J_m (Dunn 1973):

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{c} u_{ji}^m d^2 (x_i, \ v_j) \tag{17}$$

where $X = \{x_1, x_2, \ldots, x_N\} \subseteq \mathrm{IR}^m$ is the dataset in the m-dimensional vector space, N is the number of data items, c is the number of clusters with $2 \leq c < N$, u_{ji} is the degree of membership of x_i in the jth cluster, m is the weighting exponent on each fuzzy membership, v_j is the prototype of the center of cluster j, $d^2 (x_i, \ v_j)$ is distance measure between object xi and cluster center v_j.

To create an ANFIS with FCM, data are clustered by FCM algorithm and then ANFIS method is applied on the clustered data.

6 Earthquake magnitude prediction

As mentioned before, three seismic indicators for each record including logarithm of seismic moment LM0, the location of each record in grid network XO and YO as inputs, and also the difference between magnitude of the next basis earthquake and each record DMw were used as an output to modeling.

Additionally, there were 1,767 available data for ANFIS while 1,500 cases were used for training, 267 remaining data were used for testing the model. It should be mentioned that the training data were the records between 1950 and January 24, 2009. The records that occurred after this date include 267 data that were used to test the model.

During data processing, the data should be normalized to bring all of the variables into proportion with one another. Normalization is a method to classify the range or interval of values that are different to same scale and similar. If scales are very dissimilar for the

different values, the bigger one will have a higher contribution to the output error and so the error reduction algorithm will be forced on the variables of higher values, neglecting the information from the small values variables (Sola and Sevilla 1997). This action ensures that no exceptionally large-valued descriptors will have an undue effect on the network and also increases the speed and accuracy of the system in training. A simple normalization relationship within the value of 0.1–0.9 which is used to normalization in this paper is the following equation:

$$x_i = 0.8 \frac{x - x_{\min}}{x_{\max} - x_{\min}} + 0.1 \qquad (18)$$

where x_i is the normalized value of a certain parameter, x is the measured value for this parameter, $x_{\min}$ and $x_{\max}$ are the minimum and maximum values in the database for this parameter, respectively.

To find the best results based on GP, SC, and FCM methods, datasets were used randomly and many models were created. It was found that the FCM model was much faster than the other two methods and the algorithm of GP needed more time. Also, the rules to achieve the desired results that FCM takes are lower than those of both the GP and SC. Table 10 shows the results of the experiment of the model for the GP method. MF parameter in the table is Membership Function. Parameters R^2, MAE, and RMSE have been calculated based on denormal data.

It is clear from Table 10 that with increasing the number of MF, the rules and unknown parameters of the problem were increased. In ANFIS, modeling with GP algorithm was also found that the best results are obtained by Gaussian membership function with three MF (Model 1 in Table 10).

In the SC method, radius of the cluster should be defined before modeling. The smaller radius will create the greater number of unknown parameters. In Table 11, the best results obtained by the SC algorithm for the test phase are presented. According to this table, the best model was number 6. In a general comparison between the model number 1 of GP and model 6 of SC, it was found that the GP algorithm had less error but needed more rules to solve the problem.

To create ANFIS with FCM algorithm, the number of clusters was predefined for the model. Therefore, to find the proper state, many models with different number of clusters

Table 10 The results of the three best models that obtained from ANFIS by GP algorithm

Number of model	MF	Number of MF	Number of rules	R^2	MAE	RMSE
1	Gaussian	3	27	0.94	0.149	0.173
2	Gaussian	4	64	0.44	0.202	0.467
3	Gaussian	5	125	0.23	0.262	0.602

Table 11 The results of the three best models that obtained from ANFIS by SC algorithm

Number of model	Radius of cluster	Number of rules	R^2	MAE	RMSE
4	0.20	16	0.90	0.159	0.194
5	0.25	13	0.90	0.163	0.198
6	0.30	10	0.95	0.163	0.181

Table 12 The results of the three best models that obtained from ANFIS by FCM algorithm

Number of model	Number of cluster	Number of rules	R^2	MAE	RMSE
7	10	10	0.94	0.158	0.180
8	12	12	0.95	0.150	0.172
9	15	15	0.95	0.150	0.173

Table 13 The results of some test data. Here, the goal is determine the moment magnitude with amount 5.6 for the basis earthquake

