Classification of volcano events observed by multiple seismic stations

Robert P.W. Duin
Pattern Recognition Laboratory
Delft University of Technology, The Netherlands
Email: r.duin@ieee.org

Mauricio Orozco-Alzate
Departamento de Informática y Computación
Universidad Nacional de Colombia Sede Manizales
Email: morozcoa@bt.unal.edu.co

John Makario Londoño-Bonilla
Observatorio Vulcanológico y Sismológico de Manizales
INGEOMINAS, Colombia
Email: jmakario@ingeominas.gov.co

***Keywords*-volcano eruptions; combining classifier; spectrum; spectrogram;**

***Abstract*—Seismic events in and around volcanos, like tremors, earth quakes, ice quakes and strokes of lightning, are usually observed by multiple stations. The question rises whether classifiers trained for one seismic station can be used for classifying observations by other stations, and, moreover, whether a combination of station signals improves the classification performances for a single station. We study this for seismic time signals represented by spectra and spectrograms obtained from 5 seismic stations on the Nevado del Ruiz in Colombia.**

I. Introduction

Volcanic eruptions are among the most important natural disasters. They cause severe impacts on the ecology and economy, often threatening the survival of entire communities. Mitigation strategies, such as evacuations, can only be efficiently planned if geological processes as well as their changes and associated risks are understood. Such an understanding implies a permanent monitoring of geophysical and meteorological phenomena. Seismicity patterns, in particular, are a first sign of renewed volcanic activity and reveal processes such as transport of magma and gases or fracture of solid rock.

Seismic stations on a volcano surface detect many types of events, like earthquakes, icequakes, long term tremors, landslides, lightning strikes, passing cars, etcetera. Detection and classification of events is the first step in an analysis, often done manually in remote observatories that read the station signals by radiographic means.

Several stations can be positioned on a volcano that may observe the same events. Event types that cannot be detected or classified by one station may be better observed by another station. It is to be expected that combining information from several stations improves the accuracy of the analysis significantly.

As seismic stations are already in use for many years, and events are annotated manually, large datasets are available to study the possibility of automatic detection or classification. This has been undertaken in a number of previous research studies. Applied methods range from neural networks [1], [2], hidden Markov models [3], [4], [5], dissimilarity-based classifiers [6] and Bayesian networks [7]. Most of them consider representations in the frequency domain and classifiers built for single or separated seismic stations. To the best of the knowledge of the authors, none combining techniques have been used to profit from the availability of the multiple sources of information, i.e. different recording stations.

In this paper a study is presented on a part of a large seismic dataset that analyzes the above issues: do different stations observe events in the same way or not? If they do, classifiers trained by data from one station may be applied by other stations as well. If stations really observe different information, their classifications may be combined into more robust or accurate ones. We will do this for two representations: spectral features and spectrogram features. In the first, just spectral information is stored. In the second some time information is preserved as well. Their representations are discussed in section 2. Data and experiments follow in section 3.

In is not our intention to focus on the best possible classifier, but just to answer the question on the informative difference of different seismic stations on the same volcano. Results are discussed in section 4. Although we come to a clear conclusion, additional issues for further research in order to design optimal classification systems will be discussed there too.

II. Representation and classification of seismic signals

Volcano seismic signals are measured continuously, 24 hours per day. In these time signals events can be defined caused by particular environmental disturbances like earthquakes and icequakes. Short periods of maximum energy can be detected and defined as the center of the events. Appropriate time intervals before and after the central moment are considered to belong to the same event. The size of this time-window is determined by signal energy properties as

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.263

well. This is all set by an automatic event detector that is independent of the event class as recognition has still to be done. For some classes it might be appropriate, for others it might be too large or too small. This is part of future research once we have an initial event classification system.

