Improving the Classification of Volcanic Seismic Events Extracting New Seismic and Speech Features

Millaray Curilem[1(✉)], Camilo Soto[1], Fernando Huenupan[1], Cesar San Martin[1], Gustavo Fuentealba[1], Carlos Cardona[2], and Luis Franco[2]

[1] Facultad de Ingeniería y Ciencias, Universidad de La Frontera, Francisco Salazar, 01145 Temuco, Chile
{millaray.curilem, fernando.huenupan, cesar.sanmartin, gustavo.fuentealba}@ufrontera.cl
[2] Observatorio Vulcanológico de los Andes Sur, Rudecindo Ortega, 03850 Temuco, Chile
{carlos.cardona, luis.franco}@sernageomin.cl

Abstract. This paper presents a study on features extracted from the seismic and speech domains that were used to classify four groups of seismic events of the Llaima volcano, located in the Araucanía Region of Chile. 63 features were extracted from 769 events that were labeled and segmented by experts. A feature selection process based on a genetic algorithm was implemented to select the best descriptors for the classifying structure formed by one SVM for each class. The process identified a few features for each class, and a performance that overcame the results of previous similar works, reaching over that 95% of exactitude and showing the importance of the feature selection process to improve classification. These are the newest results obtained from a technology transfer project in which advanced signal processing tools are being applied, in collaboration with the Southern Andes Volcano Observatory (OVDAS), to develop a support system for the monitoring of the Llaima volcano.

Keywords: Volcanic seismicity · Pattern recognition · SVM
Feature selection

1 Volcanic Seismology and Pattern Recognition

Because of its geographical location in the pacific ring of fire, Chile has an intense volcanic activity, throughout its territory. The "Observatorio Vulcanológico de los Andes Sur" (OVDAS) is the state agency responsible for establishing systems to continuously monitor and record forty three of the most active Chilean volcanoes. This monitoring is mainly of seismological type, as seismicity is highly related to volcanic activity [1]. The most common seismic events are Tremor (TR) and Long period (LP) events, which are related to the sustained and transitory fluid flow through the volcanic conduits, respectively. Volcano Tectonic (VT) events are related to rock failure inside the volcanic structure [2]. In recent years, automatic analysis of these

M. Mendoza and S. Velastín (Eds.): CIARP 2017, LNCS 10657, pp. 177–185, 2018.
https://doi.org/10.1007/978-3-319-75193-1_22

events has grown rapidly. This task is not trivial because each volcano has a particular behavior. In many studies, the problem is addressed in two stages: extraction of signal parameters and classification.

Several approaches have been used to address the classification problem of volcanic seismicity, including Support Vector Machines [3], Self Organizing Maps [4], Hidden Markov Models (HMM) [5], Multi-Layer Perceptron (MLP) [6], Gaussian Mixture Models (GMM) [7], among others. These studies show the diversity of techniques for implementing the classification stage.

Although these works achieved good results, it is necessary to improve them to reach acceptable performances in the real time and online applications. One approach is to improve the quality of the events' descriptors (features). In this sense it is important to underline the specificities of each volcano. The feature extraction process defines which information is extracted from the signals to facilitate their discrimination. Literature on volcanoes presents many features: Linear predictor coding and a signal parameterization to extract information from the waveform can be found in [8]. In [9], autocorrelation functions, obtained using Fast Fourier Transform, represent the spectral contents, and the sum and ratios of the amplitude were used. In other works, wavelet transform gives the possibility to use multiband analysis of the frequencies [3, 10]. Cepstral coefficients were extracted in [11]. The phase information of the events was included to the analysis in [12] and the amplitude statistics (mean, standard deviation, skewness and kurtosis), together with the power of the events were studied in [13].

In addition, many methods have been applied in the literature to perform feature selection in volcanoes [7], like the discriminant method [14], principal component analysis [15] and genetic algorithms [6].

The main motivation of this study is to improve the discrimination capacity of the classifiers by improving the quality of the features. This is why a high number of features were extracted from the events and a selection process, based on a genetic algorithm was performed to evaluate which is the best feature combination for each class.