Date	Latitude	Longitude	Mw	LM0	XO	YO	LMw	Net output	Mw for the next base earthquake
6/24/2010	28.43	57.08	3.8	21.79199	1.1	1.4	1.80534	2.05438	5.8
6/30/2010	27.83	56.70	3.9	21.99557	0.7	0.8	1.66962	1.73119	5.6
7/8/2010	28.02	57.01	3.9	21.89378	1.0	1.0	1.73748	1.77828	5.7
7/11/2010	28.47	57.48	4.1	22.19915	1.5	1.5	1.53390	1.66505	5.8
7/19/2010	28.27	57.95	3.6	21.48662	2.0	1.3	2.00892	2.33016	5.9
7/24/2010	28.01	56.68	4.3	22.60631	0.7	1.0	1.26246	1.35518	5.7
7/26/2010	27.89	56.36	3.7	21.69020	0.4	0.9	1.87320	1.98801	5.7

The mean of net output in this table is the obtained output values from the neuro-fuzzy model based on denormal data

were created. The best model in the test results are shown in Table 12. As it is clear from the table, FCM compared with SC and GP has less number of rules and has more speed and also presents better results.

According to the above statements, among the models presented in Tables 10, 11, and 12, the models 7–9 have the best results and the model 8 is the most powerful system. Therefore, for prediction of basis earthquakes in the future, ANFIS with FCM algorithm with 12 clusters will be used.

The output model was the difference between the earthquake magnitude of each event and the magnitude of the basis earthquake. By adding amount of the output neuro-fuzzy model of ANFIS and magnitude of considered earthquake, the magnitude of the future earthquake of 5.5 or greater will be predicted. As it was mentioned earlier that for testing the model, records of 267 earthquakes occurring after 01/24/2009 were used. There was no basis earthquake in these 267 records. It means that earthquakes with magnitude 5.5 or greater have not occurred in the mentioned period. But the 268th record was a basis earthquake with moment magnitude 5.6 which is used to calculate DMw in the seismic indicators calculation step. Therefore, all 267 earthquakes must be predicting the earthquake with magnitude 5.6. In Table 13, information of some data with their indicators, output value of selected ANFIS, and the determined magnitudes that were calculated from the output of the model are presented. The table shows ANFIS-FCM had amount of targets with high accuracy.

The final results contain 21 cases of predicting exactly 5.6 magnitude. The minimum and maximum values obtained from the prediction, respectively, are 5.6 and 6. Therefore, the model produced results with acceptable tolerances. The summary of predicted magnitude of earthquakes has been illustrated in Table 14.

Table 14 The obtained results to predict moment magnitude with amount of 5.6 based on ANFIS-FCM

Mw	Numbers
5.6	21
5.7	132
5.8	74
5.9	35
6	5
Total	267

7 Results

In this paper, moment magnitude of the earthquakes of 5.5 or greater in the region with 2 longitude and 4 latitude that will occur in the future has been predicted by ANFIS. To do this, the three inputs and one output were used. The system was trained by 1,500 data from earthquakes that occurred in the region. Then, dataset consisting of 267 records, which is used to predict a basis earthquake with magnitude 5.6 that will occur in the future, was used. Many ANFIS models based on GP, SC, and FCM were developed, and it was found that the ANFIS-FCM predicts the earthquake magnitude with higher accuracy than other models.

Iran is one of the most earthquake prone countries in the world, and prediction of future earthquakes and earthquake risk analysis is important in this country. In order to predict the earthquake, many steps have been taken; however, more is required to be taken. This study was an attempt to provide a model based on one of the most fuzzy algorithms to predict magnitude of future large earthquakes. The ANFIS-FCM based on information that has been presented in this paper is capable of obtaining the best results with high precision.

References

Adeli H, Panakkat A (2009) A probabilistic neural network for earthquake magnitude prediction. Neural Netw 22:1018–1024

Alves EI (2006) Earthquake forecasting using neural networks: results and future work. Nonlinear Dyn 44:341–349

Andalib A, Zare M, Atry F (2009) A fuzzy expert system for earthquake prediction, case study: the Zagros range. In: 3rd International Conference on Modeling, Simulation, and Applied Optimization (IC-MSAO), American University of Sharjah, Sharjah, UAE, 2009

Bezdek JC (1981) Pattern recognition with fuzzy objective function algorithms. Kluwer Academic Publishers, New York

Bezdek JC, Pal SK (1992) Fuzzy models for pattern recognition, vol 267. IEEE Press, New York

