International Journal of Engineering

J o u r n a l   H o m e p a g e :   w w w . i j e . i r

Seismic Data Forecasting: A Sequence Prediction or a Sequence Recognition Task

A. Bali* [a], M. Mahdinejad Noori [b]

[a] *International Institute of Earthquake Engineering and Seismology, Tehran-19537-14453, Iran*
[b] *Ministry of Science, Research and Technology, Tehran, Iran*

P A P E R   I N F O

A B S T R A C T

In this paper, the multivariate adaptive regression splines (MARS) is employed to predict earthquake events based on two common approaches in sequence learning. In the first scenario, the task is defined as a sequence prediction problem, and consequently the MARS model is used as a predictor. In the second scenario, the same task is considered as a sequence recognition problem and the model of MARS, this time, is used as a binary classifier with results that could alternatively help to predict an earthquake event. Forecasting results of applying the methods to a cluster of seismic data on pacific ring of fire indicate that MARS as a binary classifier outperforms the predictor MARS. In fact, while both approaches are plausible, the best results are achieved when the earthquake prediction problem is considered as a sequence recognition task.

1. INTRODUCTION

There are a wide variety of approaches for earthquake forecasting, which are under investigation. While some of these methods consider anomalous signatures regarding specific physical quantities as precursory phenomena [1], the others employ statistical approaches for the task [2], [3]. Earthquake prediction may also incorporate time series analysis [4], [5]. In this case, a mathematical model analyzes the sequence of data points of an earthquake catalogue or some other source of information in order to extract meaningful patterns. These patterns are then used to forecast future events based on the known past events. Earthquake prediction, from this point of view, may be classified into a sequence-learning problem.

In human, sequence-learning plays a major role in skill learning, and high-level problem solving and reasoning. It is also an important part to solve many real world problems that incorporate intelligence, including time series analysis [6], speech recognition [7], DNA sequencing [8], economics, and fraud detection, to name but a few applications. There are four approaches for different sequence-learning problems [9], known as

sequence prediction, sequence generation, sequence recognition, and sequential decision making. Sequence prediction attempts to predict future elements of a sequence using historical elements. While, sequence recognition tries to determine if a sequence is legitimate according to some criteria [9]. Obviously, the problem of earthquake prediction may be formulated as a sequence prediction and/or sequence recognition task. Therefore, these two aspects of sequence learning are of interest in this paper. It has been mentioned that sequence recognition may be reformulated as a sequence prediction problem [9]. However, it is explained here how these two aspects of sequence learning may be completely different. While one is aware of the differences between the two approaches, then it is worthy to test the problem of earthquake prediction from the two viewpoints.

In this paper, a single learning machine, i.e., the model of multivariate adaptive regression splines (MARS) [10] is employed to implement the above approaches. In this way, one is able to compare the results and just find if the problem of earthquake prediction is most likely to be sequence recognition or sequence prediction. The advantage of the employed model of MARS, then, lay in the fact that it could perfectly be used for both roles. When the task is

defined as a sequence prediction problem, the MARS model is used as a predictor and when the task is considered as a sequence recognition problem, the model of MARS play the role of a classifier for finding pre-events similarities [11].

This paper is organized as follows. Section 2 reviews sequence prediction/recognition as two different aspects of sequence learning. Section 3 contains a brief introduction of multivariate adaptive regression splines model and its dual use in time series prediction. In Section 4, the forecasting method is applied to the seismic data. Section 5 contains our discussion and conclusions.

