

## CLASIFICACIÓN DE SEÑALES SÍSMICAS POR MEDIO DE ONDITAS Y MÁQUINAS DE SOPORTE VECTORIAL

J.E. Hurtado, R. Henao, G. Castellanos  
Grupo de Control y Procesamiento Digital de Señales  
Universidad Nacional de Colombia Sede Manizales  
Jhurtado14@epm.net.co, rhenao@unalmzl.edu.co, gcastell@ieee.org

**Resumen.** Un objetivo de alto interés práctico en los observatorios de la actividad de la Tierra es la clasificación automática de señales sísmicas, debido al alto número de ellas que arriban permanentemente. En este campo se destaca la tarea de asignar una señal a una clase asociada al fenómeno físico que la produce. En este trabajo se consideran las siete clases en que se clasifican las señales que llegan al Observatorio Vulcanológico y Sismológico de Colombia en Manizales. Ellas corresponden a actividades volcánicas, diversos tipos de sismicidad y movimientos masivos. La investigación desarrollada se resume en los siguientes pasos: (a) Extracción de características de la señal, para lo cual se hace la descomposición en onditas (*wavelets*) de la señal y de su función de autocorrelación y se calculan los momentos estadísticos de muy alto orden. Esto se hizo debido a que se observó que por esta vía se obtiene una buena separación; (b) Para la clasificación como tal se estudiaron tres alternativas robustas de amplia aceptación en el área de Reconocimiento de Patrones: Árboles de Clasificación, Redes Neuronales y Máquinas de Soporte Vectorial, las cuales han demostrado ser muy útiles en campos como el análisis de imágenes y el reconocimiento de voces. El conjunto de entrenamiento estuvo compuesto por 48,963 señales, de las cuales se usó una pequeña parte para el entrenamiento de los dispositivos. Para las siete clases consideradas los errores oscilan entre 1 a 15 por ciento para los árboles de clasificación, 1 a 28 por ciento para las redes neuronales y 1 a 6 por ciento para las máquinas de soporte vectorial, lo cual indica que sean éstas las recomendables para fines prácticos.

**Abstract.** The automatic discrimination of seismic signals is an important practical goal for earth-science observatories due to the large amount of information that they receive continuously. An essential discrimination task is to allocate the incoming signal to a group associated with the kind of physical phenomena producing it. In this paper seven classes of seismic signals recorded routinely in the National Volcanic and Earthquake Observatory of Colombia located in the city of Manizales are considered. They correspond to signals associated to volcanic, mass movements as well as seismic activity of different sources. The approach adopted for the development of an automatic discrimination system is composed by the following: (a) The feature extraction procedure consists in the generation of very high order moments of the wavelet decomposition of the signals and of their auto-correlation function, as this resource facilitates the separation; (b) The classification of these features are performed using three powerful Pattern Recognition devices: Classification Trees, Neural Networks and Support Vector Machines, which have proved to be very accurate in other areas such as Image Processing and Speech Recognition. The training set was composed by 6,688 and the test set by 48,963 signals, which have been obtained from 1995 to 2002 in the Observatory. For the seven classes considered, the errors in the classification vary in the range from 1 to 15 percent for Classification Trees, from 1 to 28 percent for Neural Networks and from 1 to 8 percent using Support Vector Machines. This suggests that these latter can be recommended for practical use.

**Palabras clave.** Reconocimiento de patrones, Máquinas de soporte vectorial (SVM), Árboles de clasificación (CART), Redes neuronales artificiales (RNA).

### I. INTRODUCCIÓN

Actualmente y durante todo el tiempo de funcionamiento de Ingeominas, el proceso de clasificación de los sismos observados en la estación es realizado manualmente, de manera que es un proceso muy exhaustivo hecho por expertos y, que obviamente toma demasiado tiempo. En términos generales en Ingeominas se tienen en cuenta ocho diferentes clases de eventos sísmicos tanto tectónicos como volcánicos. Así, la idea principal es implementar un sistema automático de

clasificación utilizando procesamiento digital de señales, métodos estadísticos y de inteligencia artificial sobre la base del sistema existente en Ingeominas de adquisición, que incluye el segmentador, los cuales no pueden ser modificados, por cuanto, se requiere que toda las señales tengan las mismas condiciones de entrenamiento y prueba. El objetivo principal se resume al desarrollo de un clasificador que minimice el error con un costo computacional aceptable. Los tipos de sismos que se tratan en Ingeominas y que se tienen en cuenta para el problema descrito son: Hielos (ice), Avalanchas (ava), Vulcano tectónicos (vt), Largo período (lp), tremores (tre), Tectónicos locales (tl), regionales (re) y tele-sismos (ts).

## II. ADQUISICIÓN Y SEGMENTACIÓN

### A. Adquisición

El observatorio vulcanológico de Ingeominas cuenta con una red de sismógrafos distribuidos en la zona del nevado del Ruiz. En total son 14 estaciones, cada una con un sistema de adquisición, que muestrea constantemente la componente vertical con el fin de registrar cualquier evento sísmico que se presente. Cada estación cuenta con un sistema de telemetría para el envío de las señales captadas al observatorio con sede en Manizales (Chipre).