The above-mentioned automatic detector is commonly based on the short-term average to long-term average ratio ($STA - LTA$), which is a classical algorithm used as a standard in seismic detection/segmentation [8]. It consists in a low-pass filter that averages two windows (short-term and long-term) with the last captured sample. Such an average is an estimation of the power density. When a seismic event occurs, $STA > LTA$. Averages are recursively computed as follows:

$$STA = STA + \frac{(abs(s_i - 2048) - STA)}{k_1} \quad (1)$$

$$LTA = LTA + \frac{(abs(s_i - 2048) - LTA)}{k_2}, \quad (2)$$

where s_i is the current sample, 2048 corresponds to the offset for a 12-bit analog-to-digital converter and k_1, k_2 are the sizes of the STA and LTA windows. These sizes are typically set to ranges between 0.5 – 3 s and 30–60 s, respectively. Two additional parameters must be tuned, namely thresholds for the beginning (EnThrHld) and the end (DisThrHld) of the event. See [9], [8] fur further details.

A simple way to characterize events given by a determined window of the time signal is by power spectra. Their sizes are determined by the window size. If these differ over events, alignment becomes of importance. For simplicity we selected events of the same window size. The spectra can thereby be represented by vectors of the same length: $X = \{x_1, x_2, \ldots, x_n\}'$ in which x_i is a column vector of length p representing a single spectrum and X is $m \times p$ matrix storing the spectra for all observed m events.

In the just defined spectra all time information is lost. Events however are not ergodic. The properties within event windows may change. One way to represent this is by spectrograms. These are sets of spectra defined on overlapping small windows. A single spectrogram S is thereby not a vector but has a 2-dimensional structure, e.g. $q \times n$ in case there are n windows in an event and window spectra (which we call for convenience subspectra) have a length q. Note that $p = n \times q$ for windows that don't overlap and spectra that are not reduced.

As seismic events of the same class are not perfectly aligned (e.g. two earthquake events may show different substructures) also their corresponding spectrograms are not aligned: similar subspectra may occur at different positions along the time axis. A spectrogram cannot be directly used for building classifiers. It has first to be converted into a vector. Just unfolding may not be good due to the alignment problem.

In order to deal with the variability of the spectral behavior in the spectrograms we decided to build a subspectra classifier, trained on all subspectra of all spectrograms considered in a training set. A spectrogram is classified by combining all subspectra classifications. We used the max-combiner, thereby following the classification of the subspectrum that was most confidently classified. Consequently a spectrogram is mapped on a vector with confidences for all classes.

There are various other possibilities for representing events. A few under study by us are Hidden Markov Models, Conditional Random Fields, 2D shape analysis of spectrograms and shape analysis of the trajectories followed by the subspectra as a function of time (a variation of Dynamic Time Warping). A crucial point thereby is whether a seismic volcano event can be considered as a word, a sentence or a gesture, or whether it is just a type of randomized collection of subevents. The way we deal above with the combination of subspectra classifications is following the latter interpretations.

III. Data and experimental setup

In total about 150,000 events have been collected between 1994 and 2008 for the five volcanos observed by the Volcanological and Seismological Observatory in Manizales, Colombia. Events are detected and aligned for the about 40 seismic stations that are considered. Consequently, in every event a set of seismic stations is involved, usually 1 to 10. In total about 10 different event classes are considered, some of them directly related to volcanic processes, others generated by tectonic, glaciological and environmental causes.

For the study reported in this paper we decided to simplify the dataset by taking into account just 3 classes detected by the same 5 stations and that are all given by the same event window size of 12032 time samples (120s). Moreover, we restricted ourselves to the years 2002-2006. The 3 classes are Earthquake, Icequake, and Long Term Tremor. We constructed equal class sizes of 700 events per class, neglecting the different class prior probabilities. The seismic stations for which all events are given are 'ALF', 'BIS', 'IRI', 'LIS' and 'OLL'. The dataset had thereby a size of $3 \times 700 \times 5$ event signals of 12032 time samples. Every event is for every station represented by a spectrum of 6016 points and by a spectrogram of 128 (spectrum) by 93 (time). In all experiments we split the data randomly in 50% for training and 50% for testing. This was repeated 50 times and error estimates were averaged.

For every station separately classifiers are computed using the spectra and spectrograms given for that station. For the spectra we reduced by PCA the spectral feature space from 6016 to 40. In this space a quadratic classifier was computed, assuming normal densities with different covariances per class.

Table I
CLASSIFICATION ERRORS FOR THE SPECTRAL REPRESENTATION.