2 Materials and Methods

Figure 1 depicts the block diagram of the proposed method. The signal database contains events already classified and segmented by experts. The automatic classification stage performs the discrimination between events, according to a subset of

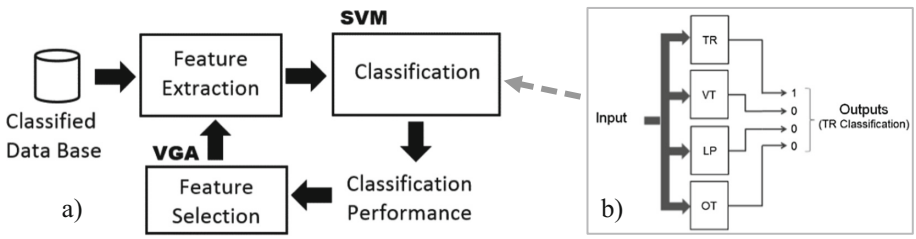

Fig. 1. (a) General working structure of the proposed method and (b) "one versus all" structure of the classification step.

features extracted from the data. The optimal subset of features is found using the validation error of the classifier as the unique fitness criterion. This error is used by the genetic algorithm to select new features, searching for the best subset.

2.1 Classified Database

The Llaima volcano is located in the Araucanía Region (38° 41'S–71° 44'W). The signals were obtained from the LAVE station, located 6 km from the crater. Only the Z component was considered. The records belong to the period 2010 to 2013. The events considered were of type LP, TR and VT. A contrast group was created, named "other" (OT) that contained events that did not belong to the first groups (like tectonic earthquakes, noise, ice breaking, etc.). The database had 769 events, 296 LP, 173 TR, 134 VT and 166 OT. The events were stored in files with variable length, according to their duration. The features were extracted from these files, thus, in this work the features were calculated from variable length windows that covered the entire events.

2.2 Feature Extraction

Most of the extracted features were obtained from our research on volcanic seismicity references, however, some of them belong to other processing domains, like speech. The features marked in gray in Table 1 were introduced as new features for this study.

All the extracted features were linearly normalized between [− 1, 1] and stored in a matrix of 769 rows (all the events) and 63 columns, each one is a specific feature, according to the order presented in Table 1. Each feature is explained next.

Energy: measures the size of the event and is calculated as the sum of the square of the samples [3].

LTA/STA: It is the ration of a long time window average (LTA) and a short time window average (STA) of the event [9]. This ratio is calculated along the signal. When it exceeds a threshold, it may define the beginning of a seismic event. Here it was calculated at the beginning of the event.

Table 1. Overview of the features extracted from each volcano's event

1	Energy	**25**	WAV Entropy
2	LTA/STA	**26-30**	LPC1-5
3-9	STA WAV1-7	**31**	Maximum Amplitude
10-16	WAV1-7 Energy	**32**	Median Amplitude
17	Variance	**33**	Mean Amplitude
18	Skewness	**34-46**	DCT1-13
19	Kurtosis	**47-59**	DCTLOG1-13
20	WAV4 Variance	**60**	Duration
21	WAV4 Skewness	**61**	Mean 5 Picks
22	WAV4 Kurtosis	**62**	Skewness Envelope
23	Pitch	**63**	Kurtosis Envelope
24	1st. Circular Moment		

STA WAV1-7: The STA indexes, calculated at the beginning of the event, were also calculated in the spectrum of 7 frequency bands, thus obtaining 7 features: STA WAV1 to STA WAV7. The considered frequency bands are presented in Table 2.

WAV1-7 Energy: This is the relative energy per wavelet band [3]. It is obtained decomposing the event in the 7 bands of Table 2, using a Daubechies mother wavelet type five. The percentage of energy is calculated as the ratio of the energy of the band over the energy of the event.

Table 2. Frequency bands for the feature extractions.

Band	1	2	3	4	5	6	7
Frequency (Hz)	25–50	12,5–25	6,25–12,5	3,13–6,25	1,56–3,13	0,78–1,56	0,39–0,78

Variance, Skewness and Kurtosis: The variance, the skewness and the kurtosis are extracted from the events in the time domain [6]. The variance is the expected value of the squared deviation from the mean of the event. The skewness is a statistical parameter that measures the asymmetry of the distribution of the seismic event. The kurtosis is a statistical parameter that measures the sharpness of the distribution of the seismic event. These features measure the shape of the event.

WAV4 Variance, WAV4 Skewness and WAV4 Kurtosis: These features are calculated for the 5^{th} frequency band (from 1.56 to 3.13 Hz).

Pitch: This feature is mainly used in speech processing but it has been applied here to the events. It is related to the lowest frequency at which an object vibrates.

