Automatic Classification of Volcanic Earthquakes in HMM-Induced Vector Spaces

Riccardo Avesani[1], Alessio Azzoni[1],
Manuele Bicego[1], and Mauricio Orozco-Alzate[2]

[1] Dipartimento di Informatica, Università degli Studi di Verona, Ca' Vignal 2, Strada Le Grazie 15, 37134, Verona, Italy
[2] Departamento de Informática y Computación, Universidad Nacional de Colombia - Sede Manizales, Kilómetro 7 Vía al Aeropuerto, Manizales (Caldas), Colombia

Abstract. Even though hidden Markov models (HMMs) have been used for the automatic classification of volcanic earthquakes, their usage has been so far limited to the Bayesian scheme. Recently proposed alternatives, proven in other application scenarios, consist in building HMM-induced vector spaces where discriminative classification techniques can be applied. In this paper, a simple vector space is induced by considering log-likelihoods of the HMMs (per-class) as dimensions. Experimental results show that the discriminative classification in such an induced space leads to better performances than those obtained with the standard Bayesian scheme.

Keywords: Automatic classification, generative embedding, hidden Markov models, model-induced feature space, seismic-volcanic signals.

1 Introduction

South America is geologically very active, having a relatively young and seismically restless mountain range —the Andes— which runs parallel to the west coast of the continent. Volcanism in the Andes results from subduction of the Nazca and Antarctic oceanic plates below South America and occurs in four separate regions named the Northern, Central, Southern and Austral volcanic zones [1]. The fist zone includes 19 volcanoes in Colombia and 55 in Ecuador, where among the largest and more hazardous ones is Nevado del Ruiz volcano (NRV), whose eruption in 1985 triggered deadly lahars that killed more than 23000 people. Since 1985 and in order to avoid further tragedies, NRV is permanently monitored by the Volcanological and Seismological Observatory at Manizales, Colombia (OVSM by its acronym in Spanish). Monitoring activities include the measurement and analysis of tilt, gas emissions, physicochemical properties of hot springs and seismic events. The latter are registered by several seismic stations that the OVSM has deployed in strategic places at NRV.

In spite of the availability of several studies on the application of pattern recognition techniques to the automated identification of seismic-volcanic events

(see [2] for a comprehensive inventory), the OVSM staff members still visually analyze and manually label every incoming seismic signal registered by the monitoring network. Class labels assigned at OVSM to the registered signals encompass volcano tectonic events, long period events, tremors, hybrid events, teleseismic events, regional and local earthquakes, rock falls, explosions, landslides, avalanches, icequakes and lightnings. The manual assignment of these labels is time and labor consuming as well as prone to human error. Such a tedious task can be significantly simplified by the design and deployment of an automatic classification system.

Several feature extraction[1] and classification approaches have been tested in previous studies [2]. Among the classification approaches, hidden Markov models (HMMs) have proved to be convenient classification tools for the sequential and unequal-length nature of the seismic signals [3–5] (conference papers and other older publications are not cited here due to space constraints, refer again to [2] for additional references). Even though these studies show the usefulness of such approaches, the full potentialities of HMM-based classification systems have not been completely exploited.

In this paper we propose an alternative approach, which explores the possibility of employing recent advancements in the HMM theory to the seismic classification problem. In particular, we use a generative embedding scheme, where the basic idea is to exploit the HMM to map the objects to be classified into a feature space (also called HMM-induced Vector Space), where discriminative techniques (e.g., kernel-based support vector machines) can be used. Actually, it has been shown in many different applications and scenarios [6–10] that the typical HMM classification scheme can be largely improved when enriched with such a discriminative step. We also compare the classical Bayes rule (representing the typical approach in the HMM-based seismic classification systems) with a simple generative embedding scheme [9], showing on a set of signals coming from the NRV the superiority of the latter approach.