2. THE BASIC DEFINITIONS: SEQUENCE PREDICTION *vs.* SEQUENCE RECOGNITION

For a deterministic state space, and for $1 \leq i \leq j < \infty$, sequence prediction may be formulated as $s_i, s_{i+1}, \cdots, s_j \rightarrow s_{j+1}$. That is, given a sequence of observations $s_i, s_{i+1}, \cdots, s_j$, the aim is to predict[2] s_{j+1}. On the other hand, a sequence recognition task may be formulated as $s_i, s_{i+1}, \cdots, s_j \rightarrow yes\, or\, no$. That is, one should recognize if the subsequence $s_i, s_{i+1}, \cdots, s_j$ is legitimate or not. It has been mentioned that sequence recognition may be reformulate as sequence prediction (Sun Giles 2001). That is, $s_i, s_{i+1}, \cdots, s_j \rightarrow yes$ (a recognition problem), if and only if, $s_i, s_{i+1}, \cdots, s_{j-1} \rightarrow s_j^p$ and $s_j^p = s_j^d$, where s_j^p is the predicted and s_j^d is the desired element.

$$\left(s_i, s_{i+1}, \cdots, s_j \rightarrow yes \right) \Leftrightarrow \left(s_i, s_{i+1}, \cdots, s_{j-1} \rightarrow s_j^p \right) \wedge \left(s_j^p = s_j^d \right) \qquad (1)$$

However, through a simple example here, it is shown that sequence prediction and sequence learning are not the same, in general. Consider optimal character recognition (OCR) problem [7], with digitized patterns, each consists of 15 black /white segments. If one assigns a 0 to any white segment and a 1 to any black segment, then any digit may be represented by a sequence of zeroes and ones, e.g. the corresponding pattern to the digit ONE can be represented by the sequence 0,1,0,0,1,0,0,1,0,0,1,0,01,0 (Figure 1). In this way, the OCR task may be defined as a sequence recognition problem, where any input observation is checked for legitimacy; e.g., the input 0,1,0,0,1,0,0,1,0,0,1,0,01,0 is legitimate to be recognized as ONE

$$0,1,0,0,1,0,0,1,0,0,1,0,01,0 \rightarrow yes \qquad (2)$$

---

[2] Sequence prediction may be extended to a stochastic state space, using a conditional probability distribution $p(s_{j+1} \mid s_i, \cdots, s_j)$.
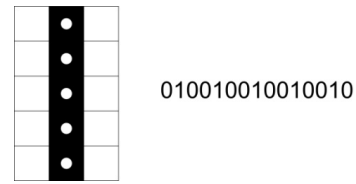

Figure 1. corresponding pattern to the digit ONE is represented by a sequence of zeroes and ones.
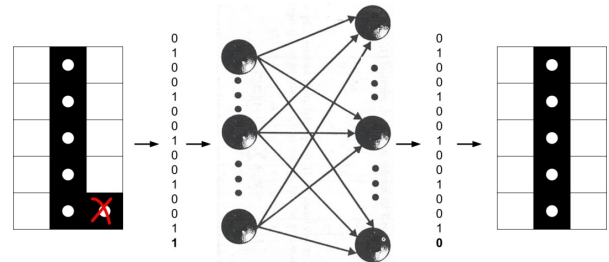

Figure 2. The (15:15) linear associative memory maps the noisy observation to a noise-free pattern for the digit ONE, which makes the noisy pattern legitimate.

However, for a real world problem, input observations may be corrupted by noise. Therefore, a successful tool should be able to recognize slightly different observations, which brings a non-strict legitimacy-check. It could be shown that a linear associative memory (LAM) with 15 neurons in each of the input and output layers, is tolerable to input noise up to a threshold. As an example the input observation 0,1,0,0,1,0,0,1,0,0,1,0,0,1,**1** is also legitimate, because this input sequence is mapped to an output sequence 0,1,0,0,1,0,0,1,0,0,1,0,0,1,**0** which is apparently legitimate (Figure 2). Therefore

$$0,1,0,0,1,0,0,1,0,0,1,0,01,1 \rightarrow yes \qquad (3)$$

According to Equation (1), Equation (2) may be reformulated as

$$0,1,0,0,1,0,0,1,0,0,1,0,0,1 \rightarrow 0 \qquad (4)$$

and Equation (3) may be reformulated as

$$0,1,0,0,1,0,0,1,0,0,1,0,0,1 \rightarrow 1 \qquad (5)$$

However, Equations (4) and (5), as prediction tasks in a deterministic world, cannot occur simultaneously. Therefore, sequence recognition is different from sequence prediction generally.