First Circular Moment: reflects the behavior of the phase part of the volcanic signals [12]. The phase is obtained using the Hilbert transform in the range of [0 2π). This feature is mainly used in biomedical signal processing.

Entropy: is a measure of the uncertainty of a distribution. It is also a measure of the "quantity of information" contained in a signal.

LPC 1-5: Linear predictive coding (LPC) is a tool used mostly in audio and speech signal processing, for representing the spectral envelope of a signal in a compressed form [8]. It uses a model based on the fact that a signal can be modeled as the linear combination of N previous samples multiplied by coefficients plus a prediction error. This work considered only 5 of all the possible LPC coefficients.

Maximum Amplitude, Mean and Median: The maximum amplitude is the maximum value of the segmented event [3, 6]. The mean of the event is obtained from its absolute value. The median calculates the median of the samples of the event. All are extracted in the time domain.

DCT1-13 and LOGDCT1-13: The discrete cosine transform (DCT) is mainly used in audio and image processing. Like the discrete Fourier transform (DFT), the DCT expresses a signal as the sum of sinusoidal signals, with different frequencies and amplitudes. However, unlike DFT that works with complex exponentials, the DCT

only works with cosines. The DCT has a good ability to compact the energy to the transformed domain, that is, it concentrates most of the information in few coefficients. 13 DCT and 13 LOG-DCT were extracted from the events.

Duration: The duration is extracted from the segmented event in the time domain, counting the number of samples from the start to the end of the event. The automatic extraction of this feature is not trivial, as it requires a start and end detection system. This system is being developed in a parallel work, so for the present study, this characteristic was manually defined by an OVDAS analyst.

Mean of the 5 Frequency Peaks: The events are transformed using the FFT [3, 6]. The 5 highest peaks are detected and their mean is calculated to obtain this feature.

Skewness Envelope and Kurtosis Envelope: The envelope of a signal is related to its external form, that is, the slow variations of the amplitude in the time domain. The envelope is obtained using the Hilbert transform of the signal and calculating its absolute value. In this work, the kurtosis and the skewness of the envelope were used as shape features.

2.3 Classification

The classification step was performed using support vector machines (SVM). SVM tackles classification problems by nonlinearly mapping input data into high-dimensional feature spaces, wherein a linear decision hyperplane separates two classes [16]. To do so, SVM transforms the input space where the data is not linearly separable into a higher-dimensional space called a feature space through functions called kernels. One of the most generally used is the RBF Kernel that was applied here.

To train SVM it is necessary to adjust two hyperparameters: σ defines the width of the Gaussian used in the RBF and c determines the trade-off between the complexity of the model and the error tolerated.

Since SVM is a two-class classifier, a "one vs. all" structure was implemented using four classifiers (one per class). The training process of the classifiers applied a 2-fold cross validation strategy (bilateral cross-validation) due to the high computational cost of the feature selection strategy. The validation error of each trained classifier is obtained as the mean of the validation error of the 2-folds.

To evaluate the performance of each classifier, a contingency table was built. Four statistical indices were calculated for each classifier: sensitivity (Se), specificity (Sp), exactitude (Ex) and error (Er), obtained from the following Eqs. 1 to 4.

$$Se = \frac{TP}{TP + FN}; \tag{1}$$

$$Sp = \frac{TN}{TN + FP}; \tag{2}$$

$$Ex = \frac{TP + TN}{n}; \tag{3}$$

$$Er = \frac{FP + FN}{n} \tag{4}$$

where TP (true positives) is the number of events correctly classified for each class; TN (true negatives) is the number of events correctly classified as negatives, for each class; FP (false positives) and FN (false negatives) is the number of events classified erroneously and n is the total number of events.

2.4 Feature Selection

A genetic algorithm (GA) [17] performed the search of the best feature subset. The chromosome of each individual was defined as a string of 63 bits, where a "1" represented the presence of the corresponding feature, according to the number of Table 1, and the "0" represented its absence. Each generation had 64 individuals and the maximum number of generations was 200. The first 63 individuals in the initial population had one of the 63 features (diagonal matrix). The last individual (number 64) had all the features set to "1". The percentage of elitism was set to 20% and mutation to 10%.

The Vasconcelos GA [18] was used, since in previous works it reached better solutions than the traditional GA approach [6]. The main characteristics of this algorithm is that it maintains a highly diverse population along the generations, because the selection and crossover operators combine the genetic information of good and bad solutions. Good individuals survive from one generation to another, through elitism.