The remaining part of the paper is organized as follows. The proposed seismic signals classification system, including the data preprocessing step and the HMM-based classification scheme, is presented in Sec. 2. A brief description of the classes in the data set and all the details about the algorithms application and their parameters are provided in Sec. 3. Experimental results are presented in Sec. 4. Finally, observations and concluding remarks are given in Sec. 5.

2   Methods

2.1   Data Preprocessing

Seismic signals recorded by OVSM instruments are acquired by using a 12-bit analog-to-digital converter that provides unsigned integers. As a result, an offset is introduced, which must be removed by subtracting the mean value before the feature extraction stage.

---

[1] More generally: representation approaches.

Second, seismic waveforms are transformed into the frequency domain. In particular we employed spectrograms for representation, mainly because spectral analysis, either in the frequency or the time-frequency domain, is widely used for both visual and computer-based [11] inspection of seismic phenomena. In particular, spectrograms are computed by using a fast Fourier transform (FFT) and a Hann (Hanning) window with a given overlap. Length of the FFT and the percentage of overlap where varied as indicated in Sec. 3.2. The magnitude in decibels is computed as $20\log_{10}|X|$, where X is a matrix containing the short-time Fourier transform of a signal x. In this way we have an observation sequence, able to be modelled with a HMM, in which every symbol is the FFT of the given window. In order to reduce the dimensionality of every symbol, we finally applied a Discrete Cosine Transform (DCT), retaining only the first four values.

2.2   Classification

Hidden Markov Models (HMMs). A discrete-time HMM $\boldsymbol{\lambda}$ is defined by the following entities [12]:

- a set $S = \{S_1, S_2, \cdots, S_N\}$ of (hidden) states;
- a transition matrix $\mathbf{A} = \{a_{ij}\}$, where $a_{ij} \geq 0$ represents the probability of going from state S_i to state S_j;
- an emission matrix $\mathbf{B} = \{b(o|S_j)\}$, indicating the probability of emission of symbol o from state S_j;
- an initial state probability distribution $\boldsymbol{\pi} = \{\pi_i\}$, representing the probability of the first state $\pi_i = P[Q_1 = S_i]$.

For a sequence $\mathbf{O}$ and a HMM $\boldsymbol{\lambda}$, there is a standard recursive procedure, called the *forward-backward* procedure [12], able to compute the probability $P(\mathbf{O}|\boldsymbol{\lambda})$. Given a set of observed sequences $\{\mathbf{O}_i\}$, the learning of the HMM parameters is usually performed using the well-known Baum-Welch algorithm [12], which is able to determine the parameters maximizing the log-likelihood: $\log P(\{\mathbf{O}_i\}|\boldsymbol{\lambda})$.

Classification with the Bayes Rule. The Bayes Decision Rule starts from the idea of determining a probability density function for each class, which can be used to find the probability that a given x element belongs to that class. The Bayes decision rule then dictates to assign that given element to the class with the higher probability (also called a posteriori probability). Now, to find those probabilities we chose a HMM approach, i.e. we represent each function with a HMM. In particular, in the training phase, a HMM $\boldsymbol{\lambda}_c$ is trained for every class c, using only the training sequences belonging to such class. At the end of the training process, we have a set of HMMs $\boldsymbol{\lambda}_1, ..., \boldsymbol{\lambda}_C$. In the testing phase, then, an unknown sequence $\boldsymbol{o} = (o_1, ..., o_T)$ is assigned to the class whose model shows the highest likelihood (assigning to each class the same prior probability), namely the label $\ell(\boldsymbol{o})$ is determined as: $\ell(\boldsymbol{o}) = \arg\max_c \log P(\boldsymbol{o}|\boldsymbol{\lambda}_c)$.