3. MEHTODOLOGY: MULTIVARIATE ADAPTIVE REGRESION SPLINES MODEL

Multivariate Adaptive Regression Splines (MARS) introduced in [10], is a nonparametric modeling strategy

that could be considered as an extension to Recursive Partitioning Regression mentioned in [10] and improves it in continuity and estimation ability .

Consider $x = [x_1, x_2, L, x_D]$ as an observation of dimension D that is corresponding to target y. MARS finds an approximation of $y = f(\mathbf{x})$, using a hockey stick spline basis functions of the form $[s(x_v - t)]_+$ or their multiplication (Interaction). Here s is ±1 and determines the polarity of the basis function, t is called a knot and $[\ ]_+$ takes the positive part of inner argument and takes zero value elsewhere and subscript v determines the component of $\mathbf{x}$ that contributes to the basis function. Figure 3 illustrates a spline basis function and its mirror image. The output of the MARS model is then a weighted sum of these basis functions

$$\hat{y} = \hat{f}(\mathbf{x}) = a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{K_m} \left[s_{k,m} \left(x_{v_{k,m}} - t_{k,m} \right) \right]_+ \qquad (6)$$

where M is number of basic functions; $v_{k,m}$ determines the dimension of $\mathbf{x}$ that contributes in k-th interaction of m-th basis and $t_{k,m}$ is the knot of m-th basis in dimension $v_{k,m}$ of $\mathbf{x}$. K_m represents maximum allowed interaction(s) for basis function m; K_m should be set to one if it is known that there is no interaction between different dimensions. In addition to constrain determined by K_m, MARS is also constrained to avoid a dimension of $\mathbf{x}$ more than one time in a basis function. To complete the MARS model, it is needed that $a_0, a_1, \cdots, a_M$ to be estimated by a Linear Least Squares or any similar method.

MARS includes a forward/backward algorithm for building the model [10]. In forward path, the model grows up to maximum number of basic functions determined by an expert. Backward procedure prunes this model and removes basis functions that their removal does not discard performance of the model. In this algorithm, performance of MARS model is measured using mean squares error criterion. This penalty increases when the number of knots in the model increases and is weighted with a factor namely penalty gain that is adjusted by the expert. In [10], it is shown that MARS has the ability to find dimensions of $\mathbf{x}$ that really contribute to y and denies those that are not necessary in modeling $y = f(\mathbf{x})$. Therefore, it is dependable in high dimensional cases. Furthermore, local nature of MARS model gives it power of fitting to nonlinear models. MARS, in this discipline, may be employed for a regression or similarly prediction task. However, behavior of MARS may be completely changed into a binary yes/no classifier using a simple threshold value on output during training process.

4. EXPERIMENTAL RESULTS

It was discussed in Section 2, that sequence prediction and sequence recognition are not always the same. Here, the idea is tested in practice. Fortunately, in the case of seismic data, there are evidences of some patterns available for both current approaches of sequence learning. Almost regular earthquakes at Parkfield on San Andreas Fault (Figures 4 and 5), suggest a very simple pattern for predictability of earthquakes events in a region. Furthermore, repeated waveform patterns for some events, in Parkfield, promise sequence recognition plausibility for earthquake data. Even more complex subsequences may be extracted by a recognition machine.
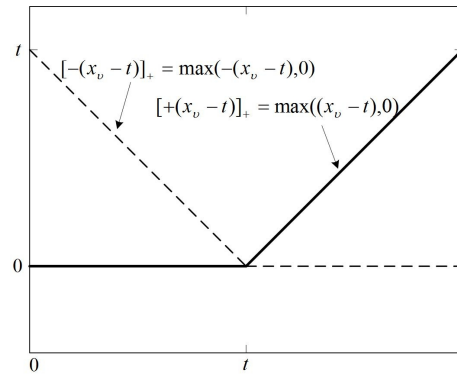

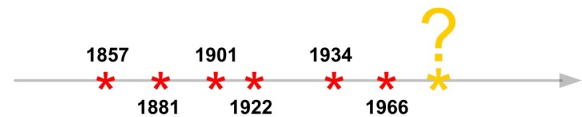**Figure 3.** The spline basis function $s = +1$ and its mirror image $s = -1$.