Las tarjetas de adquisición en cada estación posee un conversor análogo-digital (ADC) de 12 bits con frecuencia de muestreo de 100 Hz (la frecuencia máxima de la señal es 50 Hz). Las estaciones disponibles para la observación en Ingeominas son las siguientes: Olleta, Piraña, Alfombrales, Bis, Rodeo, Refugio, Nido, Recio, Cisne, Pácora, Toldafria, Pereira, Esmeralda, Mona.

### B. Segmentación

El sistema actual de segmentación utilizado por Ingeominas, a grandes rasgos funciona de la siguiente manera: continuamente se están monitoreando todas las estaciones disponibles, cuando se presenta un evento sísmico en alguna de ellas detectándolo con el algoritmo de disparo, utilizando el método STA-LTA (Short Time Analysis-Large Time Analysis), se empiezan a almacenar en un archivo las señales de todas las estaciones disponibles desde un punto anterior al sismo hasta otro instante posterior al que este termina. El archivo es almacenado en el formato binario estándar SUDS. Luego, los expertos hacen la clasifican basándose en los sismógrafos análogos y, posteriormente, se revisan cuales de ellos fueron detectados por el sistema automático y, simplemente se etiquetan dándole una extensión correspondiente a cada archivo de la siguiente manera: ava para avalanchas, ice para hielos, lp para período largo, re para regionales, tl para tectónicos locales, tre para tremores, ts para telesismos y vt para vulcano tectónicos.

El archivo guardado en formato SUDS tiene las siguientes características: nombre de la red de sismógrafos, hora de arribo del sismo, período de muestreo del ADC, número de canales, número de muestras de la traza, nombre de la estación en cada canal, señal de reloj, etc. Por último, se almacena el registro electrónico del sismo. La señal del reloj está presente en todas las trazas en el primer canal que se guarda. Ingeominas cuenta con un reloj que se actualiza utilizando satélites para garantizar la exactitud en la medida de arribo del sismo (el error es de  $\pm 1$  ms).

### C. Ensamble de análisis

El grupo de señales tanto de entrenamiento como de validación lo conforman los registros electrónicos de Ingeominas Regional Manizales, tomadas desde 1995 hasta febrero de 2002. Todas ellas recolectadas y clasificadas por expertos, cuyo número de archivos es de 6688. Sin embargo, cada archivo contiene señales separadas de varias estaciones, por lo que el número total de registros asciende a 48963

## III. PREPROCESO

El preproceso de las señales sísmicas, básicamente, se divide en las siguientes tres partes:

#### A. Eliminación del valor medio

Cuando los datos del archivo SUDS son extraídos y, teniendo en cuenta que los ADC utilizados en la adquisición son de 12 bits, se obtiene que el rango de los mismos está entre 0 y 4096, que representan valores positivos y negativos, motivo por el cual, cualquiera que sea la señal de entrada posee un valor medio que debe ser eliminado.

#### B. Filtración

La primera tarea que debe resolver la filtración, es la de conservar estrictamente aquellas componentes espectrales, en las cuales se considera que está contenida la información útil de la señal. En el caso de las señales sísmicas se considera que la frecuencia máxima de análisis es de 25 Hz. De igual manera, deben ser removidas las componentes debidas a señales de perturbación. El filtro aplicado es del tipo FIR (en orden a conservar la linealidad de la fase) [2], utilizando ventana Hamming, esto es, un filtro en cascada de orden 20 con frecuencia de corte 0.5, que equivale a 25 Hz.

#### C. Procesamiento con onditas

El objetivo principal del procesamiento con onditas en esta aplicación es:

- Suavizar las señales con el fin mejorar la calidad y estabilidad de la extracción de características.
- Evitar la saturación de algunas señales, debida a la presencia de ruido.
- Separar en dos señales las componentes de altas y bajas frecuencias, ya que algunos sismos contienen características discriminantes espectrales de los tipos: solo componentes de baja frecuencia, solo de alta frecuencia o mezclas de ambas.
- La propiedad de submuestreo (downsampling) genera señales con menor dimensionalidad, que reducen el tiempo requerido de cálculo en la etapa de extracción de características.

La onditas madre empleada es la db2 en un nivel de resolución [6][7], de modo que para cada señal se obtienen dos, una de coeficientes de aproximación o baja frecuencia y, otra de coeficientes de detalle o alta frecuencia.

### IV. EXTRACCIÓN DE CARACTERÍSTICAS

Después de haber realizado muchas pruebas con métodos de extracción de características (medidas estadísticas en el dominio del tiempo y la frecuencia) y teniendo en cuenta su capacidad de discriminación (entendiéndose como la capacidad de separación y agrupación de las clases tratadas), se seleccionó un conjunto de 12 características para la clasificación de todos los tipos, las cuales están basadas en la estimación de las funciones de autocorrelación y de los momentos de los coeficientes de descomposición.

#### A. Función de autocorrelación

En primera instancia, se extraen de los coeficientes de aproximación y detalle de las señales obtenidos del análisis onditas descrito en la sección anterior, la respectiva función de autocorrelación de cada serie de tiempo, calculándola como:

$$\hat{r}_+[x] = \frac{1}{N} \sum_{n=0}^{N-m-1} x[n+m]x^*[n] \text{ para } 0 \leq m \leq N-1$$
$$\hat{r}_-[x] = \frac{1}{N} \sum_{n=0}^{N-|m|-1} x^*[n+|m|]x[n] \text{ para } -(N-1) \leq m \leq 0$$