PREDICTION OF THE EARTHQUAKE MOMENT MAGNITUDE BY USE OF THE MULTILAYER PERCEPTRON NEURAL NETWORK

Jamal MAHMOUDI
MSc. Student, K.N.Toosi University of Technology, Tehran, Iran
jmahmoudi@mail.kntu.ac.ir

Masoud REZAEI
MSc. Student, Building and Housing Research Center, Tehran, Iran
m.rezaei@bhrc.ac.ir

Mohammad Hossein MOHAMMADI
MSc. Student, Kharazmi University, Tehran, Iran
std_h.mohammadi@khu.ac.ir

Keywords: Multilayer Perceptron, Neural Network, Earthquake Moment Magnitude, Prediction, Training Method

ABSTRACT

Because of the major disadvantages of previous methods for calculating the magnitude of the earthquakes, the neural network as a new method is examined. In this paper a kind of neural network named Multilayer Perceptron (MLP) is used to predict moment magnitude of earthquakes. MLP neural network consist of three main layers; input layer, hidden layer and output layer. Since the best network configurations such as the best number of hidden nodes and the most appropriate training method cannot be determined in advance, and also, overtraining is possible, 32 models of network are evaluated to determine the best prediction model. By comparing the results of the current method with the real data, it can be concluded that MLP neural network has high ability in predicting the moment magnitude of earthquakes and itِ s a very good choice for this purpose.

INTRODUCTION

Since ancient times, in the wake of natural events and disasters, man has always been looking for ways to prevent or control these events. The earthquake is one of these natural disasters which cause heavy losses of life and property, when it occurs. Time, location and magnitude of the earthquake are three parameters that must be a good estimate of their amounts in order to control and minimize its losses. Hence, scientists and researchers have done attempts, including many successful and unsuccessful ones, to find a relationship between these three parameters, or make a good estimation of them.

These efforts have resulted in developing a number of theoretical and empirical equations. However, applicability of equations developed for calculating the magnitude of earthquakes is affected by a lot of parameters. Most of these parameters need to be measured and entered in the equations accurately, while, in many areas, due to the lack of required equipment, these parameters mostly are measured approximately and with low precision or even sometimes assumed. Also, some parameters of the equations such as physical and functional characteristics of faults are difficult to measure. For example, geodetic strain rate of reverse faults with no apparent sign of fault strike on the earth surface is not measurable. Moreover, these equations usually are exclusive of a specific region or state, so they are not reliable enough for other new regions.

On the other hand, neural networks have been proven to be one of the most practical effects in modelling and forecasting (Lippmann, 1987). There are three major advantages of neural networks. First, neural networks are able to learn any complex non-linear mapping. Second, they do not make a priori assumption about the distribution of data. Third, they are very flexible with respect to incomplete, missing and noise data and therefore eliminate concern about this issue (Vellido et al, 1999). Moreover, neural networks, regardless of the region and country, are a general solution in all areas.

Neural networks have been used successfully to solve complicated pattern recognition and classification problems in different domains such as image and object recognition (Adeli and Hung 1993, Bourbaki et al, 2007), speech recognition (Yau W C et al, 2007), robotics and computer vision (Jorgensen Haynes and Norlund, 2008), natural language and text processing (Ruiz Pinales et al, 2008), signal processing (Cichocki and Zdunek, 2007), biomedical engineering and medical diagnosis (Adeli et al, 2008), neuroscience (Bolle and Heylen, 2007, Chakravarthy et al, 2007), optimization and nonlinear programming (Chen and Young, 2007, Mayorga and Arriaga, 2007), construction engineering (Senouci and Adeli, 2001), transportation engineering (Cyganek, 2008, Pande and Abdel Aty, 2008, Stathopoulos Dimitriou and Tsekeris, 2008), video and audio analysis (Fyfe et al, 2008), computer networking (Kimura and Ikeguchi, 2007), control (Liu and Zhang, 2008, Rigatos, 2008) and financial forecasting (Schneider and Graupe, 2008).

Because of its satisfactory history in the prediction of a variety of parameters in different fields, which is briefly mentioned in previous sentences, it was anticipated that this method would help to develop a prediction model to identify the level of moment magnitude of earthquakes.

In this regard, a type of neural network system named Multilayer Perceptron (MLP), which is one of the most influential neural network models, is used to predict the magnitude of the earthquakes, and the method and results are presented in the following sections. In Section 2, Multilayer perceptron neural network is described. In Section 3, the variables which are included in the modelling are specified. MLP network prediction results are presented and discussed in Section 4.