Figure 4. Almost regular earthquakes at Parkfield on San Andreas fault during 1857 through 1966.
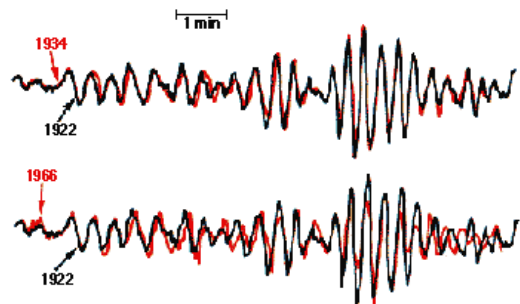

Figure 5. Recordings of the east-west component of motion made by Galitzin instruments at DeBilt, the Netherlands. Recordings from the 1922 earthquake (shown in black) and the 1934 and 1966 events at Parkfield (shown in red) are strikingly similar, suggesting virtually identical ruptures (*image courtesy* of the U.S. Geological Survey).

On the other hand, a good methodology is employed, meaning that, MARS could be used as both, a predictor and a classifier. The MARS code employed in this paper is provided by Salford Systems [12] and runs on a Pentium® 4 2.00 GHz, 512 MB of RAM computer.

A rectangular area is chosen (Figure 6) located on the pacific ring of fire around Japan with many events in a short period of time. A very interesting characteristic of data in this region is event clustering in space (Figures 6 and 7) and time (Figure 8). Data is achieved by USGS catalogue, which is publicly accessible through USGS website. In this area, there are totally 7208 events $M \geq 5$ in 17428 days from Aug. 1961 through Mar. 2009. In experiments here, data up to Aug. 2001 has served as training set, and the remaining data are used for test set. Now two different scenarios are defined for target of the model. In the first scheme, the time (in days) of next earthquake $M \geq 5.5$ is fed to the model. In this scenario, MARS is used as a predictor machine. In second scheme, the MARS model is employed to give a yes/no answer to the question: if there is an event $M \geq 5.5$ in the following 10 days, to train an immediate alarm system. Some portion of the output of this model is magnified in Figure 9.

Here, the problem is a sequence recognition task, and the machine is a binary classifier. The results over training and testing sets are presented using confusion matrices in Tables 1 and 2, respectively.
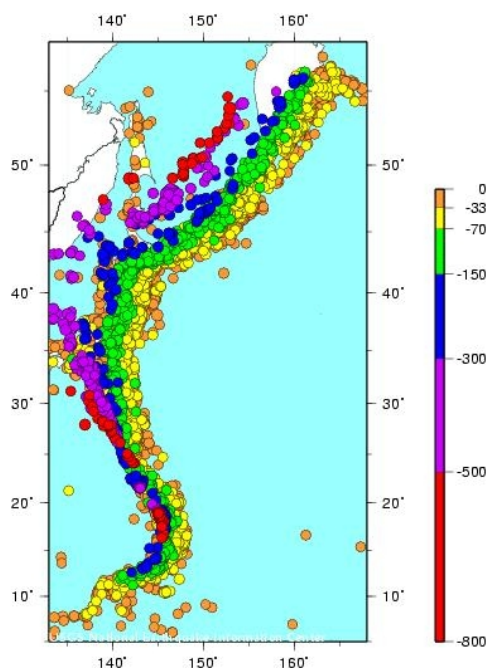
TABLE 1. The confusion matrix of the results of the earthquake prediction problem, as a sequence recognition/prediction task, over training set (Aug. 1961~March 2001).