Classifying Using the Generative Embedding Theory. The Generative Embedding Theory is based on the idea that both performance and interpretability of conventional approaches could be improved by taking into account available prior knowledge about the process generating the observed data. The term generative embedding is sometimes used to denote a particular model-induced feature space, or so-called generative score space, in which case the associated line of research is said to be concerned with generative embeddings. Here, we will use the term in singular form to denote the process of using a generative model to project the data into a generative score space, rather than using the term to denote the space itself. Generative embedding rests on two components: a generative model for the feature selection and a discriminative method for classification.

In particular, we employed the set of trained HMMs to project every sequence $\boldsymbol{o}$ (both training and testing sequences) into a feature space, using the following mapping, firstly introduced in [9]:

$$\phi(\boldsymbol{o}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \cdots, \boldsymbol{\lambda}_C) = [\log P(\boldsymbol{o}|\boldsymbol{\lambda}_1) \log P(\boldsymbol{o}|\boldsymbol{\lambda}_2) \cdots \log P(\boldsymbol{o}|\boldsymbol{\lambda}_C)] \in \mathbb{R}^C$$

Here the idea is to project a sequence in a feature space where every direction represents the log-likelihood of the model. Even if very simple, such generative embedding has been shown to be very effective in different applications. In such HMM-induced vector space, then, every standard classifier can be used. In our experiments we tried many different techniques (see Sec. 3.2), showing promising results.

3   Experimental Setup

3.1   Data Set

The data set we have been provided is composed by a total of 1359 signals, distributed per classes as follows[2]: *volcano tectonic* events (VT, 276), *long period* events (LP, 212), *regional tectonic* events (RE, 601) and *local tectonic* events: (TL, 270). VT and LP classes are related to volcanic phenomena: fracture of solid rock in the volcano and transport of fluid, e.g. gases or magma, within it; respectively. The other two classes correspond to seismic events of tectonic origin, which are localized either far (RE) or near (TL) the volcano. Even though RE and TL events are of less interest to monitor the volcanic activity, they anyway have to be distinguished and labeled.

3.2   Experimental Details

The proposed approach has been compared with the standard Bayes rule in the above described data set. HMMs have been trained using the standard

---

[2] Class names are indicated together with a pair (Abbreviation, Number of signals).

Baum-Welch procedure[3], stopping after likelihood convergence (or after 500 iterations). Choosing a higher number of iterations will probably lead to a massive waste of time and computational resources, because experiments show that the convergence threshold is achieved with less than 500 iterations. For a higher number of states, however, when the training becomes difficult, slightly increasing the number of iterations may actually be useful, even if still time consuming. Initialization has been carried out, as in many applications, with a clustering based on Gaussian mixture models. The Gaussians in the emission probabilities were assumed to have diagonal covariance matrix (due to the high dimensionality of the signal, full covariance matrices would have been poorly estimated). The number of states used is a parameter of interest, which we have diversified during the various experiments.

For the generative embedding approach, in the resulting vector space, different classifiers have been tested[4], namely the linear bayesian classifier (ldc), the quadratic bayesian classifier (qdc), the Parzen classifier (parzenc), the linear perceptron classifier (perlc), the fisher classifier (fisherc), the logistic classifier (logc), the quadratic discriminant classifier (quadr) and the k-nearest neighbor classifier (knnc).

The parameters of the preprocessing approach —the window length and the overlap— as well as the number of HMM states have been largely varied, in order to understand the impact of such parameters in the final classification. In particular, states were varied from 2 to 8, the window length was analyzed in the range [64, 128, 256, 512, 1024] and the window overlap in the range [0, 0.25, 0.5, 0.75].

Classification accuracies have been computed using cross-validation: for ten times the data set has been split in 75% for training and 25% for testing, using every split only once for training, to ensure more variability between elements of the original data set composing the training and test set. However, when conducting the experiments for the Embedding Classification System, we used the same splits and the same trained HMM used with the Bayes Classifier: this is meant to reduce the effect of the random chance as an explanation to the outcomes of the experiments and thus enhance the quality of the comparisons.

