

A Type-2 Fuzzy Neural Network Ensemble to Predict Chaotic Time Series

Victor M. Torres and Oscar Castillo

Abstract This paper presents a type-2 adaptive fuzzy neural network ensemble to predict chaotic time series in combination with the well known M8 algorithm. The chaotic time series is depicted by the register of seismic events and their seismic coordinates in a catalog. ANFIS model are used as components of the Ensemble to train and evaluate seven chaotic time series that are used by the M8 algorithm to make a prediction.

1 Introduction

A prediction is significant only to the extent it is successful beyond chance [1].

As a science, earthquake prediction is still not very mature [2]. Statistical methods have been proposed to determine the probability that something can happen, especially in the seismic field of investigation, but they are not well suited to deal with uncertain systems [3].

Prediction systems have been proposed based upon statistical methods and neural systems [1], most of them using a single adaptive neural system or an adaptive fuzzy neural system [4, 5]. The Ensemble paradigm introduces the use of several fuzzy neural networks functioning together in order to do a better prediction.

In this paper, a prediction system is described. The statistical method used is the well known M8 algorithm. This method has proven to be useful in determining

V.M. Torres

Carretera Ensenada-Tijuana No. 3918, Zona Playitas, 22860 Ensenada, B.C., Mexico

O. Castillo (✉)

Tijuana Institute of Technology, P.O. Box 4207 Chula Vista, USA

e-mail: ocastillo@hafsamx.org

seismic events of 6.0+ magnitudes, but not accurate enough. This statistical method is combined with an Ensemble of Fuzzy Neural Networks (ANFIS) to overcome the uncertainty of a real system.

2 Definitions

2.1 Neural Network

Neural networks are a form of computational multi-threaded model involving simple processing elements with a high degree of interconnection and adaptive interaction among them [4] (see Fig. 1).

A neural network is a system that approximates the operation of a human brain.

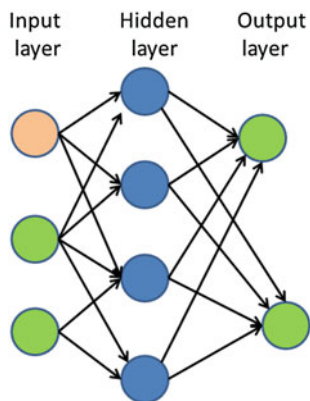
2.2 Adaptive Neural Network

An adaptive neural network is a system that processes information and makes adjustments to the network when necessary based on their own evaluation of how to most efficiently carry out their functions.

There are two main ways an adaptive neural network “learns”: supervised learning and unsupervised learning. Supervised learning requires a human counterpart who instructs the network on how to interpret and interact with various inputs in order to ensure that there are no errors in the methods that the adaptive neural network uses to process information.

Non-supervised learning relies on the central processing unit interacting with its environment and making its own decisions on how it should operate based on its original programming.

Fig. 1 Neural network diagram



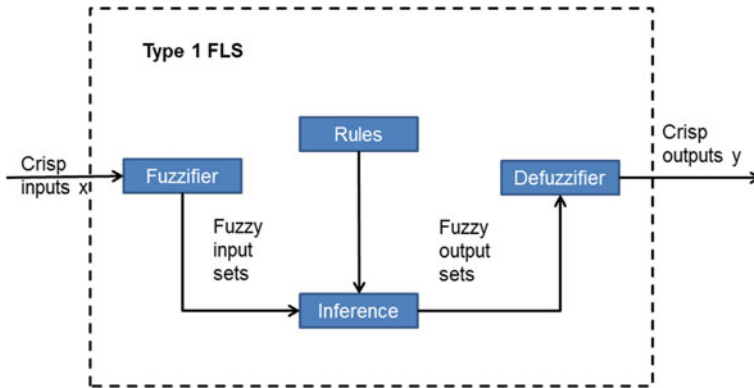


Fig. 2 Type 1 fuzzy logic system diagram

2.3 Fuzzy Logic

Fuzzy sets were first proposed by Zadeh [6] in 1965 as a way to deal with imprecise or uncertain data. Fuzzy set theory is effective in the linguistic representation of information and incorporates human knowledge dealing with uncertainty and imprecision.

A type-1 fuzzy logic system (FLS) is a system that uses fuzzy set theory to map inputs to outputs. It contains a fuzzifier block that becomes in fuzzy sets any crisp input data, an inference block that acts in the same way a person do who has do deal with imprecise data, a set of rules or expert knowledge depicted in linguistic values and an defuzzifier block that reduces the output fuzzy set in a crisp output. (see Fig. 2).