Method	Actual Class	predicted 0	predicted 1	Total correct (%)
Seq. Recognition	0	5354	3331	46.07
	1	4402	1421	
Seq. Prediction	0	4208	4477	36.11
	1	4792	1031	

TABLE 2. The confusion matrix of the results of the earthquake prediction problem, as a sequence recognition/prediction task, over testing set. (March 2001~March 2009)

Method	Actual Class	Predicted 0	Predicted 1	Total Correct (%)
Seq. Recognition	0	637	925	38.48
	1	871	487	
Seq. Prediction	0	679	856	30.92
	1	1161	224	

Figure 6. Seismicity of a cluster of events $M \geq 5.5$ around Japan.

TABLE 3. Precision of a correct alarm for an earthquake over training and testing set.

Method	Precision Over Training Set (%)	Precision Over Testing Set (%)
Seq. Recognition	24.4	35.16
Seq. Prediction	17.7	16.17

The results suggest that challenging the problem as a sequence recognition task has better results. Table 3 presents the precision of correct earthquake alarms for the two approaches. The MARS model has a better learning ability and generalization performance when it is used for earthquake prediction as a classifier. This means that to make an earthquake alarm in this case study, it is more efficient to track the sequence of events to extract earthquake patterns. On the other hand, the alternative approach, i.e., sequence (time series) forecasting has a lower efficiency.
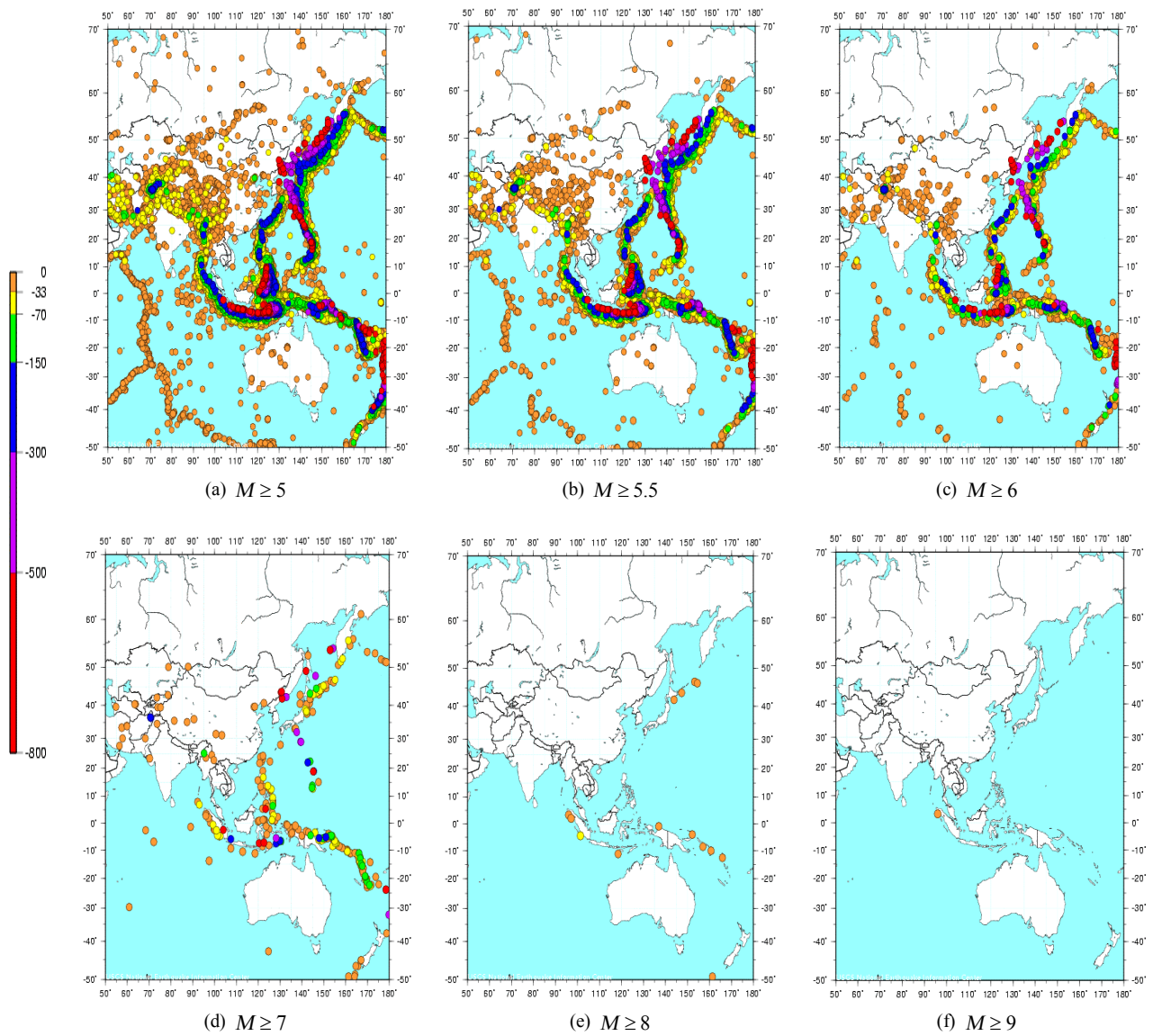
(a) $M \geq 5$ (b) $M \geq 5.5$ (c) $M \geq 6$

