Wavelet decomposition and advanced denoising techniques for analysis and classification of seismic signals

*L. Galli[1], Castellani C.[1], Pace G. [1], Saccorotti G.[2]

[1] Advanced Computer Systems, Rome, Italy
[2] National Institute of Geophysics and Volcanology (INGV), Pisa, Italy

* Corresponding author

Abstract

This work describes an automatic classification procedure for seismic signals suitable for the analysis of complex, broad-band waveforms commonly associated with fluid-rock interaction in volcanic and hydrothermal systems. Based on Discrete Wavelet Transform, a set of significant seismic signal features that characterize the type of event is identified (e.g. noise, volcano tectonic, long period). These features are initially assessed for events whose category (class) can be previously determined by an expert analyst. A Bayesian Pattern Recognition supervised technique based on these features is adopted to classify a new 'unlabelled pattern', whose class is unknown. In this way values computed for known events are used to classify events of unknown identity ('supervised classification'). A test was performed on seismological data recorded at Campi Flegrei (Italy), which was divided into three classes. Automatic classification accuracy ranges from 82% to 100% over a broad range of datasets.

Introduction

Seismic signals observed on active volcanoes reflect a variety of source processes, from brittle failure of rocks to the acoustic resonance of buried, fluid-filled fractures and cavities (e.g. Chouet, 1996; 2003). The resulting seismic waveforms depict characteristic signatures differing in both wave morphology and spectral content. The rapid and quantitative identification of these different signal types is a primary goal as it helps gain insights into the physical processes controlling the volcanic activity. Although an experienced analyst

may easily discern between these different event classes based on visual inspection or spectral analysis, more sophisticated tools are needed in order to improve both the efficiency and reliability of automatic classification procedures.

Recently, there have been several attempts to address the problem of automatic classification of signals recorded by regional and local networks. For instance, Benitez et al. (2007) used Fast Fourier Transform and Hidden Markov Modelling to classify local volcano tectonic earthquakes, long period events, hybrid events and volcanic tremor recorded at Deception Island, Antarctica. Scarpetta et al. (2005) and Del Pezzo et al. (2003) used a neural network approach to distinguish volcano tectonic earthquakes from quarry blasts and underwater explosions recorded at Vesuvius Volcano, Italy. For the analysis of the seismic signals they used the linear prediction coding (LPC). Gendron et al. (2000) used Discrete Wavelet Transform and Denoising Techniques for analysis of seismic signal, and the Bayesian approach to classify signals of the New England Seismic Network of Western Observatory of Boston College.

The work herein describes an automatic classification procedure of seismic signals suitable for the analysis of complex, broad-band waveforms commonly associated with fluid-rock interaction in volcanic and hydrothermal systems.

Methodology

In this work a supervised classification approach is adopted, which is based on a set of pre classified signals defined as the *training set.* The first step is to decide which set of signal features should be computed to obtain values that best differentiate events in different classes. This will be our *feature vector*. The classifier shall be trained based on the training set feature vector statistics assessed on a training set of signals. These statistics allows us to classify new ‘unlabelled patterns’ (new signals) whose class is unknown. Two major steps are implemented: (1) computation of the characteristics (*features*) that describe the seismic event in terms of a feature vector; (2) application of a Pattern Recognition technique based on the above features, to classify the seismic event.

Signal Analysis: Wavelet Transform

The quality of the feature vector is related to its ability to discriminate examples from different classes, i.e. to divide the features space in well separated regions. The aim is to extract a small set of features from the seismic signals that adequately characterise the seismic event. The completeness of the features space (in terms of the information needed to classify seismic events) determines the model robustness.

The main signal components are frequency, time and energy. Traditionally the seismic signals are analysed in the frequency domain using Fourier transform. This technique, though providing interesting results, has major limitations due to the so-called *time-frequency* ambiguity. Large time windows produce better frequency resolution, but at the cost of a coarse time resolution (a problem for non-stationary signals), whereas the reverse is true for small moving windows. Therefore it is impossible to clearly resolve simultaneously high- and low-frequency events in a single analysis. Considering seismic signals are very complex and occur in a wide spectral range, this is a significant limitation. Discrete Wavelet Transform (DWT) offers the following advantages: (1) resolves the time-frequency ambiguity; (2) analyses multiple scales with suitable windows enhancing phenomena on each scale; (3) allows for the use of base functions different from sin/cosine waves, and (4) can be computed efficiently via a fast recursion of convolution-decimation operations (Mallat, 2001)