In 1975, Zadeh [6] introduces inference systems diffuse type-2 (IT2FLS) as an extension to type 1 fuzzy inference systems. Mendel [7] develops the theory for these sets which have been shown as more effective than types 1.

A type-2 FLS is very similar to a type 1 FIS but a type-2 FLS incorporates a type-reduction block to reduce an interval type-2 fuzzy set into an interval-valued type-1 fuzzy set. A fuzzy set type-1 reaches the defuzzifier block to produce a crisp output (see Fig. 3).

2.4 Fuzzy Neural Network

Union of neural networks and fuzzy logic produces a fuzzy neural network. This network contains a set of nodes arranged on layers that perform the same task independently. Each node is fixed or adaptive [8] (see Fig. 4) and each layer performs different tasks, as described below:

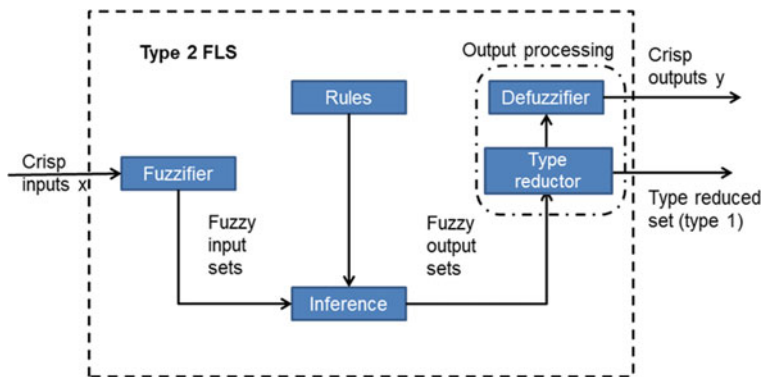


Fig. 3 Type 2 fuzzy logic system diagram

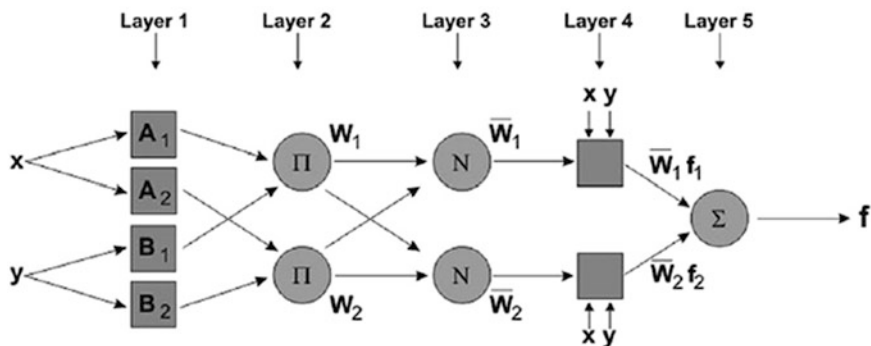


Fig. 4 Fuzzy neural network. Source Monika [8]

- Layer 1—Fuzzification layer. Each node in this layer is an adaptive node. The nodes in this layer are called I . The parameters in this layer are called assumptions.
- Layer 2—Inference layer. Each node in this layer is static. The nodes in this layer are called P , whose output is the product of all input signals. The output of each node represents the application of a rule.
- Layer 3—Implication layer. Each node in this layer is static and they are called by N . The i -th node calculates the radius of the i -input rules. The output of this layer is called normalization.
- Layer 4—Aggregation layer. Each i -th node in this layer is an Adaptive node. The parameters in this layer are referred to as the consequences.
- Layer 5—The single node in this layer is a static node, called R , which calculates the sum of all signals output.

2.5 Adaptive Neural Network Fuzzy Inference System (ANFIS)

ANFIS is an implementation of an adaptive neural network of a Takagi-Sugeno fuzzy inference system which integrates Least Square Estimation and Back-propagation algorithms for a rapid learning speed. ANFIS are widely used to explain past data and predict future behavior [9].

2.6 Ensemble

Zhou [10] proposes that the use of several systems of neural networks functioning together do better predictions than separately. This arrange of fuzzy neural systems is called Ensemble. An ensemble is a learning paradigm that includes multiple components learners that are trained to deal with the same task [11], in others words, several neural networks are attached and are trained to solve a problem [12, 13].