(d) $M \geq 7$ (e) $M \geq 8$ (f) $M \geq 9$

Figure 7. Seismicity of a region on pacific ring of fire with different threshold for magnitude. The plot demonstrates event clusters in space.
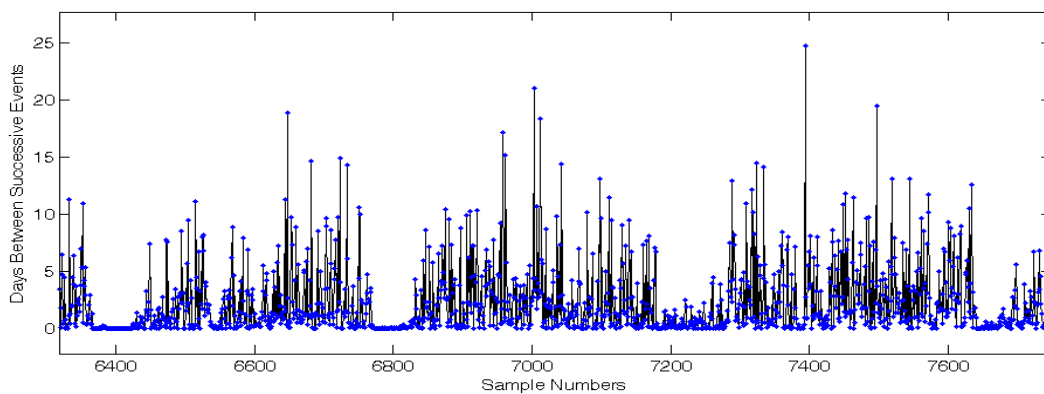

Figure 8. Time interval between subsequent events $M \geq 5$ in Japan cluster of data. The plot demonstrates event clusters in time.

5. CONCLUSION

In this paper, the problem of earthquake prediction was defined in two frameworks, namely, sequence prediction and sequence recognition. It was shown here in theory that these two apparently similar branches of sequence learning are not the same in general and may have different level of efficiency when employed. Therefore, it is very important to challenge the problem from an appropriate point of view.

To demonstrate the idea in practice, a single learning machine, i.e., the model of multivariate adaptive regression splines (MARS) was employed to solve the problem in aforementioned frameworks. The results of applying this model to the seismic data of a region on pacific ring of fire, indicates that the problem of earthquake prediction can be solved more efficiently when it is considered as a sequence recognition task.

7. REFERENCES

1. Rikitake, T., "Earthquake precursors", *Bulletin of the Seismological Society of America*, Vol. 65, No. 5, (1975), 1133-1162.