4   Results

For the sake of simplicity, we tried to analyze them from different points of view. In particular we kept a parameter fixed, averaging (or maximizing) the results over the remaining two. In this way we can inspect the impact of every parameter on the results, as well as the robustness of the two classification approaches with respect to such parameters. Results are reported in Fig. 1, for the window length, the window overlap and the number of states, respectively. In particular, for the generative embedding case, the best classifier result has been shown.

---

[3] As implemented in the Bayes Net Toolbox for Matlab by Kevin Murphy. See `http://code.google.com/p/bnt/`

[4] Implementation given in PRTools. See `http://www.prtools.org`
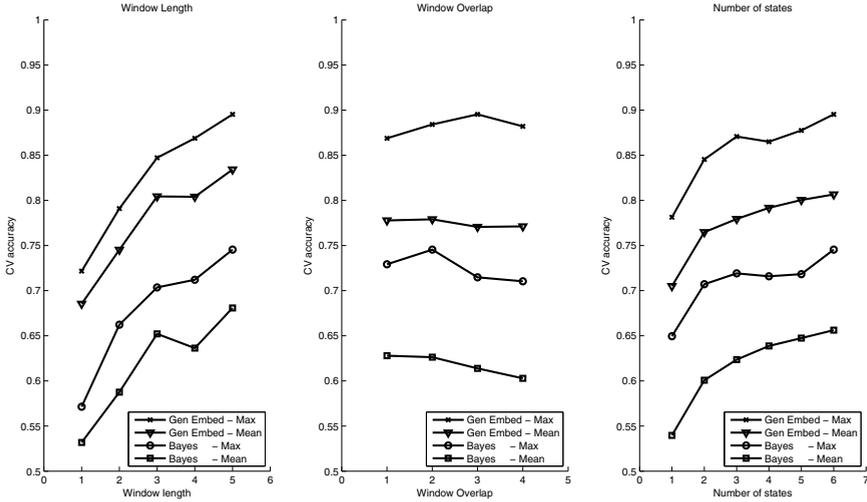
Fig. 1. Accuracy varying the different parameters. Values in horizontal axes correspond to indexes of the explored ranges as described in Sec. 3.2.

4.1   Window Length

We can see a clear trend for this parameter. From a length of 64 through a length of 1024 the efficacy of the classification keeps rising. We would also want to point out that preliminary experiments on 2048 long frames do not permit a so impressive increasing. Actually, it has been impossible to train the HMM in all cases, because the training with such higher lengths becomes impracticable: this happens because the longer the frames, the shorter the sequence. Looking both at average and maximum values of the two algorithms, it is evident that the Embedded Classification System seems to work better.

4.2   Window Overlap

Regarding this parameter we can easily notice that there is no significant trend as the experimental results are very similar. Just like with the previous parameter the average values as well as the maximum value indicate the superiority of the generative embedding approach. Eventually, we can conclude that this parameter does not seem to be that important since all the tested values delivered the same results.

4.3   Number of States

Observing this parameter, it seems clear that Embedded Classifier Systems results achieve a better efficacy when compared with the ones from the Bayes classifier; this happens not only with the maximum values, but also with the average ones.

Classifiers in the Original Space. A final experiment has been done in order to understand if the HMM-based embedding is actually useful for the classification of seismic signals if compared with standard approaches. In particular, we built a vector space by averaging the frames of every sequence, applying the same classifiers used in the HMM-induced vector spaces. Results are presented in Fig. 2, for different classifiers. It is evident the improvement obtained with the generative embedding approach.
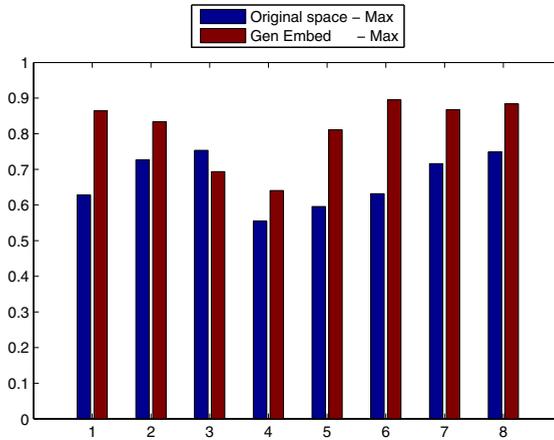
Fig. 2. Comparison with baseline classification approach. Indexes in the horizontal axis correspond to classifiers: ldc, qdc, parzenc, perlc, fisherc, logc, quadr and knnc; respectively.