The performance of each individual was measured by the validation error of the classifier, trained with the features retrieved by its chromosome. In addition to the validation error, chromosomes with a high number of features were penalized. Therefore, the performance of each individual is given by Eq. 5:

$$\boldsymbol{Performance = Validation\ Error + \frac{Number\ of\ Features}{63} * 25\%} \tag{5}$$

3 Results

The simulations were carried out with Matlab of Mathworks, in an Intel Core™ i5 CPU with 3 GHz and 8 GB RAM. The results of the best classifiers, after the 200 generations, are presented in Table 3. This table presents the contingency table and the performance of the best classifiers for each class. It also shows the values of the hyper parameters, for each classifier and the features selected for the best models.

Table 3. Results of the best classifiers after 200 generations.

LAVE Station	LP	OT	TR	VT
LP	274	1	0	9
OT	0	153	0	0
TR	0	0	170	0
VT	0	0	0	113
Not Assigned	22	12	3	12
TOTAL	296	166	173	134
Sensitivity	92,57%	92,17%	98,27%	84,33%
Specificity	97,89%	100,00%	100,00%	100,00%
Error	**4,16%**	**1,69%**	**0,39%**	**2,73%**
Exactitude	95,84%	98,31%	99,61%	97,27%
Sigma	2^{2}	2^{-2}	2^{-6}	2^{2}
C	2^{9}	2^{1}	2^{1}	2^{5}
Best Feature Set	**WAV4 Energy, WAV6 Energy, Duration, Mean Amplitude.**	**Duration DCT1.**	**Energy WAV5 Energy**	**WAV6 Energy WAV7 Energy Kurtosis WAV5 Variance WAV5 Skewness**

The "Not Assigned" events correspond to events that were classified positive by more than one classifier or negative by all the classifiers. Thus, the results presented in Table 3 may improve if a confidence step is added to the classifiers. This confidence step has to be able to give an output, even when the classifiers do not agree.

4 Discussion and Conclusions

In a previous work [3], the authors demonstrated that the features selected for the classification of the seismic events of the Villarrica volcano [6], also performed well for the Llaima volcano. However, the global performance decreased from 90 to 80%. This paper presents a study on new characteristics extracted from the seismic and speech processing areas. This study evaluates their impact in the discrimination of four groups of events from the Llaima volcano: the LP, TR, VT and a contrast group, OT.

A feature selection was necessary to define which of the 63 extracted features led to a better performance of the classifiers. It is essential that the set of features that represent the signals is appropriate to the classification problem, so that it divides the space into decision regions associated to the classes to be distinguished. In addition, the number of features has to be reduced to simplify the complexity of the classifier. This is why a feature selection process is needed. Different methods applied to the feature selection process demonstrate the complexity of this task. It is also important to observe that features that are not individually relevant may become so when used in combination with others, and features that are individually relevant may not be useful due to possible redundancies when combined with other. This is why we applied a strategy that evaluates subsets of features. The classification performance was the objective function and the GA performed the search of the best subset. The resulting classifiers reached performances superior to 95%, improving the results of previous works.

LP and VT are more difficult to classify than TR, which needed only two features to reach a high sensitivity. The energy of the low frequencies, from 0.78 to 6.25 Hz are the most important discriminators for LP and VT, while the energy of the 1.56 to 3.13 Hz is the best discriminator for the TR. It is interesting to note the importance of the duration feature, as a good descriptor for LP and OT events. It is also interesting to see that DCT, a feature from the speech domain, was considered important to discriminate the OT group.

The results obtained in this work were very promising, as the addition of new features improved the classification performance. Future works need to go-on in the study of new features and other selection methods. The performances reached in the off-line experiments were considered satisfactory by the OVDAS experts.

Acknowledgements. We would like to thank FONDEF as this study is being supported by the project FONDEF IDeA IT15I10027.