	ALF classf	BIS classf	IRI classf	LIS classf	OLL classf	combined
ALF signal	**0.398**(0.001)	0.578(0.002)	0.667(0.000)	0.557(0.002)	0.609(0.003)	0.403(0.001)
BIS signal	0.438(0.002)	**0.271**(0.001)	0.668(0.001)	0.557(0.003)	0.511(0.002)	0.281(0.002)
IRI signal	0.673(0.001)	0.664(0.001)	**0.626**(0.003)	0.667(0.000)	0.670(0.001)	0.626(0.002)
LIS signal	0.642(0.003)	0.679(0.002)	0.669(0.002)	**0.359**(0.002)	0.577(0.004)	0.373(0.002)
OLL signal	0.591(0.001)	0.572(0.001)	0.667(0.000)	0.553(0.002)	**0.362**(0.001)	0.366(0.002)
combined	0.358(0.002)	0.260(0.001)	0.622(0.004)	0.332(0.002)	0.332(0.002)	**0.248**(0.002)

Table II
CLASSIFICATION ERRORS FOR THE SPECTROGRAM REPRESENTATION.

	ALF classf	BIS classf	IRI classf	LIS classf	OLL classf	combined
ALF signal	**0.406**(0.002)	0.544(0.002)	0.662(0.003)	0.522(0.001)	0.609(0.002)	0.417(0.002)
BIS signal	0.509(0.002)	**0.287**(0.002)	0.675(0.004)	0.586(0.003)	0.495(0.003)	0.292(0.002)
IRI signal	0.671(0.001)	0.676(0.001)	**0.627**(0.002)	0.668(0.001)	0.671(0.001)	0.630(0.002)
LIS signal	0.537(0.001)	0.608(0.002)	0.654(0.004)	**0.376**(0.002)	0.547(0.002)	0.375(0.002)
OLL signal	0.556(0.001)	0.469(0.001)	0.655(0.003)	0.640(0.002)	**0.386**(0.002)	0.396(0.002)
combined	0.340(0.002)	0.269(0.002)	0.601(0.005)	0.258(0.002)	0.344(0.002)	**0.238**(0.002)

For the subspectra of the spectrograms we computed a similar quadratic classifier in the 128 dimensional spectral feature space. The 93 classified spectral vectors were combined by the max-combiner obtaining a unique classification for each spectrogram.

The five classifiers we obtained in this way (one for every seismic station) for each of the two procedures are used in several ways. First we checked the results of cross-station classifications: the performances of signals obtained from one station classified by classifiers of other stations. This yields 5×5 tables of classifications errors, see the tables 1 and 2. Recall that this is a 3-class problem with equal priors, so random assignments result in an error of 0.666. Between brackets are the standard deviations of the of the estimated mean error rates over the 50 repetitions. In bold on the diagonal are the results of test events from the same station as used for training the classifier.

We performed three ways of combing the classification results. In the vertical direction of the table the five signals of one and the same event obtained from five different stations are combined by a trained combining rule: decision templates [10]. In the horizontal direction the classifiers instead of the signals are combined, also using decision templates. Finally, in the diagonal way the sets of five station classifiers are classified by a combined combiner, again decision templates, yielding the result on the bottom right.

IV. RESULTS

From the results we draw the following observations:

- Stations show significantly different performances (in bold on the diagonals). In fact one of them is scoring about randomly. This station (IRI) has a special meaning and is used for calibration. It appeared that after removing it from the study similar performances are reached for the trained combiners, proving that our system is robust.
- The off-diagonal errors are worse than the diagonal ones showing that classifiers for individual stations cannot be applied without loss of accuracy to other stations. This is to be expected. The errors however are also not equal to the random assignment errors, which also shows that stations observe the same event in about the same way.
- With an exception of the IRI station it holds for both tables that combining signals of different stations improves the classification result of individual station classifiers (vertical combining).
- With an exception of the IRI station it holds for both tables that combining classifiers of different stations does not improve the results of individual station classifiers (horizontal combining).
- The best results are found by applying combined station classifiers to the set of station signals.
- Spectrograms including time information improve the full signal spectra just slightly.