Neural ensemble models can be formed in many different manners. One general approach to creating ensemble members is to hold the training data constant and to vary different factors/parameters of neural networks. For example, ensemble components can be networks trained with different initial random weights, networks trained with different algorithms, networks with different architectures that may include varying number of input and/or hidden nodes, or even networks with different types.

The main idea of the Ensemble is that each fuzzy neural network that integrates has a different way of train and be stimulated, so it brings a different point of view of the same task [3].

3 Data Used in This Study

3.1 Seismic Coordinates

In seismology, a tremor is “a sudden and transient shaking of the Earth’s crust which releases energy in the form of strong elastic waves” [13]. Each registered tremor has at least the same basic data, which are known as seismic coordinates. These data are:

- Magnitude
- Time of the event
- Latitude and longitude, and
- Depth.

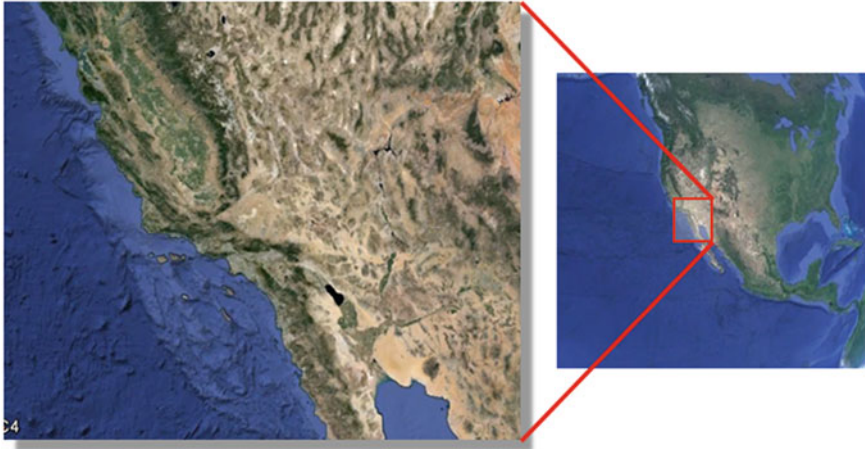


Fig. 5 Region of study

After a seismic event, it is common that events of lesser magnitude following the main arise. These events are known as aftershocks, and decreases with the time.

3.2 Seismic Catalog

Data used in this approach were acquired from the catalog of Southern California Earthquake Data Center. This catalog contains information about seismic events spanning from 1933 to present day.

Seismic coordinates listed above are present in the catalog. Data accumulated over decades allows the use predictive algorithms to estimate future events, such as M8 [14], M8S, CN among others.

An initial set of data spanning from 1993/01/01 thru 2014/09/30 is used in this study. Geographically, this catalog includes data from 30.0° to 39.0° latitude and -111.0° to -124.0° longitude (see Fig. 5). Data contained in this set were pre-processed to eliminate possible duplication and aftershocks [15].

4 Preprocessing Data and M8 Algorithm

4.1 Duplication and Aftershocks

The elimination of duplication of events in the catalog was performed by simple comparison of seismic coordinates. For the removal of aftershocks, Gutenberg-Richter and Omori law [16] were applied.

Gutenberg-Richter law [16] allows setting the scale of the tremors, determining the number of aftershocks that follow a strong land movement in a given region and over a period of time. This relationship is calculated as:

$$N = 10^{(a-b)M} \quad (1)$$

where:

N is the number of events with $M \geq M_0$
 M is magnitude
 a and b are constants

Typically, b is equal to 1 in seismic active regions and is equal to the rate of earthquakes in the region.

In the other hand, Omori law [17] expresses the decay of aftershocks in a time interval. This law is expressed as:

$$n(t) = \frac{k}{(c + t)} \quad (2)$$

where:

$n(t)$ is the probability that a seismic event shows up in time t
 c is the days after the main shock
 t is the time of the event

Utsu [5] in 1961 proposed an amendment to this rule so that the probability of the aftershock decreases exponentially within time. As shown in the next equation.

$$n(t) = \frac{k}{(c + t)^p} \quad (3)$$

Both of these laws were used to determine independent and dependent seismic events and to eliminate aftershocks from the catalog.

4.2 M8 Algorithm

The M8 algorithm [14, 18] was designed by retrospective analysis of the preceding seismicity of the great movements of the world ($M \geq 8$). This prediction scheme is implemented as follows:

- The territory to be considered is explored by overlapping circles of diameter equal to $D(M_0)$, where D is diameter of the event of M_0 magnitude.