Denoising and Feature Extraction

As shown in Gendron et al. (2000), a seismic signal *s(t)* can be modelled as the sum of a noise process *e(t)* and the part of the signal that describes the seismic event *r(t)*:

$$s(t) = r(t) + e(t)$$

Describing it in terms of wavelets coefficients $d_{(j,n)}(s)$ we use;

$$d_{(j,n)}(s) = d_{(j,n)}(r) + d_{(j,n)}(e)$$

where *j* is the scale and *n* the number of points in each scale. The goal is to minimize the contribution of the noise process *e(t).* Different techniques can be used for this purpose (Mallat 2001 and Donoho 1995). The adopted strategy can be summarized as:

- *Threshold estimator, with soft threshold preferred to hard threshold*. Values below the threshold (T) are considered as noise and are set to 0. All other values are rescaled using the threshold itself (Donoho 1995).
- *Donoho-Johnstone threshold. T* is estimated using the Donoho-Johnstone (1994) strategy for which $T = \sigma\sqrt{2logN}$ where N is the signal length and σ is the noise standard deviation.

Using this approach the threshold that we use is a soft threshold given by;

$$T = \sigma\sqrt{2logN}$$

where σ is a noise standard deviation estimation computed as the median of the absolute values of the wavelet coefficients for every wavelets scale, i.e.

$$\sigma = \frac{M_x}{0.6745}$$

where M_x is the median of the absolute value of the wavelet coefficients. This allows an estimation independent of the signal amplitude and therefore not biased by the presence of high amplitude signals or long term tremors, which would effect the estimate of the noise itself. It has been demonstrated that the above formula is a good approximation of the noise variance (Mallat, 2001). Using this threshold we let

$$\left|\left[d_{(j,n)}(s)\right]\right| < \lambda T \Rightarrow d_{(j,n)}(r) = 0$$

with λ = *const* customisable in the application configuration file.
Let

$$I_r = [j,n] : \left|\left[d_{(j,n)}\right]\right| > \lambda T$$

As in P.Gendron et al. (2000) we define:

- *Beginning of Event Signal* as the minimum time

  $min \quad n2^j$ such that $[j,n] \in I_r$

- *Peak of the waveform* as the maximum

  $max \left|\left[d_{(j,n)}\right]\right|$ such that $[j,n] \in I_r$

  and then the time linked to this peak is

  $$n_{\max} 2^{j_{\max}}$$

- *End of Event Signal* as the maximum time

  $max \quad n2^j$ such that $[j,n] \in I_r$

The above definitions allow designation of a set of signal features (as reported in Table 1), that describe the signal in terms of scale, time and energy. Moreover, estimation of these values is not computationally over-demanding.

bpw	time duration between beginning and peak waveforms
pew	time duration between peak and end waveforms
sbw	scale at beginning of waveform
spw	scale of peak waveform
sew	scale of end waveform
lsw	last scale of denoised waveform
bwe	beginning of waveform energy
pwe	peak waveform energy

Table 1. Features set extracted

Classification

A classifier tries to 'learn' the relationships between the predictors and the responses that allow a new observation, whose response is unknown, to be assigned to one of the *K* predetermined classes. Adopting a Bayesian method reduces the problem of determining the probability distribution of each of the features for each class of event. The classifier shall be *trained* using a set of known pre-labelled signals. In this way, it is possible to estimate a Probability Distribution Function (*pdf*) in features space for each class. Based on the set of features listed in the table above (applying a pattern-recognition technique), known events are described and, successively, the classification of new unknown events is performed based on the Bayes decision rule (Jain et al., 2000). Given pre-classified signals in *K* predetermined classes described by *n* features, the *training phase* comprises:

i. for each given class
   a. pre-process all signals and obtain a set of *n*-dimensional vectors *X*
   b. these vectors are viewed as observations drawn randomly from class conditional probability functions [$p(X \mid \omega_i)$, $i = 1, \ldots, k$]
      Where ω_i are the *k* predetermined classes. We use the set of this functions to estimates *pdfs*

ii. Memorise *pdfs* for future classification.

Given an unclassified signal, *classification phase* comprises:

i. Pre-processing signal, get the *n*-dimensional vector *X* of its features
ii. Use *pdfs* to associate *X* with a probability for each possible category-class ω_i
iii. Use a *decision rule* (for example getting the class with the major probability) to classify the signal.

Density estimation

Density equation is the problem of modelling a density $P(X)$ given a finite number of data points X_n, and determine a density function. Most widely used methods are.

Normal:	parametric Gaussian
LogNormal:	parametric log-normal
k-NN:	non parametric *k-nearest-neighbour*

The method that gives the best performance in our case is *k*-NN, with *k* = 10.