References

1. Chouet, B.: A seismic model for the source of long-period events and harmonic tremor. In: Gasparini, P., Scarpa, R., Aki, K. (eds.) Volcanic Seismology, pp. 133–156. Springer, Heidelberg/New York (1992). https://doi.org/10.1007/978-3-642-77008-1_11
2. Lahr, J.C., Chouet, B.A., Stephens, C.D., Power, J.A., Page, R.A.: Earthquake classification, location and error analysis in a volcanic environment: implications for the magmatic system of the 1989–1990 eruptions at Redoubt Volcano, Alaska. J. Volcanol. Geoth. Res. **62**(1–4), 137–151 (1994)
3. Curilem, M., Vergara, J., San Martin, C., Fuentealba, G., Cardona, C., Huenupan, F., Chacón, M., Khan, S., Hussein, W., Becerra, N.: Pattern recognition applied to seismic signals of the Llaima volcano (Chile): an analysis of the events' features. J. Volcanol. Geoth. Res. **282**, 134–177 (2014)
4. Langer, H., Falsaperla, S., Messina, A., Spampinato, S., Behncke, B.: Detecting imminent eruptive activity at Mt Etna, Italy, in 2007–2008 through pattern classification of volcanic tremor data. J. Volcanol. Geoth. Res. **200**(1–2), 1–17 (2011)
5. Bhatti, S.M., Khan, M.S., Wuth, J., Huenupan, F., Curilem, M., Franco, L., Becerra-Yoma, N.: Automatic detection of volcano-seismic events by modeling state and event duration in hidden Markov models. J. Volcanol. Geoth. Res. **324**, 134–143 (2016)
6. Curilem, G., Vergara, J., Fuentealba, G., Acuña, G., Chacón, M.: Classification of seismic signals at Villarrica volcano (Chile) using neural networks and genetic algorithms. J. Volcanol. Geoth. Res. **180**, 1–8 (2009)
7. Cortés, G., García, L., Álvarez, I., Benítez, C., de la Torre, T., Ibáñez, J.: Parallel System Architecture (PSA): an efficient approach for automatic recognition of volcano-seismic events. J. Volcanol. Geoth. Res. **271**, 1–10 (2014)
8. Scarpetta, S., Giudicepietro, F., Ezin, E.C., Petrosino, S., Del Pezzo, E., Martíni, M., Marinaro, M.: Automatic classification of seismic signals at Mt Vesuvius Volcano, Italy, using neural networks. Bull. Seismol. Soc. Am. **95**(1), 185–196 (2005)
9. Langer, H., Falsaperla, S., Powell, T., Thompson, G.: Automatic classification and a-posteriori analysis of seismic event identification at Soufriere Hills volcano, Montserrat. J. Volcanol. Geoth. Res. **153**, 1–10 (2006)

10. Erlebacher, G., Yuen, D.A.: A wavelet toolkit for visualization and analysis of large data sets in earthquake research. Pure. Appl. Geophys. **161**, 2215–2229 (2004)
11. Ibáñez, J., Benítez, C., Gutiérrez, L., Cortés, G., García-Yeguas, A., Alguacil, G.: The classification of seismo-volcanic signals using Hidden Markov Models as applied to the Stromboli and Etna volcanoes. J. Volcanol. Geoth. Res. **187**, 218–226 (2009)
12. San-Martin, C., Melgarejo, C., Gallegos, C., Soto, G., Curilem, M., Fuentealba, G.: Feature extraction using circular statistics applied to volcano monitoring. In: Bloch, I., Cesar, Roberto M. (eds.) CIARP 2010. LNCS, vol. 6419, pp. 458–466. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-16687-7_61
13. Joevivek, V., Chandrasekar, N., Srinivas, Y.: Improving seismic monitoring system for small to intermediate earthquake detection. Int. J. Comput. Sci. Secur. **4**(3), 308–315 (2010)
14. Álvarez, I., García, L., Cortés, G., Benítez, C., de La Torre, A.: Discriminative feature selection for automatic classification of volcano-seismic signals. IEEE Geosci. Remote Sens. Lett. **9**(2), 151–155 (2012)
15. Unglert, K., Radić, V., Jellinek, A.M.: Principal component analysis vs. self-organizing maps combined with hierarchical clustering for pattern recognition in volcano seismic spectra. J. Volcanol. Geoth. Res. **320**, 58–74 (2016)
16. Vapnik, V.: The Nature of Statistical Learning Theory. Springer, New York (1995). https://doi.org/10.1007/978-1-4757-3264-1
17. Holland, J.H.: Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. The MIT Press, Cambridge (1992). 228 p.
18. Kuri-Morales, A.: Pattern recognition via Vasconcelos' Genetic Algorithm. In: Sanfeliu, A., Martínez Trinidad, J.F., Carrasco Ochoa, J.A. (eds.) CIARP 2004. LNCS, vol. 3287, pp. 328–335. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-30463-0_40