5   Observations and Conclusions

As can be seen from the charts in Fig. 1, we can state that the Generative Embedding Classification System actually surpasses the Bayes Classification System (and thus the Bayes Theory) in its maximum values. However, in two over three cases it also shows a higher standard deviation so it is quite safe to say that the results may be better but they also suffer a little more from shifting chances and luck. The window length and the number of states really showed themselves as important parameters, whose values can abruptly change the eventual efficacy of the algorithm. The window overlap, however, does not really shine as vital. Its variation looks quite flat and while the average efficacy seems to slightly drop when the overlap is rising, the maximum efficacy actually slightly ascends. This is probably due to the fact that with longer frames its harder to train a decent HMM, but when successful the result is actually improved. As Fig. 2 shows, the Generative Embedded Classification System obtains better results than the simple application of classifiers in a standard vector space.

References

1. Stern, C.R.: Active Andean volcanism: its geologic and tectonic setting. Andean Geology 31(2), 161–206 (2004)
2. Orozco-Alzate, M., Acosta-Muñoz, C., Londoño-Bonilla, J.M.: The Automated Identification of Volcanic Earthquakes: Concepts, Applications and Challenges. In: D'Amico, S. (ed.) Earthquake Research and Analysis - Seismology, Seismotectonic and Earthquake Geology, pp. 345–370. InTech, Rijeka (2012)
3. Benítez, M.C., Ramírez, J., Segura, J.C., Ibáñez, J.M., Almendros, J., García-Yeguas, A., Cortés, G.: Continuous HMM-based seismic-event classification at Deception Island, Antarctica. IEEE Transactions on Geoscience and Remote Sensing 45(1), 138–146 (2007)
4. Beyreuther, M., Wassermann, J.: Continuous earthquake detection and classification using discrete hidden Markov models. Geophysical Journal International 175(3), 1055–1066 (2008)
5. Ibáñez, J.M., Benítez, C., Gutiérrez, L.A., 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. Journal of Volcanology and Geothermal Research 187(3-4), 218–226 (2009)
6. Tsuda, K., Kawanabe, M., Rätsch, G., Sonnenburg, S., Müller, K.R.: A new discriminative kernel from probabilistic models. Neural Computation 14(10), 2397–2414 (2002)
7. Bosch, A., Zisserman, A., Muñoz, X.: Scene Classification Via pLSA. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part IV. LNCS, vol. 3954, pp. 517–530. Springer, Heidelberg (2006)
8. Perina, A., Cristani, M., Castellani, U., Murino, V., Jojic, N.: A hybrid generative/discriminative classification framework based on free-energy terms. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 2058–2065. IEEE (2009)
9. Bicego, M., Murino, V., Figueiredo, M.A.T.: Similarity-based classification of sequences using hidden Markov models. Pattern Recognition 37(12), 2281–2291 (2004)
10. Bicego, M., Pekalska, E., Tax, D.M.J., Duin, R.P.W.: Component-based discriminative classification for hidden Markov models. Pattern Recognition 42(11), 2637–2648 (2009)
11. Lesage, P.: Interactive Matlab software for the analysis of seismic volcanic signals. Computers & Geosciences 35(10), 2137–2144 (2009)
12. Rabiner, L.R.: A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77(2), 257–286 (1989)