MULTILAYER PERCEPTRON NEURAL NETWORK MODELLING

The model neurons, connected up in a simple fashion, were given the name ِp erceptronsپby Frank Rosenblatt in 1962. He pioneered the simulation of neural networks on digital computers, as well as their formal analysis.

In a large number of complicated math problems, which their solution depends on the solving tough non-linear equations, MLP neural networks are very helpful, and can easily be employed by defining proper weights and functions.

MLP neural networks consist of several layers of nodes. It includes an input layer, an output layer, and a hidden layer, each of which contains input node(s) which are called sensory, output node(s) which are called responding nodes, and hidden node(s), respectively.

The multilayer perceptron neural networks can be built and used with arbitrary number of layers. However, it can be proven that a three layer perceptron is capable of modelling any problem adequately. This fact is referred to as Kolmogrov theorem and is a fundamental concept in neural networks modelling.

The neural network which is used in this research consists of one hidden layer. The hidden layer contains unobservable network node(s). Each hidden node is a function of the weighted sum of the inputs. The function is the activation function, and the values of the weights are determined by the estimation algorithm. The tangent hyperbolic activation function is used for the hidden layer. This activation function takes real-valued arguments and transfers them to the range (-1, 1):

$$\gamma(c) = \tanh(c) = \frac{e^{c} - e^{-c}}{e^{c} + e^{-c}} \tag{1}$$

For the output layer, softmax activation function is used. It takes a vector of real-valued arguments and transforms it to another vector whose elements fall in the range (0, 1) and their sum equals to 1.

$$\gamma(c_k) = \exp(c_k) / \sum_j \exp(c_k) \quad (2)$$

Training Methods: The training method specified how the network processes the records. There are three common methods of training which can be described as following:
- Batch training: This type of training updates the synaptic weights only after passing all training data records. This means that batch training uses information from all records in the training dataset. This method is often preferred because it directly minimizes the total error. On the other hand, batch training may need to update the weights many times until one of the stopping rules is met and consequently may need many data passes. It is most useful for ِs maller پ datasets.
- Online training: This type of training updates the synaptic weights after every single training data record. This means that online training uses information from one record at a time. This method continuously gets a record and updates the weights until one of the stopping rules is met. If all the records are used once and none of the stopping rules is met, then the process continues by recycling the data records. Online training is superior to batch training for ِ larger پ datasets. This means that if there are many records and many inputs, and their values are not independent of each other, online training can more quickly obtain a reasonable answer than batch training.
- Mini-batch training: This type of training divides the training data records into groups of approximately equal size, and then updates the synaptic weights after passing one group. This means that mini-batch training uses information from a group of records. Then the process recycles the data group if necessary. Mini-batch training offers a compromise between batch and online training, and it may be best for ِm edium-size پ datasets.

Optimization Algorithm: This is the procedure used to estimate the synaptic weights and in the current paper following two algorithms are used:
- Scaled conjugate gradient: The assumptions that justify the use of conjugate gradient methods apply only to batch training types, so this method is not available for online or mini-batch training.
- Gradient descent: For the current modelling, gradient descent algorithm is used for online training method.

Training of MLP neural network usually is done using back propagation method. A sample schematic of a multilayer perceptron network is depicted in figure 1.

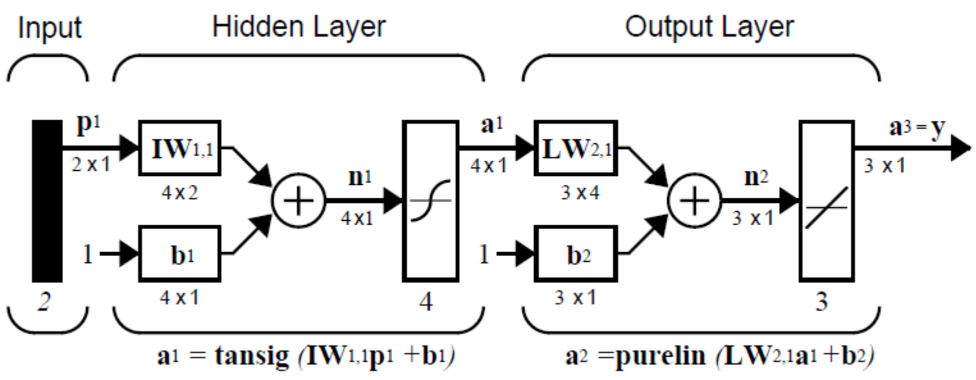

Figure1. Multilayer perceptron structure with hidden neurons and output neuron with linear function