2. Kagan, Y. Y. and Jackson, D. D., "Probabilistic forecasting of earthquakes", *Geophysical Journal International*, Vol. 143, No. 2, (2008), 438-453.

3. Kowsari, M. and Yazdani, A., "Statistical prediction of the sequence of large earthquakes in iran", *International Journal of Engineering-Transactions B: Applications*, Vol. 24, No. 4, (2011), 325-336.

4. Yazdani, A. and Komachi, Y., "Computation of earthquake response via Fourier amplitude spectra", *International Journal of Engineering*, Vol. 22, No. 2, (2009), 147-152.

5. Yazdani, A., Shahpari, A. and Salimi, M. R., "The use of monte-carlo simulations in seismic hazard analysis in tehran and surrounding areas", *International Journal of Engineering-Transactions C: Aspects*, Vol. 25, No. 2, (2012), 159-166.

6. Liu, N., Davis, R. I. A., Lovell, B. C. and Kootsookos, P. J., "Effect of initial HMM choices in multiple sequence training for gesture recognition", in Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004. International Conference on, IEEE. Vol. 1, (2004), 608-613.

7. Moshiri, B., Eslambolchi, P. and HoseinNezhad, R., "Fuzzy clustering approach using data fusion theory and its application to automatic isolated word recognition", *International Journal of Engineering Transaction B: Applications*, Vol. 16, (2003), 329-336.

8. Anastassiou, D., "Genomic signal processing", *Signal Processing Magazine, IEEE*, Vol. 18, No. 4, (2001), 8-20.

9. Sun, R. and Giles, C. L., "Sequence learning: from recognition and prediction to sequential decision making", *IEEE Intelligent Systems*, Vol. 16, No. 4, (2001), 67-70.

10. Friedman, J. H., "Multivariate adaptive regression splines", *The Annals of Statistics*, Vol. 19, (1991), 1-141.

11. Bicego, M., Murino, V. and Figueiredo, M. A. T., "Similarity-based classification of sequences using hidden Markov models", *Pattern Recognition*, Vol. 37, No. 12, (2004), 2281-2291.

12. Salford Systems. MRAS Software. executable file, http://www.salford-systems.com/mars.php/ 2004; Accessed on 10 September 2006.

Seismic Data Forecasting: A Sequence Prediction or a Sequence Recognition Task

A. Bali [a], M. Mahdinejad Noori [b]

[a] *International Institute of Earthquake Engineering and Seismology, Tehran-19537-14453, Iran*
[b] *Ministry of Science, Research and Technology, Tehran, Iran*

چکیده

دراین مقاله، برازشگر وفقی چند-متغیره با اسپلاین (مارس) برای پیش‌بینی زلزله بر اساس دو راهکار معمول در یادگیری دنباله، استفاده شده است. در سناریوی اول، کار به عنوان یک مساله‌ی پیش‌بینی تعریف شده است، و مدل مارس به عنوان یک پیش‌بینی کننده استفاده شده است. در سناریوی دوم، همان کار از دیدگاه بازشناسی دنباله مورد بررسی قرار گرفته و مدل مارس، این بار، به عنوان یک دسته‌بندی کننده‌ی دودویی به به کار رفته است که نتایج آن می‌تواند برای پیش‌بینی یک زلزله استفاده شود. نتایج اعمال روش‌ها به خوشه‌ای از داده‌های زلزله‌شناسی بر روی حلقه‌ی آتش اقیانوس آرام مشخص می‌سازد که مارس به عنوان دسته‌بندی کننده عملکرد بهتری از پیش‌بینی کننده‌ی مارس دارد. در واقع، در حالی که هر دو راهکار عملی هستند، اما بهترین نتایج زمانی حاصل می‌شود که مساله‌ی پیش‌بینی زلزله به عنوان یک کاربرد بازشناسی دنباله در نظر گرفته شود.

doi: *10.5829/idosi.ije.2013.26.02b.04*