- In each circle, the sequence of earthquakes is considered eliminating their replicas. The sequence is standardized by the lower magnitude $M_{\min}(N)$, with N as a standard value.
- Time-dependent functions characterize the sequence that is calculated in a window of time (t, t) and a range of magnitudes $M_0 > M_i \geq M_{\min}(N)$. These functions include:
 - $N(t)$, number of main shocks.
 - $L(t)$, standard deviation de $N(t)$ from the long-term trend.

$$L(t) = N(t) - [N_{\text{cum}}(t - s)] \cdot \frac{t - t_0}{t - t_0 - s} \quad (4)$$

- $N_{\text{cum}}(t)$ is the total number of main shocks with $M \geq M_{\min}$ from the beginning of the sequence t_0 to t .
 - $Z(t)$ is the concentration of major shocks estimated as the average of the diameter of the source radio l to the average distance r between them.
 - $B(t) = \max\{\text{bi}\}$, the maximum number of aftershocks, seen in a time window.
- Each N , L y Z function is calculated for $N = 10$ and $N = 20$. As a result seven functions, including the B -function, describes the earthquake sequence.
 - High values in functions are identified by a condition that exceed Q percentiles ($\geq Q$ %).
 - An alarm is declared in the following five years, when at least six functions, including B , present values that exceed Q . The prediction becomes more stable for two moments, t and $t + 05$ (in years).

5 Architecture of the Ensemble

The architecture of the ensemble is divided into five sections (Fig. 6). The first one is represented by the data base, which consists in the preprocessed catalog.

The second section consists of the ANFIS that are trained and used to validate data; three of them are enough to carry out this study. In the third section, the trained ANFIS are used to generate an initial prediction. The outcomes of the ANFIS are combined in the fourth section and the fifth section a final prediction is made.

Seven time series recorded at six months intervals are set up in the Ensemble. Two series characterize the current level of seismicity within a given region of study. Two others characterize the seismic level based on a ratio of events with magnitude $M > M_0$ over the total number of events. Two more are defined as departure values for a six year period. The seventh series represents the maximum number of aftershocks associated with a main series event.

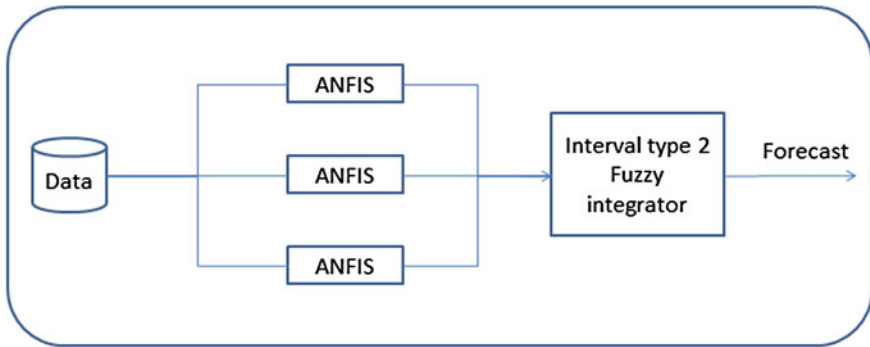


Fig. 6 Seismic hazard predictor ensemble

6 Assumptions

Due the nature of the data, the ensemble is dealing with a chaotic time series. The Database contains the seismic coordinates of all the registered events through the years.

The rates of $M > M_0$ target earthquakes are not known exactly and their empirical estimates are unreliable. Earthquakes with lower magnitude than 5.6 are being used to generate a short and medium term prediction. A variant of the M8 algorithm has to be used to predict earthquakes of magnitude 4.0 and above.

ANFIS deals with the same circles of interest in the study region. A minimum change in the architecture of the Ensemble will allow to ANFIS deal with different and overlapping circles of interest in the same research region. In that case, the resulting prediction would be only over the overlapping region of the circles of interest. Moreover, this minimum change modifies the general approach for creating ensemble members becoming the Ensemble in a modular system.

7 Conclusions

To be useful, an earthquake prediction must be precise enough to warrant the cost of increased precautions; including disruption of ordinary activities and commerce, and timely enough that preparation can be made [19].

