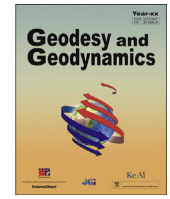
Geodesy and Geodynamics xxx (2017) 1–8

Contents lists available at ScienceDirect

Geodesy and Geodynamics

journal homepages: www.keaipublishing.com/en/journals/geog; http://www.jgg09.com/jweb_ddcl_en/EN/volumn/home.shtml



Fast magnitude determination using a single seismological station record implementing machine learning techniques

Luis H. Ochoa*, Luis F. Niño, Carlos A. Vargas

Universidad Nacional de Colombia, Carrera 45 N° 26-85 Edificio Manuel Ancizar - Of: 330, Bogotá, Colombia

A R T I C L E I N F O

Article history:
Received 11 January 2017
Accepted 22 March 2017
Available online xxx

Keywords:
Earthquake early warning
Support Vector Machine Regression
Earthquake
Rapid response
Local magnitude
Seismic event
Seismology
Bogota
Colombia

A B S T R A C T

In this work a Support Vector Machine Regression (SVMR) algorithm is used to calculate local magnitude (M_L) using only five seconds of signal after the P wave onset of one three component seismic station. This algorithm was trained with 863 records of historical earthquakes, where the input regression parameters were an exponential function of the waveform envelope estimated by least squares and the maximum value of the observed waveform for each component in a single station. Ten-fold cross validation was applied for a normalized polynomial kernel obtaining the mean absolute error for different exponents and complexity parameters. The local magnitude (M_L) could be estimated with 0.19 units of mean absolute error. The proposed algorithm is easy to implement in hardware and may be used directly after the field seismological sensor to generate fast decisions at seismological control centers, increasing the possibility of having an effective reaction.

1. Introduction

The main task in early earthquake warning systems is the estimation of magnitude and source location of an earthquake in a short period of time with the best possible accuracy [1]. Contrary to prediction, seismic early warning systems emit an alert, few seconds after the event initiates, from few seconds to a few tens of seconds before the stronger shaking movement arrives. Many early warning systems employ dense seismological networks to determine the magnitude and localization using at least 3 stations with a very good accuracy [2–6]. Although, for very sparse networks, time for single station localization is not lower enough compared with multiple station solution which is more accurate.

The density of stations in some high seismic risk areas is not enough making that the time used for localization of the event be bigger than the travel time to the area where the early warning is needed. In those cases, an alternative solution can be implemented, by using seismological records of previous events recorded at one single station to calculate magnitude and localization of the event [7]. The solution to this problem is an important step towards an efficient early warning system design for regions with a sparse network [8,9].

Data processing techniques, in a single broadband three-component station, have been developed, mainly using automatic algorithms for detection of P and S waves onsets, which allows an estimation of source location, using the back-azimuth and the apparent surface speed measurements [9–11], or to estimate the seismic moment [1,4,5,12–14].

It is possible to implement warning systems based on one three-component station inspired by some natural processes known as bio-inspired computing, natural computation or computational intelligence. These methods have been successfully used in multiple areas of knowledge, seismology among others, which has made it possible to estimate hypocentral parameters using only a few signal seconds registered at only one seismological station, with an acceptable accuracy to generate reliable alerts; this is very useful in areas with sparse seismic networks [15].

When a seismic event occurs, one of the first parameters that should be determined, as soon as possible, is the magnitude.

* Corresponding author.
E-mail address: lhochoag@unal.edu.co (L.H. Ochoa).

Peer review under responsibility of Institute of Seismology, China Earthquake Administration.

Production and Hosting by Elsevier on behalf of KeAi

http://dx.doi.org/10.1016/j.geog.2017.03.010

Multiple approaches have been used to obtain a first estimation based on one single seismic station information in a very short period of time. There exists a variety of methods ranging from those purely empirical using maximum amplitudes in first 10 s (M_{L10}) [6], up to models based on the amplitude displacement peaks (P_d) [16].

Some authors estimated magnitudes using the first complete peak arrival time, the predominant period and the Fourier amplitudes for the initial part of Ground Motion accelerometers signal. However, the error was of ±1.35 units of magnitude, calculated with a single station up to ±0.5 when eight or more stations were employed [17]. Others tried to estimate the magnitude of the initial portion of the seismogram but had larger errors [8,18]. Later, it was possible to determine the local magnitude (M_L) of large earthquakes greater than 5, based on the determination of the maximum amplitude, using the first ten signal seconds (M_{L10}) in a network of accelerometers using a linear empirical relationship from data of 23 earthquakes [6].

Other authors used the linear relationship between predominant period of the P wave and the magnitude, using the first 4 s of the signal [8], which is currently under controversy, but it is still being used with quite good results. The magnitude could be estimated with an error of ±0.33 units [19]. An interesting fact about this result is that only few recorded seconds were used despite having included events of great magnitude for which this short record did not contain the complete rupture process information.