Decision Rule

The decision rule is used to classify a new seismic signal. Given *k* classes ω_i ($i = 1, \ldots, k$), *n* features, a new event X, is described in terms of features vector $X = (x_1, \ldots, x_n)$

Then, using *k-nearest neighbor* estimation we compute the *likelihood function* $P(X \mid \omega_i)$

Using the *prior probabilities* $P(\omega_i)$

set in the configuration file, we compute the *posterior probabilities* given by the Bayes' theorem

$$P(\omega_i \mid X) = \frac{P(\omega_i \cap X)}{P(X)} = \frac{P(\omega_i)P(X \mid \omega_i)}{\sum_{i=1}^{k} P(\omega_i)P(X \mid \omega_i)}$$

The class of the new event *X* will be the ω_i for which $P(\omega_i \mid X)$ the is maxima.

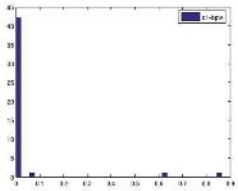
Classification Test Result

A test case was performed on a seismological data set recorded at Campi Flegrei caldera (Italy), which is inhabited by about 1.5 million people. Its magmatic system is active, as evident from the 1538 AD Monte Nuovo eruption (Di Vito et al. 1999), recent ground uplift episodes in 1969–1972 and 1982–1984 have generated a net uplift of 3.5 m around the town of Pozzuoli and there are widespread occurrences of fumaroles and thermal springs. The combination of dense urbanisation, and very active short-term deformation increases the local volcanic risk. As a consequence, major efforts are currently aimed at improving the monitoring procedures in order to ensure a prompt and reliable response during resumption of volcanic activity. The test dataset corresponds to a recent ground deformation event that occurred from early 2005 to early 2007. As observed for the previous awakening episodes, the 2005-2006 ground uplift has been accompanied by moderate seismicity, consisting of weak (M<2.5) earthquakes which, based on their spectral features, had been classified as volcano-tectonic events. However, a peculiar aspect of this latter uplift is evident in the sustained occurrence of long-period events, probably associated with destabilisation of hydrothermal fluids at shallow depths (Saccorotti et al., 2007). The dataset used for the test includes the following seismic event classes:

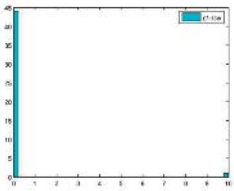
- Class 1 – noise (**ns**): 45 events 60 seconds long;
- Class 2 – volcano tectonic (**vt**): 53 events 20 seconds long;
- Class 3 – long period (**lp**): 193 events 40 seconds long.

Using this dataset we compute the *pdfs* that characterise the events in each class.

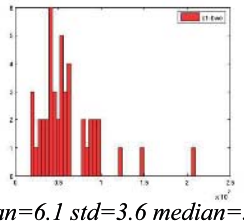
Class 1 - noise (***bpw, lsw, bwe***)



mean=0.34 std=0.16 median=0

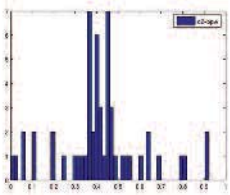


mean=0.22 std=1.49 median=0

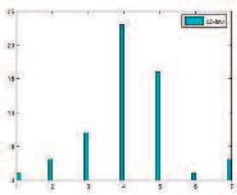


*mean=6.1 std=3.6 median=5.4*10^6*

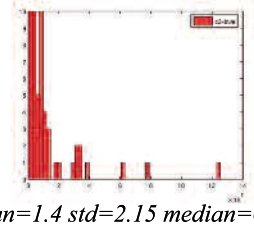
Class 2 - volcano tectonic (***bpw, lsw, bwe***)



mean=0.4 std=0.2 median=0.4

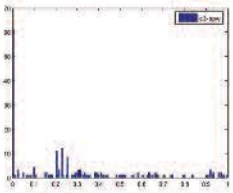


mean=4.2 std=1.17 median=4

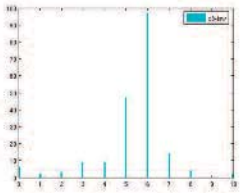


*mean=1.4 std=2.15 median=0.8*10^7*

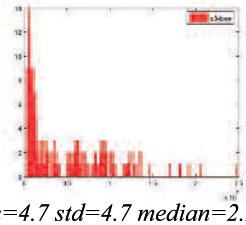
Class 3 - long period (***bpw, lsw, bwe***)



mean=0.26 std=0.28 median=0.21



mean=5.38 std=1.56 median=6



*mean=4.7 std=4.7 median=2.25*10^6*

Figure 1. Features histograms

Using a combination of features described above we obtain the following confusion matrix of results (Table 2).