The input nodes are based on some variables. In the current research, six independent variables including three spatial variables (latitude, longitude, depth), one time variable (days), and two variables related to physical characteristics (soil type, fault mechanism) are defined. The output nodes of neural networks are the prediction outputs or labels. In MLP systems it is essential to categorise dependent variable(s) into some branches. So, magnitude of earthquakes is categorized in four groups. These groups are indicated by A, B, C, and D, which represent 4-5, 5-6, 6-7 and bigger than 7 Richter, respectively.

In the hidden layer, as there is no method to decide the optimal number of hidden nodes directly, four different numbers of hidden nodes, including 8, 12, 16 and 20 are chosen. Moreover, a well-known concern with neural networks is ιιovertraining پ. To ease this problem, Roiger and Geatz (2003) suggest that the experiments could be continually conducted by different parameters. Therefore, we use a set of four different learning epochs, including 1, 2, 4 and 8. Furthermore, in training part, batch and online training methods are applicable. In order to reach to more comprehensive results, both of these methods are applied. As a result, we setup 32 different groups of parameters and shape 32 models accordingly shown in Table 1.

Table 1. Parameters settings of MLP neural networks

32 Models		Number Of Units			
Training Method	Epochs	8	12	16	20
Batch	1	N081	N121	N161	N201
	2	N082	N122	N162	N202
	4	N084	N124	N164	N204
	8	N088	N128	N168	N208
Online	1	N081	N121	N161	N201
	2	N082	N122	N162	N202
	4	N084	N124	N164	N204
	8	N088	N128	N168	N208

DEFINING THE VARIABLES

The seismic data that have been used in the current research are got from the whole instrumentally recorded earthquakes occurred in Iran from *International Institute of Earthquake Engineering and Seismology* (*IIEES*) ground motion data base. After revising and declustering the data of catalogue and omitting aftershocks and foreshocks, 11000 earthquake events were remained for consideration. In the field of structure engineering, earthquake magnitude upper than 4 Richter are more important, so the events lower than 4 are eliminated from catalogue. For better training of neural network, areas and faults with lower than 3 events were omitted, and finally 4945 event are used in research.

For each event 7 different parameters are defined. 6 of these parameters are independent variables which comprise input variables. The other one (7^{th} parameter) is the dependent variable which is the output of network. Input parameters consist of 3 spatial variables which are related to spatial features of earthquakes, one time variable, and two variables related to faults characteristic (including soil type and fault mechanism).

- **Spatial variables**: Longitude, latitude and depth of earthquake are three parameters that are allocated to each event. For earthquakes with an unknown depth, their depths are assumed to be 33 km.

- **Time variable**: This variable for each event is the time (days) between that event and the previous one in a specific fault. This variable also indicates stored strain energy in a fault. It should be mentioned that the periods of time used in the calculations are extracted from data base before eliminating the events whose magnitudes are less than 4 Richter. This is because occurrence of these small magnitude events affects the stored energy of the fault. Since for the first recorded event of any region, this variable cannot be calculated, because there is not a previous record, the variable is assumed. This assumption is made by considering the magnitude of the first event and looking for the same magnitudes in the later times. So the average time variable of the later events with the same magnitude of the first event is assumed as its time variable.

- **Faults characteristics**:
- Fault Mechanism: Three major classifications of faults (normal, reverse, strike-slip) and their combination can be applied to categorize faults more precisely into eight groups: Normal, reverse, strike-slip left lateral, strike-slip right lateral, normal-strike-slip left lateral, normal-strike-slip right lateral, reverse-strike-slip left lateral, and reverse-strike-slip right lateral.

- Soil Type: Soil type of occurrence zone of each earthquake is entered based on Iranian Code of Practice for Seismic Resistant Design of Buildings (Standard No. 2800) in four groups.

Output variable is the moment magnitude of earthquakes which has happened, and will be calculated by the MLP method. The results for outputs are divided into four qualitative groups, A, B, C and D. This is because the MLP neural networks are more compatible with qualitative variables and result in more accurate outputs. Magnitude classification is depicted in table 2.