The M8 algorithm requires a notoriously extensive research region to evaluate the possibility of a seismic event. For events of magnitude 6.0, 6.5, 7.0 and 7.5 circles of interest (CI) have a diameter of 277,078 km (172.20 miles), 384.69 km (239.08 miles), 562.11 km (349.35 miles) and 854.63 km (531.15 miles) respectively.

The above dimensions described about circles of interest shed some criticize on forecasts made by this algorithm as some researchers believe that it is likely that the

successes be more by chance than the goodness of the method. Therefore, a modification of diameter of the circles of interest or how a sequence of earthquake are associate to each circle is necessary in order to predict tremors with magnitude 6.0 and lower.

An adaptation of the M8 algorithm has to be made for various reasons: for the period from 1993/1/1 up to date, the catalog has little information about tremors with 6.0+ in magnitude. In just 20 years there have been seven (7) earthquakes in the region with these characteristics. The maximum magnitude of a tremor has been of 7.2 on April 4th 2010, and only two have been magnitude 7.0+ in 20 years.

The circles of interest should consider sequences of earthquakes that allow predicting tremors magnitudes 4.0 and above that are more numerous in the catalog and are also of interest in the field of seismic prediction.

The amendments to the M8 algorithm modify the behavior of ANFIS that must respond to new requirements to establish more reliable predictions.

Also, using the Ensemble as a base system, a modular approach can be tested bringing a different perspective over the same problem.

References

1. Zamani, A., Sorbi, M.R., Safavi, A.A.: Application of neural network and ANFIS model for earthquake occurrence in Iran. *Earth Sci. Inform* **6**, 71–85 (2013). Springer, Berlin
2. Bolt, B.A.: *Earthquakes and Geological Discovery*. Scientific American Library, New York (1993)
3. Soto, J., Melin, P., Castillo, O.: Time series prediction using ensembles of neuro-fuzzy models with interval type-2 and type-1 fuzzy integrators. In: Conference: IJCNN 2013, Dallas Texas (2013)
4. Tsunekawa, H.: *A Fuzzy Neural Network Prediction Model of the Principal Motions of Earthquakes Based on Preliminary Tremors*. IEEE, New Jersey (1998). ISBN: 0-7803-4503-7
5. Utsu, T.: A statistical study of the occurrence of aftershocks. *Geophys. Mag.* **30**, 521–605 (1961)
6. Zadeh, L.A.: Fuzzy sets. *Inf. Control* **8**(3), 338–353 (1965)
7. Karnik, N.N., Mendel, J.M.: Introduction to type-2 fuzzy logic systems. In: Proceedings in 1998 IEEE FUZZY Conference, pp. 915–920. Anchorage, May 1998
8. Monika, A.K.: Comparison of mamdani fuzzy model and neuro fuzzy model for load sensor. *Int. J. Eng. Innovative Technol. (IJEIT)* **2**(9) (2013)
9. Chen D.W., Zhang, J.-P.: Time series prediction based on ensemble ANFIS. In: Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, 18–21 August 2005
10. Zhou, Z., Wu, J., Tang, W.: Ensembling neural networks: many could be better than all. *Artif. Intell.* **137**(1–2), 239–263 (2002)
11. Dietterich, T.G.: Machine learning research: four current directions. *Artif. Intell.* **18**(4), 97–136 (1998)
12. Yang, J., Yu, P.S.: Mining asynchronous periodic patterns in time series data. *IEEE Trans. Knowl. Data Eng.* **15**(3) (2003)
13. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986)

14. Keilis-Borok, V.I., Kossobokov, V.G.: Premonitory activation of seismic flow: algorithm M8. *Phys. Earth Planet. Int.* **61**, 73–83 (1990)
15. Keilis-Borok, V.I., Knopoff, L., Rotvain, I.M.: *Nature* **283**, 259–263 (1980)
16. Gutenberg, B., Richter, C.F.: *Seismicity of the Earth and Associated Phenomena*, 2nd ed. Princeton University Press, Princeton (1954)
17. Omori, F.: On the aftershocks of earthquakes. *J. Coll. Sci. Imperial Univ. Tokyo* **7**, 111–200 (1894)
18. Keilis-Borok, V.I.: The Algorithm M8. Russian Academic of Sciences. <http://www.mitp.ru/en/m8pred.html> (2009)
19. Thomas, A.M.: Economic impacts of earthquake prediction. In: *Proceedings of the Seminar on Earthquake Prediction Case Histories*, pp. 179–185. UNDP/PRO, Geneva, 12–15 October 1982 (1983)