	* ns	* vt	* lp
* | ns	100%	0%	0%
* | vt	7%	89%	4%
* | lp	9%	9%	82%

Table 2. Confusion matrix of the results

The confusion matrix indicates how many test data of the different classes of **noise** (*ns*), **volcano tectonic** (*vt*) and **long period** (*lp*) events have been classified correctly (main diagonal) and how have been misclassified (the entries outside the main diagonal). As indicated, noise is correctly detected in all cases, while accuracies of 89% and 82% are obtained for volcano tectonic and long period events, respectively.

Conclusions and Future Work

The DWT and Advanced Denoising Techniques can be used to extract features that characterize a class of seismic events. This characterization, together with a Bayesian pattern recognition method, allows an automatic seismic events classifier to be constructed. This algorithm successfully classifies an unknown seismic event starting from an arbitrary number of event classes. The future plan is to robustly test the classifier to develop an operational system for automatic signal classification to be used on volcanic monitoring integrated systems.

Acknowledgements

Mario La Rocca, Danilo Galluzzo and the Seismological staff at INGV-Napoli are kindly acknowledged for having made the Campi Flegrei data available.

References

Anant, A., K., Dowla, F., U., 1997. *Wavelet Transform Methods for Phase Identification in Three-Componenet Seismograms.*

Anant, A., K., Duin, R., P., W., Mao, J., 2000. Statistical Pattern Recognition: a Review. IEEE Transaction on Pattern Analysis and Machine Intelligence.

Benitez, M., C., Ramirez, J., Segua, J., C., Ibanez, J., M., Almendros, J., Garcia-Yeguas, A., Cortes, G., 2007. Continuous HMM-Based Seismic-Event Classification at Deception Island, Antarctica. IEEE Transaction on Geoscience and Remote Sensing, vol.45 no.1.

Del Pezzo, E., Esposito, A., Giudicepietro, F., Marinaro, M., Martini, M., Scarpetta, S., 2003 Discrimination of Earthquakes and Underwater Explosions Using Neural Networks. Bullettin of the Seismological Society of America, vol.93, no.1, pp. 215-223.

Di Vito, M., Isaia, R., Orsi, G., Southon, J., D'antonio, M., De Vita, S., Pappalardo, L. & Piochi, M. 1999. Volcanic and deformation history of the Campi Flegrei caldera in the past 12 ka. Journal of Volcanology and Geothermal Research, 91, 221–246.

Donoho, 1995 D. De-noising by Soft-thresholding. IEEE Trans. On Inf. Theory, vol.41, 3, pp 613-627.

Donoho, D., Johnstone, J., 1994. Ideal spatial adaptation via wavelet shrinkage. Biometrica, 81:425-455.

Gendron, P, Ebel, J., Manolakis, D., 1998. *Automatic Seismic Event Detection, Characterization and Classification: A Probabilistic Approach.*

Gendron, P, Ebel, J., Manolakis, D., 2000. Rapid Joint Detection and Classification with Wavelet Bases vis Bayesian Theorem. Bulletin of the Seismological Society of America, 90, 3, pp. 764-774.

Jain, A.K., Duin, R.P.W., Mao, J.,2000. Statistical Pattern Recognition: A Review. IEEE Transactions of Pattern Analysis and Machine Intelligence, vol.22,1.

Makhoul, J., 1975. Linear Prediction: a tutorial review. Proc. IEEE 63, 561-580.

Mallat,S., 2001. A Wavelet Tour of Signal Processing. Academic Press second edition.

Ooninccx, P., J., 1998. *Automatic Phase Detection in Seismic Data Using the Discrete Wavelet Transform.*

Rabiner, L., Juang, B., 1993. Fundamentals of Speech Recognition. Englewood Cliffs, N.J.: Prentice-Hall

Saccorotti, G., Petrosino, S., Bianco, F., Castellano, M., Galluzzo, D., La Rocca, M., Del Pezzo, E., Maccarelli, L., Cusano, P., 2007. Seismicity associated with the 2004–2006 renewed ground uplift at Campi Flegrei

Caldera, Italy. Physics of the Earth and Planetary Interiors.
Scarpetta,S., Giudicepietro, F., Ezin, E.C., Petrosino, S., Del Pezzo, E., Martini, M., Marinaro, M., 2005. Automatic classification of seismic signals at Mt. Vesuvius volcano, Italy, using neural networks. Bulletin of the Seismological Society of America, 95, 1, pp. 185-196.
Shimshoni, Y., 1995. *Introduction to Classification Methods*.
Wornell, G., 1996. Signal Processing with Fractals: a Wavelet-based Approach. Prentice Hall PTR.