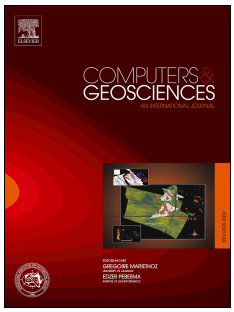
Accepted Manuscript



A novel tree-based algorithm to discover seismic patterns in earthquake catalogs

E. Florido, G. Asencio–Cortés, J.L. Aznarte, C. Rubio-Escudero, F. Martínez–Álvarez

PII: S0098-3004(17)31169-X

DOI: 10.1016/j.cageo.2018.03.005

Reference: CAGEO 4104

To appear in: *Computers and Geosciences*

Received Date: 11 November 2017

Revised Date: 16 February 2018

Accepted Date: 5 March 2018

Please cite this article as: Florido, E., Asencio–Cortés, G., Aznarte, J.L., Rubio-Escudero, C., Martínez–Álvarez, F., A novel tree-based algorithm to discover seismic patterns in earthquake catalogs, *Computers and Geosciences* (2018), doi: 10.1016/j.cageo.2018.03.005.

A novel tree-based algorithm to discover seismic patterns in earthquake catalogs

E. Florido[a], G. Asencio–Cortés[a], J. L. Aznarte[b], C. Rubio-Escudero[c], F. Martínez–Álvarez[a]

[a] *Department of Computer Science, Pablo de Olavide University of Seville, ES-41013 Spain*
eflonav@alu.upo.es, guaasecor@upo.es, fmaralv@upo.es
[b] *Department of Artificial Intelligence, Universidad Nacional de Educación a Distancia, Spain*
jlaznarte@dia.uned.es
[c] *Department of Computer Science, University of Seville, Spain*
crubioescudero@us.es

Abstract

A novel methodology is introduced in this research study to detect seismic precursors. Based on an existing approach, the new methodology searches for patterns in the historical data. Such patterns may contain statistical or soil dynamics information. It improves the former version in several aspects. First, new seismicity indicators have been used to characterize earthquakes. Second, a machine learning clustering algorithm has been applied in a very flexible way, thus allowing the discovery of new data groupings. Third, a novel search strategy is proposed in order to obtain non-overlapped patterns. And, fourth, arbitrary lengths of patterns are searched for, thus discovering long and short-term behaviors that may influence in the occurrence of large earthquakes. The methodology has been applied to seven different datasets, from three different regions, namely the Iberian Peninsula, Chile and Japan. Reported results show a remarkable improvement with respect to the former version, in terms of all evaluated quality measures. In particular, the number of false positives has decreased and the positive predictive values increased, both of them in a very remarkable manner.

Keywords: Seismic time series; earthquake prediction; pattern discovery; clustering.

1. Introduction

The discovery of earthquake precursors is a task of utmost relevance in order to take precautionary measures and prevent human losses. According to Ishibashi [11], such events can be classified into two categories: physical (irreversible rupture process) and tectonics (tectonic slide). Physical precursors are mainly considered for the short and intermediate term.

It is well-known that certain precursory events are correlated to large earthquakes. Actually, a vast majority of major earthquakes exhibit anomalous seismic activity just before they occur. The features include changes in regional activity rate and changes in the pattern of small earthquakes, including alignments on unmapped linear features near the (future) main shock. It has long been suggested that large earthquakes are preceded by observable variations in regional seismicity [25].

The main objective of this work is to generalize the methodology introduced in [17] and extended in [8]. In it, authors applied unsupervised learning to discover significant precursory anomalies. Although the results they obtained were relevant in terms of accuracy, the approach itself exhibited several limitations:

1. Only b-value and time occurrence were considered to discover meaningful anomalies. That is, only two features were considered to characterize seismicity. There exist many other features that may be used.
2. The search strategy was not exhaustive and some patterns were just sub-patterns or shorter patterns of other patterns. Therefore, an improved search strategy must be developed in order to avoid overlapping in discovered patterns.