Table 2. Magnitude classification

A	B	C	D
4-5	5-6	6-7	Bigger than 7

From whole data, 70% of them are used for network training, 20% for network testing and revising and the remained 10% are dedicated to derive the final prediction of the magnitudes of earthquakes. Then, these predictions have been compared with real values to assess the network prediction ability.

RESULTS AND DISCUSSION

As it can be seen in Figure 2, the average correct prediction of the models is about 70%. The results clarify that both batch and online training methods have good ability of prediction but MLP neural network with online training method has higher prediction power and maybe this is because a large data base is considered. Also, according to the results, models with16 nodes in the hidden layer have the highest percentage of correct predictions, so it can be concluded that the optimal number of nodes in the hidden layer is 16 nodes. Moreover, about the training period, results show that all models have their best performance in their first epoch. Overall, the best prediction is the prediction made by the model N161 using online training method.

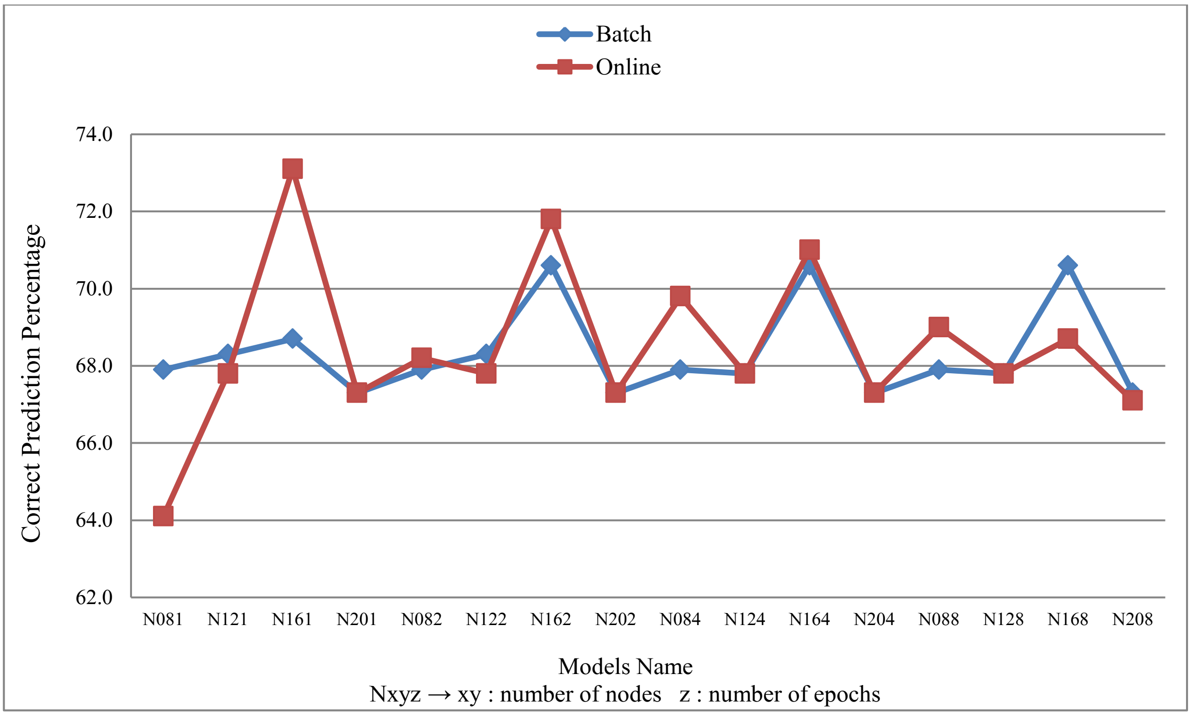

Figure 2. Models correct prediction percentage

Another output of MLP neural network is independent variables importance, which shows the impact of each independent variable on the earthquakes moment magnitude. Independent variables importance of model N161 using online training method is shown in table 4. As it can be seen in this table, and also same table of all the other models, the time variable (days) has highest impact on the moment magnitude of earthquakes.

Table 3. Independent Variable Importance

	Importance	Normalized Importance
Latitude	0.2	94.40%
Longitude	0.197	93.20%
Depth	0.185	87.50%
Days	0.211	100.00%
Soil Type	0.064	30.30%
Fault Type	0.143	67.70%

CONCLUSION

To conclude, according to the results, the MLP neural network is a functional device in predicting the magnitude of the earthquake of a region in an arbitrarily considered time.