The use of 93 time steps in constructing the spectrograms may hamper our results. It removes lower frequencies from the spectra. With larger windows, and thereby broader spectrograms (at the cost of less time steps) better results might be obtained.

V. CONCLUSION AND DISCUSSION

We studied a large set of seismic events all observed by the same 5 seismic stations. It was shown that event classifiers for one station could be applied to observations by other stations as well. Although their individual classification results were significantly worse than for observations made by the station for which the classifiers were trained, combining all observations from different stations yields improved results. This shows that observations made by different stations are at the same time similar (the classifier performs better than random) as well as informatively different (combining improves). Training separate classifiers for each station and combining them improves results further, which confirms our conclusion.

Spectrograms, representing time as well as spectral information, perform slightly better than just spectral features. Future research may improve this further.

Acknowledgements. We thank the Volcanological and Seismological Observatory in Manizales, Colombia for supplying and annotation the data. We acknowledge support from Universidad Nacional de Colombia, under the project "Sistemas múltiples de clasificación para el reconocimiento automático de eventos sísmicos en el complejo volcánico Cerro Machín-Cerro Bravo (code Hermes-8970)".

References

[1] S. Scarpetta, F. Giudicepietro, E. C. Ezin, S. Petrosino, E. D. Pezzo, M. Martini, and M. Marinaro, "Automatic classification of seismic signals at Mt. Vesuvius volcano, Italy, using neural networks," *Bulletin of the Seismological Society of America*, vol. 95, no. 1, pp. 185–196, 2005.

[2] G. Curilem, J. Vergara, G. Fuentealba, G. Acuña, and M. Chacón, "Classification of seismic signals at villarrica volcano (chile) using neural networks and genetic algorithms," *Journal of Volcanology and Geothermal Research*, vol. 180, no. 1, pp. 1 – 8, 2009.

[3] L. Gutiérrez, J. Ramírez, C. Benítez, J. Ibañez, J. Almendros, and A. García-Yeguas, "HMM-based classification of seismic events recorded at Stromboli and Etna volcanoes," in *IEEE Int. Geoscience & Remote Sensing Symposium, IGARSS 2006*, 2006, pp. 2765–2768.

[4] M. C. Benítez, J. Ramírez, J. C. Segura, J. M. Ibañez, J. Almendros, A. García-Yeguas, and G. Cortés, "Continuous HMM-based seismic-event classification at Deception Island, Antarctica," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 45, no. 1, pp. 138–146, Jan. 2007.

[5] J. M. Ibañez, C. Benítez, L. A. Gutiérrez, G. Cortés, A. García-Yeguas, and G. Alguacil, "The classification of seismo-volcanic signals using Hidden Markov Models as applied to the Stromboli and Etna volcanoes," *Journal of Volcanology and Geothermal Research*, vol. 187, no. 3-4, pp. 218 – 226, 2009.

[6] M. Orozco-Alzate, M. E. García-Ocampo, R. P. W. Duin, and C. G. Castellanos-Domínguez, "Dissimilarity-based classification of seismic volcanic signals at Nevado del Ruiz volcano," *Earth Sciences Research Journal*, vol. 10, no. 2, pp. 57–65, Dec. 2006.

[7] C. Riggelsen, M. Ohrnberger, and F. Scherbaum, "Dynamic bayesian networks for real-time classification of seismic signals," in *PKDD 2007: Proceedings of the 11th European conference on Principles and Practice of Knowledge Discovery in Databases*, ser. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 565–572.

[8] C. A. Vargas-Jimenez and S. Rincón-Botero, "Portable digital seismological AC station over mobile telephone network and internet," *Computers & Geosciences*, vol. 29, pp. 685–694, 2003.

[9] M. Basseville and A. Benbeniste, *Detection of Abrupt Changes in Signals and Dynamical Systems*. Berlin: Springer, 1986.

[10] L. I. Kuncheva, J. C. Bezdek, and R. P. W. Duin, "Decision templates for multiple classifier fusion: An experimental comparison," *Pattern Recognition*, vol. 34, no. 2, pp. 299–314, Feb. 2001.