Chattopadhyay G, Chattopadhyay S (2009) Dealing with the complexity of earthquake using neurocomputing techniques and estimating its magnitudes with some low correlated predictors. Arab J Geosci 2:247–255

Chiu SL (1994) Fuzzy model identification based on cluster estimation. J Intell Fuzzy Syst 2:267–278

Cong P, Bo J, Liu H, Wu W (2006) Application of artificial neural networks FAM model to prediction of seismic trend after main shock. South China J Seismol

Dunn JC (1973) A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. J Cybern 3:32–57

Hanks TC, Kanamori H (1979) A moment magnitude scale. J Geophys Res 84:2347–2350

Jang J-SR (1993) ANFIS: adaptive network based fuzzy inference systems. IEEE Trans Syst Man Cybern 23:665–685

Jang J-SR, Sun C-T (1996) Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence. Prentice-Hall Inc, Upper Saddle River

Liu Y, Liu H, Zhang B, Wu G (2004) Extraction of if-then rules from trained neural network and its application to earthquake prediction. In: Proceedings of the 3rd IEEE International Conference on Cognitive Informatics Columbia, Canada, 2004, pp 109–115

Liu Y, Li Y, Li G, Zhang B, Wu G (2005) Constructive ensemble of RBF neural networks and its application to earthquake prediction. In: Advances in neural networks—ISNN 2005. Springer, Berlin, Heidelberg, pp 532–537

Mohsin S, Azam F (2011) Computational seismic algorithmic comparison for earthquake prediction. Int J Geol 5:53–59

Panakkat A, Adeli H (2007) Neural network models for earthquake magnitude prediction using multiple seismicity indicators. Int J Neural Syst 17:13–33

Panakkat A, Adeli H (2009) Recurrent neural network for approximate earthquake time and location prediction using multiple seismicity indicators. Comput Aided Civ Infrastruct Eng 24:280–292

Shah H, Ghazali R, Nawi NM (2011) Using artificial bee colony algorithm for mlp training on earthquake time series data prediction. J Comput 3:135–142

Sharma M, Tyagi A (2010) Cyclic behavior of seismogenic sources in India and use of ANN for its prediction. Nat Hazards 55:389–404

Sheng-Zhong H (2010) The prediction of the earthquake based on neural networks. In: International Conference on Computer Design and Applications (ICCDA) Qinhuangdao, Hebei, China, 2010, pp 517–520

Sola J, Sevilla J (1997) Importance of input data normalization for the application of neural networks to complex industrial problems. Nuclear Sci IEEE Trans 44:1464–1468

Sri Lakshmi S, Tiwari R (2009) Model dissection from earthquake time series: a comparative analysis using modern non-linear forecasting and artificial neural network approaches. Comput Geosci 35:191–204

Wang W, Wu G-F, Huang B-S, Zhuang K-Y, Zhou P-L, Jiang C-X et al (1997) The FAM (fuzzy associative memory) neural network model and its application in earthquake prediction. Acta Seismol Sin 10:321–328

Wang W, Wu G-F, Song X-Y (2000) The application of neural networks to comprehensive prediction by seismology prediction method. Acta Seismol Sin 13:210–215

Xie J, Qiu J-F, Li W, Wang J-W (2011) The application of neural network model in earthquake prediction in East China. In: Advances in computer science, intelligent system and environment. Springer, Berlin, Heidelberg, 2011, pp 79–84

Yaghmaei-Sabegh S (2012) A new method for ranking and weighting of earthquake ground-motion prediction models. Soil Dyn Earthq Eng 39:78–87

Yi C, Jinkui Z, Jiaxin H (2010) Research on application of earthquake prediction based on chaos theory. In: International Conference on Intelligent Computing and Integrated Systems (ICISS), 2010, pp 753–756

Ying W, Yi C, Jinkui Z (2009) The application of RBF neural network in earthquake prediction. In: 3rd International Genetic and Evolutionary Computing (WGEC'09), 2009, pp 465–468

Yudong C, Jiawen G, Junren G, Linsheng Y (1994) Artificial neural network method for prediction of earthquake sequence type. J Seismol Res 1, pp 40–45

Zhong M, Zhang Q (2010) Prediction of reservoir-induced earthquake based on fuzzy theory. In: Proceedings of the Second International Symposium on Networking and Network Security (ISNNS'10), Jinggangshan, P. R. China, 2010, pp 101–104