On the other hand, Kernel-based methods have become a very powerful tool for mathematicians, scientists and engineers, providing a very rich and surprising solution in areas such as signal processing and pattern recognition [20]. Its implementation is quite simple by applying a function that combines the input variables as a combination of themselves applying a function of dot products, obtaining an enhanced new space with more dimensions, mapping variables in a hyperspace where separation of classes (in the case of classification) by a linear function or hyperplane can be achieved.

In this work, we present the results obtained with the implementation of kernel methods, in local magnitude determination for seismic events based on the information of a signal's few seconds after the arrival of the P wave, recorded on a single three component seismological station, which allows the generation of one initial alert that can be used for the beginning of first prevention tasks such as the activation of alternative sources of energy, detention of critical activities and decreasing speed in transport systems, among others. We proposed the use of a Support Vector Machine Regression (SVMR) algorithm to obtain the value of local magnitude based on some signal features as input variables also known as descriptors.

2. Study area and data and methods

This study was developed using data form the National Seismological Network of Colombia that belongs to the "Servicio Geológico Colombiano" (SGC). It has a main network composed by 42 stations transmitting in real time and is presented in Fig. 1. The average distance between stations is 162 km and this is one of the reasons why this study is pertinent to generate early warnings with on single station.

The Area of study corresponds to the Bogota's Savannah and its surroundings. This area is the most important region of the country because presents a high seismicity that can affect Bogota-Colombia City which is the capital of the country and the most important social and economic populated center (Fig. 2).

The data set used in this research belongs to the "*El Rosal*" seismological station, which is part of the Colombian Seismic Network administrated by The "*Servicio Geológico Colombiano* − *SGC*" (Colombian Geological Service). The station is located near to *Bogotá* − *Colombia* city. It uses a brand Guralp CMG − T3E007 sensor, in the three components, and a Nanometrics RD3-HRD24 digitizer, which provides simultaneous sampling of three channels with a 24-bit resolution [21].

The data used correspond to the three component raw waveforms recorded directly at seismic station, and the seismic catalogue with complete characterization of 2164 seismic events, selected between January 1st, 1998 and October 27th, 2008, located at less than 120 km from the seismic station. Each waveform file was converted to the ASCII format, using a tool of SEISAN package [22]. Earthquakes with magnitudes lower than 2 were ignored, therefore the processing was performed on the remaining 1011 events.

Since the selected seismic records present variable levels of noise, it was necessary to filter them out with both high and low frequency filters. Low frequencies correspond to instrumental noise that can be easily eliminated through the implementation of a high-pass filter with a cut-off frequency of 0.075 Hz [16], while high frequencies were removed with a low-pass filter with a cut-off frequency of 150 Hz.

Statistical distribution of the local magnitude values is presented in Fig. 3, showing the magnitude distribution in the dataset. This distribution shows that available data does not have a uniform distribution with a very high number of low magnitude events and vice versa. This is a natural condition of this phenomena and, therefore, it is not possible to obtain a statistical homogeneous distribution, but this is a feature that must be managed in the model since this is the way this occurs in the real world. In fact, it obeys the well-known magnitude and number of seismic events relation proposed by Gutenberg and Richter [23], established as one of the parameters for seismic behavior analysis [24] (Fig. 4).

In this research, some parameters that have been successfully used for magnitude estimation were selected as input variables or descriptors for the SVMR algorithm.

First, we used the relationship between the maximum amplitude of the wave in a short period of time, with the local magnitude of the earthquake [25], which is obtained directly taking the maximum amplitude recorded until the time of the signal under consideration. Consecutive maximum peaks were pointed out and a linear regression was performed, for each one of the three components, to correlate, not only the maximum peak, but the way it changes while energy is arriving to the sensors. Three basic parameters were taken from the linear regressions which correspond to slope (M), independent term (B) and correlation coefficient (R), for each of the three components. The maximum amplitude values (Mx) obtained for each component's time window were used as descriptors as well. Therefore, each event had 12 descriptors associated with this concept.

To take into account the effect of the travel path in the magnitude determination, we calculate the parameters A and B previously used in epicentral distance estimations, by adjusting by linear regression of an exponential expression in time ($Bt\exp(-At)$) which belongs to the envelope of the seismic record in a logarithmic scale [1] determined by linear regression and its respective correlation coefficient (R), for each of the three components, adding nine additional descriptors.

In a similar way, some parameters used for previous back-azimuth determination were used to include information about the source location of the seismic event into the model. Maximum eigenvalues of the two-dimensional covariance matrix were used as input descriptors, calculated as described in Refs. [10,26]. A windowing scheme, with one second time windows was performed, obtaining consecutive values for which a linear regression was calculated, in a similar way as described above, determining