The average correct prediction of the models was about 70%. The results clarified that both batch and online training methods have good ability of prediction, however, in the case considered in the current paper, MLP neural network with online training method showed a little higher prediction power. Also, according to the results, it can be concluded that, in the considered case, the optimal number of nodes in the hidden layer is 16 nodes. About the training period, results showed that all models have their best performance in their first epoch. Variables importance output of MLP neural network clarified that the time variable has highest impact on the moment magnitude of earthquakes. Overall, the best prediction was the prediction made by the model N161 using online training method. Since the proposed method is a comprehensive method and needs no priori assumption, similar modelling approach can be applied for other case studies in earthquakes magnitude prediction and it is anticipated that the results for the other cases will be as good as the results of the current study.

REFERENCES

Adeli H, Ghosh-Dastidar S anr Dadmehr N (2008) A spatio-temporal wavelet-chaos methodology for EEG-based diagnosis of Alzheimer's disease, *Neuro- science Letters*, 444(2): 190-194

Adeli H and Hung SL (1994) An adaptive conjugate gradient learning algorithm for effective training of multilayer neural networks, *Applied Mathematics and Computation*, 62(1): 81-102

Beale R and Jackson T (1991) Neural Computing: An Introductionپ,*IOP Publishing Ltd*

Bolle D and Heylen R (2007) Adaptive thresholds for neural networks with synaptic noise, *International Journal of Neural Systems*, 17(4): 241-252

Bourbakis N, Kakumanu P, Makrogiannis S, Bryll R and Panchanathan S (2007) Neural network approach for image chromatic adaptation for skin colour detection, *International Journal of Neural Systems*, 17(1): 1-12

Cichocki A and Zdunek R (2007) Multilayer nonnegative matrix factorization using projected gradient approaches, *International Journal of Neural Systems*, 17(6): 431-446

Chen YY and Young KY (2007) an SOM-based algorithm for optimization with dynamic eight updating, *International Journal of Neural Systems*, 17(3): 171_181

Chih Fong T and Yen Jiun Ch (2009) Earnings Management Prediction: A Pilot Study of Combining Neural Networks and Decision Trees, *Journal of Expert Systems with Applications*, 36: 7183ت7191

Cyganek B (2008) Color image segmentation with support vector machines: Applications to road signs detection, *International Journal of Neural Systems*, 18(4): 339_345

Fyfe C, Barbakh W, Ooi WC and Ko H (2008) Topological mappings of video and audio data, *International Journal of Neural Systems*, 18(6): 481-489

International Institute of Earthquake Engineering and Seismology (IIEES) ground motion data base

Iranian Code of Practice for Seismic Resistant Design of Buildings (Standard No. 2800) 3rd edition

Kimura T and Ikeguchi T (2007) an optimum strategy for dynamic and stochastic packet routing problems by chaotic neurodynamics, *Integrated Computer-Aided Engineering*, 14(4): 307-322

Lippmann RP (1987) an Introduction to Computing with Neural Nets, *IEEE ASSP Magazine April*: 4ت22

Liu M and Zhang S (2008) An LMI approach to design H1 controllers for discrete time nonlinear systems based on unified models, *International Journal of Neural Systems*, 18(5): 443-452

Rodrکguez-Sعnchez AJ, Simine E and Tsotsos JK (2007) Attention and visual search, *International Journal of Neural Systems*, 17(4): 275-288

Roiger RJ and Geatz MW (2003) Data Mining: A Tutorial-Based Primer, *Boston MA*: *Addison Wesley*

Ruiz-Pinales J, Jaime-Rivas R, Lecolinet E and Castro-Bleda MJ (2008) Cursive word recognition based on interactive activation and early visual processing models, *International Journal of Neural Systems*, 18(5): 419-431

Schneider NC and Graupe D (2008) a modified LAMSTAR neural network and its applications, International Journal of Neural Systems, 18(4): 331-337

Senouci AB & Adeli H (2001) Resource scheduling using neural dynamics model of Adeli and Park, *Journal of Construction Engineering and Management, ASCE*, 127(1): 8-34

SPSS Neural Networks 17.0 manual

Vellido A, Lisboa PJG & Vaughan J (1999) Neural Networks in Business: A Survey of Applications (1992ت1998), *Expert Systems with Applications*, 17: 51ت70

Yau WC, Kumar DK and Arjunan SP (2007) Visual recognition of speech consonants using facial movement features, *Integrated Computer-Aided Engineering*, 14(1): 49-61