Chapter 10

Data Mining Techniques on Earthquake Data: Recent Data Mining Approaches

Negar Sadat Soleimani Zakeri
University of Tabriz, Iran

Saeid Pashazadeh
University of Tabriz, Iran

ABSTRACT

Active faults are sources of earthquakes and one of them is north fault of Tabriz in the northwest of Iran. The activation of faults can harm humans’ life and constructions. The analysis of the seismic data in active regions can be helpful in dealing with earthquake hazards and devising prevention strategies. In this chapter, structure of earthquake events along with application of various intelligent data mining algorithms for earthquake prediction are studied. Main focus is on categorizing the seismic data of local regions according to the events’ location using clustering algorithms for classification and then using intelligent artificial neural network for cluster prediction. As a result, the target data were clustered to six groups and proposed model with 10 fold cross validation yielded accuracy of 98.3%. Also, as a case study, the tectonic stress on concentration zones of Tabriz fault has been identified and five features of the events were used. Finally, the most important points have been proposed for evaluation of the nonlinear model predictions as future directions.

INTRODUCTION

Active faults are counted as seismological active regions. Many researchers are interested to study these faults and destructive historical earthquakes (Zhang, Zhang, Yang, & Su, 2014). Occurrence of earthquakes on one hand and daily development of urban regions and critical facilities in seismic zones on the other hand, are sufficient factors for showing importance of reviewing seismo-tectonic analysis methods for predicting earthquake magnitude and its future occurrence time (Ennis, 2010), (Elmi, Ganjpour Sales, Tabrizi, Soleimanpour, & Mohseni, 2013). Earthquake is sudden motion of the ground caused by re-DOI: 10.4018/978-1-4666-8513-0.ch010
leasing energy of stored stress in the fault (Peng, Yang, Zheng, Xu, & Jiang, 2014). After stop of fault’s motion, its waves continue to move along the ground. Many important and populous cities are placed in active faults’ zones. Significant earthquakes have been reported in the recent centuries, while geodesy studies show a significant amount of deformation on both slip sides of fault that emphasize risks of new activities of fault (Orihara, Kamogawa, & Nagao, 2014). Therefore, the analysis of earthquake data is so important (Peng et al., 2014). The kinesiology of the earthquake and its computer modeling is one of the most challenging issues in modern geophysics (Ruiz et al., 2014). After large earthquakes, seismologists are able to express many phenomena associated with earthquake. They will be able to provide successful predictions using information of events and their occurrence order, but it is still difficult to predict earthquakes because of its high complex structure (Dologlou, 2008). Seismic data mining by computer modeling, computer simulation and artificial intelligence techniques can be used to study this phenomenon. Automatic learning techniques can be used to extract relationships between different phenomena and earthquakes (Uyeda, 2013). They can be used to obtain more accurate forecasting based on automated reasoning algorithms (Peng et al., 2014). Different countries like China and Japan which are located in earthquake-prone regions of the world have paid much attention to earthquake predicting sites for dealing with earthquake disasters (Kako, Arbon, & Mitani, 2014; Y. Wang, Li, Chen, & Zou, 2014).

In following part of this chapter basic and well known analyzing methods of seismic data are introduced. In related works, brief review classification and different methods that are used for analysis of the earthquake data are investigated. Then proposed method is presented. A short history behind the earthquakes, the types of data which are used for processing the occurrence of earthquake, databases of a case study that the valid seismic data can be taken from is presented. Also, preprocessing techniques such as normalization the data, data mining methods such as k-means clustering algorithm as a preprocessing step for categorizing the data that can be used for modeling the earthquake function based on related parameters and features to predict the magnitude of earthquake in terms of time and location by ANNs are discussed in section of proposed approach. In section results and discussion section, results of study is presented and discussed. In future research direction section some notes those are important and critical and must be considered in study of seismic data is discussed and finally conclusion is presented.

**PRELIMINARY CONCEPTS AND BASIC METHODS**

According to literature investigation, learning algorithms or data mining methods and in some cases combinations of these two methods are used as high-level classification methods in studies of earthquake data in the seismic regions. Each method has its own subdivisions. The learning algorithms are also known as machine learning algorithms in computer science. Most common methods that are used for this issue are ANNs, fuzzy systems and SVMs. Different types of neural networks have been used in different literature. Recurrent neural networks, PNNs, feed forward neural networks, SVMs and fuzzy based algorithms were used for predicting time and location of earthquakes. Collected data from various parts of the world and different extracted features such as sequence number, occurrence time are used by these methods while existing noises were removed along the preprocessing steps.

Clustering is the most important category of methods with different types of algorithms for clustering time series which have sequential nature or by using agglomerative hierarchical clustering algorithm to extract a temporal pattern for predicting an event. Most of the methods took the advantage of two universal laws including the Gutenberg-Richter (G-R) and the Omori to demonstrate the increasing effect